UNIVERSITI PUTRA MALAYSIA

QUERY DISAMBIGUATION APPROACH USING TRIPLE-FILTER

ALIYU ISAH AGAIE

FSKTM 2018 11

QUERY DISAMBIGUATION APPROACH USING TRIPLE-FILTER

By

ALIYU ISAH AGAIE

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

December 2017

DEDICATION

In loving memory of my parents,

Alhaji Isa Agaie

Hajiya Salamatu

Hajiya Ramatu

Hajiya Saratu

whom I absolutely adore.

And to my wife, **Zainab Muhammad**, a dependable teammate on this research journey.

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

QUERY DISAMBIGUATION APPROACH USING TRIPLE-FILTER

By

ALIYU ISAH AGAIE

December 2017

Chairman : Masrah Azrifah Azmi Murad, PhD
Faculty : Computer Science and Information Technology

In order to effectively deal with structured information, some technical skills are needed. However, most users do not possess these skills. Natural Language Interfaces (NLIs) are therefore built to provide everyday users who lack the needed technical knowledge, with some means of gaining access to the information stored in knowledge bases. They are designed to deal with the natural language articulation of what the user wants, and then transform it into a computer language that specifies how to accomplish it. Each NLI system also has a scope and limitation which everyday users are unaware of. It is therefore understandable that the users may not see errors in their queries or even know how to write appropriate queries (according to the system's limitation and scope) in order to retrieve the correct information.

The effective retrieval of any piece of information depends wholly on the correct mapping of queries made in natural language to machine understandable form. However, most of the existing NLIs are lacking in terms of being able to provide support for users to formulate their queries. Once queries are wrongly formulated, there is tendency for the system to retrieve wrong answers. As a result, a user may have to reformulate his query severally before the required answer is retrieved (if at all).

In this thesis therefore, it is proposed that the best way to formulate appropriate queries is by guiding the user through the query writing process and helping the user to resolve ambiguities by providing suggestions that are easy to understand. The proposed approach, referred to as triple-filter query disambiguation approach, has been implemented into a prototype as a proof of concept. The prototype, referred to as QuFA (Query Formulation Assistant), is intended to serve as an upper layer for NLIs. It is equipped with an authoring service that guides the user to write his query, a disambiguation module that resolves ambiguities in order to ascertain the user's intention and finally, a query rewriting module that transforms the user's input query

into an intermediate query that will suit the underlying search system's perspective of the user's question.

Extensive experimental evaluations were conducted in order to validate the proposed approach, using the developed prototype. The proposed triple-filter query disambiguation approach was directly compared with the approach in FREyA (Feedback, Refinement and Extended vocabulary Aggregation) that also provides support to users when formulating queries. The evaluation was based on the usability and performance of the approaches. In terms of usability, the results show that the proposed approach has the potential of being more acceptable in the field; and in terms of effectiveness, it also shows a high performance based on precision and recall. The proposed approach helps users to conceive and articulate more effective queries, and facilitates information search activities.

The main contributions of this research work include the introduction of an approach that enables users without knowledge of formal computer languages to formulate useful queries while effectively expressing themselves using natural language. The approach utilizes the effectiveness of human-computer dialogue to effectively retrieve desired information from ontologies. The proposed triple-filter disambiguation approach inculcates a learning mechanism that continues to automatically learn from user queries and continuously improves its performance capability. The triple-filter disambiguation algorithm was also developed, along with two documents (terms equivalence catalogue (TEC) and the enhanced concepts store (ECS)) that represent the thesaurus and the lexicon for use with the Mooney Geoquery dataset. All of these are available for use by other researchers.

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

PENDEKATAN KETIDAKTENTUAN PERTANYAAN MENGGUNAKAN TAPISAN GANDA TIGA

Oleh

ALIYU ISAH AGAIE

Disember 2017

Pengerusi : Masrah Azrifah Azmi Murad, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Untuk mengendalikan maklumat berstruktur dengan efektif, memerlukan kemahiran teknikal. Walau bagaimanapun, tidak semua pengguna mempunyai kemahiran tersebut. Maka, antaramuka bahasa tabii atau *Natural language Interfaces (NLIs)* dibangunkan bertujuan memberi kemudahan kepada pengguna yang kurang kemahiran tenikal untuk mendapatkan maklumat daripada pangkalan data pengetahuan. Antaramuka ini direkabentuk dalam bahasa tabii untuk mengendalikan keperluan penguna seterusnya ditukar kepada bahasa komputer. Ia menggerakkan kepada penyelesaian untuk memenuhi kepada keperluan pengguna tersebut. Untuk makluman, setiap sistem NLI ini mempunyai skop dan batas keupayaan yang tidak disedari pengguna. Maka, pengguna berkemungkinan tidak menyedari wujudnya ralat dalam setiap pemprosesan pertanyaan. Maka, disebabkan kelemahan ini, kaedah pemprosesan pertanyaan mungkin tidak tepat dalam mendapatkan maklumat yang betul.

Kaedah yang berkesan bagi mendapatkan maklumat yang tepat bergantung sepenuhnya kepada ketepatan pemetaan pertanyaan yang dibuat dalam bahasa tabii kepada bahasa yang difahami oleh mesin (komputer). Walau bagaimanapun, kebanyakan NLI sedia ada yang dibangunkan, kurang berkesan dalam memberi bantuan untuk pengguna membuat formulasi pertanyaan mereka. Kebarangkalian untuk mendapatkan jawapan yang salah adalah tinggi apabila formulasi pertanyaan adalah tidak tepat. Oleh yang demikian, pengguna mungkin terpaksa untuk membuat formulasi semula pertanyaan mereka beberapa kali sehingga mendapat jawapan yang tepat (untuk semua pertanyaan).

Sehubungan dengan itu, tesis ini telah mencadangkan kaedah yang terbaik membimbing pengguna dalam membuat formulasi pertanyaan yang sesuai di samping

menyediakan beberapa cadangan yang mudah difahami untuk membantu pengguna. Kaedah yang dicadangkan menggunakan pendekatan ketidaktentuan pertanyaan tapisan ganda tiga telah dibangunkan sebagai prototaip untuk pembuktian konsep. Prototaip tersebut dirujuk sebagai *QuFA (Query Formulation Assistant*) atau Bantuan Formulasi Pertanyaan. QuFA ini adalah kaedah paling tertinggi untuk NLI. Ia telah dilengkapkan dengan perkhidmatan pengarangan yang membimbing pengguna memasukkan pertanyaan mereka; modul ketidaktentuan bagi menyelesaikan masalah kekeliruan dalam memastikan pertanyaan pengguna difahami dan akhirnya, modul kemasukan semula pertanyaan yang mengubah input pertanyaan pengguna kepada pertanyaan yang sederhana dan disejajarkan dengan perspektif sistem ke atas pertanyaan pengguna.

Penilaian eksperimental secara meluas telah dijalankan menggunakan prototaip yang dibangunkan untuk menilai pendekatan tersebut. Pendekatan ketidaktentuan pertanyaan tapisan ganda tiga telah dibandingkan dengan pendekatan *FREyA (Feedback, Refinement and Extended vocabulary Aggregation)* yang juga menyokong pengguna dalam membuat formulasi pertanyaan. Penilaian ini adalah berdasarkan kebolehgunaan dan prestasi kaedah-kaedah yang dibangunkan tersebut. Dari segi kebolehgunaan, keputusan kajian menunjukkan kaedah yang dicadangkan mempunyai potensi yang meyakinkan dalam bidang tersebut. Dari segi keberkesanan, pendekatan yang dicadangkan menunjukkan prestasi yang tinggi berasaskan ketepatan dan panggilan balik pertanyaan pengguna. Oleh itu, kaedah tersebut juga telah membantu pengguna membentuk pertanyaan yang lebih efektif dan tepat, di samping mempermudahkan aktiviti gelintaran.

Sumbangan utama dalam kajian ini, adalah pengenalan kepada pendekatan yang membolehkan pengguna tanpa pengetahuan untuk memformulasi pertanyaan berkenaaan bahasa komputer yang formal di samping mempamerkan kemahiran mereka menggunakan bahasa tabii dengan berkesan. Pendekatan yang diperkenalkan menggunakan keberkesanan dialog komputer-manusia untuk mengekstrak maklumat yang diperlukan daripada ontologi. Pendekatan ketidaktentuan tapisan ganda tiga yang diperkenalkan ini, telah menyemai mekanisma pembelajaran menerusi pembelajaran daripada pertanyaan pengguna dan penambahbaikan berterusan kepada prestasi kebolehannya. Algoritma ketidaktentuan tapisan ganda tiga dibangunkan bersama dengan dua dokumen iaitu katalog kesalingbolehtukaran kata (*terms equivalence catalogue (TEC))* dan konsep storan yang dipertingkat (*enhanced concepts store (ECS)* yang mewakili tesaurus dan leksikon, untuk digunakan dengan set data *Mooney Geoquery*. Kesemua tersebut adalah tersedia untuk digunakan oleh para penyelidik.

ACKNOWLEDGEMENTS

Alhamdulillah for His endless blessings and mercies. He provided me with the strength and inspiration to remain steadfast in this research journey.

I wish to express my sincere gratitude to my supervisor, Ass. Prof. Masrah Azmi Murad, and the other members of the supervisory team, Ass. Prof. Nurfadhlina Mohd Sharef and Dr. Aida Mustapha. Their guidance and encouragement saw to the successful completion of this research work.

My association with members of the Applied Informatics Research Group (AiRG) over these past years has greatly enriched me. I am indeed indebted to both staff and student members of the group, all other staff and students of the Faculty of Computer Science and Information Technology, and also to the many (known and unknown to me) who contributed towards the success of this research work. May you all be rewarded in a greater measure.

A special and sincere thank you to Dr. Danica Damljanovic and the team at GATE for granting me access to the FREyA system, to enable me conduct the validation experiments. I truly appreciate your support.

My eternal gratitude goes to my families (the Isa Agaies, my wife and the extended family): I sailed on your prayers, love and support to attain this feat. I am blessed to have you all in my life.

I certify that a Thesis Examination Committee has met on 14 December 2017 to conduct the final examination of Aliyu Isah Agaie on his thesis entitled "Query Disambiguation Approach using Triple-Filter" 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 Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Hazura binti Zulzalil, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Azreen bin Azman, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

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

Professor Juan Luis Castro Pena, PhD
Professor
University of Granada
Spain
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 28 March 2018

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

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

Nurfadhlina Bt. Mohd Sharef, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

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

ROBIAH BINTI YUNUS, 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 the 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 software.

Signature: _____ Date: _____

Name and Matric No.: Aliyu Isah Agaie, GS35475

Declaration by Members of Supervisory Committee

This is to confirm that:

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

Signature : _____
Name of Chairman of Supervisory Committee : Associate Professor Dr. Masrah Azrifah Azmi Murad

Signature : _____
Name of Member of Supervisory Committee : Associate Professor Dr. Nurfadhlina Bt. Mohd Sharef

Signature : _____
Name of Member of Supervisory Committee : Dr. Aida Mustapha

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ACE	Automatic Concept Extractor
AT	Ambiguous Terms
ATS	Ambiguous Terms Store
CTS	Conceptual Terms Store
ECS	Enhanced Concepts Store
FREyA	Feedback, Refinement and Extended vocabulary Aggregation
IR	Information retrieval
IT	Identified Term
ITS	Identified Terms Store
NLI	Natural Language Interface
NLIDB	Natural Language Interface to Relational Database
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
OWL	Web Ontology Language
PCT	Potential Conceptual Terms Store
POS	Part-of-Speech Tagging
QuFA	Query Formulation Assistant
RDF	Resource Description Framework
SPARQL	Simple Protocol And RDF Query Language
SQL	Structured Query Language
TEC	Terms Equivalence Catalogue
URI	Uniform Resource Identifier
UT	Unidentified Term
UTS	Unidentified Terms Store
Wup	Wu and Palmer Similarity Measure
W3C	World Wide Web Consortium
XML	Extensible Mark-up Language

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Present day search engines provide access to information storages and return huge amounts of information in response to queries posed to them. Some of the results retrieved are usually irrelevant, requiring the users to peruse through the returned documents in order to get the information they require. In many instances, the information required is contained in more than one document, perhaps because the query posed is made up of multiple sub-topics. If the information required that are of relevance to the sub-topics are found amongst different documents, this will lead to the retrieval of all such documents (Guan, 2013).

In traditional keyword search systems, the systems are usually not aware of the context in which a question is being asked. The systems basically perform string similarity matching and thus retrieve all documents that contain the entered keywords. Therefore the user needs to know the exact keywords to use in order to obtain some relevant information. As an example, when a user enters the query "how much is apple?", the search engine is unaware of the context of the question and therefore returns all results containing apple as a fruit, apple as a computer product and even the company Apple. The user then has to sieve through the retrieved results in order to get the information desired. Presenting the user with so many documents to peruse may lead to lack of satisfaction.

The World Wide Web Consortium (W3C) introduced the notion of the semantic web, to help in overcoming the limitations found in traditional keyword systems. The W3C is an international community that develops open standards to ensure the long-term growth of the web. It is a group that sets the standard format for the semantic web. The W3C provides representations that model ontology into machine-understandable format, known as RDF, for computer applications to use in making inferences (Tauberer, 2008). The present version of the web was proposed to be extended to cater for the semantic web. In the proposed approach, data are made explicit by adding meaning to them and stored in a well-defined structure that models the meaning of information on the web (Zou et al., 2004). This is to permit semantic search of information, whereby computers will understand users' information needs and return only relevant results; such results will be based on the concepts contained in the query and not keywords (Solskinnsbakk, 2012). Semantic Web Technology allows computers and humans to interact seamlessly in such a way that when a user poses a query, it is first translated into the same format as the stored information to enable the computer to understand and process the query. The computer then retrieves the information desired, which is highly relevant to the query unlike in the case of the traditional search systems. The semantic web concept is applicable in various fields

such as machine learning, databases, information retrieval, natural language processing, etc.

The semantic web data is stored in Resource Description Framework (RDF) format. The RDF is a language recommended by the W3C for representing data on the semantic web. RDF data are stored as ontologies and represented in a triple structure of "subject, predicate, object". An ontology can be viewed as basically a collection of objects (or concepts) that exist in a given domain and the intra relationships that exist between them. Ontology concepts are usually annotated, and stored in repositories known as knowledge repositories or knowledgebases, to permit manipulation and querying. Concept notation involves various procedures of annotation that make the concept explicit e.g. addition of meaning and relationship between concepts (Ciccarese et al., 2011).

Some knowledge of formal computer languages, such as SPARQL, is required in order to work with structured information. Just as SQL is the standard query language for manipulating and retrieving information from relational databases, so also is SPARQL for knowledgebase. For information to be retrieved from the knowledgebase, the queries posed by users have to be transformed into the same format as the information contained in the knowledgebase. The semantic representation of user queries gives the search systems a better understanding of the needs of the users and therefore allows them to retrieve very relevant information from the knowledgebase. The implication is therefore that users will have to learn to use the complex syntax of structured queries in order to retrieve information from knowledgebases; this certainly is a huge challenge.

Presently, the information world is moving towards the integration of these different knowledgebases, which contain a lot of structured information. In order to facilitate access to, and to permit the utilization of the massive information stored in these ontologies, a natural language interface (NLI) is used (Habernal & Konopík, 2013; Kaufmann & Bernstein, 2010). A natural language interface (NLI) provides the platform for man and machine to interact. A user enters a query in his language and this is translated into a form understandable by the computer. The computer then processes the user's query and retrieves the exact information desired by the user. The studies in (Kaufmann & Bernstein, 2010; Tablan et al., 2008) show users preference of NLIs over formal languages. It is noteworthy that unlike search engines that return a list of web pages or some documents that may have answer to the query, an NLI provides precise answers since it is designed to locate targeted answers; therefore the results obtained are highly relevant.

Besides not having knowledge of the structure of the knowledgebase from which information is to be retrieved, majority of users do not also possess the relevant technical skills needed to deal effectively with such structured information. NLIs are therefore built to provide casual users who lack the needed technical knowledge, with some means of gaining access to the information stored in knowledge bases (Lei et al.,

2006). They are designed to deal with natural language articulation of what the user needs, and then they translate this need into a formal machine language that will determine how the desired task is to be achieved. However, language ambiguity which arises as a result of the complex nature of human language remains an issue. All NLIs have scopes and limitations that are not known to everyday users (K. Elbedweihy et al., 2015; Rangel & Aguirre, 2014). It is thus understood that users may be blind to errors that may appear in their questions. To retrieve the correct information, the queries must be written appropriately and within the scope and limits of the NLI. Some of these include adherence to syntax, challenges of inconsistencies in knowledge base and web data, limited support for negation queries, lack of full expressiveness, user guidance, etc.

The effective retrieval of any piece of information from an ontology depends wholly on the correct mapping of queries made in natural language to machine understandable form. More often than not, the query terms are expected to correctly match the ontology concepts and instance labels to be identified (the sequence in a text string must exactly match that of the backend, including whether the character is in upper or lower case). While some NLI systems attempt to automatically fix the errors in spellings, some other systems permit users to choose the ontology property names close to their intention from a list of suggestions. Both approaches have their setbacks. In the first, automatically fixing errors may lead to wrong interpretation where the supposed correction is also not right; whereas in the second, users may be confused with the property names used which result in them choosing from the suggestion list randomly and lead to inaccurate results (Calì et al., 2011; Revuelta-Martínez et al., 2013; Sharef & Noah, 2012).

This thesis describes an approach that guides the user to formulate his query in order to retrieve the correct information. In the proposed triple-filter query disambiguation approach, the user is guided to write his query right from the beginning in order to avoid errors that may be introduced into the query during the writing process. As the user keys in the words, a list of suggestions similar to what is being typed is presented to him. As the typing progresses, the words that best match the word being typed are continuously shown to the user in a wild card format. The user is able to choose from the list by clicking on the preferred word or select by highlighting with the cursor and pressing enter. The process is repeated for all the terms until the query construction is completed. After the query construction process is completed, the disambiguation process is activated. This involves a three-step process (triple-filter) that ensure that the intention of the user is clearly understood. The process clarifies all ambiguities before rewriting the query to a form that is understandable by the underlying search engine.

In order to validate the proposed triple-filter query disambiguation approach, a prototype known as Query Formulation Assistant (QuFA) was developed. The proposed triple-filter query disambiguation approach in QuFA was directly compared with the approach in FREyA (Damljanović et al., 2013) that also provides support to

users when formulating queries. Two experiments were conducted to validate the proposed approach: usability study and performance evaluation. The set of queries for the Mooney Geoquery data set were used in the experiments. The original dataset is available from ("Natural Language Learning Data," n.d.), while the data set used in this research is available from ("Datasets - Natural Language Interfaces," n.d.). Since search is an activity that is user-centric, the usability study focused on the users' experience. The user experience is quite important because it will lead to acceptance or rejection of the system by the users. On the other hand, the performance evaluation was based on an objective assessment of the effectiveness of the proposed approach to retrieve answers either automatically or by involving the user in a clarification dialogue in order to retrieve the correct answer. The effectiveness was measured in terms of precision and recall.

1.2 Motivation

The need to provide users who do not possess technical skills with the ability to access information stored in knowledge bases is the primary motivation for this research. The massive information stored in ontologies will be of no use if users have difficulty in accessing them. Since the information is structured, users usually do not know how to ask their questions in order to retrieve appropriate answers (Habernal & Konopík, 2013). There is therefore the need to pay attention to how the user inputs a query, in order to ease the users' information retrieval tasks.

Existing NLIs are plagued with expressivity and cognitive burden issues. Some of the few existing systems that attempt to provide guidance to users, only end up compounding the issues. The systems restrict users to the use of special terms of entities and properties (labels) which are not consistent with the understanding of the users. Users should be able to use their own vocabularies and still retrieve correct information from knowledgebases, in response to their queries. The input interface should support user expressiveness, allowing the use of terms that are familiar to the user, and not restricting the user to terms existing in the systems' vocabulary alone. This research work will attempt to provide a solution to these challenges.

Another major motivation for the research in this thesis is the need to effectively interpret user queries in order to retrieve correct answers. Most of the existing NLIs attempt to automatically correct all errors found in a user query, which sometimes lead to entity and property mismatch during the process of query translation and retrieval of answers. Sacrificing clarity in an attempt to automate all the processes of an NLI is not worth it, since this may lead to the retrieval of wrong answers. Besides, the effectiveness of clarification dialogues between humans and computers should not be disregarded in an attempt to achieve automation. This research therefore intends to propose an approach, and provide a tool based on the proposed approach, that is effective in interpreting user query. Since search is a user centric activity, the user needs to be incorporated into the search process in order to ascertain his intention. This will lead to correct interpretation of the user query, and result in the effective retrieval of desired information.

1.3 Problem Statement

The rise in the volume of structured data on the web poses a challenge of how to access and utilize the information in knowledgebases. It also threatens the very existence of present day search engines that are primarily keyword-based. These search engines retrieve a lot of documents that are irrelevant to the queries posed by users because they do not take the context of the queries into consideration. If they are to survive, they will have to adapt to the paradigm of the semantic web.

The semantic web technology was introduced in response to the limitations of the present search engines that are based on keyword. It converts data into structured format that are stored in knowledge repositories. However, knowledge of SPARQL, a formal computer language, is required in order to effectively retrieve the information stored in knowledgebases. Users will need to know the structural representation of the documents in the knowledgebase, so that they can formulate their queries in the same pattern, in order to retrieve the desired information (Jarrar & Dikaiakos, 2012). To overcome this challenge, users query input in natural language has to be transformed into the same format as the documents in the knowledgebase.

Natural language interfaces (NLI) were introduced to provide a platform for users to use natural language to retrieve information from knowledgebases. Over the years, several natural language interfaces have been built in order to facilitate access to ontologies by casual users. Although a good interface is expected to support user expressiveness (K. Elbedweihy et al., 2015; Pazos R. et al., 2013), these NLIs have scopes and limitations which casual users are unaware of, and they also impose some restrictions on users. This results in a mismatch between what the user expects of the NLI and the actual capabilities of the system (Kaufmann & Bernstein, 2010; Abraham Bernstein & Kaufmann, 2006). As such, there is the need to overcome these constraints in order to ease users' information retrieval tasks.

During the process of query construction, some of the existing NLIs provide user guidance features such as the use of auto-complete and predictive text writing (Ferré, 2016; Llopis & Ferrández, 2013; Revuelta-Martínez et al., 2013; Calì et al., 2011). All of these approaches have their shortcomings. The use of auto-complete feature is best suited for information retrieval tasks in a domain dependent system where the choices are limited: it provides assistance to carry out configured tasks. To adapt this approach to another domain will require heavy customization. Also, providing suggestions which are made up of special terms of entities and properties (labels) that are not consistent with the understanding of the users, will only lead to more confusion on the part of the users.

Ambiguity in natural language is an issue for natural language interfaces. Although some existing NLI systems permit users to choose the ontology property names close to their intention from a list of suggestions (Damljanović et al., 2013; Lopez et al., 2012) the hype in automation has led to attempts by recent researches (Kadir & Yauri,

2017; Khiroun et al., 2014) in natural language interfaces to automatically resolve all ambiguities found in user queries. Both approaches have their setbacks. In the first, users may be confused with the property names used which result in them choosing from the suggestion list randomly and lead to inaccurate results (Sharef & Noah, 2012). In some cases (Damljanović et al., 2013; Llopis & Ferrández, 2013) suggestions that are supposed to help the dialogue between the computer and the user add cognitive burden on the user; this is counter-productive. Whereas in the second case, automatically fixing errors may lead to wrong interpretation where the supposed auto-correction is also not right (Kadir & Yauri, 2017; Khiroun et al., 2014); it results in entity and property mismatch. This research is concerned about guiding the user to achieve accurate spellings in the first place, since automatically correcting all errors found in a user query sometimes lead to entity and property mismatch during the process of query translation and retrieval of answers.

Correctly interpreting user input queries remains a major challenge for NLIs. Although the goal of semantic search is to give users the capability of expressing their information needs using full natural language, the inherent ambiguity in natural language is still a hindrance to this realization. Therefore, in order to correctly interpret the input query, it is rewritten (Craswell et al., 2013; Purnamasari et al., 2016). Query rewriting is aimed at the correct interpretation of input queries in order to effectively retrieve the correct information. In some approaches such as in (Damljanović et al., 2013), the user is put in control of the query rewriting process, however, this is not suitable for casual users. The users will have to know about semantic technologies and be familiar with the ontology being queried in order to effectively rewrite the query. To rewrite queries in other approaches such as (Craswell et al., 2013) require the use of a large anchor graph that serves as a linguistic resource. The demerit here is that the input query needs to exist within the anchor data. Consequently, if there is no match, then the reformulation fails. While query rewriting improves search relevance, it is actually improving recall over precision (Daniel Tunkelang, 2017; Hugh E. Williams, 2012). There is therefore the need to ensure improved precision through query reformulation so that users will be satisfied with natural language search systems (NLIs).

1.4 Research Objectives

Some challenges have been outlined in the previous section. To overcome these problems, this research focusses on the following objectives:

(1) To enhance correct query formulation by guiding the user through the query writing process, using domain space based query authoring service.
(2) To propose a new disambiguation approach that will improve the effectiveness of query interpretation.
(3) To evaluate the overall approach of the research through a prototype, by testing its usability and its effectiveness (in terms of precision and recall).

1.5 Research Scope

For the purpose of this thesis, the proposed approach is limited to the provision of support for free text queries for domain based NLIs. It will serve as a bridge between the user and the NLI. The proposed approach will support users during query writing process by providing the users with domain based suggestions. It will also attempt to resolve ambiguities found in their queries through clarification dialogues and/or automatically. The proposed approach will neither translate input queries into structured queries nor retrieve answers; instead, the input queries will be reformulated (after resolving ambiguities) and passed on to the underlying search engine (NLI) that will retrieve the required information.

1.6 Research Contribution

The major contributions from this research are as described below:

(1) The proposed approach provides a means of exploring the domain space without the need for knowing the exact conceptual terms used in storing the information in knowledgebases. Before a domain knowledge is stored in a knowledgebase, the domain concepts are first annotated. The concept notation includes adding entity properties and relationships using labels that are not readily understood by humans. In the proposed approach, these labels are completely hidden but the user could have a good grasp of the content of the ontology by exploring the equivalent knowledgebase terms that are used to enrich the conceptual terms.

(2) The proposed approach permits the engagement of the user in clarification dialogues in order to ascertain the intention of the user, and does not require users to accept system based suggestions. When ambiguities such as spelling errors (typos) are allowed to be automatically corrected, they may lead to entity and property mismatch during the process of retrieving an answer, due to wrong interpretation of the input query. The effectiveness of the human-computer dialogue leads to the effective retrieval of desired information, as such, it is not worth sacrificing clarity in order to achieve automation.

(3) The proposed approach inculcates a learning mechanism that continue to automatically learn from user queries and continuously improves its performance capability. Once an ambiguity is successfully resolved, it is automatically associated with its equivalent concept in the knowledgebase. Therefore, when next the same term that was earlier considered to be ambiguous in a query is encountered, it will automatically be recognized.

(4) In the course of this research, a tool (besides the proposed prototype system), some documents and some algorithms were developed in order to attain the objectives of the research. The tool, automatic concept extractor (ACE) as the name implies, extracts concepts from a list of competency questions. It is available for use by other researchers that may need to extract concepts from

documents for the purpose of their work. The documents that were developed are the terms equivalence catalogue (TEC) and the enhanced concepts store (ECS). These two documents represent the thesaurus and the lexicon for use with the Mooney Geoquery dataset. The algorithms for concepts identification, triple-filter disambiguation, computing intermediate query and overall approach were also developed. All of these are available for use by other researchers.

1.7 Thesis Organization

The structure of remaining parts of this thesis is as follows:

Chapter 2 provides an overview of the semantic web and information retrieval. It then discusses natural language interfaces as tools for semantic search. It also identified and analyzed some of the challenges faced by natural language interfaces in supporting casual users to retrieve desired information.

In Chapter 3, the methodology of the approach used in this research is provided. It describes how the research problems were identified, the design of the proposed solution to the identified challenges and the process of validating the proposed solution. Chapter 4 presents a detailed discussion of the proposed triple-filter query disambiguation approach. The details of how user intention is ascertained and how a query is interpreted using the proposed approach are provided.

A detailed elaboration of the evaluation of the proposed triple-filter query disambiguation approach is presented in Chapter 5. The proposed approach is compared with the approach presented in FREyA, based on system usability and performance evaluation. Comprehensive experiments and the results obtained are presented here, along with detailed discussion and analysis of the results.

Finally, Chapter 6 presents the conclusion reached in regards to the proposed triple-filter query disambiguation approach described in this thesis. It summarizes the contributions of this research and provides plans for future work.

REFERENCES

Agirre, E., López de Lacalle, O., & Soroa, A. (2014). Random Walks for Knowledge-Based Word Sense Disambiguation. *Computational Linguistics*, *40*(1), 57–84. https://doi.org/10.1162/COLI_a_00164

Ahmed, Z., & Gerhard, D. (2010). Web to Semantic Web & Role of Ontology. In *National Conference on Information and Communication Technologies, (NCICT-2007)* (pp. 100–102). Pakistan. Retrieved from https://pdfs.semanticscholar.org/3b29/cc72beed0877da2b1de84c721e915b8db219.pdf

Al-Shboul, B., & Myaeng, S.-H. (2014). Wikipedia-based query phrase expansion in patent class search. *Information Retrieval*, *17*(5–6), 430–451. https://doi.org/10.1007/s10791-013-9233-4

Allainé, D., & Theuriau, F. (2004). Is there an optimal number of helpers in Alpine marmot family groups? *Behavioral Ecology*, *15*(6), 916–924. https://doi.org/10.1093/beheco/arh096

Alroobaea, R., & Mayhew, P. J. (2014). How many participants are really enough for usability studies? In *Science and Information Conference* (pp. 48–56). IEEE. https://doi.org/10.1109/SAI.2014.6918171

Andoni, A., Krauthgamer, R., & Onak, K. (2010). Polylogarithmic approximation for edit distance and the asymmetric query complexity. In *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS* (pp. 377–386). IEEE. https://doi.org/10.1109/FOCS.2010.43

Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). Natural language interfaces to databases – an introduction. *Natural Language Engineering*, *1*(1), 29–81. https://doi.org/10.1017/S135132490000005X

Anick, P. (2003). Using terminological feedback for web search refinement. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 88–95). New York, USA: ACM. https://doi.org/10.1145/860435.860453

Aouicha, M. Ben, Taieb, M. A. H., & Hamadou, A. Ben. (2016). SISR: System for integrating semantic relatedness and similarity measures. *Soft Computing*, 1–25. https://doi.org/10.1007/s00500-016-2438-x

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4825 LNCS, pp. 722–735). https://doi.org/10.1007/978-3-540-76298-0_52

Azad, H. K., & Deepak, A. (2017). Query Expansion Techniques for Information Retrieval: a Survey. *arXiv Preprint arXiv:1708.00247*. Retrieved from http://arxiv.org/abs/1708.00247

Backurs, A., & Indyk, P. (2015). Edit Distance Cannot Be Computed in Strongly Subquadratic Time (unless SETH is false). In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* (pp. 51–58). ACM. Retrieved from https://arxiv.org/pdf/1412.0348.pdf

Baeza-Yates, R., Hurtado, C., & Mendoza, M. (2004). Query recommendation using query logs in search engines. In *EDBT workshops* (Vol. 3268, pp. 588–596). Retrieved from https://s3.amazonaws.com/academia.edu.documents/33297174/clustwebLNCS.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1508392162&Signature=7xouBPcPlAkMinTaYr0czICVZ%2Fo%3D&response-content-disposition=inline%3B filename%3DQuery_Recommendation_Using_Query_

Ballatore, A., Bertolotto, M., & Wilson, D. C. (2014). An evaluative baseline for geo-semantic relatedness and similarity. *GeoInformatica*, *18*(4), 747–767. https://doi.org/10.1007/s10707-013-0197-8

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Bard, G. V. (2007). Spelling-error tolerant, order-independent pass-phrases via the Damerau-Levenshtein string-edit distance metric. In *Proceedings of the fifth Australasian symposium on ACSW frontiers* (Vol. 68, pp. 117–124). Australian Computer Society, Inc. Retrieved from https://eprint.iacr.org/2006/364.pdf

Belkin, N. J. (1996). Intelligent information retrieval: Whose intelligence? In *ISI '96: Proceedings of the Fifth International Symposium for Information Science* (pp. 25–31). Retrieved from https://core.ac.uk/download/pdf/11540431.pdf#page=23

Benabderrahmane, S., Smail-Tabbone, M., Poch, O., Napoli, A., & Devignes, M.-D. (2010). IntelliGO: a new vector-based semantic similarity measure including annotation origin. *BMC Bioinformatics*, *11*(1), 588. https://doi.org/10.1186/1471-2105-11-588

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43. https://doi.org/10.1038/scientificamerican0501-34

Bernstein, A., & Kaufmann, E. (2006). GINO – A Guided Input Natural Language Ontology Editor. In *The Semantic Web-ISWC 2006* (Vol. LNCS 4273, pp. 144–157). https://doi.org/10.1007/11926078_11

Bernstein, A., Kaufmann, E., Kaiser, C., & Kiefer, C. (2006). Ginseng: A guided input natural language search engine for querying ontologies. In *2006 Jena User Conference* (pp. 2–4). Bristol, UK. Retrieved from https://www.merlin.uzh.ch/contributionDocument/download/2188%0A

Börner, K. (2003). Visual Interfaces for Semantic Information Retrieval and Browsing. In V. Geroimenko & C. Chen (Eds.), *Visualizing the Semantic Web* (pp. 99–115). Springer London. https://doi.org/10.1007/978-1-4471-3737-5_7

Boytsov, L. (2011). Indexing methods for approximate dictionary searching: Comparative Analysis. *Journal of Experimental Algorithmics*, *16*(1–1). https://doi.org/10.1145/1963190.1963191

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*. Retrieved from https://pdfs.semanticscholar.org/13dd/d0ede672d91d905e56c52ed73216b17cfc81.pdf

Budanitsky, A., & Hirst, G. (2001). Semantic distance in WordNet : An experimental , application-oriented evaluation of five measures. In *Workshop on WordNet and Other Lexical Resources* (Vol. 2, pp. 2–2). Retrieved from ftp://www-vhost.cs.toronto.edu/public_html/public_html/pub/gh/Budanitsky+Hirst-2001.pdf

Cai, F., & de Rijke, M. (2016). A Survey of Query Auto Completion in Information Retrieval. *Foundations and Trends® in Information Retrieval*, *10*(4), 273–363. https://doi.org/10.1561/1500000055

Calì, D., Condorelli, A., Papa, S., Rata, M., & Zagarella, L. (2011). Improving intelligence through use of Natural Language Processing. A comparison between NLP interfaces and traditional visual GIS interfaces. *Procedia Computer Science*, *5*, 920–925. https://doi.org/10.1016/j.procs.2011.07.128

Camacho-Collados, J., Pilehvar, M. T., & Navigli, R. (2015a). A Unified Multilingual Semantic Representation of Concepts. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Vol. 1, pp. 741–751). ACL. Retrieved from http://www.aclweb.org/anthology/P15-1072

Camacho-Collados, J., Pilehvar, M. T., & Navigli, R. (2015b). NASARI: a Novel Approach to a Semantically-Aware Representation of Items. In *Human Language Technologies: The Annual Conference of the North American Chapter of the ACL* (pp. 567–577). ACL. Retrieved from http://lcl.uniroma1.it/nasari/.

Carpineto, C., & Romano, G. (2012). A Survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys*, *44*(1), 1–50. https://doi.org/10.1145/2071389.2071390

Chen, H. (2012). *String Metrics and Word Similarity applied to Information Retrieval*. University of Eastern Finland. Retrieved from http://epublications.uef.fi/pub/urn_nbn_fi_uef-20120382/urn_nbn_fi_uef-20120382.pdf

Ciccarese, P., Ocana, M., Garcia Castro, L. J., Das, S., & Clark, T. (2011). An open annotation ontology for science on web 3.0. *Journal of Bomedical Semantics*, *2*(Suppl 2), S4. https://doi.org/10.1186/2041-1480-2-S2-S4

Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string metrics for matching names and records. *KDD Workshop on Data Cleaning and Object Consolidation*, *3*, 73–78. Retrieved from https://www.cs.cmu.edu/afs/cs/Web/People/wcohen/postscript/kdd-2003-match-ws.pdf

Copestake, A., & Jones, K. S. (2009). Natural Language Interfaces to Databases. *Journal of Natural Language Engineering*, (709), 50. https://doi.org/10.1017/S0269888900005476

Couto, F. M., & Pinto, H. S. (2013). The Next Generation of Similarity Measures that Fully Explore the Semantics in Biomedical Ontologies. *Journal of Bioinformatics and Computational Biology*, *11*(5), 1371001. https://doi.org/10.1142/S0219720013710017

Couto, F. M., & Silva, M. J. (2011). Disjunctive shared information between ontology concepts: application to Gene Ontology. *Journal of Biomedical Semantics*, *2*(1), 5. https://doi.org/10.1186/2041-1480-2-5

Craswell, N., Billerbeck, B., Fetterly, D., & Najork, M. (2013). Robust query rewriting using anchor data. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 335–344). New York, USA: ACM. https://doi.org/10.1145/2433396.2433440

Croft, W. B. (1986). User-specified domain knowledge for document retrieval. In *Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '86* (pp. 201–206). New York, USA: ACM Press. https://doi.org/10.1145/253168.253211

D'Amato, C. (2007). *Similarity-based Learning Methods for the Semantic Web*. *Doctoral Dissertation*. Universita delgi Studi di Bari, Italy. Retrieved from https://pdfs.semanticscholar.org/b274/8ccf769794d68ef6e07f60b54a25df9233de.pdf

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171–176. https://doi.org/10.1145/363958.363994

Damljanović, D., Agatonović, M., Cunningham, H., & Bontcheva, K. (2013). Improving habitability of natural language interfaces for querying ontologies with feedback and clarification dialogues. *Web Semantics: Science, Services and Agents on the World Wide Web*, *19*, 1–21. https://doi.org/10.1016/j.websem.2013.02.002

Damljanovic, D. D. (2011). *Natural Language Interfaces to Conceptual Models*. The University of Sheffield, England. Retrieved from http://etheses.whiterose.ac.uk/1630/2/Damljanovic%2C_Danica.pdf

Daniel Tunkelang. (2017). Query Rewriting: An Overview. Retrieved January 11, 2018, from https://queryunderstanding.com/query-rewriting-an-overview-d7916eb94b83

Datasets - Natural Language Interfaces. (n.d.). Retrieved August 25, 2017, from https://sites.google.com/site/naturallanguageinterfaces/freya/data

Decker, S., Sintek, M., Billig, A., Henze, N., Harth, A., Leicher, A., … Neumann, G. (2005). TRIPLE - an RDF Rule Language with Context and Use Cases. In *W3C Workshop on Rule Languages for Interoperability* (pp. 1–6).

Di Santo, G., McCreadie, R., Macdonald, C., & Ounis, I. (2015). Comparing Approaches for Query Autocompletion. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 775–778). New York, USA: ACM. https://doi.org/10.1145/2766462.2767829

Diaz, F., Mitra, B., & Craswell, N. (2016). Query Expansion with Locally-Trained Word Embeddings. *arXiv Preprint arXiv:1605.07891*. Retrieved from http://arxiv.org/abs/1605.07891

Dua, M., Jindal, S., Kumar, R., & Vidyapith, B. (2014). An architectural overview of Natural Language Interface to knowledge base. In *2014 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)* (pp. 437–441). IEEE. https://doi.org/10.1109/ICCPEIC.2014.6915404

El-Deeb, R. H., Hegazy, A. F. A., & Fahmy, A. A. (2012). Semantic form-based guided search system. In *2012 22nd International Conference on Computer Theory and Applications (ICCTA)* (pp. 85–93). IEEE. https://doi.org/10.1109/ICCTA.2012.6523552

Elbedweihy, K. M., Wrigley, S. N., Clough, P., & Ciravegna, F. (2015). An overview of semantic search evaluation initiatives. *Web Semantics: Science, Services and Agents on the World Wide Web*, *30*, 82–105. https://doi.org/10.1016/j.websem.2014.10.001

Elbedweihy, K., Mazumdar, S., Wrigley, S. N., & Ciravegna, F. (2014). NL-Graphs: A Hybrid Approach toward Interactively Querying Semantic Data. In T. A. Presutti V., d'Amato C., Gandon F., d'Aquin M., Staab S. (Ed.), *The Semantic Web: Trends and Challenges. ESWC 2014. Lecture Notes in Computer Science* (pp. 565–579). Springer, Cham. https://doi.org/10.1007/978-3-319-07443-6_38

Elbedweihy, K., Wrigley, S. N., & Ciravegna, F. (2012). Evaluating Semantic Search Query Approaches with Expert and Casual Users. In P. Cudré-Maurouxet (Ed.), *The Semantic Web – ISWC 2012. ISWC 2012. Lecture Notes in Computer Science, vol 7650* (pp. 274–286). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35173-0_18

Elbedweihy, K., Wrigley, S. N., Ciravegna, F., Reinhard, D., & Bernstein, A. (2015). Evaluating Semantic Search Systems to Identify Future Directions of Research. In *Proceedings of the Second International Workshop on Evaluation of Semantic Technologies (IWEST 2012)* (Vol. 843, pp. 148–162). https://doi.org/10.1007/978-3-662-46641-4_11

Elbedweihy, K., Wrigley, S. N., Ciravegna, F., & Zhang, Z. (2013). Using BabelNet in bridging the gap between natural language queries and linked data concepts. *CEUR Workshop Proceedings*, *1064*(October 2013).

Ermakova, L., Mothe, J., & Ovchinnikova, I. (2014). Query expansion in information retrieval : What can we learn from a deep analysis of queries ? In *International Conference on Computational Linguistics - Dialogue 2014*. Moscow, Russian Federation. Retrieved from http://oatao.univ-toulouse.fr/13097/1/Ermakova_13097.pdf

Fähndrich, J., Weber, S., & Ahrndt, S. (2016). Design and Use of a Semantic Similarity Measure for Interoperability Among Agents. In M. Klusch, U. R., S. O., P. A., & A. S. (Eds.), *Multiagent System Technologies. MATES 2016. Lecture Notes in Computer Science* (Vol. 9872, pp. 41–57). Springer, Cham. https://doi.org/10.1007/978-3-319-45889-2_4

Fan, J., Wu, H., Li, G., & Zhou, L. (2010). Suggesting Topic-Based Query Terms as You Type. In *2010 12th International Asia-Pacific Web Conference* (pp. 61–67). IEEE. https://doi.org/10.1109/APWeb.2010.13

Fazzinga, B., & Lukasiewicz, T. (2010). Semantic Search on the Web. *Semantic Web*, *1*(1,2), 89–96. https://doi.org/10.3233/SW-2010-0023

Feng, Y., Bagheri, E., Ensan, F., & Jovanovic, J. (2017). The state of the art in semantic relatedness: a framework for comparison. *The Knowledge Engineering Review*, 1–30. https://doi.org/10.1017/S0269888917000029

Ferré, S. (2016). Sparklis: An expressive query builder for SPARQL endpoints with guidance in natural language. *Semantic Web*, *8*(3), 405–418. https://doi.org/10.3233/SW-150208

Ferreira, J. D., & Couto, F. M. (2010). Semantic similarity for automatic classification of chemical compounds. *PLoS Computational Biology*, *6*(9), e1000937. https://doi.org/10.1371/journal.pcbi.1000937

Ferreira, J. D., & Couto, F. M. (2011). Generic semantic relatedness measure for biomedical ontologies. In *International Conference on Biomedical Ontology* (Vol. 833, pp. 117–123). Retrieved from http://ceur-ws.org/Vol-833/paper16.pdf

Finin, T., Mayfield, J., Joshi, A., Cost, R. S., & Fink, C. (2005). Information Retrieval and the Semantic Web. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (p. 113a–113a). IEEE. https://doi.org/10.1109/HICSS.2005.319

Frakes, W. B., & Baeza-Yates, R. (1992). *Information Retrieval Data Structures & Algorithms*. Englewood Cliffs, New Jersey: Prentice Hall. Retrieved from https://theswissbay.ch/pdf/Gentoomen Library/Information Retrieval/Information Retrieval Data Structures And Algorithms_FRAKES WB %282004%29.pdf

Fuhr, N. (2004). Information Retrieval Methods for Literary Texts. Retrieved from http://www.is.inf.uni-due.de/bib/pdf/ir/Fuhr_03.pdf

Fung, W. J. (2010). A Predictive Text Completion Software in Python. In *Proceedings of PyCon Asia-Pacific 2010*. Python Papers Monograph, 2.

Gilleland, M. (2009). *Levenshtein Distance, in Three Flavors*. *Merriam Park Software*. Retrieved from http://people.cs.pitt.edu/~kirk/cs1501/Pruhs/Fall2006/Assignments/editdistance/Levenshtein Distance.htm

Gracia, J., & Mena, E. (2008). Web-Based Measure of Semantic Relatedness. In *Web Information Systems Engineering* (pp. 136–150). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85481-4_12

Guan, D. (2013). *Structured Query Formulation and Result Organization for Session Search*. *PhD Thesis*. Georgetown University, Washington, DC. Retrieved from http://infosense.cs.georgetown.edu/publication/dongyi_guan_thesis.pdf

Guessoum, D., Miraoui, M., & Tadj, C. (2016). A modification of Wu and Palmer Semantic Similarity Measure. In *UBICOMM 2016 : The Tenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies* (pp. 42–46). Retrieved from https://www.researchgate.net/publication/310572659_A_modification_of_Wu_and_Palmer_Semantic_Similarity_Measure

Guha, R., McCool, R., & Miller, E. (2003). Semantic search. In *Proceedings of the Twelfth International Conference on World Wide Web - WWW '03* (pp. 700–709). Budapest, Hungary: ACM Press. https://doi.org/10.1145/775152.775250

Guzzi, P. H., Mina, M., Guerra, C., & Cannataro, M. (2011). Semantic similarity analysis of protein data: assessment with biological features and issues. *Briefings in Bioinformatics*, *13*(5), 569–585. https://doi.org/10.1093/bib/bbr066

Habernal, I., & Konopík, M. (2013). SWSNL: Semantic Web Search Using Natural Language. *Expert Systems with Applications*, *40*(9), 3649–3664. https://doi.org/10.1016/j.eswa.2012.12.070

Hakimov, S., Tunc, H., Akimaliev, M., & Dogdu, E. (2013). Semantic question answering system over linked data using relational patterns. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops* (pp. 83–88). New York, USA: ACM. https://doi.org/10.1145/2457317.2457331

Hakimov, S., Unger, C., Walter, S., & Cimiano, P. (2015). Applying Semantic Parsing to Question Answering Over Linked Data: Addressing the Lexical Gap. In M. E. Biemann C., Handschuh S., Freitas A., Meziane F. (Ed.), *Natural Language Processing and Information Systems. NLDB 2015. Lecture Notes in Computer Science* (Vol. 9103, pp. 103–109). Springer, Cham. https://doi.org/10.1007/978-3-319-19581-0_8

Hamming, R. W. (1950). Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, *29*(2), 147–160. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x

Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2015). Semantic Similarity from Natural Language and Ontology Analysis. *Synthesis Lectures on Human Language Technologies*, *8*(1), 1–254. Retrieved from https://arxiv.org/pdf/1704.05295.pdf

Hazrina, S., Sharef, N. M., Ibrahim, H., Murad, M. A. A., & Noah, S. A. M. (2017). Review on the advancements of disambiguation in semantic question answering system. *Information Processing & Management*, *53*(1), 52–69. https://doi.org/10.1016/j.ipm.2016.06.006

Hildebrand, M., Ossenbruggen, J. Van, & Hardman, L. (2007). *An Analysis of Search-based User Interaction on the Semantic Web*. *CWI report. INS-E* (Vol. 706). Amsterdam: CWI. Retrieved from https://pure.tue.nl/ws/files/2343605/601804781996420.pdf

Hitzler, P., Krötzsch, M., & Rudolph, S. (2009). *Foundations of Semantic Web Technologies* (1st ed.). Chapman & Hall/CRC. Retrieved from https://dl.acm.org/citation.cfm?id=1655817

Hugh E. Williams. (2012). Query Rewriting in Search Engines. Retrieved January 11, 2018, from https://hughewilliams.com/2012/03/19/query-rewriting-in-search-engines/

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM*, *53*(5), 130–133. https://doi.org/10.1145/1735223.1735255

Ingason, A. K., Helgadóttir, S., Loftsson, H., & Rögnvaldsson, E. (2008). A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI). In B. Nordström & A. Ranta (Eds.), *Advances in Natural Language Processing. Lecture Notes in Computer Science* (Vol. 5221, pp. 205–216). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-85287-2_20

Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., & Bäumer, B. (2007). Algorithm, Implementation and Application of the SIM-DL Similarity Server. In *GeoSpatial Semantics* (Vol. 4853, pp. 128–145). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76876-0_9

Janowicz, K., Raubal, M., & Kuhn, W. (2011). The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, *2*(2011), 29–57. https://doi.org/10.5311/JOSIS.2011.2.3

Jarrar, M., & Dikaiakos, M. D. (2012). A Query Formulation Language for the Data Web. *IEEE Transactions on Knowledge and Data Engineering*, *24*(5), 783–798. https://doi.org/10.1109/TKDE.2011.41

Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proceedings of International Conference Research on Computational Linguistics* (pp. 19–33). Retrieved from http://arxiv.org/abs/cmp-lg/9709008

Jin, L., & Li, C. (2005). Selectivity estimation for fuzzy string predicates in large data sets. In *Proceedings of the 31st international conference on Very large data bases* (pp. 397–408). VLDB Endowment. Retrieved from http://www.vldbarc.org/archives/website/2005/program/paper/wed/p397-jin.pdf

Jones, R., Rey, B., Madani, O., & Greiner, W. (2006). Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web* (pp. 387–396). New York, USA: ACM. https://doi.org/10.1145/1135777.1135835

Kadir, R. A., & Yauri, A. R. (2017). Automated Semantic Query Formulation Using Machine Learning Approach. *Journal of Theoretical and Applied Information Technology*, *95*(12), 2761–2775.

Kassim, J. M., & Rahmany, M. (2009). Introduction to Semantic Search Engine. In *2009 International Conference on Electrical Engineering and Informatics* (Vol. 2, pp. 380–386). IEEE. https://doi.org/10.1109/ICEEI.2009.5254709

Kaufmann, E., & Bernstein, A. (2010). Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web*, *8*(4), 377–393. https://doi.org/10.1016/j.websem.2010.06.001

Kaufmann, E., Bernstein, A., & Zumstein, R. (2006). Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)* (pp. 980–981). Athens, GA. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.6633

Kaur, I., & Hornof, A. J. (2005). A comparison of LSA, wordNet and PMI-IR for predicting user click behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems CHI 05* (pp. 51–60). https://doi.org/10.1145/1054972.1054980

Kharkevich, U. (2010). *Concept Search : Semantics Enabled Information Retrieval*. *PhD Dissertation*. University of Trento. Retrieved from http://eprints-phd.biblio.unitn.it/278/1/PhD-Thesis.pdf

Khiroun, O. Ben, Elayeb, B., Bounhas, I., Evrard, F., & Saoud, B. N. Ben. (2014). Improving Query Expansion by Automatic Query Disambiguation in Intelligent Information Retrieval. In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (pp. 153–160). https://doi.org/10.5220/0004822401530160

Kraft, R., & Zien, J. (2004). Mining anchor text for query refinement. In *Proceedings of the 13th international conference on World Wide Web* (pp. 666–674). New York, USA: ACM. https://doi.org/10.1145/988672.988763

Kuck, G. (2004). Tim Berners-Lee's Semantic Web. *South African Journal of Information Management*, *6*(1). https://doi.org/10.4102/sajim.v6i1.297

Lawson, T. (2004). *A Conception of Ontology*. *The Cambridge Social Ontology*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.3252&rep=rep1&type=pdf

Leacock, C., & Chodorow, M. (1998). Combining Local Context and WordNet Similarity for Word Sense Identification. *WordNet: An Electronic Lexical Database.*, *49*(2), 265–283.

Lehmann, J., & Bühmann, L. (2011). AutoSPARQL: Let Users Query Your Knowledge Base. In G. Antoniou (Ed.), *The Semantic Web: Research and Applications* (Vol. 6643 LNCS, pp. 63–79). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21034-1_5

Lei, Y., Uren, V., & Motta, E. (2006). SemSearch: A Search Engine for the Semantic Web. In S. S. & S. V. (Eds.), *Managing Knowledge in a World of Networks. EKAW 2006. Lecture Notes in Computer Science* (Vol. 428, pp. 238–245). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11891451_22

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710). Retrieved from https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf

Li, C., Lu, J., & Lu, Y. (2008). Efficient Merging and Filtering Algorithms for Approximate String Searches. In *2008 IEEE 24th International Conference on Data Engineering* (pp. 257–266). IEEE. https://doi.org/10.1109/ICDE.2008.4497434

Li, M., Zhang, Y., Zhu, M., & Zhou, M. (2006). Exploring distributional similarity based models for query spelling correction. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics* (pp. 1025–1032). Association for Computational Linguistics. https://doi.org/10.3115/1220175.1220304

Li, Y., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, *15*(4), 871–882. https://doi.org/10.1109/TKDE.2003.1209005

Lin, D. (1993). Principle-based parsing without overgeneration. In *Proceedings of the 31st annual meeting on Association for Computational Linguistic* (pp. 112–120). Columbus, Ohio: Association for Computational Linguistics. https://doi.org/10.3115/981574.981590

Lin, D. (1998). An Information-Theoretic Definition of Similarity. In *Proceedings of the 15th International Conference on Machine Learning* (pp. 296–304). Morgan Kaufmann Publishers. Retrieved from http://l2r.cs.uiuc.edu/~danr/Teaching/CS546-09/Papers/Lin-Sim.pdf

Llopis, M., & Ferrández, A. (2013). How to make a natural language interface to query databases accessible to everyone: An example. *Computer Standards & Interfaces*, *35*(5), 470–481. https://doi.org/10.1016/j.csi.2012.09.005

Lopez, V., Fernández, M., Motta, E., & Stieler, N. (2012). PowerAqua: Supporting users in querying and exploring the Semantic Web. *Semantic Web*, *3*(3), 249–265. https://doi.org/10.3233/SW-2011-0030

Lopez, V., Uren, V., Motta, E., & Pasin, M. (2007). AquaLog: An ontology-driven question answering system for organizational semantic intranets. *Web Semantics: Science, Services and Agents on the World Wide Web*, *5*(2), 72–105. https://doi.org/10.1016/j.websem.2007.03.003

Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., & Bujnicki, J. M. (2014). The RNase H-like superfamily: New members, comparative structural analysis and evolutionary classification. *Nucleic Acids Research*, *42*(7), 4160–4179. https://doi.org/10.1093/nar/gkt1414

Meng, L., Huang, R., & Gu, J. (2013). A Review of Semantic Similarity Measures in WordNet. *International Journal of Hybrid Information Technology*, *6*(1), 1–12. Retrieved from https://pdfs.semanticscholar.org/da95/ceaf335971205f83c8d55f2292463fada4ef.pdf

Mitra, M., Singhal, A., & Buckley, C. (1998). Improving automatic query expansion. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 206–214). New York, USA: ACM. https://doi.org/10.1145/290941.290995

Monge, A. E., & Elkan, C. P. (1996). The field matching problem: Algorithms and applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 267–270). https://doi.org/10.1.1.23.9685

Moore, M. (2012). The semantic web: an introduction for information professionals. *The Indexer*, *30*(1), 38–43. Retrieved from https://innotecture.files.wordpress.com/2011/04/olc-april-2011_moore_semantic-web.pdf

Moreo, A., Castro, J. L., & Zurita, J. M. (2017). Towards portable natural language interfaces based on case-based reasoning. *Journal of Intelligent Information Systems*, *49*(2), 281–314. https://doi.org/10.1007/s10844-017-0453-8

Muchemi, L. (2008). Towards Full Comprehension of Swahili Natural Language Statements for Database Querying (pp. 50–58). Fountain Publishers. Retrieved from http://aflat.org/files/muchemi.pdf

Mustafa, J., Khan, S., & Latif, K. (2008). Ontology based semantic information retrieval. In *2008 4th International IEEE Conference Intelligent Systems* (pp. 22-14-22–19). IEEE. https://doi.org/10.1109/IS.2008.4670473

Natural Language Learning Data. (n.d.). Retrieved August 25, 2017, from http://www.cs.utexas.edu/users/ml/nldata.html

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, *33*(1), 31–88. https://doi.org/10.1145/375360.375365

Navigli, R. (2009). Word sense disambiguation. *ACM Computing Surveys*, *41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

Navigli, R., & Lapata, M. (2007). Graph connectivity for unsupervised Word Sense Disambiguation. In *International Joint Conference on Artificial Intelligence* (pp. 1683–1688). Retrieved from http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-272.pdf

Nihalani, N., Silakari, S., & Motwani, M. (2011). Natural language Interface for Database: A Brief review. *International Journal of Computer Science Issues*, *8*(2), 600–608. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.9531&rep=rep1&type=pdf#page=621

Palmer, D. D. (2010). Text Preprocessing. In N. Indurkhya & F. J. Damerau (Eds.), *Handbook of Natural Language Processing, 2nd Ed.* (pp. 9–30). CRC Press. Retrieved from https://karczmarczuk.users.greyc.fr/TEACH/TAL/Doc/Handbook Of Natural Language Processing, Second Edition Chapman & Hall Crc Machine Learning & Pattern Recognition 2010.pdf

Patel-Schneider, P. F. (2005). A Revised Architecture for Semantic Web Reasoning. In F. F. & S. S. (Eds.), *Principles and Practice of Semantic Web Reasoning. PPSWR 2005. Lecture Notes in Computer Science* (pp. 32–36). Springer Berlin Heidelberg. https://doi.org/10.1007/11552222_3

Pazos R., R. A., González B., J. J., Aguirre L., M. A., Martínez F., J. A., & Fraire H., H. J. (2013). Natural Language Interfaces to Databases: An Analysis of the State of the Art. *Recent Advances on Hybrid Intelligent Systems*, *451*, 463–480. https://doi.org/10.1007/978-3-642-33021-6_36

Ponzetto, S. P., & Navigli, R. (2010). Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems. In *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics* (pp. 1522–1531). Uppsala, Sweden: Association for Computational Linguistics. Retrieved from http://www.academia.edu/download/30772065/P10-1154.pdf

Princeton University. (2010). About WordNet - WordNet. Retrieved November 25, 2017, from https://wordnet.princeton.edu/

Purnamasari, D., Wulandari, L., Thantawi, A. M., & Wicaksana, I. W. S. (2016). Concept Similarity Searching for Support Query Rewriting. *International Journal of Computer Theory and Engineering*, *8*(6), 490–494. https://doi.org/10.7763/IJCTE.2016.V8.1094

Python NLTK Documentation. (n.d.). Natural Language Toolkit — NLTK 3.2.4 documentation. Retrieved September 10, 2017, from http://www.nltk.org/

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man and Cybernetics*, *19*(1), 17–30. https://doi.org/10.1109/21.24528

Rangel, R., & Aguirre, M. (2014). Features and Pitfalls that Users Should Seek in Natural Language Interfaces to Databases. *Recent Advances on …*, *547*, 617–630. https://doi.org/10.1007/978-3-319-05170-3

Ranjan Pal, A., & Saha, D. (2015). Word Sense Disambiguation: A Survey. *International Journal of Control Theory and Computer Modeling*, *5*(3), 1–16. https://doi.org/10.5121/ijctcm.2015.5301

Resnik, P. (1999). Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research*, *11*, 95–130. Retrieved from https://www.jair.org/media/514/live-514-1722-jair.pdf

Revuelta-Martínez, A., Rodríguez, L., García-Varea, I., & Montero, F. (2013). Multimodal interaction for information retrieval using natural language. *Computer Standards & Interfaces*, *35*(5), 428–441. https://doi.org/10.1016/j.csi.2012.11.002

Roy, D., Paul, D., Mitra, M., & Garain, U. (2016). Using Word Embeddings for Automatic Query Expansion. In *SIGIR Workshop on Neural Information Retrieval - Neu-IR '16*. Pisa, Italy. Retrieved from http://arxiv.org/abs/1606.07608

Royo, J. A., Mena, E., Bernad, J., & Illarramendi, A. (2005). Searching the Web: From Keywords to Semantic Queries. In *Third International Conference on Information Technology and Applications (ICITA'05)* (Vol. 1, pp. 244–249). IEEE. https://doi.org/10.1109/ICITA.2005.246

Sanderson, M. (2000). Retrieving with good Sense. *Information Retrieval*, *2*(1), 49–69. https://doi.org/10.1023/A:1009933700147

Schiff, J. (2011). Semantic Web Technologies Effects on Information Retrieval in Digital Libraries : An Annotated Bibliography. Retrieved from http://www.pages.drexel.edu/~js3634/eport/docs/AnnotatedBibliography.pdf

Seddah, D., Chrupala, G., Çetinoğlu, Ö., van Genabith, J., & Candito, M. (2010). Lemmatization and Lexicalized Statistical Parsing of Morphologically Rich Languages: the Case of French. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages* (pp. 85–93). Association for Computational Linguistics. Retrieved from http://doras.dcu.ie/15987/1/Lemmatization_and_Lexicalized_Statistical_Parsing.pdf

Sensebot Search. (n.d.). Retrieved August 18, 2017, from http://sensebot.com/

Sharef, N. M., & Mohd, S. A. (2012). Soft Queries Processing In Natural Language. *International Conference on Ubiquitous Information Management and Communication*, 1460–1466. Retrieved from https://www.scopus.com/record/display.uri?eid=2-s2.0-84904488556&origin=resultslist

Sharef, N. M., & Noah, S. A. M. (2012). Semantic Search Processing in Natural Language Interface. In *7th International Conference on Computing and Convergence Technology (ICCCT)* (pp. 1436–1442). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6530567

Shekarpour, S., Hoffner, K., Lehmann, J., & Auer, S. (2013). Keyword Query Expansion on Linked Data Using Linguistic and Semantic Features. In *2013 IEEE Seventh International Conference on Semantic Computing* (pp. 191–197). California, USA: IEEE. https://doi.org/10.1109/ICSC.2013.41

Singh, J., & Sharan, A. (2017). A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach. *Neural Computing and Applications*, *28*(9), 2557–2580. https://doi.org/10.1007/s00521-016-2207-x

Sivic, & Zisserman. (2003). Video Google: a text retrieval approach to object matching in videos. In *Proceedings of The Ninth IEEE International Conference on Computer Vision* (pp. 1470–1477 vol.2). IEEE. https://doi.org/10.1109/ICCV.2003.1238663

Slimani, T. (2013). Description and Evaluation of Semantic Similarity Measures Approaches. *International Journal of Computer Applications*, *80*(10), 25–33. https://doi.org/10.5120/13897-1851

Smith, C. L., Gwizdka, J., & Feild, H. (2017). The use of query auto-completion over the course of search sessions with multifaceted information needs. *Information Processing & Management*, *53*(5), 1139–1155. https://doi.org/10.1016/j.ipm.2017.05.001

Solskinnsbakk, G. (2012). *Contextual Semantic Search Navigation*. Norwegian University of Science and Technology, Trondheim. Retrieved from https://www.idi.ntnu.no/research/doctor_theses/geirsols.pdf

Stageberg, N. C. (1968). Structural Ambiguity for English Teachers. In *Selected Addresses Delivered at the Conference on English Education* (Vol. 6, pp. 29–34). National Council of Teachers of English. Retrieved from http://www.jstor.org/stable/40171840?seq=1#page_scan_tab_contents

Stubinz, J., & Whighli, S. (2008). Information Retrieval System Design for Very High Effectiveness, 1–7. Retrieved from http://goanna.cs.rmit.edu.au/~jz/sci/p3.pdf

Šukys, A., Nemuraitė, L., & Butkienė, R. (2017). SBVR Based Natural Language Interface to Ontologies. *Information Technology And Control*, *46*(1), 118–137. https://doi.org/10.5755/j01.itc.46.1.13998

Tablan, V., Damljanovic, D., & Bontcheva, K. (2008). A Natural Language Query Interface to Structured Information. In M. Koubarakis, S. Bechhofer, M. Hauswirth, & J. Hoffmann (Eds.), *The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science* (Vol. 5021, pp. 361–375). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68234-9_28

Tauberer, J. (2008). Why we need a new standard for the Semantic Web. Retrieved from https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md?section=1#why-we-need-a-new-standard-for-the-semantic-web

Tejada, S., Knoblock, C. A., & Minton, S. (2001). Learning object identification rules for information integration. *Information Systems*, *26*(8), 607–633. https://doi.org/10.1016/S0306-4379(01)00042-4

The Definitive Guide. (n.d.). Learn Python (Programming Tutorial for Beginners). Retrieved September 10, 2017, from https://www.programiz.com/python-programming

Tim Berners-Lee. (2000). Semantic Web - XML2000. Retrieved January 27, 2018, from https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

Tversky, A. (2013). Features of Similarity. In *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence* (pp. 290–302).

Ukkonen, E. (1992). Approximate string-matching with q-grams and maximal matches. *Theoretical Computer Science*, *92*(1), 191–211. https://doi.org/10.1016/0304-3975(92)90143-4

Unger, C., Bühmann, L., Lehmann, J., Ngonga Ngomo, A.-C., Gerber, D., & Cimiano, P. (2012). Template-based question answering over RDF data. In *Proceedings of the 21st international conference on World Wide Web* (pp. 639–648). New York, USA: ACM. https://doi.org/10.1145/2187836.2187923

Wagner, R. A., & Fischer, M. J. (1974). The String-to-String Correction Problem. *Journal of the ACM*, *21*(1), 168–173. https://doi.org/10.1145/321796.321811

Wagner, R. A., & Lowrance, R. (1975). An Extension of the String-to-String Correction Problem. *Journal of the ACM*, *22*(2), 177–183. https://doi.org/10.1145/321879.321880

Wang, C., Xiong, M., Zhou, Q., & Yu, Y. (2007). PANTO: A Portable Natural Language Interface to Ontologies. In E. Franconi, M. Kifer, & W. May (Eds.), *The Semantic Web: Research and Applications* (pp. 473–487). Berlin,

Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72667-8_34

Wellner, B., Castaño, J., & Pustejovsky, J. (2005). Adaptive string similarity metrics for biomedical reference resolution. In *Proceedings of the ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics* (pp. 9–16). Morristown, NJ, USA: Association for Computational Linguistics. https://doi.org/10.3115/1641484.1641486

Whiting, S., & Jose, J. M. (2014). Recent and robust query auto-completion. In *Proceedings of the 23rd international conference on World wide web* (pp. 971–982). New York, USA: ACM. https://doi.org/10.1145/2566486.2568009

Winkler, W. E. (1999). *The State of Record Linkage and Current Research Problems*. *Statistical Research Division, US Census Bureau*. https://doi.org/10.1.1.39.4336

Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 133–138). Association for Computational Linguistics. https://doi.org/10.3115/981732.981751

Wubben, S. (2008). *Using free link structure to calculate semantic relatedness*. *ILK Research Group Technical Report Series, (08-01)*. Retrieved from https://ilk.uvt.nl/downloads/pub/papers/wubben2008-techrep.pdf

Yahya, M., Berberich, K., Elbassuoni, S., Ramanath, M., Tresp, V., & Weikum, G. (2012). Natural Language Questions for the Web of Data. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 379–390). Jeju Island, Korea: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2390948.2390995

Yahya, M., Berberich, K., Ramanath, M., & Weikum, G. (2013). On the SPOT: Question Answering over Temporally Enhanced Structured Data. In *Proceedings of Workshop on Time-aware Information Access* (pp. 1–4). Dublin, Ireland. Retrieved from https://people.mpi-inf.mpg.de/~kberberi/publications/2013-taia2013b.pdf

Yauri, A. R., Kadir, R. A., Azman, A., & Murad, M. A. A. (2013). Ontology semantic approach to extraction of knowledge from Holy Quran. In *2013 5th International Conference on Computer Science and Information Technology* (pp. 19–23). IEEE. https://doi.org/10.1109/CSIT.2013.6588752

Yauri, A. R., Kadir, R. A., Azmi Murad, M. A., & Azman, A. (2012). Quranic-based concepts: Verse relations extraction using Manchester OWL syntax. *2012 International Conference on Information Retrieval & Knowledge Management*, 317–321. https://doi.org/10.1109/InfRKM.2012.6204998

Zenz, G., Zhou, X., Minack, E., Siberski, W., & Nejdl, W. (2009). From keywords to semantic queries—Incremental query construction on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, *7*(3), 166–176. https://doi.org/10.1016/j.websem.2009.07.005

Zhou, D., Wu, X., Zhao, W., Lawless, S., & Liu, J. (2017). Query Expansion with Enriched User Profiles for Personalized Search Utilizing Folksonomy Data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(7), 1536–1548. https://doi.org/10.1109/TKDE.2017.2668419

Zhou, Z., Wang, Y., & Gu, J. (2008). New model of semantic similarity measuring in wordnet. In *Proceedings of 2008 3rd International Conference on Intelligent System and Knowledge Engineering* (pp. 256–261). IEEE. https://doi.org/10.1109/ISKE.2008.4730937

Zou, Y., Finin, T., & Chen, H. (2004). F-OWL: An Inference Engine for Semantic Web. In M. G. Hinchey, J. L. Rash, W. F. Truszkowski, & C. A. Rouff (Eds.), *Formal Approaches to Agent-Based Systems* (pp. 238–248). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30960-4_16