

**WestminsterResearch**

<http://www.westminster.ac.uk/westminsterresearch>

**Semantic Selection of Internet Sources through SWRL Enabled  
OWL Ontologies**

**Almarri, H.**

This is an electronic version of a PhD thesis awarded by the University of Westminster.  
© Miss Hamda Almarri, 2017.

---

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

---

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail [repository@westminster.ac.uk](mailto:repository@westminster.ac.uk)

**Semantic Selection of Internet Sources  
through SWRL Enabled OWL  
Ontologies**

**Hamda Mohammad O. S. Binghubash  
Almarri**

**University of Westminster**

**June 2017**

A thesis submitted in partial fulfilment of the requirements of the University of  
Westminster for the degree of Doctor of Philosophy

## ***ABSTRACT***

This research examines the problem of Information Overload (IO) and give an overview of various attempts to resolve it. Furthermore, argue that instead of fighting IO, it is advisable to start learning how to live with it. It is unlikely that in modern information age, where users are producer and consumer of information, the amount of data and information generated would decrease. Furthermore, when managing IO, users are confined to the algorithms and policies of commercial Search Engines and Recommender Systems (RSs), which create results that also add to IO. this research calls to initiate a change in thinking: this by giving greater power to users when addressing the relevance and accuracy of internet searches, which helps in IO. However powerful search engines are, they do not process enough semantics in the moment when search queries are formulated. This research proposes a semantic selection of internet sources, through SWRL enabled OWL ontologies. the research focuses on SWT and its Stack because they (a) secure the semantic interpretation of the environments where internet searches take place and (b) guarantee reasoning that results in the selection of suitable internet sources in a particular moment of internet searches. Therefore, it is important to model the behaviour of users through OWL concepts and reason upon them in order to address IO when searching the internet. Thus, user behaviour is itemized through user preferences, perceptions and expectations from internet searches. The proposed approach in this research is a Software Engineering (SE) solution which provides computations based on the semantics of the environment stored in the ontological model.

*First and foremost, praises and thanks to Allah almighty, for His showers of blessings and lightning of my way through my research work to complete this research successfully*

*"I am so grateful for having the drive to succeed..."*

*This PhD is dedicated to my deceased mother and to my farther....*

*For all the comfort, caring and honest concern from the first day to her last breath;*

*When I felt that I reached my limits and she said "I understand";*

*To my mom who listened to all my complaints with patience;*

*I would like to thank her for believing in me and not giving up.*

*To my mom and my sister Asma, with whom I shared the happiest moments of my life in  
London.*

*I'm regretful that I did not finish my studies sooner to draw a smile on your face, mom...*

## ACKNOWLEDGEMENTS

*I'd like to sincerely thank almighty Allah for all his grants that he bestowed upon me.*

*To my dear deceased mother and my father,*

*I would like to extend my sincere thankfulness, for their great role in my life, their continued encouragement and numerous sacrifices for me to achieve my goals and for how far I've come.*

*To my dad who believed in me even when I gave up....*

*Thank you for all the love and support*

*I do not know how to put into words my gratitude for all valuable time, patience, and support my mother and sisters dedicated during the period of my studies. I cannot imagine how it could have been without my family when I think of all the obstacles I went through to complete my studies.*

*I'm deeply grateful for the continuous support from every single member of my family.*

*I am also extremely appreciative, for the love and generosity my sisters have showed me, and all the happy memories I had with them in the past four years.*

**IT IS IMPOSSIBLE WITHOUT YOU MY DEAR SISTERS**

*Special thanks to my sister Ayesha for spending eight months with me in London to help and support me when her only son needed her the most.*

*Khalifa, YOU ARE MY SUNSHINE. I owe you 100 wishes.*

*I appreciate all the advice I received from my sister Fatma. She answered all my questions patiently every time I needed an opinion. Her comments helped me to re-think and reshape my ideas in many aspects of my studies.*

*I would like to thank my sister Moza, who offered her help and concern on every matter or problem I faced during my stay in London. Who accompanied me to my first conference in my PhD journey.*

*Thank you very much...It was an unforgettable time ...*

*To my sister Asma, who spent four years with me in London, accompanied me to conferences in Berlin and Brazil, saved me a lot of time to focus on my research and never gave up on me every time I struggled during my studies.*

*Many thanks for my dear brothers Obaid and Essa for their support and for being truly brothers when I needed them...*

*To my family for making it possible with gratitude and appreciation...*

*I am so blessed to have a truly warm and loving family....*

*I love you all and I wish I will have a chance to return as much as I can of all the support I received during my studies...*

*Sincerely, Hamda*

*I would like to thank my government who sponsored me and gave me a great chance to learn from experienced people.*

*I wish I can give back as much as I can.*

*My dear friend Eiman, you are great, we have laughed and cried many times together...*

*I wish you all the best, my dear friend....*

*I would also like to thank my previous director of studies, Dr Radmila Juric, who supervised me before her retirement. I enjoyed all the time I spent working with you. You were a great support and you made things possible. I learned a lot from you. Thank you for giving me the chance to show my abilities...*

*I wish you a healthy life...*

*Special thanks to Dr. Alexander Bolotov, who took charge and supervised me for the remaining period of my studies. I can't express how grateful I am for your understanding and advice...*

*I also wish to thank Dr. Reza, my friend and colleague who became my internal examiner. Thank you for all the advice.*

*Dr Taj Keshavarz and Dr Andrzej Tarczynski thank you for helping me solve many small and big problems at the university...*

*Finally, I would like to thank my friend Awajimam for her help during the period of my write up.*

## ***AUTHOR'S DECLARATION***

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..Hamda Binghubash Almarri.....      DATE:....21<sup>st</sup> June 2017.....

## ***TABLE OF CONTENTS***

Abstract.....	i
Acknowledgements.....	iii
Author’s declaration .....	v
Table of Contents.....	vi
List of Publications .....	x
Conference Papers .....	x
Journal Paper Publications.....	x
List of Tables .....	xi
List of Figures.....	xii
List of Abbreviations .....	xiv
CHAPTER 1. Introduction.....	1
1.1 Overview of the research methodology .....	4
1.2 Research Motivation .....	4
1.3 Research objectives.....	5
1.4 Contributions .....	8
1.5 Research Limitations .....	9
1.6 Research Methodology .....	10
1.7 Thesis Structure .....	13
CHAPTER 2. Literature Review.....	15
2.1 Information Overload .....	17
2.1.1 Overview of Information Overload .....	17
2.1.2 Concepts of Information .....	18
2.1.3 Knowledge Overload .....	20
2.1.4 Technology Overload .....	23
2.1.5 Communication Information Overload .....	25

2.1.6	Information Overload in Businesses and organizations .....	27
2.1.7	Information Overload in Healthcare .....	31
2.1.8	Modern Information Overload.....	34
2.1.9	Information Overload or Underload .....	37
2.1.10	Summary.....	39
2.2	Information Retrieval Systems .....	41
2.3	Recommender Systems .....	44
2.3.1	Overview of Recommender Systems .....	45
2.3.2	Recommender Systems Techniques .....	51
2.3.2.1	Collaborative Filtering .....	52
2.3.2.2	Content-Based Filtering .....	56
2.3.2.3	Hybrid Filtering.....	57
2.3.2.4	Tagging / Annotations and Folksonomies.....	59
2.3.3	Summary.....	67
2.4	Search Engines.....	69
2.4.1	Overview of Search Engines .....	69
2.4.2	History of Search Engines .....	71
2.4.3	Search Engines Techniques .....	74
2.4.4	User Search Behaviour on the Internet.....	84
2.4.5	Summary.....	93
2.5	the Semantic Web Technology .....	98
2.5.1	Ontology definition.....	101
2.5.2	OWL Ontology .....	103
2.5.3	Semantic Search Engine .....	105
2.6	Chapter Summary .....	111

CHAPTER 3. Related Work .....	113
3.1 Accuracy and Relevance in Search Queries .....	114
3.2 Results of Recommender Systems.....	119
3.3 Semantic Web Solutions:.....	120
3.4 Chapter Summary .....	125
CHAPTER 4. The Proposal .....	126
4.1 establishment what the Problems are.....	126
4.2 The Proposed Computational Model .....	133
4.3 Generic Model- Overview of the Proposed Method.....	140
4.3.1 The Ontological Model.....	145
4.3.2 Constraints in the OWL Model .....	148
4.3.3 Semantic Overlapping in the OWL Model and the Reasoning Process .....	153
4.3.4 Restrictions imposed on the OWL ontological model: .....	155
4.4 SWRL rules and reasoning process.....	158
4.5 Examples of Moments/situations (OWL Model).....	162
4.6 Chapter Summary .....	163
CHAPTER 5. Illustration and Implementation of the Proposed Model .....	165
5.1 Illustration of the Proposed Model .....	167
5.1.1 The Domain of LLL in Healthcare .....	167
5.1.2 Case Study: Selection of Online Sources for a Medical Student .....	170
5.1.2.1 The Ontological Model for the Case Study.....	171
5.1.2.2 OWL Constraints .....	177
5.1.2.3 SWRL Rule for the Case Study .....	182
5.1.3 Discussion on the illustration of the generic model .....	185
CHAPTER 6. EVALUATION, CONCLUSION AND FUTURE WORK .....	187

6.1	Review of Research Objectives .....	187
6.2	Research Evaluation/ Impact .....	194
CHAPTER 7. Conclusions and Future Work .....		201
7.1	Research Conclusion.....	201
7.2	Future Work.....	202
Appendix.....		204
Reference .....		215

## ***LIST OF PUBLICATIONS***

The work described in this thesis has been presented in the following publications:

### **CONFERENCE PAPERS**

1. BINGHUBASH, H. & JURIC, R. 2011. Ontology Based Recommendation of Social Networks According to the semantics of Member's Requests. *In Proceedings of 16th International Conference of the Society for Design and Process Science (SDPS '11)* Jeju Island, South Korea.
2. ALMARRI, H. B., RAHMAN, T. & JURIC, R. 2012. Semantic Recommendation of Information Sources for Lifelong Learning. *In Proceedings of 16th International Conference of the Society for Design and Process Science (SDPS' 12)* Berlin, German.
3. JURIC, R., ALMARRI, H. B., ARNTZEN, A. A., SUH, S. C. & KONJHODZIC, A. 2013. Towards Dynamic Creation of Interdisciplinary Curricula. *In Proceedings of the 18th International Conference of Society for Design and Process Science (SDPS '13)*. 27th -31st October 2013.
4. MAHMOOD, S., ALMARI, H. B., JURIC, R. & KIM, I. 2013. Extracting Tumblr Posts Through Ontological Reasoning for Detecting Alarming Suicidal Notes. *In Proceedings of the 18th International Conference of Society for Design and Process Science (SDPS '13)*. Campinas, São Paulo, Brazil.
5. ALMARRI, B. H. & JURIC, R. 2013. Generic OWL Enabled Model for the Selection of Online Learning Sources. *In Proceedings of 17th International Conference of the Society for Design and Process Science (SDPS' 13)* São Paulo, Brazil.
6. ALMARRI, H. B. & JURIC, R. 2014. Modern Information Overload. *In Proceedings of 18th International Conference of the Society for Design and Process Science (SDPS' 14)*. Kuching, Sarawak, Malaysia.
7. ALMARI, H. B. & RADMILA, J. 2015. Modeling User Behaviour When Addressing Information Overload. *In Proceedings of 19th International Conference of the Society for Design and Process Science (SDPS' 15)*. Dallas, TX, USA.
8. ALMARRI, B. H., RADMILA, J. & MUGHAL, B. 2016. Semantic Selection of Healthcare Apps. *In Proceeding of the 49th Hawaii International Conference on System Science (HICSS 49)* Kauai, Hawaii, USA.

### **JOURNAL PAPER PUBLICATIONS**

1. ALMARRI, H. B., RAHMAN, T., JURIC, R. & PARAPADAKIS, D. 2013. Semantic Recommendation of Information Sources for Lifelong Learning. *Journal of Integrated Design and Process Science*, 17, 55-78.

## *LIST OF TABLES*

Table 2.1 Overview of Search Engines.....	96
Table 2.2 Semantic Search Engines.....	110
Table 4.1 the truth table for $p \rightarrow q$ .....	160
Table 5.1 Excerpts from the ontology: a selection of Characteristics Classes and Individuals.....	176
Table 5.2 THE RELATIONSHIPS BETWEEN SOURCES CLASS INDIVIDUALS (RELAXDOC) AND ANY CSI CLASSES INDIVIDUALS THROUGH OBJECT PROPERTY.....	179
Table 5.3 The Relationships between SOURCES class individuals (SurgeryTech) and ANY CSi Classes individuals through Object property.....	180
Table 5.4 The Relationships between SOURCES class individuals (BioCrowd) and ANY CSi Classes individuals through Object property.....	181

## *LIST OF FIGURES*

Figure 2.1The Hierarchy of Data, Information and Knowledge .....	19
Figure 2.2 THREE MAIN ELEMENTS OF THE PROPOSED NSRS MODEL IN (ULLMAN, 2012) .....	48
Figure 2.3MAIN COMPONENTS OF RATING-BASED CF RECOMMENDATION MODEL (HU AND PU, 2011) .....	55
Figure 2.4 DEPICTS THE MAIN CONCEPTS AND INSTANCES OF SPREADR MODEL. ....	63
Figure 2.5 Main elements of the re-ranking algorithm proposed in (Chidlovskii et al., 2000) .....	79
Figure 2.6 USER ENGAGEMENT MODEL PROPOSED IN (SHORT ET AL., 2015) .....	89
Figure 2.7 THE SEMANTIC WEB TECHNOLOGY STACK (SWT-STACK, 2008) ..	99
Figure 2.8 THE USE OF RDFS TO DESCRIBE RESOURCE(S) CLASSES AND PROPERTIES (W3C-RDFS, 2002) ..	100
Figure 2.9 Approaches to semantic web (Sudeepthi et al., 2012) ..	107
Figure 2.10 Depicts the Process of Providing Search Results in Powerset SSE. ....	108
Figure 3.1 MULTI VIEW RECOMMENDATION ENGINE PROCESS (OUFAIDA AND NOUALI, 2009) .....	123
Figure 4.1 Information Sources on the Internet.....	139
Figure 4.2 The Universal Set of Information Sources on the Internet.....	141
Figure 4.3 describes some of the characteristics an information source can have through has-Purpose and has-Features relationship. ....	142
Figure 4.4 shows the relation between Domain elements and Range elements.....	143
Figure 4.5 THE GENERIC MODEL WITH CONSTRAINTS.....	147
Figure 4.6 GENERIC MODEL WITH OBJECT PROPERTY AS CONSTRAINTS...148	
Figure 4.7 CHOICE OF CONSTRAINTS FOR EACH CSK AND SOURCES. ....	150

Figure 4.8 CHOICE OF CONSTRAINTS IN THE GENERIC MODEL.....	151
Figure 4.9 GENERIC MODEL WITH HORIZONTAL HIERARCHIES AND THEIR CONSTRAINTS.....	152
Figure 4.10 Proposed Generic Model (The Reasoning Process with Semantic Overlapping).....	155
Figure 4.11 The Restriction "has-Features" some Features Indicates that Individuals of the First class has at least one Feature or more.....	157
Figure 5.1 THE ONTOLOGICAL MODEL BASED ON FIGURE 4.10. AND DERIVED FROM THE SCENARIO IN the CASE STUDY.....	172
Figure 5.2 DETAILING THE ONTOLOGICAL MODEL FROM FIGURE 5.1.....	173
Figure 5.3 Individuals of the SOURCES Class for the Case Study	177
Figure 5.4 Constraints Imposed on the Ontology for Case Study 1 .....	178
Figure 5.5 SWRL Rule 1: Selection of Online Sources for the Case Study.....	183
Figure 5.6 Result of Running Rule 1 for the Case Study.....	183
Figure 5.7 SWRL Rule 2 for Selecting Technology.....	184
Figure 5.8 First Part of Rule 2 Result for the Case Study. ....	184
Figure 5.9 First Part of Rule 2 Result for the Case Study. ....	184
Figure 6.1 semantic overlapping through individuals.....	200

## ***LIST OF ABBREVIATIONS***

<b>(AA)</b> Application Agents	<b>(DS)</b> Distributed Search
<b>(ACA)</b> Anonymous Contributions Architecture	<b>(FOAF)</b> Friend Of A Friend
<b>(ACF)</b> Automated Collaborative Filtering	<b>(FTP)</b> File Transfer Protocol
<b>(ACM)</b> Association for Computing Machinery	<b>(GB)</b> Graph-Based Recommender
<b>(AI)</b> Artificial Intelligence	<b>(GNN)</b> Graph Neural Network
<b>(ANA)</b> American Nurses Association	<b>(GP)</b> General Practitioners
<b>(API)</b> Application Program Interface	<b>(GSR)</b> Generic Search Results
<b>(AS)</b> Automated Scripts	<b>(GUI)</b> Graphical User Interface
<b>(B2C)</b> Business To Consumer	<b>(HB)</b> Human Behaviour
<b>(BSR)</b> Branded Search Results	<b>(HTML)</b> Hyper Text Markup Language
<b>(CARS)</b> Context-Aware recommender systems	<b>(HTTP)</b> Hyper Text Transfer Protocol
<b>(CBF)</b> Content Based Filtering	<b>(IBL)</b> Involved Brand Loyalists
<b>(CF)</b> Collaborative Filtering	<b>(IBS)</b> Uninvolved Brand Switchers
<b>(CQ)</b> Competency Question	<b>(IDE)</b> Integrated Development Environment
<b>(CS)</b> Computer Science	<b>(IDF)</b> Inverse Document Frequency
<b>(CT)</b> Collaborative Tagging	<b>(IIS)</b> Involved Information Seekers
<b>(CTR)</b> Click-Through Ranking	<b>(IPC)</b> Personalized Information Resource Cloud
<b>(DB)</b> Databases	<b>(IPR)</b> Popularity Ranking
<b>(DL)</b> Description Logic	<b>(IRS)</b> Information Retrieval System
<b>(DM)</b> Data Mining	<b>(IRI)</b> Internationalized Resource Identifier
	<b>(IS)</b> Information System

<b>(ISB)</b> Information Seeking Disciplines	<b>(RDFS)</b> Resource Description Framework Schema
<b>(IT)</b> Information Technology	<b>(RR)</b> Recency Ranking
<b>(IU)</b> Information Underload	<b>(RS)</b> Recommender Systems
<b>(KB)</b> Knowledge-Based	<b>(SA)</b> Spreading Activation
<b>(KMS)</b> Knowledge Management System	<b>(SB)</b> Search Behaviour
<b>(LDA)</b> Latent Dirichlet Allocation	<b>(SCF)</b> Shard Collaborative Filtering
<b>(LLL)</b> Lifelong Learning	<b>(SE)</b> Software Engineering
<b>(MA)</b> Management Agent	<b>(SEO)</b> Search Engine Optimization
<b>(MAS)</b> Multi-Agent System	<b>(SIOC)</b> Interlinked Online Communities
<b>(MIS)</b> Management Information Systems	<b>(SKOS)</b> Simple Knowledge Organization Scheme
<b>(NFL)</b> Non-Formal Learning	<b>(SL)</b> System Logs
<b>(NG)</b> Next-generation	<b>(SOAP)</b> Simple Object Access Protocol
<b>(NLSE)</b> Natural Language Search Engine	<b>(SPA)</b> Secure Processing Architecture
<b>(OBA)</b> Online Behavioural Advertising	<b>(SQL)</b> Structured Query Language
<b>(OSR)</b> Organic Search Results	<b>(SSEs)</b> Semantic Search Engines
<b>(OSRD)</b> Office of Scientific and Development	<b>(ST)</b> Social Tagging Systems
<b>(OWL)</b> Web Ontology Language	<b>(SW)</b> Semantic Web
<b>(PCC)</b> Pearson Correlation Coefficient	<b>(SWRL)</b> Semantic Web Rule Language
<b>(PCE)</b> Prevised Computing Environment	<b>(SWSE)</b> Semantic Web Search Engine
<b>(PPC)</b> Pay-Per-Click	<b>(SWT)</b> Semantic Web Technology
<b>(PSR)</b> Paid Search Results	<b>(TF)</b> Term Frequency
<b>(PTR)</b> Personalised Tag Recommendation	<b>(TFIDF)</b> Term Frequency and Inverse Document Frequency
<b>(QL)</b> Query Language	<b>(TMS)</b> Tapestry Mail System
<b>(RBB)</b> Routine Brand Buyers	
<b>(RDF)</b> Resource Description Framework	

**(TSE)** Traditional Search Engines

**(UB)** User Behaviour

**(UIM)** User Interest Model

**(URI)** Uniform Resource Identifier

**(URL)** Uniform Resource Locator

**(VSS)** Vector Space Similarity

**(W3C)** World Wide Web Consortium

**(Web)** World Wide Web

**(WIRS)** Web Information Retrieval  
Systems

**(WWW)** World Wide Web

**(XML)** eXtensible Markup Language

# CHAPTER 1. INTRODUCTION

The problem of Information Overload (IO) (Bawden et al., 1999) is not new, it is a complicated issue that continually grows and becomes difficult to manage. From one century to another, IO manifested in many aspects of our daily lives. Throughout time, researchers have attempted to find solution for IO. However, these solutions were temporary and did not last long. The nature of IO has changed through time. Therefore, it is comprehensible that solutions to IO throughout time cannot be an answer to modern IO. Furthermore, most researches suggest to cope with IO, rather than try to eliminate it. Hence, we have to agree that IO is part of our lives and not a problem that cannot be solved.

Users are bombarded with excessive amount of search results and retrieved internet sources of information – by both search engines and RSs (Resnick and Varian, 1997 ). Thus, users feel confused and overwhelmed, and not sure which internet source is most relevant to their queries. The confusion is prolonged to the point that users are unsure whether the chosen information sources can offer relevant information or whether they are required to seek information from other sources. Furthermore, the huge list of retrieved information sources can negatively influence users' decision, in the sense that they are not sure when to stop reading and what is the right amount of information to satisfy their information needs.

Researchers addressed IO through RSs, which have been the focus of interest of the research community since late 90s. They have become famous because they deploy Collaborative Filtering (CF) and Content-Based Filtering (CBF) techniques (Goldberg et al., 1992) in order to guide users in their personalised way to interesting “*items*” in a large space of possible options.

Search Engines and their techniques (Seymour et al., 2011) were also exploited to understand the causes of IO. Studies indicate that information retrievals from the internet are worsened by either excessive amount of search engine results or caused by a mashup of search engine techniques to provide new search results. Thus, many researchers believe that a lot must be done in this area in order to remedy some of search engines search results problems.

The investigation of IO in this research indicates that it is necessary to initiate a shift in thinking when addressing modern IO. IO that is caused by advancements in technology on the internet and information seekers/users of these technologies in the process of creating and retrieving information on the internet. Hence, this research suggests that it is very important to first understand the semantics of the environment where the selection of information sources happen. Secondly, it is equally important to consider the information seeker/user's role and influence in the process of information creation on the internet. Thirdly, to provide "a moment" of internet searchers based on similarities between information seekers/user queries and characteristics of information sources on the internet.

In this research the environment can be defined as the problem domain which is investigated in the process of selecting information sources according to information seekers/users' queries. In the environment, different situations of information seeking can happen. Hence, it can change according to information seekers/users information needs. For example, early experiments in this research investigated information sources in the domain of social intensive environment on the internet in general (SDPS 2011). The investigation was extended to understand certain information needs of information seekers/users in the domain of healthcare (SDPS 2012). By modeling the environment and information seekers/user preference, the proposed computational model delivers "a moment" of internet searches. Each "moment" of internet searches can change based on the characteristics of the situation.

Therefore, this research proposes a generic model that will collect the semantics of information sources and user preferences on the internet to deliver most relevant search results according to the semantics of "a moment" of internet searches. The deployment of the proposed generic computational model utilises Semantic Web Technology (SWT) and its Stack (Horrocks et al., 2005). This is to reason upon the semantics of internet sources and information seeker/user preferences through SWRL (SWRL, 2004c) enabled OWL ontologies (OWL, 2004b).

The implementation of the proposed solution is a SE solution because:

1. Ontologies are defined as a formal specification of conceptualisation. This can be interpreted as perceiving a segment of the world which describe our

views of concepts and their interrelationship. Therefore, ontologies provide formal representation of consensus of domain of discourse(Gruber, 1995)

2. Ontologies plays an important role in SE because they can become a source of precisely defined concepts and provide shared understanding of certain domain of interest (Gruninger and Lee, 2002)
3. If a domain of interest in an ontology was well defined, then it can allow re-usability as present in formal ontologies.
4. Ontologies are largely used in the domain of Artificial intelligence (AI) (Shapiro, 1992)and knowledge representation and sharing.

The connection and role ontologies play in SE field is basically as follows:

- The relation of ontologies in the SE process is depicted in the ways they are used to reduce the complexity of a system such as business process, organisational structure, IT application and so forth.
- The distribution, reuse and integration of software components and systems is some of the priorities of SE issues present in the process of developing different types of ontologies.
- In SE solutions Entity Relationship (ER) modelling (Barker, 1990)and Object Oriented (OO) modelling (Jacobsen et al., 1992) were used to describe the relation between class and objects of diagrams. The same principles can be followed in the modelling of ontologies because both ER an OO allows the description of entities and association and includes the description of behaviour expression as well (Hesse, 2005).

There are different types of ontologies that are developed for certain reasons. Therefore, the categorisation of ontologies differs in the research community. For example high-level ontologies describe general concepts such as space, time material and objects. Domain ontologies provides vocabulary related to domain such as Information Systems (IS). Task ontologies describes vocabulary of tasks or activities and so forth. Some other ontologies groups the earlier mentioned ontologies which involves role of domain entities. Furthermore, there exist heavy and light weight ontologies which indicates the richness of internal structure (Ruiz and Hilera, 2006).

This chapter, provides an overview of the proposed solution in this research. Thus, it starts with a short description of the problem domain (see section 1.1), research motivation (see section 1.2), research objectives (see section 1.3), and contributions can be found in (see section 1.4). A list of the research limitations is present in (see section 1.5), section 1.6 describes the research methodology. This chapter is concluded with the thesis structure (see section 1.7).

## 1.1 OVERVIEW OF THE RESEARCH METHODOLOGY

An empirical analysis of IO and its different types was conducted in this research to understand its causes. This by investigating previously proposed solutions to reduce IO in environments where internet searches happen, such as education, business, and healthcare. The investigation also included modern IO caused by technological advancements. Thus, an itemisation of technologies was needed to understand the core of the problem. Based on the investigation, this research proposes a generic computational model that will collect the semantics of internet sources and users' preferences and provide most relevant search results on an ad-hoc basis.

## 1.2 RESEARCH MOTIVATION

This research was initially interested in investigating methods to discover information sources on the internet that can help information seekers/users who attempts to maintain their competency level to achieve certain lifelong goals. Unfortunately, it was very difficult to find learning sources on the internet required in various situations of lifelong learning (LLL) (Almarri et al., 2012b) (Almarri et al., 2013). Subsequently information seekers overload themselves with excessive amount of retrieved information sources, because they are not clear about what they need. Consequently, generate un required IO (situations that is caused by information seeker/user). The domain of LLL is a typical example of IO, where a learner in LLL environment is not able to get and choose relevant, or at least the best possible, learning sources on the internet, based on current mechanisms offered by search engines, including those customised.

This therefore was a motivation to study IO and the nature of the information sources the information seeker/user is surrounded with on the internet. This accompanied concerns of everlasting problems of IO and its nature, causes and environment because

IO is in information seekers/users everyday life. Consequently, information seekers/users are bombarded with wide range of problems in their professional, personal, or educational life on a daily basis, which results into IO. Thus, information seekers/users find themselves still confused and not able to make appropriate decision. Moreover, information seekers/users feel that they lack knowledge to solve problems in certain situations.

In all experiments in this research, IO manifested in search results in variety of situations. Furthermore, most of the attempts resulted into an appropriate search results because of either their irrelevance or the excess amount of them. Currently, information seekers/users first port on the internet is search engines. Information seekers/users submit to search engines simple questions which might create unintentional IO. If information seekers/users are not clear about what they want, then it is a typical consequence to suffer from IO. information seekers/users expect that search engines can understand the meaning behind their queries. Consequently, most of search engines problems manifested through their attempts to process search queries and utilise filtering and ranking techniques to deliver search results. Google's ranking (Su et al., 2010), for example, in all explorative studies, did not help and even created more problems in terms of judging how relevant search results were and how they are prioritised (Yong et al., 2008).

### 1.3 RESEARCH OBJECTIVES

- (OB 1) Propose (define and create) a computational model that will address IO in modern retrievals, i.e. internet searches and take into account the need to tailor the organization of search results, personalize search processes and influence the way algorithms are used to deliver search results. Hence, the model must clearly define what is needed for computations to be performed and which output is expected from it.
- (OB 2) Give more power to users. Because users are the one who are producers and consumers of information at the same time. Therefore, the proposed model must demonstrate how and where user take over in order to address the problems of selecting the most relevant search result in a particular moment of internet searches.

(OB 3) Address “a moment” in a particular situation when internet searches happen. This immediately implies that the proposed model might not be interested or involved with “the past”. It would build computations based on capturing as much semantics as possible in that “a moment”. The reason for that is simple, remembering the results of retrievals from previous “moments” might not be advisable because

- a. (a) each “moment” carries different semantics in terms of the reasons and needs for retrieval and
- b. (b) users very often change their mind while searching the internet.

Therefore, the proposed solution avoids storing the semantics of the results of past retrievals, because it may be wrong for the next “moment”.

(OB 4) The proposed generic model in this research utilise the SWT stack because it allows to interpret the meaning of a particular situation where retrievals happen. Furthermore, exploit a set of rich languages which gives the opportunity to reason upon the semantics of the situation in a particular internet search as mentioned in (OB 1). Therefore, SWT stack was used as a technology of choice when deploying models from (OB 1) above, because most of its components can be modelled through SWT. Obviously, the computational model deployed with the SWT would be a step forward in ensuring that the exact understanding and interpretation of the environment where internet searches happen: thus model the semantics of these environments, secure reasoning upon them, whilst taking into account objectives (OB 2) and (OB 3) above. This emphasises that the computational model, which is based on reasoning, will secure that the user receives the most relevant search results.

Consequently, SWRL enabled OWL ontologies might be suitable for implementing the computational model that could be deployed in any Integrated Development Environment (IDE) for Java and Android based platforms. A software application which houses the proposed computational model has been developed and tested. The issues, which are solely associated with its implementations, are as follows:

- (I). The feeding of the ontological model could be done through either

- (a). modern and intelligent interfaces (i.e. smart phone app which are built to collect user input) would allow information seekers/users to input their search query of the required information sources and infer/deliver information sources based on seeker/user preferences and the characteristics of information sources or
- (b). a set of drop-down menus in traditional User Interface (UI) that determine the format and content of user's inputs (search query) to extract information seekers/user preferences and reason upon the semantics of information sources.

For the purpose of testing the proposed generic computational model, a manual input of domain knowledge into the ontological model is followed. This to demonstrate that the proposed ideas and concepts work, but it is possible to direct the results of this research towards highly pervasive environments where voice and multimedia inputs are equally welcome as inputs planned in (a) and (b) above. In such cases, an interdisciplinary team is needed to combine multiple technologies and connect hardware and UI advances with application implementation.

The proposed model is solely placed upon internet search results in order to make them more relevant to the user's preference. However, it is important to assess how efficient the same model can be within any search engine in order to improve current ranking of search results. Therefore, it is suggested to use the model upon ranked results because it might allow the selection of more relevant results. However, what would happen if the result of ranking does not give a single relevant result? There will be nothing to select. This is an unusual situation, but it may happen. Therefore, the model should be completely effective as a part of any search engine or RSs.

(OB 5) Conduct an empirical investigation of the environment where the selection of internet sources happens. Furthermore, provide a chronological literature review of previous work in the areas of IO, Search Engine results, RSs and their techniques. The investigation also covers User Behaviour (UB) on the internet that influences searches on the internet.

## 1.4 CONTRIBUTIONS

- The novelty of the proposed solution in this research is in the automated selection of internet sources based on the semantics of the environment where selection happens. This research promotes decision making upon the most suitable internet source(s) based on reasoning process depicted in figure 4.6 (see section 4.3.3). One promising way of building such a tool would be to use the Semantic Web Technology stack (Horrocks et al., 2005) which produces a computational model based on SWRL (SWRL, 2004c) enabled OWL ontologies (OWL, 2004b) (see sections 5.1.3)
- The illustration of the proposed approach in this research is focused on specific domain situation where internet searches results into excessive amount of information sources.
- Earlier experiments (Binghubash and Juric, 2011), (Almarri et al., 2012b) in this research indicated that SWT stack (SWT, 2004) as technology of choice can help address users' preference extracted from search queries when either creating search engines/ Recommender Systems or tailoring (filtering / ranking) search result (user queries might not be sufficient!). Hence, utilising SWT to model user's intentions, expectations and demands, would definitely be seen as "sine qua non" when proposing a new solution in this research. This means that through the choice of SWT languages the model:
  - (I). capture and manipulate the semantics of the environment where internet searches happen and
  - (II). include in it users' expectations, intentions and demands.
- The proposed approach in this research is a SE solution that focuses on the selection of internet sources for "users" from dissimilar education level, age range, experience in handling advanced technologies in both forms (software and hardware), purpose of learning or retrieving the information and preferences. Moreover, it provides better approach to alleviate modern IO and minimise the problems in previously existing solutions. Furthermore, the proposed approach emphasise the importance of providing the users with results of "a moment", that will collect current preferences from the user; rather than

providing results based on past behaviour and stored demographic information of users.

- This research contributes with a rich investigation and analysis of main aspects of the research problem domain. Chapter 2 of this thesis, covers important events of causes of IO, enriched with range of examples from domains where IO happens. It also provides a chronological order of events and developments in the area of Information Retrieval Systems (IRS) (Baeza-Yates and Ribeiro-Neto, 1999 ), RSs, Search Engines and their techniques.
- This research also contributes in itemising:
  - IO's causes and problems,
  - Proposed solutions to IO, either by using RSs or search engines, and their techniques,
  - The shortcomings of these techniques,
  - Users' role in IO and proposed solutions to improve the performance of user based solutions to alleviate IO.

Consequently, the proposed approach deals only with the information that is important in a certain “moment” when the internet search happens and help in delivering relevant internet searches according to the captured semantics of the environment and user involvement.

## 1.5 RESEARCH LIMITATIONS

Some of the limitations in this research are the assertion methods of feeding the ontological model with domain-specific knowledge. At the time this research was conducted I was not aware of any existing technique which will collect domain knowledge in the process of constructing the OWL model and its components. Therefore, in this research, a manual assertion domain knowledge was followed to construct concepts of the computational model.

Furthermore, it was very difficult to obtain commonly utilised system and query logs in the process of extracting and analysing the behaviour of both IRS and information seekers/users on the internet due to privacy and confidentiality issues. and automatically. Hence, the collection and extraction of domain knowledge for the propose of examining the proposed solution was done by itemising possible information sources on the internet

(social networks, websites, blogs, online social media) and how these information sources delivers domain specific information to information seekers/users on the internet (such as patients, students, service providers, and manufactures in the domain of healthcare).

Although this research investigates RSs and search engines problems in the process to deliver recommendations/search results to information seekers/user. It is important to note that it is impossible claim that the proposed solution in this research can address all discussed problems of RSs and search engines and their techniques. This is due to time limitation, manpower and resources. Furthermore, RSs and search engines employs a wide range of techniques in the process of information retrieval which requires team work in order to enhance and solve such problems.

## 1.6 RESEARCH METHODOLOGY

This section, highlights some important aspects and methodology followed in this research. Firstly, this research is conducted on the basis of the combination of two distinctive methodologies: theoretical and experimental methodology. The decision behind this choice is that the proposed solution in this research is done in two stages.

The theoretical approach was selected to develop logic and prove the correctness of the proposed solution, define the limits of computations and computational paradigms, and model new systems.

### **Theoretical Methodology:**

- Helps researchers identify/understand research problems,
- Helps in the setup of the research objectives,
- Helps to identify the methods and techniques to start the research investigation and data collection,
- The analysis of the collected data for the purpose of producing new knowledge,

### **Experimental Methodology:**

To examine the proposed solution, the experimental methodology allows researchers to show through experiments and extract of results from real world implementations. Thus, it serves to:

- Test the veracity of theories,
- Analyse performance and behaviour, etc,

- Test the accuracy of the results and their reproducibility.

By combining these techniques, a strict process is followed in both data collection and the implementation of the proposed solution. Thus, main research components is conducted as follows:

1. *A thorough investigation of the impact of IO when retrieving information from the Internet.*

The investigation provides a definition of IO in different problem domains such as education, business, government, and healthcare. Furthermore, the investigation will also provide the readers with an overview of different types of IO in different problem domains. Furthermore, investigation highlights the sequence of providing the reader with a brief history of IO and go through a series of events that lead into listing modern IO problems. Last but not least, this thesis briefly elaborate on information Underload and its causes and impact on information retrieval on the internet.

2. *Provide an overview of existing solutions in the domain of internet retrieval systems in association to the proposal in this research.*

IO is not a new problem; much work was done to remedy and alleviate IO. it is important to note that this research deals with modern IO caused by retrieval systems on the internet. Consequently, investigation on internet retrieval systems is limited to search engines, RSs and their techniques only. Thus, this research attempt to itemise problems of existing solutions to alleviate IO. furthermore, understand the reasons behind the creation/ development of these solutions. Hence, provide a sequence events of research; that starts with a brief history of IRS, list distinguished work in the area of RSs and their techniques. Users also heavily rely on search engines to retrieve information on the internet. Therefore, it is undeniable to investigate its role in modern IO; thus start from the history of search engines and go through important events in the timeline of search engines development.

Not all technologies/solutions discussed in this research are directly related to IO. the reason behind selecting these technologies and solution is either related to their role in the improvement of these technologies or their existence solve different problems in certain domains. Sections 2.3, 2.4 and 2.5 provide an explicit review of these technologies and solutions based on situations and problem domains.

3. *Further extend the investigation in this research to understand the role of UB on the internet.*

Technologies listed in 2 above were also concerned with the user's role in the improvement or development of all these solutions. Hence, it is very important to define the user's role in both increasing modern IO and users' continued disappointment in existing information retrieval systems and their techniques.

4. *The buzz word Web 2.0 (O'Reilly, 2009) enabled users to become creators and consumers of information on the internet. Sometime later, Semantic Web Technology (SWT) surfaced and enabled software engineers to develop web contents that can be interpreted by both humans and computers at the same time.*

This research is interested in solutions in relation to IO and connected to information retrieval methods on the internet. Therefore, investigates SWT solutions, such as Semantic Search Engine (SSE) (Khan et al., 2014), that aims to improve search results based on the semantic specification of information sources. Thus, elaborate on SWT solutions in sections 2.5

5. *The proposed generic computational model in this research defines three important concepts: i) the semantics of users' role in internet search retrievals, ii) the semantics of internet sources and iii) a moment of internet searches.*

The proposed model in this research is generic in the sense that it:

- Can be used in different problem domains.
- Can provide results based on the characteristics of information sources on the internet and user preferences.
- Provides results in an ad-hoc basis; results are deleted at the next moment.
- Become domain specific once a domain of interest is present.

6. *The illustration of the proposed generic computational model in this research.*

The generic model is illustrated by developing an ontological model based on SWRL enabled OWL ontologies, that examines the proposed generic computational model in this research and follows a strict rule:

- Collects data about the problem domain manually and stores them in a spreadsheet.
- We then create an ontological model which will store the semantics of the domain.
- The ontological model then collects the semantics of the environment and users' preferences. This is done in two steps: we first investigate the environment

where the selection of Internet sources happens, to model the semantics of the problem domain. Secondly, we design a competency question that will provide a scenario of user preferences and present as case studies.

## 1.7 THESIS STRUCTURE

### **The Thesis is Laid out as Follows:**

Chapter 2, is literature review that provides an overview of IO and approaches to addressing it. The chapter spans the published work over a few decades and focuses on the latest attempts to manage diversity, relevance and amount of search results given by modern retrieval systems (see section 2.2). This chapter also provides an investigation and elaboration on Information Retrieval systems and their role in managing and finding relevant information based on users' queries (see section 2.3). furthermore, provide a discussion of Recommender Systems and their techniques, Tagging, Annotations and Folksonomies on the internet (see section 2.4). Search Engines one of the tools that are directly associated with modern IO. Therefore, an in-depth investigation of search engines and their techniques and their progress through time is present in this research (see section 2.5). Furthermore, this research also highlights some of the well-known search engines techniques such as the ranking of internet searches, which have become an important vehicle in managing enormous amounts of information on the internet through search engines. This section also explores the role of users when retrieving information on the internet (see Section 2.4.4).

Chapter 3, go through potentially related work which offers' approaches that either investigate IO or solutions that consider SWT. Furthermore, discussion on the listed work in this chapter is supported with comments and personal opinion is present. This to discuss the similarities and differences in these works to the proposed generic computational model in this research. This chapter paves the way to setup the problem in this research as described in chapter 4.

Chapter 4, starts by systemising and setting up the research problems encountered in the literature review (see section 4.1). It also addresses method to achieve the research objectives from chapter 1 (see section 4.2). furthermore, this chapter introduce the proposed generic computational model and justify reasons behind the technology of choice in this research (see sections 4.3). This chapter discusses a novel computational model that delivers the research objectives to alleviate modern IO.

Chapter 5, presents an illustration of the proposed generic computational model in this research. The illustration examine the proposed generic model in a domains-specific situation of information retrieval on the internet. This chapter also provides an overview of the domain of interest (LLL in healthcare) (see sections 5.1). The illustration shows how the generic model is defined for domain specific situation, and how the reasoning upon these models delivers a moment of internet searches. The illustration shows how the semantics of the environment is modelled, the process of collecting the semantics of users' request and preferences, and how the ontological model performs the selection of internet sources.

Chapter 6 and Chapter 7 concludes this thesis and discusses/comments on the research approaches and achievements of the research objectives (see section 6.1 and 6.2). Thus, provides a detailed evaluation/impact of the proposed approach. Furthermore, elaborates on possible extensions or future work of this research (see section 7.1 and 7.2).

## CHAPTER 2. LITERATURE REVIEW

Chapter 1, gave an overview of topic of discourse in this research, the motivation behind choosing the certain research topics in relation to the research problem such as understand the IO phenomena throughout the past 3 decades and its role in modern IO and information retrievals from the internet. Consequently, to understand modern IO it was essential to investigate how information seekers/users' behaviour and interact with modern technologies on the internet. Hence this encouraged the investigation of the two well known information retrieval techniques on the internet; search engines, RSs and their techniques. furthermore, chapter 1 also provided the reader with a detailed list of the research objectives, contribution, limitations and the methodology that was followed in this research and was concluded with the thesis structure.

The reason behind introducing different types of IO in this research is to provide an overview of IRS which were created to help information seekers/users to easily find information, itemise information retrieval problems and categorise systems and technologies used to both search and retrieve information and that possibly create IO in “a moment” of internet searches. A sequence of intensive work in the domain of IRS provided solutions to traditional IRS. Hence, it is important to highlight that not all IR systems and IO problems are applicable to modern IO but they somehow participated into it.

- **Quality and quantity of information on the internet is effected by information seekers:**

Almost three decades ago, information creation was exclusive to information owners, knowledge or service providers only (educational buddies, professionals from businesses and organisations who provide services to people, healthcare buddies who spread awareness and informative sources of information such as encyclopaedia and so forth). However, advances in technology enabled information seekers/users to become creators and consumers of these services. This role influenced users' everyday life, because they became participants in creating different types of contents on the Internet. Users are from different backgrounds, knowledge, level of education, age, gender, interests. Hence, creates and views information on the Internet accordingly. On the other hand, this influenced the quality and quantity of information available for other

information seekers/users on the internet. moreover, variety of tools and technologies on the internet Consequently, information seekers/users on the Internet started to suffer from modern IO. Therefore, service providers and researchers felt the urge to develop systems that can manage the excessive amount of information on the Internet.

- **Insight on traditional and modern systems and technologies in the process of information retrieval:**

Attempts to provide information seekers/users with relevant information started almost three decades ago with the introduction of IRS for libraries. A sequence of improvements followed to further include accuracy and precision and other traits in the process of information retrievals. The birth of networks, hypertext and the internet required new methods of information retrievals from the internet. Hence, search engines surfaced to help ease the process of information search and retrievals on the internet. Unfortunately, early search engine solutions carried some problems in search results. Some of these problems are related to keyword matching, spelling, indexing, and information structure. Hence, search engine developers attempted to solve these problems through many well known techniques such as storing problems, web crawlers, hotbots, phrase recommenders and so forth. Most of these problems does not exist anymore, but it was worth mentioning to show sequence of improvements in search engine through time.

The rapidly changing internet environment allowed the introduction of online services such as e-commerce. Hence, new ways of information presentation can be generated by both service providers and users of these services. In e-commerce, the famous technique of information retrievals can be found in the domain of RSs and their techniques such as CF and CBF. These techniques were originally developed to solve the problem of IO in business and organisations in early 90s. Developments in RSs resulted into combining CF and CBF to create a hybrid recommendations. Hybrid RSs rapidly advanced to include many other traits such as profiling of both information seekers/users and items, demographic data, location and so forth.

Both search engines and RSs were improved to provide better search results according to the environment such as education, businesses, healthcare, e-commerce and entertainment. Furthermore, a mash up of both search engines and RSs techniques were experimented to refine search results to include location, time, ranking and prediction of

user behaviour. The way users create information, the different types of information and the ease of accessing them through linked data forced researchers and service providers to combine techniques in order provide better representation of information and services. Consequently, it was very difficult to separate search engines and RSs information retrieval problems and solutions. Therefore, it was essential to provide an overview of both search engines, RSs and their techniques in order to help the read understand the contribution of the proposed solution in this research.

This chapter is organized as follows: Section 2.1 covers a chain of events of IO, its types, and solutions that have attempted to solve the problem of IO in the past three decades. Section 2.2, provides an overview of IRS and their role in solving the problem of IO in the 70s. Section 2.3 covers RSs and their techniques and the impact of developing RSs that is built on top of Collaborative Filtering (CF) (see section 2.3.2.1), Content Based Filtering (CBF) (see section 2.3.2.2) and Hybrid Filtering techniques (see section 2.3.2.3). This section also provides an empirical analysis of other RSs techniques, such Tagging, Annotations and Folksonomies (see section 2.3.2.4). Section 2.4 starts by providing the readers with an overview of the chronological order of the development of search engines (see section 2.4.2). Subsections of this chapter analyse some of the search engine techniques that influenced the development of search engines and improved their search results. The widely used technique Search Engine Optimisation (SEO) (see section 2.4.3.1) and User Behaviour (UB) on the internet (see sections 2.4.4) on the Internet. This chapter also includes investigation SWT (see section 2.5), solutions which claims the use of SWT such as Semantic Search Engines (SSE) (see section 2.5.2) The chapter is concluded with discussion of relevance in query results (see section 2.6) and examples from both RSs and search engines, and relevance and accuracy's role in search results.

## 2.1 INFORMATION OVERLOAD

### 2.1.1 OVERVIEW OF INFORMATION OVERLOAD

In the digital information age, IO (Toffler, 1970) (Schick et al., 1990) (LaPlante, 1997) (Bawden et al., 1999) (Himma, 2007) (Bergamaschi et al., 2010) (Vossen, 2012), (Remund and Aikat, 2012) (Strother et al., 2012) is considered as constant companion of

users. Since users started to produce excessive amount of information on a daily basis, IO problem continually inflate. In order to manage information, users heavily depend on search engines that are supposed to help in retrieving the right information at the right time. At the same time search engines overload users with multiple search results that do not necessarily meet their expectation (Doomen, 2009) (Kovach and Rosenstiel, 2010) (Lincoln, 2011) and contribute towards IO in any environment. Furthermore, IO also exists everywhere and it is not confined solely to retrievals from the internet (Farhoomand and Drury, 2002) (Lee and Lee, 2004) (Lohr, 2007) (Russell et al., 2007) (Baez et al., 2010) (Simperl et al., 2010) (Ulijn and Strother, 2012). Next section, provides a definition of common terms in this research (information, information Sources, and IO), give a comprehensive and chronological order of the events of IO and examples of multiple environments where IO happens. Moreover, this chapter introduce some of the most popular technology solutions that aimed at identifying causes of IO in different environment and on the internet in specific. Further discuss some of the previous solutions to solve information retrieval and IO problems in domains such as e-commerce, businesses and organizations, healthcare and online learning communities which are user intensive environment on the internet.

The next section begins with an attempt to define the term “Information” and followed with a detailed description of IO, its causes and previously proposed approaches to IO. furthermore, discuss technology role in both providing users with information sources and at the same time overloading them with excessive amounts of information. Hence, provide discussion to support it with examples of IO in two important online environments; healthcare and business that allows intensive exchange of information and can create IO for a variety of reasons. Lastly, the influence of perceived IO on users’ participation and knowledge construction in computer-mediated communication (Chen et al., 2012).

### 2.1.2 CONCEPTS OF INFORMATION

It is common that people confuse the meaning of *Data, Information and Knowledge*. Therefore, it is necessary in this research to clarify this confusion and provide the reader with a definition of these three terms and use them accordingly in this thesis. The figure 2.1 below shows the perspectives on data, information and knowledge.

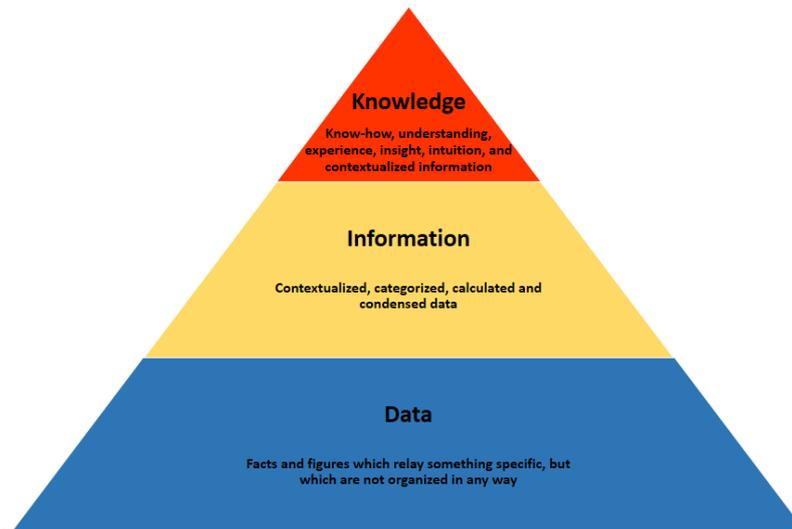


FIGURE 2.1 THE HIERARCHY OF DATA, INFORMATION AND KNOWLEDGE  
(FROST, 2010).

**Data** are unorganized collections of conditions, ideas, or objects. To be able to use data, they must go through the process of polishing them to be accurate, presented in a timely manner, specific, organized for a purpose, presented within a context that gives it a meaning and relevance, and can lead into understanding. *If data fulfil these conditions, they become information* (BusinessDictionary, 2016) (Cooper). Ackoff described data as raw that exists and has no signified beyond its existence, and can exist in any form and has not meaning (Ackoff, 1989).

The term **Information** was originated in late Middle English age, and was known as the formation of mind and learning (Oxford Dictionaries, 2016). Information was studied for many centuries, the study of information is known as “Information Science” (Borko, 1968) which is focused on the analysis, collection, classification, manipulation, storage, retrieval, movement, dissemination, and protection of information (Borko, 1968), (Stock and Stock, 2013). Information can also affect behaviour, decision and outcome. According to Oxford Dictionaries, information is the representation of a sequence of data in a particular way; it is facts about something or someone. Thus, researchers from the information science discipline agreed that information science is a multidisciplinary science which is associated with a wide range of other sciences. Some of the many fields heavily studied in correspondence with information science are Information Access, Information Architecture, Information Management, Information Retrieval, Information Seeking, Information Society, and Knowledge Representation and Reasoning. Königler

and Janowitz defined information as a set of data that are organized in a way that makes sense; thus, always depends on the purpose they are created for (Königer and Janowitz, 1995). Cooper defined information as data together with a context that results into a meaning. Thus, information answers the “who, what, where, and when”(Cooper, 2014). Ackoff also described information as data that has been given some meaning that can be useful (Ackoff, 1989).

**Knowledge**, according to Ackoff is the appropriate collection of information that are useful and requires a true cognitive and analytical ability to process information for a given problem (Ackoff, 1989). Cooper also defined knowledge as information that is structured and organised and can be processed; thus answer the “how” questions(Cooper, 2014).Hence, knowledge is the outcome of information seekers (people in general or users on the internet) ability to understand the collection of information they obtained through available means of information retrieval (in hard/softcopy); and transferred into skills, experience and education acquired from certain information (Oxford-Dictionaries-Language-matters, 2016).

In research community, Simpson and Prusak work implies that information exists to serve the needs of a function, or to provide more reasonable thinking of business process or customers. Thus, value of information is reflected in the process of creating new knowledge and benefiting from it (Simpson and Prusak, 1995)

From this point forward the report will use these three terms as defined above. The following sections in this chapter (2.1.3 to 2.1.9) will discuss some distinctive opinions and definition of IO and categorise IO according to users’ involvement and environment where IO happens.

### 2.1.3 KNOWLEDGE OVERLOAD

While a variety of opinions exist on what knowledge brings to people, some researchers labelled information as “Information Glut or Smog”. these are terms which rise controversial arguments when describing and characterizing information quality. These two phrases are associated and paired with modern information age (Denning, 2006), (Shenk, 1998), (Fox, 1998) that involves information retrievals through modern technologies and effected information seekers at either workplace or users on the internet.

As discussed in the previous section (Concepts of Information), knowledge is the outcome of users ability to process information. This to gain skills or experience that will help them to fulfil and achieve pre-set goals or help them to complete a task or business process and so forth. What if the knowledge users seek is the cause of their anxiety and deterioration in their progress at work, school or their day to day tasks?.

This section discusses the negative impact on user of knowledge disseminated in two knowledge intensive environment (academia and workplace). The discussion looks at the unexpected information that resulted into Information Pollution and knowledge Overload (KO).

Concerns of information pollution and KO loom as pair in the modern information age. Bray in his work examines IO from a different angle, and questions whether ISs professionals can think of a method that will help them address the challenge of information pollution and KO. Bary investigated the problem of Knowledge workers, who are victims of IO at workplace (Bray, 2008).

Studies indicates that in 1900 there were 9000 scientific articles published, in 1950 the number of articles is 10 times more (90000 article in a year) and by the end of the 20<sup>th</sup> century 900000 articles were published. All these articles are digitised to allow easy access to information. Hence, Bray presents an interesting argument in his article that claims information pollution is viewed as “*positive global movement empowering individuals to access and produce knowledge globally*”.

Consequently, any individual can find any information on the internet and analyse it, and produce a mix of new media or a bit of information (Bray, 2008). Thus, the slight change in the original piece of information created new information that will trigger information consumers to think that they need to obtain this new information to feed their knowledge. Therefore, users pollute their cognitive ability with additional information processing that they believe they need.

For several years ISs professionals were occupied by their obsession to design usability and user expectations, in insure successful transaction of knowledge and organizational performance. The dissemination methods of information was the main cause of information pollution. Therefore, at workplace KO resulted into lost productivity, working overtime and decline leisure time.

Accordingly, Bray suggests to study ISs artefacts and not only include augmented effects on the cognitive capabilities of individuals, but also those of groups and organizations must be taken into account. Thus, detail aspects of the IS artefact that conserve memories and process capabilities of individuals. Furthermore, Bray advises to utilise IS in better ways of use and design to attract human interactions and decisions when confronted with knowledge overload (Bray, 2008).

Controlling the flood of incoming information is very challenging in the 21st century. The available range of information on thousands of news websites, blogs, forums, photo/video sharing websites, academic journals, and the non-stop google digitalized books creates a huge pile of unread information. Moreover, new inventions, online services, the influence of major sources of information (email, Youtube, Facebook, e-libraries) on users' daily activities of information retrieval and new mobile technologies (Twitter, Snapchat, Instagram) makes it impossible for anyone to stay current.

The accommodation of web 2.0 tools enabled information to be available at a blink of an eye. Ken Coatos in his work raised some concerns about academics who struggle to stay current in their discipline. It is very rare to find a discipline that is not affected by the massive growth of information. Thus, academics in any given discipline suffer from inability to keep current and it is even more challenging to those in multi-disciplinary or inter-disciplinary areas. As a result, reading level goes down, the academics becomes more focused on selected readings that will participate in their career rather than keeping up with the latest academic publications (Coates, 2009).

A very simple answer to IO in academia was to rollback to previous methods of knowledge acquisition; (old-knowledge system such as academic journals, books and conferences). Coates suggests to shift users attention from keeping current to focusing on publications in academia to control Knowledge Overload. Although, this solution can work well for academics the problem prolongs because of annual national and international conferences that range in quality, type of audience and attendees results into producing new findings in academia. The point is that such events overload the reader with knowledge and with current technologies. Proceedings of these events are immediately indexed and can now be viewed instantly by reader.

To sum up, knowledge overload is mainly caused by the avalanche of low quality writing that drives good writing out of the picture, and thus it takes over the attention of

the reader. It's even more disappointing, says Ken Coates, "***how we currently research without having impact, speaking but not being heard, and writing without being read. Even worse is our capacity to read deteriorated we read indexes rather than journals, abstracts rather than papers, review essays rather than books***" (Coates, 2009).

This section covered two important groups of users of environment where intensive information dissemination caused by modern technologies on the internet. The reason behind choosing these two examples is that academics and workers are victims of knowledge they produce. Furthermore, KO is directly related to utilisation of modern technologies. Last, this research aimed to discuss different type of IO and its influence on information retrievals from the internet.

#### 2.1.4 TECHNOLOGY OVERLOAD

The previous section slightly mentioned users who uses technology in general and specifically on the internet and suffer from KO. Technology enabled instant information creation and retrieval, allowed users to connect in unprecedented ways i.e. through SNS, email, skype and devices like smart phones, tablets and computers. Hence, information travelled smoothly and effortlessly from one user to another. However, this created unpleasant feeling of Technology Overload (TO). Therefore, this section attempts to discuss TO that is caused by information dissemination and retrieval and discuss the impact of rapidly changing mean of communication.

In 1989 Roe in his article "letters" described IO as a common phenomenon caused by the evolving technological society (Roe, 1989). The abundance of information was a sinister sign that carried serious economic and social implications. Collecting data became a habit of many readers, learners, and researchers. Roe thus aimed to raise awareness to the few important problems such as (gathering information, storage, retrieval, analysis, interpretation and usage of data).

Technology solutions to resolve IO included IR systems in early 70s. IR systems attempted to solve IO in libraries and helped in organization and sorting of books, documents, journals, and audio/video records (G. Semeraro, 2000).

Hypertext technology is yet another technology which aimed to make information handy to user at any time. And was followed by many attempts to filter and retrieve information sources on the internet (Wittenburg, 1996).

Search engines were the next generation of technology solutions that aimed to retrieve information sources on the internet. Mitchell et al., in “*Fishing for Information on the Internet*” discuss continual growth of the internet and online resources. The amount of information user can access nowadays becomes increasingly vast. All these technologies added positively to users capabilities to teach ourselves and answer users questions with a huge list of results. Thus, users are now able to view more information through online resources than they expect in the blink of an eye (Mitchell et al., 1995).

Technological developments aimed to ease people’s lifestyle; thus, sequence of developments resulted into devices that allows the delivery of multiple functionalities to the users in an easy way. To survive the high market competitions, manufacturers develop new technology in no time. Hence, all these technologies and gadgets overwhelm the users with multiple ports of incoming information which was not expected or aimed to achieve on the first place. Grandhi et al., in his work claims that this can be considered as TO accompanied with the cognitive and physical burden placed on users who use these technologies in their everyday activities (Grandhi et al., 2005).

Mitchell et al., research was concerned with quality of the information on the internet and time needed to go through the huge list of retrieved results (Mitchell et al., 1995). Wittenburg highlights the problem accompanying hypertext technology (the web). Although hypertext technology enabled the growth of WWW and ease to publish contents on the web, it also participated in information inflation problem. Information far exceeds individuals or organizations processing capacities’ (Wittenburg, 1996). Grandhi et al., criticized the non-stop advances of technology because the development did not take into account any systematic process determine which features and functionalities should be placed in which device/s in order to reduce TO (Grandhi et al., 2005).

Levy in his discussion of the consequence of IO on our daily lives argued that: although technology tools made information accessible by people equally; it also reserved a vast amount of their daily time invest in new knowledge obtained through flood of information through different means of communication (Levy, 2005).

To overcome TO, Mitchell et al. introduces search visualization tool called FISH (Forager for the information Super Highway) that provide the users with better visualized search outcome and will allow the user to view refined list of contents once they click the search results (Mitchell et al., 1995). Wittenburg work attempts to find strategic directions for the future of Human Computing Interaction (HCI), for two reasons: 1) to find ways to use computers to allow people collaborate with each other to accomplish tasks and 2) to integrate search, browsing, and filtering technologies that were in isolation so far and means of accessing information on the digital age (Wittenburg, 1996).

To sum up, many researchers defined IO and its problems according to the environment where IO happens. TO is one way of IO to manifest. New technologies enabled users to access information instantly and permitted the them to easily participate in the creation and generation of information through the many available collaboration, sharing and discussion platforms (Wittenburg, 1996). ‘21st century technological revolution’ also made it possible to find what users want as technology (features and services) all in one device. Users of these technologies i.e. non-professionals are now able to contribute to knowledge just by single button click to either capture a scene or video, create an instant power point or poster and use these materials in an ad-hoc bases. A mashup of technology is provided in smart phones which integrates functionalities and requires minimum effort when users produce information (Grandhi et al., 2005).

#### 2.1.5 COMMUNICATION INFORMATION OVERLOAD

This section provides analysis of two distinctive works that illustrate the problem of Communication Information Overload (CIO). This type of overload can result from any information dissemination means i.e. traditional means such as TV, Radio, Mail Post ads and Banners. It also includes modern ones such as websites, blogs and social media tools on the internet. Furthermore, CIO is also the consequence of the existence of business services, e-commerce, marketing, decision making and which were carried to the internet. The reason behind this choice is their connection to Technology Overload (TO) and because they investigate CIO in two social intensive environments (marketing and the internet as the host of marketing and technologies that allow information dissemination). In the first example, Van Zandt attempts to defines CIO in relation to users and environments (that involves marketing strategies and decision making) (Van Zandt,

2004). Whereas the second example, Jonathan's work discusses the impact of modern communication methods that may result into IO (Jonathan, 2014).

Van Zandt experimental research highlights the problem of low cost of generating and transmitting information; thus, proposes a model to analyse the behaviour of targeted communication among senders and receivers of information in a university environment. This research attempts to understand the interests of senders and receivers of information who are not fully aware of the contents of a communication (message). Furthermore, study a network of targeted communication rather than broadcast based on individual target. Therefore, the constructed model consists of a finite number of senders, indexed messages, and large population of receivers. The scenarios examine the reaction of three groups of senders and receivers in different situations through intramural mail or email.

Given the main elements of the model, the experiment collects and analyses both the sender's and receiver's interactions with the system. The outcome of the experimental work indicates that the senders are not able to observe which messages a receiver would be interested in and. Furthermore, the model base assumptions on collected demographic and marketing data. Thus, creates a receiver's profile based on types, sets and interests (Van Zandt, 2004).

Jonathan in his article quoted Tony Robins when he said that: ***“the quality of our lives is in direct proportion to quality of our communication”*** (Jonathan, 2014). Users depend on many communication means to deliver information to audience. Thus, they write, speak and use visual methods to leave a lasting impression on the audience. Therefore, researchers investigated IO caused by modern communication methods.

Mass communication is yet another form of IO which was carried from traditional means of communication to the internet. Consequently, users are overwhelmed and not able to focus on all incoming information. If users begin to discard or delete emails without viewing or watch a TV program and tune out in unprecedented rate, then these are signs of being bombarded by too much information. Hence, almost all users suffer from CIO symptoms. Jonathan supported his opinion on CIO by claiming that plots in movies always have some information conveyance problems that leads to misunderstanding a situation. Thus, misuse of communication tools can create unrequired IO that mislead the users.

Consequently, Van Zandt attempted to define IO as “*IO is the consequence of the strategic interaction among senders of a messages and that they also are directly harmed by their collective overexploitation of the receiver’s attention*”(Van Zandt, 2004). Hence, Jonathan comments on CIO (in the age of information, users avoid facing IO problems and resort to continues attempts of information filtering (Jonathan, 2014).

In addition to the two examples of CIO in this section, it is important to note that, modern technologies that allow users to create information instantly on the internet further worsened the problem of CIO. The amount of information on the internet keeps growing whereas quality of information is not to the standards. Quality of information on the internet is huge problem but it is out of scoop of this research (Oliver et al., 1997, Library, 2009).

#### 2.1.6 INFORMATION OVERLOAD IN BUSINESSES AND ORGANIZATIONS

Management Information Systems (MIS) (O'Brien and Marakas, 2007) were developed to help employees manage and organize internal and external communication for businesses and organizations. MIS solution was followed by many other attempts that shard common target -overcome IO in business environment- such as controlling decision making, project management, finance and accounting, and human resource management and later were packaged into Enterprise Management Planning (ERP) systems (Ragowsky and Somers, 2002). IO in professional environment is mostly the consequence of market competition, work load and pressure from upper management. Hence, employees at work place start to seek for more information to match work demands. In addition, employees are also overload by both internal and external communications whither they are business related or not the employee is required to go through all communications to make sure that he/she is not missing out any important business related information (Denning, 2006).

Information seekers who belong to the businesses and organisations environment are another group of users who are heavily involved with information creation and dissemination to accomplish certain business task, process or for decision making reasons. In early 90s the internet witnessed revolutionary e-commerce industry which was created by e-Bay (eBay, 2017) and Amazon (Amazon, 2017). Technology enabled

the transmission of business and organizations services from real life into the internet. Hence, in addition to ERP systems, businesses and organizations conduct almost all communications and information generations and distribution through adapted new technologies such as emails and social media tools on the internet. Therefore, businesses and organizations are two vulnerable domains where users may suffer from IO

Information seekers who suffers from IO at organizations or any other environment such as healthcare and online learning are always full of complaints about service providers' information representation techniques. furthermore, information seekers criticise the amount of information received on daily basis that requires immediate processing and their inability to do so (because they are confined by a strict timeline followed with deadlines). Receiving information through multiple channels of communication -such as emails, newsletters or read circulars from the management- are useless, time wasting tasks and add more to their to-do-list (Königer and Janowitz, 1995).

In the domain of e-commerce, Lee and Lee described the impact of IO on online consumers' decision making (Lee and Lee, 2004) and Ulijn and Strother tagged users of online retail websites as "DROWNING IN DATA" (Ulijn and Strother, 2012). Russell et al., further claims that IO in business manifests through incoming information from external sources such as emails; thus, employees starts to suffer from IO (Russell et al., 2007). Lohr highlighted the negative impact of IO on organizations financial situations (Lohr, 2007), Janssen and Poot labelled societies that suffers from information overflow from multiple sources as information intensive society (Janssen and Poot, 2006),.and Farhoomand and Drury brought forward the problem of IO in managerial decision making in business (Farhoomand and Drury, 2002).

Shenk claims that user of technology and communication means are disadvantages. For instance: 1) users can access news 24/7, but are still uninformed, and 2) users miss-use technology by sending spam and scam emails, viruses, hijack websites by ads, and phishing. Nevertheless, users continue to seek for more with claims that it is needed for performance purposes, functionality and business requirements (shenk, 1997).

The Lueg and Fisher further criticised the advances of technology used to spread awareness, such as CSNET (NSF) and discussion systems such as USENET. Many of the existing technologies gets our attention and we get impressed, but once we get close

to them and start figuring our way to use them beneficially, we end up adding more to our load of to-do list (Lueg and Fisher, 2003)

Denning argues that employees of all hierarchies at companies and organizations suffers from IO as a chronic disease. The desire to obtain more technology solutions such as management systems and software resulted into electronic junk and uncontrollable growth of business related reports. Even worse is keeping up-to-date with frequent software updates that further demand training to insure proper user of the tool/software for business requirement (Denning, 2006).

Hence, approaches to resolve IO in businesses and organizations involved an intensive work to highlight main causes of IO as follows:

Simpson and Prusak's research explore three important aspects of information in professional environment to understand reasons of information detention and inability of managers to benefit from information available through MIS systems. The influence of IO on employees and failure to create high quality information and how can managers utilise information for profit making? (Simpson and Prusak, 1995).

Janssen and Poot investigates Knowledge Management Systems (KMS) which are main sources of industrial information which grows constantly. Hence, employees are expected to face abundance of information that creates IO and may slow down work productivity. This triggered the management attention to the problem of IO at workplace. Thus, they requested to investigate and analyse this problem to eliminate the cause of IO (Janssen and Poot, 2006).

Denning research was triggered by his inability to coop with nonstop information flow through emails from multiple senders. Furthermore, all attempts to filter/sort emails on daily basis and redirect the important ones to the responsible person to process the contents of the email. Denning still suffered from IO attacks. Moreover, piles of unread reports that continually grow on his workspace (either printed from email attachment or circulated in hardcopy) even worsened the situation (Denning, 2006).

Whereas, Simperl et al., work investigates the impact of IO on knowledge workers who are central to an organization's success. Furthermore, discuss the possibilities of exploiting the enterprise information to increase productivity knowledge works (Simperl et al., 2010).

Consequently, some of these researchers attempted to propose solutions that will help users in the domain of businesses and organisations to cope with IO. Simpson and Prusak for example, proposes a conceptual model based on a combination of value added models of information. The Model consists of five universal elements of value in information (Truth, Guidance, Scarcity, Accessibility and Weight) and utilise business application and context. The model take in to account to role of i) information sources in business communities and ii) the role of creators and consumers of the information. Last, the model provides decision making process that requires data collected from top and middle layer managers (Simpson and Prusak, 1995).

Hayes-Roth purposes to utilise Valued Information at The Right Time (VIRT) (Hayes-Roth, 2006) concept and combine workflow systems and decision making to bring human dimension IT; thus, aimed to reduce IO. furthermore, add (Push/Pull and Smart Push) techniques of online subscriptions to deliver the required piece of information created by the producer to the right consumer at the right time. According to Hayes-Roth, VIRT can control IO that grows with advanced technologies (Hayes-Roth, 2006).

Janssen and Poot work proposes to examine the impact of IO on daily work productivity of employees from different hierarchy and measure employees' perception and ability of information intake. Hence, utilised critical incident collection technique and textual interpretation and the affinity diagram technique to establish interview questions and collect data about frequency of IO manifestation in managers' daily work, approaches to resolve it and potential solutions to it. General data includes employees job-related information, role, tasks, responsibilities, experience, number of team members, frequency of direct reports and business trips, and description of several situations of IO (Janssen and Poot, 2006).

Whereas, Simperl et al., proposes *The Active Project approach* to address the challenge of IO through an integrated knowledge management workplace in three steps. First, sharing information through tagging, wikis, and ontologies. The second, prioritize information delivery by understanding users' current-task context. Third, leverage informal processes that are learned from user behaviour (Simperl et al., 2010).

Simpson and Prusak's investigation indicates that the problem of IO was generated by managers who are not able to process the information correctly. Consequently,

suffered from underloaded. On the other hand, employees at the same suffers from IO because the available information exceeds the ability of employees of information processing. Simpson and Prusak also complained that although the so called electronic service “THE EMAIL” aimed to provide a solution for IO in businesses. It actually worsened the situation and increased IO in commerce, marketing, floods of internal resources of information, and later e-commerce (Simpson and Prusak, 1995).

According to Janssen and Poot data analyses, IO is caused by a) ambiguous emails, b) email cascades and avalanche, c) email workload, d) poor accessibility to information sources and e) fragmentation, too many information exchange either internally and/or externally. Few participants mentioned social pressure, inefficient meetings and unwelcome notifications. Further examination of IO resulted into employees being annoyed rather than suffering from real problems.

The outcome of Janssen and Poot research indicates the severity of IO at work place differs based on employee’s hierarchies an job responsibilities and decision making techniques. Hence, suggests to:

1. Enforced communication rules which deals with email communication, i.e. TO is for action and the CC is to inform only.
2. Increased problem awareness will help in measurement and giddiness of people, awareness and contribution to IO.
3. Training/coaching and procedures to release stress caused by IO, improve information management skills and encourage face-to-face meetings
4. Better Tooling involves improved use of organization tools (intranet, knowledge exchange communities, and remote collaboration tools) and personal information management tools (email, file and task management) (Janssen and Poot, 2006).

#### 2.1.7 INFORMATION OVERLOAD IN HEALTHCARE

In healthcare domain information availability is not an issue anymore, thus information seekers can retrieve information instantly with minimum efforts. Gregory the editor of physical therapy in sport highlights the importance of healthcare information dissemination across the internet. Hence, the use of advanced technologies on the internet eased the process of offering users on the internet with plethora of journals and other published media easy to access. However, searching for information can be a tedious task to most academics. Furthermore, searching in the wrong direction can make us face a great deal of irrelevant information. (Gregory S. Kolt, 2006).

In scientific communities IO manifests through the non-stop publications and new additions of previously published books. (Baez et al., 2010). Many researches in healthcare investigated the symptoms of IO for almost 4 decades even before modern IO manifestation. Thus, indicated that IO in healthcare is the consequence of the lack of time management and continuous increase of information volume.

Healthcare is one of the crucial environments on the internet where information is published rapidly and is used by a wide range of information seekers from different background and educational level. Hence, this section discusses IO in healthcare environment and its influence on healthcare information seekers. of high priority. In our research, we would like to highlight some of the published work in healthcare IO.

Therefore, IO in healthcare can be divided into two categories: first, traditional IO problems caused by (i.e. books, published conference and journal papers, medical reports, handouts, and healthcare technologies). Second, role of technology and ISs in the available of healthcare information on the internet.

Lock work investigate the problem of publication avalanche that exceeds the ability of the enthusiastic reader, inferior papers buries vital ones and explore methods to raise awareness to set a limit to the frequency of articles or journals publications.

Hence, lock emphasise to focus on quality rather than quantity pointing out few techniques such as (scanning relative journals, supplementing review articles citation lists, printouts, and conversations with colleagues) (Lock, 1982). These techniques were also suggested by Bernal and Fox to control exponential growth of publications and apply better divisions of journals and archival methods. Thus, suggested the replace traditional journals with distributing systems that would provide an abstract form of the article. IO in healthcare also prevented information seekers from crossing the boundaries of on discipline to another. This therefore, restricted information seekers from enriching their knowledge with research outcome from cross disciplines (Lock, 1982).

Another interesting attempt to minimise IO in healthcare was to control research publication through peer reviewing. This attempt aimed to prevent the avalanche of inferior papers and give change to vital ones. Hence, enforcing strict evaluation criteria, academics managed to temporarily control IO (Lock, 1982).

The progression of IO continues, and information seekers are still struggling with floods of publications in both forms (hard/soft copies). In the editorial message, Faber highlights his dilemma of not being able to keep informed with current research and findings his own discipline but also general knowledge in healthcare literature. Furthermore, Faber complains that information seekers in healthcare data will never be able to keep current with floods of information unless they are retired. Faber concluded his discourse by sharing his experience in coping with IO; thus, says that he only keeps records of three important journals of healthcare and gives out the other journals or throws them away (Faber, 1993).

In the next addition of BMJ, Westerman et al., revisits Faber article (Faber, 1993) and reopen the discourse on IO in healthcare domain. Although technology can solve some of IO problems in healthcare, the digital age it is very difficult to try to stay current. The rapid publication of knowledge in periodical timeline in parallel with one's responsibilities makes it difficult if not impossible to stay current.

Westerman came up with two possible solutions to IO. Information seekers in healthcare can use remote access to a DB that contains an abstract of all articles. Alternatively, use "Mentor" a decision support system that allows access to information sources and provides a summary of a summary, prioritise/ filter results into top five important topics to minimise time sent on information seeking (Westerman et al., 1993).

Information seekers in healthcare can range from being doctors, nurses, patients, insurance companies, pharmaceutical industry, equipment suppliers, medical students and so forth. Mikulencak and Turner in their investigation focused on the impact of IO on nurses who are bombarded with information from multiple sources (newspapers, journals, postal and electronic mail, and the internet). Therefore, the American Nurses Association (ANA) suggested to collect data of its users to create a profile that will help improve services and provide better techniques and research topics based on interests and preferences (Mikulencak and Turner, 1997).

On the other hand, Lyons and Khot in their investigation highlighted the problem of General Practitioners (GP) who complains about their inability to access accurate information sources to support clinical practices; thus, affects daily tasks progress. Therefore, Lyons and Khot suggest to develop an electronic directory which would consist of organized data of healthcare services for GPs and based on the WAX Active

Library software ([www.medinfo.cam.ac.uk/wax](http://www.medinfo.cam.ac.uk/wax)) that will allow access to shared information (Lyons and Khot, 2000).

which was created for primary care in the first place. The tool uses a combination of multiple existing repositories and mostly newly created data. Some of the issues the authors complained about are: an abundance of poor quality information, lack of awareness of the importance of quality of information in healthcare, lack of trust from beneficiaries, hesitation of recipients to share information among themselves, and lack of computerization of information in general. The system was examined among a group of users and, according to the authors, the testing was successful and the users were satisfied with the services and contents of the system. Using the directory helped in improving shared information among its users (Lyons and Khot, 2000).

#### 2.1.8 MODERN INFORMATION OVERLOAD

Many solutions aimed to solve IO through time, researchers resorted to books and encyclopaedias and, later, data bank. Still, IO is worse than ever. So, where do sufferers go from this point? (Lock, 1982). So far, research in the domain of information Science (Borko, 1968), (Stock and Stock, 2013) indicates that although many available solutions were provided to solve IO, the flood is increasing and researchers are not able to control the problem.

Locks complains are still valid because information seekers are now in even more critical situations when seeking for information on the internet. IO has been a problem for few decades. Many people suffered, and many are still suffering. In the past decades, IO manifested itself in many ways, it ranged from simple issues of being overloaded with information, and also include serious problems such as inability to cope with information floods. Hence, information seekers always feel the anxiety of not having enough knowledge to fulfil certain information need.

Technology advancements in the 21<sup>st</sup> century have certainly been vast: being able to access information anytime on the go in an ad-hoc basis exceeds what Bush's envisioned in his article "*As We May Think*" (Bush, 1945). Information seekers are now able to access any type of information instantly; thus, sharing information is even easier than imagined. Cutting edge technology changes in the blink of an eye, competition in technological aspects are rapid to the point that technology volumes and new releases are

made on a weekly basis which makes it hard if not impossible to keep up. All these advances are good, but increased the pressure on information seekers' ability to process the incoming information through all these communication means and channels. Hence, information seekers suffer more of IO than ever.

This leads the discussion to new type of IO "Modern Information Overload (MIO)". MIO eliminated the boundaries between different types of IO; thus, MIO include all of them and adds them problems of information retrieval techniques on the internet. It further includes information (knowledge representation, education, healthcare, businesses and services) transformation into the internet.

Hence, researchers attempt to resolve this problem includes investigating the domain where IO happens, characterise causes, understand the behaviour of the effected group of information seekers and most importantly investigate the short comings of technology in use to either improve it or propose new solutions.

Bawden and Robinson attempts to identify potential factors of modern information communication and the role of pathologies of information in the quality and quantity of available information on the internet. Furthermore, the investigation also includes the impact of quantity and diversity on IO and information anxiety, and changing information in parallel with advent web 2.0 tool. Results indicate that these factors directly influence the quality of generated information though modern communication tools, thus, issues such as identity and authorship, novelty, and impermanence of information cannot be guaranteed (Bawden, 2009).

Williamson and Eaker, work describes the relationship between IO and psychometrics science, which studies and measures mental capacity and process of information seekers with regards to demographic data such as gender, age, and life satisfaction. The investigation examined information seekers (librarians, information science and psychology students) ability to manage information, mental states of feeling overwhelmed, focus and attention to process huge amount of information or learn new topics, decision making technique in the process of choosing topics of interest, the role of technology (email, fax, phone, messages) which creates floods of information, sheer volumes, continuous development in domain of expertise and search results that pressures information seekers ability of information processing (Williamson and Eaker, 2012).

Results indicated that information seekers tend to demand especially when they progress in academic settings. Most of the information-related demands are associated with increasingly complex assignments or job environment. Therefore, it is normal to experience higher levels of IO (Williamson and Eaker, 2012).

Technology brought a lot of benefits to information science. Hence, IT, IS, and information communication in parallel enhanced the way information can be searched and retrieved. The improved information communication techniques -which can be found in the domain of search engines, RSs web 2.0 tools and social media on the internet- further participated in speeding up the process of information creation, dissemination, retrieval and availability. Hence, information seekers live in an information centric age. An information explosion that exceeds information seekers ability to process. Therefore, Modern Information Overload (MIO) is associated with four important aspects of information: storing, sorting, selecting and summarizing. If existing tools and technologies are not able to provide information seekers with search results with regards to these aspects, then information seekers will be overloaded with un required information. Consequently, MIO occurs through abundance of accessible information that or not organized, filtered, or presented in an appropriate way to maximize access to accurate and relevant use of this information (Alexander et al., 2016). Therefore, researchers still show interest in exploring the phenomenon of IO. many studies indicate that IO affects information seeker productivity when learning and/or at work.

There are many approaches that address modern IO. The most influential are algorithms behind search engines such as Google (<https://www.google.co.uk>), Yahoo (<http://uk.yahoo.com/>), Bing (<http://www.bing.com/>), Semantic Search Engines (SSE) CarrotSearch (<http://search.carrot2.org/stable/search>), Clusty (<http://clusty.com/>), and Yippy (<http://yippy.com/>). They have various mechanisms and utilities embedded techniques in them, which generate search results according to user's queries (i.e. keywords). It is difficult to find their search algorithms, because they are business secrets of these companies. However, there are numerous works which improved the performance of search engines through either new algorithms, or ranking and filtering of retrieved search results (Baeza-Yates, 2006) (Cortes et al., 2007) (Baeza-Yates et al., 2007) (Su et al., 2010). Approaches also include RSs found in e-commerce such as (Ullman, 2012) (Felfernig et al., 2007) (Adomavicius and Tuzhilin, 2005) and attempts

to improve RSs performance through CF, CBF and Hybrid filtering techniques (Zhao et al., 2011b) (Pazzani and Billsus, 2007) (Herlocker et al., 2004) (Burke, 2002).

### 2.1.9 INFORMATION OVERLOAD OR UNDERLOAD

The previous sections provided an insight on different types of IO and variety of situations when IO occurs and examples of problems information seekers who suffered from IO face when attempted to search and retrieve information. Furthermore, previous sections also described some of the proposed solutions to solve IO. The discourse in this section attempts to deliver a contradicting opinion on IO; the influence of IO on information seekers information acquisition, information processing and decision making. Hence brings forward the problem of Information Underload (IU) instead and possible causes.

IO is a two-faced coin, being overloaded with information is one extreme and being under loaded is the other extreme of information seekers problems. Many researchers adapted the problem to locate the core of the predicament information seekers face when searching for information.

Königer and Janowitz work “Drowning in Information, but Thirsty for Knowledge” discuss IO and IU in organizations. This research was triggered to itemise causes of IO in organization. Surprisingly the investigation indicated that information seekers at organizations do not only suffer IO, but also IU. Information seekers complained that “Too much information is thrown at us” and “We are not receiving enough information” at the same time. Some suffered from abundance of information and others suffered limited access to required information (Königer and Janowitz, 1995).

Janssen and Poot attempts to investigate IO at businesses and organisations also indicated that low hierarchy employees suffer from IU rather than IO. The analysed interview data claim that about 17% of employees of the complained about their inability to access the right information at the right time or not knowing that the information exists on the first place (Janssen and Poot, 2006).

Information seekers in organizations suffer from IU is justified by managers complain about the quality and lack of adequate knowledge in the delivered information. Consequently, it effects managers’ decision making, executives who need reliable

information, ordinary workers who feel that they are badly informed. IU also participate in the work progress i.e. secretaries who feel stressed when they need to learn how to use new technologies to process their daily tasks, and require help with running their word processor (Quirke, 2007).

Alexander et al., defined IU as information seekers inability to process retrieved information; thus, suffer from being un informed. IU is also influenced by irrelevant information that obscure the one information seekers need. Furthermore, IU can also mean that information seekers were not able to reach the needed information for variety of reasons (does not exist, not supplied, lack of access or inability to discover information even if they exist). Hence, IU is the consequence of under delivering meaningful information and mostly occurs between scholars, academics and the public information seekers (Alexander et al., 2016).

All the above listed situations either shows overload or underload of information and shows underutilisation of information. However, in both situations, information seekers find it necessary to stay informed in one way or another.

Therefore, a common conclusion of both IO and IU are caused by the lack of information structure, and not the results of the amount of information. Moreover, technology participated in adding to the problem. When information and technology loses its harmony, this is an indicator that there must be a problem somewhere in the process of information delivery. Hence, requires immediate action to avoid further complications that could result in problems such as financial loss, especially for profit making organizations that are entirely dependent on information quality and decision making (Königer and Janowitz, 1995) (Janssen and Poot, 2006) (Quirke, 2007) (Alexander et al., 2016).

These two contradicting problems of information retrievals can be simply solved by reducing the amount of information people create in daily activities to help control IO. Königer and Janowitz also argues that creating more information may help those who suffer from IU but add more into the load of employees who already suffer from IO.

Königer and Janowitz attempts to solve IU and IO by occupying four universal information structuring dimensions (selection, time, hierarchy and sequence) to improve information quality regardless of its carrier/technology (Königer and Janowitz, 1995).

## 2.1.10 SUMMARY

In this section, the discourse started with defining three interchangeable terms people use when describing “Information” (see section 2.1.2). For the purpose of this research, it was necessary to clarify this matter before discussing main issues of IO in general and on the internet. Sections 2.1.3 to 2.1.9, provided an empirical investigation of the different types of IO: Knowledge Overload, Technology Overload, IO in business, IO in the healthcare domain, and Modern IO and was concluded with opinions on IO or IU.

The discussion on IO and IU remains open, in both situations information causes dilemma to information seekers. The vouge impact of IO is caused by amorphous piles of documents in both forms soft/hard copies which complicates the situation even more. Therefore, negatively influence peoples’ decision making ability, task prioritisation and adequacy of processed tasks quality. Advanced technology enabled information transparency, one bit of information leads to another information (can be same or related information and interlinked). Hence, structured information is heavily recommended by researchers in early problems of IO and was labelled as “An Information Paradise” (Königer and Janowitz, 1995). Unfortunately, such an ideal information society does not exist, but putting bits and pieces of information into a structured form can make sense and help reduce the load of information the information seekers suffer from (Königer and Janowitz, 1995).

All the work that was done in the domain of Information Science aimed to provide better ways of information representation to information seekers. Even though, people continue to suffer from IO in both personal and professional life. For this reason, many researchers attempted to examine the problem of IO to figure out the causes and consequences of IO. To sum up, most of these research outcomes either led to identify IO problems or to propose a solution which will help in minimizing it. Therefore, this research provides the readers with a view of different types of IO, its causes, how IO influenced the people’s everyday life and thus, helped in forming a definition of IO.

### Definition 1:

IO is controlled by user’s desire to obtain information, and process it to gain certain knowledge that the user needs to either solve a problem or fulfil certain learning requirement in personal or professional situations. Since the user is not aware of what is

exactly needed to solve the problem or satisfy learning requirements to accomplish certain tasks. The user starts a random information seeking process; thus, create an intentional IO. Another factor which participates in IO is the information seeker educational level and background. Information seekers are confused and not clear about their information needs, this results into retrieving information more than what is needed and creates crucial IO. Therefore, no matter how experienced are information seekers they are always attempted to retrieve and store the information keeping in mind that they might need the information sometime. Consequently, the information seeker's unconscious mind creates anxiety and stress that leads to think that there exists an information that he/she still did not read. This led to define knowledge overload that influences the information seeker cognitive ability to control information floods.

Definition 2:

IO is also associated with technologies that are used as means of information management, information retrieval and information dissemination. These technologies were originally developed to help information seekers to search and retrieve information in an organised manner. Unfortunately, these technologies created IO to the information seeker in two ways. First, the information seeker is now able to view information easily, but at the same time views a huge range of the retrieved information (which is the consequence of search techniques). Second, because information seekers are from different educational level and background, their ability to cope with technological features differs. Furthermore, technology overload also manifests in new versions of software and hardware. This is one of the most annoying types of overload to information seekers because rapid development results into adding new features and services that requires from information seekers to learn additional skills before they can properly use the tool that is used for information retrieval. Hence, this technology overload is a combination of overload caused by traditional IO and adds to its technology overload problems.

Definition 3:

IO also involves information communication problems. This type of IO is dependent on information distribution strategies. Modern communication methods on the internet enabled information seekers/users to be create and disseminate information on the

internet; thus, resulted into information explosion that is out of control. This type of IO can only be controlled by users of information (audience of information).

To conclude modern IO is a combination of all previously mentioned types of IO. furthermore, it belongs to social intensive environments on the internet. Furthermore, the definition presented in section 2.1.8 is very relevant to problems discussed in this research. As described in (Alexander et al., 2016), IO can be a two faced coin. If information seekers/users and information/ services providers are obsessed with controlling IO, they may end up suffering from UI.

Next sections of this chapter, will provide an empirical analysis of early techniques of IRS (see section 2.2). discussion on RSs and their techniques will be provided in (see section 2.3). This section will be followed by investigation of search engines and their role in providing the users with excessive amount of retrieved information sources on the internet. Discussion on search engines and search results will be enriched by analysis of techniques that attempted to provide relevance in search results (see section 2.4). Information seeker/user behaviour and their role in enhancing search engines techniques to provide improved search results. section 4.5 discuss semantic web technology and delivers an elaboration of some of the researches that aimed to accommodate SWT to enhance search engine performance

## 2.2 INFORMATION RETRIEVAL SYSTEMS

Information Retrieval systems (IRS) were used by libraries to help people in finding books from the library catalogue in the 70s. Through time IRS went through intensive work to deliver to information seekers/user efficient query results, accuracy and relevance. To deliver relevant material IR systems collects metadata such as (title, year, author, publisher, subject etc) about each source of information such as (books, journals, conference proceedings, video/ audio records, and so forth).

Online libraries were the next step in IRS. They became web-based, deliver improved catalogue features based on keyword and Boolean functions, and became widely spread and available to information seekers/users in late 80s (Antelman, 2006).

Ballard labelled classical catalogues as inventory of items and described some basic problems of traditional IRS such as complex search Interfaces, limited scope, in-ability to include digital contents, lack of functionalities such as ranking, relevance of search

results, limited search techniques, and isolation from modern computing such as (social networking and so forth). Furthermore, provided a description of Next-generation catalogues which aimed to avoid classical catalogues ways of information indexing. Hence, basic functionality of any next-generation catalogue depends on discovery platforms (Ballard, 2012). Moreover, deliver user friendly, simple interfaces, improved query results and refined ranking of query results to insure relevance of retrieved information. Some of them support social integration, data visualization and some of them aims to mimic popular search engines such as google and Amazon to attract users (Ballard and Blaine, 2011).

Next-generation (NG) products cover the above listed classical catalogues problems. Some of the popular NGs are Ex Libris' Primo, WorldCatLocal (Breeding, 2007), Serials Solutions', and AquaBrowser (Karr-Wisniewski and Lu, 2010). Each provides a variety of features, most of them are open source applications, aimed to innovative features. As mentioned previously, some NG's mimics search engines in the sense of provided search results, this making use of text indexers projects. NGs enrich its contents with data such as (indexing images, cover art, and cases), table of contents, summaries and reviews, links to external sources such as google books, list results based on tags, and facet functionality to narrow results by (author name, topic, subject).

To survive the high competition in the modern digital age, NG catalogues adapted some of search engines features (keyword matching) into their solution such as AquaBrowser (Karr-Wisniewski and Lu, 2010), Koha, Evergreen, Web Voyage, Primo (Breeding, 2007). Enocre pioneered in delivering relevancy ranking in query results by embedding RightResults to refine query results. Endeca combined priority of library specification to form relevancy in their results, Primo sort's dates instead of relevance, and WorldCatLocal gives additional weight to items in local libraries.

Enocre and Koha also utilised spell check technique which appears as "Did you mean" option if proceeded query returns null. Endeca utilised similarity recommendations that features a "more titles like this" tool. Primo, Koha, WorldCatLocal and Encore took the challenge and embedded web 2.0 or Social network features into NGs to allow information seekers/users to supply the catalogue with tags, comments, and reviews.

During the past 20 years libraries expanded its services to embrace as much as possible collection of sources. However, they were not able to included electronic contents in their search techniques. Therefore, AquaBrowser (Karr-Wisniewski and Lu, 2010), Koha and Evergreen (Breeding, 2007) started to deliver additional service through linking to external resources (to incorporate article contents, provide local and remote access to electronic information).

Hence, IRS reached perfect results and is now matured system (LibraryOfCongress, 2006). However, with all advances and new technologies IRS lost its value and declined usage of IRS was noticed especially with the arrival of modern IRS on the internet (known as Search engines and RSs) (Way, 2010).

## 2.3 RECOMMENDER SYSTEMS

Another approach to addressing IO today can be found in Recommender Systems (RS) (Ullman, 2012) (Felfernig et al., 2007) (Adomavicius and Tuzhilin, 2005), which have been in the focus of interest of the research community since late 90s. They have become famous because they deploy CF and CBF techniques (Zhao et al., 2011b) (Pazzani and Billsus, 2007) (Herlocker et al., 2004) (Burke, 2002) in order to guide users *in their personalised way to interesting “items” in a large space of possible options* (Lops et al., 2011b). It is obvious that RSs were launched in parallel with e-commerce applications, where information seekers/users (i.e. buyers) were guided on how to buy items of their interest. However, RSs helped providers of “items” to manage their own IO through RSs techniques in order to target each customer with adequate “items”. RSs have penetrated many aspects of information seekers/users lives today and the use of their techniques has spread outside the e-commerce domain, particularly in situations when information seekers/users can minimize IO through filtering (Ricci et al., 2010). One of the most obvious domains where RSs is currently thriving is the field of entertainment, which heavily relies on RSs techniques in order to manage excessive information (which can be viewed as IO!) when seeking music (Braunhofer et al., 2011) (Koenigstein et al., 2011) (Celma and Lamere, 2011), videos (Davidson et al., 2010), news (Mayer et al., 2010) (Prawesh and Padmanabhan, 2011), places of interests when travelling, social relationship and many more (Pazzani and Billsus, 2007) (Lee and Lee, 2011).

Section 2.3 and its sub-sections provides an overview of RSs, issues RSs solved and a range of examples and implementations of RSs in a variety of environments on the internet (see section 2.3.1). furthermore, discussion on CF technique that gained the attention of many researchers for its distinctive role in improving RSs (see section 2.3.2.1). Section 2.3.2.2, discusses CBF technique and supported with examples of the advantages and usage of CBF to improve RSs and understand the contents of the web. This is to provide better recommendations of users of SNS on the internet. Moreover, explore Hybrid filtering techniques and the benefits it brought to RSs (see section 2.3.2.3). Last section discusses Tagging, Annotations, and Folksonomies and their role in improving RSs recommendations.

### 2.3.1 OVERVIEW OF RECOMMENDER SYSTEMS

RSs are popular because they are heavily used for information filtering and retrieval on the internet. Hence, one of the most exploited solutions which have addressed modern IO in the last 2 decades and were created to help users find items of interest. Early RSs application in mid 90s were developed to help both information seekers and providers to manage information and services in the domain of e-commerce. Therefore, many RSs utilise range of modeling techniques of information, items, and services to derive refined recommendations to the information seeker/user. Hence, attempts to improve RSs performance covered many of basic information retrieval techniques and were used in variety of situations where information seekers/users' need items, services and information to be recommended (Felfernig et al., 2007).

RSs became popular because they were based on well known information filtering techniques (CF and CBF) which were developed to manage IO in businesses in early 90s. Hence, many solutions in the domain of RSs utilised CF and CBF (Adomavicius and Tuzhilin, 2005) (Ullman, 2012). Both techniques solely depended on interpretations of user profiles and interactions with other users or the system. Therefore, both filtering techniques utilise information seekers/ users and items/services modeling techniques to deliver filtered recommendations (Felfernig et al., 2007).

Some of the heavily used models of both user and item profiling in RSs can be found in the work of Felfernig et al. that utilised nearest-neighbour method for ratings matrix (Felfernig et al., 2007) (Bourke et al., 2011), Top-K model to profile users activities (Tayebi et al., 2011), close proximity model for co-occurrence of names (Kautz et al., 1997), prediction models for business process (Ullman, 2012), probabilistic model user profiling on SNs (He and Chu, 2010), learn and aggregate the weighted average of ratings model (Hoens et al., 2010a), Latent Dirichlet Allocation (LDA) model and unified probabilistic model for topics discovery and prediction of missing data in user profiles (Hariri et al., 2013). Vector space similarity model (VSSM) is a CBF based content model that index terms (Pazzani and Billsus, 2007), Pearson Correlation Coefficient model (PCCM) in hybrid RSs (Yu et al., 2011), user learning model (Pazzani, 1999).

Other than RSs in e-commerce, attempts to deliver recommendations to information seekers/ users can be found in variety of other domains. Adomavicius and Tuzhilin

investigates RSs to solve problems of domain specific recommendations (Adomavicius and Tuzhilin, 2005). Felfernig et al. examines the role of RSs in IO manifestation in situations where excessive and complex amounts of information are available and outstrip the user's capability to go through them and reach a proper decision (Felfernig et al., 2007). Similarly, Ricci et al. also investigates the relation between IO and RSs and their techniques in a range of domain specific situations (Ricci et al., 2010). Ullman attempts to enhance RSs belongs to distinctive online communities: news article and item recommendations in e-commerce (Ullman, 2012). Hornung et al., research of RSs attempts to model list of correct process fragments for business process modelling to reduce frequency of structural errors in business processes and minimise manual process modelling time (Hornung et al., 2007).

He and Chu collects social intensive environment (SNs) data to deliver enhanced RSs recommendations with personalised contents (He and Chu, 2010). Similarly, Bourke et al. attempts to generate recommendations by incorporating social information in their recommendation process (Bourke et al., 2011). Whereas, Nunes and Hu investigates the role of personal information in RS (Nunes and Hu, 2012).

Continuous development in the domain of RSs also included security and privacy of information seekers/ users. Since privacy is a crucial issue when delivering personalised contents on the internet. Hence, Hoens et al. research of healthcare information sources on the internet pressures to consider information seekers/ users privacy as top priority in the process of recommendations. Likewise, Tayebi et al.'s extends the research on RSs to investigate potential crime suspect in cybercrimes (Tayebi et al., 2011).

There are also many attempts to create group RSs. Jameson and Smyth, for instance, investigates RSs techniques to understand essential requirements to model user centric RSs (Jameson and Smyth, 2007). Whereas, Guzzi et al also explores possible methods with regards to information seekers/ users preferences and similarity between group members to improve RSs recommendations (Guzzi et al., 2011). Furthermore, Seko et al., extends the investigation of RSs by analysing user to user interactions to model group recommendation (Seko et al., 2011).

Gradually, RSs advanced to include patterns of usage (Pazzani and Billsus, 2007) in the process of recommending mutual friends (Felfernig et al., 2007), contextual information (Adomavicius and Tuzhilin, 2005), topics of interest (Ullman, 2012).

Moreover, RSs recommendation process was extended to include items, services and information rating, tagging, and annotation of (Zhao et al., 2011b), similarity traits (Herlocker et al., 2004), and so forth.

RSs also played a crucial role in the delivery of personalised contents and services to information seekers/ users on the internet. Therefore, users' activities and query logs on the system were collected and analysed to identify information seekers/ users interest. Consequently, RSs take into account demographic data of users in the process of recommendation. Modern technology on the internet changed the nature of information creation and dissemination on the internet. Therefore, RSs developers showed some interest in analysing contents on (SNs, social media and collaborative methods) to deliver better recommendations.

Hence, cross dicipline paragdims discussed above indicates that RSs can be used to solve many informtion filtering issues in domains other than e-commarce. Consequently, reseachers suggested variety of solutions to informtion filtering based on problems and situations. For instence, Adomavicius and Tuzhilin proposed approach aimed to improve RSs recommendations by interpreting users-items relationship and include contextual information, layered rating criteria and deliver more flexible recommendations. (Adomavicius and Tuzhilin, 2005).

Ullman introduced the long-tail model which is based on CBF and CF functionalities to enhance the performance of RSs. The model utilise utility matrix of items and user preferences in online communities to assign value and marks of degree of preference. The recommendation process takes into account user-items rating information and sparsity in utility matrix. Hence, utilise prediction techniques in such situations. Furthermore, the long-tail model in this proposal was used to help in the decision making process. to deliver popularity in an online communities (Ullman, 2012).

Thus, Ullman's RS model adopted the long-tail to enforces online communities to recommend items to individual users. The proposed solution in the work of Ullman can be applicable in domains where RS can recommend products, movies and news article as described in the figure 2.2 below (Ullman, 2012).

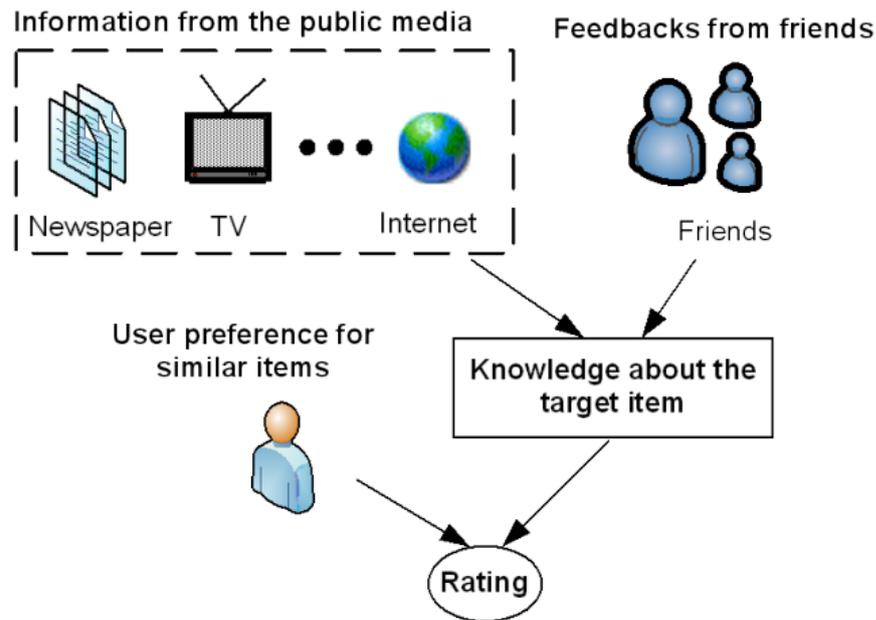


FIGURE 2.2 THREE MAIN ELEMENTS OF THE PROPOSED NSRS MODEL IN (ULLMAN, 2012) .

Hornung et al., proposed method enforces transparency of business process models when defining process elements and must be correlated to the choice of fitting process fragments to reduce modeling time. Furthermore, allow flexibility in the process of selecting fragments and consider user competency level to deliver flexible recommendations (Hornung et al., 2007).

To guarantee information seekers/ users' privacy, Hoens et al. proposed to include aspects of trustworthiness and relevance in recommendation process through controlled shared personal information on online SNs (Hoens et al., 2010a). Hoens et al. further extended the investigation on privacy to healthcare information seekers (patients) in the process of recommending appropriate physicians to diagnose and treat medical condition. Hence Hoens et al. two frameworks Secure Processing Architecture (SPA), to secure submission process of patients ratings and Anonymous Contributions Architecture (ACA) to hide patients identity in the rating process (Hoens et al., 2010b).

He and Chu's proposed a probabilistic model that profiles users on SNs to deliver personalised recommendations and help in the process of product marketing. And further apply semantic filtering of social networks to improve the performance of RSs (He and Chu, 2010).

Bourke et al. proposed to utilise neighbourhood formation approach to incorporate social information to guarantee accuracy in the recommendation process. to do so the RS will allow to select users through manual selection of friends from social graph, communication frequency will create simple metric to automatically select users through frequent communication patterns and similarity among targeted users, and global similarity corresponds to the neighbourhood formation approach. (Bourke et al., 2011). While, Tayebi et al. proposed to utilise Top-K model potential suspects with the association of rule-based methods to secure accurate recommendations (Tayebi et al., 2011).

Jameson and Smyth proposed to take into account in the process of recommendation the characteristics of members of a group to address decisions problem in RSs (Jameson and Smyth, 2007). Whereas, Guzzi et al proposed an interactive multi-partying critique for group recommendations. This technique learns about individual preference in a group of users and makes recommendations based on similarities among other group members (Guzzi et al., 2011). Seko et al. suggested to use behavioural role, tendency and power balance between group members to create group recommendations. Therefore, the proposed RS algorithm to estimate appropriate or novel content for groups of people (Seko et al., 2011).

Scarceness of information were one of the early problems of RSs. He and Chu experimental research attempted to improve prediction accuracy of RSs, and cold start issues from CF technique (He and Chu, 2010) Whereas, Tschersich's research investigated the role of unavailable, inaccessible, or incomplete user information and include location information in an ad-hoc basis. this to locate causes of inaccurate and reduced quality recommendations in mobile group RSs (Tschersich, 2011). Hariri et al. exploited the Latent Dirichlet Allocation (LDA) model to collect user-item data and unified probabilistic model to interpret the data in the process of latent topics discovery and prediction of missing data in user profiles (Hariri et al., 2013).

It is very difficult to define context in modern computing because it is not solely confined to a location which requires to run context aware applications. Therefore, information seekers/ users' queries have become an important source of contextual information for creating context aware recommendation. Hence, utilisation of contextual information from user profiling techniques can be found in many disciplines, including

e-commerce service personalization, information retrieval, ubiquitous and mobile computing, data mining, marketing, and management (Adomavicius and Tuzhilin, 2008).

Abowd et al., attempted to improve RSs analysed information seekers/ users role in mobile and ubiquitous computing (Abowd et al., 1999). Anand and Mobasher also emphasised on important role of contextual information and takes into account long and short term memory in modelling user centric RSs (Anand and Mobasher, 2007).

Adomavicius and Tuzhilin extended their recommendation techniques to include context aware models to deliver highly relevant items recommendations based on personality traits utilised in the RSs (Adomavicius and Tuzhilin, 2008). Similarly, Adomavicius et al. investigates relevance levels in RSs recommendations. Hence, utilise several context aware model to enhance RSs performance (Adomavicius et al., 2011).

Therefore, Abowd et al.'s proposed a framework for the development of context-aware applications that interprets users profiles in the process of recommendations (Abowd et al., 1999). Adomavicius and Tuzhilin proposed three different algorithmic paradigms – contextual pre-filtering, post-filtering, and modelling – for incorporating contextual information in the recommendation process (Adomavicius and Tuzhilin, 2008). Adomavicius et al. suggested Context-Aware Recommender Systems (CARS) applicable in variety situations which requires contextual information in the process of recommendations (Adomavicius et al., 2011). Whereas, Hariri et al., proposed a model that integrate user profiles, item representations, and contextual information to compute the probability of each item and user profile for music recommendations (Hariri et al., 2013).

In addition to the above mentioned RSs solutions, researchers also considered adding prediction of recommendations (Bellogín, 2011), (Bellogín and Castells, 2010), (Hauff et al., 2008), (Katz et al., 2011), testing the RSs on missing data (Steck, 2010) and solving the problem of diversity, similarity and accuracy in RSs (Zhou et al., 2010), (Jojic et al., 2011) to recommending item popularity (Steck, 2011) and interactions between users and RSs (Knijnenburg et al., 2011).

This section, provided an overview of RSs in general and gave examples of distinctive researches that attempted to identify RSs applications in multiple problem domains: e-commerce, e-government, online learning, business, and so forth.

Furthermore, it also described successful implementations of RSs solutions that aimed, somehow, to solve modern IO, by either incorporating information seekers/users and domain information, users' interaction with other users and user-items relationship on the system into the recommendation process. Hence, the presented review in this section highlighted some of the main problems of information retrieval on the internet and the solutions RSs offered, such as:

- enhance RSs recommendation process by utilising of CBF and CF techniques.
- deliver relevance in recommendations based on user-item profiling.
- Take into account privacy and security of information seekers/ user.
- Improve accuracy of recommendations based on contextual information.

The next subsections of this chapter, will provide an elaboration on RSs techniques that heavily participated in the proliferation of RSs in many disciplines. Therefore, the next section, 2.3.2 itemise RSs techniques into four main sub sections; section 2.3.2.1 will give a detailed description of Collaborative Filtering (CF), section 2.3.2.2 will provide the reader with detailed descriptions of CBF, section 2.3.2.2 will provide the reader with detailed descriptions of Hybrid Filtering, and section 2.3.2.3 will discuss Tagging, Annotations and Folksonomies role in RSs.

## 2.3.2 RECOMMENDER SYSTEMS TECHNIQUES

The previous section provided an overview of research in the domain of RSs in domains other than e-commerce. It briefly highlighted that almost all RSs are solely based on CF and CBF techniques deliver recommendations to information seekers/ users. Furthermore, RSs were also used to alleviate modern IO. However, they were also used to filter information in intensive social communities on the internet. Therefore, this section like to start with discussion of three popular RSs techniques that occupied two decades of intensive research in the domain of information filtering and recommendation on the internet. The reason behind the selection of techniques was based on our research objectives in **section (1.3)**; thus, encouraged the investigation of the role of RSs and their technique in the creation and retrieval of information in modern IO. Therefore, the following subsections, will provide the reader with an overview of RSs techniques, problem domains, technologies used to improve RSs and the shortcomings of these techniques in relation to OB in this research.

### 2.3.2.1 COLLABORATIVE FILTERING

At the heart of all these RSs are various techniques, often associated with the filtering of data. One of the most famous and frequently used is collaborative filtering (CF) which was developed to solve the problem of data flow through emails (Goldberg et al., 1992). Goldberg et al. was concerned about the excessive amounts of e-document circulation among people that created unnecessary IO. Hence, suggested that IO can be controlled by identifying the role and influence of users who create and distribute e-documents. Consequently, introduced CF technique to understand the information seeker/user's role in the process of information dissemination in electronic environments (The Email) (Goldberg et al., 1992).

Goldberg et al. pioneered in developing the first RSs that is based on CF recommendation technique for information filtering in the workplace. Hence, defined CF as:

***“people collaborate to help one another perform filtering by recording their reactions to documents they read. Such reactions may be that a document was particularly interesting (or particularly uninteresting). These reactions, more generally called annotations, can be accessed by others' filters. One...”*** (Goldberg et al., 1992).

Therefore, CF technique participated in developing Tapestry Mail System (TMS) which identifies the relationship between minimum of two documents on the system (email and reply). The tapestry architecture components of documents modeling are “indexer” that reads documents from external sources and pars them into indexed fields for query reference. “Document store” is a DB that stores and maintains indexes of documents. “Annotations” stores document annotations, tags and implicit feedback for the recommendation process (Goldberg et al., 1992).

In parallel to the development of CF technique in (Goldberg et al., 1992). Kautz et al. proposed ReferralWeb, an agent-based programs that addressed practical communication needs and allow the reconstruction of the web to improve visualization and searching the contents of the social web (Kautz et al., 1997).

The ReferralWeb allowed the identification of existing SNs and creation of new communities that will help individuals effectively benefit from large existing networks

of colleagues. Furthermore, ReferralWeb also enabled searching people who are closely associated with some known and trusted expert on the system. Co-occurrence of names in close proximity in documents publicly available on the Web were utilised in relation to sources (links, authors, publications, citation) in the recommendation process (Kautz et al., 1997).

The successful implementation of CF in TMS and ReferralWeb triggered researchers' interest to further exploit CF in other information intensive environments on the internet (Kautz et al., 1997). Furthermore, CF technique was also used to improve information representation (Schafer et al., 2007), and was combined with knowledge base information to deliver similarity (Tran, 2007). Automation of CF techniques was another attempt that involved prediction of user-item level of interest (Herlocker et al., 2000),(Herlocker et al., 2004), Shared CF that takes into account shared rating and annotation and privacy of users at the same time (Hu and Pu, 2011) (Zhao et al., 2011b), improved accuracy (Lathia et al., 2009) and incorporate tagging, annotation and folksonomies information to deliver personalised recommendations in SNs (Kim and El Saddik, 2011).

Although CF technique was developed to manage information problems in email systems, it was also one of the early techniques used to control IO problems in e-commerce. Hence, researchers utilised CF techniques with other information storing, presentation and retrieval methods. Tran research investigates the possibilities of exploiting CF to identify similarity between users and knowledge based approaches (to reason whether a product can meet user requirements) to deliver better RS that will help narrow selection options to buyers in e-commerce (Tran, 2007). Moreover, Kautz et al. investigate CF technique to ease the process of information retrieval on SNs (Kautz et al., 1997). Whereas, Schafer et al. adopts CF to personalise web contents and enhance information representation (Schafer et al., 2007).

Herlocker et al. investigates methods to utilise CF technique in the process of identifying user-item relationship and model predictions level of users interest in particular item (Herlocker et al., 2000). Furthermore, to measure the quality of CF based predictions Herlocker et al. investigate the role of user tasks and datasets to deliver refined CF based recommendations (Herlocker et al., 2004).

Hu and Pu investigate cold start, sparsity problem, lack of user rating and new user problem which prevents RS from delivering personalised recommendations (Pu et al., 2011). Likewise, Lathia investigates the possibility of utilising CF technique and Adoptive Information Sources (AIS) to solve the problem of sparsity (Lathia et al., 2009). Whereas, Zhao et al. investigate CF technique to utilise shared information in the recombination process (Zhao et al., 2011b).

Hence, a sequence of intensive research on CF technique aimed to deliver solutions for many information filtering problems. Schafer et al., for instance, proposed a system that analyse users interactions in online communities and allows to create user profiles that describe users' involvement with the system, analyse the contents of users' pages, and identify the relationship between users and the rating of items (Schafer et al., 2007).

Tran proposed an architecture design for hybrid RS that will take into account users and products information from collaborations on the system and add to it ratings and description of products from the knowledge base to provide buyers with shortlist of products based on similarity between interest profiles of users. (Tran, 2007).

Whereas, Herlocker et al. RSs proposed to utilise CF to develop Automated Collaborative Filtering (ACF) systems and design user-item prediction system. It also takes into account user-item relationships, users' interest level an item to calculate similarity interest among users (Herlocker et al., 2000) (Herlocker et al., 2004).

Instead of building new RS, Zhao et al. suggested to build a Shard Collaborative Filtering (SCF) that utilised user-item profile model (item neighbourhood), user matrix model (users shared rating and annotations of items) and prediction algorithm to overcome sparsity of data. Hence, utilised data from other parties (contributors and benefactors) participated in improving CF performance and took in consideration privacy traits of other parties at the same time. (Zhao et al., 2011b).

Pu and Hu proposed a model is a rating-based CF that incorporate both linear user personality information, rating information and cascade mechanism to leverage resources and enhance RSs performance and recommendations. Furthermore, major similarities between users and rating prediction by collecting neighbourhood formation and compare them to other user's rating patterns (Pu et al., 2011).

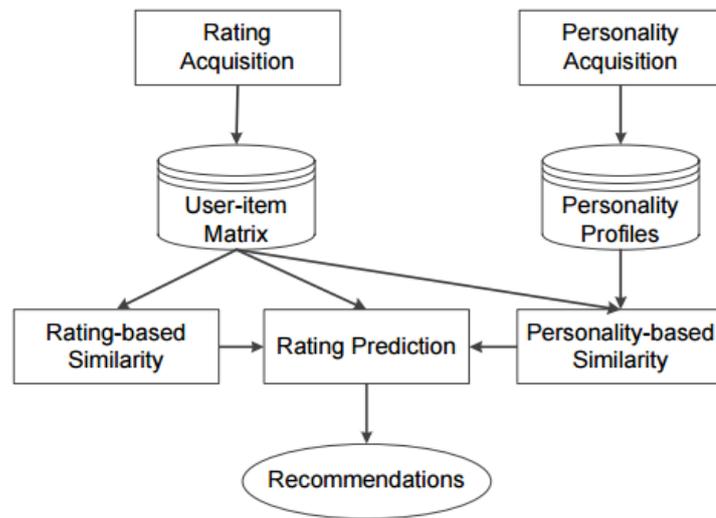


FIGURE 2.3 MAIN COMPONENTS OF RATING-BASED CF RECOMMENDATION MODEL (HU AND PU, 2011)

Lathia et al. suggested to mine web data collected from external data sources to eliminate the sparsity problem and categorise users based on external communities to generate accurate recommendations for users of video streaming websites. Hence, improve prediction accuracy and data quality in parallel (Lathia et al., 2009).

Kim and El Saddik proposed to utilise CF technique to interpret the semantics of user collaborations in SNs where tagging and annotation is implemented to deliver personalised tagging and annotation based recommendations in folksonomies (Kim and El Saddik, 2011).

To sum up, CF techniques were heavily exploited due to their role in reducing excessive amount of information traffic caused by early technology solutions such as email systems. Therefore, CF build users profiles by utilising profiling model (see 2.4.1) that allow to collect data of users and help enforce criteria that majors similarity traits between user to user, user to items/services, and items to items on the system. Hence, RSs based on CF techniques creates user and items/services profiles which contains explicit interaction details.

User profiling models mainly contains data such as demographic information, viewing patterns, rating, tagging, and annotation of items, and so forth. Whereas, items/services profile will provide details about rating, annotation, items/service

description, viewing time, number of clicks, and information about users who viewed common items/services.

By profiling users and items/services the RSs analyse the profiles for future recommendations. Hence, CF technique relays on users' past behaviour to recommend new items by comparing two users who rated the same item and so forth. This technique is also known as similarity based recommendations (Herlocker et al., 2004), (Zhao et al., 2011b).

### 2.3.2.2 CONTENT-BASED FILTERING

Content Based filtering (CBF) technique is another RSs essential method of information filtering and retrieval in the process of recommendation. CBF was discussed in many scientific research but was properly defined for the first time in Goldberg et al. (Goldberg et al., 1992). The discussion in Goldberg et al. also introduced CBF as RSs technique which indicated that CF technique utilised CBF technique. Another root of CBF technique can be found in Balabanović and Shoham which Proposed Feb a content based collaborative recommendation RS that aimed to eliminate many weaknesses found in CF and CBF approaches. Furthermore, help information seekers/users (online readers) cope with the massive content available on the internet. Hence, Balabanović and Shoham delivered new opportunities to reduce IO (Balabanović and Shoham, 1997).

Similarly, Pazzani and Billsus proposed a CBF based RS that recommended items based on their descriptions and utilised user profile in the process of recommendations. Pazzani and Billsus claimed that the proposed RS was experimented in a variety of domains (e.g., Cuisine or Service, news article, book, and entertainment) to examine its applicability in understanding domain specific item recommendations. Hence, CBF can be a common mean to create profiles of users and items because CBF is based on models that are able to analyse not only structured data as presented in DB, but also its ability to draw conclusions from contents such as rating, user comments and feedback, etc. Therefore, CBF RSs process three main elements to guarantee successful recommendation as follows:

- collect knowledge of an item and how its represented,
- create user profile which contains explicit information of user's preferences and activities.

- collected information on items and users are analysed to feed the machine learning model to create user model.
- clusters and groups data from the user model.
- runs the nearest neighbour algorithm of similarity model to find users with similar profiles.
- utilise the Vector model algorithm is then used to major relevance feedback (Pazzani and Billsus, 2007).

Hence, CBF technique was originally developed to help information seekers/ users in early systems such as (email) and later in SNs to:

1. solve problems of an experienced information seekers/users when searching for information (i.e. short queries which resulted into too much results or long queries which eliminates everything).
2. Deliver filtered information based on utilised user and items information from past activities to model diverse aspects in the process of recommendation and reduce IO.

### 2.3.2.3 HYBRID FILTERING

Filtering techniques have also been combined to improve the retrieval, organisation and management of information in Social Networks (SNs) and on the internet in general. Almost all previously discussed RSs in (see sections 2.3.1, 2.3.2.1 and 2.3.2.2) delivered Hybrid RSs because both early attempts described in (Goldberg et al., 1992) build a CF Recommendation based analysed data collected through CBF technique. Similarly, Balabanović and Shoham also combined CF and CBF to enhance recommendations and reduce electronic IO caused by email and online information resources (Balabanović and Shoham, 1997).

Hence, this section elaborates a little more on few other Hybrid filtering based techniques in RSs which placed CF and CBF together in various frameworks to deliver improved RSs recommendations.

Researches interested in RSs and their techniques attempted to develop Hybrid RSs which can be used in diverse situations when recommending items and services to information seekers/users on the internet. Pazzani investigates available information on the internet to evaluate whether a webpage or news article can be recommended to information seekers/users to help minimise IO (Pazzani, 1999). Burke explores possible methods to integrate Knowledge-Based (KB) technique and CF technique in RSs to prune

large repositories in the domain of e-commerce and address information seekers/users preferences in order to find the right product (Burke, 1999). Burke extends the investigation on hybrid RSs to test whether they can be altered to fit the process of recommending restaurants (Burke, 2002).

Whereas, Tran attempts to investigate ways to assist the online buyer in decision making in the process of selecting the preferred product from a massive collection (Tran, 2007). Yu et al. attempts to investigate the possibility to enhance recommendations in SNs. Hence, incorporates CF algorithm with similarity models (VSSM) and (PCCM) to guarantee successful recommendations based on user preferences (Yu et al., 2011).

All solutions in the domain of RSs meant to enhance the understanding of information in relation to information seekers/ users preferences, similarity to past behaviour, activities, viewing and so forth. Accordingly, researchers and service providers focused on profiling users and items, products and services on the internet.

Hence Pazzani proposed to utilise models that mine webpages contents, item rating by users, and user demographic data to improve the performance of the RSs to recommend webpages and news articles (Pazzani, 1999). Pazzani extended his work and attempted to examine the proposed solution in the domain of online recommendation of restaurants to information seekers/users (Pazzani, 1999).

Burke proposed EntreeC Hybrid RS that guaranteed successful recommendation, based on minimum data collected by both techniques (Burke, 1999). Furthermore, Burke extended the proposed model to include semantics of ratings obtained from KB technique to fit in online restaurants recommendation based on CF and a knowledge-based technique (Burke, 2002).

Hence, Tran proposed an architecture design that utilised CF and a KB technique which can be applicable in variety of information, service and product recommendation. The Hybrid RSs selects one of the two techniques in the recommendation process after examining user's behaviour to support to enhance recommendations (Tran, 2007).

Yu et al.'s proposed using an Adaptive Social Similarity (ASS) function based on the matrix factorization technique (Koren et al., 2009) deliver improved item prediction based on user-item matrix through social relations data (Yu et al., 2011).

Researchers also combined filtering techniques for the purpose of solving domain specific problems. Borger and Bosch proposed a hybrid filtering technique based on the two famous techniques – CF and CBF – to allow users of Delicious, CiteULike and BibSonomy to manage their social bookmarks of websites. This was to allow information seekers/users to store, organize, and search the bookmarks of preferred webpages. This proposal was developed and employed in the domain of tagging/ annotation, to understand information seekers/users' decisions when tagging/annotating their favourite bookmarks.

Another example of hybrid filtering in the domain of tagging is the work of Wu et al., which utilized CF technique to create a new filtering technique – Collaborative Tagging Systems (CT) – to support knowledge management activities in the domain of tagging. Moreover, the hybrid CT systems will allow information seekers/users to search content previously tagged by the information seekers/user, find users with shared interests, and explore tags created by other users on the system.

#### 2.3.2.4 TAGGING / ANNOTATIONS AND FOLKSONOMIES

The pervious section briefly mentioned attempts that took into account modern information filtering techniques more responsive to information seeker/user queries (Tagging and Annotations) in the process of delivering Hybrid RSs. They come in the format of Tagging of things in internet sources and allows grouping and categorisation to make contents easy to view by topic. Tagging was originally inspired by the traditional document repositories or digital libraries method of organization and assignment of keywords. Tagging/ Tags are metadata keywords used to describe, identify or classify items on the internet and are usually informal or personal description of the content creator. Tags are commonly used to label online contents such as blogs, webpages, pictures, videos, users, topics and so forth. They are used almost everywhere on the internet, and can be linked to allow the search of similar contents. Publicly available tags can act as bookmarks allowing the indexing of web links of these tags with common acronyms or numbers. Many online platforms provide tag service to their users as sidebars or tag clouds for easy content discovery. Tagging/Tags are also known as Blog Tagging, Social Bookmarking (Marinho et al., 2011), Social Tagging (Belen et al., 2010), Annotations (Marshall and Brush, 2004) and Folksonomies (Mathes, 2004).

Folksonomies are collection of user annotated resources with their tags, network of interrelated users, resources and tags, which is very valuable for understanding the nature of internet sources (Mathes, 2004).

Tagging/ Folksonomies, in general, appeared as a consequence of having Web 2.0 (O'Reilly, 2009) applications. The influence of web 2.0 technologies (wikis, blogs, podcasts, folksonomies, mashups, SNs, virtual worlds, and crowdsourcing) on information seekers/users everyday live is obviously undeniable. Web 2.0 technologies eased communication and collaboration in online environments, especially businesses and rapidly paved their way into corporate technologies.

However, concerns always accompanied these web 2.0 technologies that directly deal with information on the internet. Andriole suggests that it is important to examine these technologies thoroughly before introducing them as tools to represent, manage and retrieve information to organization to insure full utilisation of the tool(s) (Andriole, 2010).

A survey on web 2.0 technologies that was conducted by Andriole indicates that tagging is a solution that could positively affect productivity (Churchill, 2009), (Wrightemail et al., 2008), where users are becoming both producers and consumers of information on the Internet.

Almost all Web 2.0 platforms allow information seekers/users to post various resources and tag them freely according to their own perception of “what an appropriate tag is for a particular internet source”. However, technology can also deliver numerous problems. Free tagging allows tags to have multiple meanings, they are affected by synonymies, and they confuse different levels of abstraction of specific topic. Despite these issues, tagging is useful when satisfying information seekers/user’s needs when publishing and retrieving information from the internet.

Hence, researchers examined the role of tagging in the process of information delivery to information seekers/users on the internet. Golder and Huberman, attempts to analyse the structure of CT systems such as (Del.icio.us platform ) as well as their dynamic aspects, user activities, tag frequencies, types of tags used and popularity in bookmarking (Golder and Huberman, 2006).

Tags were also utilised to deliver Personalised Tag Recommendation (PTR) as described in Landia and Anand. The usefulness of PTR is that they are utilised by document management application and social bookmarking websites. Hence, they are heavily indexed by systems or generated to deliver new document (Landia and Anand, 2009).

Similarly, Musto et al. investigates tag RSs, indexed resources, methods of retrieving information based on similarity and relevance, and techniques to exclude irrelevant tags to enhance delivery of relevant recommendations of keywords for annotation (Musto et al., 2009).

Whereas, Zhao et al. investigates user profiles and tagging techniques that allows semantic description of shared resources to enhance CF based recommendations in social intensive communities on the internet to major similarity of rated items (Zhao et al., 2008). Troussov et al., encouraged to interpret the domain where recommendations are needed to improve the quality of recommendations (Troussov et al., 2009).

Thus, they proposed a dynamic model of CT that predicts the relative proportions of tags patterns within a given URL and relates them to imitation and shared knowledge (Golder and Huberman, 2006).

Landia and Anand proposed to utilise clustering of existing documents in order to identify sets of similar documents and as a result identifies sets of users whose tags may be propagated to the current target user-document pair (Landia and Anand, 2009).

Hence, Musto et al. proposed Tag RSs (STaR) which suggests a set of relevant keywords for annotation based on two assumptions: 1) the more resources, the more similar to share common tags, 2) TR should be able to exploit and extract tags from previously used tags by the users (Musto et al., 2009).

Whereas, Zhao et al. proposed collaborative RSs which interprets semantic distance among tags assigned by different user to improve the effectiveness of neighbour selection in social platforms. Thus, calculates the semantic similarity of tags and produce similarity metric to deliver refined recommendation to information seekers/users (Zhao et al., 2008).

Spreading Activation (SA) model interprets human semantic memory (Quillian, 1966). SA was enhanced for task retrievals and spread across domains ISs developers

employed SA to model users as nodes in directed graphs and links, sparsity problems in CF, and the development of an SA based algorithm that gives better results than the traditional memory-based approach.

Similarly, Hussein and Zigler investigates SA techniques to understand the semantics of the environment and information seekers/users and attempts to model interests and weights for relations (Hussein and Ziegler, 2008).

Hence, Trousov et al. proposed SA approach that utilised the asymmetric measure of relevancy (proximity) of two nodes, weight of multiple connections between two nodes and included paths and graph-structure to connect the nodes (Trousov et al., 2009).

Whereas, Hussein and Zigler proposed SPREADR a model-based technique that creates context-adaptive web applications and utilised SWT ontologies to stores concepts, instances of the domain, context factors (such as location, time, and user role) to form a semantic network based on SA approach. The figure 2.4 below depicts the process of modelling the collected semantics of the environment to trigger an activation flow through the network and generate a website and control the adaptive behaviour of the system. Furthermore, learn user preferences through implicit feedback (Hussein and Ziegler, 2008).

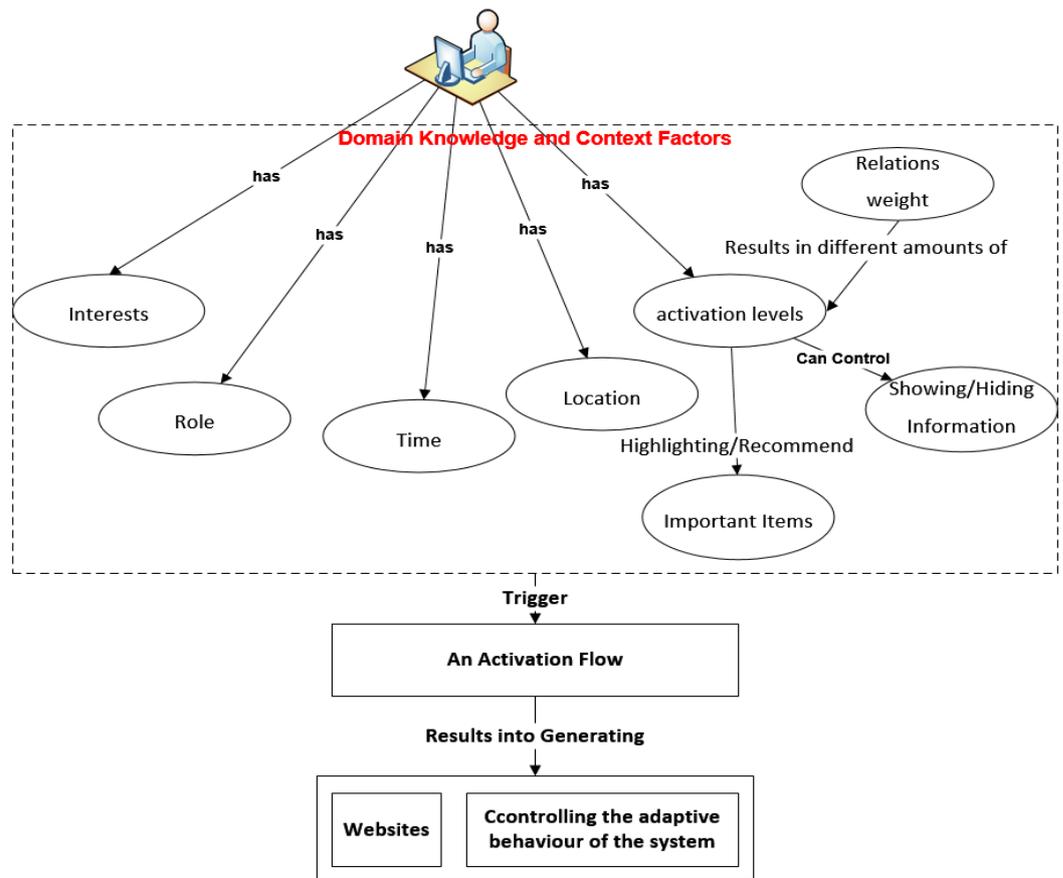


FIGURE 2.4 DEPICTS THE MAIN CONCEPTS AND INSTANCES OF SPREADR MODEL.

The hierarchical nature of Social Tagging systems (ST) used in RSs might not be the best counterpart to traditional RSs techniques (Marinho et al., 2011), (Milicevic et al., 2010). Hence, Bar-Ilan et al. attempts to examine structured and unstructured tagging of images in situations which allows users to either use existing tags or insert pre-defined metadata elements to describe them (Bar-Ilan et al., 2008).

The research community also attempted to extend user-item environments towards the contexts that requires more attention to their interconnectivities and create some kind of semantic relationship between users and items.

Belen et al. investigates possible methods to take advantage of the Web 2.0 applications and ST data and regarded as folksonomy that describes users interactions on the system (Belen et al., 2010).

Belen et al. proposed Queveo.tv RS that utilised both CF and CBF techniques and data obtained from social tagging (ST) to model user profiles that learns about each user's

viewing and VSSM to measure users-program similarity. Hence, generate CB recommendations to improve the coverage and diversity of the suggestions of digital TV systems and deliver customised TV content recommendations (Belen et al., 2010).

Another stream of tagging is present in the research of Collaborative Tagging (CT). Since CT is associated with folksonomies and mostly considered as an additional mechanism of sharing, annotating and information discovery on the internet. Hence, it was in the focus of many researchers who attempted to enhance RSs performance because CT (Wal, 2007).

Therefore, the possibilities of using CT are numerous. Gemmell et al. investigates CF techniques and CT to enhance graph based recommendation (Gemmell et al., 2009a). Whereas, Macgregor and McCulloch explores methods that aims to use CT as a knowledge organisation and resource discovery tool (Macgregor and McCulloch, 2006).

Whereas, Zhao et al. investigates the possibility to employ users tagging and annotation data collected from webpages contents to enhance ranking algorithms and deliver personalisation based tag-annotation aware search engine results (Zhao et al., 2011a).

Hence, Gemmell et al. proposed Hybrid TRS that utilised item based CF, user based CBF, informational channel in folksonomies, item popularity to deliver enhanced recommendations (Gemmell et al., 2009a).

Hotho et al. investigates information seekers/users ability to retrieve relevant information and FolkRank algorithm to deliver personalisation based on ranking of items from folksonomies and recommend users, tags and resources (Hotho et al., 2006b), (Hotho et al., 2006a)..

Furthermore, Folksonomies and CT enabled information seekers/users to share, annotate and search for information sources on the internet. Accordingly, Gemmell et al. investigates methods to simplify the annotation process and promote tagging to reduce noise in data and eliminate discrepancy in redundant tags, and avoid ambiguous tags (Gemmell et al., 2009b).

Whereas, Borgers and van-den-Bosch investigates folksonomies to predict hidden bookmarks users might like based on the user profile to recommend items in social bookmarking websites (Bogers and van-den-Bosch, 2009).

Hence, Hotho et al. proposed a formal model algorithm that collects data and models the relationship of a folksonomy tuple and utilised PageRank a search algorithm to develop FolkRank technique. FolkRank exploited structure of folksonomies to determine overall ranking, specific topic-related rankings to discover communities within folksonomies and structure search results (Hotho et al., 2006b), (Hotho et al., 2006a).

Whereas, Gemmell et al. proposed to utilise K-Nearest Neighbour algorithm and user modeling techniques to incorporate users, resource and tag information and calculate similarity among users. Hence, promote users, boost tags and improve the coverage and accuracy of K-Nearest Neighbour algorithm (Gemmell et al., 2009b).

Limpens and Gandon suggest to employ formal languages and ontologies from SWT to overcome problems in folksonomies (Limpens et al., 2008). Whereas, Wu et al. proposed a model that extends existing CT systems capability and address methods to identify communities of users who share common interests and create scalable structures to overcome IO (Wu et al., 2006). Borgers and van-den-Bosch proposed to interpret Folksonomies and item metadata to waver CF algorithms problems, tag overlapping to deliver recommendations based on the difference between two probability distributions of data and calculate user/item similarities (Bogers and van-den-Bosch, 2009).

Web 2.0 tools delivered a mashup of technologies which became a common practice in Computer Science (CS) and ISs in the process of information representations, management, retrieval and information discovery information seekers/users on the internet.

Nowadays, information seekers/users mostly acquire reading material from information sources on the internet. Hence, service providers heavily relayed on annotation for information discovery. Hence, RSs employs multiple approaches such as tagging that resulted with the emergence of a multitude of solutions to improve RSs services.

Web 2.0 allowed social tagging to rapidly spread, creating an additional dimension to the world of RSs techniques. The traditional two dimensions (user-items) have been augmented with “tags” as a third element, i.e the third dimension, which converted RSs into Tag-Aware Recommendation (TAR).

By paying more attention to CT/annotations on the web, researchers significantly reduced the time information seekers/users required to retrieve relevant information and address IO. Hence, CT based RSs enabled intensive users' participation in exciting, highly interactive online services, which also demonstrated possible user role in the creation and organization of knowledge and construction of controlled vocabularies in resource discovery.

It is also worth mentioning that RSs techniques may be used for a range of predictions in RSs: from creating domain- specific and mediated predictions (Rosenthal et al., 2010), rated predictions (Tang et al., 2013), (Steck, 2013), (Campos et al., 2011) to predicting personalised item distribution (Koren and Sill, 2011), social tags for cold start book recommendations (Givon and Lavrenko, 2009) query intent prediction (Baeza-Yates, 2010) and predicting the desirability of social match (Mayer et al., 2010) and personality traits (Gao et al., 2013a). However, in this research predictions in RS systems are out of scoop.

Although Tagging, CT and folksonomies were helpful in enhancing RSs some researchers criticised TRSs because they were unable to exploit social tagging and folksonomies and claimed that:

- Structured tagging were not well-defined, caused confusion and hold several meaning (Golder and Huberman, 2006) (Bar-Ilan et al., 2008).
- Tagging suffered from complex discovery and retrieval of information.
- Tagging lacks quality in its contents (Bar-Ilan et al., 2008).
- Tagging also suffered from duplications (Bogers and van-den-Bosch, 2009)
- Tags can be ambiguous and suffer from spelling errors.
- Folksonomies are difficult to use in order to retrieve or exchange information (Limpens et al., 2008).
- CT systems lacks the ability to extract social knowledge from tags identify communities in social intensive environment.
- CT creates unintentional IO because of its structure (Wu et al., 2006).
- CT is far from being a good technique for knowledge organisation and a resource discovery tool on the internet.
- CT, as noted, has low precision and lack of collocation compared to Metadata or structured data created by professionals.
- CT suffers from absent properties, which characterise controlled vocabularies.

- CT is inadequate to scale and sustain long term the level of user confidence or be considered as general resource discovery (Macgregor and McCulloch, 2006).

Information seekers/users also participated in degrading possibilities of utilising tagging, annotations, ST and CT in RSs for the follows:

- information seekers/users' willingness to participate in public annotations and share personal annotations in online collaborative environment.
- influence of information seekers/users educational level and background in the quality of annotations.
- Lack of annotations include anchors and are more public commentary than other types of annotations
- Frequency of changes personal annotations undergo before and after publication influence their applicability to be shared publicly (Marshall and Brush, 2004).

### 2.3.3 SUMMARY

RSs and their techniques discussed in section 2.4 are directly involved in the process of information presentation, management, retrieval and filtering on the internet. hence, participated in identifying some basic problems and the shortcomings of these technologies and their influence in inflating modern IO on the Internet.

To sum up early RSs attempted to:

- weaver the IO problem in electronic mail systems in the 90s (Goldberg et al., 1992).
- CF allowed RSs to learn from others' experiences rather than relying recommendations on missing or inaccurate content-analysis methods.
- CF managed to recommend items to users who did not rate the same items.
- CBF allowed items discovery by other users.
- created profiles from extracted contents of items to improved recommendations to users without depending on similarity of users.
- Personalization can be achieved by utilizing group feedback (Balabanović and Shoham, 1997).
- control excessive amount of information by developing a variety of models to profile both information seekers/users and products, services and information sources.
- Incorporate information seekers/users activities, views and relationships to deliver personalised recommendations.
- Incorporate Web 2.0 technologies to enhance RSs performance.

Still, RSs and their techniques suffered from problems in the process of delivering recommendations to information seekers/users itemised as follows:

- It delivered shallow analysis of items.
- Difficulty to extract features of some domain items, and in some cases extracted features are not enough to improve recommendations.
- IR techniques ignores the quality of retrieved items.
- Over-specialization can be a major problem of CBF recommendations; this can be the result of recommending items scoring highly.
- Elicitation of user feedback (Balabanović and Shoham, 1997).
- Inability of RSs to recommend new items to a user due to lack of item information and users' ratings and lack of common items features of profile of a similar item.
- Inability of RSs to recognise items which users did not rate. Thus, the system will not be able to create a user profile and match them to items on the system, leading to poor recommendations (Balabanović and Shoham, 1997)

This section, provided an overview of RSs and variety of implementation of RS in the domain of e-commerce, and e-marketing (see section 2.3). the discourse was further extended to include itemisation of most commonly used RSs techniques. Section 2.3.2.1, discussed CF technique and its types, methods to enhance of CF and existing issues of CF technique. Section 2.3.2.2, discussed CBF technique and its implementations to reduce IO in email systems, news and article recommendation and enhance information recommendation in the domain of SNS. Section 2.3.2.4, discussed research community and service providers' attempts to combine RSs technique to deliver hybrid RSs to improve the performance of RSs recommendations. The discussion on RSs and their techniques is enriched with researchers attempts to incorporate web 2 tools and technologies such as tagging, annotation and folksonomies techniques to enhance RSs and used specifically for recommendations of the contents of social intensive environments on the internet.

## 2.4 SEARCH ENGINES

This section, provide a comprehensive history of search engines and their techniques. furthermore, it also give examples of different types of existing search engines, and list some of the attempts to solve different issues with search engines. This section also investigate the role of semantic search engines in improving search engine results. Furthermore, discussion on user behaviour is present in section 2.4.4 from two different views: psychological studies of user behaviour and CS empirical analyses of user behaviour on the Internet.

### 2.4.1 OVERVIEW OF SEARCH ENGINES

In 1999 only 3 million websites existed on the Internet. Today, (InternetLiveStats) statistics indicate that the number of websites that exists on the Internet is approximately 980 million (InternetLiveStats, 2016); and the number continues to grow at a rate of 12 thousand pages per/day. With just 3 million websites (InternetLiveStats, 2016) search engines covered only 15% of these websites. With six top search engines at that time, only 60% of websites were covered. When Bradford and Marshall conducted their research, search engines were still primitive. A lot of current search engine techniques did not see the light. Consequently, problems with search engines that required a solution at the time the research was conducted are not considered a problem in current search engines and their functionalities (Bradford and Marshall, 1999a).

Most research on search engines is divided into two areas: analytical approaches that focus on analysing the mechanisms of search engines, and empirical/experimental research that focuses on collecting data through experiments and testing.

Search engines are the second generation of Information Retrieval systems (IR) (Kelly, 2009) and techniques which were used by libraries in the early 70's (Crestani, 1997). With all advances in technology and the conversion of services into the Internet, search engines became handy to allocate information on the Internet. Within the past two decades, search engines went through a very long journey to improve search engine results.

Nowadays, search engines are based on a huge collection of techniques to provide the most relevant search results to its users (see section 2.4.3 and 2.4.4). Some of these

techniques are convenient and others are not. For example, one of the early search engine techniques is the generic keyword matching level. This technique is dependent on collecting words from user's queries and delivering results constructed based on similarities between keywords and the contents of Internet sources (Jerath et al., 2014).

Ranking is a very renowned search engine technique that delivers results to users based on multiple methods, such as Click-Through Ranking (CTR) (Wing et al., 2004) , Recency Ranking (RR) (Dong et al., 2010), Item Popularity Ranking (IPR) (Vojnovic et al., 2009) , and so forth. Moreover, these ranking techniques are further refined accordingly to provide search results according to users' location (shaikh and Kharat, 2015), situation, time, trust and relevance (Metzler et al., 2009), and domain specific ranking of documents (Aleman-Meza et al., 2010).

Another popular technique is the Tagging of web contents, either by the system or users on the Internet can be found in the work of (Golder and Huberman, 2006), (Macgregor and McCulloch, 2006), (Lamere, 2008), (Siersdorfer et al., 2009) and (Cai and Li, 2010) (see section 2.3.2.4). Behaviour analysis of search results (Agichtein et al., 2006), (Hannak et al., 2013), service providers and users on the Internet (Lichy, 2011) influence Search Engines Optimization (SEO) on search results such as: Generic Search Results (GSR) (Collins-Thompson et al., 2011), Organic Search Results (OSR) (Ratliff and Rubinfeld, 2014), Branded Search Results (BSR) (Rutz and Bucklin, 2011), and Paid Search Results (PSR) (Laffey, 2007) (see section 2.4.4.1).

Search engines can be considered as a discipline of their own. This is explained by the wide range of services, techniques and business models that are used to support search engine services financially in e-Commerce, online marketing, and search engine optimization techniques that targets business partners as a source of income and so forth (Google-SEO, 2010). The combination of these elements in any search engine is the criteria to a successful delivery of search Results. In this section we would like to highlight some of the corner stones of some of the search engines that deliver search results, to consider a criterion such as efficiency and relevance and so forth in search results. We will also provide criteria to evaluate any search engine performance and compare it to our way of delivering search results. Like all solutions in CS and IS, the basic concept of search engines was to help solve some of the many problems of Information Retrieval (IR) systems. Search engines started with primitive and basic

techniques to ease the process of Information Retrieval from the Internet based on user's queries. We cannot deny that search engines helped a lot and have advanced a lot so far. A series of intensive work can be seen in the area of search engines. For Instance, research on Search engines started with the Problem of primitive keyword matching and proceeded to many problem domains such as (Ranking of search results in Recency Ranking of search results based on time, location, events and so forth, Precision in search results, Accuracy in search results, Relevance in search results) (Dong et al., 2010) . Moreover, to provide accurate search results, search engines used techniques from the domain of RSs, which were developed in the domain of e-commerce in order to allow maximum relevance in search results to users' queries.

On the other hand, most research in the domain of search engines and their techniques indicates that Information Retrievals from the Internet is exacerbated by either excessive amount of search engine results or caused by the combination of search engine techniques Thus, many researchers believe that a lot must be done in this area in order to remedy some of the problems encountered in search results (Seymour et al., 2011).

#### 2.4.2 HISTORY OF SEARCH ENGINES

The development of search engines began in early 90s to help users on the internet to find information sources. A group of CS students at McGill University developed the first tool "Archie" a UNIX command line program that collects/downloads data files from a file transfer Protocol (FTP) directories into a DB of hundreds of systems (Archie, 1990). Main functionality of Archie allowed users to search huge directories of indexed sites. Indexing websites at that time was an easy task to perform and manage, thus Archie did not provide the users with a query technique to search for information. However, the web continues to grow and the indexing process became a tedious task (Wright, 1998).

Likewise, "Gopher" was developed at University of Minnesota in 1991. An indexer and a menu system which attempted to distribute searches and simplify the retrieval of documents. It allowed to select items from the menu and customize them. and text organisation into hierarchies, user friendly and supported simple graphical UI. Furthermore, provided a limited access to the system administrator. (GoodGopher, 1991).

In 1992 Gopher, utilised Veronica and Jughead tools to discover information on the web. Veronica and Jughead shared features with Archie such as searched for files by

names and titles. Two main functionalities of Veronica and Jughead: 1) keyword search that runs an ad-hoc process to create new menus based on user queries to deliver customized Gopher menu. 2) limit the search to single server to build indexing DB (netlingo, 1995).

Attempts to improve the performance of search engines advanced with the growth of the web. since 1993 many search engines surfaced to deliver certain features and solve some problems of existing search engines. A range of techniques were utilised in the process of enhancing search engine results. For example, Oscar Nierstrasz wrote a Perl Script that periodically mirrored the websites into a standard format catalogue (Nierstrasz, 1996). Similarly, between 1993 and 1995 Mathew Gray developed “Wanderer “ a robot like Perl-based script, used to generate an index named ‘wandex’ and measure the size of the World Wide Web (WWW) (Wanderer, 1993-1995). AliWeb is another search engine allowed users to submit indexed file locations on their websites, and permitted them to add user scripts (AliWeb, 1997) The difference between the two search engines was that one used an indexing technique and the other waited for recognition by website owners.

JumpStation developed an UI web robot indexer to run their query program and was the first resource-discovery tool that crawled, indexed, and searched the web (JumpStation, 1993). In 1994, the first WebCrawler is a full text web-based Crawler allowed keyword level query that gives the users the freedom to search any webpage by keywords. Onwards keyword query technique became a standard technique in every search engine (WebCrawler, 1994) .

Lycos a commercial search engine was the first profitable internet businesses and web portals that featured email server, web hosting and social networking (Lycos, 1994). Whereas, in 1995 Daniel Dreilinger developed the first algorithm based Meta-Crawler that allowed to crawl 20 search engines and directories in parallel.

Gradually, search engines started to mine crawled contents in order to enhance search results. In 1996, HotBot utilised Inkomi a database and directory (HotBot, 1996). AltaVista was the first search engine that used a natural language processing algorithm (Chowdhury, 2003) . Yahoo later operated upon web directories rather than full-text index of websites (Seymour et al., 2011).

Many search engines were established between 1996 to 2001, and some survived for a year or two, whilst others survived little bit longer (Seymour et al., 2011). In 1998, Microsoft released MSN Search, which provided search results based on Inktomi and later in 1999 combined their techniques and based results on Looksmart and AltaVista. In 2004, Microsoft utilised web crawler “msnbot” in the search process (MsnBot, 2015). Microsoft made a final transmission in 2009 and launched their final product, MSN Search, which is now known as Bing (Bing, 2009).

Google which was established in 2000 is currently pioneering search engine services because it utilised PageRank algorithm which had a revolutionary impact on the advancements of search engines. The ranking algorithm ranked search results based on website ranking and their connection to other website. Thus, ranking search results become a basic technique in current search engines (Grehan, 2002), (Brin and Page, 1998).

Currently, Google is used by 1 billion and 100 million visitor p/m and ranked first among search engines. Bing, with 350 thousand visitor p/m ranked second best search engine, and Yahoo, with 300 thousand visitor p/m ranked third best search engine (eBizMBA, 2015). These three commercial search engines gained popularity because they dazzled users with their simplicity, user friendly interfaces, speed and wide spread all over the world (Fabos, 2005). There are other domain specific search engines such as tourism web portals that focus on online tourism services, and particular types of search engines known as Personalized Content Discovery portals that deliver services based on the user role on the system.

Bradford and Marshall research was concerned by the increasing number of published webpages (Bradford and Marshall, 1999a). Research results indicates that in the past two decades the number of webpages increased from one million to 800 million webpages in daily bases (InternetLiveStats, 2016).

## 2.4.3 SEARCH ENGINES TECHNIQUES

### 2.4.3.1 SEARCH ENGINES OPTIMIZATION

Statistics indicates that both the number of websites and user on the internet continually growing (InternetLiveStats, 2016). The transformation of many services into web-based, e-commerce and e-marketing was one of the reasons that made users become solely dependent on information on internet. Consequently, users find it easier, more efficient and fast to retrieve information from the internet. Hence, search engines became a primary method to search and allocate information on the internet.

To meet user expectations, search engines continue to develop new techniques to enhance search performance and results (Seymour et al., 2011). According to studies on Web Information Retrieval systems (WIRS), search engines help minimize information retrieval problems and can also help in discovering new websites or contents on the Internet (Schmidt-Maenz and Koch, 2006).

Some of the commonly used techniques in search engines to deliver services to business partners/clients (who deliver services to information seekers/ users) are business model Pay-Per-Click (PPC) that allowed a certain extent of information discovery on the internet. PPC technique increased the possibility of finding people, products, companies, and so forth on the internet. Google Ad word, Yahoo and Bing utilised PPC to help advertise certain information/service to clients by reserving a container that display poster or video that link to external sources as search results (GoogleAds, 2000), (YellBUSINESS, 2016) (WEBCREATIONUK, 2005) (Bing-ads, 2016).

Furthermore, PPC technique also attempted to included ranking of search results to improve users' experience on the internet and deliver personalised search results (Su et al., 2010).

The work of Anderson and Cheng investigates the possibility of combining models of ads rank and performance of hospitality related keyword searchers in paid searches (Anderson and Cheng, 2014). This research studies the behaviour of search engine results and ranking techniques by analysing the user queries, search engine ranking and service providers on the Internet. Anderson and Cheng's work compared branded versus generic

keyword searches from the domain of SEO solutions, which aimed to remove bias search results ranking by estimating keyword bidding performance.

Therefore, Anderson and Cheng proposed creating a framework model that will consist of:

- An estimation of click-through ads,
- A click function model which utilises binary logit model to estimate the click behaviour,
- a rank function which assigns a slot to selected ads,
- a joint probability function to accurately model click-through ranking, advertisers' ads,
- an estimation of the joint probability function which uses traditional maximum likelihood, and
- tests the model fit and parameter estimates.

This was to develop a consumer click-through model based upon the prior described proposal. Furthermore, the proposed approach will help in the untangling of advertisers' characteristics search specifics and advertisers' behaviour to determine individual impact upon CTR. The model was experimented on data sets collected from a Chinese search engine and the collected data described:

- users' involvement is searching for information on the internet,
- information about advertisers' hospitality services, and
- prices and users' click-through ads (Anderson and Cheng, 2014).

#### 2.4.3.2 RANKING OF SEARCH RESULTS

Ranking search results is another method which addressed modern information retrieval problems on the internet. Search engines employ ranking techniques and algorithms to deliver refined search results based on certain criteria and situations such as time, location and so forth. Furthermore, search engines also re-rank search results to provide relevance, accuracy, and precision in search results. Research in the domain of search engines mainly aimed at improving search engines performance in the process of information retrieval on the internet. Although ranking techniques aimed to remedy some of search engines retrieval problems, it also participated in modern IO on the internet. Hence, the discourse in this section discuss some distinctive attempts that participated in controlling excessive amount search engines retrievals on the internet.

Ranking in search engines also depends on profiling both user and information sources on the internet. Hence, search engines crawl the web to index information sources, and store the collected information in structured DBs (Page et al., 1999). Other attempts utilised data sets collected from large volumes of Web query. (Collins-Thompson et al., 2011). Some re-ranking algorithm is tested on public dataset ODP239 and a real search result dataset collected from commercial search engines such as Google (Burges et al., 2005) (Lin et al., 2010).

Ranking techniques were investigated and utilised by researchers to control the performance of search engines, solve some of the problems such as excessive amount of them, relevance, accuracy and consistency in search results to deliver refined search results. Variety of approaches were followed to improve ranking techniques in search engines. Page et al. attempts to understand users' role in the process of ranking search results (Page et al., 1999). Likewise, Chidlovskii et al. analyse users profiles in online communities to deliver ranking results based on users and communities (Chidlovskii et al., 2000).

Recent investigation on the role of users in the process of ranking techniques attempts to understand the behaviour and personality of users on the internet. Agichtein et al. also investigates users' behaviour and experience when searching information on the internet (Agichtein et al., 2006). Whereas, Collins-Thompson et al. investigates methods to combine personality information and reading level to re-rank search results (Collins-Thompson et al., 2011). relativity calculation of keywords model (Chidlovskii et al., 2000) and Markov's chain random model (Lin et al., 2010).

Research on ranking techniques took into account types of information users seek on the internet. Hence, Dong et al. investigates problems of documents ranking in search results to deliver relevance in search results based recency ranking (Dong et al., 2010). Zhao et al. extends the investigation to understand the role of modern technologies and web 2.0 tools (i.e. social networks and other social intensive environment on the internet) role in the process of ranking search results (Zhao et al., 2011a). Research attempts to enhance ranking techniques considered diversification in search results as an important aspect in search results. Hence, Lin et al. investigates redundancy in top ranked search results to understanding causes of user disappointment in search results (Lin et al., 2010).

Clustering search results is another approach to refine search results. Wen et al. attempts to control modern IO by investigating methods to clustering techniques of users' queries and precision in frequently asked questions (Wen et al., 2001). Wen et al. extends the investigation on clustering techniques to include information discovery of popular topics on the internet and short length queries in the process of data discovery (Wen et al., 2002). Furthermore, Zhang and Dong investigates methods to improve existing clustering techniques of ranked search results (Zhang and Dong, 2004). Whereas, Zeng et al. investigates problems of current clustering techniques that influence the speed of information exploration of search results (Zeng et al., 2004). Research on clustering ranked results also took into account possibility to cluster websites in order to narrow down search results. Carpineto et al. investigates the possibility to develop web clustering technique with regards to acquisition and pre-processing of search results (Carpineto et al., 2009).

Relevance of retrieved information denotes how well information retrieval tools understood information seekers information needs. Furthermore, it involves both information seeker and information sources in the since of time, subject, ownership and so forth. Hence, research in the domain of information retrieval systems heavily relay on delivering maximum relevance in retrieved information. in search engines. relevance of search results is a crucial factor that directly involves modern IO on the internet. Hence, many researchers were focused on investigating methods to maintain relevance in search results. Järvelin and Kekäläinen research investigates ranking techniques to major relevance degrees in search results (Järvelin and Kekäläinen, 2000). Borlund research investigates relevance measures to include multidimensional and dynamic nature of relevance in IR context (Borlund, 2003).

Some other attempts to enhance ranking techniques belong machine learned ranking which are semi-supervised or fully dependent in the process of ranking search results. Burges et al. attempts to enhance ranking techniques involves utilising Gradient Descent algorithm that utilise learning models in ranking functions (Mason et al., 2000, Burges et al., 2005). Yong et al. investigates approaches to machine learning algorithms of ranking techniques to deliver relevance based ranking of search results (Yong et al., 2008).

Recent attempts to investigate the role of ranking techniques in search results considered commercial search engines such as Google, Yahoo and Bing. Su et al. research analyse Google ranking algorithm and design (Su et al., 2010). Similarly, Beel and Gipp based their investigation of Google's ranking algorithm to understand how google delivers ranked results to the users (Beel and Gipp, 2009).

The outcome of the intensive research in the domain of both ranking techniques and search engines suggested range of solutions. These solutions aimed to either improve the performance of search engines by utilising ranking techniques to deliver refined search results or improve ranking techniques to deliver re-ranked search results.

Page et al. proposed a PageRank technique that provides methods to measure the rating of Web pages objectively and mechanically through human interest and attention. Page et al. attempted to address users' decision making when viewing web pages on the internet (Page et al., 1999).

Chidlovskii et al. suggested a ranking algorithm can secure certain level of relevance of search results based on users and their community profiles. Furthermore, the ranking algorithm can run in any context where users' involvement is required in online communities. moreover, it allows implicitly or explicitly re-weighting profile terms to enhance the relevance in search results. Thus, the proposed system architecture can be utilised by both RSs and meta-search engines to provide refined search results to users on the internet. The figure 2.5 below provides an overview of the proposed system architecture to re-rank documents on the internet (Chidlovskii et al., 2000).

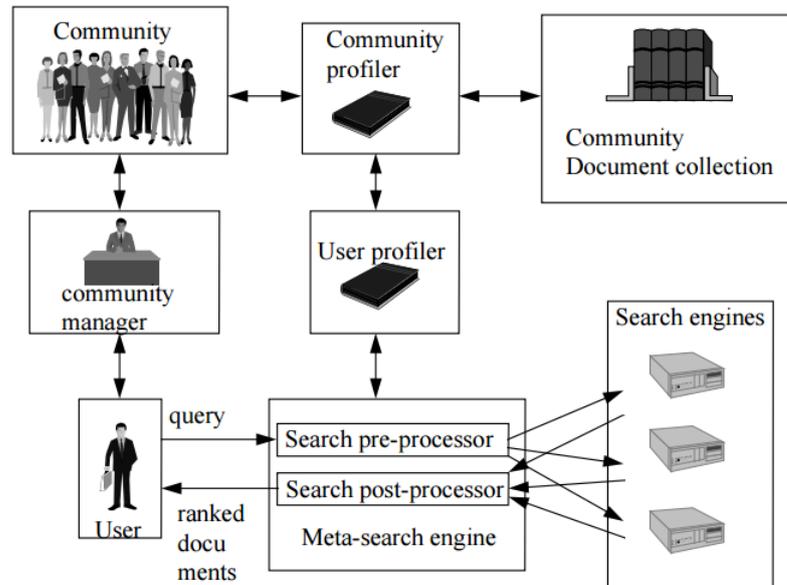


FIGURE 2.5 MAIN ELEMENTS OF THE RE-RANKING ALGORITHM PROPOSED IN (CHIDLOVSKII ET AL., 2000).

More modern solutions which involved users on the internet can be found in the work of Agichtein et al. which suggested to incorporate user behaviour in the process of ranking web searches. Furthermore, utilised implicit users' feedback model into the ranking process (Agichtein et al., 2006).

The concept of re-ranking of ranked search results can be found in the work of Collins-Thompson et al. which aimed to deliver personalised web search results based on general users reading level and in particular children. Hence, the suggested approach adjust ranking to match children's competence level and needs. Moreover, the proposed solution involved UI design, content filtering, and results presentation techniques and estimation of user proficiency (Collins-Thompson et al., 2011).

Dong et al. proposed a retrieval system that automatically detect and responds to search queries that delivers ranked search results based relevance to recent activities in real life such as breaking news is dependent on effectiveness of retrieval technique. Thus, evaluate queries sensitivity using a high precision classifier. Moreover, the collected data is processed by machine learned ranking model and several features to provide temporal evidence to represent document recency (Dong et al., 2010).

Zhao et al. proposed new search ranking algorithm based on clustering webpages and tag clustering approach that involved information problems of modern technologies.

Thus, combine the contents of the web pages and the social annotations, then clusters the web pages and the corresponding tags and user interests when delivering search results (Zhao et al., 2011a).

Lin et al. proposed a novel re-ranking algorithm: “DATAR” based on GRASSHOPPER (Zhu et al., 2007) a framework that model and absorb random walks. The results indicated that the DATAR algorithm outperforms GRASSHOPPER, by insuring diversity when re-ranking internet search results (Lin et al., 2010).

Wen et al. suggested a new clustering approach of similar queries according to their contents, and the feasibility of an automatic tool to detect FAQs. Hence, utilised user logs to count user’s clicks of a document, times of retrieval the same document from different queries, count times the document was selected and terms similarity in users’ queries (Wen et al., 2001). The extended work of Wen et al. suggested to utilise data discovery and combine it keyword approach in the process of delivering search results (Wen et al., 2002).

Zhang and Dong proposed a novel approach that combined key phrase discovery technique and orthogonal clustering to automate the organization of web search results into groups and generate hierarchal clusters of search results as semantically interpreted (Zhang and Dong, 2004). Whereas, Zeng et al. suggested to restructure clustering process by extracting and ranking salient phrases as candidate cluster names, based on a regression model learned from human labelled training data (Zeng et al., 2004).

Web Clustering engines are useful in order to narrow users search results and, as many other uses of clustering in information management, would definitively alleviate IO. Another distinctive work in the domain of search results clustering is the work of Carpineto et al. proposed a cluster based search engine that will process acquired and pre-processed search results in addition to core system specification to enhance existing cluster based ranking of search results (Carpineto et al., 2009).

Relevance in search results is a crucial factor in the process of ranking search engine results and internet searches in general; thus, the definition of relevance and its levels or degrees differs when deciding whether ranking techniques are successful or not. Hence, Järvelin and Kekäläinen proposed evaluation criteria that utilize non-dichotomous

relevance judgements in IR process to guarantee highly relevant documents retrieval. Thus, insure:

- a novel application of P-R curves and average precision computations based on separate recall bases for documents of different degrees of relevance, and
- two novel measures computing the cumulative gain the user obtains by examining the retrieval result up to a given ranked position (Järvelin and Kekäläinen, 2000).

Whereas, proposed to study multiple faces of relevance concept in IR on the internet to identify different concepts of relevance. Thus, Borlund also paid attention to the situational relevance concept which is considered as the most realistic type of user relevance from its potential dynamic nature (Borlund, 2003).

Users are aware that search engines can provide them with their preferred method of viewing huge collections of information. Furthermore, search engines apply a variety of techniques to label highly ranked web pages and make them visible to people. Thus, ranking web pages gained priority as a technique of refining search results.

Burges et al. proposed RankNet that is based on machine learned ranking algorithm (Gradient Descent ) and utilised probabilistic cost function (Burges et al., 2005). Similarly, Yong et al. used machine learning approaches such as Graph Neural Network (GNN) to successfully discover Google, and other search engine's, ranking algorithm (Yong et al., 2008).

Su et al. extended the investigation on Google's ranking techniques because its influence the way users access information on the internet. Hence, suggested to combine linear learning models and recursive partitioning ranking scheme to deliver high accuracy in search results (Su et al., 2010).

Ranking techniques of commercial search engines cannot be completely revealed to the public, because they remain the business secrets. *Researches can only guess how Google's ranking works, using experiments and experiences. Several researchers have analysed Google's ranking algorithm, using different methods.* Su et al. claimed that the proposed technique in their work can reveal Google's ranking function. This can be considered a step forward in the discovery of commercial search engine ranking algorithms (Su et al., 2010).

Beel and Gipp, experimental work studied of Google Scholar and investigated the interrelationship between an article's citation count and its position in Google Scholar. Furthermore, Beel and Gipp conducted second experiment to test whether further patterns existed in how rankings interrelate with citation counts. The outcome of the research indicated that range ranking algorithms were applied to deliver search results. (Beel and Gipp, 2009).

The avalanche of information on the internet added more pressures on search engines and overloaded information seekers/users on the internet unintentionally with unrequired IO. Consequently, this indicates that users are far from being satisfied with search engines that created excessive amount of search results.

Hence, Glut and smug of information became common phrases accompanied excessive amount of search engines results. The research community complained that glut and smug of information is the consequence of the open environment enabled by web 2.0 tools to users and rapid development of networking techniques on the internet (Denning, 2006), (Shenk, 1998), (Fox, 1998). Therefore, finding information became a tedious task to both service provides and users on the internet. Abundance of search engines results forced the users to only focus on viewing top listed results or the first page of search results. Consequently, relevant and good quality resources are pushed back to the following pages are rarely viewed by information seekers/users.

Hence, ranking techniques used in search engines created unintentional IO to the users by delivering important sources of information with less relevance to users' queries. Zhao et al. argues that basing ranking algorithms on the collection of data on relationships or links between user query and web pages is the beginning of search engine failure to deliver search results.

Although both the research community and service providers/search engines on the internet attempted to solve information retrieval problems to reduce IO. However, it is evident that IO manifests with changing technology, information seekers/users and the way information is generated on the internet. Hence, it is still arguable that information retrieval problem such as (relevance, ranking, accuracy and so forth) that occupied 3 decades of intensive research to deliver information to information seekers will continue to recur because both the characteristics of information seekers, and information on the internet continually changing; thus, requires continues work to keep up with changes.

Thus, Thus, some of the common problems of information retrieval in general and search engines such as relevance of search results cannot be ignored. Furthermore, degree and level of relevance should always be determined by users' queries. i.e. by keywords that are given by the user and then converted in a "query" by a search engine (Borlund, 2003) (Schmidt-Maenz and Koch, 2006). Furthermore, it is important to strike balance between relevance and diversity in internet searches as discussed in (Chidlovskii et al., 2000) and (Lin et al., 2010).

To sum up, according to the discussed topics in this section information retrieval problem on the internet can be minimised by enforcing relevance measures in the process of ranking search engine results delivered to information seekers/users. Commercial search engines' (Google, Yahoo and Bing) ranking algorithms are very difficult to discern due to the utilisation of multiple methods in the process of ranking. In addition to ranking techniques, search engines also use layered filtering of search results, the combination of multiple filtering techniques that can vary from simple ranking algorithms to complex ranking algorithms that consist of many techniques and models to refine search results. Hence, the research community can only build assumptions based on the observation of search results and examine their own ranking techniques and experiments to improve the quality of search results.

#### 2.4.4 USER SEARCH BEHAVIOUR ON THE INTERNET

The discussion in the previous section attempted to cover distinctive work in the domain of search engines and its advancements in the past 2 decades. It also itemised common problems of search engines, and the role of two intensively used techniques in the process of refining search engines results (search engine optimization and ranking technique). The previous section slightly mentioned the role of information seekers/users in the process of developing new methods that either solve information retrieval problems or enhance the performance of search engines when delivering search results to information seekers/users on the internet.

Hence, this section will discuss the influence of user behaviour on the internet and user SB in relation to search engines in general. The discourse in this section highlights some problems, proposed solutions of how users participate in generating excessive amount of information on the internet. The discussion in this section itemise range of problems of users on the internet which may not exist in current situations. However, they describe a sequence of events in the process of enhancing search engines and their techniques and attempts to model either search engine techniques or user SB on the internet.

According to many investigations, information seekers/ users are the main factors of the success of any service on the internet. Therefore, many attempts to improve internet services investigates the role of information seekers/users in the process of service enhancement and delivery. Hence, to understand User Behaviour (UB) both researchers and service providers are mostly dependent on “*what happened in the past*”. Such information can be found in system log (Hansson, 2015), user query log (Beitzel et al., 2004), user activity sessions which are an ideal source of information and cannot be ignored in current information retrievals on the internet (Su and Chen, 2015).

Search engines, RSs and their techniques, and other information retrieval systems on the internet (Seymour et al., 2011) were developed to improve information presentation, retrieval and delivery of search results based on information seeker/users queries. Therefore, search engines utilised many techniques to deliver search results based on situations when information seekers/users search for information.

#### 2.4.4.1 TOWARDS CYBERPSYCHOLOGY

Psychology, as the study of mind and behaviour (BPS, 2000-2015), has many branches, such as Human Mind, Brain Functions (Leont'eva, 1974), (Lomov, 1982), Behaviour (Harré and Secord, 1973), (Potter and Wetherell, 1987) Perception (Dember, 1960), and is further split into more focused studies, such as Cognition (Meichenbauma, 1977), Personality (Stagner, 1937), (Sanford, 1950), (Endler and Magnusson, 1976) and Social Behaviour (Sherif, 1936), to mention just a few. Psychology is a discipline, which has developed through the centuries. However, the latest advances in science and technology, which have affected our lives in the last 10 years, have brought forward “Cyberspaces Psychology”, which studies how human minds interact with each other and “machines” in cyberspaces, where people can communicate with each other via Cybertechnology (Suler, 1996), (Riva and Galimberti, 2001), (Piazza and Bering, 2009), (Attrill, 2015). They study the mental states of users of cyberspaces, their interactions and communication methods and their willingness to meet other users through the availability of social spaces (Shields and Kane, 2011).

It is evident that computer technologies on the internet created tangled multidisciplinary Sciences. Hence, it is difficult to distinguish boundaries in technologies used by Businesses, Education, Healthcare, and users of these technologies and services on the internet. Therefore, psychologists extended the discourse on UB to include UB On the Internet. It is worth mentioning that whenever a group of researchers extend their research to cross disciplines, they base current research on previous findings in certain domains. Hence, almost many of UB investigation on the internet is influenced by UB studies from the domain of psychology.

Some of the heavily studied topics in Cyberpsychology is the impact of Online Social Networks (SN) (Hampton et al., 2011). There are numerous publications that investigate the relationship of SNs usage to mental diseases such as depression (Banjanin et al., 2015), low self-esteem (Brewer and Kerslake, 2015), social isolation (Hamptona et al., 2011), negative relationships (Knight, 2015), fear of missing out (BBC, 2015), sleep deprivation (Lynette et al., 2015), addictive behaviour (Andreassen, 2015), eating disorders (Kirby, 2015), social media and attention deficit hyperactivity disorder

(Barkley, 2014), anxiety disorders (Indian and Grieve, 2014), and ostracism in cyberspace (Birchmeier et al., 2011).

One of the heavily influenced research areas with psychology is artificial intelligence which models human behaviour in the process of developing different types of machines and robot (Gobet and Erekü, 2016). However, it is still not clear so far how psychology research on UB influences the development and improvement of online services.

Studies of behaviourism are very important in the context of Cyberpsychology. because it allows to understand and analyse causes, purposes, actions, impact and outcomes of HB and systems on the internet. Hence, HB on cyberspaces was extended to investigate information seekers/users in domains like healthcare, e-commerce, education, and so forth. Hence, the next section will discuss some of the existing research in the area of information seekers/users role in both information retrievals, information presentation, IO, and information influence in improving search engines results.

#### 2.4.4.2 MANIPULATING USER BEHAVIOUR IN CYBERSPACE

The previous section provided the reader with an overview of psychological studies which attempted to understand UB. Furthermore, the discussion in the previous section also introduced new research discipline “Cyberspaces” as consequence of science and technology advancements that investigates the psychology of information seekers/ users on cyberspaces.

UB literature is very wide and includes human SB in general. However, this research is only interested in UB research on information seekers/ users SB on the internet. The research on User Behaviour (UB) on the internet began in late 90s. Early attempts focused on identifying users search behaviour (SB) on the internet (Bradford and Marshall, 1999b). Hence, Bradford and Marshall proposed to create a catalogue that classifies the contents of the web contents based on topic of interest and allowed users to view indexed data sets. Bessonov et al., proposed to allow users to create their own catalogue to enable faster information retrieval (Bessonov et al., 1999)..

This section will discuss some of the attempts that investigates the role of information seekers/users in the process of information representation and information

retrieval. Furthermore, discuss the role of both service providers and information seekers/ users in generating excessive amount of information on the internet. this section gives examples of general aspects of UB on the internet and focus on information retrievals (search engines, RSs, other information retrieval systems on the internet). the discourse in this section also highlights some attempts to personalise retrievals (Khapre and Chandramohan, 2011) from the internet based on users' interest, activities, relationships and profiles created by information retrieval systems and social intensive environment on the internet.

To investigate UB on the internet, researchers and service providers mostly utilise various data sources such as Data Mining (WDM) techniques for data discovery on the internet (Masand and Spiliopoulou, 2003), system logs (Bradford and Marshall, 1999a) (Benevenuto et al., 2009) and (Su and Chen, 2015), service logs (Younger, 2010) and query logs (Beitzel et al., 2004) to extract data about both the system and users. An alternative method to extract information seekers/users is to utilise user profiling models (Li, 2012) to collect users demographic data (Benevenuto et al., 2009), sessions of activities in social intensive communities on the internet (session start/end time, user login certificate, click-through activities, queries and so forth) (Jerath et al., 2014) and also include (viewing members' profiles, subscriptions to communities, posting/commenting on topics, and the frequency of online application usage) (Su and Chen, 2015) and cookies which are used to allow users to access online services and at the same time collect clients' information (Boland et al., 2015). Furthermore, SB models (Khapre and Chandramohan, 2011) and search engine logs that contains records of users' (clicks, queries, views, demographic data, and other system related statistics). It also describes users involvement in discussions, which may improve the recommendations of services/products based on users' opinion (Hansson, 2015).

All these data sources were utilised can help in adding semantics towards the understanding of UB in cyberspaces and ultimately discovering the exact SB of information seekers/users. Furthermore, understand usage patterns and explicit or implicit given preferences in context discovery to deliver personalised web contents or to deliver user-centric services (Khapre and Chandramohan, 2011). By collecting user SB data from search logs, researchers were able to categorise and analyse the contents of

these logs and learn about both the user and the search engine behind the activities in cyberspaces.

Researchers interest in understanding UB on the internet and the ways UB is analysed depends on the research objectives. Hence, UB research criteria differs from one group to another. Hence, Benevenuto et al. extracts UB from system logs (search queries, profile viewings and friends lists, joined communities, usage of online applications and uploaded pictures, viewing patterns, and so forth) in the process of investigating UB on SNs (Benevenuto et al., 2009). Similarly, Schmidt-Maenz and Koch extract search keywords from user query logs to analyse user SB (Schmidt-Maenz and Koch, 2006).

Whereas in e-commerce, Han investigates UB to understand information seekers/users tendency to search products, filtered information based on product information and price and correlation between consumers and product based consumers groups classification model from Coughing and Douglas-Tate (Coughing and Douglas-Tate, 1985). (Han, 2005).

Likewise, Beitzel et al. investigates possible methods to utilise large query logs to understand user changing interest of popular topics (Beitzel et al., 2004). Whereas, Spink and Jansen based their research on data collected from other research groups investigates web search and users SB to examine public SB, highlight trending search topics, growth and development of human interaction on search engines (Spink and Jansen, 2004).

Hence the research on UB on the internet in general resulted into developing behaviour models that collects UB based on set criteria to help both researchers and service providers in the process of improving information presentation, delivery of products and services on the internet, enhancement of search engines results and RSs recommendations.

Researchers also developed measurement criteria that can be beneficial in understanding UB on the internet. Schmidt-Mänz and Koch investigation of users usage patterns, frequency of using services in cyberspaces, time spent on particular tasks, and sequence of activities participated in developing UB prediction model in Cyberspaces (Schmidt-Mänz and Koch, 2005).

Gyarmati and Trinh’s developed a measurements framework to observe UB and activities on NSs enabled them to collect multiple users’ behaviour at a given time (Gyarmati and Trinh, 2010). Similarly, Li build searching behaviour analysis model on a multi-agent intelligent technology that will automatically build user profiles to understand the users' interests to deliver tailored services (Li, 2012).

Tan investigates the effectiveness of User Interest Model (UIM) which aimed to deliver personalised search results solve IO in search results. Hence, deliver relevant knowledge for different users, and allow users to enjoy the knowledge without wasting a lot of energy and time on handling online services (Tan, 2013).

Schmidt-Maenz and Koch investigates UB Web Information Retrieval Systems (WIRS) (Lycos) that utilised keywords in users’ queries to deliver information to information seekers/users on the internet.

Short et al. proposed a model that will collect data about users’ behaviour in online environment. The proposed model will also gather information about interventions to characterise user experience. Figure 2.6 below shows the main components of the proposed model. (Short et al., 2015).

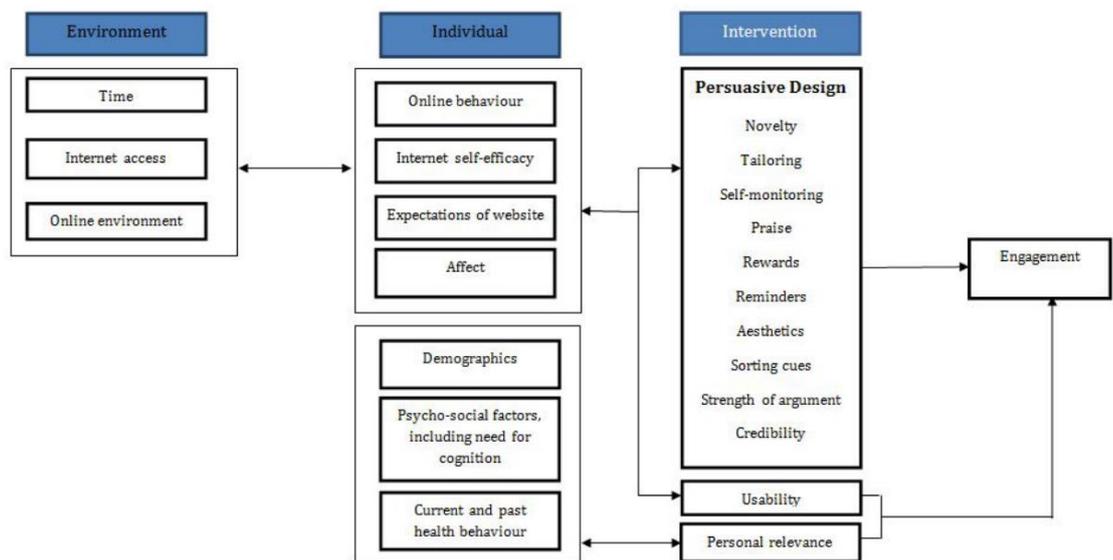


FIGURE 2.6 USER ENGAGEMENT MODEL PROPOSED IN (SHORT ET AL. , 2015) .

In e-commerce, Corbitt et al. investigates WWW business to clients (B2C) commercial services to identify key factors related to trust in the B2C context, interaction among users on the internet, their subscriptions, posts, clicks and patterns of services

usage (Corbitt et al., 2003). Likewise, Mohbeyta and Thakura investigates the role of merging Mobile e-commerce in users everyday life (Mohbeyta and Thakura, 2015).

Whereas, Junghans et al. investigates issues of complex behaviour of online services that require multiple interactions with users, such as inputs and previous actions on the web. Furthermore, attempt to understand the role of business models to deliver recommendations based on relevant online services to the users (Junghans et al., 2012).

Furthermore, Singer et al. investigates successful and unsuccessful user SB to characterise users, distinguish between simple and complex search tasks and to develop simple measures to describe task complexity (Singer et al., 2012). Whereas, March and Simon argued that it is impossible to define complex search task in literature. However, there is a definition of objective behaviour of the user conducting the search task (March and Simon, 1958), (Shaw, 1971) and subjective behaviour of the level and ability of the user to conduct the searching task (Li et al., 2011). Complex tasks can also be defined as the user's uncertainty of information type, problem solving skills of the user, and domain information.

In e-marketing Smit et al. investigates possible methods to understand Online Behavioural Advertising(OBA) and analyse surfing behaviour of users through cookies installed in clients' computers (Smit et al., 2014). Likewise, Dorčák et al. analyse perception of innovation in marketing approaches communication from the perspective of the supplier and consumer (Dorčák et al., 2015).

UB on the internet studies also investigated entertainment services such as video hosting websites which is heavily used by different types of information seekers/users. Chen et al. investigates online viewing/hosting of video platforms, to understand users' tendency to quit the video before it ends or quit before viewing the video. This to reduce pressure on platforms resources (Chen et al., 2015). Similarly, Boland et al., investigate online music libraries and user engagement level to address retrieval problems of excessive amounts of music tracks in an online retrieval system (Boland et al., 2015).

UB investigation can also be found in the domain of stock markets. Nardo et al., investigate the influence of news on the financial market and consequence of technology and information available on the internet on degree of influence they have (Nardo et al., 2015).

Whereas, another group of researchers' investigation of UB belongs to the domain of social intensive environment on the internet. Benevenuto et al. investigates and analyse SNs system logs to understand the social activities of users on SNs to identify usage patterns of all activities on the platform, sessions length which indicate diverse activities, and self-loops in all activities (Benevenuto et al., 2009). Gyarmati and Tuan investigated ways to measure user activities and usage patterns in SNs (Gyarmati and Trinh, 2010). Whereas, Nazir et al. characterised the usage of SN based applications and dynamics (Nazir et al., 2008). Gjoka et al. analysed the usage properties of available applications based on daily active users (Gjoka et al., 2008), Chun et al.'s explored user activity on Cyworld (Chun et al., 2008). Many other studies compared graphical interactions and established social connections. Cha et al., studied UB in user generated content video systems (Cha et al., 2007), Mislove et al. analysed the topological properties of SNs based on the real world measurements (Mislove et al., 2007).

Lops et al. investigate the role of SNs in proliferation of IO and possibility of utilising CBF in RSs to extract user profiles to analyse user interests (Lops et al., 2011a). similarly, Gao et al. investigates the strong correlation between the personality and personal preference of users on the internet for possible methods to enhance RSs with personal preferences (Gao et al., 2013b). Gao et al., further extends their work to investigate problems of search engines to find possible method to reduce IO (Gao et al., 2013a).

Finding from UB investigation does not require to develop any Solution. They may exist to provide evaluation of existing systems, or investigate problems and gaps in existing systems. Hence, investigation results varies depending on research objectives. Researchers that aimed to enhance information presentation, information retrievals and deliver improved services can be found in the work of Corbitt et al. who proposed a framework based on a series of underpinning relationships among trust factors in B2C context. The outcome of this research indicated that people are more likely to purchase from the web that provides a high degree of trust. Hence Corbitt et al. were successful in utilising extracted data from system logs (Corbitt et al., 2003). Tan proposed customised UIM that analyse, classify and support knowledge service and collect users personal information and search terms to deliver personalised search results (Tan, 2013).

Thus, the proposed approach collects UB when searching the WIRS and analyse collected data. Schmidt-Maenz and Koch research examined the proposed solution in four experiments. The first experiment run algorithmic formula to denote time, vocabulary, search queries, to form a dataset and describe the frequency of term appearances in a given time. The second experiment analysed search terms overtime and collected number of terms appearance in set period of time. The third experiment regarded information seeker/users information as main factor in the process of defining search terms. Furthermore, this experiment also run term clustering and trending topics from the previous experiment. The last experiment focused on detecting the topics from search terms (Schmidt-Maenz and Koch, 2006).

Research results which participated in characterising UB on the internet and information sources can be found in the work of Beitzel et al UB who aimed to deliver effectiveness and efficiency in in the process of information retrieval on the internet (Beitzel et al., 2004). Benevenuto et al. investigation helped in identifying two main groups of users with distinctive activities on the network: visible interactions and silent interaction (Benevenuto et al., 2009). Li research on UB investigates issues of relevance of search results. Hence, Li UB investigation involved collecting and analysing data of information seekers/user search keywords, data from server logs, user session logs of downloading and saving information from the internet and time spent on each website (Li, 2012).

Mohbeya and Thakura research indicated that prediction of UB models are useful because they allowed defining new services and participated in improving existing ones. Furthermore, it can help in managing service infrastructure (Mohbeya and Thakura, 2015). Smit et al. research indicated that available knowledge is still insufficient to obtain good understanding of this new advertising technique (Smit et al., 2014).

Whereas in the domain of online entertainment, Dorčák et al. managed to identify direct relationship between suppliers' activities online and tools used to promote the business. Furthermore, Dorčák et al. proposed a smart streaming strategy to improve the overall streaming service. Firstly, the system will collect user viewing data and, based on this data, the system will avoid wasting their resources by predicting user departure behaviour (Dorčák et al., 2015). Whereas, Boland et al. proposed model that analyses user behaviour on music retrieval systems in a systematic method: the model is layered

to understand different user involvement patterns (low to high involvement) (Boland et al., 2015).

The investigation of UB on the internet also included exploring problems in social intensive communities which sometimes utilise RSs techniques to recommend personalised services to information seekers/ users. Hence, Lops et al. investigation of SNS and CBF aimed to enhance recommendation to information seekers/ users of academic research papers based on users' interests (Lops et al., 2011a). Whereas, Gao et al. proposed a predicting algorithm model that automatically identify personality traits in correlation to Social Media contents (Gao et al., 2013b). Furthermore, Gao et al. suggested a user-centric approach that will consider ranking and clustering of interfaces to provide users with organised search results and allowed users to personalise and organise search results (Gao et al., 2013a).

Other attempts to study UB investigates the transformation behaviour of information seekers/users, the impact of internet technology on topics such as phenomenological and new ways of studying existing topics. Furthermore, point out emerging opportunities when using the Internet, such as:

- The viability of relying on the online population for research sampling purposes,
- The impact of emerging technologies “smartphones”, and
- The benefits the Internet brings to the research of psychology (Gosling and Mason, 2014).

#### 2.4.5 SUMMARY

According to Schmidt-Maenz and Koch some of the welcomed results from UB on the internet research was to:

- identify popular search terms.
- identify existing methods to classify search terms.
- create a basic classification of search terms.
- characterise usage of search terms.
- Possible ways to extract topic clusters in search queries (Schmidt-Maenz and Koch, 2006).

However, although researchers employed variety of techniques to collect data about both information seekers/ users and information retrieval systems on the internet. Yet,

research of UB on the internet did not deliver significant findings to change how search engines and RSs interpret and process queries to deliver improved search results or RSs recommendations (Schmidt-Maenz and Koch, 2006).

Li argues that the vast range of information sources on the internet and users' ability to access these information sources at anytime and anywhere did not help users to retrieve relevant information; thus, created IO. Li also argued that IO in these situations can be the consequence of information seekers/users inability to clearly define their information requirements when searching the internet (Li, 2012).

Junghans et al. criticised search engines which still use 2 decades old information retrieval techniques (syntactic matching of keywords, tagging of services) in the process of delivering search results to information seekers/users on the internet. Consequently, current search results do not fulfil information seekers/user information needs (Junghans et al., 2012).

The rapid growth of information on the internet resulted into two things; an information-rich environment and IO, both at the same time. Tan, investigation, argued that the huge volumes of data on the internet made obtaining information a tedious task (Tan, 2013). Similarly, Schmidt-Maenz and Koch complains that finding information became a tedious task – many users complain about their inability to find the right information at the right time (Schmidt-Maenz and Koch, 2006).

Lops et al. on the other hand argued that although SNS overload they offered users on the internet useful, accurate, and constant up to date information (Lops et al., 2011a). However, if service providers do not take into account that

- (a). users may change their mind while searching the Internet and
- (b). they can be distracted at the same time by various browser s and search engine functionalities,

then, it is definite that service providers will not be able to offer a significantly different solution to IO in Internet searches. In other words, if we do not model UB with (a) and (b) in mind, then we will not offer anything new in resolving the problem of IO in Internet searches.

Hence, the best approach would be to capture the semantics of the “moment” when a particular internet search happens and ignore everything else, which might not be

directly related to this particular search. It would not be welcome to remember what previous searches have brought to the same information seeker/user and insist that their semantics remain present in our models. If we allow for storing “old” semantics from previous searches, we will not be able to handle the changes imposed by users: they must be able to change their mind!

This section provided an overview of the history of search engines and the development stages they went through. This helped us in providing information on early types of search engines and the categorisation of search engine based on their functionalities and techniques. All these techniques participated in either improving search engines’ performances or providing search results based on relevance, accuracy or user interests (see section 2.4.1 and 2.4.2). Furthermore, we extended our investigation in this section to understanding the role of semantic search engines that claimed that they refine search results to meet the semantics of the environment and user queries. In section 2.4.4, we introduced additional types of search engine techniques. Therefore, in section 2.4.4.2, we discussed the influence of ranking search results based on recency, time, situation and so forth. We also empirically analysed User behaviour from distinctive aspects. We briefly discussed the research on user behaviour from the point of view of psychologists and enriched our analysis with research conducted by CS researchers to understand the user behaviour on the Internet. We provided an overview of different studies. From section 2.4.4 to 2.4.4.6, we also discussed the different approaches followed by the researchers to conduct their experiments and analysis of specific issues in search engines and their techniques.

The section on search engines and their techniques enriched our research as follows:

- It provided us with a deep understanding of the role of search engines in retrieving information on the Internet.
- It provided us with a critical analysis of the different techniques used to improve the performance of search engines.
- It provided us with the influence of user behaviour in enhancing search engine performance.
- User behaviour on the Internet can be advantageous in some cases. For example, when collecting data about users’ activities to refine search results. However, the lack of user involvement on the system can also negatively influence search results.

The table below summarise some search engines from 1990 to the current date.

- The discussion in the previous sections is linked to section 2.6. In this section, we discuss the importance of relevance in search results to users queries. The section provides an overview of different attempts to understand user queries through system logs, query logs, user involvement and users' activities on the Internet.

The table below summarise some search engines from 1990 to current.

TABLE 2.1 OVERVIEW OF SEARCH ENGINES

Search Engine Name	Description	Publisher	Website	Year
Archie	Collects/download's data files from FTP into databases.	McGill University	<a href="http://archie.icm.edu.pl/archie-adv_eng.html">http://archie.icm.edu.pl/archie-adv_eng.html</a>	1990
Gopher	An indexer and a menu system that aimed to distribute search and retrieve documents on the Internet.	University of Minnesota	<a href="http://www.goodgopher.com/">http://www.goodgopher.com/</a>	1991
Veronica	Keyword based searched for files by names and titles	University of Nevada	Does not exist anymore	1992
Jughead	An ad-hoc based process to create new menus from user queries and present a customized Gopher menu	/	Does not exist anymore	1992
Perl Script	Mirror the website into standard format catalogue	University of Geneva	Does not exist anymore	1993

Wanderer	Generate index named wandex that aimed to measure the size of WWW	Massachusetts Institute of Technology	Does not exist anymore	1993 to 1995
AliWeb	Allowed users to submit index file location on their websites, and enabled to add user scripts	Nexor Limited	<a href="http://www.aliweb.com/">http://www.aliweb.com/</a>	1993
Jump Station	An interface based on web robot indexer to run their query program. First resource discovery tool that crawled, indexed and searched the web.	University of Stirling in Scotland	<a href="http://www.jumpstation.com">http://www.jumpstation.com</a>	1993
WebCrawler	Text based crawler, a keyword level query tool	Blucora (formerly Infospace, Inc.)	<a href="http://www.webcrawler.com/">http://www.webcrawler.com/</a>	1994
Lycos	Commercial search engine was released	Carnegie Mellon University	<a href="http://www.lycos.co.uk/">http://www.lycos.co.uk/</a>	1994
Meta-Crawler	The first algorithm that allowed to search 20 search engines and directories at a time	Colorado State University	Does not exist anymore	1995
Excite			<a href="http://msxml.excite.com/">http://msxml.excite.com/</a>	

The discussion in the previous sections is linked to section 2.6. In this section we discuss the importance of relevance in search results to users queries. The section provides

an overview of different attempts to understand user queries through system logs, query logs, user involvement and user activities on the Internet.

## 2.5 THE SEMANTIC WEB TECHNOLOGY

In “As We May Think” which was published in “The Atlantic Monthly” July 1945 Vannevar Bush envisioned a mainframe that will deliver to information seekers/users a perfect source of information. An utopia of information reachable and accessible by everyone any time. “*The Memex*” a personal device that will contain all books, records and communications such as storing techniques (microfilm and indexing) to highly compact storing (head-mounted compact micro-cameras and voice input devices) (Bush, 1945). Bush’s vision was inspired by:

- the inability of individuals to access information and,
- scientists’ frustration of not being able to get the desired know from traditional libraries.

Since the publication of Bush’s article, the article became a reference to many technological advancements. Hence, researchers believed that Bush’s article paved the road to many inventions. David M. Levy in his review of Bush’s visionary article argued that Bush’s article paved the way to current technologies (in particular the hypertext) and it was the road map for many inventions even more than what Bush imagined and the correlation of vannevar’s vision and the manifestation of modern IO (Levy, 2007).

Consequently, many researchers attempted to map current trends in technology to the vision of Vannevar Bush to major how far technology advanced and its contribution in an ongoing development of technological inventions including cyber infrastructure and education (Bush, 1945).

Early attempts to information retrievals can be found in Gerard Salton group research who developed the SMART informational retrieval system “Magic Automatic Retriever” that took into account vector space model, Inverse Document Frequency (IDF), Term Frequency (TF), term discrimination values, and relevancy feedback mechanisms (Salton, 1971). Ted Nelson created the first project “Xanadu” which aimed to provide a network with simple interface to the users. Late, coined the term “hypertext” (PROJECT-XANADU, 1960) and ARPANET was the first step to the birth of the internet (ARPANET, 1962).

In 1989 Tim Berner-Lee invented the current World Wide Web (WWW), Tim's eagerness as a scientist at CERN (The European Organization for Nuclear Research) encouraged him to develop a method which will allow him and other scientists at CERN to share and exchange data among themselves. Tim's invention later became the internet network that is now shared among all people worldwide. A proposal was submitted that described a combination of technologies. These technologies included an improved vision of Hypertext Markup Language (HTML), Uniform Resource Identifier (URI) and Hypertext Transfer Protocol (HTTP) which are the foundation of the current web (WorldWideWeb Foundation, 2008 - 2015).

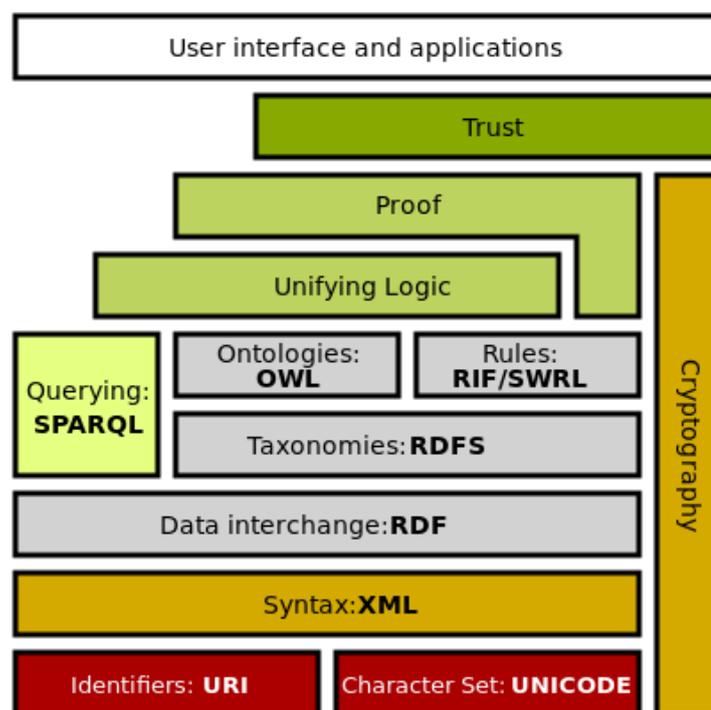


FIGURE 2.7 THE SEMANTIC WEB TECHNOLOGY STACK (SWT-STACK, 2008) .

In 2001 Tim Berner-Lee announced the new web “*The Semantic Web*”. A web that will allow human and machine interpretations. A structured and meaningful web that will create an environment for computer agents to roam from a page to another and is an extension of the current web where information is well defined. For this to happen SW required a structured collection of information and inference rules. SW vision also aimed to provide better knowledge representation which will avoid centralisation technique in traditional knowledge representation systems.

The first attempt to develop SW delivered two technologies: eXtensible Markup Language (XML) that allowed users to add arbitrary structure to their documents and

Resource Description Framework (RDF) that provided methods to express meanings in triples. Hence, triples defined subject, verb and object of a sentence that describes things people, objects, websites, and other things on the internet. Triples were further enhanced to include description of properties that defined relationship of triple elements identified by a Universal Resource Identifier (URI). Hence, URIs were used to enable interactions with representations in a network such as the WWW. Whereas, Resource Description Framework Schemas (RDFS) provided vocabulary for describing properties of RDF resource as depicted in figure 2.8 below.

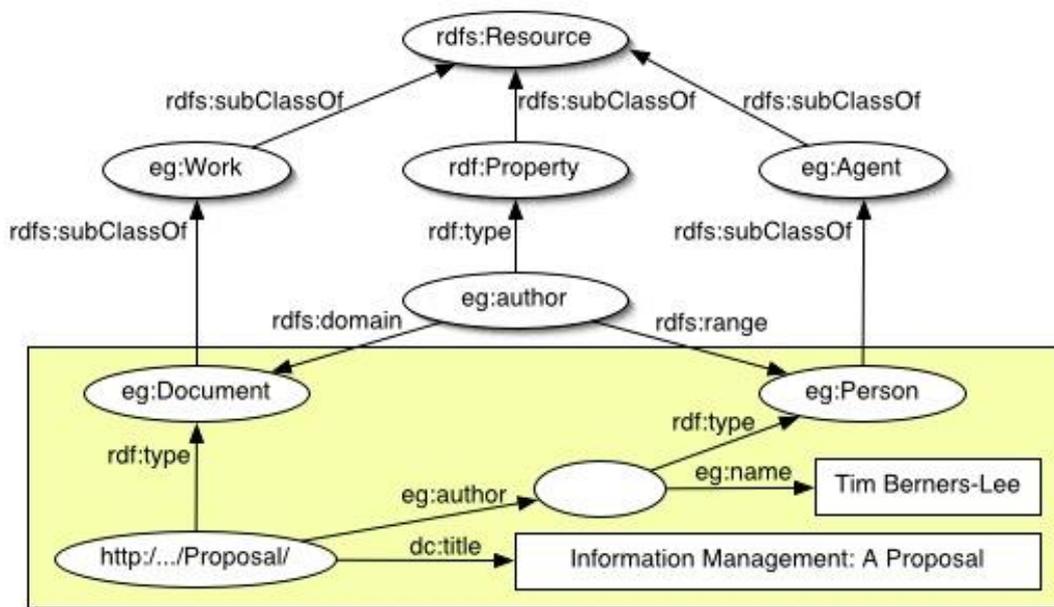


FIGURE 2.8 THE USE OF RDFS TO DESCRIBE RESOURCE(S) CLASSES AND PROPERTIES (W3C-RDFS, 2002).

Figure 2.8 above illustrated way of using RDFS to describe real world things/objects, the classes they belong to and the description of things related to each other (W3C-RDFS, 2002).

The Semantic Web Technology (SWT) now is used for variety of reasons. The huge collection of published work that adopted the SWT as technology of choice attempted to solve many domain specific problems. Furthermore, SWT allowed data on the web to be defined and linked; thus, used effectively to discover, annotate, integrate and reused across various applications on the internet (Berners-Lee et al., 2001).

Distributed extensibility feature of the SWT allowed to interlink between websites that may contribute to provide data about a particular resource. This can also mean that

SWT can extend the cumulative knowledge on the SW about resources in a distributed fashion (Guha et al., 2003).

Software developers adopted the SWT and its stack to enhance the representation of knowledge on the internet. hence, SWT was utilised across disciplines such as education, healthcare, travel and tourism online booking interlinked to other services and so forth. Gradually, SWT was slowly spreading into other domains such as government, business services, life sciences, communication and media, and variety of healthcare services (Shadbolt et al.).

Advantages of utilising SWT is that it can be interpreted by humans and computers as the same time. Moreover, it also offered an added value and comparative advantage to its users. SWT developers a very important feature which is interoperability between applications available on the Web. consequently, well know software and platform providers pioneered in the adoption of SWT to enhance their services. Oracle for example introduced the first RDF management platform focusing on application areas (life sciences, data and content integration, enterprise application integration, and supply-chain integration) (Oracle, 2010). Vodafone also mobilised RDF to improve their web search features and help their users in the retrieval of (ring tones, games, and pictures) (Smith, 2007).

The vision of web 2.0 and its technologies and the introduction of SWT and its stack encouraged Cardoso to investigate the power of the Web 2.0 tools and SWT and its stack. Thus, he surveyed current snapshot of key trends and developments and usage of SWT and its stack. Results indicated that SWT and ontologies in particular are heavily utilised to improve knowledge representation. Whereas, RDF's are deployed for data sharing and integration. The survey also indicated a rapid increase in the use of SWT and its Stack to deploy real world systems (Cardoso, 2007).

### 2.5.1 ONTOLOGY DEFINITION

The Semantic Web Technology (SWT) and its stack is widely known as the next generation of the web. Its vision was to enhance web contents with metadata; to enable human-machine readable contents. This to process, share and interpret web contents and deliver enhanced services. There are many successful implementations of general and domain specific research that utilised SWT and its stack. For the purpose of this research

we wish to focus on research that employed Web Ontology Language (OWL) Ontologies to enrich the web contents semantically. The choice of SWT and Semantic Web Rule Language (SWRL) enabled OWL ontologies in specific was based on the flexibility the technology brings to software developers. Furthermore, OWL ontologies can play a key role in the SWT. Some of the benefits noticed while using SWT and its stack is that it provides a source of shared and precisely defined domain specific knowledge. Moreover, an ontology can consist of a conceptual schema of a domain; presented in a hierarchical description of important concepts, along with the description of their properties (constraints) and enriched with the presence of domain specific knowledge. OWL ontologies also deliver added value of Semantic annotation that effective information retrieval. Hence, it enhance the internet with more capabilities of processing and understanding the semantics of information available on the internet. Consequently, it allows relevant information to be directly discovered.

*“An ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules”.* Taxonomies can include classes and has relationship defined through properties. The sub-hierarchical structure of taxonomies classes can provide explicit semantics of an environment. The ontologies can be enriched through the use of inference techniques. One advantage of using ontologies is that they enhance functioning of the web in many ways. For example, they can improve accuracy based on concepts stored in the ontology (Berners-Lee et al., 2001).

## 2.5.2 OWL ONTOLOGY

Section (2.4.2.4) discussed some of the previously explored solutions in the domain of Social Tagging systems; as a mean to classify large sets of resources shared among information seekers/users of social intensive environment on the internet known as folksonomies. Limpens and Gandon to solve the problems ambiguous tags and the problem of spelling of these tags proposed to exploit the power of ontologies in order to fully utilise folksonomies in the process of retrieving or exchanging information and overcome some of ST systems problems. Hence, utilised SWT -specifically OWL ontologies- the formal languages and ontologies (vocabularies) offered by SWT stack. This to interpret the semantics and understand the meaning of concepts in folksonomies to enhance the performance. Thus, suggested WordNet a lexical database for English; semantically understand the meaning of words and group them according to specific sense of the word. According to many research groups, WordNet can help in reducing ambiguity in tag meaning and allow better understating of domain vocabulary (University, 2010).

The spelling problem used to be a serious issue in all online environments; it directly affected the retrieval of information on the internet. A misspelled word can result into eliminating the chances of a certain source of information from being discovered. Hence, there exist solutions that explores alternative solution to guarantee full understanding of meaning of misspelled words in user's queries. Thus, reduce chances of retrieving irrelevant information sources. Martins and Silva for example, proposed an algorithm that attempted to select the best choice of word among all possible corrections for a misspelled term; to understand user quires that are submitted to a search engine, and implement corrected query based on a ternary search tree data structure (Martins and Silva).

Attempted to improve certain services also involved combining technologies to produce a refined solution to a specific problem domain. SWT and its stack was considered by many researchers as technology of choice. Limpens and Gandon for example attempted to leverage knowledge sharing on the social web by first understanding the semantics of folksonomies and then build a lightweight ontology that will describe the structure of communities, interaction between the member and contents

shared between user of the community. Thus, Limpens and Gandon utilised Semantically-Interlinked Online Communities (SIOC) vocabulary (W3C-SIOC, 2010) to provide a formal and technological framework that describes resources exchanged across the community. They also align and merge other ontologies such as Simple Knowledge Organization Scheme (SKOS) (W3C-SKOS, 2004), that describes systems of organization of knowledge, and Friend Of A Friend (FOAF) (FOAF, 2000-2015) to enrich the semantics of the environment.

Similarly, Straccia employed SWT stack to improve Distributed Search (DS). Therefore, addressed issues of DS in large number of heterogeneous and distributed information sources on the internet. Thus, suggested an algorithm based agent; that will perform selection of relevant information sources through subsets of web sources stored in multiple ontologies. And further query all sources for the purpose of selecting a resource. Furthermore, the proposed approach also used ontology alignments and last merged selected sources together (Straccia, 2003).

Guha et al. proposed an application “Semantic Search”; that utilised SWT technologies to improve traditional web searching. The power of this solution can be seen in its ability to understand the search query, and embed aggregated data from distributed sources into traditional search results.

This research is based on the following assumptions:

- Semantic Web (SW) will contain interlinked resources and data on the Semantic Web is modelled as a directed labelled graph.
- Node corresponds to a resource and each arc is labelled with a property type (also a resource).

The proposed solution utilised Resource Description Framework (RDF) and Schema Vocabulary -Resource Description Framework (RDFS); this to define resources and their relationships. Furthermore, Simple Object Access Protocol (SOAP) is used as a protocol to query and exchange RDF instances. Consequently, the proposed approach can enhance TSE technique thought augmented features of SWT (Guha et al., 2003).

SWT was also used to solve information retrieval problems from the domain of search engines, RSs and their techniques. Stojanovic et al. and Chau et al. employed Semantic Web Technology (SWT) in order to address relevance and rankings of search results (Stojanovic et al., 2003) (Chau et al., 2012) this due to SWT ability to interpret

the meaning behind user queries and environments where they are created (Almarri et al., 2013), (Almarri and Juric, 2013a). Moreover, address relationships between search results by exploiting metadata of semantic web sources, which was proposed in (Aleman-Meza et al., 2005) and (Anyanwu et al., 2005).

### 2.5.3 SEMANTIC SEARCH ENGINE

Modern technologies enabled new ways of information generation on the internet. Hence, the type and quality of information solely depends on information seekers and producers on the internet (ex: Government, Educational Institutes, Organizations, Companies, Businesses, and the most powerful source is users on the internet). Unfortunately, these modern technologies created new type of IO. Consequently, research communities' continuous attempts to develop new methods to help users on the internet to find the right information resulted into investigating the possibility to employ SWT technology to improve search engines results.

Sudeepthi et al. argued although TSE delivered results to their users, they also bombard them with abundance of results. Hence, users suffered from IO and required more time to go through search results because they were unable to judge how relevant results are to their queries. Furthermore, common reaction to abundance of retrieved results is that users are forced to view every links to view information. consequently, Sudeepthi et al. complained that TSE are not adequate to handle the increasing volume of data on the internet (Sudeepthi et al., 2012).

Sudeepthi et al., claimed that problems of TSE and the benefits SWT stack bring to the users on the internet has encouraged the development of Semantic Web Search Engine (SWSE). Semantic Search Engines (SSE) are another attempt to deliver search results to information seekers on the internet by avoiding some of Traditional Search Engines (TSE) problems such as relevance in search results, excessive amount of them and duplication in search results. Hence, SSE aimed to enhance search engine results through interpreting the meaning of the information source(s), their contents and understand the role of the use in the process of information retrieval on the internet (Sudeepthi et al., 2012). Thus, according to Khan et al., SSE played a crucial role in the enhancement of TSE (Khan et al., 2014).

The role of TSEs as powerful information Retrieval systems is undeniable but the existence of SWT influenced TSEs techniques. Consequently, search engines developers were interested in utilising SWT towards the new generation of search engines SWSEs. There are many type of SSEs in the SWSEs which claims that they are based on SWT stack and delivers results according to the semantics of users' queries.

Some of the common issues of TSEs that SSEs tries to solve are:

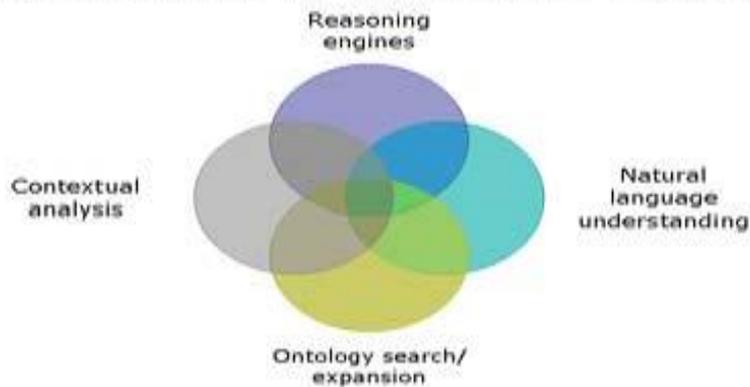
- Lack of common structure of information on the internet.
- Poor description of sources that results into excessive information retrieval.
- Increasing volume of users and information sources resulting with a lack of trust in both quality and quantity.
- Automatic information transformation.

There are four approaches to SWSEs in literature that takes into account basic functionalities of search engines and are depicted in figure 2.9 below:

- ❖ Contextual analysis, which allows to understand the meaning of things (i.e. the word 'Cat' can have multiple meaning).
- ❖ Reasoning layer, that allows to create problem domain description by utilising SWT stack to model the relationship of domain concepts and reason upon its concepts to extract further information from existing information.
- ❖ Natural language processing technique, allows to analyse the contents of users' queries to identify people, places, organizations, and so forth.
- ❖ Semantic knowledge representation in ontologies, this technique allows to create taxonomies of concept (i.e. truck is a vehicle). It also help in making a broad and precise meaning of things, which helps in delivering a unique search results to the users (Sudeepthi et al., 2012).

SSEs techniques utilise other methods such as using equivalent terms, divide search results into groups according to relevance, and they use semantic rank algorithms to deliver better search results.

## Approaches to semantic search



### Address somewhat different problems

FIGURE 2.9 APPROACHES TO SEMANTIC WEB (SUDEEPHI ET AL., 2012).

Some of the approaches that belongs to SWSW category is CARROT2, an open source text mining SSE that aimed to solve ambiguity in search queries though:

1. collecting user's queries and send them to TSEs via API pull/push queries and search results to CARROT2 users.
2. analyse and understand the user's query and compare it to those returned from TSE's search results.
3. filter top 100 TSE search results by applying CARROT2's techniques in step

Additional techniques are used that depend on each query's requirements (i.e. clustering technique that allows the collection of a max number of possible groups on a topic) (Carrot2, 2005). Furthermore, in the query panel, CARROT2 offers its users to input/ algorithm selection, number of search results to fetch and default algorithm configuration settings (Carrot2, 2005).

Kosmix focused on achieving two important goals this by exploiting SWT for (topic exploration and Deep Web crawling) by linking data on the internet to deliver relevance in search results. furthermore, allowed grouping of search results into categories such as (Video, Web, News & Blogs, Images, Forums, Twitter, Amazon and Facebook).

Powerset is a collaborative Natural Language Search Engine (NLSE) that interpreted every sentence on the Web to find target answers to user questions. NLSE also attempted to avoid keyword based searching by understanding the nature of users' questions.

For instance, TSEs will collect words from a query and try to match them with a maximum number of data/information on the Internet. Whereas Powerset SSE will try to process the query by finding an answer instead of a huge list of websites that contain part of the answer as depicted in figure 2.10 below.

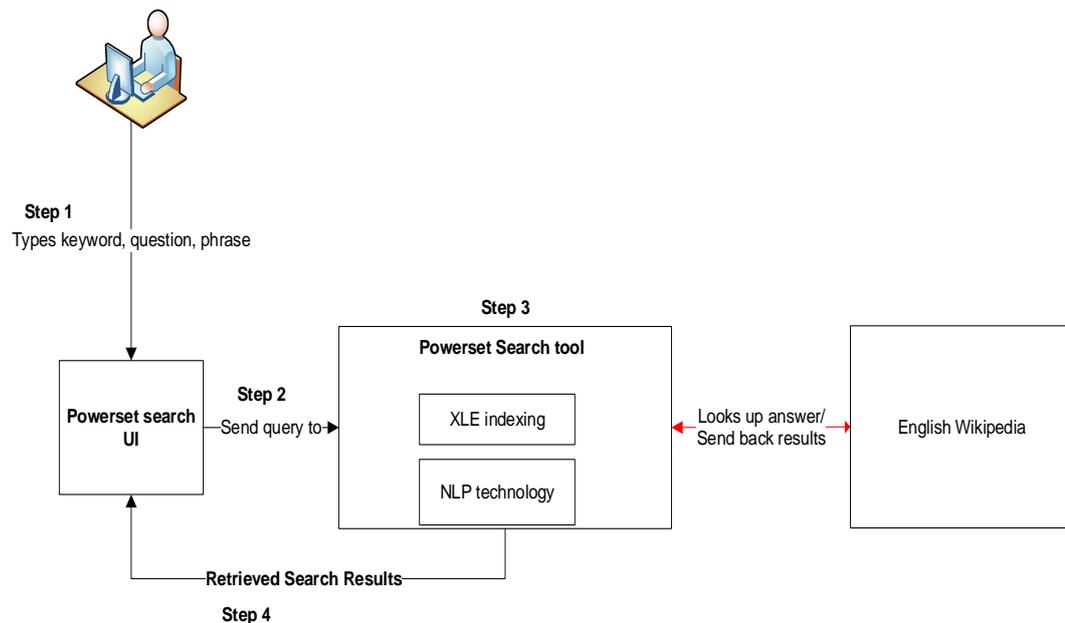


FIGURE 2.10 DEPICTS THE PROCESS OF PROVIDING SEARCH RESULTS IN POWERSSET SSE.

Furthermore, it indexes and queries the system for answers to extract semantic “facts” from previously indexed data and displays semantic connections between words and concepts. Powerset allows categorisation of search results into conventional search results with links to relevant Wikipedia pages, relevant subject/relation/object triples related to the user’s query. Hence, Powerset claimed that they automatically extract semantic facts that can be used in the creation of extended knowledge resources including lexicons, ontologies, and entity profiles (Sudeepthi et al., 2012).

There are many other techniques used to improve the performances of SSEs. For example, Sensebot is a text mining SSE that analyse the Web contents and performs multi document summarization to identify semantics of concepts on the web and understand users’ queries to deliver accurate and adequate amount search result. One of Sensebot features is that it allows to choose how to perform search query. For example, the user types in a search query, then selects to either search Sensebot, Google, news contents only, and choose a language (Sensebot, 2007), (Lei et al., 2006).

DuckDuckGo, is semantic rich directory that allows its users to search the internet for data in three different ways (classic search, information search and shopping search). Furthermore, attempted to improve search experience by avoiding excessive amount of search results and it solves term ambiguity by predicting terms in the search query. Most important is that DuckDuckGo avoided the use of a common technique of TSEs (building user profiles) (DuckDuckGo, 2008).

The table below provides of current SSE that attempts to deliver search results based on the semantics of the domain. For instance, Swoogle is a crawler that searches the web to discover and index ontologies and RDF. Whereas, Silvia and kngine are SSE that crawls the web to find similar images. Another type of SSE is Klevu that interprets the meaning of keywords in search queries to create categorisation of topics based on the semantics of the environment. Similarly, Carrot2 interprets users queries to deliver categorised search results based on topics and other user interests. Yummly and DuckDuckGo SSE interprets the semantics of users queries to deliver personalised search results and minimise excessive amount of them.

Hence, most SSEs claimed that they avoid to base their search results on keyword matching, popularity ranking, and popular search terms which can be found in TSEs and deliver search results by utilising SWT stack. However, after investigating all these SSEs it is still not clear which exact layer of the SWT stack is used in the process of developing all these SSEs. Therefore, it is important to bring forward the reader's attention that the proposed solution in this research does not below to SSEs due to the lack of understanding of how SWT is utilised in:

- modeling and interpreting the semantics of the environment where searches happen
- interpreting the semantics of users preferences to deliver search results.

TABLE 2.2 SEMANTIC SEARCH ENGINES

Semantic Search Engine Name	Description	Website	Year
Swoogle	A crawler to discover ontologies, RDF embedded in HTML	<a href="http://swoogle.umbc.edu/">http://swoogle.umbc.edu/</a>	2004
SILVIA	An indexer and image matcher	<a href="http://silvia4u.info/">http://silvia4u.info/</a>	2007
Klevu	Understand keywords and creates automatically categories based on semantic logic	<a href="http://www.klevu.com/">http://www.klevu.com/</a>	2013
Yummly	Semantically search for food recipe and create personalized recommendation	<a href="http://www.yummly.com/">http://www.yummly.com/</a>	2009
kngine	A concept based semantic search engine that divides search results into either web results or images	<a href="http://kngine.com/">http://kngine.com/</a>	2013
CARROT 2	Semantically organize search results into topics	<a href="http://search.carrot2.org/stable/search">http://search.carrot2.org/stable/search</a>	2002
Kosmix	Social media search engine	Does not exist anymore	2005
Powerset's	Natural language processing semantic search engine	Does not exist anymore	2006
Sensebot	Provides summary of search results based on text mining and multi-document summarization	<a href="http://www.sensebot.net/">http://www.sensebot.net/</a>	2007
Hakia	Semantically categorized search results based on relevance to main topic	<a href="https://web.archive.org">https://web.archive.org</a>	2004
DuckDuckGo	Aims at maintaining users privacy and avoid the personalization of search results	<a href="https://duckduckgo.com/">https://duckduckgo.com/</a>	2008

## 2.6 CHAPTER SUMMARY

Chapter 2, provided an itemisation of some of the well-known techniques that were developed to search and recommend information sources on the internet: search engines, RSs, and their techniques are solutions which were involved and aimed to:

- Manage information for both information seekers/users and service providers e-commerce.
- Understand the nature of information available in social intensive environment.
- Improve the existing techniques to deliver relevant recommendations, (e.g. RSs technique in e-commerce
- RSs incorporated tagging, annotation, and folksonomies into CF and CBF techniques to Hybrid RSs recommendations.
- Many techniques also attempted to interpret information seekers/user interests to personalise recommendations, by utilising system logs and query logs to extract user behaviour and usage patterns on the internet.
- Some other attempts to enhance recommendations were based on the information collected to create a history profile of information seekers/users' sessions of activities and interactions with other users or services on the system.

Whereas, search engines were the next generation of IRS to discover information on the internet. They were developed to manage new ear of information representation and retrieval. Hence, they aimed to make information sources easily reached by information seekers/user on the internet regardless to their competence level. Thus, search engines employed 100s of techniques to collect and index information sources, analyse the characteristics of information sources and attempted to incorporate them in the process of delivering search results to information seekers/users on the internet.

Like RSs, search engines also attempted to enhance search results by incorporating information seekers/users when searching and viewing search results. Hence, search engines aimed to:

In this chapter, we presented the current literature review of our research problem domain (IO). We presented a selection of problems, solutions, and examples of our problem, proposal and illustration in Chapters 3, 4, and 5. We also justified our selection of research background topics and the influence of them on the proposed solution. Our research is a complementary solution to much similar research which tries to solve the problem of online IO through the use of the widely spread techniques as listed in Chapter 2 sections (2.1 - 2.10) above. Although several research individuals/ groups tried to solve the problem through a variety of technique, we found that IO still exists in many domains and different groups of people still suffer from IO in their daily lives at home, at school, at Universities, and work. Research so far has not presented an ultimate solution to the problem of Online IO. Most of the existing works only attempt to combine different techniques to propose a new solution. RSs and search engines are the only method known to date that integrates all existing techniques to improve search results. These techniques are limited to past behaviour, which in many cases are not necessarily relevant to the user's queries. One specific application, i.e. is Amazon's search results, which look at the user past behaviour to recommend new results. Thus, there is a need for a method that will avoid problems of existing RSs and search engines problems.

## CHAPTER 3. RELATED WORK

Dealing with IO is especially challenging when accessing information sources on the internet. In the last decade, users have witnessed dramatic changes in mobile/wireless technologies and the proliferation of the internet. All these advancements have enabled a revolution in the way information is created on the internet. Therefore, in the current state of WWW, over a billion of webpages exist and information can now be viewed on an ad-hoc basis. Consequently, an urgent call was required to control the excessive amount of information that is created instantly. Hence, many techniques surfaced to manage information through different data aggregation techniques that required data of both information sources and information seekers/users' involvement on the internet. This to ease the process of information management, presentation, and retrieval and deliver personalised recommendations or search results.

Early approaches to information discovery can be found in library IRS. Whereas, current techniques of information discovery are RSs in e-commerce which allowed filtering of information interpreted from contents of information sources and information seekers/users' collaboration in social intensive environment and on the internet. RSs techniques were also extended by utilising many information seekers/users and products, services, and information profiling models to recommend information according to information seekers/users interaction with other users, products and services on the internet. Search engines and their techniques heavily participated in information discovery on the internet. Hence, search engines eased information seekers/users navigation process to access knowledge and services on the internet. furthermore, search engines techniques helped in controlling information presentation and took into account information relevance, accuracy and personalisation of search results.

At the time of writing this thesis, I was not aware of any research or method that attempted to deliver or select information sources according to their semantics or utilised SWT in particular SWRL enabled OWL ontologies computations select information sources according to users preferences. However, I would like to mention few research topics which has some common objective and connection to core components of the proposed solution in this research. Moreover, I would like to bring forward to the reader's attention that this work does not aim to criticise previously proposed working solutions

that existed for certain problem domains. As a matter of fact, in this chapter I would like to convey my opinion about existing work in the area of information retrieval on the internet which can be slightly similar to some the aspects of the proposed solution in this research.

Many researchers argued that the deterioration of the quality of search engine results, RSs and their techniques is mostly due to their inability to skip the imprisonment of three decade old solution that relayed on user profiling which were perfect for the purpose of information management, representation and retrieval at the time they were invented. However, modern technologies empowered information seekers/users on the internet with the most crucial tool. Information seekers/users are now creators and consumers of information sources on the internet. Hence, information scientist argued that information seekers/users are unable to survive the avalanche of information sources on the internet.

The incompatibility of most search engines and RSs techniques to modern information sources and users of these advanced technologies in the 21<sup>st</sup> century created excessive amounts information and retrievals in search results, rather than reducing it. Therefore, it heavily contributed towards modern IO.

This chapter will discuss to some extent previously proposed approaches which at some point addressed:

- The accuracy and relevance in recommendations or search results to manage excessive amount of information on the internet, **which resulted into IO.**
- Attempts to understand users' role and influence when creating and using information available on the internet.
- Approaches that attempted to deliver the best possible recommendation or selection of internet sources based on a situation or a problem domain.

### 3.1 ACCURACY AND RELEVANCE IN SEARCH QUERIES

The accuracy and relevance of search results are very critical issues in search results. Many search engines claim that their search results provide accurate and relevant search results. Furthermore, they argued that by understanding information seekers/users' queries, they are able to detect the most relevant information source on the internet to answers to information seekers/ users queries. Therefore, current search results, according to search engines, delivers accuracy and relevance. This research argues that

search engines create excessive amount of search results because search engines still delivers search results based primitive technique “keyword” matching. Consequently, information seekers/users also suffer from huge volume of retrieved information sources. Hence, information seekers/users are forced manually filter information sources to judge whether relevant and accurate information sources are retrieved in search results or not. Thus, information seekers/users are also required to go through every retrieved information sources to decide if knowledge can be gained from viewing some or all search results.

Sections on RSs and search engines discussed some of the followed practices to deliver accuracy and relevance in search results according based on certain criteria. Thus, it is important to point out that:

Most of the existing work on TSE (especially commercial search engines) hide their algorithms from competitors because they are business secrets. Therefore, researchers struggle to find any proper documented information of search engines and their techniques. Similarly, SSEs claimed that they deliver search results based on understanding the environment and utilising SWT languages. Even though it was very difficult to find any details about which language is used to deliver semantic search results.

Search engines utilises wide range of techniques to deliver search results to information seekers/ users on the internet. Thus, techniques are used in every single process from data collection to the delivery of processed data as search results. Hence, many factors determine the quality of search results delivered to information seekers/user on the internet. The investigation on search engines and their techniques in this research can hardly find clearly defined search engines process of information discovery on the internet. However, so researchers attempted to analyse search engines performance based on observing them for a certain period.

Accuracy in search results is of high importance when delivering search results. search engines use multiple methods to measure accuracy in the process of extracting terms and keywords from search queries. Some researchers claimed that accuracy in search results can be measured by employing relevance criteria in the process of information discovery. Hence, Zaragoza et al. to understand which fraction of the search queries provide excellent search results proposed takes into account lower and upper

bound analysis of information over the standard relevance measures adopted for the domain of IRS. Furthermore, they introduced the concept of disruptive sets to estimate the degree of the search engine's ability to solve unsolved queries by search engine competitors (queries that delivered minimum search results or results which has no relation to the original query). Consequently, Zaragoza et al. attempted to add new criteria to relevance measures in search engines results to enhance accuracy in delivered search results (Zaragoza et al., 2010).

Accuracy in search results are also claimed in search engine ability to deliver diversification in search results. Search engines attempts to deliver diversification was to ensure that the process of information discovery is not restricted to keyword matching in search results. Hence, in the process of extracting data from information seekers/users query, search engines attempt to extract contextual meaning of search query contents to discover alternative ways of describing the same search terms. Lin et al., proposed a novel algorithm, DATAR, which aimed to improve GRASSHOPPER (a framework of absorbing random walks to understand click through in relation to the search query) techniques (Zhu et al., 2007) of search engines and take into account the search results' diversification to deliver effectiveness and user satisfaction in search results. Furthermore, this work also considered that redundancy in top ranking results often disappoints users. Therefore, they (Lin et al., 2010).

Clustering of search results was used to ease the navigation to information sources and is considered another technique which measures accuracy in search results. in the process of information discovery, the search engine utilises clustering of matching information sources to the information seeker/user search query according to terms which describes document (title, topic, authors name, year of publications and so forth). Furthermore, some attempts also considered personalised clustering according to information seekers/users' interest and profiles. Zeng et al. for example attempted to improve the effectiveness of search engines through the organization of search results by using customised clustering techniques to enable quick browsing to information sources on the internet. Moreover, re-formalize the clustering problem as a salient phrase ranking problem (Zeng et al., 2004).

Search engines, RSs and other information retrieval systems performance are heavily influenced by many techniques when delivering search results/ recommendations to

information seekers/users on the internet. In addition to basic techniques used in the process of searching/recommending information sources, IRS on the internet also utilised hundreds of techniques to deliver relevant recommendations/ search results according to search queries.

Ranking algorithms, optimisation techniques of user queries and design of IU are some of the main factors that influenced search results or RSs recommendations. Therefore, it is important to understand what relevance means, and how to treat it in the process of delivering recommendations or search results of information on the internet.

Croft et al. for example, measures effectiveness and efficiency of search engines to assurance that the delivered search results are relevant to the analysed search query. Hence, Croft et. al. employed relevance judgement technique which process and refine search results based on the original query multiple times to calculate relevance. The process starts with running basic search of indexed information sources to analyse the query and find keyword matches in search engine repositories. The search engine analyses the retrieved results and rerun the search query to find further matching information sources based on search terms and retrievals from the first attempt. The search engine continues to run the same query to them point that no more new information sources can be retrieved. The outcome of this process selects information sources with minimum relevance to the original query (if any search term exist in the information source). This process is known as relevance judgment, relevance feedback and is used in almost all search engines techniques to measure relevance (Croft et al., 2010b).

Hence, Croft et al. argued that efficiency of search engines nowadays can be measured by the time and space for each algorithm to produce ranking for a given query. Furthermore, interpret information seekers/users' information needs regardless to their competence level when searching for information on the Internet.

Croft et al. further extended the investigation to examine search results' effectiveness and efficiency by adopting three major performance evaluation criteria (CACM, AP and GOV2). Each criteria tests the performance of search engines with a set element. For example, CACM focused on bibliographic records containing titles and abstracts, whereas AP and GOV2 focused on full text analysis.

However, it is still difficult to measure the relevance of search results based on users' evaluation of relevance because there are no reliable techniques to collect users' criteria of effectiveness of search results (Croft et al., 2010a). Furthermore, the variety of techniques used to present effectiveness in search engines are all equal. This means that quality in search engines and their results depends on the combination of these techniques to deliver effectiveness rather than one single technique; thus, one technique cannot be better than the other technique (see discussion on RS and search engines and their techniques).

Information Retrieval (IR) research focused on improving the quality of information delivered to information seekers/users. Therefore, the first factor is to establish techniques that will accurately deliver effective search results and guarantee that they are useful. Second, is to find methods to implement the proposed techniques (Croft et al., 2010a).

Whereas, Gao et al. investigate problems of search engines technique of information presentation that directly effects the performance of search engines. Hence, explores possible method to reduce IO, and minimise information seekers/users search efforts by interpreting collaborative behaviour of information seekers/users (Gao et al., 2013a).

Hence, Gao et al. suggested a user-centric approach to organise ranking and clustering of interfaces to allow information seekers/users organise search results and enhance personalisation based on mass-collaboration of aggregated views (Gao et al., 2013a).

According to the proposed solution, Gao et al. claimed that by utilising TSE technique of maintaining user profiles the system can improve delivered ranking and clustering of search results. Hence, analyse information seekers/users queries to categories queries into rare queries and very few frequent queries. Consequently, the proposed solution can act as a complementary technique to enhance ranking and clustering of TSE techniques.

Although Gao et al. attempted to allow information seekers/user to personalise the organisation of search results to enhance the performance of search engine. They

actually utilised data from past behaviour of information seekers/users (users search history) which is avoided in the proposed solution in this research (Gao et al., 2013a).

System logs are heavily utilised to identify users' role in the process of information retrievals. The research on system logs gained the focus of researchers because they allow them to understand the nature of these systems, how these systems crawl and index information sources, how these systems interpret information seekers/users queries and how these systems deliver recommendations or search results.

Similarly, query logs occur to be a rich source of data for search engines, RSs and other IRS on the internet (Croft et al., 2010b). Therefore, many attempts in IRS research focused on extracting data on search engine, RSs and information seekers/users behaviour depending on the purpose of data collection from these logs. Most important is that these logs provide data on activities, such as the priority of browsing and the results of specific queries, session details, viewing time and location, personal profile details if available and login details on search engines or customised searches which helps in calculating relevance in search results (Mei and Church, 2008).

To measure relevance in search results, search engine calculates information seekers/users click-through on search results to improve ranking for future retrievals of the same query topic. Click-through is also commonly used to build user evaluation criteria of relevance in search results. Click prioritization of search results shows users' preferences, but it can also be considered as a random click. Therefore, a lot of research also disagrees with the claims that the click-through activity of the users' can be used to build user preference profiles and improved ranking of search results. Hence, click-through belongs to the category of biased evaluation towards highly ranked, or popular search results. This is explained by the users' tendency to click the first page results in the search results list much more than the ones in the bottom of the page. Thus, these results gain more points and higher position even if it is not relevant at all (Croft et al., 2010a).

### 3.2 RESULTS OF RECOMMENDER SYSTEMS

RSs pioneered in collecting and profiling information seekers/users interactions with the other information seekers/user and the system in domains such as e-commerce, e-marketing, online booking, e-learning and social intensive environment. Hence, almost

all RSs techniques are based on users' profiling, which are built upon the experiences of collecting various user activities in the past, and any kind of submitted information about users (Resnick and Varian, 1997 ).

However, RSs prediction techniques of information seekers/users' behaviour through interactions recorded in the past proved to be an unreliable way of modelling environments where recommender techniques play a crucial role (Ziegler, 2005). As a consequence, it is not very desirable to perform CF (Takács et al., 2009), (Rennie and Srebro, 2005) because their prediction is based on the similarities of various origins (e.g. with other users), which can be accidental and completely wrong. However, content based filtering (CBF) (Pazzani and Billsus, 2007) take into account actions performed, therefore users' involvement and its suitability should be examined when creating any type of recommender techniques (Kautz et al., 1997), (Goldberg et al., 1992).

There are various attempts to combine CF and CBF known as hybrid RSs (Lathia et al., 2009), (Jannach and Hegelich, 2009), (Zhou et al., 2010). However, a classification of recommender techniques (Resnick and Varian, 1997 ), (Schafer et al., 2007) discovers numerous options of using demographics (Burke, 2000), (Burke, 1999), ranking (Gemmell et al., 2009a), (Gemmell et al., 2009b) and matching (Tran, 2007) when creating RSs techniques. Hence, relevance in RSs recommendations are based on past information seekers/users interactions.

### 3.3 SEMANTIC WEB SOLUTIONS:

At the time this research was conducted, it was very difficult to find any existing research which utilise SWT and its stack to select information sources and consider user preferences in the process of reasoning or building SWRL enabled OWL ontologies in order to reason upon the semantics of information sources on the internet and select the most relevant source(s) according to a particular user query for a particular moment of internet searches.

Therefore, it is of high importance to not confuse the proposed solution in this research which utilise the SWT stack with numerous examples of formal ontologies associated with internet search results. They are not related to this research. They contribute towards creating vocabularies for the purpose of describing various concepts in and characteristics of knowledge of domain-specific problems. Hence, the discourse

in the paragraphs below is an overview of a selection of research that use the word “semantic” or “ontology” is present.

Some of the attempts to enhance the performance of RSs attempts to utilise ontologies. Loizou and Dasmahapatra proposed a semantic-based approach to RSs by exploiting the contextual information of items to be recommended and the recommendation process in order to overcome problems with traditional RSs and their techniques (Loizou and Dasmahapatra, 2006). Ruotsalo gives a common example of using the ontology as a vocabulary to improve the functionality of the content-based RSs. Ontologies help to understand semantic similarities between “items” and features stored in user profiles. These ontological elements are given in advance and have fixed values before performing recommendations (Ruotsalo, 2010). Likewise, Fan et, al. attempts to utilise ontologies to improve CBF. They are domain specific ontologies, which are used for improving the analysis of the content and accuracy of the filter in CBF (Fan et al., 2010).

Uschold also took into account annotation in the process of developing new technique for RSs. Thus, annotations are termed “semantic” because they have additional explanations in ontologies (Uschold, 2005). Blanco-Fernandes et al. report of Intelligent RSs claims to overcome overspecialisation in recommendations by applying reasoning techniques available in the SWT stack (Blanco-Fernandez et al., 2008). Ziegler, suggests deploying RSs into Semantic Web and devise SWT RSs to perform recommendation computations locally for one given user (Ziegler, 2005).

Chang and Quiroga use ontologies based on Wikipedia’s content as a shared platform to model web pages and cross system recommendations (Chang and Quiroga, 2009). Nagypál investigates ways that participates in the effectiveness of IRS and claims that it can be improved by using domain knowledge stored in ontologies as suggested. However, they propose a framework based on an ontology-supported semantic metadata generation and an ontology query expansion, which allows the integration of results from traditional full-text engines in document retrievals (Nagypál, 2005).

Whereas, Chamiel and Pagnucco user preferences are elicited through the knowledge they describe in an ontology (which is based on the expert information and knowledge from social web resources) (Chamiel and Pagnucco, 2009). Araujo-Fontes et, al. claimed that they use ontologies to empower software agents in order to annotate various concepts

using ontological inference (Araujo-Fontes et al., 2013). Finally, if the SW ideas is used a across the web then it should allow to use them in various domains such as entertainment and business (Cantador et al., 2008), (Passant and Yves, 2008), (Costa et al., 2007), in order to manage internet sources as SW concepts and recommend them according to the knowledge stored in these concepts. Therefore, SW ideas might work very well for refining the recommendations of internet sources.

Another set of research which attempted to employ SWT can be found in the domain of search engines. Grčar et al. utilised SWT stack to interpret the semantics of information seekers/user in the process of browsing webpages to view topics of interest. Hence, Grčar et al. suggested an approach that utilised SWT to model users' profiles and create a plugin that can be installed on browsers (i.e. IE) to track users' activities, maintain dynamic user profile. Furthermore, the system automatically construct the collected data into topic ontology that delivers interest-focused browsing, hierarchical clustering and interpretation of users' current interest by analysing recently viewed webpages.

Therefore, Grčar et al. research contributed towards user modelling that increased users' efficiency through the delivery of personalised information. The attempt to interpret the semantics of webpages viewed by users allows the system to understand users' behaviour (Grčar et al., 2005).

Although, Grčar et al. attempt contradicted towards modelling UB, the model was based on past behaviour of information seekers/users on the internet (Grčar et al., 2005). According to the research objectives in this thesis, it is important to ignore past UB and only take into account the present not the past or the future. Hence, model "a moment" that will require current user preferences in the process of selecting information sources in internet searches.

Oufaida and Nouali investigated the excessive amount of information on the internet. Although search engines provided search results based on information seekers/users' queries. However, Oufaida and Nouali were not satisfied with retrieved information sources because search engines did not properly analyse and extract exact information seekers/users required information. Hence, the delivered search results are abstract "general and does not contain any accurate contents". Consequently, search engines delivered similar search results to all users regardless to details in

information seeker/user query. This means that the search results are not based on users' preferences, tastes and interests, and neither expressed by users nor profiled by RSs techniques over time.

Therefore, Oufaida and Nouali proposed a hybrid multi-view recommendation approach that integrates CF techniques from the well-known domain of RSs and adds SWT to interpret social data to deliver re-ranked search results to information seekers/users as follows:

- Users' Representation: builds users' profile through the collaborative view for either explicit or implicit ratings and store demographic data by employing socio-demographic view and the semantic view in a hierarchical items' classification.
- Neighbourhoods Generation: each user will be affiliated with a group of users to generate recommendations based on the characteristics of the group. Hence, generate three distinctive recommendations based on collaborative, social and semantic neighbourhood.

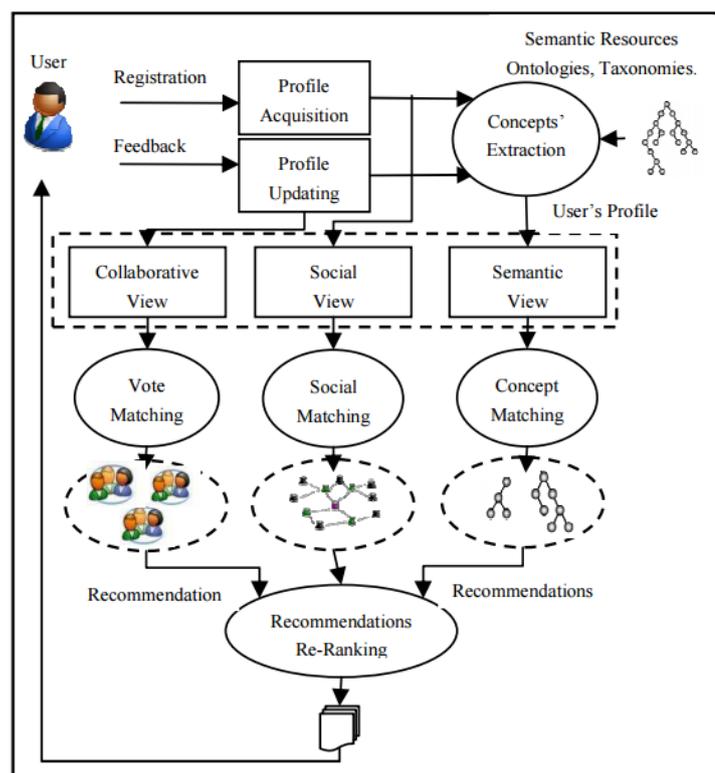


FIGURE 3.1 MULTI VIEW RECOMMENDATION ENGINE PROCESS (OUFAIDA AND NOUALI, 2009)

The system automatically generates users' profiles for each type of recommendation technique. Therefore, it is evident that the system is enriched with three different sources of data about the user (Oufaida and Nouali, 2009).

Hence, in relation to information seeker/user role in the proposed generic model in this research it is of interest to investigate recommendations based on the semantic view technique in the work of Oufaida and Nouali. Although this work exploited similarity between users to identify and group them based on their profiles. Furthermore, utilised semantic neighbourhood technique that seeks for users with similar interests and build a hierarchical concepts organization. Consequently, delivered personalised recommendations to information seekers/users on the system. Oufaida and Nouali based recommendations on information seekers/users' past interest because they employed collaborative view which collect information seekers/users' ratings from past views to deliver new recommendation (Oufaida and Nouali, 2009). Hence, it contradicts the way information seekers/users are modelled in the proposed method in this research.

More relevant research attempts to the proposed solution in this research and which utilise SWT is present in Shojanoori et al. research which investigates possible methods to reason upon the semantics of Pervasive Computing Environment (PCE) to deliver a situation specific services to users in PCE.

Hence, Shojanoori et al. utilise existing technologies (devices) in order to model the semantics of a PCE of certain problem domains (specifically selfcare home in healthcare) to deliver synergy between users and services in PCE. By capturing the semantics of both the environment and users in a PCE situation, the proposed Formal Computational Model (FCM) will then reason upon the semantics of the situation to infer new knowledge to the user. Moreover, by allowing information seekers/users to provide the system with relevant information, the extracted semantics can ease the reasoning process and help in decision making for certain situation information seeker/user of PCE require.

Furthermore, this research employs SWT stack particularly OWL ontologies to create concepts, constraints and instances of PEC and certain situations. And reason upon the semantics of the PCE by writing SWRL rule to infer new knowledge (Shojanoori et al., 2012a) (Shojanoori and Juric, 2013).

Consequently, Shojanoori et al. construct a PCE that involves both the semantics of a PCE and changing situations based on the collected user semantics. Similarly, the proposed solution in this research investigates information sources on the internet and of information seekers/users to construct an OWL model that will defines the characteristics of the environment and semantics of information seekers/users' preferences. Consequently, the proposed model will then be able to reason upon the collected semantics in information seeking situations on the internet to deliver the most relevant search results. hence, this can minimise modern IO.

### 3.4 CHAPTER SUMMARY

This chapter provided the reader with an overview of some of the previously proposed approaches that have some traces in relation to main aspects of the proposed generic computational model in this research. Therefore, his chapter itemised related work according to the research objectives. The first section of this chapter discussed previous work that looked at the issue of excessive amount of information on the internet and proposed approaches to reduce IO. It also highlighted approaches present in the domain of RSs, search engines and their techniques to provide relevant recommendation and search results. Furthermore, some approaches that aimed to personalise recommendation in social intensive environment and search results to reduce the retrieved amount of information sources on the internet were also discussed. Discussion on ranking techniques and research that aimed to provide relevant search results based on recency, time, location, situation and so forth was provided. Thus, the investigation of related work in this chapter helped to conclude that, most of the proposed solutions to reduce IO were based on techniques that depended on building profiles of both the environment and user's past behaviour on social intensive environments. Furthermore, these solutions also resorted to combine technologies and techniques from the domain of RSs and search engines; this with the aim to improve recommendation and search results.

## CHAPTER 4. THE PROPOSAL

### 4.1 ESTABLISHION WHAT THE PROBLEMS ARE

The discourse in the previous sections have indicated that both information seekers/users and information/service providers might be far from diminishing IO and securing results of internet searches which satisfy information seekers/users preferences to deliver “a moment” of internet searches (Mahoney et al., 2009), (Wang et al., 2009), (Addis et al., 2010), (Girard and Allison, 2009). IO has changed in the last decade due to the way information are created and consumed. The amount of both information seekers/users and information provider create on a daily basis will constantly grow and IO is not something that can be controlled (Pollar, 2004), (Westmead, 2013). Today information seekers/users:

- primarily want to “grab” information from the internet at the moment information is needed, and
- tend to retrieve the content of repositories available on the internet, which are not necessarily structured.

The impact of technologies, which enabled information seekers/users to become producers and consumers of information sources, is imminent. Therefore, this research proposes to initiate a shift in thinking on “*how to create modern and more effective ways of searching*” when trying to find solutions for managing modern IO.

Some critics would say that the division of the literature review in this research into RSs, interpretation of tagging, annotations, folksonomies into RSs, search engines and their techniques, ranking of search results, and solutions which takes into account UB on the internet might not be the best way of underpinning new ideas for addressing modern IO. However, the discourse in this research followed the exact chronological order of events that happened across fields of IRS, RSs and search engines in the last 20 years.

Information seekers/users are aware that an abundance of information available around them is accessible through search engines. If so, then both: the amount of information and the results of internet searches are the main reasons for having modern IO. The collected and interpreted semantics of the environment where internet searches happen could help to be more precise and deliver relevant or at least the best possible

search result to information seekers/users if techniques found in RSs and search engines are carefully investigated. Hence, few major obstacles in using them to improve the results of internet searches and reduce modern IO:

They are all focused on building profiles of information seekers/users and items, that may be recommended to other users, based on information seeker/user PAST behaviour and item's ranking. RSs sometimes use CF and CBF algorithms, which measure similarities between the item's rankings according to keywords which appear in information seeker/user's reviews of a particular item. The results of measuring these similarities are then used in building a "better" profile of the information seeker/user.

Some of the common issues in RSs technique for example can be:

- CF required a large data sets in order to make useful recommendations,
- The common problem of CF is the persistent cold start,
- The lack of basic information on new items on the system, and
- The inability of CF to accommodate with information seeker/user changing behaviour (Tran, 2007).

However efficient these ideas may have been in early days of e-commerce by telling information seeker/user "*this is what you might like*", they are very unreliable now if used to make recommendations when information seekers/users search the internet. Relying exclusively on the past behaviour, when modelling information seekers/users' profile and basing recommendations on that profile is risky. Information seekers/users often change their mind and interests, sometimes instantly when searching the internet and their profiles might have been built wrongly.

The reason is trivial, but powerful: buying a set of books on Amazon on Spiritual Healing for friends and families, and viewing them before buying do not make the buyer a spiritual healer, it should not be interpreted as the buyer's "interest" and should not be built into the user's profile because it is wrong. Furthermore, one attempt of viewing certain information, items or service on the internet to gather information does not mean that they are in the focus of information seekers/users' interest. It can happen accidentally!

Hence, it is difficult to predict exactly, through any of the available Artificial Intelligence (AI) or RSs techniques/search engines, based on the knowledge collected from the information seekers/users' past behaviour, how they would react from

situation to situation when generating and retrieving data at the same time and manipulating information on the internet.

In modern information age, information seekers/users are in charge of computational environments (Shojanoori, 2013), (Shojanoori et al., 2012b). They very often manage the data involved in such computations because they are producers and consumers of information. Consequently, attempts which involves information seekers/users in the process of information discovery are very welcomed.

Tagging, annotations and folksonomies are a huge step forward towards managing excessive information in modern information age. They have allowed information seekers/users to be involved in the classification of data they generate and consequently secure more relevant results of searches in terms of satisfying information seeker/user's preferences in the process of retrieving information on the internet. If information seekers/users are given the power to classify information they create, then any type of retrieval is expected to give more "relevant" search results to information seeker/user. However, there are a few issues with tagging/annotations/ folksonomies:

- a) They have been introduced as a consequence of extensive involvement of information seekers/users in online SNs and the proliferation of Web 2.0 applications that required tagging technique. In other words, tagging was not explicitly introduced for addressing the deficiency of RSs techniques and for reducing modern IO. From that respect, they should not be assumed that they can be efficient in solving the problem of modern IO. however, they maybe a method of information discovery in certain situations.
- b) Tagging/social tagging/annotations/folksonomies have become very complicated because they address relationships between items, users and tags. It is difficult to think that it is possible to manage the complexity of such relationships through folksonomies. They do not have enough "space" for describing fully the semantics of relationships in such situations which may make them too complicated. Modelling relationships in CS has always been sensitive, if not a problematic issue, particularly if we do not reserve a special modelling element where the semantics of relationships can be "stored" in folksonomies.

IO is also the consequence of search engines and their techniques. hence, search engines might not always fulfil information seekers/users information needs, because they might be confronted with either irrelevant search result or excessive amount of them (Mahoney et al., 2009), (Wang et al., 2009), (Addis et al., 2010), (Girard and Allison,

2009). Search engines heavily rely on ranking techniques. Hence, rankings of search results are present in all search engines today and there is no doubt that they may reduce modern IO by using various criteria (recency ranking, time, location and so forth).

However, there is one important problem there. Ranking algorithms have been, and will remain business secrets of companies which are in charge of search engines. Hence, researchers can only “guess” how Google ranking works (Google, 2012). Consequently, many companies today claim that they will make users web presence “Google-ranking-friendly”, i.e. users URL “might become highly ranked”. In other words, it appears that users’ URL might never be retrieved if the Google ranking algorithms does not “recognise certain user”, which will then affect the relevance of information retrieved by the Google engine.

Although internet search engines delivered to information seekers/users relevant search results, it was still dependent on 15 year old techniques, such as syntactic matching of keywords, or tagging of services on the internet which requires manual efforts. Consequently, internet retrievals through Google, without understanding exactly how its ranking works, will not reduce modern IO. It may even add to it. How could a ranking system in widely used search engine(s) fit the expectation of all information seekers/users? No wonder information seekers/users suffer from smog (Shenk, 1998) and glut (Bawden et al., 1999) in results of retrievals through modern search engines. Hence, its doubtful that ranking/ratings of internet search results are suitable methods of solving the problem of “relevance” when delivering “a moment” of internet searches.

On the other hand, information seekers/users on the internet heavily participated in increasing modern IO. Olston and Chi complained that information seekers/users are usually not certain of their information needs and mostly initiate internet searches with a vague notion of the type of information they seek or wish to retrieve. Hence, they set criteria and while browsing the results they often change their initial criteria as new information starts to surface (retrieved from initial search query) (Olston and Chi, 2003).

Furthermore, Olston and Chi argued that keywords matching techniques do not always deliver relevant search results, thus browsing in these situations become the focus of information seekers/users in the process of information discovery. Moreover, information seekers/users may not be aware of the correct terminology of things on the

internet. Consequently, search results are abstract and requires information seekers/users effort to discover relevant information to search query (Olston and Chi, 2003).

Whether researchers are still happy with the existing *keyword matching technique* (Chang et al., 2001) in search engines (Chowdhury and Soboroff, 2002), (Baeza-Yates, 2006), (Ian et al., 2007), or use searches which are labelled as “semantic” (Tumer et al., 2009) , (Sudeepthi et al., 2012), (Hendler, 2010), (Sheth, 2011) they still do not have answers to and solutions for (i).-(iii) below.

The discourse in this research (see chapter 2) addressed a portion of the IO problem because it is difficult, if not impossible to solve it completely (Pollar, 2004), (Allan, 1997). There is no doubt that they alleviate it. Attempts to address IO in the past are not applicable in current situations due to the nature of information available on the internet. More important is that information seekers/users should be in charge of the management of internet searches and their results, this by giving them opportunities to:

- (i). tailor the organization of search results, as elaborated in. This means that each search query must be processed separately according to the environment and situation where they belong (Gao et al., 2013a).
- (ii). personalize search processes and the use of search engines, as indicated in and avoid generalisation in the process of information discovery. This can be explained as different information seeker/user has a different background and competency level (Wen et al., 2009)
- (iii). influence the way we construct or choose mechanisms/algorithms which deliver search results. hence, interpret the semantics of both information sources and information seeker/users on the internet.

The purpose of this research is not to challenge current search engines or recommendation techniques used in social intensive environment. Furthermore, the proposed solution in this model do not aim to replace existing techniques due to:

- Both search engines and RSs utilise a wide range of models to collect and analyse information on the internet.
- Both search engines and RSs also employ variety of techniques which attempted to deliver results and recommendations as mentioned above.

However, bullet (iii). above states that it is important to rethink methods that influence the way search engines work. This research argues that information retrievals on the internet should allow flexibility to information seekers/users in the process of

evaluating the relevance and accuracy of ranked results and not take for granted the way Google search engine ranks its results without allowing them to decide about the purpose and wanted level of accuracy of ranking? Google may claim that the purpose of Google ranking may improve relevance of search results, but does it really do the job as the information seeker/user expects? Does it really address modern IO?

Hence, this research does not intent to say that more than two decades of mastering algorithms for creating perfect search results did not produce good solutions. The intention is to raise awareness that a shift in thinking on “*how to create modern and more effective way of searching*” is needed if we wish to carry on using search engines and address modern IO efficiently.

It is important to highlight that the investigation of RSs and search engines in this research itemised some of main concerns that domain of IRS on the internet. furthermore, this research does not cover all RSs and search engine problems. Hence, the proposed solution in this research is concerned with modern IO caused by modern technologies on the internet. Thus, the investigation on RSs and search engines and their techniques aimed to:

1. Understand causes of modern IO which required to investigate different types of IO as discussed in (section 2.2).
2. Extract data on characteristics of information sources on the internet
3. Investigate possible ways to address information seekers/users’ preferences in the process of information discovery on the internet.
4. The emerging SWT offered an alternative solution to minimise modern IO by allowing software engineers to interpret the semantics of the different environment on the internet.

Consequently, to deliver “a moment” of information retrievals on the internet that can address modern IO it is important to:

- 1) Take into account that the nature of IO and the power of traditional IRS has CHANGED. Hence, abandon techniques and “systems” that claimed to address IO in the past. For example, it would be inappropriate to claim that RSs techniques, developed in the 90s, for managing a surplus of information in structured repositories, would work in modern retrievals from the internet. It does not mean that these techniques are not applicable at all. It means that they need investigation to find out if they can address modern IO and contribute towards interpreting new type of information when managing it. Furthermore, modern IO has become closely

associated with internet searches and results from search engines very often overload information seekers/users.

- 2) Focus on interpreting the semantics of the environments where information seekers/users experience modern IO and consequently model both:
  - (i). user's preferences when creating or using information and
  - (ii). characteristics of the environments where IO happens, i.e. characteristics of information sources on the internet by identifying features, services, purpose etc.

The proposal in this research aims to offer slightly different way of thinking in terms of how to address the issue of irrelevance of search results and their correlation to modern IO. Thus, address differently the way to

- (A) interpret information seekers/users search queries when either creating search engines or tailoring (filtering / ranking) search results (user queries might not be sufficient!) in domain specific situations on the internet as described in (i) and (ii) above and
- (B) construct computational model(s) for supporting a) above and enriching current search engines with a complementary technique which will utilise SWT stack.

By addressing (A) and (B) above the proposed solution in this research can definitively address modern IO from a new perspective. The argument is not to eliminate IO. The amount of data information seekers/users create on daily basis will constantly grow and IO is not something that both information seekers/users or information/service provider can or wish to control (Pollar, 2004), (Westmead, 2013). However, it is of high importance to improve and change the way relevance is addressed in the process of delivering search results.

For understanding the proposals of this research it is important to note that:

- (I) The proposed solution in this research utilises SWT and stack (is SWRL enabled OWL ontologies specific) (see OB4 from chapter 1, section 1.3). Consequently, the vocabulary in this proposal is restricted to SWRL/OWL vocabulary.
- (II) The core of the proposed model is expected to accommodate a reasoning process as suggested in (OB1 and OB4 from chapter 1, section 1.3) similar to the reasoning process implemented in (Shojanoori et al., 2012a), (Shojanoori and Juric, 2013) and discussed in chapter 3 (related work) in order to understand how a reasoning process becomes a part of any computational model.
- (III) The proposed computational model in this research bears NO resemblance with formal ontologies, which constructs knowledge-basis of certain

problem domains and AI algorithms which may build them. Hence, the proposed solution in this research cannot be confused with them. Readers who has some knowledge on research which utilise the SWT stack can recognise the proposed solution in this research as a Software Engineering (SE) solution based on reasoning where its results are not made persistent. These SE solutions may be re-usable only when they play a role in the management of the semantics of a particular “moment” where internet searches happen (see OB 2 and OB 3 from chapter 1, section 1.3).

- (IV) (OB 1) stated that the proposed model could be placed as a re-ranking (or refining) mechanism, which run on top of search engine results or incorporated in the current internet search engines as a complementary technique to enhance search results with relevance. Therefore, the model should be flexible enough to fit both requirements. However, it is not realistic to expect that current search engines would welcome any change in their algorithms without marketing this proposal. Therefore, the proposed computational model will reason upon Google search engine results and perform selection of information sources on the internet which were given by the search engine.

Consequently, the terminology from the previous 3 chapters would now change as follows: coming sections will discuss *selection of information sources on the internet*, that will produce relevant results of internet searches triggered by information seeker/user queries to reduce modern IO.

## 4.2 THE PROPOSED COMPUTATIONAL MODEL

The proposed computational model in this research will address modern IO by interpreting the stored semantics of the environment and information seekers/users’ preferences (presented in this research as domain specific problem situation) as described in (a)-(c). below. This to guide information seekers/users in the process of selecting most relevant information sources according to the extracted information from information seekers/users query in the reasoning process. consequently, the proposed solution attempts reduce retrieval of excessive amount of information sources to minimise modern IO. Hence, deliver “a moment” of internet searches that only takes into account how information seekers/user described preferred information source in certain situation and change instantly as soon as information seekers/users change their preferences.

### **The proposed computational model main components**

The proposed computational model consists of three distinctive parts which are itemised in (a). – (c). below. Because of its nature, i.e. computations which are based on reasoning upon SWRL enabled OWL ontologies, it is important to address all aspects of creating OWL model and securing reasoning upon its concepts through SWRL rules.

- (a). Firstly, construct a detailed OWL ontological model with its classes, subclasses and relationships which will participate in the model. Hence, they should be generic, but their illustration through domain specific classes and subclasses should show the nature of computations performed according to the abstract model.
- (b). Secondly, create an abstract model of the reasoning process based on SWRL enabled OWL ontologies. The model must show which OWL classes should be involved in the reasoning, and how the semantic matching between them, based on their semantic overlapping, will be performed through reasoning. In other words, without semantic matching the model might not be able to secure reasoning with SWRL. Therefore, the model must show where the inference is. The model must also be generic, i.e. it should be suitable for any environment where internet searches are required.
- (c). Thirdly, a set of SWRL rules which creates inferences upon OWL concepts and would work in any domains specific situation when selecting information sources on the internet. It is difficult to predict how far we can go in creating SWRL rules, because the proposed abstract model should work in any environments (SWRL rules cannot be domain specific) and should be reusable (SWRL rules should avoid hard coding). The goal is to achieve reusability of computational solution and its applicability in *different situations and domains* where selection of information sources happens.

It is important to note that it is difficult to illustrate certain aspects of the abstract computational model from (a)-(c) above without occasionally becoming domain specific.

Another important aspect of the proposal is its role in creating computations based on SWRL enabled OWL ontologies. The model should show how the semantics of the environment, where the retrieval of information sources is being performed, is stored within it, and how the OWL ontological model, with its hierarchies and constraints, will be created. Hence, it is important to know which classes and their horizontal hierarchies are important and what will be modelled as OWL constraints.

Finally, the reasoning within OWL models creates inference: OWL individuals moved (copied) across the classes of the OWL ontological model, which is based on

*semantic overlapping*. Therefore, the model must emphasize where the overlapping semantic is and how it affects the results of reasoning. Ultimately, the selection of information sources depends on exactly that: how successfully the *overlapping semantics* are modelled.

Hence, it is important to note a subtle difference between (a), (b), and (c). The understanding of the essence of the proposed computational model should result from a and b. However, its real power, in terms of reusability could be addressed only through (c) which will require domain specific example and cannot be a part of the abstract model.

### **Information collection and extraction to construct OWL ontologies:**

Information sources on the internet can be anything and everything, it ranges from being information about human, animal, nature, science, technology, and news of recent events to information about different kinds of e-services, businesses, companies, schools, universities, institutes, research buddies and so forth. Moreover, these sources can be further categorised into two distinctive types of information sources per an environment or problem domain.

The first type of information sources are informative sources also known as read only sources on the internet; are sources which deliver to information seekers/users facts (i.e. news websites, electronic libraries, encyclopaedia and Wikipedia, government websites and so forth). Informative type of Information sources on the internet can therefore be edited and updated by information owners only.

The second type of information sources are social intensive information sources which are created as consequences of technologies available on the internet also known as read and write information sources. These technologies empowered Information seekers/Users who were only allowed to read information to become creators of these information sources. Therefore, information seekers/users have a dual role in modern computing.

This means that information is generated by utilising variety of technology means available on the internet -belongs to web 2.0 tools and technologies- that delivered collaborative environment of information sources to the information seekers/users on the internet. Consequently, the amount of information created in the past two decades exceeds information seekers/users and service providers on the internet to control.

Furthermore, web 2.0 tools and technologies can also be a standalone tools or multiple tools that allow synergy and production of new/enhanced information sources on the internet. Furthermore, each of these tools allow creation and representation of information differently. Therefore, characteristics of information sources differs according to how they are created (i.e structured data retrieved from DB, or unstructured data created by using social media tools or developer tools on the internet). Consequently, web 2.0 tools enabled the transformation of physical information into the internet. Thus, information seekers/users are now able to view any information effortlessly at any time.

Moreover, wireless and ad-hoc technologies allowed to access and create information on the internet anywhere and everywhere. Consequently, the overwhelming availability of information sources on the internet resulted into modern IO. Therefore, researchers attempt to develop techniques that will allow to improve information creation, management, presentation and retrieval on the internet. Some of these solutions were perfect for a certain IO problems at a certain time for a certain group of users. However, IO is a resilient problem which gets out of control every now and then.

Therefore, the proposal in this research is an attempt to alleviate IO and manage information sources on the internet. This means that by interpreting the semantics of both information seekers/users and information sources on the internet modern IO can be slightly controlled. Hence, the proposed generic computational model in this research can contribute towards refining search results to reduce modern IO.

The proposed generic model can work in parallel with currently existing tools. It is important to emphasise that our initial aim in this research is not to create any type of RSs or replace currently existing search engines and working techniques of information retrieval methods on the internet.

The first step in this research is modeling information sources on the internet and accessing them based on characteristics of both the environment they exist in and the preferences of information seekers/users of environment. For this purpose, this research investigates and analyse available information sources on the internet such as website, blogs, forums, electronic documents, video clips, audio files and their nature. The investigation will also include understanding the preference of information seekers/users, investigate the role of owners and consumers of these information sources on the internet

and last try to draw the relationship between information sources and users on the internet.

In a chronological order Binghubash and Juric (Binghubash and Juric, 2011), attempted to analyse social networks because they are social intensive environment on the internet and rich with information that can be accessed by information seekers/users. Hence, they allow information seekers/user to communicate, create and distribute information instantly. The investigation also attempted to understand

- which exact service these collaborations may bring to information seekers/users,
- which level of security and privacy is guaranteed to SNs members and
- which technical support members may need when collaborating through SNs.

In order to answer bullets above, it was important to classify SNs. Consequently, the first attempt was to analyse studies of SNs available on websites which provided evaluation of existing SN sites on the internet. The second step was to further extend the investigation to include information about “demographics, profile, security, networking features, search and technical help/support”. This helped us to create a basic concept of characteristic of SN sites which is one type of information sources on the internet.

In Almarri et.al, 2012. (Almarri et al., 2012a) focused on understanding the significance of LLL society on the internet. The reason behind this choice was because LLL is an environment rich of information where information seekers/users can choose to learn at any stage of their lives and anywhere or anytime. Information seekers/users are from different competence level and background. Furthermore, the wide spread of advanced communication technologies such as information sharing and dissemination through peer to peer meeting tools created new type of how information seekers/users want to share and create information in LLL environment.

Therefore, this investigation explored available communication methods to characterise functionalities/services offered by information sources for information seeker/user in LLL environment on the internet. Furthermore, this investigation specifically focused on LLL in healthcare environment because lifelong learners in healthcare requires to maintain their level of knowledge and heavily dependent on the process of LLL in their career.

The investigation on information sources on the internet also included higher education. Juric et. al, 2013 (Juric et al., 2013) was concerned with the impact of advanced technology in formal educational environment and ways to re-model formal learning practices and delivery and dissemination of knowledge. Hence, this added a new dimension to the characteristics of information sources on the internet. Thus, investigated available tools on the internet which will allow easy communication with students to share knowledge, virtually when creating interdisciplinary modules.

The motivation behind the investigation above was to model the characteristics of different types of information sources on the internet. Consequently, the results indicated that most common characteristics of information sources on the internet can be as listed below. Every internet source has these distinctive characteristics:

1. Features, services, purpose of existence, and type of information it provides.
2. Every information source must be either informative or collaborative source.
3. Every information source must have an owner and consumer.
4. Every information sources can be singular ie. One technology or tool which serves a certain cause and has identified user, or multiple tools and technologies which aims at either to improve the quality of an existing information source or create a new one.

Description of Modern Information Sources: is depicted in figure 4.1 below depicts a portion of things which can be considered as an information sources on the internet. As mentioned above an information source on the internet can be anything in figure 4.1 below. Furthermore, these information sources can exit either separately or a member of an extended information source based on 1-4 above and also include characteristics of information seekers/users on the internet. as described in 1-4 above information sources can share common characteristics according to the environment or domain they belong to. Furthermore, categorisation and grouping of characteristics of information sources also depends on personal views of the person who conducts the investigation.

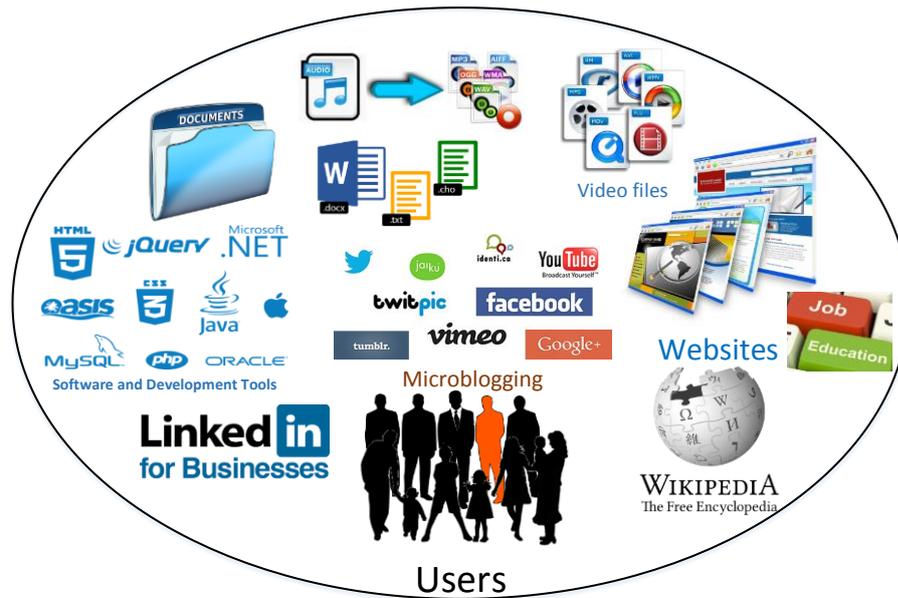


FIGURE 4.1 INFORMATION SOURCES ON THE INTERNET

Internet source(s) exist for any reason, information seekers/users create contents on the internet to (A) share personal opinions and interests or to (B) promote businesses and services through different means on the internet, (C) provide learning material and so forth. Thus, to be able to deliver relevant information source(s) to information seekers/users, it is important to understand the purpose they exist for in the first place. Hence, investigating the environment the information source(s) belongs to, features and services provided through a specific source(s), technology in use, and users of a certain information source(s) is essential. Otherwise, both information source(s) provider and information seekers/users are in uncertain environment of source(s) where everything is mixed up. Consequently, it is of high priority to properly define the characteristics of these information source(s) to be able to retrieve them.

The next three subsections of 3.4 describe various parts of the proposed computational model. Subsection 4.3.1 describes the OWL model as specified in (b) (see section 4.1) and define the way semantic of the environment are stored. In section 4.3.2 covers only *relationships* in the proposed computational model, as specified in (b) (see section 4.1) by defining OWL constraints. Section 4.3.3. describes the reasoning process as required in (a) and explain how semantic overlapping can be created. Whereas, (c) above, is part of the illustration of the proposed computational model and therefore c. can be addressed only **when** building particular *situation* and *domains* which require the selection of information sources.

### 4.3 GENERIC MODEL- OVERVIEW OF THE PROPOSED METHOD

The previous section described information extraction and analysis techniques followed in order to construct OWL ontological model. Therefore, this section elaborates and discuss how the environment is modelled. This section divides the discussion into two parts. First, present the abstract concepts of information sources environment and describe the logic behind the categorisation of the semantics of information sources. Second, define domain concepts into description logics language (DL).

1. Hence, to define an information source(s) it is important to define:
2. Basic concepts of the general environment of any information source(s) on the internet.
3. Relation or association between concepts and instance(s) of sources on the internet.
4. The Reasoning Process which delivers “a moment” of internet Searches.

Therefore, the first step is to understand the general concepts of any information source(s) on the internet. As mentioned above in (see section 4.2), there are different types of information source(s) on the internet. The universal environment of information source(s) is massive, which means it is almost impossible to cover all information source(s) to find the relevant information source in certain situation of information discovery.

Therefore, this research attempts to design the general concepts of information sources as “universal environment” to be able to recommend/ deliver relevant information sources to information seekers/ users search queries on the internet. Hence, the first step is to group information source(s) on the internet, figure 4.2 below illustrates the universal environment on the internet.

#### **Basic Concepts of any Information Source(s) on the Internet:**

The universal concept of information sources is a set defined as a collection of instance “**elements**” that if brought together can obey to a certain rule. These instances can be anything that will allow the description of an information source on the internet. Thus, by fulfilling a certain rule(s), these instances can belong to a grouped collection of a subset(s) of the universal set that will share certain features.

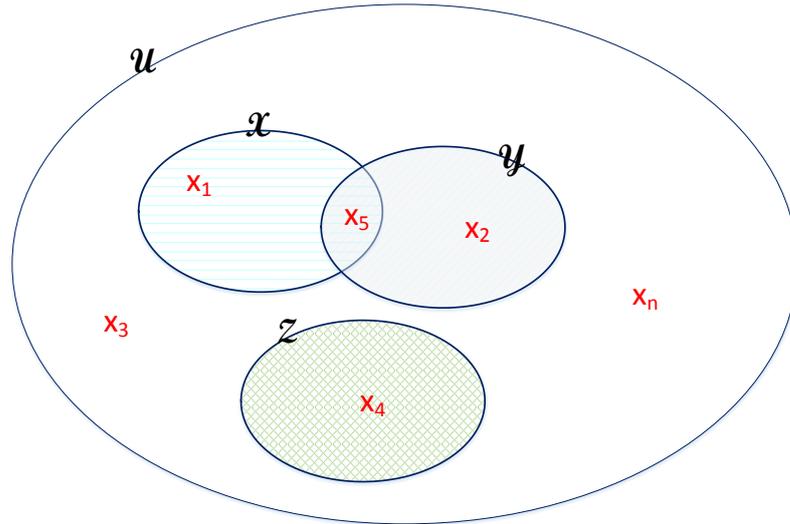


FIGURE 4.2 THE UNIVERSAL SET OF INFORMATION SOURCES ON THE INTERNET.

Furthermore, universal concept ( $\mathcal{U}$ ) can be a set which consist of collection of instances. These instances can have some common features or Not. It can also be a collection of random features that have nothing in common as well.

Thus, the universal concept can be expressed as follows:

$$\text{If } \mathcal{U} = \{x_i \in \mathcal{U} | i = 1, 2, 3, \dots n, n = (n - 1)\}$$

For example, in figure 4.2 the universal set  $\mathcal{U}$  can have an “ $n$ ” number of subsets which shares common features. Thus if  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  are sets, then  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  are called subsets of  $\mathcal{U}$ , if and only if every element of  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  are also an element of  $\mathcal{U}$ . Consequently, this also applies that all sets  $\mathcal{Y}$  and  $\mathcal{Z}$  that are subsets of  $\mathcal{U}$

$$\mathcal{X} \subseteq \mathcal{U} \text{ means that for all elements } x_n, \text{ if } x \in \mathcal{X} \Rightarrow x_n \in \mathcal{U}$$

$$\therefore \mathcal{X} \subseteq \mathcal{U} \Leftrightarrow \forall x_n, \text{ if } x \in \mathcal{X} \text{ then } x \in \mathcal{U}$$

For subsets of  $\mathcal{U}$  to be equivalent each element of any set must be an element of all other sets of  $\mathcal{U}$ . In the proposed generic computational model in this research all subsets of  $\mathcal{U}$  are disjoint. This means that elements of each set cannot be an element of any other set and each element has s distinctive features and characteristics. Thus,

$$\therefore \mathcal{X}, \mathcal{Y} \text{ and } \mathcal{Z} \in \mathcal{U} \Leftrightarrow \mathcal{X} \neq \mathcal{Y} \neq \mathcal{Z}$$

Subsets of the universal set  $\mathcal{U}$  forms the domain of all information sources on the internet which also describes information seekers/users preferences that are then

interpreted to deliver “a moment” of internet searches. These sets are divided into two distinctive sets:

1. Sets which stores the characteristics of information sources on the internet (i.e. the set of all base classes which stores the semantics of an information source), and
2. Sets that store available information sources and information seekers/users’ preference which will be utilised in the reasoning process to deliver the most relevant information source on the internet.

**Relations or Association Between Concepts and Instance(s) of Information**

**Source(s) on the Internet**

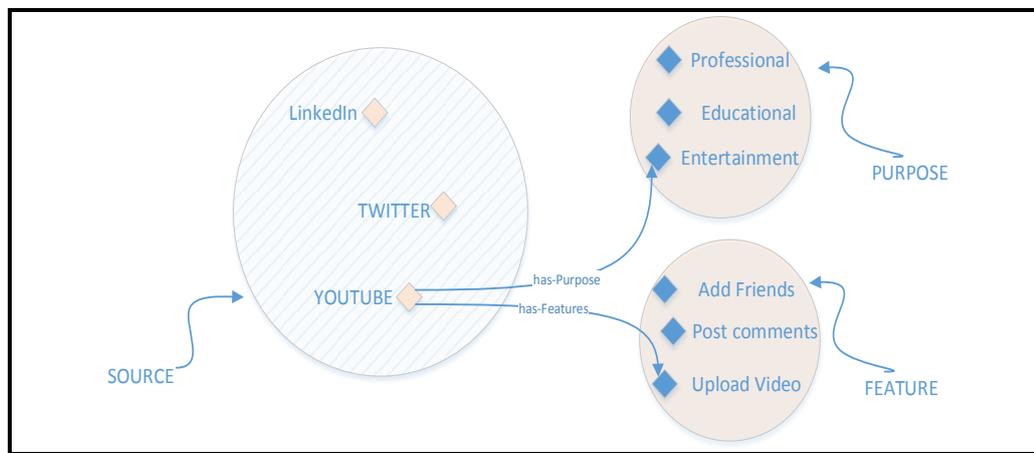


FIGURE 4.3 DESCRIBES SOME OF THE CHARACTERISTICS AN INFORMATION SOURCE CAN HAVE THROUGH HAS-PURPOSE AND HAS-FEATURES RELATIONSHIP.

To define information source(s) on the internet, it is essential to first form a relation or association between sets of information source(s) on the internet and their characteristics as depicted in figure 4.3 above. Therefore, there will be a domain concept that will contain elements “i.e. information sources on the internet” and set that describes the characteristics of information source(s) as the range set or the output of the relation between the two sets. For example, the *domain* is a set of source(s) as illustrated in figure 4.4 **Sources = {websites, blogs, forums, documents, audio, video}** and the range is a set **Purpose = {Professional, Educational, Entertainment, Personal, Sport, Regional, E-commerce, News, Games}**, the purpose of these source(s) form a relation between the information sources and its characteristics. This means that **Websites** are related to or associated with one or more **purpose** i.e. **Professional and Educational**, while **Blogs** are related to **Personal** only and **Documents** to **Educational** only as depicted in figure 4.4 below.

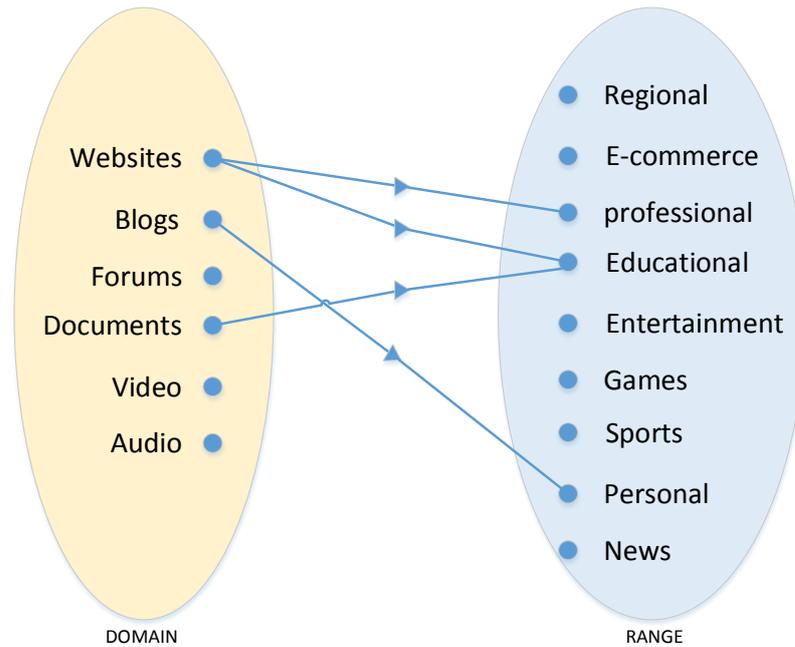


FIGURE 4.4 SHOWS THE RELATION BETWEEN DOMAIN ELEMENTS AND RANGE ELEMENTS .

Another way to express this relation is to create Cartesian Products pairs of these concepts. Thus, the outcome of this relationship creates new subset of ordered pairs drawn from all possible matches between domain elements in a relation with range elements. Therefore, we can formally present this relationship as follows:

Let  $R$  is a relation from the domain  $\mathcal{X}$  to range  $\mathcal{Y}$  (sources and purpose)

$$\mathcal{X} \xrightarrow{\mathcal{R}} \mathcal{Y}$$

$$\text{Where } R \subseteq \mathcal{X} \times \mathcal{Y} = \{(x, y) | x \in \mathcal{X}, y \in \mathcal{Y}\}$$

This therefore allows to write the subset  $R \subseteq \mathcal{X} \times \mathcal{Y}$  in ordered pairs for some elements of  $(\mathcal{X}, \mathcal{R}, \mathcal{Y})$  as follows:

$$\{(\text{Websites, Professional}), (\text{Websites, Educational}), (\text{Blogs, Personal}),$$

$$(\text{Documents, Educational})\}$$

The inverse relation of the above given is to say that any source on the internet is of a given purpose. This can be present as follows:

$$\mathcal{Y} \xrightarrow{\mathcal{R}} \mathcal{X}$$

$$R^{-1} = \{(y, x) | (x, y) \in \mathcal{R}\}$$

It is important to note that the elements of both the domain and range can change per to the environment or problem domain (situation of information retrieval) and information seekers/users preferences. This means that some elements of the Domain can be an element of the Range if and only if it is part of the characteristics of the domain element and vice versa. Therefore, characteristics of source(s) can conditionally become an information source itself (i.e. some websites allow the users to link blogs, twitter pages, and other social media based pages in profiles).

### **Difference between a relation and a function:**

Based on the relationship described in the previous paragraph, a relation can be formed between two sets, and is a collection of ordered pairs containing one element from each set. If the element  $\mathbf{x}$  is from the first set and the element  $\mathbf{y}$  is from the second set, then the elements are said to be related if the ordered pair  $(\mathbf{x}, \mathbf{y})$  is in the relation. Furthermore, a relation  $\mathcal{X} \xrightarrow{\mathcal{R}} \mathcal{Y}$  can relate the element  $\mathbf{x}$  in the first set with more than one element in the second set. This relation is known as a binary relation. Hence, the description of information sources is defined as a binary relation the proposed computational model.

Whereas, functions in relationships can map only one value from the domain set to exactly one value in the range set and implies that  $f(x) = f(y)$ . hence, the proposed computational model avoids using functional relationships to describe the characteristics of information sources because they limit the usage of each relationship (one to one relationship). Thus, it is impossible to have semantic overlapping between information sources and information seekers/user preference on the internet.

### 4.3.1 THE ONTOLOGICAL MODEL

Figure 4.5 introduces the abstract model as specified in **(b)** (see section 4.2). It must show OWL classes and relationships between them, it has to point out where the reasoning is performed and where the inference results are stored.

The ontological model consists of three major classes: SOURCES, USER-PREFERENCES and RECOMMENDED-SOURCES:

- SOURCES class contains all possible information sources on the internet which can be selected through the reasoning process.
- USER-PREFERENCES class stores the semantics of preferences information seeker/user may have when selecting information sources on the internet. These might be interpreted as various “requirements” information seekers/users may have and which, at the same time, are expected to be met by information sources on the internet if they are to be selected by information seekers/users.
- RECOMMENDED-SOURCES class stores only information sources on the internet that have been selected through the reasoning process and match information seekers/users preferences.

The individuals  $\{JND_1, JND_2, \dots, JND_a\}$  of the SOURCES class are actual URLs of information sources on the internet that are results of internet searches and possibly ranked by search engines. However, in order to improve their relevance to information seeker/user’s queries to address “a moment” of internet searches and reduce modern IO, the proposed model will have to select some of them through the reasoning process to deliver relevance based on extracted information seeker/user preference and semantics of information sources stored collected and stored in the OWL model. These are individuals which might become members of the RECOMMENDED-SOURCES class. However, each of these sources  $JND_i$ , where

$$SOURCES = \{JND \in SOURCES \mid i = 1 \leq i \leq a\}$$

must be described fully before they are selected through the reasoning process. Their description is stored in a set of ontological classes named  $\{CS_1, \dots, CS_b\}$ , shown in the left part of Figure 4.5, which actually describe the characteristics of information sources on the internet. This means that for each individual  $JND_i$  of the SOURCES class, the characteristics are stored in a set of  $\{CS_1, \dots, CS_b\}$  classes. For example, to define

characteristics of a particular information source on the internet, which is a social network, some of the specifications are:

- features, as services SN offers to their members,
- purpose of SN and the benefits its offers to their members,
- privacy policies which are available to SN members
- technical support which helps SN members to manage their presence in the social intensive online environment.

Hence, the universal set of all possible subsets which describes certain environment or a problem domain is:

$$UNIVERSAL = \{CS_i \in UNIVERSAL \mid i = 1 \leq i \leq \ell\}$$

Furthermore, an environment can have as many  $CS_i$  [ $i = 1 \leq i \leq \ell$ ] classes as needed, i.e. the number and type of characteristics of information sources can be decided according to a particular situation and domain. Also each of the  $CS_i$  classes may have a set of ‘c’ sub-hierarchies  $\{S_1 - CS_i, S_2 - CS_i \dots, S_c - CS_i\}$  as a sub-characteristic of the characteristic class  $CS_i$  for each  $JND_j$  of the SOURCES class. They may be needed in situations when the  $CS_i$  class itself is not sufficient to describe the complexity of the semantics of the characteristic of an information source.

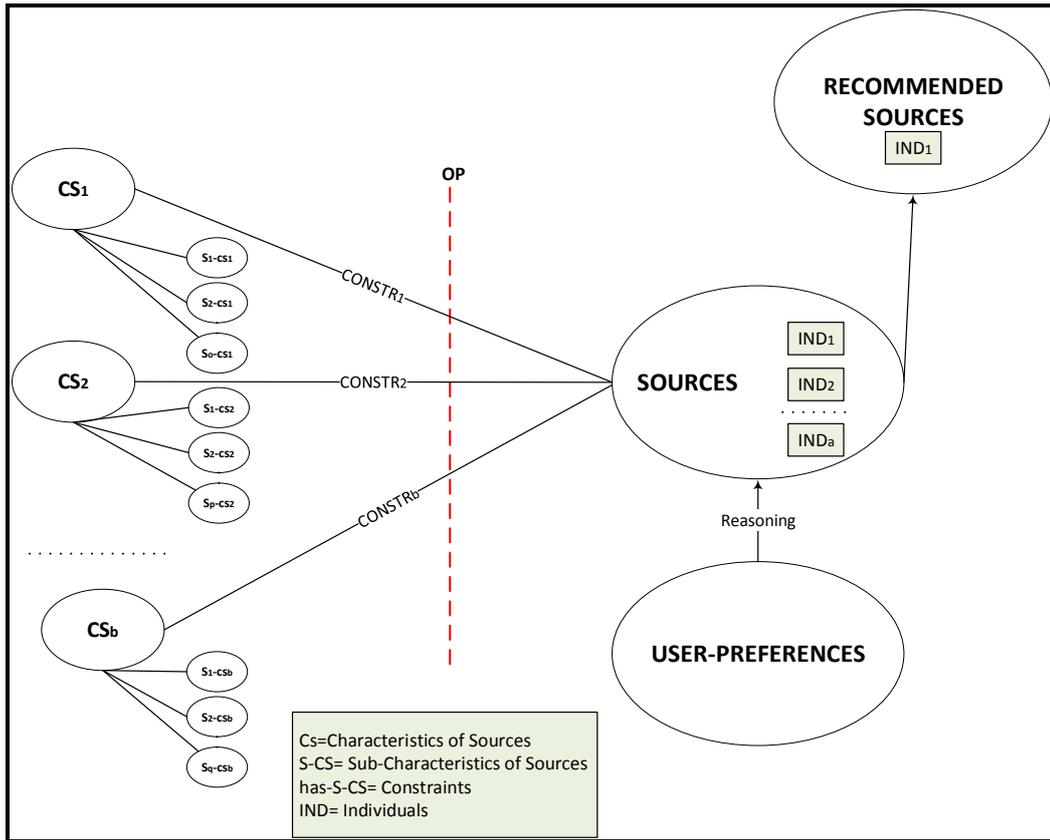


FIGURE 4.5 THE GENERIC MODEL WITH CONSTRAINTS .

In order to specify some relationships in the abstract model and define which particular characteristic  $CS_i$  is applicable to which individual  $IND_i$  from the SOURCES class, the model utilise OWL **constraints**, noted as CONSTR in Figure 4.5, which is described in the next subsection. Figure 4.5 above is different from figure 4.6 below because relationships in figure 4.5 can be either object properties or datatype properties. Whereas, relationship defined in figure 4.6 enforce the utilisation of object properties in the process of defining the characteristics of  $IND_i$  from the SOURCES class. Lines which connects the  $IND_i$  from the SOURCES class to  $CS_i$  classes are relationships which define the characteristics of information sources stored in SOURCES class.

### 4.3.2 CONSTRAINTS IN THE OWL MODEL

The relationships between concepts stored in the OWL model and are defined through **constraints** imposed on individuals of these concepts. For example, in the figure below the relationship from class SOURCES and a set of  $\{CS_1, CS_2, \dots, CS_b\}$  classes is defined by creating a set of object properties (OP). If OPs are defined between SOURCES and sets of  $\{CS_1, CS_2, \dots, CS_b\}$  classes then these OPs are to be inherited by sub-hierarchies of each characteristic, i.e. by a set of  $\{S_1 - CS_i, S_2 - CS_i \dots, S_c - CS_i\}$  classes for a particular characteristic  $CS_i$ , if sub-hierarchies  $\{S_1 - CS_i, S_2 - CS_i \dots, S_c - CS_i\}$  exist.

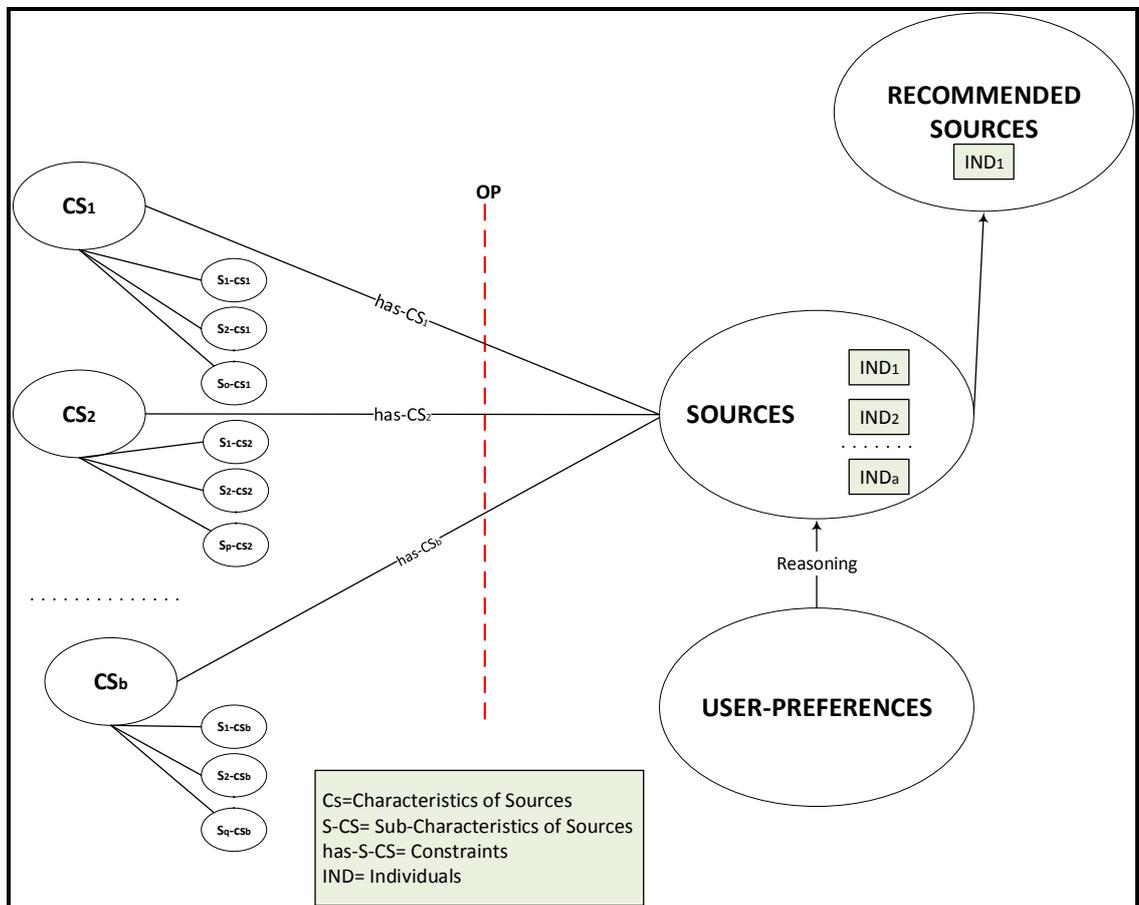


FIGURE 4.6 GENERIC MODEL WITH OBJECT PROPERTY AS CONSTRAINTS.

Figure 4.6 shows one important aspect of the proposed abstract model: if information sources are described through their characteristics, which are  $\{CS_1, CS_2, \dots, CS_b\}$  classes, then there must be a set of OPs named  $has-CS_1, has-CS_2, \dots, has-CS_i$  which are defined between the SOURCES and the classes from set  $\{CS_1,$

$\mathcal{CS}_2, \dots, \mathcal{CS}_k$ . The decision behind choosing to create  $\text{has-CS}_1$  is basically a modelling principle.

Figures 4.7, 4.8 and 4.9 are refinements of Figure 4.6 and illustrate various possibilities of defining OWL constraints when modelling the characteristics of information sources on the internet. This therefore, allows more flexibility in the process of modelling information sources domain knowledge. The difference among these three figures is that each of them allow utilising domain knowledge based on information seekers/users' preferences and how precise they are when describing preferred information sources, they seek to retrieve. Hence, according to the extracted information from information seeker/user query, a domain specific OWL ontological model is constructed and OPs are employed to perform the reasoning process.

Figure 4.7 shows that for each characteristic  $\mathcal{CS}_k$  of a particular information source, it is possible to have a set of more detailed characteristics represented through subclasses  $\{\mathcal{S}_1 - \mathcal{CS}_k, \dots, \mathcal{S}_c - \mathcal{CS}_k\}$ . Therefore, when defining OPs between SOURCES and  $\mathcal{CS}_k$  classes, it is possible to choose to either use

- i.  $\text{has-CS}_k$  property, which can be inherited by all  $\mathcal{CS}_k$  subclasses  $\{\mathcal{S}_1 - \mathcal{CS}_k, \dots, \mathcal{S}_c - \mathcal{CS}_k\}$ , or
- ii. a set of properties  $\text{has-}\mathcal{S}_1\text{-CS}_k, \text{has-}\mathcal{S}_2\text{-CS}_k, \dots, \text{has-}\mathcal{S}_c\text{-CS}_k$ , which connects each subclass of the  $\mathcal{CS}_k$  class with the SOURCES class.

For example, to describe an information source through its “features” ( $\mathcal{CS}_i$ ) then there should be the  $\text{has-features}$  property defined between the SOURCES and that  $\mathcal{CS}_i$  class (“features”) as explained in (i). above. However, if the same characteristic  $\mathcal{CS}_i$  (“features”) has been defined through subclasses  $\{\mathcal{S}_1 - \mathcal{CS}_i \text{ and } \mathcal{S}_2 - \mathcal{CS}_i\}$  which stores information such as “COMMUNICATION\_METHOD” and “MEMBERS\_PROFILE\_FEATURES”, then there should exist  $\text{has-communication-method}$  and  $\text{has-member-profile-features}$  properties defined within the model and between SOURCES and subclasses of the  $\mathcal{CS}_i$  class (“features”).

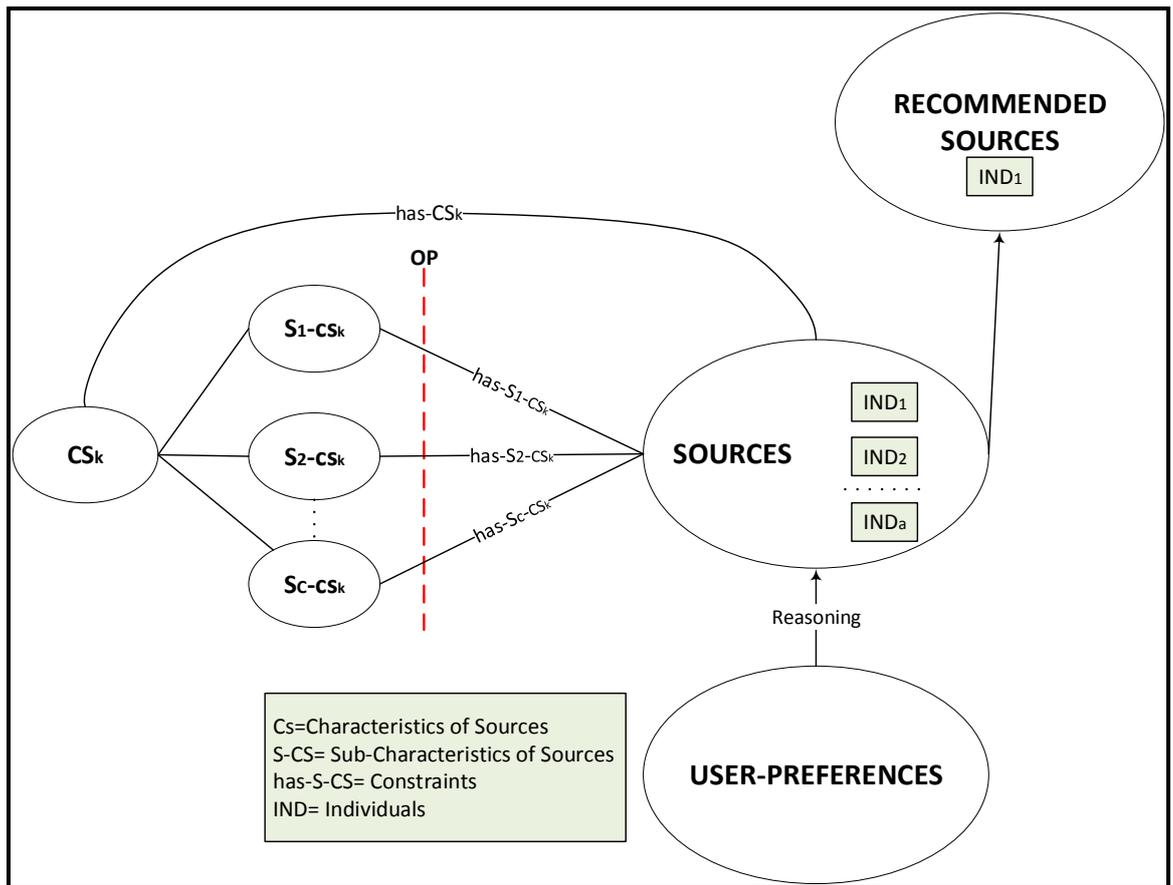


FIGURE 4.7 CHOICE OF CONSTRAINTS FOR EACH CS<sub>k</sub> AND SOURCES.

The choice of OPs illustrated in Figure 4.7 could be made by a software developer, but it might also be dictated by a situation or the domain where the selection of information sources is performed. It is important to take into account that numerous constraints imposed on OWL model might result in software application overload. Hence, extra cautious is required when defining these constraints (Almarri and Juric, 2013a).

Figures 4.8 and 4.9 show further options a developer may have when defining OWL constraints in the abstract model.

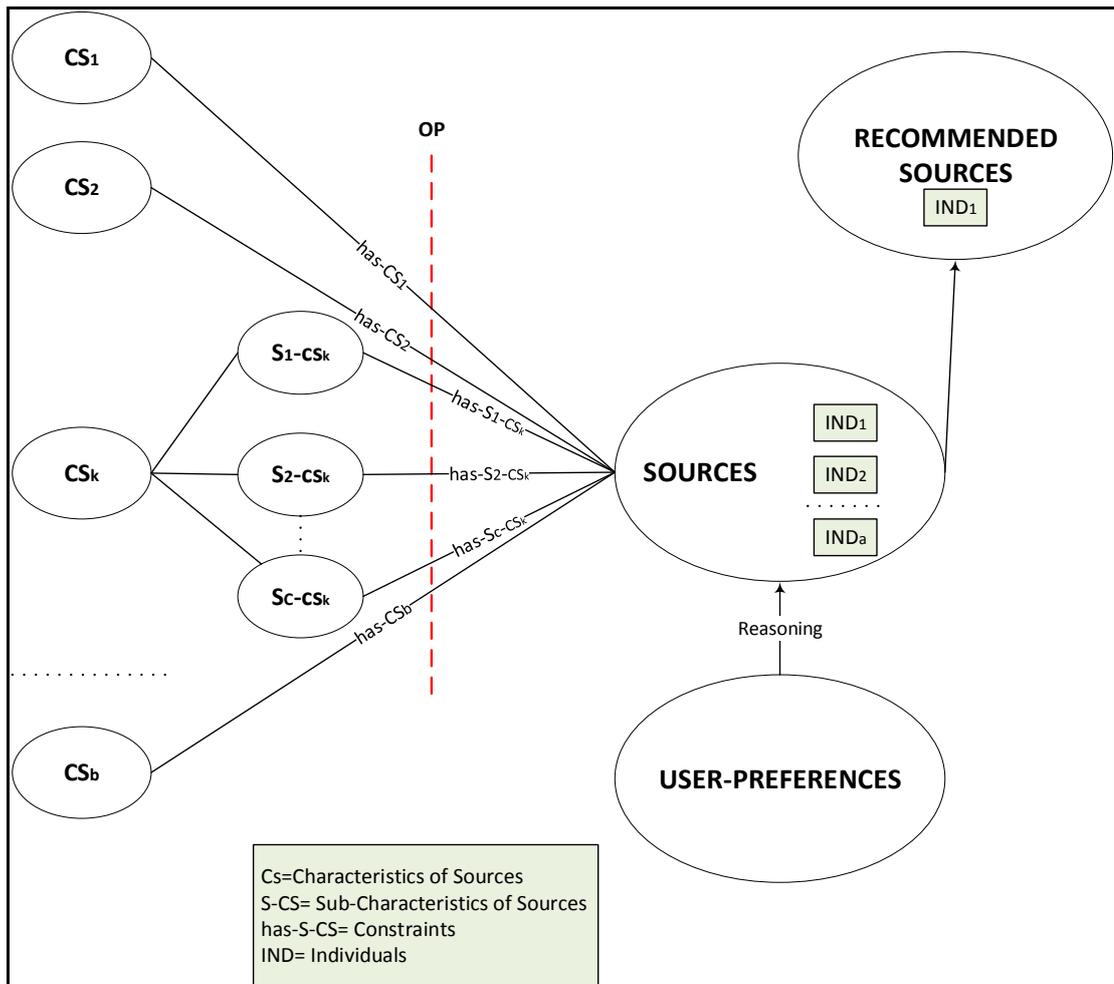


FIGURE 4.8 CHOICE OF CONSTRAINTS IN THE GENERIC MODEL.

It is also possible to choose to use a detailed description of a particular characteristic ( $CS_{\ell}$  in Figure 4.8) and use all other characteristic  $CS_i$  of information sources, without detailing them (i.e. without their further sub-hierarchies due to extracted information seekers/users queries), as shown in Figure 4.8. OPs can also be defined between SOURCES class and a detailed description of all characteristics through subclasses  $\{S_1 - CS_{\ell}, \dots, S_c - CS_{\ell}\}$  of the  $CS_{\ell}$  classes,  $[CS_{\ell} | \ell = 1 \dots n]$  as in Figure 4.9.

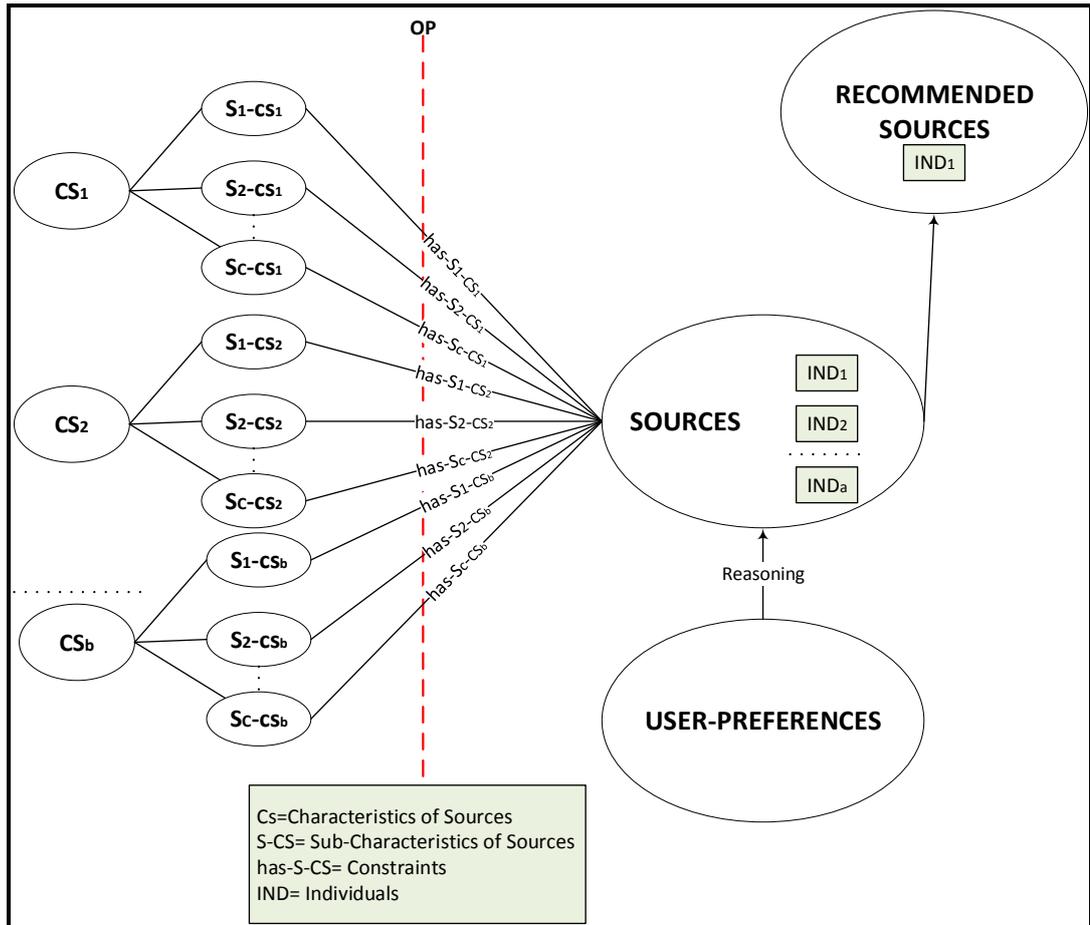


FIGURE 4.9 GENERIC MODEL WITH HORIZONTAL HIERARCHIES AND THEIR CONSTRAINTS.

OWL OPs in the proposed model has essential role in the reasoning process. Hence, the emphasis on OPs in the proposed abstract model has been underpinned with the following three facts:

- it is essential requirement of the proposed model to strengthen the semantic of OWL ontological classes in order to prepare them for reasoning which can be done though OPs. Thus, highly recommended reusability and maintain generic characteristic of the OWL model.
- exploit the natural power of inheritance when describing the semantics of information sources and apply it to its OPs.
- Flexibility of choosing to use either a detailed sub-hierarchies of the characteristics classes together with the generic OPs defined as  $has-CS_k$  which means that the reasoner will use the inheritance for OPs, or removing the inheritance completely from the domain specific OWL model by creating finely granulated  $has-S_1-CS_k$ ,  $has-S_2-CS_k$ ,  $\dots$ ,  $has-S_c-CS_k$  OPs. Powerful constraints might have an adverse effect on the performance of software application based on SWRL enabled OWL ontologies due to its one to one relationship.

Hence, (a)-(c) above means that in the process of modelling the domain of interest it is of high priority to carefully collect, analyse and store domain knowledge in  $\mathcal{CS}_i$  classes and create as much as needed of these  $\mathcal{CS}_i$  classes. Thus, construct and enrich the OWL ontological model with domain knowledge that will enable flexible description of information seeker/user preferences through Ops. The reusability of the proposed model is insured due to the use of object properties in the process of describing relationship between both information sources and  $\mathcal{CS}_i$  classes and information seeker/user preference and  $\mathcal{CS}_i$  classes that does not require any hardcoding in the process of writing SWRL rules for the reasoning process. Hence, the same rule can be reused regardless of information seeker/user preferences.

### 4.3.3 SEMANTIC OVERLAPPING IN THE OWL MODEL AND THE REASONING PROCESS

Figure 4.6 shows an important aspect of the proposed abstract model, which was deliberately omitted from Figures 4.1- 4.5.

In order to guarantee the matching between OWL ontological classes, which can secure reasoning upon their individuals in the generic OWL model, the individuals of the USER PREFERENCES class are described similarly to the same way information sources are described. Hence, secure the selection through these preferences. Readers familiar with OWL ontological matching would know that it would be difficult to select information sources on the internet, according to user preferences, if the sources are NOT described the way that will allow to match them with user preferences. Therefore, a set of  $\mathcal{CS}_i$  classes, which are characteristics of information sources on the internet should play important role when describing individuals of the USER-PREFERENCES class. According to the definitions of OPs in the abstract model from section 4.3.2 above, the same set of OPs  $has\_CS_i$  should be imposed between individuals of the  $\mathcal{CS}_i$  and USER-PREFERENCES class. Obviously, the same rationale of using subclasses of  $\mathcal{CS}_i$  classes, explained in section 4.3.2 can be used for describing the semantics of user preferences.

Therefore, Figure 4.10 presents the final version of the OWL ontological model where constraints  $has\_CS_i$  have been reused in the abstract model to secure semantic overlapping.

The reasoning process in the abstract model, emphasised with blue colour, may be performed successfully, by matching SOURCES and USER-PREFERENCES classes. It can be done through rules written in SWRL which will take all the individuals from the SOURCES class, which can be matched with characteristics of individuals from the USER PREFERENCES class, and infer individuals of the SOURCES class into the RECOMMENDED-SOURCES class. Hence, these moved individuals are results of inference through the reasoning process performed with SWRL rules.

The decision behind modelling user preferences into separate class is due to the way SWRL rule is written. If for example user preferences is an individual of SOURCES class then it would be impossible to select possible matches between user preferences and semantics of information sources stored in the same class (SOURCES). Furthermore, this will require to change how the rule is written.

Although both USER-PREFERENCES and SOURCES class utilise the same set of OPs, semantic matches only happen if there exist an individual in SOURCES class that can match the description of individuals of USER-PREFERENCES class. Hence, the reasoner will only infer both individuals of USER-PREFERENCES and SOURCES class that share common characteristics.

It is obvious that the selection of information sources is based on their characteristics and user preferences. Consequently, the semantic overlapping through the re-use of OPs and correctly interpreted semantic (through  $\mathcal{CS}_i$  sub-hierarchies and OPs of  $\mathcal{CS}_i$  classes) stored in both SOURCES and USER-PREFERENCES classes is essential for performing the semantic matching in this reasoning. In general, the nature of the reasoning upon the classes and OPs in Figure 4.10 is based on our own understanding and interpretation of user preferences and the characteristics of the environment where the selection of information sources on the internet is needed.

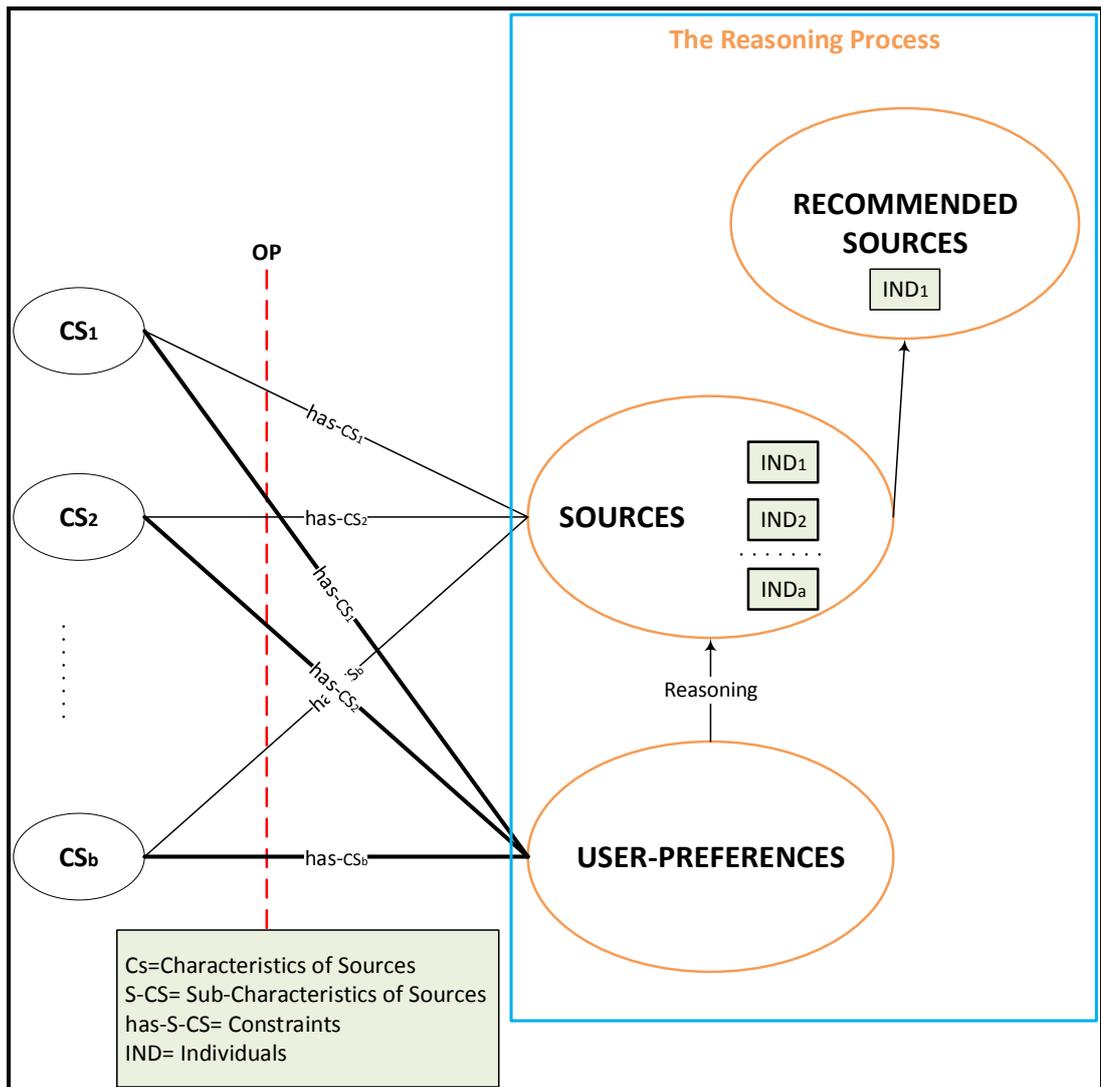


FIGURE 4.10 PROPOSED GENERIC MODEL (THE REASONING PROCESS WITH SEMANTIC OVERLAPPING) .

#### 4.3.4 RESTRICTIONS IMPOSED ON THE OWL ONTOLOGICAL MODEL:

OWL ontologies allow the use of some restriction properties on OWL classes, properties, individuals in the process of constructing OWL ontologies. In this research, some of these restrictions were utilised to control the outcome of the reasoning process. In this section some of these restrictions are listed to show the role of each in the process of the illustration of the generic computational model.

##### **Enumerated classes in the OWL Model:**

Enumerated classes expression allows to define individuals that belongs to a certain class and (only these individuals). In the process of building the OWL model, two classes

are defined as enumerated classes Sources and User-Preferences classes. Due to the definition of Sources and User-Preferences classes it was important to clarify that these two classes are distinct. Therefore, the property enumerated classes was assigned to these two classes to allow the description of membership of individuals to these classes. Therefore, by giving members of a classes the definition of enumerated classes only these individuals can belong to a certain class.

### **Other Restrictions enforced on OWL ontological model:**

In OWL ontologies, classes can be described by applying restrictions to how these classes are interpreted in the model. There are three types of restrictions – quantifier restrictions, cardinality restrictions, and hasValue restrictions- which can help in forcing certain rules in the process of building OWL ontologies.

### **Quantifiers Restrictions:**

Quantifier restriction allows two types of restrictions existential and universal restrictions on OWL classes, properties and individuals of the owl ontological model. This to imply “necessary” and “necessary and Sufficient” conditions in the process of constructing the OWL ontological model.

Each class can have at least one quantifier restriction and can be of the type of both either existential or universal and can be a combination of both. For the purpose of the illustration of the proposed generic model only existential restriction is enforced on OWL classes.

Existential quantifier is denoted with the symbol  $\exists$  and is read as “some values from” and referred to as necessary and sufficient criteria. This restriction allows the description of an individual of a certain class through a relationship that indicates at least one as “some” to another individual of a given class. For example, *has-Features* some **Features** allows to describe all individuals of both **Sources** and **User-Preferences** classes that have at least one relationship along the *has-Features* property to individuals of **Features** class. Furthermore, an individual of both Sources and User-Preferences classes can have more than one value described through a given relationship as depicted in the figure below.

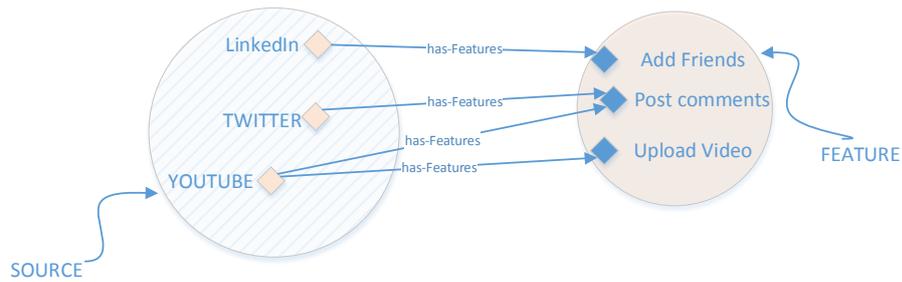


FIGURE 4.11 THE RESTRICTION "HAS-FEATURES" SOME FEATURES INDICATES THAT INDIVIDUALS OF THE FIRST CLASS HAS AT LEAST ONE FEATURE OR MORE .

It is very important to note that these properties are also inherited if for example Features class has a superclass subclass relationship.

The other type of quantifier restriction is the universal quantifier; denoted with the symbol  $\forall$  and is read as "only" and known as "All Values From". This type of restriction was not enforced on the OWL ontological model because it restricts the OWL classes to certain rules which are not applicable with the reasoning process of the proposed generic model. This type of restriction is applicable if and only if we allow the grouping of Sources into the creation of sub-hierarchies.

### **Cardinality Restrictions:**

Another type of restrictions which can be applied to the OWL ontological model is cardinality restrictions. This type of restrictions allows to describe a class of individuals to have at least, at most and exactly specified number of relationships with another individuals or data value. This type of restriction is useful if all relationships in an OWL ontologies must be utilised in the description of the environment or domain of interest. However, this type of restriction was difficult to use in the illustration of the generic computational model because as previously described in chapter 4 section 4.3.1, the OWL ontological model can have "n" number of classes and relationship imposed on the OWL ontological classes. Hence, it is impossible to restrict relationships imposed on OWL classes to utilise some or all.

Although the domain of interest is designed to cover all possible concepts and relationship between concepts. The utilisation of relationships (object properties) in the illustration process of the proposed generic model depends on the user preferences and the description of the information sources in certain situation. Therefore, this type of restriction was not used.

## 4.4 SWRL RULES AND REASONING PROCESS

The current SWT stack is widely accepted that each layer will improve the ones below. SWT language architecture can be extended with a rules component. This to maximises compatibility with existing languages; i.e. RDF and OWL to benefit the development of the SWT (Horrocks et al., 2005). Rules were developed to enhance the performance of the SWT stack. SWRL rule is a first order logic rule that is used to query RDFS and OWL ontologies to extend the existing work of both. It is based on a combination of OWL Lite and OWL DL sublanguages. with the Unary/Binary Datalog RuleML sublanguages of the Rule Markup Language (SWRL, 2004a).

As mentioned above SWT stack consist of multiple layers, each layer exists for a purpose. In this research the proposed generic model is based on SWRL enabled OWL ontologies. This means that a combination of two distinctive layer of SWT are utilised in order to construct OWL ontological model in a domain specific scenarios. Although OWL adds considerable expressive power to the semantics of any OWL ontological model and it allows flexibility in modeling domains of interest, but it does have expressivity limitations -one important issue is that OWL cannot retain decidability of key inference problems. Consequently, it is necessary to extend OWL expressive capabilities with a more powerful language. This to overcome expressivity limitation problem and utilise the ontological model to the maximum end. Therefore, for this purpose SWRL -an enhanced language of OWL Rules Language (OWL RL) (Horrocks and Patel-Schneider, 2004)- was developed to extends OWL in a syntactic and semantic coherent manner. Furthermore, SWRL rules are given formal meaning through OWL DL model theoretic semantics and adds a new axiom to the OWL model. Hence, to interpret the semantics of the domain of choice and infer new knowledge, then it is important to wirte SWRL rules on top of the OWL ontological model to reason and infer further knowledge from the domain model.

As a common knowledge SWT is a human/computer readable technology. Therefore, SWRL rules are likewise. The syntax of SWRL rules consists of two parts the antecedent which is responsible for managing the conditions part of the reasoning process. An antecedent can consist of minimum of one condition and can include an “n” number of conditions. This depends on the complexity of the problem domain. The

second part of the SWRL rule is the consequent which is the results of the truth value of the antecedent.

antecedent  $\rightarrow$  consequent

Rule 1: antecedent and consequent are conjunctions of atoms written  $a_1 \wedge \dots \wedge a_n$ . Variables are indicated using the standard convention of prefixing them with a question mark (e.g., ?x). Using this syntax, a rule asserting that the composition of SOURCES (?S) and USER-PREFERENCES through the property *has* –  $CS_i$  implies (?S) is inferred in the class RECOMMENDED-SOURCES and would be written as:

$$\text{User} - \text{Preferences} (?UP) \wedge \text{Sources} (?S) \wedge \text{has} - CS_i (?UP, ?CS_i) \wedge \text{has} - CS_i (?S, ?CS_i) \rightarrow \text{Recommended} - \text{Sources} (?S)$$

A rule is read as if the antecedent holds or is true then the consequent must also hold.

Rule 2: a SWRL based OWL ontologies contains a mixture of OWL DL constructs i.e. annotations, axiom about classes and properties and facts about individuals of certain environment.

### **Initiation of the reasoning process:**

As mentioned throughout this research, the proposed computational model is a SWRL enabled OWL ontologies. This mean that all languages used in the process of illustration is based on utilising the SWT stack; namely semantic web rule language (SWRL) and web ontology language (OWL) ontologies.

The process of constructing the OWL ontological model starts once domain knowledge is collected and analysed as mentioned in (section 4.1 and 4.2). Hence, categorised domain data is ready to be constructed as OWL ontological model component (classes, OPs and individuals). This to prepare the OWL ontological model to reason upon the semantics of the domain of interest according to user preferences.

Therefore, the listed steps blow must be carefully followed before initiating the reasoning process:

1. Create base classes and its sub-hierarchies by following one of the versions of the proposed generic model (in section 4.3.1- 4.3.3).
2. Create individuals of each class as categorised from the analysed domain knowledge. This step also involves the creation of individuals of both Sources and User-Preferences classes.

3. Create object properties to allow the definition of both the individuals of Sources and User-Preferences classes. This is done by:
  - a. assigning the domain and range for each object property and
  - b. use the object property to define the characteristics of individuals of both Sources and User-Preferences classes.
4. The population of the SWRL rule elements is dependent on the choice of the generic computational model described in (section 4.3.1- 4.3.3). This mean the choice of object properties utilised in the SWRL rule depends on:
  - a. How precise is information seeker/user in describing the desirable information sources and
  - b. This leads to the selection of one of the four versions of the proposed generic model.

Once (1)-(4) above are done correctly, the execution of the SWRL can start.

**The description of the role of the SWRL Rule in the reasoning process:**

Suppose that a information seeker/user is trying to view an information source that contains information on diabetes. Then,

If the information seeker/user provides a minimum of one preference, then a list of sources will be recommended.

if the user provides a minimum of one preference , then a list of sources will be recommended  
premises conclusion

The truth table for User-Preference  $\rightarrow$  has preference

$$\left\{ \begin{array}{l} \text{User - prefernce } \rightarrow \text{ has - Preference} \\ \text{Source } \rightarrow \text{ has - Preference} \\ \text{User - Preference } \wedge \text{ Source } \rightarrow \text{ Recommended - Sources} \end{array} \right.$$

To define that p implies q it is important to satisfy the truth values for if p then q

TABLE 4.1 THE TRUTH TABLE FOR  $P \rightarrow Q$ .

User-Preference (p)	Has preference (q)	User-Preference $\rightarrow$ has preference
T	T	T
T	F	F
F	T	F
F	F	T

Under one circumstance that this is justified as false is if the information seeker/user provides a preference which does not exist. Hence, the conclusion is false. This means that although the user provided some preferences, the condition was not fulfilled and the model did not infer any recommended source to the user.

The second row in the table above can be the justification of this claim. The inability of finding any matching between the user preferences and the description of the information source on the internet implies that although the premises is fulfilled the conclusion is false. Hence, the rule will not infer any information source -third row of the table shows that the user did not specify any preference. The rule says that the premises must be true for the conclusion to be true. The condition is only fulfilled if the user provides a preference, and gets back a recommended source. If the user provides a list of preferences, then a minimum of one source may or may not be recommended based on the reasoning process described below.

1. Assign a variable to individuals of USER-PREFERENCES class
2. Assign a variable to individuals of SOURCES classes
3. Assign a variable to individuals of the range class in each relationship between USER-PREFERENCES and range class through  $has\_CS_i$  and SOURCES and range class through  $has\_CS_i OP$ .
4. The loop starts after assigning variables as described in (1)-(3) above.
  - a. The first loop in the rule will collect number of relationships used in the description of user preferences.
    - (I). This means that even if the domain of interest uses “n” number of relationships to describe the characteristics of the domain and the user only specified preferences which utilises few then the rule will only run based on the number of relationships that describes the user preferences. For example, if user preferences utilise three object properties to describe the semantics of user preferences then the rule will assume that if these three preferences are met then an individual of SOURCES will be recommended.
    - (II). This also means that the rule is restricted to perform matching between user preferences and sources class individuals based on the number of relationships used to describe the user preferences only.
  - b. The second loop in the rule will only select individuals of sources class which utilise exactly the same relationship (OP) identified in (a.) above.
  - c. The third loop in the rule will go through each individual of SOURCES selected in (b) above and test whether the description of each individual matches the user preferences or not for each relationship separately. Note that the number of loops for each relationship will repeat based on the number

of times each relationship is used to describe user preferences as described in the table above.

- d. Results of each loop must always be true. This means that even if the individual of the SOURCES class passed both the first and second loop but failed the third loop, the rule will automatically disqualify the individuals from proceeding to the next step.
5. The rule infers results of matching between USER-PREFERENCES and SOURCES classes in RECOMMENDED-SOURCES class. Therefore, an instance of semantically matched source(s) will be inferred.

The length of the loops in the rule depends on again the user preferences. If the user provides a detailed description of preferences each loop in (a) – (c) above will increase based on number of relationships and number of the usage of each relationship (OP) (see table 4.1).

#### 4.5 EXAMPLES OF MOMENTS/SITUATIONS (OWL MODEL)

The illustration of the proposed computational model starts by first collecting the semantics of information seeker/user preferences and characteristics of the environment where selection of information sources happens. This means that the proposed computational model must become domain specific to:

- (a) interpret the semantics of the environment
- (b) interpret the semantic of user preferences,
- (c) Build an ontological model; SWRL enabled OWL ontologies to provide the user with a “moment” of Internet searches.

An elaboration on how the interpretation of semantics of semantics of both the environment and information seeker/user is present in (sections 4.1-4.2 and 4.4). Hence, in order to select the best possible information source on the Internet. Therefore, we wish to elaborate on some examples that provided a moment in their approaches to solve different problems of either selecting or providing results based on a “moment”.

In Pervasive Computing Environment (PCE) Shojanoori, attempted to model a PCE that is based on constant change in the characteristics of the environment. This means that the proposed approach takes into account actions that triggers new situations in PCE. Furthermore, the model also delivers new results of a moment on acquired information; that indicates the presence of new situation. Therefore, data collected for the previous situation are discarded because they became of no use in the next situation. Thus, this

research avoids historic information when computing the semantics of a PCE (Shojanoori, 2013).

Similarly, Almami et al. focuses on providing a moment of the best possible teaching practice(s). therefore, we can find another implementation of a moment in this work. Thus, the computational model proposed in this work allows users to decide on which teaching practices will be suitable for pupils with impairments if they have clearly defined learning goals. Therefore, the model assumes that the developed ontological model is consistent of classes, individuals, and constraints that will describe the semantics of the environment and the relationship of its concepts to each other. This to provide the users with the best possible teaching practice(s) (Almami et al., 2015). Both Shojanoori and Almami et al. provided results based on the semantics of a moment and did not take into consideration the past or future behaviour of either the environment nor the user.

Consequently, to deliver to users “a moment” of Internet searches; we would like to characterise a moment as a situation that requires:

- (1).Semantics of the environment. This means that we investigate a problem domain to collect data as described in our experiments (Almarri et al., 2016).
- (2).Semantics of the user preferences. This means that we analysis a situation where users require assistance to find the most relevant Internet sources. This can be present in the form of competency question.
- (3).Build an ontological model (classes, individuals and properties) to interpret the semantics of the environment from 1 and user preference from 2 above.

## 4.6 CHAPTER SUMMARY

In this chapter we first started by setting up the research problem this twofold two important aspects of this research:

- First, our research is focused on finding the most relevant Internet sources according to users’ goals, interest, intentions, and preferences, and
- Second, we introduced a generic computational model that will interpret the semantics of the environment where Internet searches happen. In addition, it takes into account users’ preferences and compare and semantically match them to the available sources on the Internet.

Therefore, we started this chapter by providing the reader with an overview of existing problems from previously proposed solutions and techniques to solve the problem of IO. Thus, this helped us to define the Role of users in self-oriented information sources discovery on the Internet. We further defined the relationship of the

two main concepts of the proposed approach in this research; characterising user preferences and Internet source.

Consequently, we proposed a novel computational model that will deliver our research objectives and alleviate modern IO. The model is generic, i.e. can be used in any domain of interest where we wish to receive relevant results of Internet searches. It is Semantic Web Rule Language (SWRL) (SWRL, 2004b) enabled Web Ontology Language (OWL) (OWL, 2004a) ontology-specific approach. This means that the computations proposed by the model are explained through various mechanisms that exists in SWRL enabled OWL ontologies and include reasoning through SWRL rule language an initiative of the W3C. Moreover, in this chapter we provided a detailed description of the proposed generic computational model and its main components. We also justified design decision depicted in figures 4.5- 4.10 (see section 4.7).

In section 4.8 we provided an overview of the history and proliferation of the WWW technologies. We also introduced SWT as technology of choice, advantages SWT brings to the Internet and description of different languages in the SWT. Furthermore, we enriched our choice of technology by discussing and presenting research examples that utilized SWT and its stack to solve domain specific problems. In sections 4.8.1-4.8.4 we provided definitions and justification SWT stack used in the implementation of the proposed generic computational model in this research. In sections 4.9- 4.10 we provided examples of a moment in OWL research and criterion to implement the proposed generic computational model.

## CHAPTER 5. ILLUSTRATION AND IMPLEMENTATION OF THE PROPOSED MODEL

In this section, we focus on the illustration of the proposed model from chapter 4 (see sections 4.3.1-4.3.3). We have to define a specific domain where the selection of Internet sources happens and, consequently, we will be able to address and debate (c) from chapter 4 (see section 4.2).

This section is divided into two parts because we wish to address two different case studies. These two case studies are important for two reasons:

- a) We must show that the generic proposed model is deployable and reusable. Therefore, we need at least two different scenarios (two different situations and domains) in order to illustrate the reusability.
- b) We must also look at the (c) part of the proposed model from chapter 4 (see section 4.2) and illustrate the changes in SWRL rules when we change the scenario (case study).

We have had numerous choices of domains suitable for the illustration of the proposed model. For the purpose of this document we choose the domain of LLL and the selection of learning sources for an ad-hoc learner. However, readers interested in the overall applicability of the proposed model should read some of our publication where the selection of online sources may be performed outside learning environments (Almarri et al., 2016), (Juric et al., 2013), (Mahmood et al., 2013), (Binghubash and Juric, 2011). In all cases we address the urgent need of dealing with IO and the selection of online sources that have not been successfully addressed with our search engines and various other methods of finding the right online sources at the right time.

In section 5.1, we introduce the domain of LLL. This is needed because both of our case studies have to have a rich background, which can derive scenarios where the selection of online sources takes place, after they are delivered by a search engine. Sections 5.1.1 and 5.1.2 show these two case studies. The first one focuses on the selection of online sources for a medical student and the second one is centred on the needs of a diabetic patient when managing his/her chronic condition.

Finally, any illustration of the proposed model must be SWRL enabled OWL ontologies specific. This means that we must use vocabulary, terms and descriptions

typical for SWRL enabled OWL computations. Consequently, we assume that readers are familiar with all of them and no further explanations are needed for case studies, except for the purpose of illustrating the proposal.

We also wish to clarify that case studies in sections 5.1.2 and section 5.1.3 are conducted on two steps:

- We first interview the user who is applicable, according to the problem domain we selected for our experiments, and we provide a list of questions to help us understand the problems the user faced when looking for relevant information.
- we then investigate the environment to collect the semantics of the environment where the user struggled to retrieve relevant information according to his/her preferences.

Based on the interview answer and our investigation of the domain, we create the ontological model.

## 5.1 ILLUSTRATION OF THE PROPOSED MODEL

### 5.1.1 THE DOMAIN OF LLL IN HEALTHCARE

The proliferation of web, semantic web, virtual and pervasive technologies has made an impact on the way we perceive learning, teaching, and the dissemination of knowledge today. Learning outside traditional education institutions has already gained momentum, following rapid advances in mobile and wireless, communication and software technology, which are pervasive in the learning environment (Alexander, 2006), (Graf et al., 2008), (Kemp and Livingstone, 2006), (Syvanen et al., 2005), (Yang S. J. H., 2006). Modern learners are not solely dependent on their teachers and lecturers. They tend to take charge of their learning processes and make decisions on the most suitable learning pathways. Having the opportunity to choose to learn at any stage of our lives and achieve our goals brings our societies forward in terms of gaining and exchanging knowledge, which in turn affects our private lives, businesses, governance and economy (Kokosalakis, 2000), (Silverwood et al., 2008). These views have created the term LLL, which has been present in education over the last decade (Collins, 2009), (Dali and Yongmei, 2008), (Duke, 2012), (Dunlap and Lowenthal, 2011), (Graven and MacKinnon, 2005), (Martin, 2010), (Olson et al., 2008), (Wessner et al., 2002). However, there is no widely accepted definition of LLL, but it has been debated by various interest groups in education environments and appeared in government initiatives around the world (Jannette, 2009), (Smith and Clayton, 2009). For example, the UK government report from 1997 (Dearing, 1997) introduced the term LLL as ‘thinking in the long term about education and learning’.

If learners wish to focus on LLL, which takes place outside formal institutions, for the purpose of satisfying a learner’s personalised needs at the time and place when it suits the learner, then they should agree that that LLL happens very often and in many situations in someone’s life. Each situation is a particular moment, when an individual wishes to satisfy his/her ad-hoc created learning goal(s). Obviously, learners trigger each ‘situation’ in order to participate in learning. Learner may be surrounded by various information sources, which are learning sources, full of information that can be delivered to the learner on an ad-hoc basis, possibly on mobile and wireless devices, allowing the learner to collaborate and exchange information or knowledge, post questions and learn

from individuals and groups of his/her choice and own preferences. Information sources on the internet, equipped with social media tools, that guarantee the creation of online, socially intensive and highly collaborative environments, which would support various situations in someone's LLL (Baxter et al., 2011), (Klamma et al., 2007), (Redecker et al., 2010), (Wan, 2010), (Wang, 2011). However, finding and choosing appropriate information sources, in a particular situation in LLL is a problem. It requires understanding of both: learners' goals, needs and preferences on the one hand, and characteristics, purpose and content of learning sources on another. Therefore, if we wish to address learner's participation in a situation of LLL, it is important to create a mechanism for selecting the most suitable learning source(s) that would "fit the situation", i.e. the information source(s) that would guarantee that the learner will achieve his/her ad-hoc created learning goals and preferences. These information sources may contain data, information and knowledge created by various parties for various purposes, aimed at various information seekers/users.

However, it is necessary to narrow the domain of the LLL to have a clear illustration of the proposed model. Hence, will give an example by creating a situation in an LLL by choosing a healthcare domain. The motivation is twofold:

- Firstly, healthcare is a domain where the advances of mobile and wireless communications and new computing technologies that accompanied them have been very successful in delivering pervasiveness in their environments. Pervasive healthcare has enabled us to create environments which empower both the patient and the healthcare professional in terms of disseminating information, experiences and knowledge on a daily basis.
- Secondly, the impact of web 2.0/3.0 technologies (Murugesan, 2009), tools and visions, in terms of using social interactions, supported by SNs and Social Media (SM) tools that have enabled information sharing by healthcare professionals and physicians in particular. The constant learning and exchange of experiences and knowledge enables professionals to learn from each other, from other information seekers/users interested in healthcare, such as medical students, caregivers, service providers and patients.

There are numerous examples of LLL in healthcare:

- healthcare professionals who must take learning opportunities and undertake professional developments on an almost daily basis (CSC et al., McGowan et al., 2012),

- public health workers being educated through LLL practice education which has enhanced with a competency based curriculum development (Olson et al., 2008),
- radiologists who ought to develop a habit of LLL (Collins, 2004),
- orthopaedic surgeons who are delivered educational materials by the volunteer body that oversees all practice management related initiatives at the American Academy of Orthopaedic Surgeons (AAOS, 2012),
- a social media physician's voice in the form of a blog <http://www.kevinmd.com/blog>, created by Kevin Pho, MD, for physicians' insight on breaking medical news,
- an MD who sees Twitter as a doctors' lounge where he can participate in worldwide discussions on the latest journal articles or clinical research (Dolan, 2012), and
- empowered patients who have moved beyond the social media content they can "follow" or "like" (Amednews) and many more.

### 5.1.2 CASE STUDY: SELECTION OF ONLINE SOURCES FOR A MEDICAL STUDENT

The LLL scenario focuses on the request/demand of a medical student for acquiring more knowledge for the purpose of his/her professional advancement. Let us assume that (a) the knowledge needed for professional advances of the medical student cannot be gained in the traditional manner (through any formal education) and (b) the student wishes to learn when he/she is either ready or motivated. However, it is very important to formulate this scenario in a particular manner, i.e. by using ‘competency-question-style’ (Gruninger and Fox, 1995), (Gruninger, 1996), (Noy and McGuinness, 2001). Furthermore, adhere to the OWL terminology and it would be wrong to use any other expression or term to explain the format of the scenario. Using OWL terminology and converting a particular situation in LLL into OWL ‘competency questions’ helps readers understand the way SWRL enabled OWL ontologies computations are constructed. Their purpose is to reason about the most suitable sources for the lifelong learner in the scenario.

This is our scenario:

“I am a medical student who would like to join a research group on ‘*Quality & Safety in Health Care*’, within the Faculty. However, I will not qualify for the membership of the research group if I do not learn more about ‘*Quality & Safety in Health Care*’. I have no related publications, but I have been motivated to join the research group, since I attended my tutor’s research talk, which in the long term, might determine my future professional pathway. I would like to know how to obtain access to and learn from online sources, which are most suitable in my situation. Therefore:

- (i) I wish to learn from online sources, which provide knowledge created by *experienced professionals*;
- (ii) I also wish to learn from socially intensive environments, which will allow me to post questions, use *forums* and browse *electronic libraries* where I can search from the latest *articles in journals*, using *author’(s)’ names* or the names of groups of researchers;
- (iii) I should be able to view *videos and images* uploaded by *members* of the social networks interested in ‘*Quality & Safety in Health Care*’
- (iv) I should have access to their *discussion boards* and *professional blogs*.

- (v) I will be using my *smart phone* and will have to know exactly which *App* for *iPhone* I will have to download in order to manage my access to a SN(s) which will allow me to learn about my topic of interest.

The scenario above creates two competency questions:

- CQ1: Which information source(s) on the internet is the most suitable in the situation?
- CQ2: Which App for iPhone will have to be downloaded?

They should be answered by performing reasoning through SWRL rules that utilise the stored semantics of the domain and user preferences in the OWL ontology as proposed in the abstract model from Figure 4.10.

The next 3 subsections, give an illustration of the abstract model, which was defined in 4.3, by using the scenario above. This means that the ontology and the reasoning process become a domain, i.e. scenario specific. Consequently, the part of the proposed model from section 4.3 will be enriched with the exact SWRL rule and thus bullet (c) from section 4.2 will be illustrated through this particular scenario.

#### 5.1.2.1 THE ONTOLOGICAL MODEL FOR THE CASE STUDY

Ontologies are bodies of knowledge which describes certain domain and context represented in vocabulary. The contents of ontologies are dependent on certain requirement of intended usage. Hence, ontologies are constructed to deliver logical assertions which include simple statements, facts and rules of how to reason with facts stored in the ontological model. Consequently, new knowledge is inferred/derived by utilising deductive reasoning.

The model in Figure 5.1 is an illustration of the generic model given in Figure 4.10. According to the scenario, the construction of the domain knowledge requires to create 4 different characteristics of information sources, i.e. there are four  $[\mathcal{CS}_i \mid i= 1, \dots, 4]$  classes in Figure 5.1 which define characteristics of information sources: FEATURES, TOPICS\_OF\_INTEREST, MEMBERS\_ROLE and TECHNOLOGY. The categorisations of  $\mathcal{CS}_i$  into these for classes is due to their description extracted from the competency question above.

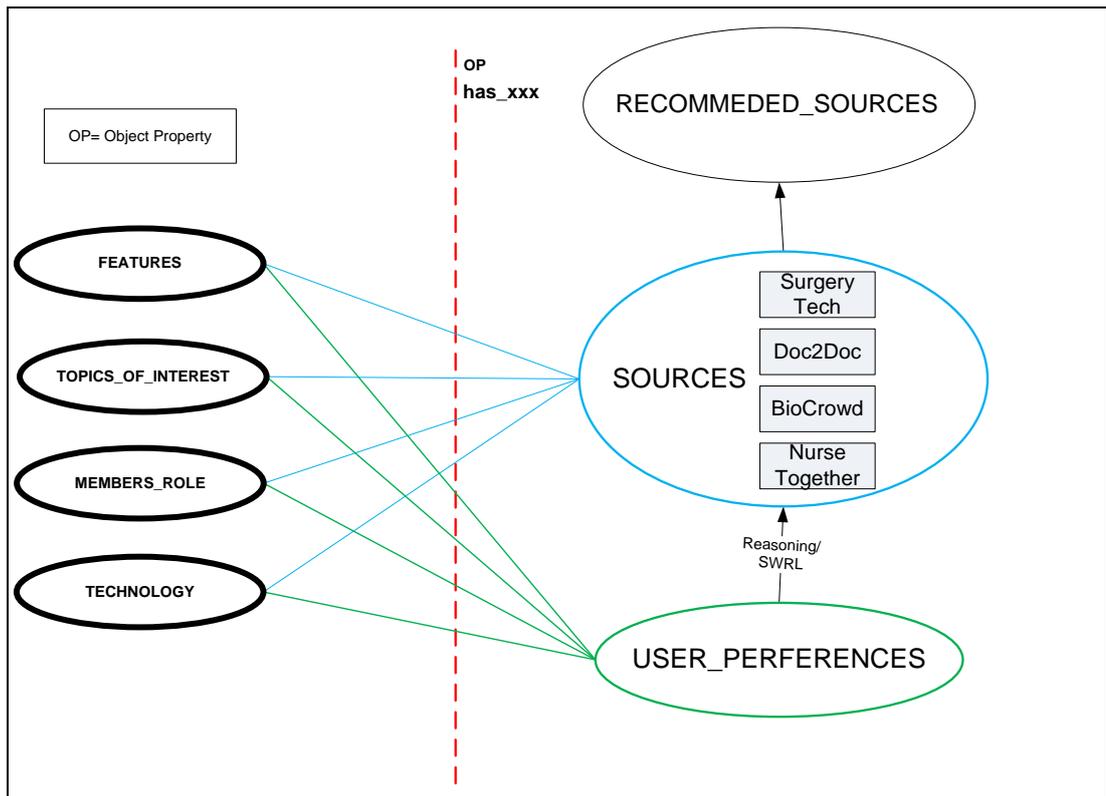


FIGURE 5.1 THE ONTOLOGICAL MODEL BASED ON FIGURE 4.10. AND DERIVED FROM THE SCENARIO IN THE CASE STUDY.

Figure 5.1 also shows a few individuals stored in the SUORCES class: SurgeryTech, Doc2Doc, BioCrowd and NursesTogether are illustrations of possible information sources which will be available for the selection, according to the scenario.

Figure 5.2 adds more horizontal hierarchies to the ontological model as  $\mathcal{CS}_i$ . For example, the FEATURES classes could be better described if SOCIALIZING, SEARCH and RESEARCH are created as subclasses to it. Furthermore, the RESEARCH subclass of the FEATURES class could be better described if ARTICLES, eLIBRARIES, JOURNALS and RESEARCH\_BLOGS are created as subclasses to it. The same applies to any  $\mathcal{CS}_i$  class which describes the characteristics of information sources.

Figure 5.2 also specifies clearly that all object properties are defined according to Figure 4.10 of the abstract model. Therefore, the OWL model:

- a) May have characteristic classes  $\mathcal{CS}_i$  further described through their subclasses (i.e. horizontal hierarchies)
- b) Should use constraints (i.e. object properties in this example) which are defined between characteristic classes  $\mathcal{CS}_i$  and SOURCES / USER-PREFERENCES classes.

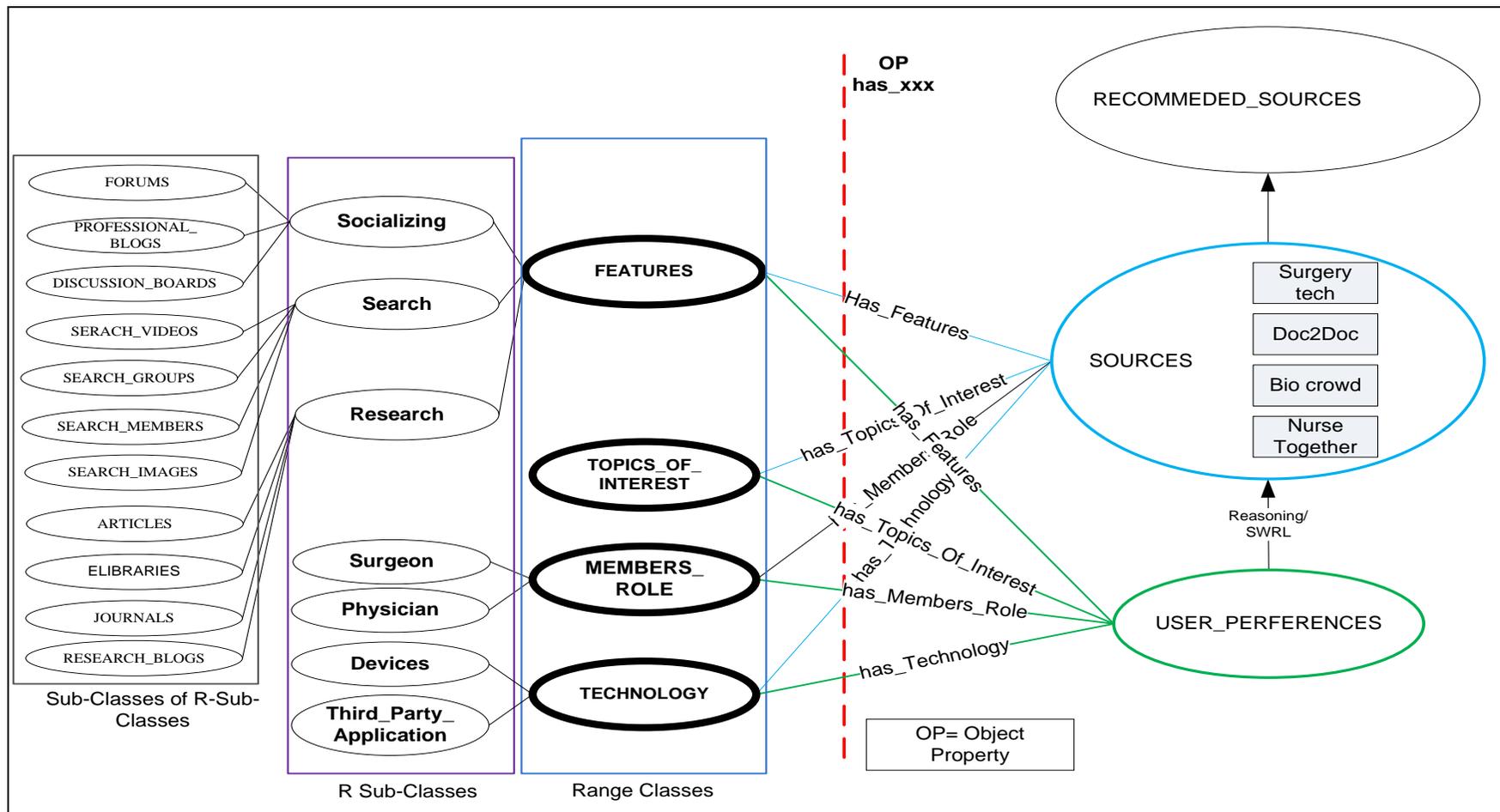


FIGURE 5.2 DETAILING THE ONTOLOGICAL MODEL FROM FIGURE 5.1.

Hence, Figure 5.2 resembles the generic model from Figure 4.6 and deploys  $has-CS_i$  properties instead of optional  $has-S_j-CS_i$ . Therefore, in the OWL model for a medical student exploit i. and not ii. from 4.5. This has further been justified with bullet (c) from section 4.2

An elaboration on how the design decision on the ontological models from Figures 5.1 and 5.2 were made is yet to be discussed. Thus, describe and explain how  $CS_i$  classes were decided upon and their horizontal hierarchies. Hence, the next few paragraphs give a rationale which underpins design choice. Table 5.1 shows how the semantic(s) from the scenario generates  $CS_i$  classes, their horizontal hierarchies and their individuals.

The right-most column in Table 5.1 contains a selection of “words” which are taken from the scenario. These are the words which are in italics in the scenario text. The choice of characteristic classes in Table 5.1 might be obvious to readers because they are clearly defined in the scenario. However, their subclasses are either:

- (i) consequences of more detailed explanations of characteristics of information sources in the scenarios or
- (ii) very well known “facts” which can be deduced from the scenario.

The contents of the column named Individuals from Table 5.1 contains more asserted individuals (as in (ii)), than individuals that are taken out to from the scenario (as in (i)). The reader would be familiar with individuals of the DEVICES and THIRD\_PARTY\_APPLICATION classes as being asserted into the ontology regardless of which USER-PREFERENCES a medical student may have. They may also be legitimate individuals in some other scenarios where DEVICES and THIRD\_PARTY\_APPLICATION characteristic classes play important role.

Table 5.1 also shows how  $CS_i$  classes are derived. For example, the FEATURES class exists because we have to model that

- Individuals such as Forums\_yes, Articles\_yes, eLibraries\_yes and similar will need classes where they can be assigned. Therefore, it is necessary to have FORUMS, ARTICLES, eLIBRARIES classes (these subclasses are listed in the middle column in Table 5.1).
- Classes from the middle column in Table 5.1 can be categorised: FORUMS class is part of socialising on the web and thus SOCIALIZNG in the second column from left in Table5.1. The same applies to the ARTICLES and eLIBRARIES

classes: they belong to research activities on the web, and thus RESEARCH in the second column from left in Table5.1.

- The SOCIALIZING and RESEARCH classes store the semantics of various features offered to members of social intensive online environments, and therefore they can be subclasses of the FEATURES class. In other words, it is a must to create class FEATURES to model various features of information sources, where socialising and research are only a few of them.

These three bullets above demonstrate that the choice of  $\mathcal{CS}_i$  classes is the result of trying to prepare classes to store the semantics of individuals from the scenario and categorise these classes to create horizontal hierarchies, according to their role and purpose in the scenario.

TABLE 5.1 EXCERPTS FROM THE ONTOLOGY: A SELECTION OF CHARACTERISTICS CLASSES AND INDIVIDUALS

Characteristics Class	Characteristics Sub-Class	Characteristics Sub-Sub-Class	Individual	Scenario
TOPICS_OF_INTEREST			Quality_&_Safety_in_Healthcare	Quality & Safety in Healthcare.
MEMBERS_ROLE	PHYSICIAN SURGEON		Physician_Yes Surgeon_Yes	experienced professionals
FEATURES	SOCIALIZING	FORUMS DISCUSSION_BOARDS PROFESSIONAL_BLOGS	Forums_Yes Discussion_Boards_Yes Professional_Blogs_Yes	Forums Discussion boards Professional blogs
	RESEARCH	ARTICLES eLIBRARIES JOURNALS RESEARCH_BLOGS	Articles_Yes eLibrary_Yes journals_Yes research_blogs_Yes	electronic libraries articles in journals
	SEARCH	SEARCH_MEMBERS SEARCH_GROUPS SEARCH_VIDEO SEARCH_IMAGES	Search_members_Yes Search_By_Research_Group_names__Yes Search_video_Yes Search_images_Yes	Search Author'(s)' names or names of groups of researchers. View videos and images
TECHNOLOGY	DEVICES		BlackBerry HTC Samsung_smart_phone ipads iphones other_smart_tablets	iphone
	THIRD_PARTY_APPLICATION		Download_All_Pro Download_Free Download_Music_Pro Download_Video-Dolt_Video Free_Music_Download MyMedia-Download VideoGet_for_Facebook_Lite Video_Downloader_Super_Lite Video_Download-iBolt_Downlaoder iDownloader_Free_Download	App

Figure 5.3 is a screen shot of a selection for possible individuals in the SOURCES class. It is important to note that URLs are individuals in the SOURCES class and not BioCrowd, DermRounds, Doc2Doc and similar. They are actually names of website of these URLs represent information sources on the internet. Names of information sources are used instead of their URLs in order to help reader understand the semantic of the model easily.

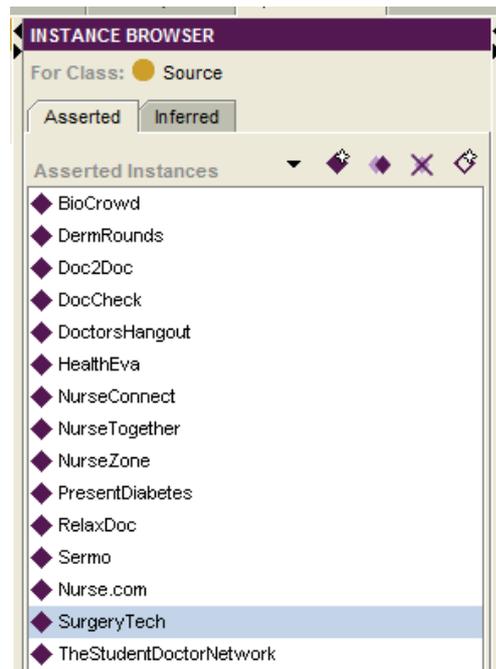


FIGURE 5.3 INDIVIDUALS OF THE SOURCES CLASS FOR THE CASE STUDY

### 5.1.2.2 OWL CONSTRAINTS

Figure 5.4 is self-explanatory: as previously described in section 4.3.2 Ops are named according to their relationship with  $\mathcal{CS}_i$ . Hence, for the purpose of illustrating the domain of interest 4 constraints, which are actually OWL OPs, that are defined according to Figure 5.2 of the abstract model. However, these properties may be defined between any individuals of SOURCES and  $\mathcal{CS}_i$  and USER-PREFERENCES and  $\mathcal{CS}_i$  classes. In order to have a clear picture on how individuals are connected through the properties created tables 5.2, 5.3 and 5.4.

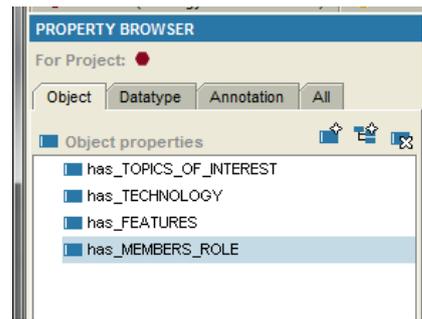


FIGURE 5.4 CONSTRAINTS IMPOSED ON THE ONTOLOGY FOR CASE STUDY 1

For example, Table 5.2 describes an individual BioCrowd from SOURCES class, property `has_FEATURES` that connects BioCrowd with individuals `Search_members_(yes/no)` and `Forums(yes/no)` in order to characterise that BioCrowd allows members to explore forums and search for other members. This is shaded in grey in Table 5.2. The Domain Class Individuals Value column is merely there to help the reader understand what is said above: BioCrowd an information source that allows “forums” and “search for members”, as a part of BioCrowd features and consequently, the `has_FEATURES` property between FEATURES and SOURCES classes must be defined as in Table 5.2. Furthermore, the lack of Domain Class Individuals Value for BioCrowd online source means two things:

- (a) BioCrowd online source has only two characteristic classes as its description FEATURES and MEMBERS\_ROLE and
- (b) There are only two properties defined between BioCrowd, as an individual of the SOURCES class and its characteristics: FEATURES and MEMBERS\_ROLE classes.

Tables 5.3 and 5.4 should be interpreted the same way as Table 5.2.

TABLE 5.2 THE RELATIONSHIPS BETWEEN SOURCES CLASS INDIVIDUALS (RELAXDOC) AND ANY CSI CLASSES INDIVIDUALS THROUGH OBJECT PROPERTY.

Domain Class	D-Class Individuals	D-Class Individuals Value	Object property	R-Class individuals	Sub-Classes of R-Sub-Classes	R Sub-classes	Range Class
SOURCES	RelaxDoc	Articles_Yes eLibraries_Yes	Has_FEATURES	Articles_(Yes/No) eLibraries_(Yes/No) Journals_(Yes/No) Research_blogs_(Yes/No)	ARTICLES eLIBRARIES JOURNALS RESEARCH_BLOGS	RESEARCH	FEATURES
		Search_Members_Yes		Search_Members_(Yes/No) Search_By_Research_Group_names_(Yes/No) Search_video_(Yes/No) Search_images_(Yes/No)	SEARCH_MEMBERS SEARCH_GROUPS SEARCH_VIDEO SEARCH_IMAGES	SEARCH	
		/		Forums_(yes/no) Descussion_Boards_(Yes/No) Professional_blogs_(Yes/No)	FORUMS DESCUSSION_BOARDS PROFESSIONAL_BLOGS	SOCIALIZING	
		Physician_Yes Surgeon_Yes	has_MEMBERS_ROLE	Physician_(Yes/No) Surgeon_(Yes/No)	PHYSICIAN SURGEON		MEMBERS_ROLE
		/	has_TOPICS_OF_INTEREST	Quality_&_Safety_in_Healthcare			TOPICS_OF_INTEREST
		/	has_TECHNOLOGY	BlackBerry HTC Samsung_smart_phone ipads iphones Other_smart_tablets		DEVICES	TECHNOLOGY
				Download_All_Pro Download_Free Download_Music_Pro Download_Video-Dolt_Video Free_Music_Download MyMedia-Download VideoGet_for_Facebook_Lite Video_Downloader_Super_Lite Video_Download-iBolt_Downlaoder iDownloader_Free_Download		THIRD_PARTY_APPLICATION	

TABLE 5.3 THE RELATIONSHIPS BETWEEN SOURCES CLASS INDIVIDUALS (SURGERYTECH) AND ANY CSI CLASSES INDIVIDUALS THROUGH OBJECT PROPERTY

Domain Class	D-Class Individuals	D-Class Individuals Value	Object property	R-Class individuals	Sub-Classes of R-Sub-Classes	R Sub-classes	Range Class	
SOURCES	SurgeryTech	Articles_Yes e-library_Yes Journals_Yes research_blogs_Yes	Has_FEATURES	Articles_(Yes/No) e-library_(Yes/No) Journals_(Yes/No) Research_blogs_(Yes/No)	ARTICLES eLIBRARIES JOURNALS RESEARCH_BLOGS	RESEARCH	FEATURES	
		Search_Members_Yes Search_By_Research_Group_names_Yes Search_video_Yes Search_images_Yes		Search_Members_(Yes/No) Search_By_Research_Group_names_(Yes/No) Search_video_(Yes/No) Search_images_(Yes/No)	SEARCH_MEMBERS SEARCH_GROUPS SEARCH_VIDEO SEARCH_IMAGES	SEARCH		
		Forums_Yes Descussion_Boards_Yes Professional_blogs_Yes		Forums_(Yes/No) Descussion_Boards_(Yes/No) Professional_blogs_(Yes/No)	FORUMS DESCUSSION_BOARDS PROFESSIONAL_BLOGS	SOCIALIZING		
		Physician_Yes Surgeon_Yes	has_MEMBERS_ROLE	Physician_(Yes/No) Surgeon_(Yes/No)	PHYSICIAN SURGEON			MEMBERS_ROLE
		Quality_&_Safety_in_Healthcare	has_TOPICS_OF_INTEREST	Quality_&_Safety_in_Healthcare				TOPICS_OF_INTEREST
		Iphones	has_TECHNOLOGY	BlackBerry HTC Samsung_smart_phone ipads iphones other_smart_tablets		DEVICES		TECHNOLOGY
				Download_All_Pro Download_Free Download_Music_Pro Download_Video-Dolt_Video Free_Music_Download MyMedia-Download VideoGet_for_Facebook_Lite Video_Downloader_Super_Lite Video_Download-iBolt_Downlaoder iDownloader_Free_Download		THIRD_PARTY_APPLICATION		

TABLE 5.4 THE RELATIONSHIPS BETWEEN SOURCES CLASS INDIVIDUALS (BIOCROWD) AND ANY CSI CLASSES INDIVIDUALS THROUGH OBJECT PROPERTY .

Domain Class	D-Class Individuals	D-Class Individuals Value	Object property	R-Class individuals	Sub-Classes of R-Sub-Classes	R Sub-classes	Range Class
SOURCES	BioCrowd	/	has- FEATURES	Articles_(Yes/No) eLibraries_(Yes/No) Journals_(Yes/No) Research_blogs_(Yes/No)	ARTICLES eLIBRARIES JOURNALS RESEARCH_BLOGS	RESEARCH	FEATURES
		Search_Members_Yes		Search_Members_(Yes/No) Search_By_Research_Group_names_(Yes/No) Search_video_(Yes/No) Search_images_(Yes/No)	SEARCH_MEMBERS SEARCH_GROUPS SEARCH_VIDEO SEARCH_IMAGES	SEARCH	
		Forums_Yes		Forums_(yes/no) Descussion_Boards_(Yes/No) Professional_Blogs_(Yes/No)	FORUMS DESCUSSION_BOARDS PROFESSIONAL_BLOGS	SOCIALIZING	
		/	has- MEMBERS _ROLE	Physician_(Yes/No) Surgeon_(Yes/No)		PHYSICIAN SURGEON	MEMBERS _ROLE
			has- TOPICS_ OF_ INTEREST	Quality_&_Safety_in_Healthcare			TOPICS_OF INTEREST
		/	has- TECHNOLOGY	BlackBerry HTC Samsung_smart_phone ipads iphones Other_smart_tablets		DEVICES	TECHNOLOGY
		/		Download_All_Pro Download_Free Download_Music_Pro Download_Video-Dolt_Video Free_Music_Download MyMedia-Download VideoGet_for_Facebook_Lite Video_Downloader_Super_Lite Video_Download-iBolt_Downlaoder iDownloader_Free_Download		THIRD_PARTY_APPLICATION	

### 5.1.2.3 SWRL RULE FOR THE CASE STUDY

The SWRL rules 1 and 2, which run upon the ontological model defined through Figures 5.2 and Tables 5.2, 5.3, 5.4) are given in Figures 5.5 and 5.7. Due to two distinctive of user preference, two competency questions clearly generated from the scenario as CQ1 and CQ2 from section 5.1.2 and they have to be answered through SWRL rules imposed on OWL ontological model.

The result of running Rule 1 is in Figure 5.6. Out of 15 online sources from Figure 5.3, only two online sources RelaxDoc and SurgeryTech are selected and they MAY be suitable for the medical student. However, the rule does not take into account if RelaxDoc and SurgeryTech are suitable for iPhone users. Therefore, we need another rule.

Rule 2 from Figure 5.7 does two different things:

- (1) It selects (again) online source from Figure 5.6, which satisfy the technology preferences of the medical student (see (v) in section 5.1.3: the medical student is an iPhone user)) and therefore it appears that only SurgeryTech would be suitable for the medical student (and RelaxDoc would NOT).
- (2) It adds that the medical student may download the Apps called Download\_ALL\_Pro on his/her iPhone, if he/she wishes to use SurgeryTech, as shown in Figure 5.9.

The competency question CQ1 is answered though Rules 1 and 2. However, CQ2 is answered though Rule 2 only. It is important to note that these two rules are part of a reasoning process which is introduced in figures 5.2 in this particular scenario, which in turn is derived from Figure 4.10.

The paragraphs below explain exactly how the reasoning process is performed. We used extra OWL classes which are NOT shown in the abstract model and in Figures 4.5 and 4.6, but they appear in the SWRL rules below. Their role is specific for the scenario and consequently cannot appear in any abstract model. For example, Rule 2 uses results of reasoning from Rule 1 and therefore these results should be temporarily placed “somewhere” within the ontology and thus these OWL classes are named RULE\_1\_RESULT.

Rule 1 selects the best possible information source for the scenario. Figure 5.6 shows the result of running Rule 1: the student is recommended RelaxDoc and SurgeryTech, which can be selected for him/her. These two sources are actually individuals from the

SOURCES class, which have been moved/inferred to the RULE\_1\_RESULT subclass of RECOMMENDED\_SOURCES class, as a consequence of running Rule 1. These two sources are not the answer to both competency questions. This is because the user also specified that the information source must be accessible through iPhone. Consequently, it is necessary to find out if these two sources ARE accessible to iPhone users, who in turn might need an App to secure access to RelaxDoc and SurgeryTech.

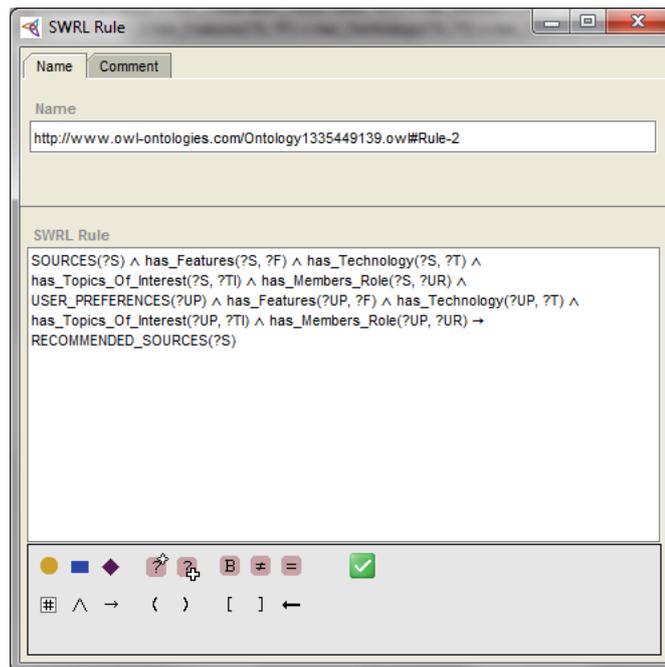


FIGURE 5.5 SWRL RULE 1: SELECTION OF ONLINE SOURCES FOR THE CASE STUDY

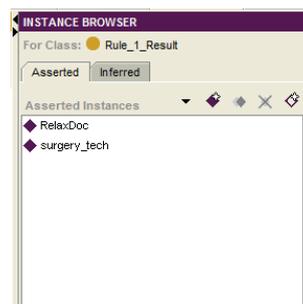


FIGURE 5.6 RESULT OF RUNNING RULE 1 FOR THE CASE STUDY.

Figure 5.7 shows Rule 2. It performs double reasoning (hence Figures 5.8 and 5.9). Thus, it determines if the RelaxDoc and SurgeryTech can be viewed on iPhone, and if so, which App should be downloaded in order to access them. Therefore, individuals of RULE\_1\_RESULT subclass of RECOMMENDED\_SOURCES class (RelaxDoc and SurgeryTech) will be moved (inferred) into the RULE\_2\_RESULT class by Rule 2 only if they can be accessed through iPhones. Figure 5.8 shows that only SurgeryTech satisfies

this criterion. This answers CQ1. The same Rule 2 also determines that suggested iPhone App, which will enable access to SurgeryTech would be Download\_All\_Pro (see Figure 5.9). Therefore, individual Download\_All\_Pro is being moved from the THIRD\_PARTY\_APPLICATION class (which is a subclass of the TECHNOLOGY class) into the SUGGESTED\_APPS\_FOR\_IPHONE class, which is a subclass of the RECOMMENDED-SOURCES class. This answers the CQ2.

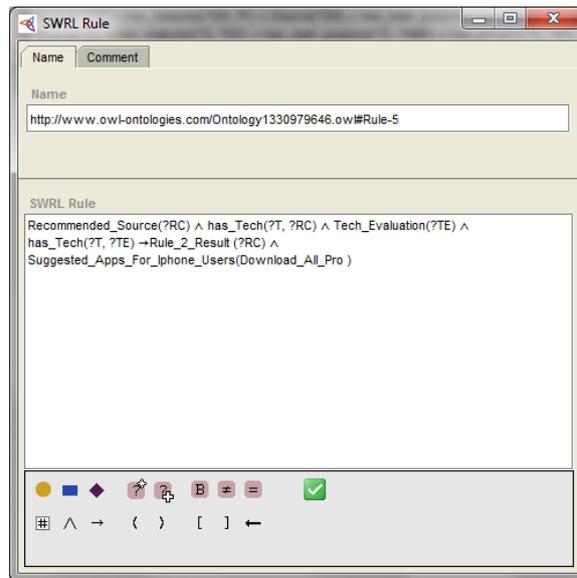


FIGURE 5.7 SWRL RULE 2 FOR SELECTING TECHNOLOGY THE CASE STUDY.

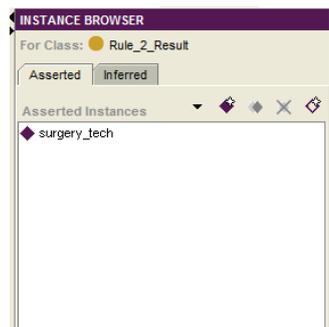


FIGURE 5.8 FIRST PART OF RULE 2 RESULT FOR THE CASE STUDY.

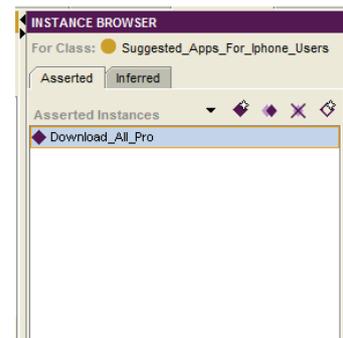


FIGURE 5.9 FIRST PART OF RULE 2 RESULT FOR THE CASE STUDY.

### 5.1.3 DISCUSSION ON THE ILLUSTRATION OF THE GENERIC MODEL

By modelling a) and b) in (5.1.2.1) above (i) and (ii) from section 4.1 are addressed. Furthermore, constructing the semantics of the environment and interpreting user preferences present in CQ1 and CQ2 the model attempted to avoid basing search results on past behaviour which is present in most IRS on the internet. The semantics of the environment is strengthened through the horizontal hierarchies present in the figure 5.1. This therefore, allowed maximum utilisation of characteristics of information sources stored in the OWL ontological model elements. Consequently addressed (1) – (3) from section 4.5. Furthermore, section 4.4 claimed that in order to infer new knowledge from the OWL ontological model, SWT developed a SWRL rules. In this section SWRL rules are utilised in the reasoning process to match the semantics of user preferences to deliver “a moment” of information sources on the internet. Consequently, reduce modern IO.

Another modelling decision present in this section is running two SWRL rules to answer CQ1 and CQ2. As described in section 4.4 (the reasoning process), SWRL rules reason upon semantics of information sources which are defined through a set of  $\mathcal{CS}_i$  classes and relationships imposed on both SOURCES and  $\mathcal{CS}_i$  through *has* –  $\mathcal{CS}_i$  and USER-PREFERENCES and  $\mathcal{CS}_i$  through the same relationship. Therefore, if *has* –  $\mathcal{CS}_i$ . Furthermore, in bullets 1-5 of the reasoning process it is clearly stated that selection of information sources is prioritised by number of Ops used to define user preferences. Therefore, if an information sources fulfils all user preferences except one the information sources will be excluded from being selected as an information sources which can also be viewed on iPhone.

Consequently, a layered processing of the same SWRL rule was an attempt to address both competency questions. Layering of SWRL rule was not experimented in any other problem domains which were modelled to examine the applicability of the proposed generic model in this research.

Another important aspect of illustrating the generic model is that according to user preferences as set of OPs are created. This to address user preference, the choice of OPs is controlled by how users are precise when defining requirements (as defined the reasoning process in section 4.4). Hence, if the user says or then inheritance is automatically applied on Features class and its sub-hierarchies. This means that all

characteristics which are stored in Features sub-hierarchies can be selected in reasoning process (described in section 4.3.2). Whereas, if the user say “and” this means that all preferences must be met and the OWL model must follow modelling principle present in figures 4.7, 4.8 and 4.9.

## CHAPTER 6. EVALUATION, CONCLUSION AND FUTURE WORK

In this section, we will provide a detailed evaluation of archived objectives in chapter 1 (see section 1.3). We will also revisit research problems itemised in chapter 4 (see section 4.1). We will further elaborate on decisions made when designing the proposed computational model in chapter 4 (see section 4.3). This section will also list some of the issues encountered in the process of investigating our research problem. Furthermore, we extend our evaluation by discussing the proposal and implementation phase of this research. We will then discuss some of the shortcomings in this research that could have been done differently. Therefore, the next subsections of this section will provide an evaluation of our research objectives, the proposed computational model, and a semantic selection of Internet sources through SWRL enabled OWL ontologies to reduce modern IO.

### 6.1 REVIEW OF RESEARCH OBJECTIVES

In chapter 1, we provided a detailed description of our research objectives. In this section, we wish to remind the users of these objectives.

In (OB 1), we proposed defining and creating a computational model that will address IO in modern retrievals, i.e. Internet searches, and take into account (i)-(iii) and (a) (see section 4.1). Therefore, the model must clearly define what is needed for computations to be performed and which output we may expect from it (Almarri and Juric, 2013a).

In (OB 2), we aimed to give more power to users. This is because, in modern computations, users are the producers and consumers of information at the same time. Therefore, the proposed computational model in this research must demonstrate how and where we give power to the user, in order to address the problems of finding the most relevant search result in a particular moment. (see section 5.2)

Our third objective (OB 3) aimed to address “a moment” in a particular situation when Internet searches happen. This was depicted in our generic computational model, by not saving the results of the selection of sources and just displaying them for that

moment in the recommended sources class. Therefore, this immediately implies that the model we proposed might not be interested or involved with “the past” and deletes the reasoning results to collect new semantics for the next moment. Furthermore, it would build computations based on capturing as much semantics as possible in that “moment”. The reason for that is simple: we do not want it to remember results of retrievals from previous “moments”, as this might not be advisable because (a) each “moment” carries different semantics in terms of the reasons and needs for retrieval and (b) users very often change their mind while searching the Internet. Consequently, we have to avoid storing the semantics of the results of past retrievals, because it may be wrong for the next “moment” (see sections 5.2.2 and 5.2.3).

In (OB 4), we choose to use the SWT stack because it allows us to interpret the meaning of a particular situation where retrievals happen. Furthermore, it enabled us to exploit a set of rich languages that give us opportunities to reason upon the semantics of the situation in a particular internet search, as mentioned in (OB 1). The flexibility of SWT stack as a technology of choice enabled a successful deployment of the computational model from (OB 1) above. Obviously the computational model deployed with the SWT would be a step forward in ensuring that we understand the environment where Internet searches happen. Therefore, we managed to model semantics of these environment where the selection of Internet sources happens, secure reasoning upon them, whilst taking into account objectives (OB 2) and (OB 3) above. Consequently securing that the user receives the most relevant search results according to the semantics of the environment and user preferences (see section 5.2).

In (OB 5), suggested to conduct an empirical investigation of the environment where a selection of internet sources happen. Furthermore, chapter 2 and its subsections provided a chronological order of IO, IRSs, RSs, Search Engine, their techniques, and search results. furthermore, investigated UB on the internet, to understand the role of information seekers/users role in the creation of excessive amount of information sources on the internet. This helped us in understanding the role of UB and its influence on searches on the internet (Almarri and Juric, 2014).

Based on the investigation present in chapter 2 a summary of data collection method is present below:

The advancement of CS and IS technologies resulted into revelation of information on the internet. Information seekers/users on the internet creates information instantly and effortlessly. This therefore triggered the need of methods that will help in controlling information storage, management, presentation and retrieval.

In this research a generic computational model was proposed that will allow to model characteristics of information sources on the internet and reason upon them by interpreting the semantics of information seeker/user preferences in a certain domain of interest. Consequently, alleviate modern IO which is caused by technology and information seekers/users on the internet. hence, it was important to investigate techniques that are used to discover and retrieve information sources on the internet (search engines and RSs).

There are many ways to collect data on the internet. Search engines for example utilise many tools to collect data about things on the internet. These tools collect data from search logs, web crawling and spiders that index information sources on the internet. Search engines also collect data about information seekers/users from (click-through stored in information seekers/users sessions and query logs) on search results, websites analytical tools, advertising tools, email contents, browsing sessions and so forth. Google search engine for instance, most of their data are provided directly by information seekers/users of google services (google apps, google translation tool, google reader, user contact networks and so forth) in addition to the common techniques mentioned in the previous paragraph.

Whereas, RSs data are collected mostly from profiling information seekers/users on the system and interpret these profiles for future recommendations. These techniques were adopted from early attempts of data filtering of email contents in early 90's. All these methods of data collection improved through time to include hybrid techniques generated from combining techniques to produce a new one. Furthermore, RSs also depends on query and session logs for data collection. However, data collection techniques are exclusive to search engines and RSs and very difficult to obtain or use them.

The internet continually grow and the amount of information becomes extremely vast. New types of information sources keep surfacing which delivers new ways of information creation and representation. This growth negatively affected current IRS on

the internet. Consequently, information seekers/users suffer from modern IO. Hence, new techniques are required that can improve information discovery and retrieval on the internet.

Some of the attempts to enhance current IRS on the internet involved combining search engines and RSs techniques to create another category of hybrid techniques. The reason behind these attempts was the consequence of web 2.0 technologies. Service providers on the internet uses every possible method to keep a hold of their users. Therefore, they keep combining different type of trending applications on the internet in addition to their main services. For example, many service providers allow their user to communicate through blogs, forums, chatrooms, and so forth. All these means of communication allows information seekers/user to create information every time they access the internet. Consequently, the volume of information got out of control.

The development of all these hybrid techniques of information retrieval on the internet diminished the boundaries between search engines and recommender systems as two distinctive techniques of IRS on the internet. Therefore, it is very difficult to decide whether modern IO is the consequence of the high volume of information created by information seekers/users in daily bases or is the result of all these hybrid services and techniques delivered by service/product provided and available on the internet. Consequently, creating hybrid techniques which utilise search engines and RSs. Thus, a common characteristic of both search results and recommendations; all results are solely based on past behaviour of users on the internet.

The evaluate of the proposed generic model was done by:

- firstly, illustrate the proposed generic model in a particular domain of interest and give a clear definition of the model's components (OWL classes and constraints). This means that a detailed OWL ontological model with its classes, subclasses, and relationships must be modelled.
- secondly running the computations from the model (i.e. performing reasoning) upon domain specific OWL elements.

It is important to note that OWL models are always domain specific and therefore OWL classes, individuals and constraints may exist only when a domain of interest is clearly defined and selected. In all experiments to examine the applicability of the proposed generic model to deliver “a moment” of internet searches to reduce modern IO, an intensive investigation of the problem domain was conducted (Binghubash and Juric,

2011), (Almarri et al., 2012b), (Juric et al., 2013), (Almarri and Juric, 2013a), (Almarri and Juric, 2015) and (Almarri et al., 2016).

Early attempts in this research focused on understanding information sources on the internet in general and information generated using web 2.0 tools and technologies.

In Binghubash and Juric, we were motivated twofold:

- 1) Primarily wanted to find out if there is a mechanism of managing automatically potential matches between member's expectation when joining SNs and the range of SN "services" and "features" which are supposed to be available to their members.
- 2) Utilise the power of SWT and SWRL enabled OWL ontologies in particular, when deciding how to address this matching between potential member's expectations from SNs and the actual values given to them by SNs.

Hence, the first experiment in this research attempted to examine the applicability of the proposed generic computational model in the domain of social intensive environment on the internet "SNs" (Binghubash and Juric, 2011). This because SNs have shaken the way of how information seekers/users and service providers create, generate, access, process, disseminate and share data and information on the internet. Hence, they paved the way for information seekers/users towards a society that immerse more and more in social intensive environments, which generate data on the go (work, play, travel, study, entertain, teach etc).

The investigation on SNs, their features and services was conducted as follows:

First, search the internet for website or published academic research that lists existing SNs on the internet. The preliminary investigation aimed to collect general information on reasons of their existence. Some of the retrieved information from websites provided ranked SNs based evaluation and certain criteria. Whereas, academic research publications in the domain of SNs and white papers provided information on SNs sites, users, purpose they exist for and so forth. There was overlapping information in websites and published papers. Thus, this helped in confirming the truthiness of the contents of both websites and research papers. Therefore, the created spreadsheet combined data from both sources to described the characteristics of SNs according to their purpose, types, domains, services they offer, audience they target, types of members they attract, and many more. Hence, the research objectives at this stage attempted to understand:

which exact service SN memberships and collaborations may bring to their members,

- (a). which level of security and privacy is guaranteed to them and
- (b). which technical support members may need when using SNs.

In order to start thinking how to answer (a)-(c) above, it is necessary to consider the classification of SNs as the first step, but this is far from a trivial task.

At the time of investigating this topic it was almost impossible to find surveys of SNs, where their general features, types, purposes, services, privacy commitments and similar factors are clearly collected and used in categorisation of SNs. What is currently available is the information provided by websites that gives general statistics and surveys of “TopTen Reviews, 2011, Love To Know” through criteria such as:

- (i) Overall rating of ten popular SNs as “excellent, very good, good, fair and poor” and
- (ii) Evaluation of SNs by looking at “demographics, profile, security, networking features, search and technical help/support”.

The criteria in (i) and (ii) have sub categories which might give more granularities when analysing results of such surveys, and probably could then become a good starting point in SNs’ classification.

Hence, a refined information collected and used to construct the first ontological model which delivered information about the purpose of the SNs as follows:

- 1) Educational /Academic SNs: have a purpose of collecting members who share their joint interests in learning an exchanging knowledge. Their members enjoy group membership either in real life or in virtual class rooms.
- 2) Hobbies SNs: have a purpose of collecting members with similar interests in life and very often offer membership to online communities connected though specific hobbies, skills, interests sand activities.
- 3) News/ informational SNs: have a purpose of collecting members who are interested in finding information about specific topic, commenting on articles, sharing knowledge and posting questions that they need to get an answer for,
- 4) Professional SNs: have a purpose of collecting members who are interested business relations relevant to their professions, which in turn mat result in improve customer relationship and increase business efficiency.

The spreadsheet also contained information about members’ roles, features and so forth (see appendix A for detailed description).

The first attempt was primitive and answered questions such as:

- Which SN would be suitable for a potential member if he/she specifies exactly what the purpose of his membership would be (i.e. the purpose of SN)
- Which features SN should have, and how the security and privacy is handled in such SNs?
- Members also have to specify their role within the SN and indicate their preferences in terms of receiving technical help and support when using SNs.

This process was followed in two more experiments (Almarri et al., 2013) (Juric et al., 2013) which attempted to model domain-specific problems according to situations when information seekers/users search for information on the internet.

Furthermore, competency question (see section 5.1.2) illustrates how the collected information on SNs are modelled to exploit the semantic of the problem domain through SWRL enabled OWL ontologies to retrieve information based on user preferences in a moment of internet searches.

Experiments in this research also investigated problems of information retrieval in relation to LLL process in the domain of healthcare. Thus, utilised findings from the previous investigation on SNs and were enhanced with data of information sources on LLL in healthcare on the internet. However, the investigation on Healthcare SNs opened doors to new research problems which required investigating information sources that allowed a mashup of multiple technologies and variety of information seekers/user from different competence level. Therefore, this experiment attempted to address the issue of choosing information sources for life-long learners To this point the proposed generic computational model showed high re-usability regardless of the problem domain (Almarri and Juric, 2013b).

later attempts to investigate information sources on the internet in this research took another direction. Mahmood et al. attempted to analyse suicidal contents posted by member of microblogging systems and have negative influence on other members of the system. Hence, this attempt utilised data extracted by OCR tool (2009). The extracted data was organised based on topic relevance, account owner, and criteria of level of how crucial these posts can be when viewed by depressed members. Hence, this experiment aimed to deliver an alert system based on the characteristics of posts (Mahmood et al., 2013).

Other attempts to model information sources on the internet examined the retrieved search results from five search engines. The appendix provides a sample of the collected data from search engine queries. The process in this attempt involved interviewing a student who wish to join piano lessons. Hence, search the internet for relevance information sources which can provide information on location, number of hours, price and so forth. Hence, a search was conducted to examine search results based on collected description of the student preferences. The same query was typed into five different search engines to:

- major relevance in search results.
- frequency of appearance of certain information sources in the five search engines.
- Analyse how search engines rank search results.

Hence, based on the retrieved results from information seekers/users queries a categorisation is followed to group retrieved results based on relevance to search results.

In the process of characterising information sources, it was important to investigate the role of information seeker/user of these sources. Therefore, Almarri and Juric, investigation explored possible ways to understand information seeker/user behaviour on the internet. thus, the research on UB attempted to itemise characteristics of how information seekers/users search the internet to discover information sources (Almarri and juric, 2015).

## 6.2 RESEARCH EVALUATION/ IMPACT

This research according to (OB 1), proposed a new type of computational model that will help information seekers/users to select relevant information sources and deliver search results based on reasoning upon the semantics of the environment where internet search happens. Therefore, the model illustrated in Chapter 5 shows how (OB 1) was addressed. Firstly, the proposed model illustrates that it is possible to select relevant information sources from the reasoning process due to semantic overlapping between users' preferences and the description of information sources (see section 4.1.3). This means that the way preferences are formulated must be semantically similar to the way information sources from search results are characterised. The technology from the SWT stack guarantees that we can create a computational model which will deliver relevant information sources in search results: input to the model would be the semantics of user

preferences and characteristic of information sources from search results (see illustration of the proposed generic model in chapter 5) and the output from the model will be a selection of relevant information sources from search results according to the semantic of user preferences. It is important to note that

- (a). We cannot talk about “computations” behind our model if the model is not domain specific, because of the nature of technologies used when deploying the model (all OWL ontologies must be domains specific);
- (b). Case studies show how our proposed model changes when we deploy it: Figures 5.1 and 5.10 show how the proposed OWL model changes according to the semantic of the environment when Internet searches happen;
- (c). The semantics of the environments where Internet searches happen are captured through the  $CS_i$  classes and the constraints of  $has\_CS_i$  see sections 4.3.1. and 4.3.2)
- (d). The way we describe the semantics of the environments where Internet search happens guarantee semantic overlapping between user preferences and search results because of the reusability of  $has\_CS_i$  constraints (see section 4.3.3), which is underpinned in Figure 4.10.

Due to (a) – (d) above, we can conclude that the input to our computational model is a set of users’ preferences, which can be captured through various interfaces available today. They can feed our ontological model by using numerous tools, which transfer persistent data into an ontological model (Golder and Huberman, 2005).

The output of our computational model would be a list of relevant search results, selected through the reasoning process based on SWRL (see sections 5.2.2 and 5.2.3).

We have a freedom of expressing users’ intentions and expectations, as mentioned in (a) (see section 4.1) through the options given in the set of  $CS_i$  classes, constraints and inheritance, as described in Figures 4.7, 4.8 and 4.9 (see section 4.3). Consequently, (OB 2) has been met: the model is based on user’s decisions on *what is relevant in a particular Internet search* for two reasons:

- (I) The user supplies the semantics which generate  $CS_i$  classes and  $has\_CS_i$  constraints (see both case studies) (see sections 5.2.2 and 5.2.3). Selections of Internet searches may require a different set of  $CS_i$  classes and  $has\_CS_i$  constraints in order to supply the ontological elements involved in the reasoning.
- (II) The characteristics of the results of Internet searches before their selection takes place should semantically overlap with the way we describe user preferences (see Figure 4.10) and therefore user input is also essential here.

Both (I) and (II) above pave the way towards a higher relevance of the selected search results because of the user's role in the selection.

In (OB 3), we claimed that we should address a "moment" in Internet searches and avoid exploiting past user behaviour when capturing the semantics of the environment when Internet searches happen. We give rationale, which explains how we achieved it.

- (A) The reader should note that in both case studies, we have to delete the result of reasoning.
- (B) Our SWRL moves (copies) individuals of OWL classes across the ontology as the result of reasoning. Consequently, these moved individuals have to be erased from the classes they are moved to, before a new reasoning take place. Therefore, we do not wish to remember the result of previous reasoning.
- (C) OB 2 almost dictates that we address a "moment". As we said in (II) above, the selections of Internet searches may require a different set of  $CS_i$  classes and has  $CS_i$  constraints in order to supply the ontological elements involved in the reasoning. We cannot assume that the semantics of one "moment" in Internet searches can be copied to another, because everything depends on how the user expresses his/her preferences and the intentions he/she may have (i.e. everything is in user's hands).

For proving that we have achieved OB 4 and OB 5, we suggest that the reader pays attention to all our publications (Binghubash and Juric, 2011), (Almarri et al., 2012b), (Almarri et al., 2013), (Almarri and Juric, 2013a), (Mahmood et al., 2013), (Juric et al., 2013) and takes into account the research output from (Kataria, 2011), (Shojanoori, 2013), (Chau et al., 2012), (Juric et al., 2012). In this research our two case studies underpin claims for OB 5, but there are other numerous experiments and examples where the proposed model has been implemented across problem domains, which can be found in our other publications and available upon request. Therefore, we are confident that our proposed model works and it remains to be seen if it will have further implications on the way we search Internet.

### **Controlling the performance of the proposed generic model through imposed restrictions on the OWL concepts, constraints and instances:**

It is common that enforcing restrictions on SE solution is a good practice, this to solve many common errors a system may have. In this research, it was very important to follow these rule and common practices were followed to keep maximum integrity of the solution. Therefore, some basic rules which were followed are

Quantifiers restrictions was used to apply some basic rules which indicates that any concept in the OWL model must have “some values from” through a relationship between individuals of two classes. In the illustration of the proposed model I also attempted to enforce cardinality restriction. However, this type of retraction did not work well along with the concept of “n” number of concepts and relationships. Therefore, it was omitted from further experiments.

Some other common practices followed in the implementation of the generic computational model is enumerated class expression was applied to two main classes of the OWL ontological model to indicate that members (individuals) of these two classes are distinct. This therefore, clarified the confusion that if two classes share relationships are assumed to be in a subsumption relationship.

**Important comments on the selection of generic model:**

- (I). The user role in the proposed generic model is very important. This means that the extracted information from the user’s competency question controls (1) - (4) listed in section 4.4. To be able to design the domain of interest correctly and create “a moment” of internet searches it is important to extract as much as possible domain knowledge from the competency question.
- (II). The categorisation of classes and their sub-hierarchies is dependent on how the domain of interest is analysed. Again, this step is also dependent on how the user described their preference but it also depends on how these preferences are interpreted. Whether the domain of interest is designed to include essentials and very precise information or to be more flexible and generalise all concepts in a given domain; depends on how each person interprets the semantics of the environment.
- (III). Some of the important aspects of the generic model is inheritance of proprieties of the generic model main components (classes and object properties).
  - (a). What happens when one of the four versions of the generic model is selected for a certain domain of interest.
  - (b). Pressure on relationships (object properties) how do they allow the selection of common characteristics of information sources according to user preferences.

(c). The relationship automatically converts from and to or in inheritance.

Which is not really good.

(IV). Although the proposed generic computational model aimed to promote simplicity of design. Most of the work is done in the background as explained in the reasoning process in (section 4.4).

(V). What happens next?

#### **Problems when interpreting user preferences from the competency question.**

Mostly information seeker/user just provides general information, this leads to ambiguity when describing information sources. Consequently, the selected information sources will have some concepts which answers to the user preferences (user queries). on the other and if the user provides a detailed description of preferences, it will help in constructing semantically rich OWL ontological model that will fully describe the semantics of the domain of interest but it will be very difficult to fulfil user's preferences. This means that the only option to match user preference and semantics of the information sources is to deploy an OWL ontological model based on the generic model depicted in figure 4.9 which stress the use of horizontal hierarchies and their constraints -which forces a one to one relationship- between both sources class and CSi classes and User-Preferences class and CSi classes (not recommended).

Inheritance of relationships is a complementary advantage of the proposed generic model. It can be useful because it can interpret shallow user preferences and deliver relevant information sources. However, peculiar and precise user preferences might end up facing IO. this is justified due to how SWRL rules will interpret the semantics of user preferences. For example of the user defines many features that utilise the same OP, the reasoning process will halt querying characteristics defined through the same OP as soon as one condition is fulfilled. Therefore, the information seeker/user will be forced to filter results according to how they are relevant to the search query.

#### **An alternative direction in this research:**

Another attempt in modeling the semantics of the environment can be done through the modeling of User Behaviour on the internet. In Almarri and Juric, 2015, the focus was directed towards modelling the semantics of user behaviour (UB) through OWL concepts, and reason upon them in order to address IO (Almarri and juric, 2015). By

focusing on users and their (intention), it was possible to come slightly closer to the issues of modelling of user behaviour on the internet. The proposed model in this experiment is slightly different from the original idea in the proposal in this research but can be a complement. The experiment takes on board the task of characterizing users behaviour on the internet and tries to capture them. This to describe the semantics of user preferences, which are often interwoven with their perceptions and expectations, and which dictate how user behaves when searching the Internet. Therefore, this attempt was to get closer to capturing constant changes in the way users use the Internet and perform searches on the internet.

One common problem in all attempts to understand and to illustrate UB in cyberspaces is that almost all solutions which claim to capture UB always relay on “what happened in the past”. Therefore, this encouraged me to look at this issue closely in parallel with previous attempts to model the characteristics of the environment where internet searches happen. As a fact, it is a common practice to collect data about UB through the analysis of system logs because they are an ideal source of information and cannot be ignored if we wish to understand user behaviour on the internet. However, if we do not take into account that (a) users may change their mind while searching the internet and (b) they can be distracted at the same time by various browser and search engine functionalities, then we will definitively not be able to offer a significantly different solution to IO caused by internet searches. In other words, if we do not model UB with (a) and (b) in mind, then we will not offer anything new in resolving the problem of IO in Internet searches.

In this research, it is core to provide a moment of internet searches to the users. Hence, most of the experiments evolved around the idea of capturing the semantic of the “moment” when a particular internet search happens and ignore everything else, which might not be directly related to this particular search. Furthermore, because in all experiments in this research it was important to avoid remembering what previous searches have brought to the same user and insist that their semantics remains present in our models. If we allow for storing “old” semantics from previous searches, we will not be able to handle the changes imposed by users and give the user the chance to be able to change their mind!

The ontological model in this experiment based on the figure below, use a generic model introduced in (Almarri and Juric, 2013b) where user preferences and source given by Internet searches (very often containing URLs) are described though characteristics of sources (CSi classes) and the way we describe user behaviour (UBi classes).

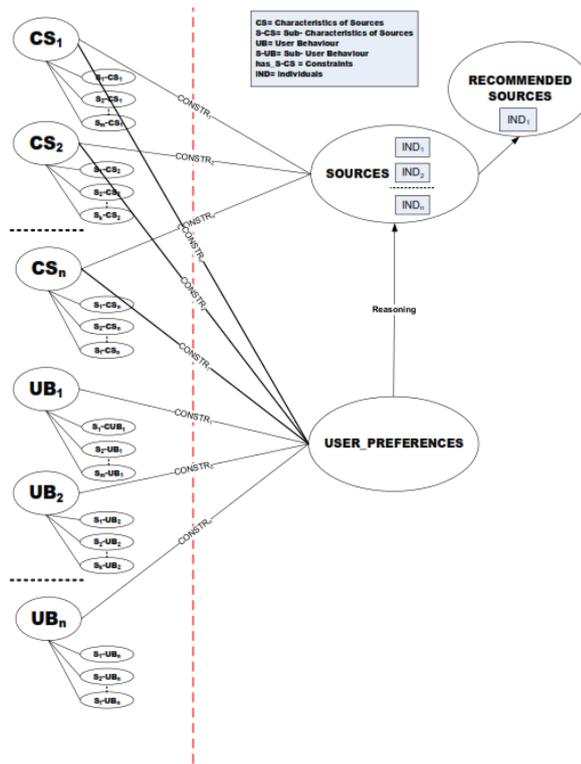


FIGURE 6.1 SEMANTIC OVERLAPPING THROUGH INDIVIDUALS

Like all past experiment, the above model in figure 6.1 must be transformed into domain specific environment in order to reason upon the semantics of both the user and the domain of interest. Thus, there must be

- a) A set of individuals, which may populate the SOURCE class
- b) A set of Object Properties, which we should define according to the generic model form Figure above 6.1.

Hence, semantic overlapping was present in a different method. This method aimed at experimenting the method of overlapping through meanings of knowledge extracted from user's query described in competency question with synonyms of words. The experiment was successful because results of the reasoning process provided an information source as a recommended source at the end of the process. Therefore, the

model guaranteed semantic overlapping through instance of the domain rather than semantic overlapping through the re-use of constraints in the semantic model.

## CHAPTER 7. CONCLUSIONS AND FUTURE WORK

### 7.1 RESEARCH CONCLUSION

This research, analysed the problem of IO and overviewed approaches to addressing it. It also focused on systems/techniques which claimed to alleviate IO, but it seems that they are not very suitable for and successful in addressing the major issues with modern IO today. The reasons are numerous but two of them are obvious:

- (a). IO has changed in the last decade, and
- (b). the way information seekers/users retrieve information in order to alleviate IO has moved on from using classical IRS to the deployment of powerful search engines in our everyday lives.

Furthermore, in current information age, IO is more related to the

- (i) vast amount of information on the internet,
- (ii) the results of internet searches delivered by search engines, and
- (iii) information seekers users who create information as soon as they are on the internet.

Therefore, information seekers/users are producers and consumers of information sources today, which is likely to be available through the internet, and the results of internet searches are likely to create a new IO in our lives. On top of that there are no available mechanisms of expressing users' preferences when either creating search engine algorithms or filtering/ranking search results, as pointed out in (a) (see section 4.1). Whichever way user's intentions and expectations were modelled in the reviewed literature, they very often rely on user profiling, their past behaviour, habits and similarities of a particular user's profile with other people. However, in order to understand user's intentions, expectations and demand it is important to focus on the semantics of the environment where internet searches happen. This may be a very complex task, but the semantics should be collected and interpreted before search engines "decide" on what exactly information seekers/users want when searching the internet.

This can reduce modern IO because it will allow modelling information seekers/users' intentions, expectations and demands in a particular situation. Hence, retrieve more relevant results of internet searches.

During the time this research was conducted, it was not possible to find any research available that solely focuses on: *how do we select or even get relevant internet search results when using modern search engines?* Currently, the queries information seekers/users impose on search engines, and the ranking of internet search results, given by search engines, might produce IO for two reasons: internet search results may be irrelevant to the queries and information seekers/user may have too many of them. furthermore, add to this problem to the fact that it is sometimes difficult to form queries, it is no wonder that we have a very low level of relevance of internet search results that in turn might not satisfy information seekers/users expectations.

## 7.2 FUTURE WORK

We believe in a software application which houses our proposed computational model. The issues which are solely associated with its implementations are as follows:

- I) The feeding of our ontological model could be done through either
  - a) modern and intelligent interfaces, which would allow users to input their intentions and expectation and infer our preference and  $CS_i$  classes or
  - b) a set of drop-down menus in traditional UI which determine the format and content of users' input

At the moment, we have manual input into our ontology in order to prove that the proposed ideas and concepts work, but it is possible to direct the results of this research towards highly pervasive environments, where voice and multimedia inputs are equally welcome. In such cases we would need an interdisciplinary team which would connect hardware and UI advances with our application implementation.

- II) We solely placed our application upon the Internet search results in order to make them more relevant to the user. However, we should assess how efficient the same model can be within any search engine in order to improve current ranking. In principle we do not mind using the model upon ranked results because we can select more relevant results. However, what would happen if

the result of ranking does not give a single relevant result? We will have nothing to select. This is an unusual situation, but it may happen. Therefore, our model should be completely effective as a part of any search engine. We hope that search engine companies would show interest in this work. Without knowing their exact search algorithms, we cannot predict where to place our proposed model within search engines.

Chapter 2 of this thesis provided the reader with an overview of state of the art in main areas of this research (see sections 2.1-2.6). We also elaborated on technologies that directly participated on the advances of our research problem. In chapter 4 (see section 4.2-4.5), we itemised the common problems of early solutions of selection of information sources on the Internet through a variety of technologies, in general and domain specific solutions. Consequently, we struggled to find a proper solution which directly solves the problem of modern IO caused by advances in technology. Our related work in chapter 3 provided a detailed description of research which participated in solving a part of objectives in this research.

**Thank you for reading**

# APPENDIX

	Facebook	YouTube	LinkedIn	Advertiserbay	Twitter	MailChimp	BaseCamp	Google Plus+	MySpace	Ning
<b>2 Target Audience</b>										
3 13-17	✓								✓	
4 18-25		✓			✓			✓		✓
5 26-32				✓						
6 32+			✓			✓				
<b>7 Location</b>	✓	✓	✓	✓	✓	✓		✓	✓	✓
<b>8 Interested groups/topics</b>	✓	✓	✓	✓	✓		✓		✓	✓
<b>9 Sign in/Register</b>										
10 Public page	✓	✓			✓				✓	
11 Professional Pages	✓	✓	✓	✓	✓	✓	✓	✓		✓
12 Free to sign in	✓	✓			✓			✓		
13 Registration	✓	✓	✓	✓		✓	✓		✓	✓
14 Accessable (do u need to sign in before you can view contents of the SN)			✓	✓		✓	✓	✓		✓
<b>15 Collaboration Tools</b>										
16 Request/invite users	✓		✓		✓			✓	✓	
17 Events	✓		✓		✓	✓	✓		✓	✓
18 Message (private inbox)	✓		✓					✓	✓	
19 Chat	✓							✓	✓	
20 Advert	✓	✓	✓	✓		✓		✓	✓	✓
21 Feedbacks	✓	✓	✓		✓			✓	✓	
22 Post comments	✓	✓	✓		✓				✓	
23 Music videos	✓	✓							✓	
24 Tags	✓	✓			✓				✓	
25 Upload/Copy/Share Photos	✓	✓	✓	✓			✓	✓	✓	
26 Instant Message	✓							✓	✓	✓
27 Groups	✓					✓	✓	✓	✓	✓
28 Create groups	✓					✓	✓		✓	
29 Additional functionality (games, video)	✓									
<b>30 Media Tools</b>										
31 Blogs					✓					
32 RRS feeds				✓						
33 Forums							✓	✓	✓	
34 Video sharing		✓		✓				✓	✓	
35 Wikis										
36 eMail						✓		✓		
37 Pod cast		✓								
38 Social network	✓		✓		✓				✓	✓
<b>39 Advertisement</b>										
40 Free	✓	✓								
41 Buy space			✓	✓		✓	✓	✓		✓
<b>42 Requires a structure for ads</b>										
43 layout for ad (choose where to place	✓	✓	✓	✓		✓		✓	✓	✓
44 Approx. of uploading content	✓	✓	✓	✓		✓	✓	✓	✓	
45 Running time (duration of ad)	✓	✓	✓	✓		✓	✓	✓	✓	
46 Able to customise the adverts				✓						
<b>47 Pre-request to be able to ad</b>										
48 Sign in	✓	✓	✓	✓		✓	✓	✓	✓	✓
<b>49 Similar techniques of market</b>										
50 Marketing the segment	✓	✓	✓	✓		✓		✓	✓	✓
51 Monitoring the activities	✓	✓		✓				✓	✓	
52 Marketing plan	✓	✓	✓	✓		✓	✓	✓	✓	✓
53 Marketing research	✓	✓	✓	✓		✓	✓	✓		
<b>54 Search Engine</b>										
55 Bookmark pages	✓	✓			✓			✓		
56 Search by email/names	✓	✓	✓		✓	✓		✓	✓	✓
57 Browse with membership		✓	✓	✓	✓	✓	✓			✓
58 Rating	✓	✓			✓			✓	✓	
<b>59 Additional Support</b>										
60 FQA	✓	✓	✓	✓				✓		
61 Technical advise	✓	✓	✓	✓		✓	✓	✓	✓	✓
62 Privacy settings	✓	✓	✓	✓			✓	✓	✓	✓

Social Network	Sermo	doc2doc bmj	Doctors Hangout	DocCheck UK	Relaxdoc	DermRounds	BioCrowd
<b>Purpose</b>							
1. phisicians	yes	yes	yes	yes	yes	yes	yes
2. Patients	No	yes	No		no	no	no
3. Caregiver		yes			no	yes	no
4. Knowdelge	yes	yes	yes	yes	yes- medical student	yes	yes
<b>Search</b>	1. customized Search 2. Save Search 3. Analyze Search Results 4. Phsycians	1. Name 2. Country 3. Speciality 4. Blog contents 5. Forum discussion contents 6. Job offers 7. Services 8. website pages 9. Fourm 10. Files 11. Events 12. news items 13. Images 14. Author	1. Phsycians 2. Groups 3.blogs 4. Events 5. Video 6. Images		1. Medical journals 2. Jobs 3. medical news 4. medical organisations 5. members	1. Members 2. Blogs and news 3. Images 4. Groups 5. Forums 6. jobs	members
<b>Services</b>							
1. News bullets	yes	Yes	yes		yes	yes	
2. Reports	1.Event Reports 2. Pharmaceutical Reports 3. Drug Report 4. Therapeutic areas Reports 5. Disease Reports	1. Jobs	No		yes		
3. View Members List			yes		yes	yes	yes
4. Monitor Descussions	yes		yes		yes	yes	yes

5. Medical Products	yes		No		yes-medical supplies	no	yes
6. New Research updates	yes		No		yes	yes	
7. Events	yes		yes		no	yes	
8. Awareness programs	yes		Yes		no	no	
9. Drug Awareness	yes		Yes		no	no	
10. Premium product services	yes		No		no	no	
11. Announcements	yes		No	yes	yes	yes	
12. promotional Activities	yes		Yes			no	
13. Posts	1. All Catigories 2. Speciality 3. Most popular 4. listing		No		no		
14. Create posts	yes		No		no		yes
15. follow posts	yes		No		no		no
16. Authored Posts	yes		yes				
17. Survey	yes		No				
18. Statistical Results	yes		No				
19. disseminate Disease Info	yes		yes				
20. On news	yes	yes	yes	yes			
21. Panels	yes		no				
22. Tips	yes	yes 1.Job tips	no	yes	no		
23. Job offers	Not mentioned		yes 1. by years experience 2. by Location 3. By catigories	yes	yes by speciality	yes	yes

24. Sort contents	yes	1. Title 2. Date	1. recent jobs 2. Physicians 3. Groups 4. Latest Posts 5. Most popular 6. Random Sort				
25. Post Jobs			yes		no	yes	yes
26. Events Calender			yes	yes	yes	yes	no
27. Tagging			yes 1. Image 2. Discussion				no
28. E-Library	no	yes	no				no
29. Recommendations	no	yes	no				
30. E-Books		yes	yes				no
31. Journal		yes	No		yes		no
32. Job listing		yes				Yes, depending on	yes
33. Classifieds		yes			1. employment opportunities 2. personals 3. help wanted 4. for sale 5. trade/exchange 6. for rent	no	
<b>Availability</b>	US only	two types of access Free or purchased access for journals 1.free access list <a href="http://group.bmj.com/products/journals/countries-with-free-access">http://group.bmj.com/products/journals/countries-with-free-access</a>	Worldwide	1.available in 30 countries (locations) Free entrence for healthcare professionals		worldwide	

<b>Role</b>							
1. Doctor	yes	yes	yes		yes	yes	yes
2. Patient	No	yes	No		no	no	yes
3. Journals	No	no	yes		no		
4. Industry and Professional Organizations	Yes		yes		no	yes	
5. Medical student	Not mentioned	yes	not mentioned		yes	yes	yes
<b>Website</b>	<a href="http://sermo.com">http://sermo.com</a>	<a href="http://doc2doc.bmj.com/">http://doc2doc.bmj.com/</a>	<a href="http://www.doctorshangout.com/">http://www.doctorshangout.com/</a>		<a href="http://www.relaxdoc.com/">http://www.relaxdoc.com/</a>	<a href="http://www.dermrounds.com/">http://www.dermrounds.com/</a>	
<b>NetWork Services</b>							
1. Smart devices version	Yes	yes	yes				
2. Customize pages	yes	yes	yes				
3. Video	yes		yes			yes	yes
4. Rating Topics	yes		no				
5. Polls	yes	yes	no				
6. Discussion Board	yes	Yes	yes		yes	yes	
7. Blogs	yes	yes	yes		yes	yes	yes
8. Groups	yes		yes		yes	yes	
9. Create Groups			yes		yes	yes	
10. Fourm	No	Yes 1. doctors only fourm (access restricted)	yes		no	yes	
11. Images			yes		yes	yes	yes
12. Advertising	no		yes		yes-allow advertising through sponsorship	yes	yes
13. Chat			yes		yes	yes	yes
14. Edit Profile	yes	yes	yes		yes	yes	yes
15. Upload photos			yes		yes	yes	yes
16. Personal webpages			yes			yes	

17. Add Blog		no	yes		yes	yes	yes
18. Add Videos		no	yes			yes	yes
19. Add Images		no	yes		yes	yes	yes
20. Featured Contents		1. Featured Jobs	yes 1. Groups 2. Images 3. Members list				
21. Facebook page	yes	no	yes		yes	yes	yes
22. Twitter	yes	no	yes		no	yes	yes
23. Orkut		no	yes		no	no	no
24. Google Buzz		no	yes		no	no	no
25. LinkedIn		no	yes		yes	yes	yes
26. Froum Types		1. Members only fourm 2. Students Fourm 3. Open Clinical fourm 4. BMJ Fourm 5. Off Duty Fourns 6. Closed Clinical Fourms			n/a	1.General forum and introductions 2.Jobs and career opportunities 3.Suggestions and feedback 4.Residents and fellows 5.Medical students 6.Attending	no forums
<b>Security</b>							
1. Report an issue	yes	no	yes			yes	no
2. Hide personal details	yes	yes	yes				no
3. Privacy settings							no
4. Block users							no
5. Report abuse							no
6. Safety Tipe							no
7. Account deletion (diffeculty 1-5)							

<b>Privacy</b>							
1.profile access permission	yes	yes	yes		yes		no
<b>Type of Technical</b>							
1. office address	yes	no			yes		no
2. phone no.	yes	yes			yes		no
3. email	yes	Yes			yes		yes
4. FAQ	yes	yes			no		no
5. most closest topic form	yes						no
6. contact us form (for further details)	yes				yes		no
Note:	1.Reports from 1-5 are non-membership access 2. registration is only for doctors and healthcare business related				only physicians and medical students allowed to access this website		

Search Engine	Piano Lessons London	Relevant?	If No, Why?
	URL		
Google	<a href="http://www.piano-composer-teacher-london.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31Q6EHbjKAULzGSp7yKo3czkKOHmKodbM4TYaWjCPimsaAmeQ8P8HAQ#!piano-lessons-london/c911">http://www.piano-composer-teacher-london.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31Q6EHbjKAULzGSp7yKo3czkKOHmKodbM4TYaWjCPimsaAmeQ8P8HAQ#!piano-lessons-london/c911</a>	Yes	
	<a href="http://www.londonpianolessons.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31Xfj984J2RcE0JTANA9Z6Zfeqk1k_zQJVrXrGzQ4cOMaAiji8P8HAQ">http://www.londonpianolessons.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31Xfj984J2RcE0JTANA9Z6Zfeqk1k_zQJVrXrGzQ4cOMaAiji8P8HAQ</a>	Yes	
	<a href="http://www.londonpianoinstitute.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31XeDVlJduEtzjPLS2tgthyUknJS4TPS-no3Mglyw4ecaAumX8P8HAQ">http://www.londonpianoinstitute.co.uk/?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31XeDVlJduEtzjPLS2tgthyUknJS4TPS-no3Mglyw4ecaAumX8P8HAQ</a>	Yes	
	<a href="http://alltalents.co.uk/Music_School_Hackney.html">http://alltalents.co.uk/Music_School_Hackney.html</a>	Yes	
	<a href="https://www.bidvine.com/services/piano-lessons?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31TEcaBnDCI-b1Emf8HW4DFkB01OqxyVachHahJi0GDygaAiZL8P8HAQ">https://www.bidvine.com/services/piano-lessons?gclid=Cj0KEQjw-tSrBRck8bzDiO_gbwBEiQAK-D31TEcaBnDCI-b1Emf8HW4DFkB01OqxyVachHahJi0GDygaAiZL8P8HAQ</a>	No	Provides different services
	<a href="http://www.musicosouth.co.uk/">http://www.musicosouth.co.uk/</a>	Yes	
	<a href="http://londonurbanartsacademy.org/creative-workshops/">http://londonurbanartsacademy.org/creative-workshops/</a>	No	Musical Workshop
	<a href="http://londonurbanartsacademy.org/">http://londonurbanartsacademy.org/</a>	No	Youth Workshop
	<a href="http://www.kijiji.ca/b-music-lessons/london/c8611700214">http://www.kijiji.ca/b-music-lessons/london/c8611700214</a>	Yes	
	<a href="http://www.genia-music.com/blog/category/adverts/">http://www.genia-music.com/blog/category/adverts/</a>	No	Piano album
	<a href="http://www.themusicbox.co.uk/teacher_adverts.html">http://www.themusicbox.co.uk/teacher_adverts.html</a>	No	
	<a href="http://london.vivastreet.co.uk/private+london">http://london.vivastreet.co.uk/private+london</a>	Yes	
	<a href="http://www.trinitylaban.ac.uk/student-experience/careers/careers-in-music/music-teaching-and-education-">http://www.trinitylaban.ac.uk/student-experience/careers/careers-in-music/music-teaching-and-education-</a>	No	Music school
	<a href="http://www.morleycollege.ac.uk/">http://www.morleycollege.ac.uk/</a>	No	College
	<a href="http://archive.friday-ad.co.uk/AdRef/RW515WSAF/FullAdDetails.aspx">http://archive.friday-ad.co.uk/AdRef/RW515WSAF/FullAdDetails.aspx</a>	No	Drum lessons
	<a href="http://www.franceinlondon.com/en-Consult-an-classified-ad-25615-Piano-lessons-by-a-Master-student-of-the-Royal-College-of-Music.html">http://www.franceinlondon.com/en-Consult-an-classified-ad-25615-Piano-lessons-by-a-Master-student-of-the-Royal-College-of-Music.html</a>	Yes	
	<a href="http://www.dailyinfo.co.uk/tuition">http://www.dailyinfo.co.uk/tuition</a>	No	School tuition
	<a href="http://www.piano-tuners.org/piano_teachers/London/Maida_Vale/1244-0.html">http://www.piano-tuners.org/piano_teachers/London/Maida_Vale/1244-0.html</a>	Yes	
	<a href="http://www.windsorpianoschool.co.uk/teachers.html">http://www.windsorpianoschool.co.uk/teachers.html</a>	Yes	
	<a href="http://www.northlondonmusicacademy.com/#!teachers/cf0u">http://www.northlondonmusicacademy.com/#!teachers/cf0u</a>	No	Music academy
<a href="https://login.yahoo.com/config/mail?&amp;.src=ym&amp;.intl=uk">https://login.yahoo.com/config/mail?&amp;.src=ym&amp;.intl=uk</a>			

Search Engine	Piano Lessons London	Relevant?	If No, Why?
	URL		
Bing	<a href="http://www.singinglelessons london.net/">http://www.singinglelessons london.net/</a>	No	Singing lessons
	<a href="https://hero.clickfunnels.com/optin?path=mm-method-video&amp;utm_campaign=Head%20Terms%20INTL&amp;utm_medium=cpc&amp;utm_term=piano%20lessons%20for%20beginners">https://hero.clickfunnels.com/optin?path=mm-method-video&amp;utm_campaign=Head%20Terms%20INTL&amp;utm_medium=cpc&amp;utm_term=piano%20lessons%20for%20beginners</a>	No	Online piano Lesson
	<a href="http://www.thetutorpages.com/piano-all-teachers">http://www.thetutorpages.com/piano-all-teachers</a>	Yes	
	<a href="http://www.onlinelessonreviews.com/piano-lessons/?utm_source=bing&amp;utm_medium=cpc&amp;utm_term=video%20lessons%20for%20piano&amp;utm_campaign=olr_gen_piano#.VXYW3FKbiCc">http://www.onlinelessonreviews.com/piano-lessons/?utm_source=bing&amp;utm_medium=cpc&amp;utm_term=video%20lessons%20for%20piano&amp;utm_campaign=olr_gen_piano#.VXYW3FKbiCc</a>	No	in location outside London
	<a href="https://curious.com/markalmond/series/piano-for-beginners?utm_source=bing&amp;utm_medium=cpc&amp;utm_campaign=Music_Piano&amp;utm_content=COURSE_markalmond_PianoBeg&amp;utm_term=playing%20piano%20chords">https://curious.com/markalmond/series/piano-for-beginners?utm_source=bing&amp;utm_medium=cpc&amp;utm_campaign=Music_Piano&amp;utm_content=COURSE_markalmond_PianoBeg&amp;utm_term=playing%20piano%20chords</a>	No	Online piano Lesson
	<a href="http://www.thepianostudio.co.uk/">http://www.thepianostudio.co.uk/</a>	Yes	
	<a href="http://piano-lessons-london.co.uk/">http://piano-lessons-london.co.uk/</a>	Yes	
	<a href="http://www.pianocourses.co.uk/">http://www.pianocourses.co.uk/</a>	Yes	
	<a href="http://www.arts-academy.co.uk/">http://www.arts-academy.co.uk/</a>	Yes	
	<a href="http://www.pianocourses.co.uk/piano-lessons-city-of-london/">http://www.pianocourses.co.uk/piano-lessons-city-of-london/</a>	Yes	
	<a href="http://londonpianolesson.co.uk/">http://londonpianolesson.co.uk/</a>	Yes	
	<a href="http://pianolessons.org.uk/">http://pianolessons.org.uk/</a>	Yes	
	<a href="http://www.keytomusic.co.uk/">http://www.keytomusic.co.uk/</a>	Yes	
	<a href="http://www.harmonica.com/lessons/?utm_source=bing&amp;utm_medium=ppc&amp;utm_campaign=hh&amp;utm_content=hh-bg-001&amp;utm_exp=17482697-26.MG2TiQwmSqGGtvl3WTtzyw.0">http://www.harmonica.com/lessons/?utm_source=bing&amp;utm_medium=ppc&amp;utm_campaign=hh&amp;utm_content=hh-bg-001&amp;utm_exp=17482697-26.MG2TiQwmSqGGtvl3WTtzyw.0</a>	No	Learn harmonica
	<a href="http://www.jazzpianolessons london.co.uk/">http://www.jazzpianolessons london.co.uk/</a>	Yes	
	<a href="http://mmq pianostudio-nw london piano lessons.co.uk/">http://mmq pianostudio-nw london piano lessons.co.uk/</a>	Yes	
	<a href="http://www.tutorhunt.com/piano.asp">http://www.tutorhunt.com/piano.asp</a>	Yes	
	<a href="http://www.learnpiano in30days.com/?hop=viewinfo">http://www.learnpiano in30days.com/?hop=viewinfo</a>	No	Online piano Lesson
	<a href="http://www.lovemusic today.com/piano/compare-piano-lessons.html">http://www.lovemusic today.com/piano/compare-piano-lessons.html</a>	No	Home study piano courses
	<a href="http://www.singtheeasyway.com/">http://www.singtheeasyway.com/</a>	No	Singing lessons
<a href="http://www.gumtree.com/all/london/piano+lessons">http://www.gumtree.com/all/london/piano+lessons</a>	Yes		

Search Engine	Piano Lessons London	Relevant?	If No, Why?
	URL		
Yahoo	<a href="http://www.singinglelessons london.net">www.singinglelessons london.net</a>	No	Singing Lessons
	<a href="http://www.PianoFool.com/Lessons">www.PianoFool.com/Lessons</a>	No	Online Piano Lesson
	<a href="http://www.gumtree.com/all/london/piano%20lessons">www.gumtree.com/all/london/piano lessons</a>	Yes	
	<a href="http://www.pianocourses.co.uk">www.pianocourses.co.uk</a>	Yes	
	<a href="http://www.londonpiano institute.co.uk/private-piano-lessons">www.londonpiano institute.co.uk/private-piano-lessons</a>	Yes	
	<a href="http://www.pianolessons-london.co.uk">www.pianolessons-london.co.uk</a>	Yes	
	<a href="http://pianolessons.org.uk">pianolessons.org.uk</a>	Yes	
	<a href="http://www.arts-academy.co.uk">www.arts-academy.co.uk</a>	Yes	
	<a href="http://preloved.co.uk/Piano">preloved.co.uk/Piano</a>	No	Sell piano lessons online
	<a href="http://MartWhiz.com/Piano-Lessons">MartWhiz.com/Piano-Lessons</a>	Yes	
	<a href="http://www.thetutortpages.com">www.thetutortpages.com</a>	Yes	
	<a href="http://curious.com/basic-piano-chords">curious.com/basic-piano-chords</a>	No	Online Piano Lesson
	<a href="http://www.Harmonica.com">www.Harmonica.com</a>	No	Learn Harmonica
	<a href="http://musiclessons london.co.uk">musiclessons london.co.uk</a>	Yes	
	<a href="http://www.keytomusic.co.uk">www.keytomusic.co.uk</a>	Yes	
	<a href="http://piano-lessons-london.co.uk/contact-us">piano-lessons-london.co.uk/contact-us</a>	Yes	
	<a href="http://www.kensingtonpianolessons.com">www.kensingtonpianolessons.com</a>	Yes	
<a href="http://www.dynamicpianolessons.co.uk">www.dynamicpianolessons.co.uk</a>	Yes		
<a href="http://www.gsokol.com/piano-lessons-london">www.gsokol.com/piano-lessons-london</a>	Yes		
<a href="http://www.tutorhunt.com/piano.asp">www.tutorhunt.com/piano.asp</a>	Yes		

Search Engine	Piano Lessons London	Relevant?	If No, Why?
	URL		
ZapMeta	<a href="http://piano-composer-teacher-london.co.uk/">piano-composer-teacher-london.co.uk/</a>	Yes	
	<a href="http://www.londonpianoinstitute.co.uk/Lessons">www.londonpianoinstitute.co.uk/Lessons</a>	Yes	
	<a href="http://www.alltalents.co.uk/Piano-Lessons">www.alltalents.co.uk/Piano-Lessons</a>	Yes	
	<a href="http://www.playgroundsessions.com/Piano">www.playgroundsessions.com/Piano</a>	No	Online Piano Lesson
	<a href="http://www.musicosouth.co.uk/">www.musicosouth.co.uk/</a>	Yes	
	<a href="http://www.gumtree.com/music-services/london/piano+lessons">www.gumtree.com/music-services/london/piano+lessons</a>	Yes	
	<a href="http://piano-composer-teacher-london.co.uk/">piano-composer-teacher-london.co.uk/</a>	Yes	
	<a href="http://www.londonpianoinstitute.co.uk/Lessons">www.londonpianoinstitute.co.uk/Lessons</a>	Yes	
	<a href="http://www.pianolessons-london.co.uk/Piano_Lessons_London.html">www.pianolessons-london.co.uk/Piano_Lessons_London.html</a>	Yes	
	<a href="http://takelessons.com/london-ky/piano-lessons">takelessons.com/london-ky/piano-lessons</a>	No	Not in London, in a City in Kentucky called London
	<a href="http://www.banksidekeys.org/">www.banksidekeys.org/</a>	Yes	
	<a href="http://www.piano-composer-teacher-london.co.uk/#!piano-lessons-london/c911">www.piano-composer-teacher-london.co.uk/#!piano-lessons-london/c911</a>	Yes	
	<a href="http://www.marccorbett-weaver.com/">www.marccorbett-weaver.com/</a>	Yes	
	<a href="http://www.markgoodwinpianos.co.uk/london">www.markgoodwinpianos.co.uk/london</a>	No	
	<a href="http://www.londonontario.worldweb.com/BusinessIndex/MusicLessons/">www.londonontario.worldweb.com/BusinessIndex/MusicLessons/</a>	No	not in London in Canada
	<a href="http://www.bespokemusic tuition.com/instruments/piano">www.bespokemusic tuition.com/instruments/piano</a>	Yes	
	<a href="http://www.jazzpianolessons london.com/#!">www.jazzpianolessons london.com/#!</a>	Yes	
	<a href="http://www.londonpianoinstitute.co.uk/private-piano-lessons/piano-programs-and-fees/">www.londonpianoinstitute.co.uk/private-piano-lessons/piano-programs-and-fees/</a>	Yes	
<a href="http://www.city-academy.com/piano-lessons">www.city-academy.com/piano-lessons</a>	Yes		
<a href="http://www.kijiji.ca/b-music-lessons/london/c861700214">www.kijiji.ca/b-music-lessons/london/c861700214</a>	Yes		

## REFERENCE

- AAOS. 2012. *Social Media Healthcare Primer, A Primer for Orthopaedic Surgeons* [Online]. Available: [http://www3.aaos.org/member/prac\\_manag/Social\\_Media\\_Healthcare\\_Primer.pdf](http://www3.aaos.org/member/prac_manag/Social_Media_Healthcare_Primer.pdf)
- ABOWD, G. D., DEY, A. K., BROWN, P. J., DAVIES, N., SMITH, M. & STEGGLES, P. 1999. Towards a Better Understanding of Context and Context-Awareness. In *Proceedings of the 1st International symposium on Handheld and Ubiquitous Computing (HUC '99)*. Karlsruhe, Germany.
- ACKOFF, R. L. 1989. From Data to Wisdom. In *Journal of Applied Systems Analysis*, 16, 3-9.
- ADDIS, A., ARMANO, G. & VARGIU, E. 2010. Using Progressive Filtering to Deal with Information Overload. In *Proceedings of the 21st International Workshop on Database and Expert Systems Applications (DEXA '10)*. Bilbao, Spain.
- ADOMAVICIUS, G., MOBASHER, B., RICCI, F. & TUZHILIN, A. 2011. Context-Aware Recommender Systems. *AI Magazine*.
- ADOMAVICIUS, G. & TUZHILIN, A. 2008. Context-Aware Recommender Systems. In *Proceedings of the 2nd International ACM Conference on Recommender Systems (RecSys '08)*. Lusanne, Switzerland.
- ADOMAVICIUS, G. & TUZHILIN, A. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *Journal of IEEE Transactions on Knowledge and Data Engineering*, 17, 734 - 749.
- AGICHTEIN, E., BRILL, E. & DUMAIS, S. 2006. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*. New York, NY, USA.
- ALEMAN-MEZA, B., ARPINAR, I. B., NURAL, M. V. & SHETH, A. P. 2010. Ranking Documents Semantically Using Ontological Relationships. In *Proceedings of IEEE 4th International Conference on Semantic Computing (ICSC '10)*. Pittsburgh, Pennsylvania, USA.
- ALEMAN-MEZA, B., HALASCHEK-WIENER, C. & ARPINAR, I. B. 2005. Ranking Complex Relationships on the Semantic Web. *Internet Computing, IEEE* 9, 37 - 44.
- ALEXANDER, B. 2006. Web 2.0: A New Wave of Innovation for Teaching and Learning? *Educause Review*, 41, 32-44.
- ALEXANDER, B., BARRETT, K., CUMMING, S., HERRON, P., HOLLAND, C., KEANE, K., OGBURN, J., ORLOWITZ, J., THOMAS, M. A. & TSAO, J. 2016. Report from the Information Overload and Underload Workgroup. In *Proceedings of the 1st Open Scholarship Initiative Proceedings*. Virginia, USA.: George Mason University.
- ALIWEB. 1997. *Aliweb* [Online]. Available: <http://www.aliweb.com/> [Accessed 12th September 2015].

- ALLAN, B. 1997. Information Needs: A Person-In-Situation Approach. *In Proceedings of the International Conference on Research in Information Needs, Seeking and Use in Different Contexts (ISIC '97)*. Tampere, Finland.
- ALMAMI, E., JURIC, R., AHMED, M. Z. & DABBOUR, M. 2015. Exploring Owl Models for Creating and Sharing Knowledge in Education Environments. *In Proceedings of 20th International Conference of the Society for Design and Process Science (SDPS' 15)*. Dallas, TX, USA.
- ALMARRI, B. H. & JURIC, R. 2015. Modeling User Behaviour When Addressing Information Overload. *In Proceedings of 20th International Conference of the Society for Design and Process Science (SDPS' 15)*. Dallas, TX, USA.
- ALMARRI, B. H. & JURIC, R. 2014. Modern Information Overload. *In Proceedings of 18th International Conference of the Society for Design and Process Science (SDPS '14)*. Kuching, Sarawak, Malaysia.
- ALMARRI, B. H. & JURIC, R. 2013a. Generic OWL Enabled Model for the Selection of Online Learning Sources. *In Proceedings of 17th International Conference of the Society for Design and Process Science (SDPS' 13)* São Paulo, Brazil.
- ALMARRI, B. H. & JURIC, R. 2013b. Generic model for OWL/SWRL enabled computations for the extraction of online contents. *In Proceedings of 17th International Conference of the Society for Design and Process Science (SDPS' 13)* São Paulo, Brazil.
- ALMARRI, B. H., JURIC, R. & MUGHAL, B. 2016. Semantic Selection of Healthcare Apps. *In Proceeding of the 49th Hawaii International Conference on System Science (HICSS 49)* Kauai, Hawaii, USA.
- ALMARRI, B. H., RAHMAN, T. & JURIC, R. 2012a. Semantic Recommendation of information Sources for Lifelong Learning. *Proceedings of the Society for Design and Process Science (SDPS '12)*. Berlin, Germany.
- ALMARRI, H. B., RAHMAN, T. & JURIC, R. 2012b. Semantic Recommendation of Information Sources for Lifelong Learning. *In Proceedings of 16th International Conference of the Society for Design and Process Science (SDPS' 12)* Berlin, German.
- ALMARRI, H. B., RAHMAN, T., JURIC, R. & PARAPADAKIS, D. 2013. Semantic Recommendation of Information Sources for Lifelong Learning. *Journal of Integrated Design and Process Science*, 17, 55-78.
- AMAZON. 2017. *About Amazon* [Online]. Available: [https://www.amazon.com/p/feature/rzekmvyjojcp6uc?ref =footer\\_aa](https://www.amazon.com/p/feature/rzekmvyjojcp6uc?ref =footer_aa) [Accessed 15th January 2017].
- ANAND, S. S. & MOBASHER, B. 2007. Contextual Recommendation. *In: BETTINA BERENDT, A. H., DUNJA MLADENIC, AND GIOVANNI SEMERARO (EDS.). (ed.) From Web to Social Web: Discovering and Deploying User and Content Profiles, Lecture Notes In Artificial Intelligence*. Berlin, Heidelberg: Springer-Verlag.
- ANDERSON, C. K. & CHENG, M. 2014. Paid Search: Modeling Rank Dependent Behavior. *In Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14)*.

- ANDREASSEN, C. S. 2015. Online Social Network Site Addiction: A Comprehensive Review. *Journal of Current Addiction Reports*, 2, 175-184.
- ANDRIOLE, S. J. 2010. Business Impact of Web 2.0 Technologies. *In Communications of the ACM Magazine*, 53, 67-79.
- ANTELMAN, K. 2006. Toward a 21st Century Library Catalog. *Journal of Information Technology and Libraries*, 25, 261-273.
- ANYANWU, K., MADUKO, A. & SHETH, A. 2005. SemRank: ranking complex relationship search results on the semantic web. *In Proceedings of the 14th international conference on World Wide Web (WWW '05)*. Chiba, Japan.
- ARAUJO-FONTES, C., CAVALCANTI, M. C. & DE-C-MOURA, A. M. 2013. An Ontology-Based Reasoning Approach for Document Annotation. *In Proceedings of the IEEE 17th International Conference on Semantic Computing (ICSC '13)*. Irvine, CA.
- ARCHIE. 1990. *Archie Query Form* [Online]. Archie. Available: [http://archie.icm.edu.pl/archie-adv\\_eng.html](http://archie.icm.edu.pl/archie-adv_eng.html) [Accessed 23rd April 2014].
- ARPANET. 1962. *Internet History 1962 to 1992* [Online]. Available: <http://www.computerhistory.org/internethistory/> [Accessed 23rd August 2015].
- ATTRILL, A. 2015. *Cyberpsychology*, UK, Oxford University Press.
- BAEZ, M., BIRUKOU, A., CASATI, F. & MARCHESE, M. 2010. Addressing Information Overload in the Scientific Community. *Internet Computing, IEEE*, 14, 31 - 38.
- BAEZA-YATES, R. 2010. Query Intent Prediction and Recommendation. *In Proceedings of the 4th International ACM Conference on Recommender Systems (RecSys '10)*. Barcelona, Spain.
- BAEZA-YATES, R. 2006. Algorithmic Challenges in Web Search Engines. *In: ÁLVAREZ, C. & SERNA, M. (eds.) In Proceedings of the 5th International Workshop on Experimental Algorithms (WEA '06)*. Cala Galdana, Menorca, Spain: Springer Berlin Heidelberg.
- BAEZA-YATES, R., HURTADO, C. & MENDOZA, M. 2007. Improving Search Engines by Query Clustering. *Journal of the American Society for Information Science and Technology*, 58, 1793-1804.
- BAEZA-YATES, R. A. & RIBEIRO-NETO, B. 1999 *Modern Information Retrieval*, Boston, MA, USA., Addison-Wesley Longman Publishing Co., Inc.
- BALABANOVIĆ, M. & SHOHAM, Y. 1997. Fab: Content-Based, Collaborative Recommendation. *In Communications of the ACM Magazine*, 40, 66-72.
- BALLARD, T. 2012. 11 - Discovery platforms. *Google This! Putting Google and Other Social Media Sites to Work for your Library A volume in Chandos Information Professional Series*. Woodhead Publishing Limited.
- BALLARD, T. & BLAINE, A. 2011. User search- limiting behavior in online catalogs: Comparing classic catalog use to search behavior in next- generation catalogs. *Journal of New Library World*, 112, 261 - 273.

- BANJANIN, N., BANJANIN, N., DIMITRIJEVIC, I. & PANTIC, I. 2015. Relationship Between Internet Use and Depression: Focus on Physiological Mood Oscillations, Social Networking and Online Addictive behavior. *Journal of Computers in Human Behavior*, 43, 308-312.
- BAR-ILAN, J., SHOHAM, S., IDAN, A., MILLER, Y. & SHACHAK, A. 2008. Structured vs. Unstructured Tagging – A Case Study. *In Journal of Online Information Review*, 32, 635 - 647.
- BARKER, R. 1990. *CASE Method: Entity Relationship Modelling*, Addison-Wesley.
- BARKLEY, R. A. 2014. *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment*, Guilford Publications.
- BAWDEN, D. 2009. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35, 180-191.
- BAWDEN, D., HOLTHAM, C. & COURTNEY, N. 1999. Perspectives on Information Overload. *In Aslib Journal of Information Management* 51, 249 - 255.
- BAXTER, G. J., CONNOLLY, T. M., STANSFIELD, M. H., TSVETKOVA, N. & STOIMENOVA, B. Introducing Web 2.0 in education: A structured approach adopting a Web 2.0 implementation framework. *In: ABRAHAM, A., CORCHADO, E., HAN, S.-Y., GUO, W. & CORCHADO, J., eds. the 7th International Conference on Next Generation Web Services Practices (NWeSP), , 2011 Salamanca, Spain. 499-504.*
- BBC. 2015. *FOMO: How the Fear of Missing Out Drives Social Media 'Addiction'* [Online]. Available: <http://www.bbc.co.uk/schoolreport/31942696> [Accessed 12th July 2015].
- BEEL, J. & GIPP, B. 2009. Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study). *In Proceedings of the 3rd International Conference on Research Challenges in Information Science (RCIS '09)* Fez, Morocco.
- BEITZEL, S. M., JENSEN, E. C., CHOWDHURY, A., GROSSMAN, D. & FRIEDER, O. 2004. Hourly Analysis of a Very Large Topically Categorized Web Query Log. *In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04)*. Sheffield, UK.
- BELEN, A., BARRAGÁNS-MARTÍNEZ, REY-LÓPEZ, M., COSTA MONTENEGRO, E., MIKIC-FONTE, F. A., BURGUILLO, J. C. & PELETEIRO, A. 2010. Exploiting Social Tagging in a Web 2.0 Recommender System. *Internet Computing, IEEE*, 14, 23 - 30.
- BELLOGÍN, A. 2011. Predicting Performance in Recommender Systems. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA. .
- BELLOGÍN, A. & CASTELLS, P. 2010. A Performance Prediction Approach to Enhance Collaborative Filtering Performance. *In Proceedings of the 32nd European Conference on IR Research (ECIR '10)*. Milton Keynes, UK.
- BENEVENUTO, F., RODRIGUES, T., CHA, M. & ALMEIDA, V. 2009. Characterizing User Behavior in Online Social Networks. *In Proceedings of the 9th ACM*

- SIGCOMM Conference on Internet Measurement Conference (IMC '09)*. Chicago, IL, USA.
- BERGAMASCHI, S., GUERRA, F. & LEIBA, B. 2010. Guest Editors' Introduction: Information Overload. *Internet Computing, IEEE*, 14, 10 - 13
- BERNERS-LEE, T., HENDLER, J. & LASSILA, O. 2001. *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* [Online]. Available: <http://www.krchowdhary.com/ai/ai16/sematic%20web-sci-am.pdf> [Accessed 23rd February 2011].
- BESSONOV, M., HEUSER, U., NEKRESTYANOV, I. & PATEL, A. 1999. Open Architecture for Distributed Search Systems. *In Proceedings of the 6th International Conference on Intelligence and Services in Networks (IS&N '99)*. Barcelona, Spain.
- BING-ADS. 2016. *Get your ad on the Bing Network Today* [Online]. Available: [http://advertise.bingads.microsoft.com/en-uk/cl/40771/coupon?s\\_cid=UK-SMB-PPC\\_mk\\_UK\\_src\\_GGL\\_cat\\_Competitive\\_mt\\_e\\_mkwid\\_sKX13BinT|dc\\_pcrId\\_9\\_0399460095\\_pkw\\_pay%20per%20click%20yahoo\\_pmt\\_e\\_slid](http://advertise.bingads.microsoft.com/en-uk/cl/40771/coupon?s_cid=UK-SMB-PPC_mk_UK_src_GGL_cat_Competitive_mt_e_mkwid_sKX13BinT|dc_pcrId_9_0399460095_pkw_pay%20per%20click%20yahoo_pmt_e_slid) [Accessed 6th April 2016].
- BING. 2009. *Bing* [Online]. Available: <https://www.bing.com/> [Accessed 12th September 2015].
- BINGHUBASH, H. & JURIC, R. 2011. Ontology Based Recommendation of Social Networks According to the semantics of Member's Requests. *In Proceedings of 16th International Conference of the Society for Design and Process Science (SDPS '11)* Jeju Island, South Korea.
- BIRCHMEIER, Z., DIETZ-UHLER, B. & STASSER, G. 2011. *Strategic Uses of Social Technology: An Interactive Perspective of Social Psychology*, UK, Cambridge University Press.
- BLANCO-FERNANDEZ, Y., PAZOS-ARIAS, J., GIL-SOLLA, A., RAMOS-CABRER, M. & LOPEZ-NORES, M. 2008. Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems. *IEEE Transactions on Consumer Electronics* 54, 727 - 735.
- BOGERS, T. & VAN-DEN-BOSCH, A. 2009. Collaborative and Content-Based Filtering for Item Recommendation on Social Bookmarking Websites. *In Proceedings of the 3rd International ACM Conference on Recommender Systems and the Social Web Workshop (RecSys '09)*. New York, NY, USA.
- BOLAND, D., MCLACHLAN, R. & MURRAY-SMITH, R. 2015. Engaging with Mobile Music Retrieval. *In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. Copenhagen, Denmark.
- BORKO, H. 1968. Information science: What is it? *American Documentation*, 19, 35.
- BORLUND, P. 2003. The Concept of Relevance in IR. *Journal of American Society for Information Science*, 54, 913 - 925.

- BOURKE, S., MCCARTHY, K. & SMYTH, B. 2011. Power to the people: exploring neighbourhood formations in social recommender system. *In Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- BPS. 2000-2015. *The British Psychological Society: Promoting Excellence in Psychology* [Online]. Available: <http://www.bps.org.uk/> [Accessed 1st June 2015].
- BRADFORD, C. & MARSHALL, I. 1999a. Analysing users WWW search behaviour. *In Proceedings of the IEE Colloquium, Navigating the Web*.
- BRADFORD, C. & MARSHALL, I. W. 1999b. A Bandwidth Friendly Search Engine. *In Proceedings of the International IEEE Multimedia Systems Conference on Multimedia Computing and Systems (Icmcs '99)*. Florence, Italy.
- BRAUNHOFER, M., KAMINSKAS, M. & RICCI, F. 2011. Recommending music for places of interest in a mobile travel guide. *In Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- BRAY, D. A. 2008. Information Pollution, Knowledge Overload, Limited Attention Spans, and Our Responsibilities as IS Professionals. *In Proceedings of the 9th World Conference on Global Information Technology Management Association (GITMA '08)*. Atlanta, Georgia, USA.
- BREEDING, M. 2007. Next-Generation Library Catalogs. *Journal of Library Technology Reports*, 43, 1-41.
- BREWER, G. & KERSLAKE, J. 2015. Cyberbullying, Self-Steem, Empathy and Loneliness. *Journal of Computers in Human Behavior*, 48, 255-260.
- BRIN, S. & PAGE, L. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *In Proceedings of the 17th International World-Wide Web Conference (WWW '98)*. Brisbane, Australia.
- BURGES, C., SHAKED, T., RENSHAW, E., LAZIER, A., DEEDS, M., HAMILTON, N. & HULLENDER, G. 2005. Learning to Rank Using Gradient Descent. *In Proceedings of the 22nd international Conference on Machine Learning (ICML '05)*. Bonn, Germany.
- BURKE, R. 2002. Hybrid Recommender Systems: Survey and Experiments. *Journal of User Modeling and User-Adapted Interaction*, 12, 331-370.
- BURKE, R. 2000. Knowledge-based Recommender Systems. *Encyclopedia of Library and Information Systems*, 69, 180-200.
- BURKE, R. 1999. Integrating Knowledge-Based and Collaborative-Filtering Recommender Systems. *In AAAI Workshop on AI in Electronic Commerce*, 69-72.
- BUSH, V. 1945. As We May Think. *The Atlantic Monthly*, 176, 1641 - 649.
- BUSINESSDICTIONARY. 2016. *Information* [Online]. Available: <http://www.businessdictionary.com/definition/information.html> [Accessed 13th July 2014].
- CAI, Y. & LI, Q. 2010. Personalized search by tag-based user profile and resource profile in collaborative tagging systems. *In Proceedings of the 19th ACM international*

- conference on Information and knowledge management (CIKM '10)*. Toronto , ON, Canada.
- CAMPOS, P. G., DÍEZ, F. & SÁNCHEZ-MONTAÑÉS, M. 2011. Towards a More Realistic Evaluation: Testing the Ability to Predict Future Tastes of Matrix Factorization-Based Recommenders. *In Proceedings of the 5th International ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- CANTADOR, I., BELLOGÍN, A. & CASTELLS, P. 2008. News@hand: A Semantic Web Approach to Recommending News. *In: NEJDL, W., KAY, J., PU, P. & HERDER, E. (eds.) In Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH '08)*. Hannover, Germany.: Springer-Verlag, Berlin, Heidelberg.
- CARDOSO, J. 2007. The Semantic Web Vision: Where Are We? *Journal of IEEE Intelligent Systems*, 22, 84-88.
- CARPINETO, C., OSÍNSKI, S., ROMANO, G. & WEISS, D. 2009. A survey of Web clustering engines. *ACM Computing Surveys (CSUR)*, 41, 38.
- CARROT2. 2005. *Carrot2: Organizes Your Search Results into Topics. With an Instant Overview of What's Available, You Will Quickly Find What You're Looking for.* [Online]. Available: <http://search.carrot2.org/stable/search> [Accessed 12th July 2012].
- CELMA, O. & LAMERE, P. 2011. Music recommendation and discovery revisited. *In Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- CHAMIEL, G. & PAGNUCCO, M. 2009. Ontology Guided Dynamic Preference Elicitation. *In: Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED'09), CEUR Workshop*. Chantilly, VA, United States.
- CHANG, G., HEALEY, M. J., MCHUGH, J. A. M. & WANG, J. T. L. 2001. Keyword-Based Search Engines. *Mining the World Wide Web*.
- CHANG, P.-C. & QUIROGA, L. M. 2009. Using Wikipedia Content to Derive an Ontology for Modeling and Recommending Web Pages across Systems. *In Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys'09), Workshop on Recommender Systems and the Social Web*. New York, USA.
- CHAU, V., KOAY, N., JACKSON, D. & JURIC, R. 2012. Addressing Information Overload by Performing Semantic Filtering After Google Ranking. *In Proceedings of the 2012 International Conference of Society for Design and Process Science (SDPS '12)* Berlin, Germany.
- CHEN, C.-Y., PEDERSEN, S. & MURPHY, K. L. 2012. The influence of perceived information overload on student participation and knowledge construction in computer-mediated communication. *Instructional Science*, 40, 325 - 349.
- CHEN, L., ZHOU, Y. & CHIU, D.-M. 2015. Smart Streaming for Online Video Services. *Journal of IEEE Transactions on Multimedia*, 17, 485-497.
- CHIDLOVSKII, B., GLANCE, N. S. & GRASSO, M. A. 2000. Collaborative Re-Ranking of Search Results. *In Proceedings of the 17th International Conference on*

*Association for the Advancement Artificial Intelligence Workshop (AAAI '00)*  
Austin, Texas, USA.

- CHOWDHURY, A. & SOBOROFF, I. 2002. Automatic evaluation of world wide web search services. *In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '02)*, 421-422.
- CHOWDHURY, G. G. 2003. Natural Language Processing. *Journal of Annual Review of Information Science and Technology*, 37, 51-89.
- CHUN, H., KWAK, H., EOM, Y.-H., AHN, Y.-Y., MOON, S. & JEONG, H. 2008. Comparison of Online Social Relations in Volume vs Interaction: A Case Study of cyworld. *In Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*. Vouliagmeni, Greece.
- CHURCHILL, D. 2009. Educational Applications of Web 2.0: Using Blogs to Support Teaching and Learning. *In British Journal of Educational Technology*, 40, 179–183.
- COATES, K. 2009. *Knowledge Overload* [Online]. Available: <https://www.insidehighered.com/views/2009/03/23/coates> [Accessed 12th March 2015].
- COLLINS-THOMPSON, K., BENNETT, P. N., WHITE, R. W., DE-LA-CHICA, S. & SONTAG, D. 2011. Personalizing Web Search Results by Reading Level. *In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11)*. Glasgow, Scotland.
- COLLINS, J. 2004. Education techniques for lifelong learning: principles of adult learning *Radiographics*, 24, 1483- 1489.
- COLLINS, J. 2009. Lifelong learning in the 21st century and beyond. *Radiographics*, 29, 613-622.
- COOPER, P. 2014. Data, Information, Knowledge and Wisdom. *In Journal of Anaesthesia & Intensive Care Medicine*, 15, 44–45.
- CORBITT, B. J., THANASANKIT, T. & YI, H. 2003. Trust and e-Commerce: A Study of Consumer Perceptions. *Journal of Electronic Commerce Research and Applications*, 2, 203-215.
- CORTES, C., MOHRI, M. & RASTOGI, A. 2007. An alternative ranking problem for search engines. . *In: CAMIL DEMETRESCU (ED.). SPRINGER-VERLAG, B., HEIDELBERG, (ed.) In Proceedings of the 6th international conference on Experimental algorithms (WEA'07), Lecture Notes in Computer Science Rome, Italy.*
- COSTA, A. E. C. M., GUIZZARDI, R. S. S., GUIZZARDI, G. & FILHO, J. E. G. C. P. 2007. CORES: Context-aware, Ontology-based Recommender system for Service recommendation. *In Proceedings of the 19th International Conference on Advanced Information Systems Engineering (CAiSE, '07)* Trondheim, Norway.
- COUSHING, P. & DOUGLAS-TATE, M. 1985. The Effect of People/Product Relationships on Advertising Processing. *In: THORSON, E. & MOORE, J. (eds.) Integrated Communication: Synergy of Persuasive Voices: Advertising and*

*Consumer Psychology Integrated Communication: Synergy of Persuasive Voices Resources for Ecological Psychology*. Psychology Press.

- CRESTANI, F. 1997. Application of Spreading Activation Techniques in Information Retrieval. *Journal of Artificial Intelligence Review*, 11, 453-482.
- CROFT, B., METZLER, D. & STROHMAN, T. 2010a. *Search Engines: Information Retrieval in Practice*, Addison Wesley.
- CROFT, B., METZLER, D. & STROHMAN, T. 2010b. Chapter 8: Evaluating Search Engines. In: HIRSCH, M., GOLDSTEIN, M., SARAH, M. & HOLCOMB, J. (eds.) *Search Engines: Information Retrieval in Practice*. Addison Wesley.
- CSC, LEE, J. S., LORINCZ, C., DRAZEN, E., MAYBERRY, T. & RICCA, L. *Should Healthcare Organizations Use Social Media?* [Online]. Available: [http://www.csc.com/health\\_services/insights/72849-should-healthcare-organizations-use-social-media](http://www.csc.com/health_services/insights/72849-should-healthcare-organizations-use-social-media).
- DALI, H. & YONGMEI, G. 2008. Analysis on the Coupling between E-learning and Lifelong Learning. In *Proceedings of the International Conference on Computer Science and Software Engineering, (CSSE '08)*. Wuhan, China.
- DAVIDSON, J., LIEBALD, B., LIU, J., NANDY, P., VLEET, T. V., GARGI, U., GUPTA, S., HE, Y., LAMBERT, M., LIVINGSTON, B. & SAMPATH, D. 2010. The YouTube video recommendation system. In *Proceedings of the 4th International ACM conference on Recommender systems (RecSys '10)*. Barcelona, Spain.
- DEARING, R. 1997. *The National Committee of Inquiry into Higher Education* [Online]. Available: <http://www.leeds.ac.uk/educol/ncihe/>.
- DEMBER, W. N. 1960. *The Psychology of Perception*, Oxford, England, Henry Holt.
- DENNING, P. J. 2006. Infoglut. *Communications of the ACM*.
- DOLAN, P. L. 2012. *Doctors tell how they use social media as professional water cooler - A survey describes how physicians check sites to filter information and gauge what developments are the most meaningful* [Online]. Available: <http://www.ama-assn.org/amednews/2012/10/22/bisc1022.htm>.
- DONG, A., CHANG, Y., ZHENG, Z., MISHNE, G., BAI, J., ZHANG, R., BUCHNER, K., LIAO, C. & DIAZ, F. 2010. Towards Recency Ranking in Web Search. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM '10)*. New York City, USA.
- DOOMEN, J. 2009. Information Inflation. *Journal of Information Ethics*, 18, 27-37.
- DORČÁK, P., ŠTRACH, P. & POLLÁK, F. 2015. Analytical View of the Perception of Selected Innovative Approaches in Marketing Communications. *Journal of Quality Innovation Prosperity*, 19.
- DUCKDUCKGO. 2008. *DuckDuckGo* [Online]. Available: <http://duckduckgo.com/> [Accessed 12th July 2012].
- DUKE, C. 2012. Global Perspectives on Higher Education and Lifelong Learners. *International Journal of Lifelong Education*, 31, 835-842.

- DUNLAP, J. C. & LOWENTHAL, P. R. 2011. *Learning, unlearning, and relearning: Using Web 2.0 technologies to support the development of lifelong learning skills.*
- EBAY. 2017. *Our History* [Online]. eBay. Available: <https://www.ebayinc.com/our-company/our-history/> [Accessed 15th January 2017].
- ENDLER, N. S. & MAGNUSSON, D. 1976. Toward an Interactional Psychology of Personality. *Journal of Psychological Bulletin*, 83, 956-974.
- FABER, R. G. 1993. Information Overload. *BMJ: British Medical Journal*, 307.
- FAN, C., SONG, J., WEN, Z. & WU, Y. 2010. Content semantic filter based on Domain Ontology. In *Proceedings of the IEEE International Conference on Progress in Informatics and Computing (PIC '10)*. Shanghai, China.
- FARHOOMAND, A. F. & DRURY, D. H. 2002. Managerial information overload. *Communications of the ACM CACM Homepage archive*.
- FELFERNIG, A., FRIEDRICH, G. & LARS SCHMIDT-THIEME, L. 2007. Guest Editors' Introduction: Recommender Systems. *IEEE Intelligent Systems*, 22, 18-21.
- FOAF. 2000-2015. *FOAF (2000-2015+)* [Online]. Available: <http://www.foaf-project.org/> [Accessed 10 February 2011].
- FOX, J. 1998. *Conquering information anxiety. Relief from your data glut starts here* [Online]. Available: <http://www.ibt-pep.com/articles/dataglutrelief.doc> [Accessed 18 September 2015].
- G. SEMERARO, F. A., N. FANIZZI AND S. FERILLI 2000. Intelligent Information Retrieval in a Digital Library Service. In *Proceedings of the 1st DELOS network of excellence workshop on information seeking, searching and querying in digital libraries (DELOS '00)*. Zurich, Switzerland.
- GAO, B. J., BUTTLER, D., ANASTASIU, D. C., WANG, S., ZHANG, P. & JAN, J. 2013a. User-Centric Organization of Search Results. *Internet Computing, IEEE* 17, 52 - 59.
- GAO, R., HAO, B., BAI, S., LI, L., LI, A. & ZHU, T. 2013b. Improving User Profile with Personality Traits Predicted from Social Media Content. In *Proceedings of the 7th International ACM Conference on Recommender Systems (RecSys '13)*. Hong Kong, China.
- GEMMELL, J., SCHIMOLER, T., RAMEZANI, M., CHRISTIANSEN, L. & MOBASHER, B. 2009a. Improving FolkRank with Item-Based Collaborative Filtering. In *Proceedings of the 3rd International ACM Conference on Recommender Systems and the Social Web Workshop (RecSys '09)*. New York, NY, USA.
- GEMMELL, J., SCHIMOLER, T., RAMEZANI, M. & MOBASHER, B. 2009b. Adapting K-Nearest Neighbor for Tag Recommendation in Folksonomies. In *Proceedings of the 7th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems (ITWP '09)*. Pasadena, California, USA.
- GIRARD, J. & ALLISON, M. 2009. Information Overload – The Tip of the Iceberg. *Inside Knowledge*.

- GIVON, S. & LAVRENKO, V. 2009. Predicting Social-Tags for Cold Start Book Recommendations. *In Proceedings of the 3rd International ACM Conference on Recommender Systems (RecSys '09)*. New York, NY, USA.
- GJOKA, M., SIRIVIANOS, M., MARKOPOULOU, A. & YANG, X. 2008. Poking Facebook: Characterization of OSN Applications. *In Proceedings of the 1st Workshop on Online Social Networks (WOSN '08)*. SEATTLE, WA, USA.
- GOBET, F. & EREKU, M. H. 2016. *Can Artificial Intelligence Make Us Happy?* [Online]. Available: <https://www.psychologytoday.com/blog/inside-expertise/201610/can-artificial-intelligence-make-us-happy> [Accessed 1st December 2016].
- GOLDBERG, D., NICHOLS, D., OKI, B. M. & TERRY, D. 1992. Using Collaborative Filtering to Weave an Information Tapestry Communications. . *In Communications of the ACM - Special Issue on Information Filtering Magazine*, 35, 61–70.
- GOLDER, S.-A. & HUBERMAN, B. A. 2005. The Structure of Collaborative Tagging Systems. *Journal of Computing Research Repository (CoRR)*, 32, 198-208.
- GOLDER, S. A. & HUBERMAN, B. A. 2006. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32, 198-208.
- GOODGOPHER. 1991. *Good Gopher: The Search Engine for Independent News and Information* [Online]. Available: <http://www.goodgopher.com/> [Accessed 23rd April 2014].
- GOOGLE-SEO. 2010. *Google: Search Engine Optimization Starter Guide* [Online]. Available: <http://static.googleusercontent.com/media/www.google.co.uk/en/uk/webmasters/docs/search-engine-optimization-starter-guide.pdf> [Accessed 10th September 2015].
- GOOGLE. 2012. *Algorithms* [Online]. Available: <https://www.google.com/insidesearch/howsearchworks/algorithms.html> [Accessed 12 November 2012].
- GOSLING, S. D. & MASON, W. 2014. Internet Research in Psychology. *Journal of Annual Review of Psychology*, 66, 877-902.
- GRAF, S., MACCALLUM, K., LIU, T. C., CHANG, M., WEN, D., TAN, Q., DRON, J., LIN, F., CHEN, S. N., MCGREAL, R. & KINSHUK. Supporting Pervasive Learning Environments. Sixth Annual IEEE International Conference on Pervasive Computing and Communications, Adaptability and Context Awareness in Mobile Learning (PerCom 2008), 2008 Hong Kong.
- GRANDHI, S. A., JONES, Q. & HILTZ, S. R. 2005. Technology Overload: Is There a Technological Panacea? *In Proceedings of the 11th Americas Conference on Information Systems (AMCIS '05)*. Omaha, Nebraska, USA.
- GRAVEN, O. H. & MACKINNON, L. 2005. A survey of current state-of-the art support for lifelong learning. *In Proceedings of the 6th International Conference on Information Technology Based Higher Education and Training (ITHET '05)*. Juan Dolio, Dominican Republic.
- GRČAR, M., MLADENIČ, D. & GROBELNIK, M. 2005. User Profiling for Interest-focused Browsing History. *In Proceedings of 2005 Conference on Data Mining*

- and Data Warehouses (SiKDD '05), 7th International Multi-Conference on Information Society (IS'05). Ljubljana, Slovenia.
- GREGORY S. KOLT 2006. Wading Our Way Through Information Overload. *Jouranl of Physical Therapy in Sport*, 7, 113-114.
- GREHAN, M. 2002. How Search Engines Work. *Search Engine Marketing: The Essential Best Practice Guide*. New York, NY. USA.: Incisive Interactive Marketing LLC.
- GRUBER, T. R. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing? *International Journal of Human-Computer Studies*, 43, 907-928.
- GRUNINGER, M. 1996. Designing and Evaluating Generic Ontologies. *In Proceedings of the 12th European Conference of Artificial Intelligence*. Budapest, Hungary.
- GRUNINGER, M. & FOX, M. S. 1995. Methodology for the Design and Evaluation of Ontologies. *In Proceedings of the International Joint Conference on Artificial Intelligence, Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Canada.
- GRUNINGER, M. & LEE, J. 2002. Applications of Ontology Design Patterns in Biomedical Ontologies. *In Communications of the ACM*, 45, 39-41.
- GUHA, R., MCCOOL, R. & MILLER, E. 2003. Semantic Search. *In Proceedings of the 12th International Conference on World Wide Web (WWW '03)*. Budapest, Hungary.
- GUZZI, F., RICCI, F. & BURKE, R. 2011. Interactive Multi-Party Critiquing for Group Recommendation. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- GYARMATI, L. & TRINH, T. A. 2010. Measuring User Behavior in Online Social Networks. *In Journal of IEEE Network*, 24, 26-31.
- HAMPTON, K. N., GOULET, L. S., RAINIE, L. & PURCELL, K. 2011. Social Networking Sites and our Lives: How People's Trust, Personal Relationships, and Civic and Political Involvement are Connected to their Use of Social Networking Sites and Other Technologies. Washington, D.C, USA.
- HAMPTONA, K. N., SESSIONSB, L. F. & HERB, E.-J. 2011. Core, Networks, Social Isolation, and New Media. *Journal of Information, Communication & Society*, 14, 130-155.
- HAN, T. 2005. Exploring Price and Product Information Search Behavior in e-Market. *In Proceeding of the 6th International Conference on Information Technology: Coding and Computing (ITCC ;05)* Las Vegas, Nevada, USA.
- HANNAK, A., SAPIEZYNSKI, P., KAKHKI, A. M., KRISHNAMURTHY, B., LAZER, D., MISLOVE, A. & WILSON, C. 2013. Measuring personalization of web search. *In Proceedings of the 22nd International Conference on World Wide Web (WWW '13) and International World Wide Web Conferences Steering Committee*. Rio de Janeiro, Brazil.
- HANSSON, L. 2015. *Product Recommendations in E-commerce Systems using Content-based Clustering and Collaborative Filtering*. Master, Lund University.

- HARIRI, N., MOBASHER, B. & BURKE, R. 2013. Query-Driven Context Aware Recommendation. *In Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13)*. Hong Kong, China.
- HARRÉ, R. & SECORD, P. F. 1973. *The Explanation of Social Behaviour*, Littlefield, Adams.
- HAUFF, C., HIEMSTRA, D. & JONG, F. D. 2008. A Survey of Pre-Retrieval Query Performance Predictors. *In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*. Napa Valley, California, USA.
- HAYES-ROTH, F. 2006. Two theories of process design for information superiority: Smart pull vs. smart push. *In Proceedings of the Command and Control Research and Technology Symposium: The State of the Art and the State of the Practice*. San Diego, CA, U.S. Department of Defense, Command and Control Research Program (CCRP '06). San Diego, CA, USA.
- HE, J. & CHU, W. W. 2010. A Social Network-Based Recommender System (SNRS). *In: MEMON, N., XU, J. J., HICKS, D. L. & CHEN, H. (eds.) Data Mining for Social Network Data*. Springer US.
- HENDLER, J. A. 2010. Web 3.0: The Dawn of Semantic Search. *IEEE Computer* 43, 77-80.
- HERLOCKER, J. L., KONSTAN, J. A. & RIEDL, J. Explaining Collaborative Filtering Recommendations. *In Proceedings of the 8th ACM Conference on Computer Supported Cooperative Work, (CSCW '00)*, November 2000 2000 Philadelphia, PA, USA., 241-250.
- HERLOCKER, J. L., KONSTAN, J. A., TERVEEN, L. G. & RIEDLE, J. T. 2004. Evaluating Collaborative Filtering Recommender Systems. *Jouanla of ACM Transactions on Information Systems (TOIS)*, 22, 5-53.
- HESSE, W. 2005. Ontologies in the Software Engineering Process. *In Proceedings of the 2nd GI-Workshop on Enterprise Application Integration (EAI '05)*. Marburg, Germany.
- HIMMA, K. E. 2007. The concept of information overload: A preliminary step in understanding the nature of a harmful information-related condition. 9, 259-272.
- HOENS, T. R., BLANTON, M. & CHAWLA, N. V. 2010a. A Private and Reliable Recommendation System for Social Networks. *In Proceedings of the 2nd IEEE International Conference on Social Computing (SOCIALCOM '10)*. Washington, DC, USA.
- HOENS, T. R., BLANTON, M., STEELE, A. & CHAWLA, N. V. 2010b. Reliable medical recommendation systems with patient privacy. *ACM Trans. Intell. Syst. Technol*, 4, 31.
- HORNUNG, T., KOSCHMIDER, A. & OBERWEIS, A. 2007. A Recommender System for Business Process Models. *In: Proceedings of the 17th Workshop on Information Technologies and Systems (WITS '07)*. Montreal, Canada.

- HORROCKS, I., PARSIA, B., PATEL-SCHNEIDER, P. & HENDLER, J. 2005. Semantic Web Architecture: Stack or Two Towers? *In Proceedings of the 3rd International Workshop, (PPSWR '05)* Dagstuhl Castle, Germany.
- HORROCKS, I. & PATEL-SCHNEIDER, P. F. 2004. A Proposal for an OWL Rules Language. *In Proceedings of the 13th international conference on World Wide Web (WWW '04)*. New York, NY, USA.
- HOTBOT. 1996. *HotBot* [Online]. Available: <http://www.hotbot.com/> [Accessed 12th September 2015].
- HOTHO, A., JÄSCHKE, R., SCHMITZ, C. & STUMME, G. 2006a. Information Retrieval in Folksonomies: Search and Ranking. *In: SURE, Y. & DOMINGUE, J. (eds.) In Proceedings of the 3rd European Conference on The Semantic Web: Research and Applications (ESWC '06)*, . Budva, Montenegro.
- HOTHO, A., JÄSCHKE, R., SCHMITZ, C. & STUMME, G. 2006b. FolkRank : A Ranking Algorithm for Folksonomies. *In Proceedings of the 14th Workshop on Adaptivity and User Modeling in Interactive Systems (ABIS '06) -In Conjunction with Workshop Information Retrieval 2006 of the Special Interest Group Information Retrieval (FGIR '06) - and Workshop on Knowledge and Experience Management (FGWM '06) - and 12th Workshop on Knowledge Discovery, Data Mining, and Machine Learning (KDML '06)*. Hildesheim, Germany.
- HUSSEIN, T. & ZIEGLER, J. 2008. Adapting Web Sites by Spreading Activation in Ontologies. *In Proceedings of the International Conference Workshop on Recommendation and Collaboration (ReColl '08)*. Maspalomas, Gran Canaria, Spain.
- IAN, H., GORI, M. & NUMERICO, T. 2007. *Web Dragons: Inside the Myths of Search Engine Technology*, USA, Morgan Kaufmann as in imprint of elsevier.
- INDIAN, M. & GRIEVE, R. 2014. When Facebook is Easier than Face-to-Face: Social Support Derived from Facebook in Socially Anxious Individuals. *Journal of Personality and Individual Differences*, 59, 102-106.
- INTERNETLIVESTATS. 2016. *Total number of Websites* [Online]. Available: <http://www.internetlivestats.com/total-number-of-websites/> [Accessed 11th January 2016].
- JACOBSEN, I., CHRISTERSON, M., JONSSON, P. & OVERGAARD, G. 1992. *Object Oriented Software Engineering*, Addison-Wesley.
- JAMESON, A. & SMYTH, B. 2007. Recommendation to groups. In The adaptive web. *In: BRUSILOVSKY, P., KOBSA, A. & NEJDL, W. (eds.) Lecture Notes In Computer Science*. Berlin, Heidelberg: Springer-Verlag.
- JANNACH, D. & HEGELICH, K. 2009. A case study on the effectiveness of recommendations in the mobile internet. *In Proceedings of the third ACM conference on Recommender systems (RecSys '09)*. New York City, NY, USA.
- JANNETTE, C. 2009. Education Techniques for Lifelong Learning: Lifelong Learning in the 21st Century and Beyond1. *Journal of Countinuing Education in Radiology, March 2009 RadioGraphics*, 29, 613-622.

- JANSSEN, R. & POOT, H. D. 2006. Information overload: why some people seem to suffer more than others. In: MØRCH, A., MORGAN, K., BRATTETEIG, T., GHOSH, G. & SVANAES, D. (eds.) *In Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles (NordiCHI '06)*. Oslo, Norway.
- JÄRVELIN, K. & KEKÄLÄINEN, J. 2000. IR evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International Conference on Research and Development in Information Retrieval ACM (SIGIR '00)*. Athens, Greece.
- JERATH, K., MA, L. & PARK, Y.-H. 2014. Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity. In *Journal of Marketing Research*, 51, 480-486.
- JOJIC, O., MANU SHUKLA, M. & BHOSAREKAR, N. 2011. A Probabilistic Definition of Item Similarity. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- JONATHAN. 2014. *Information Overload and the Art of Communication* [Online]. Available: <http://advancedlifefskills.com/blog/information-overload-the-art-of-communication/> [Accessed 23rd March 2015].
- JUMPSTATION. 1993. *JumpStation* [Online]. Available: <https://web.archive.org/web/19981212030221/http://www.jumpstation.com> [Accessed 12th September 2015].
- JUNGHANS, M., AGARWAL, S. & STUDER, R. 2012. Behavior Classes for Specification and Search of Complex Services and Processes. In *Proceedings of the IEEE 19th International Conference on Web Services (ICWS '12)*. Honolulu, HI, USA.
- JURIC, R., ALMARRI, H. B., ARNTZEN, A. A., SUH, S. C. & KONJHODZIC, A. 2013. Towards Dynamic Creation of Interdisciplinary Curricula. In *Proceedings of the 18th International Conference of Society for Design and Process Science (SDPS '13)*. Campinas, São Paulo, Brazil.
- JURIC, R., OPPONG, A., KOAY, N. & EMODI, A. 2012. REASONING WITH OWL/SWRL TO PERFORM A SELECTION OF SOFTWARE TOOLS: AN EXAMPLE OF VW TOOLS. In *Proceedings of the 2012 International Conference of Society for Design and Process Science (SDPS '12)*. Berlin, Germany.
- KARR-WISNIEWSKI, P. & LU, Y. 2010. When More is Too Much: Operationalizing Technology Overload and Exploring its Impact on Knowledge Worker Productivity. *Journal of Computers in Human Behavior*, 26, 1061-1072.
- KATARIA, P. 2011. *Resolving semantic conflicts through ontological layering*. PhD, University of Westminster.
- KATZ, G., OFEK, N., SHAPIRA, B., ROKACH, L. & SHANI, G. 2011. Using Wikipedia to Boost Collaborative Filtering Techniques. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- KAUTZ, H., SELMAN, B. & SHAH, M. 1997. Referral web: Combining social networks and collaborative filtering. *Communications of the ACM Digital Library*, 40, 63-65.

- KELLY, D. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval*, 3, 1–224.
- KEMP, J. & LIVINGSTONE, D. 2006. Massively multi-learner: Recent advances in 3D social environments. *Computing and Information Systems Journal*, 10.
- KHAPRE, S. & CHANDRAMOHAN, D. 2011. Personalized Web Service Selection. *In International Journal of Web & Semantic Technology (IJWesT)*, 2.
- KIM, H.-N. & EL SADDIK, A. 2011. Personalized PageRank Vectors for Tag Recommendations: Inside FolkRank. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA. .
- KIRBY, S. 2015. *Social Media: Feeding Eating Disorders?* [Online]. Available: <http://www.dw.com/en/social-media-eating-disorder-body-image/a-18283683> [Accessed 22nd June 2015].
- KLAMMA, R., CHATTI, M. A., DUVAL, E., HUMMEL, H., HVANNBERG, E. H., KRAVCIK, M., LAW, E., NAEVE, A. & SCOTT, P., . , 2007. Social Software for Life-long Learning. *Educational Technology & Society*, 10, 72-83.
- KNIGHT, S. 2015. *Survey: The Internet Has Had a Negative Impact on Morality, Good for Education and Relationships* [Online]. Available: <http://www.techspot.com/news/60105-internet-has-had-negative-impact-morality-survey-finds.html> [Accessed 1st April 2015].
- KNIJNENBURG, B. P., REIJMER, N. J. M. & WILLEMSSEN, M. C. 2011. Each to His Own: How Different Users Call for Different Interaction Methods in Recommender Systems. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- KOENIGSTEIN, N., DROR, G. & KOREN, Y. 2011. Yahoo! Music Recommendations: Modeling Music Ratings with Temporal Dynamics and Item Taxonomy. *In Proceedings of the 5th International ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- KOKOSALAKIS, N. 2000. Lifelong Learning: implications for Universities. (*Editorial*) *European Journal of Education*, 35, 341-359. .
- KÖNIGER, P. & JANOWITZ, K. 1995. Drowning in information, but thirsty for knowledge, . *International Journal of Information Management*, Volume 15, 5 - 16.
- KOREN, Y., BELL, R. & VOLINSKY, C. 2009. Matrix Factorization Techniques for Recommender Systems. *In Journal of Computer*, 42, 30-37.
- KOREN, Y. & SILL, J. 2011. OrdRec: An Ordinal Model for Predicting Personalized Item Rating Distributions. *In Proceedings of the 5th International ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- KOVACH, B. & ROSENSTIEL, T. 2010. How to Know What's True in the Age of Information Overload. *In: SHRIBMAN, D. M. (ed.) Blur*. USA: Bloomsbury.
- LAFFEY, D. 2007. Paid Search: The Innovation that Changed the Web. *Business Horizons* 50, 211-218.
- LAMERE, P. 2008. Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37.

- LANDIA, N. & ANAND, S. S. 2009. Personalised Tag Recommendation. *In Proceedings of the International ACM Recommender Systems & The Social Web Workshop (Recsys '09)*. New York, NY, USA.
- LAPLANTE, A. 1997. Still Drowning! *Computerworld*, 31, 69.
- LATHIA, N., AMATRIAIN, X. & M., P. J. 2009. Collaborative Filtering with Adaptive Information Sources. *In Proceedings of the 21st International Joint Conference on Artificial Intelligence in Conjunction with the 7th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (IJCAI '09)*. Pasadena, CA, USA.
- LEE, B.-K. & LEE, W.-N. 2004. The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment. *Article first published online*, 21.
- LEE, K. & LEE, K. 2011. My head is your tail: applying link analysis on long-tailed music listening behavior for music recommendation. *In Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- LEI, Y., UREN, V. & MOTTA, E. 2006. SemSearch: A Search Engine for the Semantic Web. *In Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks (EKAW '06)*. Pödebrady, Czech Republic.
- LEONTEVA, A. N. 1974. The Problem of Activity in Psychology. *Soviet Psychology*, 13, 4-33.
- LEVY, D. M. 2007. No Time to Think: Reflections on Information Technology and Contemplative Scholarship. *In Journal of Ethics and Information Technology*, 9, 237-249.
- LEVY, D. M. 2005. To grow in wisdom: vannevar bush, information overload, and the life of leisure. *In Proceedings of the Joint Conference on Digital Libraries (JCDL '05). and Proceedings of the 5th ACM/IEEE-CS (ACM/IEEE '05)*. Denver, Colorado, USA.
- LI, C.-J. 2012. Building of Searching Behavior Analysis Models on Multi-Agent Intelligent Agent Technology. *In Proceeding of the 2012 IEEE Symposium on Electrical & Electronics Engineering (EEESYM '12)*. Kuala Lumpur, Malaysia.
- LI, Y., CHEN, Y., LIU, J., CHENG, Y., WANG, X., CHEN, P. & WANG, Q. 2011. Measuring task complexity in information search from user's perspective. *In Proceedings of the American Society for Information Science and Technology Journal*, 48, 1-8.
- LIBRARY, O. 2009. *Evaluating Internet Sources: A Library Resource Guide* [Online]. Northern Michigan University. Available: <http://library.nmu.edu/guides/userguides/webeval.htm> [Accessed 15th October 2016].
- LIBRARYOFCONGRESS. 2006. *Library of Congress* [Online]. Available: <https://www.loc.gov/> [Accessed 10th September 2015].
- LICHY, J. 2011. Internet user behaviour in France and Britain: exploring socio-spatial disparity among adolescents. *International Journal of Consumer Studies*, 35, 470-475.

- LIMPENS, F., GANDON, F. & BUFFA, M. 2008. Bridging Ontologies and Folksonomies to Leverage Knowledge Sharing on the Social Web: A Brief Survey. *In Proceedings of the 23rd International Conference on Automated Software Engineering - Workshops (ASE '08)*. L'Aquila, Italy.
- LIN, G.-L., PENG, H., MA, Q.-L., WEI, J. & QIN, J.-W. 2010. Improving diversity in Web search results re-ranking using absorbing random walks. *In Proceedings of the 2010 International Conference on Machine Learning and Cybernetics (ICMLC '10)*. Qingdao, China.
- LINCOLN, A. 2011. FYI: TMI: Toward a holistic social theory of information overload *First Monday*, 16.
- LOCK, S. 1982. Information Overload: Solution By Quality? *BRITISH MEDICAL JOURNAL*, 284, 1289 - 1290.
- LOHR, S. 2007. *Is Information Overload a \$650 Billion Drag on the Economy?* [Online]. The New York Times - Bits. [Accessed 20 - August 2013].
- LOIZOU, A. & DASMAHAPATRA, S. 2006. Recommender Systems for the Semantic Web. *In: Proceedings of the 2006 European Conference on Artificial Intelligence (ECAI 2006), Recommender Systems Workshop*. Trento, Italy.
- LOMOV, B. F. 1982. The Problem of Activity in Psychology. *Soviet Psychology*, 21, 55-91.
- LOPS, P., DE-GEMMIS, M., SEMERARO, G., NARDUCCI, F. & MUSTO, C. 2011a. Leveraging the LinkedIn Social Network Data for Extracting Content-Based User Profiles. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- LOPS, P., GEMMIS, M.-D. & SEMERARO, G. 2011b. Content-based Recommender Systems: State of the Art and Trends. *In: RICCI, F., ROKACH, L., SHAPIRA, B. & KANTOR, P. B. (eds.) Recommender Systems Handbook*. US: Springer.
- LUEG, C. & FISHER, D. 2003. *From Usenet to CoWebs: Interacting With Social Information Spaces*, Springer.
- LYCOS. 1994. Lycos [Online]. Available: <http://www.lycos.co.uk/> [Accessed 12th September 2015].
- LYNETTE, V., BONNIE, B. L. & KATHRYN, M. L. 2015. Adolescent Problematic Social Networking and School Experiences: The Mediating Effects of Sleep Disruptions and Sleep Quality. *Journal of Cyberpsychology, Behavior, and Social Networking*, 18, 386-392.
- LYONS, J. & KHOT, A. 2000. Infopoints: Managing Information Overload: Developing an Electronic Directory. *BMJ: British Medical Journal*, 320.
- MACGREGOR, G. & MCCULLOCH, E. 2006. Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Journal e-Prints in Library and Information Science*, 55, 291 - 300.
- MAHMOOD, S., ALMARI, B. H., JURIC, R. & KIM, I. 2013. Extracting Tumblr Posts Through Ontological Reasoning for Detecting Alarming Suicidal Notes. *In Proceedings of the 18th International Conference of Society for Design and Process Science (SDPS '13)*. Campinas, São Paulo, Brazil.

- MAHONEY, W. R., HOSPODKA, P., SOUSAN, W., NICKELL, R. & ZHU, Q. 2009. A Coherent Measurement of Web-Search Relevance. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39, 1176 - 1187.
- MARCH, J. & SIMON, H. 1958. *Organizations*, Chichester: John Wiley & Sons.
- MARINHO, L. B., NANOPOULOS, A., SCHMIDT-THIEME, L., JÄSCHKE, R., HOTHO, A., STUMME, G. & SYMEONIDIS, P. 2011. Recommender Systems for Social Tagging Systems. In: RICCI, F., ROKACH, L., SHAPIRA, B. & KANTOR, P. B. (eds.) *Recommender Systems Handbook*. Springer Science+Business Media, LLC.
- MARSHALL, C. C. & BRUSH, A. J. B. 2004. Exploring the Relationship Between Personal and Public Annotations. In *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries* Tucson, AZ, USA.
- MARTIN, G. 2010. Motivation for lifelong learning: a biographical account of efficacy and control. *International Journal of Lifelong Education*, 31, 669-685.
- MARTINS, B. & SILVA, M. J. Spelling Correction for Search Engine Queries. In *Proceedings of the 4th International Conference in Advances in Natural Language Processing (EsTAL '04)*. Alicante, Spain.
- MASAND, B. & SPILIOPOULOU, M. 2003. *Web Usage Analysis and User Profiling: International WEBKDD'99 Workshop San Diego, CA, USA, August 15, 1999 Revised Papers*, Springer-Verlag Berlin Heidelberg.
- MASON, L., BAXTER, J., BARTLETT, P. & FREAN, M. 2000. Boosting Algorithms as Gradient Descent. In *Proceedings of the 13th Conference on Advances in Neural Information Processing Systems (NIPS '00)*. Denver, CO, USA.
- MATHES, A. 2004. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata* [Online]. Available: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> [Accessed 23rd June 2013].
- MAYER, J. M., MOTAHARI, S., SCHULER, R. P. & JONES, Q. 2010. Common Attributes in an Unusual Context: Predicting the Desirability of a Social Match. In *Proceedings of the 4th International ACM Conference on Recommender Systems (RecSys '10)*. Barcelona, Spain.
- MCGOWAN, B. S., WASKO, M., VARTABEDIAN, B. S., MILLER, R. S., FREIHERR, D. D. & ABDOLRASULNIA, M., . 2012. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information,2012. *Journal of Medical Internet Research*.
- MEI, Q. & CHURCH, K. 2008. Entropy of search logs: how hard is search? With personalization? With backoff? In *Proceedings of the 1st International Conference on Web Search and Web Data Mining, (WSDM '08)* Palo Alto, California, USA, .
- MEICHENBAUMA, D. 1977. Cognitive Behaviour Modification. In *Scandinavian Journal of Behaviour Therapy*, 6, 185-192.
- METZLER, D., JONES, R., PENG, F. & ZHANG, R. 2009. Improving Search Relevance for Implicitly Temporal Queries. In *Proceedings of the 32nd International ACM*

- SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*. Boston, Massachusetts, USA.
- MIKULENCAK, M. & TURNER, G. 1997. New Nursing World Features To Prevent User 'Overload'. *AJN The American Journal of Nursing*, 97.
- MILICEVIC, A. K., NANOPOULOS, A. & IVANOVIC, M. 2010. Social Tagging in Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *Journal of Artificial Intelligence Review*, 33, 187-209.
- MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P. & BHATTACHARJEE, B. 2007. Measurement and Analysis of Online Social Networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC '07)*. San Diego, California, USA.
- MITCHELL, R., DAY, D. & HIRSCHMAN, L. 1995. Case study: fishing for information on the Internet. *Information Visualization, 1995. Proceedings*. Atlanta, GA, USA.
- MOHBEYA, K. K. & THAKURA, G. S. 2015. Interesting User Behaviour Prediction in Mobile E-commerce Environment using Constraints. *Journal of IETE Technical Review*, 32, 16-28.
- MSNBOT. 2015. *Meet Our Crawlers* [Online]. Available: <https://www.bing.com/webmaster/help/which-crawlers-does-bing-use-8c184ec0> [Accessed 20th December 2015].
- MURUGESAN, S. 2009. *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (Advances in E-Business Research Series (Aebr) Book)*, Information Science Reference.
- MUSTO, C., NARDUCCI, F., DE-GEMMIS, M., LOPS, P. & SEMERARO, G. 2009. A Tag Recommender System Exploiting User and Community Behavior. In: JANNACH, D. G., EYER, W., FREYNE, J., ANAND, S. S., DUGAN, C., MOBASHER, B. & KOBSA, A. (eds.) *In Proceedings of the International ACM Conference on Recommender Systems & The Social Web Workshop (RecSys '09)*. New York, NY. USA.
- NAGYPÁL, G. 2005. Improving information retrieval effectiveness by using domain knowledge stored in ontologies. In: MEERSMAN, R., TARI, Z. & HERRERO, P. (eds.) *In Proceedings of the 2005 OTM Confederated International Conference on On the Move to Meaningful Internet Systems (OTM '05)*. Agia Napa, Cyprus.: Springer-Verlag, Berlin, Heidelberg.
- NARDO, M., PETRACCO-GIUDICI, M. & NALTSIDIS, M. 2015. Walking Down Wall Street With a Tablet: A survey of Stock Market Predictions Using the Web. *Journal of Economic Surveys*.
- NAZIR, A., RAZA, S. & CHUAH, C.-N. 2008. Unveiling Facebook: A Measurement Study of Social Network Based Applications. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*. Vouliagmeni, Greece.
- NETLINGO. 1995. *The lowdown on Archie, Gopher, Veronica and Jughead* [Online]. Available: <http://www.netlingo.com/more/gopher.php> [Accessed 23rd April 2014].

- NIERSTRASZ, O. 1996. *W3 Catalog History* [Online]. Available: <http://scg.unibe.ch/archive/software/w3catalog/W3CatalogHistory.html> [Accessed 12th September 2015].
- NOY, N. F. & MCGUINNESS, D. L. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University. .
- NSF. <http://www.nsf.gov/about/history/nsf0050/internet/modest.htm> [Online]. National Science Foundation. [Accessed 25th August 2014].
- NUNES, M. A. S. N. & HU, R. 2012. Personality-based recommender systems: an overview. In *Proceedings of the 6th ACM conference on Recommender systems (RecSys '12)*. Dublin, Ireland.
- O'BRIEN, J. & MARAKAS, G. 2007. *Management Information Systems*, McGraw-Hill Education.
- O'REILLY, T. 2009. *What is Web 2.0*, O'Reilly Media, Inc.
- OLIVER, K. M., WILKINSON, G. L. & BENNETT, L. T. 1997. *Evaluating the Quality of Internet Information Sources* [Online]. Available: [http://www.iicm.tugraz.at/thesis/cguetl\\_diss/literatur/Kapitel06/References/Oliver\\_et\\_al.\\_1997/Evaluating%20the Quality.html](http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel06/References/Oliver_et_al._1997/Evaluating%20the%20Quality.html) [Accessed 15th October 2016].
- OLSON, D., MARY, H., LARSON, S., EHRENBERG, A. & LEITHEISER, A. T. 2008. Lifelong Learning for Public Health Practice Education: A Model Curriculum for Bioterrorism and Emergency Readiness. *Public Health Reports*, 123, 53–64.
- OLSTON, C. & CHI, E.-H. 2003. ScentTrails: Integrating Browsing and Searching on the Web. *Journal of ACM Transactions on Computer-Human Interaction (TOCHI)*, 10, 177-197.
- ONLINEOCR. 2009. *Online OCR* [Online]. Available: <http://www.onlineocr.net/> [Accessed 11th December 2016].
- ORACLE. 2010. *Spatial Resource Description Framework (RDF)* [Online]. Available: [http://docs.oracle.com/cd/B19306\\_01/appdev.102/b19307/sdo\\_rdf\\_concepts.htm](http://docs.oracle.com/cd/B19306_01/appdev.102/b19307/sdo_rdf_concepts.htm) [Accessed 12th December 2016].
- OUFALIDA, H. & NOUALI, O. 2009. Exploiting Semantic Web Technologies for Recommender Systems. *A Multi View Recommendation Engine*, 528, 10.
- OWL. 2004a. *OWL Web Ontology Language* [Online]. Available: [www.w3.org/TR/owl-guide/](http://www.w3.org/TR/owl-guide/) [Accessed 23rd June 2010].
- OWL. 2004b. *OWL Web Ontology Language Reference* [Online]. Available: <http://www.w3.org/TR/2004/REC-owl-ref-20040210/> [Accessed 20 March 2015].
- OXFORD-DICTIONARIES-LANGUAGE-MATTERS. 2016. *knowledge* [Online]. Available: <http://www.oxforddictionaries.com/definition/english/knowledge> [Accessed 13th July 2014].
- OXFORD DICTIONARIES, L. M. 2016. *Information* [Online]. Available: [http://www.oxforddictionaries.com/definition/american\\_english/information](http://www.oxforddictionaries.com/definition/american_english/information) [Accessed 13th July 2014].

- PAGE, L., BRIN, S., MOTWANI, R. & WINOGRAD, T. 1999. The PageRank Citation Ranking: Bringing Order to the Web. *Technical Report. Stanford InfoLab.*
- PASSANT, A. & YVES, R. 2008. Combining Social Music and Semantic Web for music-related recommender systems. In: BRESLIN, J., BOJÄRS, U., PASSANT, A. & FERNÁNDEZ, S. (eds.) *In Proceedings of the 7th International Semantic Web Conference (ISWC '08) Workshop on Social Data on the Web (SDoW '08).* Karlsruhe, Germany.
- PAZZANI, M. J. 1999. A Framework for Collaborative, Content-Based and Demographic Filtering. *Journal of Artificial Intelligence Review*, 13, 393-408.
- PAZZANI, M. J. & BILLSUS, D. 2007. Content-Based Recommendation Systems. In: BRUSILOVSKY, P., KOBASA, A. & NEJDAL, W. (eds.) *The adaptive web, Lecture Notes In Computer Science.* Berlin, Heidelberg, 321: Springer-Verlag.
- PIAZZA, J. & BERING, J. M. 2009. Evolutionary Cyber-Psychology: Applying an Evolutionary Framework to Internet Behavior. *Journal of Computers in Human Behavior*, 25, 1258-1269.
- POLLAR, O. 2004. *Surviving Information Overload: How to Find, Filter, and Focus on What's Important*, Crisp Learning.
- POTTER, J. & WETHERELL, M. 1987. *Discourse and Social Psychology: Beyond Attitudes and Behaviour*, SAGE.
- PRAWESH, S. & PADMANABHAN, B. 2011. The "top N" news recommender: count distortion and manipulation resistance. In *Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- PROJECT-XANADU. 1960. *PROJECT XANADU: The Original Hypertext Project* [Online]. Available: <http://www.xanadu.net/> [Accessed 23rd August 1015].
- PU, P., CHEN, L. & HU, R. 2011. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the 5th ACM International Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- QUIRKE, B. 2007. Information Overload or Meaning Underload? In *Strategic Communication Management*, 11, 7.
- RAGOWSKY, A. & SOMERS, T. M. 2002. Enterprise Resource Planning. In *Journal of Management Information Systems*, 19, 11-15.
- RATLIFF, J. D. & RUBINFELD, D. L. 2014. Is There a Market for Organic Search Engine Results and Can Their Manipulation Give Rise to Antitrust Liability? *Journal of Competition Law & Economics*, 10, 517-541.
- REDECKER, C., ALA-MUTKA, K. & PUNIE, Y. 2010. Learning 2.0 - The Impact of Social Media on Learning in Europe. European Commission Joint Research Centre Institute for Prospective Technological Studies Luxembourg: Office for Official Publications of the European Communities.
- REMUND, D. & AIKAT, D. 2012. Drowning in Data: A Review of Information Overload Within Organizations and The Viability of Strategic Communication Principles. In: STROTHER, J., B., ULIJN, J. M. & FAZAL, Z. (eds.) *Information Overload: An International Challenge for Professional Engineers and Technical*

- Communicators*. The Institute of Electrical and Electronics Engineers. Wiley-IEEE Press.
- RENNIE, J. D. M. & SREBRO, N. 2005. Fast maximum margin matrix factorization for collaborative prediction. *In Proceedings of the 22nd international conference on Machine learning (ICML '05)*. Bonn, Germany.
- RESNICK, P. & VARIAN, H. R. 1997 Recommender Systems. *Communications of the ACM Digital Library*, 40, 56-58.
- RICCI, F., ROKACH, L., SHAPIRA, B. & KANTOR, P. B. 2010. *Recommender Systems Handbook*, Springer Science + Business Media, LLC.
- RIVA, G. & GALIMBERTI, C. 2001. *Towards Cyberpsychology: Mind, Cognition, and Society in the Internet Age*.
- ROE, K. E. 1989. LETTERS - Information age and overload. *Science, New Series*, 246.
- ROSENTHAL, S., VELOSO, M. M. & DEY, A. K. 2010. Online Selection of Mediated and Domain-Specific Predictions for Improved Recommender Systems. *In Proceedings of the 7th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems (ITWP'09)*. Pasadena, California, USA.
- RUIZ, F. & HILERA, J. R. 2006. Using Ontologies in Software Engineering and Technology. *In: CALERO, C., RUIZ, F. & PIATTINI, M. (eds.) Ontologies for Software Engineering and Software Technology*. Berlin: Springer Berlin Heidelberg.
- RUOTSALO, T. 2010. *Methods and Applications for Ontology-Based Recommender Systems*. PhD, Aalto University.
- RUSSELL, E., PURVIS, L. M. & BANKS, A. 2007. Describing the strategies used for dealing with email interruptions according to different situational parameters. *Computers in Human Behavior*, 23, 1820-1837.
- RUTZ, O. J. & BUCKLIN, R. E. 2011. From Generic to Branded: A Model of Spillover in Paid Search Advertising. *In Journal of Marketing Research*, 48, 87-102.
- SALTON, G. G. 1971. *The SMART Retrieval System - Experiments in Automatic Document Processing*, Englewood Cliffs, N.J. : Prentice-Hall.
- SANFORD, F. H. 1950. Psychology of Personality. *Journal of Psychological Bulletin*, 47, 446-447.
- SCHAFER, J. B., FRANKOWSKI, D., HERLOCKER, J. & SEN, S. 2007. Collaborative Filtering Recommender Systems. *In: BRUSILOVSKY, P., KOBSA, A. & NEJDL, W. (eds.) The Adaptive Web, Lecture Notes in Computer Science*. Berlin Heidelberg: Springer
- SCHICK, A. G., GORDON, L. A. & HAKA, S. 1990. Information overload: A temporal approach. *Accounting, Organizations and Society*, 15, 199-220.
- SCHMIDT-MAENZ, N. & KOCH, M. 2006. A General Classification of (Search) Queries and Terms,. *In Proceedings of the 3rd International Conference on Information Technology: New Generations (ITNG '06)*. Las Vegas, Nevada, USA.

- SCHMIDT-MÄNZ, N. & KOCH, M. 2005. Patterns in Search Queries. In: BAIER, D., DECKER, R. & SCHMIDT-THIEME, L. (eds.) *Data Analysis and Decision Support*. Springer Berlin Heidelberg.
- SEKO, S., YAGI, T., MOTEGI, M. & MUTO, S. 2011. Group recommendation using feature space representing behavioral tendency and power balance among members. In *Proceedings of the 5th International ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- SENSEBOT. 2007. *Sensebot: The Search Engine that finds sense in a heap of Web pages* [Online]. Available: <http://www.sensebot.net/> [Accessed 10th September 2015].
- SEYMOUR, T., FRANTSVOG, D. & KUMAR, S. 2011. History Of Search Engines *International Journal of Management & Information Systems*, 15, 47-58.
- SHADBOLT, N., BERNERS-LEE, T. & HALL, W. The Semantic Web Revisited. In *IEEE Intelligent Systems Journal*, 21, 96-101.
- SHAIKH, S. & KHARAT, S. 2015. Personalized Mobile Search Engine. *International Journal for Innovative Research in Science & Technology*, 1, 499- 502.
- SHAPIRO, S. C. 1992. Artificial Intelligence. *Encyclopedia of Artificial Intelligence*. 2nd ed.
- SHAW, M. 1971. *Group dynamics: The psychology of small group behavior*, New York, McGraw Hill.
- SHENK, D. 1998. *Data Smog: Surviving the Information Glut Revised and Updated Edition*, Harper Collins.
- SHERIF, M. 1936. *The Psychology of Social Norms*, Oxford, England, Harper.
- SHETH, A. 2011. Semantics Scales Up: Beyond Search in Web 3.0. *Internet Computing, IEEE*, 15, 3 - 6.
- SHIELDS, N. & KANE, J. 2011. Social and Psychological Correlates of Internet Use among College Students. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 5.
- SHOJANOORI, R. 2013. *Towards formalisation of situation-specific computations in pervasive computing environments*. PhD thesis, University of Westminster.
- SHOJANOORI, R. & JURIC, R. 2013. Semantic Remote Patient Monitoring System. *Tele-Medicine and e-Health Journal*, 19, 129-136.
- SHOJANOORI, R., JURIC, R. & LOHI, M. 2012a. Computationally Significant Semantics in Pervasive Healthcare. *Transactions of the SDPS: Journal of Integrated Design and Process Science*, 16, 43-62.
- SHOJANOORI, R., JURIC, R., LOHI, M. & TERSTYANSZKY, G. 2012b. Creating and Manipulating the Semantics of Assistive Self Care Homes. In *Proceedings of 16th International Conference of the Society for Design and Process Science (SDPS' 12)*. Berlin, German.
- SHORT, C. E., REBAR, A. L., PLOTNIKOFF, R. C. & VANDELANOTTE, C. 2015. Designing Engaging Online Behaviour Change Interventions: A Proposed Model of User Engagement. *Journal of The European Health Psychologist*, 17, 32-38.

- SIERSDORFER, S., SAN-PEDRO, J. & SANDERSON, M. 2009. Automatic video tagging using content redundancy. *In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '09)*. Boston, Massachusetts, USA.
- SILVERWOOD, V., SWIFT, C., PETERS, S., WHITEHURST, J., HIGGINS, A. & BRIGDEN, D. 2008. Medical students and lifelong learning: The Journey starts here! : University of Newcastle, UK.
- SIMPERL, E., THURLOW, I., WARREN, P., DENGLER, F., DAVIES, J., GROBELNIK, M., MLADENIĆ, D., GOMEZ-PEREZ, J. M. & MORENO, C. R. 2010. Overcoming Information Overload in the Enterprise: The Active Approach. *Internet Computing, IEEE*, 14, 39 - 46.
- SIMPSON, C. W. & PRUSAK, L. 1995. Troubles with Information Overload—Moving from Quantity to Quality in Information Provision. *International Journal of Information Management*, 15, 413 - 425.
- SINGER, G., NORBISRATH, U. & LEWANDOWSKI, D. 2012. Ordinary Search Engine Users Carrying Out Complex Search Tasks. *Journal of Information Science*, 1 -13.
- SMIT, E. G., VAN-NOORT, G. & VOORVELD, H. A. M. 2014. Understanding Online Behavioural Advertising: User Knowledge, Privacy Concerns and Online Coping Behaviour in Europe. *Journal of Computers in Human Behavior*, 32, 15-22.
- SMITH, K. 2007. *Case Study: Semantic Content Description to Improve Discovery* [Online]. W3C. Available: <https://www.w3.org/2001/sw/sweo/public/UseCases/Vodafone/> [Accessed 12th December 2016].
- SMITH, L. & CLAYTON, B. 2009. Recognising non-formal and informal learning: Participant insights and perspectives. National Centre for Vocational Education Research Ltd. .
- SPINK, A. & JANSEN, B. J. 2004. *Web Search: Public Searching of the Web*, Springer Netherlands.
- STAGNER, R. 1937. *Psychology of Personality*, New York, NY, USA, McGraw-Hill.
- STECK, H. 2013. Evaluation of Recommendations: Rating-Prediction and Ranking. *In Proceedings of the 7th International ACM Conference on Recommender Systems (RecSys '13)*. Hong Kong, China.
- STECK, H. 2011. Item Popularity and Recommendation Accuracy. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- STECK, H. 2010. Training and Testing of Recommender Systems on Data Missing Not at Random. *In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*. Washington DC, DC, USA.
- STOCK, W. G. & STOCK, M. 2013. *Handbook of Information Science*, Berlin, Boston, Walter de Gruyter DmbH
- STOJANOVIC, N., STUDER, R. & STOJANOVIC, L. 2003. An Approach for the Ranking of Query Results in the Semantic Web. *In Proceedings of the 2nd International Semantic Web Conference (ISWC '03)*. Sanibel Island, FL, USA.

- STRACCIA, U. 2003. Distributed Search in the Semantic Web. *In Proceedings of the 2003 International Workshop on Description Logics (DL '03)*. Rome, Italy.
- STROTHER, J. B., ULIJN, J. M. & FAZAL, Z. 2012. Information Overload: An International Challenge to Professional Engineers and Technical Communicators. *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. John Wiley & Sons, Inc.
- SU, A.-J., HU, C., KUZMANVIC, A. & KOH, C.-K. 2010. How to Improve Your Google Ranking: Myths and Reality. *In Proceeding of the International IEEE/ACM Conference on Web Intelligence and Intelligent Agent Technology (WIC '10)*. Toronto, Canada.
- SU, Q. & CHEN, L. 2015. A Method for Discovering Clusters of e-Commerce Interest Patterns Using Click-Stream Data. *Journal of Electronic Commerce Research and Applications*, 14, 1-13.
- SUDEEPTHI, G., ANURADHA, G. & PRASAD-BABU, M. S. 2012. A Survey on Semantic Web Search Engine. *In IJCSI International Journal of Computer Science Issues*, 9, 241-245.
- SULER, J. 1996. *The Psychology of Cyberspace* [Online]. Available: <http://users.rider.edu/~suler/psycyber/psycyber.html> [Accessed 12 June 2015].
- SWRL. 2004a. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML* [Online]. Available: <http://www.fmi.uni-sofia.bg/Members/marian/411430437438-43e442-43743d43043d43844f-43c43043343844144244a44044143a438-43f44043e43344043043c438-418418-418421-418422423-438-434440.-2009-2010-44344743543143d430-43343e43443843d430/41c43044243544043843043b438-43a44a43c-43b43543a446438438-6-438-7/SWRL%20%20A%20Semantic%20Web%20Rule%20Language%20Combining%20OWL%20and%20RuleM....pdf> [Accessed 12th February 2011].
- SWRL. 2004b. *Semantic Web Rule Language (SWRL)* [Online]. Available: [www.w3.org/Submission/2004/03/](http://www.w3.org/Submission/2004/03/) [Accessed 23rd June 2010].
- SWRL. 2004c. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML* [Online]. Available: <http://www.w3.org/Submission/SWRL/> [Accessed 12 May 2015].
- SWT. 2004. *Description of W3C Technology Stack Illustration* [Online]. Available: <http://www.w3.org/Consortium/techstack-desc.html> [Accessed 20 March 2012].
- SYVANEN, A., BEALE, R., SHARPLES, M., AHONEN, M. & LONSDALE, P. Supporting pervasive learning environments: adaptability and context awareness in mobile learning. *In: OGARA, H., SHARPLES, M., KINSHUK & YANO, Y., eds. Wireless and Mobile Technologies in Education, (WMTE 2005)*. IEEE International Workshop, 2005.
- TAKÁCS, G., PILÁSZY, I., NÉMETH, B. & TIKK, D. 2009. Scalable Collaborative Filtering Approaches for Large Recommender Systems. *The Journal of Machine Learning Research*, 10, 623-656.
- TAN, J. 2013. Discusses of User Interest Model in Personalized Search. *International Journal of Advancements in Computing Technology*, 5, 619-626.

- TANG, J., GAO, H., HU, X. & LIU, H. 2013. Context-Aware Review Helpfulness Rating Prediction. *In Proceedings of the 7th International ACM Conference on Recommender Systems (RecSys '13)*. Hong Kong, China.
- TAYEBI, M. A., JAMALI, M., ESTER, M., GLÄSSER, U. & FRANK, R. 2011. CrimeWalker: a recommendation model for suspect investigation. *In Proceedings of the 5th International ACM conference on Recommender systems (RecSys '11)*. Chicago, IL, USA.
- TOFFLER, A. 1970. *Future Shock*, United States, A Bantam Book / published by arrangement with Random House, Inc.
- TRAN, T. 2007. Combining Collaborative Filtering and Knowledge-Based Approaches for Better Recommendation System. *Journal of Business and Technology (JBT) Computational Science and Engineering, (CSE)*, 2, 17-24.
- TROUSSOV, A., PARRA, D. & BRUSILOVSKY, P. 2009. Spreading Activation Approach to Tag-aware Recommenders: Modeling Similarity on Multidimensional Networks *In Proceedings of the 3rd International ACM Conference on Recommender Systems and the Social Web Workshop (RecSys '09)*. New York, NY, USA.
- TSCHERSICH, M. 2011. Design Guidelines for Mobile Group Recommender Systems to Handle Inaccurate or Missing Location Data. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- TUMER, D., SHAH, M. A. & BITIRIM, Y. 2009. An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia. *In Proceedings of the 4th International Conference on Internet Monitoring and Protection (ICIMP '09)*.
- ULIJN, J. M. & STROTHER, J. B. 2012. The Influence of Culture on Information Overload. *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. John Wiley & Sons, Inc.
- ULLMAN, J. D. 2012. Chapter 9 Recommendation Systems. *Mining of Massive Datasets*. New York, USA.: Cambridge University Press.
- UNIVERSITY, P. 2010. *About WordNet* [Online]. Available: <http://wordnet.princeton.edu> [Accessed 15th September 2014].
- USCHOLD, M. 2005. Semantic Annotations for Semantic Filtering. *In Dagstuhl Seminar on Machine Learning for the Semantic Web*.
- VAN ZANDT, T. 2004. Information Overload in a Network of Targeted Communication. *The RAND Journal of Economics*, 35, 542-560.
- VOJNOVIC, M., CRUISE, J., GUNAWARDENA, D. & MARBACH, P. 2009. Ranking and Suggesting Popular Items. *In IEEE Transactions on Knowledge and Data Engineering*, 21, 1133-1146.
- VOSSSEN, P. H. 2012. The Challenge of Information Balance in the Age of Affluent Communication. *In: STROTHER, J. B., ULIJN, J. M. & FAZAL, Z. (eds.) Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. 1 ed.: John Wiley & Sons, Inc.

- W3C-RDFS. 2002. *RDF Vocabulary Description Language 1.0: RDF Schema* [Online]. Available: <https://www.w3.org/TR/2002/WD-rdf-schema-20021112/> [Accessed 12th March 2011].
- W3C-SIOC. 2010. *SIOC Core Ontology Specification* [Online]. Available: <http://rdfs.org/sioc/spec/> [Accessed 20th February 2013].
- W3C-SKOS. 2004. *SKOS Simple Knowledge Organization System* [Online]. Available: <https://www.w3.org/2004/02/skos/> [Accessed 13th July 2014].
- WAL, T. V. 2007. *Folksonomy Coinage and Definition* [Online]. Available: <http://vanderwal.net/folksonomy.html> [Accessed 12th October 2013].
- WAN, L. Application of web 2.0 technologies in e-learning context. . In: GROUP, I. C. P. M., ed. 2nd International Conference on Networking and Digital Society (ICNDS), 2010. 437-440.
- WANDERER. 1993-1995. *World Wide Web Wanderer of Matthew Gray* [Online]. Available: <http://history-computer.com/Internet/Conquering/Wanderer.html> [Accessed 15th September 2015].
- WANG, F.-Y. 2011. Social Media and the Jasmine Revolution. *Intelligent Systems, IEEE*, , 26, 2-4.
- WANG, K., WALKER, T. & ZHENG, Z. 2009. PSkip: estimating relevance ranking quality from web search clickthrough data. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09)*. Paris, France.
- WAY, D. 2010. The Impact of Web-scale Discovery on the Use of a Library Collection. *In Serials Review*, 36, 214-220.
- WEBCRAWLER. 1994. *WebCrawler* [Online]. Available: <http://www.webcrawler.com/> [Accessed 12th September 2015].
- WEBCREATIONUK. 2005. *PPC Management Pay Per Click* [Online]. Available: <https://www.webcreationuk.co.uk/ppc/?keyword=pay%20per%20click&matchtype=e&gclid=CizTh7i2-ssCFUWNGwodNXgFAQ> [Accessed 6th April 2016].
- WEN, J.-R., DOU, Z. & SONG, R. 2009. Personalized Web Search. In: LIU, LING, ÖZSU, M. & (EDS.), T. (eds.) *Encyclopedia of Database Systems*. New York, USA: Springer-Verlag.
- WEN, J.-R., NIE, J.-Y. & ZHANG, H.-J. 2002. Query clustering using user logs. *ACM Trans. Inf. Syst*, 20, , 59-81.
- WEN, J.-R., NIE, J.-Y. & ZHANG, H.-J. 2001. Clustering user queries of a search engine. In *Proceedings of the 10th international conference on World Wide Web (WWW '01)*. Hong Kong.
- WESSNER, M., HAAKE, J. M. & TIETZE, D. A. 2002. An infrastructure for collaborative lifelong learning. In: H., R. & SPRAGUE, J. (eds.) *In Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS '02)* Big Island, HI, USA.
- WESTERMAN, C. G., BROOKS, G. J. & LONGMORE, J. M. 1993. Information Overload. *BMJ: British Medical Journal*, 307.

- WESTMEAD, D. 2013. *Deal With Information Overload: Better Results in Less Time*, Moses Akinmuyiwa.
- WILLIAMSON, J. & EAKER, C. P. E. 2012. The Information Overload Scale. In *Proceedings of the 75th American Society for Information Science and Technology Annual Meeting (ASIS&T '12)*. Baltimore, Maryland.
- WING, S. C., WAI, T. L. & LEE, D. L. 2004. Clustering search engine query log containing noisy clickthroughs. In *Proceedings of the 2004 International Symposium on Applications and the Internet (SAINT '04)* Tokyo, Japan.
- WITTENBURG, K. 1996. The WWW Information Glut: Implications for Next-Generation HCI Technologies. *ACM Computing Surveys (CSUR) - Special Issue: Position Statements on Strategic Directions in Computing Research*, 28, 5.
- WORLDWIDEBE FOUNDATION. 2008 - 2015. *History of the Web* [Online]. World Wide Web Foundation. Available: <https://webfoundation.org/about/vision/history-of-the-web/> [Accessed 12 - July 2014].
- WRIGHT, J. 1998. An Overview of Indexing Methods. In *A to Z: The Newsletter of STC's Indexing SIG.*, 4-8.
- WRIGHTMAIL, A., BATES, D. W., MIDDLETON, B., HONGSERMEIER, T., KASHYAP, V., THOMAS, S. M. & SITTIG, D. F. 2008. Creating and Sharing Clinical Decision Support Content with Web 2.0: Issues and Examples. *Journal of Biomedical Informatics*, 42, 334-346.
- WU, H., ZUBAIR, M. & MALY, K. 2006. Harvesting Social Knowledge from Folksonomies. In *Proceedings of the 17th International Conference on Hypertext and Hypermedia (HYPERTEXT '06)*. Odense, Denmark.
- YANG S. J. H. 2006. Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. *Educational Technology & Society*, 9, 188-201.
- YELLBUSINESS. 2016. *Google Advertising Made Easy for Local Businesses* [Online]. Available: <http://contact.yell.com/search?gclid=CMrKvu21-ssCFclsGwodEV4Ezw> [Accessed 6th April 2016].
- YONG, S. L., HAGENBUCHNER, M. & TSOI, A. C. 2008. Ranking Web Pages Using Machine Learning Approaches,. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08)*. Sydney, NSW, Australia.
- YOUNGER, P. 2010. Internet-Based Information-Seeking Behaviour amongst Doctors and Nurses: A Short Review of the Literature. *Health Information & Libraries Journal*, 27, 2-10.
- YU, L., PAN, R. & LI, Z. 2011. Adaptive Social Similarities for Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- ZARAGOZA, H., CAMBAZOGLU, B. B. & BAEZA-YATES, R. 2010. Web Search Solved?: All Result Rankings the Same? In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*. Toronto, Ontario, Canada.

- ZENG, H.-J., HE, Q.-C., CHEN, Z., MA, W.-Y. & MA, J. 2004. Learning to cluster web search results. *In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '04)*. University of Sheffield, UK.
- ZHANG, D. & DONG, Y. 2004. Semantic, Hierarchical, Online Clustering of Web Search Results. *In: YU, J. X., LIN, X., LU, H. & ZHANG, Y. (eds.) In Proceedings 6th Asia-Pacific Web Conference on Advanced Web Technologies and Applications (APWeb '04)* Hangzhou, China.
- ZHAO, C., ZHANG, Z., LI, H. & XIE, X. 2011a. A Search Result Ranking Algorithm Based on Web Pages and Tags Clustering. *In Proceedings of the IEEE International Conference on Computer Science and Automation Engineering (CSAE '11)*. Zhangjiajie, China.
- ZHAO, S., DU, N., NAUERZ, A., ZHANG, X., YUAN, Q. & FU, R. 2008. Improved Recommendation Based on Collaborative Tagging Behaviors. *In Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI '08)*. Canary Islands, Spain.
- ZHAO, Y., FENG, X., LI, J. & LIU, B. 2011b. Shared Collaborative Filtering. *In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. Chicago, IL, USA.
- ZHOU, T., KUSCSIK, Z., LIU, J.-G., MEDO, M., WAKELING, J. R. & ZHANG, Y.-C. 2010. Solving the Apparent Diversity-Accuracy Dilemma of Recommender Systems. *In National Academy of Sciences of the United States of America, (PANS)*, 107, 4511-4515.
- ZHU, X., GOLDBERG, A. B., GAEL, J. V. & ANDRZEJEWSKI, D. 2007. Improving Diversity in Ranking Using Absorbing Random Walks. *In Proceeding of NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics. Assoc. for Computational Linguistics*. Rochester, New York, NY, USA.
- ZIEGLER, C.-N. 2005. Semantic Web Recommender Systems. *In: W., W., LINDNER, M., MESITI, C., TÜRKER, Y., TZITZIKAS, A. & VAKALI (eds.) EDBT 2004 Workshops Lecture Notes in Computer Scienc*. Berlin Heidelberg: Springer-Verlag.