SPATIAL AND TEMPORAL-BASED QUERY DISAMBIGUATION FOR IMPROVING WEB SEARCH

SHAHID KAMAL

UNIVERSITI TEKNOLOGI MALAYSIA

SPATIAL AND TEMPORAL-BASED QUERY DISAMBIGUATION FOR IMPROVING WEB SEARCH

SHAHID KAMAL

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
University Teknology Malaysia

MAY, 2016

DEDICATION

Dedicated to my caring father: May Allah SWT, be pleased with him! My mother (late): May ALLAH SWT awards her Al Jannah (Aameen)!

ACKNOWLEDGEMENT

First and Foremost praise is to ALLAH, the Almighty, the greatest of all, on whom ultimately we depend for sustenance and guidance. I would like to thank Almighty Allah for giving me opportunity, determination, and strength to do my research.

Firstly, I wish to express my sincere gratitude to my supervisor Dr. Roliana Ibrahim for the continuous support of my Ph.D. study and related research, for her patience, motivation, and immense knowledge. Besides my supervisor, I would like to thank and express my deepest appreciation to my co-supervisor Dr. Imran Ghani for his insightful comments, encouragement, support, and personal kindness. His guidance helped me in all the time of research and writing of this thesis.

I owe my thanks Dr. Ziauddin from ICIT, Gomal University for his support, motivations, and guidelines. Hafiz Fida Hussain, a sincere and dedicated colleague in the Gomal University for his endless support in all the matters related back to my office.

A Ph.D. candidate also has a life besides studies. Besides my research activities, I also enjoyed my stay in UTM. I thank my fellow colleagues and friends especially Fasee Ullah for spending time with me and express my gratitude to my friends back home for supporting and encouraging me throughout my study.

I would like to acknowledge my family, my parents, Brother Dr. Sajjad Ahmed, sisters especially the youngest one Azra Baloch, foster-brother Yasir Jamal Shani and niece Umaima Baloch for supporting me spiritually throughout writing this thesis and my life in general.

The acknowledgment would remain incomplete without thanking Universiti Teknologi Malaysia and Gomal University Dera Ismail Khan, Pakistan for providing all necessary facilities that helped in conducting this research.

ABSTRACT

Queries submitted to search engines are ambiguous in nature due to users' irrelevant input which poses real challenges to web search engines both towards understanding a query and giving results. A lot of irrelevant and ambiguous information creates disappointment among users. Thus, this research proposes an ambiguity evolvement process followed by an integrated use of spatial and temporal features to alleviate the search results imprecision. To enhance the effectiveness of web information retrieval the study develops an enhanced Adaptive Disambiguation Approach for web search queries to overcome the problems caused by ambiguous queries. A query classification method was used to filter search results to overcome the imprecision. An algorithm was utilized for finding the similarity of the search results based on spatial and temporal features. Users' selection based on web results facilitated recording of implicit feedback which was then utilized for web search improvement. Performance evaluation was conducted on data sets GISQC_DS, AMBIENT and MORESQUE comprising of ambiguous queries to certify the effectiveness of the proposed approach in comparison to a well-known temporal evaluation and two-box search methods. The implemented prototype is focused on ambiguous queries to be classified by spatial or temporal features. Spatial queries focus on targeting the location information whereas temporal queries target time in years. In conclusion, the study used search results in the context of Spatial Information Retrieval (S-IR) along with temporal information. Experiments results show that the use of spatial and temporal features in combination can significantly improve the performance in terms of precision (92%), accuracy (93%), recall (95%), and f-measure (93%). Moreover, the use of implicit feedback has a significant impact on the search results which has been demonstrated through experimental evaluation.

ABSTRAK

Di dalam dunia sebenar, pertanyaan yang dikemukakan oleh pengguna kepada enjin carian adalah masih samar kerana input tidak relevan yang memberikan cabaran kepada enjin carian web untuk memahami pertanyaan dan memberikan keputusan. Banyak maklumat yang tidak relevan dan samar menyebabkan para pengguna merasa kecewa. Oleh itu, kajian ini mencadangkan satu proses evolusi kesamaran diikuti penggunaan bersepadu ciri-ciri ruang dan masa diambil kira untuk mengurangkan ketakpersisan hasil carian. Untuk meningkatkan keberkesanan capaian maklumat web, kajian ini membangunkan Pendekatan Nyahkabur Adaptif dipertingkat untuk carian pertanyaan web bagi mengatasi masalah yang disebabkan oleh pertanyaan yang samar. Satu kaedah klasifikasi pertanyaan telah digunakan untuk menapis hasil bagi mengatasi ketakpersisan itu. Algoritma telah digunakan untuk mencari persamaan hasil carian berdasarkan kepada ciri-ciri ruang dan masa. Hasil maklum balas tersirat yang dihasil melalui pilihan pengguna digunakan untuk penambahbaikan carian web. Penilaian prestasi diuji pada set data GISQC_DS, AMBIENT dan MORESQUE yang terdiri daripada pertanyaan yang samar-samar untuk memperakui keberkesanan pendekatan yang dicadangkan berbanding dengan kaedah penilaian masa dan kaedah gelintaran dua-kotak. Prototaip yang dibangunkan tertumpu kepada pertanyaan yang samar-samar untuk diklasifikasi berdasarkan ciri-ciri ruang atau masa. Bagi pertanyaan ciri ruang, sasaran pertanyaan memfokus kepada maklumat lokasi manakala sasaran pertanyaan bagi ciri masa tertumpu kepada maklumat berkaitan tahun. Kesimpulannya, kajian ini menggunakan hasil carian dalam konteks Capaian Maklumat Spatial (S-IR) bersama-sama dengan maklumat sementara. Eksperimen hasil kajian menunjukkan bahawa penggunaan kombinasi ciri-ciri ruang dan masa dapat meningkatkan prestasi dari segi kepersisan (92%), ketepatan (93%), ingat kembali (95%), dan pengukuran-f (93%). Tambahan pula, penggunaan maklum balas tersirat juga mempunyai impak yang signifikan ke atas keputusan carian yang telah ditunjukkan melalui eksperimen pengujian.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADA	-	Adaptive Disambiguation Approach
AMBIENT	-	Ambiguous Entries
BPMW	-	Business Process Management Workshops
CbP	-	Constraint-based Precision
CSK	-	Cuckoo Search K-means
Ctx	-	Contextual
GISQC_DS	-	Google Insights for Query Search-Data set
GTE	-	Genereic Temporal Evaluation
IDE	-	Integrated Development Environment
IJCSI	-	Journal of Computer Science Issues
IJWesT	-	Journal of Web and Semantic Technology
IR	-	Information Retrieval
MORESQUE	-	MORE Sense-tagged QUEries
NWESP	-	Next Generation Web Services Practices
PsAQCM	-	Post-search Ambiguous Query Classification Method
QDA	-	Query Disambiguation Approach
SW	-	Semantic Web
T-IR	-	Temporal Information Retrieval
Tmp	-	Temporal
WDC	-	Web Document Clustering
WWW	-	World Wide Web

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

In this section, an overview of the domain knowledge and its associated problems are described with defending the establishment of the study. Furthermore the subsequent sections are the ephemeral elaboration of the problem background, problem statement, and research questions separately, to be answered in this study.

In the web search, the elementary research is related with the searched queries disambiguation in order to get information with respect to the user needs so as to enhance the performance. These objectives are considered by many people from different outlooks. The process that is initiated to search desired information from a collection according to user needs is known as information retrieval. This searching process can either be based on metadata or on full-text indexing. For example, a user expresses a query and there are several documents which are related with it found suitable in respect of his /her information desires. The user then will be required to analyze all those recovered results and will keep the most relevant ones while others will need to be rejected. This situation will be called as optimum information retrieval. The solution adopted by the user is clearly impossible because he/she will neither have enough time to check all the documents including irrelevant ones.

In furtherance of solving this problem after the invention of high-speed computers, this thought has been established that the computer would be able to read an entire collection of documents while choosing the relevant ones. Specifically the reading involves making attempts to get information either syntactic or semantic from the retrieved text and then make a decision about its relevance according to the given query. In this regard, difficulty refers to the process of getting information and finding its relevance with respect to its use (Bar-Hillel, 1964).

The importance for searching information and its related problems has been increased because of rapid growth and use of internet in all fields associated with information world. The activities from scientific information, looking for advancements and research needs are useful for significant information retrieval. World Wide Web (WWW) and the search engines have become essential components of our everyday life.

The individuals are reinforced in using their abilities for obtaining and utilizing the knowledge. With the advent of internet, the organizations as well as individuals are producing information in huge quantity for the sharing with others. This results in discovering the useful information from this huge and diverse quantity of information without the support of information systems.

The users need to give precisely their information needs while making communication with information systems. But however, natural language limitations in terms of synonyms of the words and lacking of information in knowledge domain cause difficulties for user to express their queries in an effective ways. When 40 years ago, the Information Systems (IR) systems were developed, it was assumed that users will clearly express their queries i.e. might be information professionals, that should be appropriate for the IR systems to process those queries. While, modern IR systems are not restricted to professional searchers. Somewhat, these are to carry out a several different new tasks, e.g. book search, social media search, to serve a large number of users with different needs (Zhang, 2013).

Time has an important role in the domain of web search. Because many of the web pages contain temporal information and most of the user queries for search are time-related. The temporal information has gained an important position in different web related fields like web search, information extraction, topic detection, answering to the queries, and analysis of query log. It is commonly expressed in web pages as temporal expressions either in the form of explicit e.g., September 3, 2015 or implicit, e.g., Today. Within the scope of web search, various issues have been imposed because of different forms of temporal expressions, e.g., determination of exact temporal information for the implicit expressions, focused time that is needed for a web page, integration of temporal information into a web search (Lin *et al.*, 2014).

Disambiguating the search intent and improving the accuracy of resulting information is a crucial issue in the domain of IR systems, especially when most of the users are unable to clearly express their information needs. For this purpose, IR system should be able to identify ambiguous user intents and then transform poorly expressed queries into effective ones. Thus improving the effectiveness of user queries by disambiguating and then query enhancement is a critical task for modern IR systems (Manning *et al.*, 2008; Zhang, 2013).

The background of the problem that has been attempted to answer in this study is given in the following section.

1.2 Background of the Problem

This section describes in detail the problem background in the specified domain along with the concept of ambiguity with the examples and also the issues associated with ambiguity, for the better understanding of the readers.

Despite tremendous improvements being made for the web search optimization, the ample efforts are still required to enhance the user experience that

emphases on having deep understanding of the users such that their needs, abilities, limitations. One of the main challenges in modern Information Retrieval (IR) is web search optimization (Anastasiu *et al.*, 2013) and has gained remarkable attention of the experts from both the industry and the academia. In pursuit of web search optimization, an emerging research area known as Temporal Information Retrieval (T-IR) has been gaining increasing importance in recent years (Joho *et al.*, 2013) within the search context. The T-IR refers to the process of document retrieval mainly predicated on time, because of its crucial role in assessing the document relevance in web search (Joho *et al.*, 2014). Generally, T-IR intends to gratify the search needs by combining the traditional concept of document relevance with the temporal relevance. Significant numbers of user search queries have strong temporal components or characteristics for example, the queries about past facts, most recent information, weather forecasts or about future related events. For instance referring to the "World Cup" example, the user might be interested in information about FIFA World Cup 2014 at Brazil. In this regard, if the user issues a query phrase "World Cup 2014", it will make use of the temporal feature and will produce 12 ambiguous results i.e. ICC T20 World Cup at Bangladesh, FIFA World Cup at Brazil, Men's Hockey World Cup at Netherlands, Alpine Skiing World Cup at Austria, FIBA 2014 at Spain and so on.

Temporal characteristics can be useful for a wide range of information retrieval such as similarity search, summarization, and document exploration (Henry *et al.*, 2015). Furthermore, other characteristics related to temporal information include well-definition of two intervals in time, normalization of temporal expressions in standard format, and mapping of temporal information hierarchically. By the use of these characteristics, the temporal information about documents can be utilized as time-specific information retrieval. In order to determine the quality of the document, timeliness or currency play an important role but however, there are another aspects namely; coverage, objectivity, accuracy and relevance need due consideration in T-IR. Different ways have been described to express temporal information such as explicit, implicit and relative for the types date and time in the documents (Alonso *et al.*, 2011).

Majority of the existing literature such as rule-based (Song *et al.*, 2007), topological (Song *et al.*, 2009), and ontological (Page *et al.*, 1999; Song *et al.*, 2007) approaches are based on temporal information retrieval. The T-IR based approaches whereas; somehow refine the search results by exploiting various temporal features: date, time, duration, and set. However, due to lacking of spatial information, it results into a large proportion of irrelevant information retrieval (Palacio *et al.*, 2015).

Context is an important source of information in computing environments. The term context is defined by the authors of (Dey, 2001) as "any information that can be used to characterize the situation of an entity". An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. According to the authors of (Anastasiu *et al.*, 2013), the query disambiguation can be greatly improved by applying spatial information. For instance, referring to our example, if we add the spatial information (Brazil as place) and rephrase our search query as "World Cup 2014 Brazil," this would produce more accurate results according to our intent. Hence, it is observable that context plays an important role in resolving the queries ambiguity (Patil and Keole, 2014).

Ambiguity is considered as the most important problem that exists in inappropriate search results (Zahariadis, 2014). As the web size is mounting at growing rate, the ambiguity turn out to be universal and the users need active means of disambiguating the information that is retrieved in response of queries. The ambiguity can be resolute by defining knowledge of the available domain and then applying refinement process over query with the addition of spatial terms as well as temporal features. The main reason for the ineffectiveness of the previous approaches was use of spatial and temporal features independently (Campos *et al.*, 2014a; Joho *et al.*, 2013). Commonly the query terms are short in nature, comprising one to three terms only (Roul and Sahay, 2012), recognized as naturally ambiguous due to polysemy i.e., different possible meanings associated with a word or phrase. Consequently, numerous inappropriate web pages are retrieved in a response of the ambiguous queries expressed in the form of user intents and information needs. The major question arises here is that how to get the relevant information for the

ambiguous queries. With the high growth in the size of the web, the ambiguity becomes ever-present and hence, the users seek for the active means that would cover their needs to meet disambiguation of the searched results accordingly (Winokur, 2015).

We give an improved disambiguation approach that makes use of both the temporal information i.e., year and spatial information i.e., location, thereby retrieving the most accurate results in accordance with the search queries. The proposed approach is comprised of five stages namely: query input, query classification, sub-query construction, results integration and improved results through feedback. Experimentally based evaluation of ADA reveals improved performance in terms of accuracy, precision, recall, and F1-measure as compared to the existing work (Campos *et al.*, 2014b; Cobos *et al.*, 2014).

To conclude the description, the query disambiguation in web information retrieval is an extensive field and much effort has been made to improve the mechanisms in the context of the information retrieval. Even with having different approaches for query disambiguation (Bennett *et al.*, 2015; Chowdhury and Pass, 2014), the user needs to put more efforts for the specific cases while using spatial and temporal information.

The web search optimization is intended to retrieve relevant information while using different search engines. But it is getting more multifaceted and challenging to retrieve accurate information according to user needs because of high and fast growth of internet size and its complexity (Hannak *et al.*, 2013). To do so, it needs user queries in an accurate manner for retrieving information. While user queries are known to be ambiguous in nature (Song et al., 2009); cause low performance of the system in terms of accuracy. Additionally, leading towards ambiguous queries identification, which is also a laborious task. We need at this point to give an example for the better understanding about the problem such that if a search query “Cultural Show” is put in a famous search engine Google, it will generate 3400 thousands outcomes while taking only 0.39 seconds for processing the search query. However, there will be lot of irrelevant information because of unclear

query given earlier which is describing clearly that what we mean to retrieve about "cultural show" at which place, year, or event.

The query disambiguation process support the user not only in receiving the significant information but also contributes in enhancing the search engine performance (Bunescu and Pasca, 2006). In this perspective, the techniques dealing with query disambiguation such as (Anastasiu et al., 2013; Bunescu and Pasca, 2006; Mihalcea and Csomai, 2007; Song et al., 2009) along with support of identifying ambiguous queries were introduced. In order to get accurate information, the search processes are enhanced previously with the addition of different time and context related features. These enhancements headed towards temporal search introduction in search processed (Campos *et al.*, 2012; Drew and Wolfe, 2012; Lan *et al.*, 2013) and spatial search(Anastasiu *et al.*, 2013; Kraft *et al.*, 2006; Mizzaro and Vassena, 2011) as well. Ricardo, at al. highlighted the disambiguation of text queries with respect to temporal feature time in terms of year (Campos *et al.*, 2012). The approach was based on clustering the search results based on the temporal features that were previously neglected by some clustering engines i.e., iBoogie[1], Yippy[2]. They proposed a two-stage process where documents were grouped together into a single cluster while sharing a common year i.e., temporal expression. Their approach was based on the idea of finding a non-trivial term in text and focused on temporal clustering. The temporal clustering is firstly introduced by (Alonso *et al.*, 2011) on the basis of topics and time. Their work was conceding the result accuracy because of exclusive dependency over temporal features. Link Text Topic Model (LTTM) based disambiguation approach has been proposed by (Skaggs, 2011), however, it resolves the link disambiguation problem only thereby lacking the capability to disambiguate the user queries. (Boston *et al.*, 2014) developed a system (called "Wikimantic") for link disambiguation and query expansion in response to user queries for the retrieval of information graphics. In the developed system, they first disambiguate short text strings, followed by determination of the instant when the sequence of words should be disambiguated. The main limitation of their system is that it only entertains short queries and the performance is greatly deteriorated when

[1] http://www.iboogi.com

[2] http://search.yippy.com

exposed to large queries. Furthermore, it attains low precision and recall as compared to other approaches.

Given a large text string, it's always possible to find at least one non-trivial term to start the process. (Ferragina and Scaiella, 2010) addressed this problem by employing a voting system that resolved all ambiguous terms simultaneously. Their system makes use of various characteristics associated with different fragments of the input strings and completely overlooks the temporal and spatial features. Furthermore, due to the unavailability of non-trivial terms in short text strings, its performance is greatly affected. More recently (Anastasiu *et al.*, 2013) investigated the problem of query disambiguation by making use of keywords search and spatial information. First the articles were retrieved on the basis of both combined fragments of the query as well as spatial terms. Next they retrieved the articles based on only query terms and finally similarities were computed. Eventually, the commonly retrieved results were presented to users for their selection.

All such approaches are based on disambiguating the search intents, meaning they use temporal as well as spatial features as additional elements in the user given queries to find the most relevant results according to user needs. However, each of these approaches uses features in a way that supports certain specific feature of the queries while neglecting others (e.g., (Campos *et al.*, 2012) neglects the use of spatial features in terms of location in their approach). Therefore the query disambiguation to get relevant and accurate search results, temporal features cannot solely be exploited to an extent that these actually results relevant and accurate information based on queries. Consequently, there is a need for an approach that bases upon temporal features in terms of time as well as spatial features in terms of location and eventually produces highly relevant and accurate search results to realize a better information retrieval in the larger context of web search.

Taking into account earlier described research gaps; this research study aims to deal with the problem of ambiguous queries that affect the information relevancy using the spatial as well as temporal features. The ambiguous queries classification on the afore-mentioned features is employed over ambiguous datasets. The query

classification and then applying post search results filtering is expected to result in better search performance. Post search ambiguous query classification method, results similarity based on spatial and temporal features is proposed along with implicit feedback collection to get the better performance in this study.

1.3 Problem Statement

In this study, we intend to deal with problem of disambiguation of the ambiguous queries being input for searching according to user intents. The problem is defined as:

The search engines generate thousands of web pages in reply of user queries. Among these results, many results are irrelevant, known as ambiguous results and are triggering confusion towards query understanding and its results. The major reason of having a lot of irrelevant information is unclear contents of the queries given by the users. This irrelevancy creates disappointment among the users and is deliberated as one of the vigorous problems, mainly instigated due to search queries ambiguity. Prior to get back the relevant and more accurate results, it is required to have ambiguity evolvement process and then to have measures to solve the issues related with it. In order to solve the afore-mentioned problem, different associated sub-problems can be solved for the improvement of overall search process. These are as follows:

1.3.1 Ambiguous Queries Investigation

The first problem is to investigate the nature of queries, whether the queries being input are ambiguous or clear? This problem leads to diversified results, when ambiguous or broad queries are received by search engine by knowing little about the user while covering several interpretations of the query. Therefore, prior to apply any method for the disambiguation and sorting out the relevant results from the response,

it is mandatory to have procedures for the investigating the nature of the queries being input for search.

1.3.2 Post Search Results Filtering

The web search can be significantly improved, and the efforts involved in resolving the ambiguity of the queries for search can be reduced by employing query disambiguation techniques using different features i.e., spatial and temporal. The approaches that consider temporal features do not generate the better results, related with spatial features such that created the space for use of combined features for the disambiguation and better search results as well. The past research has mainly focused on temporal features i.e., year and spatial features i.e., author name independently. Hence, in post-search processes, using different features, we can have more accurate results that are being responded back. Therefore in this study, we intend to propose an approach for the disambiguation of user queries in order to improve the search results using the spatial information i.e., location and temporal information i.e., year.

1.3.3 Implicit User Feedback

Using Internet, it is common to collect user feedback information by Internet contents providers. This feedback may be either explicit or implicit. Consequently, improving the search established on user feedback has not been given much attention. By collecting user feedback and then improving the search contents can significantly increase the information exactness as well as will cause improvement in the performance of the webs search. Furthermore, all previous approaches being identified in literature review process, suffer from at least one common problem, which is their inability to generate better relevant search results according to user needs.

1.4 Research Question

This study aims to resolve the afore-mentioned problems by using the spatial and temporal features being identified in the post search results. Based on these problems, we can come up to synthesize our main research question that is:

"How can we improve the web information retrieval by using the spatial as well as temporal information in combination, for disambiguating the user queries that are more clearly expressed in terms of user intents based on the post search results and exploiting user feedback?"

Based on the problem statement given above, the following questions that need to be answered are pointed out as follows:

(i) How can we classify the ambiguous queries being input for searching specific information?

(ii) How these ambiguities in the user given queries can be reduced by using spatial information and temporal information being existed in post search results, in order to get better search results?

(iii) How to specify an improved approach for the web search under ambiguous queries that are causing retrieval of irrelevant information in response?

(iv) How to measure performance of proposed approach using the user feedback?

After pointing out the research questions, the following section describes the research aim and objective to be achieved.

1.5 Research aim

The purpose of the research is to develop an approach for web search query disambiguation to improve the accuracy of the results according to user needs. Hence, this research attempts to use ambiguous queries that are given to search engines by users in order to find the specific information. Besides, it builds the execution of the approach by enhancing the substance of the user input focused around the determination of results made by users. Moreover, this study is to propose an improved approach that should be implemented to overcome the issues related with the accuracy of the search results in a response of the ambiguous queries lacking the spatial as well as temporal features, in the domain of web search results.

1.6 Research Objectives

Towards achievements of the research aim, some research objectives that are being identified are given below:

(i) To give an ambiguous queries classification method for accurate information retrieval in response.

(ii) To define a post search results-based implementation method that will disambiguate the search queries in order to increase the accuracy in terms of relevance of the retrieved information.

(iii) To give a method based on a collection of the feedback through users to increase the performance of the query disambiguation approach.

The following section describes in detail the research scope of the study which includes the limitations of the study in terms of data sets and methodology as well.

1.7 Research Scope

In this thesis, an improved methodology has been created containing relevant spatial (location) and temporal (year) data to be transformed for improved web information retrieval. So as to assess the methodology, two previous methodologies have been chosen for the benchmarking. An algorithm has additionally been proposed for the better hunt as indicated by user needs. The methodology is focused on the algorithm and the results are defined in such a path, to the point that it can understand the logical and additionally transient data accordingly of user queries. These augmentation unobtrusive components are given underneath.

(i) For the experimental validation of the proposed approach, three publically available data sets namely; GISQC_DS data set that comprise of 450 (220 Ambiguous) queries which are manually extracted from Google Insights for Search; AMBIENT data set with 44 ambiguous queries and MORESQUE data set with 114 ambiguous queries have been used.

(ii) With a specific end goal to evaluate the execution of the proposed methodology, the common IR measures namely; Precision, Accuracy, Recall, and F1-Measure have been considered as these four are considered as common IR measures in the literature.

(iii) The Proposed technique will provide measurable enhancement in overall performance regarding web information retrieval with the addition of above-mentioned features.

The subsequent section describes in detail the organization of this dissertation i.e. chapter-wise detail about included contents in the specified chapters.

1.8 Thesis Organization

This thesis is organized into seven different chapters described as follows:

Chapter 1, *Introduction*, presents an overview of the web information retrieval, input queries for search and ambiguity issues related to queries and search results generated in response. The chapter also publicizes the problem statement and introduces the research objectives and highlights of the contributions.

Chapter 2, *Literature review*, presents the detailed background about the latest work done in the field of information retrieval and also the literature about ambiguities found in queries and their effects on the retrieved results. We discussed different research contributions in the area of web information retrieval and also different approaches that had been used in different researches. The chapter also gives an overview of the previous studies and further discussion includes the concept of disambiguation, web search related issues and evaluation metrics.

Chapter 3, *Research methodology*, discusses the methods, which are used in this proposed approach that has been adapted during the entire research process. In the subsequent sections of the chapter, research framework, problem analysis, disambiguation approach, and information retrieval processes are discussed. Furthermore, how the disambiguation process takes place, has been discussed. The complete structure of the methodology and proposed algorithm is displayed in subtle element by utilizing pictorial representation. Finally, a complete rundown of the chapter is concluded at the end.

Chapter 4, *Implementation of approach; Ambiguous queries investigation*, describes different steps of the approach in detail that how the disambiguation approach is categorized into different steps. Additionally, how the investigation process takes place and different categories of the queries that are being investigated. A complete rundown of the chapter is concluded at the end.

Chapter 5, *Implementation of approach: exploitation of user feedback*, in this chapter, implementation of the approach in terms of exploiting user feedback is discussed to give an overall picture of the procedure.

Chapter 6, *Experiments and results analysis,* discusses the results by explaining the evaluation processes used to validate the results and by comparing the proposed disambiguation approach with the other existing approaches.

Chapter 7, *Conclusion and future work,* concludes the research, provides the description of contributions along with limitation associated with this study and finally the future bearings are given for further study.

1.9 Summary

The fundamentals of the research and the necessary parts of this study are discussed in this chapter. An overview of the research domain, problem background, and statement of the problem resulting into research questions, objectives, and research scope are introduced. The basic idea providing this chapter is to give an overall detail of the major parts of this research study so that readers can get clear understanding about domain, its associated problems, and the solution being proposed.

REFERENCES

Agirre, E., and Edmonds, P. (2007). *Word Sense Disambiguation: Algorithms and Applications.* (1st ed.) Netherlands: Springer.

Agosti, M., and Melucci, M. (2001). Information Retrieval on the Web. *Lectures on Information Retrieval* (pp. 242-285). Berlin Heidelberg: Springer.

Alonso, O., Gertz, M., and Baeza-Yates, R. (2009). *Clustering and Exploring Search Results using Timeline Constructions*. 18th ACM Conference on Information and knowledge management (CIKM'09). November 2-6. Hong Kong, 97-106.

Alonso, O., Strötgen, J., Baeza-Yates, R. A., and Gertz, M. (2011). *Temporal Information Retrieval: Challenges and Opportunities*. 1st Temporal Web Analytics Workschop. March 28. Hyderabad, India, 1-8.

Anastasiu, D. C., Gao, B. J., Jiang, X., and Karypis, G. (2013). A novel two-box search paradigm for query disambiguation. *World Wide Web*, *16*(1), 1-29.

BAEZA-Yates, R., and Ribeiro-Neto, B. (1999). *Modern Information Retrieval* (1st ed.). New York, USA: ACM press.

BAEZA, Y., and Ribeiro-Neto, B. (2011). *Modern Information Retreival-The Concepts and Technology behind Search* (2nd ed.) New York, USA: ACM Press.

Ballesteros, L., and Croft, W. B. (1998). *Resolving Ambiguity for Cross-language Retrieval*. 21st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '98). August 24-28. Melbourne, Australia, 64-71.

Bar-Hillel, Y. (1964). *Language and Information: Selected Essays on Their Theory and Application* (3rd ed.). United States: Addison-Wesley.

Barathi, M., and Valli, S. (2013). Query Disambiguation Using Clustering and Concept Based Semantic Web Search For efficient Information Retrieval (QDC-CSWS). *Life Science Journal*, *10*(2), 147-155.

Bekkerman, R., and McCallum, A. (2005). *Disambiguating Web Appearances of People in a Social Network*. 14th International Conference on World Wide Web (WWW'05). December 12-15. Amsterdam, The Netherlands, 463-470.

Bennett, P., Collins-thompson, K., Sarkizova, S., Shokouhi, M., and Sloan, M. (2015). Learning And Using Spatial Content Retrieval Rules For Query Disambiguation: US Patent 20,150,363,485.

Bennett, P. N., Chickering, D. M., Collins-Thompson, K. B., Dumais, S., and Liebling, D. J. (2014). Washington DC, USA Patent No. 8719249. U. S. P. a. T. Office.

Berkhin, P. (2006). Bookmark-Coloring Algorithm for Personalized PageRank Computing. *Internet Mathematics, 3*(1), 41-62.

Bortnikov, E., Donmez, P., Kagian, A., and Lempel, R. (2012). Modeling transactional queries via templates. *Advances in Information Retrieval* (pp. 13-24). Berlin Heidelberg: Springer.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). *A Training Algorithm for Optimal Margin Classifiers*. fifth annual workshop on Computational learning theory (COLT'92). July 27 - 29. Pittsburgh, PA, USA 144-152.

Boston, C., Fang, H., Carberry, S., Wu, H., and Liu, X. (2014). Wikimantic: Toward effective disambiguation and expansion of queries. *Data & Knowledge Engineering, 90*, 22-37.

Bota, H., Zhou, K., Jose, J. M., and Lalmas, M. (2014). *Composite Retrieval of Heterogeneous Web Search*. 23rd international conference on World Wide Web (WWW'14). April 7-11. Seoul, Korea, 119-130.

Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems, 30*(1), 107-117.

Broder, A. (2002). A Taxonomy of Web Search. *ACM SIGIR FORUM, 36*(2), 3-10.

Brody, S., Navigli, R., and Lapata, M. (2006). *Ensemble Methods for Unsupervised WSD*. 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. 17th–21st July. Sydney, Australia, 97-104.

Bunescu, R. C., and Pasca, M. (2006). *Using Encyclopedic Knowledge for Named entity Disambiguation*. 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06). 3-7 April. Trento, Italy, 9-16.

Callan, J., and Moffat, A. (2012). Panel on use of proprietary data. *ACM SIGIR Forum, 46*(2), 10-18.

Campos, R., Dias, G., Jorge, A. M., and Jatowt, A. (2014a). Survey of temporal information retrieval and related applications. *ACM Computing Surveys (CSUR), 47*(2), 15.

Campos, R., Dias, G., Jorge, A. M., and Nunes, C. (2014b). GTE-Cluster: A temporal search interface for implicit temporal queries. *Advances in Information Retrieval* (pp. 775-779). Switzerland: Springer International Publishing.

Campos, R., Gaël , D., and Alípio , M. J. (2011a). *What is the Temporal Value of Web Snippets?.* 1st International Temporal Web Analytics Workshop (TWAW2011) associated to the 20th International World Wide Web Conference(WWW2011). 28th March. Hyderabad, India, 9-16.

Campos, R., Jorge, A., and Dias, G. (2011b). *Using Web Snippets and Query-logs to Measure Implicit Temporal Intents in Queries*. 2nd Workshop on Query Representation and Understanding of the 34th ACM Annual SIGIR Conference (SIGIR 2011), Jul 2011, Pekin, China. pp.4 Pages.

Campos, R., Jorge, A. M., Dias, G., and Nunes, C. (2012). *Disambiguating Implicit Temporal Queries by Clustering Top Relevant Dates in Web Snippets*. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on 4-7 Dec. Macau, 1-8.

Campos, R. N. T. (2013). *Disambiguating implicit temporal queries for temporal information retrieval applications.* Doctor of Computer Science, Faculty of Sciene, University of Porto, Portugal.

Carpineto, C., and Romano, G. (2008). Ambient dataset. *http://credo.fub.it/ambient/*.

Carpineto, C., and Romano, G. (2012). A survey of automatic query expansion in information retrieval. *ACM Computing Surveys (CSUR), 44*(1), 1-50.

Chirita, P. A., Nejdl, W., Paiu, R., and Kohlschütter, C. (2005). *Using ODP Metadata to Personalize Search*. 28th annual international ACM SIGIR conference on Research and development in information retrieval. August 15-19. Salvador, Brazil, 178-185.

Chowdhury, A. R., and Pass, G. S. (2014). Washington, DC Patent No.: U. S. P. a. T. Office.

Cobos, C., Muñoz-Collazos, H., Urbano-Muñoz, R., Mendoza, M., León, E., and Herrera-Viedma, E. (2014). Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion. *Information Sciences, 281*, 248-264.

Craswell, N., and Hawking, D. (2009). *Web Information Retreival.* In A. Goker , and J. David (Eds.). *Information Retrieval: Searching in the 21st Century(pp.85-100.).* London, UK:Wiley.

D'Angelo, C. A., Giuffrida, C., and Abramo, G. (2011). A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. *Journal of the American Society for Information Science and Technology, 62*(2), 257-269.

Daelemans, W., Van Den Bosch, A., and Zavrel, J. (1999). Forgetting Exceptions is Harmful in Language Learning. *Machine Learning, 34*(1-3), 11-41.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing, 5*(1), 4-7.

Dijksta, T. (2005). Bilingual Visual Word Recognition and Lexical Access. *Handbook of bilingualism psycholinguistic approaches, 54,* 179-201.

Dijkstra, T., and Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. In G. Jonathan, and J. Arthur M (Eds.). *Localist Connectionist Approaches to Human Cognition* (pp. 189-225). University of Nijmegen: Psychology Press.

Dijkstra, T., and Van Heuven, W. J. (2002). The Architecture of the bilingual Word Recognition System: From Identification to Decision. *Bilingualism: Language and Cognition, 5*(03), 175-197.

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V., and Sachs, J. (2004). *Swoogle: A Search and Metadata Engine for the Semantic Web*. Thirteenth ACM international conference on Information and knowledge management (CIKM'04). 8-13 November. Washington, DC, USA, 652-659.

Dongen, S. v. (2000) *Graph clustering by flow simulation*. Doctor Philosophy, University of Utrecht, Utrecht, The Netherlands.

Drew, T., and Wolfe, J. M. (2012). Hybrid search in the temporal domain: Monitoring an RSVP stream for multiple targets held in memory. *Journal of Vision, 12*(9), 1276.

Duffy, S. A., Morris, R. K., and Rayner, K. (1988). Lexical Ambiguity and Fixation Times in Reading. *Journal of Memory and Language, 27*(4), 429-446.

Enss, M. J. R. (2006). *An Investigation of Word Sense Disambiguation for Improving Lexical Chaining.* Master of Mathematics in Computer Science, University of Waterloo, Ontario, Canada.

Erkan, G., and Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research, 22*(1), 457-479.

Escudero, G., Màrquez, L., and Rigau, G. (2000a). Boosting Applied to Word Sense Disambiguation. *Machine Learning: ECML* (pp. 129-141). Berlin Heidelberg: Springer.

Escudero, G., Màrquez, L., Rigau, G., and Salgado, J. G. (2000b). *On the Portability and Tuning of Supervised Word Sense Disambiguation Systems* joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, EMNLP/VLC. October 7-8. Hong Kong University of Science and Technology, China, 1-18.

Ferragina, P., and Scaiella, U. (2010). *Tagme: on-the-fly annotation of short text fragments (by wikipedia entities).* 19th ACM International Conference on Information and Knowledge Management (CIKM'10). October 26 - 30. Toronto, Canada 1625-1628.

E.Tanaka (2012). Syntactic and Structural Pattern Recognition Ferraté, G., Pavlidis, T., Sanfeliu, T., and Bunke, H. (Eds). *A String Correction Method Based on The Context-Dependent Similarity (pp.3-17). Berlin: Springer-Verlag.*

FitzPatrick, I., and Indefrey, P. (2014). Head start for target language in bilingual listening. *Brain Research, 1542*, 111-130.

Florian, R., Cucerzan, S., Schafer, C., and Yarowsky, D. (2002). Combining Classifiers for Word Sense Disambiguation. *Natural Language Engineering, 8*(04), 327-341.

Gao, B. J., Anastasiu, D. C., and Jiang, X. (2010). *Utilizing User-input Spatial Terms for Query Disambiguation.* 23rd International Conference on Computational Linguistics: Posters. August. Beijing, China, 329-337.

Garg, D., and Sharma, D. (2012). Information Retrieval on the Web and its Evaluation. *International Journal of Computer Applications, 40*(3), 26-31.

Gauch, S., Speretta, M., Chandramouli, A., and Micarelli, A. (2007). User Profiles for Personalized Information Access. *The Adaptive Web* (pp. 54-89). Berlin Heidelberg: Springer-Verlag.

Glucksberg, S., Kreuz, R. J., and Rho, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 323-335.

Göker, A., and Myrhaug, H. (2008). Evaluation of a Mobile information System in Context. *Information Processing & Management, 44*(1), 39-65.

Gupta, Y., Saini, A., Saxena, A., and Sharan, A. (2014). Fuzzy logic Based Similarity Measure for Information Retrieval System Performance Improvement. *Distributed Computing and Internet Technology* (pp. 224-232). Switzerland: Springer International Publishing.

Hachey, B., Radford, W., Nothman, J., Honnibal, M., and Curran, J. R. (2013). Evaluating Entity Linking with Wikipedia. *Artificial Intelligence, 194*, 130-150.

Han, X., and Zhao, J. (2009). *Named Entity Disambiguation by Leveraging Wikipedia Semantic Knowledge*. 18th ACM Conference on Information and Knowledge Management (CIKM'09). November 2-6. Hong Kong,, 215-224.

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., and Wilson, C. (2013). *Measuring Personalization of Web Search*. Proceedings of the 22nd International Conference on World Wide Web. 527-538.

He, T., Li, F., and Ma, L. (2010). *Document Relevance Identifying and its Effect in Query-Focused Text Summarization*. Granular Computing (GrC), 2010 IEEE International Conference on. 14-16 Aug. San Jose, CA, 206-211.

Hearst, M. (1991). *Noun Homograph Disambiguation Using Local Ccontext in Large Text Corpora*. 7th Annual Conference of the University of Waterloo Centre for the New OED and Text Research. October. Oxford, UK, 185-188.

Henry, M. J., Herrmann, B., and Obleser, J. (2015). Selective Attention to Temporal Features on Nested Time Scales. *Cerebral Cortex, 25*(2), 450-459.

Hindle, D., and Rooth, M. (1993). Structural Ambiguity and Lexical Relations. *Computational Linguistics, 19*(1), 103-120.

Hogaboam, T. W., and Perfetti, C. A. (1975). Lexical Ambiguity and Sentence Comprehension. *Journal of Verbal Learning and Verbal Behavior, 14*(3), 265-274.

Jansen, B. J., Spink, A., and Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management, 36*(2), 207-227.

Joho, H., Jatowt, A., and Blanco, R. (2014). *NTCIR Temporalia: A Test Collection for Temporal Information Access Research*. companion publication of the 23rd International Conference on World Wide Web Companion (WWW'14 Companion). April 7-11. Seoul, Korea, 845-850.

Joho, H., Jatowt, A., and Roi, B. (2013). *A Survey of Temporal Web Search Experience*. 22nd International Conference on World Wide Web Companion. 13-17 May. Rio de Janeiro, Brazil, 1101-1108.

Jones, R., and Diaz, F. (2007). Temporal Profiles of Queries. *ACM Transactions on Information Systems (TOIS), 25*(3), 14.

King-Sun, F. (1983). Introduction to Syntactic Pattern Recognition. In Fu, K.S (Ed.) *Syntactic Pattern Recognition and Applications* (pp.1-30). Berlin Heidelberg: Springer-Verlag.

Klein, D., Toutanova, K., Ilhan, H. T., Kamvar, S. D., and Manning, C. D. (2002). *Combining Heterogeneous Classifiers for Word-sense Disambiguation*. SIGLEX/SENSEVAL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. July. Philadelphia, 74-80.

Kobayashi, M., and Takeda, K. (2000). Information Retrieval on the Web. *ACM Computing Surveys (CSUR), 32*(2), 144-173.

Kopliku, A., Pinel-Sauvagnat, K., and Boughanem, M. (2014). Aggregated Search: A New Information Retrieval Paradigm. *ACM Computing Surveys (CSUR), 46*(3), 1-31.

Koutrika, G., and Ioannidis, Y. (2005). *A Unified User Profile Framework for Query Disambiguation and Personalization*. Proceedings of workshop on new technologies for personalized information access (PIA 2005). 25-26 July. Edinburgh,United Kingdom, 44-53.

Kraft, R., Chang, C. C., Maghoul, F., and Kumar, R. (2006). *Searching with Context*. 15th International Conference on World Wide Web (WWW'06). 23-26 May. Edinburgh Scotland, 477-486.

Lan, R., Lee, H., Fong, A., Monroe, M., Plaisant, C., and Shneiderman, B. (2013). *Temporal Search and Replace: An Interactive Tool for the Analysis of Temporal Event Sequences.* Technical Report HCIL-2013-TBD, HCIL, University of Maryland, College Park, Maryland.

Lee, Y. K., and Ng, H. T. (2002). *An Empirical Evaluation of Knowledge Sources and Learning Algorithms for Word Sense Disambiguation*. Empirical Methods in Natural Language Processing (EMNLP '02). July. Philadelphia, 41-48.

Lesk, M. (1986). *Automatic Sense Disambiguation Using Machine Readable Dictionaries: how to tell a pine cone from an ice cream cone*. 5th Annual International Conference on Systems Documentation (SIGDOC'86). 8-11 June. University of Toronto, 24-26.

Lewandowski, D. (2011). The Retrieval Effectiveness of Search Sngines on Navigational Queries. *ASLIB Journal of Infomation Management, 63*(4), 354-363.

Li, F., Yi, K., and Le, W. (2010). Top-k Queries on Temporal Data. *The VLDB Journal—The International Journal on Very Large Data Bases, 19*(5), 715-733.

Li, Y., Wen, A., Lin, Q., Li, R., and Lu, Z. (2014). Name Disambiguation in Scientific Cooperation Network by Exploiting User Feedback. *Artificial Intelligence Review, 41*(4), 563-578.

Lin, D. (1998). *Automatic Retrieval and Clustering of Similar Words*. 17th International Conference on Computational Linguistics, Vol. 2 (COLING'98). August 10-14. Université de Montréal, Montreal, Quebec, Canada, 768-774.

Lin, S., Jin, P., Zhao, X., and Yue, L. (2014). Exploiting Temporal Information in Web Search. *Expert Systems with Applications, 41*(2), 331-341.

Littorin, J., and Taslaman, N. (2009). Supervised Word Sense Disambiguation. *http:// fileadmin.cs.lth.se*

Liu, S., Yu, C., and Meng, W. (2005). *Word Sense Disambiguation in Queries*. 14th ACM International Conference on Information and Knowledge Management (CIKM'05). October 31 - November 05. Bremen, Germany, 525-532.

Luo, C., Liu, Y., Zhang, M., and Ma, S. (2014). Query Ambiguity Identification Based on User Behavior Information. *Information Retrieval Technology* (pp. 36-47). Switzerland: Springer International Publishing.

Maeda, A., Sadat, F., Yoshikawa, M., and Uemura, S. (2000). *Query Term Disambiguation for Web Cross-language Information Retrieval Using a Search Engine*. 5th International Workshop on Information Retrieval with Asian Languages (IRAL'00). September 30 - October 01. Hong Kong, China 25-32.

Makris, C., Plegas, Y., and Stamou, S. (2012). Web Query Disambiguation Using PageRank. *Journal of the American Society for Information Science and Technology, 63*(8), 1581-1592.

Makvana, K., Shah, P., and Shah, P. (2014). *A Novel Approach to Personalize Web Search Through User Profiling and Query Reformulation*. Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on. 5-6 Sept. New Delhi, India, 1-10.

Mallery, J. C. (1988). *Thinking About Foreign Policy: Finding an Appropriate Role for Artificially Intelligent Computers.* Master's Thesis. MIT Political Science Department, United States.

Manica, E., Dorneles, C. F., and Galante, R. (2010). Supporting Temporal Queries on XML Keyword Search Engines. *Journal of Information and Data Management, 1*(3), 471.

Manica, E., Dorneles, C. F., and Renata Galante, R. (2012). Handling Temporal Information in Web Search Engines. *ACM SIGMOD Record, 41*(3), 15-23.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval* (1st ed.). New York, USA: Cambridge University Press

McCulloch, W. S., and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115-133.

McRae-Spencer, D. M., and Shadbolt, N. R. (2006). *Also by the Same Author: AKTiveAuthor, A Citation Graph Approach to Name Disambiguation*. 6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'06). 11-15 June. Chapel Hill, NC, USA, 53-54.

Mihalcea, R. (2004). *Co-training and Self-training for Word Sense Disambiguation*. Conference on Computational Natural Language Learning (CoNLL-2004). May 6-7. Boston, MA, USA, 33-40.

Mihalcea, R., and Csomai, A. (2007). *Wikify!: Linking Documents to Encyclopedic Knowledge*. 16th ACM Conference on Information and Knowledge Management (CIKM'07). 6-8 November. Lisboa, Portugal, 233-242.

Mihalcea, R., and Tarau, P. (2004). *TextRank: Bringing Order into Texts*. Conference on Empirical Methods in Natural Language Processing (EMNLP 2004). 25-26 July. Barcelona, Spain, 404-411.

Mihalkova, L., and Mooney, R. (2009). Learning to Disambiguate Search Queries from Short Sessions. *Machine Learning and Knowledge Discovery in Databases* (pp. 111-127). Berlin Heidelberg: Springer

Milne, D., and Witten, I. H. (2008). *Learning to Link with Wikipedia*. 17th ACM conference on Information and knowledge management (CIKM'08). October 26 - 30. Napa Valley, CA, USA, 509-518.

Minkov, E., Cohen, W. W., and Ng, A. Y. (2006). *Spatial Search and Name Disambiguation in Email using Graphs*. 29th annual international ACM SIGIR conference on Research and development in information retrieval. 6-11 August. Seattle, Washington, USA, 27-34.

Mizzaro, S., and Vassena, L. (2011). A Social Approach to Context-aware Retrieval. *World Wide Web: Internet and Web Information Systems, 14*(4), 377-405.

Navigli, R. (2009). Word Sense Disambiguation: A Survey. *ACM Computing Surveys (CSUR), 41*(2), 10.

Ng, H. T. (1997). *Getting Serious About Word Sense Disambiguation*. ACL SIGLEX ' 97 Workshop on Tagging Text with Lexical Semantics: Why, What, and How? 04 April. Washington, D.C., USA, 1-7.

Nguyen, H. T., and Cao, T. H. (2010). Exploring Wikipedia and Text Features for Named Entity Disambiguation. *Intelligent Information and Database Systems* (pp. 11-20). Berlin Heidelberg: Springer.

Nunes, S. (2007). Exploring Temporal Evidence in Web Information Retrieval. *Future Directions in Information Access (FDIA).*

Nunes, S., Ribeiro, C., and David, G. (2007). *Using Neighbors to Date Web Documents*. 9th Annual ACM International Workshop on Web Information and Data Management (WIDM'07). 9 November, 129-136.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web *Techreport (Technical Report), Stanford InfoLab* (pp. 1-17).

Palacio, D., Derungs, C., and Purves, R. (2015). Development and Evaluation of a Geographic Information Retrieval System Using Fine Grained Toponyms. *Journal of Spatial Information Science*. 11(2015). 1-29.

Pantel, P., and Lin, D. (2002). *Discovering Word Senses from Text*. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'02). July 23-26. Edmonton, Alberta, Canada, 613-619.

Patil, M. S. C., and Keole, R. (2014). The Role of Web Content Mining and Web Usage Mining in Improving Search Result Delivery. *International Journal of Computer Science and Mobile Computing (IJCSMC).* 3(3), 7-14.

Pedersen, T. (2006). Unsupervised Corpus-Based Methods for WSD. In E. Agirre and P. Edmonds (eds.), *Word Sense Disambiguation: Algorithms and Applications* (pp. 133-166). Netherlands: Springer

Pedersen, T., Banerjee, S., and Patwardhan, S. (2005). Maximizing Semantic Relatedness to Perform Word Sense Disambiguation (Vol. 25): University of Minnesota Supercomputing Institute Research Report UMSI 2005/25. USA.

Ponzetto, S. P., and Navigli, R. (2010). *Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems*. 48th Annual Meeting of the Association for Computational Linguistics. 11-16 July. Uppsala, Sweden, 1522-1531.

Preotiuc-Pietro, D., and Hristea, F. (2014). Unsupervised Word Sense Disambiguation with N-gram Features. *Artificial Intelligence Review, 41*(2), 241-260.

Purandare, A., and Pedersen, T. (2004). *Improving Word Sense Discrimination with Gloss Augmented Feature Vectors*. Workshop on Lexical Resources for the Web and Word Sense Disambiguation. 22 November. Puebla, Mexico, 123-130.

Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. *New directions in question answering, 3*, 28-34.

Quinlan, J. R. (2014). *C4. 5: Programs For Machine Learning* (1st ed.). California, USA: Morgan Kaufmann.

Ratinov, L., Roth, D., Downey, D., and Anderson, M. (2011). *Local and Global Algorithms for Disambiguation to Wikipedia*. 49th Annual Meeting of the

Association for Computational Linguistics: Human Language Technologies-Volume 1 (HLT'11). June 19-24. Portland, Oregon, USA, 1375-1384.

Regazzi, J. J. (1988). Performance Measures for Information Retrieval Systems—An Experimental Approach. *Journal of the American Society for Information Science, 39*(4), 235-251.

Rivest, R. L. (1987). Learning Decision Lists. *Machine learning, 2*(3), 229-246.

Roberto, N., and Giuseppe, C. (2010). *Inducing Word Senses to Improve Web Search Result Clustering*. Conference on Empirical Methods in Natural Language Processing. 9-11 October. MIT, Massachusetts, USA 116-126.

Rose, D. E., and Levinson, D. (2004). *Understanding User Goals in Web Search*. 13th International Conference on World Wide Web (WWW'04). May 17 - 22. New York, NY, USA, 13-19.

Roul, R. K., and Sahay, S. K. (2012). An Effective Information Retrieval for Ambiguous Query. *International Journal of Computer Science and Information Technologies (IJCSIT), 2*(3), 26-30.

Salton, G. (1971). *The SMART retrieval system—Experiments in Automatic Document Processing*: Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Salton, G., and Buckley, C. (1988). Term-weighting Approaches in Automatic Text Retrieval. *Information Processing & Management, 24*(5), 513-523.

Sanderson, M. (2008). *Ambiguous Queries: Test Collections Need More Sense*. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 20-24 July. Singapore, 499-506.

Saygin, A. P., Cicekli, I., and Akman, V. (2003). Turing Test: 50 Years Later. *Minds and Machines* (pp. 463-518). Hingham, USA: Kluwer Academic.

Schutze, H. (1992). *Dimensions of Meaning*. Supercomputing '92. Proceedings of the 1992 ACM/IEEE Conference on Supercomputing. 16-20 November. Minneapolis, MN, USA. 787-796.

Shang, S. S., Li, E. Y., Wu, Y.-L., and Hou, O. C. (2011). Understanding Web 2.0 Service Models: A Knowledge-creating Perspective. *Information & Management, 48*(4), 178-184.

Shekarpour, S., Ngomo, A.-C. N., and Auer, S. (2013). Keyword-Driven Resource Disambiguation over RDF Knowledge Bases. *Semantic Technology* (pp. 159-174). Berlin Heidelberg: Springer.

Simpson, G. B., and Kang, H. (1994). Inhibitory Processes in the Recognition of Homograph Meanings. *Inhibitory Processes in the Recognition of Homograph Meanings* (pp. 359-381). San Diego: Academic Press.

Singhal, A. (2001). Modern information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24(4), 35-43.

Skaggs, B., and Getoor, L. (2014). Topic Modeling for Wikipedia Link Disambiguation. *ACM Transactions on Information Systems (TOIS), 32*(3), 1-24.

Skaggs, B. A. (2011). *Topic Modeling for Wikipedia Link Disambiguation.* Doctor of Computer Science, University of Maryland (College Park, Md.).

Song, R., Luo, Z., Nie, J.-Y., Yu, Y., and Hon, H.-W. (2009). Identification of Ambiguous Queries in Web Search. *Information Processing & Management, 45*(2), 216-229.

Song, R., Luo, Z., Wen, J.-R., Yu, Y., and Hon, H.-W. (2007). *Identifying Ambiguous Queries in Web Search.* 16th International Conference on World Wide Web (WWW '16). May 8-12. Banff, Alberta, Canada, 1169-1170.

Speretta, M., and Gauch, S. (2005). *Personalized Search based on User Search Histories*. Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on. 19-22 Sept. Compiègne University of Technology, France, 622-628.

Spink, A., Wolfram, D., Jansen, M. B., and Saracevic, T. (2001). Searching the Web: The Public and their Queries. *Journal of the American Society for Information Science and Technology, 52*(3), 226-234.

Strube, M., and Ponzetto, S. P. (2006). *WikiRelate! Computing Semantic Relatedness Using Wikipedia*. 21st National Conference on Artificial Intelligence (AAAI'06). 16 -20 July. Boston, Massachusetts, 1419-1424.

Struber, D., Rubin, J., Taentzer, G., and Chechik, M. (2014). Splitting Models Using Information Retrieval and Model Crawling Techniques. *Fundamental Approaches to Software Engineering* (pp. 47-62). Berlin Heidelberg: Springer.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 645-659.

Sy, M.-F., Ranwez, S., Montmain, J., Regnault, A., Crampes, M., and Ranwez, V. (2012). User Centered and Ontology based Information Retrieval System for Life Sciences. *BMC Bioinformatics, 13*((Suppl 1):S4), 1-12.

Takase, W., Hasan, A., Matsumoto, Y., and Sasaki, T. (2014). *Capturing User-Generated Metadata*. The International Symposium on Grids and Clouds (ISGC). 23-28 March. Academia Sinica, Taipei, Taiwan.

Tamine-Lechani, L., Boughanem, M., and Daoud, M. (2010). Evaluation of Spatial Information Retrieval Effectiveness: Overview of Issues and Research. *Knowledge and Information Systems, 24*(1), 1-34.

Tamine-Lechani, L., Boughanem, M., and Zemirli, N. (2008). Personalized Document Ranking: Exploiting Evidence from Multiple User Interests for Profiling and Retrieval. *Journal of Digital Information Management, 6*(5), 354-365.

Thada, V., and Jaglan, V. (2013). Web Information Retrieval. *International Journal of Computer Applications, 76*(1), 29-32.

Van Den Brand, M. G., Scheerder, J., Vinju, J. J., and Visser, E. (2002). Disambiguation Filters for Scannerless Generalized LR Parsers. *Compiler Construction* (Vol. 2304, pp. 143-158). Berlin Heidelberg: Berlin Heidelberg.

Véronis, J. (2004). Hyperlex: Lexical Cartography for Information Retrieval. *Computer Speech & Language, 18*(3), 223-252.

Vieira, V., Tedesco, P., Salgado, A. C., and Brézillon, P. (2007). Investigating the Specifics of Spatial Elements Management: The CEManTIKA Approach. *Modeling and Using Context* (pp. 493-506). Berlin Heidelberg: Springer

Wang, X., Xu, M., Ren, Y., Xu, J., Zhang, H., and Zheng, N. (2014). A Location Inferring Model Based on Tweets and Bilateral Follow Friends. *Journal of Computers, 9*(2), 315-321.

Weikum, G., Ntarmos, N., Spaniol, M., Triantafillou, P., Benczúr, A. A., Kirkpatrick, S., Rigaux, P., and Williamson, M. (2011). *Longitudinal Analytics on Web Archive Data: it's about time!* Conference on Innovative Data Systems Research (CIDR). 9-12 January. Asilomar, California, 199-202.

Whang, S. E., Menestrina, D., Koutrika, G., Theobald, M., and Garcia-Molina, H. (2009). *Entity Resolution with Iterative blocking*. ACM SIGMOD International Conference on Management of data (SIGMOD '09). June 29–July 2. Providence, Rhode Island, USA., 219-232.

Widdows, D., and Dorow, B. (2002). *A Graph Model for Unsupervised Lexical Acquisition*. 19th International Conference on Computational Linguistics-Volume 1 (COLING '02). August 24 - September 1. Taipei, Taiwan, 1-7.

Wikipedia. (2014). Wikipedia :Statistics English, from http://stats.wikipedia.org/EN/TablesWikipediaEn.htm

Winokur, M. (2015). The Ambiguous Panopticon: Foucault and the Codes of Cyberspace. *CTheory*, 3-13/2003.

Xue, G.-R., Zeng, H.-J., Chen, Z., Yu, Y., Ma, W.-Y., Xi, W., and Fan, W. (2004). *Optimizing Web Search Using Web Click-through Data*. 13th ACM International Conference on Information and Knowledge Management (CIKM'04). 8-13 November. Washington, DC, USA, 118-126.

Yih, W.-T., and Meek, C. (2007). *Improving Similarity Measures for Short Segments of Text*. 22nd Conference on Artificial Intelligence (AAAI-07). 22-26 July. Vancouver, British Columbia, 1489-1494.

Zahariadis, N. (2014). *Ambiguity and Multiple Streams*. In Paul A. Sabatier and Christopher M Weible (Eds.) *Theories of the Policy Process*. (pp. 25-58). University of California, USA: WESTVIES Press.

Zhai, C. X., Cohen, W. W., and Lafferty, J. (2003). *Beyond Independent Relevance: Methods and Evaluation Metrics for Subtopic Retrieval*. 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval (SIGIR'03). July 28–August. Toronto, Canad, 10-17.

Zhang, H. (2013). *Query Enhancement with Topic Detection and Disambiguation for Robust Retrieval*. Doctor of Computer Science, Indiana University, United States.

Zhao, K., Cai, Z., Sui, Q., Wei, E., and Zhu, K. Q. (2014). Clustering Image Search Results by Entity Disambiguation. *Machine Learning and Knowledge Discovery in Databases* (pp. 369-384). Berlin Heidelberg: Springer.