UNIVERSITI PUTRA MALAYSIA

QUESTION ANALYSIS MODEL USING USER MODELLING AND RELEVANCE FEEDBACK FOR QUESTION ANSWERING

SYARILLA IRYANI BINTI AHMAD SAANY

FSKTM 2015 45

QUESTION ANALYSIS MODEL USING USER MODELLING AND RELEVANCE FEEDBACK FOR QUESTION ANSWERING

By

SYARILLA IRYANI BINTI AHMAD SAANY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor Philosophy

December 2014

DEDICATION

To my family, parents, friends,
FIK and UniSZA

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy.

QUESTION ANALYSIS MODEL USING USER MODELING AND RELEVANCE FEEDBACK FOR QUESTION ANSWERING

By

SYARILLA IRYANI BINTI AHMAD SAANY

December 2014

Chairman: **Associate Professor Ali Mamat, PhD**

Faculty: **Computer Science and Information Technology**

Accessing vast volume of information quickly and easily through the Internet has become a major challenge. One way to access information on the web is through a question answering (QA) mechanism. A Question Answering system aims to provide relevant answers to users' (natural language) questions (queries) by consulting its knowledge base. Providing users with the most relevant answers to their questions is an issue. Many answers returned are not relevant to the questions and this issue is due to many factors. One such factor is the ambiguity yield during the semantic analysis of lexical extracted from the user's question. The existing techniques did not consider some of the terms, called modifier terms, in the user's question which are claimed to have a significant impact of returning correct answer.

The objective of this research is to propose the question analysis model of the QA system that would correctly interpret all the modifier terms in the user's question in order to yield correct answers. In question analysis model, a combination of user modelling (UM) and relevance feedback (RF) is used to increase the accuracy of the returned answer. On the one hand, UM helps the QA system to understand the user's question, manage for question adjustment and increase robustness of the question. Additionally, RF provides an extended framework for QA system to avoid or remedy the ambiguity of the user's question. The proposed model which utilizes Vector Space Model (VSM) is able to semantically interpret and correctly convert modifier terms into a quantifiable form. These modifier terms in user's question may involve with evaluation, computation and/or comparison.

The proposed model is implemented in a prototype of QA system called QAUF (**Q**uestion **A**nswering system with **U**ser Modelling and Relevance **F**eedback). The Answer Retrieval Module of QAUF is adopted from FREyA QA system. The experiments are conducted by using the Raymond Mooney gold standard dataset of Geoquery and run on QAUF. The results are then compared with those of the previous existing QA systems, namely Aqualog and FREyA. The proposed model shows a relatively increased in the F-measure where QAUF is 94.7%, FREyA is 92.4% and Aqualog is only 42%.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah.

MODEL ANALISIS SOALAN MENGGUNAKAN PEMODELAN PENGGUNA DAN MAKLUMBALAS RELEVAN UNTUK MENJAWAB SOALAN

Oleh

SYARILLA IRYANI BINTI AHMAD SAANY

Disember 2014

Pengerusi: **Profesor Madya Ali Mamat, PhD**

Fakulti: **Sains Komputer Dan Teknologi Maklumat**

Mengakses jumlah besar maklumat dengan cepat dan mudah melalui Internet telah menjadi satu cabaran yang besar. Salah satu cara untuk mengakses maklumat di web adalah melalui mekanisme Menjawab Soalan (QA). Sistem Menjawab Soalan bertujuan untuk memberi jawapan tepat untuk soalan pengguna berbahasa asli (pertanyaan) dengan merujuk pangkalan pengetahuan. Menyediakan jawapan yang paling tepat dengan soalan-soalan pengguna adalah merupakan satu isu. Kebanyakkan jawapan yang dikembalikan adalah tidak tepat dengan soalan-soalan dan isu ini adalah disebabkan oleh banyak faktor. Salah satu daripada faktor tersebut adalah wujudnya kekaburan semasa analisis semantik terhadap leksikal yang diekstrak daripada soalan pengguna. Teknik-teknik yang sedia ada tidak mengambilkira beberapa terma yang dirujuk sebagai terma pengubahsuai, yang terdapat di dalam soalan pengguna di mana terma ini mempunyai kesan yang ketara dalam proses mengembalikan jawapan yang betul.

Objektif kajian ini adalah untuk mencadangkan satu model analisis soalan bagi sistem QA yang akan mentafsir dengan betul semua terma pengubahsuai didalam soalan pengguna bagi menghasilkan jawapan yang betul. Didalam model analisis soalan, gabungan pemodelan pengguna (UM) dan maklum balas relevan (RF) digunakan untuk meningkatkan ketepatan jawapan yang dikembalikan. Dalam pada itu, UM membantu sistem QA untuk memahami soalan pengguna, menguruskan pelarasan soalan dan meningkatkan kekukuhan soalan. Selain itu, RF menyediakan satu rangka kerja tambahan untuk sistem QA bagi mengelakkan atau

memulihkan kekaburan soalan pengguna. Model yang dicadangkan ini yang mana menggunakan model ruang vektor (VSM), mampu mentafsir dengan betul dan menukarkan terma pengubahsuai ke dalam bentuk yang boleh diukur. Terma pengubahsuai yang terdapat didalam soalan pengguna mungkin terlibat untuk penilaian, pengiraan dan/atau perbandingan.

Model yang dicadangkan ini dilaksanakan didalam prototaip sistem QA yang dikenali sebagai QAUF (Sistem Menjawab Soalan dengan Pemodelan Pengguna dan Maklum balas Relevan). Modul Dapatan Jawapan bagi QAUF diambil daripada sistem QA FREyA. Eksperimen dijalankan dengan menggunakan set data piawai Geoquery daripada Raymond Mooney dan dilaksanakan pada QAUF. Hasil keputusan tersebut dibandingkan dengan sistem QA terdahulu yang sedia ada, iaitu Aqualog dan FREyA. Model yang dicadangkan menunjukkan keputusan yang agak meningkat dalam Pengukuran-F di mana QAUF adalah 94.7%, FREyA adalah 92.4% dan Aqualog hanya 42%.

ACKNOWLEDGEMENTS

In the name of Allah, Most Gracious, Most Merciful.

All praise, gratitude and thanks are to Almighty, the Lord of the worlds, the Cherisher. All prayers and blessings are upon our prophet Muhammad, his family, Companions and followers.

This work would not have been possible without the prayers, support and encouragement of my loving family and parents whom I share the hardest and the happiest moments with. My deepest gratitude goes to them.

Foremost, I would like to convey my special appreciation to my supervisor, Associate Professor Dr. Ali Mamat, who shared with me a lot of his expertise and research insight. His faithful encouragements, keen insight, worthy guidance, and valuable suggestions throughout the academic period have helped me immensely to achieve success in both my studies and accomplishing this thesis. I am also grateful to my committee members, Dr. Aida Mustapha and Associate Professor Dr. Lilly Suriani Affendey for their helpful suggestions and comments on my research. My acknowledgement also goes to all academics and staffs of Faculty of Computer Science and Information Technology, UPM for their support, cooperation and knowledge.

Thanks must also be extended to all friends and members of Faculty of Informatics and Computing at UniSZA, for their sincere cooperation and supports throughout my long period of studies. Finally, I am grateful to everyone who helped me in anyway to get this research done.

To them I dedicate this thesis.

APPROVAL

I certify that a Thesis Examination Committee has met on 17 December 2014 to conduct the final examination of Syarilla Iryani Binti Ahmad Saany on her thesis entitled "Question Analysis Model using User Modelling and Relevance Feedback for Question Answering" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Degree of Doctor Philosophy.

Members of the Thesis Examination Committee were as follows:

Marzanah binti A. Jabar, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hamidah binti Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Masrah Azrifah binti Azmi Murad, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Eyas El-Qawasmeh, PhD
Professor
Faculty of Computer and Information Science
King Saud University
Saudi Arabia
(External Examiner)

_____

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia
Date: 19 March 2015

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirements for the degree of Doctor Philosophy. The members of the Supervisory Committee were as follows:

Ali Mamat, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Aida Mustapha, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Lilly Suriani Affendey, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____

BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification / fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: 7 May 2015

Name and Matric No.: Syarilla Iryani Binti Ahmad Saany, GS19011

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- a supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:	
Name of Chairman of Supervisory Committee:	**Ali Mamat, PhD**
Signature:	
Name of Member of Supervisory Committee:	**Aida Mustapha, PhD**
Signature:	
Name of Member of Supervisory Committee:	**Lilly Suriani Affendey, PhD**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IR	Information Retrieval
KB	Knowledge base
NL	Natural Language
NLP	Natural Language Processing
QA	Question Answering
RF	Relevance Feedback
TREC	Text REtrieval Conference
UM	User Modelling
VSM	Vector Space Model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IR	Information Retrieval
KB	Knowledge base
NL	Natural Language
NLP	Natural Language Processing
QA	Question Answering
RF	Relevance Feedback
TREC	Text REtrieval Conference
UM	User Modelling
VSM	Vector Space Model

CHAPTER 1

INTRODUCTION

1.1 Research Background

The promising of World Wide Web (WWW) and the increase in web popularity has tremendously contributed to all levels of society. With the growing number of digital documents populating in cyberspace in these days, it is a challenging task to locate needed information based on the given user's query. Rather than returning a list of potentially related documents users demand for systems that can help them to find relevant information easily and quickly. One of the systems is question answering (QA) system which aims to provide precise textual answers to specific natural language users' questions by consulting its knowledge base (Burger, et al., 2001; Voorhees, 2001; Vargas-Vera et al., 2003). As opposed to information retrieval systems that return ranked lists of documents only.

QA system attempts to answer a natural language question in an assortment of question types such as facts, lists, definition, how, why, hypothetical, semantically-constrained and cross-lingual questions (Burger, et al., 2001). Search collections may consist of small local document collections, internal organization documents, knowledge bases or pages in the World Wide Web. To successfully provide an answer to a natural language question, QA system uses combination techniques of information retrieval (IR), information extraction (IE), and natural language processing (NLP).

QA system continues the research trend of natural language interface data base (NLIDB) which was introduced during sixties to seventies era (Androutsopoulos et al., 1995). NLIDB provides access to the databases in natural language. In recent years, QA system has received an extensive attention in research activities which has been largely driven by the TREC[1] (Text REtrieval Conference) QA Track (Voorhees, 2001). In TREC QA track, a lot of QA systems manage to understand questions in natural language and produce answers in the form of selected paragraphs extracted from very large collections of text. Generally QA systems are either focusing in the restricted (closed) or open domain. Restricted (closed) domain QA system deals with questions under a specific domain. This QA system only allows a limited type of questions. As in the question answering of clinical medicine, it utilized a specific domain semantic to understand a user's question and document as well as to extract and generate answers (Demner-Fushman, 2006). In (Kangavari et. al., 2008), a QA system for weather forecasting domain uses syntax and semantic relation among words, dynamic pattern of english language grammar and previous asked question to find the exact answer.

[1] http://trec.nist.gov/

Alternatively, open domain QA system takes questions about everything and usually will have much data available from to extract the answers. Open domain QA system potentially returns the answers to a broad range of questions since no restriction is imposed on the user's special vocabulary or on the type of question (Cooper & Ruger, 2000; Agichtein et al., 2007; Saias & Quaresma, 2007; Dwivedi & Singh, 2013). This less or no restriction makes the processing tasks even more complex.

Subsequently, recent researches have shown increased emphasis in the use of the ontology, which is known to promote semantic capability of a QA system. These ontology-based QA systems are also claimed to perform significantly better as compared to classical QA systems (Zajac, 2001; Atzeni et al., 2004; Saias & Quaresma, 2007; Lopez et al., 2007b; Guo & Zhang, 2008; Damljanovic et al., 2010a; Iqbal et al., 2012). In general, ontology is used to provide and share domain knowledge which is also routed through for finding answers to the questions. For complex questions, ontology is used for reasoning. Ontology can be used to study the ontology "neighbourhood" of other terms in the question if a word is lexically different to the word used by the user (Lopez et al., 2007a). Instead of failing to provide an answer, ontology may assist in finding the value or semantic meaning of the term or relation that are looked for. Restricted (closed) domain ontology-based QA system exploits domain-specific knowledge frequently formalized in ontology.

Ontology-based QA system capitalizes ontology as the semantic model to deduce questions, retrieve and extract the target information from knowledge repositories. Ontology has the facilities in providing the semantic information based on its knowledge of a specific domain. Here, ontology assists in analysing user's natural language question semantically (Saias & Quaresma, 2007; Lim, et al., 2009b; Damljanovic et al., 2010c; Iqbal et al., 2012).

Since the emergence of ontology, research activities in QA systems are progressing towards solving research issues in handling complex questions[2] such as comparative, evaluative, superlative and negation question types (Lopez et al., 2007b; Lim, et al., 2009a; Damljanovic et al., 2010a; Iqbal et al., 2012). This is not an easy task to automatically capture the semantics lies in a complex question structure. The semantic information contained in the user NL question may miss or lose during the question analysis process. As in Demner-Fushman (2006), complex questions are analysed and parsed to multi simple questions which later the existing techniques are used to answer them. Handling complex question may comprise inferences in terminology, analyses the properties or attributes involved before an answer can be drawn (Lim et al., 2009a). The semantic meaning of this question structure has to be thoroughly analysed, semantically interpreted and converted into an executable query

[2] In the context of this research a question is considered to be *simple* if the answer is a piece of information that has been located and retrieved directly as it appears in the information source. On the other hand, a question is considered *complex* if its answer needs more on elaboration (Mollá-Aliod & Vicedo, 2010).

so that the potential answers can be obtained from the corpus or knowledge base (Lim et al., 2009). In Damljanovic et al., (2010a), the FREyA (Feedback Refinement and Extended Vocabulary Aggregation) QA system relies on the user feedback to disambiguate any ambiguity exists in the complex question.

Apart from the brief discussion on the complex question type, an approach of user modelling is also examined. User modelling (UM) is the process of gathering knowledge about a specific user in order to generate a personalized answer from the knowledge base accordingly to the specific requirements (Quarteroni & Manandhar, 2009; Quarteroni, 2010). UM involves the process of developing, retaining and maintaining the user profiles of the systems (Quarteroni, 2010). Once the users have been classified, the QA system will begin the inference process on the basis of that classification. UM has been widely applied in cross disciplinary research including human-computer interaction (Fischer, 2001), artificial intelligence (Fink & Kobsa, 2002), and philosophy (Fischer, 2001; Fink & Kobsa, 2002). Besides UM, relevance feedback (RF) is another mechanism used to improve the performance of QA system. RF is used for rating the relevancy of answers with respect to the query (Burger et. al., 2001; Fink & Kobsa, 2002). Based on the feedback given by users, the QA system will re-rank the answers and present the results again to the user. RF is useful to users who do not possess prior knowledge of the knowledge base (KB). RF is either applied after the system has produced the results (answers) based on the natural language question submitted by the user or is exploited to interpret the questions.

1.2 Problem Statement

Forming the right queries from a given question is crucial. These queries are responsible to get the right set of documents or information which contains potential answers. Since the advent of ontology, recent studies in QA systems have utilized ontology in the development of QA systems. This can be seen in the works of (Atzeni et al., 2004; Lopez et al., 2007b; Guo & Zhang, 2008; Damljanovic et al., 2010c; Chandra et al., 2011; Iqbal et al., 2012) and many more. During the early time, many QA systems used ontology as a mechanism to support query expansion (Lopez et al., 2007b). Most of the former research efforts in QA and ontology-based QA were concentrated on finding the correct answer thus very much attention was paid to the question analysis and processing (Cunningham et. al., 2002). Most research agrees that the question analysis and processing component is one of the core engines in the QA systems. From the analysis of the AquaLog QA system (Lopez et al., 2007b), some user questions such as "Which research areas bring in the **most** funding?"; "Who are the **main** researchers in the semantic web research area?"; "What are the **new** projects in KMi?" and "Who works on the **same** project **as** Enrico Motta?" failed to return correct answer that match with user's intent question. Currently, the major focus in ontology-based QA research

is advancing towards complex question handlers such as in negation questions (Iqbal et al., 2012) or comparative and evaluative questions (Lim et al., 2009a; Damljanovic et al., 2010a). Processing the comparative question or the evaluative question involve comparison and evaluation of one or more criteria. It involves inferences in terminologies before the ontology-based QA system is able to return an answer. The inferencing process includes determining the properties or attributes required for evaluation; computing the associated values of entities or objects and comparing the entities or objects and their values depending on the evaluation of one or more criteria (Lim et al., 2009b).

This means the semantics of comparative and evaluative questions have to be correctly interpreted and converted into a representation based on quantifiable criteria, before the answer could be retrieved from the knowledge base (KB). Ambiguity is one of the main concerns during the evaluation of semantic dimensions in both comparative and evaluative questions. Specific terms in the natural language question structure need to be disambiguated based on its syntax and semantic interpretation in order to generate an executable query. The executable query will manage to obtain answer from the corpus or the knowledge base based on the intent of user's question. One of the terms in the user's question structure is known as "*modifier term*". *Modifier term* is any term that changes other term (Huddleston & Geoffrey, 2002). As in the example of AquaLog's user questions set given above, *'most', 'main', 'new'* and *'same as'* are modifier terms contained in the user NL questions.

In natural language (NL) interface of the QA, the user's question can be very specific or not quite yet clear either to the user himself or to the QA system. This scenario has become another main concern in ontology-based QA. For an example, user may provide question with imprecise representations. This requires expansion, modification or adding information to the question specification. For example, to interpret a question of "`How big is Alaska`" will depend on either, the context of question (user's intent) or the structure of the knowledge base. The word *big* may refer to the size of Alaska or *big* can also mean the population of Alaska. In this scenario, QA needs to have equipped with complex question handler on how to quantify and evaluate specific term ('*big' is a specific term*) contained in the user's NL question.

As mentioned above, the specific term here is known as modifier term. Handling modifier term correctly is important so that the answer drawn from the knowledge base (KB) will match with the user's intent question. Among FREyA's main objectives is to improve understanding of the question's semantic meaning so that FREyA may provide a concise answer to the user's question (Damljanovic et al., 2010c). Nevertheless, questions such as "`Give me the number of rivers in California?`" and "`count the states which have elevations lower than what Alabama has?`" have returned no answer and incorrect answer, respectively. Ambiguities of term's semantic meaning has resulted the incorrect answers to be drawn from the KB. Accordingly,

the ontology-based QA system must be able to correctly interpret all the modifier terms in the user's question so that the returned answer will be successful and accurate. The associated modifier term needs to be identified, interpreted and quantified.

User modelling (UM) is performed to retain individual user information, experiences, common goals, and requirement behaviours. User modelling can characterize the user to certain style of question since users mostly have little knowledge on the contents and structures of the knowledge base. Through user modelling, the context of question may also be emphasized. In YourQA, it utilizes user modelling technique together with a web search engine to generate answers from a KB (Quateroni & Manandhar, 2009). Meanwhile, many researchers also consider Relevance Feedback (RF) technique in QA systems, and it is proven that RF can improve the answers ranking process and the performance of QA systems (Pizzato et al., 2006; Lopez et al., 2007b; Damljanovic et al., 2010a). Despite their individual strengths and contribution in QA systems mentioned above, combination of UM and RF has not been manipulated in the area question analysis for ontology-based QA systems. To fill in this gap, this research proposes a new formulation strategy for analysing a complex question especially containing a specific term known as "modifier term". This is as an effort to increase the accuracy of the returned answer to the user NL question by a QA system. Details on the modifier term will be discussed in the next chapters. The intention of this research is to have both the UM and RF approach which act as a formulation strategy in analysing, interpreting and converting user question into further-processed queries for QA system.

The investigations of this research are embedded in the following research questions:

i. How users' needs and information seeking shall be understood from the user's NL question?
ii. How shall users' needs that are expressed in a question be analysed, interpreted and processed?
iii. How shall the user's question be matched for the answers on the knowledge base?

1.3 Research Objectives

The main research objective is as follows:

i. To design a new question analysis model for an ontology-based QA system that enables to return the answer to the user's intent question.

To achieve this objective, a sub-objective is set which is:

a. To implement the new question analysis model for question answering system using user modelling and relevance feedback.

1.4 Scope of the Research

The scope of this research focuses on the following aspects:

i. The research is on question analysis module which will receive natural language (NL) question that contains modifier terms.
ii. The proposed model is for question analysis module in ontology-based QA system. The ontology-based QA system with proposed question model will return the answer by consulting the gold-standard ontology and knowledge base (KB) taken from Raymond Mooney Dataset (MooneyData, 1994). A set of question used to query the KB is based on Mooney Dataset. The dataset contains 607 annotated user questions with modifier terms from the total of 880 user questions. The Geobase ontology and KB is on US geographical information.
iii. The proposed model will be evaluated with selected existing ontology-based QA systems for the performance comparison purposes.

1.5 Contribution of the Research

The following are the main contributions of this research:

i. Theoretical contributions: investigating and exploiting a theoretical framework for question analysis in ontology-based question answering system.
ii. Empirical contributions: developing a formulation strategy for analysing NL user's question that contains modifier terms. This type of question may denote a comparative and/or evaluative question.
iii. Methodological contributions: proposing a model to formulate a strategy in analysing a NL question containing modifier term for question analysis in an ontology-based QA system.
iv. Application contributions: The performance evaluation of the proposed model using the formal information retrieval metrics calculator which are precision, recall and F measure.

1.6 Organization of the Thesis

The thesis is organized in six chapters, including the introductory chapter which discusses the background of the research, problem statement, objectives, scope and contributions of the research. The remaining of the chapters is as follows:

Chapter 2 discusses about Question Answering and outlines previous research on QA systems which start with a brief of Natural Language Interface Database (NLIDB) systems. It follows with discussion on general architecture of QA systems. Chapter 2 also investigates the methodological and theoretical of related methods and techniques in

question analysis for QA systems. Trends and direction of question analysis for QA system concludes this chapter by addressing the issues and shortcomings exist in the research area.

Chapter 3 describes the research methodology process involves in this research. The chapter further discusses all the activities of the three phases. The final part of the chapter summarizes the research methodology employed.

Chapter 4 presents a new model of question analysis for the ontology-based QA system. The chapter discusses the strategy and approach in analysing the natural language question containing modifier terms. The chapter concludes with a set of detailed processing algorithms and specifications.

Chapter 5 discusses the experimental and result analysis of the implemented model. The experiments use the Raymond Mooney Gold Standard dataset of Geoquery dataset.

Chapter 6 makes the concluding remarks about improving the proposed model of the research. Recommendations for future works are presented as guidelines for further research.

1.7 Concluding Remarks

Apparently, with vast volume of information accessible through the Internet, it has become a major challenge to access information quickly and easily. To provide users with the most relevant answers to his queries in less time and/or resources has turned into complex tasks. Evidently, with the combination of highly availability of web document collections, improvements and advancements in information technology has prioritized the demand for better information access which can be exploited through Question Answering (QA) System. Ambiguity exists when interpreting, comparing and evaluating the semantic dimension of the specific term contains in the user's question. This specific term is referred as *modifier term*. Disambiguation and correct analysis of the intent user's question is critical. Many ontology-based QA systems return partially correct or irrelevant answer to these users' NL questions. The purpose of this research is to formulate a general strategy in analysing the user NL question before it can be transformed into further-processed queries. These queries will consult the knowledge base (KB) in order to return the answer. The correct returned answer has to match the user's intent question without considering partial or incorrect answer. Therefore, there are several components that need to be thoroughly understood before a better performance QA system can be formulated. Here, a better performance QA means the system is able to return a correct answer based on user's intent question. The first essential task is the understanding of the users' needs and users' information seeking behaviour. The second essential task is the analysing and processing of

the users' needs expressed in a question (request). Lastly, the third task is providing a strategy for matching of the user's question to data or information on the document collections or knowledge base.

REFERENCES

(n.d.). Retrieved from The Stanford Natural Language Processing Group : http://nlp.stanford.edu/index.shtml

Resource Description Framework (RDF). (2004). Retrieved April 1, 2012, from W3C Semantic Web: http://www.w3.org/RDF/

Dublin Core Metadata Initiative. (2010, October 11). Retrieved April 3, 2012, from Dublin Core Metadata Element Set, Version 1.1: http://dublincore.org/documents/dces/

IDC Predicts 2012 Will Be the Year of Mobile and Cloud Platform Wars as IT Vendors Vie for Leadership While the Industry Redefines Itself. (2011, Dec 1). Retrieved Apr 11, 2013, from IDC: http://www.idc.com/

Merriam-Webster Dictionary. (2013). Merriam-Webster, Incorporated.
Oxford English Dictionary. (2013). Oxford University Press.

Agichtein, E., Burges, C., & Brill, E. (2007). Question Answering Over Implicitly Structured Web Content. *IEEE/WIC/ACM International Conference on Web Intelligence.* Silicon Valley.

Ahn, D., Fissaha, S., Jijkoun, V., Muller, K., Rijke, M., & Sang, E. (2005). Towards a Multi-Stream Question Answering -As-XML-Retrieval Strategy.

Amaral, C., Figueira, H., Martins, A., Mendes, A., Mendes, P., & Pinto, C. (2005). Priberam's Question Answering System for Portuguese. *CLEF Lecture Notes in Computer Science.4022*, 410-419. Springer.

Androutsopoulos, I., Ritchie, G., & Thanisch, P. (1993). MASQUE/SQL - An Efficient and Portable Natural Language Query Interface for Relational Databases . *Proc of the 6th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 327–330.

Androutsopoulos, I., Ritchie, G., & Thanisch, P. (1995). Natural Language Interfaces to Databases - an Introduction. *Journal of Natural Language Engineering, 1*, 29-81.

Attardi, G., Cisternino, A., Formica, F., Simi, M., & Tommasi, A. (2001). PIQASso: Pisa Question Answering System. *Text REtrieval Conference (TREC 2001).*

Atzeni, P., Basili, R., Hansen, D. H., Missie, P., Paggio, P., Pazienza, M. T., et al. (2004). Ontology-based Question Answering in a Federation of University Sites: The MOSES Case Study. *9th International Conference on Applications of Natural Language to Information Systems (NLDB '04).* Manchester (United Kingdom).

Baeza-Yates, R., & Ribeiro, B. (1999). *Modern Information Retrieval.* Harlow: Addison-Wesley.

Baral, C., Vo, N.H., Liang, S. (2012). Answering Why and How Questions with respect to a Frame-based Knowledge Base: APreliminary Report. *Technical Communications of the 28th International Conference on Logic Programming (ICLP'12)* 17 (2012) 26-36.

Barker, K., Chaudhri, V. K., Chaw, S. Y., Clark, P., Fan, J., Israel, D. J., ... & Yeh, P. Z. (2004). A Question-Answering System for AP Chemistry: Assessing KR&R Technologies. In *KR* (pp. 488-497).

Benamara, F. (2004, July). Cooperative question answering in restricted domains: the WEBCOOP experiment. In *Proceedings of the Workshop Question Answering in Restricted Domains, within ACL*.

Bhaskar, P., Pal, B. C., & Bandyopadhyay, S. (2012). Answer Extraction of Comparative and Evaluative Question in Tourism Domain. *International Journal of Computer Science and Information Technologies (IJCSIT), 3*(4), 4610-4616.

Brants, T. (2003, September). Natural Language Processing in Information Retrieval. In *CLIN*.

Brickley, D. (2002, August 2). *Understanding the Striped RDF/XML Syntax*. Retrieved April 1, 2012, from http://www.w3.org/2001/10/stripes/

Burger, J., Cardie, C., Chaudri, V., Gaizauskas, R., Harabagiu, S., Israel, D., et al. (2001). *Issues, Tasks and Pogram Structures to Roadmap Research in Question & Answering (Q&A).*

Carreras, X., & M`arquez, L. (2004). Introduction to the CoNLL-2004 shared task: Semantic role labeling. *Proceedings of CoNLL 2004*.

Cassan, A., Figueira, H., Martins, A., Mendes, A., Mendes, P., Pinto, C., et al. (2006). Priberam's question answering system in a cross-language environment. *CLEF, Lecture Notes in Computer Science.4730*, pp. 300-309. Springer.

Chandra Pal, B., Bhaskar, P., & Bandyopadhyay, S. (2011). A Rule Based Approach for Analysis of Comparative or Evaluative Questions in Tourism Domain. *Proceedings of the KRAQ11 Workshop*,. Chiang Mai, Thailand 29-37.

Choi, K., Pacana, R. M., Tan, A. L., Yiu, J., & Lim, N. R. (2011). Processing Comparisons and Evaluations in Business Intelligence: A Question Answering System. *International Conference on Uncertainty Reasoning and Knowledge Engineering IEEE,* 137-140.

Cimiano, P., Haase, P., Heizmann, J., Mantel, M., & Studer, R. (2008). Towards portable natural language interfaces to knowledge bases–The case of the ORAKEL system. *Data & Knowledge Engineering*, *65*(2), 325-354.

Cooper, J. R. & Ruger, M. (2000). A Simple Question Answering System. Proceedings of TREC-9.

Cunningham, H., Maynard, D., Bontcheva, K., & Tablan, V. (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *The 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).*

Damljanovic, D., Agatonovic, M., & Cunningham, H. (2010a). Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. *Proceedings ESWC-2010.Part I, LNCS 6088*, Springer, 106–120.

Damljanovic, D.,Agatonovic, M., and Cunningham, H. (2010b). FREyA: an Interactive Way of Querying Linked Data Using Natural Language, *Proceedings of the European Semantic Web Conference*.

Damljanovic, D., Agatonovic, M., and Cunningham, H. (2010c). Identification of the Question Focus: Combining Syntactic Analysis and Ontology-based Lookup through the User

Interaction, *Proceedings of the 7th Language Recourses and Evaluation Conference*.

Demner-Fushman, D. (2006). *Complex Question Answering Based on Semantic Domain Model of Clinical Medicine.* OCLC's Experimental Thesis Catalog, University of Maryland (United States), College Park, Md.

De Roeck, A. N., Ball, R., Brown, K., Fox, C.,et. al. (1991). Helpful Answers To Modal And Hypothetical Questions. *EACL*, 257-262.

Doan-Nguyen, H., & Leila, K. (2004). The Problem of Precision in Restricted-Domain Question Answering. Some Proposed Methods of Improvement. *the ACL 2004 Workshop on Question Answering in Restricted Domains*. Barcelona, Spain: Publisher of Association for Computational Linguistics, 8-15.

Dwivedi, S and Singh, V. (2013). Research and Reviews in Question answering System. *Procedia Technology, First International Conference on Computational Intelligence: Modeling Techniques and Applications*, 417-424.

Fautsch, C. & Savoy, J. (2010). Adapting the tf idf Vector-Space Model to Domain Specific Information Retrieval. *Proc. of Symposium on Applied Computing*, 1708 – 1712.

Fernandez, O., Izquierdo, R., Ferrandez, S., and Vicedo, J.L., (2009). Addressing Ontology-based question answering with collections of user queries.,*Journal of Information Processing and Management*, 45(2), 175-188.

Ferrés, D., & Rodríguez, H. (2006, April). Experiments adapting an open-domain question answering system to the geographical domain using scope-based resources. *Proceedings of the Workshop on Multilingual Question Answering*, 69-76.

Fink, J & Kobsa, A., (2002) User Modeling for Personalized City Tours., *Artificial intelligence review,* 18(1), 33-74.

Fischer, G., (2001). User Modeling in Human–Computer Interaction., *User Modeling and User-Adapted Interaction,* 11(1), 65-86.

Frank, A., Krieger, H. U., Xu, F., Uszkoreit, H., Crysmann, B., Jorg, B., & Schafer, U. (2007). Question Answering from Structured Knowledge Sources. *Journal of Applied Logic*, *5*(1), 20-48.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational Research: Competencies for Analysis and Application* (Sixth Edition ed.). Upper Saddle River, New Jersey: Pearson College Division.

Geun-hae, L., & Lofgren, K. (n.d.). The Survey of Knowledge based Question Answering Systems.

Gildea, D., & Jurafsky, D. (2002). Automatic labeling of semantic roles.*Computational linguistics*, *28*(3), 245-288.

Green, B. F., Wolf, A. K., Chomsky, C., & Laughery, K. (1961). BASEBALL: An automatic question answerer. *Proceedings Western Joint Computer Conference*, 219 - 224.

Gruber, Thomas R. (1993). A translation approach to portable ontology specifications (PDF). *Knowledge Acquisition* 5 (2), 199–220.

Gruber, T. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies* 43 (5-6), 907–928.

Guda, V., Sanampudi, S. K., & Manikyamba, I. (2011). Approaches for Question Answering. *International Journal of Engineering Science and Technology (IJEST), 3*, 990-995.

Guo, Q., & Zhang, M. (2008). Question Answering System Based on Ontology and Semantic Web. *RSKT.*

Harabagiu, S., Moldovan, D., Clark, C., Bowden, M., Hickl, A., Wang, P. (2005) Employing Two Question Answering Systems in TREC-2005. *Proceedings of the Fourteenth Text REtrieval Conference, 2005*.

Heflin, J. (2004). OWL Web Ontology Language Use Cases and Requirements. Retrieved 2010, from http://www.w3.org/TR/2004/REC-webont-req-20040210/

Hendrix, G., Sacerdoti, E., Sagalowicz, D., & Slocum, J. (1978). Developing a Natural Language Interface t oComplex Data. *ACM Transactions on Database Systems*, 105–147.

Hirschman, L., & Gaizauskas, R. (2001). Natural Language Question Answering : The View From Here. *Journal of Natural Language Engineering, 7*, 275-300.

Huddleston, R. D., and Geoffrey K. P. (2002). *The Cambridge Grammar of the English Language*, Cambridge University Press

Huysamen, G. K. (1997). Parallels Between Qualitative Research and Sequentially Performed Quantitative Research. *South African Journal of Psychology, 27*, 1-8.

Iqbal, R., Murad, M. A. A., Selamat, M. H., & Azman, A. (2012). Negation Query Handling Engine For Natural Language Interfaces to Ontologies. *International Conference on Information Retrieval and Knowledge Management (CAMP '12)*, 294-253.

Johansson, P. (2002). User Modeling In Dialog Systems. *St. Anna Report SAR*, 02-2.

Jones, K. S. (1989). *Realism About User Modeling*. Springer Berlin Heidelberg, 341-363.

Kalyanpur, A., Patwardhan, S., Boguraev, B., Lally, A., & Chu-Carroll, J. (2012). Fact-based question decomposition in DeepQA. *IBM Journal of Research and Development 56(3): 13*.

Kangavari, M. R., Ghandchi, S., & Golpour, M. (2008). A New Model for Question Answering Systems. *World Academy of Science, Engineering and Technology, 42*.

Kantz, J. (2001). *Open Domain Question Answering on the WWW.* Retrieved 2008, from http://kantz.com/jason/writing/question.htm

Katrin, E. & Sebastian, P. (2008). A Structured Vector Space Model for Word Meaning in Context. *Proc. Of the 2008 Conference on Empirical Methods in Natural Language Processing,* 897 – 906.

Katz, B. (1997, June). Annotating the World Wide Web using Natural Language. *RIAO*, 136-159.

Khorasani, E. S., Rahimi, S., & Gupta, B. (2009). A Reasoning Methodology for CW-Based Question Answering Systems. *WILF 2009*, 328-335.

Kian, W.K., (2005). Improving answer precision and recall of list questions. *Master's Thesis*, School of Informatics, University of Edinburgh.

Kwok, C., Etzioni, O., & Weld, D. S. (2001). *Scaling Question Answering to the Web.*

Laurent, D., Séguéla, P., & Négre, S. (2006) *Cross Lingual Question Answering using QRISTAL for CLEF 2006*. Working Notes for the CLEF 2006 Workshop.

Leady, P. D., & Ormrod, J. E. (2012). *Practical Research Planning and Design* (Eighth Edition ed.). New Jersey: Pearson Merrill Prentice Hall.

Lee, D.L., Chuang, H. & Seamons, K. (1997). Document Ranking and the Vector-Space Model. *IEEE Software,* March/April, 67 – 75.

Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D., & Shepherd, M. (1990). Cyc: toward programs with common sense. *Communications of the ACM*, *33*(8), 30-49.

Li, C-Y. & Hsu, C-T. (2008). Image Retrieval with Relevance Feedback Based on Graph-Theoritic Region Correspondence Estimation. *IEEE Transactions Multimedia*, 10(2), 447 – 456.

Li, X & Roth, D. (2002). Learning question classifiers. *Proceedings of the 19th international conference on Computational linguistics - Volume 1 (COLING '02), 1, Association for Computational Linguistics,* Stroudsburg, PA, USA, 1-7.

Lim, N. R., Saint-Dizier, P., & Roxas, R. (2009a). Some challenges in the design of comparative and evaluative question answering systems. *KRAQ '09 Proceedings of the 2009 Workshop on Knowledge and Reasoning for Answering Questions*, 15-18.

Lim, N. R., Saint-Dizier, P., Gay, B., & Roxas, R. E. (2009b). A preliminary study of comparative and evaluative questions for business intelligence. *Eighth International Symposium on Natural Language Processing, 2009. SNLP '09.* , 35 - 41.

Lopez, V., Motta, E., Uren, V., & Sabou, M. (2007a). *Literature Review and State of the art on Semantic Question Answering.*

Lopez, V., Uren, V., Motta, E., & Pasin, M. (2007b). AquaLog: An Ontology-driven Question Answering System for Organizational Semantic Intranets. *Journal of Web Semantics, 5*, 72-105.

Lopez, V., Uren, V., Sabau, M., & Motta, E. (2011). Is Question Answering fit for the Semantic Web?: a Survey. *Semantic Web Journal, 2*(2), 125-155.

Love, T. (2000). Theoretical Perspectives, Design Research and the PhD Thesis. In D. Durling, & K. Friedman (Eds.), *Doctoral Education in Design, Foundations for the Future.* Staffordshire, UK: Staffordshire University Press.

Lyman, P., & Varian, H. R. (2003). *How Much Information*. Retrieved Apr 11, 2013, from http://www.sims.berkeley.edu/how-much-info-2003

Mackinnon, L., & Wilson, M. (1996, November). User Modelling For Information Retrieval From Multidatabases. *Proc. 2nd ERCIM Workshop on " User Interfaces for All". Prague, Czech Republic*.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a Large Annotated Corpus Of English: The Penn Treebank. *Computational linguistics*,*19*(2), 313-330.

Martin, P., Appelt, D. E., Grosz, B., & Pereira, F. (1985). TEAM: An Experimental Transportable Natural-Language Interface. *IEEE Database Eng. Bull., 8(3)*, 10-22.

Melucci, M. (2005). Context Modeling and Discovery using Vector Space Bases. *Proc. Of The ACM 14th Conference on Information and Knowledge Management*, 808 – 815.

Mirizzi, R. Di Noia, T., Di Sciascio, E. & Ragone, A. (2012). Web 3.0 in Action: Vector Space Model for Semantic (Movie) Recommendations. *Proc. Of Symposium on Applied Computing*, 403 – 404.

Mizzaro, S., Nazzi, E. & Vassena, L. (2009). Collaborative Annotation for Context-Aware Retrieval. *Proc. Of the Workshop on Exploiting Semantic Annotation in Information Retrieval*, 42 – 45.

Mollá, D., & Vicedo, J. L. (2007). Question answering in restricted domains: An overview. *Computational Linguistics*, *33*(1), 41-61.

Mollá-Aliod, D., & Vicedo, J. (2010). Question answering. *Indurkhya and Damerau (eds) Handbook of Natural Language Processing*, 485-510.

Moldovan, D., & Novischi, A. (2002, August). Lexical chains for question answering. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1* (pp. 1-7). Association for Computational Linguistics.

Moldovan, D., Harabagiu, S., Girju, R., Morarescu, P., Lacatusu, F., Novischi, A., et al. (2002). LCC Tools for Question Answering. *Proceedings of TREC.*

Mondal, D., Gangopadhyay, A. & Russel, W. (2010). Medical Decision Making using Vector Space Model. *Proc. Of the 1st ACM International Conference on Health Informatics*, 386 – 390.

Monz, C. (2003). Document Retrieval in the Context of Question Answering. *Proc. Of the 25th European Conference on Information Retrieval Research*, 571 – 579.

MooneyData. (1994). Retrieved from http://www.cs.utexas.edu/~ml/nldata/geoquery.html.

Mooers, C. N. (1951). Zatocoding Applied to Mechanical Organization of Knowledge. *American Documentation, 2*, 20-32.

Moussa, A. M., & Abdel-Kader, R. F. (2011). QASYO: A Question Answering System for YAGO Ontology. *International Journal of Database Theory and Application, 4*, 99-112.

Narayanan, S., & Harabagiu, S. (2004, August). Question answering based on semantic structures. In *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, 693.

Necib, C. B., & Freytag, J. C. (2005). Query Processing Using Ontologies . *Conference on Advanced Information Systems Engineering (CAiSE '05).* Porto, Portugal.

Nihalani, N., Motwani, M., & Silakari, S. (2010). An Intelligent Interface for relational databases. *International Journal of Simulation Systems, Science & Technology, 11*(1), 1473-8031.

Niu, Y., & Hirst, G. (2004, July). Analysis of semantic classes in medical text for question answering. In *Proceedings of the ACL 2004 Workshop on Question Answering in Restricted Domains*, 54-61.

Noy, N. F. & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. *Stanford*

Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Parton, K., & McKeown, K. (2010, August). MT Error Detection For Cross-Lingual Question Answering. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Association for Computational Linguistics, 946-954.

Pasca, M. (2002, May). Answer Finding Guided by Question Semantic Constraints. In *FLAIRS Conference*, 67-71.

Pazienza, M. T., & Stellato, A. (2006). An Open and Scalable Framework for Enriching Ontologies with Natural Language Content. *19th International Conference on Industrial, Engineering, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE'06), special session on Ontology & Text.* Annecy, France.

Pazienza, M. T., Stellato, A., Henriksen, L., Paggio, P., & Zanotto, F. M. (2005). Ontology Mapping to Support Ontology-based Question Answering. *2nd MEANING Workshop.* Trento.

Perez-Carballo, J., & Strzalkowski, T. (2000). Natural language information retrieval: progress report. *Information processing & management*, *36*(1), 155-178.

Pizzato, L. A., Molla, D. & Paris, C. (2006). Pseudo Relevance Feedback Using Named Entities for Question Answering.*Proc. Australasian Language Technology Workshop*, 83 – 90.

Popescu, A. M., Etzioni, O., & Kautz, H. (2003). Towards a theory of natural language interfaces to databases. *International Conference on Intelligent User Interfaces*, 149-157.

Porter, B. W., Lester, J., Murray, K., Pittman, K., Souther, A., Acker, L., & Jones, T. (1988). *AI research in the context of a multifunctional knowledge base: The botany knowledge base project*. Artificial Intelligence Laboratory, University of Texas at Austin.

Quarteroni, S. (2010). Personalized Question Answering.*TAL*, 51(1), 97 – 123.

Quarteroni, S. & Manandhar, S. (2009). Designing an Interactive Open-Domain Question Answering System. *Journal Language Engineering*, 15(1), 73 – 95.

Razmara, M., & Kosseim, L. (2007). A Little Known Fact Is ... Answering Other Question Using Interest-Markers. *CICLing 2207* (pp. 518-529). Springer-Verlag Berlin Heidelberg 2007.

RDF Primer. (n.d.). Retrieved from W3C Recommendation: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Rieck, K., Wressnegger, C. & Bikadorov, A. (2012). Sally: A Tool for Embedding Strings in Vector Spaces. *Journal on Machine Learning Research*, 13, 3247 – 3251.

Rinaldi, F., Dowdall, J., Kaljurand, K., Hess, M., & Mollá, D. (2003, July). Exploiting paraphrases in a question answering system. In *Proceedings of the second international workshop on Paraphrasing-Volume 16* (pp. 25-32). Association for Computational Linguistics.

Robertson, S. E. (1981). The Methodology of Information Retrieval Experiments. In K. S. Jones, *Information Retrieval Experiments* (pp. 9-12). London: Butterworths.

Rocchio, J.J. (1971). Relevance Feedback in Information Retrieval.*In G. Salton (ed.), The Smart Retrieval System: experiments in automatic document processing, Prentice Hall*, 313 – 323.

Rose, N. T., Saint-Dizier, L, P., Gay, B., and Roxas, R.E., (2009). A preliminary study of comparative and evaluative questions for business intelligence. *Natural Language Processing, SNLP '09. Eighth International Symposium*, 35-41.

Saias, J., & Quaresma, P. (2007). A Proposal for a Web Information Extraction and Question-Answer System. *Advance in Intelligent Web ASC 43.*

Salloum, W. (2009). A Question Answering System based on Conceptual Graph Formalism. *2009 Second International Symposium on Knowledge Acquisition and Modeling.3*, 383-386. IEEE.

Salton, G., Wong, A. & Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11), 613 – 620.

Saquete, E., Martínez-Barco, P., Muñoz, R., & Vicedo, J. (2004). Multilayered question answering system applied to temporality evaluation. *SEPLN (ed.) XX Congreso de la SEPLN.* Barcelona, España.

Saxena, A. K., Sambhu, G. V., Kaushik, S., & Subramaniam, L. V. (2007, October). IITD-IBMIRL System for Question Answering Using Pattern Matching, Semantic Type and Semantic Category Recognition. In *TREC*.

Shaban-Nejad, A., & Haarslev, V. (2008). Web-based dynamic learning through lexical chaining: a step forward towards knowledge-driven education. *ACM SIGCSE Bulletin*, *40*(3), 375-375.

Shaban-Nejad, A. (2010). *A Framework for Analyzing Changes in Health Care Lexicons and Nomenclatures* (Doctoral dissertation, Concordia University).

Shen, D., & Lapata, M. (2007, June). Using Semantic Roles to Improve Question Answering. In *EMNLP-CoNLL*, 12-21.

Stellato, A., & Oltramari, A. (2008). Enriching Ontologies with Linguistic Content: An Evaluation Framework . *4th Workshop on Interfacing Ontologies and Lexical Resources for Semantic Web Technologies (OntoLex2008).* Marrakech, Morocco.

Strzalkowski, T., Lin, F., Perez-Carballo, J., & Wang, J. (1997, November). Natural language information retrieval TREC-6 report. In *TREC,* 347-366.

Strzalkowski, T., Carballo, J. P., Karlgren, J., Hulth, A., Tapanainen, P., & Lahtinen, T. (1999, November). Natural Language Information Retrieval: TREC-8 Report. In *TREC*.

Strzalkowski, T., Stein, G. C., Wise, G. B., & Bagga, A. (2000, April). Towards the Next Generation Information Retrieval. In *RIAO* , 1196-1207.

Sun, R., Jiang, J., Fan Tan, Y., Cui, H., Chua, T., Kan, M. (2005) Using Syntactic and Semantic Relation Analysis in Question Answering. *Proceedings of the Fourteenth Text REtrieval Conference*.

Tablan, V., Damljanovic, D., & Bontcheva, K. (2008). A Natural Language Query Interface to Structured Information. *European Semantic Web Conference (ESWC 2008)*, 361-375.

Tribbey, W. & Mitropoulos, F. (2012). Construction and Analysis of Vector Space Models for Use in Aspect Mining. *Proc. Of ACMSE 2012*.

Trigui, O., Belguith, H.L., Rosso, P. (2010). DefArabicQA: Arabic Definition Question Answering System. *Workshop on Language Resources and Human Language Technologies for Semitic Languages, 7th LREC*, Valletta, Malta .

Tsatsaronis, G. & Panagiotopoulou, V. (2009). A Generalized Vector Space Model for Text Retrieval Based on Semantic Relatedness. *Proc. Of the EACL 2009 Student Research Workshop,* 70 – 78.

Turing, A. (1950), Computing Machinery and Intelligence, *Mind LIX* (236): 433 – 460.

Uren, V., Lei, Y., Lopez, V., Liu, H., Motta, E. & Giordanino, M. (2007) The Usability Of Semantic Search Tools: A Review, *Knowledge Engineering Review*, 22, Cambridge University Press. 361-377.

VanRijsbergen, C. J. (1979). *Information Retrieval* (2nd ed.). London: Butterworths.

Vargas-Vera, M., & Motta, E. (2004). AQUA - Ontology-Based Question Answering System. *Third Mexican Internation Conference on Artificial Intelligence.* Mexico City, Mexico.

Vargas-Vera, M., Motta, E., & Dominigue, J. (2003). An Ontology-Driven Question Answering System (AQUA). *Knowledge Media Institute, The Open University.*

Vargas-Vera, M., Motta, E., & Domininge, J. (2003). AQUA: An Ontology-Driven Question Answering System. *AAAI Spring Symposium New Directions in Question Answering.* Stanford University.

Vassilvitskii, S. & Brill, E. (2006). Using Web-Graph Distance for Relevance Feedback in Web Search. *Proc. of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 147 – 153.

Vicedo, J. L., & Molla, D. (2001). Open-Domain Question-Answering Technology: State of the Art and Future Trends. *ACM Journal Name*, *2*(3), 09.

Voorhees, E. (2004). Overview of the TREC 2004 Question Answering Track.

Voorhees, E. M. (2001). The TREC question answering track. *Natural Language Engineering, 7*(4), 361-378.

Wahlster, W & Kobsa, A. (1989). User Models in Dialog Systems. Heidelberg: Springer.

Wang, C., Xiong. M., Zhou, Q., & Yu, Y. (2007). PANTO: A Portable Natural Language Interface to Ontologies. *European Semantics Web Conference(ESWC '07)*, 473-487.

Warren, D., & Pereira, F. (1982). An Efficient Easily Adaptable System for Interpreting Natural Language Queries. *Computational Linguistics, 8*(3-4), 110–122.

Wegner, P. (1976). Research Paradigms in Computer Science *2nd International Conference Proceedings on Software Engineering*, , 322–330.

Woods, W. (1973). Progress in natura llanguage understanding - anapplication to luna rgeology. *American Federation of Information Processing Societies (AFIPS) Conference Proceedings*, 441 - 450.

Wu, J-W. & Tseng, J.C.R. (2008). A Hierarchical Relevance Feedback Algorithm for Improving the Precision of Virtual Tutoring Assistant Systems. *WSEAS Transactions on Information Science*, 5(3), 94 – 103.

Wu, M., Duan, M., Shaikh, S., Small, S., & Strzalkowski, T. (2006). ILQUA An IE-Driven Question Answering System System.

Zajac, R. (2001). Towards Ontological Question Answering . *Proceedings of ACL-2001 Workshop.*

Zhou, X. S. & Huang, T.S. (2003). Relevance Feedback in Image Retrieval: A Comprehensive Review. *Multimedia Systems*, 8, 536 – 544.