Premier Reference Source

Digital Technology Advancements in Knowledge Management

Albert Gyamfi and Idongesit Williams

IGI Global
PUBLISHER of TIMELY KNOWLEDGE

Digital Technology Advancements in Knowledge Management

Albert Gyamfi
Regina University, Canada

Idongesit Williams
Aalborg University, Denmark

A volume in the Advances in Knowledge Acquisition, Transfer, and Management (AKATM) Book Series

Published in the United States of America by
IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Gyamfi, Albert, 1974- editor. | Williams, Idongesit, editor.
Title: Digital technology advancements in knowledge management / Albert Gyamfi and Idongesit Williams, editors.
Description: Hershey : Information Science Reference, 2021. | Includes bibliographical references and index. | Summary: "This book discusses and addresses how various new and emerging digital technologies either can or currently support Knowledge Management processes within organizations by offering case studies and practical tips based on research on the emerging possibilities for knowledge management using these technologies"-- Provided by publisher.
Identifiers: LCCN 2020052965 (print) | LCCN 2020052966 (ebook) | ISBN 9781799867920 (hardcover) | ISBN 9781799867937 (paperback) | ISBN 9781799867944 (ebook)
Subjects: LCSH: Knowledge management. | Management--Technological innovations.
Classification: LCC HD30.2 .D54 2021 (print) | LCC HD30.2 (ebook) | DDC 658.4/0380285--dc23
LC record available at https://lccn.loc.gov/2020052965
LC ebook record available at https://lccn.loc.gov/2020052966

This book is published in the IGI Global book series Advances in Knowledge Acquisition, Transfer, and Management (AKATM) (ISSN: 2326-7607; eISSN: 2326-7615)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Advances in Knowledge Acquisition, Transfer, and Management (AKATM) Book Series

Murray E. Jennex
San Diego State University, USA

ISSN:2326-7607
EISSN:2326-7615

Mission

Organizations and businesses continue to utilize knowledge management practices in order to streamline processes and procedures. The emergence of web technologies has provided new methods of information usage and knowledge sharing.

The **Advances in Knowledge Acquisition, Transfer, and Management (AKATM) Book Series** brings together research on emerging technologies and their effect on information systems as well as the knowledge society. **AKATM** will provide researchers, students, practitioners, and industry leaders with research highlights surrounding the knowledge management discipline, including technology support issues and knowledge representation.

Coverage

- Cognitive Theories
- Cultural Impacts
- Information and Communication Systems
- Knowledge Acquisition and Transfer Processes
- Knowledge Management Strategy
- Knowledge Sharing
- Organizational Learning
- Organizational Memory
- Small and Medium Enterprises
- Virtual Communities

IGI Global is currently accepting manuscripts for publication within this series. To submit a proposal for a volume in this series, please contact our Acquisition Editors at Acquisitions@igi-global.com or visit: http://www.igi-global.com/publish/.

The Advances in Knowledge Acquisition, Transfer, and Management (AKATM) Book Series (ISSN 2326-7607) is published by IGI Global, 701 E. Chocolate Avenue, Hershey, PA 17033-1240, USA, www.igi-global.com. This series is composed of titles available for purchase individually; each title is edited to be contextually exclusive from any other title within the series. For pricing and ordering information please visit http://www.igi-global.com/book-series/advances-knowledge-acquisition-transfer-management/37159. Postmaster: Send all address changes to above address.

Titles in this Series

For a list of additional titles in this series, please visit:
http://www.igi-global.com/book-series/advances-knowledge-acquisition-transfer-management/37159

Promoting Qualitative Research Methods for Critical Reflection and Change
Viktor Wang (University of Montana Western, USA)
Information Science Reference • © 2021 • 367pp • H/C (ISBN: 9781799876007) • US $195.00

Enhancing Academic Research and Higher Education With Knowledge Management Principles
Suzanne Zyngier (Zyngier Consulting, Australia)
Information Science Reference • © 2021 • 323pp • H/C (ISBN: 9781799857723) • US $195.00

Role of Information Science in a Complex Society
Elaine da Silva (São Paulo State University (UNESP), Brazil) and Marta Lígia Pomim Valentim (São Paulo State University (UNESP), Brazil)
Information Science Reference • © 2021 • 297pp • H/C (ISBN: 9781799865124) • US $195.00

Theoretical and Practical Approaches to Social Innovation
Chamindika Weerakoon (Swinburne University of Technology, Australia) and Adela McMurray (RMIT University, Australia)
Information Science Reference • © 2021 • 265pp • H/C (ISBN: 9781799845881) • US $195.00

Developing Knowledge Societies for Distinct Country Contexts
Nuno Vasco Lopes (University of Minho, Portugal) and Rehema Baguma (Makerere University, Uganda)
Information Science Reference • © 2021 • 296pp • H/C (ISBN: 9781522588733) • US $185.00

Approaches and Processes of Social Science Research
Icarbord Tshabangu (Leeds Trinity University, UK) Stefano Ba' (Leeds Trinity University, UK) and Silas Memory Madondo (CeDRE International Africa, Zimbabwe)
Information Science Reference • © 2021 • 260pp • H/C (ISBN: 9781799866220) • US $195.00

Practice-Based and Practice-Led Research for Dissertation Development
Robin Throne (University of the Cumberlands, USA)
Information Science Reference • © 2021 • 288pp • H/C (ISBN: 9781799866640) • US $185.00

Optimizing Data and New Methods for Efficient Knowledge Discovery and Information Resources Management Emerging Research and Opportunities
Susan Swayze (The George Washington University, USA)
Information Science Reference • © 2020 • 198pp • H/C (ISBN: 9781799822356) • US $155.00

701 East Chocolate Avenue, Hershey, PA 17033, USA
Tel: 717-533-8845 x100 • Fax: 717-533-8661
E-Mail: cust@igi-global.com • www.igi-global.com

This book is dedicated to Profs. Knud Erik Skouby, Reza Tadayoni, Morten Falch, and Anders Henten

Editorial Advisory Board

Table of Contents

Detailed Table of Contents

Foreword

The organization generates a significant amount of data and information, which is insufficient in the absence of a knowledge management system. KM effectively manages resources and establishes a consistent link with the organization's aims and missions. When Knowledge management is paired with intelligent practices, the organizational structure gains a new degree of intensity, expertise, and purpose.

This book brings a clear perspective on knowledge management, emerging technology enhancement, innovative instructional approaches, automated strategies, and rational choices to maximize system efficacy. The book is grouped into a range of solutions for knowledge management counting information visualization, analysis, and accessibility practices.

The book highlights a variety of KM solutions, most often used in the enterprise, data quality evaluation, digital technology progress, and the development of a knowledge management platform. Besides, the chapters also discuss information management solutions in the fields of healthcare, extraction of data, and big data analytics. This book also addresses adequate information-sharing strategies for social media networks, travel systems, e-government services, and intelligent teaching-learning solutions. The authors further address the barriers to process innovation and knowledge acquisition.

The authors provide a concise outline of knowledge management, operational capability, and content collaboration. The book is edited by experts who provide insights and opportunities for discussion and growth in the strategic aspects and collaborative use of knowledge management.

More aptly, the book offers the importance of knowledge management across smart advances in KM networks, health care, education, automation, government entities and the internet of things. It includes the quantification of knowledge, preservation, knowledge design, analysis, and administration methods including numerous applications.

Finally, I would like to express my appreciation to the authors, expert team, editor, and publisher who have contributed to this book's success.

Jyoti Malhotra
MIT School of Engineering, MIT ADT University, Pune, India

Preface

The idea behind the development of this book was to provide discussions and insights into the value of current and future digital technologies for Knowledge Management within organizations. The book provides both empirically driven and expert input on how digital technologies such as Internet of Things (IoT), Artificial Intelligence (AI) and Machine Learning, etc. can be used to implement KM in organizations.

The book was inspired by the new working situation foist on organizations by the COVID-19 pandemic. Last year, in the bid to curb the spread of the pandemic, Governments around the world have instituted lock down measures and banned the assembly of persons beyond 50, 10 and 5 persons depending on the scenario. This has resulted in employees of organizations working from home using the Internet. The good news in some countries was that they had Broadband Internet services to enable them work from home. In countries where the penetration of the Internet infrastructure was low, working from home became a challenge. However, before the world was struck by the COVID-19 pandemic, some business organizations, public organizations, educational organizations etc., in the west had implemented knowledge management systems. In business organizations, such systems were embedded in existing technical systems used in product life cycle management, Project management, Customer Relationship Management (CRM), Enterprise Resource Planning (ERP) and supply chain management. In public organizations and educational institutions, KM systems were part of the administrative and operational Information Technology systems within these organizations. Hence when the stay at home orders and lock downs were instituted, employees of such organizations could access those systems using the Internet. Hence, organizations could still function, even though employees worked remotely. Although KM systems were not implemented to serve pandemics, the use of digital technologies during the pandemic has revealed the value of digital technologies to KM.

The thought that arose from this inspiration is the fact that organizations will now rely more on Digital technologies in their Knowledge management activities. This in part is because of the huge savings organizations have incurred as a result of not spending money to facilitate offline Knowledge Management activities. They have learnt from experience that training, workshops, seminars, and other forms of Knowledge sharing actives can be easily organized online as a much lesser cost. Another reason is the growing efficiency in the use of the Internet alongside organizational Knowledge Management systems. This causes organizations such as Twitter and Microsoft, to name, a few to contemplate on promoting the work from home culture.

Bearing these advantages in mind, as one accesses the digital technology landscape, one can identify a great deal of digital technologies that can enhance Knowledge management in different organizations. Examples of such technologies being used today include, augmented reality, robots, sensor networks, cloud computing, and big data to name a few. Bearing in mind the rapid rate at which digital technologies

evolve future digital technologies will support more advance Knowledge Management systems. Hence, the editors felt it necessary to commission a book where academics and industry experts could provide insight into how current advanced digital technologies support Knowledge management systems and activities in different organizations.

This book consist of knowledge and practical approaches on how different digital technologies can be integrated and used in various KM processes within organizations. The digital technologies covered in the book include Artificial Intelligence (AI), Machine learning, big data, robots, digital twin, cloud computing, Block chain and Internet of things. The book also consist of case studies from the tourism, healthcare, education and public sectors of the economy. The book also presents propositions and models that will enable the integration of KM processes in organizations using digital technologies. Based on the contents therein, readers of this book will gain knowledge and develop competences needed to facilitate the implementation of KM in organizations using advanced digital technologies.

The language of the book is mostly academic, but there are expert opinions as well. The book is designed primarily for academics, policy makers, KM practitioners, and operators of digital technologies, developers of digital technology, integrators of digital technology, knowledge institutions, consultants and knowledge workers. Bearing in mind that KM transcends formal organizations the book will also be relevant for Non-Governmental Organizations and persons operating informal organizations.

OVERVIEW OF CHAPTERS

There are 12 chapters in this book. A very brief summary of the content of each chapters are as follows.

Chapter 1 is a case study of a Portuguese stock management firm involved in gas distribution. The case study involves the analysis and improvement of process innovation within a firm. The work highlights a comparison between current process and its proposed redesign. DMAIC method (Define-Measure-Analyze-Improve-Control) is systematically applied, and new data emerge from tests made in the ERP (Enterprise Resource Planning system) of the company. The chapter posits that the use of the DMAIC method reduce execution time, as well as the number of actors and amount of information circulating outside the system of the process Innovation.

Chapter 2 presents a machine learning (ML) system that can be used in operational cloud-based websites. The system predicts the appropriate social media that supports knowledge sharing. The chapter is inspired by the COVID-19 outbreak and foresees the possibility of using the proposed KM system hereafter. The Media Richness Theory supports the empirical work in the chapter. The media richness theory (MRT) is used in establishing the relationship between the richness of a social media platform, which is based on four criteria: ability to provide feedback, multiple cues, language variety, and personal focus, and the level of equivocality and uncertainty in the knowledge-sharing task associated with the ML predictions.

Chapter 3 presents an AI driven KM system that supports Mergers and Acquisition known as the Guidance for Business Merger and Acquisition (GBMA). The chapter describes the system, how it is used to support merger and acquisition processes. The computational architecture of the system is design to automate intelligently business scenarios that support mergers and acquisitions.

Chapter 4 provides solutions to key word indexing in text documents. It presents insight into how to use future engineering, evaluation metrics and benchmark dataset systems to solve this challenge.

Chapter 5 tackles challenges related to knowledge sharing in software teams. These could be teams within the same company or teams from different companies collaborating to develop a product. The author's presents an ontology-based web portal framework called the Collaborative Knowledge Integration Architecture (CKIA). A framework can be implemented with software teams.

Chapter 6 provides a review of how cognitive robots support Knowledge Management within organizations. The authors contextualize the interaction within the boundaries of the robot-human interactions within organizations. Based on a review of literature conducted, the chapter provides a framework that will guide Robot-human interactions in KM processes within organizations.

Chapter 7 investigates and provides practical implementation guidelines on how software modules, Artificial Intelligence Algorithms can be used to manage the internal lifecycle of Russia's tour product use. The study is an outcome of research conducted at the laboratory for digitalization and management. Multi-agent models of intelligent digital twin-avatars were used to for the study.

Chapter 8 provides discussion on the potential impact of Bog-data analytics in developing economies. The discussion is supported by an analysis on the potentials and challenges facing the adoption of big-data technologies in developing economies. It also provides recommendations on how organizations in developing countries can adopt and harness Big-data to extract value for such organizations.

Chapter 9 provides an insight into the importance of IoT data in KM processes in travel services. The authors examine the challenges and the opportunities for the delivery of personalized services in the sector and how IOT can be used to enable these personalized services. The chapter argues that the implementation of these IoT enabled services will result in the improvement on customer services delivery in the travel sector.

Chapter 10 is an expert opinion on the Future of Artificial Intelligence and Machine learning in Education. The chapter provides an insight into the importance of AI in education, and the future of AI in education to provide information to educators on the AI transformation in education.

Chapter 11 is a case study on the quality of data transferred in a knowledge sharing process. The study was on the possibility towards assessing quality data from the Expanded Programme on Immunization (EPI) in Ghana the six childhood killer diseases. The chapter adopts the exploratory research approach to elicit empirical data from the EPI tally books, monthly reports and the District Health Information Management System (DHIMS II) using Data Quality Self-Assessment (DQS) tool of WHO. The chapter documents high availability and completeness of data in the EPI tally books and the monthly EPI reports. The accuracy and currency of data on all antigens from EPI tally books compared to reported number issued were comparatively low.

Chapter 12 presents an argument on the correlation between the desire towards enabling KM systems and e-government development. The chapter presents an analysis of e-government implementation in Europe using the Mobilization-Decision Theory. Based on the analysis the chapter proposes that the reason e-government was implemented in Europe was because policy makers in Europe saw the KM potentials of the digital solutions they planned on adopting.

In each of these chapters, the various approaches as to which advanced digital technologies could support KM activities in organizations are discusses in details.

CONCLUSION

This book contributes most to practice but also to theory. It is a book that does not just provide knowledge but assists in the development of competence. It is suitable for development of course work, training and organizational learning activities. As the world changes around us, organizations need the manual for the future. This is one of such manuals.

Albert Gyamfi
Regina University, Canada

Idongesit Williams
Aalborg University, Denmark

Acknowledgment

We the editors want to specially thank the contributors to this book: Kamalendu Pal, Dr. Dário Ribeiros, Prof. Paula Ventura, Prof. Silvia Fernandes, Prof. Vardan Mkrttchian, Dr. Viacheslav Voronin, Dr. Forgor Lempogo, Dr. Ezer Osei Yeboah-Boateng, Dr. William Leslie Brown-Acquaye, Mr. Furkan Goz, Dr. Alev Mutlu, Dr. M.Fevzi Esen, Dr. Serkan Polat, Fati Tahiru, Dr. Samuel Agbesi, Dr. Patrick Ohemeng Gyaase, Joseph Tei Boye-Doe, Ulkem Basdas, and Christiana Amoanteng for their immeasurable contribution to make this book a success. We were privileged to have worked with you on the dissemination of knowledge in this area of research.

We are also grateful to the Editorial Advisory Board, the team of peer reviewers and the Editorial Team from IGI for their assistance in making this book a reality. We are grateful to the assistant development editors from IGI and the project managers for the book, Ms. Katie McLoughlin, and Mr. Eric Whalen for their guidance, and supervision in the course of this project.

Chapter 1
A Challenge in Organizations: The Process Innovation – Main Results of a Case in Portugal

Dário Ribeiros
FE, University of Algarve, Portugal

Paula Ventura
https://orcid.org/0000-0002-3199-4517
FCT & Cinturs, University of Algarve, Portugal

Silvia Fernandes
https://orcid.org/0000-0002-1699-5415
FE & Cinturs, University of Algarve, Portugal

ABSTRACT

The chapter intends to exemplify process innovation challenges and trends. A case is studied—stock management—at an important enterprise in Portugal. It involves the analysis and improvement of this process within a firm related with gas distribution. Stock management is critical to deliver value to other processes such as sales. This issue has led to a focus on improving the inventory process. As it involves sub-value chains, this work highlights a comparison between current process and its proposed redesign. DMAIC method (define-measure-analyze-improve-control) is systematically applied, and new data emerge from tests made in the ERP (enterprise resource planning system) of the company. The improved process tends to greatly reduce execution time, as well as the number of actors and amount of information circulating outside the system. Other aspects are studied in line with new trends in ERP platforms due to cloud computing.

DOI: 10.4018/978-1-7998-6792-0.ch001

INTRODUCTION

This work proposes a process improvement, in the scope of stock management, by reviewing this process in a large gas company of Portugal. Inventory management is critical to the supply chain and involves three key activities: customer service, product availability and cost of inventory. Large amounts of stock are considered a waste by companies, leading managers in recent decades to focus on improving the efficiency of the stock existent in their warehouses (Shen et al., 2017).

It aims to compare the efficiency of the process in its current management and execution with its re-design in the context of the company's ERP. Why? Because the current process uses intermediate processes in Excel and/or manual processes. So, the company's ERP (Enterprise resource planning system, from SAP) is not used in its whole potential. How can this company still use intermediate processes outside the ERP? Several authors have studied this issue, such as Lund-Jensen et al. (2016), who refers it as a factor affecting information systems' projects. When these do not meet the deadline or all the expected requirements, a solution is using intermediate or manual procedures. Thus, a goal of this work is raising the following issues:

- Business processes research and design
- Continuous improvement of business processes (innovation)
- Comparison between the current stock process and the new process.

Therefore, the state of process management systems in Portugal is addressed. Some of main related concepts to be invoked are: business process management (BPM), business process management notation (BPMN), and business process maturity model (BPMM). The latter has to do with the area of continuous process improvement, an important issue in the current context of increasing digitization and organizational change. Then the current process of stock management in the company is discussed, presenting its three phases, tasks performed and involved stakeholders. This is followed by the results from a comparison between the current process and the proposed process in the company's ERP. This work then concludes the assessment made, giving some management recommendations.

DATA INTEGRATION IN PORTUGAL

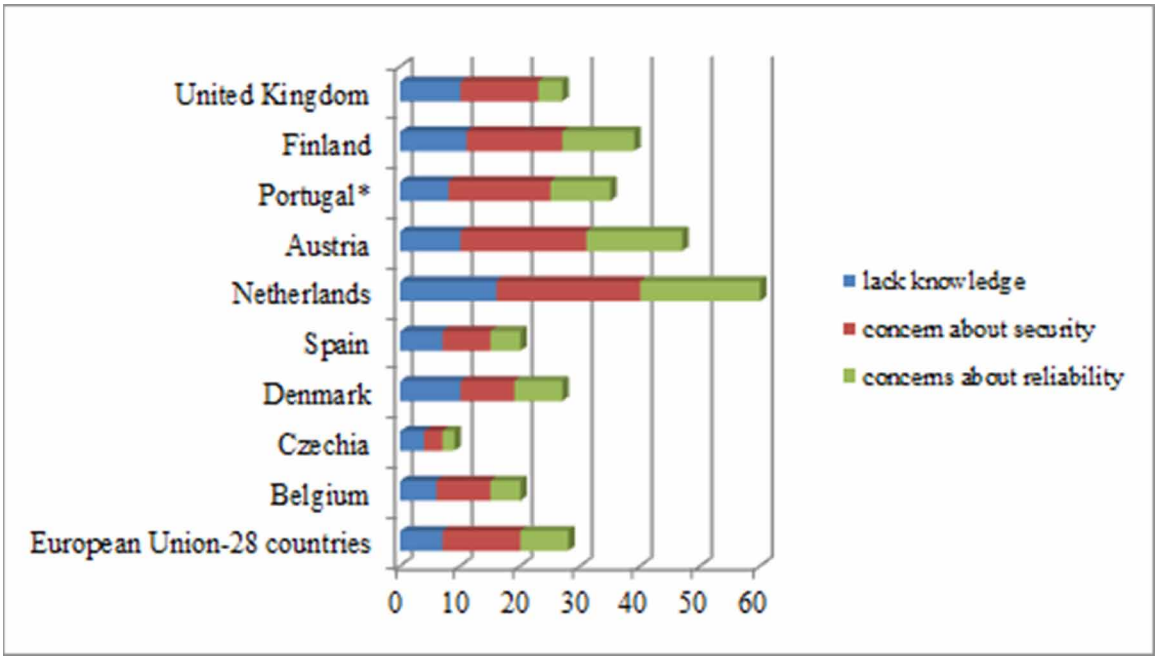
An ERP is an information system, working hard to stay relevant in this new age of cloud services. In an ERP adoption, firms start to choose which modules they need to address (integrate) every organizational aspect of their business. When compared with new agile cloud-delivered apps, the issues inherent to this approach usually involve price and customization (Rist and Martinez, 2018).

Many ERP suppliers are addressing the price problem by dropping the cost of their functional modules to compete with cloud services. And customization is not easy as they rely on complex scripting languages to manage it. These suppliers are dealing with competition from new SaaS startups (Software-as-a-Service, a type of cloud computing), as well as new trends that threaten to disrupt how data are gathered and processed. Big data, together with data analytical tools, top the list of new technologies that threaten the way ERP systems are built and used. Indeed, the shift to SaaS will accelerate over the next years becoming a necessary option for many business models.

Organizations have become more dependent on digital platforms. As a result, they have accumulated an amount of data that has been underutilized. As the internet connects tools and employees across many levels, this data generation will grow exponentially. That is why is so important to review (re-design) processes and their supporting systems. Another reason is that more devices and products have become connected. This concern forms the basis for a more intelligent approach to business (Matthews, 2014).

Figure 1 shows how is Portugal, regarding awareness about cloud services and main reasons for its lower use.

Figure 1. Awareness about cloud services
Source: Eurostat (2019)



We can acknowledge that, in Portugal, main issues underlying its lower use of cloud services are related to security concerns. Therefore, like other charts show, Portugal maintains its use of ERP systems over new cloud services. According to data from the National Institute of Statistics (INE) in Portugal, 70% of companies have online presence, either with a simple webpage or with a global digital strategy. But the percentage related to cloud ERP and smart data analytics is much lower.

For instance, a sample of SME (small and medium enterprises) in the technology sector of the Algarve region (91 enterprises analyzed through their websites) is resumed in the next table, regarding cloud adoption:

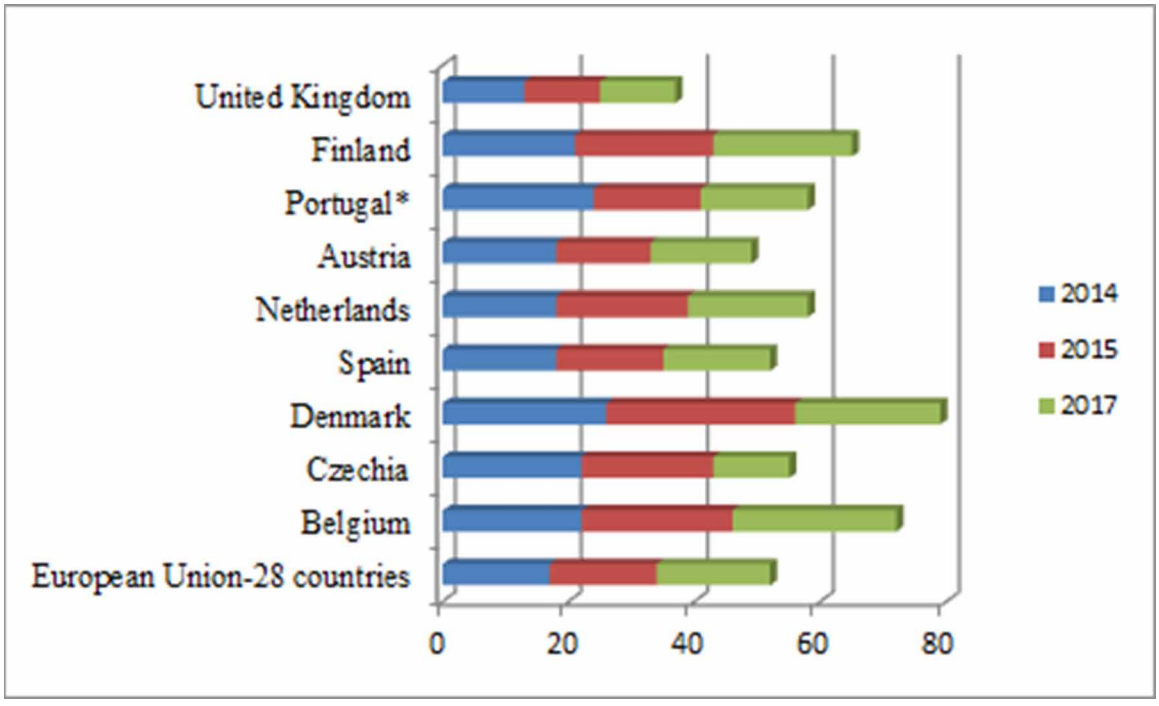
This table shows that there are few firms in the IT sector that develop cloud ERP. The majority has developed ERP systems tailored to serve hospitality and restauration firms in Algarve (this a touristic region in the south of Portugal). Those are mainly based on modules to assist several business processes what explains their partners: the ERP suppliers (such as PHC, SAGE, Wintouch, among others). Only three companies have developed cloud ERP. These curiously have a diverse workforce, international

Table 1. Analysis of a sample regarding cloud adoption

More than half of the sample is operating in sectors such as electronic security, telecommunications, technical assistance, and technological equipment (54)
12 firms have or develop ERP systems, but with modules that support the enterprise at operational and management levels. The few that have cloud computing, develop it as a solution of backup, storage and/or invoice processing (ex: 'Cloudlink', 'Acelink')
Only 3 firms develop cloud ERP, then at a level that can support the entire business model of the company and its flexibility and intelligence (BI)
8 firms are currently updating their websites
9 firms are centered in developing mobile apps, digital mkt, web design, and analytics
5 firms are centered in consulting and auditing the areas of technology and innovation in businesses

presence, and relations with the local university (through projects, teaching, training). In this context, Portuguese companies are looking for more experts in the area (Freire, 2017). This reality can be related to a certain delay of the country in external business process integration, i.e., between enterprises and their suppliers/customers (figure 2).

Figure 2. Enterprises with business processes linked to suppliers/customers
Source: Eurostat (2019)



Importance of Business Processes

According to Laudon and Laudon (2014), a business process is a set of logically related tasks to produce specific business results, as well as the mechanism as they are organized. According to the Association of Business Process Management Professionals (ABPMP), a business process is an integration of activities and behaviors, performed by people or machines, to achieve one or more outputs. Business processes can be subdivided into three main types:

- Primary processes: usually internal, responsible for adding value to the product. These are considered first-rate processes, as they represent essential activities to fulfill the firm's mission. They can flow through several business areas, giving a complete view of a value creation chain. Examples are the launch of new products, the sales process and after-sales process;
- Supporting processes: serve not only to support primary processes, but also to aid other processes. The main difference lies in delivering value. If primary processes deliver value directly to customers, supporting processes deliver value to other processes. They can be strategic for business through increasing the ability to perform primary processes. Examples are computing, human resources, among others;
- Management processes: have the objective of measuring/monitoring the activity and managing present and future of the organization. Like supporting processes, they do not deliver value directly to the customer, but are important in ensuring that the organization meets its goals. An example is management control, a function that measures and monitors company data, making it available to management for decisions that impact on the organization.

Processes are becoming increasingly important because they can highlight value chains, eventual problems, and relationships deepened in process design/mapping (Osei et al., 2016).

BPM (Business Process Management)

BPM is a management area which integrates organizational objectives, focusing on processes that relate various organizational areas and positions (ABPMP, 2013). For Hammer (2010), BPM is a comprehensive system for transforming organizations, based on a set of ideas on organizational performance.

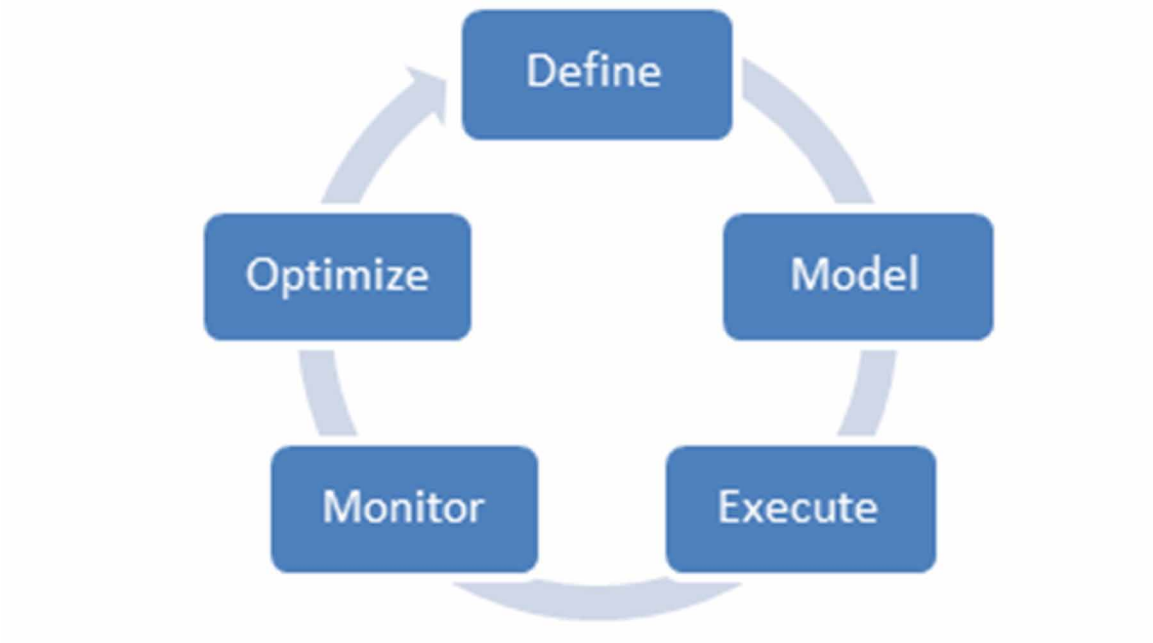
The history of BPM is divided into three periods (Smith and Fingar, 2006): 1st) early 20th century, with Taylor's principles focused on scientific management where the ability of each worker was measured to improve performance (producing more at lower costs); 2nd) the 50's, having works developed with focus on managing statistical process control; and 3rd) recently, giving priority to the optimization and (re)design of business processes as means of organizational transformation.

Business processes must be managed based on a cycle that maintains their integrity and allows their transformation (ABPMP, 2013). This reminds a feedback loop to ensure that they are aligned with business strategy. Figure 3 illustrates the functions <Define, Model, Execute, Monitor, Optimize> as phases of the BPM life cycle (for processes with predictable behavior).

Legend:

- Define: this step identifies the existing process and create an evaluation plan

Figure 3. BPM life cycle
Source: Adapted from Munk (2015)



- Model: while the new process is still on paper, the project team can test various changes to the original plan and evaluate all combinations of variables
- Execute: the project team creates, improves or acquires an application that executes the process
- Monitor: this step tracks the process by measuring its performance
- Optimize: the project team identifies opportunities for redesign/innovation.

BPMN Notation (Business Process Model and Notation)

According to ABPMP (2013), BPMN is a standard created by BPMI (Business Process Management Initiative) that sets standards for process design. BPMN usage has grown in many areas, including modeling tools. This notation has a robust set of symbols for building different aspects of processes and their design. The symbols clearly describe the relationships such as: activity flows and order of precedence; indication of events (initial, intermediate, final); message flows; etc.

Key advantages of using BPMN are: easy to read and understand workflow in the organization, versatility in modeling several process situations, and easy interaction with other tools. This notation gives companies the ability to understand their internal procedures and communicate them in a standardized way (OMG, 2011). Standard graphical notation makes it easy to understand procedures and transactions within organizations. This method ensures understanding by the employees and allows quick adjustments facing constant internal business or business-to-business changes. Within the basic categories of notation graphics, variations and information can be added to support more complex requirements.

Besides process design or mapping, the area of process improvement/innovation is increasingly important, mainly due to the new challenges of digital transformation.

Maturity Models for Process Improvement

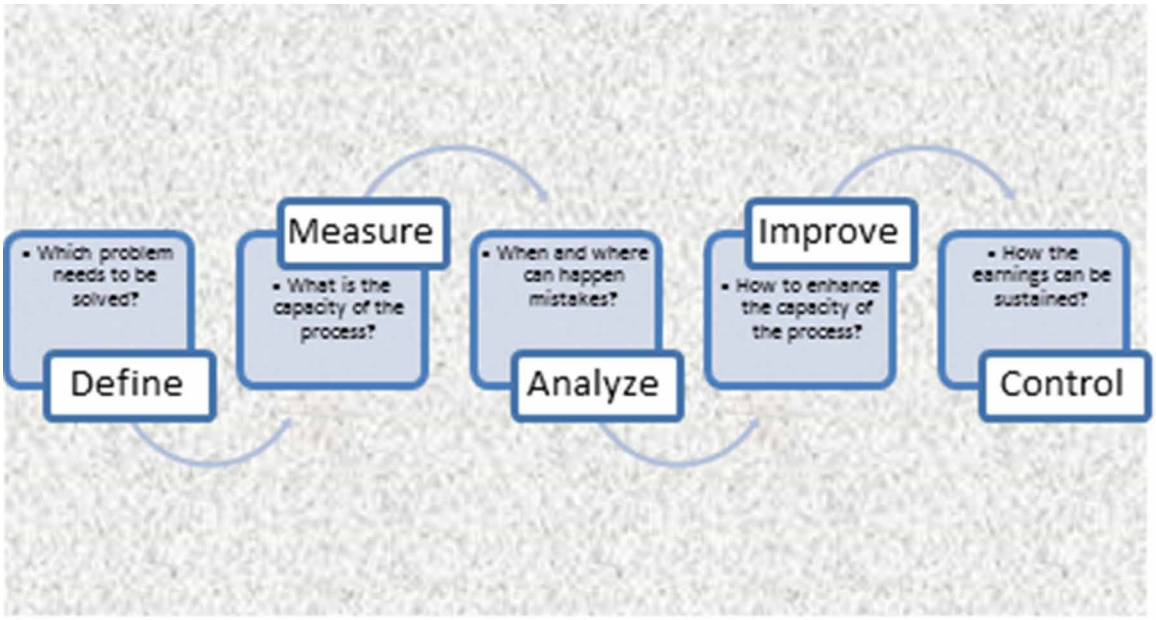
BPMM (Business Process Maturity Model) is a conceptual model based on best practices applied to a domain (OMG, 2008), i.e., applied to a set of activities where people involved share interests (ex: marketing, banking, assembly lines, finance). BPMM describes the essential elements of effective processes for one or more selected domains. Such elements also provide a basis for quantitative control of the process, which is crucial for its improvement. Therefore, BPMM can be considered a reference for continuous process design.

There is another kind of maturity models such as CMMI (Capability Maturity Model Integration, 2010) for software, which has obtained significant benefits over the years. This model has been mainly used in IS (Information Systems) projects. For many organizations such as banks, retail, equipment manufacturers and consulting services, such maturity models address about 10 to 20% of their business.

BPMM describes a path of constant improvement that guides organizations from inconsistent to disciplined processes. It includes several steps, where enhancements in each step provide a basis for refinements in the next step. The approach helps to identify process deficiencies and drives improvements through logical incremental steps. Thus, a maturity level is an evolutionary strategy defined as process improvement. Reaching each level establishes a different component in the process, resulting in increased efficiency and responsiveness of the organization itself.

On the other hand, the reduction in process variation leads to an enhancement that allows centralizing the process and ensuring accuracy and reliability. For example, the average delivery time for a product is 5 days, but there is a possible variation between 2 and 8 days. This wide variation can lead to customer confidence breakdown and then a loss in sales. This is one of the goals of DMAIC <Define, Measure, Analyze, Improve, Control> approach, i.e., reduce the variation of negative effects of the process (figure 4).

Figure 4. The phases of DMAIC approach
Source: Adapted from Goldsby and Martichenko (2005)



After exploring main concepts around business processes and the basis of process innovation, follows an analysis of the stock's management process in the company under study. Its main driver is the noticed under-use of the company's ERP system. BPMM usually refers to the whole organization, but here it is applied to a single process. As it involves 5 levels of maturity (figure 5), its application to this case-study resides on the second and third levels (related to goals such as extended organization and meaningful automation).

Case Study – Stock Management Process

Figure 5. How BPMM works
Source: http://biz-performance.blogspot.com/2018/02/business-process-maturity-model.html

Why a stock is necessary? What are the advantages from keeping stocks? How can a stock be controlled? What is the best stock replenishment method? The answers to these questions help to discuss how the company can manage its stock replenishment more effectively.

Generally, the stock represents items of manufacturing process that are available for sale, or that are in the form of unused materials. According to Slack et al. (2010), stock is the accumulation of transformed resources of an operation. For Kumar and Suresh (2008), stock generally refers to materials in stock. For Chiavenato (2005) stock is the composition of raw materials, products in process, semi-finished products, or finished products which, although not used at a given moment, need to exist in function of future needs.

For this research, we refer to stock as all finished products that the company stores within its delegations, in order to serve customers daily. As this is not a production company, but a product marketing company, there is no process of product creation. A good stock control brings benefits to the company, such as:

- improving daily relationship with the customer due to timely delivery of goods/services (Kumar and Suresh, 2008)
- dealing better with delays in the supply network. A safe stock management reduces the impact of unforeseen events
- eliminating the possibility of duplicate orders (Slack et al., 2010)
- providing economies of scale (Chiavenato, 2005) and increased bargaining capacity
- making effective use of company equity (Slack, et al., 2010). An effective management of business needs avoids the purchase of unnecessary products.

A key question is how managers decide which control levels to apply to different product stock (Slack et al., 2010). The most common method of doing it is ABC stock classification, which relies on the Pareto principle to distinguish stock types (Plinere and Borisov, 2015). In this approach, the classification of existing inventory is based on the consumption and annual value of products. The amount of inventory products consumed during the year multiplied by the unit cost results in the annual usage cost. The products are then organized in descending order of this cost. The analysis is based on a chart with the cumulative use of the consumption cost.

Challenges to Present Stock Process

The focus of this work is the redesign of the stock replenishment process in the company, due to certain inconsistencies. According to Rushton et al. (2010), different demand forecasting methods are used to estimate future product requirements against customer needs. Forecasting helps the manager make decisions such as: which materials to order for stock; how much to order; what facilities are needed to allocate stocks; etc. It is important to mention that all forecasting errors end up as inventory problems (Rushton et al., 2010).

There are several reasons why the demand for a product may vary over time. The global search pattern can be divided into the following patterns:

- a trend line over several months or years
- a seasonal fluctuation (approximately the same, year after year)
- random fluctuations that may occur at any time.

Each of those patterns must be considered in a stock control system. This section covers the steps of the current stock replenishment process in the company. It presents the three phases of the process execution, where average execution times of tasks are mentioned, as well as the resources involved. Three departments intervene directly in the process: provisioning, Centre coordination, and South coordination (regional level). Then, a proposal to change the process to the context of company's ERP is discussed as this useful system is under-used. The approach considers the BPMM premises, mainly the two first levels (figure 5), here resumed in: activities, roles (players), products, and inputs/outputs.

Phase 1: Data collection

- Main players: the employees of the supply and logistics area.
- Roles: each employee has the task of managing daily purchases to a specific group of suppliers and/or the product ranges.

- Activities: the process begins on the first business day of each week, where each employee consults the planning map to confirm the product groups that require product rotation.
- Inputs: key information to aid decision making includes purchase conditions, last 13-month exits, last 6-month exits, last 3-month exits, suggested highs and lows, current stock, and purchase suggestion. As the task of analyzing a list of 260 columns and about 1000 product lines in a text file becomes unworkable, this information is extracted from ERP to an excel file before moving to the next task.
- Output: at the end of this phase, the employee sends the excel file to the Center and South coordinators for analysis.

Phase 2: Data analysis
The second phase of the process is mainly done by Center and South zone coordinators.

- Roles and players: the Center coordinator receives the file sent by the procurement employee and uses it as basis of his work.
- Activities and inputs: with his experience and input from the external business team, he adjusts the ERP information to the present needs of the coordinated delegations. Once future needs are identified, he forwards the file to the procurement department. The South coordinator uses his analysis on the selling trend of last 13 months, the sales of last 3 months and the maximum/ minimum buying suggestions.

Each one of 40 product groups includes a database where the information in the file sent by the procurement department is stored.

- Outputs: the database calculates future sales' trend which, together with the levels suggested by the ERP, serves to achieve the maximums and minimums required by the South coordinator for each product in each delegation. Within the same database, stock requirements are calculated. Whenever the current stock is less than the minimum stock, the need for replenishment up to the maximum value is registered. The quantity of stock that exceeds the effective maximum is available for internal redistribution. After the purchase file is created, it is forwarded to the procurement department.

Phase 3: Execution
At this stage, files returned by both coordinators are received by the provisioning area.
Activities and inputs: here the file sent by the South coordinator serves as a working base, as it includes information on the internal distribution of the surplus product.

- Roles: the response from the Center coordinator is added to the file and purchase decisions are evaluated on this new increment. After gathering the answers, the procurement developer checks the ERP for any pending orders for a one-off sale. If so, this need is added to the file.
- Outputs: once information on purchasing needs is made available, we move on to consult suppliers. This sub-process starts with sending the query to the supplier via email, requiring the analysis and processing of a proposal sent via email. Upon its reception, the proposal is reviewed and one of the following decisions is taken: 1) the proposal fits company's expectations and is accepted;

or 2) the proposal is out of expectation and is sent back to the supplier for improvement. After the purchase conditions are agreed with the supplier, they are registered in the ERP from which the purchase files are extracted.

- Regarding the description of current stocks' process in the company, it is possible to draw some lessons that lead to a decision of reviewing/improving it:
  - one concerns the number of participants (roles) which is excessive (7 in total)
  - a second is the average processing time, which is currently 10 working hours per week
  - a third is that information flowing through the process risks to be lost, as it is not centralized in a single point, common to all stakeholders.

The Process Redesign

This section discusses the solution that best fits the real needs of the company under study, regarding the referred process. In the stock replenishment function, it is possible to create a model that combines material management with production planning.

The first phase, in the reviewed process, allows the procurement and logistics coordinator to prioritize the most important materials for the daily activity of each of the company's eleven delegations. In the second phase, the focus of the supply and logistics coordinator is to parameterize the material codes in the ERP database, automatically and independently of the user. It then calculates: the replenishment point, safety stock, and needs for subsequent periods. Lastly, the third phase presents the tasks performed by the procurement and logistics employee, from the forecasting of needs to the purchase order for the supplier.

The reviewed process strives to capture all business rules of stock management function and integrate them in the existing ERP to support the whole related activities and roles (Laudon and Laudon, 2014). An important goal is that the redesigned process better contributes to meet its required KPI (key performance indicators) for business performance.

Comparison: Current Process vs. Improved Process

For an evaluation between these two situations, it is essential to compare phase 1 of the current process with phases 1 and 2 of the reviewed process (as they have in common data collection and processing). On the other hand, the phases 2 and 3 of the current process are grouped and compared with phase 3 of the reviewed process (as these are phases of analysis and processing of results). In this comparison, documents and resources involved in both processes are considered, as well as an estimation of execution time of each phase. The execution times include the time registered during the tests developed.

Resuming, for phase 1 of current process, three rotation files are created per week and the stock replenishment map, all in excel. A total of 157 documents are generated during a year. For the same process, 5 procurement people are allocated devoting about an hour per week to process the tasks needed. By analyzing both phases 1 and 2 of the reviewed process, zero documents are generated, as the whole process is managed and stored in the ERP system. Only 1 person is required to carry out the tasks of both phases (the procurement coordinator). The tests have defined that this coordinator would have to dedicate 2 hours per year in phase 1 and 4 hours per semester in phase 2 (thus, 10 hours per year for both phases).

In phases 2 and 3 of current process, 156 excel documents are entered and then sent to the Center and South coordinators, to be returned to the procurement collaborator, who groups responses into those documents. The tasks in these phases take about 8 hours per week. Analyzing phase 3 of the reviewed

process, zero documents are generated (again the whole process is managed and stored in the ERP). Only 1 person is required to perform the tasks (the procurement/logistics employee). The tests have defined that this employee would have to dedicate 3 hours per month to each business area. Table 2 summarizes these comparative issues:

Table 2. Comparative indicators: current process vs. improved process

Indicators/process	Process now	Improved process
Documents in process	157	0
Persons involved	7	2
Hours spent in process/year	468	226

Thus, we acknowledge that the proposed process will require few human resources, as well a simpler flow due to less documents circulating in several departments of the company. Other important result is the reduction of time spent on stock replenishment, which is almost half the time spent in the current process.

This redesign exercise also prepares the company to cope with the increasing trend of cloud platforms. Using the right cloud provider, a company can rapidly scale business productivity software as its business grows. It also gives companies access to their business applications at any time and location. An ERP, which can be cloud-supported, together with process design tools, easily allows to align business needs with process integration.

CONCLUSION

This work highlights some issues around process management in Portuguese companies. It also points out that business can be improved or innovated through process review/redesign. BPMM (Business Process Maturity Model) was considered to define a methodology that could manage the need to improve the stock process of an important company. Aspects such as trend, seasonality and randomness of demand were addressed to analyze how the company's ERP answers to its users. This helped to foresee an advancement to the current process and derive lessons to guide its redesign. Issues such as number of documents generated and circulating in the process flow, number of players, and number of working hours to fulfill the process were the main motivating factors of this change.

The DMAIC (Define-Measure-Analyze-Improve-Control) method was systematically applied in this project, from which emerged data that helped to rethink the output involved. Part of these data was based on tests made in the company's ERP. Regarding continuous process performance, BPMM maturity levels 2 and 3 were particularly relevant to this process improvement. Identifying products to be purchased and customer requirements were crucial in building the tasks to be performed by the new process. Once implemented, it will greatly reduce the execution time and number of actors, as well as the amount of information circulating outside the system.

This work shows that process management should have a continuous approach, including new technological trends and their synergies. Without doing so, this can result in a big threat to enterprise

performance (Köhne & Sawyer, 2018; Fasna & Gunatilake, 2020). This redesign effort also prepares a company to cope with cloud systems and service-oriented architectures (SOA). A cloud ERP/CRM (of which Salesforce is a relevant supplier) provides access to full-function applications at a reasonable price. These trends also support the design of customer-centric business processes (Frank et al., 2020). It is important to create conditions for this qualitative 'jump' as it means enhancing business intelligence by leveraging its system's effectiveness. Also, an enterprise needs to determine which model will best align with its strategy and security needs.

ACKNOWLEDGMENT

This paper is financed by National Funds provided by FCT- Foundation for Science and Technology through project UIDB/04020/2020.

REFERENCES

ABPMP. (2013). *Guia para o Gerenciamento de Processos de Negócio - Corpo Comum de Conhecimento* (1st ed.). LECOM Tecnologia.

Chiavenato, I. (2005). *Administração de Materiais - Uma Abordagem Introdutória* (3rd ed.). Elsevier Editora, Lda.

CMMI. (2010). *CMMI for development, version 1.3*. https://resources.sei.cmu.edu/asset_files/TechnicalReport/2010_005_001_15287.pdf

Fasna, M., & Gunatilake, S. (2020). Towards successful strategies to overcome BPR implementation issues: Case of Sri Lanka. *Business Process Management Journal*, *26*(6), 1241–1259. doi:10.1108/BPMJ-03-2019-0087

Frank, L., Poll, R., Roeglinger, M., & Lea, R. (2020). Design heuristics for customer-centric business processes. *Business Process Management Journal*, *26*(6), 1283–1305. doi:10.1108/BPMJ-06-2019-0257

Freire, M. (2017). *Especialistas em digital marketing e data analytics são das profissões mais procuradas pelas empresas*. https://www.bit.pt/especialistas-em-digital-marketing-e-data-analytics-sao-as-profissoes-que-as-empresas-mais-procuram/?inf_by=59cba655671db80f3d8b46cb

Goldsby, T. J., & Martichenko, R. (2005). *Learn Six Sigma Logistics - Strategic Development to Operational Success* (1st ed.). J. Ross Publishing, Inc.

Hammer, M. (2010). What is business process management? In J. Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management 1*. Springer. doi:10.1007/978-3-642-00416-2_1

ISO/TEC. (2004). *Information technology - process assessment - Part 4: guidance on use for process improvement and process capability determination*. https://www.iso.org/obp/ui/#iso:std:iso-iec:15504:-4:ed-1:v1:en

Köhne, H., & Sawyer, A. (2018). Process innovation: Requirements and opportunities. *European Journal of Management Issues*, *26*(1-2), 22–28. doi:10.15421/191803

Kumar, S. A., & Suresh, N. (2008). Production and Operations Management (2nd ed.). New Delhi: New Age International (P) Ltd.

Laudon, K. C., & Laudon, J. P. (2014). Management Information System - Managing the Digital Firm (13th ed.). Pearson Education Limited.

Lund-Jensen, R., Azaria, C., Permien, F., Sawari, J., & Bækgaard, L. (2016). Feral information systems, shadow systems, and workarounds - A drift in IS terminology. *Procedia Computer Science*, *100*, 1056–1063. doi:10.1016/j.procs.2016.09.281

Matthews, D. (2014). *Five trends that will shape the future of ERP*. https://www.industryweek.com/technology-and-iiot/article/21964701/five-trends-that-will-shape-the-future-of-erp

Munk, J. (2015). *Six sigma daily*. https://www.sixsigmadaily.com/six-sigma-vs-business-process-management/

OMG. (2008). *Business Process Maturity Model (BPMM), version 1.0*. OMG.

OMG. (2011). *Business Process Model and Notation (BPMN), version 2.0*. OMG.

Osei, A., Yunfei, S., Appienti, W., & Forkuoh, S. (2016). The antecedents of process innovation and Smes growth: Empirical evidence from shoe manufacturing sector in the Ashanti region Of Ghana. *European Scientific Journal*, 163–173.

Plinere, D., & Borisov, A. (2015). Case study on inventory management improvement. *Information Technology and Management Science*, *18*(1), 91–96. doi:10.1515/itms-2015-0014

Rist, O., & Martinez, J. (2018). *5 Enterprise resource planning (ERP) trends to watch in 2019*. https://www.pcmag.com/news/5-enterprise-resource-planning-erp-trends-to-watch-in-2019

Rushton, A., Croucher, P., & Baker, P. (2010). *The Handbook of Logistics and Distribution Management* (4th ed.). Kogan Page Limited.

SAP. (2012). *News SAP*. https://news.sap.com/brazil/2012/05/10/gartner-coloca-sap-na-lideranca-mundial-de-solucoes-erp/

Shen, H., Deng, Q., Lao, R., & Wu, S. (2017). A case study of inventory management in a manufacturing company in China. *Nang Yan Business Journal*, *5*(1), 20–40. doi:10.1515/nybj-2017-0003

Slack, N., Chambers, S., & Johnston, R. (2010). Operations Management (6th ed.). Pearson Education Limited.

Smith, H., & Fingar, P. (2006). *Business Process Management: The Third Wave*. Meghan Kiffer Press.

KEY TERMS AND DEFINITIONS

BPM: Discipline or approach where various methods are used to model, analyze, measure, improve, optimize, and automate business processes. It sees processes as important assets of an organization which must be managed and developed to deliver value-added products/services.

BPMM: Consists of a standard to assess maturity and guide organizations to improve processes. Based on the concept of continuous improvement, it leads organizations to evolve from inconsistent low-maturity activities to structured and organized processes.

DMAIC: An iterative tool used to improve processes. It can be used in any situation where to implement improvements. It is an acronym for five steps: Define, Measure, Analyze, Improve, Control. D-M-A-I-C cycle must be repeated until the desired improvement is achieved.

ERP: System for integrating organizational processes to produce information for decision support in real time. This business management system provides an integrated and continuously updated view of core business processes.

Process Design: The function of reviewing or creating a new process workflow. It may involve: process analysis; best practices; process models from industry-standards; or ideas coupled with the experience of the process design team.

Process Innovation: The application of a new technology/method or the redesign of a process that executes an activity. This occurs when the organization solves an existing problem or performs an existing business process in a different way that adds value to it.

Stock Process: Shows how much stock the firm has, at any time, and how it keeps track of it. This applies to every item used to produce a product/service. Efficiently managed, it allows the right amount of stock in the right place at the right time, and serves production if problems arise.

Chapter 2
Determining Appropriate Social Media Sites for Knowledge Sharing:
A Data Science Approach

Albert Gyamfi
https://orcid.org/0000-0002-3839-4178
University of Regina, Canada

ABSTRACT

The study aims at developing a fully developed and operational cloud-based published website for predicting appropriate social media for sharing knowledge during an outbreak. The media richness theory (MRT) is used in establishing the relationship between the richness of a social media platform, which is based on four criteria—ability to provide feedback, multiple cues, language variety, and personal focus—and the level of equivocality and uncertainty in the knowledge sharing task. A survey is used to gather data on the use of four social network sites (SNSs) (Facebook, Twitter, YouTube, and Instagram) that were mostly used for sharing knowledge during the COVID-19 pandemic. Data science techniques are used to analyze the data and develop a system for selecting appropriate social media for sharing knowledge.

INTRODUCTION

The use of social media for knowledge sharing has been a hot research theme in recent times. For example, the COVID-19 outbreak has sparked up several activities on social media beyond anything the world has ever seen. Social media has been used in previous disaster responses including Hurricane Sandy and SARS epidemic (Mirbabaie et al., 2019) (Reuter et al., 2019). In each of these outbreaks, the critical challenge has been how to transfer knowledge of most recent best practices to people like frontline workers who needed them the most at a faster pace than the spread of the pandemic (Chan et al., 2020) (Zhang et al., 2017) (Pennycook et al., 2020). In other disaster responses, social media has been used in

DOI: 10.4018/978-1-7998-6792-0.ch002

sharing information such as areas with electricity, which pharmacies are opened, where flooding was intense and so on. In the current pandemic various public health institutions have been relying on social media to sensitize the public and creating the needed awareness. Messages warning the public on the use of unapproved drugs were seen coming from the CDC, when they observed many people talking about possible drugs for treating the virus. Moving forward, there is a high anticipation that governments, medical experts and public health authorities to share knowledge with people on almost everything associated with the outbreak of pandemics will continuously use social media (Chan et al., 2020).

An important determinant for successful knowledge sharing during a pandemic could be attributed to the careful selection of appropriate media and for that matter, social media application to match the task of involved in knowledge sharing. Various modalities for selecting social media for knowledge sharing purposes have been prescribed in the literature(Ahmed et al., 2019) (Qi and Leung, 2015) (Ghazali et al., 2016), however, these approaches are based on conventional information systems used for resolving issues related to transferring information which is readily available at a source to a destination where it's needed the most. Meanwhile issues of knowledge sharing, which pertains during pandemics are novel and nonrecurring just like the virus itself and require innovative ways not only to describe and prescribe but to also predict appropriate media, which could be determined through data driven approaches involving the use of machine learning techniques.

The use of big data technologies such as neural nets, machine learning, dynamic algorithms etc. are providing novel forms of evidence leading to theory development and advanced understanding of both science and business (Abbasi et al., 2016) (Gover et al., 2020) (Maass et al., 2018).

The current study uses a blended approach of theory-based and data-driven research by combining traditional survey research with machine learning techniques in developing a data-driven information system for the selection of appropriate social media platform for effective knowledge sharing. The study will contribute to the ongoing discussion on the use of information systems research as an enabler for bringing together big data research and theory-based research in the era of data-driven economy (Abbasi et al., 2016) (Gover et al., 2020) (Maass et al., 2018).

PROBLEM STATEMENT

Knowledge transferring and Knowledge sharing plays a pioneering role and have increasingly gained a pivotal position in organizations especially in the data science communities (Carvalho and Gomes, 2017). In order to succeed in the exponentially increasing competitive market, organizations are required to implement diverse strategies to acquire, maintain, store, and disseminate strategic knowledge to be able to reuse when the need arises. As a result, many organizations feel the urge to acquire and use different technologies to enhance efficiency in sharing and accessing knowledge in efficient manner (Barros et al., 2015) (Diamantini et al., 2013). Among the technologies widely chosen for knowledge sharing are social media websites. However, with the plethora of social media platforms with different functionalities and media characteristics, selecting the appropriate medium for a giving communication tasks becomes a challenge, especially in the course of a disaster or a pandemic.

Meanwhile knowledge-sharing tasks come in different forms depending on the nature of knowledge being shared, and each form require media with certain matching properties. Consequently, not all social media platforms are suitable for a given knowledge-sharing task since they have varying degree of media richness. Therefore, there should be criteria for selecting the most suitable platform that will

be capable of conveying the exact meaning to the information that is being shared in order to make it easier for the recipients to understand and use it to accomplish the intended tasks. Using inappropriate media for a given knowledge sharing tasks could result in misinterpretation of content, due to increase in ambiguity, which in turn can lead to unaccomplished tasks with disastrous consequences especially during a pandemic.

BACKGROUND AND RELATED WORK

Using Social Media for Knowledge Sharing in a Pandemic

On March 11, 2020, COVID-19, which is the disease caused by the novel coronavirus (SARSCoV-2), was declared a pandemic with most patients (81%) developing mild illness, 14% developing severe illness, and 5% developing critical conditions, which included acute respiratory failure, septic shock, and multiorgan dysfunction (Matthews et al, 2020). As at the time of writing this article, over 3 million people had been diagnosed with more than 200,000 deaths recorded worldwide (Cho et al., 2020). The health threat caused to the entire world and healthcare systems by the outbreak that was originally detected in Wuhan, China is unprecedented. Furthermore, the crisis has expanded across all aspects of everyday life including social relations and economic activities causing governments to place tight restrictive measures that restricts access to work, travels, meetings which has necessitated and motivated distance work, education, entertainment and other social activities. As the crisis rapidly evolved, these measures triggered an increase use of technological support especially regarding the sharing of knowledge.

Knowledge sharing enabled the delivery of reliable information that addresses critical infection control issues to healthcare workers (Chan et al., 2020). Ineffective knowledge dissemination and transference play a major role in the existing gap between public health knowledge discovery and their applications in practice and policy development (Brownson et al., 2018).

In order for knowledge sharing to be effective, the content of the message, source, audience, and media should be taken into consideration (Panahi et al., 2013). Improper way of media selection for the message has the potential of damaging relationships (Ishii et al., 2019). The study focuses on understanding the relationship between the level of equivocality of the message and the media richness of the channel and based on that relationship, develop a recommendation system that could predict appropriate social media platform for a given knowledge transfer task.

Media Richness of Social Media Applications

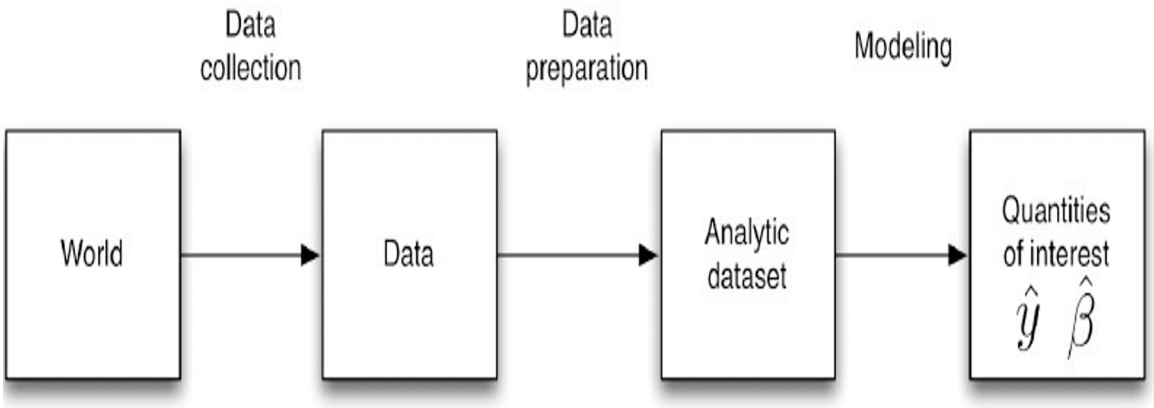
Kaplam and Haenlein, (2020) defines social media "as a group of internet-based applications that build on the ideological and technological foundations of web 2.0, and that allow the creation and exchange of User Generated Content". Social media builds on web 2.0 technologies which uses the internet as a platform to support applications that harness network effect to get better the more people use them (Irani et al, 2017) (Gyamfi, 2016) (O'Reilly,2007). This move of the internet has shifted the focus of end users from being passive consumers of web content into active user participation where they are free to create, collaborate, communicate, control, and share content by using the web as a communication medium (Aldahdouh et al., 2020). There are different forms and classes of web 2.0 tools including blogs, collective intelligence, digital content management, social networks, mashups, virtual worlds,

Really Simple Syndication (RSS), tagging, peer-to-peer programs and so on. There is a rising popularity among Social Network Sites (SNS) such as Facebook, Twitter, Instagram, and YouTube etc., making it easier to socially interact, and communicate (Chen et al., 2020). During Emerging Infectious Diseases (EIDs) such as the current COVID-19 outbreak, SNS and microblogging are among the main channels used by individuals for sharing feelings, stories, opinions, judgments and evaluations about COVID-19 (Xu et al, 2020).

These SNS differ in their capacities to facilitate understanding and convey meanings to messages shared through them. They can be characterized as high or low in richness based on a blend of four criteria: ability to provide feedback, multiple cues, language variety, and personal focus (Ishii et al.,2019). Instant feedback allows users to ask questions and receive corrections. Multiple cues refer to an array of cues such as physical presence, voice inflection, body gestures etc. that may be included in the message. The third criterion is language variety, which represent the scope of meaning that can be conveyed with language symbols. Lastly personal focus, which refers to the infusion of personal feelings and emotions into communication in order to fully convey a message. Based on the media richness criteria, face-to-face communication was originally placed at the highest position on the media richness hierarchy, while other media such as telephone, memos, impersonal written documents and numeric documents were followed in decreasing order of media richness (Ishii et al.,2019) (Daft et al., 1987). The exclusion of e-mail, and video conferencing on the media richness hierarchy was considered among the shortfalls of the MRT (El-Shinnawy and Markus, 1997). Consequently, other modalities have been theorized in literature for determining the media richness hierarchy for Computer Mediated Communication (CMC) such as E-mail, Instant Messaging (IM), Video and Voice over Internet Protocol (VVOIP), Voice over, Internet Protocol (VoIP), forums, and portals (Schwartz, 2007).

As new media such as SNSs continue to emerge as channels for sharing knowledge during EIDs there is a call to revisit the MRT to understand how the media richness of these platforms could be matched with the task characteristics of HCWs for effective knowledge sharing for enhanced task performance (Ishii et al.,2019).

Figure 1. Data pipeline in survey research
Adopted from Kindel et al., 2018



RESEARCH APPROACH

The current research is situated at the intersection between data-driven and theory driven perspective by mapping domain theory to data analytics as presented in (Mirbabaie et al., 2019) (Chan et al., 2020) (Aldahdouh et al., 2020). In the era of data economy, conducting research with an integrated data-driven and theory-driven approach provides many opportunities for impactful research. Applying the two approaches independently can be problematic, in the sense that emphasis on big data analytics without considering domain theory has been argued as leading to identifying correlations, trends, and patterns that may provide answers to situated questions without meaningful contributing to lasting scientific knowledge. On the other hand, emphasizing domain theory may also lead to missing opportunities from making discoveries from big data usage. The study applies survey and data analytics techniques as the research methods as presented in (Delen et al., 2013).

Data Pipelines in Survey Research

Data pipeline refers to the process of defining, what, where, and how data are collected, transform and loaded. There are key challenges that need to be understood when incorporating machine-learning techniques into social science research. A data pipeline for understanding how data is processed in a survey research with machine learning is presented in Kindel et al., (2018). Information gathered about the SNSs usage for knowledge sharing through survey are preprocessed and transformed into analytical dataset. The goal of data analytics is to estimate certain quantities that are of interest using both regression analysis and classification for predictive modelling (see figure.1)(Hofman et al., 2017) (Mullainathan and Spiess, 2017).

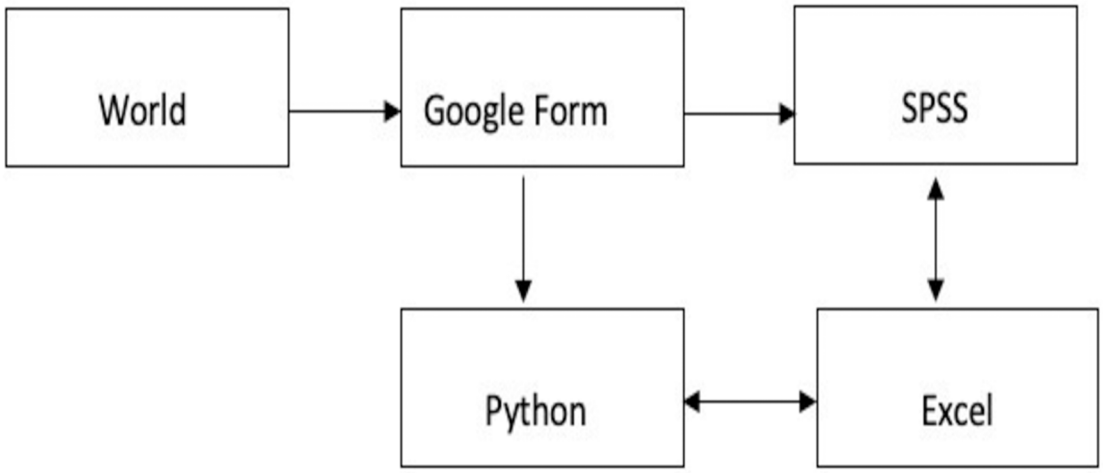
Data Collection

Data source and Survey Development: The data for the study were collected through a cross-sectional mail survey with questions comprising of four constructs: media richness, knowledge ambiguity, and knowledge sharing.

Based on their experience with the use of SNS (Facebook, Twitter, Instagram, and YouTube) during the COVID-19 Pandemic respondents were required to indicate their level of agreement on 7-point Likert scale ranging from 1- “strongly disagree” to 7“strongly agree” on each of the items measuring the three constructs.

The content of the survey was divided into three sections comprising of seven (7) question on knowledge ambiguity, and four (4) questions on media richness, and two (2) on knowledge sharing. The measures of knowledge ambiguity included tacitness (T), Complexity (C), Specificity (S), Partner Protectiveness (PP), Cultural Distance (CD) and Organizational Distance (OD) as presented in (Simonin (a), 2019) (Simonin (b), 2019) (Uygur, 2013). Measures of Media richness were deduced from previous studies (Chao et al., 2020) consisting of Availability of Feedback (AF), Use of Multiple Cues (MC), Language Variety (LV) and personal Focus of the medium (PF). The knowledge sharing measures were also applied as used in (Simonin (b), 2019). The data were then collected through self-administration of the questionnaire to the selected sample population. Overall, 1200 invitations were sent for the questionnaires to be filled using Google form and out of that, 1136 were filled and processed representing a response

Figure 2. Tools for data extraction, storage and analytics



rate of 94%. The descriptive statistics of the survey response are recorded in table 1 and excerpt of the distribution chart in figure 3).

Data extraction and storage: The data was extracted in excel format and saved as csv file. However, since SPSS and Excel are traditional tools for managing and analysing survey data, it was convenient to extract and populate the data to SPSS and excel for further examination and preparation before loading them up to Python (see Figure. 2).

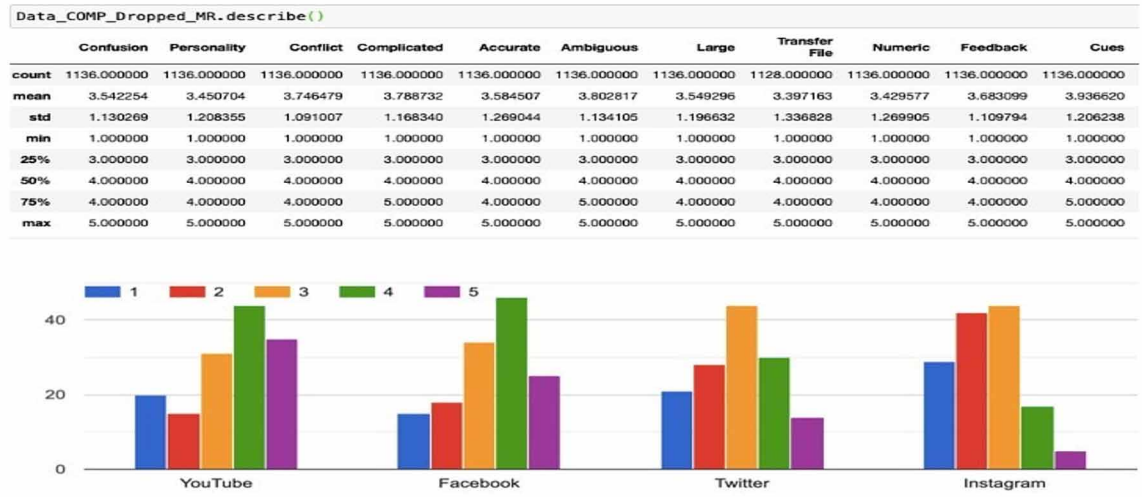
Data Preparation (Cleaning and Preprocessing)

It was discovered that the data, which was gathered for the study, were incomplete in the sense that attributes of several tuples, which were of interest to the study, had no recorded values. Moreover, some of the recorded values were inaccurate and therefore represented as noise since they contained errors and others, too, deviated from the normal or expected values making them inconsistent (Han et al., 2021). The last data cleaning tasks was related to the fact that the tuples of the answers provided contained both numerical values and actual texts provided in each cell. Thus, the main data cleaning tasks involved filling the missing values, leveling out the noise in the data, and resolving the inconsistencies.

Factor Analysis

In this step a series of data analytics methods and applications were used on the dataset to derive conclusions on the proposed hypotheses and to develop a classification model predicting appropriate SNS for knowledge sharing. The first step is performing an exploratory factor analysis (EFA) followed by a confirmatory factor analysis (CFA).

Figure 3. Descriptive statistics



```
Data_COMP_Dropped_MR.describe()
```

	Confusion	Personality	Conflict	Complicated	Accurate	Ambiguous	Large	Transfer File	Numeric	Feedback	Cues
count	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1128.000000	1136.000000	1136.000000	1136.000000
mean	3.542254	3.450704	3.746479	3.788732	3.584507	3.802817	3.549296	3.397163	3.429577	3.683099	3.936620
std	1.130269	1.208355	1.091007	1.168340	1.269044	1.134105	1.196632	1.336828	1.269905	1.109794	1.206238
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
75%	4.000000	4.000000	4.000000	5.000000	4.000000	5.000000	4.000000	4.000000	4.000000	4.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Performing the Test of Adequacy

Prior to conducting the factor analysis, there was the need to check to see if the dataset contains all the factors as proposed in the model. This could be assessed with either the Bartlett's sphericity test or the Kaiser-Meyer-Olkin (KMO) test. The Bartlett's test is used to check if there exists intercorrelation between the observe variables by using the observed correlation matrix against the identity matrix. It gives an indication whether it has all right to perform factor analysis. The results from the dataset are nearly 0 (3.2261750798e-147), indicating that the test is significant, statistically and the observed correlation matrix is not an identity matrix.

On the other hand, the KMO tests the adequacy of each of the observed variables and for the complete model. In order for KMO value to be considered adequate it should be above 0.6. The dataset passed both tests as seen in the code snippets above, with KMO value=0.869, which means the factor analysis could be performed.

Exploratory Factor Analysis

In order to understand how well each of the measures loaded on the various corresponding factors. The assumption is that each of the observed variables could be directly associated with any factor. Two methods used are factor rotation using Varimax method and estimate communalities using Maximum Likelihood method (ML).

Determine the Number of Factors from the Dataset

This was obtained by using the Kaiser Criterion and scree plot using eigenvalues. From the code snippet and the scree plot below, it could be seen that on three-factor had Eigen value above 1. Which confirms the number of factors as proposed in the research model (Uncertainty, Equivocality, and Media richness).

Figure 4. Results of test of adequacy

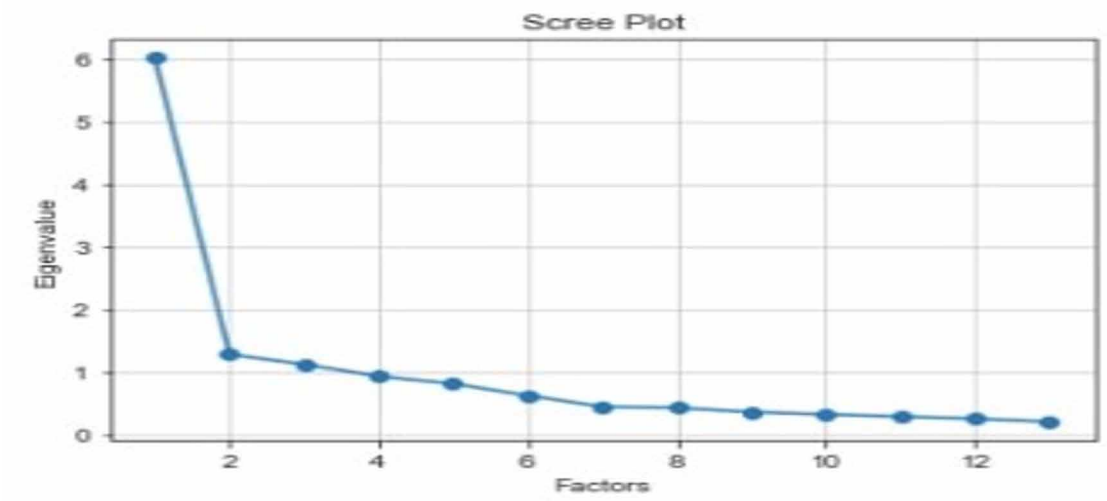
```
#Bartlett's test of sphericity checks

from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity
chi_square_value,p_value=calculate_bartlett_sphericity(Data)
chi_square_value, p_value
```

```
(934.4525718927435, 3.226175079830245e-147)
```

```
#Kaiser-Meyer-Olkin Test
from factor_analyzer.factor_analyzer import calculate_kmo
kmo_all,kmo_model=calculate_kmo(Data)
kmo_model
```

```
0.869353823238433
```

Figure 5. Scree plot



Confirmatory Factor Analysis (CFA)

After performing the EFA, the next step is the CFA which is used to confirm the number of factors and the loadings of the corresponding indicator variables to agree with what was expected as proposed in theory (Confirmatory factor analysis, 2019). The basic idea with CFA is that any of the factors is connected with a particular set of observed variables.

The factor loadings show how the various variables loaded on the three main factors as indicated on the scree plot.

Table 1. Factor loadings

Measures	Factor1	Factor2	Factor3
To present some confusing information to twenty friends			
To work out a personality problem that has affected the working relationship between you and your boss	0.37073165	-0.088187	0.16488422
To explain information with conflicting interpretations to a group of friends	0.42269855	0.09888693	0.61578538
To get explanation from a peer of a complicated technical matter in which you have no experience of	0.09506073	0.10472814	0.80398362
To get clarification of an ambiguous information from your friends	-0.00330152	0.05415273	-0.00499137
To exchange a large number of messages with your friends and work groups	0.64063603	-0.09788218	0.87543258
To exchange important information that needs to be conveyed accurately	-0.10952096	0.15962237	0.05134689
To transfer files and reports	0.65448406	0.14966918	-0.20619117
To exchange numerical information	0.85011952	-0.05258108	-0.03079282
Makes it easier for me to communicate and exchange ideas with others	0.87480244	0.11902186	0.13633334
Allows me to use non-verbal cues when communicating with others	0.15094916	-0.00549492	0.13650707
Provides flexible way of expression that allows me to freely use the language I am familiar	-0.11599147	0.57473789	-0.11362825
Is a medium where I can exhibit my personal features such as looks, form etc.	0.05700944	0.83017753	0.74479554
	0.04765294	0.7202375	-0.01524012

The results obtained from the CFA were used as the basis for determining the most important factors for each of the three constructs that could use to predict the suitability of the SNSs. The CFA results depicted in tables 2 for YouTube, Facebook, and Twitter respectively confirms what was proposed in the research model. Now that the variables and the factors have been confirmed, the next stage will be developing and building the prediction model using by going through the stages of the data analytics life-cycle (see figure 6).

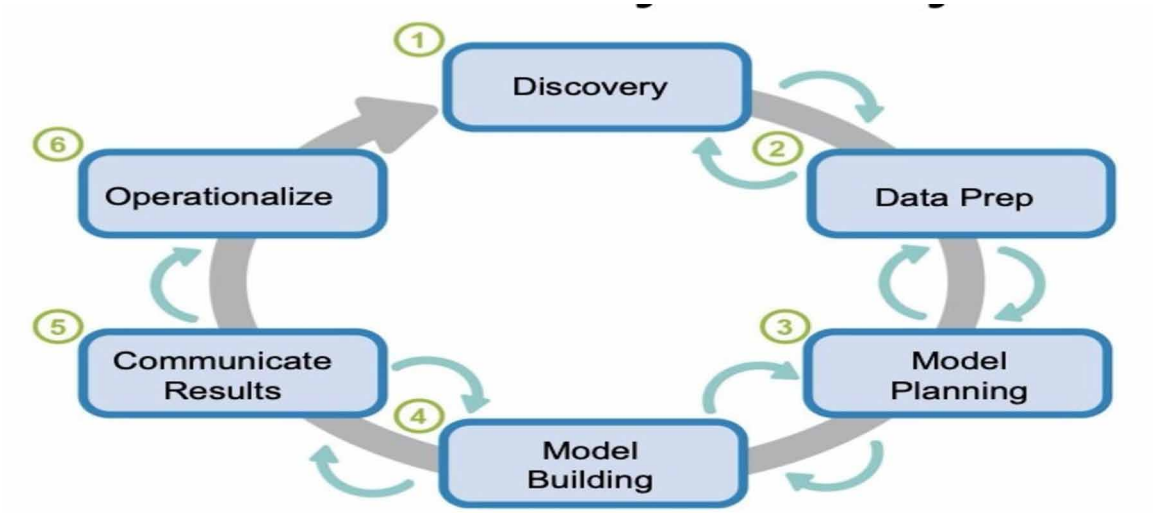
Table 2. Results of confirmatory factor analysis

Equivocality			
To present some confusing information to twenty friends			
To work out a personality problem that has affected the working relationship between	**YouTube**	**Facebook**	**Twitter**
you and your boss	0.76707626	0.56348166	0.72532141
To explain information with conflicting interpretations to a group of friends	0.93671536	0.74462947	0.94997076
To get explanation from a peer of a complicated technical matter in which you have no	0.89689475	0.99444639	1.00874173
experience of	0.88914664	0.9708842	0.98526505
To get clarification of an ambiguous information from your friends	0.98931636	1.14596963	0.92373821
Uncertainty	0.69369861	0.70215231	0.84840067
To exchange a large number of messages with your friends and work groups	0.85634457	0.83109785	1.05261164
To exchange important information that needs to be conveyed accurately	1.06685243	0.95052735	0.98898291
To transfer files and reports	1.09961721	0.86440092	0.76183117
To exchange numerical information	0.88996081	0.61953285	0.75546715
Media Richness	1.04635731	0.79655348	0.96951364
Makes it easier for me to communicate and exchange ideas with others	0.82039776	0.74113269	0.90613254
Allows me to use non-verbal cues when communicating with others	0.95152822	0.71547274	0.9002454
Provides flexible way of expression that allows me to freely use the language I am			
familiar Is a medium where I can exhibit my personal features such as looks, form etc.			

Data Preparation for Classification Model

The key activities at this stage of the data analytics life cycle include establishing analytic sandbox, performing ETLT (Extract, Transform, Load, Transform), removing outliers and missing data, as well

Figure 6. Data analytics lifecycle



as, summarizing and visualizing the data. Since the dataset was collected directly through survey with Google form, no specialised sandbox was created; instead, the responses were extracted from the google form from time to time on the personal computer as csv file using Microsoft Excel Spreadsheet, as stated previously. After performing the data exploration via factor analysis, the data was further conditioned by merging the four datasets, which represented the four SNSs used for the study. Consequently, the media preference column emerged based on how respondents indicated for each of the four SNSs. Moreover, the media richness feature was also created from the aggregation of the choices of the respondents regarding the use of the four selected SNSs for the tasks that were specified. For example:

Using YouTube makes it easier for me to communicate and exchange ideas with others

Using Facebook makes it easier for me to communicate and exchange ideas with others Using Twitter makes it easier for me to communicate and exchange ideas with others

Using Instagram makes it easier for me to communicate and exchange ideas with others

Summary and Visualization of the Data

Humans are could be described as visual beings in the sense that visualizing data facilitates understanding, and enhances the ability to discover new knowledge and gain deeper insights. It also enhances analysis and communication of results.

Skewness is used to measure the degree of symmetry of the distribution. If the distribution of the dataset stretches towards the tail ends (right and left) of the distribution, then the distribution is considered as skewed. In general, the distribution is skewed if the value is greater than +1. The skewness of the current dataset is approximately 0.5 and hence passed the test. Kurtosis is used to measure whether or

Figure 7. Accessing the dataset

```
#Accessing the data
Data_COMP=pd.read_csv('/Users/bruger/Desktop/Social_Media_Preferences (COMP2).csv')
Data_COMP.head()
```

	Confusion	Personality	Conflict	Complicated	Accurate	Ambiguous	Large	Transfer File	Numeric	Feedback	Cues	Language	P. Focus	Media Richness	Media Preference
0	5	1	2	1	4	4	3	3.0	1	2	5	4	4	15	YouTube
1	5	1	5	2	5	5	1	1.0	5	5	5	5	5	20	YouTube
2	4	4	4	5	1	3	2	1.0	1	1	1	2	1	5	YouTube
3	3	4	4	4	2	3	2	2.0	2	2	2	2	2	8	Facebook
4	2	3	4	3	3	3	4	3.0	3	4	3	3	3	13	Facebook

not the distribution is too peaked. A kurtosis value of less than -1 is considered non-normal since it is too flat. The distribution of the current dataset passed this test with a kurtosis value of 0.16 (see Figure 9).

The distributions of the media preference and media richness attributes of the dataset are represented in different plots in countplot, boxplot and bar plots in that order (see Figure 11)). The countplot shows the distribution of the media preferences, whereas both the box and bar plots illustrate the relationships between media richness and media preferences.

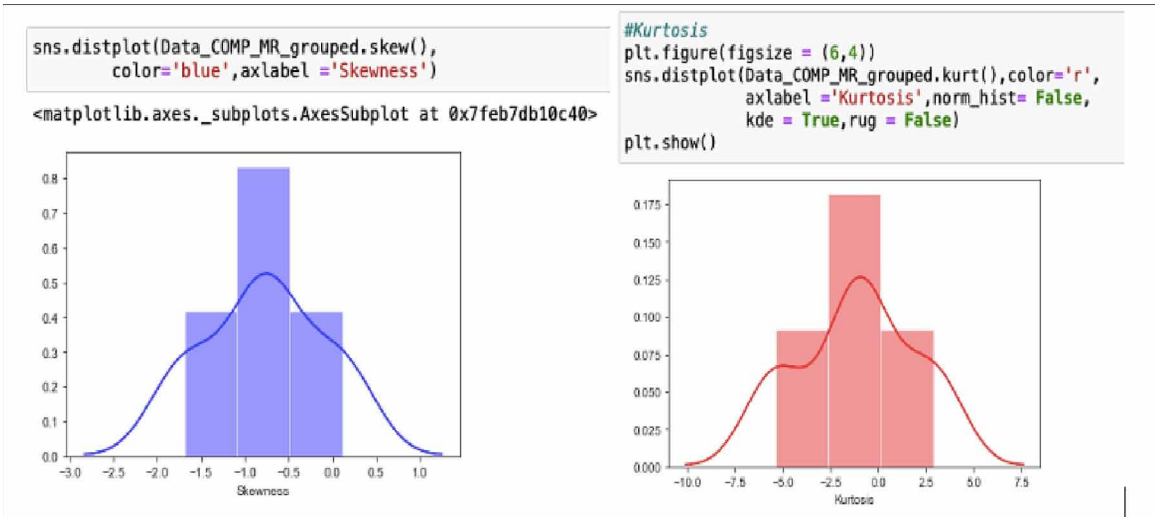
As depicted in figure 11), the most important criterion the respondents consider in selecting a SNS if the ability of the media to support multiple cues, followed by language variety across all the four selected SNSs. Regarding feedback and personal focus, it was rated differently on individual SNSs. For example,

Figure 8. Descriptive statistics

```
Data_COMP.describe()
```

	Confusion	Personality	Conflict	Complicated	Accurate	Ambiguous	Large	Transfer File	Numeric	Feedback	Cues
count	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1136.000000	1128.000000	1136.000000	1136.000000	1136.000000
mean	3.542254	3.450704	3.746479	3.788732	3.584507	3.802817	3.549296	3.397163	3.429577	3.683099	3.936620
std	1.130269	1.208355	1.091007	1.168340	1.269044	1.134105	1.196632	1.336828	1.269905	1.109794	1.206238
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
75%	4.000000	4.000000	4.000000	5.000000	4.000000	5.000000	4.000000	4.000000	4.000000	4.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Figure 9. Measuring skewness and kurtosis of the data



providing instant feedback was rated higher in Facebook but was the least considered in YouTube. The basic rule of the correlation between 2 quantities is that the correlation ranges between -1 and +1 and values closer to zero means no correlation, closer to 1 means positive and strong correlation, which means as one quantity increases so does the other. Correlation closer to -1 means one quantity decreases same way as the other decreases. The correlation between the features for media richness determinants is represented on a heatmap in figure 12).

Figure 10. Examining distribution of media richness

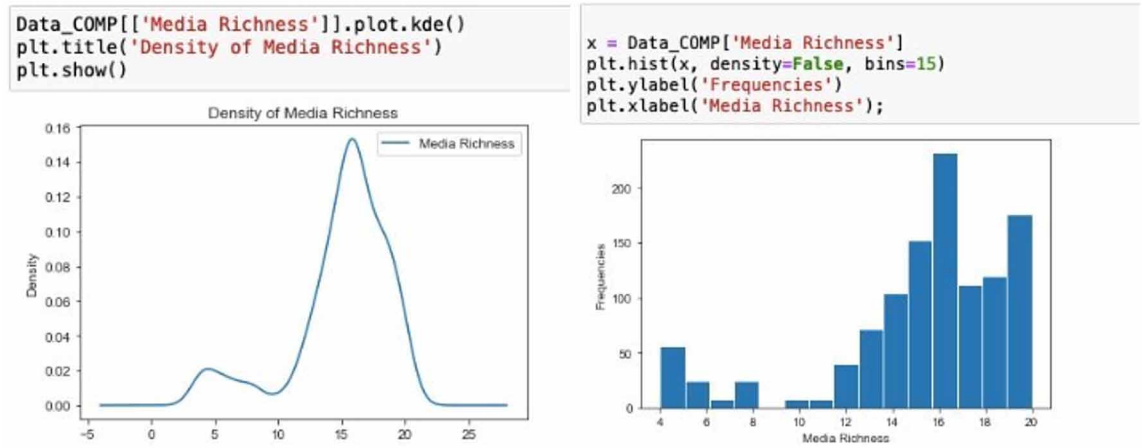
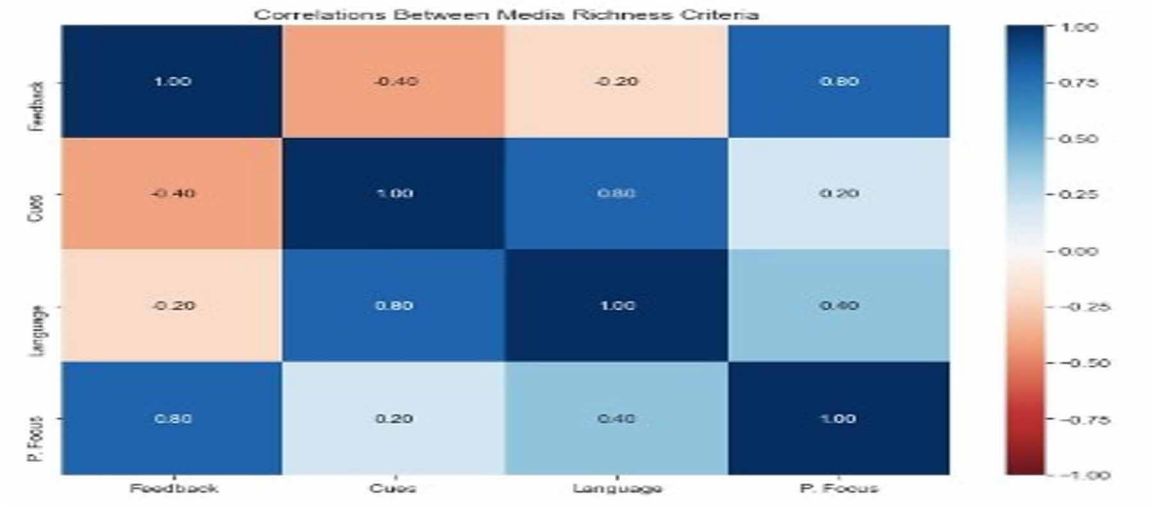


Figure 11. Observing media preference with count, box, and bar plots

Figure 12. Examining correlation among media richness criteria



Model Planning

The goal of the section is to answer the question: "What kind of problem do I want to solve and how can I solve it?" That narrows down to choosing the appropriate classifiers and briefly explaining how they work and their suitability for the problems being dealt with.

The problem is a multiclass classification problem as it involves predicting the appropriate SNS using the media richness criteria (Feedback, Cues, Language, and P.Focus) as input features, with four SNSs (YouTube, Facebook, Twitter, and Instagram) being the labels/class. Multiclass problems are different from binary classification in the sense that the latter involves predicting an object from one of two classes (e.g., "Yes" or "No"), whereas the former involves making prediction from one of N classes (e.g., "Apple", "Orange" or "Grape"). That requires a bit of caution since not all models that support

binary classification inherently supports multiclass classification. In a multiclass classification, each training point is a member of N different classes and the goal to build a function for which if given a new data point, would be able to predict correctly the class to which the new point belongs (Rafkin, 2018). There are several approaches for solving multiclass problems and the two simple ones are One-vs-All and All-vs-All classification.

In One-vs-All technique, a binary classifier is selected and N different classifiers are built. For the ith classifier, represent the positive examples by all the points in class i, and the negative examples with all the points not in class i. Let *fi* be the ith classifier and then use *f(x)* for the classification, where:

With All-vs-All classification, *N (N-1)* classifiers are built and one classifier is used to distinguish each pair of classes *i* and *j*. Use f_{ij} as the classifier where class I represent positive examples and class j as negative. With f_{ji}=-f_{ij} classify with f*(x)* where:

Four models known to be capable of handling multiclass problem were selected, trained and evaluated. These include: K-NN, Random forest, and SVM.

K-Nearest Neighbors (K-NN): The K-NN algorithm is described as a non-parametric lazy learner. Non-parametric in the sense that it does not take into consideration any assumption regarding the underlying distribution and lazy because it does not require training set of data for model building. The algorithm works in three basic steps (see figure 13). Suppose S1 is the point required to be classified:

1. Find the k points that are close to S1.
2. Calculate the distance (e.g., Euclidean distance) between S1 and the k-points to determine those that are closest to S1.
3. Classify S1 by the majority votes of its k-neighbors (each point votes for their class and the class with the most votes is taken as the prediction).
4. The value k is determined by the scientist in a way that minimizes error (Usually an odd number is chosen for an even number of classes).

Figure 13. How K-NN algorithm works
Source: (Navlani, 2018)

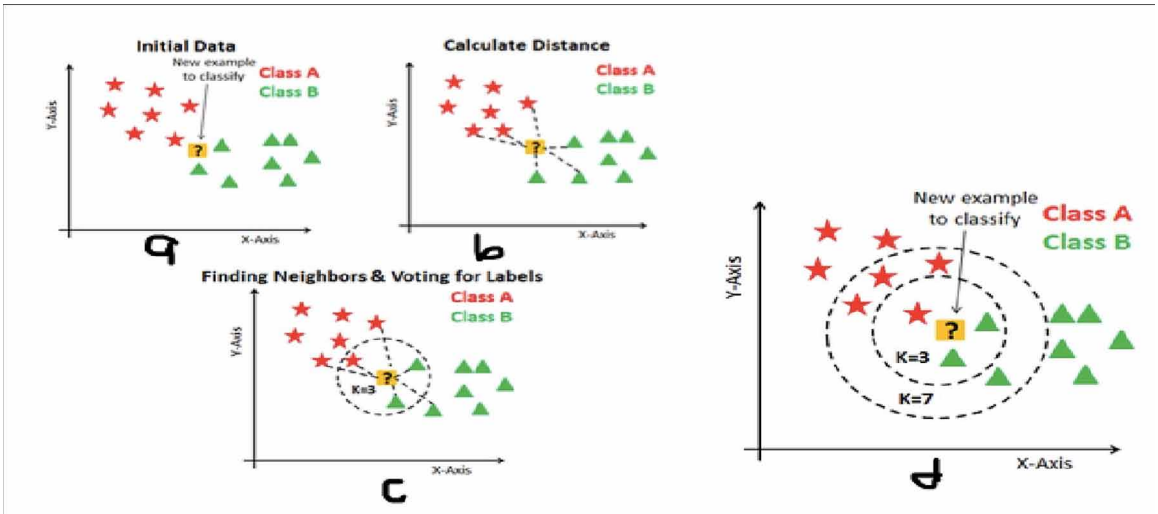
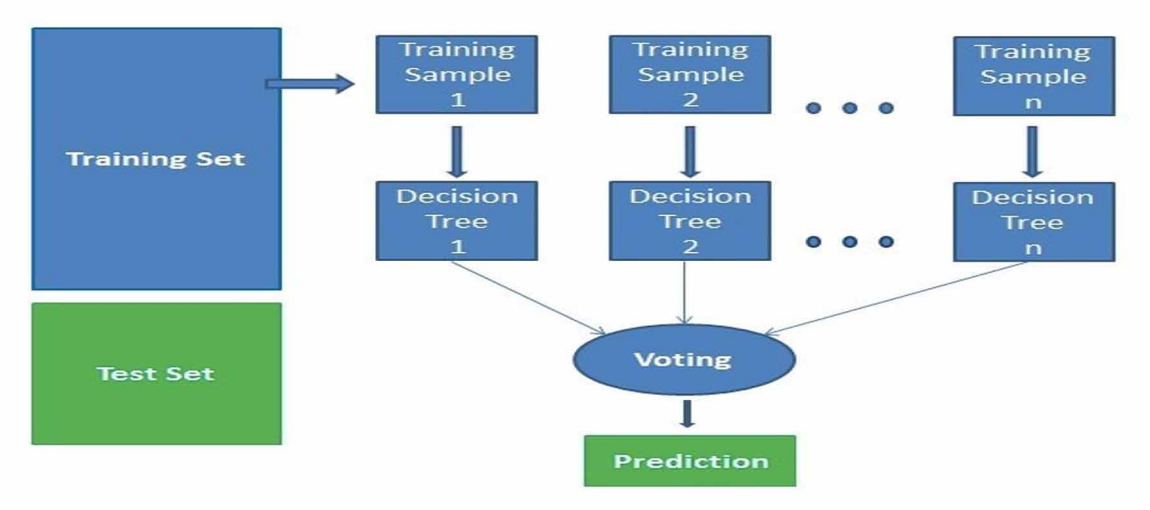


Figure 14. How random forest algorithm works
Source: (Navlani (b), 2018)



Random Forest: Technically, the algorithm is an ensemble method comprising of a collection of decision tree classifiers known as the *forest*. The individual decision trees are generated by calculating the attribute selection indicators such as information gain, gain ratio, and Gini index for each attribute. During a classification problem, each of the trees votes and the most popular voted class is used as the outcome.

The Random Forest Classifier works in four steps:

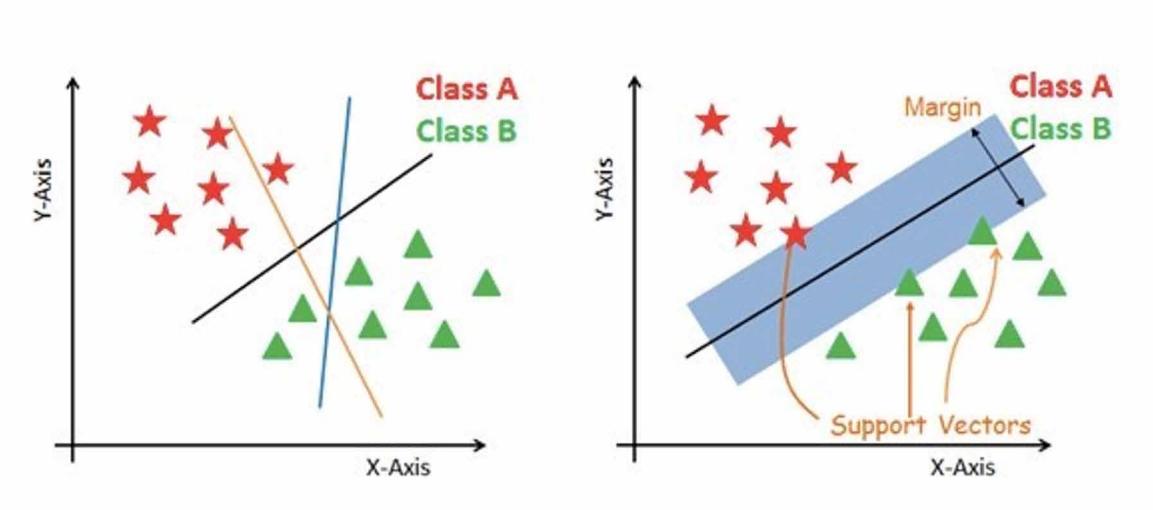
- Select random samples from a given dataset
- Construct a decision tree for each sample and obtain prediction results from each decision tree
- Perform a vote for each predicted result
- Select the prediction results with the most votes as the final prediction

Support Vector Machine (SVM): The main goal of SVM is to segregate a given dataset into classes in the best way possible. It accomplishes this goal by constructing an optimal hyperplane in a multidimensional space to separate the different classes.

The primary ideas are to find a maximum marginal hyperplane (MMH) that best divides the dataset into classes with minimum error and maximum possible margin between support vectors. The algorithm determines MMH in these steps:

- SVM generates three hyperplanes (black, blue and orange). However, the blue and orange hyperplanes show high classification errors, while the black line correctly separates the two classes.
- The right hyperplane is determined as having maximum segregation from the nearest datapoints from either side (See right-hand side figure 15).

Figure 15. How SVM algorithm works
Source: (Navlani, 2019)



Model Building

The model-building phase involves training the selected models which includes:

- Evaluating the fitted model and
- Adjusting the model accordingly

As explained earlier, the three selected models were trained and evaluated with a train: test split ratio of 80:20 for SVM and K-NN and 70:30 for Random Forest. The reason for the difference in the ration was because in each case the hyperparameters were fine-tuned for the best results possible. The F1 scores indicate accuracies of 66%, 65%, and 64% for SVM, Random Forest, and K-NN respectively on the training set with corresponding 61%, 60%, and 60% on the test sets (see code snippets for SVM, K-NN, and Random Forest, respectively). A detailed outcome of the evaluation will be discussed next.

Classification Report - Metrics

The metrics include precision, Recall, F1-Score, and Support scores.

Precision score is defined as the ratio of the true positives to the total predicted positives.

Recall score is defined as the ratio of true positives to the total actual positives. *Accuracy* is defined as total number of correct predictions divided by the total number of samples (See Figure 16).

Other metrics recorded were F1-score and support. F1 score is determined by multiplying precision and recall and dividing the result by the sum of precision and recall. It is useful for interpreting the balance between the precision score and the recall score. Support is simply the size of the test data used for the evaluation.

Figure 16. Classification report metrics

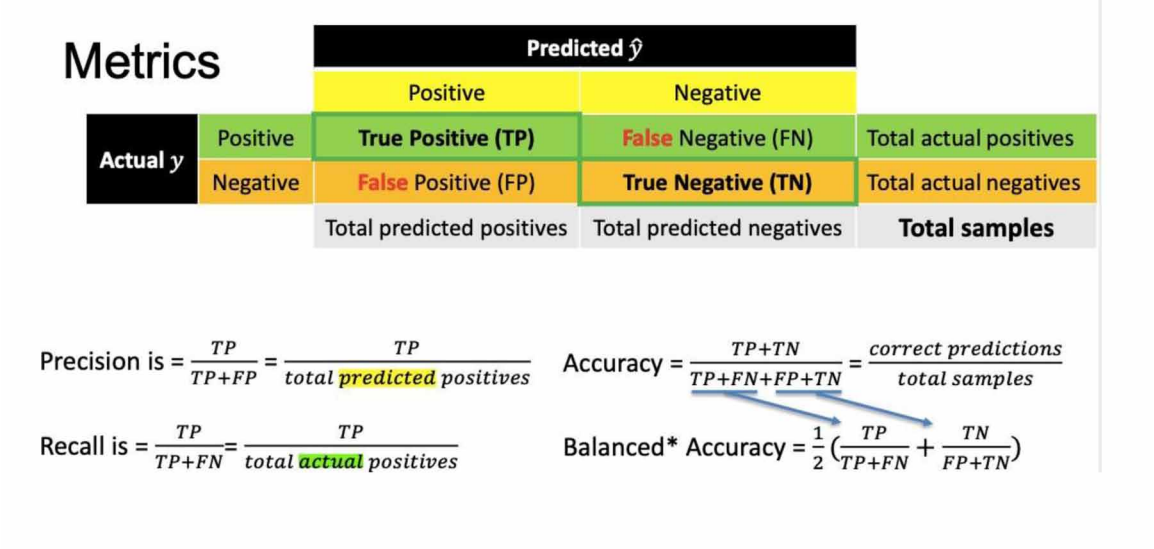
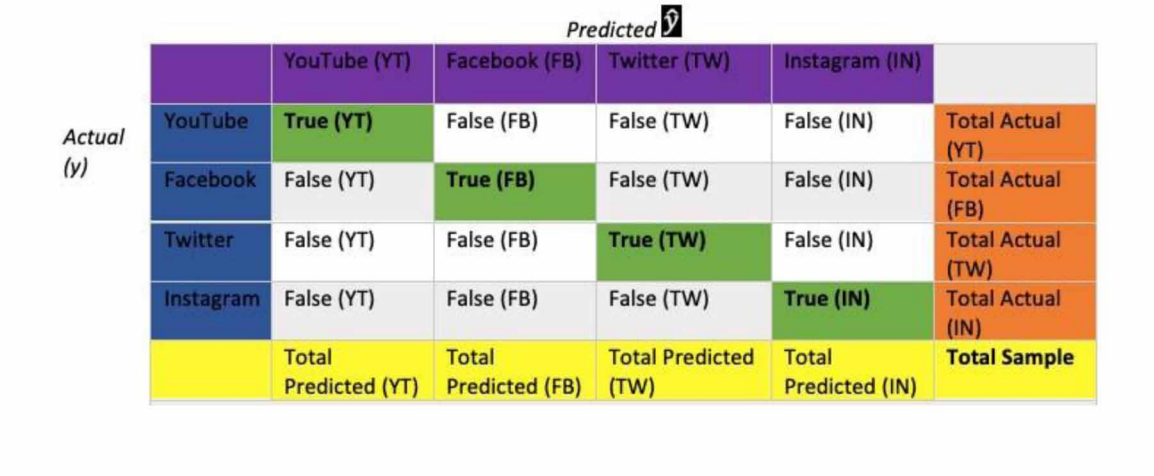


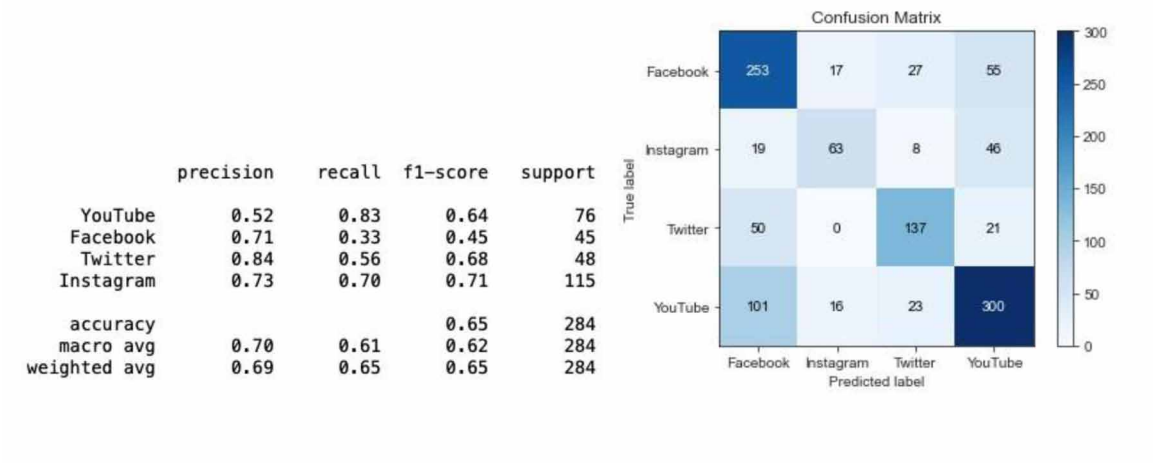
Figure 17. Confusion matrix for snss



Confusion Matrix

The confusion matrix is used to check the ability of the model to rightly predict the SNSs as it was trained to do. The results as recorded in the confusion matrix indicates that random forests got 253 predictions correct for Facebook, 63 for Instagram, 137 for Twitter, and 300 for YouTube, with an accuracy of 65%. Similarly, SVM rightly predicted 169 for Facebook, 25 for Instagram, 33 for Twitter, and 385 for YouTube, resulting in an accuracy of 55%. Finally, K-NN performed better than SVM by obtaining an accuracy of 63%, by correctly predicting 207 for Facebook, 70 for Instagram, 138 for Twitter and 280

Figure 18. Results of random forest performance evaluation



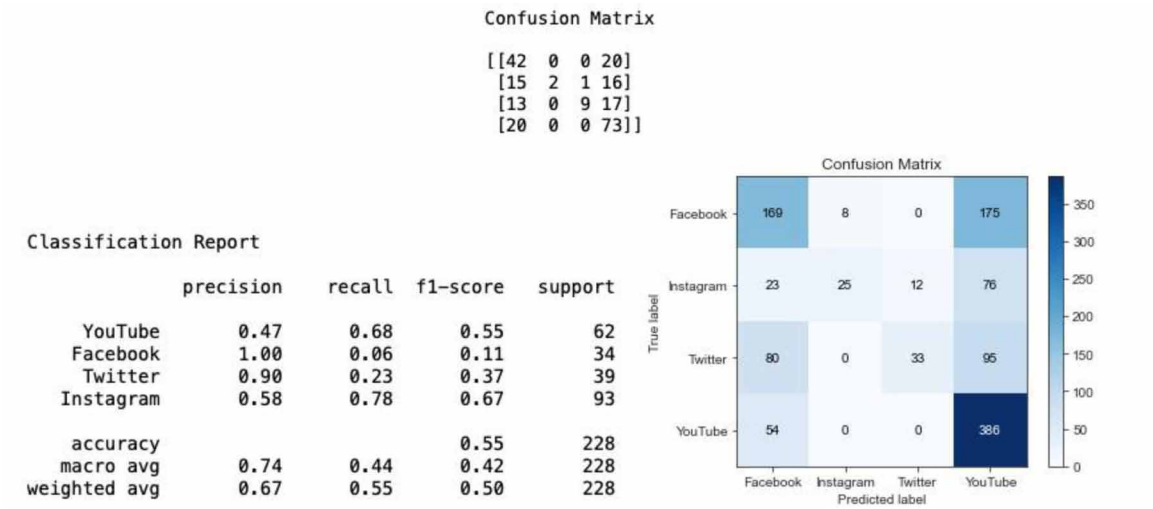
	precision	recall	f1-score	support
YouTube	0.52	0.83	0.64	76
Facebook	0.71	0.33	0.45	45
Twitter	0.84	0.56	0.68	48
Instagram	0.73	0.70	0.71	115
accuracy			0.65	284
macro avg	0.70	0.61	0.62	284
weighted avg	0.69	0.65	0.65	284

for YouTube. Since random forest performed better than SVM and K-NN, it was selected for building the model for the social media predictions, which is discussed in the next section.

Communication of Results

Potentially, the implementation of the system would definitely lead to some changes in the business processes of organizations. For example, if currently there are no restrictions on the use of SNSs in the organization so long as members use these platforms for business purposes. With the implementation of the systems some SNSs may no longer be used, some may rarely be used, while others would be used

Figure 19. Results of SVM performance evaluation



Confusion Matrix

```
[[42  0  0 20]
 [15  2  1 16]
 [13  0  9 17]
 [20  0  0 73]]
```

Classification Report

	precision	recall	f1-score	support
YouTube	0.47	0.68	0.55	62
Facebook	1.00	0.06	0.11	34
Twitter	0.90	0.23	0.37	39
Instagram	0.58	0.78	0.67	93
accuracy			0.55	228
macro avg	0.74	0.44	0.42	228
weighted avg	0.67	0.55	0.50	228

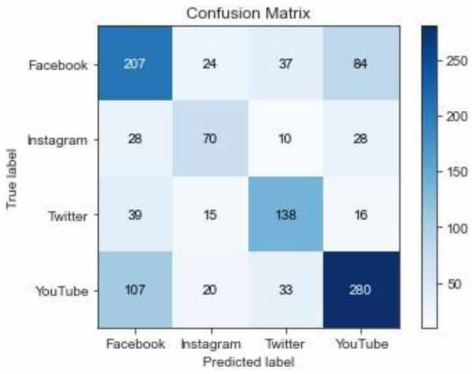
Figure 20. Results of K-NN performance evaluation

Confusion Matrix

```
[[48 10  4 14]
 [11 27  2  5]
 [ 6  0 32 10]
 [27 15  1 72]]
```

Classification Report

	precision	recall	f1-score	support
YouTube	0.52	0.63	0.57	76
Facebook	0.52	0.60	0.56	45
Twitter	0.82	0.67	0.74	48
Instagram	0.71	0.63	0.67	115
accuracy			0.63	284
macro avg	0.64	0.63	0.63	284
weighted avg	0.65	0.63	0.64	284



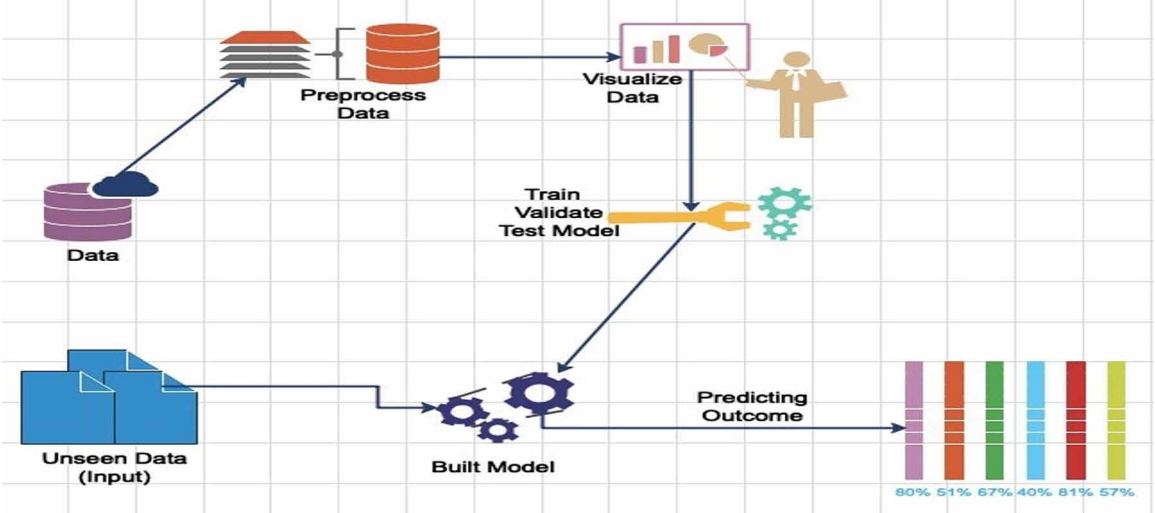
very often since the system determines will focus on those SNSs that are more useful to the organization. Unlike previously when members could select any SNS to share their knowledge, this time round you need to specify the criteria for the system to determine what SNS would be most appropriate.

Technical Specifications for Implementing the Code

The aim of the system is to allow users to select the most suitable SNSs based on their knowledge needs using the media richness criteria. As previously explained, the data for the study was gathered through surveys. It was then preprocessed, explored and visualized to expose any form anomalies in the data as well as any imbalances in the classes. The dataset was split into training and testing sets and three carefully selected predictive classifiers were trained and evaluated (See figure 21). After rigorous analysis of the metrics, Random Forest Classifier was selected to be used for the system development. Random forest is known to be robust for learning multiclass problems with capacity of handling high-dimensional features spaces, which a characteristics of classification trees (Rafkin, 2018). As a bagged classifier, random forest combines a collection of classification or regression trees and uses both random input and feature selection to yield favorable error rates.

Moreover, random forest is more robust to noise, which makes it an accurate classifier in diverse domains including the current study (Rafkin, 2018). At this point, a well-documented code with meaningful variables names has been provided. As explained in the sections following, there are four key input data elements defined as the media richness criteria: Feedback, cues, language, and personal focus. Users would be required to provides these inputs according to the degree/or how much of each they require in the knowledge they are sharing. For instance, “I need a quick feedback, mild cues, very good language variety, and mild personal focus”. The system uses those as input for the predictions. These input data are ordinal (qualitative) and don’t require any units. No default conditions are required except

Figure 21. System overview

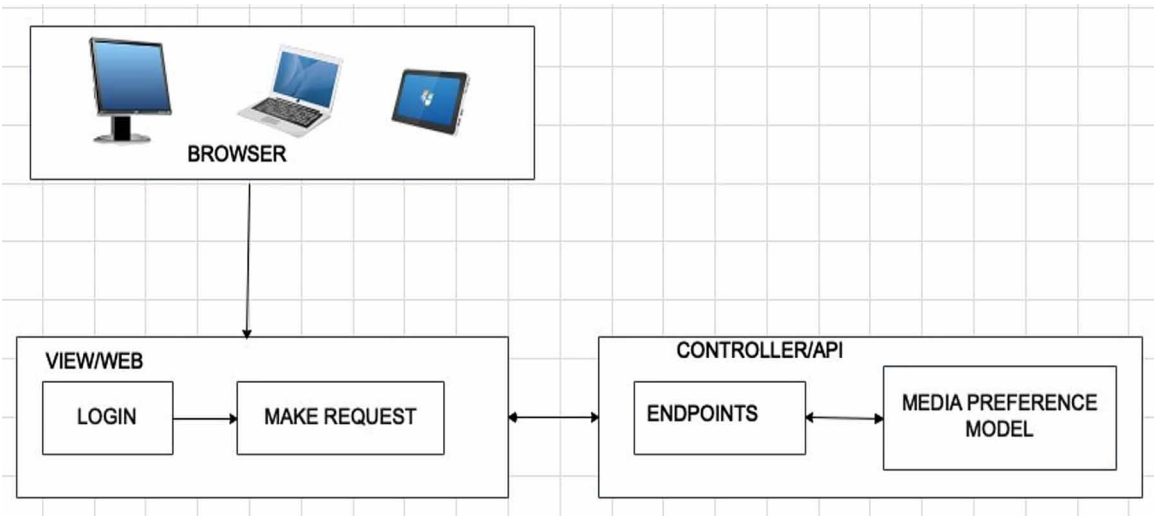


those specified in the hyperparameter within the models themselves. The model outputs are simple and straightforward and do not require any further explanations on how to apply them. However, the model needs to be retrained continuously to improve on its accuracy and precision.

Operationalization

The system comprises of three components: Model, View and Controller (MVC) as prescribed in the MVC architecture. The system has been developed to hide all the technical details of the multiclass classification and only present users with a simplified web-based user interface to enhance their knowledge

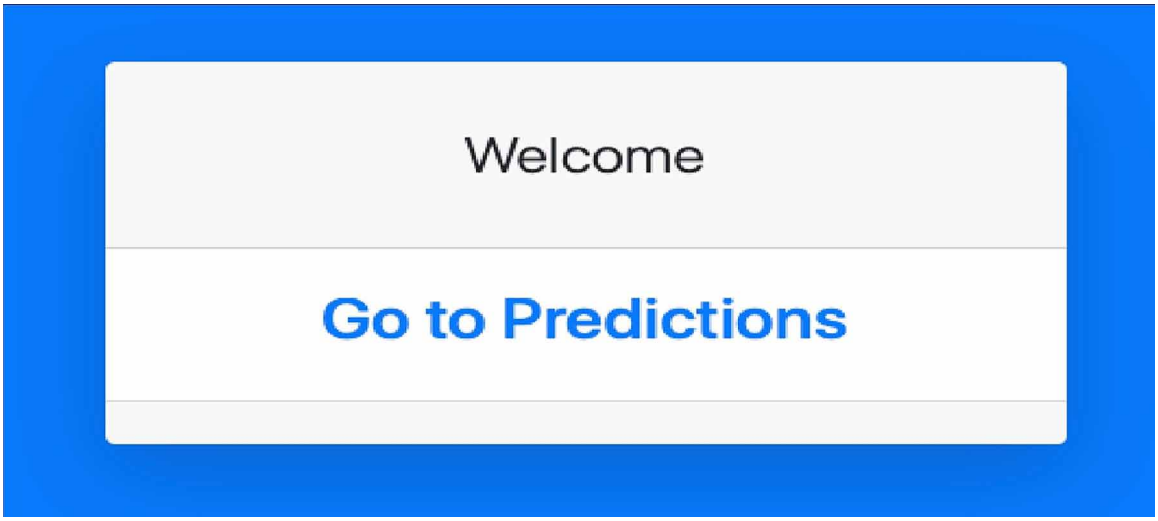
Figure 22. System overview of user interaction with GUI



sharing experience. In effect, users are able to interact with the system via the front-end application interface provided. After building the prediction model, it is pickled and linked up to the controller, which is responsible for processing the request from the users through an API. The system is initialized when a user makes an HTTP request from the browser. The controller then fetches the predictions from the machine learning model and displays the output via the view to the user (see figure 23).

1. When the user clicks the link to the website, he/she is directed to the landing page. The user can then begin the predictions by selecting "Go to Predictions".

Figure 23. Social media landing page



2. That brings the user to the social media preference page. The user is then supposed to select inputs for all the media richness criteria, with each criterion having five options ("None", "Mild", "Good", "Very Good", "Excellent")

3. After selecting all the options, the user could then select "Get Preference" to receive the SNS that will be most suitable depending on the degree of richness specified via the media richness criteria.

4. As an example, the user choose the following preferences: Feedback-mild, LanguageGood, Cues-Excellent, and P.focus-Very Good. The outcome was Twitter.

Training*:* As part of the operationalization phase, there will be the need to organize some form of training for new onboard team members both technical and analysts. This particular group of users would need the detailed information about the data source, code and requirement and so that would be made available to them. Since a project of this caliber would impact on organizational policies, especially for knowledge sharing, there would be the need to describe the current business situation and how the

Figure 24. Social media preference page

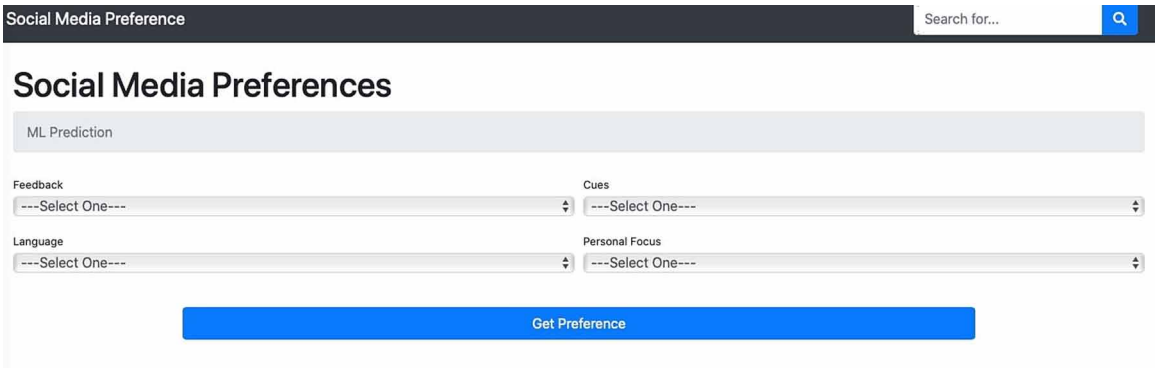


Figure 25. Social Media preferences

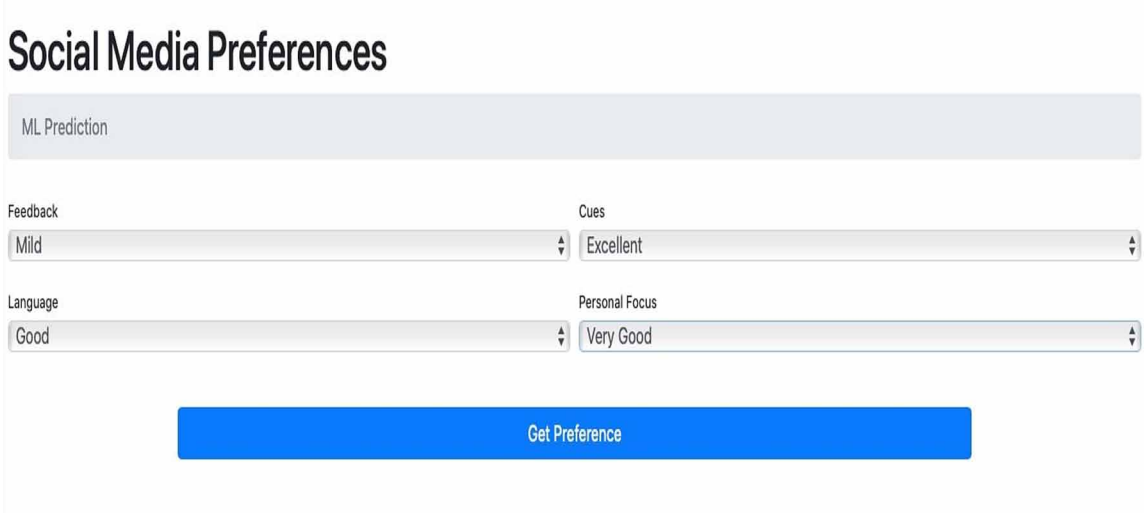
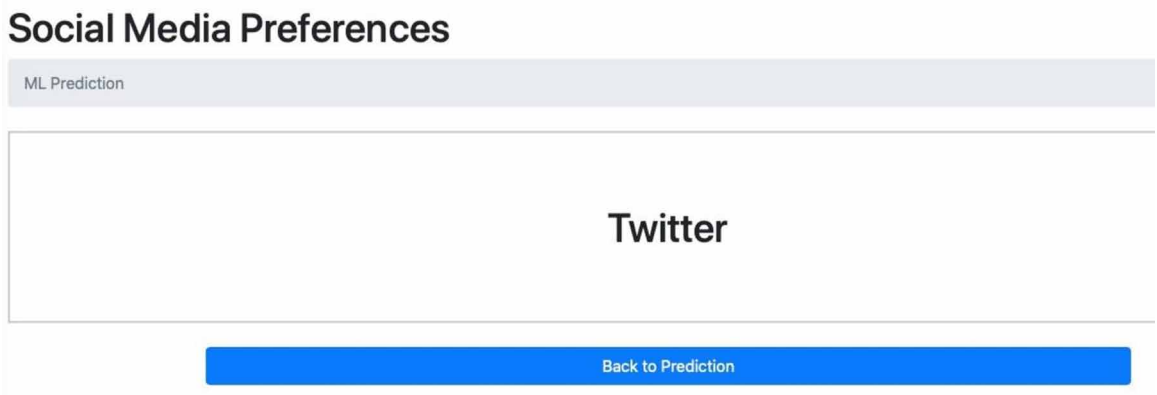


Figure 26. Social media preferences (Twitter)



system stands to benefit the organization, changes in reporting with these groups, as well as providing them the technical specifications for implementing the code.

Pilot Deployment and Testing*:* In order to transition the project from development environment to production, there is the need to follow the change procedures and approval within the organization. Normally, such projects are allowed to operate alongside the existing systems while testing its robustness and reliability. When the pilot period is over, the old would give way to the new through the laid down procedures.

Monitoring and Re-training Model*:* The solution requirements may require changes from time to time and so some changes may be required in the nearest future such as emergence of new SNSs, as well as, including more training dataset. In order to facilitate smooth transitions of the project due to some of these changes, a technical document detailing every step required to make such changes has been provided.

Limitations of the Study

The study was limited to only four SNSs (YouTube, Facebook, Twitter and Instagram), whilst there still remains a plethora of other SNSs that could have been considered. Which means the study is limited to only these four SNSs.

The performance accuracies of the model were very poor, which may be due to underlying distribution of the dataset, hence a more robust data collection strategies may help in alleviating some of these shortcomings.

The system was developed using only media richness criteria, meanwhile other social influence factors could have been included as well.

FUTURE RESEARCH

- More SNSs could be included in the project in the future.
- More data and more training data could improve the model performance in the future.
- Additional functionalities could be added to the application to enhance its usefulness in the future.

Link to the Application:

http://smpreferences.ny-2.paas.massivegrid.net/smpreference/pages/social_media/sm_preferences.xhtml

Link to Video Presentation:

https://www.youtube.com/watch?v=lyTHxBA38qY&list=PLc9JP_23neAhrb1h1w70jp0I1sA6vbiU

REFERENCES

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), i–xxxii. doi:10.17705/1jais.00423

Ahmed, Y. A., Ahmad, M. N., Ahmad, N., & Zakaria, N. H. (2019). Social media for knowledge-sharing: A systematic literature review. *Telematics and Informatics*, *37*, 72–112. doi:10.1016/j.tele.2018.01.015

Aldahdouh, T. Z., Nokelainen, P., & Korhonen, V. (2020). Technology and Social Media Usage in Higher Education: The Influence of Individual Innovativeness. *SAGE Open*, *10*(1), 215824401989944. doi:10.1177/2158244019899441

Barros, V. F. D. A., Ramos, I., & Perez, G. (2015). Information systems and organizational memory: A literature review the last 20 years. *Journal of Information Systems and Technology Management*, *12*(1), 1–5. doi:10.4301/S1807-17752015000100003

Brownson, R. C., Eyler, A. A., Harris, J. K., Moore, J. B., & Tabak, R. G. (2018). Getting the Word Out: New Approaches for Disseminating Public Health Science. *Journal of Public Health Management and Practice*, *24*(2), 102–111. doi:10.1097/PHH.0000000000000673 PMID:28885319

Carvalho, N., & Gomes, I. (2017). Knowledge Sharing between Enterprises of the Same Group. *International Journal of Knowledge Management*, *13*(1), 34–52. doi:10.4018/IJKM.2017010103

Chan, A. K. M., Nickson, C. P., Rudolph, J. W., Lee, A., & Joynt, G. M. (2020). Social media for rapid knowledge dissemination: Early experience from the COVID -19 pandemic. *Anaesthesia*, *75*(12), 1579–1582. doi:10.1111/anae.15057 PMID:32227594

Chao, S. H., Jiang, J., Hsu, C. H., Chiang, Y. T., Ng, E., & Fang, W. T. (2020). Technology-Enhanced Learning for Graduate Students: Exploring the Correlation of Media Richness and Creativity of Computer-Mediated Communication and Face-to-Face Communication. *Applied Sciences (Basel, Switzerland)*, *10*(5), 1602. doi:10.3390/app10051602

Chen, J. V., McBush Hiele, T., Kryszak, A., & Ross, W. H. (2020). Predicting Intention to Participate in Socially Responsible Collective Action in Social Networking Website Groups. *Journal of the Association for Information Systems*, 342–363. doi:10.17705/1jais.00604

Cho, H. J., Feldman, L. S., Keller, S., Hoffman, A., Pahwa, A. K., & Krouss, M. (2020). Choosing Wisely in the COVID-19 Era: Preventing Harm to Healthcare Workers. *Journal of Hospital Medicine*, *15*(6), 360–362. doi:10.12788/3457 PMID:32490804

Confirmatory Factor Analysis — A Case study. (2019, May). https://arxiv.org/abs/1905.05598

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. *Management Information Systems Quarterly*, *11*(3), 355. doi:10.2307/248682

Delen, D., Zaim, H., Kuzey, C., & Zaim, S. (2013). A comparative analysis of machine learning systems for measuring the impact of knowledge management practices. *Decision Support Systems*, *54*(2), 1150–1160. doi:10.1016/j.dss.2012.10.040

Diamantini, C., Potena, D., & Storti, E. (2013). A virtual mart for knowledge discovery in databases. *Information Systems Frontiers*, *15*(3), 447–463. doi:10.100710796-012-9399-0

El-Shinnawy, M., & Markus, M. L. (1997). The poverty of media richness theory: Explaining people's choice of electronic mail vs. voice mail. *International Journal of Human-Computer Studies*, *46*(4), 443–467. doi:10.1006/ijhc.1996.0099

Ghazali, S., Sulaiman, N., Zabidi, N., Omar, M., & Alias, R. (2016, October). The impact of knowledge sharing through social media among academia. *AIP Conference Proceedings*, *1782*(1), 030003. doi:10.1063/1.4966060

Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The Perils and Promises of Big Data Research in Information Systems. *Journal of the Association for Information Systems*, 268–293. doi:10.17705/1jais.00601

Gyamfi, A. (2016). The Impact of Media Richness on the Usage of Web 2.0 Services for Knowledge Transfer. *International Journal of E-Services and Mobile Applications*, *8*(2), 21–37. doi:10.4018/IJESMA.2016040102

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324), 486–488. doi:10.1126cience.aal3856 PMID:28154051

Irani, Z., Sharif, A. M., Papadopoulos, T., & Love, P. E. D. (2017). Social media and Web 2.0 for knowledge sharing in product design. *Production Planning and Control*, *28*(13), 1047–1065. doi:10.1080/09537287.2017.1329955

Ishii, K., Lyons, M. M., & Carr, S. A. (2019). Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies*, *1*(2), 124–131. doi:10.1002/hbe2.138

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Kindel, A. T., Bansal, V., Catena, K. D., Hartshorne, T. H., Jaeger, K., Koffman, D., McLanahan, S., Phillips, M., Rouhani, S., Vinh, R., & Salganik, M. J. (2019). Improving Metadata Infrastructure for Complex Surveys: Insights from the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World*, *5*, 237802311881737. doi:10.1177/2378023118817378

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-Driven Meets Theory-Driven Research in the Era of Big Data: Opportunities and Challenges for Information Systems Research. *Journal of the Association for Information Systems*, 1253–1273. doi:10.17705/1jais.00526

Mathews, B. K., Koenig, S., Kurian, L., Galen, B., Mints, G., Liu, G., & Soni, N. J. (2020). Clinical Progress Note: Point-of-Care Ultrasound Applications in COVID-19. *Journal of Hospital Medicine*, *15*(6), 353–355. doi:10.12788/jhm.3454 PMID:32490807

Mirbabaie, M., Bunker, D., Stieglitz, S., Marx, J., & Ehnis, C. (2020). Social media in times of crisis: Learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *Journal of Information Technology*, *35*(3), 195–213. doi:10.1177/0268396220929258

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *The Journal of Economic Perspectives*, *31*(2), 87–106. doi:10.1257/jep.31.2.87

Navlani, A. (2018a, May 16). *Understanding Random Forests Classifiers in Python*. Datacamp. https://www.datacamp.com/community/tutorials/random-forests-classifier-python

Navlani, A. (2018b, August 2). *KNN Classification using Scikit-learn*. Datacamp. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

Navlani, A. (2019, December 27). *Support Vector Machines with Scikit-learn*. Datacamp. https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python

O'Reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *International Journal of Digital Economics*, *65*, 17–37. https://mpra.ub.uni-muenchen.de/4580/

Panahi, S., Watson, J., & Partridge, H. (2013). Towards tacit knowledge sharing over social web tools. *Journal of Knowledge Management*, *17*(3), 379–397. doi:10.1108/JKM-11-2012-0364

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, *31*(7), 770–780. doi:10.1177/0956797620939054 PMID:32603243

Prinzie, A., & Van den Poel, D. (2008). Random Forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications*, *34*(3), 1721–1732. doi:10.1016/j.eswa.2007.01.029

Qi, C., & Leung, H. (2015, June). *Exploring the role of social media and knowledge management processes in organizational learning*. https://research.polyu.edu.hk/en/publications/exploring-the-role-of-social-media-and-knowledge-management-proce

Rafkin, R. (2008, February 25). *Multiclass Classification*. MIT. https://www.mit.edu/~9.520/spring09/Classes/multiclass.pdf

Reuter, C., Kaufhold, M. A., Schmid, S., Spielhofer, T., & Hahne, A. S. (2019). The impact of risk cultures: Citizens' perception of social media use in emergencies across Europe. *Technological Forecasting and Social Change*, *148*, 119724. doi:10.1016/j.techfore.2019.119724

Schwartz, D. G. (2007). Integrating knowledge transfer and computer-mediated communication: Categorizing barriers and possible responses. *Knowledge Management Research and Practice*, *5*(4), 249–259. doi:10.1057/palgrave.kmrp.8500153

Simonin, B. L. (1999a). Ambiguity and the process of knowledge transfer in strategic alliances. *Strategic Management Journal*, *20*(7), 595–623. doi:10.1002/(SICI)1097-0266(199907)20:7<595::AID-SMJ47>3.0.CO;2-5

Simonin, B. L. (1999b). Transfer of Marketing Know-How in International Strategic Alliances: An Empirical Investigation of the Role and Antecedents of Knowledge Ambiguity. *Journal of International Business Studies*, *30*(3), 463–490. doi:10.1057/palgrave.jibs.8490079

Uygur, U. (2013). Determinants of Causal Ambiguity and Difficulty of Knowledge Transfer within the Firm. SSRN *Electronic Journal*, 742–755. doi:10.2139srn.2336794

Xu, L., Qiu, J., Gu, W., & Ge, Y. (2020). The Dynamic Effects of Perceptions of Dread Risk and Unknown Risk on SNS Sharing Behavior During Emerging Infectious Disease Events: Do Crisis Stages Matter? *Journal of the Association for Information Systems*, *21*(3), 545–573. doi:10.17705/1jais.00612

Zhang, Z., Gonzalez, M. C., Morse, S. S., & Venkatasubramanian, V. (2017). Knowledge Management Framework for Emerging Infectious Diseases Preparedness and Response: Design and Development of Public Health Document Ontology. *JMIR Research Protocols*, *6*(10), e196. doi:10.2196/resprot.7904 PMID:29021130

Chapter 3
A Business Merger and Acquisition Knowledge Management System Using Artificial Intelligence Techniques

Kamalendu Pal
https://orcid.org/0000-0001-7158-6481
City, University of London, UK

ABSTRACT

Knowledge is recognized as a strategic resource, with major key drivers being the need to cut time to market and gain the business opportunities in a global market with new products and services. This chapter presents a knowledge management system known as guidance for business merger and acquisition (GBMA) process. This application uses a hybrid knowledge-based system to place bidding on the target company, formulating a strategy, and modification of the initial strategy if necessary, for the business acquisition processes. Legal knowledge for GBMA is represented in two forms, as rules and cases. Besides distinguishing the two different forms of knowledge representation, the chapter outlines the actual use of these forms in a computational architecture that is designed to generate a suitable solution, for a given new business scenario, using different reasoning mechanisms (e.g., rule-based reasoning, case-based reasoning). Business scenarios are used to show the functionalities of the presented architecture.

INTRODUCTION

In the business world, there is a specific governing rule – survive and thrive or die. Business organizations (e.g., merger and acquisition consulting firm, law firm, accounting firm, academic institution) that continuously pursue growth and innovation will take away market share from competitors, create profits, and provide returns to shareholders. Those who do not grow or innovate tend to stagnate, lose customers,

DOI: 10.4018/978-1-7998-6792-0.ch003

and market share, and destroy shareholder value. The business strategy of merger and acquisition plays a vital role in enabling strong companies to grow faster than competitors, providing entrepreneurs rewards for their efforts, and swallowing weaker companies, or worse, making them irrelevant through exclusion and ongoing share erosion. Merger and acquisitions are prevalent in most economies and can be a lever that business organizations use to provide returns to owners and investors. Due to the potential for contributing to returns, mergers and acquisitions can be attractive to entrepreneurs and owners to create value in a company. The motivation for mergers and acquisitions played a variety of roles in corporate history, ranging from the "*greed is good*" corporate raiders buying companies in a hostile manner to the world trend to use mergers and acquisitions for external growth and industry consolidation.

A merger and acquisition can be classified as either horizontal, vertical, conglomerate, or concentric. Horizontal or related mergers and acquisitions combine two similar companies in a related line of business in the same industry. This can be the case of a merger between direct competitors. Vertical mergers and acquisitions unite companies from successive processes within the same industry. They refer to organization in a supplier-customer relationship. Conglomerate mergers and acquisitions occur between companies in completely unrelated business fields. They are often a consequence of diversification strategies. Concentric mergers and acquisitions combine organizations from different but related industries. This often happens when an acquirer tries to expand into other fields of business acidity.

The motivations for a merger or acquisition can be manifold, and the fundamental underlying principle is to create shareholder value that is greater than the sum of that for the two companies. That two business organizations together are more valuable than two separate companies is the belief behind mergers and acquisitions. This justification is mainly alluring to business organizations when times are tough. Dominating business will act to buy other businesses to do a more cost-effective, competitive business. One of the promising goals of mergers and acquisitions is to gain a more significant efficiency or gain a considerable market share. Because of these leading advantages, target companies will often agree to be purchased when they know they cannot survive alone. Many companies embrace mergers and acquisitions as a growth strategy, and they are widely believed to increase the net value of the business. Nevertheless, it is a fact that many mergers and acquisitions ended in financial disappointment.

Mergers and acquisitions matchmakers frequently cite synergy to justify their strategic actions. The world synergy is originated from the Greek linguistic term '*synergos*' and its literal interpretation is "working as one unit". Incorporating business usage and synergy refers to the ability of two or more companies to generate more excellent value working together than they could be working individually. More intuitively, one plus one makes three is the unique alchemy of synergy. Synergy is the magic force that drives many mergers and acquisition. Realizing the synergy potential in a merger or business acquisition is an uphill battle. Achieving synergy is simpler said than done – it is not automatically realized once two businesses merge. Sure, there ought to be the economics of scale when two organizations unified, but sometimes an acquisition (or a merger) does just the opposite. In many cases, one and one add up to less than two. Hence, the prudent guidance is an essential element of merger and acquisition.

Business consulting companies (specialized in merger and acquisition) often provide strategic guidance in corporate amalgamation process. The strategic guidance provides help for the legislative part of merger and acquisition, financial synergies, and human resource-related issues. The legal departments of these business consulting companies are knowledge-based organizations, and knowledge management (KM) becomes crucial to their continuing success. A KM related automated decision support system allows consulting professionals to work more efficiently and facilitate consulting services quicker than

ever. By making operational processes to support and identify, capture, and disseminate a consulting firm's knowledge, KM systems leverage the collective organizational wisdom.

The business merger and acquisition consulting industry has faced significant pressure in recent years, making KM a business imperative. In the age of faster digital communication, legal consultants have been forced to find quicker ways to deliver traditional merger and acquisition legal services. Consulting company clients have become very sophisticated buyers of legal and related services, and therefore, they expect a faster turnaround time. Innovative law firms however are already working to find a more efficient way to work, leveraging the knowledge of their experts by delegating work to more junior staff and hence looking for better ways to improve knowledge sharing and exploitation.

The work of professionals within a business merger and acquisition consulting firm, or a legal department in a business, ultimately leads to consultation and review of legal documents (e.g., statutes, previously decided law cases, contractual agreements). In this sense, consultants work processes are document centric. Hence, the KM solutions targeting legal professionals focus on document management issues. Hence, from knowledge building and sharing, what is relevant is the intellectual process carried on delivering the outcome. These intellectual processes are often use rule-based reasoning (RBR) and case-based reasoning (CBR) mechanisms to deliver the ultimate outcome.

Bespoke KM systems represent a category of Intelligent Decision Support System (IDSS) for teaching and corporate learning purpose. Often IDSS provides some forms of intelligent tutoring systems (ITSs) functionalities with adequate domain knowledge and its one of main objectives is to share that knowledge to the users (i.e. students) for learning purpose using appropriate interaction sessions, trying to emulate the way a human tutor guides the student in his/her learning path. Designing and implementing an IDSS is a straightforward task since the required technology often implies most of the areas of Artificial Intelligence (AI): knowledge representation, diagnosis, cognitive modelling, qualitative processing, and causal modelling process. Besides, it is necessary that the designers need to have a good knowledge of the domain or topic selected to be taught. The IDSS intelligence is represented by the analysis process and the tutoring process adaptation, according to the student profile. In this sense, a challenging research goal is the development of IDSS with adaptive characteristics that ushers effective KM practice.

Adaptive IDSSs can be obtained at several levels: (a) at the level in which the material or the help is presented, (b) considering the difficulty of the problems proposed, or (c) during the selection of the suitable instructional strategy according to its capacities, abilities and learning styles preferred. In response to this challenge, this chapter describes argument structures to generate plausible explanations for the conclusions reached by rule-based reasoning (RBR) and provides a means of integrating with case-based reasoning (CBR). In this way, it proposes a CBR approach to design Intelligent Tutoring Systems able to personalize the teaching process in different domains. This approach has three important advantages: (i) it provides a learning method, which uses knowledge-acquired from past experiences, (ii) it allows the retrieval of similar student models from multi-organizational distributed datasets and the adaptation of teaching strategies according to the student characteristics and (iii) it preserves all the major pedagogical features associated with cognitive tutoring systems, a highly effective subtype of IDSS.

The reusable problem-solving method permits scalability, ease acquisition and maintenance of knowledge. The system, Knowledge Guidance for Business acquisition Process (GBMA) presents a hybrid knowledge-based system architecture to create two interlacing components: (i) RBR and (ii) CBR. The area of application covers the legislative part of the merger and acquisition-related issues, for which a hybrid RBR-CBR knowledge-based system was built for the business community.

The most valuable resources of business communities (e.g., consulting companies) are the knowledge and expertise of consultants; but the knowledge of senior consultants is not easily available to less experienced consultants, interns, and clients. The technology of knowledge-based systems (KBS) and IDSSs are relevant for this problem because in principle it should allow scarce expertise and knowledge to be more widely available and easily accessible. For junior business consultants and interns to have access to such information at the touch of a few keys could improve productivity, quality, and performance. Also, the use of these technologies could give merger and business acquisition consulting firms a competitive advantage over others that would be slower to adopt the AI-based IDSSs. Therefore, the construction of automated knowledge-based system (KBS) is a valuable exercise. It is not unusual to find this view among those involved in research and the parts of business merger and acquisition consulting practice where a sympathetic view of research exists. In automating the business acquisition-system one must understand the nature of merger and acquisition legal reasoning processes, which goes beyond the simple rule-based approach of the earliest and best-known legal expert systems. A generation of *knowledge guided IDSS* challenge the research community.

The GBMA software system considers only the common law areas of the United Kingdom (UK) and its former colonies. The common law has the tradition of using both statutory laws and previously decided cases. In common law, decision-making takes place according to specified regulations (i.e. statues). Most of the cases that appear in the legal system will be dealt with by whatever rules appear to arise most straightforwardly from the statutes. Given a case, a judge must interpret that case in the light of the statutes in the selected area of law. Consequently, the judge's justification for a decision is specific to that case. Therefore, it is not surprising that even the simplest rules in the statutes may be open to interpretation (or *open-textured*). In a simplistic sense, open-textured legal concepts are not clearly defined, or not defined at all, or may have unspoken exceptions or prerequisites. Often, they are context-dependent and need proper interpretation for a particular-situation.

Academics and practitioners have taken different approaches to design and build the human legal reasoning process in forms suitable for computational software. The earliest of them have tried to automate the legal knowledge in an IF <*condition*(*s*)> THEN <*conclusion*> format (Sergot et al., 1986a, b). Advising in each situation consists of inferring more information about the case in hand using deductive or inductive reasoning using the rules. This paradigm of legal reasoning is rule-based reasoning (RBR).

It is natural to think that a knowledge-based system that tries to encapsulate merely *rules* of a domain of legal enquire is of limited use as an IDSS tool. Such an automated tool would be one person's theory of the law in an area of application. One might imagine oneself going to some single well-known legal professional in business merger and acquisition law and constructing a knowledge-based system which comprises his or her version of that area of law. A business merger and acquisition lawyer in private practice might appreciate having access to such an encapsulation of finest thinking on merger and acquisition law, but he or she would not necessarily accept the 'answers' provided by such a system. The business acquisition lawyer has the goal of maximizing the client's interest – which inevitably guides an interpretation of the law in his or her own special and interest-dependent way. But the history of litigation in merger and acquisition law – as in most areas of law shows that the interpretation of lawyers is not always accepted by company law judges.

Apart from status, the next most significant influence on the common law is previously decided cases in the form of *case reports*. Thus, when a judge decides on a new problem in a case, similar cases in the past may be referred to as a guide. A given case may be related to several previous cases with conflicting decisions. Consequently, a lot of interpretation is demanded of the decision-maker. Even when one

is deciding upon which previous cases are like a current case, the process is open to personal judgement and there can be significant differences between the choices of different specialists. Initially, a judge will choose those cases from the case reports that deal with the same area of law as the current case. Then a search is made for cases that have similar case histories to produce arguments or decisions for the present case. Though previous decisions have considerable influence on the future judicial handling of comparable situations, legal authorities that have made decisions in the past retain full freedom to decide differently in similar cases in the future; but, in fact, these reversals hardly ever take place. The precedents if they do not bind the judges, inevitably inspired them (vonMehren & Gordley, 1977). In other words, the precedents are followed in a diluted form.

The reported decisions in previous cases can also be used to argue in a new case by analogy (Bain, 1986) (Rissland & Ashley, 1989) (Pal & Campbell, 1995) (Pal & Campbell, 1996) (Pal & Campbell, 1997) (Pal & Campbell, 1998); the legal reasoning mechanism here can be described as case-based reasoning (CBR). RBR and CBR are both ubiquitous in the human problem-solving process, so it is not surprising that these techniques have been used by legal professionals in varying degrees. Rules and cases have complementary strengths. Rules capture broad trends in the domain, while cases are good at filling in small pockets of exceptions in the rules. As one can see, the benefit of the present research is that it takes on board all these characteristics of legal decision-making in common law and exploits both RBR and CBR to realize results that legal practitioners can use naturally in their work.

The chapter is organized as follows. Section 2 introduces the basic idea of RBR, CBR, and a theory of argumentation used in this chapter. Section 3 gives a short description of the domain of the research and the knowledge representation scheme is provided in the same section. In particular, the model of similarity and the similarity assessment technique of the CBR part of GBMA is presented. How the system predicts the outcome of a case is described by an example in Section 4. Finally, Section 5 compares other related research and Section 6 provides the concluding remarks.

BASIC IDEAS OF RBR AND CBR

This section gives a basic outline of RBR and CBR. It indicates how GBMA's knowledge base makes use of these reasoning methods.

Rule-Based Reasoning

It is often possible to represent declarative knowledge in the form of IF < preconditions> THEN <conclusions> rules. Therefore, a typical rule has two parts: the precondition and the conclusion. The precondition contains information on which facts or situations must be true for the rule to be used. If these, match exactly to facts of the current applications, the rule fires, and the conclusion then presents information on the consequences of the match. For example, a small set of rules is shown in Table 1.

These rules determine how the status 'raider should send a copy of the notice to the company' may be obtained (Ryan, 1997). There is one rule having 'raider may give notice for the rest of the shareholders' as the consequence, and another rule having as its precondition 'raider may give notice for the rest of the shareholders'. This is a simple instance of how one can reason about even limited rule structures.

Rules have enabled knowledge base designer to represent problem-solving knowledge as models that could be implemented computationally. In rule-based systems, knowledge is represented as facts

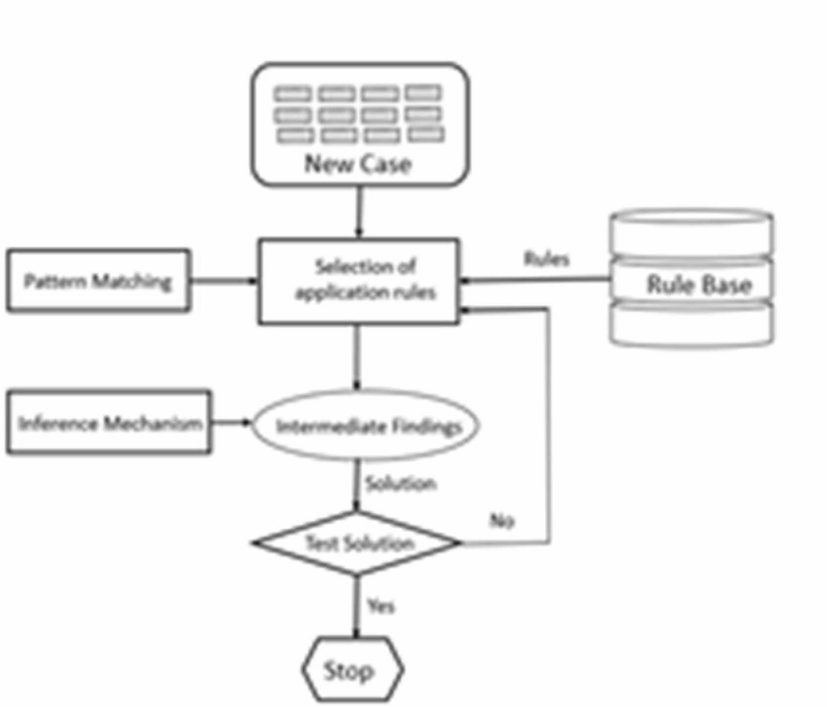
Table 1. A small set of rules relevant to business acquisitions

IF <preconditions> THEN <conclusions>	Symbolic representation
IF raider has contracted ninety per cent of the shares THEN raider may give notice for the rest of the shareholders	An → B
IF raider may give notice for the rest of the shareholders THEN raider should send a copy of the notice to the company	B → C

about the world i.e., relationships between entities, e.g., A, B, C, and mechanisms known as inference engine for manipulating the facts. One might construct a network of rules that interconnect to form a repository of knowledge, known as a rule base. The reasoning architecture of rule-based systems has such a rule base plus an inference engine that performs inferences e.g., forward, or backward chaining, or a combination of these two.

The general reasoning process of a rule-based system is shown in Figure 1. Given a new case, applicable rules are first found by matching against the rules of the rule base. Then intermediate results are generated by the chosen inference mechanism, and the process is repeated until the desired solution state is reached.

Figure 1. Rule-based reasoning process



In the present business merger and acquisition application, four different categories of rules are used. The syntax and purpose of these categories are discussed in the latter part of this chapter. Next, this chapter is going to explain how to develop knowledge-based systems that use previously solved cases to perform some of the CBR task (Benbasat, DeSanctis & Nault, 1993) (Blanning & King, 1989).

Case-Based Reasoning

The main idea of CBR is to adapt solutions that were used to solve old problems and use them for solving new problems or cases. In other words, people reuse all their past problem-solving experience to deal with a new case (Kolodner, 1993) (Pal & Campbell, 1996) (Pal & Campbell, 1997). A CBR system consists of a case base which is the set of all cases that are known to the system and an inferencing mechanism to derive a solution from the stored cases. Moreover, CBR follows a different process than RBR and when a new case is presented to the system.

CBR is an approach to problem-solving that emphasizes the role of prior experience (i.e., new problems are solved by reusing and, if necessary, adapting the solutions to similar problems that were solved in the past). Solving a problem by CBR involves obtaining a problem description, measuring the similarity of the current problem with previous problems stored in a case base (or memory) with their known solutions, retrieving one or more similar cases and attempting to reuse the solution of one of the retrieved cases, possibly after adapting it to account for differences in problem descriptions. The solution proposed by the system is then evaluated (e.g., by being applied to the initial problem or assessed by a domain expert). Following the revision of the proposed solution, the problem description and its new solution can then be retained as a new case. Thus - the system has learned how to solve a new problem. Figure 2 shows the CBR cycle.

This cycle works as follows: (i) Retrieve previously experienced cases related to the current problem. (ii) Reuse these cases in one way or another. (iii) Revise the solution based on re-using previous cases. (iv) Retain the new solution (as a new case) by adding it into the existing case-based database. Then, a CBR system will gradually grow larger and become a valuable resource.

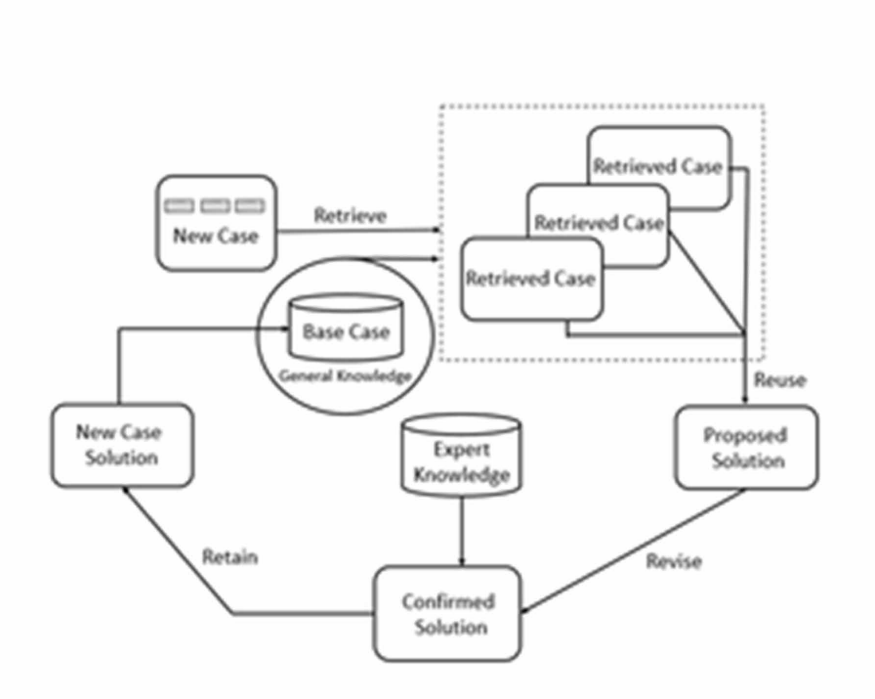
More formally, one can say that a previously solved case consists of a pair (H, D) where H stands for a case history and D is a decision for H. A case-base is a set of such cases. Considering the discretionary nature of business acquisition-related reasoning, one can assume that (H, D_1) and (H, D_2) does not imply $D_1 = D_2$. GBMA's CBR part operates as follows:

1. Receive a new case description H_n.
2. Select the case or cases $(H_1, D_1) \ldots (H_n, D_n)$ from the case-base such that $H_1 \ldots H_n$ is 'similar'.
3. Use the past decisions $D_1 \ldots D_n$ for $H_1 \ldots H_n$ to form the final decision D_n for H_n.

In the business decision-making process, specialists refer often to previously solved cases. Although much of the business acquisition law has now been superseded by statute, even where there is a statute there are probably also cases interpreting the legislation. Business acquisition practitioners' reason with such cases (known as *precedent*) when they exist. A precedent is a decision of a court, considered as providing an example or authority for an ideal or similar case arising later from a similar question of business acquisition law.

To assist this process, an automated business acquisition CBR system should find the case most like the new case and each of them may, therefore, act a potential candidate for the new case. But this

Figure 2. Case-based reasoning process



choice, which is cognitively a complex activity, is usually approximated by some simple metric scheme in case-based computing. Such a metric scheme for the present research work is described in the latter part of this chapter.

Reasoning from precedents to interpret a new case (much as business acquisition practitioners do) or creating an equitable solution to a new problem (much as business mediation do) is business acquisition CBR. The use of CBR has been considered in the past to enhance ITSs with learning abilities. This activity raises a variety of research issues, which researchers are addressing.

CBR and Intelligent Decision Support Systems

Intelligent decision support systems started to be developed in the 1980s, they were designed with the idea of providing knowledge based on some form of intelligence to guide the end-user in the process of learning (Gonzalez, et. al., 2006). An intelligent decision support system that uses Artificial Intelligence techniques to represent the knowledge and interacts with the students to teach them (VanLehn, 1988). In the (Giraffa, et. al., 1997) authors add the consideration of different cognitive styles of the students who use the system according to (Cern, 2002). In the 90s, with the advances of cognitive psychology and the new programming paradigms, ITSs have evolved from a mere instructional proposal to the design of environments of new knowledge discovery application (Sierra et. al., 2006).

In (Han, Lee, & Jo, 2005) the authors propose the use of CBR as a technology for student modelling in it. That approach follows the steps of the CBR cycle, and it can build concrete student models by combining rule-based reasoning. However, such approximation only supports the retrieval and reusing

phases of the cycle. Other approaches recommend the use of CBR for instructional and route planning (McGinty, & Smyth, 2001). In (Kharrat, Reyhani, & Badie, 2003) an Intelligent Tutoring System based on the CBR methodology was developed. This system can produce novel courseware arrangements for new students, based on a process of case adaptation. Elorriaga and colleague (Elorriaga, & Fernandez-Castro, 2000) propose an approach for producing case-based instructional planners that are integrated into ITS to enhance the pedagogical model. The works mentioned above only use CBR as a technology for building isolated ITS modules, but they do not consider CBR as a methodology that integrates all the components of the ITS architecture, which challenges the new frontier of intelligent knowledge-based systems.

These Systems are computer-based software tools that help its use for decision-making (or learning how to decide) by presenting information and interpretations for various alternatives. Such systems can assist managers in making a strategic decision (Benbasat, DeSanctis & Nault, 1993) (Turban, 1990). The ability of such systems in processing knowledge has led to cost savings, faster decision process, good payoff, and significant competitive advantage (Blanning & King, 1989) (Bonarini & Maniezzo, 1991) (Finlay, 1992) (Guida, Marchesi & Basaglia, 1992) (King, 1990) (Meyer, et al., 1992) (Schutzer, 1990). There are three important approaches in the development of current business ITSs:

1. Rule-based reasoning RBR (James, 1997) (Kim & Lee, 1995) (Lee & Wu, 1995) (Michaelsen, 1984) (Steinbart, 1984);
2. Case-based reasoning CBR (Ashley, 1987) (Curet & Jackson, 1996) (Huang & Cross, 1989) (Rissland & Ashley, 1989); and
3. Hybrid i.e., a combination of RBR and CBR (Branting, 1991) (Golding & Rosenbloom, 1991) (McIvor & Mulvenna & Humphreys, 1997) (Rissland & Skalak, 1991) or an integration of other reasoning methods (Quah, et al., 1996) (Shin & Han, 1999). Each focus on enriching some aspects of the traditional knowledge-based business ITSs. In RBR systems, the specialized domain knowledge is represented as a set of IF < preconditions> THEN <conclusions> rule format.

It is often extremely difficult to obtain an appropriate set of rules in an application domain to cover all possible eventualities. Moreover, in certain circumstances where the situation is not so clear, decision-makers have often relied on previously solved cases that are like the new case in hand. This reasoning process, where the past experiences are used in solving a new case, is known as CBR. The previously resolved cases can bridge at least some of the problems encountered in an RBR system by representing exceptions to the rule in the form of cases. This is because the facts of the previously resolved cases, like those of new cases, are expressed in the case-description language.

What is required, then, is a knowledge-based intelligent tutoring system (ITS) that respects relevant information expressed in rules but that also comes closer to the task that confronts the business professional: the analysis of the present business situation in terms of previous experiences. This requires the use of both RBR and CBR in an integrated environment (Pal & Campbell, 1995). There are numerous w x domains in which it is important to combine RBR and CBR, for example, the legal (Branting, 1991) (Rissland & Skalak, 1991) and financial domains (Dutta & Bonissone, 1993). This chapter is concerned with an implemented prototype system, Decision-support System for the Business Acquisition Process GBMA, which also integrates RBR and CBR.

Integration of RBR and CBR

In general, there are three different ways to combine RBR and CBR: [i] keep RBR and CBR as two. [ii] let the CBR module use the inference capability of rules when needed; and [iii] have the RBR module dominate and make use of CBR categorically at points in the reasoning process. The research of Rissland and Skalah (Rissland & Skalak, 1991) is an example of the first approach. They have built a system called CAse-BAsed REasoning Tool, CABARET, as a hybrid system, which integrates RBR and CBR. CABARET's application domain is that of income tax law concerning the deduction for expenses relating to an office maintained in one's home. It deals with the circumstances under which a taxpayer may legitimately deduct, on a United States income tax return, expenses relating to an office return, expenses relating to an office maintained at the taxpayer's expense. The integration of RBR and CBR methods is performed by using control heuristics. These control heuristics suggest how to interleave RBR and CBR to produce an argument to support a certain interpretation. CABARET interleaves CBR and RBR dynamically.

The second paradigm is exemplified by the work of Bonissone et al. in the development of the CARS system (Blau, Bonissone & Ayub, 1991) (Bonissone & Blau, 1990). The CBR module in CARS is the dominant system and it activates PRIMO (Bonissones, Aragones & Stillman, 1990), a rule-based reasoner that contain plausible rules of abstraction, evaluation, and modification. The rules in the PRIMO system are used in case indexing to augment the case. In the case of retrieval, the rules combine the similarity measures computed across the abstract features of cases. The research presented in MARS (Dutta & Bonissone, 1993) is an example of the third approach to combining RBR and CBR. In MARS, rules are used to represent the domain expertise that is required for structuring various parts of company mergers and acquisition deals or deciding upon the best course of action to determine which company to select as a possible target. CBR is activated by selective rules that state the need for integration. The RBR module dominates but activates CBR explicitly at specific points in the reasoning process. The present research i.e., GBMA is also an example of the third approach to integrating RBR and CBR.

In GBMA, the business acquisition domain knowledge is represented mainly in different types of rule form. The system covers three aspects of a business acquisition process, namely, company formulation, strategy formulation, and strategy modification. The CBR module is explicitly used at the time of strategy modification only. The MARS project was developed for the domain of mergers and acquisitions. However, GBMA deals with British business acquisitions only. In MARS, the individual cases are represented using rule templates. Hence, the integration of CBR and RBR was achieved without altering the inference engine of RBR. The GBMA system research uses an object-oriented knowledge representation technique instead of abstracting the case description into rule form. The main differences between GBMA and MARS are: [i] GBMA support a partial rule-based advice facility, whereas MARS does not support any such facility and [ii] GBMA provides an argument structure to generate plausible explanations for conclusions reached by RBR, and a means of integration with CBR. But in MARS, there is no argument generation facility and the integration technique of RBR and CBR is different from that of GBMA.

OVERVIEW OF GBMA

GBMA has been developed as a pure research product. The knowledge base consists of a rule base and a case base. The system covers three aspects of a business acquisition process, namely company valuation, strategy formulation, and strategy modification. It provides a text-based user interface and collects facts for a new case in a question-answering session with its user. The user inputs a case description to GBMA and can select any of the appropriate options i.e., company valuation, strategy formulation, and strategy modification from the system menu. In the. case of the first two options, the system provides advice for the user by using rules from the rule base only.

But in the case of the strategy modification option, the system uses both RBR and CBR in an integrated environment. The computational framework for GBMA is shown in Figure 3. The different aspects of the system are discussed in the following subsections.

Company Valuation

An acquisition process starts when the raider who usually initiates an acquisition attempt chooses a target which is the company of interest to the raider for takeover. The next important consideration in any acquisition is the ability of the raider company to place a value on the target company. However, placing a value on a company is a very complicated process and there are no straightforward rules for doing this. It is possible to come up with several different valuations for a single company. Moreover, two raider companies can arrive at with different valuations of the same target company due to the different plans that each has for the target. Each method of valuation has its associated advantages and disadvantages and will be appropriate according to the intentions of the raider towards its target. The different methods of company valuation used for the present project are stock market valuation, net asset valuation, capitalized earning value method, profit earning. E ratio valuation, Gordon growth model valuation, and discounted cash flow valuation. The detailed treatments for these methods can be found in any standard textbook on corporate finance (Brealey & Myers, 1996) (Pike & Neale, 1999).

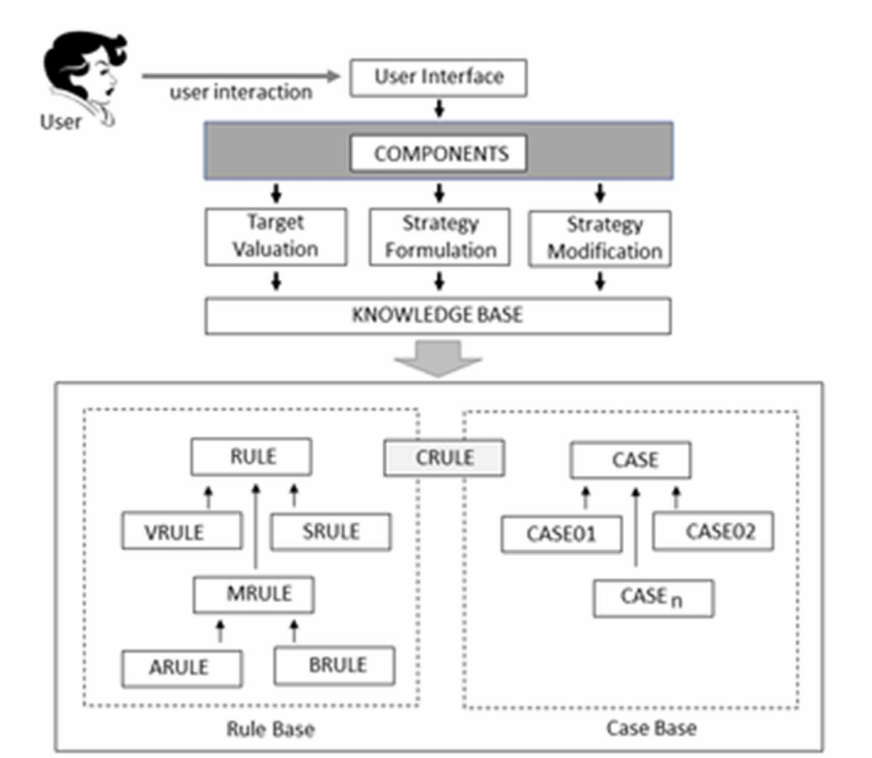
Strategy Formulation

In any acquisition process, strategy and tactics of the raider company play a very important role. Before a raider becomes involved in an acquisition process, it must satisfy itself that acquisition represents a more efficient alternative than organic growth or the independent purchase of required assets.

Once a company has satisfied itself on these points, the strategic process that it should follow towards acquiring a target company can be summarized as follows: Select appropriate target companies. Find as much information about the target companies as possible. Value each of the possible target companies using the collected information.

Determine which of the possible target companies is most suitable. Decide upon the best way to finance the acquisition, considering which methods of payment are agreeable to the shareholders. Once a raider company has gone through this strategic process, it must then decide upon the acquisition tactics it will use. Failure to employ the right tactics can result in a predator paying over the odds or, in the worst-case scenario, failing to acquire its target altogether. Also, companies must be aware of the rules and regulations governing acquisitions.

Figure 3. The computational framework for GBMA



Strategy Modification Process

The target's response acceptance or defiance to the raider's acquisition offer determines how the raider subsequently modifies the chosen strategy. For example, a minority shareholder may refuse to sell his or her shares to the raider. In that situation, the raider may look for an alternative strategy e.g., an approach involving exploitation of legal issues to acquire the minority shares. Keeping this brief introduction and the functionality in mind, we are now going to discuss the knowledge representation scheme for GBMA.

Knowledge Representation in GBMA

Object-oriented knowledge representation method is used in GBMA. The object-oriented knowledge representation approach reflects the natural perceptions of the world as being composed of objects, classifiable into general types. This involves thinking about the world as a set of entities or objects that are related to and communicate with one another. Each real-world entity e.g., anything, idea, or concept is modelled by an object. Each object is associated with a unique identifier that makes the object distinguishable from other objects. Each object has a set of attributes and methods operations. The value of an attribute can be an object or a set of objects. The set of attributes of an object and the set of methods represent the object structure and behaviour, respectively. The attribute values represent the object's state. This state is accessed or modified by sending messages to the method.

A case report, as shown in Figure. 6, may have a name, source, court name, participants, facts, main surface features, and so on. Thus, a specific case report e.g., CASE04 is an instance of the case report class. Individual object instances can be distinguished from other instances by differences in the actual values of the attributes and by associations with other object classes and object instances. Object instances that are members of the same class share a common real-world meaning in addition to their shared attributes and relationships.

Object models can be simplified by defining hierarchies of data structures. Thus, generalization is a kind of hierarchy defines a relationship among classes. Another type of hierarchies is aggregation. Aggregation relationships depict 'part of' hierarchies, as shown in Figure. 6 for a case description. In GBMA, all items of interest in the application domain, such as different aspects of case reports, main surface features, case description, and case outcome rules, and methods to manipulate. they are modelled as objects. Among them are aggregation association and generalization association. For example, the 'rule base' consists of aggregation association of different classes of rules: V-type rules company valuation rules, S-type rules strategy formulation rules, M-type rules strategy modification rules, and C-type rules control rules. The rule base consists of 70 rules related to the different function of the system.

The Organization of Rule Base

In the present business application, four main types of rules are used: valuation rule, strategy formulation rule, control rule, strategy modification rule. The strategy modification rule set consists of two further sub-categories of rules: available actions rule and prediction rule. The descriptions and examples of the valuation rule, strategy formulation rule, and control rule are shown in Table 2.

Strategy Modification Rules

The strategy modification rules are used to modify the initial strategy whenever a problem arises in the acquisition process.

Available-Actions Rules

The available actions type of rule involves a straightforward transformation of business statute law to rule form. An example of this type of rule is shown in Table 3. The label of the rule in Table 3 begins with 'A', to denote the type. An 'A' rule determines whether a court has the power to act or take a specific action. For example, an English court has a range of options available in a minority shareholder protection order case as itemized below: – raider shall not be entitled to acquire the minority shares; – alteration of the terms of the acquisition. The available-action s rules determine which of the option is applicable in a case. For rules to be applied for a new problem, and, hence, for a rule-based system to be of any use, the system will need to have access to facts of that situation.

Prediction Rules

In the rule-based analysis, the valuation rules, the strategic rules, the control rules, the plan modification rules, and the available-action s rules are used as in a conventional expert system in one of two modes: response or no response. A response is produced when all the preconditions of a rule are matched by

the facts of the new problem in hand. However, when this situation does not hold, such rules are not considered in determining any conclusion.

Prediction Action Rules

A prediction rule predicts actions a decision-maker is likely to take. When all the preconditions of a prediction rule are matched by the facts of a problem, the rule can give unconditional advice. The predictive rule-based part can produce some tentative or partial advice, which may be helpful or informative for the user. The prediction rules can generate any of three types of output: clear-prediction, speculation, or no-prediction. Clear prediction is possible when at least one of the prediction rules has fired because of the facts of the new problem. The system presents what one can call speculation by applying a weighting criterion to the true rule preconditions, even when none of the rules is fired.

Table 2. Description of different types of rule

Different categories of rule		Example
Valuation Rule: The valuation rules are used for the target company valuation purpose. Each of these rules has a unique name (e.g., VRULE01, VRULE02, etc.) beginning with 'V'. **Strategy Formulation Rule:** The strategy formulation rules are used to formulate a plan for the proposed business acquisition. The label of strategic rule starts with 'S' (e.g., SRULE01, SRULE02, etc.). **Control Rule:** The control rules are used for rule-execution and other knowledge-manipulation purposes. Each of these rules has a unique name (e.g., CRULE01, CRULE02, etc.).		**VRULE01** IF fixed assets? fa net current assets? nca long term debt? ltd net asset value (nav) = fa + nca – ldt THEN The net asset value of the company is nav. **SRULE01** IF (business type is a restaurant) (target restaurant is well known for its food quality) (the restaurant has been taken over by a raider) THEN In a strategic restructuring process, it is recommended not to replace the chef of the target restaurant. **CRULE01** IF Selection? type = 'company valuation' THEN Choose all company valuation rules.

Speculation consists of conclusions that would have followed if all the preconditions of a rule that has some relevance had been true, and output focusing on the failed preconditions i.e., reasons why a conclusion cannot be accepted without reservations. The first step in generating speculation is to identify the rules that are nearly fired. A scoring mechanism is used to determine which rules are closest to firing. For this scoring mechanism, the preconditions of the prediction rules can be divided into three subclasses: peripheral, significant, and essential. The justification for the scheme exemplified by equation 1 is as follows. In examining the previously decided case reports and the acquisition-related legal text sources (Pike & Neale, 1999) (Rabinowitz, et al., 1997), it was observed that some of the preconditions

of the rule base were of secondary importance in drawing a conclusion from a rule and some were of little significance.

Hence, the preconditions here fall into the above categories. Peripheral preconditions are of secondary importance in concluding and are helpful in practice to provide information about the context. Essential preconditions are those that are critical in concluding. Significant preconditions are those that fall between essential and peripheral, in that though they are important when concluding, they are not critical on their own. Each of these rules has a unique name e.g., BRULE01, BRULE02, etc. beginning with 'B'. At the time of specifying these rules, proper interpretation of the different business norms (Pike & Neale, 1999) (Rabinowitz, Weinberg & Bank, 1997) and experts' specific domain knowledge were taken into consideration. For example, the structure of a prediction rule is shown in Table 4.

Table 3. An available-action(s) rule for a minority-shareholder-protection order

ARULE05 IF (application is one of the shareholders) (respondent is the raider) (offer notice has been given to the applicant) (applicant is not ready to sell the share to the raider) THEN The court may make one of the following orders: [1] an order that the raider shall not be entitled and bound to acquire the shares; or [2] an order specifying terms of acquisition different from those of the offer.

It is found by experiment that there is a consistent threshold i.e., 0.40 in our score, below which any information that GBMA may give is unhelpful. A value of 0.40 or above in a scoring range between 0 and 1 indicates that the rule has some significant ability to contribute advice that the user is likely to find helpful. The threshold value was determined by actual checks of the relevance of retrieved material at a late stage of the knowledge acquisition process.

The system, therefore, offers no information unless at least one of its prediction rules has a score above the threshold. If there is no such score, one can say that the output is of a no-prediction type. The score Score of a predictive rule Ri can be defined as follows:

$$Score_{Ri} = \frac{Score_u}{Score_l} \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

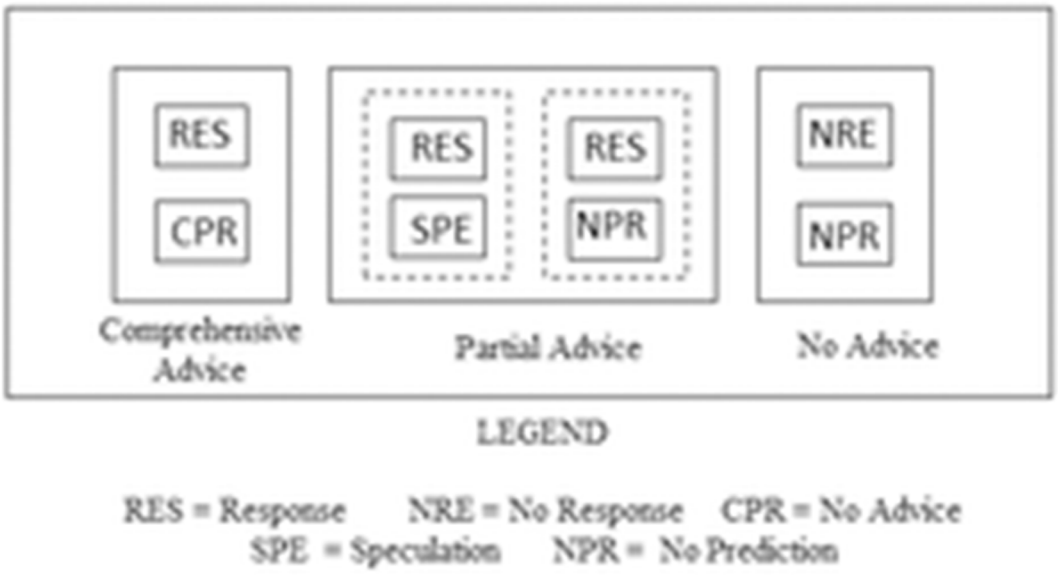
where $Score_u = w_1 N_e + w_2 Ns + w_3 N_p$; and $Score_l$ is the total number of preconditions of the rule in the equation; N_e, N_s, N_p are the numbers of essential, significant, and peripheral preconditions that are true for the current case. The weighting factors w_1, w_2, and w_3 are for essential, significant, and peripheral 3 categories of preconditions. It has been found that the most convincing behaviour of GBMA occurs when $w_1 = 0.75$, $w_2 = 0.62$, and $w_{3=}$ 0.25. Moreover, these values were also agreed upon by the domain expert based on the retrieved information from the prototype system. The prediction rules explain how a decision-maker is likely to act within the range of option available, which is circumscribed by the

Table 4. A 'prediction' rule for a minority shareholders protection order

BRULE02 IF *peripheral*: (right of minority shareholders) (business type is a restaurant) (applicant is one of the shareholders) (respondent is the raider) *significant*: (offer related to purchasing the share) (offer relates to a particular class of shares) *essential*: (respondent contracted ninety per cent of the share) (four months has passed after the date of the offer) (acquisition offer has not been accepted) THEN If the case proceeds to court, an order authorizing to the shareholder to give notice against the offer.

available-actions rules. One can say, for distinction, that the available-action s rules give available actions concerning the available options and the predictions rules provide a prediction about what a decision-maker may conclude for a situation. When a user asks for rule-based advice, GBMA can provide one of three possible options: comprehensive advice, partial advice, and no ad comprehensive advice, the system provides the possible available action s plus a predictive decision, provided that at least one of the prediction rules has fired.

Figure 4. Different types of behaviour leading to rule-based output



The partial rule-based advice can be in one of two categories. For category one of partial rule-based advice, GBMA offers the relevant available actions and presents speculation. Category two of partial rule-based advice produces no prediction or speculation but does suggest some valid available actions. Finally, the system provides no rule-based advice at all when it fails to come up with available action s or any kind of predictive information. These different types of behaviour leading to rule-based output are shown diagrammatically in Figure. 4. Additionally, the strategic modification module has a facility to generate an argument for the advice.

Argument Mechanism in GBMA

In general, an argument consists of preconditions and conclusion of an RBR when at least one of the rules has fired. There is a vast literature on theories of argumentation, but we have used a simple theory proposed by Toulmin (Toulmin, 1958). Toulmin [Ref. (Toulmin, 1958)], p.126 presents an interesting and commonsensical example of defensible reasoning: (i) Anne is one of Jack's sisters; (ii) All Jack's sisters have previously been. observed to have red hair; (iii) So, presumably, Anne now has red hair, unless Anne has dyed her hair, gone white, lost her hair, etc. Toulmin has his diagrammatic approach to representing arguments. In its simplest form, Toulmin's model states that an argument can be thought of as a statement that a given set of data lead to a claimed result the claim as shown in Figure. 5.

In Toulmin's theory, all arguments consist of four basic components: claim, data, warrant and backing. The assertion of an argument stands as the claim of the argument. Knowing the data and the claim does not necessarily convince one that the claim follows from the data. A mechanism is required to justify the claim given the data. This justification is known as the warrant. Useful warrants are based on analogies between cases, or even on the authority of a given speaker.

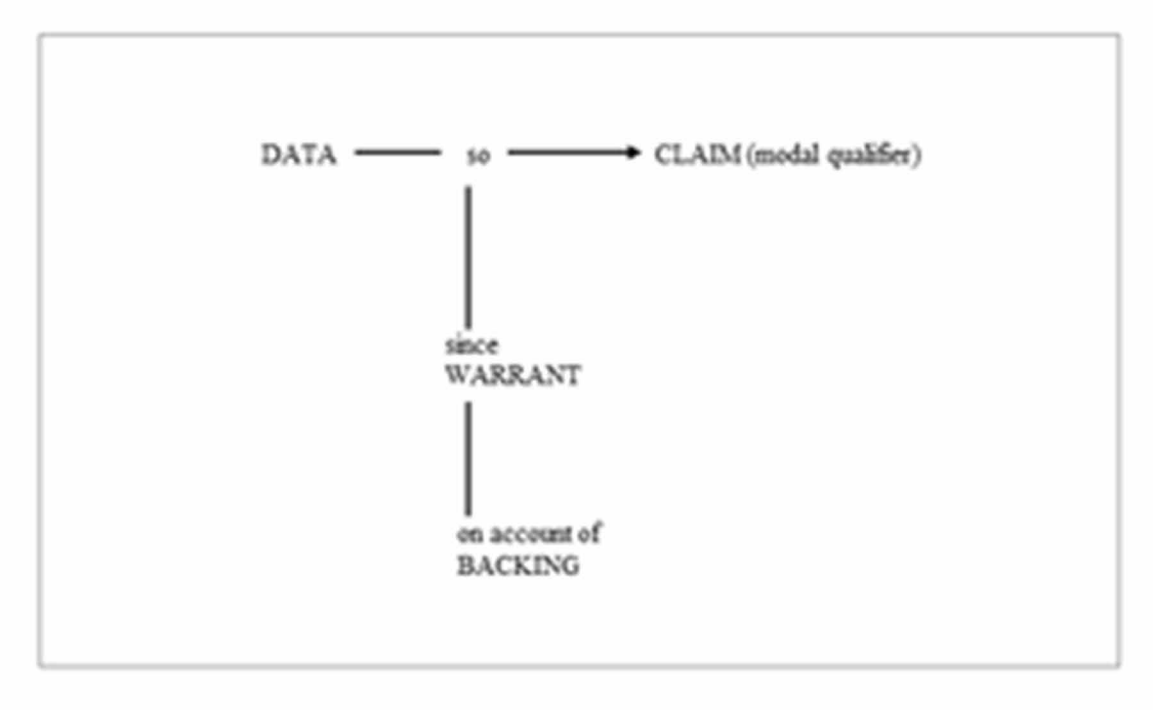
Moreover, a warrant is a true reflection of the expertise used by a subject in solving a problem. The backing of an argument supports the validity of the warrant. GBMA's argument mechanism uses legislative information and previously resolved case s for the backing. As this body of knowledge involves both cases and rules, the system uses both RBR and CBR in an integrated environment to produce an argument. A later part of this chapter describes how this argument structure is used in GBMA, with an example. First, it is important to explain how to develop business knowledge-based systems that use previous experience or previously solved cases to perform CBR tasks.

The Organization of Case Base

The present case base is comprised of 20 manually coded cases, labelled CASE01, CASE02, and so on. This can be expanded incrementally as new cases are collected. The overall organization of case-based design and implementation is based on object-oriented modelling. In this model, individual case is indexed by important attributes (e.g., CASE04 the index is INDEX04 points to its characteristics features *'right of minority shareholders', 'business type is restaurant', 'application is one of the shareholders', 'respondent is the raider', 'offer related to purchase of the share', 'offer related to a particular case of shares', 'respondent contracted 90% of the share', '4 months has passed since the date of the offer'*). The court name attribute is the name of the court – that ruled on the case. In the present circumstance, COURT001 represents the Chancery Division companies court.

The participants attributes contain all. the information associated with the participants and the relationships among them. The fact attribute represents the history of the case briefly. Similarly, the appeal

Figure 5. Toulmin's argument structure



data structure describes who has applied for the present case, how many shares he or she has and the type of the shares e.g., ordinary or preference shares. In the case base side of GBMA, the cases that have the highest similarity rating concerning a current problem are retrieved and used for generating the argument for rule-based advice. The similarity is judged by comparing the main facts and events recorded in the histories of the cases. The similarity measure is based on numerical taxonomy (Sneath & Sokal, 1976).

To illustrate the functionality of GBMA's target valuation, strategy formulation, and its strategy modification facilities, consider the example of the Ravenna Restaurants case, which involves a minority-shareholder-protection order issue. The bare facts of this case are stated as follows:

Example: Ravenna Restaurants Ltd

Ravenna Restaurants Ltd is an Italian restaurant situated in the south-west London. It was established 20 years ago, owned by two brothers, and is a well-known restaurant for its authentic Italian food. One brother is the chef, and the other is the business manager. The chef is Nicholas Ladenis and business manager are Peter Ladenis. In the same locality, Valentino Restaurants is another Italian restaurant, which was set up in 1993. Ravenna Restaurants has distributable earnings of £727,000, a weighted average cost of capital of 14% and a P/E ratio of 18.7. It is in the process of acquisition of Valentino Restaurants whose financial details are as follows:

Valentino Restaurants — Key Financial Information:

Profit before interest and tax PBIT £660,000
Interest paid £72,000 Corporation tax £176,400
Distributable earnings £411,600
Earnings per share EPS 16.7p
P/E ratio of 12.87
The market price of ordinary shares £2.15
Equity beta 1.17

Gross Profit Forecast					Dividends	
Year	Sales	Increase			Year	Dividend
1995	205,000	-			1998	15p
1996	250,000	21.96%			1997	14p
1997	310,000	24.00%			1996	12p
1998	411,600	32.72%			1995	10p

Valentino Restaurants Plc is optimistic that it will be able to maintain an annual increase in distributable earnings of 5 per cent per annum due to anticipated synergy because of the takeover. The company will also be able to sell duplicated assets which will release £60,0000 in one year. The risk-free rate of return is 9 per cent and the return on the market is 15 per cent.

On 31 January 2019, Nicholas and Peter wrote to the petitioner, a shareholder in Valentino Restaurants, repeating an invitation that had been made before to all shareholders, that the petitioner offers his shares to them for purchase and stating that the invitation would remain open until 28 February 2019. The date within which shareholders could take up the invitation was subsequently extended to 10 March.

Valentino Restaurants plc – Balance Sheet 1998		
	£000	£000
Fixed assets		265
Current assets	60	
Current liabilities	43	
Net current assets		17
		282
Long term debt		72
		210
Financed by:		
Ordinary shares(50p)		123
reserves		87
		210

The petitioner was served with a notice under Section 429 of the Companies Act 1985 to buy out his shares at 40 pence. The Section 429 notice sent to the petitioner was not signed by the Ladenis brothers but by their solicitor and the statutory declaration under Section 429 (4) was not made until two weeks after the first notice had been given.

In a question-answering session, GBMA gathers the facts of this case. The facts are as follows:

'business type is the restaurant', 'applicant is one of the shareholders', 'the respondent is the raider', 'right of minority shareholders', 'offer relates to purchasing of the shares', 'offer relates to a particular class of shares', 'respondent contracted ninety percent of the share', 'offer notice has been given to the applicant', 'four months has passed since the date of the offer'.

When a user selects the company valuation mode of GBMA for this case, the system provides the target valuation analysis as follows:

The valuation of Valentino Restaurants that has been obtained:
Stock market valuations £529,000
Net asset valuation using book value £210,000
Capitalised earning values £528,000
P/E ratio valuations£528,000 or £682,000 or £769,000
Gorden growth model valuations £771,000
Discounted cash flow valuations£861,000

The accuracy of different valuations will depend on the reliability of the information used. Which valuation method is most appropriate will depend upon the information available to a user of the system? The above output is designed to help the user to arrive at a reasonable decision. Using the same case, we now describe how GBMA formulates the initial strategy by the following. The strategy for the present case includes the following.

1. [1] This restaurant business has demonstrated very good growth and exhibits that it is under excellent management. Moreover, it is well established that small restaurants have rather traditionally considered their approximate values to be slightly under 1 times gross income. For this restaurant, the approximate initial bidding price would be: 0.751 x £441,600 = £331,200.
2. [2] There are two options to finance the present acquisition:
    a. Firstly, share-for-share offers, and
    b. Secondly, case offers.

In the present case, ninety percent of the target company's share are held, so the bidder must make a cash offer to all remaining shareholders at a price no less than the highest price paid in the preceding 12-month period.

3. [3] The target restaurant is well known for its food quality. If there is any organizational restructuring of the target, then it is recommended not to replace the chef of the target restaurant in the initial phase.

To show how the strategy modification works, let us consider that a minority shareholder is not ready to sell his share to the raider. The shareholder has applied for an order under Section 430c that he should not be under a duty to sell his shares to Ladenis brothers.

GBMA can provide certain relevant information to its user here. This information consists of legislative information and the appropriate precedent. Moreover, according to the facts, GBMA presents a comprehensive predictive rule-based advice, which consists of the output of the rules ARULE05 and BRULE02. The respective outputs are shown by the following.

The present situation has triggered the available action s rule ARULE05. Its preconditions and possible actions are as follows:

Preconditions of ARULE05:

applicant is one of the shareholders

respondent is the raider

offer notice has been given to the applicant

applicant is not ready to sell the share to the raider

date of application is within six weeks of the notice period.

Available action(S):

The court may make one of the following orders:

1. [1] an order requiring that the raider shall not be entitled and bound to acquire the shares; or
2. [2] an order specifying terms of acquisition different from those of the offer.

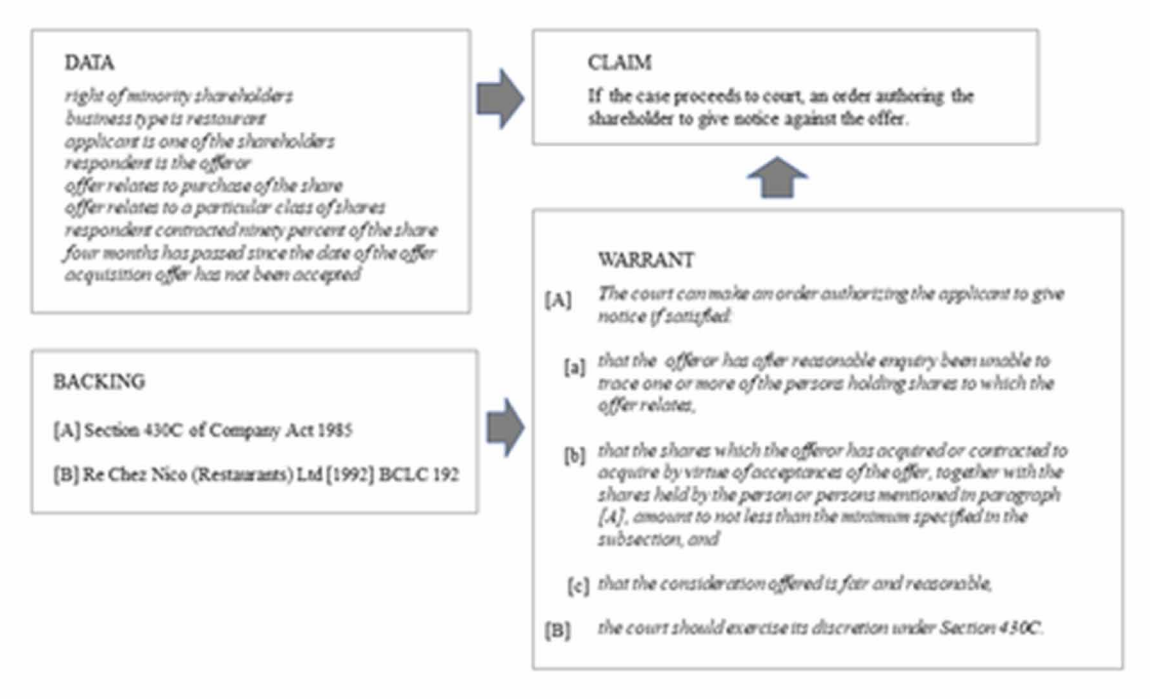
All the preconditions of BRULE02 are matched because of the facts of the Ravenna (Restaurants) case. Due to this, GBMA provides a clear prediction that is a possible outcome of this case. The preconditions, prediction and argumentation or justification for the prediction of BRULE02 are shown in Figure. 6.

GBMA can provide further information and justification for the above argumentation scheme. On demand, the system can provide the text of the section or subsection of the legislation. e.g., Section 429 of the Companies Act 1985. Furthermore, GBMA can justify its previously decided case selection process and show how a case amplifies a predicted rule-based outcome. For example, the selection of a case Re Chez Nico Restaurants is justified to its user by displaying both the common and non-shared surface features of the new case with respect to that case. It also shows the legal implications of Re Chez Nico Restaurants in the light of the new case. It is now helpful to discuss how a case has been selected in GBMA for supporting the system argument scheme.

Case Selection Process

The represented cases are given unique identifications (e.g., CASE01, CASE02) for use in the case base. All the cases are indexed. The INDEX is id used for similarity assessment between cases. From the case

Figure 6. Argument structure of GBMA



base, two cases have been selected to show the similarity assessment between them. To illustrate by an example, we consider the legal problem of protection of minority shareholders in acquisition-related cases.

The represented cases are given unique identifications for use in the case base, a part of which is shown in Table 5. Let INDEX01, and INDEX02 be the two indexes for CASE01 and CASE02, respectively.

The indexes attribute of CASE01 refers to the main surface features. For example, in Re BUGLE PRESS i.e., CASE01, the indexes INDEX01 refers to the surface features 'right of minority shareholders', 'business type is publishing and selling', 'applicant is one of the shareholders', 'respondent is the raider', 'offer relates to purchase of the share', 'offer relates to a particular class of shares', 'respondent contracted 90% of the share'.

Table 7. Cases relating to a minority shareholders protection order

Case no	Source
CASE01	Re BUGLE PRESS [1960] 1 All ER 768
CASE02	Re PRESS CAPS [1949] 1 All ER 1013
CASE03	Re HOARES [1933] All ER 105
CASE04	Re Chez Nico Restaurants [1992] BCLC 192
CASE05	Re GRIERSON, OLDHAM and ADAMS [1967] 1 All ER 192

Similarly, in Re PRESS CAPS i.e., CASE02 the indexes INDEX02 refers to the surface features 'right of minority shareholders', 'business type is engineering company', 'applicant is one of the shareholders', 'respondent is the raider', 'offer relates to purchase of the share', 'offer relates to a particular class of shares', 'respondent contracted 97% of the share'.

The simplest of all association measures is CASE01 ∩ CASE02, which produces five shared main surface features. Considering the number of main surface features, the total number of matched features for CASE01 and CASE02, the similarity coefficient of CASE01 and CASE02 is calculated as 6/8; and the similarity coefficient of CASE02 and CASE01 is calculated as 6/8. Therefore, the mutual similarity coefficient is 0.75.

Using the above similarity assessment method, GBMA can calculate the similarity between the new case in hand and the stored cases in the case base. The cases that have the highest similarity rating with respect to a new case are retrieved and used in the GBMA argument scheme. Only, the cases with scores above 0.5 are selected for this purpose. We have determined this threshold value by experiment. A value of 0.5 or above indicates that the case has some significant ability to provide a contribution towards the generated argument.

CONCLUSION

In this chapter, a KM system is presented which uses both RBR and CBR for solving a business merger and acquisition problem. It has been shown how a complex business situation can be represented using an object-oriented scheme. The performance of the system as judged by criteria such as completeness, relevance, etc. applied by business specialist users is generally good. However, the implemented prototype is functional. The production of partial rule-based advice and the argument generation facilities reflect the intelligent ability of the implemented system to use the rule and case knowledge in ways that correspond to how humans use it.

The company valuation methods, in the project, are very much theoretical in nature and therefore require further work. This future work should be conducted in three areas. Firstly, consultation with a real-world company valuation expert to build an appropriate model is needed which should produce many more rules and even cases. Secondly, there is a huge amount of data on company acquisition and valuation that are accessible for research. It is intended to analyze these data to find out a pattern if there are any to form a model and possible application of CBR method to do this valuation. Thirdly, the other planning to do research work on company valuation based on machine learning e.g., Neural network to determine whether it can provide any better valuation model for the present system. Lastly, the presented software architecture has used different threshold values for case and rule selection process. These values are very appropriate for the present knowledge base. But the number of cases and rules in the knowledge base will increase in the future and, hence, these threshold values may change accordingly to reflect the retrieved information.

The other important things need to be considered in the future, is the student module in the presented model. On the one hand, the student model describes the knowledge that the student has acquired in the domain to be learnt. Different types of techniques can be used: vectors, semantic networks, Bayesian networks, affirmation repositories, etc. On the other hand, the diagnostic process is in charge of updating the student model based on the current student model and the student performance during the learning

process, according to diverse variables previously defined (problem evaluation, answers to questions, time spent in studying each explanation, and so on).

REFERENCES

Ashley, K. (1987). *Modelling Legal Argument: Reasoning with Cases and Hypothetical* (PhD thesis). Department of Computer and Information Science, University of Massachusetts, Amherst, MA.

Benbasat, I., DeSanctis, G., & Nault, R. (1993). Empirical research in managerial support systems: a review and assessment. In Ž. C. W. Holsapple & B. Whinston (Eds.), *Recent Developments in Decision Support Systems* (pp. 383–437). Springer-Verlag. doi:10.1007/978-3-662-02952-7_16

Blanning, R. W., & King, D. R. (1989). Decision support and knowledge-based systems. *Journal of Management Information Systems*, *6*(3), 3–6. doi:10.1080/07421222.1989.11517860

Blau, L., Bonissone, P., & Ayub, S. (1991). Planning with dynamic cases. In *Proceedings of Case-Based Reasoning Workshop*. Morgan Kaufmann.

Blau, L., Bonissone, P. P., & Ayub, S. (1991). Planning with dynamic cases. In *Proceedings of Case-Based Reasoning Workshop*. Morgan Kaufmann.

Bonarini, A. & Maniezzo, V. (1991). Integrating expert systems and decision-support systems: principles and practice. *Knowledge-Based Systems*, 172–176.

Bonissone, P. P., Aragones, J., & Stillman, J. (1990). PRIMO: a tool for reasoning with incomplete and uncertain information. *Proceedings of the Third International Conference on Information Processing and Management of Uncertainty in Knowledge-Based System, IPMU 90*, 325–327.

Bonissone, P. P., Blau, L., & Ayub, S. (1990). Leveraging the integration of approximate reasoning systems. *Proceedings of 1990 AAAI Spring Symposium in Case-Based Reasoning*,11–6.

Branting, L. (1991). *Integrating Rules and Precedents for Classification and Explanation: Automating Legal Analysis* (PhD thesis). Department of Computer Science, University of Texas, Austin, TX.

Brealey, R., & Myers, S. (1996). *Principles of Corporate Finance*. McGraw-Hill.

Cern, S. (2002). *Intelligent tutoring systems*. Springer Verlag Publication.

Curet, O., & Jackson, M. (1996). Evaluting a case-based learning and reasoning applications: the top management fraud diagnostic tool TMFDT. In *Proceeding of Fourth German Workshop on Case-Based Reasoning — System Development and Evaluation*. Computer Science Department, Humboldt University.

Dutta, S., & Bonissone, P. P. (1993). Integrating case- and rule-based reasoning. *International Journal of Approximate Reasoning*, *8*(3), 163–203. doi:10.1016/0888-613X(93)90001-T

Elorriaga, J., & Fernandez-Castro, I. (2000). Using case-based reasoning in instructional planning. towards a hybrid selfimproving instructional planner. *International Journal of Artificial Intelligence in Education*, 416–449.

Finlay, P. N. (1992). IT for competitive Advantage: The place of expert systems. *The Journal of Strategic Information Systems*, *1*(3), 126–133. doi:10.1016/0963-8687(92)90025-R

Giraffa, L. (1997). Multi-ecological: a learning environment using multi-agent architecture. *Multi-Agent System: Theory and Application Proceedings*.

Golding, R., & Rosenbloom, P. S. (1991). Improving rule-based systems through case-based reasoning. *Proceedings of the Ninth National Conference on Artificial Intelligence AAAI*, *91*, 22–27.

Gonzalez, C., Burguillo, J., & Llamas, M. (2006). A qualitative comparison of techniques for student modeling in intelligent tutoring systems. *ASEE/FIE Frontiers in Education Conference*. 10.1109/FIE.2006.322537

Guida, M., Marchesi, P., & Basaglia, G. (1992). Knowledge-based decision support systems for manufacturing decision-making. *Information and Decision Technology*, *18*, 347–361.

Han, S. G., Lee, S. G., & Jo, G. S. (2005). Case-based tutoring system for procedural problem solving on the www. *Expert Systems With Applications*, 573–582.

Huang, W., & Cross, G. (1989). Reasoning about Trademark infringement cases. *Proceedings of DARPA Case-Based Reasoning Workshop*, 270–274.

James, R. N. (1997). DISXPERT: A rule-based vocational rehabilitation risk assessment system. *Expert Systems with Applications*, *12*(4), 465–472. doi:10.1016/S0957-4174(97)00006-7

Kharrat, M., Reyhani, N., & Badie, K. (2003). A case-based reasoning approach to intelligent tutoring system by considering learner style model. *Proceedings of the 2003 Systems and Information Engineering Design Symposium*. 10.1109/SIEDS.2003.158002

Kim, B., & Lee, S. (1995). A bond Rating expert system for industrial companies. *Expert Systems with Applications*, *91*(1), 63–70. doi:10.1016/0957-4174(94)00049-2

King, D. (1990). Intelligent decision Support: Strategies for integrating decision support, database management, and expert system technologies. *Expert Systems with Applications*, *1*(1), 23–38. doi:10.1016/0957-4174(90)90066-4

Kolodner, J. (1993). *Case-based Reasoning*. Morgan Kaufmann. doi:10.1016/B978-1-55860-237-3.50005-4

Lee, S., & Wu, C. (1995). CLXPERT: A rule-based scheduling system. *Expert Systems with Applications*, *22*(2), 153–164. doi:10.1016/0957-4174(94)00058-4

McGinty, L., & Smyth, B. (2001). Collaborative case-based reasoning: Applications in personalized route planning. *4th International Conference on Case Based Reasoning, ICCBR. Lecture Notes in Computer Science, 2080*, 362.

McIvor, R., Mulvenna, K., & Humphreys, P. (1997). Hybrid knowledge-based system for strategic purchasing. *Expert Systems with Applications*, *12*(4), 497–512. doi:10.1016/S0957-4174(96)00009-7

Meyer, H., Tore, A. D., Siegel, F., & Curley, F. (1992). The strategic use of expert systems for risk management in the insurance Ž. industry. *Expert Systems with Applications*, *5*(1-2), 15–24. doi:10.1016/0957-4174(92)90090-F

Michaelsen, R. (1984). An expert system for federal tax planning. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, *1*(2), 149–167. doi:10.1111/j.1468-0394.1984.tb00440.x

Pal, K., & Campbell, J. A. (1995). A hybrid system for decision-making about assets in English divorce cases. In *Advances in Case Based Reasoning: First UK CBR Workshop*. Springer-Verlag. 10.1007/3-540-60654-8_30

Pal, K., & Campbell, J. A. (1996). A hybrid Legal decision-support system using both rule-based and case-based reasoning. *Information & Communications Technology Law*, *5*(3), 227–245. doi:10.1080/13600834.1996.9965747

Pal, K., & Campbell, J. A. (1997). An application of rule-based and case-based reasoning within a single legal knowledge-based system. *The Data Base for Advances in Information Systems*, *28*, 48–63.

Pike, R., & Neale, B. (1999). *Corporate Finance and Investment: Decisions and Strategies*. Prentice-Hall Europe.

Quah, T., Tan, C., Raman, K. S., & Srinivasan, B. (1996). Towards integrating rule-based expert systems and neural networks. *Decision Support Systems*, *17*(2), 99–118. doi:10.1016/0167-9236(95)00016-X

Rabinowitz, L., Weinberg, M. A., & Blank, M. V. (1997). *Weinberg and Blank on Takeovers and Mergers*. Sweet & Maxwell.

Rissland, E. L., & Skalak, B. D. (1991). CABARET: Rule interpretation in a hybrid architecture. *International Journal Man–Machine Studies*, *34*, 839–887.

Rissland, L., & Ashley, K. (1989). HYPO: a precedent-based legal Ž. reasoner. In G. Vandenberghe (Ed.), *Advanced Topics of Law and Information Technology* (pp. 213–234). Kluwer.

Ryan, C. (1997). *Company Law*. Butterworths.

Schutzer, D. (1990). Business expert systems: The competitive edge. *Expert Systems with Applications*, *1*(1), 17–21. doi:10.1016/0957-4174(90)90065-3

Shin, K., & Han, I. (1999). Case-based reasoning supported by genetic algorithms for corporate bond rating. *Expert Systems with Applications*, *16*(2), 85–95. doi:10.1016/S0957-4174(98)00063-3

Sierra, E., Martnez, R., & Cataldi, Z. (2006). *Towards a methodology for the design of intelligent tutoring systems*. *Research in Computing Science Journal*.

Sneath, P. H. A., & Sokal, R. R. (1976). *Numerical Taxonomy*. Freeman.

Steinbart, P. (1984). *The Construction of a Rule-based Expert System as a Method for Studing Materiality Judgements* (PhD thesis). Michigan State University.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge Univ. Press.

Turban, E. (1990). Decision Support and Expert Systems — Management Support Systems. Macmillan UK.

VanLehn, K. (1988). *Student Modelling*. M. Polson. Foundations of Intelligent Tutoring Systems.

Yegge, W. M. (1996). *A Basic Guide for Valuing a Company*. Wiley.

ACKNOWLEDGMENT

The author sincerely thanks late Mr Owen Palmer, at Greenwich Management School, late Professor John A Campbell, at University College London, Computer Science Department, Professor Stephen Guest of the Law Faculty, University College London, and Prof. Dan Hunter, Executive Dean, QUT Faculty of Law and Chief Investigator, ARC Centre of Excellence for Automated Decision-Making and Society, Queensland University of Technology, Brisbane, Australia, for their constructive suggestions on legal decision support systems.

Chapter 4
Automatic Keyword Extraction From Text Documents

Furkan Goz
Kocaeli University, Turkey

Alev Mutlu
Kocaeli University, Turkey

ABSTRACT

Keyword indexing is the problem of assigning keywords to text documents. It is an important task as keywords play crucial roles in several information retrieval tasks. The problem is also challenging as the number of text documents is increasing, and such documents come in different forms (i.e., scientific papers, online news articles, and microblog posts). This chapter provides an overview of keyword indexing and elaborates on keyword extraction techniques. The authors provide the general motivations behind the supervised and the unsupervised keyword extraction and enumerate several pioneering and state-of-the-art techniques. Feature engineering, evaluation metrics, and benchmark datasets used to evaluate the performance of keyword extraction systems are also discussed.

INTRODUCTION

Keywords[1] are salient words that best describe the content of a document. Keywords are particularly important as they provide readers with the concept of a document and have applications in document clustering (Kang, 2003; Christy, Gandhi, and Vaithyasubramanian, 2019), searching (Liu, Do, and Cao, 2020), meta-data enrichment (Al-Natsheh, Martinet, Muhlenbach, Rico, and Zighed, 2018), and summarizing (Choi, Kim, and Lee, 2020).

The number of online documents is increasing at a high rate and a few of them come with keywords. Turney (2002) cites that some 20% of full-text academic journals do not contain author-assigned keywords, which goes up to 30% for web pages. As manual keyword extraction of keywords is a time-consuming, error-prone, and expensive task, automatic keyword indexing has become an active research topic. Automatic indexing of documents with keywords can be achieved in two ways, either via keyword

DOI: 10.4018/978-1-7998-6792-0.ch004

assignment or keyword extraction. Keyword assignment is concerned with indexing a document with keywords that are not necessarily present in the document. It requires external resources such as ontology and thesaurus (C. Zhang and Xu, 2009). This approach is cited as particularly useful for short texts such as microblog feeds, as they do not contain enough structural, linguistic, and statistical information (Singhal, Kasturi, Sharma, and Srivastava, 2017). The primary advantage of keyword assignment is that documents can be indexed with keywords that are not present in the text but better describe the content of the document than in-text words. Primary limitations of keyword assignment, on the other hand, include their dependence on external resources and being language-dependent. However, multilingual vocabularies such as EUROVOC and AGROVOC may help overcome the second limitation (Steinberger, 2001; Balaji et al., 2010).

Keyword extraction methods aim to index a document with words that are present in the document. Keyword extraction systems are classified based on several aspects. One classification is based on the learning algorithm they employ: supervised, unsupervised. Supervised keyword extraction can be considered a binary classification problem, where words in a document are assigned either into the keyword class or the non-keyword class (Witten, Paynter, Frank, Gutwin, and Nevill-Manning, 1999). Unsupervised keyword extraction methods use unsupervised learning methods for keyword extraction. Simple statistic approaches make use of n-gram statistics (Suleiman and Awajan, 2017), term frequency (Hulth, 2003), and term frequency-inverse document frequency (Sun, Wang, and Xia, 2017). Linguistic approaches incorporate NLP tools such as lexical analysis (Ercan and Cicekli, 2007), syntactic analysis (Hulth, 2003).

Keyword extraction systems also differ by the number of documents they work on. Some keyword extraction systems (Matsuo and Ishizuka, 2004; Campos et al., 2020) consider the content of a single document to extract keywords, while some others (C. Zhang and Xu, 2009; T. D. Nguyen and Luong, 2010) consider the content of multiple documents. Multi-document approaches benefit from a richer source of information, i.e., learn structures of documents to improve keyword extraction (T. D. Nguyen and Luong, 2010), verify single document-based findings against other documents in the corpus (Wartena, Brussee, and Slakhorst, 2010), and build domain knowledge (C. Zhang and Xu, 2009). However, building and annotating, in case of supervised learning, a corpus is a difficult task.

Keyword extraction algorithms yet differ by the type of document they focus on. The type of a document is fundamental, as microblog posts may not contain enough structural and linguistic information, language is causal, texts contain typos, slang, and abbreviations (Marujo et al., 2015). Moreover, keyword extraction from such texts requires modeling temporal and social features (Bellaachia and Al-Dhelaan, 2014). Webpages may be unstructured. News and scientific articles generally cover multiple topics; hence keywords relating to every topic need to be extracted (Z. Liu, Huang, Zheng, and Sun, 2010).

In the literature, several features are proposed to describe words. These features can be grouped as statistical, linguistic, and semantic (Siddiqi and Sharan, 2015). As keyword indexing has applications in several domains and a number of methods have been proposed, some properties that keywords should conform are also defined. Firoozeh, Nazarenko, Alizon, and Daille (2020) classify keywords based on their information content, presentation, and domain conformity. Although classical machine learning performance metrics such as precision, recall, and f-score are used to evaluate keyword extraction performance, their definitions are different from standard applications. Furthermore, performance metrics specific to information retrieval such as mean average precision, precision@K are also defined to evaluate keyword indexing.

In this study, we introduce various keyword extraction methods and cite their advantages and disadvantages. We also present properties of keywords, several features to describe words, and evaluation metrics for keyword extraction.

FEATURE ENGINEERING

Several features are proposed to represent words in the literature, and there is still a great effort in feature engineering. Below we list some of the most commonly used features.

Term Frequency (tf): Term frequency indicates the number of occurrences of a word in a document. A frequent word is assumed to be more relevant to a document.

Term Frequency - Inverse Document Frequency (tf-idf): Inverse document frequency of word t, idf_t is the inverse ratio of the number of documents t appears in over the number of total documents in a corpus. $tf - idf_{t,d}$ score of word t for document d is obtained by $tf_{t,d} x\, idf_t$. The score is highest when *t* is frequent for a small number of documents and lowest when it appears in almost every document. A word with low *tf-ifd* score is considered non-discriminative.

Position: Position is the offset of a word within a document. Several studies consider that words appearing in early positions are more important than others.

Co-Occurrence Count: This metric indicates how often two words coappear in a predefined window size.

Term Relatedness to Context: If a word frequently co-occurs with different words, it is assumed to be unimportant (Machado, Barbosa, Pais, Martins, and Dias, 2009). Term relatedness to context is defined as

$$\frac{WLR(t)}{2\,x\,tf(t)}$$

where WLR is the number of the words that appear with word t, either on the left or right hand side, within a specified window size. Campos et al. (2020) normalizes this value with $tf(t) = maxTF$ to penalize non-frequent words.

Centrality Measures: In graph-based keyword extraction approaches centrality measures are used extensively. The most commonly used centrality measures are degree, clustering coefficient, structural diversity index, eigenvector, closeness, betweenness, coreness, and PageRank. Degree centrality indicates the number of edges incident to a node. For directed graphs, in-degree centrality and out-degree centrality are also defined. They are defined as the number of edges going into a node and the number of edges leaving a node, respectively. Degree, in-degree and out-degree centrality measures can be normalized by dividing them by the maximum degree of the graph. Nodes with high degree centrality are assumed to have an impact on the dynamics of the graph. Clustering coefficient of node *v* is the fraction of the number of triangles centered on *v* over the maximum number connections of *v*. Structural diversity index of node *v* is the normalized entropy of edges incident to *v*. Eigenvector centrality of a node is calculated with respect to centrality of its neighbors. The eigenvector centrality for node i is the i^{th} element of the vector x, $Ax = \lambda G$ where A is the adjacency matrix of the graph G with eigenvalue λ.

Eigenvector centrality is used to distinguish nodes with the same degree centrality. Closeness centrality of node v is the reciprocal of the average shortest path distance to v from all nodes reachable to v. Betweenness centrality of node v is the fraction of the shortest paths that pass through v over the number of all pairs shortest paths. Coreness of node v indicates the highest core v belongs to. PageRank of node v is the probability of arriving v in a random walk manner. Vega-Oliveros, Gomes, Milios, and Berton (2019) classify these measures as local, intermediate and global based on the information needed to calculate them. The local centrality measures of a node such as degree centrality are calculated using the information provided by the node. Intermediate centrality measures of a node such as clustering coefficient and structural diversity index are calculated using information provided by neighbors of the node. Global centrality measures of a node such as betweenness, closeness, eigenvector, coreness, and PageRank are calculated using the information provided by all nodes of the graph.

In addition, POS tags, phrase length, suffix sequences, semantic ratio, Boolean features indicating whether a word starts with a capital letter or is all capitalized, whether it appears in the title or in abstract are also studied.

Properties of Keywords

Several properties have been proposed for keywords (Z. Liu,Li, Zheng, and Sun, 2009; Firoozeh et al., 2020). Below we define some of them briefly and provide studies that aim to satisfy these properties.

Specificity: Keywords should be as specific as possible to the domain. To this aim, KIP (Wu, Li, Bot, and Chen, 2005) utilizes domain specific dictionaries to score candidate keywords.

Exhaustivity: Keywords should reflect all the subjects presented in a document. To achieve this property, Z. Liu et al. (2009) cluster candidate keywords and select seeds from these clusters to extract final keywords.

Minimality: All keywords should be different in meaning. To achieve diversity of the keywords, Habibi and Popescu-Belis (2013) propose a reward function whose value decreases when multiple keywords from the same topic are chosen.

Representativity: Keywords should reflect major aspects of a document rather than minor ones. Xiong and Guo (2019) adjust PageRank's dumping factor to indicate relatedness of a word to the topic of the document.

Well-Formedness: Keywords should be in human interpretable form. To achieve well-formed keywords Le, Le Nguyen, and Shimazu (2016) propose to include participles in keyword post-processing.

Other keyword properties discussed in the literature include impartiality (keywords should not contain any bias), citationess (keywords should be in their uninflected form); conformity (keywords should be natural to the domain), homogeneity (more than one keyword reflect the same concept, only one of them should be used) and univocity (keywords should not be ambiguous, at least for the domain expert).

EVALUATION METRICS AND DATASETS

In the literature, several metrics are used to evaluate keyword extraction systems' performance, some of which are adapted from machine learning and some others from information retrieval. These metrics can be broadly classified as human evaluation and automatic evaluation. Human evaluation employs

human judgments to evaluate extracted keywords. This approach is impractical as it is both expensive and subjective (Zesch and Gurevych, 2009).

Automatic keyword evaluation metrics are based on matching extracted keywords against gold standard keywords, i.e., keywords assigned by authors or human annotators. The matching can be achieved in two ways:

Exact Matching: An extracted keyword and a gold standard keyword are compared in their stemmed version. If they are equal, exact matching is achieved.

Partial Matching: Partial matching occurs when a gold standard keyword is a substring of an extracted. Hence, it compares the extracted keywords and the gold standard keywords in a word-wise manner. In this step, both the extracted and the gold standard words are stemmed, as well. This matching is cited as more suitable for keyword extraction systems (Zesch and Gurevych, 2009).

Below we define and formulate some of the most commonly used automatic keyword evaluation metrics. In these definitions, K_{ext} donates the set of keywords extracted by a keyword extraction system, K_{match} donates the set of extracted keywords that match the gold standard keywords, K_{gold} donates the set of gold standard keywords. $|A|$ donates cardinality of set A.

Precision indicates the fraction of the number of correctly extracted keywords over the number of extracted keywords.

$$recall = \frac{|K_{match}|}{|K_{ext}|}$$

Recall indicates the fraction of the number of correctly extracted keywords over the number of gold standard keywords.

$$recall = \frac{|K_{match}|}{|K_{gold}|}$$

F1-Score is the harmonic mean of precision and recall.

$$recall = \frac{2 \text{ x precision x recall}|}{\text{precision x recall}}$$

Precision, recall, and F1-score are calculated for a fixed cutoff value, and are reported as P@X, R@X and F1@X, respectively. X indicates the number of keywords extracted by a keyword extraction system. However, these metrics may be misleading when X > $|K_{gold}|$ as X - $|K_{gold}|$ number of keywords will always be wrong.

R-Precision is proposed to overcome the problems due to differing cardinalities of K_{ext} and K_{gold} R-precision picks $|K_{gold}|$ keywords from K_{ext} and calculates P-precision as follows:

$$R-precision = \frac{|K_{ext}|}{|K_{gold}|}, st.|K_{ext}| = |K_{gold}|$$

Mean Reciprocal Rank is the harmonic mean of ranks and is calculated as

$$MRP = \frac{1}{|D|}\sum_{d \in D}\frac{1}{rank_d}$$

D indicates a document set and *d* is a specific document in *D*, $rank_d$ is the position of the first matched extracted keyword in rank-ordered list of extracted keywords.

Mean Average Precision indicates mean of the average precisions. Average precision is defined as

$$MAP = \frac{1}{|D|}\sum_{i=1}^{|D|} AP_{d_i}$$

where, *D* is a document set, d_i is a document in *D*, and AP_{d_i} is the average precision of d_i. AP is calculated as follows:

$$AP = \sum_{r=1}^{|K|}\frac{P(r)\,x\,rel(r)}{|K_{match}|}$$

where, P(r) indicate the first r keywords in K, and *rel(r)* is a function that returns if r^{th} element of *P(r)* is in the gold standard keyword set. MRP and MAP work on rank-order list of keywords.

Benchmark datasets to evaluate keyword extraction systems' performance include Inspec, SemEval2010, SemEval2017, KDD, WWW, Marujo2012, and Krapivin2012. In Table 1, we provide some statistics related to the datasets.

Table 1. Properties of some benchmark datasets

Dataset	Lang.	Type	# Docs.	Avg. Length	Avg. Gold Keywords	Absent keywords	Domain
Inspec	Eng.	Abstract	2000	128.2	14.6	37.7%	CS
KDD	Eng.	Abstract	755	76	6	53.2%	CS
WWW	Eng.	Abstract	1330	84.1	5.8	55%	CS
SemEval2010	Eng.	Paper	243	8332	16.5	11.3%	CS
SemEval2017	Eng.	Paragraph	493	178.2	18.2	0%	Misc
Krapvin2012	Eng.	Paper	2304	8041	6.3	15.3%	CS

PRE- AND POST-PROCESSING IN KEYWORD EXTRACTION

Pre-processing in keyword extraction generally concerns two distinct tasks. The first task is transforming a document into a format that is processable by a keyword extraction program. This transformation includes several NLP tasks, some of which are as follows:

Stemming: Stemming converts words into their base form, e.g., prefer, preferable, preferably, preferments, preference will map to common form prefer. This is particularly important when the frequency of a word is of concern during keyword extraction.

Sentence Splitting and Chunking: Sentence splitting is the process of extracting sentences from a text, and chunking is the process of extracting grammatical constructs, such as noun phrases, from a text. These two processes are essential as it is commonly agreed that a keyword cannot span over two sentences or chunks (Campos et al., 2020).

POS Tagging: POS tagging is annotating a word with a part-of-speech tag. Several studies (Mihalcea and Tarau, 2004; Rabby, Azad, Mahmud, Zamli, and Rahman, 2020) assume that keywords are generally nouns and noun phrases and build their systems using words of a particular type of POS tags. Hence, POS tagging enables filtering out words of irrelevant POS tags from a text.

Named Entity Recognition: Named entity recognition identifies text units that define entities of various types such as person, organization, and location. Studies have demonstrated that named entities can improve keyword extraction success (Huang, Wang, and Wang, 2020; Nagy, Berend, and Vincze, 2011).

Annotation: Annotation refers to tagging words in a text. Such annotations are later used to build models. For example, Alzaidy, Caragea, and Giles (2019) annotate each word with a label that indicates either the word appears in a keyword or not and build CRF models using annotated word sequences. C. C. Zhang (2008) follows a similar approach; however, the author uses a richer set of tags. Campos et al. (2020) annotate each word in a text with binary tags such as acronyms, uppercase, and digit and calculates several statistics with the help of these tags.

Removal of Stop Words and Text Cleaning: Stop words are defined as the most commonly used words in a language. Such words are assumed uninformative; hence they are removed from a text. Moreover, some studies remove words that contain fewer characters than a threshold (Campos et al., 2020) and or a non-alphabetic symbol (Rabby et al., 2020).

The second task of pre-processing involves extracting candidate keywords. During the pre-processing step of YAKE! (Campos et al., 2020), all 1-gram tokens are extracted and considered as candidate keywords. In SGRank (Danesh, Sumner, and Martin, 2015), all n-grams, n ranging from 1 to 6, are extracted during the pre-processing step, and those that contain stop words, punctuation marks, or any word with POS tag other than noun, adjective, and verb are removed. T. D. Nguyen and Luong (2010) analyzed the distribution of keywords across different logical sections of scientific papers and cited that headers and abstracts are most densely populated with keywords and introductions contain most of the keywords. Furthermore, they also cited that the first sentences of paragraphs are more likely to contain keywords. Based on these observations, the authors introduced a keyword extraction system, namely WINGUS, where candidate keywords are extracted from the most covering and dense sections.

Post-processing in keyword extraction includes generating multiple-word keywords from single-word keywords and transforming keywords into a human-readable format. Post-processing is argued to have a significant impact on the performance of a keyword extraction system (Hasan and Ng, 2010).

SUPERVISED KEYWORD EXTRACTION

Supervised approaches consider the keyword extraction problem as a binary classification problem: assigning terms of a document either to the keyword class or the non-keyword class (Turney, 1999). Supervised approaches in keyword extraction are cited to build good models for the domains they are trained for (Papagiannopoulou and Tsoumakas, 2020); however, they perform poorly for any other domain. To build supervised models that can perform well on various domains, Turney (1999) suggests using manually arranged training data that reflects test data properties.

Initial works on supervised keyword extraction used traditional machine learning algorithms such as C4.5 and Naive Bayes. In an application of C4.5 to keyword extraction (Turney, 1999), phrases were represented with nine features such as length of the phrase, frequency of the most frequent word in the phrase, and position of the first word in the phrase. The leaves of the tree represented the class label of the phrase. As one cannot control the number of items to be assigned to a certain class in C4.5, the author implemented the tree as a soft decision C4.5, which calculates the probability estimate for each class of a test instance. To obtain N-keywords, the author retrieved N-keywords with the highest probability estimate for the class *keyword*. Several experiments were conducted with varying parameters, and the author concluded that parameter setting drastically affects the performance. In a subsequent work, called GenEx (Turney, 2000), the author proposed a supervised keyword extraction method where a genetic algorithm was employed to adjust parameters. In this study, the author also cited that domain knowledge could improve supervised learning keyword extraction performance.

The application of Naive Bayes for keyword extraction was firstly discussed in KEA (Witten et al., 1999). In this initial implementation, words were represented with two features: *tfxidf* score and the position of the first occurrence of the word. (T. D. Nguyen and Kan, 2007) improved performance of KEA by introducing new features such as POS tag sequence and suffix sequences. Medelyan, Frank, and Witten (2009); T. D. Nguyen and Luong (2010) further improved the performance of KEA by introducing new features.

Learning to rank algorithms are also employed in supervised keyword extraction. Learning to rank differs from classification algorithms so that rather than assigning instances to classes, they provide an ordering of the instances based on their degree of preference. Furthermore, classification techniques work on a single instance while learning to rank models work on a set of items at a time. Broadly speaking, learning to rank methods are trained on pairs $S = \{(x_i, x_k), y_{i,k})\}_{i,k=1}^{m}$ where x_i and x_k indicate objects, $y_{i,k} \in \{l_1, l_2 \ldots l_n | l_1 \succ l_2 \ldots \succ l_n\}$, indicates grade and $\succ$ implies total order. Objective function is $f(a,b) = f(x)$ assigning a score to instance a and b (Li, 2011). In context of supervised keyword extraction, such methods are trained on $(x_i - x_j, y_{i,j})$, where x_i and y_j are feature vectors describing two terms, and $y_{i,j} = 1$ if $x_i > x_j$ and -1 otherwise. To extract keywords, words of a document are ranked and sorted according to the learned model, and top N-ranking words are considered keywords. Jiang, Hu, and Li (2009) use Ranking SVM, Wang and Li (2011) use Naive Bayes based ranking to extract keywords.

Another direction in supervised keyword extraction is based on sequence labeling. Sequence labeling can be considered as a multiclass classification problem that predicts labels of a given observation sequence from a set of labels (N. Nguyen and Guo, 2007). C. Zhang (2008) tags words of each sentence with KWB, KWY, KWN, which indicate, respectively, that the word is the first word of a keyword, the

word itself is a keyword, and the word is not a keyword or a part of a keyword. The tagged sequences are then used to train a CRF model. Lu and Chow (2019) propose a method based on generalized CRF (Semi-CRF) for keyword extraction. Different than conventional CRF, Semi-CRF assigns tags to n-gram structures rather than individual words.

UNSUPERVISED KEYWORD EXTRACTION

Unsupervised keyword extraction methods can be categorized into three categories: simple statistics, graph-based, and NLP-based.

Simple statistics-based approaches extract keywords from text based on various local and global statistics regarding words. Graph-based approaches represent text documents as graphs and generally implement node ranking algorithms. NLP-based approaches make use of linguistic features such as POS tagging, lexical analysis, and syntactic analysis.

Simple Statistics-Based Keyword Extraction

Simple statistic approaches aim to extract keywords from text based on various statistics regarding words. Most commonly used statistics include term position, term frequency, term frequency-inverse document frequency, and word co-occurrence counts. Below we summarize some studies based on simple statistics.

Matsuo and Ishizuka (2004) propose keyword extraction for a single document, and cite that a word that co-occurs with frequent words of a document has a biased distribution. However, high biases may occur for infrequent words due to lexical, semantic, and other relations. The authors use the X^2 test to measure the statistical significance of biases. The authors count frequencies of each word in a document and select the top 30% as frequent words. Then they calculate the X^2 value of words that co-occur with frequent words and output words with the highest X^2 as keywords.

Papagiannopoulou and Tsoumakas (2018) formulate the keyword extraction problem as an outlier detection problem. The authors assume that the non-keyword words' embeddings would come from the same distribution, and outliers should be keywords. To this aim, GloVe is used to learn vector representations of the words, and the Minimum Covariance Determinant estimator is used to model the distribution of the words. As the size of a document is usually small, the authors cite that learning vectors of small size would avoid overfitting. The authors rank the outliers based on their position in the document and pick the top 100 ranking unigrams as candidate keywords. Multi-word keywords are constructed by concatenating adjoining unigrams. A post ranking based on *tf-idf* is implemented for uni-, bi-, and trigrams to get the final set of keywords. Papagiannopoulou, Tsoumakas, and Papadopoulos (2020) extend this work and rank candidate keywords based on the distance between their local vector representation and the center of the main distribution in addition to their first occurrence in the text.

TeKET (Rabby et al., 2020) is an unsupervised keyword extraction system that uses term frequency to prune candidate keywords during preprocessing and, furthermore, to rank final keywords. *TeKET* firstly generates a list of candidate keywords by extracting noun phrases from the document. This initial list is pruned by removing candidate keywords that contain words that rarely appear, those with frequency less than some threshold. TeKET associates each word in a keyword with a cohesiveness index, indicating

its frequency in other keywords. The weight of a keyword is calculated by multiplying the summation of each word's frequencies in the keyword by the summation of each word's cohesiveness index.

YAKE! (Campos et al., 2020) is a recent keyword extraction method that aggregates multiple local statistical text features. These features include casing, term position, normalized term frequency, relatedness to context, term different sentence. With casing, the authors give more weight to words that start with a capital letter than those starting with a lower case letter. The position of a term in this study reflects the position of the sentence the term appears in. More weight is assigned to words that appear in the early sentences of a text. Term frequency is an aspect that reflects the importance of a term. However, in long documents, the frequency may yield misleading bias; hence the authors use normalized term frequency. Relatedness to context measures how often a term appears with different terms either on its left side or right side. If the number of terms is mostly different, then the significance of the term decreases. The last metric the authors use is the number of different sentences a term occurs. They assume that a term is more likely to be important if it appears in many different sentences.

Node Ranking-Based Keyword Extraction

In node ranking-based keyword extraction, documents are represented as word-of-graphs (GoW). In this representation, the candidate keywords of a document constitute the vertex set of the graph. The candidate keywords that appear in a window of a size n, i.e., located in n consecutive positions, are connected via edges. Graphs may be directed or undirected, as well as weighted or unweighted. Nodes are assigned with a weight indicating their initial importance of the candidate keywords. In node ranking-based keyword extraction, a node ranking algorithm is implemented to extract the most important nodes, which are later post-processed to form keywords. The PageRank algorithm is extensively used for node ranking in graph-based keyword extraction. The PageRank algorithm based ranking is formulated below, where *S(v)* is the weight of node *v*, *d* is a dumping factor, *In(v)* indicates the set of nodes connected to *v*, and *Out(v)* indicates the set of nodes *v* is connected to, and w_{ij} indicates the weight of edge between nodes i and j. The ranking is assumed to converge when the value of a node differs by ε between two iterations.

$$S\left(V_i\right)=\left(1-d\right)xW\left(v_i\right)+d\,x\sum_{V_j\epsilon\,In\left(V_i\right)}\frac{w_{ij}}{\sum_{V_k\epsilon\,Out\left(V_j\right)}w_{jk}}S\left(V_j\right)$$

Below, we introduce some of the most cited graph-based keyword extraction studies.

TextRank (Mihalcea and Tarau, 2004) is one of the earliest studies on graph-based keyword extraction that extracts keywords for a single document. It builds a GoW on a pre-processed text where all words other than nouns and adjectives are removed. It assumes all candidate keywords have equal importance hence assigns initial weight 1 to all nodes. The nodes are ranked according to the PageRank algorithm and sorted based on their rank in descending order. The top 33% percent of the nodes are selected as candidate keywords. If any two or more top-ranking words form a sequence in the original text, they are concatenated to form a final keyword. The experimental analysis shows that TextRank achieves the best result for an undirected graph representation formed using a window size equal to 2. The authors do not report any results considering weighted graphs.

ExpandRank (Wan and Xiao, 2008) aims to improve single document keyword extraction systems by enriching GoWs with the content of topically related documents. For a specific document, *d*, ExpandRank selects a small number of topically related documents, *D*, from a corpus based on cosine similarity. It builds a single GoW representing the content of d and D in a way similar to TextRank but assigns weights to the edges. The weight of an edge is calculated by multiplying the two nodes' co-occurrence count and the similarity score of document *d* and the document in which the nodes appear. The weight of a node is the *tf-idf* score of the word it represents. After implementing a node-ranking algorithm, nodes that appear in *d* are sorted based on their final ranks. Adjacent nodes are merged to form multi-word keywords with a final score equal to the summation of scores the individual words.

PositionRank (Florescu and Caragea, 2017) preprocesses texts and builds a GoW similar to TextRank. However, it determines the nodes' an initial weights according to the positions of the words they represent. A node is assigned with initial weight m_i / N where m_i is the position of the word in the text the node represents and N is the total number of words in the text. If a word appears in multiple positions, it aggregates weights for each position to get the overall weight. It implements the PageRank algorithm to rank nodes. Once convergence is achieved, PositionRank builds candidate keywords up to length three and assigns a score to each candidate by summing scores of the individual words in the candidate keyword. Top 10 ranking candidates are considered as keywords. The authors state that aggregating position weights improves performance compared to using weight of the first appearance only. Furthermore, the authors cite that window size does not have statistically significant effect on the performance.

HG-Rank (Bellaachia and Al-Dhelaan, 2014) differs from the studies mentioned above as it models documents as hypergraphs and is particularly proposed to extract keywords from microblog posts. Hypergraphs are a generalization of graphs where an edge, called hyperedge, can connect any number of vertices. In the hypergraph representation of the microblog posts, nodes correspond to unique words in the posts, and hyperedges connect those nodes that appear in the same post.

TopicRank (Bougouin, Boudin, and Daille, 2013) differs from the literature as it builds a graph based on topics, i.e., clusters of candidate keywords, rather than individual candidate keywords. The authors assume that the longest noun phrases are candidate keywords. To identify topics, they compare each candidate keyword against the others and assume that any candidate keyword pairs that overlap by 25% of their stemmed words belong to the same cluster. TopicRank builds a complete graph where nodes represent topics and edges are weighted according to the position of the keywords in the document. Topics are ranked based on the random surfer model, and top-ranking topics are chosen to extract final keywords. The authors propose three strategies to select final keywords from the top-ranking topics. The first strategy is to select the candidate keyword that appears earliest in the document. The second strategy is to select the most frequent candidate keyword, and the last one selects the candidate keyword that is the centroid.

NLP-Based Approaches

Ercan and Cicekli (2007) propose a supervised keyword extraction system using lexical chains. A lexical chain is a list of related words that captures a portion of the cohesive structure of the document. The authors use WordNet to infer word senses and relations between words while building lexical chains. In the study, synonym, hypernym, and meronym relations are used to build lexical chains. As WordNet does not provide such information for phrases, the authors limit their study to extract single-word keywords.

The authors propose four features derived from lexical chains: lexical chain score of a word, direct lexical chain score, lexical chain span score and direct lexical chain score. Along with these features, frequency, the first and the last position of a word are used to train C4.5.

Zu, Xie, and Liu (2020) propose a node ranking-based method for keyword extraction that utilizes word and document embeddings. The authors propose to assign initial weights to candidate keywords based on their similarity to the document. The authors use cosine similarity to calculate a candidate keyword and the document. Sen2Vec model is used to generate embeddings. The authors propose two different graph models to represent a document. The first model follows traditional GoW model, where each node represents a distinct word and words co-occurring are connected. Edge weights in this representation indicate co-occurrence count. In the second model, noun phrases are considered as candidate keywords. A fully connected graph is constructed, where nodes represent candidate keywords, and edge weights are calculated based on the cosine similarity of the vector representations of candidate keywords. A modified PageRank that considers semantic preference is implemented to rank candidate keywords.

Other Approaches

Other unsupervised machine learning methods have also found applications in keyword extraction. Below we introduce some of these studies.

Loukam, Hammouche, Mezzoudj, and Belkredim (2019) propose an association rule mining based approach for keyword extraction. The authors represent a pre-processed text as a matrix where rows indicate sentences, columns indicate the words of the text, and each cell indicates whether a word appears in a sentence or not. The Apriori algorithm is applied on the matrix to find frequent itemsets, i.e. words appearing frequently in a sentence. Association rules generated out of these frequent itemsets that conform minimum support and minimum confidence are considered keywords.

Tixier, Malliaros, and Vazirgiannis (2016) consider the keyword extraction problem as an influence prediction problem and cite that nodes forming dense and cohesive subgraphs with other central nodes are influential. The authors argue that, in a word graph representation of a text, such influential nodes are likely to be the keywords of the text. To discover such nodes, the authors implement the K-truss algorithm.

RAKE (Rose, Engel, Cramer, and Cowley, 2010), Rapid Automatic Keyword Extraction, is an unsupervised, language and domain independent algorithm to extract keywords from single documents. RAKE differs from the related literature as it inputs predefined word and phrase delimiters to extract candidate keywords from the document. A GoW is built based on unique words and the authors assume that words co-occur with those that appear in the same phrase. This assumption eliminates setting window size, a parameter that is used widely in graph-based keyword extraction. To score a word three features are used: frequency(w), degree(w), and degree(w)/frequency(w). The first feature rewards frequent words of a document, the second one rewards those that appear in phrases, and the last one favors frequent words that appear in phrases. These scores are summed to get the final score of a word, and final score of a phrase is calculated by summing scores of individual words in a phrase.

WikiRank (Yu and Ng, 2018) aims to improve the graph-based keyword extraction problem by incorporating background knowledge. As the name indicates, source for the background is the open encyclopedia Wikipedia. The proposed graph representation consists of two types of nodes: candidate keyword nodes and concept nodes. An edge is established between a candidate keyword node and a concept node, if the keyword contains that concept according to TAGME. Concept nodes are weighted to indicate the number of candidate keywords they belong to. The authors formulate the keyword extraction

problem as a subgraph discovery problem. Subgraph should contain candidate keyword nodes such that the concepts the keywords connected to are as important as possible and as large in number as possible.

Vega-Oliveros, Gomes, Milios, and Berton (2019) discuss applicability of various graph centrality measures for keyword extraction. The authors consider nine centrality measures, namely degree, clustering coefficient, structural holes, eigenvector, pagerank, k-core, eccentricity, closeness, and betweenness, and conclude that they are correlated. However, the authors have also experimentally witnessed that combination of the centrality measures may improve the performance. The authors determined the best combination of centrality measures using Principle Feature Analysis and Multi-Cluster Feature Selection and concluded that performance of the combination of centrality statistically differs from the performance of the individual centrality measures.

Several studies (Papagiannopoulou and Tsoumakas, 2018; Papagiannopoulou et al., 2020; Florescu and Caragea, 2017) consider word positions are important in keyword extraction and select the position of the first occurrence of a word while ranking it. However, Lynn, Lee, Choi, and Kim (2017) argue that words occurring at the beginning and at the end of a text are more informative compared to those that appear in the middle. In order to statistically analyze effect of the word positions in keyword extraction, Kabasakal and Mutlu (2020) conducted a set of experiments. In the first experiment, namely WP, words are scored analogous to PositionRank, i.e., the first word has an initial weight 1, the second word 1/2, and word at position n has rank $1/n$. In the second experiment, namely WPB, words appearing at the beginning and at the end of a document are equally weighted, i.e., the first word and the last word have initial weight 1, the second word and the second to the last word have initial weight 1/2. In the third experiment, namely SP, words are assigned with initial weight according to the position of the sentence they appear in, regardless of their position in the sentence. In SP, words appearing in the first sentence have initial weight 1, words appearing in the n^{th} sentence have initial weight $1/n$. In the last experiment, namely, SPB, words appearing in the first and last sentences of a document are considered equally important and words appearing in those sentences have initial weight 1, words appearing in the second sentence and second to the last sentence have initial weight 1/2. The PositionRank algorithm is implemented using these settings and experiments are conducted on four benchmark datasets. In Table 2, we report the experimental results, where the highest results are in bold. As the results indicate, sentence position based rankings achieve higher results when documents get longer. Moreover, sentence position based rankings tend to perform better as documents contain more gold keywords.

Graph based keyword extraction relies on GoW and these graphs are built according to a predefined window size, and in the literature experiments are reported for varying window sizes. In order to statistically analyze the effect of window size, we conducted a set of experiments using TextRank with window sizes 2, 3, 5, and 10. We statistically analyzed the F1 scores using Friedman and Nemenyi post-hoc tests. Friedman test is a non-parametric test to detect difference of variance by ranks across multiple attempts. The null hypothesis for the Friedman test is that there are no differences among the attempts. Nemenyi post-hoc test is employed to detect differing attempts, if there is. The statistical results are presented in Figure 1. In the visual representation of the statistical analysis, CD donates the critical difference and approaches that do not statistically differ are connected with straight lines. As the results indicate, there is a statistical difference between attempts, hence setting window size is important.

Table 2. Position-based node ranking results

Dataset	Method Name	F1@2	F1@4	F1@6	F1@8	F1@10	F1@15
Nguyen	WP	**0.066**	**0.106**	**0.112**	**0.143**	**0.146**	**0.142**
	SP	0.063	0.102	0.112	0.135	0.143	0.145
	WPB	0.045	0.095	0.106	0.126	0.132	0.134
	SPB	0.059	0.098	0.111	0.13	0.138	0.143
WWW	WP	**0.073**	**0.103**	**0.108**	**0.107**	**0.103**	**0.09**
	SP	0.057	0.086	0.094	0.094	0.092	0.084
	WPB	0.052	0.08	0.088	0.091	0.088	0.081
	SPB	0.058	0.086	0.09	0.092	0.091	0.084
SemEval2010	WP	0.032	0.05	0.067	0.077	0.084	0.089
	SP	**0.034**	**0.057**	**0.07**	**0.082**	**0.087**	0.094
	WPB	0.03	0.043	0.062	0.075	0.081	0.091
	SPB	0.033	0.055	**0.07**	**0.082**	0.086	**0.095**
SemEval2017	WP	0.085	0.144	0.184	0.216	0.238	0.272
	SP	**0.091**	**0.153**	**0.198**	**0.227**	0.248	**0.281**
	WPB	0.085	0.145	0.184	0.217	0.237	0.271
	SPB	0.09	0.149	0.195	0.222	**0.249**	**0.281**

Deep Learning-Based Keyword Extraction

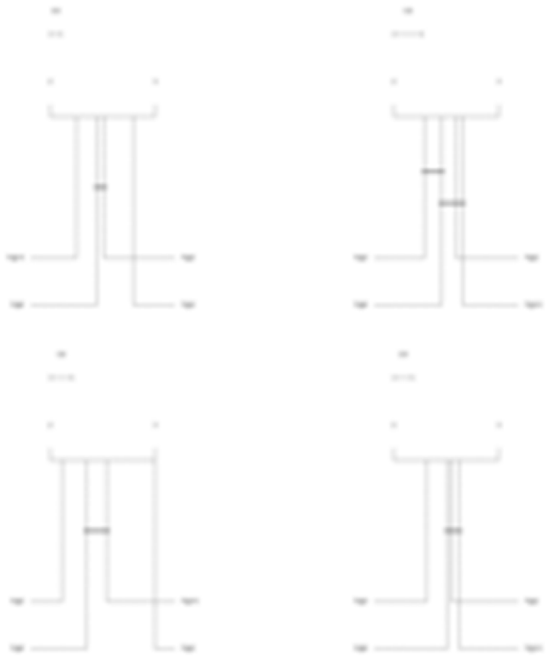
The performance of keyword extraction systems, and many other learning problems, greatly depends on feature engineering. However, proposing new features is a difficult task and requires extensive domain knowledge. Deep learning models have become popular recently and attract researchers from various domains as such models learn to perform feature extraction in an automated way. Deep learning found applications also in keyword extraction. Below we summarize some of them.

Q. Zhang, Wang, Gong, and Huang (2016) introduce a novel deep recurrent neural network (RNN) model to extract keywords from single tweets. The model includes the combination of keyword ranking, candidate keyword generation, and keyword ranking. The proposed model consists of two hidden layers. The first hidden layer discriminates one-word keywords and the second layer multi-word keywords. These two sub-objectives are combined into a final objective. The proposed architecture is evaluated on a Twitter dataset. Hashtags are used as golden keywords.

Meng et al. (2017) introduce a generative model for keyword prediction with RNN encoder-decoder framework. The encoder RNN convers arbitrarily long input sequences into hidden representations and the decoder RNN is used to generate keywords. The authors implement copying mechanism to incorporate positional information in keyword extraction.

Alzaidy et al. (2019) address keyword extraction as a sequence labeling problem. They propose Bi-LSTM-CRF model with two layers. The Bi-LSTM network captures the semantics of the input sequence, each word is annotated either KP (positive example) or Non-KP (negative example), and the CRF network

Figure 1. Nemenyi test results for Inspec, KDD, SemEval 2017, WWW



produces a probability distribution over the tag sequence using the dependencies among labels of the entire sequence. The Viterbi algorithm is used to find the best sequence of labels.

FUTURE RESEARCH DIRECTIONS

As to our knowledge multigraphs are not studied for graph based-keyword extraction. Multigraphs may enable to establish multiple relations between candidate keywords where aggregating multiple features is not possible.

Window size is a crucial parameter for several graph based keyword extraction systems. Research on GoW representation that is free of window size is still not extensively explored.

Recent studies incorporate word embeddings in keyword extraction. However, as to our knowledge, structural embeddings are yet not discussed in keyword extraction.

ACKNOWLEDGMENT

This work is supported by TUBITAK under the project number 117E566.

REFERENCES

Al-Natsheh, H. T., Martinet, L., Muhlenbach, F., Rico, F., & Zighed, D. A. (2018). Metadata enrichment of multi-disciplinary digital library: a semantic-based approach. In *International Conference on Theory and Practice of Digital Libraries* (pp. 32–43). 10.1007/978-3-030-00066-0_3

Alzaidy, R., Caragea, C., & Giles, C. L. (2019). Bi-lstm-crf sequence labeling for keyphrase extraction from scholarly documents. In *Proceedings of the The World Wide Web Conference* (pp. 2551–2557). 10.1145/3308558.3313642

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., ... Yadav, V. (2010). Agrotags–a tagging scheme for agricultural digital objects. In *Research Conference on Metadata and Semantic Research* (pp. 36–45). 10.1007/978-3-642-16552-8_4

Basaldella, M., Antolli, E., Serra, G., & Tasso, C. (2018). Bidirectional lstm recurrent neural network for keyphrase extraction. In *Italian Research Conference on Digital Libraries* (pp. 180–187). 10.1007/978-3-319-73165-0_18

Bellaachia, A., & Al-Dhelaan, M. (2014). Hg-rank: A hypergraph-based keyphrase extraction for short documents in dynamic genre. In # msm (pp. 42–49). Academic Press.

Boudin, F. (2016, December). pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System demonstrations* (pp. 69–73). Academic Press.

Bougouin, A., Boudin, F., & Daille, B. (2013). *Topicrank: Graph-based topic ranking for keyphrase extraction*. Academic Press.

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., & Jatowt, A. (2020). Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, *509*, 257–289. doi:10.1016/j.ins.2019.09.013

Choi, Y., Kim, D., & Lee, J.-H. (2020). Neural attention model with keyword memory for abstractive document summarization. *Concurrency and Computation*, *32*(18), e5433. doi:10.1002/cpe.5433

Christy, A., Gandhi, G. M., & Vaithyasubramanian, S. (2019). Clustering of text documents with keyword weighting function. *International Journal of Intelligent Enterprise*, *6*(1), 19–31. doi:10.1504/IJIE.2019.100029

Danesh, S., Sumner, T., & Martin, J. H. (2015). Sgrank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction. In *Proceedings of the 4th Joint Conference on Lexical and Computational Semantics* (pp. 117–126). 10.18653/v1/S15-1013

Ercan, G., & Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing & Management*, *43*(6), 1705–1714. doi:10.1016/j.ipm.2007.01.015

Ferragina, P., & Scaiella, U. (2010). Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1625–1628). 10.1145/1871437.1871689

Firoozeh, N., Nazarenko, A., Alizon, F., & Daille, B. (2020). Keyword extraction: Issues and methods. *Natural Language Engineering*, *26*(3), 259–291. doi:10.1017/S1351324919000457

Florescu, C., & Caragea, C. (2017). Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 1105–1115). 10.18653/v1/P17-1102

Habibi, M., & Popescu-Belis, A. (2013). Diverse keyword extraction from conversations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (vol. 2, pp. 651–657). Academic Press.

Hasan, K. S., & Ng, V. (2010). Conundrums in unsupervised keyphrase extraction: making sense of the state-of-the-art. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 365– 373). Academic Press.

Huang, H., Wang, X., & Wang, H. (2020). Ner-rake: An improved rapid automatic keyword extraction method for scientific literatures based on named entity recognition. *Proceedings of the Association for Information Science and Technology*, *57*(1), e374. doi:10.1002/pra2.374

Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (pp. 216–223). 10.3115/1119355.1119383

Jiang, X., Hu, Y., & Li, H. (2009). A ranking approach to keyphrase extraction. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 756–757). ACM.

Kabasakal, O., & Mutlu, A. (2020). *On effect of word positions in graph-based keyword extraction.* Submitted.

Kang, S.-S. (2003). Keyword-based document clustering. In *Proceedings of the 6th International Workshop on Information Retrieval with Asian Languages* (pp. 132–137). Academic Press.

Le, T. T. N., Le Nguyen, M., & Shimazu, A. (2016). Unsupervised keyphrase extraction: Introducing new kinds of words to keyphrases. In *Australasian Joint conference on Artificial Intelligence* (pp. 665–671). 10.1007/978-3-319-50127-7_58

Li, H. (2011). A short introduction to learning to rank. *IEICE Transactions on Information and Systems*, *94*(10), 1854–1862. doi:10.1587/transinf.E94.D.1854

Liu, Q., Do, T. D. T., & Cao, L. (2020). Answer keyword generation for community question answering by multi-aspect gamma-poisson matrix completion. *IEEE Intelligent Systems*, 1. doi:10.1109/MIS.2020.2997714

Liu, Z., Huang, W., Zheng, Y., & Sun, M. (2010). Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 366–376). Academic Press.

Liu, Z., Li, P., Zheng, Y., & Sun, M. (2009). Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 257–266). 10.3115/1699510.1699544

Loukam, M., Hammouche, D., Mezzoudj, F., & Belkredim, F. Z. (2019). Keyphrase extraction from modern standard arabic texts based on association rules. In *International Conference on Arabic Language Processing* (pp. 209–220). 10.1007/978-3-030-32959-4_15

Lu, X., & Chow, T. W. (2019). Duration modeling with semi-markov conditional random fields for keyphrase extraction. *IEEE Transactions on Knowledge and Data Engineering*, 1. doi:10.1109/TKDE.2019.2942295

Lynn, H. M., Choi, C., Choi, J., Shin, J., & Kim, P. (2016). The method of semi-supervised automatic keyword extraction for web documents using transition probability distribution generator. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems* (pp. 1–6). 10.1145/2987386.2987399

Lynn, H. M., Lee, E., Choi, C., & Kim, P. (2017). Swiftrank: An unsupervised statistical approach of keyword and salient sentence extraction for individual documents. *Procedia Computer Science*, *113*, 472–477. doi:10.1016/j.procs.2017.08.305

Machado, D., Barbosa, T., Pais, S., Martins, B., & Dias, G. (2009). Universal mobile information retrieval. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 345–354). Academic Press.

Marujo, L., Ling, W., Trancoso, I., Dyer, C., Black, A. W., Gershman, A., . . . Carbonell, J. G. (2015). Automatic keyword extraction on twitter. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (pp. 637–643). Academic Press.

Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal of Artificial Intelligence Tools*, *13*(01), 157–169. doi:10.1142/S0218213004001466

Medelyan, O., Frank, E., & Witten, I. H. (2009). Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 1318–1327). Academic Press.

Meng, R., Zhao, S., Han, S., He, D., Brusilovsky, P., & Chi, Y. (2017). Deep keyphrase generation. arXiv preprint arXiv:1704.06879. doi:10.18653/v1/P17-1054

Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404–411). Academic Press.

Nagy, I., Berend, G., & Vincze, V. (2011). Noun compound and named entity recognition and their usability in keyphrase extraction. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011* (pp. 162–169). Academic Press.

Nguyen, N., & Guo, Y. (2007). Comparisons of sequence labeling algorithms and extensions. In *Proceedings of the 24th International Conference on Machine Learning* (pp. 681–688). 10.1145/1273496.1273582

Nguyen, T. D., & Kan, M.-Y. (2007). Keyphrase extraction in scientific publications. In *International Conference on Asian Digital Libraries* (pp. 317–326). Academic Press.

Nguyen, T. D., & Luong, M.-T. (2010). Wingnus: Keyphrase extraction utilizing document logical structure. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 166–169). Academic Press.

Papagiannopoulou, E., & Tsoumakas, G. (2018). *Unsupervised keyphrase extraction based on outlier detection.* CoRR, abs/1808.03712.

Papagiannopoulou, E., & Tsoumakas, G. (2020). A review of keyphrase extraction. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *10*(2), e1339. doi:10.1002/widm.1339

Papagiannopoulou, E., Tsoumakas, G., & Papadopoulos, A. N. (2020). *Keywords lie far from the mean of all words in local vector space*. Academic Press.

Rabby, G., Azad, S., Mahmud, M., Zamli, K. Z., & Rahman, M. M. (2020). Teket: A tree-based unsupervised keyphrase extraction technique. *Cognitive Computation*, *12*(4), 1–23. doi:10.100712559-019-09706-3

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory, 1*, 1–20.

Siddiqi, S., & Sharan, A. (2015). Keyword and keyphrase extraction techniques: A literature review. *International Journal of Computers and Applications*, *109*(2).

Singhal, A., Kasturi, R., Sharma, A., & Srivastava, J. (2017). *Leveraging web resources for keyword assignment to short text documents.* arXiv preprint arXiv:1706.05985.

Steinberger, R. (2001). Cross-lingual keyword assignment. *Procesamiento del lenguaje natural, 27,* 273-280.

Suleiman, D., & Awajan, A. (2017). Bag-of-concept based keyword extraction from arabic documents. In *Proceedings of the 2017 8th International Conference on Information Technology* (pp. 863–869). 10.1109/ICITECH.2017.8079959

Sun, P., Wang, L., & Xia, Q. (2017). The keyword extraction of chinese medical web page based on wf-tf-idf algorithm. In *2017 International Conference on Cyber-enabled Distributed Computing and Knowledge Discovery* (pp. 193–198). 10.1109/CyberC.2017.40

Tixier, A., Malliaros, F., & Vazirgiannis, M. (2016). A graph degeneracy-based approach to keyword extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1860–1870). 10.18653/v1/D16-1191

Turney, P. D. (1999). Learning to extract keyphrases from text (Tech. Rep.). National Research Council, Institute for Information Technology, Canada.

Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information Retrieval*, *2*(4), 303–336. doi:10.1023/A:1009976227802

Turney, P. D. (2002). Extraction of keyphrases from text: Evaluation of four algorithms. *Clinical Orthopaedics and Related Research*. https://arxiv.org/abs/cs/0212014

Vega-Oliveros, D. A., Gomes, P. S., Milios, E. E., & Berton, L. (2019). A multi-centrality index for graph-based keyword extraction. *Information Processing & Management*, *56*(6), 102063. doi:10.1016/j.ipm.2019.102063

Wan, X., & Xiao, J. (2008). *Single document keyphrase extraction using neighborhood knowledge* (Vol. 8). AAAI.

Wang, C., & Li, S. (2011). Corankbayes: Bayesian learning to rank under the co-training framework and its application in keyphrase extraction. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 2241–2244). 10.1145/2063576.2063936

Wartena, C., Brussee, R., & Slakhorst, W. (2010). Keyword extraction using word co-occurrence. In *Proceedings of the 2010 Workshops on Database and Expert systems Applications* (pp. 54–58). 10.1109/DEXA.2010.32

Witten, I. H., Paynter, G. W., Frank, E., & Gutwin, C., & NevillManning, C. G. (1999). KEA: practical automatic keyphrase extraction. In *Proceedings of the Fourth ACM Conference on Digital Libraries* (pp. 254–255) 10.1145/313238.313437

Wu, Y.-B., Li, Q., Bot, R. S., & Chen, X. (2005). Domain-specific keyphrase extraction. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 283–284). ACM.

Xiong, A., & Guo, Q. (2019). Chinese news keyword extraction algorithm based on textrank and topic model. In *International Conference on Artificial Intelligence for Communications and Networks* (pp. 334–341). 10.1007/978-3-030-22968-9_29

Yu, Y., & Ng, V. (2018). *Wikirank: Improving keyphrase extraction based on background knowledge.* arXiv preprint arXiv:1803.09000.

Zesch, T., & Gurevych, I. (2009). Approximate matching for evaluating keyphrase extraction. In *Proceedings of the International Conference RANLP-2009* (pp. 484–489). Academic Press.

Zhang, C. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computer Information Systems*, *4*(3), 1169–1180.

Zhang, C., & Xu, H. (2009). Using citation-knn for automatic keyword assignment. In *Proceedings of the 2009 International Conference on Electronic Commerce and Business Intelligence* (pp. 131–134). 10.1109/ECBI.2009.25

Zhang, Q., Wang, Y., Gong, Y., & Huang, X.-J. (2016). Keyphrase extraction using deep recurrent neural networks on twitter. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 836–845). 10.18653/v1/D16-1080

Zu, X., Xie, F., & Liu, X. (2020). Graph-based keyphrase extraction usingword and document embeddings. In *Proceedings of the 2020 IEEE International Conference on Knowledge Graph* (pp. 70–76). 10.1109/ICBK50248.2020.00020

ADDITIONAL READING

Firoozeh, N., Nazarenko, A., Alizon, F., & Daille, B. (2020). Keyword extraction: Issues and methods. *Natural Language Engineering*, *26*(3), 259–291. doi:10.1017/S1351324919000457

Aydın, Ö., Sakallı, F. B., & Mesut, A. Ş. (2020). A Review of Approaches for Keyphrase Extraction. *Journal*. *Fundamental Sciences and Applications*, *26*(1), 1–6.

Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, *56*(6), 102088. doi:10.1016/j.ipm.2019.102088

Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: A review. *EURASIP Journal on Wireless Communications and Networking*, *2017*(1), 1–12. doi:10.118613638-017-0993-1 PMID:29263717

KEY TERMS AND DEFINITIONS

Co-Occurrence: Is a statistical feature to describe the semantic relation of two words based on their occurrence in a text document.

Keyword: Is salient word that best describes content of a document.

Keyword Assignment: Is the process of indexing a document with keywords using external resources.

Keyword Extraction: Is the process of extracting keywords from a text document.

Word-Graph: Is a graph representation of a text document.

ENDNOTE

[1] In literature the terms *keyword* and *keyphrase* are used interchangeably, however some studies assume that keywords are one-word structures while keyphrases are multiple-words structures. In this study we assume both terms refer to the same concept.

Chapter 5
Software Development Knowledge Management System Using Web Portal

Kamalendu Pal
https://orcid.org/0000-0001-7158-6481
City, University of London, UK

Idongesit Williams
https://orcid.org/0000-0002-5398-3579
CMI, Denmark & Aalborg University, Copenhagen, Denmark

ABSTRACT

Software development is a knowledge-intensive practice. Software development teams rely on human resources and systematic approaches to share knowledge on system design. This collaborative knowledge sharing and preserving mechanism is known as "knowledge management" in software industries. In the software development process, coordination of system design functionalities requires knowledge-sharing infrastructure within the team members. Semantic web service computing (SWSC) provides opportunities and value-added service capabilities that global software development team requires to exchange information. This chapter describes the features of an ontology-based web portal framework, called CKIA (Collaborative Knowledge Integration Architecture), for integrating distributed knowledge in a global software development project. The CKIA framework uses a hybrid knowledge-based system consisting of Structural Case-Based Reasoning (S-CBR), Rule-Based Reasoning(RBR), and an ontology-based concept similarity assessment mechanism. A business scenario is used to present some functionalities of the framework.

DOI: 10.4018/978-1-7998-6792-0.ch005

INTRODUCTION

Software system development has transitioned from a predominantly solo activity of designing standalone development activities to a mainly distributed and collaborative production that needs a team-based software developer's contribution. Many software project staff now contribute to multiple projects. Due to this work practice, project boundaries blur, not just in terms of their work and how they design and develop software but also their communication channels and knowledge management practice. This way, software development is a knowledge-intensive practice. People work in software development teams to bundle the man-powers and use the systematic approach to share system design knowledge. This collaborative knowledge sharing mechanism is known as '*knowledge management'* in software industries.

Modern software systems play a vital role in shaping significant social challenges (Pal, 2019). Software is an essential value-adding component of most consumer products (e.g., mobile phones, digital music systems, automobile). Moreover, software systems are also heavily used in the aerospace industry, industrial business process automation, and control systems (Pal, 2020) (Pal & Karakostas, 2020). In these software applications, malfunction or error can cause loss of life or injury, and error-free software is crucial to the safety and wellbeing of people and business. Hence, there is an increasing requirement for applying strict engineering discipline to the development of software systems.

Consequently, software products, like any engineering products, must be verified against their requirements throughout the development process. Different software development process models (e.g., Waterfall, Spiral) have evolved over the decades to accommodate challenges in software development practice. This way, software development process models play a crucial role to provide a systematic and organized approach to software development (Sommerville, 2019). According to Kevin Roebuck (Roebuck, 2012), a traditional Software Development Life Cycle (SDLC) provides the framework for planning and controlling the development or modification of software products, along with the methodologies and process models are used for software development.

The design and development of software is a complex process consisting of many interdependent activities that involve many stages such as inception, initial design, detailed design and development, implementation and testing, operation, maintenance, and retirement. In this way, it also includes requirements analysis, technical development, project management, quality assurance, and customer support activities. Requirement analysis has always been with any human act of design, so it may seem strange that they have been singled out for study in computer science and created a subject area known as requirement engineering.

Requirement engineering is a term used to describe the business processes involved in eliciting, documenting, and maintaining of requirements for a software system. It is about discovering what the users need the system to do for them. It is worth noting that incomplete requirements and lack of user involvement are the two fundamental causes of software development project failure. These issues are related to the failure of inappropriate requirement gathering engagement.

One can define a requirement as "a specification of what should be implemented". There are mainly two types of requirements: (i) functional requirements – what behaviour the system should offer; and (ii) non-functional requirements – a specific property related to quality assurance, time, or cost of development related issues. Requirements are the basis of all software systems. They are a statement of *what* the software system should do and not *how* it should do it.

In requirement elicitation process, building different analysis models is also essential to understand the system under development. Unified modelling language (UML) supports this multiple viewed model-

ling technique by providing different types of diagrams (e.g., analysis class diagram, use case diagram, sequence diagram) for realization of appropriate requirements. There may be many different stakeholders for a given software development system: many users, maintenance engineers, system support staff, sales engineers, and other relevant works. Requirement engineering is about eliciting and prioritizing the requirements these users have for the system. It is a process of negotiation as there are often conflicting requirements that must be balanced.

When trying to make software requirements elicitation and specification, requirement engineers share their experiences; and these experiences (i.e., knowledge) need to be used in the software development process. In a simplistic sense, knowledge is the outcome of perception, realization, rational thinking, experience, or innate reasoning ability. Knowledge is intuitive and exists within people, part and parcel of human reasoning and decision making. Therefore, knowledge is hard to capture in words or understand completely in logical terms. Knowledge is a framework for "evaluating and incorporating new experiences and information". This is the basis for the process, which continues along time (Davenport & Prusak, 1998).

It is vital to have a digital infrastructure to manage software development knowledge, which will help the software development team effectively do their work. This chapter will focus on software requirement engineering (SRE) related knowledge management issues. There are several research works (Wouters, Deridder & Van Paesschen, 2000) (Mayank, Kositsyna & Austin, 2004) (Lasheras et al., 2009) (Kaiya & Saeki, 2006) published in recent years that relates to knowledge management in software requirement engineering. A new era of SRE related management practices is gaining massive interest in knowledge management of software requirements gathering and analysis practice. Such knowledge must include not only the generic knowledge of individual speciality fields of the domain of SRE and the project application domain (e.g., Enterprise Resource Planning – ERP modules – manufacturing, sales, payroll, transportation, and so on) but also knowledge of best practices captured from the previous and similar software development projects.

Requirement engineering business activities typically involve people from at least two distinct fields: (1) business consulting area (e.g., clients and other stakeholders); and (2) software development area (e.g., business engineers, system architects, software project managers). These different groups of people often produce information flows and knowledge exchange that need to be captured in an automated way by which global participants can interact with this newly formed system. This automated system needs to accommodate the different UML models of proposed software development and to critique these models at different business meetings. In other words, this automated environment will help to the gathering, analyzing, and documenting software requirements. It will help the global software development actors managing collaboratively and exchanging their knowledge sharing through this global digital platform.

This global software platform will use the Internet, Intranet, or any other forms of computer networks to connect with the different stakeholders for software development purpose. Its central constituent part is web service-based technology, and web service can strengthen communication and information exchange within a community. Different web service-based 'web portal' infrastructures appeared to provide an open and effective communication forum for their members. A web portal collects and presents relevant information for the community in a simplistic sense, and users can publish and access events or information to the community.

However, standard web service presents many limitations to information access to its users. One of the main limitations is that information search in simple web service application often produces imprecise results, and precise retrieval of information is virtually impossible. Machine processable semantics

of data enables better access mechanism in web service applications. This new semantic enhanced web service is known as semantic web. Ontologies are the backbone technology for Semantic Web, and more generally – for the management of formalized knowledge in the context of distributed systems. In an ontology-based system, information is made better understandable for the computer application, thus assisting end-users to search, extract, interpret and process information efficiently.

Therefore, semantic web technologies can considerably improve the information sharing process by overcoming the problems of standard web portals. This chapter presents the main features of an ontology-based web portal framework, known as CKIA (Collaborative Knowledge Integration Architecture), for integrating distributed business information systems in a global software development project.

The rest of the chapter is organized as follows. Section 2 outlines software requirement engineering related issues. It includes what requirements are and why they are needed in the software development process. This section also describes the steps of the requirement process and briefly explains requirement gathering techniques. Section 3 presents the use of ontologies in the software requirement engineering domain. Section 4 describes some of the issues of knowledge management in software development and highlights its use in this domain. Section 5 illustrates the main constituent parts of the CKIA framework through a business scenario related to a web portal-based application to a virtual software project in the domain of ERP. Section 6 concludes the chapter with concluding remarks and points out directions for future work.

RELATED RESEARCH WORKS

Several researchers have tried to address different issues of global software development. A detailed account of the issues/problems being faced by global software development and their solutions proposed by different researchers is provided in this section. Sangwan and Ros (Sangwan & Ros, 2008) stress that the communication, coordination, and control mechanisms are major issues in global software development. Therefore, the role of an architect is very important to share the common context of the system for all the involved teams in global software development. The research emphasizes on adaptation of Architectural Description Languages (ADL), but this approach suffers a major limitation as conflict among teams may occur in the understanding of ADL due to availability of several ADLs.

Korkala et al. (Korkala et al., 2010) applied traditional and agile methods in global software development. The findings of the study show that agile methods provide better results in global software development. However, the study suffers from the limitation that the data was collected from only one group and other teams were ignored. Agile methods have their own limitations like due to too many meetings with the customers, they may lose their interest. This technique is beneficial in large projects where rapid application is required. But in case of small projects, the project managers may not feel comfortable to maintain a team on customer site for communication and coordination due to budget constraints.

Grechanik et al. (Grechanik et al., 2010) discussed communication and coordination challenges among developers and testers and their impacts on the overall project. The research identifies many issues that can arise due to poor communication and coordination between testers and developers as well as highlighting their impact on project. However, a fitting framework or model is required for the communication and coordination between developers and testers to resolve these issues. Rammasubbu and Balan (Rammasubbu and Balan, 2008) point out certain research directions to enhance the governance scheme for Distributed Software Development projects.

Software engineering (SE) knowledge is dynamic and evolves with technology, organizational culture, and the changing needs of an organisation's software development practices. Kess and Haapasalo (Kess & Haapasalo, 2002) argue that software processes are essentially knowledge processes, structured within a KM framework. Aurum et al. (Aurum et al., 2003) point out that software development can be improved by recognising related knowledge content and structure, as well as appropriate knowledge and engaging in planning activities. Basili et al. (Basili et al., 1994) (Basili et al., 2001) acknowledge that for an organisation to implement the 'Experience Factory' (EF) approach for KM, several potential barriers to success must be overcome. They argue that while the EF is aimed at instituting a learning organisation, it requires a significant investment of time and effort. They stress the need to leverage alternate approaches to distribute knowledge quickly. The 'Answer Garden' approach is depicted as a short-term solution to questions that may not require extended responses.

Johansson et al (Johansson et al., 1999) apply an 'Experience Engine' approach to KM in SE, as a subset of the EF. They list problems identified with the EF approach, such as its experimental nature, the organisational restructuring it prompts as well as its reliance upon an experience base containing a vast amount of written documentation. They assert that experience is best transferred when the receiver is "actually doing something related to the experience being transferred" (Johansson et al., 1999). The researchers claim that written documentation is generally not referred to when problems occur, as well as emphasising the short life span of software engineering knowledge.

Kess and Haapasalo (Kess & Haapasalo, 2002) advocate the use of project reviews to improve software quality. The results of a case study into a telecommunications organisation are disclosed, revealing the centrality of knowledge creation, and sharing to improving the software development process. It is argued that project reviews enable both tacit and explicit knowledge to be managed effectively. Inspection metrics are portrayed as being integral to brainstorming sessions, which in turn deliver feedback to various phases in the software development process.

Dingsøyr et al. (Dingsøyr et al., 2001) provide an insight into problems faced by small to medium organisations in addressing KM in SE. They consider postmortem reviews and experience reports as two approaches suitable for collecting software development knowledge. They conclude that lightweight postmortem reviews perhaps reveal more about software development practices, while experience reports are more suited to client relationships and interaction.

Rus and Lindvall (Rus & Lindyall, 2002) declare organisations must facilitate both formal and informal knowledge sharing between software developers. They assert that KM complements existing approaches to software process improvement, rather than seeking to replace them. KM activities designed to support SE are grouped into three categories: purpose of outputs, scope of inputs and effort required to process inputs. Several options for implementing and using KM systems for SE are advanced, such as expert identification, the creation of KM champions, document management and using predictive modeling to direct decision-making.

SOFTWARE DEVELOPMENT AND REQUIREMENT ENGINEERING

Global software development is an undertaking that requires appropriate planning and coordination of the involved resources. The inception, design, implementation, examination, and maintenance of a software product are a team effort, organized and executed to satisfy the product customer. Following the separation of concerns principle, software life-cycle models distinguish between different phases or

activities in the production of software, linking them by feed-forward and feed-back loops. This separation reduces the complexity of each single phase or activity, however, at the same time poses needs for an efficient and effective project management.

The management of software development projects is a well-known part of management science. But until now purely quantitative and "hard" project management techniques like the critical path method and the project evaluation and review technique have been dominant. With this mainstream approach, only simply structured software development projects can normally be managed. The focus has been on isolated performance indicators. This approach can be characterised as: (i) data driven, (ii) primarily business process-based operative, (iii) focused on limited quantitative objectives, and (iv) developed for "hard" business criteria.

In this section, we investigate one phase in the software development lifecycle, requirements engineering (RE), that benefit particularly from a coordinated functioning. In the requirement phase, global software development team produces the software requirements specification. The requirement specification provides an unambiguous description of what the software application is to be, and it need to include exact conditions for evaluating the final software product to ensure that it does what it is supposed to do.

Requirement elicitation plays an important role in software development. Requirement elicitation is defined as the process of obtaining a comprehensive understanding of stakeholder's requirements. It is the initial and main process of requirements engineering phase. Elicitation process usually involves interaction with stakeholders to obtain their real needs. It helps users to express their need and expectations from the new system. RE is a complex process as it constitutes seeking, determining, learning, acquiring, discovering, and elaborating requirements of potential stakeholder. It is impossible to get all solutions from an individual or a particular group. Based on project type and stakeholders involved, appropriate technique needs to be used for eliciting customer requirements. Some of the well-known requirement techniques are interviews, questionnaires, observation, Joint Application Development (JAD), brainstorming, and so on.

Individual requirements need to be specified in an appropriate way. Thus, the requirements process consists principally of asking and answering questions. After the requirement engineer understand what is wanted and how it will be used, requirement engineer can rewrite the requirements in his/her words and check back with the users to verify that this is what they want. To do an effective job of developing requirements, requirement engineer need the following: (i) an initial statement of the customer's need, and (ii) someone who can explain why the software product is wanted.

In an actual software development project, requirement engineer would need access to an end-user or end-user representative. Without such access, requirement engineer has no way to get reliable answers to many requirements questions. And without reliable requirements information, requirement engineer cannot produce a quality software product. In fact, with poor requirements information, requirement engineer will almost inevitably produce an unusable and often an unwanted software product.

Although the mechanics of producing and inspecting a software requirement specification (SRS) are much the same, most industrial requirements phases involve a great deal of uncertainty and debate and usually take a substantial amount of time. However, the requirement development process is more important. The reason is that during requirements development, requirement engineer review the customer's needs and formulate questions about how the software product is to work. Then requirement engineer discusses these questions with his/her teammates and decide which ones is clear and which questions need extra clarification. In this way, the global requirement engineering team needs collaboration.

Once the initial requirements are gathered, they need to store in a globally accessible repository for future consultations. In addition, changes in requirements are often a problem because users generally cannot know precisely what they need until they try to use the finished software product. They often think that they know, but this knowledge is based on how they do their tasks today. Introducing a new system, however, usually changes the way they work. Thus, as the development work progresses, they begin to appreciate how the product will affect their environment. The better they understand, the more likely they are to think of new functions and features. Thus, the requirements almost always change, and they keep on changing until they are frozen in a product.

The difficult part of the software requirements process is to learn what the users believe they need and to help them define their needs in terms of useful software product functions. Requirement engineer should so this as quickly as you can and make this understanding as specific as possible. The next step is to agree with the users on a software requirement specification that represents requirement engineering team and customer common understanding of what they need. But because these requirements will start to change as soon as requirement engineer agree to them, requirement engineer must manage all the requirements changes. If requirement engineer does not, requirement engineer may never get anything built, and if requirement engineer does, it would not be right.

This raises one key question: how does requirement engineer distinguish between requirements clarifications and significant functional additions? The only way to do this is to start with a precise agreement on what is to be in the product so that requirement engineering team have a mechanism for settling requirement disagreements. Requirement engineering team can best reach this agreement by producing a clear and precise software requirement specification document. The software requirement specification helps software project management team to manage changes, and it is software development company only protection against a customer insisting on a new interpretation of a function software designers have already designed. With a software requirement specification, software developers' words describe how they interpreted the need statement. After the customers have read the software requirement specification and agreed with its contents, software development company can argue that any changes will cost time and/or money. That is why the software requirement specification is important.

In this way, software requirement specification end-ups with legal agreement between software development company and the customer. This Agreement constitutes the whole agreement between the parties and supersedes any previous arrangement, understanding or agreement between them relating to the subject matter of the Agreement. Subject matter often includes service level agreements, and intellectual property rights.

Moreover, in every software development project the software engineers must understand the proposed system specifications and development tools (e.g., case tools, programming languages, databased development software), and company's guidelines and inhouse practice policies. Accordingly, software development is very knowledge-intensive process. To this end, more complex, cognitive structures are required. These are generally denoted as "knowledge". For being able to organize the knowledge beneficially, knowledge management becomes a vital task. The final product is software and every step leading there is or can be realized through a computer. Hence, literally everything is digital. This aspect makes it particularly interesting to study computer-based knowledge management in software development.

KNOWLEDGE MANAGEMENT IN REQUIREMENT ENGINEERING

Most global software companies agree that knowledge is an important asset for success and survival on an increasingly competitive and global market. Knowledge management is not a product, nor a solution that organizations can buy of-the-shelf or assemble from various components. It is a process implemented over a period, which has as much to do with human relationships as it does with business practice and information technology (IT).

Knowledge Management Processes

Alavi and Leidner (2001) describe the knowledge management processes of a company, based on what they call knowledge system, the individuals and groups that share their knowledge in a company, as shown in Figure 1. In this knowledge system the authors elaborate on the different activities related to knowledge management. This whole web of knowledge management activities is constructed on top of the modes of knowledge creation as outlined by Nonaka (1994). On an individual level, every person has tacit and explicit knowledge. The knowledge is being transferred back and forth through the mode's externalization and internalization within the individual or through combination and socialization between individuals. However, similarly to tacit and explicit knowledge of individuals, every group of individuals (e.g., team members) has two types of memory: Episodic and semantic.

A group's semantic memory represents the available explicated knowledge, for example a document on a file server. The explicit knowledge of an individual can be made available for the rest of the group by transferring it to the semantic memory of the group. Also, an individual can increase its explicit knowledge by accessing the group's semantic memory. For this learning from the group's semantic memory, the group's episodic memory is a critical necessity.

An episodic memory represents the collection of shared experiences of the group. Every individual contributes with parts of their tacit knowledge to it. Beyond the interaction of people is the utilization of knowledge. The knowledge application is always based on an individual's tacit knowledge. At the same time, when applying knowledge, the individual learns from that, which feeds back to the tacit knowledge of the individual. Additionally, the application of knowledge can also be based on the semantic memory of a group directly, which feeds back to the group's episodic memory. This system of knowledge sharing among individuals in a group occurs in different areas of a company. Each of these groups then shares their knowledge, via a group-dialogue. The process of managing knowledge involves the following actions:

- Knowledge gathering - acquisition and collection of the knowledge to be managed.
- Knowledge organization and structuring -imposing a structure on the knowledge acquired to manage it effectively.
- Knowledge refinement - correcting, updating, adding, defining knowledge, in short, maintaining knowledge.
- Knowledge distribution **-** brining the knowledge to the professionals who need it.

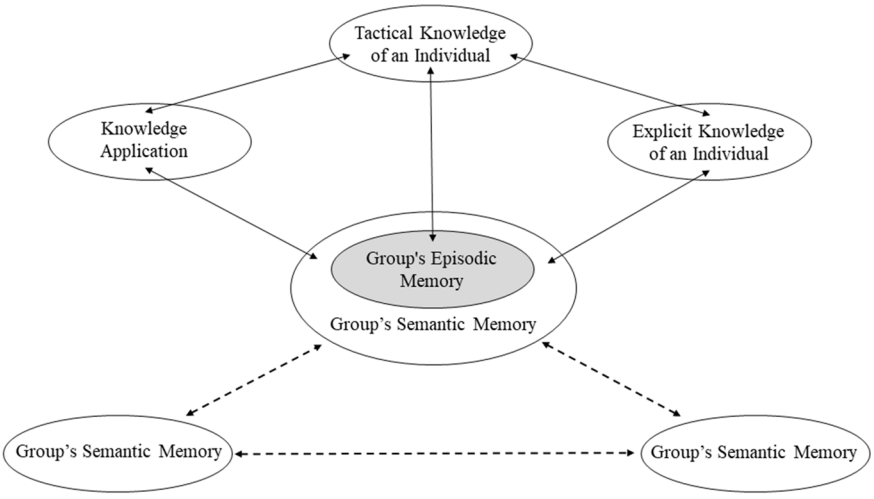
Relevance of Knowledge Management in Software Development

As software development is a very abstract engineering discipline, knowledge management is an important issue. When developing software, a high degree of coordination (Kraut and Streeter, 1995)

and management (Sommerville, 2001; Pressman, 2000) become vital tasks. Because the focus is to solve specific problems, the organization of software projects often differs enormously from one to the other (Mockus et al., 2002). Sveiby (1997) points out that most companies face similar problems when it comes to administrating the own intellectual capital. He explains that employees are usually highly educated and qualified professionals whose everyday job is using their competence to develop software. Their major resource is their knowledge; therefore, they are called knowledge workers. And because knowledge is such an important asset in these companies, so called knowledge organizations, the knowledge management becomes a crucial activity. This counts for software development. Hence, knowledge management for software development companies is a wide field with a variety of different approaches (Aurum et al., 2008).

Rus and Lindvall (2002) describe three aspects of software development to be supported by knowledge management: Core software engineering activities, product & project memory, and learning & improvement. The core activities of software engineering contain the management of documents or competences as well as software re-use.

Figure 1. Knowledge transfer among individuals in a group



With product and project memory the authors refer to the evolution of software, e.g., with the help of systems for version control, change management or design documentation. Finally, the learning and improvement includes a recording of results and experiences. The reason is to learn from that and improve future decisions or activities. The desire to improve in these three areas of concern is the motivation for knowledge management in software development.

To conduct knowledge management successfully, many different approaches are possible and documented. Liebowitz and Megbolugbe (2003) propose a framework for implementing knowledge manage-

ment, which combines an activity cycle of knowledge management levels and the resulting knowledge objects. The different knowledge management levels are conceptualization, reflection, acting and review. Each of these lead to the four knowledge objects: Goals, risks, constraints, or measures. The diversity of dimensions illustrates the complexity of knowledge management.

Information systems that are applied to manage a company's knowledge or to support the managing of a company's knowledge are referred to as knowledge management systems. Alavi and Leidner (2001) conducted a literature review and illustrate six perspectives on knowledge with their implications (see table 1). They identify the differences in perception of knowledge and describe the influence of a perspective on knowledge management and the knowledge management system.

Table 1. Knowledge perspectives and their implications

Perspectives		Implications for Knowledge Management (KM)	Implications for Knowledge Management Systems (KMS)
Knowledge (data and information)	Data is facts, raw numbers. Information is processed / interpreted data. Knowledge is personalized information.	KM focuses on exposing individuals to potentially useful information and facilitating assimilation of information.	KMS will not appear radically different from existing information system (IS) but will be extended towards helping in user assimilation of information.
State of mind	Knowledge is the state of knowing and understanding.	KM involves enhancing individual's learning and understanding through provision of information.	Role of information technology (IT) is to provide access to sources of knowledge rather than knowledge itself.
Object	Knowledge is an object to be stored and manipulated.	Key KM issue is building and managing knowledge stocks.	Role of IT involves gathering, storing, and transferring knowledge.
Process	Knowledge is a process of applying expertise.	KM focuses on knowledge flows and the process of creation, sharing, and distributing knowledge.	Role of IT is to provide link among sources of knowledge to create wider breadth and depth of knowledge flows.
Access to Information	Knowledge is a condition of access to information.	KM focuses organized access to and retrieved of content.	Role of IT is to provide effective search and retrieval mechanisms for locating relevant information.
Capability	Knowledge is the potential to influence action.	KM is about building core competencies and understanding strategic know-how.	Role of IT is to enhance intellectual capital by supporting development of individual and organizational competencies.

In this chapter, we present a knowledge management approach that is grounded in research on knowledge engineering, and semantic web technology. Knowledge engineering is a field that during past few decades – has been concerned with capturing, analysing, organizing, structuring, representing, manipulating, and maintaining knowledge to obtain intelligent solutions for industrial problems. It is therefore no surprise that knowledge engineering techniques (e.g., rule-based reasoning, cased-based reasoning) can be of high value for knowledge management, which is exactly concerned with the issues for global software development.

This awareness is one of the main reasons for the exponential growth of knowledge management in software development industry in the recent decades. Our approach to knowledge management is based on ontologies and makes knowledge assets intelligently accessible to people in global software devel-

opment companies. Most company-vital knowledge resides in the heads of people, and thus successful knowledge management does not only consider technical aspects, but also social ones. In the following sections, we describe an approach to intelligent knowledge management that explicitly use of ontology, and semantic web-based portal technology. The proof of concept is given by a protype development involving knowledge management of a virtual organization.

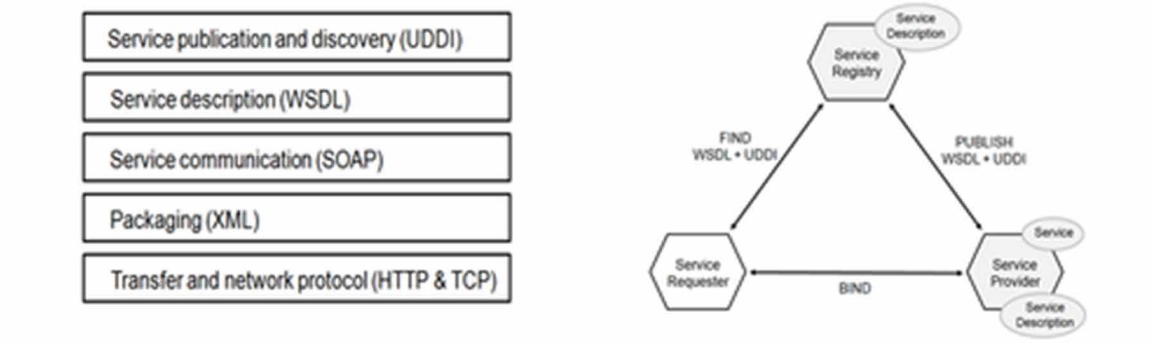
OVERVIEW OF WEB SERVICE AND MOTIVATION

Web services have become the popular choice for the implementation of service delivery systems which are distributed and interoperable. These services are built by a set of core technologies that cater the functionalities for communication, description, and discovery of services. The standards that provide these functionalities are Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL), and Universal Description, Discovery, and Integration (UDDI) (OASIS, 2004). These XML-based standards use common Internet Protocols for the exchange of service requests and responses. (Extensible Markup Language, XML, is a common platform-independent data format across the enterprise) Figure 2A shows the relationship of these technologies as a standard stack for web services, and Figure 2B briefly describes service publishing, service requesting and service finding mechanisms using a simple diagrammatic representation.

When a service provider creates a new service, it describes the service using standard WSDL, which defines the service in terms of the messages to be exchanged between services and how they can be found by specifying the location of the service with an appropriate Universal Resource Locater (URL). To make the service available to consumers, the provider registers the service in a UDDI registry by supplying the details of the service provider, the category of the service, and technical details on how to bind to the service. The UDDI registry will then maintain pointers to the WSDL description and to the service. When a consumer wants to use a service, it queries the UDDI registry to find the service that matches its needs and obtains the WSDL description of that service, as well as the access point of the service. The consumer uses the WSDL description to construct a SOAP message to be transported over HTTP (Hyper Text Transmission Protocol) with which to communicate with the service.

Web services are loosely coupled software components that are published, located, and invoked across a network computing infrastructure. Software-based web services are the building blocks for Service Oriented Computing (SOC), and they can be composed to provide a coarse-grained functionality and to automate business processes. In addition, technological improvements are providing more advanced communication facilities (e.g., online vendor managed inventory replenishment, payment using mobile hand-held devices). Business service facilities are providing more flexibility to its end-users and at the same time managing these business processes are becoming more complex. In SCM, many applications can be built by calling different web services available on the web or corporate intranets. These applications are highly dependent on discovering of correct web services. In particular, the description of web service consists of the technical parameters, constraints and policies that define the terms to invoke the web service. A web service definition needs four important things – *name, description, input,* and *output*. *Name* provides business service name, and it is used as a unique identifier; *description* represents the brief outline of the service; the *input* consists of number of parameters; and the *output* is also represented by a set of service parameters. SOAP based protocol provides the mechanism to exchange structured information in a decentralized and distributed information system.

Figure 2. Diagrammatic representation of web service technologies



In this way, web services aim to use the Web as a world-wide infrastructure for distributed computation purposes to carry out seamless integration of business processes. However, as the set of available web services increases, it becomes crucial to have automated service discovery mechanisms to help in finding services that match a requester's requirement.

Finding appropriate web services depends on the facilities available for service providers to describe the capabilities of their services, and for service requesters to describe their needs in an unambiguous form that is ideally machine-readable. To achieve this objective, ordinary web service descriptions need to be enriched using domain ontology (or semantic mark-up). The next section introduces the concept of semantic annotation mechanisms of web services.

Semantic Web Service and Ontology

Semantic web service is an emerging information technology (IT) paradigm in which the main goal is to realize the development of distributed applications in a heterogeneous business environment. It is built on top of the Web Services technology that provides means for software development that enable dynamic, execution-time discovery, composition, and invocation of Web Services. The underlying technology is extended with rich semantic representations developed in the Semantic Web and with capabilities for automatic reasoning developed in the field of artificial intelligence.

Semantic Web is the new vision of the Web whose main goal is to make Web contents not only human readable but also machine readable and processable. The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The essential property of the World Wide Web is its universality. The power of a hypertext link is that "anything can link to anything". Web technology, therefore, must not discriminate among different types of information. On the contrary, information varies along many axes. One of these is the difference between information produced primarily for human consumption and that produced mainly for machines.

The Semantic Web brings structure to the meaningful contents of Web pages, creating an environment where software agents, roaming from page to page, can readily carry out sophisticated tasks for users. So, for example, such an agent coming to the Web page of the Department of Logistic Science will know not only that the page has keywords such as "Department, people, research activities, publications" but also that Professor Kulkarni works at this Department on Tuesdays, Thursdays, and Fridays and that the script takes a date range in dd-mm-yyyy format and returns his appointment times. These semantics were encoded into the Web page using off-the-shelf software for writing Semantic Web pages along with resources listed on the Company's site.

Like the Internet, the Semantic Web must be as decentralized as possible. Decentralization requires compromises: The Web had to throw away the idea of the total consistency of all its interconnections, ushering in the information message "Error 1001: Not Found" but allowing unchecked exponential growth. Traditional knowledge-representation systems typically have been centralized, requiring everyone to share the same definition of common concepts such as "parent" or "vehicle". But central control is stifling and increasing the size and scope of such a system rapidly becomes unmanageable. On the contrary, in a Semantic Web context, paradoxes and unanswerable questions are accepted as the price that must be paid to achieve versatility.

However, decentralization raises an important issue: different identifiers could be used for the same concept, consequently it is necessary to have a way to know that different identifiers mean or refer to the same thing. For example, consider two databases that refer to *'jack'* and '*coat'* respectively, a program that wants to compare or combine information across the two databases must know that these two terms are being used to mean the same thing. Ideally, the programs must have a way to discover such common meanings for whatever databases it encounters. A solution to this problem is provided using collections of information called *ontologies*. In philosophy, an ontology is a theory about the nature of existence, of what types of things exist, ontology as a discipline studies such theories. Artificial intelligence and web service researchers have co-opted the term for their own jargon. For them, an ontology is a formal definition of the semantics of the resources and relationships. In information management, the term ontology has a particular meaning: "An ontology is an explicit specification of a conceptualization" (Gruber, 1993).

Ontology: Meaning, Usage, and Representation

Ontologies have shown to be the right answer to the Semantic Web vision, by providing a formal conceptualization of a domain that is shared and reused across domains, tasks, and group of people. Their role is to make semantics explicit. Particularly, ontologies describe domain theories for the explicit representation of the semantics of the data. The semantic structuring achieved by ontologies differs from the superficial composition and formatting of information (as data) offered by relational and XML (eXtensible Markup Language) databases. Within a database context, virtually, all the semantic content must be captured in the application logic. On the country, ontologies can provide an objective specification of domain information, by representing a consensual agreement on the concepts and relationships characterizing the way knowledge in that domain is expressed. The result is a common domain of discourse available on the Web, that can be interpreted further by inference rules and application logic. Note that ontologies put no constraints on publishing (possibly contradictory) information on the web, only on its (possible) interpretations.

Ontologies may vary not only in their content but also in their structure and implementation. Building an ontology means different things to different practitioners. Indeed, an ontology could be used for describing simple lexicons or controlled vocabularies to categorically organized thesauri and taxonomies where terms are given distinguishing properties, to full-blown ontologies where these properties can define new concepts and where concepts have named relationships. Ontologies also differ with respect to the scope and purpose of their content. The most prominent distinction is between the domain ontologies describing specific fields of a retail textile product, and upper-level ontologies describing the basic concepts and relationships invoked when information about any domain is expressed in natural language. The synergy among ontologies (exploitable by a vertical application) springs from the cross-referencing between upper-level ontologies and various domain ontologies.

Building an ontology means to distinguish between two different structural components: The Terminological component and the *Assertional* component. The terminological component is roughly analogous to what it is known as the schema for a relational database or XML document. It defines the terms and structure of the ontology's area of interest. The assertional component populates the ontology further with instances or individuals that manifest that terminological definition. To build an ontology, several possible languages could be used, including general logic programming language like Prolog (PROgramming in LOGic). However, in the last few decades, the Web Ontology Language (OWL) has become the de-facto standard for the knowledge representation in the Semantic Web. It is based on a logic thought to be especially computable, known as Description Logics (DLs) (Baader & Nutt, 2003). It is a fragment of First-Order Logic.

Web Service Description of a Business Scenario

In inventory management system, different materials need to be procured for garment manufacturing supply chain management purpose. Material attribute ontology design can be viewed from higher perspectives, such as semantic meanings or logical reasoning. However, in this chapter, focus is on one pragmatic perspectives: as a definition of concepts (or taxonomies) in the domain and related relations. To illustrate the functionalities of domain ontologies, a simple business scenario has been used to demonstrate the activities.

An Example of Business Case for Matchmaking Algorithm

A simple 'retail sales scenario' is used to describe the implemented system functionalities. It envisions an application running on a mobile computer that allows its user to purchase Jackets from an online business. This example considers how a request is matched with service advertised for jacket selling service. An algorithm tries to perform semantic matching for relevant Jersey. The algorithm is shown in ALGORITHM 1, and it takes two ontological concepts, root node ($Root_N$), and the concepts graph (G) as input and computes a semantic similarity between the concepts as output. The part of ontology hierarchy used in this example is shown in Figure 4. Each node of this hierarchy represents a concept. In the experimental comparison, semantic similarity among Jersey, Waistcoat, Sweater, Vest, Cardigan, Pullover, and Jumper are considered.

Semantic Web Services and Case Based Reasoning

Semantic web service initiatives define information systems infrastructure, which enrich the human-readable data on the Web with machine-readable annotations thereby allowing the Web to evolve into the world's biggest information repository which can be accessible from anywhere, anytime throughout the world. To achieve these objectives, one main issue would be the *markup* of web services to make them computer interpretable. Within this markup and semantically enhanced service descriptions, powerful tools should be facilitated across the *web service lifecycle* (Papazoglou, 2012). Web services lifecycle includes automatic web service discovery to find either a web service that offer a service, or a web service that is sufficiently like be used to the current service request; and automatic web service composition and interoperation that involves the run-time service selection, composition, and interoperation of appropriate web services to complete some business activity, given a high-level abstraction of service description.

At the same time, another research community has intensively been working about similarity-based *retrieval* and *adaptation* of past solutions to match new problems: two main aspects in the working semantic web service lifecycle. Case-Based Reasoning (CBR) is one of thriving applied computing community is propagating the idea of finding a solution of a problem based on experience of similar type of problems. CBR systems are an analogical reasoning system (Liang & Konsynski, 1993). It has got diverse applications in many fields, such as classification system for credit card transactions (Reategui & Campbell, 1994) and decision support systems for business acquisitions (Pal & Palmer, 1999). The aim of CBR systems is to infer a solution for a problem in hand from solutions of a set of previously solved similar problems. Attempting to imitate the way human reason, this technique solves problems by using or adopting solutions of previously solved old problems to solve new ones. A CBR system consists of a case base, which is the set of all previously solved cases that are known to the system. The case base can be thought of as a specific kind of knowledge that contains only *cases* and their *solutions*. There are mainly four main stages in CBR life cycle, and they are:

- **Case Representation**: A case is a contextualized piece of knowledge representing an experience. Since a problem is solved by recalling an experience suitable for solving the new problem in hand, the case search and matching processes need to be both effective and reasonably time efficient. Moreover, since the experience from a problem just solved must be retained in some way, these needs also apply to the method of integrating a new case into the case collection. In this way, CBR is heavily dependent on the structure and content of its collection of cases.
- **Case Storage and Indexing**: Cases are assigned indices that express information about their content, then stored in a case library.
- **Case Retrieval**: An important step in the CBR cycle is the retrieval of previous cases that can be used to solve the target problem. Whenever a new problem needs to be solved, the case library index is searched for cases which can be a potential solution. The first phase of this search is case retrieval with the aim of finding the cases which are contextually like the new problem. The case retrieval task starts with a problem description and ends when a suitable matching previous case has been found. Its subtasks are referred to as Identify Features, Search, and Select best possible cases from the system's repository.
- **Case Matchmaking and Use**: Matchmaking performs the comparison between the similar cases and the new request to verify if the possible solution is the one applied to prior cases. The past solutions may be reused, directly or through adaptation, in the current situation.

CBR systems typically apply retrieval and matching algorithms to a case base of past problem-solution pairs. Many successful research and industry results are paving the way of CBR in software development and deployment practice. In recent years, *ontologies, and descriptive logics* (DLs) have become systems of interest for the CBR community. Many multinational organizations (e.g., IBM, British Airways, Volkswagen, NASA, and so on) are using CBR technique for their knowledge intensive business operations (Watson, 1997) (Pal & Campbell, 1997). Moreover, some the real-world CBR applications are taking advantage of the descriptive logics (DLs) reasoning mechanisms for the processes involved in the CBR cycle. However, several different approaches are considered, they all focus on the intuition that the formal semantics and the capabilities of DLs to maintain a terminological taxonomy are interesting properties to measure similarity and to manage a case base.

In CKIA, efforts of the semantic web services lifecycle management and CBR cycle are trying to find synergies between both. Given a certain requirement describing the user goals, automatic web service discovery typically uses a dedicated inference mechanism to answer queries conforming to the logic formalism and the terms defined in the ontology.

PROPOSED SYSTEM FRAMEWORK

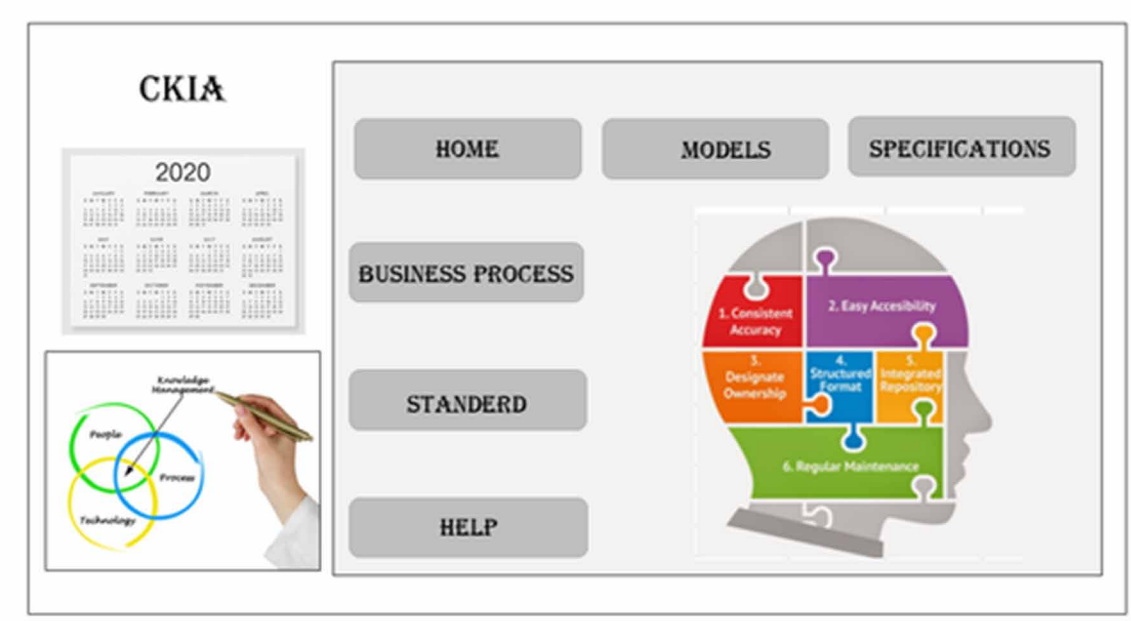
This section briefly presents the overall architecture of CKIA system and illustrates the interplay of the different components. The computational framework of CKIA is shown in Figure 3. It uses a relational similarity assessment measure between implicitly stated concepts. The proposed framework accepts the service consumer request which consists of the requirements of new service (e.g., input, output, precondition, and so on). Next, the user requirement information is parsed for further processing; and final semantically ranked web services are presented to service consumer.

The dynamics of CKIA are as follows:

- Initially, the service repository is populated with semantically enriched web service descriptions for specific application area within software requirement engineering related issues.
- The service requester inputs the service requirements using CKIA's interface.
- The service matchmaking module takes the retrieved cases and the annotation of problem description from the semantic description generator module (within the system framework), run them through a matchmaking algorithm and forward the closest match web service to the requester.

The ontologically enhanced web service descriptions are manually encoded in CKIA service repository. In the processing of ontological concept matching, when dealing with similarity between concepts, it not only considers inheritance (i.e., the relationship between supper-class and subclass) relations, but also considers the distance relationship between concepts. In CKIA, based on the comprehensive consideration of the inheritance relations and semantic distance between concepts, a concept similarity matching method based on semantic distance has been used. The CKIA uses structural case-based reasoning (S-CBR) for services and the relevant ontological concepts storage purpose; and it uses a rule-based reasoning (RBR) for service similarity assessment. The algorithm, as shown in Figure 4, is used to discover semantic web services advertised within CKIA.

Figure 3. Diagrammatic representation of the CKIA



In CKIA, the similarity between two concepts C_i, C_j can be expressed by a number, and it values can fall somewhere between 0 and 1. It may be viewed as a one-directional relation, and its larger values imply higher similarity between the concepts. The concept similarity is described as follows:

Concept Similarity: An ontological concept (C) similarity (∂) is considered as a *relation* and it can be defined as ∂: C x C → [0, 1]. In simple, it is a function from a pair of concepts to a real number between *zero* and *one* expressing the degree of similarity between two concepts such that:

1. $\forall C_1 \in G, \partial(C_1, C_1) = 1$
2. $\forall C_1, C_2 \in G, \quad 0 \le \partial(C_1, C_2) \le 1$
3. $\forall C_1, C_2, C_3 \ \diamondsuit G, \quad IF\ Sim_d(C_1, C_2) > Sim_d(C_1, C_3) THEN \quad \partial(C_1, C_2) < \partial(C_1, C_3)$

The above properties provide the range of semantic similarity function. $\partial(C_i, C_j)$.. For exactly similar concepts the similarity is $\partial(C_1, C_1) = 1$; when two concepts have nothing in common, their similarity is $\partial(C_1, C_2) = 0$. In this way, the output of similarity function should be in closed interval [0, 1]. Here Sim_d represents the semantic distance and (C_1, C_2, C_3) represent three concepts of graph G. In CKIA, the following semantic similarity (∂) function has been used for computation purpose:

$$\partial(C_1, C_2) = \frac{1}{\deg * Sim_d(C_1, C_2) + 1}$$

Where C_1 and C_2 represent two concepts and 'deg' represents the impact degree of semantic distance on semantic similarity, and it should be between $0 < \deg \leq 1$. A weight allocation function is used, as shown below, to compute the semantic similarity between concepts:

$$w(C_m, \ C_n) \ = \ [\ \max(depth(C_m)) \ + \ \frac{OrderNumber(C_n)}{TNodes(G) \ +1} \ +1 \ ]^{-1}$$

Where, $C_m \, and \ C_n$ represent two nodes directly connected, $\max(depth(C_m))$ represents the maximum depth of the node C_m (the depth of the root node is equal to 0 and 1 for the nodes directly connected to the root node and so on), TNodes(G) and OrderNumber(C_n) represent the total number of nodes in concept graph G and the order number of the node (C_n) between their siblings.

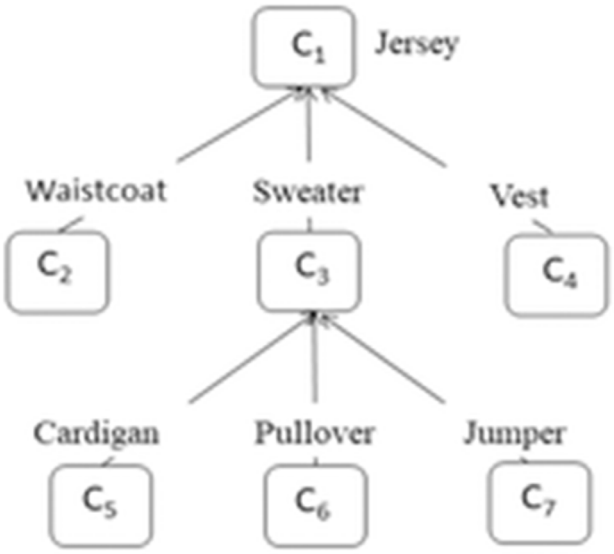
Algorithm I – Algorithm for Semantic Similarity Computation

```
input: two concepts (C1,C2), root node (root), concepts graph (G)
output: semantic similarity value between two concepts
 1:   begin
 2:  if  C1 and C2 are same concept then Simd  = 0
 3:   else
 4:       if C1 and C2 are directly connected then Simd  = w (C1 , C2)
 5:        else
 6:           if idirect path connection exist then
 7:                Spath01 = ShortestPath (G, C1, RootN)
 8:                Spath02 = ShortestPath (G, C2, RootN)
 9:                Simd = w(Spath01) + w(Spath02) - 2*w(CSPath]
 10:           end if
11:            ∂(C1,C2)  =  1 / (deg* Sim_d + 1)
12:        end if
 13:  end if
 14:  return ∂
 15:  end
```

In Table 1, (a) is the result of synonymy similarity (Giunchiglia et al, 2004), (b) tabulates the results of Jian and Conrath similarity (Jiang & Conrath, 1997) results, (c) tabulates the results of path similarity (Varelas et al, 2005), and (d) tabulates the results of the proposed Algorithm-I used in CKIA. In this experiment a suitable value for *deg* parameter is considered.

As shown in Tables 2, synonymy similarity measure can only find similarity between the same concepts, and Jian and Conrath' similarity measure is better than the synonymy similarity measure. The path similarity measure and CKIA's used method are better than the above two methods. The path similarity

Figure 4. The hierarchical concept relationships



measure can find the semantic similarity between concepts, but the similarity score is low. The CKIA's similarity method can also get the semantic similarity between concepts, and the similarity score is high.

CONCLUSION

This chapter reviews some of the business and technology challenges that companies face today in software design and development business activities for information sharing between business partners and describes how several of these difficulties could be overcome with the use of semantic web service. Knowledge management evolved to a sevice management discipline that aims to integrate into the orchestration of the existing management approach. Current knowledge management approaches merge only partly with existing management disciplines like strategic management (in the context of business intelligence), process management (in the context of process-oriented knowledge management) or human resource management (in the context of skill and competence management). In contrast to relatively weak integration on the management level, the technological integration of knowledge – and information management is well advanced.

In this chapter, a software development information sharing architecture between business partners is described using a semantic web service framework. The architecture is based on a hybrid knowledge-based service matchmaking framework, which uses structural case-based reasoning (S-CBR) and rule-based reasoning (RBR). It uses ontology enhanced web service descriptions, object-oriented S-CBR knowledge representation, description logic (DL) for service formalization, and an algorithm to measure ontological concept similarity based on semantic distance. This algorithm considers the inheritance relation between

Table 2. The results of various similarity measures, (a) synonymy similarity, (b) Jian & Conrath similarity, (c) path similarity, (d) the proposed method

	C_1	C_2	C_3	C_4	C_5
C_1	1.00	0.00	0.00	0.00	0.00
C_2	0.00	1.00	0.00	0.00	0.00
C_3	0.00	0.00	1.00	0.00	0.00
C_4	0.00	0.00	0.00	1.00	0.00
C_5	0.00	0.00	0.00	0.00	1.00

(a) Synonymy similarity

	C_1	C_2	C_3	C_4	C_5
C_1	1.00	0.60	0.41	0.97	0.52
C_2	0.42	1.00	0.81	0.60	0.36
C_3	0.97	0.81	1.00	0.68	0.44
C_4	0.60	0.60	0.68	1.00	0.53
C_5	0.52	0.36	0.44	0.53	1.00

(b) Jian & Conrath similarity

	C_1	C_2	C_3	C_4	C_5
C_1	1.00	0.25	0.50	0.20	0.20
C_2	0.25	1.00	0.50	0.33	0.16
C_3	0.50	0.50	1.00	0.25	0.16
C_4	0.20	0.33	0.25	1.00	0.20
C_5	0.20	0.16	0.16	0.20	1.00

(c) Path similarity

	C_1	C_2	C_3	C_4	C_5
C_1	1.0	0.48	0.65	0.51	0.38
C_2	0.48	1.0	0.65	0.51	0.38
C_3	0.65	0.65	1.0	0.71	0.48
C_4	0.51	0.51	0.71	1.0	0.59
C_5	0.38	0.38	0.48	0.59	1.0

(d) The proposed method

concepts and the level of concepts in ontology hierarchy; an experimental evaluation of the proposed algorithm is presented. In this architecture, ontological concepts play an essential role in developing the semantic web as a means for defining share terms in web resources in global software development.

The main benefit of this approach is to be able to model a portal structure with ontology. As shown in previous work, ontologies are suitable to represent consensus knowledge. Precisely that is needed to exchange information with a community of interest and enable automated processing of information items. Conventional portals try to tackle this problem with various structuring methods like content type, view, proprietary metadata elements, and so on. However, this often ends up in user confusion and incompatibility with other portals. Several methodologies exist to model ontologies, which can be used to create a conceptual structure for a web portal in the form of an ontology as the formal representation of a user consensus. The benefit of a semantic web portal is that it can load this initial ontology and build a system out of the box that can satisfy user needs.

Considering the novelty of the semantic features provided by unique semantic web technologies in this chapter, the work done constitutes only the starting point of a more-wide research initiative.

For example, semantic web technologies are appropriately exploited, providing enhanced capabilities concerning standard web portals, such as improved information creation, maintenance, and access. The ontology concepts are very well integrated into creating, maintaining, and accessing steps, guiding, and supporting the user in these tasks. The usability and general assessment of the portal need to be evaluated. The portal also lacks community features to support and improve communication between community members. Indeed, many improvements and open points need to be solved. Further development of successful semantic web portals should focus on the above criteria and international standards, and global software project management legal issues.

REFERENCES

Akkiraju, R., Farrell, J., Miller, J., Nagarajan, M., Schmidt, M., Sheth, A., & Verma, K. (2005). *Web Service Semantics - WSDL-S*. A joint UGA-IBM Technical Note, version 1.0. http://lsdis.cs.uga.edu/projects/METEOR-S/WSDL-S

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *Management Information Systems Quarterly*, *25*(1), 107–136. doi:10.2307/3250961

Antoniou, G., & Harmelen, F. V. (2008). *A Semantic Web Primer*. The MIT Press.

Aurum, A., Daneshgar, F., & Ward, J. (2008, May). Investigating Knowledge Management practices in software development organisations An Australian experience. *Information and Software Technology*, *50*(6), 511–533. doi:10.1016/j.infsof.2007.05.005

Aurum, A., Jeffery, R., Wohlin, C., & Handzic, M. (2003). *Managing Software Engineering Knowledge*. Springer.

Baader, F., & Nutt, W. (2003). *Basic Description Logics, The description logic handbook*. Cambridge University Press.

Basili, V., Costa, P., Lindvall, M., Mendonca, M., & Seaman, C. (2001). An Experience Management System for a Software Engineering Research Organisation. *Proceedings of 26th Annual NASA Goddard Software Engineering Workshop*, 29-35.

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). Experience Factory. In J. J. Marciniack (Ed.), *Encyclopedia of Software Engineering*. John Wiley and Sons.

Bergmann, R., & Schaaf, M. (2003). Structural Case-Based Reasoning and Ontology-Based Knowledge Management: A Perfect Match? *Journal of Universal Computer Science, UCS*, *9*(7), 608–626.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, (May), 34–43.

Bianchini, D., Antonellis, V. A., Melchiori, M., & Salvi, D. (2006). Semantic Enriched Service Discovery. *International Conference in Data Engineering Workshop*, 38.

Cai, M. Zhang, W. Y., Zhang, K. & Li, S. T. (2010). SWMRD: A Semantic Web-based manufacturing resource discovery system for cross-enterprise collaboration. *International Journal of Production Research, 48*(12), 3445-3460.

Cardoso, J. (2006). Discovering Semantic Web Services with and without a Common Ontology Commitment. *IEEE Service Computing Workshop*, 183-190.

Cardoso, J., & Sheth, A. (2003). Semantic e-Workflow Composition. *Journal of Intelligent Information Systems*, *21*(3), 191–225.

Cassar, G., Barnaghi, P., & Moessner, K. (2014). Probabilistic matchmaking methods for automated service discovery. *IEEE Transactions on Services Computing*, *7*(4), 654–666.

Copacino, W., & Anderson, D. (2003). Connecting with the Bottom Line: A Global Study of Supply Chain Leadership and its Contribution to the High-Performance Business. *Accenture, 1*.

Dingsøyr, T., Brede Moe, N., & Nytro, O. (2001). Augmenting Experience reports with Lightweight Postmortem Reviews. *Third International Conference on Product Focused Software Process Improvement, PROFES 2001*.

Domingue, J., Cabral, L., Galizia, S., Tanasescu, V., Gugliotta, A., Norton, B., & Pedrinaci, C. (2008). IRS-III: A Broker-based Approach to Semantic Web Service. *Journal of Web Semantics*, *6*(2), 109–132.

Farrag, T. A., Saleh, A. I., & Ali, H. A. (2013). Toward SWSS discovery: Mapping from WSDL to OWL-S based on ontology search and standardization engine. *IEEE Transactions on Knowledge and Data Engineering*, *25*(3), 1135–1147.

Fensel, D., & Bussler, C. (2002). The Web Service Modeling Framework WSMF. *Electronic Commerce Research and Applications*, *1*(2), 113–137.

Fugate, B., Sahin, F., & Mentzer, J. T. (2006). Supply Chain Management Coordination Mechanisms. *Journal of Business Logistics*, *27*(2), 129–161.

Ganeshan, R., & Harrison, T. P. (1995). An introduction to supply chain management. *Supply Chain Management, 1*. Available from http://silmaril.smeal.psu.edu/misc/supply_chain_

Gibbs, W. (1994, September). Software's chronic crisis. *Scientific American*, *271*(3), 72–81. doi:10.1038cientificamerican0994-86 PMID:8091191

Giunchiglia, F., Shvaiko, P., & Yatskevich, M. (2004). S-Match: an algorithm and an implementation of semantic matching. *Proceedings of 1st European Semantic Web Symposium (ESWS)*, 3053, 61-75.

Grechanik, M., Jones, J. A., Orso, A., & van der Hoek, A. (2010). Bridging Gaps between Developers and Testers in Globally, Distributed Software Development. *Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research*, 149-154.

Grimm, S., Monk, B., & Preist, C. (2006). Matching Semantic Service Descriptions with Local Closed-World Reasoning. *European Semantic Web Conference*, 575-589.

Gruber, T. R. (1993). *A Translation Approach to Portable Ontology Specifications*. Stanford University, Computer Science Department, Knowledge Systems Laboratory, Technical Report KSL 92-71.

Guarino, N., & Giaretta, P. (1995). *Ontologies and Knowledge Base: Towards a Terminological Classification Toward Very Large Knowledge Base: Knowledge Building and Knowledge Sharing*. IOS Press.

Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. *Proceedings of International Conference Research on Computational Linguistics*.

Johansson, C., Hall, P., & Coquard, M. (1999). Talk to Paula and Peter – They are Experienced. *Proceedings of the Workshop of Learning Software Organizations*, 69-76.

Kaiya, H., & Sacki, M. (2006). Using Domain Ontology as Domain Knowledge for Requirement Elicitation. *Proceedings of the 14th IEEE International Requirements Engineering Conference*, 189-198.

Kalakota, R., & Whiston, A. (1997). *Electronic commerce: a manager's guide*. Addison Wesley.

Kaufer, F., & Klusch, M. (2006). WSMO-MX: A Logic Programming Based Hybrid Service Matchmaker. *European Conference on Web Services*, 161-170.

Kess, P., & Haapasalo, H. (2002). Knowledge creation through a project review process in software production. *International Journal of Production Economics*, *80*(1), 49–55.

Kess, P., & Haapasalo, H. (2002). Knowledge creation through a project review process in software production. *International Journal of Production Economics*, *80*(1), 49–55.

Kiefer, C., & Bernstein, A. (2008). The Creation and Evaluation of iSPARQL Strategies for Matchmaking. *European Semantic Web Conference*, 463-477.

Klusch, M., Fries, B., & Sycara, K. (2008). OWLS-MX: A Hybrid Semantic Web Service Matchmaker for OWL-S Services. *Journal of Web Semantics*.

Korkala, M., Pikkarainen, M., & Conboy, K. (2010). A Case Study of Customer Communication in Globally Distributed Software Product Development. *Proceedings of the 11th International Conference on Product Focused Software*, 43-46.

Lajmi, S., Ghedira, C., & Ghedira, K. (2006a). How to apply CBR method in web service composition. In *Second International Conference on Signal-Image Technology and Internet Based Systems (SITI'2006).* Springer Verlag.

Lajmi, S., Ghedira, C., Ghedira, K., & Benslimane, D. (2006b). Web_CBR: How to compose web service via case-based reasoning. *IEEE International Symposium on Service-Oriented Applications, Integration and Collaboration held with the IEEE International Conference on eBusiness Engineering (ICEBE 2006)*.

Laliwala, Z., Khosla, R., Majumdar, P., & Chaudhary, S. (2006). Semantic and Rule Based Event-Driven Dynamic Web Service Composition for Automation of Business Processes. *Proceedings of the IEEE Service Computing Workshop (SCW06)*.

Lambert, D. M., & Cooper, M. C. (2000). Issues in Supply Chain Management. *Industrial Marketing Management*, *29*(1), 65–83.

Lasheras, J., Valencia-Garcia, R., Fernandez-Breis, J. T., & Toval, A. (2009). Modeling Reusable Security Requirements Based on Ontology Framework. *Journal of Research and Practice in Information Technology*, *4*(2), 119–133.

Li, J., Sikora, R., Shaw, M. J., & Woo Tan, G. A. (2006). Strategic analysis of inter-organizational information sharing. *Decision Support Systems*, *42*, 251–266.

Li, L., & Horrocks, I. (2003). A Software Framework for Matchmaking Based on Semantic Web Technology. *International Conference in World Wide Web*, 331-339.

Li, M., Yu, B., Rana, O. F., & Wang, Z. (2008). Grid Service Discovery with Rough Sets. *IEEE Transactions on Knowledge and Data Engineering*, *20*(6), 851–862.

Li, S. H. (2002). *An Integrated Model for Supply Chain Management Practice, Performance and Competitive Advantage* (PhD Dissertation). University of Toledo, Toledo, OH.

Liang, T., & Konsynski, B. R. (1993). Modeling by analogy: Use of analogical reasoning in model management systems. *Decision Support Systems*, *9*, 113–125.

Liebowitz, J., & Megbolugbe, I. (2003, April). A set of frameworks to aid the project manager in conceptualizing and implementing knowledge management initiatives. *International Journal of Project Management*, *21*(3), 189–198. doi:10.1016/S0263-7863(02)00093-5

Martin, D., Burstein, M., Mcdermott, D., Mcilraith, S., Paolucci, M., Sycara, K., Mcguinness, D. L., Sirin, E., & Srinivasan, N. (2007). Bringing Semantics to Web Services with OWL-S. *World Wide Web (Bussum)*, *10*(3), 243–277.

Martin, D., Paolucci, M., McIlraith, S., Burstein, M., McDermott, D., McGunness, D., Barsia, B., Payne, T., Sabou, M., Solanki, M., Srinivasan, N., & Sycara, K. (2004). Bringing Semantics to Web Services: The OWL-S Approach. *Proceeding of First International Workshop Semantic Web Services and Web Process Composition*.

Matskin, M., Maigre, R., & Tyugu, E. (2007). Computational logical semantics for business process language. In *Proceedings of second international conference on Internet and Web applications and services (ICIW 2007)*. IEEE Computer Society.

Mayank, V., Kositsyna, N., & Austin, M. (2004). *Requirements Engineering and Semantic Web, Part II. Representation, Management, and Validation of Requirements and System-Level Architectures, ISR Technical Report*. University of Maryland.

McIlraith, S., & Martin, D. (2003). Bringing Semantics to Web Services. *IEEE Intelligent Systems*, *18*(1), 90–93.

Mockus, A., Fielding, R. T., & Herbsleb, J. D. (2002, July). Two Case Studies of Open-Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, *11*(3), 309–346. doi:10.1145/567793.567795

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37. doi:10.1287/orsc.5.1.14

OASIS. (2004). *Introduction to UDDI: Important Features and Functional Concepts*. Organization for the Advancement of Structured Information Standards.

OMG. (2009). *Business Process Model and Notation*. https://www.omg.org/spec/BPMN/1.2/(2009)

Pal, K. (2019). Markov Decision Theory-Based Crowdsourcing Software Process Model. In Crowdsourcing and Probabilistic Decision-Making in Software Engineering: Emerging Research and Opportunities. IGI Global.

Pal, K. (2020). Framework for Reusable Test Case Generation in Software System Testing. In Software Engineering for Agile Application Development. IGI Global.

Pal, K. & Campbell, J. A. (1997). An Application of Rule-Based and Case-Based Reasoning within a Single Legal Knowledge-Based System. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 28*(4), 48-63.

Pal, K., & Karakostas, B. (2020). Software Testing Under Agile, Scrum, and DevOps. In Agile Scrum Implementation and Its Long-Term Impact on Organizations. IGI Global.

Pal, K., & Palmer, O. (2000). A decision-support system for business acquisition. *Decision Support Systems*, *27*, 411–429.

Paliwal, A. V., Shafiq, B., Vaidya, J., Xiong, H., & Adam, N. (2012). Semantics-based automated service discovery. *IEEE Transactions on Services Computing*, *5*(2), 260–275.

Papazoglou, M. (2012). *Web Services and SOA: Principles and Technology*. Pearson.

Pathak, J., Koul, N., Caragea, D., & Honavar, V. G. (2005). A Framework for Semantic Web Services Discovery. *ACM International Workshop on Web Information and Data Management*, 45-50.

Patil, A., Oundhaka, S., Sheth, A., & Verma, K. (2004). METEOR-S Web service Annotation Framework. *The Proceedings of the Thirteenth International World Wide Web Conference*, 553-562.

Rammasubbu, N., & Balan, R. K. (2008). Towards Governance Scheme for Distributed Software Development Projects. *Proceedings of the 1st International Workshop on Software Development Governance*, 11-14.

Reategui, E. B., & Campbell, J. A. (1995). A Classification System for Credit Card Transactions. In *Advances in Case-Based Reasoning: Second European Workshop (EWCBR-94)*. Springer.

Roger, S. (2000). *Pressman. Software engineering: A Practitioner's Approach (European Adaption)* (5th ed.). McGraw-Hill International.

Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., & Fensel, D. (2005). Web service modeling ontology. *Applied Ontology*, *1*(1), 77–106.

Roman, D., Kopecky, J., Vitvar, T., Domingue, J., & Fensel, D. (2015). WSMO-Lite and hRESTS: Lightweight semantic annotations for Web services and RESTful APIs. *Journal of Web Semantics*, *31*, 39–58.

Rus, I., & Lindvall, M. (2002, June). Knowledge Management in Software Engineering. *IEEE Software*, *19*(3), 26–38. doi:10.1109/MS.2002.1003450

Rus, I., & Lindvall, M. (2002). Knowledge Management in Software Engineering. *IEEE Software*, *19*(3), 26–38.

Sangwan, R. S., & Ros, J. (2008). Architectural Leadership and Management in Globally Distributed Software Development. *Proceedings of the 1st International Workshop on Leadership and Management in Software Architecture*, 17-22.

Shingo, S. (1988). *Non-Stock Production*. Productivity Press.

Skoutas, D., Simitsis, A., & Sellis, T. (2007). A Ranking Mechanism for Semantic Web Service Discovery. *IEEE Congress on Services*, 41-48.

Sommerville. (2001). *Software Engineering* (6th ed.). Addison-Wesley.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1-2), 161–197.

Sveiby. (1997). *The New Organizational Wealth: Managing & Measuring Knowledge-Based Assets*. Berrett-Koehler Publishers, Inc.

Sycara, K., Widoff, S., Klusch, M., & Lu, J. (2002). LARKS: Dynamic Matching Among Heterogeneous Software Agents in Cyberspace. *Autonomous Agents and Multi-Agent Systems*, *5*(2), 173–203.

Sycara, K. P., Paolucci, M., Ankolekar, A., & Srinivasan, N. (2003). Automated Discovery, Interaction and Computation of Semanticmweb Services. *Journal of Web Semantics*, *1*(1), 27–46.

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. G. M., & Milios, E. (2005). Semantic Similarity methods in WordNet and their application to information retrieval on the Web. *Proceedings of the 7th annual ACM international workshop on web information and data management*.

Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S., & Miller, J. (2005). METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Information Technology and Management*, *6*(1), 17–39.

Vrijhoef, R., & Koskela, L. (1999). Role of supply chain management in construction. *Proceedings of the Seventh Annual Conference of the International Group for Lean Construction*, 133-146.

Wang, L., & Cao, J. (2007). Web Services Semantic Searching enhanced by Case Based Reasoning. *18th International Workshop on Database and Expert Systems Applications*.

Wang, P., Jin, Z., Liu, L., & Cai, G. (2008). Building Toward Capability Specifications of Web Services Based on an Environment Ontology. *IEEE Transactions on Knowledge and Data Engineering*, *20*(4), 547–561.

Watson, I. (1997). *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufman.

Wouters, B., Deridder, D., & Van Paesschen, E. (2000). The Use of Ontologies as a Backbone for Use Case Management. *Proceedings of the European Conference on Object Object-Oriented Programming ECOOP, Workshop: Object and Classifications, a Natural Convergence*.

WSDL-S. (2005). http://www.w3.org/Submission/WSDL-S/

Zhang, W. Y., Cai, M., Qiu, J., & Yin, J. W. (2009). Managing distributed manufacturing knowledge through multi-perspective modelling for Semantic Web applications. *International Journal of Production Research*, *47*(23), 6525–6542.

Chapter 6
Knowledge Management in Human–Robot Interaction Approaches and Processes

William Leslie Brown-Acquaye
https://orcid.org/0000-0001-9679-5976
Ghana Communication Technology University, Ghana

Ezer Osei Yeboah-Boateng
https://orcid.org/0000-0002-1355-8586
Ghana Communication Technology University, Ghana

Forgor Lempogo
Ghana Communication Technology University, Ghana

ABSTRACT

Cognitive robots, exhibiting cognitive characteristics and synthesizing knowledge to perform tasks and interacting with humans in both industrial and social settings, have become a big part of modern societies. In this chapter, the authors review the processes and approaches to knowledge management in cognitive robot agents for effective human robot interaction. They present the current state of the art in current robotics technology and human-robot interaction. They state current requirements of cognitive robot agents in human-robot interaction and examine the role of knowledge in human-robot interaction. They finally propose a knowledge management framework for cognitive robots that consist of three main stages: knowledge acquisition and grounding, knowledge representation and knowledge integration, and instantiation into robot architectures.

DOI: 10.4018/978-1-7998-6792-0.ch006

INTRODUCTION HUMAN ROBOT INTERACTION

Robots "experts" also known as cognitive robots, exhibiting cognitive characteristics, synthesizing knowledge to perform tasks, and interacting with humans in various capacities both industrial and social have come to stay. This has brought about the need for formalized tools and processes to acquire, ground, represent and integrate and instantiate knowledge that is used by these cognitive robots agents in sustaining some form of human robot interactions and performing tasks (Lemaignan S., Ros, Mösenlechner, Alami, & Beetz, 2010) (Lemaignan, 2012). According to a United Nations survey on robotics, robot agents can be grouped, into three major categories, namely industrial robotics, professional service robotics, and personal service robotics (U.N. & I.F.R.R., 2002). The categorization is mainly based on the specific use and environment in which the robots work and also represent different technologies and correspond to different evolutionary phases of robotic development and commercialization (Thrun, 2004).

Over the year's developments in robotics have seen the design of robot agents that have become better at interaction with human agents and at planning and performing tasks that are more complex. The main aim of human robot interaction is to achieve some level of competencies, or as some might argue, some cooperation with each other in the performance of tasks. To achieve this however we need to contextualize the processes and challenges of knowledge acquisition, representation, integration and instantiation for service, social and interactive robotic agents.

In this chapter, we seek to review the knowledge management processes in cognitive robot agents used to achieve effective human robot interaction. Current existing literature hardly look at the entirety of the process of knowledge management for cognitive robotic, and so this study seeks to provide in-depth perspectives at all the stages involved in knowledge management for robotics. Areas to be covered in the study include specifying the needs of robots in terms of knowledge acquisition, representation and reasoning techniques, integration, and instantiation in robotic architectures to achieve effective human-robot interaction.

OPEN ISSUES IN LITERATURE

Research in Human–robot interaction goes hand in hand with research into the robot's degree of autonomy, as a determinant in the level of complexity of the interaction that the robot agent can handle and the kinds of tasks it can perform successfully. Automation and interaction levels differ from system to system and are defined by both requirements of the human and the system. (Sheridan, 2012). With systems that have high levels of automation, human robot interaction is minimal and the robot agents in the system carries out most tasks and decision making; whereas in systems with low automation, the human robot interaction is higher and the supervisor is more involved in making decisions while controlling the system through inputs (Sheridan, 2012).

Human robot interaction can be categorized into these 4 general areas of application. (Sheridan, 2016).

The main characteristics of robot agents in human-robot interaction includes their ability to exhibit cognitive skills such as

1. The ability to understand and process dialogue from humans through verbal and non -verbal cues. This takes the form of speech, slang, dialect, intonation etc., for verbal cues and – gaze orientation, postures, gestures, facial expressions etc. for non-verbal cues.

2. The ability to perceive understand and manipulate the environment and potentially anticipate and predict the human intentions and actions to help modulate the environment. This must be done not only from the robot point of view, but also from the other agents' points of view
3. The ability to plan, control and monitor the human agents' actions in furtherance of the stated goal of the interaction

The first type of interaction is what can be described as human supervisory control of robots in performance of routine tasks. Here the human supervisor plans, defines and controls the performance of often routine and highly formalized tasks by robots. These tasks are mostly automated and can be further optimized by using intelligent robotic agents that can learn from the experience gained from performing them over and over and through side by-side interactions with humans in industrial settings. Application can be found in industry including aircraft assembly that has shown examples of cognitive robot agent observing human agents in the performing of tasks, as described in (Gombolay, Huang, & Shah, 2015) (Shah, Wiken, Williams, & Breazeal, 2011). Further examples of human robot interaction can be found in production and assembly lines where robots perform various processes of production with parameters set by the human supervisor, as well as delivery of parts and packages.

The second type of interaction involves the remote supervisory control of robot agents such as unmanned aerial and undersea robotics vehicles, as well as other tasks that take place in hazardous environments and inaccessible areas. Here, the interaction is achieved by remote control with force feedback. Increasingly with such systems new technologies such as haptic feedback over the tactile internet is gaining traction (Ferrell & Sheridan, 1967) (Brown-Acquaye, Arthur, & Forgor, 2019). They have found applications in all sectors of life, including remote search, rescue and recovery, remote exploration (c.f. Mars Rover) and remote clean-up of hazardous materials like oil and other chemical spillages. Other promising areas of application include robotic avatars for use by police and border patrols in surveillance work, as well as in, firefighting, rescue, and military operations (Barnes, Jentsch, Chen, & Redden, 2011), where unmanned aerial vehicles used in military warfare (Burke & Murphy, 2011). Interestingly, Ghana launched a Drone Delivery Service on April 24, 2019 to deliver medical supplies within designated areas in remote, deprived and hard-to-reach areas of the country. The key underlying principle for this type of human robot interaction is that the human supervisor and the robotic agent are geographically separated (Goodrich, Lin, & Morse, 2012).

The third type of interaction is with autonomous automated robot agents where human set certain goals and the robot agents exhibiting high levels of autonomy to perform the task assigned. Example of this include simultaneous localization and mapping systems which provide intelligent robot movement to natural language processing and natural-language generation systems which allow for natural, humanlike interaction which meet well-defined psychological benchmarks. Other applications include autonomous unmanned aerial vehicles used in military warfare (Burke & Murphy, 2011) and human-vehicle interaction in automated driving (Biondi, Alvarez, & Jeong, 2019).

The final type of interaction takes the form of human–robot social interaction, including robot devices to provide entertainment, companionship and assistance for disabled people and people needing emotional or physical assistance. Examples of such robots include rehabilitation robot proposed by (Aggogeri, Mikolajczyk, & O'Kane, 2019) and (Edwin, Garcia-Haro, Jardón, & Balaguer, 2019) which is an example of a robot-aided system implemented in health care. Other examples include nursing robots which provide assistance to elderly people with physical and cognitive disabilities (Feil-Seifer & Matarić, 2011) (Robinson, MacDonald, & Broadbent, 2014) (Fasola & Matarić, 2013). Over the past

decade, human-robot interaction has shown promising outcomes in autism intervention (Begum, Serna, & Yanco, 2016). Remote-controlled disinfection robots were used to disinfect residential areas during the outbreak of the pandemic in Wuhan, China (Hornyak, 2020).

At the heart of this human robot interaction is the concept of a knowledge base, which is a formalized mechanism for organizing and storing knowledge about the environment and the object, terms and Meta knowledge. Based on the acquired knowledge, the interaction takes place and is made available to the robot agent in sustaining an interaction with a human. Understanding how robots and human formulate and apply knowledge and investigating effective ways to represent that knowledge in a machine-processable way is of upmost importance in achieving useful and effective human- robot Interaction (Alili, Warnier, Ali, & Alami, 2009). The way humans and robots acquire and process knowledge varies; it largely depends on the sort of interaction that takes place and the level of automation the robots exhibit in performing tasks. Different types of knowledge are required for different kinds of interaction and this leads to very different modes of managing that knowledge. This is outlined in the framework in the Table 1 below.

Table 1. Levels of automation and knowledge requirements in robot agents

Type of interaction	Level of automation	Type of knowledge and environment
Human supervisory control of robots	Low to medium	Highly formalized knowledge used in performing repetitive task in a well-defined operational environment
Remote supervisory control	Medium to high	Some highly formalized knowledge but also sensor processing and integrating new knowledge in a generally undefined operational environment.
Autonomous automated control	Very high	Predominantly acquiring and ground new unformalized knowledge and integrating them as well as some formalized knowledge in generally undefined environments
Human–robot social interaction	Low to medium	Predominantly acquiring and ground new unformalized knowledge and integrating them as well as some formalized knowledge in highly undefined environments

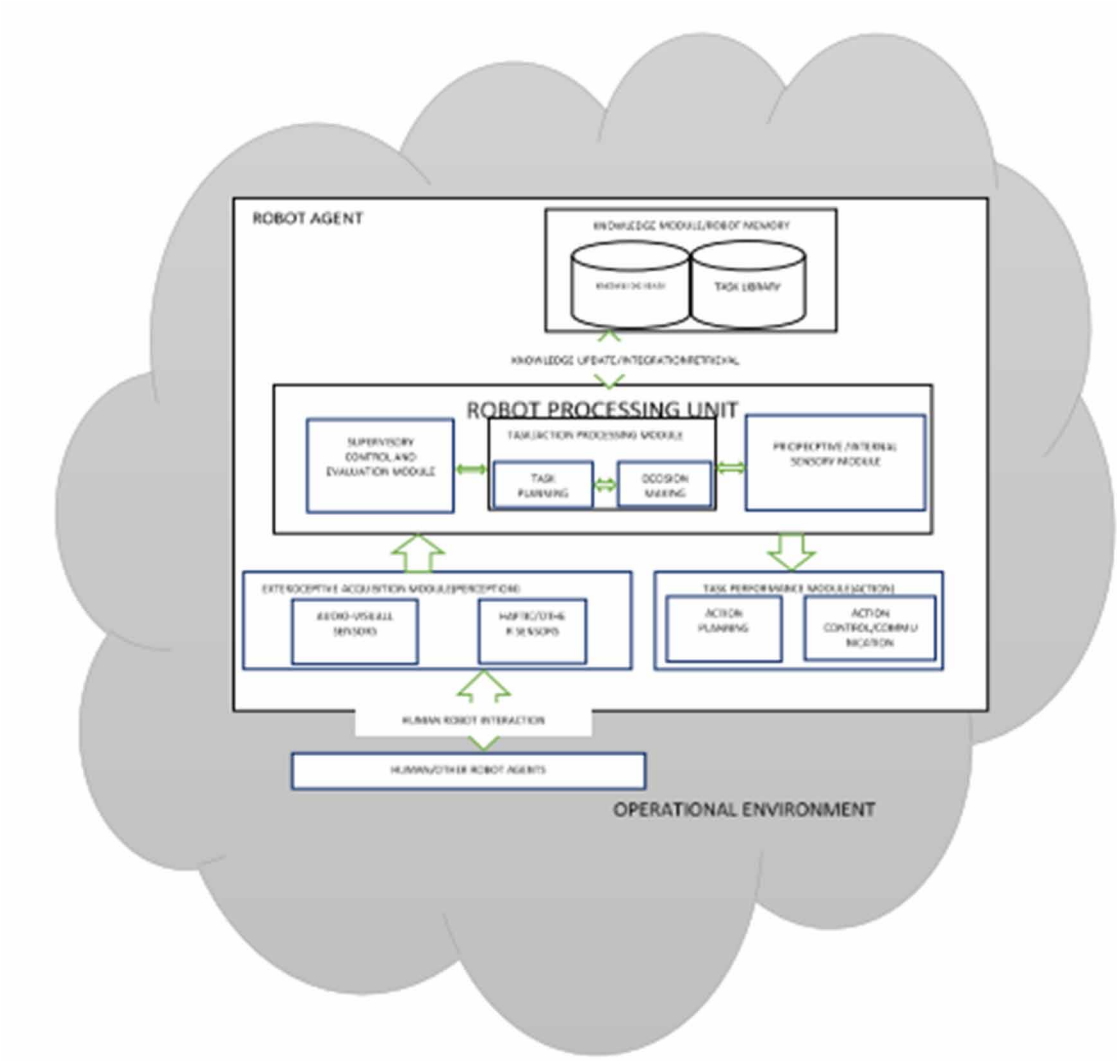
Although there has been diverse and extensive research over the years in Human–robot interaction and the integration of knowledge in the robot architectures, there are still several open issues that currently require active and collaborative research to resolve. One of the most notable open issues in human robot interaction research include a lack of standardized approaches and frameworks for knowledge management for achieving effective and useful human robot interaction (Vasconez, Kantor, & Auat Cheein, 2019) (Lemaignan, Warnier, Sisbot, Clodic, & Alami, 2017). This is largely due to the diverse nature of interaction types and context as well as the varied applications of human robot interaction in different environments and scenarios.

COGNITIVE ROBOT AGENT DESIGN REQUIREMENTS

Cognitive robots are robotic agents (Figure 1) which are endowed with the ability to exhibit intelligent behavior through the process of learning new knowledge and reasoning on how to achieve complex goals and effective interaction within a specified environment (Mutlu, Roy, & Šabanović, 2016). They

do this by acquiring, synthesizing and representing new knowledge and use that knowledge to plan and accomplish tasks and interact with human and other robot agents within their environment scope. To develop cognitive robot agents a clear requirement analysis is required to set the scope of the use of the robot agent within their environment in terms perception, reasoning, learning and interaction.

Figure 1. High level architecture of cognitive robotic agents and its operational environment



The requirements of cognitive robot design and use for human robot interaction usually follow two approaches. The first is the biologically inspired approach where cognitive robots are designed to internally emulate and simulate the actions and capabilities of the human agent. This is grounded in the notion that for useful interaction to happen, robot agents must mimic human actions and speech as much as possible (Adams, Breazeal, Brooks, & Scassell, 2000). The second approach known as the functional

approach develops robot agents that do not internally replicate the human reasoning but only outwardly exhibit the level of reasoning and intelligence (Persson, Laaksolahti, & Lonnqvist,, 2001).

The functional approach to cognitive robot agent design has fewer requirements and allows engineers to design robots using requisites determined by the level of interaction and task performance needed, whilst avoiding unnecessary complexity. This design approach mainly uses existing solutions to prove effectiveness in meeting the requirements of the human robot interaction where the biologically inspired approaches often require novel design approaches. The design approach used is largely dependent on the level of cognitive abilities exhibited by the robot agent and the kind of interaction that it must exhibit. The design requirements can be distilled into three main areas:

1. First, the requirements deal with the ability of cognitive robots to perceive objects, humans, and other robot agents as well as their actions within its operational environment using some localization capabilities. Cognitive robot agents as well should be able to perceive human dialogue, both verbal and non-verbal. These include speech, gestures, facial expressions, body language and posture, etc. The robot agents must also do this within the context of a specified knowledge domain (Fong, Nourbakhsh, & Dautenhahn, 2003).
2. Secondly, the requirement deals with the ability of cognitive robot agents to acquire new knowledge and learn from its environment, human agents, other robot agents as well as the knowledge sources (Feil-Seifer & Mataric, 2011).These require that robot agent have cognitive systems inbuilt into them that aid in the acquisition, representation, integration and instantiation of new knowledge into its architectures.
3. Thirdly, the requirement deals for the design of cognitive robot agents deals with the level of interaction it must possesses. This is largely dependent on the stated goal and actions these robot agents must exhibit in furthering human robot interaction. Cognitive robots must possess natural language processing tools and be able to communicate in natural language a well. This requirement of interaction must happen in real time (Fong, Nourbakhsh, & Dautenhahn, 2003).

THE CONCEPT OF KNOWLEDGE IN HUMAN ROBOT INTERACTION

Knowledge in the context of robotics can be defined as a set of interrelated logical facts that are meaningful to the robot executive controller and can be used in performing task and sustain useful interaction (Lemaignan, Grounding the Interaction: Knowledge Management for Interactive Robots., 2012). This knowledge is generated from low-level data that is generated from sensory devices, and information available from various sources including the internet as well as symbolic facts that have not been contextualized.

Robots receive commands as inputs (both verbal and non-verbal) and plan its course of action to achieve the desired outcome of the interaction. This however, is only achievable based on the knowledge base available to the robot in planning out and executing the interaction. The knowledge is normally in the form of a set of terms commonly referred to as ontology (Lemaignan S., Ros, Mösenlechner, Alami, & Beetz, 2010). Ontology here defines all the knowledge and terms that are used in describing objects in the environment in which the robot agent functions. These include all concepts both verbal and non-verbal objects and the knowledge must be understandable to both the robot agent and human. The knowledge can be classified as low-level features that can be acquired by the robot agent independently

and high-level ones that has to be thought and understood by the robot agent as well as the human agent. The main challenge while designing an architecture for cognitive robot agents and integrating knowledge is making it as explicit as possible. Where explicit means that the knowledge is readily articulated, codified, stored, and accessed by the robot agent. Knowledge in the context of human robot interaction has three main characteristics:

1. The knowledge must be contextualized; where contextualizing means the ability for a cognitive system to connect a fact with a cultural context, an interpretive frame and the set of other facts previously acquired by the agent. This means that the knowledge is based on generally accepted concepts that is acceptable to both robot agent and the human and also that the same concept may have different interpretation ascribed to it in different environments where robot agents function and in different timeframes.
2. The knowledge must be grounded which deals with how to create a link with newly created or acquired knowledge in its syntactic form and its semantics, i.e. its meaning, anchored in the world. Grounding knowledge deals with how the semantic interpretation of a formal symbol system can be made intrinsic to the system (Harnad, 1990).
3. The knowledge must be limited to a domain of validity. Where the domain of validity denotes the scope in which a fact or information can be used in performing a task or interaction in a specified context.

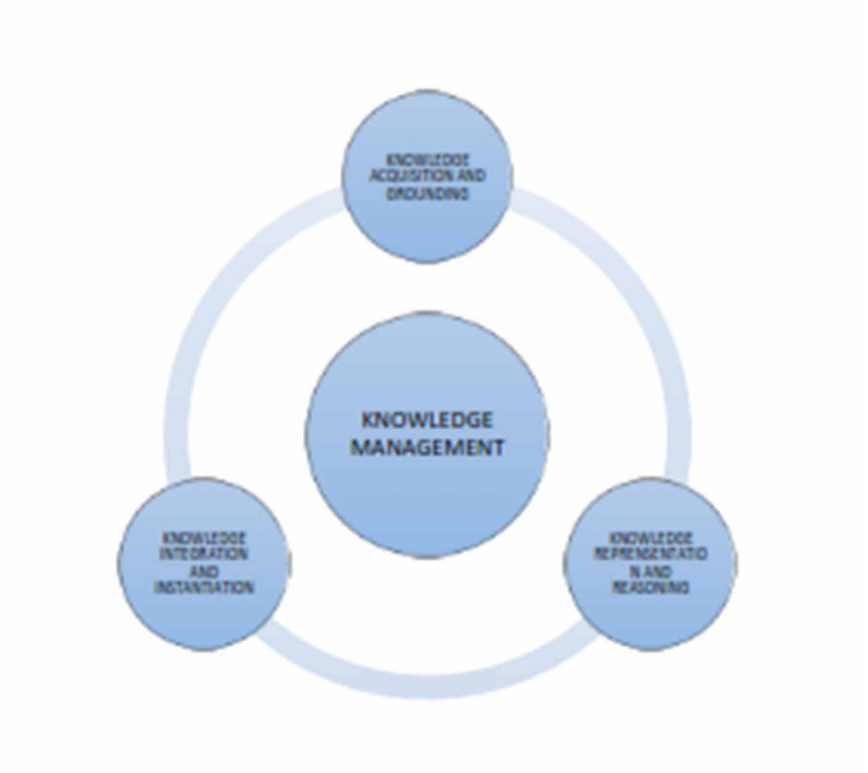
Knowledge is therefore contingent on its ability to represent the links between bits and pieces of knowledge and ground it into the robot agent's knowledge base, as well as an agreed upon context in which to consider that knowledge in general and the scope of validity of the knowledge. These are aided by several knowledge management processes which are reviewed in the ensuing section. The processes include the acquisition, representation, and integration of knowledge in order to maintain the validity and integrity of the knowledge used by robot agents, as well as ensure the logical consistency of the knowledge and the knowledge base.

A FRAMEWORK FOR KNOWLEDGE MANAGEMENT IN HUMAN ROBOT INTERACTION

In this section, we propose a comprehensive novel framework for knowledge management in robot architectures to foster human-robot interaction. Our proposed framework combines various activities that we have identified front extant literature and have combined into a comprehensive approach for acquiring, grounding, representing and integration of knowledge into robot architectures. In the subsections below we delve into each stage of the knowledge management process in our framework and describe the process and activities performed in each stage.

The process of knowledge management in robotic agents from we propose can be distilled into three main sub processes as shown in figure 2 below. These processes are the acquisition and grounding of new knowledge in robot architectures, the representation of the knowledge for effective reasoning in human robot interaction, the integration and instantiation of the knowledge in the robot agent's knowledge base.

Figure 2. Knowledge management processes in robotic agents



Knowledge Acquisition and Grounding

Cognitive robot agents acquire new knowledge from the manifold sensors which they are equipped with. These include audiovisual sensors, haptics sensory devices amongst others. Another very useful source of new knowledge acquisition is from interaction with humans and other robot agents. Here the robot acquires new knowledge by learning about and from others and tapping into their existing knowledge bases. Robot agent do not acquire knowledge directly into the knowledge representation system but use external components are usually required to convert percepts into symbolic facts and to ground them (Tenorth, Jain, & Beetz, 2010). The key challenge in knowledge acquisition is establishing a semantic correspondence between the objects and features in the operational environment of the robot and the symbols used to represent them in the robot architectures, which is also known as grounding the knowledge. This is further complicated by the multi-modal channels in which robotic agents acquire data and information to transform into knowledge. Robotic agents use multi-modal perception capabilities to understand gestures, and natural language and to integrate them into the robot architecture within the context of the operational environment. The interpretation of the perceived content can take place at various levels within the architecture of the robotic agent, but for integration will require some symbolic-level reasoning.

The acquisition and grounding of different types of knowledge from sensory devices, interaction and from other linked knowledge resources is the first key stage in knowledge management in human-robot

interaction. Acquiring and grounding knowledge through sensory devices is the most common and arguably the easiest way that robots acquire knowledge. Exteroceptive (perception of the environment) sensing is the most obvious and largely studied means of knowledge acquisition (Underwood, Hill, Peynot, & Scheding, 2010) (Soter, Conn, Helmut, & Rossiter, 2018) (Morato, Kaipa, Zhao, & Gupta, 2014). Robots use sensing devices like infrared (IR), cameras, laser-scanners, and more recently developed synthetic sensors in perceiving their environments. These synthetic sensors include post-processing to provide higher-level percepts that ease the grounding. The prototypical example of such a device is the Kinect sensor (Lemaignan, 2012).

Interaction with other intelligent agents (humans or robots) is another important source of knowledge acquisition. It relies obviously on some form of sensing (from speech recognition to gesture recognition) but we distinguish it from the previous section because interaction implies a form of communication and learning. We distinguish between two main interaction channels: verbal and deictic

The field of verbal interaction processing for robots spans from pattern-based, constrained sentences recognition to natural, bidirectional, unconstrained verbal communication. Because the interactors, both the robot and the human, are establishing a communication within a shared physical context, the verbal communication channel is complemented by deictic channels, back channels and possibly shared physical experiences.

Deictic (used in the literal meaning of "display, demonstration, reference") interaction is also an established field of research in human-robot interaction. Common deictic forms of communication include attentional focus (via face and gaze tracking) and joint attention, pointing, emotional expressions (based on face expressions, postures, emotional gestures) (Brooks & Breazeal, 2006) (Li, 2012). Another very useful source for new knowledge are Linked Knowledge Resources. This includes the World Wide Web and remote knowledge stores that robot agents have some form of access to.

The concept of grounding deals with the ability of the cognitive robot agents to contextualize the newly acquired knowledge and linking language in the interaction to action in the real world. Grounding the acquired knowledge is concerned with robot agents understanding the knowledge as it applies to the operational environment in which it operates. This is a critical ability that cognitive robot agents must possess in order to learn and apply newly acquired knowledge in human robot interaction.

Knowledge Representation in Human Robot Interaction

The main challenge of knowledge management is the mapping of acquired higher kevel knowledge to the robot's knowledge base for it to interpret and use in furthering useful interactions with human within a specified environment. This challenge deals with the concept of knowledge representation and reasoning.

The knowledge representation process involves presenting information about a specified environment in a format that can be accessed and used by a robot agent to perform tasks and sustain some form of interaction between the human agent and the robot. To do this the robot agent draws on information from various sources including visual, haptic (touch), auditory information gathered by the robot itself as well as the world wide web and data sets. Robot agent as well build on their existing knowledge by accessing past experiential knowledge from other robot agents in a form community of learning (Huang, Bianchi, Liarokapis, & Sun, 2016). All concepts and terms present within the cognitive robot agents' environment should be represented in the robot agent's architecture, so that robot agents can understand, reason, plan and perform tasks and react to other actions on its own.

To achieve effective knowledge representation certain features should be present:

1. the robot agent should be able to represent motions or skills for task planning at the semantic level of understanding.
2. the robot should have a perception system, that allow the integration of input like vision, haptics, natural language, audio amongst others to identify all entities, within the robot agents environment, and to perceive observable states as well as possible obstacles, and to communicate with other robots or humans;
3. the grounding of perception to logical statements, symbols, or values, which can be used to connect concepts and solve problems through logical reasoning (Lemaignan S., Ros, Sisbot, Alami, & Beetz, 2011);
4. Experiences as beliefs should be retained to represent uncertainty, as this allows the robot agent to evaluate other possibilities as it plans its performance of task and is able to tailor the robot's actions to the human beings or operational environment.
5. Newly acquired knowledge and concepts should be integrated so that robot agent can use and learn from them.
6. The representation of a well-defined environment should be a basis for effective interaction.

The key aim of knowledge representation is to represent certain key cognitive capabilities that allow for effective interaction. The first of these cognitive abilities is the representation of roles, which include spatial relations (space, time, and actions) at the symbolic level.

Specifically, space as concept deals with knowledge about environment in which the robot agent operates in terms of its topology and placement. The topology deals with the abstraction of the environment using graphs, with nodes as locations and edges representing the connections between them. Placement on the other hand deals with the positioning of all the entities within the specified environment in absolute and or relative terms.

The representation of time deals with the knowledge and reasoning about time at specific points and the passing of time (actions that span specific durations) at the symbolic level.

Actions performed by both by human and robot agents must be represented in a way that can allow their parameters to be semantically qualified. The roles in that action, like who performs the action (agent), who receives the action (e.g. patient) and the goal of the action need to be represented.

The next cognitive ability that needs to be represented is the context withing which to situate facts that are used in the interaction. Context is important because it set the parameters for the interaction to happen and help to set the domain of validity for certain facts and knowledge. This can be very difficult to represent and current literature defines context mainly as a set of beliefs that initiate a reasoning frame (Lemaignan S., Ros, Sisbot, Alami, & Beetz, 2011).

Linked to the context representation, is the likely support of the concept of parallel beliefs models (or interpretations) that can be independently accessed by the robot agents in their interaction. other cognitive abilities that need to be represented include memory as a pool of facts that are not forgotten by the robot until it is halted commonly denoted as a working memory as wells as introspection which is the ability of the robot agent to self-describe its innate capabilities, and state in the performance of tasks or in state of idle etc.

To be able to represent symbolic knowledge effectively all cognitive robots use some knowledge representation systems that are symbolic systems and are suited for abstract reasoning (Baader, Horrocks, & Sattler, 2008). Examples of the knowledge representation tools for cognitive robots include: KnowRob which is a framework that uses knowledge representation and reasoning methods to acquire

and ground all kinds of knowledge in robot architectures including common sense knowledge, the descriptions of tasks performed by robots, the environment in which it operates and the object(both human and non-human) within it and make the semantic knowledge available robot agent (Tenorth & Beetz, 2017). It represents the knowledge using the OWL ontology language and an extended first-order logic knowledge representation with computable predicates. The process of reasoning in KnowRob is done with an inference procedure beyond the capabilities of OWL to extract information about tasks executions (Tenorth & Beetz, 2013). The framework allows the robot agent to find missing information in somewhat fuzzy instructions in the performance of tasks. Applications of the KnowRob framework can be found in (Saigol, Wang, Ridder, & Lane, 2015) (Constantin & Constantin, 2017) (Yuichiro, Ogawa Yuhei, Akatsuki, & Takahira, 2014) (Kim, Chang, & Suk Choi, 2017).

Other well-known examples of knowledge representation systems include the Grounded Situation Model (GSM) which uses layered data structure that organizes the surrounding world into agents and relations between agents and has been used (Mavridis & Roy, 2006.).

Another well-known knowledge representation tool is CAST (CoSy Architecture Schema) Toolkit which builds cognitive architectures for robots through a set of interconnected sub architectures and represents instead knowledge as unrooted proxies (Hawes, Zillich, & Wyatt., 2007)

The Ontology-based Unified Robot Knowledge (OUR-K) framework is a knowledge representation system which proposes a layered approach to knowledge representation that allows to integrate the grounding process to the knowledge representation process. (Lim, Suh, & Suh., 2011) (Suh, Lim, Hwang, Suh, & Choi, 2007)

The Narrative Knowledge Representation Language (NKRL) is a conceptual language, which provides normalized, pragmatic description of the semantic contents (in short, the "meaning") of NL narrative documents and has bad some application to the robotic field (Sabri, Chibani, Amirat, & Zarri, 2011)

Knowledge Integration

Knowledge representation in robot agent architecture are meaningless if they cannot be synthesized by the robot all the acquired knowledge for performing tasks and decision-making. Knowledge integration is the process of synthesizing multiple knowledge representations in a robot agent into a chain of reasoning, decision-making processes actions. It focuses more on synthesizing the understanding of a given subject from different perspectives.

Knowledge integration involves incorporating newly acquired and represented knowledge into a body of existing knowledge and creating a mechanism for the new knowledge and the existing knowledge to accommodate and interact with each other and to apply within a given context. Knowledge integration happens with both the sensory-motor layers where: passive knowledge repositories that process symbolic facts produced by lower-level sensory-motor layers (push flow) and active knowledge managers that directly query (possibly by polling or on-demand) low-level layers and executive layers where, the knowledge management module needs a tight integration with the decision-making processes as well.

The actual integration techniques vary also widely, from language extensions (like the integration of CRAM with KnowRob (Beetz, Mösenlechner, & Tenorth, 2010)and client-server architectures, to event driven models (SHARY and ORO) (Alami, Warnier, Guitton, Lemaignan, & Sisbot, 2011)Choices at this level have notable consequences on the whole design of the upper control architecture of the robot, in particular regarding its modularity and the ease of addition of new components.

This last of the known as instantiation deals with content of the knowledge base: and refers to the actual, practical knowledge available to the robot agent in planning and performing meaningful tasks and interaction and it comes either from some variant of perception plus grounding, or from knowledge that the developer considers as already meaningful for the robot (Lemaignan S., Ros, Sisbot, Alami, & Beetz, 2011).There are two main approaches to knowledge instantiation namely the a top-down design or a bottom-up design.

A bottom-up approach to knowledge instantiation takes the output of sensors as the primary source of knowledge: instances of objects, agents, and their relations directly result from what is perceived.

The top-down approach on the other hand allows the robot agents to represent integrate and instantiate high level knowledge. The bottom-up and top-down strategies also reflect how the knowledge structure itself is constructed: either by successive classification and refinement of percepts, or from generic categories, typically extracted from standard upper-ontologies.

CONCLUSION

As cognitive robot agents become ubiquitous with time in our daily industrial service and social life, it is imperative that there is a drive for the standardization of the process of knowledge acquisition and management in order to formalize their design and development. In this chapter, we reviewed the processes and approaches to knowledge management in cognitive robot agents for effective human robot interaction. We presented the current state of the art in cognitive robotics technology and human robot interaction and presented some modern applications of robotic technology in both industrial and social settings. We identified the current requirements of modern cognitive robot agents in human robot interaction and examine the role of knowledge in human robot interaction. We finally proposed a knowledge management framework for cognitive robots which consist of three main stages: knowledge acquisition and grounding, knowledge representation and knowledge integration and instantiation into robot architectures and presented a review of some existing knowledge representation tools currently in use.

REFERENCES

Adams, B., Breazeal, C., Brooks, R., & Scassell, B. (2000). Humanoid robots: A new kind of tool. *IEEE Intelligent Systems Magazine*, *15*(4), 25–31. doi:10.1109/5254.867909

Aggogeri, F., Mikolajczyk, T., & O'Kane, J. (2019). Robotics for rehabilitation of hand movement in stroke survivors. *Advances in Mechanical Engineering*, *11*(4). Advance online publication. doi:10.1177/1687814019841921

Alami, R., Warnier, M., Guitton, J., Lemaignan, S., & Sisbot, E. A. (2011). When the robot considers the human. *Proceedings of the 15th International Symposium on Robotics Research*.

Alili, S., Warnier, M., Ali, M., & Alami, R. (2009). Planning and plan-execution for human-robot cooperative task achievement. *19th International Conference onAutomated Planning and Scheduling*.

Baader, F., Horrocks, I., & Sattler, U. (2008). Handbook of Knowledge Representation (Vol. 3). Elsevier.

Barnes, M. J., Jentsch, F. G., Chen, J. Y., & Redden, E. S. (2011). Designing Effective Soldier-Robot Teams in Complex Environments: Training, Interfaces, and Individual Differences. *International Conference on Engineering Psychology and Cognitive Ergonomics, 6781*. 10.1007/978-3-642-21741-8_51

Beetz, M., Mösenlechner, L., & Tenorth, M. (2010). CRAM — A Cognitive Robot Abstract Machine for everyday manipulation in human environments. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 10.1109/IROS.2010.5650146

Begum, M., Serna, R. W., & Yanco, H. A. (2016). Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review. *International Journal of Social Robotics*, *8*(2), 157–181. doi:10.100712369-016-0346-y

Biondi, F., Alvarez, I., & Jeong, K.-A. (2019). Human–Vehicle Cooperation in Automated Driving: A Multidisciplinary Review and Appraisal. *International Journal of Human-Computer Interaction*, *35*(11), 932–946. doi:10.1080/10447318.2018.1561792

Brooks, A. G., & Breazeal, C. (2006). Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. *1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. 10.1145/1121241.1121292

Brown-Acquaye, W., Arthur, J. K., & Forgor, L. (2019). Developing A Framework for A Tactile Internet Enabled Robot Assisted Real-Time Interactive Medical System. *International Conference on Communications, Signal Processing and Networks (ICCSPN)*.

Burke, J. L., & Murphy, R. R. (2011). The safe human–robot ratio. In *Human–robot interaction in future* (pp. 31–49). Ashgate.

Constantin, F., & Constantin, F. (2017). Collaborative Robots and Knowledge Management - A Short Review. *Acta Uiversitatis Cibiniensis Technical Series*, *49*, 143–147.

Edwin, D., Garcia-Haro, J. M., Jardón, A., & Balaguer, C. (2019). Robotics in Health Care: Perspectives of Robot-Aided Interventions in Clinical Practice for Rehabilitation of Upper Limbs". 9 (13): 2586. *Applied Sciences (Basel, Switzerland)*, *9*(13).

Fasola, F., & Matarić, M. J. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, *2*(2), 3–32. doi:10.5898/JHRI.2.2.Fasola

Feil-Seifer, D., & Mataric, M. (2011). Socially assistive robotics. *IEEE Robotics & Automation Magazine*, *18*(1), 24–31. doi:10.1109/MRA.2010.940150

Feil-Seifer, D. J., & Matarić, M. J. (2011). Ethical principles for socially assistive robotics. *IEEE Robotics & Automation Magazine*, *18*(1), 24–31. doi:10.1109/MRA.2010.940150

Ferrell, W. R., & Sheridan, T. B. (1967). Supervisory control of remote manipulation. *IEEE Spectrum*, *4*(10), 81–88. doi:10.1109/MSPEC.1967.5217126

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3-4), 143–166. doi:10.1016/S0921-8890(02)00372-X

Gombolay, M. C., Huang, C., & Shah, J. A. (2015). Coordination of human–robot teaming with human task preferences. *AAAI Fall Symposium Series on AI-HRI.* Retrieved September 2020, from https://interactive.mit.edu/sites/default/files/documents/Gombolay_2015_AAAI_FSS_AI-HRI.pdf

Goodrich, M. A., Lin, L., & Morse, B. S. (2012). Using camera-equipped mini-UAVS to support collaborative wilderness search and rescue teams. *International Conference on Collaboration Technologies and Systems (CTS)*. 10.1109/CTS.2012.6261008

Harnad, S. (1990). The symbol grounding problem. *Physica D. Nonlinear Phenomena*, *42*(1-3), 335–346. doi:10.1016/0167-2789(90)90087-6

Hawes, N., Zillich, M., & Wyatt, J. (2007). BALT & CAST: Middleware for cognitive robotics. *16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2007)*, 998–1003. 10.1109/ROMAN.2007.4415228

Hornyak, T. (2020, March 18). What America can Learn from China's Use of Robots and Telemedicin to Combat the Coronavirus. *Tech Drivers - Robotopia*.

Huang, Y., Bianchi, M., Liarokapis, M., & Sun, Y. (2016). Recent data sets on object manipulation: A survey. *Big Data*, *4*(4), 197–216. doi:10.1089/big.2016.0042 PMID:27992265

Kim, K., Chang, D. S., & Suk Choi, Y. (2017). Ontology based Human-Robot Interaction Knowledge for Intelligent Services. *Workshop on Social Human-Robot Interaction of Service Robots*.

Lemaignan, S. (2012). *Grounding the Interaction: Knowledge Management for Interactive Robots*. INSA de Toulouse.

Lemaignan, S., Ros, R., Mösenlechner, L., Alami, R., & Beetz, M. (2010). ORO, A Knowledge Management Platform for Cognitive Architectures in Robotics. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3548-3553. 10.1109/IROS.2010.5649547

Lemaignan, S., Ros, R., Sisbot, E. A., Alami, R., & Beetz, M. (2011). Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics*, 1–19.

Lemaignan, S., Warnier, M., Sisbot, A. E., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45–69. doi:10.1016/j.artint.2016.07.002

Li, Z. (2012). *Multimodal Human-Robot Interaction in an Assistive Technology Environment*. Monash University.

Lim, G., Suh, I., & Suh, H. (2011). Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Transactions on Systems, Man, and Cybernetics*, 1–18.

Mavridis, N., & Roy, D. (2006.). Grounded situation models for robots: Where words and percepts meet. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 10.1109/IROS.2006.282258

Morato, C., Kaipa, K., Zhao, B., & Gupta, S. K. (2014). Safe Human Robot Interaction by Using Exteroceptive Sensing Based Human Modeling. *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE).*

Mutlu, B., Roy, N., & Šabanović, S. (2016). Cognitive Human–Robot Interaction. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics*. Springer. doi:10.1007/978-3-319-32552-1_71

Persson, P., Laaksolahti, J., & Lonnqvist, P. (2001). Understanding socially intelligent agents - a multilayered phenomenon. *IEEE Transactions on Systems, Man, and Cybernetics*, *31*(5), 349–360. doi:10.1109/3468.952710

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, *6*(4), 575–591. doi:10.100712369-014-0242-2

Sabri, L., Chibani, A., Amirat, Y., & Zarri, G. (2011). Narrative reasoning for cognitive ubiquitous robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop: Knowledge Representation for Autonomous Robots*.

Saigol, Z., Wang, M., Ridder, B., & Lane, D. M. (2015). The Benefits of Explicit Ontological Knowledge-Bases for Robotic Systems. *Towards Autonomous Robotic Systems*, 229-235.

Shah, J. A., Wiken, J., Williams, B., & Breazeal, C. (2011). Improved human–robot team performance using Chaski, a human-inspired plan execution system. *6th ACM/IEEE International Conference on Human–Robot Interaction*. 10.1145/1957656.1957668

Sheridan, T. B. (2012). *Human Supervisory Control*. John Wiley & Sons, Inc. doi:10.1002/9781118131350.ch34

Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. *Human Factors*, *58*(4), 525–532. doi:10.1177/0018720816644364 PMID:27098262

Soter, G., Conn, A., Helmut, H., & Rossiter, J. (2018). Bodily aware soft robots: integration of proprioceptive and. *IEEE International Conference on Robotics and Automation (ICRA)*. 10.1109/ICRA.2018.8463169

Suh, I., Lim, G., Hwang, W., Suh, H., & Choi, J. (2007). Ontology-based multilayered robot knowledge framework (OMRKF) for robot intelligence. *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Tenorth, M., & Beetz, M. (2013). Knowrob: A knowledge processing infrastructure for cognition-enabled robots. *The International Journal of Robotics Research*, *32*(5), 566–590. doi:10.1177/0278364913481635

Tenorth, M., & Beetz, M. (2017). Representations for robot knowledge in the knowrob framework. *Artificial Intelligence*. *Supplement C*, *247*, 151–169.

Tenorth, M., Jain, D., & Beetz, M. (2010). Knowledge Processing for Cognitive Robots. *KI, 24*, 233-240.

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human-Computer Interaction*, *19*(1), 9–24. doi:10.120715327051hci1901&2_2

U.N. & I.F.R.R. (2002). *United Nations and The International Federation of Robotics*. World Robotics.

Underwood, J. P., Hill, A., Peynot, T., & Scheding, S. J. (2010). Error modeling and calibration of exteroceptive sensors for accurate mapping applications. *Journal of Field Robotics*, *27*(1), 2–20. doi:10.1002/rob.20315

Vasconez, J. P., Kantor, G. A., & Auat Cheein, F. A. (2019). Human–robot interaction in agriculture: A survey and current challenges. *Biosystems Engineering*, *179*, 35–48. doi:10.1016/j.biosystemseng.2018.12.005

Yuichiro, M., Ogawa Yuhei, O., Akatsuki, H., & Takahira, Y. (2014). Multi-robot Coordination Based on Ontologies and Semantic Web Service. *Knowledge Management and Acquisition for Smart Systems and Services*, 150-164.

Chapter 7
Digital Technology Advancements in Knowledge Management in Domestic Tour Products in the Russian Federation:
Theoretical and Methodological Aspects

Vardan Mkrttchian
HHH University, Australia

Viacheslav Voronin
https://orcid.org/0000-0001-8114-6383
Moscow State Technological University "Stankin", Russia

ABSTRACT

This chapter discusses the capabilities with problem-oriented digital twin avatars, supply chain, volumetric hybrid, and federated-consistent blockchain use to the nature of knowledge. The goal of this chapter is a theoretical study and practical implementation in the form of basic models and software modules and artificial intelligence algorithms in managing the life cycle of an internal Russian tour product. A laboratory for digitization and management, using multi-agent models of intelligent digital twins-avatars, is created. The purpose of these studies is to solve a scientific problem.

INTRODUCTION

The chapter fits into a moment of operational uncertainty and theoretical redevelopment of the nature of tourism in a society marked by geopolitical turmoil and declining international security, as well as rapid changes at the global level, including the pandemic (COVID-19), which is currently posing new

DOI: 10.4018/978-1-7998-6792-0.ch007

challenges for the sector. Today, it is more relevant and appropriate than ever to reflect on them, with the new, digital energy of blockchain technology, using a fundamental approach to digitalizing the decentralized lifecycle management of the domestic Russian tour product with problem-oriented digital twin avatars, supply chain, volumetric hybrid and federated-consistent blockchain. The goal of the chapter is theoretical study and practical implementation, in the form of basic models and software modules, artificial intelligence algorithms in managing the life cycle of an internal Russian tour product. Why at the State Sochi University, using the scientific potential of the head and responsible executors of the project, the Laboratory for digitalization and management of tour products, using multi-agent models of intelligent digital twins-avatars, is being created, the purpose of these studies is to solve a scientific problem in terms of creating an integrated scientific and methodological approach to modeling and design of monitoring systems, diagnostics and management of distributed cyber-physical objects and processes in the network segments of the Industrial Internet of Things based on the convergence of engineering technologies, data mining and in-depth analysis of processes, predictive modeling and machine learning. The objectives of the research are related to the development of new models, methods and a set of tools for digital transformation of monitoring, diagnostics and management of distributed cyber-physical objects during the transition to the digital economy within the framework of the fourth industrial revolution (Industry 4.0). The results of design research are needed to synthesize the architecture of a new generation of intelligent cyber-physical systems, which represents a multi-agent computing ecosystem. It is designed to provide decision support processes based on monitoring events and processes at distributed cyber-physical objects of the Russian tourism industry. In such systems, there are many cyber-physical objects that receive a huge amount of sensory data that cannot be processed by humans in real time. Currently, there are no ready-made integrated solutions for modeling and designing distributed monitoring and control systems for cyber-physical objects. Despite advances in engineering and knowledge management, the use of this approach for the synthesis of cyber-physical monitoring and control systems is still poorly developed. Such systems work with a variety of distributed cyber-physical objects, which are, in most cases, measuring devices with sensors that collect and accumulate sensor data for transmission to a processing center via a telecommunications network. The results of data analysis are used for predictive modeling of the dynamics of the development of processes at cyber-physical objects and for making management decisions. Cyber-physical monitoring and control systems are needed to automate the decision-making process based on data mining. The relevance of the project is associated with the need to develop and develop new universal mechanisms for modeling and designing cyber-physical systems using new control technologies and in-depth analysis of processes at controlled objects of the Russian tour product. For in-depth analysis of processes, it is necessary to develop automated technologies for collecting, storing and intelligent analysis of data obtained from controlled cyber-physical objects of the Russian tour product. The scientific novelty of design research consists in the creation of a new scientific and methodological approach to the modeling and design of cyber-physical systems for monitoring and controlling distributed objects and processes in the network segments of the Industrial Internet of Things, as well as the methodology for distributed monitoring, diagnostics and recovery of these systems during their operation. Scientific and practical significance lies in the creation of new technologies and software and tools for the synthesis of cyber-physical systems for monitoring and controlling distributed objects and processes on the Internet of Things. For in-depth analysis of processes, it is necessary to develop automated technologies for collecting, storing and intelligent analysis of data obtained from controlled cyber-physical objects of the Russian tour product. The scientific novelty of design research consists in the creation of a new scientific and methodological approach to the modeling

and design of cyber-physical systems for monitoring and controlling distributed objects and processes in the network segments of the Industrial Internet of Things, as well as the methodology for distributed monitoring, diagnostics and recovery of these systems during their operation. Scientific and practical significance lies in the creation of new technologies and software and tools for the synthesis of cyber-physical systems for monitoring and controlling distributed objects and processes on the Internet of Things. A new generation cyber-physical monitoring and control system is implemented in the form of a hyper-converged component-based architecture of a reconfigurable ecosystem, which performs the functions of multi-agent processing of large amounts of sensor data in a computing grid of sensor node controllers based on a fog (edge) computing model.

MATERIALS AND METHODS

The Block chain technology won the interest of many individuals and corporations due to its technological capabilities and scalability for various use cases. This led towards the disruption of traditional internet / intranet, business models alongside services such as the way we conduct business, transactions and managing information in effective and secure ways. These use cases clearly communicate a message for systems and experience designers to get equipped with relevant skills and to keep polishing them as the technology grows. Following the evolution brought by this rapid introduction, the blockchain technology nowadays consists of three types being the 'public', 'private' and 'federated / consorted'. In a nutshell these block chains share similar functionalities. In terms of differences, pretty much they rely on the use cases, permission levels and privacy. The blockchain within the business context brings several advantages such as time saving over work processes, minimizing costs, risk reduction and increase in trust. By learning these values and benefits, we designers will have the ability to foresee how this technology can reshape our clients' businesses notwithstanding the knowledge and confidence we need to guide and proposing right solutions fitting their needs. However for that to happen, the business must have a network of some kind in order to ensure a solid foundation of a good blockchain use case. It is often said that 'with great power comes great responsibility'. This statement is heavily applicable when it comes to this technology. The blockchain restores control and ownership of information back to its rightful owner thus eliminating dependencies on central authorities and third parties.

BACKGROUND

In analysis of world experience in transforming industry into the digital economy and the transition to cyber-physical systems of the industrial Internet of Things shows that work is being carried out in accordance with the concept of the fourth industrial revolution Industry 4.0 as part of the transition to the digital economy and the introduction of Smart Manufacturing, Digital Manufacturing, Internet of Manufacturing, Open Manufacturing technologies.

The foundations of the transition to the digital economy have been described for a long time in the writings of Tapscott, D. (1996). The digital economy: Promise and peril in the age of networked intelligence (Vol. 1). Digital economy - various the way to effective application of technologies (BIM, PLM, CAD, IOT, Smart City, BIG DATA and others). The International Journal of Open Information Technologies, 4 (1)) on the digital economy was devoted to the importance of the integrated application

of technology. It noted, for example, that the joint use of building information modeling (BIM) technology and geographic information technology (GIS) is the path to building systems that work efficiently in the life cycle of the design, construction and operation of a building. This is a conclusion made by leading world experts and practitioners.

A very important element for the digital economy are cyber-physical systems (Kupriyanovsky, V.P., Namiot, D.E., & Sinyagov, S.A. (2016). Cyber-physical systems as the basis of the digital economy. International Journal of Open Information Technologies, 4 (2)). The essence of cyber-physical systems is that they connect the physical processes of production or other other processes (for example, transmission and distribution control of electric power), which require the practical implementation of continuous control in real time, with software and electronic systems. This is a rather little studied topic in Russian literature. At the same time, its importance is obvious.

Cyber physical systems are characterized by multidimensionality, structural and functional complexity. Research in this area lies at the intersection of many disciplines and is still in the initial stage of development. All this determines the need to develop adequate methods for their design. The most promising approach is the model-based approach. A review of the design methods, modeling and integration of CFS, as well as signs of their use are presented. The diversity of descriptions of such systems, consisting of physical, cybernetic and communication parts, requires a certain unified approach to the description, which would allow simple integration of parts into a single whole, reuse of parts, and also support portability and interoperability. These requirements can be satisfied to some extent using the languages UML, SysML, XML. However, the disadvantage of these languages is either a focus on the presentation of syntactic information, or limited semantics. Semantic Web technologies offer much greater opportunities. For example, descriptions based on ontologies make it possible to present semantic information in addition to syntactic information. In addition, there is the possibility of ontological reasoning, which is useful in the analysis, verification and validation of the ontological model.

The models and methods proposed in the project correspond to the world level of scientific research, which is confirmed by the increase in the number of scientific publications in recent years in the field of creation and development of intelligent cyber physical systems such as "Smart Energy Grid", "Smart Road", "Smart city" (Smart City). These systems are developed on the basis of technologies for distributed processing of big data, methods of intellectual analysis and machine learning, M2M interaction of cyber physical devices in the network of the industrial Internet of things, etc. In recent years, many foreign publications have appeared in a similar field.

An example of the smart environment for Smart City is the Smart Road. Therefore, research and development in this area occupy a large place. For such environments, the main component is an intelligent transport infrastructure monitoring system; Smart City Road Monitor by Imagem and Antea Group wins Geospatial World Excellence Award. https://geoinformatics.com/smart-city-road-monitor-wins-award/]. An example of research on the analysis of streaming information and forecasting the development of situations in this area is the development of a road traffic modeling system for forecasting traffic incidents with coordinate reference to digital map layers. To process sensory data in cyber physical systems, it is proposed to use nodes of sensory networks (a layer of "foggy" computing). Wireless sensor networks are now widely used in various fields of human activity, which determines the huge interest in them from scientists and research to create new innovative developments. Leading developers of software and hardware, for example, NXP, offer innovative developments in terms of integrating several network technologies and protocol stacks (technological and network convergence), for example, the connected ZigBee and Bluetooth modules in the sensor node. Examples of publications in the field of wireless

sensor network application research. The fog computing platform is a variation of the cloud computing model, which differs in that the computing nodes for distributed data processing are not servers, but sensor nodes with limited computing and energy resources. Ontologies and ontological models began to be used in modeling cyber physical systems relatively recently proposes a semantic framework based on the use of models for system design, tracking requirements, simulation and assessment of cyber physical systems. In this paper, domain ontologies are used for computing and decision making.

Work is associated with the development of knowledge structures to support the correct ("correct-by-design") design of cyber-physical systems (CFS). This chapter presents a new ontological knowledge base and logical conclusion to support decision-making for cyber-physical systems. This allows the development of deterministic, provable and feasible models of cyber physical systems supported by reliable semantics, which strengthens the approach to the design of cyber physical systems based on model management.

An approach is proposed to develop a digital representation of all information available about an object and from an object, which can be a hardware system or software platform. Digital presentation is based on semantic knowledge presentation formalisms such as RDF, RDF Schema, and OWL. In this paper, we also introduce the concept of the Semantic I4.0 component, which solves the problems of communication and understanding in the scenarios of Industry 4.0 using semantic technologies.

The brief overview given above testifies to the facts of using ontological models for modeling structures of cyber physical systems, contexts of cyber physical systems, verification of projects of cyber physical systems, decision-making in the design process, and presentation of information about an object. At the same time, it can be stated that at the moment there are no works that use dynamic ontological models in which changes would be incorporated into their semantics.

Currently, the ontological approach is used in combination with model-oriented design, and in the future it can completely replace. In general, ontology-based software development refers to new methods by which ontologies can help improve models, techniques and software development processes. Benefits include optimal verification of program code, reusability of artifacts, and increased levels of interaction and integration of software system components.

The direction of development of management methods and forecasting the behavior of cyber physical systems and processes are technologies for extracting knowledge and the intellectual analysis of big data. Big data has a number of properties. High speed of generation and processing of data in real time, which allows you to make the most appropriate decisions regarding specific impacts on the control process. C) Diversity - a wide range of information generated from various sources in various formats, with different structure and size, sorted into different categories related to all aspects of the management process, which allows preparing classifications, groupings, correlations. D) The complexity of processing and data management - the heterogeneity of data that are taken from various sources requires a comprehensive and heterogeneous data processing methodology.

The method of deep analysis of processes (Process mining) can be considered as the development of the method of deep analysis of data (Data mining), but as a result of the use of Process mining, an output is obtained that describes the dynamics of the system. The ancestor of the Process mining method is Wil van der Aalst. This method is actively developing in its group. The basic principles of the Process mining method are described. The process mining method begins to be actively used in monitoring systems.

A new approach to the automatic generation of trust properties obtained as a result of studying and analyzing the system using the Process mining method and comparing with the formal specification of the tested system is proposed.

The use of the Process mining method in an online and traditional audit system is proposed. Moreover, a continuous information monitoring system is proposed, which can identify and prevent risks in the big data environment in advance by monitoring risk factors in organizations and enterprises. The aim of this work is to develop a preliminary risk factor verification system using practical examples of sales audits.

A promising direction is the expansion of the Process mining method of the semantic component. An attempt in that direction was made. This study examines the learning process - how data from various process areas can be extracted, semantically prepared and converted into executable mining formats to support real-time detection, monitoring and improvement of processes. At the same time, the proposed method allows predicting individual patterns / behavior through further semantic analysis of the generated models. The chapter proposes the formalization of the so-called "Semantic Learning Process Mining (SLPM)", technically implemented as a "fuzzy semantic miner" (Semantic-Fuzzy Miner).

An analysis of the literature on the use of the Process mining method in monitoring systems showed that a promising, but not yet explored segment is the extension of the Process mining method of the semantic component based on the use of ontologies.

In conclusion, we note that a review of the scientific literature and existing design studies in this field of knowledge showed that there are a number of problems, which include the lack of adequate mathematical models for the analysis of large sensory data and the in-depth analysis of processes in cyber physical systems, the imperfection of technologies for accounting for hidden patterns in time series and event logs, taking into account the influence of external factors and random fluctuations on the behavior of the cyber physical system, the complexity of automation to accept I make decisions in the process of monitoring and process control in cyber-physical systems. This confirms the relevance of the proposed project, aimed at developing new models and methods for modeling and designing cyber physical systems.

The most widespread research in the field of knowledge related to the development of cyber physical systems and technologies of the Internet of things is conducted in the countries of the USA, EU, China, Japan and Korea by research institutes, universities with the support of public authorities. Let's consider some projects in this area.

1. Intel is developing models, methods and technologies of cloud and GRID computing to increase the efficiency of distributed computing, reduce the complexity of the practical use of cloud solutions and increase the security and stability of distributed computing systems. The main focus is the development of solutions and tools for cloud data centers.
2. Toshiba is working in the field of combining cloud computing, big data and smart technologies to support the work of the energy sector, healthcare and services as part of the implementation of the "Smart Human Community" concept.
3. Cisco is developing Internet of Things technologies using the network infrastructure of multiple sensors and distributed data processing systems based on a fog and cloud computing model. Toshiba and Cisco are conducting joint research on the Internet of Thing, creating the ubiquitous wireless Internet, machine communications, fog and cloud computing in a wide range of devices for managing multimodal transport and smart cities. The main goal is to increase the efficiency of technological processes, productivity and functional capabilities of production, transport and the urban environment. The basis is the development of the ubiquitous Internet (Internet of Everything) by combining the infrastructure of the Cisco Fog Computing network with Toshiba Group technologies in the area of network point management. This will allow you to track and maintain geographically remote devices, to develop distributed computing technologies.

MAIN FOCUS OF THE CHAPTER

Issues, Controversies, Problems, Solutions and Recommendations

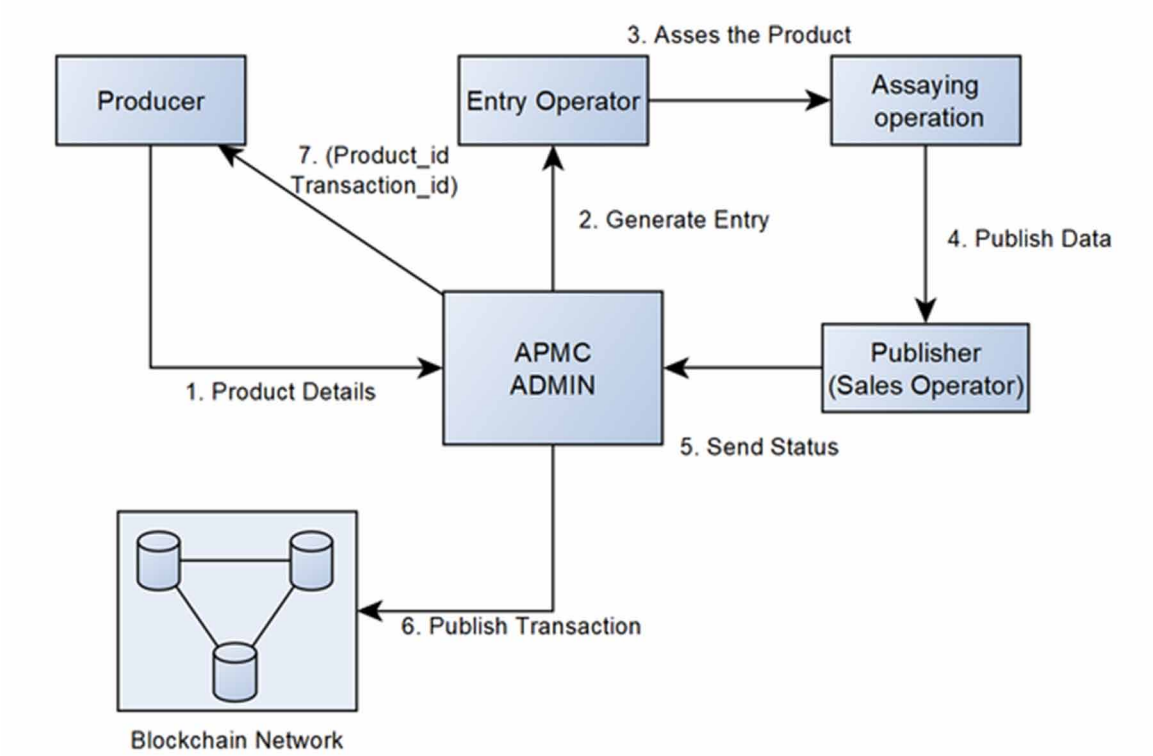
The main scientific problem solved in the research process is associated with the synthesis of a new approach to modeling and designing cyber-physical systems for monitoring, diagnostics and control of distributed objects and processes in the network segments of the Industrial Internet of Things. Optimization of management is one of the central tasks facing the Russian economy. Currently, there is a gradual transition in control systems from simple automation to technologies of "smart" or "smart", and the concept of "digital twin" is central to the development of the corresponding systems. The existing experimental systems have a number of obvious bottlenecks - cyber vulnerability, fragmentation, binding to a specific tour product, etc. The use of intelligent avatar technology for the development of twins can eliminate bottlenecks, which is detailed in three fundamental monographs of the project manager. As a result of the implementation of the proposed scientific research, new scientific, scientific, technical and technological digital solutions will be created that will provide an innovative and digital transformation of product tour management, as well as the development of a typical multi-agent system of intelligent avatars for effective management. The expected results correspond to the world level of scientific research in this area, as they relate to the development of new approaches to monitoring and control of complex geographically distributed cyber-physical systems, which are the basis for the implementation of intelligent cyber-physical systems of a new generation. The results of the project are components for the creation and implementation of new technologies and systems within the framework of the fourth industrial revolution, the transition to a digital economy, digital transformation of management processes and decision support. The public and social significance of the project is determined by the fact that the results of the project are intended for the implementation and development of new intelligent cyber-physical systems for managing the tour product of the Russian Federation (Figure1).

The main scientific problem solved in the research process is related to the synthesis of a new approach to modeling and designing cyber-physical systems for monitoring, diagnostics and management of distributed objects and processes of creating and implementing an internal tour product. in the network segments of the industrial Internet of Things.

The expected results correspond to the world level of scientific research in this area, as they relate to the development of new approaches to monitoring and managing complex geographically distributed cyber-physical systems, which are the basis for the implementation of intelligent cyber-physical systems of a new generation. During the implementation of the project, a new multi-agent approach will be developed. modeling and design of modern cyber-physical systems in the industrial Internet of Things. The results of the project are components for the creation and implementation of new technologies and systems within the framework of the fourth industrial revolution, the transition to a digital economy, digital transformation of management processes and decision support. The results can be used to synthesize new intelligent monitoring systems, which prove the practical significance of design research, as well as the versatility of the developed models, methods and technologies, which in turn will allow the use of tools for creating various geographically distributed cyber-physical systems for creating and implementing an internal tour product (Figure 2).

In our studies, we found out that "signs" of information on the Internet act as symptoms or show problems. The researchers took the hidden information and knowledge available in blogs, and so on, and began to determine the nature of the "signs" in these media in order to make the signs and knowledge

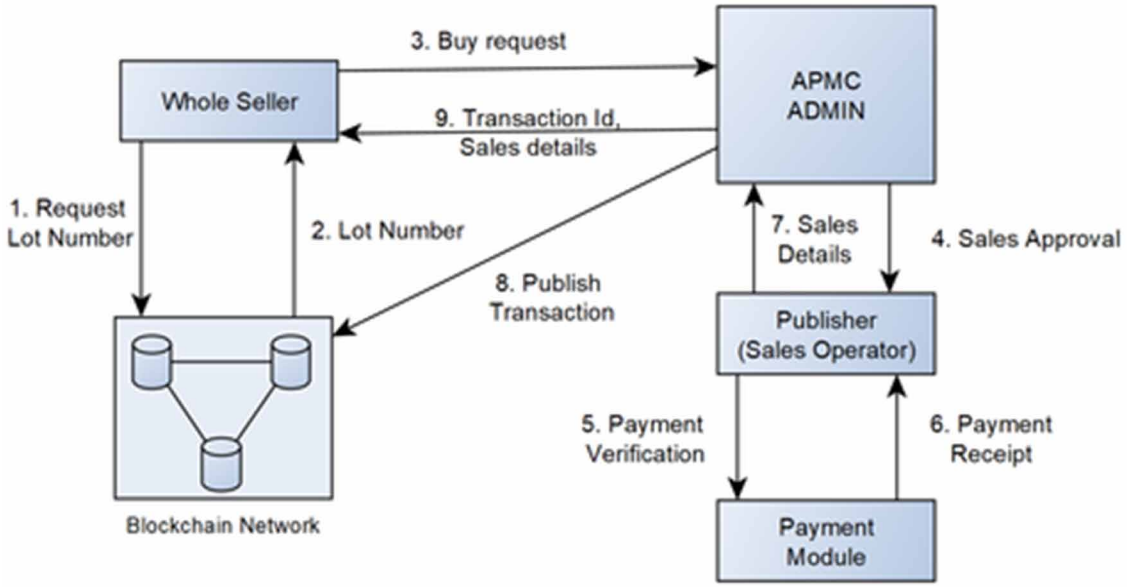
Figure 1. The intended for the implementation and development of new intelligent cyber-physical systems for managing the tour product of the Russian Federation



about them more explicit. Accordingly, these data sources record indicators and potential indicators of activity. This study can be interpreted as the result of the generation of signs from what is sometimes regarded as a data source in "Big Data".

As a result of the search, the context was defined (Dictionary.com) as a set of circumstances or facts that surround a particular event, situation, etc. First, consider the concepts of circumstances or facts. In the case of "big data" or "Internet of things" there would be significant data that could be used to describe circumstances or facts. Such data will precede, occur or occur with a specific event or situation. This would provide essential data as a basis for characterizing the context. Secondly, in the definition, the context is defined around an event or situation. As a result, one approach would be to define a model of the world surrounding specific events or situations that would help define the context. For example, Schilit and Theimer (1994) defined the context as consisting of a location, the identity of nearby objects and people, and changes in these objects. As another example, Schilit et al. (1994) show that an important part of the context / events includes the resources (resources) with whom you (the adjacent agents) and where you are (location). In his work, Day (2001) suggests that these definitions are too specific and that it is difficult to list the entire set of interesting variables a priori. Accordingly, he suggests such a definition "Context" is any information that can be used to describe a situation in essence. An entity is a person, place, or object that is considered relevant to the interaction between the user and the application, including the user and the application it.As a result, the definition of Day (2001) is that

Figure 2. The allow the use of tools for creating various geographically distributed cyber-physical systems for creating and implementing an internal tour product



it is consistent with the use of the essential data available in the "Big Data" and "Internet of Things" to collect and analyze data about a particular event or situation.

Our research has shown that at one level the context is determined by a set of "preceding", "accompanying" and "subsequent" data. But as events occur, the data is generated from different sources. Capturing more different data leads to the fact that more "context" limits. If the context is captured using all available data, then "Big Data" should also be able to provide "Big Context". In this parameter, the "Big Context" will refer to access to significant amounts of data in different formats from different sources, situations or events, but integrated and available for use.

Accordingly, "Big Data" provides an opportunity to provide "more" context than traditional settings. As a result, the recent development in the "Big Data" was to try and integrate the data into context. As an example, as Earnandez notes (Eernandez, 2012), in the case of business transactions, a new perspective is to store "each transaction in the context of business activity, for example, pay, search or purchase, how well it was performed, who initiated it, where the user was, and much more. "

In the case of business settings, there may be theoretical structures or schemes that can facilitate the identification of the relevant variables and the expected relationships between them (for example, O'Leary, 1999). Thus, contextual identification variables are likely to require some consideration of events, situations or settings of interest.

Our research has shown that any discussion of the context in semiotics gives an idea of the context. Eco (1981) in his work emphasizes the importance of contexts, noting that the sign becomes only. Completely meaningful when it is inserted into a broader context" As another example, also noted by Eco (1981), "... I need to look for possible contexts that can make an expression ... understandable and reasonable. The very nature of signs postulates an active role on the part of their translator. " Our research has shown

that semiotics offers terms that make information part of the context: meaningful, understandable and reasonable. "Value" implies that there is a model of how the functions of the world allow us to understand data in both the local and global contexts. "Intelligent" means that the relationship between data and the model can be understood. "Reasonable" indicates that the behavior of the data in the model corresponds to the required parameters. The new designs we have obtained, the Internet of Things and the Internet of Signs, can contribute to the definition of both local and larger contexts. For example, a context can be defined as a set of other "things" within some epsilon of a "thing" of concern. Since the Internet of Things forms a network, classical network approaches can be used to facilitate analysis. Alternatively, "things" can be grouped according to some model. Such models can have several relationships, such as a cascading grouping element. Such cascading groupings can be used to determine local and larger contexts.

FUTURE RESEARCH DIRECTIONS

This study can be expanded in the future in a number of ways. This chapter did not explore the role of business intelligence or the more classic data warehouse. This collection and analysis of data is usually more associated with traditional transaction data. However, data from the "Internet of things" and "large data", which include social media and other forms of unstructured data, can be integrated into business intelligence and data warehouse, including Internet Signs.

CONCLUSION

As a result of the research, key technologies of the of "Industry 4.0" era were identified, their characteristics and role in use were given. Conclusions are made that the introduction of these technologies will favorably affect productivity, revenue growth, employment and investment. In the conclusion the detailed description of various areas of using the Internet of things in activity of the agricultural organizations is resulted. The study allows us to conclude that the digitalization of the agricultural sector will entail the release of better products. In addition, Industry 4.0 will lead to the creation of more flexible systems, the participants of which will exchange information via the Internet, which in turn will significantly increase labor efficiency and reduce costs in production processes.

Digitalization is an absolutely logical process that takes place in all areas of the economy: in marketing, in retail, and in service. Modern information systems and neural networks will be able to analyze more factors and significantly increase the efficiency of any business process. Of course, this also applies to touristic product.

Any agricultural producer in the competitive market has two main tasks: to minimize the cost of production and increase the resulting net revenue, while maintaining product quality at a consistently high level. To solve them, at all stages of the production process must be fully manageable and transparent. For example, you need to clearly, gradually monitor the value chain for each unit of production. For this, a single information space is being created at the agricultural enterprise, where high-tech equipment, analytical and management IT systems non-stop exchange data.

The study showed that using the Internet of things technology can radically change farm management. The introduction of various kinds of sensors and sensors, the introduction of big data technology, as well

as the use of unmanned aerial vehicles and self-propelled tractors and machines today can transform traditional farms into new generation farms, Smart farms.

Blockchain solutions are applicable for different touristic product operations where farmers can directly interact with regulatory authorities. This will help them to gain information on prices, weather, and market trends. They can also use the services to develop co-op farms where the secure and transparent platform will benefit them. The technology will be helpful for nearly all farmers, either small or large-scale. Subsequently, the technology will be used to track food through supply chain-based contracts.

ACKNOWLEDGMENT

The reported study was funded DFG-RFBR cooperation according by Joint German-Russian Project Nº 21-57-12002 Avatar-Based Control Technologies for Decision Making in E-Learning and Cyber Security Systems, was funded RFBR cooperation according by Joint Russian-Armenia Project Nº 20-510-05007 Development of the fundamental principles of human capital management in the Eurasian Economic Union through the introduction of digital intelligent systems to support the decision-making process in the conditions of economic instability (the example of the Russian Federation and the Republic of Armenia), and was funded by Russian Foundation for Basic Research (RFBR) according to the project Nº 18-010-00204. The results of parts «Materials and methods “, ” In analysis of world experience in transforming industry into the digital economy and the transition to cyber-physical systems of the industrial Internet of Things shows that work is being carried out in accordance with the concept of the fourth industrial revolution Industry 4.0 as part of the transition to the digital economy and Convergent information-analytical platform” were obtained within the RSF grant (project No. 21-71-30007).

REFERENCES

Abubakar, I., Khalid, S. N., Mustafa, M. W., Shareef, H., & Mustapha, M. (2017). Application of load monitoring in appliances' energy management—A review. *Renewable & Sustainable Energy Reviews*, *67*, 235–245. doi:10.1016/j.rser.2016.09.064

Babu, V. S., Kumar, U. A., Priyadharshini, R., Premkumar, K., & Nithin, S. (2016). An intelligent controller for smart home. In *Proceedings of international conference on advances in computing, communications and informatics (ICACCI)* (pp. 2654–2657). IEEE.

Baliga, J., Ayre, R., Hinton, K., & Tucker, R. (2011). Energy consumption in wired and wireless access networks. *IEEE Communications Magazine*, *49*(6), 70–77. doi:10.1109/MCOM.2011.5783987

Blanco-Novoa, Ó., Fernández-Caramés, T. M., Fraga-Lamas, P., & Castedo, L. (2017). An electricity price-aware open-source smart socket for the internet of energy. *Sensors (Basel)*, *17*(3), 643. doi:10.339017030643 PMID:28335568

Caivano, D., Fogli, D., Lanzilotti, R., Piccinno, A., & Cassano, F. (2018). Supporting end users to control their smart home: Design implications from a literature review and an empirical investigation. *Journal of Systems and Software*, *144*, 295–313. doi:10.1016/j.jss.2018.06.035

Guo, X., Shen, Z., Zhang, Y., & Wu, T. (2019). Review on the application of artificial intelligence in smart homes. *Smart Cities*, *2*(3), 402–420. doi:10.3390martcities2030025

Kailas, A., Cecchi, V., & Mukherjee, A. (2012). A survey of communications and networking technologies for energy management in buildings and home automation. *Journal of Computer Networks and Communications*, *2012*, 1–12. doi:10.1155/2012/932181

Kopytko, V., Shevchuk, L., Yankovska, L., Semchuk, Z., & Strilchuk, R. (2018). Smart home and artificial intelligence as environment for the implementation of new technologies. *Traektoriâ Nauki = Path of Science, 4*(9), 2007–2012.

Leslie, P., Pearce, J. M., Harrap, R., & Daniel, S. (2012). The application of smartphone technology to economic and environmental analysis of building energy conservation strategies. *International Journal of Sustainable Energy*, *31*(5), 295–311. doi:10.1080/1478646X.2011.578746

Li, M., & Lin, H. J. (2015). Design and implementation of smart home control systems based on wireless sensor networks and power line communications. *IEEE Transactions on Industrial Electronics*, *62*(7), 4430–4442. doi:10.1109/TIE.2014.2379586

Mkrttchian, V. (2019a). New Tools for Cyber Security Using Blockchain Technology and Avatar-Based Management Technique. In *Machine Learning and Cognitive Science Applications in Cyber Security* (pp. 105–122). IGI Global., doi:10.4018/978-1-5225-8100-0.ch004

Mkrttchian, V. (2019b). Machine Learning With Avatar-Based Management of Sleptsov Net-Processor Platform to Improve Cyber Security. In *Machine Learning and Cognitive Science Applications in Cyber Security* (pp. 139–153). IGI Global.

Mkrttchian, V. (2020). Human Capital Management in the Context of the Implementation of Digital Intelligent Decision Support Systems and Knowledge Management: Theoretical and Methodological Aspects. In Knowledge Management, Innovation, and Entrepreneurship in a Changing World (pp.123-148). IGI Global. doi:10.4018/978-1-7998-2355-1.ch006

Mkrttchian, V. (2021a). Avatars-Based Decision Support System Using Blockchain and Knowledge Sharing for Processes Simulation a Natural Intelligence: Implementation of the Multi Chain Open Source Platform. *International Journal of Knowledge Management*, *17*(1), 5. Advance online publication. doi:10.4018/IJKM.2021010105

Mkrttchian, V., & Aleshina, E. (2017). *Sliding Mode in Intellectual Control and Communication: Emerging Research and Opportunities*. IGI Global. doi:10.4018/978-1-5225-2292-8

Mkrttchian, V., & Chernyshenko, S. (2021b). Digital Intelligent Design of Avatar-Based Control with Application to Human Capital Management. *International Journal of Human Capital and Information Technology Professionals*, *12*(1), 2. Advance online publication. doi:10.4018/IJHCITP.2021010102

Nagy, Z., Yong, F. Y., & Schlueter, A. (2016). Occupant centered lighting control: A consumer study on balancing comfort, acceptance, and energy consumption. *Energy and Building*, *126*, 310–322. doi:10.1016/j.enbuild.2016.05.075

Peffer, T., Blumstein, C., Culler, D., Modera, M., & Meier, A. (2015). Software-defined solutions for managing energy use in small to medium sized commercial buildings (No. EE0006351). Berkeley, CA: Regents of the University of California.

Philips Lighting Catalog. (2014). *Document Number 919002151397*. Amsterdam: Koninklijke Philips Electronics N.V.

Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, *57*(10), 2266–2279. doi:10.1016/j.comnet.2012.12.018

Shie, M. C., Lin, P. C., Su, T. M., Chen, P., & Hutahaean, A. (2014). Intelligent energy monitoring system based on ZigBee-equipped smart sockets. In *Proceedings of IEEE international conference on intelligent green building and smart grid (IGBSG)* (pp. 1–5). IEEE. 10.1109/IGBSG.2014.6835281

Singh, P. P., Khosla, P. K., & Mittal, M. (2019). Energy conservation in IoT-based smart home and its automation. In M. Mittal, S. Tanwar, B. Agarwal, & L. M. Goyal (Eds.), *Energy conservation for IoT devices. Studies in systems, decision and control* (Vol. 206, pp. 155–177). Springer. doi:10.1007/978-981-13-7399-2_7

Stojkoska, B. L. R., & Trivodaliev, K. V. (2017). A review of Internet of Things for smart home: Challenges and solutions. *Journal of Cleaner Production*, *140*, 1454–1464. doi:10.1016/j.jclepro.2016.10.006

Tsai, K. L., Leu, F. Y., & You, I. (2016). Residence energy control system based on wireless smart socket and IoT. *IEEE Access: Practical Innovations, Open Solutions*, *4*, 2885–2894. doi:10.1109/ACCESS.2016.2574199

Yaqoob, I., Ahmed, E., Hashem, I. A. T., Ahmed, A. I. A., Gani, A., Imran, M., & Guizani, M. (2017). Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges. *IEEE Wireless Communications*, *24*(3), 10–16. doi:10.1109/MWC.2017.1600421

Yu, H., Sun, Q., Sheng, K., & Wang, Z. (2016). Intelligent street lamp control system using ZigBee and GPRS technology. *International Journal of Simulation-Systems, Science & Technology, 17*(35), 36.1–36.13.

Zaidan, A. A., Zaidan, B. B., Qahtan, M. Y., Albahri, O. S., Albahri, A. S., Alaa, M., Jumaah, F. M., Talal, M., Tan, K. L., Shir, W. L., & Lim, C. K. (2018). A survey on communication components for IoT-based technologies in smart homes. *Telecommunication Systems*, *69*(1), 1–25. doi:10.100711235-018-0430-8

KEY TERMS AND DEFINITIONS

Big Context: Is defined as a better understanding of how entities.

Big Data: Is extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

Context: Is the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood.

Internet of Everything: Is a broad term that refers to devices and consumer products connected to the internet and outfitted with expanded digital features.

Internet of People and Things: Is a system of interrelated computing devices, mechanical and digital machines, objects, animals, or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

Internet of Signs: Is categories of signs, including written language, natural language, cultural codes, aesthetic codes, codes of tastes and a number of others.

Internet of Things: Is the interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data.

Semiotics: Is the study of signs and symbols and their use or interpretation.

Chapter 8
Big Data Analytics in Developing Economies:
Harnessing Insights and Creating Value

Forgor Lempogo
Ghana Communication Technology University, Ghana

Ezer Osei Yeboah-Boateng
https://orcid.org/0000-0002-1355-8586
Ghana Communication Technology University, Ghana

William Leslie Brown-Acquaye
https://orcid.org/0000-0001-9679-5976
Ghana Communication Technology University, Ghana

ABSTRACT

In a world increasingly driven by data, most developed economies are leveraging big data to achieve greater feats in various sectors of their economies. From advertisement, commerce, healthcare, and energy to defense, big data has given new insights into the huge volume of data accumulated over the past few decades that is helping reshape our knowledge and understanding of these sectors. Unfortunately, the same cannot be said about the state of big data in the developing world, where investments in IT infrastructure are dangerously low, keeping huge proportions of the population offline. This chapter discussed the challenges that exist in developing countries, which affect the smooth take-off of big data and data science as well as recommendations as to how countries and companies in the developing world can overcome these challenges to harness the benefits and opportunities presented by this technology.

DOI: 10.4018/978-1-7998-6792-0.ch008

INTRODUCTION

It is generally agreed among industry players that data is the new oil, since various parallels can be drawn between the two. Like oil, even though data exist in abundance, extracting and mining to find meaningful information is messy, expensive and excruciatingly difficult. That aside, a large chunk of the world's data is controlled by a few multi-national technology oriented organizations and very expensive and elaborate infrastructure has been built to refine, transport and distribute it. Furthermore, organizations who control the world's data will eventually control a significant portion of the world's wealth.

Data is not an end in its self, but a valuable raw material, which when refined, mined and organized appropriately can bring out valuable insights (Englmeier, 2015). To become really useful, these insights must be processed into information, and the information transformed to knowledge. The bigger the data, the more resource-intensive it is to work with, the better the value of the information must be to make the trade-off a wise business decision (Dean, 2014).

The past decade has seen a revolution in the area of data generation, processing, transmission and storage. This revolution is largely as a result of advancement in technology, most prominent of which include social networking, the Internet-of-things (IoT) such as wearable devices, reduction in the cost of IT equipment such as storage, increased processing power and miniaturization (EMC Education Services, 2015). Even though data has always been in abundance, these new technologies have led to data being generated at a pace never experienced before. Which has led to the accumulation of huge volumes of data at very short periods of time. Necessitating the coining of a new term "Big Data", to describe the sheer volume, velocity and complexity of the new data.

Formally, Romero, et al., (2016) define Big Data as a set of non-structured and complex data, which come from sensors, social media, applications and devices that work with the Internet and require technologies to store, manage, analyze and visualize the information. In a world increasingly driven by behavioral economics, developing economies need to leverage big data analytics for inclusive growth, to enable all citizens to contribute towards growth with the potential to benefit the entire society (Hagstroem, 2015). The main challenge in the field is to organize these complex data structures into formats that computers can be used to augment our understanding of the domain knowledge and to timely discover previously unknown insights such as trends, patterns and clusters (Dean, 2014).

In a very fast paced world, the time to solution, thus timely execution of these tasks is as important as the accuracy of the solution (Gorelik, 2019). For instance, a recommendation system must be able to quickly profile a visitor to a site and recommend advertisement before the visitor leaves the site. A system that arrives at a right solution after a visitor has already left the site is as good as that which gives the wrong recommendation.

Big data presents society with a new opportunity to shift from a world in which we think we know how to improve our living conditions into a world in which we know exactly how to do it and can prove that we can (Hagstroem, 2015). Consequently, organizations in the financial, insurance, commerce, and health sectors heavily rely on big data for operational and management decision support. Research has proven that firms that make decisions based on data will over time make better decisions, which leads to a stronger, more viable business than those who use their instincts. To truly harness the benefits of big data, companies must invest in both technology and human capital and none of these will come cheap. Apart from that, the national information technology (IT) infrastructure must be mature and robust enough to support and complement the investments of the private sector. Also countries must formulate and implement IT policies and Regulatory frameworks to create an environment where these companies

can safely collect and use data without violating the safety and privacy of the users of these systems as well as policies to clearly define ownership and areas of jurisdiction in case there is a dispute.

In developing economies such as those in sub-Saharan Africa, where national budgets are thinly spread across all primary sectors of the economy, IT spending among governments is always among the least and in most countries, exist in the books only, leaving most IT infrastructure decades behind those in the developed world. Also, small and medium size enterprises (SMEs) turn to be the drivers of most of these economies, with very few large enterprises. For most SMEs the kind of resources needed to invest in their own IT infrastructure for big data is nonexistent, due to the fact that the capital market in most of these economies is very tight. It is also a common complaint from industry in most of these countries that even when companies do invest in such technologies, finding graduates with such skills become a very daunting task, as a result of mismatch between industry skills needs and university curricula. Additionally, most countries lack basic legislation and appropriate regulatory framework regulating electronic transactions and data protection as well as the policies prioritizing information technology as a driver of economic development. As a result, even the little investment in the sector is mostly misapplied or put in areas that do not really bring the needed impact.

This chapter looks at the problems in developing economies and how organizations operating in these economies can navigate these challenges to harness the benefits of big data. Areas covered include, big data infrastructure needs, state of National IT infrastructure, state and corporate investment in IT sector and infrastructure, Internet penetration, bandwidth and cost, human capital and capacity building and recommendation on how to navigate these challenges to harness the benefits and opportunities presented by this technology.

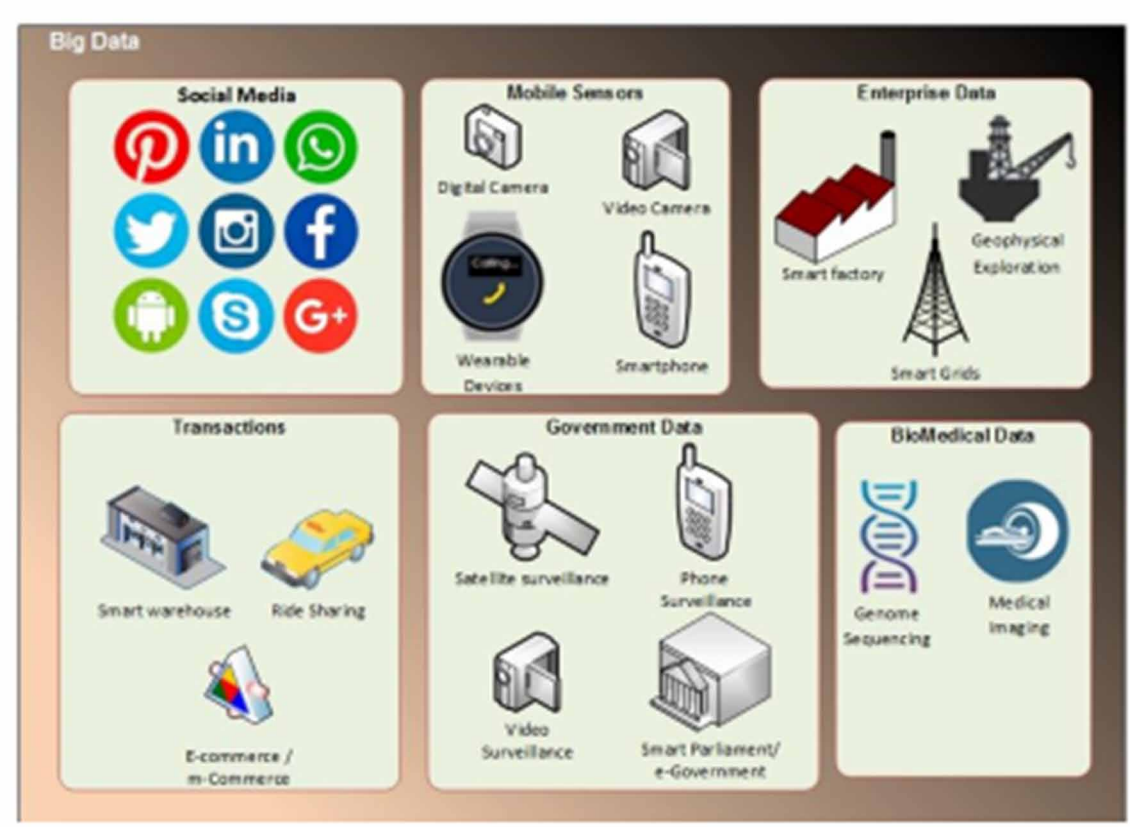
BIG DATA INFRASTRUCTURE

The overwhelming increase in the volume, speed and complexity of data, due to the wide adoption and high penetration of smart phones and mobile technologies, social media technologies, etc. (see fig. 1), has caused an absolute paradigm shift and hence the need for disruptive techniques and technologies for data processing and management. For instance an average smartphone today has sensors to tell where we are (GPS), how fast we are moving (accelerometer), what the weather is like around us (barometer), what force we are using to press the touch screen (touch sensor), etc. (Marr, 2016). Even though these huge volumes of data present great opportunities to transform businesses, governments, science, and our society at large, it also poses a great challenge for existing Information technology (IT) infrastructure, which in most instances have reached its limits in terms of robustness and scalability, requiring new investments in the sector. Investment in infrastructure is not limited to only high-end hardware and software, but expensive maintenance efforts as well (Sharma, 2016). The main reasons why companies invest in big data are to improve insight into user behavior, speed-up the time to solution and improve service availability, which according to (Otava, 2017), correlates directly to having a robust and high performing hardware infrastructure.

Over the years, various definitions have been given for big data infrastructure. One of the most comprehensive is given by (Tozzi, 2019) as follows:

Big data infrastructure entails the tools and agents that collect data, the software systems and physical storage media that store it, the network that transfers it, the application environments that host the

Figure 1. Key drivers of big data
(Source: Authors, 2020)



analytics tools that analyze it and the backup or archive infrastructure that backs it up after analysis is complete." (Tozzi, 2019)

According to Tozzi's definition, big data infrastructure actually starts from the devices (tools and agents) that collect the data. These devices are mostly used to capture data provided 'freely or voluntarily", directly or indirectly by individuals (subscribers, users, clients, customers, etc.) using personal or public systems. As seen in fig. 1, devices used to access technologies in the categories of Social Media, Device Sensors and Transactions are mostly personal devices and therefore are not within the control of the organizations implementing big data. Hence, their quality, speed and robustness cannot be directly influenced by these organizations. Same is also true for the infrastructure needed to transmit the data (Internet), which is highly dependent on the national telecommunication infrastructure. However, the hardware and software systems, utilities and tools needed to store, process, analyze and back-up the data fall directly under these organizations' control.

For an organization looking to invest in Big Data, it is always a good idea to start at the hardware infrastructure level. Investing in high-performance processing and robust storage hardware infrastructure will solve some of the most pertinent issues of eliminating bottlenecks and improving performance throughout the life span of the system (Otava, 2017). Once a robust and efficient hardware infrastructure

is in place, then the attention can turn to software, utilities and tools needed to analyze the data. It therefore goes without saying that, successfully harnessing the benefits of big data goes beyond having the right data analytics tools and applications, but also having a cleverly optimized hardware infrastructure fit for that purpose (Gupta & Shilpi, 2016).

In developing appropriate use cases to harness the potential benefits of Big Data, (Yeboah-Boateng & Nwolley, 2019) proposed the use of wizard-based data integration techniques. The approach is meant to receive the inputs of datasets, from say, WhatsApp instant messaging, and then activating the MapReduce functionality to integrate, clean and prepare the datasets for analytics. In essence, various unstructured datasets from disparate sources undergo virtual data integration, and then Big Data analytics to harness the insights, such as trends, patterns and even outliers.

NATIONAL IT INFRASTRUCTURE

The reliance of big data technologies on national information technology (IT) infrastructure cannot be over emphasized. Numerous research on this subject have established a direct and strong correlation between national information technology infrastructure and the availability, affordability and improved quality of service (QoS) of the Internet, which is the greatest enabler of big data and its associated technologies. Fig. 2 clearly illustrates this relationship between infrastructure and affordability in the African continent. Citizens in countries with improved infrastructure index turn to enjoy a much more affordable Internet compared to those who live in countries with low infrastructure score.

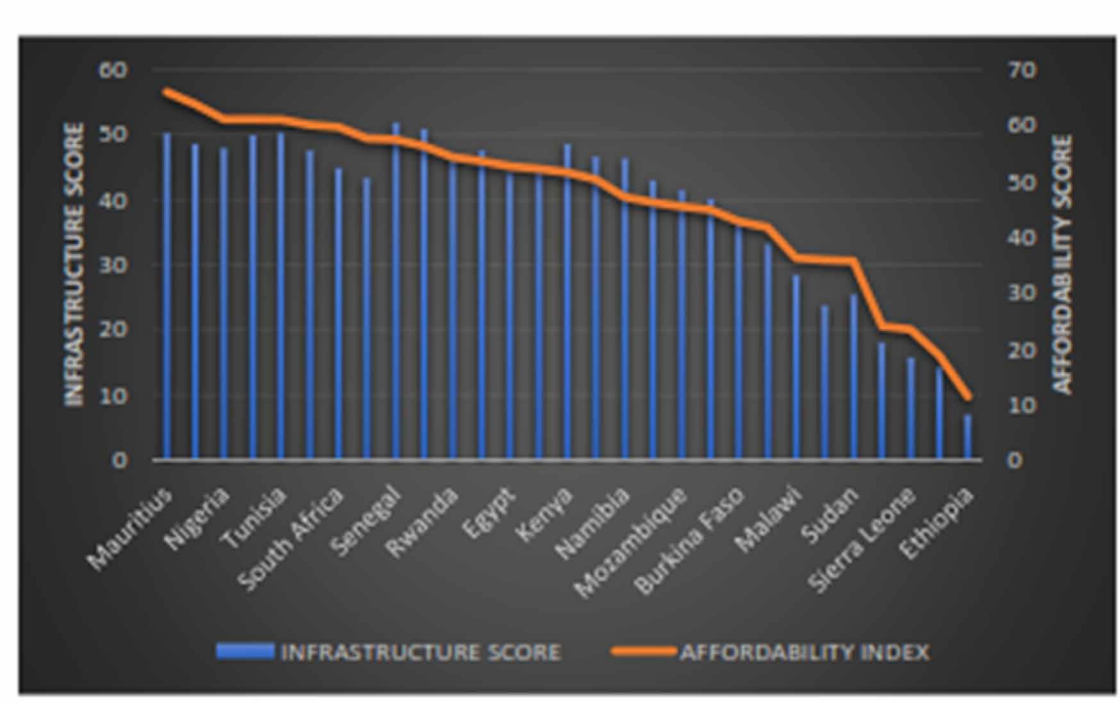
Like other types of infrastructure such as roads, electricity, etc., the national IT infrastructure is a cooperative affair between government and the private sector. In this regard, the networks of the telecommunication companies turn to be the core, keeping all these together. Hence, information technology Infrastructure is a distributed technical framework that encompasses the unseen realm of protocols, networks, and middleware that bind the computing enterprise/state together and facilitate efficient data flows, as well as the people providing support and services (Kordha, et al., 2011).

There are currently various infrastructural options available to nations, such as fixed network technologies, Broadcast infrastructures, such as Digital terrestrial television (DTT), Mobile networks, such as 3G, 4G and now 5G, that can be blended to collectively provide a nationwide communications transport layer. However, these Infrastructure choices must be made with a clear understanding of the nation's overall information technology strategy. Improving quality of service (QoS) depends not only on these technical aspects of infrastructure, but also on having the policies and regulatory framework in place to guide and compel mobile network operators (MNOs) to improve networks and operations, promote competition in the market, and increase capacity to provide independent and accurate data on quality of service (Woodhouse & Thakur, 2018). A nation's big data infrastructure usually has two components, thus the state IT infrastructure and the corporate IT infrastructure, hence requiring investments from both side.

STATE IT INVESTMENTS

With about 33 percent of the Africa's 1.2 billion population living in extreme poverty and governments with debt averaging more than 50 percent of GDP, most African countries have very little room to invest in economic and social infrastructure, such as what is needed for big data and digitalization

Figure 2. National IT infrastructure and internet affordability
Source: (Alliance for Affordable Internet, 2020)



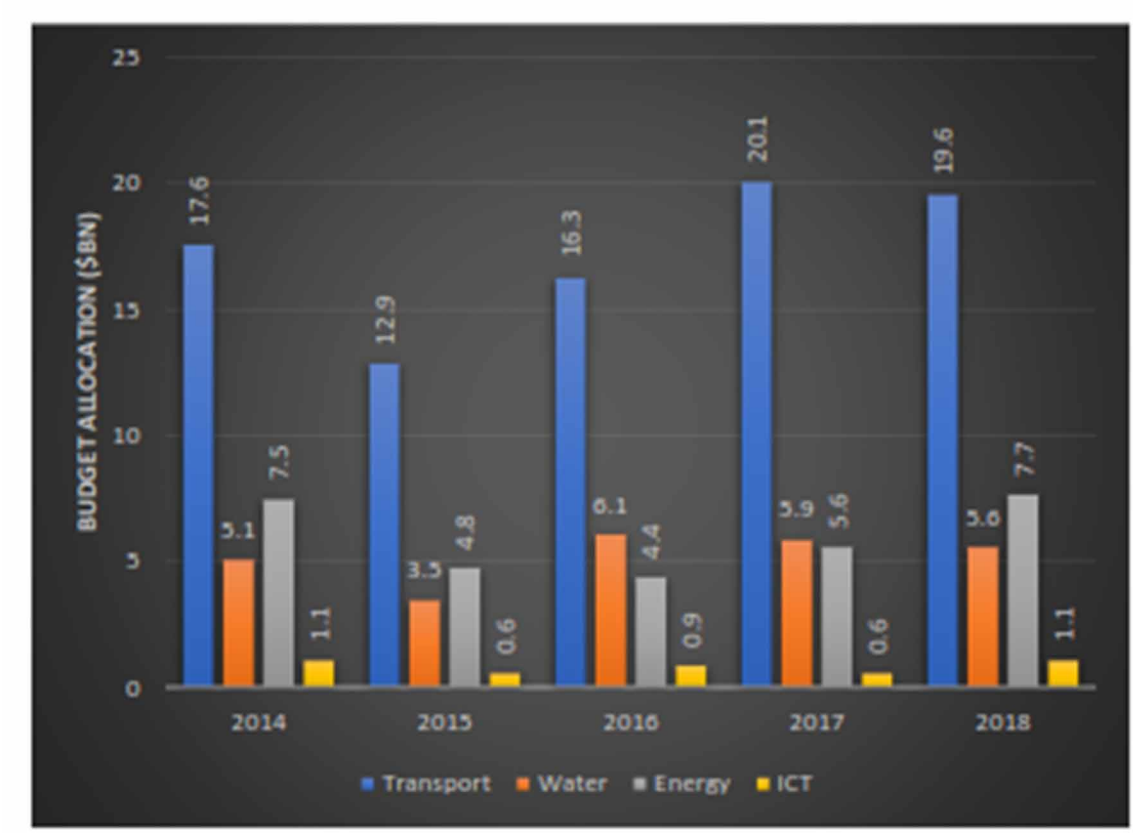
in general (Songwe, 2019). In their 2019 report, (Alliance for Affordable Internet, 2020) accessing the communication infrastructure of some selected African countries, concluded that only a couple of African countries scored up to 50 on the infrastructure index, which ranges from zero (lowest score) to 100 (highest score) (Seen in fig. 2). As shown on fig. 3, investment data for sub-Saharan Africa indicate a very low infrastructural investment, especially so for the IT sector. According to data from ICA, national budget allocations to the ICT sector increased by 85%, reaching $1.1bn in 2018, $514m more than in 2017, with the most noticeable changes being a $193m increase in allocation in West Africa and a $109m increase in Central Africa (ICA, 2019). Even though the increase is a welcome development in the sector, $1.1 billion investment shared among 53 countries is still not going to be enough to close the huge infrastructure gap.

In some cases, investment in the ICT sector is perceived as a private sector activity. Even though this perception may be appropriate for cities and metropolitan areas, it is however counter-productive in solving the problem of connecting the huge unconnected mostly poor rural dweller.

In an effort to complement the low national investment in infrastructure and other areas of the economies of these countries, multilateral development banks (MDBs) such as the World Bank, intentional monetary fund, Africa development bank, etc., have stepped in to pick up the slug.

For decades, MDBs have contributed significantly to infrastructural development around the world, especially in the developing world, through direct investments, collaboration of government and private sector to execute projects, and by facilitating the development of appropriate policies to create an enabling environment to attract private sector investment. Even though a cumulative amount of $525 billion was

Figure 3. National government budget allocations by sector ($bn) - Africa
Source: (ICA, 2019)



committed by MDBs for the funding of development projects in low-to middle-income countries worldwide between 2012 and 2016, only 1% of these project funds went toward ICT projects (Zibi, 2018).

These low levels of public sector, private sector and MDBs investment in the ICT sector over an extended period has led to the generally low infrastructure index, which has also led to the low internet accessibility and affordability index in most African countries today.

CORPORATE INVESTMENT IN IT

According to (Kordha, et al., 2011), a firm's information technology portfolio is made up of the total investment in computing and communications technology, encompassing investments in hardware, software, telecommunications, electronically stored data, devices to collect and represent that data, and the people who provide IT services. Since IT infrastructure is the backbone of a successful big data project, a company that want to harness the benefits of big data must be capable and willing to invest in creating a suitable comprehensive IT infrastructural foundation that will provide adequate levels of speed, robustness and agility.

In most African economies, more than 60 percent of the population lacks access to financial services (Songwe, 2019), and even when these services are available, companies must contend with interest rates of up to 30 percent per annum. This does not only leave the IT sector with a huge investment deficit, but also stifle much needed investment and kills many potentially great ideas. By the estimates of (Zibi, 2018), $10 billion a year is needed to close the universal access gap, and expanding digital inclusion will require more capital from MDBs, the private and public sectors alike.

As a result, in most developing countries such as those in Africa, the IT space in dominated by multi-nation technology giants who have access to external financing. Additionally, mobile network operators turn to dominate telecommunication space for both voice and data communication.

One of the ways out for SMEs who in most cases lack the financial muscle to invest in the required expensive IT infrastructure necessary for big data analytics is to subscribe for cloud services and solutions. However, adopting cloud technologies in isolation might not be enough to solve these problems, since issues of slow Internet connections and poor QoS issues due to infrastructural limitations may render these cloud solutions inappropriate.

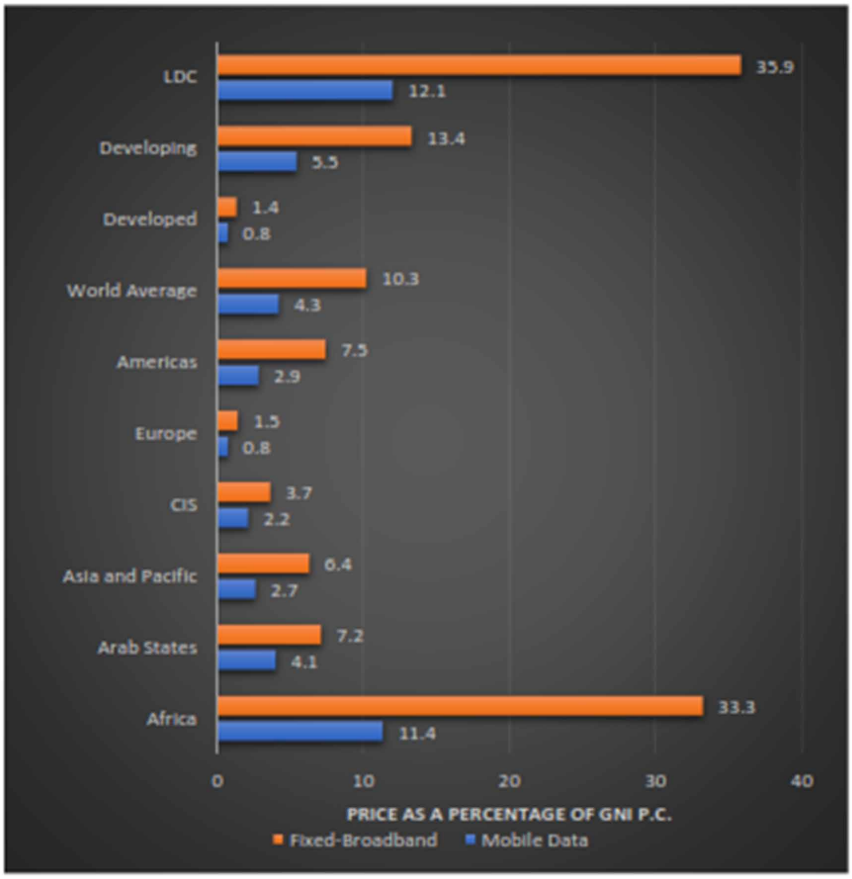
INTERNET PENETRATION, BANDWIDTH AND COST

The volume and velocity of big data for most of the industry players such as Google, Twitter, Facebook, etc., depend largely on the number of users and how long or often they use these systems and services. This is because data is mostly generated by these users through the use of mobile sensors, social media and transactions such as e-commerce. For instance, in order to profile a potential customer, the user must spend some time online, to voluntarily give out his data or involuntarily leave enough digital traces of his location and actions, while using other services. For this to happen, the user must have a device that can connect to the internet, such as smartphone, tablet or PC, the infrastructure to connect, such as Broadband or mobile network and the means to afford to connect. In most developing economies such as those in sub-Saharan Africa, any of these three stages could be a potential barrier of participation.

In accessing the affordability of mobile devices in 70 low and middle-income countries (Woodhouse, 2020) concluded that, even with the consistent, drastic decline in the prices of smart phones, as much as 2.5 billion inhabitants of poor countries must spend at least a quarter of their monthly average income, in order to acquire the cheapest smartphone in their country. This he noted, is equal to the same percentage of monthly income a European household spends on housing and utilities combined. A much gimmer picture was painted when it came to countries in Africa. The same research concluded that a simple mobile phone in Africa cost up to 62.8% of average monthly income compared with 11.7% in the Americas. Coupled with the fact that consumer must pay for these products upfront, since most of these countries lack any other alternative financing models, they warned that with the twin effect of increasing poverty and disrupted supply chain, Covid-19 could make phones even much less affordable in the months ahead.

Even when citizens are able to overcome this equipment cost barrier to acquire the devices, they still have to be able to purchase mobile or fixed broadband subscription, which are mostly out of their reach due to the exorbitant prices they have to pay. According to the (ITU, 2020), Africa is the region with the least affordable mobile data and fixed-broadband relative to income, which stands at a staggering average of 11.4 per cent and 33 percent of Gross national income per capita (GNI p.c.), respectively, as show in fig. 4.

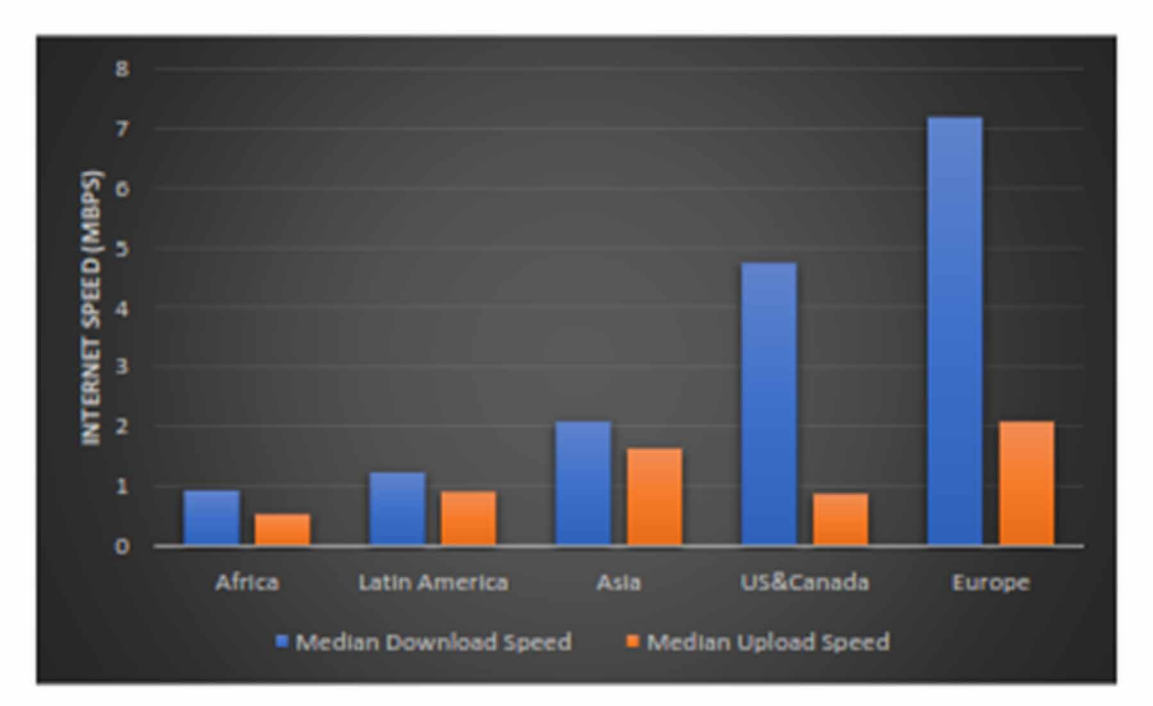
Figure 4. Average regional prices of internet as a percentage of GNI p.c.
Source: ITU 2019



This serves as a huge barrier and has serious effect on the number of citizens who are able to go online from these countries. ITU (2020) puts the total internet users in Africa at a paltry 28 percent as against 82 percent and 77 percent for Europe and the Americas respectively.

Beyond this barrier lies another important limitation of service quality and the bandwidth itself. As seen in figure 5 and figure 6, mobile data and fixed-broadband are not only out of the reach of most Africans (costing close to 100% GNI p.c. in some countries), but also have very limited bandwidth. In most of these African countries internet service is not only limited to a few kilobytes per second, but also the service itself is very erratic and is only available and stable in certain parts of the city at certain times of the day. This greatly limits the length of time users spend online and the type of websites and service they are able to access, as well as the type of applications they run on their devices and how they run them. For instance, a user who is economizing her/his data might not have the luxury to keep a data-intensive application such as a location sensor running for an extended period of time, or frequently upload/download large files such as videos and pictures or even watch YouTube videos. They might also be less likely to click on advertisements, as they are unwilling to "waste" their data. This will greatly

Figure 5. Median download and upload speeds (Mbps) by Region
Source: (Woodhouse & Thakur, 2018)



limit the volume and intensity of data such users are able to generate, leading to a potentially inaccurate digital profiling of such users.

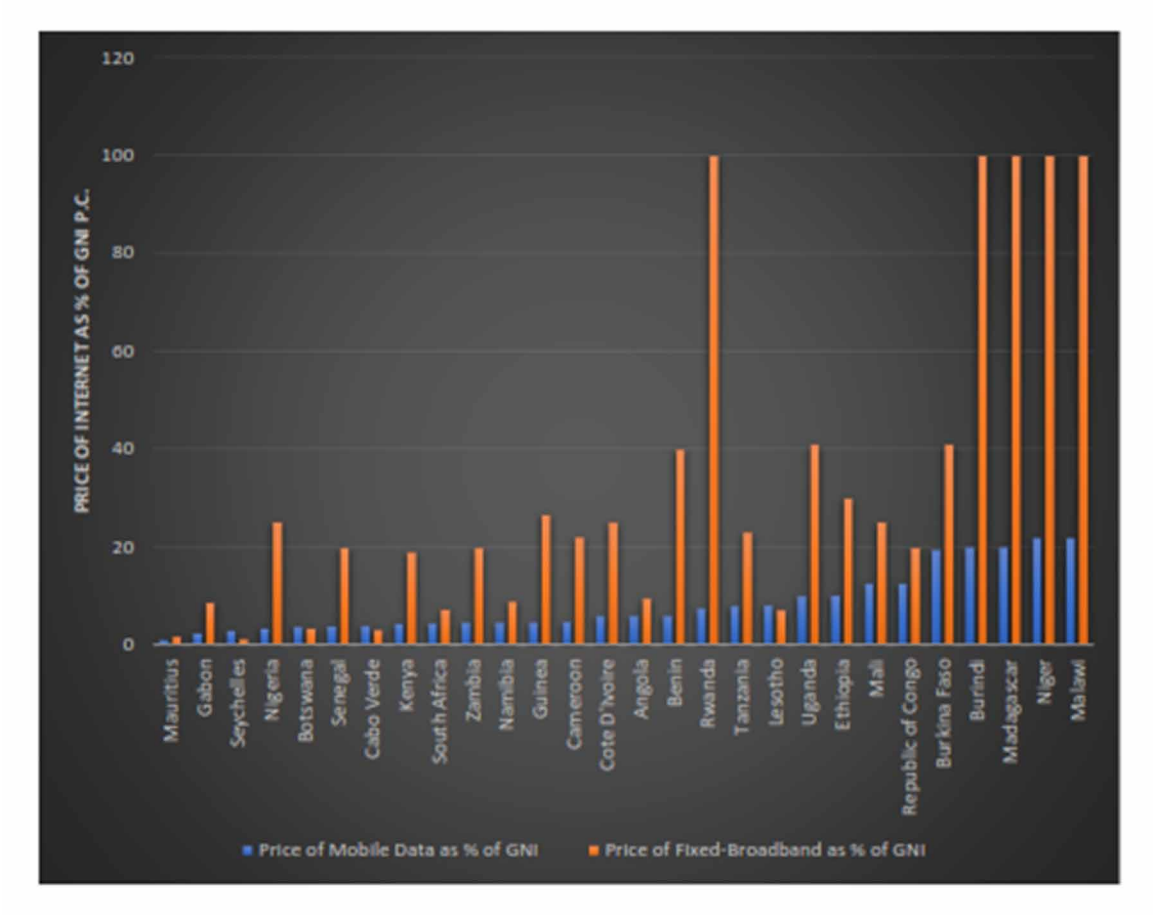
As illustrated on figure 5, mobile Internet users in Africa experience the longest delays, especially compared with developed regions such as Europe, US and Canada. Coupled with the fact that most Africans are offline, online Africans are faced with connectivity options that are much lower in quality compared to other regions (Woodhouse & Thakur, 2018). These low internet speeds surely negatively affect the implementation of advanced computing systems, processes and technologies, such as those required for the implementation of big data analytics.

The main reason most organizations, particularly SMEs report very low levels of participation in big data may not be related only to the high initial investments involved in setting up the infrastructure, but also, the potential lack of real benefits, due to low participation of the general populations of these countries in online activities and interactions. This low participation is largely because a very high percentage of their population are virtually not online. Consequently, even the data these companies accumulate from the online population may turn out to be unreliable.

HUMAN CAPITAL AND CAPACITY BUILDING

In times of drastic industry transformation, such as the current phenomenal big data explosion, it is inevitable to experience a skills gap needed to remain competitive. This problem is compounded by the

Figure 6. Cost of mobile data and fixed-broadband as a percentage of GNI p.c. in Africa, 2019
Source: ITU 2019



extremely technical nature of data science positions, requiring potential employees to be well skilled in science, technology, engineering, and mathematics (STEM), precisely the skills that are currently already lacking (Hagstroem, 2015). Buttressing this point (Luaces, 2019) stated that, the four main technical skills needed by a data scientist are statistics, data science, computer science and machine learning. In his opinion, the main challenge for companies is finding professionals who possess all four, including the soft skills.

A report by The Royal Society (The Royal Society, 2019), stipulates that the period from 2013 to 2018, saw a sharp rise in UK job-listings for Data Science positions by over 231%, driven predominately by increased numbers of vacancies for Data Scientists (over 1287%) and Data Engineers (over 452%). Similarly, LinkedIn chose Data Scientist as their 2019's Most Promising Job in the U.S, indicating that the skills gap is already stifling digital transformation in an estimated 54% of companies and predicted a global skill shortage of 4.3 million in the field by 2030. LinkedIn (LinkedIn, 2018) puts the US National data science skill shortage at an estimated 151,717 people in 2018. Also, (Fayyad, 2019) reports of the prevalence of mismatch between skill expectations of employers seeking Data Scientists and what true data science and analytics professionals actually put on their resumes.

Bringing this to the African context, (World Economic Forum, 2017), estimated that for the next couple of decades, between 15 and 20 million young people will join the African workforce each year. Indicating that by 2030, Africa will be home to more than a quarter of the world's population under 25 and that 15% of the world's working-age population will be in Africa. Essentially, this is great news for the continent. These huge unfilled job positions in data science present a great opportunity for African countries, currently facing unprecedented youth unemployment in all sectors.

However, with universities that are mostly ill equipped for technology education and curricula that hardly equip graduates with the real skills needed for industry, Africa is facing a worst skills deficit in the field of data science. Even the few skilled professionals are lured away by much wealthier multinationals outside the continent with the financial muscle. The persistent history of brain drain presents a peculiar challenge for African companies and even government**(s)**, having to compete in the global stage in recruiting and maintaining these highly skilled and scares human resource. Thus, companies that manage to overcome the huge infrastructure investment barrier still face an even greater challenge of attracting and maintaining the qualified workforce necessary to effectively operate and maintain these very complex and expensive big data infrastructure.

SOLUTIONS AND RECOMMENDATION

Having laid bare the complex nature of the challenges and difficulties affecting big data participation in developing economies, it begs the simple question: why then bother? The answer is very simple: The potential benefits of big data are just too powerful to ignore. It is however quite clear that conventional methods will not be feasible for most companies in these countries. Navigating these challenges to achieve a viable and successful big data project will take tact and technological appropriation. This will also require close collaboration between industry, universities and government in deciding where, what, and when to invest to make the most impact. While governments work on creating policies to facilitate and stimulate participations, companies must also device ways to navigate the investment challenges through the adoption of technologically affordable and viable methods, as well as engaging the universities on the areas of emphasis in curriculum development.

Use of Cloud Solutions

With very limited financial and human resources available, one of the most logical ways for SMEs in developing economies to overcome the investment cost barrier is by subscription to cloud technologies and solutions. At the very least, cloud deployments are associated with numerous benefits, including but not limited to CAPEX-free investment requirement**s**, low operational and maintenance cost, the ability to focus on the outcomes instead of the infrastructure and technological aspects, less time and resources needed (including skilled data scientists), etc.. The accrued benefits that can be derived from the combination of big data and cloud technologies has been so great that a new terminology: Big Data as a Service (BDaaS), has been coined to describe it.

BDaaS is a broad range of technologies, encompassing platforms, managed toolsets or managed service that facilitates management or running big data on the cloud (Maayan, 2018). For SMEs willing to opt for BDaaS, there are different models depending on the need. Most common BDaaS models

include Big Data Infrastructure as a Service (BDIaaS), Big Data Platform as a Service (BDPaaS) and Big Data Software as a Service (BDSaaS).

BDIaaS provides the basic infrastructure for data storage and processing on the cloud, with very basic data services being provided. For this model, the customer is required to provide and maintain the tools such as software and data analytics tools. There is therefore the need for expertise in data analytics and computer programming. However, this model is most flexible and allows for maximum control by the customer in managing the tools and utilities as well as the systems as a whole. Currently BDaaS systems that have properties to run this model include Amazon's EC2 storage platform and Azure Blob Storage.

BDPaaS on the other hand, provides the basic infrastructure together with preliminary functionalities for application development and deployment. A customer opting for this model must also be aware of the need for programming and data science expertise necessary to maintain the infrastructure and systems in the model. It is also worth noting that in most BDPaaS platforms the customer turns to have limited control over the hardware and storage due to the levels of virtualization. Currently BDaaS systems that have properties to run BDPaaS model include Heroku, Google App Engine, Amazon S3, Redshift, EMR, etc.

Finally, BDSaaS provides a comprehensive package of tools within a platform, running on an infrastructure. Thus, an end-to-end data stack in one complete package. A comprehensive set of tools enveloped in a system of very low complexity, low cost, enabling data analysis without the need for expertise in computer programming and data science. All the technical expertise in creating algorithms, programming, installation and maintenance of the system are yielded to the vendor while the customer concentrates on their core business. The downside with this is that the system is not very flexible since the customers has no control over infrastructure layers and the raw data itself. This limited access to the raw data might be a very serious drawback for future data projects. Currently BDaaS systems that have properties to run BDSaaS model include Looker, Chartio, Tableau, Periscope, Geckoboard, GoodData, etc.

Table 1 presents a summary of how these models stack up against each other, in terms of complexity, cost, flexibility and expertise requirement. Depending on the data storage and processing needs as well as financial muscle and technical expertise, a decision will be made as to which model is most suitable and appropriate. For SMEs who typically have very simple organizational dependencies or data processes, BDSaaS model is mostly the best choice for their big data needs.

Reduce Taxes on Low-Cost Devices and Internet Services

Often justified by a misguided and decades old idea that the ICT sector is a luxury, most countries still put so much taxes and levies on basic communication devices such as smartphones. Taxes in this sector in most countries include Import Duties, VAT, Excise tax, Sector-specific tax, etc., which in some countries can add up to 35% to the cost of the devices. While such taxes may be justified to an extent, their blanket application to all smartphone**(s)** however cannot be justified.

Instead of blanket taxes, governments must be willing to create tax regimes that selectively apply tax exemptions to devices of certain categories and classes. Government at some point must have to agree that smartphones are a necessity in the modern era to improve the wellbeing and quality of life of citizens. Each country must then identify the point beyond which a device can be considered a luxury, with taxes applying only to the luxury category.

Apart from taxes on the devices themselves, some governments also put taxes on the use of these devices. African countries like Ghana has since 2008 introduced The Communications Service Tax (CST), putting levies on charges for the use of communications services, such as voice calls, mobile

Table 1. BDaaS models – how they compare

Criteria	IaaS	PaaS	SaaS
Complexity	Technically complex	Moderate complexity	Least complex
Cost	Most Expensive	Moderate	Low
Need for Technical Expertise	Advanced expertise needed in Programming and data analytics	Requires expertise in programming and data science	No needed for expertise in Programming and data science
Organizational suitability	Suitable for organizations with very complex data pipelines	Suitable for organizations with moderately complex organizational dependencies or data processes	Suitable for organizations without complex organizational dependencies or data processes
Flexibility	- Most flexible - Allows for maximum control in managing the infrastructure, tools and utilities.	- Moderate flexibility - Access to hardware and storage is reduces due to virtualization	- Least flexible - Access to the infrastructure layers is limited. - Limits the type of data that can be used
Current Tools and platforms	- Amazon's EC2 storage platform - Azure Blob Storage.	- Heroku, - Google App Engine, - Amazon S3 - Redshift - EMR	- Looker - Chartio - Tableau - Periscope - Geckoboard - GoodData

data and broadband internet, that are provided by electronic communications service providers (GRA, 2020). CST has since 2008 steadily move from 6% to 9% of the service charge in 2020, adding to the overall prices of the service, which is already out of the reach of most citizens.

Another area of special interest when it comes to the affordability barrier is financing the purchase of the devices. In most developing economies, purchasing a device can only be achieved by paying for it upfront, compelling most citizens to end up purchasing used/second-hand devices of poor quality with the associated problems of e-waste. Governments must come in with policies, creating an enabling environment to facilitate the acquisition through credit facilities from the banks and microfinances, or even through hire purchases, co-operatives, and support projects to help people spread the cost of devices over a reasonable period. It might even make sense for governments to consider subsidizing low-end smartphones, as some already do for selected essential basic commodities.

Removing these taxes from basic-featured or low cost smartphone**(s)** and electronic communication services will significantly reduce their cost and increase affordability, putting more devices in the hands of poorer household, increasing accessibility and connectivity. Getting connected devices into the hands of the citizens is the beginning of digital participation. There is no doubt that removing these taxes from these devices and services will bring about some shortfall in governments' tax revenue in the short term. However, it is believed that the potential long-term benefits of digital inclusion, including gains from big data will definitely make up for any such shortfall in the long term.

Increase Spending in IT Sector

Apart from the tax aspect, government also need to increase spending in the ICT sector to improve infrastructure as well as creating and resourcing agencies to implements key regulations in the sector. Most developing countries through the guidance and assistance of multilateral banks and donors have

developed ICT regulatory and policy frameworks covering most important aspects of the field. However, in most of these countries, these laws and policies only exist in the books, without the will or in some cases the means to implement or enforce them.

To achieve any of such feats in this sector, African governments must stop looking at investment in ICT as a private sector affair. ICT infrastructure must carry the same importance and must be seen to be significantly contributing to economic growth as well as the general well-being of the population, as much as the sectors of transportation, energy, water and security. They must increase budgetary allocation and consistently remain committed for a reasonable period to achieve real rewards.

In addition, investments and operational policies in the sector need to be able to persuade or if necessary coerce telecommunication companies to invest in rural areas. This can be done by giving companies tax incentives or making it a requirement for issuance or renewal of operating licenses, or even such investments can be counted as part of their corporate social responsibility. This will help improve access and participation in the rural area.

Apart from that, the regulators must also be properly setup and adequately resourced to carry out their mandate in ensuring strict adherence to quality standards on the part of the telecommunication companies. That will not only improve the quality of service (QoS) of the networks, but also go a long way to ensure fairness and trust in the sector. There is no doubt that improving QoS of a network will definitely require some investment from the service providers and therefore may meet stiff resistance from the industry. Therefore, the regulators must be firm to ensure the right thing is done to give customers their moneys' worth.

Rethink the Education System

Prognosis in the data science job market has been positive and from all indication will continue to be for a very long time to come. With most countries facing the greatest challenges of massive youth unemployment, Africa is presented with an unprecedented opportunity to harness the great benefits this field of data science currently presents and will continue to present in the future. It does not only promise to solve this unemployment problem, but also an opportunity to exploit the possibility to export skilled labour. However, these benefits are not automatic. The only countries that will really benefit are those who will proactively plan and implement the necessary policies as well as dedicate the needed investments in the future of their youth in this field.

One measure that will set a country on this path is the restructuring of the educational curricula, placing more emphasis on STEM subjects at the early levels of education. This is necessary to equip students with the necessary technical and analytical foundation to study data science programmes and later become successful data scientist.

One of the symptoms of a weak educational system is that it leaves huge numbers of graduates without jobs and industry yearning for graduates with requisite skills for employment. Industry must therefore closely collaborate with academia in developing curricula that will produce graduates with skills that match their needs. This partnership should not only be technical in nature, but also financial in terms of providing the needed financing for the acquisition of technical and technological equipment, which are mostly out of the reach of universities on the continent due to low financial investments, but are seriously needed for quality training.

CONCLUSION

Big data presents a great opportunity for the world today. Giving us the opportunity to draw insights and conclusions from vast volumes of data which until recently was thought to be impossible. Drivers of big data are social media, mobile sensors and transactions on systems such as e-commerce and content management systems, working through the internet to generate, transmit and store huge volumes of data for analysis. Citizens of developing countries are faced with multiple challenges and barriers preventing them from getting online and staying online to generate the data needed to create these insights. Leaving companies with 'big data' that might not truly reflect the real situation on the ground. Even the most sophisticated data analytic system or tool cannot generate reliable insights if the data on which it is operating is not suitable.

This chapter discussed the challenges that exist in developing countries, which affect the field of big data and data science as well as recommendations on how countries and companies in the developing world can overcome these challenges to harness the benefits of these technologies. Some of the problems discussed include inadequate national IT infrastructure due to low IT spending among governments. In addition, most developing economies are dominated by SMEs with no financial muscle to invest in the IT infrastructure required for big data implementation. Lastly, shortage of job seekers with expertise in the field of data science due to mismatch between industry skills needs and university curricula.

Even though problems and barriers exists, the potential benefits of big data outweigh them all. The solution lies with the governments and the corporations alike. Some solutions are technological; others financial, while others are political. SMEs for instance can overcome the initial investment requirement barrier by investing in BDaaS technologies, which require less initial and operational investments, as well as medium to no programming and data science expertise needed. Some of the financial solutions include increasing investments in the national IT infrastructure, compelling the telecommunication companies to expand into the rural areas, while making sure that they improve service quality and adhere to laid down quality standard. On the policy side, governments must come up with policies to make smartphones and mobile data and fixed-broadband internet services more affordable to the average citizen, through the removal of taxes put on devices and services, as well as providing alternative financing for such devices.

It is believed that these steps will greatly improve access to the internet, leading to improvement in participation. This in turn will lead to the generation of enough data and digital footprints to improve the quality, volume, velocity and veracity of data, which is what, makes big data.

REFERENCES

Alliance for Affordable Internet. (2020). *Affordability Report 2019*. Alliance for Affordable Internet.

Alliance for Affordable Internet. (2020). *From luxury to lifeline: Reducing the cost of mobile devices to reach universal internet access*. Web Foundation.

Callahan, S., & Glass, R. (2015). The big data-driven business - how to use big data to win customers, Beat competitors, and boost profits. John Wiley & Sons.

Dean, J. (2014). Big Data, Data Mining, and Machine Learning - Value Creation for Business Leaders and Practitioners. John Wiley & Sons, Inc.

EMC Education Services. (2015). Data Science and Big Data Analytics - Discovering, Analyzing, Visualizing and Presenting Data. John Wiley & Sons.

Englmeier, K. (2015). Role and Importance of Semantic Search in Big Data Governance. In *Big-Data Analytics and Cloud Computing* (pp. 21–35). Springer International Publishing. doi:10.1007/978-3-319-25313-8_2

Fayyad, U. (2019). *Skill-Set Analysis on LinkedIn: Data Scientists, Engineers, Analysts vs. Job Postings.* Available at: https://www.iadss.org/post/skill-set-analysis-on-linkedin-data-scientists-engineers-analysts-vs-job-postings

Frempong, G. (2012). *Evidence for ICT Policy Action - Understanding what is happening in ICT in Ghana: Policy Paper*. Research ICT Africa.

Gorelik, A. (2019). The Enterprise Big Data Lake - Delivering the Promise of Big Data and Data Science. O'Reilly Media.

GRA. (2020). *Communication-service-tax.* Available at: https://gra.gov.gh/communication-service-tax/

Gupta, S. & Shilpi, S. (2016). *Real-time big data analytics - Design, process, and analyze large sets of complex data in real time.* Packt.

Hagstroem, M. (2015). Big Data Analytics for Inclusive Growth:How Technology Can Help Elevate the Human Condition. In S. Dutta, T. Geiger & B. Lanvin (Eds.), *The Global Information Technology Report 2015: ICTs for Inclusive Growth.* Geneva: World Economic Forum and INSEAD.

ICA. (2019). *Spending by African governments on Infrastructure.* Available at: https://www.icafrica.org/en/topics-programmes/spending-by-african-governments-on-infrastructure/

ITU. (2020). *Measuring Digital Development: ICT Price Trends 2019*. ITU Publications.

Kordha, E., Gorica, K. & Ahmetaj, L. (2011). Managing IT Infrastructure for Information Society Development. The Albanian Case. *Romanian Economic and Business Review,* 6(2), 122 - 131.

LinkedIn. (2018). *LinkedIn Workforce Report: United States.* Available at: https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018

Maayan, D. G. (2018). *Big data as a service (BDaaS) solutions: comparing IaaS, PaaS and SaaS.* Available at: https://hub.packtpub.com/big-data-as-a-service-bdaas-solutions-comparing-iaas-paas-and-saas/

Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley and Sons Ltd.

Otava. (2017). *The importance of infrastructure to Big Data analytics.* Available at: https://www.otava.com/blog/the-importance-of-infrastructure-to-big-data-analytics/

Romero, C. D. G., Barriga, J. K. D. & Molano, J. I. R. (2016). *Big Data Meaning in the Architecture of IoT for Smart Cities.* Academic Press.

Sharma, S. (2016). Expanded cloud plumes hiding Big Data ecosystem. *Future Generation Computer Systems*, *59*, 63–92.

Songwe, V. (2019). A digital Africa: Technology can be a springboard for faster, more inclusive growth. In Finance and Development. IMF.

The Royal Society. (2019). *Dynamics of data science skills How can all sectors benefit from data science talent?* The Royal Society.

Tozzi, C. (2019). *4 Big Data Infrastructure Pain Points and How to Solve Them.* Available at: https://www.precisely.com/blog/big-data/4-big-data-infrastructure-points-solve

Woodhouse, T. (2020). *Mobile devices are too expensive for billions of people and it's keeping them offline.* Available at: https://a4ai.org/mobile-devices-are-too-expensive-for-billions-of-people-and-its-keeping-them-offline/

Woodhouse, T., & Thakur, D. (2018). *Improving Mobile Broadband Quality of Service in Low- and Middle-Income Countries*. Web Foundation.

World Economic Forum. (2017). *The Future of Jobs and Skills in Africa: Preparing the Region for the Fourth Industrial Revolution World Economic Forum.* Geneva: World Economic Forum.

Yeboah-Boateng, E. O., & Nwolley, S. J. (2019). Developing Use Cases of Big Data Analytics: Data Integration with Social Media Metrics. In K. Skouby, I. Williams & A. Gyamfi (Eds.), Handbook on ICT in Developing Countries: Next Generation ICT Technologies. Rivers Publishers.

Zibi, G. (2018). *Closing the Investment Gap: How Multilateral Development Banks Can Contribute to Digital Inclusion.* Washington, DC: A4AI, Web Foundation, and Xalam Digital Analytics.

Chapter 9
Internet of Things for Travel Services

Serkan Polat
Istanbul Medeniyet University, Turkey

M. Fevzi Esen
https://orcid.org/0000-0001-7823-0883
University of Health Sciences, Turkey

ABSTRACT

A variety of data sources are available from smart devices which are connected to each other via different communication protocols. These devices are designed to be used in human-centric environments through a distributed physical-virtual interaction. Internet of things (IoT) is a concept of gathering the variety of devices through wired or wireless connections anytime and anyplace. This helps to create new services by integrating the physical world into virtual systems within various domains of tourism. In this chapter, the authors discuss the importance of IoT data for travel services. In the study, the major challenges and opportunities of IoT that allow tour operators and travel agencies to improve the customer experience and provide personalized services are examined. It is concluded that although there are studies on the use of IoT applications within tourism industries, there have been very limited studies conducted on integrated applications of IoT, especially for tour operators and travel agencies.

INTRODUCTION

IoT has the ability of capturing the all sources of value and it creates the systems that can maximize business value by new business models and management innovations. According to Gartner (2017), 8.4 billion connected devices are in use worldwide and total spending on IoT services is about $2 trillion in 2017. As an ever-increasing number of devices become associated with the internet, there are in excess of 21.7 billion active and interconnected devices worldwide at the end of 2020 and it is anticipated to attain 70 billion by 2025 (IoT Analytics, 2020). More than 65% of the businesses are expected to adopt IoT products including a combination of devices such as sensors, gateways, mobile devices, and cloud

DOI: 10.4018/978-1-7998-6792-0.ch009

applications (Gartner, 2019). According to Vodafone Report of 2020, IoT enabled businesses to decrease their operating costs by an average of 21% and it creates new revenue streams (Vodafone, 2020). It is also estimated that one trillion sensors will be connected to the global network by 2025 and $14 trillion will be generated via IoT devices for the global economy by 2030 (WEF, 2021).

The use of IoT produces massive amount of unstructured data. Almost all devices collect and send data, which have hidden and valuable knowledge. IoT technologies are a good option to discover knowledge from large amounts of data that have been produced from daily activities of people, tourists, business operations and all facilities relating activities. For this reason, IoT has become a very important mediator for various industries especially for tourism. IoT help businesses to improve their service quality by integrating business services, operations and visitors'/tourists' experiences with an effective planning.

To deliver goods and services to the customers, the sectors with instant data sharing such as travel services and accommodation need the use of IoT services more than other sectors. Since the nature of touristic services are not eligible for before-you-buy-experience, it may not be possible to meet expectations of the customers without digital interaction and IoT technologies. From this point of view, IoT has a crucial role to produce huge amount of customer data and process data flows for creating value with new insights.

Digitalization and informatization are the future trends of the integration of the IoT technologies into the tourism industry (Gretzel, 2011). The data generated from IoT devices related to daily key activities of habitants of a destination city, tourists or even other people, can be used in investigating the impact of certain events or decisions for tour operators and travel agencies. These data must be processed using various analytic tools for high quality of product and service delivery to customers. IoT technologies can enhance creating tour packages and alternative routes in a destination, service differentiation, revenue forecasting as well as and an efficient decision making on travel services and tour operations.

In tourism industry, travel agencies and tour operators are two different types of businesses operating on the basis of mutual cooperation and dependency. Tour operators are mostly wholesalers that create mass package tours and make collective agreements with service providers while travel agencies are retail businesses that mediate in the sale of the package tours. Although the activities of both businesses differ from each other in terms of scope and content, their mutual dependency is high. For this reason, tour operators and travel agencies must have a technological infrastructure in order to adapt to technological developments as soon as possible. IoT technology is a suitable tool that integrates different forms of interaction both through walk-in operations and online services in travel services and tour operations domain. Valuable data can be acquired from real-time, near real-time or batch processes through IoT devices. In case of considering the importance of the big data and analytics which are the key tools to extract valuable information from the data, IoT environments have a crucial role to improve tourism services for tour operators and travel agencies (Polat et al., 2019).

In this study, we try to discuss the importance of IoT for travel services and tour operations, in a subfield of tourism operations. We firstly describe the concept of travel services and tour operations in tourism terminology. Then, we discuss the types and the sources of data, which is generated by IoT applications towards tourism services and operations. Finally, we conclude the chapter with major challenges and opportunities of IoT for tour operators and travel agencies.

CHARACTERISTICS OF TRAVEL SERVICES SECTOR

Intangibility, heterogeneity, perishability, inseparability, and lack of ownership are the characteristics of travel, tourism and hospitality service sector businesses (Robinson, 2009). The travel industry has evolved along certain well-defined lines that have increased demand worldwide for their services. The travel impetus has intensified over the years and has resulted in the spread of travel agencies around the world. Travelers want to get rid of the difficulties of making travel arrangements in advance and having to deal with the various pre-travel arrangements for which they have limited knowledge (Bhatia, 2012).

Sector features help to define travel industry better than some definitions. It usually involves traveling away from home and traveling with a different type of spending. It is likely to be an enjoyable or interesting event for the participants and at the same time may include all aspects of tourism, travel, hospitality, activities and leisure, including accommodation, activities and attractions. All of these definitions show important relationships. Travel, however, includes some other components that are easier to identify. In many cases, travel includes tour operations and holiday packages (Robinson, 2009).

Travel agencies are the result of the impulse of organized individual travelers. That situation fundamentally signifies using the means of travel companies holding a position to offer travelers a relatively low price, since these companies have an agreement to collectively purchase these means/products and to transfer them to potential travelers through payment. Moreover, these means/products provide transportation, accommodation, visiting museums, city tours and other products/services potential travelers want to receive by making the appropriate payments (Bhatia, 2012).

The distribution channel in tourism, where different businesses interact and are connected to each other, enables a travel business that produces travel-related products within a system and structure to sell products and services to potential customers and make arrangements based on customer demands (Leuterio, 2007). Distribution channels are complex systems of behaviour in which people and companies interact to reach their goals (Cooper and Lewis, 2004). The distribution channel in tourism provides interaction with the tourism industry, which includes many units such as accommodation, travel businesses, food and beverage businesses, visiting places, and entertainment places and those who will demand the products and services offered by this industry is a very important process (Cooper and Lewis, 2004).

Availability and accessibility can be defined as the two main objectives of tourism distribution (Lubbe, 2000). According to Lubbe (2000), the former expresses the non-storability and abstraction characteristics of the tourism product, unlike other products. Therefore, travel businesses fulfil the function of promoting and presenting tourism products to potential customers. The latter refers to different levels of distribution channels that enable the purchase of tourism products and services other than where they are offered. On the other hand, the North American Industry Classification System (NAICS) defines travel operations and services as destination marketing organizations (DMOs), online travel agencies (OTAs), tour operators, travel agencies, and other institutions (Knowles and Westcott, 2015).

According to Singh (2008), a travel business prepares and sells travelers individual or full packages including transportation, accommodation, tours, and other related products and services for travelers' needs and expectations. Lavery (1992) defines a retail travel agency as a business that performs the following functions (Bhatia, 2012):

- selling prepared package tours, preparing individual itineraries, personally escorted tours and group tours;

- arranging transport; selling airline tickets, rail, coach and cruise trips and arranging car hire abroad;
- arranging hotels, motels, sightseeing trips, music festivals, transfer of passengers between terminals and hotels;
- handling and advising on many details involved in travel especially foreign travel such as travel and luggage insurance, medical insurance, travelers cheques, visa requirements and so on;
- providing information and advice on airline, rail and coach schedules and fares; hotel rates; whether rooms have baths; whether their rates include local taxes. All of this information can take days of the intending tourist's time or weeks of endless phone calls and letters;
- arranging reservations for special interest activities such as business travel, sporting holidays, religious travel, pilgrimage, etc.
- in the case of legitimate complaints from customers writing to the principal (tour operator or airline) to try and get a refund or a written statement or apology for any mishaps that may have occurred;
- interpreting and advising clients of the many complex discounted fares offered by the airlines and to warn clients of 'overbooking'.

Tour operators create and offer packages consisting of at least two of transportation, accommodation and other important services. Moreover, travel organizations offer a lot of services for business and leisure tourism, from package holidays to individual accommodation, transportation and other arrangements. At the same time they provide secondary products such as advice and currency exchange, travel insurance (King et al., 2005). According to Cavlek (2006), while a tour operator is a wholesaler, a travel agency is a retailer. For this reason, travel businesses are not confronted with unsold capacity risks because they do not pre-purchase suppliers' services. In other words, a travel agency only sells a product or service when a purchaser occurs. Nevertheless, a travel business sometimes can be as a particular tour organizer.

The most important information systems are in the tourism sector by airlines such as American Airlines and United Airlines computer reservation systems (CRSs) that are developed to make it easier operations, ticketing, tariffs and price quotations (Lubbe, 2000). American Airlines, in 1963, introduced the first online CRS with its Semi-Automated Business Reservation Environment (SABRE), which was the largest computer network outside the U.S. government because of its expanding from Canada to Mexico within American Airlines (Maurer, 2003).

Technological developments are also responsible for the growth of a new type of travel agency specialized in e-commerce (electronic business). For example, ebookers and Expedia provide online accommodation, travel, car rental, etc. opportunities for travelers (King et al., 2005). In other words, advances in technology have led to the emergence of new distribution channels. Internet, which is one of these, brings producers and consumers together very quickly and easily all over the world. Moreover, that reduces the need for travel agencies, which are traditional distribution channels. Electronic channels offer more products and service options to consumers. The use of these new channels depends on the consumer's knowledge of the technological know-how and the knowledge of travelers (Lubbe, 2000).

Offutt (2013) states that OTAs and other technology providers can benefit operators and the travel services sector as a whole. Keeping in mind that travel services pertain to the planning and reserving of trip components, recent beneficial technologic improvements, these innovations will probably increase as more progress is made, include the following (Knowles and Westcott, 2015):

- real-time and automated inventory management, ensuring operators and travelers alike are working with accurate information when planning and booking
- a pollution and weather detection chip that would help tour operators, transportation providers, and visitors anticipate, and plan for changes in conditions
- personalized information presented to visitors to help them narrow their choices in the trip planning process, ensuring users are not overwhelmed with information, and making the most of limited screen size on mobile devices and tablets
- social technologies and on-the-go information sharing, allowing users to plan at the last minute as they travel
- virtual assistant holograms and tablets carrying information that can replace humans during the travel experience (for instance, at airport arrivals and visitor centres)

EasySABRE^{SM} was created when Internet emerged to provide convenience to those who travel frequently and those who have easy access through computers to the SABRE (Maurer, 2003). Moreover, the first travel businesses were linked to Apollo and SABRE in 1976 and one year later, American Airlines declared that it has three hundred online agencies worldwide. CRS's revenue-generating capacities (commissions) and ease-of-use make it possible for travel agencies to sell their tickets to major national airlines inexpensively and efficiently through travel agents, as well as their dependence on airlines (Maurer, 200).

Typically, some functions of a computer reservation system are (Lubbe, 2000):

- airline reservations, up-to-date availability of airlines,
- car-rental reservations,
- travel documents tailored to client requirements,
- selling and booking cruises and tour packages,
- fare displays (published and discounted) and itinerary pricing,
- hotel bookings, availability of hotels, rates, descriptions and special information,
- seamless ticketing,
- discount fares information and booking of travel-related services such as theatre tickets, opera, limousine or parking services,
- direct distribution to clients for booking travel arrangements but controlled by the travel agency,
- personalising the reservation system through integration with the capabilities of devices.

Information services use a variety of media including written, audio-visual systems, the Internet and virtual tours and verbal information (King et al., 2005).

THE INTERACTION OF TRAVEL SERVICES AND TOUR OPERATORS

One of the biggest changes that have affected the sector lately is the information and communication technologies (ICT). ICT dramatically changes the way many people book holidays, and tour operators and travel agencies have to respond to survive. Today, tour operators and travel agencies have their own websites to help them increase their market (King et al., 2005). Kim et al. (2006) argue that it is increasingly becoming common for travelers to plan and book a trip process. In addition, it is often repeated that electronic markets are more efficient than traditional markets due to technological means. The

common hypothesis was that as the cost incurred by people to search for information decreased, they would further expand their search for information and go beyond comparisons made in the traditional marketplace (Öörni, 2015).

Online marketing and e-commerce, which have gained momentum since the mid-1990s, have caused revolutionary change on all sectors. It would not be wrong to mention the hospitality industry as one of the sectors most affected by this transformation. Accommodation businesses, brokerage businesses, food and beverage businesses and other related businesses operating in the hospitality industry have taken advantage of this technology very quickly and tried to directly market their products and services to their customers. In addition, these businesses wanted to take advantage of this transformation in order to facilitate transactions with customers, receive feedback from them, collect and analyse data (Freeman and Glazer, 2015). The travel industry is complicated by its component parts: tour operators, travel agencies, transportation, accommodation, activities, events and attractions. It can be defined by the size of the businesses in the travel industry (Robinson, 2009):

- multinationals are usually leading to the markets. Many independents (see below) are owned by these major operators and fulfil the more niche-market offerings.
- national multiples are those businesses which are based across the UK, and include Leger and Midland Counties Cooperative.
- regional multiples are those businesses operating within a region, such as the West Midlands or London.
- independents are the individual operators of generally very small businesses and often specialists.

Unlike consumers of manufactured products, potential tourists cannot test or "try" a destination before buying it. This places a great responsibility on travel agencies and tour operators having extensive information about services and products presented by producers and these intermediaries take into account the interests of their customers. This is the more information and responsibility level that distinguishes the travel agent from the manufacturer (Lubbe, 2000). Table 1 shows the changes in travel retailing.

According to Robinson (2009) the travel industry can also be defined according to the type of business it is involved in. These are as follows:

Business Travel: This includes travel for business purposes. International business travel is threatened by the growth of technology and the affordability of communication technology.

Mass-Market Travel: As the most popular travel type it is still trying to improve the image of this sector. This includes a range of business areas from clubs to high-level skiing holidays.

Niche-Market Travel: This means more specialized travel movements. They can be packages or independent travel and usually focus on specific interests. General niche market products include trips designed to learn new skills, such as language, art, or adventure skills, or discover new emerging places.

Sustainable, Green and Eco-Travel: When the growing awareness of environmental issues is combined with the desire to be more environmentally friendly, it is enough to say that it brings all kinds of problems in the sector to the agenda.

Independent and Specialist Travel: The independent market has gained in popularity since the early 1980s, led by specialized companies such as Travelbag.

Domestic Travel: A simple description here will suffice, domestic travel is within the country of origin of travelers using local, regional and national suppliers.

Table 1. Changes in travel retailing (Source: Lubbe, 2000)

Period	Trading environment	Type of travel retailing
1950s	• Limited demand for holidays or other travel. • Advent of more air travel for leisure purposes.	• Establishment of travel agents located in major urban areas. • Limited competition.
1960s	• Development of demand for leisure travel, particularly in air travel, with decline in sea travel.	• Gradual increase in city centre travel retailers.
1970s	• Rapid expansion in demand for holidays.	• Successful retailers expand the number of outlets – proliferation of city street front retailers.
1980s	• Development of out-of-town shopping malls and large-scale town centres. • Many city centres suffer from shop closures.	• Computerised reservations system for individual and inclusive holidays. • Larger travel agency chains grow by acquiring smaller agencies, consolidating ownership and putting pressure on independents. • Development of specialised holiday shops, and decline of full-service travel agencies.
1990s	• Increasing financial pressure on travel retailers, increasing rate of acquisition and mergers.	• Increasingly selective racking policies. • Technological developments enable customers to create their own holiday packages by booking direct from home.

Travel and tourism agents act as representatives of the interests of those who demand tourism-related services and as coordinators of suppliers in the market. Tourism consumers want to use the services of intermediary institutions because they help them organize their travels and many other services they will use in touristic establishments (Cavlek, 2006).

The most popular way for consumers to purchase travel tickets and other related facilities is to use online travel services. This, in turn, causes traditional travel agencies to take the last place in the customer's preference (Kim et al., 2006).

Suppliers of tourism services should decide how to deliver their services most effectively to potential tourists. Each airline, accommodation provider, destination marketing organization and travel agency within the tourism distribution system wants to reach the maximum level of knowledge of potential tourists. The aim is to encourage purchasing decisions so that reservations and payments can be made (Lubbe, 2000).

Since internet and online platforms eliminate the need for intermediaries, travel businesses are stated as the businesses that are most at risk in this change in technology (Anckar and Walden, 2001). In addition, through mobile technology that allows interactive interaction, many hotel businesses and other tourism businesses had the opportunity to reach existing and potential customers through their websites. Therefore, tourism businesses are now able to communicate with their customers 24/7, reduce their costs, and exist all over the world (Freeman and Glazer, 2015). The ICT is the preferred platform after being popular rapidly for many business models, as well as providing different ways of obtaining information and providing an efficient and easy distribution opportunity for retail manufacturers, causing many business concepts to change (Kim et al., 2006).

In 1996 SABRE launched Travelocity as the first full service offered to web users for online travel bookings and information (Maurer, 2003). Kim et al. (2006) investigated nine features used by customers in order to select an OTA and found those features including useful content, speed, website design, ease of use, flexibility in booking, security, finding low fares, ability to book all travel services in one transaction, and sorting option. According to the results of their study, finding low fares is the most fundamental feature followed by security.

Platforms such as expedia and travelocity are used by millions of internet users to meet the expectations and needs of tourism and travel. The volume of online bookings in 1997 increased almost four times compared to the previous year, as many airline companies also provided the opportunity to book and purchase online for their passengers. According to another indicator announced by the American Travel Industry Association, the monetary value of reservations made via the internet, which was $827 million in 1997, will reach $9 billion in 2002 (Maurer, 2003). Experts also believe that e-commerce has great potential for travel businesses to use. More and more travel businesses are trying to communicate with customers using internet communication technology. These developments have led to travel businesses to be among the most successful e-commerce businesses (Kim et al., 2006).

Rapid and radical changes in technology have made it easier for consumers to cope with their travel suppliers and this situation has given more momentum to the changing face of travel distribution. Along with these trends, consumer needs have also changed fundamentally. Travelers are looking for different kinds of experiences than their predecessors, and this is the biggest challenge facing the travel industry as well as more accessible information (Lubbe, 2000). In today's e-commerce market, it is of great importance that potential customers reach satisfactory information about the products and / or services they are looking for. Therefore, the fact that people are not given the information they are looking for by the companies or they cannot access this information due to the system significantly limits the customer efficiency (Öörni, 2015).

The role that intermediaries such as travel agencies and tour operators play on the market today is not being left out with the individualization of tourism travel or the new possibilities offered by modern information technology to customers of travel services. The reason for this is that they are rapidly adapting to the changing business environment of travel and tourism agents (Cavlek, 2006). Travel businesses did not remain indifferent to these changes caused by the Internet and online platforms and developed three significant methods (Lubbe, 2000):

- the consolidation of travel companies through mergers, acquisitions, franchising and the formation of consortiums.
- the increasing use of technology to benefit both the client and travel company.
- the restructuring of their revenue base, where a strong move is emerging away from the traditional compensation model of commissions to a more flexible service-fee model.

Along with the explosive increase in online travel bookings made by consumers, many companies have begun offering online bookings to take advantage of a new distribution channel from the internet and reduce costs such as labour and commission fees. Intensive competition is a natural result of maturation. Because of the growing competition, marketing executives of online travel agencies should know their positions according to their competitors. The various aspects of the customer's purchase decision regarding online travel products are taken into account by the customer's perception of a particular brand (Kim et al., 2006).

Law, Leung, and Wong (2004) wanted to reveal the perception that travel businesses will disappear due to the internet. According to the results of this study on tourists, people believe that travel businesses will survive. It is also believed that professional advice and service is provided through travel businesses. Therefore, it can be said that travel businesses and the internet will continue to exist together.

IOT TECHNOLOGIES FOR TRAVEL SERVICES

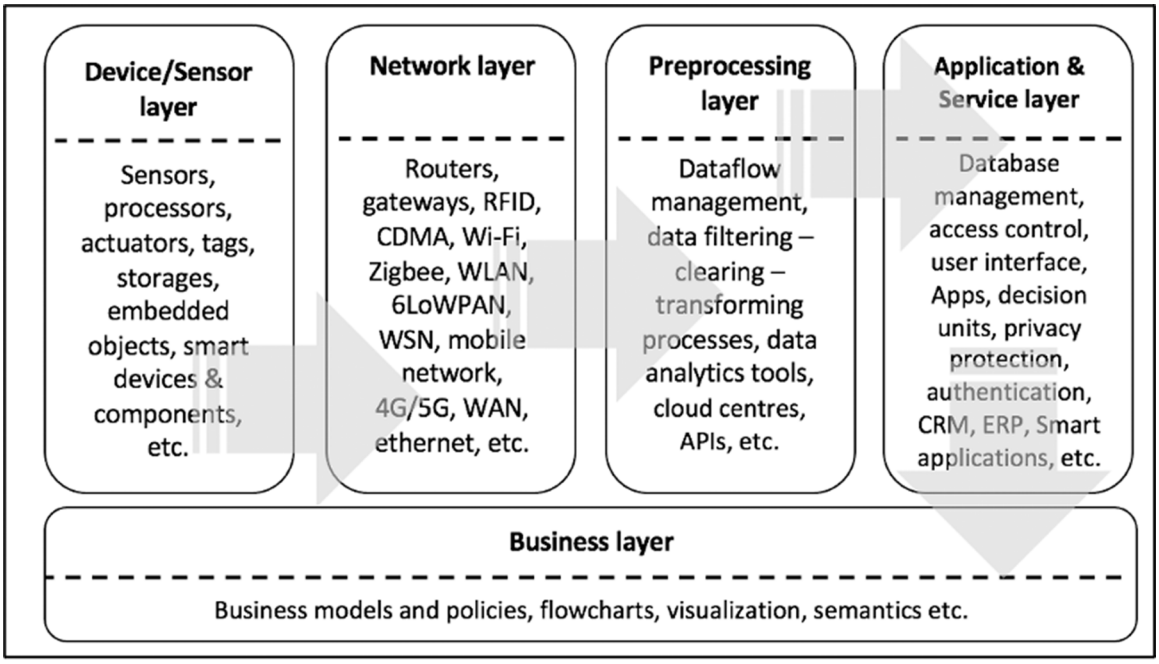
IoT refers to all objects that can communicate with another such as devices, sensors and modules. Today, as a widely accepted definition for the concept of IoT is more than one object or platform that can store, process and analyse data. Kranenburg (2008) states that IoT is a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols in which "Physical and Virtual Objects" are identifiable. IoT has physical qualities and virtual personalities and it uses smart interfaces, which are integrated into the information networks. IoT can also be expressed as the systems that allow objects to connect to the internet via embedded sensors, either wired or wireless (Banger, 2018).

Today, computers have been integrated into IoT devices or objects This makes IoT become practical and state-of-the-art alternative for businesses. Ease of installation and instant reaction to the needs with real-time data are the most important factors that widens the usage of IoT applications (Lee and Lee, 2015). The main features of an IoT system can be listed as follows:

- Connectivity: The main role of IoT is to provide data exchange between systems, networks and people. Objects must be capable of being connected to the global information and communication network.
- Heterogeneity and interoperability: It refers to the diversity of IoT devices and subsystems that can interact with different platforms. The operability feature indicates that two or more devices, systems, networks or platforms communicate with each other seamlessly and share data over IoT system.
- High scalability: There are many connected devices in IoT systems. Big amounts of data can be generated through the systems and evaluated in real-time. The data can be transformed into knowledge by using big data analytics for deep and sophisticated perspective.
- Cost advantage: IoT applications have a cost-reducing effect in many businesses as they can continuously monitor the actual performance and status of the equipment involved in business processes. According to McKinsey report on IoT and its economic impacts, it will generate $1.2 trillion to $3.7 trillion saving per year, including productivity improvements and avoiding failures (McKinsey, 2015).

The organization and configuration of the physical components in the IoT system vary according to the procedure of the operation and the specification of the data. Objects are physically and virtually connected in IoT. It is also stored, processed and accessible virtually. However, many IoT architectural taxonomies have been proposed in the literature according to architectural elements and application areas. Basically, IoT architecture is composed of five layers as device/sensor layer, networking layer, preprocessing layer, application and service layer and business layer (see Figure 1).

Figure 1. IoT layers
(Source: Formed by authors)



Device/Sensor Layer

In this layer, objects are easily tagged with wireless sensor nodes, processors, and actuators. These devices can measure the mechanical, thermal, chemical, electrical or magnetic properties of their environment or the object and they are classified according to their functions and application areas. The sensitivity of a sensor shows how much the sensor's output is changed when the amount of inputs change. In particular, the layer acts as a bridging tool in terms of energy efficiency, performance, reliability and interoperability. A suitable topology for the network equipment that provides data transmission with sensors, is essential.

Network Layer

This layer is also known as communication layer. The things interact with each other or with an IoT gateway and the large amount of data produced in the device layer is transferred to the system. Due to the configuration differences of the devices in the IoT system, the need for high speed communication and context sensitive access, multiple networking technologies and access protocols have been developed such as 4G, RFID, Wi-Fi, Zigbee, and so on. Large number of objects that need to be included in the communication on the IoT system and security policies entail various protocols such as IP, IPv6, ICMP etc. to exchange data between the layers.

Preprocessing Layer

Real-world data is imperfect and inscrutable which consist of errors, noise and missing values. Data preprocessing layer is an important step to convert the raw data into useful information. In this layer, quality of data is assured via cleansing, transforming, binning, and extracting processes.

Application and Service Layer

This layer is responsible for providing services regarding data transmission and information management. It also provides a cost-effective platform for IoT-based applications where hardware as well as software platforms can be redesigned. In this layer, the management of services required for the storage of information in databases, access control, privacy and protection, and smart applications are also managed. The layer also ensure the accuracy, integrity, and confidentiality of IoT data.

Business Layer

The business layer manages applications, models and policies of the IoT systems in a user-friendly way. It is also responsible for integrating IoT-generated information into the business processes towards business strategies and day-to-day transactions. Privacy concerns and trust challenges of users are addressed in this layer.

IoT technologies integrate data resources and provide customized and personalized services notably in a variety of fields related to tourism (e.g. travel, transportation, safety, housing, cultural facilities, and healthcare). It ensures a smooth integration of different types of services to establish innovative applications to improve overall experience of people as well as achieving an efficient collaboration between individuals, organizations and communities. Although there are significant overlaps between travel and tourism industries, the scope of travel industry generally covers tourism agencies' and tour operators' activities that cater for travel purposes such as physical, cultural, personal needs and pursuit of hobbies. The industry provide services from transportation (e.g., airlines, railways, car rental, local services), accommodation (hotels, camping, cruises,agri-tourism etc.), food and beverage (e.g., restaurants, catering, cafes), entertainment (casinos, bars and clubs, gaming), recreation (e.g., exhibits, festivals, museums, historical sites, performing arts, amusement) sectors and so on (Camilleri, 2018).

Travel agencies and tour operators are the businesses responsible for delivering tourism products or services to the consumers. Considering the fact that consumers are faced with a large number of options that complicate the decision process in purchasing products and services, IoT holds an enormous potential to facilitate automation for businesses and boosts customer experience by personalization of services. (Stylos and Zwiegelaar, 2019). It also comes up with novel approaches to simplify processes and tasks in business operations and minimize the effort and cost for product development and marketing. Based on this fact, IoT technologies also have another aspect that establishes information integration and network infrastructure which summarize every concept related to tourism product and services. This phenomenon is defined as smart cities that connect physical, social, business and IT infrastructure and creates value and pleasure for tourists and service providers (Buhalis and Amaranggana, 2015). IoT is the one of the major components of smart cities that are driven by heterogeneous systems and devices to employ complex tasks (Kim et al., 2017).

There are three different goals in providing communication in IoT. Human-to-human communication refers to a natural communication when humans interact through talking, body language, media or simpler devices/objects. In human to machine communication, people use a variety of devices such as sensors, virtual and augmented devices or networks to transfer information. For example, vacation planning over internet search engine or booking airport transfers via online reservation systems can facilitate human-machine communication to fulfil travel needs. Machine to machine communication is automated exchange of data between devices without human interaction. It can connect millions of wired or wireless objects through a single network and objects can be physical devices and products, as well as logical content and resources. A wearable sensor can obtain movement based parameters or biomedical data of a tourist during his walking tour and transmits data from the sensor to a mobile device. Then, a mobile application can analyse data and make suggestions for a better tour planning or sightseeing adventures in various kinds and length. A tour operator or service provider can create holiday packages considering a combination of such kind of data and tourism business information and devises the whole trip.

Tour operators prepare travel, accommodation, culinary, entertainment and guidance services as packages and sell them directly or through travel agencies. On the other hand, travel agencies are the distribution agents who provide information about travel opportunities, touristic service prices, quality and conditions, resulting a certain commission from their sales. They play an intermediary role between customers and service providers. Travel agencies make arrangements according to the circumstances that occurred prior to the tour or last minute changes, as well as accommodation, food, transportation and guiding services. Unlike travel agencies, tour operators both make agreements with suppliers and also create marketing activities for customers by organizing tour packages. The emergence of ever-increasing amounts of large-scale diversified data available from various sources have provided significant opportunities for tour operators and travel agencies to have information about competitors and the market, create demand for services and products, provide information about the existence of goods and services, plan business functions, and communicate with consumers as in many businesses. Acquisition and analysis of such data depicting complex dynamics (e.g. real-time, social network, text, sensor, spatial or UTC) is likely to reveal a rich context for tourism interventions. Interdisciplinary approaches combining with complex data types, physical or virtual devices with embedded sensors, database systems, data management disciplines, and analysis are urgently needed to boost transformative innovation for enabling effective and timely solutions to the businesses operating in the tourism industries such as tour operators and travel agencies.

Traditionally, travel agencies and tour operators are intermediary businesses in the tourism industry. Among these businesses working in a wholesaler-retailer relationship, commissions are reflected on customers. Different kinds of commissions and fees arise, in case of more businesses involve in the distribution channel. This also affects the price that end users have to pay. In addition, the effectiveness of information and product flow between businesses gains importance according to the size of tourism supply chain. IoT enabled supply chain can allow an enhanced system for data and process flow resulting in cost savings (Kumar et al., 2019). Services to be provided through IoT by the tour operators (supplier) to travel agencies (intermediaries) will create satisfaction and trust for customers At this point, IoT applications completely change the structure of the mutual interactions of the businesses operating in the tourism industry (Skelia, 2018; Car et al., 2019).

It is important that consumers must have all available information about tourism services and products before purchasing since they are intangible and produced - consumed at the same time. Tourism distribution channel is a structure formed by internal business units or network of intermediaries for the

service and product passes from the supplier to the consumer. IoT technologies can provide distribution channel optimization in tourism industries by providing high volume and real-time data flow. Data can be transferred from interconnected devices across augmented systems in a synchronized manner and consumers can be constantly aware of products and services anytime and anywhere. For businesses, IoT-connected operations boosts efficiency with real-time updates and minimise costs to resolve issues. As a result, IoT technologies not only stimulates tourist's satisfaction and helps them make better decisions but also offers innovative opportunities to the businesses (Huang et al., 2017; Brous et al., 2020).

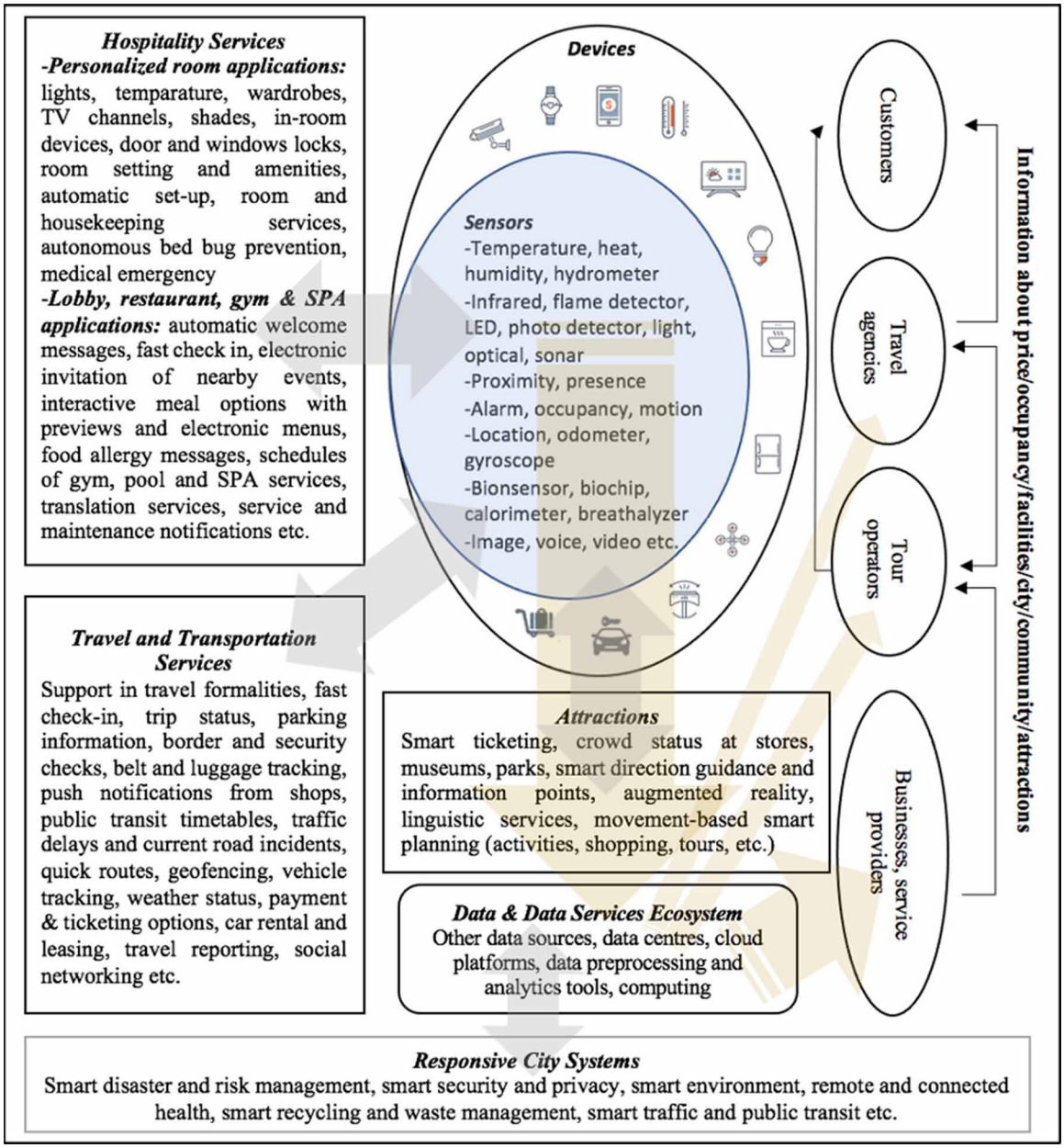
Proliferation of smart devices and sensors play an important role in changing the perspective of the businesses towards following a human-centred approach. Businesses develop applications which help customers spend more time at mobile applications, websites or blogs and prefer their services. Users' interaction with those devices through touching, speaking, sensing or gazing has extended the paradigm of interaction that makes this new technology usable by everyone, not just experts or researchers. In this way, businesses can find chance to acquire raw data (e.g. IoT sensor data, graphics, sound, video, text) from various tangible data sources that prove faster and effective transfer of information. For example, sensors augmented with everyday objects can collect vibration or movement data of the guests to give them advice on food consumption according to the user's food logs and how to reach the user's goals with its smart coach. A smart t-shirt based on IoT can collect energy consumption, skin temperature or respiration rate data of the guests. Service providers can offer personalized services (e.g. restaurant, fitness and spa, medical) services for the guests. Domestic and foreign tourism demands of a particular region, occupancy rates of hotels, number of overnight stays and average length of stay can be estimated with data obtained from IoT devices. These are collaborative devices that work well with products, customers, information providers (e.g. tour operators and travel agencies) and suppliers.

As stated in Figure 2, tourism products and services are shaped by mobility and technological advancements that brought a new form of transition and communication. Most of IoT data are unstructured and require preprocessing and storage services, analytic tools and computing services compared to traditional technologies. To achieve competitive advantage over the parties, businesses also get benefit of responsive city systems by merging and processing raw data to discover knowledge and develop the innovative methods. For example, the term "smart living" focuses on touristic attractions, housing options, cultural facilities and sport activities, safety and social cohesion in a city. It exposes operational data on city's environment such as traffic congestion, public safety and availability of public transportation systems by instrumented and interconnected IoT devices (Westraadt and Calitz, 2020). Those physical and virtual sensors provide real time information source via computing platforms in order to convey this information to travel agencies, tour operators, policy makers and business networks.

Urban data are generated by social and wearable devices or other sensors that are embedded into buses, taxis, roads, railways, streets, shopping centers, museums, buildings even services and products. This massive amounts of data is generally referred as crowdsourcing. Unlike traditional data sources (e.g., newspapers, annual reports, press releases, flyers, surveys), crowd-sourced data is generated by the people and city entities when the devices (e.g., smartphones, RFID readers, NFC-enabled devices, sensors) are activated or involved in social, economic or civic activities. Mobile applications also collect web search queries, customer ratings and trends that are created upon people's advices, opinions, habits and concerns. Such data can be easily implemented in information flow between service providers and tourism agencies – tour operators.

On the tourism services side, the guests can check-in/check-out by kiosks or mobile applications. In either case, the guest does not need the assistance of the receptionist. However, in some hotels, the guests

Figure 2. The use of IoT in travel services
(Source: Formed by authors)



perform these operations through interactive sensors in the room. The request of the guest is transmitted via wireless communication to the hotel management system. This procedure is completed upon approval of the request. At the same time, guests can make room orders, restaurant and spa reservations via interactive smart TVs or hotel room controllers for room services, lighting and curtain control, daily weather forecasts, city attractions, local restaurants or timetables. In addition, it is possible to process personal

data obtained through the wristband to suggest personalized and exclusive attractions and contact with the personnel in emergency situations when the tourist's safety is in danger.

Through the integration of different data sources, new business models can be formed spatially and temporally. Tourism intermediaries can develop personalized services and products through a variety of sources or privately owned sensor systems. To the best of our knowledge, there has been very limited literature on integrated applications using IoT, especially for tour operators travel agencies. Potential applications are presented in Table 2. For example, a tourist may not be aware of baggage carousel problem or crowd at baggage reclaim area that became constant at a service provider. An airline company may be losing luggage more than usual or not showing enough attention to luggage in some airport terminals. The data gathered from sensors can enable the assessment of the current physical status of the luggage and the airport. Or, people generally may not have information about the maintenances or long-term repairs at the swimming pool, recreational areas or casino services of the hotel in which they plan to stay before purchasing a package tour. IoT technologies with data services can be used effectively by travel agencies or tour operators to become aware of such problems in real time.

CONCLUSION

People requesting tourism products and services are entitled to temporary use of these products instead of owning them. For example, a double room booked for one week's accommodation must be ready to serve to another person at the end of the specified time. Also, there is no chance to try the reserved room in advance. Therefore, customers have to rely on the information provided by the intermediaries, such as tour operators and travel agencies, about the requested room. This shows the importance of distribution channels in the travel industry. On the other hand, advances in information technology have provided customers with direct access to information on tourism products and services. In addition to internet sites accessed through traditional electronic devices, some travel applications enable customers to compare alternative products and services and to choose between businesses that meet expectations. Mostly, tour operators and travel agencies are responsible for facilitating this process. This has led to the questioning of the function and future of traditional travel agencies and tour operators.

Companies are experiencing challenges more complex than ever before, and they have been looking for efficient ways to solve business problems. Recent technological advances led to methods that can be successfully used to solve such problems. As the number of technological sensors and devices have been growing rapidly day by day, IoT has been transforming every aspect of businesses and has eventually become a relatively critical concept for tourism industry. This technology provides a better insight and deeper understanding and allow real-time automated services for the intermediaries in tourism distribution system. Efficient usage of IoT devices with data services and responsive city can radically change the way of supplying tourism services and products for tour operators, travel agencies and service providers. It became possible to track human activities with electronic devices (e.g., smart phones, smart watches, wristbands), to monitor physical actions in hotels, restaurants, shopping centers, museums or cities by sensors, and also to integrate smart city systems to distribution channels in the form of networks and interactions in real-time while offering new services and products.

IoT technologies has begun spreading its presence in various business operations as well as penetrating travel industry by automated hotel rooms, lobby, restaurant and facilities, transportation services, attractions and so on. Similarly, cities have also been affected by IoT with smart applications. Adaptation

Table 2. Wearable IoT devices and potential applications for tourism industry

Device	Sensors	Network	Description	Other Applications
Wristband, bracelet	Heat, temperature, proximity, biosensor, fototransistor,GPS	RFID – WSN and NFC integration	The activity data of a tourist can be collected through the sensors inside the wristband. Service providers can offer city attractions, personalized tours or health benefits (insurance offers, social well-being packages, emergency, etc.) to the customer during the tour. In addition, NFC technology enables tourist to make contactless payments. Tour operators, hotels, or service providers with this expense data can create economic or customised packages (tour, attraction, shopping, culinary, etc.) for the tourists.	Doycheva et al. (2019); Ito (2020);
Smartglasses	Infrared, electromagnetic, CMOS, camera	ZigBee, bluetooth, Wi-Fi	Smartglasses can collect data regarding what people track. Augmented reality with smartglasses may provide benefits to enhance tourist experience by integrating societal behaviors of individuals/norms/influences during a cultural visit (e.g., art gallery, historical monuments). Moreover, smartglasses are able to correct visual deficiencies/impairments (e.g., blurred vision, colour blindness, myopia). It can be used to adapt customers to tourism services.	Obeidy et al. (2017); Sedarati and Baktash (2017); Han et al. (2019);
Embedded skin patches, fingertips	Tactile, pressure, infrared	Bluetooth, NFC, passive RFID	Electronic skin patches and fingertips have a multi-functional capability, which is utilized to collect real-time medical (e.g., vital signs, chemical indicators, metabolic processes) and sensing data (e.g., voice, breathing, movement, chewing, swallowing). For example, travel agencies or tour operators can plan interactive online tours or experiential tourism products to enhance virtual sensation of a touristic activity by obtaining and analysing such data.	Lee and Sharma (2016); Kim et al. (2017)
E-textiles (t-shirts, sweatpants, hats, scarf, glove)	Heat, humidity, biosensor, infrared, hydrometer, GPS	Bluetooth, Wi-Fi, RFID	Smart textiles bring a new challenge to IoT applications with electronic measurement and storage features. For example, a wearable clothing can receive biomechanical and geographical data by absorbing, spreading and transmitting. It also allows remote detection and monitoring of physiological measurements of the people to meet rapidly changing needs. For example, a tour operator can navigate their customers to green spaces or cool/warm urban/rural areas through electronic t-shirt or hat based measurements during their trips in real-time.	Singha et al. (2019); Tyler et al. (2020);
Sensing headbands	EEG, wavelet-based	Bluetooth, Wi-Fi	Sensing headbands read and assess brain signals to evaluate state of brain in real-time. In neuromarketing research, the devices have a wide range of use for detection of motivational responses to services and products. For example, service providers or tour operators can provide personalized packages or hotel selection options to their customers or enhance event or product based experiences by analysing neuroscientific data and intersubject synchronies of brainwave activities while getting the services. It can be also a useful tool for managing stress and mental well-being.	Yadava et al. (2017); Hsu and Chen (2020)

of IoT into smart city applications is essential for planning and management processes in travel services. The use of cloud platforms, computing, data services, and analytics is also a requirement while working with IoT technologies. To achieve these goals, it is critical to evaluate business goals, cost of operations, chain of intermediaries and data management processes in travel and tourism companies. The innovations in IoT revolutionize the travel industry. As long as the advantages of innovations are unfolding, it is crucial for tourism businesses to incorporate the IoT in their current strategies.

Tour operators and travel agencies have begun to use information technologies in an adaptive manner very quickly and effectively. That has resulted in the emergence of the use of IoT in travel services. While customers share their personal information only with tour operators or travel agencies in the pre-internet period, personal information with the internet has been shared directly and/or indirectly with many parties. The data of each customer (e.g., biomedical, mechanical, text, media, device, applications) are collected

and stored by various devices. It is important to ensure the safety of collected customer information and to use this information by businesses to meet customer expectations and demands.

REFERENCES

Anckar, B., & Walden, P. (2001). Self-booking of high- and low-complexity travel products: Exploratory findings. *Information Technology & Tourism*, *4*(3-4), 151–165. doi:10.3727/109830501108750958

Banger, G. (2018). *Endüstri 4.0 ve Akıllı İşletme, Dorlion Yayınları, 2*. Baskı.

Bhatia, K. A. (2012). *The business of travel agency and tour operations management*. Sterling Publishers.

Bhatia, K. A. (2012). *The business of travel agency and tour operations management*. Sterling Publishers.

Brous, P., Janssen, M., & Herder, P. (2020). The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations. *International Journal of Information Management*, *51*, 101952. doi:10.1016/j.ijinfomgt.2019.05.008

Buhalis, D., & Amaranggana, A. (2015). Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015*. Springer. doi:10.1007/978-3-319-14343-9_28

Camilleri, M. A. (2018). The Tourism Industry: An Overview. In *Travel Marketing, Tourism Economics and the Airline Product* (pp. 3–27). Springer Nature. doi:10.1007/978-3-319-49849-2_1

Car, T., Stifanich, L. P., Simunic, M. (2019). Internet OF Things (IoT) in tourism and hospitality: Opportunities and Challenges. *ToSEE – Tourism in Southern and Eastern Europe, 5*, 163–175.

Cavlek, N. (2006). Travel and tourism intermediaries. In L. Dwyer & P. Forsyth (Eds.), *International handbook on the economics of tourism*. Edward Elgar Publishing. doi:10.4337/9781847201638.00014

Cooper, C., & Lewis, J. (2004). Transformation and trends in the tourism industry: implications for distribution channels. In Tourism distribution channels: Practices, issues and transformations. Thomson Learning.

Doycheva, A., Glushkova, T., Ivanova, V., Doukovska, L., & Stoyanov, S. (2020). A Multi-agent Environment Acting as a Personal Tourist Guide. In O. Castillo, P. Melin, & J. Kacprzyk (Eds.), *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications. Studies in Computational Intelligence* (Vol. 862). Springer.

Freeman, R., & Glazer, K. (2015). Services marketing. In *Introduction to Tourism and Hospitality in BC* (pp. 167-190). BCcampus. Retrieved July 13, 2020, https://opentextbc.ca/introtourism/

Gartner. (2017). *Leading the IoT.* Retrieved November 12, 2020, https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

Gartner. (2019). *IoT press release*. Retrieved December 11, 2020, https://informationmatters.net/internet-of-things-statistics/

Gretzel, U. (2011). Intelligent systems in tourism: A social science perspective. *Annals of Tourism Research*, *38*(3), 757–779. doi:10.1016/j.annals.2011.04.014

Han, D. D. M., Dieck, C. T., & Jung, T. (2019). Augmented Reality Smart Glasses (ARSG) visitor adoption in cultural tourism. *Leisure Studies*, *38*(5), 618–633. doi:10.1080/02614367.2019.1604790

Hsu, L., & Chen, Y. J. (2020). Neuromarketing, subliminal advertising, and hotel selection: An EEG study. *Australasian Marketing Journal*, *28*(4), 200–208. doi:10.1016/j.ausmj.2020.04.009

Huang, C. D., Goo, J., Nam, K., & Yoo, C. W. (2017). Smart tourism technologies in travel planning: The role of exploration and exploitation. *Information & Management*, *54*(6), 757–770. doi:10.1016/j.im.2016.11.010

IoT Analytics. (2020). *IoT market outlook*. Retrieved November 22, 2020, https://iot-analytics.com/

Ito, E. (2020). Understanding Cultural Variations in Outdoor Tourism Behaviours for Outdoor Sport Tourism Development: A Case of the Blue Mountains National Park. *Tourism Planning & Development*, 1–7. Advance online publication. doi:10.1080/21568316.2020.1807401

Kim, D. J., Kim, W. G., & Han, J. S. (2006). A perceptual mapping of online travel agencies and preference attributes. *Tourism Management*, *28*(2), 591–603. doi:10.1016/j.tourman.2006.04.022

Kim, S.-W., Lee, Y., Park, J., Kim, S., Chae, H., Ko, H., & Kim, J. (2017). A Triple-Mode Flexible E-Skin Sensor Interface for Multi-Purpose Wearable Applications. *Sensors (Basel)*, *18*(2), 78. doi:10.339018010078 PMID:29286312

King, C., Kerr, A., Jefferies, M., & Brombley, D. (2005). *Travel and Tourism*. Heinemann Educational Publishers.

Knowles, H., & Westcott, M. (2015). Travel services. In *Introduction to Tourism and Hospitality in BC* (pp. 147-166). BCcampus. Retrieved from https://opentextbc.ca/introtourism/

Kranenburg, R. V. (2008). *The Internet of Things: A Critique of Ambient Technology and the All-Seeing Network of RFID*. Institute of Network Cultures. Retrieved December 17, 2020, https://www.networkcultures.org/_uploads/notebook2_theinternetofthings.pdf

Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: A review. *Journal of Big Data*, *6*(1), 111. doi:10.118640537-019-0268-2

Lavery, P. (1992). The financing and organisation of national tourist offices. *EIU Travel & Tourism Analyst*, *4*, 84–101.

Law, R., Leung, K., & Wong, R. (2004). The impact of the Internet on travel agencies. *International Journal of Contemporary Hospitality Management*, *16*(2), 100–107. doi:10.1108/09596110410519982

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431–440. doi:10.1016/j.bushor.2015.03.008

Lee, W., & Sharma, A. (2016). Smart sensing for IoT applications. *13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, 362-364. 10.1109/ICSICT.2016.7998921

Leuterio, F.C. (2007). *Introduction to Tourism*. Manila, Philippines: Book Store.

Lubbe, B. (2000). *Tourism distribution: Managing the travel intermediary*. JUTA.

Maurer, E. (2003). *Internet for the retail industry*. Thomson Delmar Learning.

McKinsey. (2015). *The internet of things: mapping the value beyond the hype*. Retrieved January 10, 2021, https://www.mckinsey.com/

Obeidy, W. K., Arshad, H., & Huang, J. Y. (2017). An acceptance model for smart glasses based tourism augmented reality. AIP Conference Proceedings, 1891(1). doi:10.1063/1.5005413

Offutt, B. (2013). *PhoCusWright's travel innovations & technology trends: 2013 and beyond*. Academic Press.

Öörni, A. (2003). Consumer search in electronic markets: An experimental analysis of travel services. *European Journal of Information Systems*, *12*(1), 30–40. doi:10.1057/palgrave.ejis.3000450

Polat, S., Esen, F., & Bilgic, E. (2019). Analysis of the 5Vs of Big Data in Virtual Travel Organizations. In *Big Data and Knowledge Sharing in Virtual Organizations* (pp. 43–70). IGI Global. doi:10.4018/978-1-5225-7519-1.ch002

Robinson, P. (2009). Travel and management: An introduction. In P. Robinson (Ed.), *Operations management in the travel industry*. CAB International. doi:10.1079/9781845935030.0001

Sedarati, P., & Baktash, A. (2017). Adoption of Smart Glasses in Smart Tourism Destination: A System Thinking Approach. *Travel and Tourism Research Association: Advancing Tourism Research Globally, 13*. Retrieved January 09, 2021, https://scholarworks.umass.edu/ttra/2017/Grad_Student_Workshop/13

Singh, L. K. (2008). *Management of travel agency*. ISHA Books.

Singha, K., Kumar, J., & Pandit, P. (2019). Recent Advancements in Wearable & Smart Textiles: An Overview. *Materials Today: Proceedings*, *16*(3), 1518–1523. doi:10.1016/j.matpr.2019.05.334

Skelia. (2018). *How Iot Is Changing The Travel Industry Today And Tomorrow*. Retrieved December 18, 2020, from https://skelia.com/articles/how-iot-is-changing-the-travel-industry-today-and-tomorrow/

Stylos, N., & Zwiegelaar, J. (2019). Big Data as a Game Changer: How Does It Shape Business Intelligence Within a Tourism and Hospitality Industry Context? In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*. Springer. doi:10.1007/978-981-13-6339-9_11

Tyler, D., Wood, J., Sabir, T., McDonnell, C., Sayem, A. S. M., & Whittaker, N. (2019). Wearable electronic textiles. *Textile Progress*, *51*(4), 299–384. doi:10.1080/00405167.2020.1840151

Vodafone. (2020). *IoT spotlight report*. Retrieved January 02, 2021, https://www.vodafone.com/business/news-and-insights

WEF (World Economic Forum). (2021). *Shaping the future of IoT and urban transformation*. Retrieved January 25, 2021, https://www.weforum.org/platforms/shaping-the-future-of-the-internet-of-things-and-urban-transformation

Westraadt, L., & Calitz, A. (2020). A modelling framework for integrated smart city planning and management. *Sustainable Cities and Society*, *63*, 102444. doi:10.1016/j.scs.2020.102444

Yadava, M., Kumar, P., Saini, R., Roy, P. P., & Prosad Dogra, D. (2017). Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications*, *76*(18), 19087–19111. doi:10.100711042-017-4580-6

KEY TERMS AND DEFINITIONS

IoT: IoT has physical qualities and virtual personalities, and it uses smart interfaces, which are integrated into the information networks.

IoT Layers: IoT layers are specialized modules that constitutes IoT architecture.

Sensor: Sensors are smart tools that collect, store, integrate and share data in various applications.

Smart Applications: The term represents different characteristics that is associated by a certain number of applications such as environment, economy, people, mobility and so on.

Tour Operator: A tour operator is mostly a wholesaler that creates mass package tours and makes collective agreements with service providers.

Travel Agency: Travel agencies are retail businesses that mediate in the sale of the package tours.

Wearable Devices: Wearable technology or wearable device refers to the technology that users place on their body without using their hands, etc.

Chapter 10
The Future of Artificial Intelligence in Education

Fati Tahiru
https://orcid.org/0000-0003-0874-0428
Durban University of Technology, South Africa

Samuel Agbesi
https://orcid.org/0000-0002-9527-1924
Aalborg University, Copenhagen, Denmark

ABSTRACT

The key accelerating factor in the increased growth of AI is the availability of historic datasets, and this has influenced the adoption of artificial intelligence and machine learning in education. This is possible because data can be accessed through the use of various learning management systems (LMS) and the increased use of the internet. Over the years, research on the use of AI and ML in education has improved appreciably, and studies have also indicated its success. Machine learning algorithms have successfully been implemented in institutions for predicting students' performance, recommending courses, counseling students, among others. This chapter discussed the use of AI and ML-assisted systems in education, the importance of AI in education, and the future of AI in education to provide information to educators on the AI transformation in education.

INTRODUCTION

This chapter presents the application of Artificial Intelligence (AI) technology in the field of education and its use cases in various educational activities. The current machine learning algorithms that have been implemented in different areas in the educational field were further discussed. Then finally the importance and future of Artificial intelligence in education were discussed.

The field of Artificial Intelligence (AI) is growing exponentially and its impact on all livelihood cannot be overemphasized (Sandu & Gide, 2019). The use of artificial intelligence has gradually elevated from fiction movies and now a reality (Tahiru, 2021). The continuous investments and innovations of

DOI: 10.4018/978-1-7998-6792-0.ch010

AI systems among Technological Giants such as Google, IBM, Microsoft, etc. have contributed to its popularity (Li et al., 2020; Tahiru, 2021). According to the Artificial Intelligence Market in the US 2020 report, the increasing use of the internet and connected devices are part of the major factors impacting the market growth of AI (Fortune Business Insight, 2019). Currently, AI use cases are spread across different industries and services (Li et al., 2020). For example, Microsoft's Cortana, Amazon's Alexa, Google Assistant, Apple Siri, and IBM Watson have become part of our everyday lives (Kandlhofer et al., 2016). The use of AI also features in e-commerce websites to track customers buying trends and suggest or recommend other items that customers may be interested in based on their historical data. Data is considered the new oil in the 21st century, as AI and machine learning(ML) models are in the ascendancy in every sector (Sandu and Gide, 2019). Artificial intelligence is the science of designing machines that can behave like humans. Machine learning is the process of extracting useful patterns from data and applying an algorithm to machines that automatically learns and improves from experience (Hoffer, 2005).

The main building force of all AI systems uses machine learning (ML) to create models which in turn depends on data for a successful outcome (Ciolacu et al., 2019). Machine learning whether supervised, unsupervised, and reinforced learning uses an algorithm to create models that learn by themselves (Ciolacu et al., 2019). When we come to the educational environment the enormous data that are captured in online course registrations, Massive Open Online Courses (MOOC), and the use of Learning Management Systems (LMS), has given rise to the use of machine learning in education as well (Conati et al., 2018). Machine learning is being used by some institutions for predicting students' performance, recommending courses, counseling students among others (Li et al., 2020).

Over the years research on the use of AI and ML in education has improved appreciably and studies have also indicated success in the use of AI in education. Sandu et al (2019) researched the adoption of AI chatbots to enhance the student learning experience in higher education in India. And the findings indicated that 48.9% of students prefer using the chatbot for assistance on educational related issues than chatting with a human. Likewise, to foster collaborative learning Prof Ashok of Georgia Tech University used IBM's Watson as a teaching assistant for an online distance learning program (Benedetto et al., 2019). AI has been used to solve the problem of students who lose concentration during online classes using face recognition algorithms. This encouraged the students to be conscious during online lessons (Conati et al., 2018).

This chapter seeks to address the current situation of academic instruction in Education in the light of AI transformation. Also, the chapter discussed the importance and future of education in the era of AI transformation. This knowledge will equip teachers and learners in the educational field about the emerging trends in AI in education and its importance in order to be ready for the expected massive AI transformation in the educational sector.

The rest of the sections in the chapter is presented as follows section 2 discussed the current trend of AI and ML techniques that are used in education. Section 3 presents the importance of AI in education and then section 4 discussed the future of AI in education.

CURRENT TREND OF AI TECHNIQUES USED IN EDUCATIONAL FIELD

This section discussed the current trend of AI and ML techniques that are used in education.

There is much previous literature about the use of Artificial intelligence in education however this section focused on the current trends of Machine learning that are used for designing AI-assisted systems in Education.

AI-powered systems have been widely used in various areas of education to infer learners' grades using Bayesian Networks and k-nearest neighbor (K-NN) Instance-Based algorithms (Sciarrone, 2018). The Bayesian network algorithm is one of the widely used inductive algorithms which learns based on how well the hypothesis can explain the observed training data. It is a technique where a graph model is used to compute uncertainties by using the concept of probability. (García et al., 2007)Whereas, the k-nearest algorithm makes choices based on the closest similar examples previously stored. To classify a data instance, first identify its k nearest neighbor, count the number of instances of the subset of k, the class with the highest count will be the example to be classified (Romero et al., 2013).

Similarly, in predicting student grade and retention rate (determining the rate of dropouts and those that stay till completion), the Bayesian network together with the decision tree algorithms have been used to predict students' grade and their retention rates based on their extracted features. A decision tree is a hierarchical structure consisting of nodes. Nodes can have two or more child nodes and are referred to as internal nodes whereas nodes without child nodes are known as leaves. The nodes split determines the value of an expression of an attribute. Decision trees make predictions by following the path of satisfying conditions from the root to the leaf of the tree based on the correspondent class labels. (Romero et al., 2013). The algorithms determined whether a student will drop out of school or not. The Bayesian algorithm had a high accuracy rate of 98%, which suggests that the Bayesian network outperformed the decision tree algorithm. The WEKA open source software was used to implement the algorithm (Khasanah and Harwati, 2017). In the study (Pal, 2017) a decision tree algorithm was used to assess the performance of students in higher institutions. The decision tree based on an if-then rule to classify students resulted in the authors' ability to predict students end of semester performance and also able to identify students who needed special attention(Pal, 2017).

The Neural Networks algorithm has been used to detect student distraction in online live lessons. The neural network is a computational model that performs classification similar to the human brain operation. Its basic unit is the neuron and works as the switch with input and output connections to other neurons through synapses (path through which neurons transmit signals) (Romero et al., 2013). Every unit receives an output from all units (the previous layer) and produces its output which is then fed in the nodes in the next layer (Romero et al., 2013). Learners' emotions were determined during live lessons to achieve better-personalized education. The procedure includes pictures taken every five seconds and cached on a server. After which, human body segmentation and face detection are run and priori features designed manually to generate a priori feature image which serves as an input to the AlexNet to determine distraction results. AlexNet is an image classification system based on neural network architecture. With its five convolutional layers, it's able to learn the semantics of an image and decide the label of the image with three connected layers using the features. Convolution is the application of a filter to a set of inputs to activate it (Brownlee Jason, 2020). The system achieved good precision in detecting destruction in the online live lesson(Zhou, 2020). Though the algorithm has good predictive power, the smaller size of the dataset may not produce better precision, and therefore the need for further study to improve on the size of the dataset to achieve better results.

Again, the Neural networks (NN) together with Support Vector Machine (SVM) have been used to reduce the failure rate in students' examinations. The SVM is a supervised ML algorithm that classifies numerical data(data with numerical features x and y) by drawing a single line that maximizes the margin

between positive and negative points (Hoffer, 2005). It works by plotting data in an n-dimensional space where n is the features. The value of each feature corresponds to a coordinate. Classification is performed by finding the hyperplane of the two classes (Ciolacu et al., 2019). The two types of experimental studies that were conducted include predicting the score for a new student in a semester and conducting a live recognition of students at risk with the focus of determining failure or success for each given student. The procedure involves the extraction of the number of clicks by the student from log files and grouped it in months. All the responses from the quiz in an Intelligent learning environment (ILE) were also aggregated in months. The ILE is a learning model constituting an entry knowledge questionnaire and personal data. The knowledge and performance get updated dynamically anytime a student interacts with the model. The neural networks learn from the data that were clicked, as well as the score of the online quiz and the grades for classification. The neuralnet library based on the R package was used to train the NN models. The hyper-parameter "threshold" was fixed at 0.01 and the learning rate at 0.01 or 0.001. Polynomial kernel and Radial basis function were considered for adding more features to the data to make it linearly separable. Linearly separable means inputs are split into two spaces where the first space contains half points and the second space contains the other half points. The neuron in the hidden layer spanned from 1 to 9. Manual training was performed and the internal weights of the trained neural network were stored on the Moodle server to generate the forecasts. True positive rate(sensitivity), true negative (specificity), and general performance were the three performance measures employed. It was indicated that NN showed better results on non-linear decisions better than SVM (Ciolacu et al., 2019). Though the study made use of the NN-model and SVM it failed to measure the accuracy rate of the algorithms and how well the predictions reflected the real scenario.

Another area in education where AI and ML learning techniques have been used is in language learning. AI identification technologies such as sound recognition, graph identification, and facial recognition have been applied in the field of language learning. The use of emotion recognition assists teachers to understand the emotions of students to interact and provide solutions to them. Also, facial recognition serves as feedback systems as well as correcting composition and translation. Voice translation assisted students in oral speech (Zhu, 2020).

The use of intelligent tutoring systems (ITS) has been in existence for the past five (5) decades (Kokku et al., 2018) and has proven to have a significant impact on instructional delivery to students. It offers guided learning support to students by engaging them in problem-solving (Conati et al., 2018). Currently, AI is leveraging the education sector beyond just instruction delivery but encouraging and creating a personalized and adaptive learning environment. This is evidenced in most scholarly works on ML algorithms. The Neural network with its good predictive power is used to infer students' grades and also ensuring students remain conscious during online lessons. Bayesian network and decision trees are among the two most popular classification algorithms that are used for predicting students' performance and retention rate however, the Bayesian algorithm has proven to have a high prediction rate. From the literature, it can be inferred that Neural Networks records high AI implementation in education and this is as a result of its high predictive power and its ability to eliminate irrelevant attributes in a dataset. The Bayesian algorithm also measures higher performance when combined with other ML algorithms as suggested by (Khasanah and Harwati, 2017; Romero et al., 2013). The Bayesian algorithm however works well with large data set to make reliable estimations.

The next section discusses the importance of AI in education and some benefits that have been achieved since its deployment in the field of education.

IMPORTANCE OF AI IN EDUCATION

Looking at the current state of education where the teacher to student ratio is low, the use of AI in education has become a necessity to augment teacher shortage and inequality (Luckin, 2020), and also to improve teacher-student interaction (Swain, 2020). AI is transforming how we teach students and also how students learn (Plitnichenko, 2020). With the use of AI in education we can address some of the educational challenges we are facing globally (Luckin, 2020). According to Luckin (Luckin, 2020), AI can assist in teaching STEM subjects and other languages to students of all ages (Luckin, 2020). Furthermore, AI can provide some form of feedback to students, which is something that is lacking in the education sector in globally (Luckin, 2020). The adoption of AI in education can also improve the monitoring of student progress in real-time (Swain, 2020). Currently, teachers can only monitor student progress only after the compilation of end-of-semester or term reports. But with AI the progress of students can be monitored in real-time (Swain, 2020). Once we are able to monitor student progress in real-time, it will also help in introducing personalized learning for such students who need assistance (Swain, 2020). In our current educational setup, it is difficult to adopt personalized learning due to the class size, but with AI we can customized learning materials to suit the specific needs of a student (*The Role of Artificial Intelligence in the Future of Education | GetSmarter Blog*, n.d.).

Personalized learning is achieved through the concept of "hyper-personalization" using a machine learning (ML) algorithm (Rangaiah, 2020). Hyper-Personalization is the process of using AI technology and real-time data based on user profiles to create content that is more customized and relevant to the end-user (Lebo, 2019; Maayan, 2020). With personalized learning, content is tailor-made to suit the student's learning needs, ability, and experience (Rangaiah, 2020). There is also a Voice Assistance AI application that can assist a student in learning educational materials without the help of a teacher (Rangaiah, 2020). Some of the most popular Voice Assistance AI applications include Alexa developed by Amazon, Apple Siri, and Microsoft Cortana (Rangaiah, 2020). Voice assistance can also be adopted and deploy in an educational environment to answer common questions and inquiries within an educational setup in other to ease the burden on support staff (Rangaiah, 2020).

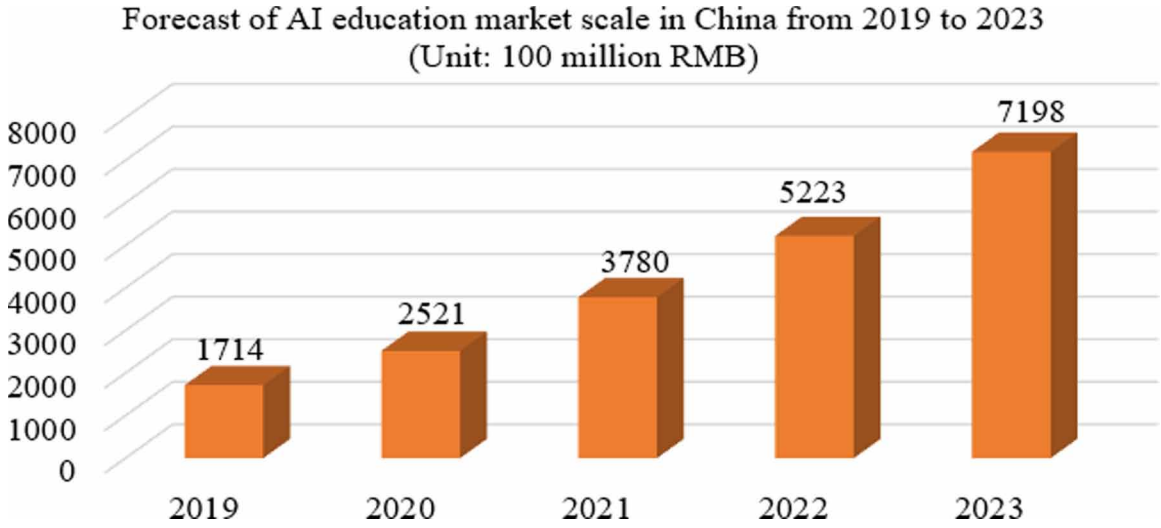
Another importance of AI in education is that AI as a discipline is something that needs to be introduced in our education. The current global economy is moving along the lines of AI technologies. Hence it needs to be taught in schools for a better understanding of the principles of the technology and what it can do (Luckin, 2020). Currently, there are demands for AI researchers and developers in the various application domain, but there is a lack of qualified researchers to fill such positions.

THE FUTURE OF AI IN EDUCATION

This section addresses the future of AI in education in the era of Industry 4.0 transformation. And the needed preparation for its adoption. Industry 4.0 denotes the 4th industrial revolution, where innovations in AI and robotics are rapidly taken over most industries. About 3000 companies in the US and over 1000 companies in China continuously invest in AI-assisted education(Zhu, 2020). According to the AI education market scale forecast of China, AI uses in education will attain a higher market scale in 2023 as indicated in Figure 1 (Li et al., 2020).

As the market scale of AI increases proportionally with huge investment by the Technology Giants, collaboration with stakeholders (Government, Teachers, and learners) becomes a necessity. The fear of AI

Figure 1. Forecast of AI education market scale in China from 2019 to 2023
Source (Li et al., 2020)



displacing teachers continues to linger with educators and tutors. In as much as AI technology provides a greater advantage in education in the Industry 4.0 era, its adoption remains a challenge (Ciolacu et al., 2019). However, it has been proven that AI use in education has created an enabling environment for both learners and educators to have a level playing field (Grosz et al., 2016). Thus, AI and ML algorithms have been used to create students' instructional programs that assist learners in achieving personalized education. Teachers have also been relieved of their extra duties with the use of AI which infers students' grades and suggests materials to assist students in the different study areas.

According to the 2020 report of the World Economic Forum (WEF), there is going to be an increased demand for a workforce with non-cognitive soft skills and STEM education (science, technology, engineering, and mathematics) by 2022, this is as a result of companies adopting AI and ML technologies into their operation(World Economic Forum, 2020). This suggests that Government and Educational leaders must develop a curriculum that can train students to master the relevant skillset needed for industries in the light of AI transformation.

According to the American Artificial intelligence market, AI use in education is anticipated to rise by 47.5% from 2017 to 2021. The rapid pace of AI transformation in education may change the way teaching and learning are conducted. However, research continues to conclude that AI can teach and perform other relevant activities in education yet the emotional and social impact can only be delivered by teachers. The ubiquitous nature of AI makes its use in education phenomenal, and distance is no more considered a barrier to education.

CONCLUSION

The focus of AI is universalizing all nations hence the focus of AI impacts on economic and Social growth has extended from Countries such as Europe (EU), United States (US), and United Kingdom (UK) to cover most low and middle-income countries such as China and India. The Global workforce

is undergoing digital transformation in this industry 4.0 era. The recent WEF report indicates that most work workforce would need digital skills to keep them in employment by 2025. The education sector has not been an exception in this transformation. AI has disrupted the teaching and learning from pedagogy design through to the mode of delivery. Interactive learning is achieved by the use of AI systems. Educators are assisted by AI systems with extra duties such as grading and assessment of students. AI is a reality now and the question remains how are we ready to embrace this fundamental transformation in Education. A reform of the educational curriculum to include AI training to produce AI required skillset for industries and academia. Innovation has driven economics to new ways of productivity and this has created a massive change in the skills required in current education. Teaching and learning must change from giving instructions and memorizing to a more interactive and problem-solving skills development to ensure student preparedness for the Technological world. These initiatives, however, require support for Government and Institution leaders in providing the requisite infrastructure to facilitate implementation and curricula development in various institutions.

REFERENCES

Benedetto, L., Cremonesi, P., & Parenti, M. (2019). A virtual teaching assistant for personalized learning. *CEUR Workshop Proceedings*, *2482*(June).

Brownlee. (2020). *Deep Learning for Computer Vision*. Academic Press.

Ciolacu, M., Tehrani, A. F., Binder, L., & Svasta, P. M. (2019). Education 4.0 - Artificial Intelligence Assisted Higher Education: Early recognition System with Machine Learning to support Students' Success. *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging, SIITME 2018 - Proceedings,* 23–30. 10.1109/SIITME.2018.8599203

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in education needs interpretable machine learning: Lessons from open learner modelling. *ArXiv, Whi*.

Fortune Business Insight. (2019). *Artificial Intelligence Market*. https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, *49*(3), 794–808. doi:10.1016/j.compedu.2005.11.017

Hoffer, E. (2005). Machine Learning. *Seoul Studies*, *6*(1), 71–80.

Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., & Huber, P. (2016). Artificial intelligence and computer science in education: From Kindergarten to university. *Proceedings - Frontiers in Education Conference, FIE,* 1–9. 10.1109/FIE.2016.7757570

Khasanah, A. U., & Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. *IOP Conference Series. Materials Science and Engineering*, *215*(1), 012036. Advance online publication. doi:10.1088/1757-899X/215/1/012036

Kokku, R., Sundararajan, S., Dey, P., Sindhgatta, R., Nitta, S., & Sengupta, B. (2018). Augmenting Classrooms With AI for Personalized Education. *IEEE Global Engineering Education Conference, EDUCON*, 6976–6980.

Lebo, T. (2019). *Hyper-Personalization: What It Is and Why You Need It in Your 2019 Marketing*. Academic Press.

Li, S., Yan, M., Zhang, X., & Li, Z. (2020). *Analysis on the Application of AI Technology in Online Education Under the Public Epidemic Crisis*. Springer International Publishing. doi:10.1007/978-3-030-63885-6_34

Luckin, R. (2020). *Hey Siri – why is AI important for education?* Academic Press.

Maayan, G. (2020). *Hyper Personalization: Customizing Service With AI*. Academic Press.

Pal, A. K. (2017). Analysis and Mining of Educational Data for Predicting the Performance of Students Analysis and Mining of Educational Data for Predicting the Performance of Students. *International Journal of Electronics Communication and Computer Engineering*, *4*(June).

Plitnichenko, L. (2020). *5 Main Roles Of Artificial Intelligence In Education*. Academic Press.

Rangaiah, M. (2020). *4 Major Applications of Artificial Intelligence in Education Sector*. Academic Press.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. doi:10.1002/cae.20456

Sandu, N., & Gide, E. (2019). Adoption of AI-chatbots to enhance student learning experience in higher education in india. *2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019*, 1–5. 10.1109/ITHET46829.2019.8937382

Sciarrone, F. (2018). Machine learning and learning analytics: Integrating data with learning. *2018 17th International Conference on Information Technology Based Higher Education and Training, ITHET 2018*. 10.1109/ITHET.2018.8424780

Swain, A. (2020). *Advantages and Challenges of AI in Education for Teachers and Schools*. Academic Press.

Tahiru, F. (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology*, *23*(1), 1–20. doi:10.4018/JCIT.2021010101

The Role of Artificial Intelligence in the Future of Education. (n.d.). *GetSmarter Blog*.

Zhou, Z. (2020). An intelligent teaching assistant system using deep learning technologies. *ACM International Conference Proceeding Series*, 18–22. 10.1145/3409073.3409079

Zhu, A. (2020). *Application of AI Identification Technology in Foreign Language Education.* doi:10.1109/ICAIE50891.2020.00024

Chapter 11
An Assessment of the Quality of Data From the EPI for Knowledge Management in Healthcare in Ghana

Patrick Ohemeng Gyaase
https://orcid.org/0000-0002-4711-6380
Catholic University College of Ghana, Fiapre, Ghana

Joseph Tei Boye-Doe
Catholic University College of Ghana, Fiapre, Ghana

Christiana Okantey
University of Cape Coast, Ghana

ABSTRACT

Quality data from the Expanded Immunization Programme (EPI), which is pivotal in reducing infant mortalities globally, is critical for knowledge management on the EPI. This chapter assesses the quality of data from the EPI for the six childhood killer diseases from the EPI tally books, monthly reports, and the District Health Information Management System (DHIMS II) using the Data Quality Self-Assessment (DQS) tool of WHO. The study found high availability and completeness of data in the EPI tally books and the monthly EPI reports. The accuracy and currency of data on all antigens from EPI tally books compared to reported number issued were comparatively low. The composite quality index of the data from the EPI is thus low, an indication poor supervision of the EPI programme in the health facilities. There is therefore, the need for effective monitoring and data validation at the point of collection and entry to improve the data quality for knowledge management on the EPI programme.

DOI: 10.4018/978-1-7998-6792-0.ch011

INTRODUCTION

Vaccines have been successfully used in the history of healthcare to prevent communicable diseases through immunization. Data from such immunization programmes has been used as the basis for effective decision concerning such diseases. Knowledge management is thus, critical for successful eradication of communicable diseases (Meri, 2020). Technological ecosystems, which provides general framework to develop solutions in which quality data is the backbone, have emerged to facilitate knowledge management in healthcare delivery (Vázquez-Ingelmo, et al, 2019). The World Health Organization (WHO) estimation shows about three million children below the age five die each year from vaccine preventable diseases globally, with sub-Saharan Africa having the highest mortality rate in spite of the availability of vaccines (WHO, 2019). The Expanded Immunization Program (EPI) in Ghana, based on the World Health Organization guidelines, is organized along the hierarchical organizational framework of the Ghana Health Service (GHS) with the immunization services being rendered at the national, regional, district, sub-district and community levels using different strategies to reach every child (Yawson, et al., 2017). Ghana is believed to have achieve substantial coverage in the immunization programme thereby reducing morbidity and mortality among children under the age of five. For instance, since 2003, there has been no death caused by measles, while in 2011, Ghana was certified as having attained elimination status for maternal and neonatal tetanus (Unicef, 2020). These successes are partly due to the knowledge acquired from the implementation of the EPI over the years. Ghana's EPI has been acclaimed by the WHO to the extent that Ghana was the first country of sub-Saharan Africa to receive COVID-19 vaccines. Knowledge management is thus critical in healthcare delivery.

The foundation of knowledge management in healthcare is quality data. Poor quality data resulting from the use of outdated source documents, lack of reporting guidelines and errors when transposing data unto reporting forms affect the effective knowledge management in the healthcare sector since data are unreliable for effective decision-making (Brown, et al., 2020). This chapter therefore, assesses the availability, completeness and accuracy of the routine EPI data in healthcare facilities in Ghana for effective decision-making and knowledge management on the six childhood killer diseases, which are covered by the Expanded Programme on Immunization (EPI). The EPI data that was used in this paper covers immunization exercises for Bacillus Calmette Guerin (BCG), Oral Polio Virus 0 (OPV 0), Pentavalent Vaccine Dose 3 (Penta 3), Pneumococcal Conjugate vaccine dose 3 (PCV 3), Measles Rubella dose 1 (MR1) and Yellow Fever (YF).

PROBLEM STATEMENT

Data quality plays a fundamental role in the success of an immunization with the Global Vaccine Action Plan (GVAP) identifying poor data quality as a major hindrance to reaching the Decade of Vaccines mission towards strengthening immunization systems programme (Wetherill, Lee, & Dietz, 2017).

Countries receiving support from Global Alliance for Vaccines and Immunization (GAVI) are required to develop an Immunisation Data Improvement Plan. This periodic plan should include a root cause analysis of unavailability and poor quality of data so as to define actionable recommendations to increase data availability, data quality, and usage (GAVI, 2019).

Although Ghana has been praised for its EPI programme, the quality of data from the programme to support knowledge management has not been well studied. There are inconsistencies in the variables

used and results obtained from various studies although institutions such as Ghana Health Service and International organizations such as UNICEF and WHO have depended on these data for knowledge management on the EPI and the eradication of the six childhood killer diseases (Unicef, 2020). Assessing the quality of the data from the three sources of the data from the EPI programme would inform the development of various strategies as well as information and communication framework to support knowledge management in the healthcare delivery systems in Ghana. It is therefore, imperative to assess the quality of routine data collected at healthcare facilities using EPI tally books, EPI Monthly reports and data from the District Information System for Health Management (DHMIS 2) to establish the availability, completeness, accuracy and overall quality of EPI data using Data Quality Self-Assessment (DQS) tool from the World Health Organization (WHO, 2006). Ethical approval for the study was obtained from the Ghana Health Service (GHS) Ethics Review Committee (ERC) with approval number (GHS-ERC: 040/08/20).

THE CONCEPT OF KNOWLEDGE MANAGEMENT IN HEALTHCARE

Knowledge refers to the information that enables action and decisions or information with direction. The fundamental problem of knowledge management (KM) is to explicate tacit knowledge and make it available to others for use. It is possible to convert explicit knowledge into tacit and vice versa. An essential characteristic of knowledge is its explicitness to enable storing and transferring to others (Gyaase, Anane, & Armah, 2015). Healthcare organizations are adopting digital systems to enhance patient care, improve efficiency and productivity, and minimizing the risk of errors.

Solving problems and making optimal decisions in healthcare is heavily dependent on access to knowledge since healthcare is a knowledge-driven industry. The increasing complexities associated with the healthcare provisioning makes it imperative for healthcare organizations to effectively manage both internal knowledge and externally generated knowledge (Haughom, 2014). A well-organized and effective strategy for knowledge management in healthcare is therefore, indispensable to any healthcare delivery system.

Healthcare systems are complex involving diverse stakeholders working in tandem to provide the necessary healthcare. The complexity of the healthcare sector provides the needed reason for the adoption of an effective knowledge management system to facilitate the collection, processing, storage and transmission for quality data. Presents (Karamitri, Talias, & Bellali, 2017).

BENEFITS OF KNOWLEDGE MANAGEMENT SYSTEMS FOR HEALTHCARE

Knowledge management (KM) systems are increasingly becoming crucial to healthcare organizations since they have been successfully employed by other sectors to improve efficiency and organizational learning. These knowledge-based systems are gradually becoming an integral part of the IT infrastructure for healthcare organizations (Nor'ashikin, et al, 2017). This is because knowledge management systems have under-exploited benefits for healthcare delivery. Knowledge management systems are credited with creating a more efficient information flow between the players of the healthcare ecosystem thereby facilitating effective decision-making processes. Information overload and lack of the needed knowledge for decision-making makes clinicians and healthcare manager handicapped in decision-making. Effective

knowledge management system can thus, facilitate information and knowledge flow resulting in better and more-informed decisions (Lee, Shin-Yuan, & Chau, 2011).

Knowledge Management allows medical professionals to access patients' electronic health records to facilitate quick and efficient diagnosis and reduce malpractices cases. An institutionalized knowledge management system will empower healthcare providers to partake in knowledge sharing among healthcare professionals to cut down on fatal misdiagnoses and ultimately streamline medical decision-making (Svensson & Hedman, 2021).

The utilization of knowledge management systems in healthcare organizations facilitates ongoing collaboration and innovation. By enabling a culture of knowledge sharing, healthcare professionals are more likely to engage in continuous learning and education. It builds learning organizations by making learning routine. This requires the creation a data-driven continuous-learning environment that supports organizational learning based on experience. Learning from experience builds knowledge that can then be used to improve care and streamline operations over time (Mohammed, Elrehail, Alatailat, & Elçi, 2019).

Finally, knowledge management stimulates cultural change and innovation in healthcare delivery. Actively managing knowledge drives free flow of ideas and an innovative culture, which begins with accepting the changing world of healthcare delivery and be open-minded about doing things differently. If top management of healthcare organization fosters the innovation mindset, it would ultimately permeate every level of an increasingly complex healthcare environment. An effective knowledge management system therefore, helps players in the healthcare ecosystem to embrace change and encourage ideas and insights leading to innovation (Alexander, 2020).

Quality Characteristics of EPI Data to Support Knowledge Management

The advances in healthcare technologies have increased the availability and access to healthcare data. However, availability and accessibility of data does not mean that all information is correct and reliable since poor quality data negatively impact the quality of information required for healthcare decision making and strategies (Alipour & Ahmadi, 2017). Data quality can be viewed from various perspectives namely, availability, timeliness, reliability, relevance and accuracy and completeness. For the purpose of this study, the main attributes for the EPI data quality to be examined are data availability, accuracy and completeness.

Data Availability

Data availability refers to the ability to ensure that required data is always accessible when and where needed in a healthcare organization since quality data is a prerequisite for patient safety and successful outcomes (Ahanhanzo, et al., 2015). Data availability impact directly on data completeness. Hence, the availability and use of immunization data are generally considered as the bedrock of successful immunization programme (Scott et al., 2017). Data availability is measured by checking for the availability of registers and reporting forms used in health facilities at the time of the assessment.

Data Completeness

Data completeness measures the degree to which all major steps in data processing, data entry, data cleaning, and data analysis have been executed thoroughly, ensuring that no data is missing, no details or

response are incomplete and uncollected. According to WHO, (2012), data completeness measures the proportion of the total number of reports submitted, as against the total number of reports that a given reporting unit is expected to submit at a given point in time or assessing all the number of all mandatory fields on a reporting form filled. Data completeness is evaluated by identifying the gaps in the data entered into the Health Management Information System (HMIS) in a health facility.

Data Accuracy

Data accuracy is defined as the closeness of agreement between a data value and the true value, in other words, how closely the data correctly captures what it was designed to capture (Audit Commision, 2012). One major defect to the accuracy and reliability of healthcare data is missing data, especially at the health center level resulting from under reporting or over-reporting of data, which can skew decision making processes (Sychareun, et al., 2014).

Data Management of the Extended Programme on Immunization in Ghana

In Ghana, all data on immunization are recorded into the District Health Information System (DHIMS II) which is the official online database for reporting. DHIMS II was adopted by the Ministry of Health (MOH) and Ghana Health Service (GHS) in 2011 to enable data collection, recording, compilation and storage of health data (GHS, 2018). These data from health-facilities serve as a primary source for measuring health sector performance through reported achievement and trends on key health indicators. However, data from the District Health Information Management System (DHIMS) usually reveals data quality problems which makes it difficult for its users to trust these data (Nisingizwe, et al., 2014).

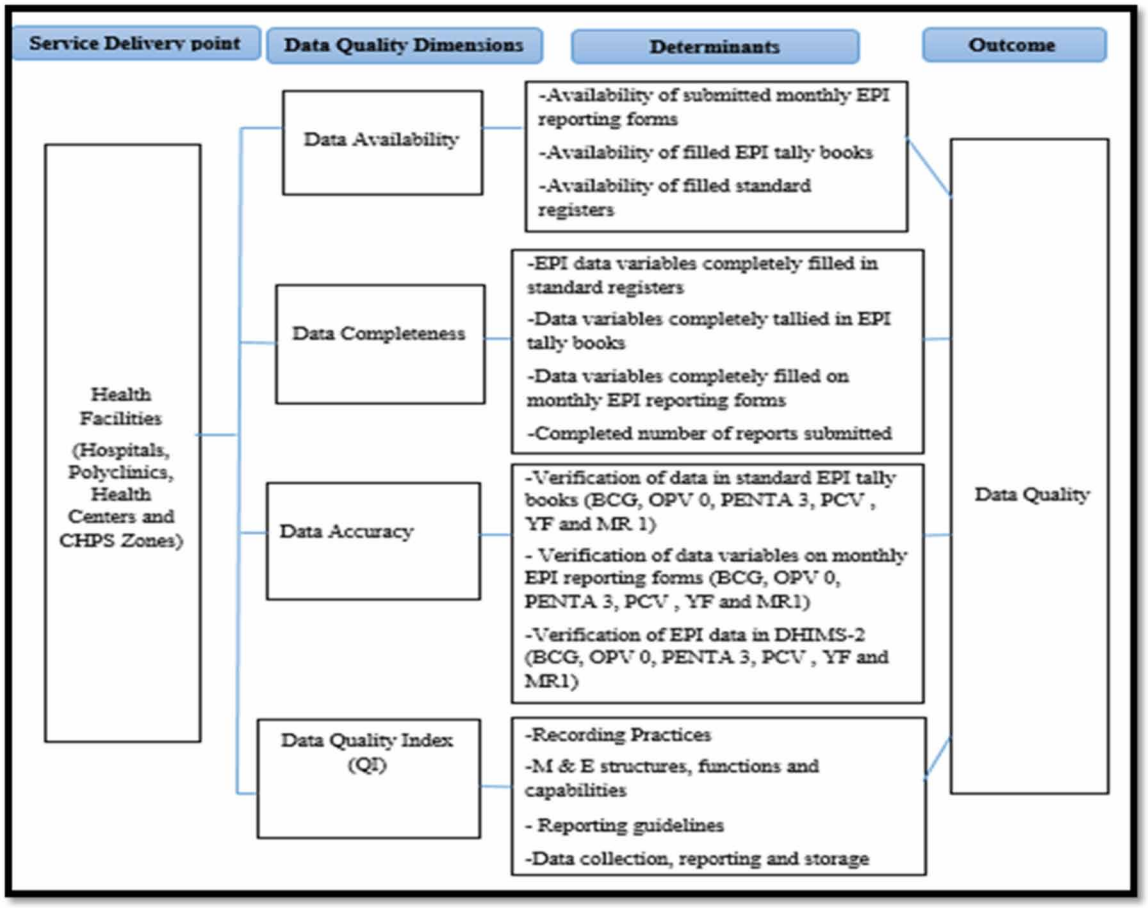
The staff of healthcare facilities collect routine primary data into registers and tally books from which the summary reports are compiled. These aggregated and primary data are electronically captured into DHIMS II. However, due to resource constraints, most health facilities lack the capacity to enter their primary data into the DHIMS II hence send reports to the District level for data entry (GHS, 2018). It is therefore imperative to ensure that EPI data collected, collated and entered into the DHIMS II is examined periodically to rectify errors identified and enhance the quality of data in the system for accurate and reliable decision-making and knowledge management.

Conceptual Framework for the Assessment of the EPI Data Quality

Figure 1 below presents the conceptual framework for the assessment of routine EPI data quality. The framework shows the service delivery points where routine EPI service is rendered and data is collected, collated and analyzed. It also presents the data quality dimensions being measured and the determinants of each of these dimensions that affect EPI data quality. Data availability, completeness, accuracy and the quality index of EPI data are used to determine the quality of EPI data generated at the service delivery points.

The quality of EPI data is dependent on the availability of submitted monthly EPI reports, filled EPI tally books and the standard registers which are used to capture data at the service delivery points. The variables of interest under availability are the number of source documents such as standard registers and EPI tally books available for the assessment period and the availability of EPI reports that was aggregated and submitted for the assessment period.

Figure 1. Conceptual framework for the assessment of routine EPI data quality



Data accuracy is an important ingredient for data quality. Data gathered from EPI tally books is expected to be the same as aggregated data compiled on monthly reporting forms and the data entered into DHIMS II. The accuracy of data generated in the health facility is crosschecked from EPI tally books and compared to the reported variables and the entries made into DHIMS II. The variables of interest for data accuracy are the number of children vaccinated from 0-11months for BCG, OPV 0, PENTA 3, PCV 3, YF and MR1 (GHS, WHO, Gavi, 2018).

The quality of data is influenced by data completeness. This a measured by the absence of omissions and gaps in the three sources of data for the EPI.

The data quality index (QI) assesses the data recording practices, M & E structures, functions and capabilities, reporting guidelines and data collection, reporting and storage procedures in the health facility which affect data quality. Recording practices such as immediate documentation after service is rendered and the adoption of daily tallying is vital to reduce errors in data management at the service delivery point. The QI enables the identification of the moderating factors for quality of data generated at the service delivery point and the effective measure to improve the quality if necessary.

RESEARCH METHODOLOGY

A descriptive cross-sectional design was used for the study by looking at data collected from a defined population at a particular point in time (Levin, 2006). Secondary and primary data were collected from 30 healthcare facilities selected purposively due to their provision of immunization services under the EPI. These data were quantitively analyzed using, data visualization, descriptive and inferential statistical analysis. The key healthcare personnel directly involved in the immunization exercise, documentation and reporting in the selected health facilities including disease control officers, community health nurses, public health nurses, field technicians and health information officers were present during the collection of the data for clarifications where necessary.

The sample includes hospitals, polyclinics, health centers and CHPS Zones that report through DHIMS II. The use of purposive sampling technique ensured that all the facilities selected had reported on BCG, OPV 0, PENTA 3, PCV 3, YF and MR1 in DHIMS II for the year 2019.

Data Analysis Methods

The DQS used by WHO for routine EPI data quality audit was adopted and modified using Spreadsheet. This was used for the quantitative data entry, processing and analysis. The data collected on accuracy was entered into the Spreadsheet template to generate the accuracy ratio and the discrepancy rate. Data on availability and completeness were also assessed for percentage data availability and data completeness rate. The following methods were used to analyses data availability, completeness, accuracy and the overall quality;

1. Data availability was assessed by checking for the availability of reporting registers and reporting forms per corresponding months for January to December, 2019. This was calculated by assessing the total number of reports physically available and retrieved at the time of assessment (A), divided by the total number of expected reports for the same period (B). Data availability was also calculated for the availability of standard registers by assessing the total number of EPI tally books physically available and retrieved at the time of the assessment (A), divided by the total number of expected EPI tally books or source document for the same period (B). Hence data availability for source documents and reporting forms was calculated as (A ÷ B) X 100 (WHO, 2006).
2. The computation data completeness was done based on the formulae used by Amoakoh-Coleman and Kayode (2015). Data completeness assesses whether the fields in a report filled at the facility, sub-district and district levels satisfy the required fields filled expected to be completed in the reports. The formula used is; Completeness = (md ÷ di) X 100. Where md= total number of filled fields on a report; di =total number of expected required fields on a report. The required fields assessed from the EPI tally books were the name of facility, the dates of immunization, BCG, OPV 0, Penta 3, PCV 3, YF and MR1. The same required fields were assessed in the EPI monthly reporting in addition to the signature of head of the reporting facility (WHO, 2017).
3. Data discrepancy and accuracy ratio were used to assess the accuracy of EPI data using the immunization Data Quality Self-assessment (DQS) tool recommended by the World Health Organization (WHO) to evaluate the different aspects of the immunization monitoring system. The accuracy ratio referred to as data verification factor is the main quantitative measure of data accuracy. It is the ratio of the number of vaccinations verified from a source (tally book) at one level (numerator),

compared to the number of vaccinations reported by that level (facility summary report) to higher level (denominator). This ratio gives the proportion of reported numbers that could be verified expressed as a percent. If the accuracy ratio is less than (<) 100 percent (%) then the data is over reported. On the other hand, if the accuracy ratio is greater than (>) 100%, then it is under reported (WHO, 2006).

4. For the overall quality of data from the EPI, the Immunization Quality Index checklists used in the health facilities were assessed. Responses from the responsible officers and observations from the checklist were categorized into four components namely, recording practices; monitoring and evaluation structures; storage and reporting practices; data collection, Reporting Forms, Tools and Storage. A scale of 1-3 was used according to the merit of each sub-question. A score of zero was indicated for a negative response of "No". the quality index use was measured using the maximum possible score and acquired score for each sub-item. Using composite scores, the QI was calculated for each sub-component by using the formula "QI =acquired scores/ sum of maximum scores that could be obtained" (WHO, 2006).

Data Analysis and Presentation

The secondary data from the three sources of the data from the extended programme for immunization (EPI) in addition to the data collected from the responsible reporting officers of the health facilities were analyzed using the respective methods above.

Categories of Healthcare Institutions for the Study

Out of the 30 healthcare facilities sampled, 6 were district hospitals making 20%, 4 were polyclinics constituting 13%. 3 were health centers constituting 10%. 8 were classified as clinics making up 27% whilst 9 are CHIPS compound which is 30%. The chips compounds are mostly located in the rural areas and are at the forefront of the annual scheduled immunization month in Ghana.

Data Availability

Figure 3 below presents a comparative descriptive analysis of the data availability of EPI tally books and the availability of EPI monthly summarized reports from the various categories of health facilities. The overall percentage availability of EPI tally books was 88.4% for EPI tally books and 98.6% for Monthly reports. Majority of the facilities assessed had all EPI tally books for the period under consideration. CHPS zones had the lowest data availability rate for tally books averaging 65%. On the other hand, all categories health facilities assessed recorded almost100% availability of monthly EPI returns.

Data Completeness

Figure 4 below depicts none of the health facilities assessed recorded 100% data completeness in terms of data elements in both tally books and EPI monthly summarized reports. The overall data completeness in EPI tally books was relatively low representing 73.75% (354/480), while on the other hand data completeness in EPI reports was 94.44% (510/540). In relation to data completeness in EPI tally books, Pentecost Hospital recorded the least representing 41.67% (40/96), while Rawlings Circle Polyclinic

Figure 2. Categories of Health Facilities

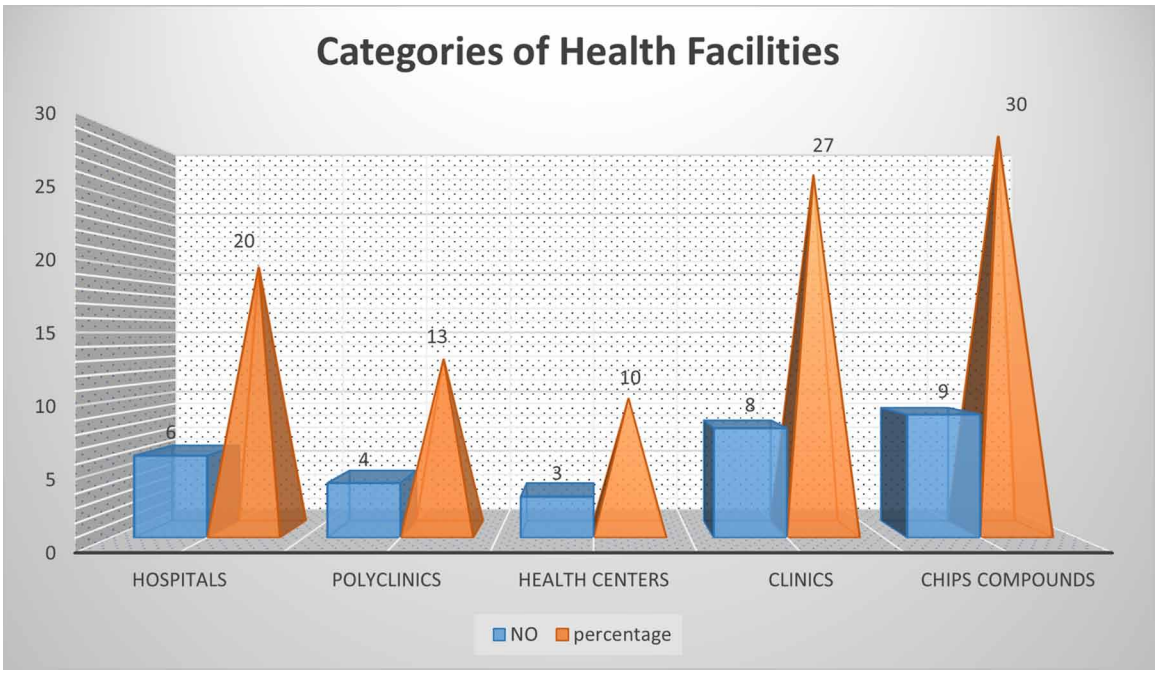


Figure 3. Percentage availability of EPI tally books and monthly EPI reports

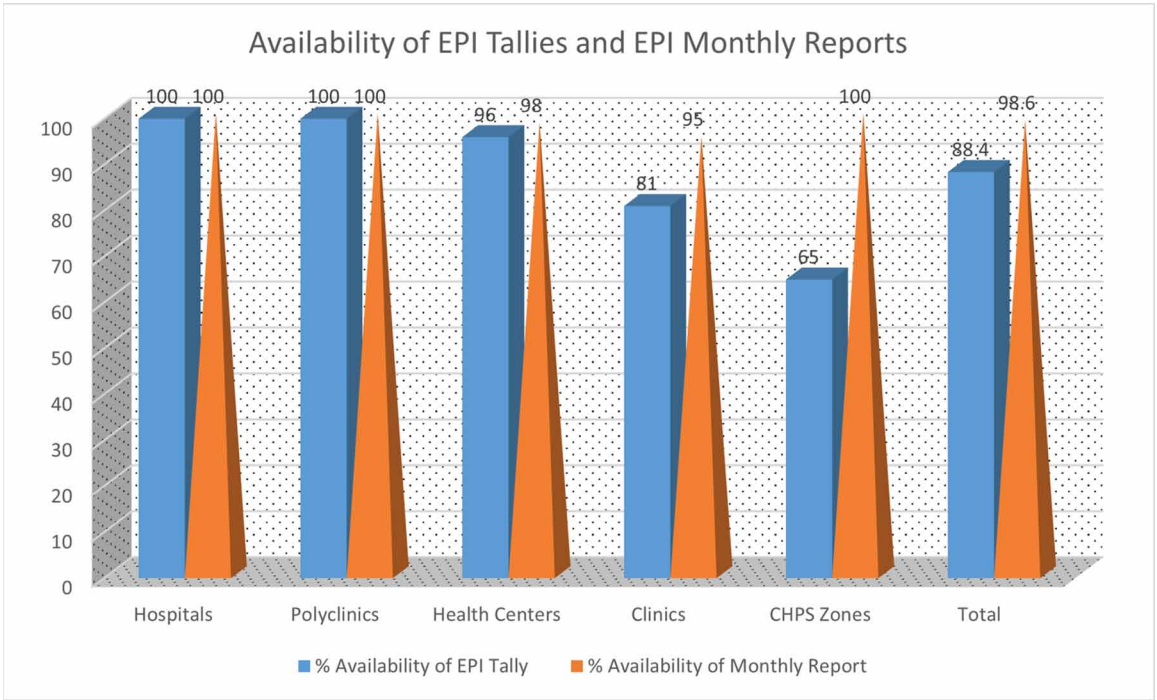
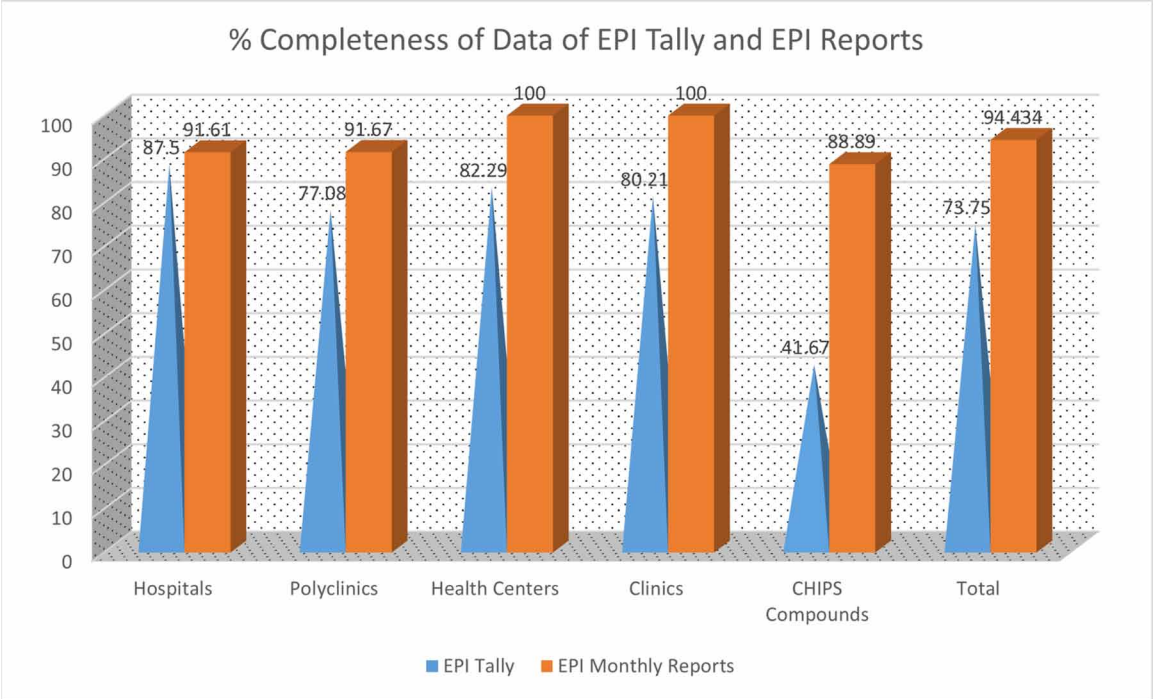


Figure 4. Percentage data completeness of EPI tally books and monthly EPI reports



recorded the highest data completeness in EPI tally books representing 87.50% (84/96). Two facilities recorded 100% for data completeness of EPI monthly reports, while the least was 88.89% (96/108).

Data Accuracy

Table 1 below presents the level of accuracy of data on all the BCG vaccine across the three data sources. A total of 27,216 children received the BCG vaccines from health facilities studied during the period under consideration according to their register. The monthly report forms revealed 40,962 BCG vaccinations during the period. However, DHIMS II captured 40,932 as the total number of children who received the same vaccine during the same period. The overall accuracy rate of the vaccines received was 66.44% comparing the register and the monthly reporting forms, indicating over reporting whilst comparing the aggregated data to the DHIMS II data, the accuracy rate of BCG vaccinations is 100.7% indicating under reporting.

Table 2 below presents the level of data accuracy for OPV0 vaccination across the three data sources. The register revealed that 27036 children received OPV0 vaccines for the period under consideration. The monthly report forms showed 40920 OPV0 vaccination for the same period whilst 40398 OPV0 vaccinations were captured by DHIMS II. The data accuracy rate of OPV0 vaccination is 66.1% comparing the register and the monthly reporting form, indicating under-reporting whereas comparing the aggregated data to DHIMS II data showed an accuracy rate of 101.29% indicating under-reporting.

Table 3 below presents the level of data accuracy for PENT 3 vaccinations across the three data sources. The total number of children who received PENTA 3 was 20,496 during the period under consideration

Table 1. Analysis of the reports on BCG vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100-E]**	**Result interpretation (Under/Over)**
January	2250	2892	642	77.80	22.20	Over	2892	0	100	0	Accurate
February	1080	3090	2010	34.95	65.05	Over	3102	12	99.61	0.38	Under
March	1338	3498	2160	38.25	61.75	Over	3672	174	95.26	4.73	Over
April	1734	4590	2856	37.78	62.22	Over	4590	0	100	0	Accurate
May	1758	3732	1974	47.11	52.89	Over	3732	0	100	0	Accurate
June	1356	3420	2064	39.65	60.35	Over	3420	0	100	0	Accurate
July	1896	3396	1500	55.83	44.17	Over	3180	-216	106.79	-6.79	Under
August	3114	3102	-12	100.39	-0.39	Under	3102	0	100	0	Accurate
September	3648	3768	120	96.82	3.18	Over	3768	0	100	0	Accurate
October	3528	3642	114	96.87	3.13	Over	3642	0	100	0	Accurate
November	2442	2760	318	88.48	11.52	Over	2760	0	100	0	Accurate
December	3072	3072	0	100.00	0.00	Accurate	3072	0	100	0	Accurate
Total	27216	40962	13746	66.44	33.56	Over	40932	-30	100.07	-0.07	Under

Note: A=facility counted data; B=Reported Data; C=Accuracy Ratio for facility counted data; D=Discrepancy rate for facility counted data; DH=DHIMS-2 reported data; E=Accuracy ratio for DHIMS2 data; F=Discrepancy rate for DHIMS2 data.

Table 2. Analysis of the reports on OPV0 vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100-E]**	**Result interpretation (Under/Over)**
January	2136	2808	672	76.1	23.93	Over	2808	0	100	0	Accurate
February	1050	3072	2022	34.2	65.82	Over	3072	0	100	0	Accurate
March	1320	3672	2352	35.9	64.05	Over	3672	0	100	0	Accurate
April	1794	4590	2796	39.1	60.92	Over	4590	0	100	0	Accurate
May	1968	3834	1866	51.3	48.67	Under	3834	0	100	0	Accurate
June	1266	3372	2106	37.5	62.46	Over	3372	0	100	0	Accurate
July	1836	3378	1542	54.4	45.65	Over	3162	-216	106.83	-6.83	Under
August	3084	3060	-24	100.8	-0.78	Under	3084	24	99.22	0.77	Over
September	3588	3708	120	96.8	3.24	Over	3378	330	109.76	-9.76	Under
October	3480	3630	150	95.9	4.13	Over	3630	0	100	0	Accurate
November	2442	2730	288	89.5	10.55	Over	2730	0	100	0	Accurate
December	3072	3066	-6	100.2	-0.20	Under	3066	0	100	0	Accurate
Total	27036	40920	13884	66.1	33.93	Over	40398	-530	101.29	1.29	Under

Table 3. The analysis of the reports PENTA 3 vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100-E]**	**Result interpretation (Under/Over)**
January	2232	2424	192	92.08	7.92	Over	2400	-24	101	-1	Under
February	1350	2172	822	62.15	37.85	Over	2178	6	100	0	Accurate
March	1176	1974	798	59.57	40.43	Over	1950	-24	101	-1	Under
April	1230	2388	1158	51.51	48.49	Over	2262	-126	106	-6	Under
May	1170	2052	882	57.02	42.98	Over	1956	-96	105	-5	Under
June	1332	1962	630	67.89	32.11	Over	1962	0	100	0	Accurate
July	1794	2118	324	84.70	15.30	Over	2454	336	86	14	Over
August	2358	2268	-90	103.97	-3.97	Under	2388	120	95	5	Over
September	2532	2334	-198	108.48	-8.48	Under	2358	24	99	1	Over
October	1824	2100	276	86.86	13.14	Over	2094	-6	100	0	Accurate
November	1638	1698	60	96.47	3.53	Over	1770	72	96	4	Over
December	1860	2034	174	91.45	8.55	Over	2034	0	100	0	Accurate
Total	**20496**	**25524**	**5028**	**80.30**	**19.70**	**Over**	**25806**	**282**	**99**	**1**	**Under**

according to the register. The monthly report forms however recorded 25,524 children received PENTA 3 vaccines whilst the records on DHIMS II showed 25,806 PENTA 3 vaccinations for the same period. This shows an accuracy rate of 80.30%, indicating over-reporting comparing the register and the monthly report form. The accuracy rate is however, 99% indicating over-reporting comparing the DHIMS II data and the aggregated data.

Table 4 below present the data accuracy rate for PCV 3 recipients across the three data sources for the period under consideration. 20,550 was recorded as the recipients of PCV 3 vaccinations according to the register. The monthly report forms however, showed 25,560 PCV 3 recipients whilst the records on DHIMS II showed 25,794 PCV3 recipients. The accuracy rate of PCV 3 was 80.40% indicating over-reporting when the register is compared with the monthly report forms. The accuracy rate is however, 99.1%, also showing over-reporting comparing the DHIMS II data to the aggregated data.

Table 5 below presents the data accuracy rate for yellow fever vaccination across the three data sources for the period under consideration. The records from the registers shows 18,036 recipients whereas DHIMS II records showed 21,312. The accuracy rate of 84.11%, indicating over-reporting comparing the register and the monthly report forms. The accuracy rate is 100.62%, indicating under-reporting comparing DHIMS II data with the aggregated data.

Table 6 shows the data accuracy on Measles Rubella (MR) vaccinations across the three data sources. According the data from the registers, a total of A total of 18,162 recipients of MR vaccinations, the monthly reports showed 21,444 recipients of MR for the same period whilst the data from DHIMS II showed 21,276 recipients of the MR vaccination for the period. The accuracy rate is 84.70% showing over-reporting the data from the registers is compared with the data from the monthly report forms. The accuracy rate was 100.79% showing under-reporting, comparing DHIMS II with the aggregated data.

Table 4. The analysis of the reports of PCV 3 vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100-E]**	**Result interpretation (Under/Over)**
January	2238	2424	186	92.33	7.67	Over	2400	-24	101.0	-1.0	Under
February	1350	2172	822	62.15	37.85	Over	2178	6	99.7	0.3	Over
March	1176	1974	798	59.57	40.43	Over	1950	-24	101.2	-1.2	Under
April	1212	2388	1176	50.75	49.25	Over	2262	-126	105.6	-5.6	Under
May	1164	2052	888	56.73	43.27	Over	1956	-96	104.9	-4.9	Under
June	1278	1962	684	65.14	34.86	Over	1962	0	100.0	0.0	Accurate
July	1794	2118	324	84.70	15.30	Over	2454	336	86.3	13.7	Over
August	2358	2256	-102	104.52	-4.52	Under	2376	120	94.9	5.1	Over
September	2532	2334	-198	108.48	-8.48	Under	2358	24	99.0	1.0	Over
October	1824	2100	276	86.86	13.14	Over	2094	-6	100.3	-0.3	Under
November	1602	1746	144	91.75	8.25	Over	1770	72	98.6	1.4	Over
December	2022	2034	12	99.41	0.59	Over	2034	0	100.0	0.0	Accurate
Total	**20550**	**25560**	**5010**	**80.40**	**19.60**	**Over**	**25794**	**282**	**99.1**	**0.9**	**Under**

Table 5. The analysis of the reports of yellow fever vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100-E]**	**Result interpretation (Under/Over)**
January	2100	2442	342	86.00	14.00	Over	2424	-18	100.74	-0.74	Under
February	1392	2034	642	68.44	31.56	Over	2034	0	100	0	Accurate
March	1662	2142	480	101.84	-1.84	Under	2094	-48	102.29	-2.29	Under
April	1146	1632	486	67.97	32.03	Over	1632	0	100	0	Accurate
May	1002	1686	684	79.15	20.85	Over	1632	-54	103.30	-3.30	Under
June	984	1266	282	49.10	50.90	Over	1266	0	100	0	Accurate
July	1716	2004	288	89.10	10.90	Over	1974	-30	101.51	-1.51	Under
August	1866	1926	60	96.88	3.12	Over	1902	-24	101.26	-1.26	Under
September	1398	1404	6	99.57	0.43	Over	1404	0	100	0	Accurate
October	1782	1908	126	93.40	6.60	Over	1914	6	99.68	0.31	Over
November	1644	1698	54	96.82	3.18	Under	1698	0	100	0	Accurate
December	1344	1302	-42	103.23	-3.23	Under	1338	36	97.30	2.69	Over
Total	18036	21444	3408	84.11	15.89	Over	21312	-132	100.61	0.61	Under

Table 6. The analysis of the report on MR vaccination

	Facility data accuracy and Discrepancy						DHIMS-2 Data Accuracy and Discrepancy				
Month	**[A]**	**[B]**	**Gap**	**C=[A/B]*100**	**D=[100-C]**	**Result Remarks**	**[DH]**	**Gap**	**E=[B/D]*100**	**F=[100- E]**	**Result interpretation (Under/Over)**
January	2100	2442	342	85.99	14.00	Over	2424	-18	100.74	-0.74	Under
February	1398	2034	636	68.73	31.26	Over	2034	0	100	0	Accurate
March	1662	2142	480	77.59	22.40	Over	2094	-48	102.29	-2.29	Under
April	1152	1632	480	70.59	29.41	Over	1632	0	100	0	Accurate
May	1020	1686	666	60.50	39.50	Over	1632	-54	103.30	-3.30	Under
June	984	1266	282	77.73	22.27	Over	1266	0	100	0	Accurate
July	1752	2004	252	87.43	12.57	Over	1974	-30	101.51	-1.51	Under
August	1866	1926	60	96.89	3.12	Over	1902	-24	101.26	-1.26	Under
September	1398	1404	6	99.57	0.43	Over	1404	0	100	0	Accurate
October	1836	1908	72	96.23	3.78	Over	1914	6	99.68	0.31	Over
November	1650	1698	48	97.17	2.83	Over	1698	0	100	0	Accurate
December	1344	1302	-42	103.23	-3.23	Under	1302	0	100	0	Accurate
Total	18162	21444	3282	84.70	15.31	Over	21276	-168	100.78	0.79	Under

Figure 5 below presents the comparative analysis of the data accuracy for the vaccines. The overall data accuracy of immunization data when data from the registers are compared with the monthly report was 74.8% which indicates over-reporting and between DHIMS II and the monthly report, the data accuracy rate of 100.2 was reported which indicates under reporting. None of the antigens reported an accuracy ratio of 100%. MR 1 recorded the highest accuracy ratio of 84.7% indicating over-reporting while OPV0 recorded the lowest accuracy ratio of 66.1% indicating over-reporting. All the antigens recorded more than 100% indicating under-reporting, expect for PENTA 3 and PCV 3 which recorded 98.9% and 99.1 indicating over-reporting, respectively when DHIMS 2 is compared with the monthly reports.

Figure 6 below presents the comparative analysis of data accuracy from the data sources according to the categories of facilities. Comparing the EPI Tally registers with the monthly report, CHPS Zones had the lowest accuracy ratio of 47.2% whilst hospitals recorded the highest accuracy rate of 99.9%. Comparing the DHIMS II with the monthly reports, hospitals recorded the highest accuracy rate of 100.6% whilst CHPS Zones recorded the lowest accuracy ratio of 98.8%.

Quality Index of EPI Data

Table 7 presents the EPI data quality assessment using the Quality Index Checklists. The analysis shows that majority of health personnel input data into the standard register immediately immunization service is provided with an average QI score of 60%. Also, a QI score of 60% was obtain for the number of health facilities that tally into EPI tally book immediately service is rendered. The highest QI score under recording practices was 80%. This result shows majority of health facilities assessed properly documenting all data variables into source documents.

The average QI score for monitoring and evaluation structure, functions and capabilities component was just 50%. Staff training on the data management processes and tools under this component scored only 40% indicating majority of the healthcare staff had not received the appropriate training although

Figure 5. Comparison of accuracy of data on antigens used

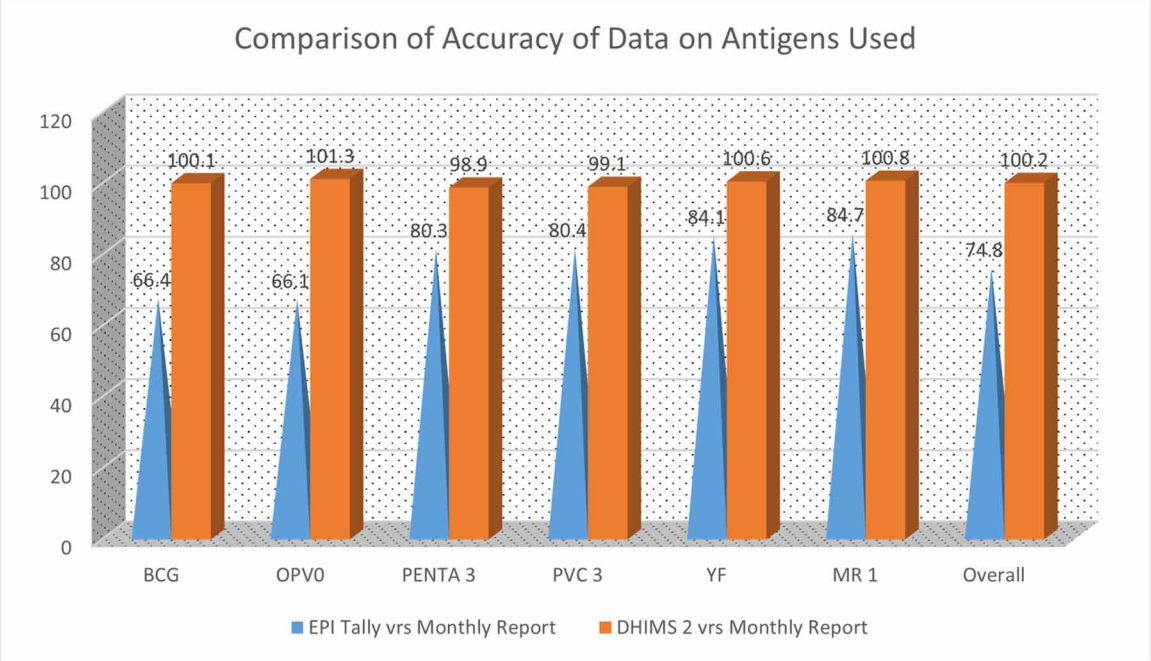
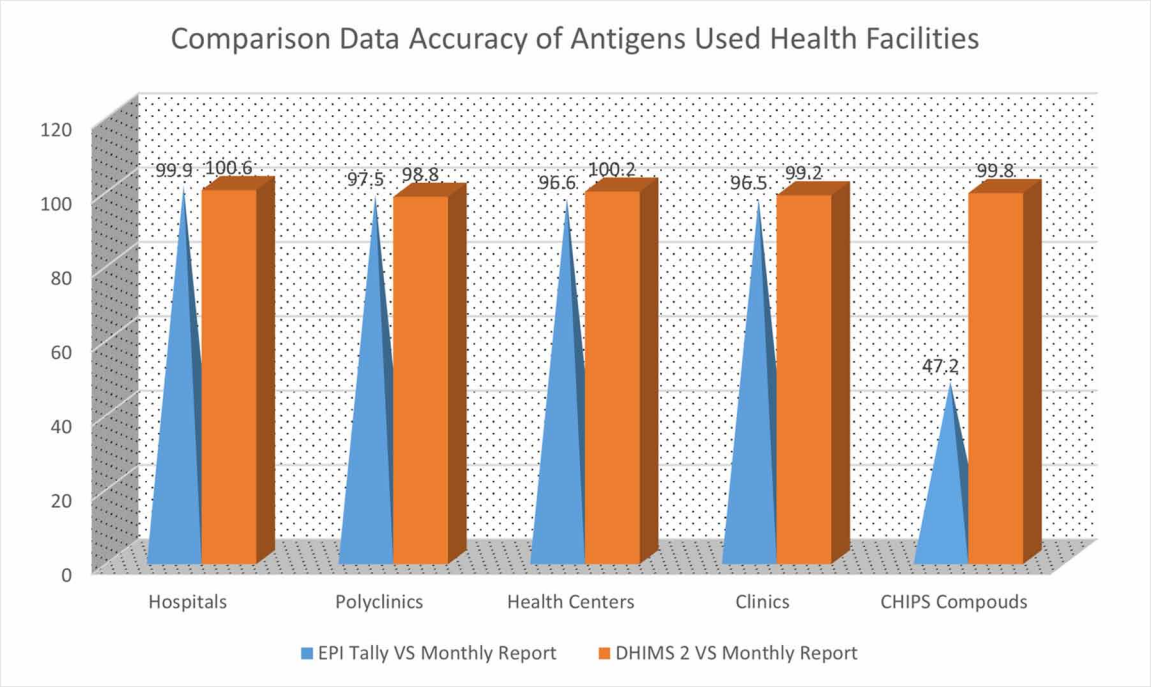


Figure 6. Comparison of data accuracy between facility categories



majority of health facilities (QI=60%) had appointed a staff responsible for reviewing periodic reports prior to submission to the next level. A greater number (QI=80%), of the facilities reported that, sometimes their health facilities receive regular supportive supervisory visits from district and/or regional level staff according to the guidelines. However, majority of these service provision sites indicated lack of regular feedback on the quality of data generated and reports submitted and entries made into DHIMS II according to the national guidelines. In relation to reporting guidelines, there was a very low score of existence of written guidelines on EPI indicator definitions and existence of GHS standard operating procedure in health information management. A low score of 20% was recorded for both components. In contrast all staff of the health facilities visited knew the deadlines for reporting to the next level.

A low QI score of 40% was recorded for the sufficiency of standard registers indicating majority of health facilities did not have sufficient stocks of standard registers at the health facility. However, a QI score of 100% was recorded for sufficiency of monthly reporting forms. A QI score of 60% for the facilities having a designated place for storing previously filled immunization reports and source documents. A higher proportion (QI=80%), of the facilities have filed copies of reports submitted to District and reports entered into DHIMS II. The overall quality index score was 58.6%. which is quiet low.

DISCUSSIONS

The explosion of information technology applications supporting knowledge management in healthcare, quality data would play a critical in healthcare knowledge management systems to improve decision making, continues leaning and healthcare innovations (Meri, 2020). In assessing the quality of EPI data for knowledge management, the study focused on three (3) main dimension of data quality which was data availability, completeness and accuracy as well as the quality index (QI).

Availability of Routine EPI Data

The overall availability of immunization data was 91.67% and 100% for EPI tally books and EPI monthly reports respectively. The study found 20% of facilities studied did not have standard EPI tally books used for five (5) months. This as results of the remoteness of their location. Majority of those facilities were CHPS compounds some of whom have no designated places to keep records leading to misplacement of the tally books and monthly records forms. The low availability of immunization data recorded by the CHIPs compound was also attributed to the fact that staff walk to surrounding villages for immunization and transporting this books day to day leads deterioration, misplacement and mishandling.

Completeness of Routine EPI data

The results showed an overall data completeness in EPI tally books stood at 73.75% and 94.44% completeness of data elements was recorded in EPI monthly reports. All facilities assessed had incomplete data filled in EPI tally books, while 60% of the facilities had incomplete data on the monthly EPI returns. One category of facility recorded 41.67% of data completeness in EPI tally books. The least data completeness of EPI monthly reports was 88.89%. The higher level of data incompleteness of immunization data in this study was due to missing EPI tally books mostly in CHPS compounds. The major omissions were dates of the reporting period in EPI tally books and also most reporting forms assessed did not have

Table 7. Quality Index (QI) score of EPI data

Selected EPI Quality variables	Maximum Score	Acquired Score	QI Score
Recording Practices			
Data recorded into register immediately service is rendered.	15	9	60.0
Tally is done in EPI tally book immediately service is rendered.	15	9	60.0
All data variables properly documented in register.	15	12	80.0
Sub-Total	**45**	**30**	**66.7**
M&E Structure, Functions and Capabilities			
Received appropriate training on the data management	15	6	40.0
Designated staff responsible for reviewing periodic reports	15	9	60.0
The health facility received regular supportive supervisory visits	15	12	80.0
The health facility receives regular feedback	15	3	20.0
Sub-Total	**60**	**30**	**50.0**
Reporting guidelines			
Existence of written guidelines on EPI indicator definitions	15	3	20.0
Existence of GHS standard operating procedure	15	3	20.0
Staff know reporting deadline	15	15	100.0
Sub-Total	**45**	**21**	**46.7**
Data collection, reporting forms, tools and Storage			
Sufficient stocks of standard registers	15	6	40.0
Sufficient stocks of reporting forms at the service delivery site	15	15	100.0
Designated place for storing historical data	15	9	60.0
Reports have been filed	15	12	80.0
Sub-Total	**60**	**42**	**70.0**
Total Score	**210**	**123**	**58.6**

the signature of the head of the facility on the forms. Furthermore, most facilities did not have the GHS standard operating procedure on data management which serves as a guide on how to properly fill the source documents. Data provides the source for information the facilitates decision making. Incomplete data thus, is an anathema for effective decision making and knowledge management. Incomplete tally books and monthly reports, which are the source documents for EPI data is therefore an unacceptable situation.

Accuracy of Routine EPI Data

Data quality is influenced by the rate of discrepancies resulting from inaccuracies from the sources of data. The higher the discrepancy rate, either positive or negative, the less quality the data. According to the DQS tool standard, a discrepancy of more than 10% render the data unreliable for decision making and planning of programme. The result from this study showed over-reporting when data from source document was compared to reported data. This was largely due to missing tally books in the some of

the facilities, transposition errors and the lack of data validation in health facilities. In contrast data was of high quality when monthly reported data is compared to DHIMS-2 values.

The overall accuracy rate of the BCG vaccines received was 66.4% comparing the register and the monthly reporting form whilst comparing the aggregated data to the DHIMS II data, the overall accuracy ratio of BCG vaccines received is 100.7%. The findings thus, puts the reliability of the data from the EPI tally books from the CHPS compounds into question.

The overall OPV0 data accuracy rate was 66.1% when data from the tally book is compared with a discrepancy of 33.93%. The accuracy of data on the same antigen when the aggregated data was compared to DHIMS II data was significantly better with an accuracy rate of 101%. The discrepancy level between data from tally books and monthly report can be ascribed to missing data in mostly one category of facilities and transcription errors.

The overall accuracy rate of PENTA 3 was 83.3% comparing the data from tally books to data on EPI monthly reports. A discrepancy rate of 19.70% is an indication of poor data quality in the health facilities. Data accuracy again was high when the aggregated data was compared to DHIMS II data with an accuracy rate of 99% and a discrepancy of 1%. This clearly signifies that data transposed from the monthly reports form into DHIMS II is of high quality.

PCV-3, which is administered at the same time as PENTA 3 had similar results. The overall accuracy rate of PCV 3 was 80.40%, the data from IPI tally book is compared with the aggregated monthly reports. The accuracy rate aggregated data entered into DHIMS II is very high at 99.1%. Some reasons given for the data discrepancy at the health facility level were inaccurate data recording, arithmetic errors, the use of improvised tally books and documentation on non-standardized pieces of papers.

The overall accuracy ratio of Yellow fever was 84.11%, comparing EPI tally books data to aggregated monthly reports, with a discrepancy ratio of 15.89. The accuracy ratio for aggregated monthly data against DHIMS II was 100.6%. Similarly, the overall accuracy ratio was 84.7% for MR 1 comparing data from tally books to monthly reported figures. The discrepancy of 15.3 is similar to results of other antigens accessed in this study. The concordance between monthly aggregated data and DHIMS II was at 0.78% with accuracy ratio of 100.79%. This could be attributed to the availability of all reporting forms and low transcription errors during data entry.

Quality Index (QI) of EPI Monitoring System

The overall QI score of the immunization monitoring system recorded for the facilities accessed was 58.6%. This was as a result of the different variations of the quality index components accessed in the health facilities. In relation to recording practices, the study indicated a QI of 60%, of the health facilities assessed document services rendered into standard register immediately service is rendered. This implies significant number of service providers do not document immediately service is rendered.

Also, a QI of 60 was recorded in relation to facilities that tally immediately service is provided. This is one of the contributory reasons for missing data is standard registers and tally books in health facilities. Also, a significant number of facilities assessed had a QI of 80% for availability of data entry sources, which implies most data variables were properly documented in registers. These findings clearly indicate that majority of the health facilities assessed either did not document immediately into standard registers or did not document properly. This can be attributed to high workload, lack of supervision and lack of feedback given to service providers on the quality of data they produce. This could lead to data discrepancies which affects the quality of EPI data generated and ultimately lead to wrong healthcare decisions.

In relation to QI on the data collection, reporting forms and data storage, the current study revealed that majority 24 (80%) of the facilities did not have sufficient stock of standard registers and designated place for storage and filling of reports and completed registers. The lack of registers in the health facilities can lead to the use of improvised registers which might lead to poor documentation practices. This can ultimately affect the quality of data generated and entered into DHIMS II. Also, lack of storage place for archiving threatens data storage for future reference and usage. This can be attributed to the loss of standard registers in some of the facilities visited which contributed to the poor correlation between source data, aggregated data and DHIMS II data. On the other hand, all 30 (100%) facilities accessed had sufficient stock of reporting forms.

Under monitoring and evaluation structure, functions and capabilities, this study indicated a QI of only 40%. Majority of the staff interviewed during the study have not undertaken training on any form of data management processes and reporting tools. This shows that staff are not given any orientation or in-service training on data management, hence might contribute to poor documentation in registers and reporting forms. Also, majority (QI=80%) of the health facilities although received supportive supervisory visit from the district level, the QI rate of 80% showed that they did not receive any feedback on the quality of EPI data submitted to the district or on the data entered into DHIMS II. The study clearly revealed the monitoring and evaluation component as the second worst, when compared to the other QI components assessed.

One key requirement of the data management is the availability of up-to-date written guidelines to serve as a reference material and a guide to service providers who document and report at the health facility. The study result showed QI of 20%, which implies majority of service providers did not have any copy of written guidelines on indicator definitions. Similarly, QI of 20% reflected majority of service providers did not have the GHS standard operating procedure for data management. The SOP entails relevant information and data management process, variable and indicator definitions, as well as reporting guidelines. This document is mostly kept at the district level with most health facilities not having access to it.

RECOMMENDATIONS

In conclusion, the show found significantly high data availability of source documents used to capture data as well as high completeness of data variables in standard registers and reporting forms. However, it revealed weak correlation between data from source documents (EPI tally books) and data on monthly EPI reports. Also, there were disparities between aggregated reports and DHIMS-2 data. These inaccuracies were progressively worse as the source document data was captured and comparted with aggregated data (EPI tally book and EPI monthly report).

The findings above signify poor data quality in the health facilities assessed. This was as a result of missing EPI tally books at the CHPS compounds. Another contributory factor was the lack of storage facilities to archive historical data which was a major issue in all the study sites. Findings from the study also highlighted the known fact from studies such as (Aaron, et al., 2017) and (Heaton, et al., 2017) that EPI data is characterized by under-reporting and over-reporting which tend to affect the quality of data. Hence, utilization and decision made based on data generated will not be a true reflection of the services rendered.

In addition, the quality of the immunization system index revealed poor documentation practices, which tend to affect the quality of immunization data. The study brought to very weak M&E systems in the health facilities, as well as lack of written guidelines on M&E processes. These reasons explain why EPI data generated in most of the health facilities was of poor quality.

To ensure the reliability of EPI data for knowledge management in the healthcare deliver in Ghana, there is the need to put in appropriate measures to improve the quality of routine EPI data that is used in the planning and decisions making in the health sector.

Management of health facilities should earmark a storage unit and provide appropriate storage materials such as cabinets and files to improve reports storage and retrieval. Management should provide service providers with the Ghana Health Service Standard Operating Procedures on data management to serve as a reference material.

Hospital staff should be trained on good data management practices and the importance of data quality, as well as conducting orientation for newly posted staff. Immediate migration from paper-based data collection electronic capturing of data using tablets, mobile phones and laptops should be implemented to prevent the lag between immunization service provisioning and data capture. This approach is being used for the current COVID-19 vaccinations by the Ghana Health Service.

Regular supervisory visit should be carried out by hospital management, district and regional teams who conduct supervisory visits to health facilities should give a written feedback on the quality of data they generate. Facilities should be provided with standard register and EPI tally books by the District Health Directorate to avoid the use of improvised registers where there are no internet connectivity or electronic means to capture data electronically.

Data validation teams should be made mandatory in all health facilities by the District Management Team. Monthly data validation should be made organized during data compilation, data entry and after it has been completed in DHIMS II. Data quality checks that will prevent data entry errors should be incorporated into DHIMS II database. Head of units should assess monthly reports by checking for data quality inconsistencies before endorsing.

CONCLUSION

In conclusion, quality healthcare data is non-negotiable in the delivery of quality healthcare. Like other industries, there is growing information technology solutions that are enabling smart decision making, organizational leaning and innovation in service delivery. In a developing country such as Ghana, the healthcare authorities should harness the competencies of growing technologically savvy youth, the available low-cost hardware and often opensource applications such the DHIMS II to improve healthcare data management processes in order to ensure data quality for effective and efficient knowledge management systems to solve the growing and complex healthcare needs of its citizens.

REFERENCES

Aaron, S. W., Willis, F., Nwaze, E., Dieng, B., Sipilanyambe, N., Daniels, D., . . . Rymane, T. K. (2017, October 21). *Vaccine wastage in Nigeria: An assessment of wastage rates and related vaccinator knowledge, attitudes and practices*. doi:10.1016/j.vaccine.2017.09.082

Ahanhanzo, Y. G., Ouedraogo, L. T., Kpozèhouen, A., Coppieters, Y., Makoutodé, M., & Wilmet-Dramaix, M. (2015). Data quality assessment in the routine health information system: an application of the Lot Quality Assurance Sampling in Benin. *Health Policy and Planning, 30*(7), 837-843.

Alexander, R. (2020). *5 Reasons Knowledge Management Is Crucial in the Healthcare Industry.* Retrieved from https://bloomfire.com/blog/3-reasons-knowledge-management-critical-healthcare-industry/: https://bloomfire.com

Alipour, J., & Ahmadi, M. (2017). Dimensions and assessment methods of data quality in health information systems. *Acta Medica Mediterranea*, *33*(2), 313–320.

Audit Commission. (2013). *The Audit Commission for Local Authorities and the National Health Service in England Annual report and accounts 2012/13*. NHIS.

Brown, D. W., Warrener, L., Scobie, H. M., Donadel, M., Waku-Kouomou, D., Mulders, M. N., & Rota, P. (2020). Rapid diagnostic tests to address challenges for global measles surveillance. *Current Opinion in Virology*, *41*, 77–84. doi:10.1016/j.coviro.2020.05.007 PMID:32615510

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*(0), 2. doi:10.5334/dsj-2015-002

GAVI. (2019, March). *Data improvement requirements for all types of Gavi support*. GAVI.

Ghana Health Service. (2015). *Annual Report, 2014.* Author.

GHS, WHO, & GAVI. (2018). Expanded Program on Immunization, Ghana field Guide. Accra: WHO.

Gyaase, P. O., Anane, E. T., & Armah, I. N. (2015). The use of information and communication technology (ICT) for knowledge management in the second cycle educational institutions in Ghana. *International Journal of Computers and Applications*, *128*(7), 7–13. doi:10.5120/ijca2015906600

Haughom, J. (2014). *Knowledge Management in Healthcare: It's More Important Than You Realize*. Retrieved 04 20, 2021, from https://www.healthcatalyst.com/enable-knowledge-management-in-healthcare: https://www.healthcatalyst.com

Heaton, A., Krudwig, K., Lorenson, T., Burgess, C., Cunningham, A., & Steinglass, R. (2017). Doses per vaccine vial container: An understated and underestimated driver of performance that needs more evidence. *Vaccine*, *35*(17), 2272–2278. doi:10.1016/j.vaccine.2016.11.066 PMID:28162822

Karamitri, I., Talias, M. A., & Bellali, T. (2017). Knowledge management practices in healthcare settings: A systematic review. *The International Journal of Health Planning and Management*, *32*(1), 4–18. doi:10.1002/hpm.2303 PMID:26153388

Lee, W.-T., Shin-Yuan, H., & Chau, P. Y. (2011). Influence of knowledge management infrastructure on innovative business processes and market-interrelationship performance: An empirical study of hospitals in Taiwan. *Journal of Global Information Management*, *19*(2), 67–89. doi:10.4018/jgim.2011040104

Levin, K. A. (2006). Evidence-based dentistry. *Study Design III: Cross-Sectional Studies, 7*(1), 24-25.

Meri, M. (2020). New Trends in HRM & Knowledge Management in the Health Sector beyond COVID-19-A Practical Model. *Business Excellence and Management, 10*(SI 1), 5-21.

MOH. (2016). *Minstry of Health National Policy Guidelines*. MOH.

Mohammed, A., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation & Knowledge*, *4*(2), 104–114. doi:10.1016/j.jik.2017.07.003

Nisingizwe, M. P., Iyer, H. S., Gashayija, M., Hirschhorn, L. R., Amoroso, C., Wilson, R., & Hedt-Gauthier, B. (2014). Toward utilization of data for program management and evaluation: Quality assessment of five years of health management information system data in Rwanda. *Global Health Action*, *7*(1), 25829. doi:10.3402/gha.v7.25829 PMID:25413722

Nor'ashikin, A., Tretiakov, A., Whiddett, D., & Hunterb, I. (2017). Knowledge management systems success in healthcare: Leadership matters. *International Journal of Medical Informatics*, *97*, 331–340. doi:10.1016/j.ijmedinf.2016.11.004 PMID:27919392

Scott, C. C. (2017). *Country Immunization Information System Assessments*. Academic Press.

Svensson, A., & Hedman, E. (2021). Operations Knowledge Management in Health Care. In Operations Management - Emerging Trend in the Digital Era. Intechopen. doi:10.5772/intechopen.93793

Sychareun, V., Hansana, V., Phengsavanh, A., Chaleunvong, K., Eunyoung, K., & Durham, J. (2014). Data verification at health centers and district health offices in Xiengkhouang and Houaphanh Provinces, Lao PDR. *BMC Health Services Research*, *14*(1), 255. doi:10.1186/1472-6963-14-255 PMID:24929940

UNICEF. (2020). Retrieved April 20, 2020, from https://www.unicef.org/ghana/immunization

Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., & Therón, R. (2019). Dashboard meta-model for knowledge management in technological ecosystem: A case study in healthcare. *Multidisciplinary Digital Publishing Institute Proceedings*, *31*(1), 44. doi:10.3390/proceedings2019031044

Wetherill, O., Lee, C., & Dietz, V. (2017). Causes of Poor Immunisation Data Quality and Proven Interventions : A Systematic Literature Review. *Ann Infect Dis Epidemiol*, *2*, 1–7.

WHO. (2006). *The Immunization Data Quality Self-assessment (DQS) Tool*. World Health Organization.

WHO. (2006). *Vaccine Preventable Diseases and the Global Immunisation vision and Strategy, 2006-2015*. WHO.

WHO. (2017). Data quality review: a toolkit for facility data quality assessment Module 2. Desk review of data quality. Geneva: WHO.

World Health Organization. (2012). *Global vaccine action plan 2011-2020*. World Health Organization.

Yawson, A. E., Bonsu, G., Senaya, L. K., Yawson, A. O., Eleeza, J. B., Awoonor-Williams, J. K., Banskota, H. K., & Agongo, E. E. (2017). Regional disparities in immunization services in Ghana through a bottleneck analysis approach: Implications for sustaining national gains in immunization. *Archives of Public Health*, *75*(1), 1–10. doi:10.118613690-017-0179-7 PMID:28293422

Yeh, Y. (2011). The implementation of knowledge management system in Taiwan's higher education. *Journal of College Teaching and Learning*, *2*(9). Advance online publication. doi:10.19030/tlc.v2i9.1861

Chapter 12
Knowledge Management is Why E-Government Exists

Idongesit Williams
https://orcid.org/0000-0002-5398-3579
Aalborg University, Copenhagen, Denmark

ABSTRACT

There are many countries in the world where e-government services are underdeveloped. In e-government literature, numerous reasons are attributed to the failures in the implementation of e-government services. A reason often overlooked is the fact that government agencies may not see the value of existing ICTs to the current knowledge management processes supporting the delivery of government services. In this chapter, the Mobilization-Decision theory is used to explain how the perceived knowledge management value that can be enabled using information and communication technologies resulted in the implementation of e-government services in Europe.

INTRODUCTION

This chapter explains why Knowledge Management (KM) is the reason government agencies implement e-government services. The aim of this chapter is not to invalidate other factors that impede the implementation of e-government services as identified in e-government literature. Rather this chapter indicates that the challenges identified in e-government literature exist because the government agency in question does not see the KM value in implementing the Information and Communication Technologies (ICT). In other words, the government agency perceives the technology as a disruption to the current process of delivering government services – resulting in their reluctance to implement e-government agencies.

Government agencies rely on information and knowledge in order to govern. As is the case with any organization, information and knowledge can be described as the life wire of government agencies. The information and knowledge input into government agencies are derived either from the interaction within and between government agencies or the interaction between government agencies (Kraemer & King, 1977), businesses and citizens (Nixon & Koutrakou, 2007). Knowledge can be derived from the information received via either various forms of explicit or implicit KM extraction techniques. In other

DOI: 10.4018/978-1-7998-6792-0.ch012

words knowledge derived from data within the information received can be externalized (codified) or internalized, (Nonaka, Ikujiro & von Krogh, 2009). The captured Knowledge can be stored, shared, transferred and used (Gyamfi & Williams, 2018), to support government activities and policies.

The advent of ICTs revealed the possibility towards facilitating KM digitally. ICTs enabled Information and Knowledge societies. The growth of the information and knowledge societies also revealed the value in the use ICT to enable and support KM processes. Furthermore, it revealed the value in ICT as an enabler of KM processes in the delivery of government services as well. However, government agencies in countries and regions that did identify the value enabled by ICT on their KM processes made the decision to mobilize resources in order to develop ICTs that can enable KM processes in government service delivery. In other words these government agencies implemented because the ICT supported KM processes in their agencies. Once such agencies identify the "KM value of the ICT", they always found a way to mobilize resources to implement e-government. However, government agencies that could not perceive such value were either reluctant or refused to implement an e-government services.

Europe is a region where most countries have, over the centuries, mobilized resources to develop e-government services using different ICTs (UN DESA, 2020). This has resulted in their very high E-Government Development Index (EGDI). See figure 1. "*The EGDI is a composite measure of three important dimensions of e-government, namely: provision of online services, telecommunication connectivity and human capacity*" (UN DESA, 2020). The EGDI scale consist of very high EGDI (0.75 -1.0), High EGDI (0.5- 0.7499), Middle EGDI (0.25 - 0.4999) and low EGDI (0.0 - 0.2499) (ibid). Countries in Europe rank between the high and Very high EGDI. In other regions there are countries that rank from either low to very High EGDIs, medium to very high EGDI or low to High EGDI. More on this will be discussed in section 2. Nevertheless based on the analogy made so far, one could say that the countries in Europe, and countries with high EGDI in other regions, perceived the value of KM in the delivery of their government services. Hence, they made the decision to mobilize resources to implement e-government services. To explain this point further, the Mobilization-Decision theory (Williams, 2021) is used to analyses different initiatives and settings that led to e-government development in Europe. The focus will be examples extracted from national initiatives spanning a little over two centuries. These examples will be pointers to the decision to mobilize resources to implement e-government services.

This chapter builds on previous research by the researcher on e-government and telecom infrastructure development. The chapter is divided into 7 sections. Section 1: introduction; second 2: explains the rational for choosing Europe as a case; Section 3: describes the level of e-government infrastructure development in Europe using the EGDI index; Section 4: explains the theory used for analysis; section 5: is the analysis; Sections 6 & 7 sixth: are the discussion and conclusion.

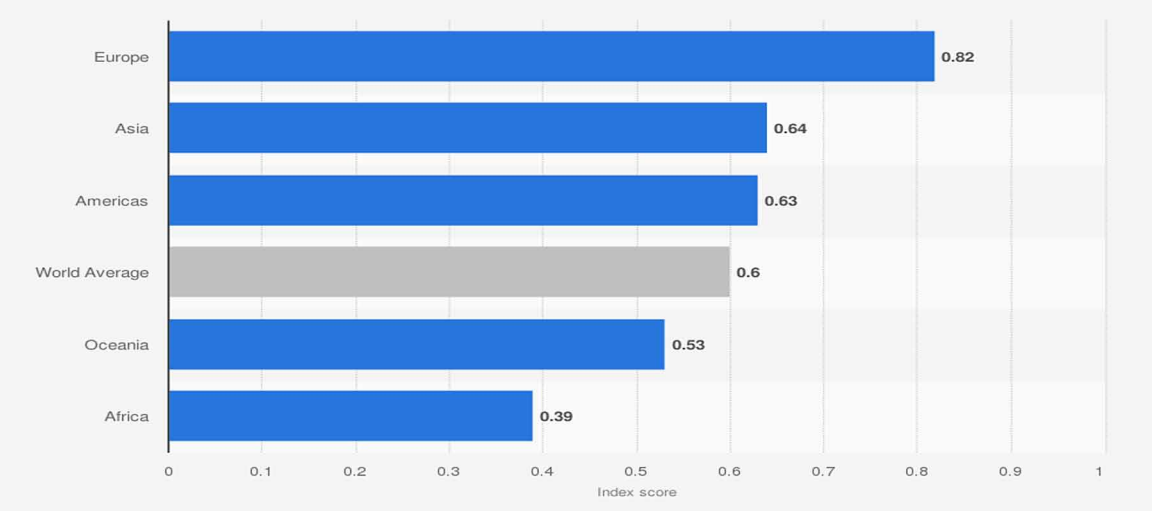
WHY EUROPE IS CHOSEN AS A CASE

This section provides a background on why Europe was selected as a case for this research. The state of e-government implementation in each region varies. Each year, from 2012, the United Nations Department of Economic and Social Affairs publishes the UN e-government survey. The survey, among other, things, provides an insight into the state of e-government development around the world, nationally and regionally. The index used in evaluating the state of e-government implantation is the EGDI index, mentioned earlier in the introduction. The EGDI consists of the Online Service Index (OSI), Human Capital Index (HCI) and The Telecom Infrastructure Index (TII) (UN DESA, 2020). The TII

index indicates of the level of penetration and usage of the Internet, mobile cellular telephony, mobile broadband and fixed broadband infrastructure within either a country or region (ibid). The HCI consist of the level of adult literacy in the country/region, expected years of schooling, mean years of schooling and gross enrolment ratio (ibid). Human capacity ultimately denotes the ability of the stakeholders in the e-government service ecosystem to interact with the e-government services. The assumption is that the higher the education level of the user, the greater the skill they possess in order to use e-government services. The OSI denotes the percentage of government services that are delivered digitally and the type of services they provide (ibid). The type of service could be either informational or transactional.

Based on the UN E-government survey, it is evident that, as a region, the state of development of e-government infrastructure is higher in Europe than that of other regions.

Figure 1. Regional EGDI outlook 2020
Source (UN DESA, 2020)

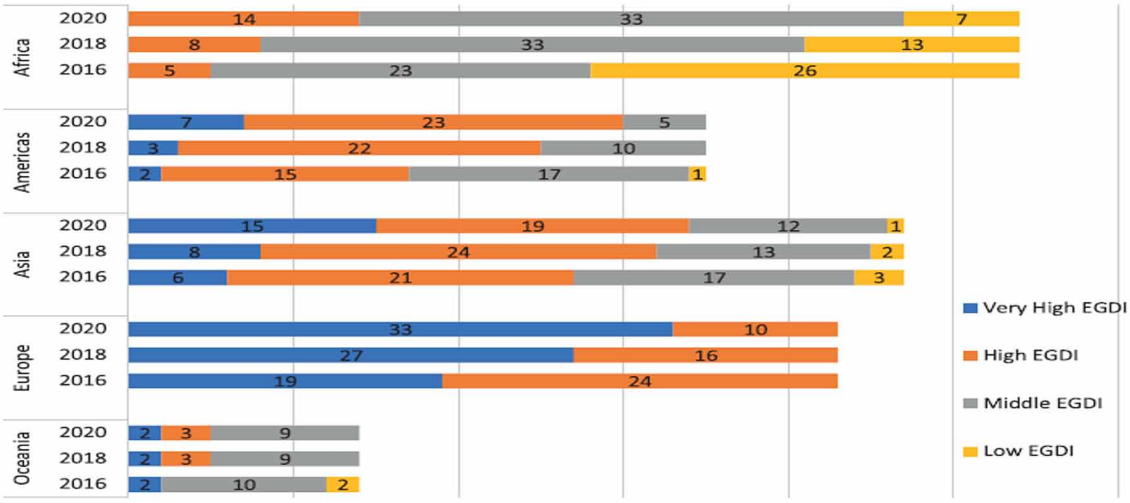


Although Europe has a high EGDI, that does not imply that countries in Europe are without their own unique challenges in the delivery of e-government services. The fact that the European EGDI is not 1, as seen in the figure above, points to the fact that there are challenges in either the provision of online services, HCI or the delivery of telecom infrastructure in some European countries. The outlook in figure 1 does not indicate that countries in Europe are without either equals or superiors in other region. Rather it indicates that countries in Europe possess either high or very high EGDI as mentioned in the introduction.

Having cleared that fact, it is evident in figure 1 that the EGDI of Asia, Americas, Oceania and Africa respectively are less than that of Europe. This has been the case from 2012 when the UN survey was first introduced (UN DESA, 2012). The only difference has been that the Americas in some of the previous surveys recorded slightly higher EGDI than Asia. However, as seen in figure 1, Asia had a very slight lead in 2020. Nevertheless, if one looks at figure 2 below, it is evident that the number of very high EGDI has been on the increase for the past three years. This has resulted in the shrinkage on the number of countries with high EGDI. Similar increase in the number of countries with very high

EGDI in Asia and a slight increase in the Americas is noticeable. However, the shrink is noticeable in the number of countries with middle and low EGDI in the Americas, but a very slight shrink in Asia. In Africa, there has been a noticeable increase in the number of countries with high EGDI and a shrink in the number of countries with low EGDI but the number of countries with middle EGDI has been constant between 2018 and 2020.

Figure 2. Three-year EGDI comparison between Europe, Americas, Asia, African and Oceania
Source (UN DESA, 2020)



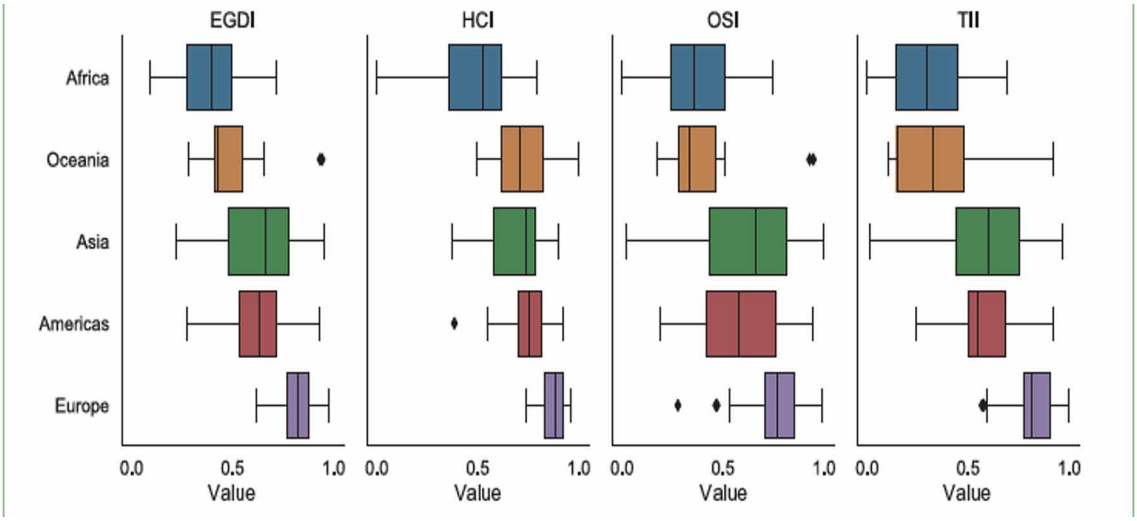
However, in Oceania, Australia and New Zealand have always been countries with high EGDI. The EGDI of other countries in the region has been stagnant since 2018.

The outlook from figures 1 and 2 points to the fact that there are countries that still have challenges with either their TII, HCI, OSI or in all the three indices in the other regions, aside Europe. Figure 3 provides a breakdown of the TII; HCI and the OSI at the regional level.

The challenges in Africa as seen in figure 3 are those of HCI, and TII. The African TII is below 0.5 pointing to a greater digital divide (both telecoms and Internet) in more African countries as compared to the number (percentage wise) of countries with similar problems in other regions. In the same vain, the African HCI outlook implies that the literacy rate and the rate of enrolment in the various layers of education is behind that of the rest of the world. Nevertheless, it is important to note that by 2010 between 65% and 70% of Africans were enrolled in education (Roser & Ortiz-Ospina, 2018). Therefore, the situation percentagewise is not that bad. However, the uneducated 30% to 35% of Africans is still a huge population as per the population of Africa. Nevertheless, these TII and HCI related challenges obviously has an impact on the availability, of online services, and that of the Internet infrastructure required to access such services.

The HCI of Oceania is almost at par with the rest of the world but its OSI and TII. As in the case of Africa, the TII and OSI index of Oceania is lower than average. This denotes a low number of online services and low availability of telecom infrastructure in most Oceanian countries. The low penetra-

Figure 3. TII, OSI and HCI comparison per region
Source (UN DESA, 2020)



tion of telecom infrastructure makes it challenging to deliver online services. However, as mentioned in the chapter, the lack of prioritization in the implementation of e-government still boils down to the inability of the government agency to perceive the KM value of the existing ICT. In a situation where that value was identified, then such Government agency would adopt different policy initiatives that would improve the TII index much further. Australia and New Zealand are the countries are exceptions to the case (UN DESA, 2020).

However, the regions where the increasing value of KM in e-government services is gaining traction are the Americas and Asia. Countries in the Americas, North American possess very high EGDI. However, where the America and Asian lag more, as compared in Europe, is in the TII and HCI. Asia has a slight edge and a greater edge on the Americas in the OSI and TII indices respectively. This implies that the percentage of countries with higher TII and OSI is greater than a similar percentage of countries in the Americas. However, a fewer percentage of countries face these challenges as compared to the percentages of countries in Africa and Oceania.

Hence, briefly, there are countries in other regions, outside Europe, where the value of KM in e-government service delivery is not perceived in most government services. Hence, the decision was made to use Europe as the case for analysis. In the next section, the EGDI snapshot of Europe is presented.

EGDI OUTLOOK OF EUROPE

In the previous section, an EGDI comparison across regions was used to make a case for why Europe is used as a case in this chapter. In this section, an overview of the TII, OSI and HCI in Europe is presented. The presentation of the outlook is based on statistics and data highlighting the TII, OSI and HCI of the countries in Europe. The source of data include, academic literature, reports from ITU, the EU and UN sources cited in this section.

The TII Outlook in the EU

Different European countries have made giant strides, as most countries in Europe possess high and very high TII (UN DESA, 2020). A brief outlook of this stride is visible in the demand side and supply side of the telecommunications market. On the supply-side, the level of penetration of mobile and fixed broadband telecom and internet network in European countries is very high (IHS, Omdia, & PointTopic, 2020). However, on the demand side, the level of subscription is greater for Internet and mobile broadband subscription than for fixed broadband network subscriptions. The number of active fixed broadband subscription in Europe is 33% (ITU, 2020). While the percentage of active mobile broadband subscribers is, 100% of European citizens and 86.5% of EU citizens have access to the internet in their homes (ibid). Furthermore, 82% of households in Europe own computers (ibid). The average use of the Internet in the Europe stands at 83% (ibid). However when it comes to the regularity in the use of the Internet, that varies per country. For example, the use of Internet services is high countries such as Denmark Finland, Netherlands, and UK (DESI, 2020). More than 70% of the population use the Internet in these countries (ibid). On the other side of the coin, less than 50% of the population use Internet services in Portugal, Greece, Italy, Bulgaria and Romania, less than least 50% of the inhabitants of use the Internet (ibid). But when it comes to the regularity of the use of the internet by existing users, more than 60 of active access the Internet in the European Union (not the entire Europe this time) at least once a week (ibid). Now just as in other regions of the world, there are part of the population in different European countries who have never used the Internet. The lack of usage of the Internet is not because of the lack of neither the Internet nor telecom networks, rather it is either because the user either possesses a low educational background, is a senior citizen or a pensioner whose interest is not the internet (ibid).

Based on the outlook presented, it is evident why the TII is high. However, it is important to note that Europe has areas that are not commercially viable for telecom network operators. Hence, different innovative top-down and bottom-up approach to Broadband infrastructure was adopted to extend technology neutral broadband connectivity to rural areas (Williams, 2015). Hence, in nature and data rates from the Broadband connectivity found in some provincial areas might not be the same as that found in bigger cities. But as a result of the fact that government agencies realized the KM value in the digitization of government services (Nixon & Koutrakou, 2007), efforts were made to extend Broadband connection to non-commercially viable areas, even though most people might access the Internet rarely.

The HCI Outlook in the EU

The level of adult literacy is high. However, when it comes to the level of education completed, the number of persons that pursue higher education is relatively smaller than those that completed basis and secondary education. Nevertheless, an insight from UNESCO in 2015 indicates that the literacy rates of Europeans, for ages 15-24 is 100% (Roser & Ortiz-Ospina, 2018). While the literacy rate of Europeans from Ages 65 and above was between 80% and 100% (ibid). The expansion of basic education began in Europe. As a result, the literacy rate in Europe has been higher than that of other regions, except for the North Americans. For example, between the 1860s and 1880s, almost 30% of Eastern Europeans and 60% of Western Europeans ages 15 were enrolled in education (ibid). Other regions aside North America had less number of enrolments. By 1990s, the rate of enrolment in education was 100% (ibid). Aside the huge number of enrolment and being able to read and write, in 2019 the number of persons between 25 and 54 years in the European Union (this time around) who have completed high school was

80.8% (Eurostat, 2020). Furthermore in the by the same year, 66.2% of persons in the EU between 55 and 74 years had at least a high school degree (ibid). As it pertains to the attainment of tertiary education, 34.6% of those aged 25–54 and 20.9% of those aged 55–74 completed tertiary education. This outlook has an impact on the statistics on the average number of Internet users mentioned in the TII. However, the outlook indicates that the average European can interact with at least informational e-government services provided over the Internet. However being able to interact with information provided online does not necessary mean that the person can perform a transaction. This is especially so if the user does not know how to navigate the internet. But experiences in Denmark, UK, Netherlands, Estonia, Norway, Finland and Sweden to mention a few points to the fact that, making the use of e-government service compulsory forces people to learn how to use the Internet and its services. This is an area where countries in the EU could learn from one another.

However, based on the UN 2020 e-government survey, their measurement is not on how many persons that have attained different levels of Education but on the gross enrolment ratio, the expected number years for schooling and the mean years of schooling. Based on that assessment they rate the European HCI index as very high. As mentioned earlier, the aim of the HCI is to measure the ability of the stakeholders in the e-government service ecosystem to interact with the e-government services. Hence, after assessing the level of literacy in Europe, this chapter agrees with the survey-even though different methods of assessment is used - that the HCI in Europe high. Hence, government agencies are able to deliver e-government services bearing in mind that the user is educated enough to learn how to use the service.

The OSI Outlook in the EU

The final outlook is the OSI. The focus in this chapter is not about the sophistication of the government services delivered online. Rather the focus is on the number of online services, the percentage of e-services that are digitized per country in the EU can be seen in figure 4 below. The focus in this section is on the EU, because of the scarcity of data on the overall European outlook- as at the time this chapter was developed. However, as seen in the figure below, apart from Romania, every EU Member state delivers more than 50% of their government services online.

This implies that there are government services in EU member state that are not online. Furthermore, some e-government services still require paper work. Hence, part of the process is performed online while other parts of the process are performed offline. Figure 5 provides an insight into countries where the percentage of e-government services are delivered either fully online or partially online.

In figure 5, it is evident that at least 60% of transactions in existing e-government services in EU member states performed online as at 2020. However, it is only Malta where all transactions with their existing e-government services are performed online.

Nevertheless, in principle the delivery of e-government services in the European Union is high. The approach used in this assessment differs from that used by the UN 2020 E-government survey. Nevertheless, it provides a foundational basis for the insight provided in the survey as it pertains to the sophistication in the delivery of e-government services in Europe.

Based on the TII, HCI and OSI outlooks presented in this chapter, it was imperative that the level of e-government implementation and service delivery in Europe is very high. Hence, one of the major drivers for this outlook is the fact that Governments in Europe can identify the KM value that can be

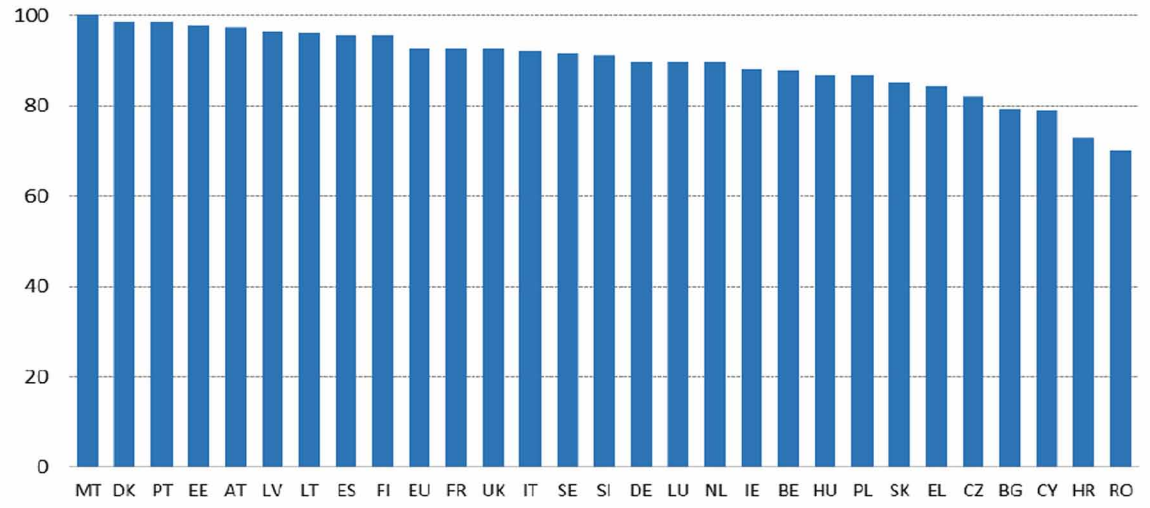
Figure 4. Percentage of government services digitized in each EU country
Source (EU, 2020)

derived in the use of ICTs to deliver government services. In the next section, a brief overview of the theory used to analyze the decision to deliver e-government services in European countries is presented.

MOBILIZATION-DECISION THEORY

The premise for Mobilization-decision theory hinges on why mobilization occurs. To mobilize means to assemble often for performing an action. It would be strange to witness a mobilization effort if an

Figure 5. Percentage of existing e-government services delivered fully or partially online
Source (DESI, 2020)



action is not looming in the horizon. Mobilization itself is an act, in this case an act of preparation. Hence, mobilization can be viewed as an action that precedes another action. Hence, mobilization has two aspects, namely the recruitment aspect and the deployment aspect.

Mobilization is often preceded by a decision. The action to be performed is the object of the decision. Decision is a conscious rational activity or process driven by different factors. The mobilization-decision theory highlights the factors that result in the desire to either recruit or be recruited in order to perform an action (Williams, 2021). The recruitment can be human centric, machine centric, human-to-machine recruitment, machine-to-machine recruitment or machine-to-human recruitment. The recruitment of humans by machines has been discussed as a possibility (Williams, 2020) (Leslie, 2019) (Fugener, 2019). Nevertheless, in each of these forms of recruitment, the mobilization-decision theory posits that the same factors that result in the decision to mobilize exists. Before proceeding in the discussion, it is important to relay the theory.

Mobilization-Decision theory has its origins in the social sciences. It owes the origin of its concept to social theories such as Actor Network theory (Latour, 1996), social movement theories and the logic of collective action (Smelser, 1963). However, its application goes beyond the boundaries of the social sciences as earlier alluded to. The commonality within these theories is the concept of mobilization. The observations within the theory originate from earlier models developed in 2015 (Williams, 2015). These were models developed based on the grounded theory approach and later theorized in 2020. The models provided an insight into the process that results in the decision to mobilize resources aimed at performing an action. The emerging theory was the Mobilization-decision theory.

The theory states: "*the decision process towards either mobilizing a person or group of persons or being mobilized to perform a certain action will occur if the following conditions are satisfied. These are: the evidence of needed resources, the mental congruence between the mobilizer and the mobilized; the evidence of either actual or potential cooperativeness between the mobilizer and the mobilized and the perception of individual value to be derived between the mobilizer and the mobilized from the action to be performed*".

Source (Williams, 2021)

Hence, the factors that result in the decision to mobilize are: "resource availability", "congruence", "cooperativeness" and "value" (ibid). The absence of any of these factors will result in the failure to mobilize. Furthermore, each factor could be a driver resulting in the other factors becoming relevant. However, as mentioned earlier, mobilization involves recruitment or the assembly of resources. This could be self-recruitment (self-mobilization), recruitment of others (mobilizing others) and being recruited (being mobilized). Let us take a look at how the recruitment process works.

Scenario 1: Self-Mobilization

If agent "A" decides to mobilize his/herself to perform an action. The **resources or vital resources** needed to perform the action should be either available or he/she can see the possibility of harnessing such a resource. He or she should possess some form of self-belief that he or she can perform that action. This implies that there is an agreement or similarity (**congruence**) in thought process between his/essence or soul and his mind- a convergence of rationality. In other words, there is conviction in his/her mind. He or she should be willing to perform the action. This implies that he/she cooperate (**cooperativeness)** with himself/herself and makes a move. Finally, he/she should see the **value** in performing the action.

The theory posits that agent "A" will mobilize his or herself to perform the action. However, if one of the factors is absent, the theory posits that agent "A" will not perform the action.

Scenario 2: Mobilization and Being Mobilized

Now assuming agent "A" is self-mobilized and decides to mobilize agent "B". In this case Agent B is the one being organized. Agent "B" will decide to become mobilized if he or she is convinced by agent "A" that the four factors are present. First, Agent "B" has to be convinced that the **resources or vital resources** needed to perform the action exist. This could be either resources owned by Agent "A" that would accommodate Agent "B" or Resources owned by Agent "B" which could assist in performing the action. Secondly, Agent "B" has to possess either the same or similar (**congruence**) line of thought as agent "A". This could either as a result of being convinced by agent "A" or they already think in the same way. Thirdly, Agent "B" has to be willing to work with (**cooperativeness**) Agent "A". Finally, Agent "B" must see the **value** in being mobilized by Agent "A". Their values could be either similar or dissimilar. Just as in the first scenario, if one of the factors is absent, the theory posits that Agent "B" will decide not to be mobilized to perform the action promoted by Agent "A".

There will be obvious minor nuances between human centric recruitment and machine related recruitment. For human related recruitment, the scenarios will work as it is. However, for machine related recruitment, machines are not rational beings. Their rationality is programmed but in a more restricted way than the way, we humans are programmed. Even in the case of machine learning or AI where the machines evolve, the evolution is based on the information it was programmed to assimilate in the environment it finds itself. Once the machine operates outside its natural habitat, it malfunctions. Nevertheless, congruence and cooperativeness is identified by the feedback received when trying to engage the machine. For example, the congruence between humans and cars is in either the movement or stoppage of the car. If the driver performs the wrong action, cooperativeness will not be possible. The break in cooperativeness will result in the car not moving or in the car crashing in an accident if it is still moving when the breaks are applied. In machines, the resources are in most cases inherent. If one of the resources malfunction, the machine is likely to malfunction. Value on the other hand is what the machine has been programmed to value.

If we juxtapose this theory with the context of this chapter, national governments are agents. They either can be self-mobilized, or are mobilized by other agencies such as regional bodies and International governmental organizations to develop. The relevant action here is the implementation of e-government services. The value derived in the implementation is the facilitation of KM practices in government service delivery. Hence, in order to make the decision to mobilize resources for this action, they ought to:

1. Possess the **resource**s needed to achieve this goal;
2. Be convinced of the usefulness and task-fitness of the ICT to their current KM processes (**congruence**),
3. Be willing to devise the necessary steps to work towards this goal (**cooperativeness)**
4. **Realize the value of the ICT to the enhancement of their KM processes** (**value**).

This theory will be used in the next section to analyze the factors that resulted in the self-mobilization (national centric) of European countries to implement e-government services.

MOBILIZATION-DECISION THEORY, KM AND E-GOVERNMENT IMPLEMENTATION IN THE EU.

Europe as highlighted earlier in this chapter has a long history with ICTs. This dates back to the 19th century. The period can be divided into 3 sub-periods. The first being the Electronic Data Processors, telegraph and telephony Era; the convergence of telephony and Electronic Data Processors (now called computers) era; and the mobile cellular telephony, broadband and future ICT. In each era, the pointers to congruence and cooperativeness will be discussed. The value and resources were the same across the three era, so that will be treated separately.

Era 1: The Electronic Data Processors (EDP), Telegraph and Telephony (Fixed Line Telephone) Era

Congruence: The 19th century saw the extensive development of telegraph networks in European countries and their colonies (for those with colonies) (Williams, 2015). Such networks could be seen in Scandinavia (See (Derdak & Hast, 1992) (Tomas, 2014)), the United Kingdom, Finland, Germany and France to mention a few (Williams, 2015). By the late 19th century and early 20th century, led by American and European innovators and investors, telephony networks were also expanding in Europe as well (ibid). In the early 20th century, the UK census office (Agar, 2003) and private European companies (Calhoun, 1956) (IBM, 2017) were adopting electronic tabulators (Electronic Data Processors (EDP)), developed by US Inventor Hollerith (Hollerith, 1894), to handle electronic data processing. By 1910 and the 1930s, Hollerith's company IBM was present in Germany, Central and Eastern Europe (IBM, 2017). These technologies enabling limited knowledge creation (EDP only), knowledge transfer and knowledge sharing within and between government agencies. The reported observations were on the emergence and use of the telegraph, the telephone and Electronic Data Processors (Williams, 2020)

The invention and eventual adoption of these ICT, by innovators and the early adopters (mostly private companies and national census bureaus in the US and Europe) (Cortada, 2018), enabled government agencies in Europe to become familiar with the KM value proposition presented by these technologies. The telegraph and telephone networks revealed the potential for information and knowledge transfer as well as information and Knowledge sharing within and between government agencies. It also revealed the potential for either information or knowledge sharing and transfer between government agencies, and from businesses and citizens to government agencies. Hence, these network technologies (telegraph and telephone) has the potential towards kick starting an information society. The telegraph and telephone (see (Daft, Lengel, & Trevino, 1987) (Gyamfi & Williams, 2018)) enabled the transfer of explicit and implicit knowledge respectively. The only challenge with the telegraph was that one needed to understand Morse code in order to achieve a successful transfer of knowledge. In the same vain, the EDPs enabled knowledge creation activities using semi-automated data processing techniques. For the first time, voluminous explicit knowledge could be created digitally as opposed to the limited explicit knowledge shared via the telegraph systems. Nevertheless, data processing could be achieved at a reduced cost, as the number of hands required for the process would reduce drastically.

These developments resulted in decision-makers within government agencies in Europe to be convinced of the usefulness and the need to adopt these technologies for some of their KM activities (Williams, 2020). Hence, there was congruence that ICTs had the potential of enabling KM in their e-government service delivery processes.

Cooperativeness: By the 1960s and 70's, there is evidence of cooperativeness identified by the move by some countries in Europe adopting EDP in their G2G e-government service delivery. Cooperativeness implying that government agencies were willing to adopt the EDP, despite regulatory barriers. The agencies that adopted EDPs were interested in efficient approaches towards Knowledge creation. In Europe, one of the early adoption of EDPs were adopted by Post, telegraphs and telephone (Calhoun, 1956). There are also examples of countries where EDP were adopted and used in few government agencies were France (Baquiast and van de Donk, 1989); Germany, the then Soviet Union, Italy and Sweden (Schumacher, 1967); and Croatia (Kliček and Vukovac, 2007) to name a few. In Germany the adoption of EDPs, for public administration were limited due to limitations posed by data privacy regulations (Reinermann, 1997) (also see (Kliček & Vukovac, 2007)). This is as opposed to the US where there was wider usage of the EDPs at the state and federal level (Price & Mulvihill, 1965). The challenge in the adoption of the EDP were due to high cost of renting/leasing EDPs by government agencies and the increasing fear from government employees on losing their jobs to computers (See (Reinermann, 1997)). Aside the adoption of EDP, there were innovative public-private funding approach aimed at funding the developments of EDP in Europe (Calhoun, 1956). Some countries instituted computer policies with one of the aims being the usage of computers in public administration (Baquiast & van de Donk, 1989). Hence, public administration could rent the EDPs locally rather than rely on the ones from the US.

In the case of the network technologies, the telegraph had very limited knowledge sharing, knowledge transfer functionality as mentioned earlier. Furthermore, the tele density of telephony in the early 20th century Europe were low, and the cost of access to the telephone then was high. The high cost in some cases were due to interconnectivity wars from different telephone operators (Williams, 2015). In order to deal with some of the challenges pertaining to cost of access and interconnectivity of network providers, countries in Europe initiated telecom policy aimed at promoting the tele density of their existing networks. Majority of these policies resulted in the consolidation of numerous telephone networks into either a natural or a national monopoly (Williams, 2015). Such monopolies enabled the delivery of unified services and the reduction in the gradual cost of tariff (ibid). These networks existed well into the 1990s.

These initiatives paved the way for the expansion of the use of EDPs and network technologies in government agencies and society. The fact that these initiatives among many others were adopted, despite the legal challenges of those days, points to the fact that KM value of the ICT outweighed its limitations. It could also have been that they ICTs in existence then were limited, so there was no luxury for choice. However, the existence of this value resulted in the motivation of using innovative approaches to the ICT in their government services delivery processes.

Era 2: The Convergence of Telephony and Electronic Data Processors (Now Called Computers)

Congruence: Evidence of congruence in this era is seen in the increasing adoption of computers and the expansion in the delivery of telecom networks. One could say that the evidence of congruence in this era is a spillover of the congruence in the previous era. What made this congruence stronger was the upgrade in the KM capabilities enabled by the convergence of Computer and telephony (network) technologies. This convergence began in the 1970s (Dickey, 2017). During this period the evolution of ICT of value was more rapid than in the previous era. This is because, the convergence of computer and telephony technologies opened up the possibility of remote computer connectivity. Hence there was value is the gradual digitization of more G2B and G2C government service.

The convergence was enabled by peer-to-peer EDP (now computers) communication using Local Area Networks (LAN), and Wide Area Networks (WAN). From the 1970s upwards, computers were increasingly equipped with functionalities that enables it to host applications, databases as well as store data (CHM, 2019). The computing abilities of computers enabled not only the computation of complex statistics but:

1. the storage of explicit knowledge in word processors,
2. The ability to query different forms of data, information and explicit knowledge from databases.
3. And the ability to use different forms of applications to combine knowledge in order to create new knowledge.

It was in at this point in history where countries such as France saw the need for "informatization" or "computerization" as identified in the Nora Minc report (Nora & Minc, 1978).

The convergence also coincided with the greater expansion of the telephony network. By the late 1970s, the tele density in some European countries were high. In some countries such as Sweden, Denmark and Luxemburg some households could boast of more than one telephones (Minges, 1999). In this landscape, the idea for the remote delivery of e-government service to citizens and businesses looked feasible. Hence, there was a push for every household to own a telephone. This push extended to the 1990s, as seen in table 1 below. Furthermore, it still exists until date but with different technologies.

Table 1. Percentage of household with telephone in Europe between 1970 and 1990

Country	Year	Percentage of Household with telephone
Sweden	1975	100
Denmark	1982	100
France	1985	97
Finland	1997	90
Luxembourg	1989	100

Source (Minges, 1999)

However, so far the transfer of tacit or implicit knowledge from businesses and citizens over the telephone was the value government agencies could identify. Other countries were not far behind but most countries in Europe. Most European countries were able to provide more than 90% of households in their country with telephony in the 1990s.

Nevertheless, the coincidence in the convergence of ICTs and the expansion of telecom networks opened up new possibilities for remote transfer and sharing of explicit knowledge. The new possibilities occurred because of Videotex online technologies and much later with the World Wide Web. Hence for the first time, knowledge creation, knowledge creation, Knowledge sharing and the transfer of codified or explicit knowledge in G2G, G2C, and G2E government services became possible.

The emerging technological landscape were accompanied with access devices that were becoming simpler and suitable for KM activities and Knowledge storage was becoming possible. Hence, task-fitness and perceived ease of use (Oliver, 1980) of ICTs for KM were revealed by the day. This led to stronger

congruence in the mind of policy makers, which led to their desire to eventually adopt the computer and telephony services.

Cooperativeness: By the 1970s to the 1990s, the barrier to cooperativeness was lower in Europe. This is because of the growing congruence on the usefulness of the converged ICT. Hence, by the early 1980s, before the commercialization of the internet, countries such as France, Germany, Belgium, Finland, Ireland, Italy and Netherlands, to name a few, deployed online videotext systems or Minitel systems (see examples (Schofield, 2012) (DER SPIEGEL, 1983)). In France, the system was used to delivery government informational services (Schofield, 2012). However, as a sign of cooperativeness resulting in positive congruence, European countries, such as Norway, Croatia and Cyprus, began developing computerization and digital public administration policies. The aim was to create a roadmap that will enable the, taking advantage of the KM potentials presented by the new technologies available then. Another point where evidence of cooperativeness can be identified is with the commercialization of the Internet in the 1990s. Countries such as Germany and Greece embarked on pilot projects to see the feasibility and potentials of these new technologies in public service administration etc. (see example (SYZEFXIS Project, 2004)). Other countries such as Austria ("IT-Kooperation zwischen Bund und Landern, 1998) and Estonia (Estonian Informatics center, 1998) among others were developing policies and national visions aimed at developing information society -with e-government at the core of these plans. There are many other initiatives that one could point to but this is a snapshot pointing to actions that denote the existence of cooperativeness towards implementing e-government services.

Era 3: The Growth of Mobile Cellular Telephony, Broadband and Future Digital Technologies

Congruence: As ICTs evolve, so was the growth in congruence in Europe. By the 1990s, there was another convergence. This was the convergence of broadcast, the existing converged ICT and electronic communications networks. The World Wide Web was one of the drivers of this convergence. Hence, at that time, e-government policies (addressed in ICT policies) were aimed at enabling the infrastructure that will transmit the Internet between government agencies, to homes and businesses. However, this evidence of cooperativeness became successful when countries in Europe and in the West embraced the liberalization and the regulation of the ICT market. Above all technology neutrality was promoted. This led to the standardization and delivery of digital cellular, mobile networks and within few years' Broadband networks.

The entrance of Broadband networks made the dream of enabling Knowledge sharing, knowledge aggregation, knowledge transfer and knowledge storage in much more government services possible. Hence, ICT was compatible with the transfer of rich media, such as video and multimedia applications. Hence, government agencies in the EU saw the possibility of the simultaneous transfer of tacit and explicit knowledge on the internet to citizens, businesses and governments using knowledge bases. They also had the possibility towards engage actively in remote collaborations that support simultaneous knowledge transfer and knowledge exchange activities. Furthermore, existing Knowledge sharing activities such as workshops, seminars, and meetings can be held online. In the same vain Knowledge transfer, activities can be held online. These digital KM possibilities enabled by broadband has been the driving force behind current e-government implementation in European countries. The European Union has been on Board and working on facilitating cross border knowledge sharing within Europe. Hence, the congruence on the value of ICT for KM has been strengthened.

The current use of future ICTs such as Machine Learning and Artificial Intelligence (AI), is on the rise globally and there is evidence that congruence of its KM value in the delivery of e-government services is growing among government agencies. An example will be cited when discussing cooperativeness.

Cooperativeness: From the 1990s, the desire to implement e-government services in Europe has been on a fever pitch. There is growing cooperativeness in the region. This is evident in policy initiatives developed by countries in the Europe aimed at the diffusion of either mobile cellular telephony, mobile broadband, fixed broadband technologies, Next Generation Networks etc. (see (Williams, 2015)). These policies, be they ICT, broadband or digitalization policies, AI policies etc, are often pragmatic, reflective and proactive. The pragmatic aspect of the policies include the policy implementation timeline (mostly 3 -5 years), clear policy objectives, clear policy funding initiatives and a clear roadmap for the e-government implementation and national infrastructure development (see example (Regeringskansliet, 2014)). The reflective aspects provides insights in the achievements and challenges in the implementation of previous policies and the road ahead (ibid). The proactive element provides an insight into the near future, the possible upcoming technologies of interest and the plan to support the growth of the market (ibid). Today, in some European countries, the current policies are open data policies and AI policies ((Williams, 2020)). Countries like Estonia, are working on implementing AI in e-government (MKM, 2020) (Plantera, 2017). Hence, currently, in Estonia, there is cooperativeness towards AI. Based on the historical track record of Europe, these initiatives that point to cooperativeness will continue. Although the focus of this chapter is on national centric examples, one cannot fail to mention that with the mobilizing effort of the EU, the cooperativeness will result in more sophisticated e-government infrastructure in EU member states (see (Williams, Falch, & Tadayoni, 2018) (Williams, Falch, & Tadayoni, 2020)).

Value Across the Three Era

In each Era, different KM values were identified were identified by government agencies in Europe. Some technologies mentioned in this chapter are presents in table 2 below. In the table below, the technologies that are emboldened are those that still exist. Others are either extinct, defunct or less used.

Table 2. Different ICTs and their KM value

	Computer technologies				Network technologies				
KM value	**EDP**	**Extant Computer***	**Videotex**	**World wide web**	**Telegraphy**	**Fixed line Telephony ****	**Mobile telephony *****	**Broadband**	**Future tech**
Knowledge creation	E	E,I	E	E,I	E	I	E,I	E,I	E,I
Knowledge sharing		E,I	E	E,I	E	I	E,I	E,I	E,I
Knowledge transfer			E	E,I		I	E,I	E,I	E,I
Knowledge aggregation/use		E,I		E,I				E,I	E,I
Knowledge storage		E,I				I	E,I	E,I	E,I

E=Explicit Knowledge I=Implicit Knowledge *Extant computers can store sound and videos. **modern telephone have voice mailboxes. ***SMS stores explicit information

When describing the evidence of congruence and cooperativeness in different era most of KM value were mentioned. These values, among others, served as drivers for the continuous upgrade and adoption of different ICT. These technologies were of great value during the lockdown attributed to the COVID 19 pandemic, as the various KM activities within government agencies were not disrupted because of the pandemic. All activities in most European countries occurred online. What is of interest in table 2, is that technologies that exist today support the creation, sharing, transfer, aggregation, and storage of both explicit and implicit knowledge to and from government agencies. These values will continue to be enhanced as new technologies emerge. Furthermore, based on the evidence of growing congruence and cooperativeness described earlier, government agencies in European countries will continue to adopt future technologies for their e-government services.

Resources Across the Eras

The presence of value, congruence and cooperativeness is not enough. The resources that should be mobilized should be available. When discussing congruence and cooperativeness, there were mention of resources that supported the policies, the development, and the evaluation of the technological landscape. To aid in this description of this section, the seven resources needed to develop technological systems was consulted (Krar) ICTs are technological systems. The seven resources include people, information, materials, tools and machines, energy, capital, and time (Ibid).

People: In each era, on the supply side. Competent Europeans universities, institutions and Agencies have been involved in either inventing, innovating, financing or promoting ICTs (CHM, 2019). On the demand side, business entities have been eager to adopt ICTs. Just as government services, they found that ICT supported their data processing and communication needs. Without their enthusiasm to adopt ICTs, maybe it would have been difficult for government agencies to see the need to adopt it later. Furthermore, on the demand side, European citizens have been open to the telegraph, the telephone and subsequent network technologies that came thereafter. Hence, in principle Europe has always had people either consciously or unconsciously operating the chain of events that results in the adoption of e-government. Finally, what has also worked for Europe is the collaboration within the continent and across the Atlantic, resulting in the diffusion of innovation to Europe.

Information: According to Steve Krar, technology had the need for information in order to solve problems (Krar). The source of knowledge is information. One of the values of ICT is its ability to transform information to either implicit or explicit knowledge –depending on the type of ICT. As ICTs evolves, so did the type of information they could produce. The information can be either internalized or combined and stored in the ICT as a new form of knowledge. In Europe, just as in any other region, they had information they wanted to digitalize, store or distribute. Therefore, this one resource exists everywhere.

Materials, Tools and Machines: This means natural resources. However, in the context of this chapter, it means the material, tools and machines needed to create and use ICTs. Most ICTs used in Europe are developed in Europe. In the late 19th century American inventors, such Alexander Bell (AT&T), Hollerith (IBM), Thomas Edison all ventured to extend their ICT inventions to Europe. However, they faced competitors in Lars Ericsson (Ericsson- Sweden), C.F. Tietgen (TDC Denmark), William Cooke and John Ricardo (British telegraph company-now British telecoms to mention a few. Hence, European inventors and businesspersons produced most ICTs in Europe. It is also most likely that most of the re-

sources used in the development of these ICTs originate from Europe. In 2018, the ICT sector in Europe contributed 3.7% of the regions GDP (Eurostat, 2018).

Energy: This implies the source of energy needed by the technology. The development of energy to power the ICTs in Europe required ingenuity. When the telephone and telegraph was invented, the use of electricity was not wide spread either in the US and Europe (Kim, 2007) (Iskhakov, 2013). Charles Wheatstone, one of the British inventors of the telegraph powered the telegraph with a battery-powered magnet (McNiel, 1990). Energy is a resource the Americans also harnessed to power their early ICT. Alexander Bell powered his first telephone with sound energy (Borthe, 2021). However, much as the telephone developed, the telephone exchange (switching center) was powered by 48-volt direct current (ibid). However, this was replaced by electricity over time and electricity has been the main source of powering current exchanges. The same ingenuity was required for the early EDPs until there was electricity. Hollerith's initial data processor was electromechanical (Hollerith, 1894). It suited that day and time. Hence, there was an energy resource to power the ICT. These resources were available in Europe as well. However, to from the early 20th century European countries began their electrification projects. Although hampered by the world wars, electricity has been made available to European households. This implies that European citizens can power their ICTs.

Capital: the investment of European countries into inventions in not a new idea. For example, the British government to produce his invention financed Charles Babbel, at some point (Tuck, 2021). A mention earlier in this section there were direct public funding initiatives into the development of the eDP and later telecom networks. Therefore, that culture exists. However, in the EU capital has been raised by direct private investment. At different times, in many centuries private enterprises have been involved in financing, expanding and developing telecom networks. Aside private funding, very recent approach to financing the e-government initiatives has been through Public-Private Partnerships. In the delivery of G2G services, government agencies acquired their own computers, LAN networks and WAN networks (mostly Wi-Fi these days) instead of leasing them as it were back in the days. It is not uncommon to visit a government office in Europe and not have access to Wi-Fi. To enable the delivery of G2C and G2B, in recent times, policy initiatives that enabled the affordability of access devices have been implemented. Hence, countries in the Europe have device different means of raising the capital they need, either directly or indirectly. The approach varies from country to country.

Time: Before the 1980s ICT were not adopted to boost economic competitiveness. However, after the 1980's but mostly in the 1990, competitiveness has been the key word. The more efficient the KM initiatives, the more the country will transform into an information and knowledge society. This would open up opportunities for innovation, economic growth and in recent time's inflow in foreign direct investment. Hence, countries in Europe are taking the first movers advantage seriously.

Section Summary

In this section, pointers have been made to evidence of value, congruence, cooperativeness and resource availability. Value has been the most critical in this case. Furthermore, explanation have been made as to how the desire to facilitate KM in the delivery of government services led to e-government service delivery. This implies that the absence of this value would have resulted in government agencies either ignoring or adopting the ICTs to deliver government services.

DISCUSSION

Based on the analysis made in the previous section, one could say that e-government development in Europe was driven by the quest to manage knowledge digitally. ICT were identified as tools that could support this endeavor. As technology improved, the potential for developing an information society emerged and the journey towards the digitization of e-government services has been ongoing since then.

Nevertheless, if one is to look back at countries with either low or medium EGDI, one could point to the fact that government services that are not digitized in these countries are those in which the value for KM is not eminent to decision makers. Hence, in such countries, such government agents will only identify such value if external agents either inspire them. Such agents could be international development agencies, donor agencies etc. (see (Salifu-Siddi & Williams, 2016) (Hafkin, 2009)). Actually, the same could be said on some countries in Europe that are yet to digitize some of their e-government services.

There are lessons from Europe that could be of value to countries with low to medium EDGI in other in the Americas, Asia, Africa and Oceania where digital divides still exist. An interesting finding in the case of Europe is their proactivity in updating their ICTs. This is evident in the various innovative policy and technological initiatives they adopted to ensure that most government agencies, citizens and businesses have access to various ICTs over the years. The desire to facilitate KM (OECD, 2017), in the delivery of e-government services led to their interest in the penetration of ICTs in rural areas. This lesson from European countries seems to be the missing link in the discussion on digital divide and e-government service delivery as well.

However, one of the reasons government agencies are unable to identify the value of KM is because they are not aware of the various KM processes in their agencies. If we use Europe as an example, as early adopters of early ICTs, they had centuries to see how various computer technologies enable KM in their e-government processes. This is also the case of countries in North America; Argentina and Brazil in South America (Williams, 2015); and Japan, and China in Asia (Yang, 1996) (Knuesel, 2017), to mention a few. Hence as technologies evolved, the were able to identify value which helped them to develop congruence and cooperativeness. This resulted in their harnessing of the resources to develop small sectors per time. This is a sharp contrast to countries mobilized to adopt e-government services in the 1990s and 2000. Then the technology had matured, but the size and processes of governments had expanded from the 1950's, 1960's and 1970's when most nations became independent. Nevertheless, in the past 15 to 20 years, different countries have been able to identify the value of KM in some areas of government operations. Such services are mostly informational and in recent times increasingly transactional services (UN DESA, 2020). However, there are still challenges in digitizing more services.

CONCLUSION

Hence, there is the need to audit the KM processes within government agencies to identify the services that require ICT; as well as identify the ICT that will enhance KM the most. This should be followed by consultations and brainstorming with the relevant stakeholders on the needed technology, its cost, how to finance it and the citizens/businesses that require access for such a service. If it is a G2G service, then find out the nature of work, the type of knowledge to be created, shared, stored or transferred; the digital technology that will support such process and how to finance it. Financing does not always require direct investment (see (Williams, 2015)). Using this approach will assist in the development of

policies that cater to the government needs as well as deliver value both to the government agencies, the citizens and businesses.

REFERENCES

Agar, J. (2003). *The Government Machine: A Revolutionary History of the Computer*. MIT Press.

Baquiast, J. P., & van de Donk, W. (1989). Some policy Issues regarding Informattization in Pulic administation: French experiences. *International Review of Administrative Sciences*, *55*, 559–568.

Borthe, D. E. (2021). *Telephone.* Britannica. Retrieved from https://www.britannica.com/technology/telephone

Calhoun, E. S. (1956). New computer developments around the world. *Proceedings in AIEE-IRE '56 (Eastern) Papers and discussions presented at the December 10-12, 1956*, *eastern joint computer conference: New developments in computers* (pp. 5 - 9). New York: ACM.

CHM. (2019). *Computer History Museum*. Retrieved from Timeline of Computer History: https://www.computerhistory.org/timeline/1951/

Cortada, J. W. (2018). Change and Continuity at IBM: Key Themes in Histories of IBM. *Business Review History*, *92*(1), 117–148.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. *Management Information Systems Quarterly*, *11*(3), 355–366.

Der Spiegel. (1983). *Bildschirmtext: Computing for all?* Retrieved from Der Spiegel: https://www.spiegel.de/wirtschaft/bildschirmtext-computerwissen-fuer-alle-a-6dd9ba4b-0002-0001-0000-000014019866?context=issue

Derdak, T., & Hast, A. (1992). *International Directory of Compnay Histories*. St James Press.

DESI. (2020). *Digital Economy and Society Index (DESI) 2020: Use of internet services.* Brussels: EU Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi

Dickey, J. (2017). *The History of WAN*. Retrieved from Oracle Communications Blog: https://blogs.oracle.com/oracle-communications/post/the-history-of-wan

DIGST. (2016). *Danish digital Strategy 2016-2020*. Retrieved from DIGST: https://en.digst.dk/policy-and-strategy/digital-strategy

Estonian Informatics center. (1998, May 13). *Principles of Estonian Information policy*. Retrieved from https://issuu.com/e-governanceacademy/docs/eesti_infopoliitika_p__hialused

EU. (2020). *Digital Economy and Society Index Report 2020 - Digital Public Services*. Retrieved from Digital Public Services: https://ec.europa.eu/digital-single-market/en/digital-public-services

Eurostat. (2018). *ICT sector - value added, employment and R&D.* EU Commission. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_sector_-_value_added,_employment_and_R%26D#:~:text=The%20total%20value%20added%20of,of%20the%20EU%20Member%20States

Eurostat. (2020, April). *Education Attainment Statistics*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php/Educational_attainment_statistics#Main_statistical_findings

Fang, Z. (2002). E-Government in Digital Era: Concepts, Practice, and Development. *International Journal of the Computer. The Internet and Management*, *10*(2), 1–22.

Fugener, A. (2019). *Collaboration and Delegation Between Humans and AI: An Experimental Investigation of the Future of Work* (No. ERS-2019-003-LIS). ERIM report series research in management Erasmus Research Institute of Management. Retrieved from https://repub.eur.nl/pub/115830/

Gyamfi, A., & Williams, I. (2018). *Evaluating Media Richness in Organizational Learning*. IGI Global.

Hafkin, N. J. (2009). *E-government in Africa: Progress made and challenges ahead*. Conference on Electronic/Mobile Government in Africa: Building capacity in Knowledge management through partnership, Addis Ababa, Ethiopia. Retrieved from http://unpan1.un.org/intradoc/groups/public/documents/un/unpan033526.pdf

Hollerith, H. (1894). The Electrical Tabulating Machine. *Journal of the Royal Statistical Society*, *57*(4), 678–689.

IBM. (2017). *IBM in Europe: A Long History and Deep Presence.* Retrieved from IBM.com: https://www.ibm.com/blogs/policy/wp-content/uploads/2017/03/IBM-in-Europe-2017_Brochure_pdf-FINAL.pdf

IHS, Omdia, & PointTopic. (2020). *Broadband coverage in Europe 2019: Mapping progress towards the coverage objectives of the Digital Agenda.* Luxembourg: EU.

Iskhakov, R. (2013). *History of Electricity in the United States.* Stanford. Retrieved from http://large.stanford.edu/courses/2013/ph240/iskhakov2/

IT-Kooperation zwischen Bund und Landern. (1998, June). [IT cooperation between the federal government and the federal states]. Retrieved from https://archive.md/20121215162234/http://reference.e-government.gv.at/uploads/media/IT-Koop--Bund-Laender.html

ITU. (2020). *Measuring digital development: Facts and figures 2020.* ITU. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf

Kim, K. (2007). *History of Electrification in Europe.* Retrieved from https://www.zum.de/whkmla/sp/0809/kyungmook/km2.html

Kliček, B., & Vukovac, D. (2007). Information Society and eGovernment Developments in Croatia. *Informatica (Vilnius)*, *31*, 367–372.

Knuesel, A. (2017). British Diplomacy and the Telegraph in Nineteenth-Century China. *Diplomacy and Statecraft, 18*(2007), 517-537.

Kokai, T. R. (1974). Measures towards EDP- Appropriate legislation. *Jurimetrics Journal*, *15*(37), 37–42.

Kraemer, K. L., & King, J. L. (1977). *Computers and Local Government*. Praeger.

Krar, S. (n.d.). *Technology Resources - Manufacturing Automation.* https://www.automationmag.com/images/stories/LWTech-files/83%20Technology%20Resources.pdf: Automation mag.

Latour, B. (1996). On Actor-Network Theory: A few clarifications. *Soziale Welt*, *47*(4), 369–381.

Leslie, D. (2019). Raging robots, hapless humans: The AI dystopia. *Nature*, *574*(7776), 32.

McNiel, I. (1990). *An Encyclopedia of the History of Technology*. Routledge.

Minges, M. (1999). *Universal Service/Access Indicators - measuring access to telecommunications: Universal service and access indicators.* ITU. Retrieved from https://www.itu.int/ITU-D/ict/WTIM99/PDF/PartVIFinalR_e.pdf

MKM. (2020). *Real-Time Economy Vision (2020 -2027).* Tallin: MKM Estonia. Retrieved from https://www.mkm.ee/sites/default/files/real-time_economy_vision_2020-2027.pdf

Nixon, P. G., & Koutrakou, V. N. (2007). *E-government in Europe*. Routledge, Taylor and Francis Group.

Nonaka, I., & von Krogh, G. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science*, *20*(3), 635–652.

Nora, S., & Minc, A. (1978). *L'informatisation de la société.* Paris: La Documentation Francaise: Présidence de la République. Retrieved from https://www.vie-publique.fr/rapport/34772-linformatisation-de-la-societe

OECD. (2017). *Digital Government Review of Norway: Boosting the digital transformation of the publlic sector.* OECD. Retrieved from https://www.oecd.org/gov/digital-government/digital-government-review-norway-recommendations.pdf

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *JMR, Journal of Marketing Research*, *17*(4), 460.

Plantera, F. (2017). *Artificial Intelligence is the next step for e-governance in Estonia", State adviser reveals*. E-Estonia. Retrieved from https://e-estonia.com/artificial-intelligence-is-the-next-step-for-e-governance-state-adviser-reveals/

Price, D. G., & Mulvihill, D. E. (1965). The Present and Future Use of Computers in State Government. *Public Administration Review*, *25*(2), 142–150.

Regeringskansliet. (2014). *Broadband Strategy for Sweden.* Stockholm: Ministry of Enterprise Energy and Communications Sweden.

Reinermann, H. (1997). The utilisation of information technology in German public administration. *International Review of Administrative Sciences*, *2*, 207–223.

Roser, M., & Ortiz-Ospina, E. (2018). *Literacy*. Retrieved from Our world in data: https://ourworldindata.org/literacy

Salifu-Siddi, F., & Williams, I. (2016). The challenges of replicating Western E-government Structures in Ghana. *The 1st Africa Regional ITS Conference in Accra, Ghana, March 10-11. 2016.* Accra: ITS.

Schofield, H. (2012, June). *Minitel: The rise and fall of the France-wide web*. Retrieved from BBC News Magazine (Paris): https://www.bbc.com/news/magazine-18610692

Schumacher, B. G. (1967). *An EDP approach in public administration*. University of Oklahoma.

Smelser, N. (1963). *Theory of collective behaviour.* Free Press of Glencoe.

SYZEFXIS Project. (2004). *Information Society*. Retrieved from SYZEFXIS: https://web.archive.org/web/20090414225658/http://www.infosoc.gr/infosoc/en-UK/epktp/Parakoloythish_ajiologish/EthsiaEktheshEkteleshs/Ethsiesektheseis/kales_praktikes/syzeuksis.htm

Tomas, S. (2014). *Tomas Söderblom's telephone museum*. Retrieved from History of Telegrafverket: http://www.soderblom.se/eng_tvhistoria.html

Tuck, S. (2021). *Charles Babbage.* Wichita State University. Retrieved from https://www.math.wichita.edu/history/Men/babbage.html

UN DESA. (2012). *UN E-government Survey.* New York: UN. Retrieved from https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2012-Survey/Complete-Survey.pdf

UN DESA. (2020). *United Nations E-Government Survey 2020: Digital Government in the decade of Action for Sustainable development*. United Nations.

Wagner, C. (2003). Knowledge Management in E-Government. *Americas Conference on Information Systems 2003, Proceedings.* Retrieved from https://aisel.aisnet.org/amcis2003/105

Williams, I. (2015). *Analysis of Public Private Interplay Frameworks in the Development of Rural Telecommunications Infrastructure: A Multiple-Case Study*. Aalborg University Press.

Williams, I. (2020). E-Government, Yesterday, Today and in the Future. In A. Sieben (Ed.), *E-Government: Perspectives, Challenges and Opportunities*. Nova Publishers.

Williams, I. (2020). The Robots are here. In I. Williams (Ed.), *Contemporary applications of Actor Network Theory* (pp. 51–75). Palgrave Macmillan.

Williams, I. (2021). Mobilization - Decision Theory. In *M. Khosrow-Pour, Encyclopedia of Information Science and Technology* (5th ed., vols. 1–3). IGI Global.

Williams, I., Falch, M., & Tadayoni, R. (2018). Internationalization of e-Government Services. In *11th CMI International Conference: Prospects and Challenges Towards Developing a Digital Economy within the EU* (pp. 19-31). Copenhagen Denmark: IEEE.

Williams, I., Falch, M., & Tadayoni, R. (2020). The proposed Baltic Sea Region Cross-border Business Registration Service. In *2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges.* IEEE.

Yang, D. (1996). *The Technology of Japanese Imperialism: Telecommunications and Empire-Building, 1895-1945*. Harvard University.

Compilation of References

Aaron, S. W., Willis, F., Nwaze, E., Dieng, B., Sipilanyambe, N., Daniels, D., . . . Rymane, T. K. (2017, October 21). *Vaccine wastage in Nigeria: An assessment of wastage rates and related vaccinator knowledge, attitudes and practices*. doi:10.1016/j.vaccine.2017.09.082

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), i–xxxii. doi:10.17705/1jais.00423

ABPMP. (2013). *Guia para o Gerenciamento de Processos de Negócio - Corpo Comum de Conhecimento* (1st ed.). LECOM Tecnologia.

Abubakar, I., Khalid, S. N., Mustafa, M. W., Shareef, H., & Mustapha, M. (2017). Application of load monitoring in appliances' energy management—A review. *Renewable & Sustainable Energy Reviews*, *67*, 235–245. doi:10.1016/j.rser.2016.09.064

Adams, B., Breazeal, C., Brooks, R., & Scassell, B. (2000). Humanoid robots: A new kind of tool. *IEEE Intelligent Systems Magazine*, *15*(4), 25–31. doi:10.1109/5254.867909

Agar, J. (2003). *The Government Machine: A Revolutionary History of the Computer*. MIT Press.

Aggogeri, F., Mikolajczyk, T., & O'Kane, J. (2019). Robotics for rehabilitation of hand movement in stroke survivors. *Advances in Mechanical Engineering*, *11*(4). Advance online publication. doi:10.1177/1687814019841921

Ahanhanzo, Y. G., Ouedraogo, L. T., Kpozèhouen, A., Coppieters, Y., Makoutodé, M., & Wilmet-Dramaix, M. (2015). Data quality assessment in the routine health information system: an application of the Lot Quality Assurance Sampling in Benin. *Health Policy and Planning, 30*(7), 837-843.

Ahmed, Y. A., Ahmad, M. N., Ahmad, N., & Zakaria, N. H. (2019). Social media for knowledge-sharing: A systematic literature review. *Telematics and Informatics*, *37*, 72–112. doi:10.1016/j.tele.2018.01.015

Akkiraju, R., Farrell, J., Miller, J., Nagarajan, M., Schmidt, M., Sheth, A., & Verma, K. (2005). *Web Service Semantics - WSDL-S*. A joint UGA-IBM Technical Note, version 1.0. http://lsdis.cs.uga.edu/projects/METEOR-S/WSDL-S

Alami, R., Warnier, M., Guitton, J., Lemaignan, S., & Sisbot, E. A. (2011). When the robot considers the human. *Proceedings of the 15th International Symposium on Robotics Research*.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *Management Information Systems Quarterly*, *25*(1), 107–136. doi:10.2307/3250961

Aldahdouh, T. Z., Nokelainen, P., & Korhonen, V. (2020). Technology and Social Media Usage in Higher Education: The Influence of Individual Innovativeness. *SAGE Open*, *10*(1), 215824401989944. doi:10.1177/2158244019899441

Alexander, R. (2020). *5 Reasons Knowledge Management Is Crucial in the Healthcare Industry.* Retrieved from https://bloomfire.com/blog/3-reasons-knowledge-management-critical-healthcare-industry/: https://bloomfire.com

Alili, S., Warnier, M., Ali, M., & Alami, R. (2009). Planning and plan-execution for human-robot cooperative task achievement. *19th International Conference onAutomated Planning and Scheduling*.

Alipour, J., & Ahmadi, M. (2017). Dimensions and assessment methods of data quality in health information systems. *Acta Medica Mediterranea*, *33*(2), 313–320.

Alliance for Affordable Internet. (2020). *Affordability Report 2019*. Alliance for Affordable Internet.

Alliance for Affordable Internet. (2020). *From luxury to lifeline: Reducing the cost of mobile devices to reach universal internet access*. Web Foundation.

Al-Natsheh, H. T., Martinet, L., Muhlenbach, F., Rico, F., & Zighed, D. A. (2018). Metadata enrichment of multi-disciplinary digital library: a semantic-based approach. In *International Conference on Theory and Practice of Digital Libraries* (pp. 32–43). 10.1007/978-3-030-00066-0_3

Alzaidy, R., Caragea, C., & Giles, C. L. (2019). Bi-lstm-crf sequence labeling for keyphrase extraction from scholarly documents. In *Proceedings of the The World Wide Web Conference* (pp. 2551–2557). 10.1145/3308558.3313642

Anckar, B., & Walden, P. (2001). Self-booking of high- and low-complexity travel products: Exploratory findings. *Information Technology & Tourism*, *4*(3-4), 151–165. doi:10.3727/109830501108750958

Antoniou, G., & Harmelen, F. V. (2008). *A Semantic Web Primer*. The MIT Press.

Ashley, K. (1987). *Modelling Legal Argument: Reasoning with Cases and Hypothetical* (PhD thesis). Department of Computer and Information Science, University of Massachusetts, Amherst, MA.

Audit Commission. (2013). *The Audit Commission for Local Authorities and the National Health Service in England Annual report and accounts 2012/13*. NHIS.

Aurum, A., Daneshgar, F., & Ward, J. (2008, May). Investigating Knowledge Management practices in software development organisations An Australian experience. *Information and Software Technology*, *50*(6), 511–533. doi:10.1016/j.infsof.2007.05.005

Aurum, A., Jeffery, R., Wohlin, C., & Handzic, M. (2003). *Managing Software Engineering Knowledge*. Springer.

Baader, F., Horrocks, I., & Sattler, U. (2008). Handbook of Knowledge Representation (Vol. 3). Elsevier.

Baader, F., & Nutt, W. (2003). *Basic Description Logics, The description logic handbook*. Cambridge University Press.

Babu, V. S., Kumar, U. A., Priyadharshini, R., Premkumar, K., & Nithin, S. (2016). An intelligent controller for smart home. In *Proceedings of international conference on advances in computing, communications and informatics (ICACCI)* (pp. 2654–2657). IEEE.

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., ... Yadav, V. (2010). Agrotags–a tagging scheme for agricultural digital objects. In *Research Conference on Metadata and Semantic Research* (pp. 36–45). 10.1007/978-3-642-16552-8_4

Baliga, J., Ayre, R., Hinton, K., & Tucker, R. (2011). Energy consumption in wired and wireless access networks. *IEEE Communications Magazine*, *49*(6), 70–77. doi:10.1109/MCOM.2011.5783987

Banger, G. (2018). *Endüstri 4.0 ve Akıllı İşletme, Dorlion Yayınları, 2*. Baskı.

Baquiast, J. P., & van de Donk, W. (1989). Some policy Issues regarding Informattization in Pulic administation: French experiences. *International Review of Administrative Sciences*, *55*, 559–568.

Barnes, M. J., Jentsch, F. G., Chen, J. Y., & Redden, E. S. (2011). Designing Effective Soldier-Robot Teams in Complex Environments: Training, Interfaces, and Individual Differences. *International Conference on Engineering Psychology and Cognitive Ergonomics, 6781*. 10.1007/978-3-642-21741-8_51

Barros, V. F. D. A., Ramos, I., & Perez, G. (2015). Information systems and organizational memory: A literature review the last 20 years. *Journal of Information Systems and Technology Management*, *12*(1), 1–5. doi:10.4301/S1807-17752015000100003

Basaldella, M., Antolli, E., Serra, G., & Tasso, C. (2018). Bidirectional lstm recurrent neural network for keyphrase extraction. In *Italian Research Conference on Digital Libraries* (pp. 180–187). 10.1007/978-3-319-73165-0_18

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). Experience Factory. In J. J. Marciniack (Ed.), *Encyclopedia of Software Engineering*. John Wiley and Sons.

Basili, V., Costa, P., Lindvall, M., Mendonca, M., & Seaman, C. (2001). An Experience Management System for a Software Engineering Research Organisation. *Proceedings of 26th Annual NASA Goddard Software Engineering Workshop*, 29-35.

Beetz, M., Mösenlechner, L., & Tenorth, M. (2010). CRAM — A Cognitive Robot Abstract Machine for everyday manipulation in human environments. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 10.1109/IROS.2010.5650146

Begum, M., Serna, R. W., & Yanco, H. A. (2016). Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review. *International Journal of Social Robotics*, *8*(2), 157–181. doi:10.100712369-016-0346-y

Bellaachia, A., & Al-Dhelaan, M. (2014). Hg-rank: A hypergraph-based keyphrase extraction for short documents in dynamic genre. In # msm (pp. 42–49). Academic Press.

Benbasat, I., DeSanctis, G., & Nault, R. (1993). Empirical research in managerial support systems: a review and assessment. In Ž. C. W. Holsapple & B. Whinston (Eds.), *Recent Developments in Decision Support Systems* (pp. 383–437). Springer-Verlag. doi:10.1007/978-3-662-02952-7_16

Benedetto, L., Cremonesi, P., & Parenti, M. (2019). A virtual teaching assistant for personalized learning. *CEUR Workshop Proceedings*, *2482*(June).

Bergmann, R., & Schaaf, M. (2003). Structural Case-Based Reasoning and Ontology-Based Knowledge Management: A Perfect Match? *Journal of Universal Computer Science, UCS*, *9*(7), 608–626.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, (May), 34–43.

Bhatia, K. A. (2012). *The business of travel agency and tour operations management*. Sterling Publishers.

Bianchini, D., Antonellis, V. A., Melchiori, M., & Salvi, D. (2006). Semantic Enriched Service Discovery. *International Conference in Data Engineering Workshop*, 38.

Biondi, F., Alvarez, I., & Jeong, K.-A. (2019). Human–Vehicle Cooperation in Automated Driving: A Multidisciplinary Review and Appraisal. *International Journal of Human-Computer Interaction*, *35*(11), 932–946. doi:10.1080/10447318.2018.1561792

Blanco-Novoa, Ó., Fernández-Caramés, T. M., Fraga-Lamas, P., & Castedo, L. (2017). An electricity price-aware open-source smart socket for the internet of energy. *Sensors (Basel)*, *17*(3), 643. doi:10.339017030643 PMID:28335568

Blanning, R. W., & King, D. R. (1989). Decision support and knowledge-based systems. *Journal of Management Information Systems*, *6*(3), 3–6. doi:10.1080/07421222.1989.11517860

Blau, L., Bonissone, P. P., & Ayub, S. (1991). Planning with dynamic cases. In *Proceedings of Case-Based Reasoning Workshop*. Morgan Kaufmann.

Blau, L., Bonissone, P., & Ayub, S. (1991). Planning with dynamic cases. In *Proceedings of Case-Based Reasoning Workshop*. Morgan Kaufmann.

Bonarini, A. & Maniezzo, V. (1991). Integrating expert systems and decision-support systems: principles and practice. *Knowledge-Based Systems*, 172–176.

Bonissone, P. P., Aragones, J., & Stillman, J. (1990). PRIMO: a tool for reasoning with incomplete and uncertain information. *Proceedings of the Third International Conference on Information Processing and Management of Uncertainty in Knowledge-Based System, IPMU 90*, 325–327.

Bonissone, P. P., Blau, L., & Ayub, S. (1990). Leveraging the integration of approximate reasoning systems. *Proceedings of 1990 AAAI Spring Symposium in Case-Based Reasoning*,11–6.

Borthe, D. E. (2021). *Telephone.* Britannica. Retrieved from https://www.britannica.com/technology/telephone

Boudin, F. (2016, December). pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System demonstrations* (pp. 69–73). Academic Press.

Bougouin, A., Boudin, F., & Daille, B. (2013). *Topicrank: Graph-based topic ranking for keyphrase extraction*. Academic Press.

Branting, L. (1991). *Integrating Rules and Precedents for Classification and Explanation: Automating Legal Analysis* (PhD thesis). Department of Computer Science, University of Texas, Austin, TX.

Brealey, R., & Myers, S. (1996). *Principles of Corporate Finance*. McGraw-Hill.

Brooks, A. G., & Breazeal, C. (2006). Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. *1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. 10.1145/1121241.1121292

Brous, P., Janssen, M., & Herder, P. (2020). The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations. *International Journal of Information Management*, *51*, 101952. doi:10.1016/j.ijinfomgt.2019.05.008

Brown-Acquaye, W., Arthur, J. K., & Forgor, L. (2019). Developing A Framework for A Tactile Internet Enabled Robot Assisted Real-Time Interactive Medical System. *International Conference on Communications, Signal Processing and Networks (ICCSPN)*.

Brown, D. W., Warrener, L., Scobie, H. M., Donadel, M., Waku-Kouomou, D., Mulders, M. N., & Rota, P. (2020). Rapid diagnostic tests to address challenges for global measles surveillance. *Current Opinion in Virology*, *41*, 77–84. doi:10.1016/j.coviro.2020.05.007 PMID:32615510

Brownlee. (2020). *Deep Learning for Computer Vision*. Academic Press.

Brownson, R. C., Eyler, A. A., Harris, J. K., Moore, J. B., & Tabak, R. G. (2018). Getting the Word Out: New Approaches for Disseminating Public Health Science. *Journal of Public Health Management and Practice*, *24*(2), 102–111. doi:10.1097/PHH.0000000000000673 PMID:28885319

Buhalis, D., & Amaranggana, A. (2015). Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015*. Springer. doi:10.1007/978-3-319-14343-9_28

Burke, J. L., & Murphy, R. R. (2011). The safe human–robot ratio. In *Human–robot interaction in future* (pp. 31–49). Ashgate.

Cai, M. Zhang, W. Y., Zhang, K. & Li, S. T. (2010). SWMRD: A Semantic Web-based manufacturing resource discovery system for cross-enterprise collaboration. *International Journal of Production Research, 48*(12), 3445-3460.

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*(0), 2. doi:10.5334/dsj-2015-002

Caivano, D., Fogli, D., Lanzilotti, R., Piccinno, A., & Cassano, F. (2018). Supporting end users to control their smart home: Design implications from a literature review and an empirical investigation. *Journal of Systems and Software*, *144*, 295–313. doi:10.1016/j.jss.2018.06.035

Calhoun, E. S. (1956). New computer developments around the world. *Proceedings in AIEE-IRE '56 (Eastern) Papers and discussions presented at the December 10-12, 1956*, *eastern joint computer conference: New developments in computers* (pp. 5 - 9). New York: ACM.

Callahan, S., & Glass, R. (2015). The big data-driven business - how to use big data to win customers, Beat competitors, and boost profits. John Wiley & Sons.

Camilleri, M. A. (2018). The Tourism Industry: An Overview. In *Travel Marketing, Tourism Economics and the Airline Product* (pp. 3–27). Springer Nature. doi:10.1007/978-3-319-49849-2_1

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., & Jatowt, A. (2020). Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, *509*, 257–289. doi:10.1016/j.ins.2019.09.013

Car, T., Stifanich, L. P., Simunic, M. (2019). Internet OF Things (IoT) in tourism and hospitality: Opportunities and Challenges. *ToSEE – Tourism in Southern and Eastern Europe, 5*, 163–175.

Cardoso, J. (2006). Discovering Semantic Web Services with and without a Common Ontology Commitment. *IEEE Service Computing Workshop*, 183-190.

Cardoso, J., & Sheth, A. (2003). Semantic e-Workflow Composition. *Journal of Intelligent Information Systems*, *21*(3), 191–225.

Carvalho, N., & Gomes, I. (2017). Knowledge Sharing between Enterprises of the Same Group. *International Journal of Knowledge Management*, *13*(1), 34–52. doi:10.4018/IJKM.2017010103

Cassar, G., Barnaghi, P., & Moessner, K. (2014). Probabilistic matchmaking methods for automated service discovery. *IEEE Transactions on Services Computing*, *7*(4), 654–666.

Cavlek, N. (2006). Travel and tourism intermediaries. In L. Dwyer & P. Forsyth (Eds.), *International handbook on the economics of tourism*. Edward Elgar Publishing. doi:10.4337/9781847201638.00014

Cern, S. (2002). *Intelligent tutoring systems*. Springer Verlag Publication.

Chan, A. K. M., Nickson, C. P., Rudolph, J. W., Lee, A., & Joynt, G. M. (2020). Social media for rapid knowledge dissemination: Early experience from the COVID -19 pandemic. *Anaesthesia*, *75*(12), 1579–1582. doi:10.1111/anae.15057 PMID:32227594

Chao, S. H., Jiang, J., Hsu, C. H., Chiang, Y. T., Ng, E., & Fang, W. T. (2020). Technology-Enhanced Learning for Graduate Students: Exploring the Correlation of Media Richness and Creativity of Computer-Mediated Communication and Face-to-Face Communication. *Applied Sciences (Basel, Switzerland)*, *10*(5), 1602. doi:10.3390/app10051602

Chen, J. V., McBush Hiele, T., Kryszak, A., & Ross, W. H. (2020). Predicting Intention to Participate in Socially Responsible Collective Action in Social Networking Website Groups. *Journal of the Association for Information Systems*, 342–363. doi:10.17705/1jais.00604

Chiavenato, I. (2005). *Administração de Materiais - Uma Abordagem Introdutória* (3rd ed.). Elsevier Editora, Lda.

CHM. (2019). *Computer History Museum*. Retrieved from Timeline of Computer History: https://www.computerhistory.org/timeline/1951/

Cho, H. J., Feldman, L. S., Keller, S., Hoffman, A., Pahwa, A. K., & Krouss, M. (2020). Choosing Wisely in the COVID-19 Era: Preventing Harm to Healthcare Workers. *Journal of Hospital Medicine*, *15*(6), 360–362. doi:10.12788/3457 PMID:32490804

Choi, Y., Kim, D., & Lee, J.-H. (2020). Neural attention model with keyword memory for abstractive document summarization. *Concurrency and Computation*, *32*(18), e5433. doi:10.1002/cpe.5433

Christy, A., Gandhi, G. M., & Vaithyasubramanian, S. (2019). Clustering of text documents with keyword weighting function. *International Journal of Intelligent Enterprise*, *6*(1), 19–31. doi:10.1504/IJIE.2019.100029

Ciolacu, M., Tehrani, A. F., Binder, L., & Svasta, P. M. (2019). Education 4.0 - Artificial Intelligence Assisted Higher Education: Early recognition System with Machine Learning to support Students' Success. *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging, SIITME 2018 - Proceedings,* 23–30. 10.1109/SIITME.2018.8599203

CMMI. (2010). *CMMI for development, version 1.3*. https://resources.sei.cmu.edu/asset_files/TechnicalReport/2010_005_001_15287.pdf

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in education needs interpretable machine learning: Lessons from open learner modelling. *ArXiv, Whi*.

Confirmatory Factor Analysis — A Case study. (2019, May). https://arxiv.org/abs/1905.05598

Constantin, F., & Constantin, F. (2017). Collaborative Robots and Knowledge Management - A Short Review. *Acta Uiversitatis Cibiniensis Technical Series*, *49*, 143–147.

Cooper, C., & Lewis, J. (2004). Transformation and trends in the tourism industry: implications for distribution channels. In Tourism distribution channels: Practices, issues and transformations. Thomson Learning.

Copacino, W., & Anderson, D. (2003). Connecting with the Bottom Line: A Global Study of Supply Chain Leadership and its Contribution to the High-Performance Business. *Accenture, 1*.

Cortada, J. W. (2018). Change and Continuity at IBM: Key Themes in Histories of IBM. *Business Review History*, *92*(1), 117–148.

Curet, O., & Jackson, M. (1996). Evaluting a case-based learning and reasoning applications: the top management fraud diagnostic tool TMFDT. In *Proceeding of Fourth German Workshop on Case-Based Reasoning — System Development and Evaluation*. Computer Science Department, Humboldt University.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. *Management Information Systems Quarterly*, *11*(3), 355. doi:10.2307/248682

Danesh, S., Sumner, T., & Martin, J. H. (2015). Sgrank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction. In *Proceedings of the 4th Joint Conference on Lexical and Computational Semantics* (pp. 117–126). 10.18653/v1/S15-1013

Dean, J. (2014). Big Data, Data Mining, and Machine Learning - Value Creation for Business Leaders and Practitioners. John Wiley & Sons, Inc.

Delen, D., Zaim, H., Kuzey, C., & Zaim, S. (2013). A comparative analysis of machine learning systems for measuring the impact of knowledge management practices. *Decision Support Systems*, *54*(2), 1150–1160. doi:10.1016/j.dss.2012.10.040

Der Spiegel. (1983). *Bildschirmtext: Computing for all?* Retrieved from Der Spiegel: https://www.spiegel.de/wirtschaft/bildschirmtext-computerwissen-fuer-alle-a-6dd9ba4b-0002-0001-0000-000014019866?context=issue

Derdak, T., & Hast, A. (1992). *International Directory of Compnay Histories*. St James Press.

DESI. (2020). *Digital Economy and Society Index (DESI) 2020: Use of internet services.* Brussels: EU Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi

Diamantini, C., Potena, D., & Storti, E. (2013). A virtual mart for knowledge discovery in databases. *Information Systems Frontiers*, *15*(3), 447–463. doi:10.100710796-012-9399-0

Dickey, J. (2017). *The History of WAN*. Retrieved from Oracle Communications Blog: https://blogs.oracle.com/oracle-communications/post/the-history-of-wan

DIGST. (2016). *Danish digital Strategy 2016-2020*. Retrieved from DIGST: https://en.digst.dk/policy-and-strategy/digital-strategy

Dingsøyr, T., Brede Moe, N., & Nytro, O. (2001). Augmenting Experience reports with Lightweight Postmortem Reviews. *Third International Conference on Product Focused Software Process Improvement, PROFES 2001*.

Domingue, J., Cabral, L., Galizia, S., Tanasescu, V., Gugliotta, A., Norton, B., & Pedrinaci, C. (2008). IRS-III: A Broker-based Approach to Semantic Web Service. *Journal of Web Semantics*, *6*(2), 109–132.

Doycheva, A., Glushkova, T., Ivanova, V., Doukovska, L., & Stoyanov, S. (2020). A Multi-agent Environment Acting as a Personal Tourist Guide. In O. Castillo, P. Melin, & J. Kacprzyk (Eds.), *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications. Studies in Computational Intelligence* (Vol. 862). Springer.

Dutta, S., & Bonissone, P. P. (1993). Integrating case- and rule-based reasoning. *International Journal of Approximate Reasoning*, *8*(3), 163–203. doi:10.1016/0888-613X(93)90001-T

Edwin, D., Garcia-Haro, J. M., Jardón, A., & Balaguer, C. (2019). Robotics in Health Care: Perspectives of Robot-Aided Interventions in Clinical Practice for Rehabilitation of Upper Limbs". 9 (13): 2586. *Applied Sciences (Basel, Switzerland)*, *9*(13).

Elorriaga, J., & Fernandez-Castro, I. (2000). Using case-based reasoning in instructional planning. towards a hybrid selfimproving instructional planner. *International Journal of Artificial Intelligence in Education*, 416–449.

El-Shinnawy, M., & Markus, M. L. (1997). The poverty of media richness theory: Explaining people's choice of electronic mail vs. voice mail. *International Journal of Human-Computer Studies*, *46*(4), 443–467. doi:10.1006/ijhc.1996.0099

EMC Education Services. (2015). Data Science and Big Data Analytics - Discovering, Analyzing, Visualizing and Presenting Data. John Wiley & Sons.

Englmeier, K. (2015). Role and Importance of Semantic Search in Big Data Governance. In *Big-Data Analytics and Cloud Computing* (pp. 21–35). Springer International Publishing. doi:10.1007/978-3-319-25313-8_2

Ercan, G., & Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing & Management*, *43*(6), 1705–1714. doi:10.1016/j.ipm.2007.01.015

Estonian Informatics center. (1998, May 13). *Principles of Estonian Information policy*. Retrieved from https://issuu.com/e-governanceacademy/docs/eesti_infopoliitika_p__hialused

EU. (2020). *Digital Economy and Society Index Report 2020 - Digital Public Services*. Retrieved from Digital Public Services: https://ec.europa.eu/digital-single-market/en/digital-public-services

Eurostat. (2018). *ICT sector - value added, employment and R&D*. EU Commission. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_sector_-_value_added,_employment_and_R%26D#:~:text=The%20total%20value%20added%20of,of%20the%20EU%20Member%20States

Eurostat. (2020, April). *Education Attainment Statistics*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php/Educational_attainment_statistics#Main_statistical_findings

Fang, Z. (2002). E-Government in Digital Era: Concepts, Practice, and Development. *International Journal of the Computer. The Internet and Management*, *10*(2), 1–22.

Farrag, T. A., Saleh, A. I., & Ali, H. A. (2013). Toward SWSS discovery: Mapping from WSDL to OWL-S based on ontology search and standardization engine. *IEEE Transactions on Knowledge and Data Engineering*, *25*(3), 1135–1147.

Fasna, M., & Gunatilake, S. (2020). Towards successful strategies to overcome BPR implementation issues: Case of Sri Lanka. *Business Process Management Journal*, *26*(6), 1241–1259. doi:10.1108/BPMJ-03-2019-0087

Fasola, F., & Matarić, M. J. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, *2*(2), 3–32. doi:10.5898/JHRI.2.2.Fasola

Fayyad, U. (2019). *Skill-Set Analysis on LinkedIn: Data Scientists, Engineers, Analysts vs. Job Postings.* Available at: https://www.iadss.org/post/skill-set-analysis-on-linkedin-data-scientists-engineers-analysts-vs-job-postings

Feil-Seifer, D., & Mataric, M. (2011). Socially assistive robotics. *IEEE Robotics & Automation Magazine*, *18*(1), 24–31. doi:10.1109/MRA.2010.940150

Fensel, D., & Bussler, C. (2002). The Web Service Modeling Framework WSMF. *Electronic Commerce Research and Applications*, *1*(2), 113–137.

Ferragina, P., & Scaiella, U. (2010). Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1625–1628). 10.1145/1871437.1871689

Ferrell, W. R., & Sheridan, T. B. (1967). Supervisory control of remote manipulation. *IEEE Spectrum*, *4*(10), 81–88. doi:10.1109/MSPEC.1967.5217126

Finlay, P. N. (1992). IT for competitive Advantage: The place of expert systems. *The Journal of Strategic Information Systems*, *1*(3), 126–133. doi:10.1016/0963-8687(92)90025-R

Firoozeh, N., Nazarenko, A., Alizon, F., & Daille, B. (2020). Keyword extraction: Issues and methods. *Natural Language Engineering*, *26*(3), 259–291. doi:10.1017/S1351324919000457

Florescu, C., & Caragea, C. (2017). Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 1105–1115). 10.18653/v1/P17-1102

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3-4), 143–166. doi:10.1016/S0921-8890(02)00372-X

Fortune Business Insight. (2019). *Artificial Intelligence Market*. https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

Frank, L., Poll, R., Roeglinger, M., & Lea, R. (2020). Design heuristics for customer-centric business processes. *Business Process Management Journal*, *26*(6), 1283–1305. doi:10.1108/BPMJ-06-2019-0257

Freeman, R., & Glazer, K. (2015). Services marketing. In *Introduction to Tourism and Hospitality in BC* (pp. 167-190). BCcampus. Retrieved July 13, 2020, https://opentextbc.ca/introtourism/

Freire, M. (2017). *Especialistas em digital marketing e data analytics são das profissões mais procuradas pelas empresas*. https://www.bit.pt/especialistas-em-digital-marketing-e-data-analytics-sao-as-profissoes-que-as-empresas-mais-procuram/?inf_by=59cba655671db80f3d8b46cb

Frempong, G. (2012). *Evidence for ICT Policy Action - Understanding what is happening in ICT in Ghana: Policy Paper*. Research ICT Africa.

Fugate, B., Sahin, F., & Mentzer, J. T. (2006). Supply Chain Management Coordination Mechanisms. *Journal of Business Logistics*, *27*(2), 129–161.

Fugener, A. (2019). *Collaboration and Delegation Between Humans and AI: An Experimental Investigation of the Future of Work* (No. ERS-2019-003-LIS). ERIM report series research in management Erasmus Research Institute of Management. Retrieved from https://repub.eur.nl/pub/115830/

Ganeshan, R., & Harrison, T. P. (1995). An introduction to supply chain management. *Supply Chain Management, 1*. Available from http://silmaril.smeal.psu.edu/misc/supply_chain_

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, *49*(3), 794–808. doi:10.1016/j.compedu.2005.11.017

Gartner. (2017). *Leading the IoT.* Retrieved November 12, 2020, https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

Gartner. (2019). *IoT press release.* Retrieved December 11, 2020, https://informationmatters.net/internet-of-things-statistics/

GAVI. (2019, March). *Data improvement requirements for all types of Gavi support*. GAVI.

Ghana Health Service. (2015). *Annual Report, 2014.* Author.

Ghazali, S., Sulaiman, N., Zabidi, N., Omar, M., & Alias, R. (2016, October). The impact of knowledge sharing through social media among academia. *AIP Conference Proceedings*, *1782*(1), 030003. doi:10.1063/1.4966060

GHS, WHO, & GAVI. (2018). Expanded Program on Immunization, Ghana field Guide. Accra: WHO.

Gibbs, W. (1994, September). Software's chronic crisis. *Scientific American*, *271*(3), 72–81. doi:10.1038cientificamerican0994-86 PMID:8091191

Giraffa, L. (1997). Multi-ecological: a learning environment using multi-agent architecture. *Multi-Agent System: Theory and Application Proceedings*.

Giunchiglia, F., Shvaiko, P., & Yatskevich, M. (2004). S-Match: an algorithm and an implementation of semantic matching. *Proceedings of 1st European Semantic Web Symposium (ESWS)*, 3053, 61-75.

Golding, R., & Rosenbloom, P. S. (1991). Improving rule-based systems through case-based reasoning. *Proceedings of the Ninth National Conference on Artificial Intelligence AAAI*, *91*, 22–27.

Goldsby, T. J., & Martichenko, R. (2005). *Learn Six Sigma Logistics - Strategic Development to Operational Success* (1st ed.). J. Ross Publishing, Inc.

Gombolay, M. C., Huang, C., & Shah, J. A. (2015). Coordination of human–robot teaming with human task preferences. *AAAI Fall Symposium Series on AI-HRI.* Retrieved September 2020, from https://interactive.mit.edu/sites/default/files/documents/Gombolay_2015_AAAI_FSS_AI-HRI.pdf

Gonzalez, C., Burguillo, J., & Llamas, M. (2006). A qualitative comparison of techniques for student modeling in intelligent tutoring systems. *ASEE/FIE Frontiers in Education Conference*. 10.1109/FIE.2006.322537

Goodrich, M. A., Lin, L., & Morse, B. S. (2012). Using camera-equipped mini-UAVS to support collaborative wilderness search and rescue teams. *International Conference on Collaboration Technologies and Systems (CTS)*. 10.1109/CTS.2012.6261008

Gorelik, A. (2019). The Enterprise Big Data Lake - Delivering the Promise of Big Data and Data Science. O'Reilly Media.

GRA. (2020). *Communication-service-tax.* Available at: https://gra.gov.gh/communication-service-tax/

Grechanik, M., Jones, J. A., Orso, A., & van der Hoek, A. (2010). Bridging Gaps between Developers and Testers in Globally, Distributed Software Development. *Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research*, 149-154.

Gretzel, U. (2011). Intelligent systems in tourism: A social science perspective. *Annals of Tourism Research*, *38*(3), 757–779. doi:10.1016/j.annals.2011.04.014

Grimm, S., Monk, B., & Preist, C. (2006). Matching Semantic Service Descriptions with Local Closed-World Reasoning. *European Semantic Web Conference*, 575-589.

Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The Perils and Promises of Big Data Research in Information Systems. *Journal of the Association for Information Systems*, 268–293. doi:10.17705/1jais.00601

Gruber, T. R. (1993). *A Translation Approach to Portable Ontology Specifications*. Stanford University, Computer Science Department, Knowledge Systems Laboratory, Technical Report KSL 92-71.

Guarino, N., & Giaretta, P. (1995). *Ontologies and Knowledge Base: Towards a Terminological Classification Toward Very Large Knowledge Base: Knowledge Building and Knowledge Sharing*. IOS Press.

Guida, M., Marchesi, P., & Basaglia, G. (1992). Knowledge-based decision support systems for manufacturing decision-making. *Information and Decision Technology*, *18*, 347–361.

Guo, X., Shen, Z., Zhang, Y., & Wu, T. (2019). Review on the application of artificial intelligence in smart homes. *Smart Cities*, *2*(3), 402–420. doi:10.3390martcities2030025

Gupta, S. & Shilpi, S. (2016). *Real-time big data analytics - Design, process, and analyze large sets of complex data in real time.* Packt.

Gyaase, P. O., Anane, E. T., & Armah, I. N. (2015). The use of information and communication technology (ICT) for knowledge management in the second cycle educational institutions in Ghana. *International Journal of Computers and Applications*, *128*(7), 7–13. doi:10.5120/ijca2015906600

Gyamfi, A. (2016). The Impact of Media Richness on the Usage of Web 2.0 Services for Knowledge Transfer. *International Journal of E-Services and Mobile Applications*, *8*(2), 21–37. doi:10.4018/IJESMA.2016040102

Gyamfi, A., & Williams, I. (2018). *Evaluating Media Richness in Organizational Learning*. IGI Global.

Habibi, M., & Popescu-Belis, A. (2013). Diverse keyword extraction from conversations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (vol. 2, pp. 651–657). Academic Press.

Hafkin, N. J. (2009). *E-government in Africa: Progress made and challenges ahead*. Conference on Electronic/Mobile Government in Africa: Building capacity in Knowledge management through partnership, Addis Ababa, Ethiopia. Retrieved from http://unpan1.un.org/intradoc/groups/public/documents/un/unpan033526.pdf

Hagstroem, M. (2015). Big Data Analytics for Inclusive Growth:How Technology Can Help Elevate the Human Condition. In S. Dutta, T. Geiger & B. Lanvin (Eds.), *The Global Information Technology Report 2015: ICTs for Inclusive Growth.* Geneva: World Economic Forum and INSEAD.

Hammer, M. (2010). What is business process management? In J. Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management 1*. Springer. doi:10.1007/978-3-642-00416-2_1

Han, S. G., Lee, S. G., & Jo, G. S. (2005). Case-based tutoring system for procedural problem solving on the www. *Expert Systems With Applications*, 573–582.

Han, D. D. M., Dieck, C. T., & Jung, T. (2019). Augmented Reality Smart Glasses (ARSG) visitor adoption in cultural tourism. *Leisure Studies*, *38*(5), 618–633. doi:10.1080/02614367.2019.1604790

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann.

Harnad, S. (1990). The symbol grounding problem. *Physica D. Nonlinear Phenomena*, *42*(1-3), 335–346. doi:10.1016/0167-2789(90)90087-6

Hasan, K. S., & Ng, V. (2010). Conundrums in unsupervised keyphrase extraction: making sense of the state-of-the-art. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 365– 373). Academic Press.

Haughom, J. (2014). *Knowledge Management in Healthcare: It's More Important Than You Realize.* Retrieved 04 20, 2021, from https://www.healthcatalyst.com/enable-knowledge-management-in-healthcare: https://www.healthcatalyst.com

Hawes, N., Zillich, M., & Wyatt, J. (2007). BALT & CAST: Middleware for cognitive robotics. *16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2007)*, 998–1003. 10.1109/ROMAN.2007.4415228

Heaton, A., Krudwig, K., Lorenson, T., Burgess, C., Cunningham, A., & Steinglass, R. (2017). Doses per vaccine vial container: An understated and underestimated driver of performance that needs more evidence. *Vaccine*, *35*(17), 2272–2278. doi:10.1016/j.vaccine.2016.11.066 PMID:28162822

Hoffer, E. (2005). Machine Learning. *Seoul Studies*, *6*(1), 71–80.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324), 486–488. doi:10.1126cience.aal3856 PMID:28154051

Hollerith, H. (1894). The Electrical Tabulating Machine. *Journal of the Royal Statistical Society*, *57*(4), 678–689.

Hornyak, T. (2020, March 18). What America can Learn from China's Use of Robots and Telemedicin to Combat the Coronavirus. *Tech Drivers - Robotopia*.

Hsu, L., & Chen, Y. J. (2020). Neuromarketing, subliminal advertising, and hotel selection: An EEG study. *Australasian Marketing Journal*, *28*(4), 200–208. doi:10.1016/j.ausmj.2020.04.009

Huang, C. D., Goo, J., Nam, K., & Yoo, C. W. (2017). Smart tourism technologies in travel planning: The role of exploration and exploitation. *Information & Management*, *54*(6), 757–770. doi:10.1016/j.im.2016.11.010

Huang, H., Wang, X., & Wang, H. (2020). Ner-rake: An improved rapid automatic keyword extraction method for scientific literatures based on named entity recognition. *Proceedings of the Association for Information Science and Technology*, *57*(1), e374. doi:10.1002/pra2.374

Huang, W., & Cross, G. (1989). Reasoning about Trademark infringement cases. *Proceedings of DARPA Case-Based Reasoning Workshop*, 270–274.

Huang, Y., Bianchi, M., Liarokapis, M., & Sun, Y. (2016). Recent data sets on object manipulation: A survey. *Big Data*, *4*(4), 197–216. doi:10.1089/big.2016.0042 PMID:27992265

Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (pp. 216–223). 10.3115/1119355.1119383

IBM. (2017). *IBM in Europe: A Long History and Deep Presence.* Retrieved from IBM.com: https://www.ibm.com/blogs/policy/wp-content/uploads/2017/03/IBM-in-Europe-2017_Brochure_pdf-FINAL.pdf

ICA. (2019). *Spending by African governments on Infrastructure.* Available at: https://www.icafrica.org/en/topics-programmes/spending-by-african-governments-on-infrastructure/

IHS, Omdia, & PointTopic. (2020). *Broadband coverage in Europe 2019: Mapping progress towards the coverage objectives of the Digital Agenda.* Luxembourg: EU.

IoT Analytics. (2020). *IoT market outlook*. Retrieved November 22, 2020, https://iot-analytics.com/

Irani, Z., Sharif, A. M., Papadopoulos, T., & Love, P. E. D. (2017). Social media and Web 2.0 for knowledge sharing in product design. *Production Planning and Control*, *28*(13), 1047–1065. doi:10.1080/09537287.2017.1329955

Ishii, K., Lyons, M. M., & Carr, S. A. (2019). Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies*, *1*(2), 124–131. doi:10.1002/hbe2.138

Iskhakov, R. (2013). *History of Electricity in the United States.* Stanford. Retrieved from http://large.stanford.edu/courses/2013/ph240/iskhakov2/

ISO/TEC. (2004). *Information technology - process assessment - Part 4: guidance on use for process improvement and process capability determination*. https://www.iso.org/obp/ui/#iso:std:iso-iec:15504:-4:ed-1:v1:en

IT-Kooperation zwischen Bund und Landern. (1998, June). [IT cooperation between the federal government and the federal states]. Retrieved from https://archive.md/20121215162234/http://reference.e-government.gv.at/uploads/media/IT-Koop--Bund-Laender.html

Ito, E. (2020). Understanding Cultural Variations in Outdoor Tourism Behaviours for Outdoor Sport Tourism Development: A Case of the Blue Mountains National Park. *Tourism Planning & Development*, 1–7. Advance online publication. doi:10.1080/21568316.2020.1807401

ITU. (2020). *Measuring digital development: Facts and figures 2020.* ITU. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf

ITU. (2020). *Measuring Digital Development: ICT Price Trends 2019*. ITU Publications.

James, R. N. (1997). DISXPERT: A rule-based vocational rehabilitation risk assessment system. *Expert Systems with Applications*, *12*(4), 465–472. doi:10.1016/S0957-4174(97)00006-7

Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. *Proceedings of International Conference Research on Computational Linguistics*.

Jiang, X., Hu, Y., & Li, H. (2009). A ranking approach to keyphrase extraction. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 756–757). ACM.

Johansson, C., Hall, P., & Coquard, M. (1999). Talk to Paula and Peter – They are Experienced. *Proceedings of the Workshop of Learning Software Organizations*, 69-76.

Kabasakal, O., & Mutlu, A. (2020). *On effect of word positions in graph-based keyword extraction.* Submitted.

Kailas, A., Cecchi, V., & Mukherjee, A. (2012). A survey of communications and networking technologies for energy management in buildings and home automation. *Journal of Computer Networks and Communications*, *2012*, 1–12. doi:10.1155/2012/932181

Kaiya, H., & Sacki, M. (2006). Using Domain Ontology as Domain Knowledge for Requirement Elicitation. *Proceedings of the 14th IEEE International Requirements Engineering Conference*, 189-198.

Kalakota, R., & Whiston, A. (1997). *Electronic commerce: a manager's guide*. Addison Wesley.

Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., & Huber, P. (2016). Artificial intelligence and computer science in education: From Kindergarten to university. *Proceedings - Frontiers in Education Conference, FIE,* 1–9. 10.1109/FIE.2016.7757570

Kang, S.-S. (2003). Keyword-based document clustering. In *Proceedings of the 6th International Workshop on Information Retrieval with Asian Languages* (pp. 132–137). Academic Press.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Karamitri, I., Talias, M. A., & Bellali, T. (2017). Knowledge management practices in healthcare settings: A systematic review. *The International Journal of Health Planning and Management*, *32*(1), 4–18. doi:10.1002/hpm.2303 PMID:26153388

Kaufer, F., & Klusch, M. (2006). WSMO-MX: A Logic Programming Based Hybrid Service Matchmaker. *European Conference on Web Services*, 161-170.

Kess, P., & Haapasalo, H. (2002). Knowledge creation through a project review process in software production. *International Journal of Production Economics*, *80*(1), 49–55.

Kharrat, M., Reyhani, N., & Badie, K. (2003). A case-based reasoning approach to intelligent tutoring system by considering learner style model. *Proceedings of the 2003 Systems and Information Engineering Design Symposium*. 10.1109/SIEDS.2003.158002

Khasanah, A. U., & Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. *IOP Conference Series. Materials Science and Engineering*, *215*(1), 012036. Advance online publication. doi:10.1088/1757-899X/215/1/012036

Kiefer, C., & Bernstein, A. (2008). The Creation and Evaluation of iSPARQL Strategies for Matchmaking. *European Semantic Web Conference*, 463-477.

Kim, K. (2007). *History of Electrification in Europe.* Retrieved from https://www.zum.de/whkmla/sp/0809/kyungmook/km2.html

Kim, B., & Lee, S. (1995). A bond Rating expert system for industrial companies. *Expert Systems with Applications*, *91*(1), 63–70. doi:10.1016/0957-4174(94)00049-2

Kim, D. J., Kim, W. G., & Han, J. S. (2006). A perceptual mapping of online travel agencies and preference attributes. *Tourism Management*, *28*(2), 591–603. doi:10.1016/j.tourman.2006.04.022

Kim, K., Chang, D. S., & Suk Choi, Y. (2017). Ontology based Human-Robot Interaction Knowledge for Intelligent Services. *Workshop on Social Human-Robot Interaction of Service Robots*.

Kim, S.-W., Lee, Y., Park, J., Kim, S., Chae, H., Ko, H., & Kim, J. (2017). A Triple-Mode Flexible E-Skin Sensor Interface for Multi-Purpose Wearable Applications. *Sensors (Basel)*, *18*(2), 78. doi:10.339018010078 PMID:29286312

Kindel, A. T., Bansal, V., Catena, K. D., Hartshorne, T. H., Jaeger, K., Koffman, D., McLanahan, S., Phillips, M., Rouhani, S., Vinh, R., & Salganik, M. J. (2019). Improving Metadata Infrastructure for Complex Surveys: Insights from the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World*, *5*, 237802311881737. doi:10.1177/2378023118817378

King, C., Kerr, A., Jefferies, M., & Brombley, D. (2005). *Travel and Tourism*. Heinemann Educational Publishers.

King, D. (1990). Intelligent decision Support: Strategies for integrating decision support, database management, and expert system technologies. *Expert Systems with Applications*, *1*(1), 23–38. doi:10.1016/0957-4174(90)90066-4

Kliček, B., & Vukovac, D. (2007). Information Society and eGovernment Developments in Croatia. *Informatica (Vilnius)*, *31*, 367–372.

Klusch, M., Fries, B., & Sycara, K. (2008). OWLS-MX: A Hybrid Semantic Web Service Matchmaker for OWL-S Services. *Journal of Web Semantics*.

Knowles, H., & Westcott, M. (2015). Travel services. In *Introduction to Tourism and Hospitality in BC* (pp. 147-166). BCcampus. Retrieved from https://opentextbc.ca/introtourism/

Knuesel, A. (2017). British Diplomacy and the Telegraph in Nineteenth-Century China. *Diplomacy and Statecraft, 18*(2007), 517-537.

Köhne, H., & Sawyer, A. (2018). Process innovation: Requirements and opportunities. *European Journal of Management Issues*, *26*(1-2), 22–28. doi:10.15421/191803

Kokai, T. R. (1974). Measures towards EDP- Appropriate legislation. *Jurimetrics Journal*, *15*(37), 37–42.

Kokku, R., Sundararajan, S., Dey, P., Sindhgatta, R., Nitta, S., & Sengupta, B. (2018). Augmenting Classrooms With AI for Personalized Education. *IEEE Global Engineering Education Conference, EDUCON*, 6976–6980.

Kolodner, J. (1993). *Case-based Reasoning*. Morgan Kaufmann. doi:10.1016/B978-1-55860-237-3.50005-4

Kopytko, V., Shevchuk, L., Yankovska, L., Semchuk, Z., & Strilchuk, R. (2018). Smart home and artificial intelligence as environment for the implementation of new technologies. *Traektoriâ Nauki = Path of Science, 4*(9), 2007–2012.

Kordha, E., Gorica, K. & Ahmetaj, L. (2011). Managing IT Infrastructure for Information Society Development. The Albanian Case. *Romanian Economic and Business Review,* 6(2), 122 - 131.

Korkala, M., Pikkarainen, M., & Conboy, K. (2010). A Case Study of Customer Communication in Globally Distributed Software Product Development. *Proceedings of the 11th International Conference on Product Focused Software*, 43-46.

Kraemer, K. L., & King, J. L. (1977). *Computers and Local Government*. Praeger.

Kranenburg, R. V. (2008). *The Internet of Things: A Critique of Ambient Technology and the All-Seeing Network of RFID*. Institute of Network Cultures. Retrieved December 17, 2020, https://www.networkcultures.org/_uploads/notebook2_theinternetofthings.pdf

Krar, S. (n.d.). *Technology Resources - Manufacturing Automation.* https://www.automationmag.com/images/stories/LWTech-files/83%20Technology%20Resources.pdf: Automation mag.

Kumar, S. A., & Suresh, N. (2008). Production and Operations Management (2nd ed.). New Delhi: New Age International (P) Ltd.

Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: A review. *Journal of Big Data*, *6*(1), 111. doi:10.118640537-019-0268-2

Lajmi, S., Ghedira, C., & Ghedira, K. (2006a). How to apply CBR method in web service composition. In *Second International Conference on Signal-Image Technology and Internet Based Systems (SITI'2006).* Springer Verlag.

Lajmi, S., Ghedira, C., Ghedira, K., & Benslimane, D. (2006b). Web_CBR: How to compose web service via case-based reasoning. *IEEE International Symposium on Service-Oriented Applications, Integration and Collaboration held with the IEEE International Conference on eBusiness Engineering (ICEBE 2006).*

Laliwala, Z., Khosla, R., Majumdar, P., & Chaudhary, S. (2006). Semantic and Rule Based Event-Driven Dynamic Web Service Composition for Automation of Business Processes. *Proceedings of the IEEE Service Computing Workshop (SCW06).*

Lambert, D. M., & Cooper, M. C. (2000). Issues in Supply Chain Management. *Industrial Marketing Management*, *29*(1), 65–83.

Lasheras, J., Valencia-Garcia, R., Fernandez-Breis, J. T., & Toval, A. (2009). Modeling Reusable Security Requirements Based on Ontology Framework. *Journal of Research and Practice in Information Technology*, *4*(2), 119–133.

Latour, B. (1996). On Actor-Network Theory: A few clarifications. *Soziale Welt*, *47*(4), 369–381.

Laudon, K. C., & Laudon, J. P. (2014). Management Information System - Managing the Digital Firm (13th ed.). Pearson Education Limited.

Lavery, P. (1992). The financing and organisation of national tourist offices. *EIU Travel & Tourism Analyst*, *4*, 84–101.

Law, R., Leung, K., & Wong, R. (2004). The impact of the Internet on travel agencies. *International Journal of Contemporary Hospitality Management*, *16*(2), 100–107. doi:10.1108/09596110410519982

Lebo, T. (2019). *Hyper-Personalization: What It Is and Why You Need It in Your 2019 Marketing*. Academic Press.

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431–440. doi:10.1016/j.bushor.2015.03.008

Lee, S., & Wu, C. (1995). CLXPERT: A rule-based scheduling system. *Expert Systems with Applications*, *22*(2), 153–164. doi:10.1016/0957-4174(94)00058-4

Lee, W., & Sharma, A. (2016). Smart sensing for IoT applications. *13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, 362-364. 10.1109/ICSICT.2016.7998921

Lee, W.-T., Shin-Yuan, H., & Chau, P. Y. (2011). Influence of knowledge management infrastructure on innovative business processes and market-interrelationship performance: An empirical study of hospitals in Taiwan. *Journal of Global Information Management*, *19*(2), 67–89. doi:10.4018/jgim.2011040104

Lemaignan, S. (2012). *Grounding the Interaction: Knowledge Management for Interactive Robots*. INSA de Toulouse.

Lemaignan, S., Ros, R., Mösenlechner, L., Alami, R., & Beetz, M. (2010). ORO, A Knowledge Management Platform for Cognitive Architectures in Robotics. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3548-3553. 10.1109/IROS.2010.5649547

Lemaignan, S., Ros, R., Sisbot, E. A., Alami, R., & Beetz, M. (2011). Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics*, 1–19.

Lemaignan, S., Warnier, M., Sisbot, A. E., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45–69. doi:10.1016/j.artint.2016.07.002

Leslie, D. (2019). Raging robots, hapless humans: The AI dystopia. *Nature*, *574*(7776), 32.

Leslie, P., Pearce, J. M., Harrap, R., & Daniel, S. (2012). The application of smartphone technology to economic and environmental analysis of building energy conservation strategies. *International Journal of Sustainable Energy*, *31*(5), 295–311. doi:10.1080/1478646X.2011.578746

Le, T. T. N., Le Nguyen, M., & Shimazu, A. (2016). Unsupervised keyphrase extraction: Introducing new kinds of words to keyphrases. In *Australasian Joint conference on Artificial Intelligence* (pp. 665–671). 10.1007/978-3-319-50127-7_58

Leuterio, F.C. (2007). *Introduction to Tourism*. Manila, Philippines: Book Store.

Levin, K. A. (2006). Evidence-based dentistry. *Study Design III: Cross-Sectional Studies, 7*(1), 24-25.

Li, S. H. (2002). *An Integrated Model for Supply Chain Management Practice, Performance and Competitive Advantage* (PhD Dissertation). University of Toledo, Toledo, OH.

Liang, T., & Konsynski, B. R. (1993). Modeling by analogy: Use of analogical reasoning in model management systems. *Decision Support Systems*, *9*, 113–125.

Liebowitz, J., & Megbolugbe, I. (2003, April). A set of frameworks to aid the project manager in conceptualizing and implementing knowledge management initiatives. *International Journal of Project Management*, *21*(3), 189–198. doi:10.1016/S0263-7863(02)00093-5

Li, H. (2011). A short introduction to learning to rank. *IEICE Transactions on Information and Systems*, *94*(10), 1854–1862. doi:10.1587/transinf.E94.D.1854

Li, J., Sikora, R., Shaw, M. J., & Woo Tan, G. A. (2006). Strategic analysis of inter-organizational information sharing. *Decision Support Systems*, *42*, 251–266.

Li, L., & Horrocks, I. (2003). A Software Framework for Matchmaking Based on Semantic Web Technology. *International Conference in World Wide Web*, 331-339.

Li, M., & Lin, H. J. (2015). Design and implementation of smart home control systems based on wireless sensor networks and power line communications. *IEEE Transactions on Industrial Electronics*, *62*(7), 4430–4442. doi:10.1109/TIE.2014.2379586

Li, M., Yu, B., Rana, O. F., & Wang, Z. (2008). Grid Service Discovery with Rough Sets. *IEEE Transactions on Knowledge and Data Engineering*, *20*(6), 851–862.

Lim, G., Suh, I., & Suh, H. (2011). Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Transactions on Systems, Man, and Cybernetics*, 1–18.

LinkedIn. (2018). *LinkedIn Workforce Report: United States.* Available at: https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018

Li, S., Yan, M., Zhang, X., & Li, Z. (2020). *Analysis on the Application of AI Technology in Online Education Under the Public Epidemic Crisis*. Springer International Publishing. doi:10.1007/978-3-030-63885-6_34

Liu, Z., Huang, W., Zheng, Y., & Sun, M. (2010). Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 366–376). Academic Press.

Liu, Q., Do, T. D. T., & Cao, L. (2020). Answer keyword generation for community question answering by multi-aspect gamma-poisson matrix completion. *IEEE Intelligent Systems*, 1. doi:10.1109/MIS.2020.2997714

Liu, Z., Li, P., Zheng, Y., & Sun, M. (2009). Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 257–266). 10.3115/1699510.1699544

Li, Z. (2012). *Multimodal Human-Robot Interaction in an Assistive Technology Environment*. Monash University.

Loukam, M., Hammouche, D., Mezzoudj, F., & Belkredim, F. Z. (2019). Keyphrase extraction from modern standard arabic texts based on association rules. In *International Conference on Arabic Language Processing* (pp. 209–220). 10.1007/978-3-030-32959-4_15

Lubbe, B. (2000). *Tourism distribution: Managing the travel intermediary*. JUTA.

Luckin, R. (2020). *Hey Siri – why is AI important for education?* Academic Press.

Lund-Jensen, R., Azaria, C., Permien, F., Sawari, J., & Bækgaard, L. (2016). Feral information systems, shadow systems, and workarounds - A drift in IS terminology. *Procedia Computer Science*, *100*, 1056–1063. doi:10.1016/j.procs.2016.09.281

Lu, X., & Chow, T. W. (2019). Duration modeling with semi-markov conditional random fields for keyphrase extraction. *IEEE Transactions on Knowledge and Data Engineering*, 1. doi:10.1109/TKDE.2019.2942295

Lynn, H. M., Choi, C., Choi, J., Shin, J., & Kim, P. (2016). The method of semi-supervised automatic keyword extraction for web documents using transition probability distribution generator. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems* (pp. 1–6). 10.1145/2987386.2987399

Lynn, H. M., Lee, E., Choi, C., & Kim, P. (2017). Swiftrank: An unsupervised statistical approach of keyword and salient sentence extraction for individual documents. *Procedia Computer Science*, *113*, 472–477. doi:10.1016/j.procs.2017.08.305

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-Driven Meets Theory-Driven Research in the Era of Big Data: Opportunities and Challenges for Information Systems Research. *Journal of the Association for Information Systems*, 1253–1273. doi:10.17705/1jais.00526

Maayan, D. G. (2018). *Big data as a service (BDaaS) solutions: comparing IaaS, PaaS and SaaS.* Available at: https://hub.packtpub.com/big-data-as-a-service-bdaas-solutions-comparing-iaas-paas-and-saas/

Maayan, G. (2020). *Hyper Personalization: Customizing Service With AI*. Academic Press.

Machado, D., Barbosa, T., Pais, S., Martins, B., & Dias, G. (2009). Universal mobile information retrieval. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 345–354). Academic Press.

Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley and Sons Ltd.

Martin, D., Burstein, M., Mcdermott, D., Mcilraith, S., Paolucci, M., Sycara, K., Mcguinness, D. L., Sirin, E., & Srinivasan, N. (2007). Bringing Semantics to Web Services with OWL-S. *World Wide Web (Bussum)*, *10*(3), 243–277.

Martin, D., Paolucci, M., McIlraith, S., Burstein, M., McDermott, D., McGunness, D., Barsia, B., Payne, T., Sabou, M., Solanki, M., Srinivasan, N., & Sycara, K. (2004). Bringing Semantics to Web Services: The OWL-S Approach. *Proceeding of First International Workshop Semantic Web Services and Web Process Composition*.

Marujo, L., Ling, W., Trancoso, I., Dyer, C., Black, A. W., Gershman, A., . . . Carbonell, J. G. (2015). Automatic keyword extraction on twitter. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (pp. 637–643). Academic Press.

Mathews, B. K., Koenig, S., Kurian, L., Galen, B., Mints, G., Liu, G., & Soni, N. J. (2020). Clinical Progress Note: Point-of-Care Ultrasound Applications in COVID-19. *Journal of Hospital Medicine*, *15*(6), 353–355. doi:10.12788/jhm.3454 PMID:32490807

Matskin, M., Maigre, R., & Tyugu, E. (2007). Computational logical semantics for business process language. In *Proceedings of second international conference on Internet and Web applications and services (ICIW 2007)*. IEEE Computer Society.

Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal of Artificial Intelligence Tools*, *13*(01), 157–169. doi:10.1142/S0218213004001466

Matthews, D. (2014). *Five trends that will shape the future of ERP*. https://www.industryweek.com/technology-and-iiot/article/21964701/five-trends-that-will-shape-the-future-of-erp

Maurer, E. (2003). *Internet for the retail industry*. Thomson Delmar Learning.

Mavridis, N., & Roy, D. (2006.). Grounded situation models for robots: Where words and percepts meet. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 10.1109/IROS.2006.282258

Mayank, V., Kositsyna, N., & Austin, M. (2004). *Requirements Engineering and Semantic Web, Part II. Representation, Management, and Validation of Requirements and System-Level Architectures, ISR Technical Report*. University of Maryland.

McGinty, L., & Smyth, B. (2001). Collaborative case-based reasoning: Applications in personalized route planning. *4th International Conference on Case Based Reasoning, ICCBR. Lecture Notes in Computer Science, 2080*, 362.

McIlraith, S., & Martin, D. (2003). Bringing Semantics to Web Services. *IEEE Intelligent Systems*, *18*(1), 90–93.

McIvor, R., Mulvenna, K., & Humphreys, P. (1997). Hybrid knowledge-based system for strategic purchasing. *Expert Systems with Applications*, *12*(4), 497–512. doi:10.1016/S0957-4174(96)00009-7

McKinsey. (2015). *The internet of things: mapping the value beyond the hype*. Retrieved January 10, 2021, https://www.mckinsey.com/

McNiel, I. (1990). *An Encyclopedia of the History of Technology*. Routledge.

Medelyan, O., Frank, E., & Witten, I. H. (2009). Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 1318–1327). Academic Press.

Meng, R., Zhao, S., Han, S., He, D., Brusilovsky, P., & Chi, Y. (2017). Deep keyphrase generation. arXiv preprint arXiv:1704.06879. doi:10.18653/v1/P17-1054

Meri, M. (2020). New Trends in HRM & Knowledge Management in the Health Sector beyond COVID-19-A Practical Model. *Business Excellence and Management, 10*(SI 1), 5-21.

Meyer, H., Tore, A. D., Siegel, F., & Curley, F. (1992). The strategic use of expert systems for risk management in the insurance Ž. industry. *Expert Systems with Applications*, *5*(1-2), 15–24. doi:10.1016/0957-4174(92)90090-F

Michaelsen, R. (1984). An expert system for federal tax planning. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, *1*(2), 149–167. doi:10.1111/j.1468-0394.1984.tb00440.x

Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404–411). Academic Press.

Minges, M. (1999). *Universal Service/Access Indicators - measuring access to telecommunications: Universal service and access indicators.* ITU. Retrieved from https://www.itu.int/ITU-D/ict/WTIM99/PDF/PartVIFinalR_e.pdf

Mirbabaie, M., Bunker, D., Stieglitz, S., Marx, J., & Ehnis, C. (2020). Social media in times of crisis: Learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *Journal of Information Technology*, *35*(3), 195–213. doi:10.1177/0268396220929258

MKM. (2020). *Real-Time Economy Vision (2020 -2027).* Tallin: MKM Estonia. Retrieved from https://www.mkm.ee/sites/default/files/real-time_economy_vision_2020-2027.pdf

Mkrttchian, V. (2020). Human Capital Management in the Context of the Implementation of Digital Intelligent Decision Support Systems and Knowledge Management: Theoretical and Methodological Aspects. In Knowledge Management, Innovation, and Entrepreneurship in a Changing World (pp.123-148). IGI Global. doi:10.4018/978-1-7998-2355-1.ch006

Mkrttchian, V. (2019a). New Tools for Cyber Security Using Blockchain Technology and Avatar-Based Management Technique. In *Machine Learning and Cognitive Science Applications in Cyber Security* (pp. 105–122). IGI Global., doi:10.4018/978-1-5225-8100-0.ch004

Mkrttchian, V. (2019b). Machine Learning With Avatar-Based Management of Sleptsov Net-Processor Platform to Improve Cyber Security. In *Machine Learning and Cognitive Science Applications in Cyber Security* (pp. 139–153). IGI Global.

Mkrttchian, V. (2021a). Avatars-Based Decision Support System Using Blockchain and Knowledge Sharing for Processes Simulation a Natural Intelligence: Implementation of the Multi Chain Open Source Platform. *International Journal of Knowledge Management*, *17*(1), 5. Advance online publication. doi:10.4018/IJKM.2021010105

Mkrttchian, V., & Aleshina, E. (2017). *Sliding Mode in Intellectual Control and Communication: Emerging Research and Opportunities*. IGI Global. doi:10.4018/978-1-5225-2292-8

Mkrttchian, V., & Chernyshenko, S. (2021b). Digital Intelligent Design of Avatar-Based Control with Application to Human Capital Management. *International Journal of Human Capital and Information Technology Professionals*, *12*(1), 2. Advance online publication. doi:10.4018/IJHCITP.2021010102

Mockus, A., Fielding, R. T., & Herbsleb, J. D. (2002, July). Two Case Studies of Open-Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, *11*(3), 309–346. doi:10.1145/567793.567795

MOH. (2016). *Minstry of Health National Policy Guidelines*. MOH.

Mohammed, A., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation & Knowledge*, *4*(2), 104–114. doi:10.1016/j.jik.2017.07.003

Morato, C., Kaipa, K., Zhao, B., & Gupta, S. K. (2014). Safe Human Robot Interaction by Using Exteroceptive Sensing Based Human Modeling. *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE).*

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *The Journal of Economic Perspectives*, *31*(2), 87–106. doi:10.1257/jep.31.2.87

Munk, J. (2015). *Six sigma daily*. https://www.sixsigmadaily.com/six-sigma-vs-business-process-management/

Mutlu, B., Roy, N., & Šabanović, S. (2016). Cognitive Human–Robot Interaction. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics*. Springer. doi:10.1007/978-3-319-32552-1_71

Nagy, I., Berend, G., & Vincze, V. (2011). Noun compound and named entity recognition and their usability in keyphrase extraction. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011* (pp. 162–169). Academic Press.

Nagy, Z., Yong, F. Y., & Schlueter, A. (2016). Occupant centered lighting control: A consumer study on balancing comfort, acceptance, and energy consumption. *Energy and Building*, *126*, 310–322. doi:10.1016/j.enbuild.2016.05.075

Navlani, A. (2018a, May 16). *Understanding Random Forests Classifiers in Python*. Datacamp. https://www.datacamp.com/community/tutorials/random-forests-classifier-python

Navlani, A. (2018b, August 2). *KNN Classification using Scikit-learn*. Datacamp. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

Navlani, A. (2019, December 27). *Support Vector Machines with Scikit-learn*. Datacamp. https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python

Nguyen, T. D., & Kan, M.-Y. (2007). Keyphrase extraction in scientific publications. In *International Conference on Asian Digital Libraries* (pp. 317–326). Academic Press.

Nguyen, T. D., & Luong, M.-T. (2010). Wingnus: Keyphrase extraction utilizing document logical structure. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 166–169). Academic Press.

Nguyen, N., & Guo, Y. (2007). Comparisons of sequence labeling algorithms and extensions. In *Proceedings of the 24th International Conference on Machine Learning* (pp. 681–688). 10.1145/1273496.1273582

Nisingizwe, M. P., Iyer, H. S., Gashayija, M., Hirschhorn, L. R., Amoroso, C., Wilson, R., & Hedt-Gauthier, B. (2014). Toward utilization of data for program management and evaluation: Quality assessment of five years of health management information system data in Rwanda. *Global Health Action*, *7*(1), 25829. doi:10.3402/gha.v7.25829 PMID:25413722

Nixon, P. G., & Koutrakou, V. N. (2007). *E-government in Europe*. Routledge, Taylor and Francis Group.

Nonaka, I., & von Krogh, G. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science*, *20*(3), 635–652.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37. doi:10.1287/orsc.5.1.14

Nor'ashikin, A., Tretiakov, A., Whiddett, D., & Hunterb, I. (2017). Knowledge management systems success in healthcare: Leadership matters. *International Journal of Medical Informatics*, *97*, 331–340. doi:10.1016/j.ijmedinf.2016.11.004 PMID:27919392

Nora, S., & Minc, A. (1978). *L'informatisation de la société.* Paris: La Documentation Francaise: Présidence de la République. Retrieved from https://www.vie-publique.fr/rapport/34772-linformatisation-de-la-societe

O'Reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *International Journal of Digital Economics*, *65*, 17–37. https://mpra.ub.uni-muenchen.de/4580/

OASIS. (2004). *Introduction to UDDI: Important Features and Functional Concepts*. Organization for the Advancement of Structured Information Standards.

Obeidy, W. K., Arshad, H., & Huang, J. Y. (2017). An acceptance model for smart glasses based tourism augmented reality. AIP Conference Proceedings, 1891(1). doi:10.1063/1.5005413

OECD. (2017). *Digital Government Review of Norway: Boosting the digital transformation of the publlic sector.* OECD. Retrieved from https://www.oecd.org/gov/digital-government/digital-government-review-norway-recommendations.pdf

Offutt, B. (2013). *PhoCusWright's travel innovations & technology trends: 2013 and beyond*. Academic Press.

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *JMR, Journal of Marketing Research*, *17*(4), 460.

OMG. (2008). *Business Process Maturity Model (BPMM), version 1.0*. OMG.

OMG. (2009). *Business Process Model and Notation*. https://www.omg.org/spec/BPMN/1.2/(2009)

OMG. (2011). *Business Process Model and Notation (BPMN), version 2.0*. OMG.

Öörni, A. (2003). Consumer search in electronic markets: An experimental analysis of travel services. *European Journal of Information Systems*, *12*(1), 30–40. doi:10.1057/palgrave.ejis.3000450

Osei, A., Yunfei, S., Appienti, W., & Forkuoh, S. (2016). The antecedents of process innovation and Smes growth: Empirical evidence from shoe manufacturing sector in the Ashanti region Of Ghana. *European Scientific Journal*, 163–173.

Otava. (2017). *The importance of infrastructure to Big Data analytics.* Available at: https://www.otava.com/blog/the-importance-of-infrastructure-to-big-data-analytics/

Pal, K. & Campbell, J. A. (1997). An Application of Rule-Based and Case-Based Reasoning within a Single Legal Knowledge-Based System. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 28*(4), 48-63.

Pal, K. (2019). Markov Decision Theory-Based Crowdsourcing Software Process Model. In Crowdsourcing and Probabilistic Decision-Making in Software Engineering: Emerging Research and Opportunities. IGI Global.

Pal, K. (2020). Framework for Reusable Test Case Generation in Software System Testing. In Software Engineering for Agile Application Development. IGI Global.

Pal, K., & Campbell, J. A. (1995). A hybrid system for decision-making about assets in English divorce cases. In *Advances in Case Based Reasoning: First UK CBR Workshop*. Springer-Verlag. 10.1007/3-540-60654-8_30

Pal, K., & Karakostas, B. (2020). Software Testing Under Agile, Scrum, and DevOps. In Agile Scrum Implementation and Its Long-Term Impact on Organizations. IGI Global.

Pal, A. K. (2017). Analysis and Mining of Educational Data for Predicting the Performance of Students Analysis and Mining of Educational Data for Predicting the Performance of Students. *International Journal of Electronics Communication and Computer Engineering*, *4*(June).

Paliwal, A. V., Shafiq, B., Vaidya, J., Xiong, H., & Adam, N. (2012). Semantics-based automated service discovery. *IEEE Transactions on Services Computing*, *5*(2), 260–275.

Pal, K., & Campbell, J. A. (1996). A hybrid Legal decision-support system using both rule-based and case-based reasoning. *Information & Communications Technology Law*, *5*(3), 227–245. doi:10.1080/13600834.1996.9965747

Pal, K., & Campbell, J. A. (1997). An application of rule-based and case-based reasoning within a single legal knowledge-based system. *The Data Base for Advances in Information Systems*, *28*, 48–63.

Pal, K., & Palmer, O. (2000). A decision-support system for business acquisition. *Decision Support Systems*, *27*, 411–429.

Panahi, S., Watson, J., & Partridge, H. (2013). Towards tacit knowledge sharing over social web tools. *Journal of Knowledge Management*, *17*(3), 379–397. doi:10.1108/JKM-11-2012-0364

Papagiannopoulou, E., & Tsoumakas, G. (2018). *Unsupervised keyphrase extraction based on outlier detection.* CoRR, abs/1808.03712.

Papagiannopoulou, E., Tsoumakas, G., & Papadopoulos, A. N. (2020). *Keywords lie far from the mean of all words in local vector space*. Academic Press.

Papagiannopoulou, E., & Tsoumakas, G. (2020). A review of keyphrase extraction. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *10*(2), e1339. doi:10.1002/widm.1339

Papazoglou, M. (2012). *Web Services and SOA: Principles and Technology*. Pearson.

Pathak, J., Koul, N., Caragea, D., & Honavar, V. G. (2005). A Framework for Semantic Web Services Discovery. *ACM International Workshop on Web Information and Data Management*, 45-50.

Patil, A., Oundhaka, S., Sheth, A., & Verma, K. (2004). METEOR-S Web service Annotation Framework. *The Proceedings of the Thirteenth International World Wide Web Conference*, 553-562.

Peffer, T., Blumstein, C., Culler, D., Modera, M., & Meier, A. (2015). Software-defined solutions for managing energy use in small to medium sized commercial buildings (No. EE0006351). Berkeley, CA: Regents of the University of California.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, *31*(7), 770–780. doi:10.1177/0956797620939054 PMID:32603243

Persson, P., Laaksolahti, J., & Lonnqvist, P. (2001). Understanding socially intelligent agents - a multilayered phenomenon. *IEEE Transactions on Systems, Man, and Cybernetics*, *31*(5), 349–360. doi:10.1109/3468.952710

Philips Lighting Catalog. (2014). *Document Number 919002151397*. Amsterdam: Koninklijke Philips Electronics N.V.

Pike, R., & Neale, B. (1999). *Corporate Finance and Investment: Decisions and Strategies*. Prentice-Hall Europe.

Plantera, F. (2017). *Artificial Intelligence is the next step for e-governance in Estonia", State adviser reveals*. E-Estonia. Retrieved from https://e-estonia.com/artificial-intelligence-is-the-next-step-for-e-governance-state-adviser-reveals/

Plinere, D., & Borisov, A. (2015). Case study on inventory management improvement. *Information Technology and Management Science*, *18*(1), 91–96. doi:10.1515/itms-2015-0014

Plitnichenko, L. (2020). *5 Main Roles Of Artificial Intelligence In Education*. Academic Press.

Polat, S., Esen, F., & Bilgic, E. (2019). Analysis of the 5Vs of Big Data in Virtual Travel Organizations. In *Big Data and Knowledge Sharing in Virtual Organizations* (pp. 43–70). IGI Global. doi:10.4018/978-1-5225-7519-1.ch002

Price, D. G., & Mulvihill, D. E. (1965). The Present and Future Use of Computers in State Government. *Public Administration Review*, *25*(2), 142–150.

Prinzie, A., & Van den Poel, D. (2008). Random Forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications*, *34*(3), 1721–1732. doi:10.1016/j.eswa.2007.01.029

Qi, C., & Leung, H. (2015, June). *Exploring the role of social media and knowledge management processes in organizational learning*. https://research.polyu.edu.hk/en/publications/exploring-the-role-of-social-media-and-knowledge-management-proce

Quah, T., Tan, C., Raman, K. S., & Srinivasan, B. (1996). Towards integrating rule-based expert systems and neural networks. *Decision Support Systems*, *17*(2), 99–118. doi:10.1016/0167-9236(95)00016-X

Rabby, G., Azad, S., Mahmud, M., Zamli, K. Z., & Rahman, M. M. (2020). Teket: A tree-based unsupervised keyphrase extraction technique. *Cognitive Computation*, *12*(4), 1–23. doi:10.100712559-019-09706-3

Rabinowitz, L., Weinberg, M. A., & Blank, M. V. (1997). *Weinberg and Blank on Takeovers and Mergers*. Sweet & Maxwell.

Rafkin, R. (2008, February 25). *Multiclass Classification*. MIT. https://www.mit.edu/~9.520/spring09/Classes/multiclass.pdf

Rammasubbu, N., & Balan, R. K. (2008). Towards Governance Scheme for Distributed Software Development Projects. *Proceedings of the 1st International Workshop on Software Development Governance*, 11-14.

Rangaiah, M. (2020). *4 Major Applications of Artificial Intelligence in Education Sector*. Academic Press.

Reategui, E. B., & Campbell, J. A. (1995). A Classification System for Credit Card Transactions. In *Advances in Case-Based Reasoning: Second European Workshop (EWCBR-94)*. Springer.

Regeringskansliet. (2014). *Broadband Strategy for Sweden.* Stockholm: Ministry of Enterprise Energy and Communications Sweden.

Reinermann, H. (1997). The utilisation of information technology in German public administration. *International Review of Administrative Sciences*, *2*, 207–223.

Reuter, C., Kaufhold, M. A., Schmid, S., Spielhofer, T., & Hahne, A. S. (2019). The impact of risk cultures: Citizens' perception of social media use in emergencies across Europe. *Technological Forecasting and Social Change*, *148*, 119724. doi:10.1016/j.techfore.2019.119724

Rissland, E. L., & Skalak, B. D. (1991). CABARET: Rule interpretation in a hybrid architecture. *International Journal Man–Machine Studies*, *34*, 839–887.

Rissland, L., & Ashley, K. (1989). HYPO: a precedent-based legal Ž. reasoner. In G. Vandenberghe (Ed.), *Advanced Topics of Law and Information Technology* (pp. 213–234). Kluwer.

Rist, O., & Martinez, J. (2018). *5 Enterprise resource planning (ERP) trends to watch in 2019*. https://www.pcmag.com/news/5-enterprise-resource-planning-erp-trends-to-watch-in-2019

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, *6*(4), 575–591. doi:10.100712369-014-0242-2

Robinson, P. (2009). Travel and management: An introduction. In P. Robinson (Ed.), *Operations management in the travel industry*. CAB International. doi:10.1079/9781845935030.0001

Roger, S. (2000). *Pressman. Software engineering: A Practitioner's Approach (European Adaption)* (5th ed.). McGraw-Hill International.

Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., & Fensel, D. (2005). Web service modeling ontology. *Applied Ontology*, *1*(1), 77–106.

Roman, D., Kopecky, J., Vitvar, T., Domingue, J., & Fensel, D. (2015). WSMO-Lite and hRESTS: Lightweight semantic annotations for Web services and RESTful APIs. *Journal of Web Semantics*, *31*, 39–58.

Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, *57*(10), 2266–2279. doi:10.1016/j.comnet.2012.12.018

Romero, C. D. G., Barriga, J. K. D. & Molano, J. I. R. (2016). *Big Data Meaning in the Architecture of IoT for Smart Cities.* Academic Press.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. doi:10.1002/cae.20456

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory, 1*, 1–20.

Roser, M., & Ortiz-Ospina, E. (2018). *Literacy*. Retrieved from Our world in data: https://ourworldindata.org/literacy

Rushton, A., Croucher, P., & Baker, P. (2010). *The Handbook of Logistics and Distribution Management* (4th ed.). Kogan Page Limited.

Rus, I., & Lindvall, M. (2002). Knowledge Management in Software Engineering. *IEEE Software*, *19*(3), 26–38.

Rus, I., & Lindvall, M. (2002, June). Knowledge Management in Software Engineering. *IEEE Software*, *19*(3), 26–38. doi:10.1109/MS.2002.1003450

Ryan, C. (1997). *Company Law*. Butterworths.

Sabri, L., Chibani, A., Amirat, Y., & Zarri, G. (2011). Narrative reasoning for cognitive ubiquitous robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop: Knowledge Representation for Autonomous Robots.*

Saigol, Z., Wang, M., Ridder, B., & Lane, D. M. (2015). The Benefits of Explicit Ontological Knowledge-Bases for Robotic Systems. *Towards Autonomous Robotic Systems*, 229-235.

Salifu-Siddi, F., & Williams, I. (2016). The challenges of replicating Western E-government Structures in Ghana. *The 1st Africa Regional ITS Conference in Accra, Ghana, March 10-11. 2016.* Accra: ITS.

Sandu, N., & Gide, E. (2019). Adoption of AI-chatbots to enhance student learning experience in higher education in india. *2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019*, 1–5. 10.1109/ITHET46829.2019.8937382

Sangwan, R. S., & Ros, J. (2008). Architectural Leadership and Management in Globally Distributed Software Development. *Proceedings of the 1st International Workshop on Leadership and Management in Software Architecture*, 17-22.

SAP. (2012). *News SAP*. https://news.sap.com/brazil/2012/05/10/gartner-coloca-sap-na-lideranca-mundial-de-solucoes-erp/

Schofield, H. (2012, June). *Minitel: The rise and fall of the France-wide web*. Retrieved from BBC News Magazine (Paris): https://www.bbc.com/news/magazine-18610692

Schumacher, B. G. (1967). *An EDP approach in public administration*. University of Oklahoma.

Schutzer, D. (1990). Business expert systems: The competitive edge. *Expert Systems with Applications*, *1*(1), 17–21. doi:10.1016/0957-4174(90)90065-3

Schwartz, D. G. (2007). Integrating knowledge transfer and computer-mediated communication: Categorizing barriers and possible responses. *Knowledge Management Research and Practice*, *5*(4), 249–259. doi:10.1057/palgrave.kmrp.8500153

Sciarrone, F. (2018). Machine learning and learning analytics: Integrating data with learning. *2018 17th International Conference on Information Technology Based Higher Education and Training, ITHET 2018*. 10.1109/ITHET.2018.8424780

Scott, C. C. (2017). *Country Immunization Information System Assessments*. Academic Press.

Sedarati, P., & Baktash, A. (2017). Adoption of Smart Glasses in Smart Tourism Destination: A System Thinking Approach. *Travel and Tourism Research Association: Advancing Tourism Research Globally, 13.* Retrieved January 09, 2021, https://scholarworks.umass.edu/ttra/2017/Grad_Student_Workshop/13

Shah, J. A., Wiken, J., Williams, B., & Breazeal, C. (2011). Improved human–robot team performance using Chaski, a human-inspired plan execution system. *6th ACM/IEEE International Conference on Human–Robot Interaction*. 10.1145/1957656.1957668

Sharma, S. (2016). Expanded cloud plumes hiding Big Data ecosystem. *Future Generation Computer Systems*, *59*, 63–92.

Shen, H., Deng, Q., Lao, R., & Wu, S. (2017). A case study of inventory management in a manufacturing company in China. *Nang Yan Business Journal*, *5*(1), 20–40. doi:10.1515/nybj-2017-0003

Sheridan, T. B. (2012). *Human Supervisory Control*. John Wiley & Sons, Inc. doi:10.1002/9781118131350.ch34

Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. *Human Factors*, *58*(4), 525–532. doi:10.1177/0018720816644364 PMID:27098262

Shie, M. C., Lin, P. C., Su, T. M., Chen, P., & Hutahaean, A. (2014). Intelligent energy monitoring system based on ZigBee-equipped smart sockets. In *Proceedings of IEEE international conference on intelligent green building and smart grid (IGBSG)* (pp. 1–5). IEEE. 10.1109/IGBSG.2014.6835281

Shingo, S. (1988). *Non-Stock Production*. Productivity Press.

Shin, K., & Han, I. (1999). Case-based reasoning supported by genetic algorithms for corporate bond rating. *Expert Systems with Applications*, *16*(2), 85–95. doi:10.1016/S0957-4174(98)00063-3

Siddiqi, S., & Sharan, A. (2015). Keyword and keyphrase extraction techniques: A literature review. *International Journal of Computers and Applications*, *109*(2).

Sierra, E., Martnez, R., & Cataldi, Z. (2006). *Towards a methodology for the design of intelligent tutoring systems*. *Research in Computing Science Journal*.

Simonin, B. L. (1999a). Ambiguity and the process of knowledge transfer in strategic alliances. *Strategic Management Journal*, *20*(7), 595–623. doi:10.1002/(SICI)1097-0266(199907)20:7<595::AID-SMJ47>3.0.CO;2-5

Simonin, B. L. (1999b). Transfer of Marketing Know-How in International Strategic Alliances: An Empirical Investigation of the Role and Antecedents of Knowledge Ambiguity. *Journal of International Business Studies*, *30*(3), 463–490. doi:10.1057/palgrave.jibs.8490079

Singha, K., Kumar, J., & Pandit, P. (2019). Recent Advancements in Wearable & Smart Textiles: An Overview. *Materials Today: Proceedings*, *16*(3), 1518–1523. doi:10.1016/j.matpr.2019.05.334

Singhal, A., Kasturi, R., Sharma, A., & Srivastava, J. (2017). *Leveraging web resources for keyword assignment to short text documents.* arXiv preprint arXiv:1706.05985.

Singh, L. K. (2008). *Management of travel agency*. ISHA Books.

Singh, P. P., Khosla, P. K., & Mittal, M. (2019). Energy conservation in IoT-based smart home and its automation. In M. Mittal, S. Tanwar, B. Agarwal, & L. M. Goyal (Eds.), *Energy conservation for IoT devices. Studies in systems, decision and control* (Vol. 206, pp. 155–177). Springer. doi:10.1007/978-981-13-7399-2_7

Skelia. (2018). *How Iot Is Changing The Travel Industry Today And Tomorrow*. Retrieved December 18, 2020, from https://skelia.com/articles/how-iot-is-changing-the-travel-industry-today-and-tomorrow/

Skoutas, D., Simitsis, A., & Sellis, T. (2007). A Ranking Mechanism for Semantic Web Service Discovery. *IEEE Congress on Services*, 41-48.

Slack, N., Chambers, S., & Johnston, R. (2010). Operations Management (6th ed.). Pearson Education Limited.

Smelser, N. (1963). *Theory of collective behaviour.* Free Press of Glencoe.

Smith, H., & Fingar, P. (2006). *Business Process Management: The Third Wave*. Meghan Kiffer Press.

Sneath, P. H. A., & Sokal, R. R. (1976). *Numerical Taxonomy*. Freeman.

Sommerville. (2001). *Software Engineering* (6th ed.). Addison-Wesley.

Songwe, V. (2019). A digital Africa: Technology can be a springboard for faster, more inclusive growth. In Finance and Development. IMF.

Soter, G., Conn, A., Helmut, H., & Rossiter, J. (2018). Bodily aware soft robots: integration of proprioceptive and. *IEEE International Conference on Robotics and Automation (ICRA)*. 10.1109/ICRA.2018.8463169

Steinbart, P. (1984). *The Construction of a Rule-based Expert System as a Method for Studing Materiality Judgements* (PhD thesis). Michigan State University.

Steinberger, R. (2001). Cross-lingual keyword assignment. *Procesamiento del lenguaje natural, 27,* 273-280.

Stojkoska, B. L. R., & Trivodaliev, K. V. (2017). A review of Internet of Things for smart home: Challenges and solutions. *Journal of Cleaner Production*, *140*, 1454–1464. doi:10.1016/j.jclepro.2016.10.006

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1-2), 161–197.

Stylos, N., & Zwiegelaar, J. (2019). Big Data as a Game Changer: How Does It Shape Business Intelligence Within a Tourism and Hospitality Industry Context? In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*. Springer. doi:10.1007/978-981-13-6339-9_11

Suh, I., Lim, G., Hwang, W., Suh, H., & Choi, J. (2007). Ontology-based multilayered robot knowledge framework (OMRKF) for robot intelligence. *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Suleiman, D., & Awajan, A. (2017). Bag-of-concept based keyword extraction from arabic documents. In *Proceedings of the 2017 8th International Conference on Information Technology* (pp. 863–869). 10.1109/ICITECH.2017.8079959

Sun, P., Wang, L., & Xia, Q. (2017). The keyword extraction of chinese medical web page based on wf-tf-idf algorithm. In *2017 International Conference on Cyber-enabled Distributed Computing and Knowledge Discovery* (pp. 193–198). 10.1109/CyberC.2017.40

Sveiby. (1997). *The New Organizational Wealth: Managing & Measuring Knowledge-Based Assets*. Berrett-Koehler Publishers, Inc.

Svensson, A., & Hedman, E. (2021). Operations Knowledge Management in Health Care. In Operations Management - Emerging Trend in the Digital Era. Intechopen. doi:10.5772/intechopen.93793

Swain, A. (2020). *Advantages and Challenges of AI in Education for Teachers and Schools*. Academic Press.

Sycara, K. P., Paolucci, M., Ankolekar, A., & Srinivasan, N. (2003). Automated Discovery, Interaction and Computation of Semanticmweb Services. *Journal of Web Semantics*, *1*(1), 27–46.

Sycara, K., Widoff, S., Klusch, M., & Lu, J. (2002). LARKS: Dynamic Matching Among Heterogeneous Software Agents in Cyberspace. *Autonomous Agents and Multi-Agent Systems*, *5*(2), 173–203.

Sychareun, V., Hansana, V., Phengsavanh, A., Chaleunvong, K., Eunyoung, K., & Durham, J. (2014). Data verification at health centers and district health offices in Xiengkhouang and Houaphanh Provinces, Lao PDR. *BMC Health Services Research*, *14*(1), 255. doi:10.1186/1472-6963-14-255 PMID:24929940

SYZEFXIS Project. (2004). *Information Society*. Retrieved from SYZEFXIS: https://web.archive.org/web/20090414225658/http://www.infosoc.gr/infosoc/en-UK/epktp/Parakoloythish_ajiologish/EthsiaEktheshEkteleshs/Ethsiesektheseis/kales_praktikes/syzeuksis.htm

Tahiru, F. (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology*, *23*(1), 1–20. doi:10.4018/JCIT.2021010101

Tenorth, M., Jain, D., & Beetz, M. (2010). Knowledge Processing for Cognitive Robots. *KI, 24*, 233-240.

Tenorth, M., & Beetz, M. (2013). Knowrob: A knowledge processing infrastructure for cognition-enabled robots. *The International Journal of Robotics Research*, *32*(5), 566–590. doi:10.1177/0278364913481635

Tenorth, M., & Beetz, M. (2017). Representations for robot knowledge in the knowrob framework. *Artificial Intelligence. Supplement C*, *247*, 151–169.

The Role of Artificial Intelligence in the Future of Education. (n.d.). *GetSmarter Blog*.

The Royal Society. (2019). *Dynamics of data science skills How can all sectors benefit from data science talent?* The Royal Society.

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human-Computer Interaction*, *19*(1), 9–24. doi:10.120715327051hci1901&2_2

Tixier, A., Malliaros, F., & Vazirgiannis, M. (2016). A graph degeneracy-based approach to keyword extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1860–1870). 10.18653/v1/D16-1191

Tomas, S. (2014). *Tomas Söderblom's telephone museum*. Retrieved from History of Telegrafverket: http://www.soderblom.se/eng_tvhistoria.html

Toulmin, S. (1958). *The Uses of Argument*. Cambridge Univ. Press.

Tozzi, C. (2019). *4 Big Data Infrastructure Pain Points and How to Solve Them.* Available at: https://www.precisely.com/blog/big-data/4-big-data-infrastructure-points-solve

Tsai, K. L., Leu, F. Y., & You, I. (2016). Residence energy control system based on wireless smart socket and IoT. *IEEE Access: Practical Innovations, Open Solutions*, *4*, 2885–2894. doi:10.1109/ACCESS.2016.2574199

Tuck, S. (2021). *Charles Babbage*. Wichita State University. Retrieved from https://www.math.wichita.edu/history/Men/babbage.html

Turban, E. (1990). Decision Support and Expert Systems — Management Support Systems. Macmillan UK.

Turney, P. D. (1999). Learning to extract keyphrases from text (Tech. Rep.). National Research Council, Institute for Information Technology, Canada.

Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information Retrieval*, *2*(4), 303–336. doi:10.1023/A:1009976227802

Turney, P. D. (2002). Extraction of keyphrases from text: Evaluation of four algorithms. *Clinical Orthopaedics and Related Research*. https://arxiv.org/abs/cs/0212014

Tyler, D., Wood, J., Sabir, T., McDonnell, C., Sayem, A. S. M., & Whittaker, N. (2019). Wearable electronic textiles. *Textile Progress*, *51*(4), 299–384. doi:10.1080/00405167.2020.1840151

U.N. & I.F.R.R. (2002). *United Nations and The International Federation of Robotics*. World Robotics.

UN DESA. (2012). *UN E-government Survey.* New York: UN. Retrieved from https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2012-Survey/Complete-Survey.pdf

UN DESA. (2020). *United Nations E-Government Survey 2020: Digital Government in the decade of Action for Sustainable development*. United Nations.

Underwood, J. P., Hill, A., Peynot, T., & Scheding, S. J. (2010). Error modeling and calibration of exteroceptive sensors for accurate mapping applications. *Journal of Field Robotics*, *27*(1), 2–20. doi:10.1002/rob.20315

UNICEF. (2020). Retrieved April 20, 2020, from https://www.unicef.org/ghana/immunization

Uygur, U. (2013). Determinants of Causal Ambiguity and Difficulty of Knowledge Transfer within the Firm. SSRN *Electronic Journal*, 742–755. doi:10.2139srn.2336794

VanLehn, K. (1988). *Student Modelling*. M. Polson. Foundations of Intelligent Tutoring Systems.

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. G. M., & Milios, E. (2005). Semantic Similarity methods in WordNet and their application to information retrieval on the Web. *Proceedings of the 7th annual ACM international workshop on web information and data management*.

Vasconez, J. P., Kantor, G. A., & Auat Cheein, F. A. (2019). Human–robot interaction in agriculture: A survey and current challenges. *Biosystems Engineering*, *179*, 35–48. doi:10.1016/j.biosystemseng.2018.12.005

Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., & Therón, R. (2019). Dashboard meta-model for knowledge management in technological ecosystem: A case study in healthcare. *Multidisciplinary Digital Publishing Institute Proceedings*, *31*(1), 44. doi:10.3390/proceedings2019031044

Vega-Oliveros, D. A., Gomes, P. S., Milios, E. E., & Berton, L. (2019). A multi-centrality index for graph-based keyword extraction. *Information Processing & Management*, *56*(6), 102063. doi:10.1016/j.ipm.2019.102063

Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S., & Miller, J. (2005). METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Information Technology and Management*, *6*(1), 17–39.

Vodafone. (2020). *IoT spotlight report*. Retrieved January 02, 2021, https://www.vodafone.com/business/news-and-insights

Vrijhoef, R., & Koskela, L. (1999). Role of supply chain management in construction. *Proceedings of the Seventh Annual Conference of the International Group for Lean Construction*, 133-146.

Wagner, C. (2003). Knowledge Management in E-Government. *Americas Conference on Information Systems 2003, Proceedings.* Retrieved from https://aisel.aisnet.org/amcis2003/105

Wang, C., & Li, S. (2011). Corankbayes: Bayesian learning to rank under the co-training framework and its application in keyphrase extraction. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 2241–2244). 10.1145/2063576.2063936

Wang, L., & Cao, J. (2007). Web Services Semantic Searching enhanced by Case Based Reasoning. *18th International Workshop on Database and Expert Systems Applications*.

Wang, P., Jin, Z., Liu, L., & Cai, G. (2008). Building Toward Capability Specifications of Web Services Based on an Environment Ontology. *IEEE Transactions on Knowledge and Data Engineering*, *20*(4), 547–561.

Wan, X., & Xiao, J. (2008). *Single document keyphrase extraction using neighborhood knowledge* (Vol. 8). AAAI.

Wartena, C., Brussee, R., & Slakhorst, W. (2010). Keyword extraction using word co-occurrence. In *Proceedings of the 2010 Workshops on Database and Expert systems Applications* (pp. 54–58). 10.1109/DEXA.2010.32

Watson, I. (1997). *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufman.

WEF (World Economic Forum). (2021). *Shaping the future of IoT and urban transformation*. Retrieved January 25, 2021, https://www.weforum.org/platforms/shaping-the-future-of-the-internet-of-things-and-urban-transformation

Westraadt, L., & Calitz, A. (2020). A modelling framework for integrated smart city planning and management. *Sustainable Cities and Society*, *63*, 102444. doi:10.1016/j.scs.2020.102444

Wetherill, O., Lee, C., & Dietz, V. (2017). Causes of Poor Immunisation Data Quality and Proven Interventions : A Systematic Literature Review. *Ann Infect Dis Epidemiol*, *2*, 1–7.

WHO. (2006). *The Immunization Data Quality Self-assessment (DQS) Tool*. World Health Organization.

WHO. (2006). *Vaccine Preventable Diseases and the Global Immunisation vision and Strategy, 2006-2015*. WHO.

WHO. (2017). Data quality review: a toolkit for facility data quality assessment Module 2. Desk review of data quality. Geneva: WHO.

Williams, I., Falch, M., & Tadayoni, R. (2020). The proposed Baltic Sea Region Cross-border Business Registration Service. In *2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges.* IEEE.

Williams, I. (2015). *Analysis of Public Private Interplay Frameworks in the Development of Rural Telecommunications Infrastructure: A Multiple-Case Study*. Aalborg University Press.

Williams, I. (2020). E-Government, Yesterday, Today and in the Future. In A. Sieben (Ed.), *E-Government: Perspectives, Challenges and Opportunities*. Nova Publishers.

Williams, I. (2020). The Robots are here. In I. Williams (Ed.), *Contemporary applications of Actor Network Theory* (pp. 51–75). Palgrave Macmillan.

Williams, I. (2021). Mobilization - Decision Theory. In *M. Khosrow-Pour, Encyclopedia of Information Science and Technology* (5th ed., vols. 1–3). IGI Global.

Williams, I., Falch, M., & Tadayoni, R. (2018). Internationalization of e-Government Services. In *11th CMI International Conference: Prospects and Challenges Towards Developing a Digital Economy within the EU* (pp. 19-31). Copenhagen Denmark: IEEE.

Witten, I. H., Paynter, G. W., Frank, E., & Gutwin, C., & NevillManning, C. G. (1999). KEA: practical automatic keyphrase extraction. In *Proceedings of the Fourth ACM Conference on Digital Libraries* (pp. 254–255) 10.1145/313238.313437

Woodhouse, T. (2020). *Mobile devices are too expensive for billions of people and it's keeping them offline.* Available at: https://a4ai.org/mobile-devices-are-too-expensive-for-billions-of-people-and-its-keeping-them-offline/

Woodhouse, T., & Thakur, D. (2018). *Improving Mobile Broadband Quality of Service in Low- and Middle-Income Countries*. Web Foundation.

World Economic Forum. (2017). *The Future of Jobs and Skills in Africa: Preparing the Region for the Fourth Industrial Revolution World Economic Forum.* Geneva: World Economic Forum.

World Health Organization. (2012). *Global vaccine action plan 2011-2020*. World Health Organization.

Wouters, B., Deridder, D., & Van Paesschen, E. (2000). The Use of Ontologies as a Backbone for Use Case Management. *Proceedings of the European Conference on Object Object-Oriented Programming ECOOP, Workshop: Object and Classifications, a Natural Convergence*.

WSDL-S. (2005). http://www.w3.org/Submission/WSDL-S/

Wu, Y.-B., Li, Q., Bot, R. S., & Chen, X. (2005). Domain-specific keyphrase extraction. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 283–284). ACM.

Xiong, A., & Guo, Q. (2019). Chinese news keyword extraction algorithm based on textrank and topic model. In *International Conference on Artificial Intelligence for Communications and Networks* (pp. 334–341). 10.1007/978-3-030-22968-9_29

Xu, L., Qiu, J., Gu, W., & Ge, Y. (2020). The Dynamic Effects of Perceptions of Dread Risk and Unknown Risk on SNS Sharing Behavior During Emerging Infectious Disease Events: Do Crisis Stages Matter? *Journal of the Association for Information Systems*, *21*(3), 545–573. doi:10.17705/1jais.00612

Yadava, M., Kumar, P., Saini, R., Roy, P. P., & Prosad Dogra, D. (2017). Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications*, *76*(18), 19087–19111. doi:10.100711042-017-4580-6

Yang, D. (1996). *The Technology of Japanese Imperialism: Telecommunications and Empire-Building, 1895-1945*. Harvard University.

Yaqoob, I., Ahmed, E., Hashem, I. A. T., Ahmed, A. I. A., Gani, A., Imran, M., & Guizani, M. (2017). Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges. *IEEE Wireless Communications*, *24*(3), 10–16. doi:10.1109/MWC.2017.1600421

Yawson, A. E., Bonsu, G., Senaya, L. K., Yawson, A. O., Eleeza, J. B., Awoonor-Williams, J. K., Banskota, H. K., & Agongo, E. E. (2017). Regional disparities in immunization services in Ghana through a bottleneck analysis approach: Implications for sustaining national gains in immunization. *Archives of Public Health*, *75*(1), 1–10. doi:10.118613690-017-0179-7 PMID:28293422

Yeboah-Boateng, E. O., & Nwolley, S. J. (2019). Developing Use Cases of Big Data Analytics: Data Integration with Social Media Metrics. In K. Skouby, I. Williams & A. Gyamfi (Eds.), Handbook on ICT in Developing Countries: Next Generation ICT Technologies. Rivers Publishers.

Yegge, W. M. (1996). *A Basic Guide for Valuing a Company*. Wiley.

Yeh, Y. (2011). The implementation of knowledge management system in Taiwan's higher education. *Journal of College Teaching and Learning*, *2*(9). Advance online publication. doi:10.19030/tlc.v2i9.1861

Yu, H., Sun, Q., Sheng, K., & Wang, Z. (2016). Intelligent street lamp control system using ZigBee and GPRS technology. *International Journal of Simulation-Systems, Science & Technology, 17*(35), 36.1–36.13.

Yu, Y., & Ng, V. (2018). *Wikirank: Improving keyphrase extraction based on background knowledge*. arXiv preprint arXiv:1803.09000.

Yuichiro, M., Ogawa Yuhei, O., Akatsuki, H., & Takahira, Y. (2014). Multi-robot Coordination Based on Ontologies and Semantic Web Service. *Knowledge Management and Acquisition for Smart Systems and Services*, 150-164.

Zaidan, A. A., Zaidan, B. B., Qahtan, M. Y., Albahri, O. S., Albahri, A. S., Alaa, M., Jumaah, F. M., Talal, M., Tan, K. L., Shir, W. L., & Lim, C. K. (2018). A survey on communication components for IoT-based technologies in smart homes. *Telecommunication Systems*, *69*(1), 1–25. doi:10.100711235-018-0430-8

Zesch, T., & Gurevych, I. (2009). Approximate matching for evaluating keyphrase extraction. In *Proceedings of the International Conference RANLP-2009* (pp. 484–489). Academic Press.

Zhang, C. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computer Information Systems*, *4*(3), 1169–1180.

Zhang, C., & Xu, H. (2009). Using citation-knn for automatic keyword assignment. In *Proceedings of the 2009 International Conference on Electronic Commerce and Business Intelligence* (pp. 131–134). 10.1109/ECBI.2009.25

Zhang, Q., Wang, Y., Gong, Y., & Huang, X.-J. (2016). Keyphrase extraction using deep recurrent neural networks on twitter. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 836–845). 10.18653/v1/D16-1080

Zhang, W. Y., Cai, M., Qiu, J., & Yin, J. W. (2009). Managing distributed manufacturing knowledge through multi-perspective modelling for Semantic Web applications. *International Journal of Production Research*, *47*(23), 6525–6542.

Zhang, Z., Gonzalez, M. C., Morse, S. S., & Venkatasubramanian, V. (2017). Knowledge Management Framework for Emerging Infectious Diseases Preparedness and Response: Design and Development of Public Health Document Ontology. *JMIR Research Protocols*, *6*(10), e196. doi:10.2196/resprot.7904 PMID:29021130

Zhou, Z. (2020). An intelligent teaching assistant system using deep learning technologies. *ACM International Conference Proceeding Series*, 18–22. 10.1145/3409073.3409079

Zhu, A. (2020). *Application of AI Identification Technology in Foreign Language Education.* doi:10.1109/ICAIE50891.2020.00024

Zibi, G. (2018). *Closing the Investment Gap: How Multilateral Development Banks Can Contribute to Digital Inclusion.* Washington, DC: A4AI, Web Foundation, and Xalam Digital Analytics.

Zu, X., Xie, F., & Liu, X. (2020). Graph-based keyphrase extraction usingword and document embeddings. In *Proceedings of the 2020 IEEE International Conference on Knowledge Graph* (pp. 70–76). 10.1109/ICBK50248.2020.00020

About the Contributors

Albert Gyamfi has a PhD in Knowledge Management from Aalborg University, Denmark. He is a consultant and an expert who has researched into media richness and web applications for knowledge management process and organisational learning. He is currently an Instructor (Artificial Intelligence and Data Analytics Message) at Saskatchewan Polytechnic. He also worked at the University of Regina as a Graduate Teaching assistant. He is also attached to Centre for Communication, Media and Information Technologies (CMI), Aalborg University Copenhagen as a guest researcher

Idongesit Williams is an Assistant Professor at Aalborg University Copenhagen. He holds a Bachelor in Physics, a Master degree in Information and Communications Technologies and a Ph.D. He has since 2010 researched into Knowledge Management, Organizational Learning, socioeconomic, sociotechnical related to Information and Communications Technologies. He has authored and co-authored more than 30 research publications, including journal papers, books, book chapters, conference papers and magazine articles. He is the co-editor of the Books, The African Mobile Story, and the Handbook on ICT for developing countries:5G perspectives. He has organized conferences like the CMI annual conference and the CMI/GTUC conferences.

* * *

Samuel Agbesi is a PhD fellow at the Center for Communication, Media and Information technologies (CMI), Aalborg University Copenhagen. He received his Master of Philosophy in Information Technology from Kwame Nkrumah University of Science and Technology (KNUST). Samuel has over eight (8) years experience in Information Technology with specialization in database administration, sql pl/sql development, optimization, performance tuning and support for varied flavors (MySQL, Oracle 9i, 10g) of databases and Computer application programming with .Net framework (VB and C# .Net), Java and Web application development, and his research area is in e-government.

Joseph Tei Boye-Doe is a Public Health Officer (Health Information) who has been working with the Greater Accra Regional Health Directorate of the Ghana Health Service (GHS) over a decade. He completed his MPH at the Catholic University College of Ghana, Fiapre. His Area of research interest is in the application of information technology, Health Information Systems and the use of health information management to improve health care delivery services. In recent years he has been working with both local and international organizations such as GHS Policy Planning Monitoring and Evalua-

tion, USAID Systems for health, USAID SHOPS, among others to improve the quality of health data used to make decisions.

William Leslie Brown-Acquaye is a Senior Lecturer and currently the Head of Department at the Information Systems & Technology department, Ghana Communication Technology University (GCTU).

Fevzi Esen is working as assistant professor of information technologies in University of Health Sciences, Turkey.

Silvia Fernandes is Assistant Professor at the Faculty of Economics, University of Algarve. She holds a PhD and is member of Cinturs (Research Centre for Tourism, Sustainability and Well-being). She lectures Information Technology and Innovation Economics in undergraduate and master courses. She has several publications in books, journals, and communications in international conferences in themes such as: information systems, tourism development, innovation, etc.

Silvia Fernandes assisted process analysis and development in several enterprises and has a master degree in Management from the Faculty of Economics, University of Algarve.

Furkan Goz received his B.S. degree in Computer Engineering from Anadolu University and M.S. degree in Computer Engineering from Kocaeli University. He is currently pursuing the Ph.D. degree in Computer Engineering department at Kocaeli University. His research interests include graph databases, and graph mining.

Forgor Lempogo is a Senior Lecturer and currently the Head of Department at the Emerging Computing Technologies department, Ghana Communication Technology University.

Paula Martins is assistant professor at the Faculty of Science and Technology, University of Algarve. She holds a PhD from IST (Lisbon) and is member of Cinturs (Research Centre for Tourism, Sustainability and Well-being). She lectures Computer systems in undergraduate and master courses. She has several publications in books, journals, and communications in international conferences in themes related with computer systems, process management, etc.

Vardan Mkrttchian graduated from State Engineering University of Armenia, the Faculty of Cybernetics as an Electronics Engineer & received his Doctorate of Sciences (Engineering) in Control Systems from Lomonosov Moscow State University (former USSR). Dr. Vardan Mkrttchian taught for undergraduate and graduate student's courses of control system, information sciences and technology, at the Astrakhan State University (Russian Federation), where he was is the Professor of the Information Systems (www.aspu.ru) six years. Now he is full professor CAD&Economy departments of Penza State University (www.pnzgu.ru). He is currently chief executive of HHH University, Australia and team leader of the international academics (www.hhhuniversity.com). He also serves as executive director of the HHH Technology Incorporation. Professor Vardan Mkrttchian has authored over 400 refereed publications. He is the author of over twenty books published of IGI Global, included ten books indexed of SCOPUS in IT, Control System, Digital Economy, and Education Technology, and 43 works in SCOPUS and 47 in WoS. He also is Editor-in-Chief of International Journal of Applied Research in Bioinformatics (IJARB).

Alev Mutlu received his BS degree from Kocaeli University and Ph.D. degree from Middle East Technical University, both in computer engineering. He has spent a year at Case Western Reserve University, OH, USA as a researcher, and is currently with Kocaeli University, Kocaeli, Turkey. His research focuses on multi-relational data mining with particular interest in concept discovery and concept discovery in graph databases.

Christiana Amoanteng is a Lecturer, Department of Nursing, University of Cape Coast.

Kamalendu Pal is with the Department of Computer Science, School of Mathematics, Computer Science and Engineering, City University London. Kamalendu received his BSc (Hons) degree in Physics from Calcutta University, India, Postgraduate Diploma in Computer Science from Pune, India; MSc degree in Software Systems Technology from Sheffield University, Postgraduate Diploma in Artificial Intelligence from Kingston University, MPhil degree in Computer Science from University College London, and MBA degree from University of Hull, United Kingdom. He has published widely in the scientific community with research articles in the ACM SIGMIS Database, Expert Systems with Applications, Decision Support Systems, and conferences. His research interests include knowledge-based systems, decision support systems, blockchain technology, software engineering, and service-oriented computing. He is a member of the British Computer Society, the Institution of Engineering and Technology, and the IEEE Computer Society.

Serkan Polat is working as assistant professor of tourism in Istanbul Medeniyet University, Turkey.

Fati Tahiru is a Lecturer at the Department of Computer Science, Ho Technical University, Ghana. As an expert in Technology adoption in government institutions, her research focuses on Artificial intelligence (AI) in Education, E-Government and Cyber Security. She has over 10 years of experience in the Telecommunication sector, holding the position of Project Manager. Her current research is on the application of AI in Education, Learning Analytics and Blockchain Technology.

Viacheslav Voronin is the head of the Center for Cognitive Technology and Machine Vision at Moscow State University of Technology "STANKIN," Moscow, Russian Federation. He received his BS (2006), MS (2008) in the communication system from the South-Russian State University of Economics and Service, and his Ph.D. in technics from Southern Federal University (2009). Voronin is a member of the Program Committee of the conference SPIE. His research interests include image processing, inpainting, and computer vision.

Ezer Yeboah-Boateng is a Telecoms Engineer and professional ICT Specialist with a strong background in cyber-security, digital forensics, business development, change management, knowledge management, strategic IT-enabled business value creation, digital transformation and capabilities to develop market-oriented strategies aimed at promoting growth and market share. An Executive with over 25 years of domestic and global experience conceptualizing ideas, seizing opportunities, building operations, leading highly successful new business development initiatives and ventures. A Consultant with emphasis on Cyber-security, digital forensics, telecommunications, Internet, and network integration technologies, with additional experience related to dealing with refurbished and overstock equipment, and manufacturing. Strong presentation, negotiation, and team building skills. Ezer currently serves a

member of the Electronic Communications Tribunal in Ghana and a governing board member of the National Information Technology Agency (NITA).

Index

M

N

O

P

R

S

T

U

V

W