Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 244-248

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

An Experimental Analysis of the Impact of Domain Independent Ontology over Navigational Queries

[1]Ruban S and [2]Behin Sam

[1]*Research Scholar, Dept of Computer Science, Bharathiar University.*
[2]*Asst Professor, Rajeswari Vedachalam College, Chengalpattu, TamilNadu, India.*

ARTICLE INFO

Article history:
Received 20 January 2015
Accepted 02 April 2015
Published 20 May 2015

Keywords:
Query, WordNet, Query Expansion, Navigational query, Information Retrieval.

ABSTRACT

Information Retrieval Applications has been part of our day to day lives. They play a vital role in handling various aspects of our human life such as Information seeking, capturing the information available in different formats and information analysis. One of the greatest benefits that these applications offer is to retrieve relevant information from a mountain of unstructured data that is available. The common and widely used strategy that most of the search engines adopt is to search using the keywords. Though this strategy helps us to retrieve data, the precision of the data retrieved can be improved further by enhancing the query that is given by the user. To enhance the query it is essential to understand the user intent behind this. There are different classifications of queries that have been proposed over the period of time. One such classification, categorizes queries into Navigational, informational and transactional. In this experimental study we analyze the impact of Domain independent ontology when expanding the navigational queries.

To Cite This Article: Ruban S and Behin Sam., An Experimental analysis of The Impact of Domain Independent Ontology Over Navigational Queries. ***Aust. J. Basic & Appl. Sci.,*** *9(16): 244-248, 2015*

INTRODUCTION

Information Retrieval plays a vital role in handling various aspects of our human life such as Information seeking, capturing the information available in different formats and the information analysis. One of the greatest benefits that these applications offer is to retrieve relevant information from a mountain of unstructured data that is available. The common and widely used strategy that most of the search engines adopt is to search using the keywords. Though this strategy helps us to retrieve data, the precision of the data retrieved can be improved further by enhancing the query that is given by the user. To enhance the query it is essential to understand the user intent behind this.

Query:

In brief query can be considered as a user information need. There are many works that were aimed to address the concept of query. One assumption that is being shared by many literatures in this area conveys that, a thought pattern generated by the user is being put forward as a query. There is therefore no standard defined way how a query has to be written. Hence different search engines engage in different ways by which they try to understand what the user means by what he has expressed as the user information need. One of the areas by which they process the user information need is by adding new related terms to the query hence making the user query much more responsive to the information retrieval system. The process of adding new terms to the query is called as query expansion or query enhancement or query refinement or query augmentation. Though Query expansion has always seen as a way of improving the relevant results, it may not be always. In this experimental study we have undertaken, we will be pointing out instances of queries which may not be improved no matter how you enhance the query.

Related Work:

Search engines have been looking out for methods through which they can improve the behavior of the system. They try to understand the behavior of the user and the background i.e., intent behind the search. Based on the user intent the following classification of web queries are found in the literature.

In the year 1992 (Carmel, Crawford and Chen, 1992) in their study on browsing, they divided browsing into three types such as a) Browsing oriented towards search which is a process of finding

Corresponding Author: Ruban S., Research Scholar, Dept of Computer Science, Bharathiar University.

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 244-248

relevant information for a particular task b) Browsing oriented towards review, which is the way of looking out for some information and c) Browsing oriented towards scan, which is towards finding information without involving review. Later in the year 1995 (Marchionini, 1995) in his work on Information seeking also classified the browsing patterns into directed, semi directed and undirected browsing.

(Andrei Broder , 2002) at IBM Research [3] proposed the widely known classification based on user intent as navigational web queries, Informational web queries and Transactional web queries. His classification was however based on some search logs that he collected when he was working with AltaVista Corporation and later was supported by the surveys. His experiments revealed that based on query log analysis 30% are transactional web queries, whereas 48% are informational and 20% are navigational web queries. Out of the user survey he conducted he reveals that an estimate of 39% is informational whereas an estimate of 36% is transactional and 24.5% are navigational.

(Jansen, Spink and Pedersen, 2005) based on their experiment conducted with AltaVista states that there is an increased use of search engines for navigation. Because of the hypermedia capability of the web, it provides a very unique way of browsing where we use search engines to move to another site of interest or of need. Another classification of queries was made by (Rose and Levinson, 2004) based on the query user used to search, the results which the user viewed. They sub categorized as informational queries, navigational queries and then resource queries.

The studies that were quoted above were based on query logs from the Search Engines and many of the classifications were manually done. Some of the works that followed automatic classification are as follows.(Lee, Liu and Cho, 2005) in their work based on the automatic identification of user goals classified the queries into informational and navigational queries. They did this experiment with 50 queries collected from a group of students in a university. They could experience a success rate of over 50%.

(Dai et al, 2006) in their work on detecting commercial intention studied whether the user information need given by user has any commercial motivation or not. They later revealed that 38% of the queries that were taken for the study have commercial intention.(Baeza-Yates, Calderon Benavides and Gonzalez, 2006) in their work on the intention behind the web queries used supervised and unsupervised learning to divide the queries as informational or not informational or ambiguous. They declared that they were able to achieve 50% of precision.

(Bernard J. Jansen, Danielle L. Booth and Amanda Spink,2008) in their work on determining the user intent of queries defined and presented three levels of hierarchical classification of user intent for searching web. In the next section we will be explaining about the material and methods in which we will be talking about the research objectives specifically with respect to this experimental study, which will be followed by Results, then Discussion and finally the conclusion, followed by the References.

MATERIALS AND METHODS

In this chapter let me narrate the materials and methods that we have used as part of this experimental study. The research objectives that were formed as part of out experimental study are listed out below.

Research Objectives:

1. Develop a Query Expansion methodology to enhance the query:

For research objective one we analyzed prior work in this area with an analysis of numerous actual query expansion approaches that has been developed and proposed over the period of time. It is very difficult to compare results across the different studies that have happened. But we would like to use the query Expansion that used ontology.

2. Identify the Navigational Queries:

For Research objective two, we followed the classification that was laid down by (Andrei Broder, 2002). This study was later redefined by (Bernard J.Jansen, Danielle L.Booth and Amanda Spink, 2008) where they did a automatic classification of queries that were used by the users. Based on the user intent, the queries are divided into navigational, informational and transactional. we identified and selected few queries which are navigational in nature, i.e., the user intent was to reach a particular site of interest.

3. Implement the Query Expansion methodology for the navigational queries:

For this research objective, we took the navigational queries that we have identified in the previous research objective and passed through query expansion method identified in the first research objective. The query is passed through the Stanford parser, where the initial parsing is done. Later it is passed to the Query expansion phase where the identified keywords of the query are taken into consideration and all its relevant terms which are called as synsets are displayed. Once the user selects the relevant synsets, they are added to the original query hence forms the refined query which is then passed to the Search API.

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 244-248

4. Evaluate the performance of the expanded queries using the comparative study with any Search API:

The evaluation of any Information Retrieval system is based on the precision and recall values. Since we evaluate our system directly in the web resource, it is impossible to calculate the correct the recall value because the environment in which we test is a dynamic and fluid environment. Hence we will be calculating the precision value for every individual query then we together calculate the Average Precision value. We evaluate our system in two levels: Case i : Navigational Queries directly given to the Search API. Case ii: Navigational Queries that has been refined using Ontology which is then given to the Search API. The queries were given to the Google API and first 100 results were manually checked to know whether they are relevant.

Results:

The Experimental evaluation was conducted in the span of 3 months from the month of December 2014 to February 2015. The above queries were executed in two ways. We choose to use Google's Search API for our experiment and overall coding was done using JENA a Java based Framework.

Case i: Navigational Queries directly given to the Search API.

Case ii: Navigational Queries that has been refined using Ontology which is then given to the Search API.

In the first case the queries were directly given to the search API, the first 100 results were manually evaluated for relevance, and the precision was calculated for every query.

Table 1: Navigational Queries with their precision values

1.	Anand Bus Travels Mangalore	55
2.	Oriental Bank Of Commerce	99
3.	Malaysian Airlines	97
4.	World Health Organization	99
5.	The Times Of India	66
6.	Vellore Institute Of Technology	97
7.	Software Engineering by Roger Pressman	100
8.	Indira Gandhi National Open University	98
9.	Euro Kids School Mangalore	78
10.	Ginger Hotel Mysore	87
11.	Young Men Christian Organization	94
12.	First book	93
13.	The Hindu	77
14.	ChetanBhagat	82
15.	Dell Computers	86

In the second case, the queries were given to the Stanford parser, then they were given as input to the ontology i.e., Wordnet, the terms that were generated are called as synsets, the relevant terms are added to the original query hence the new refined query is formed, which is then passed to the Search API. The first 100 results were manually evaluated for relevance and the precision is calculated. The following table displays the original value that was got during the experiment when the queries were passed to the Search API.

Table 2 : Refined Navigational Queries with their precision

1.	Anand Bus or Omni Bus Mangalore	8
2.	Oriental Bank Or Banking Company Of Commerce	66
3.	Malaysian Airlines or Airlines Business	87
4.	World Or Universe Health Or Welfare Organization	44
5.	The Times Or Clock Of India Or Bharat	21
6.	Vellore Institute Or Engineering Of Technology	64
7.	Software Engineering Or Computer Software by Roger Pressman	48
8.	Indira Gandhi Or Gandhi National Open University	85
9.	Euro Kids School Or Schooling Mangalore	40
10.	Ginger or Gingery Hotel Mysore	62
11.	Young Men Or Gentlemen Christian Organization	35
12.	First book or good book	20
13.	The Hindu	77
14.	ChetanBhagat	82
15.	Dell Computers or computing machine	44

Discussion:

The following figures display the comparative results of the navigational queries without applying ontology and applying ontology to it.

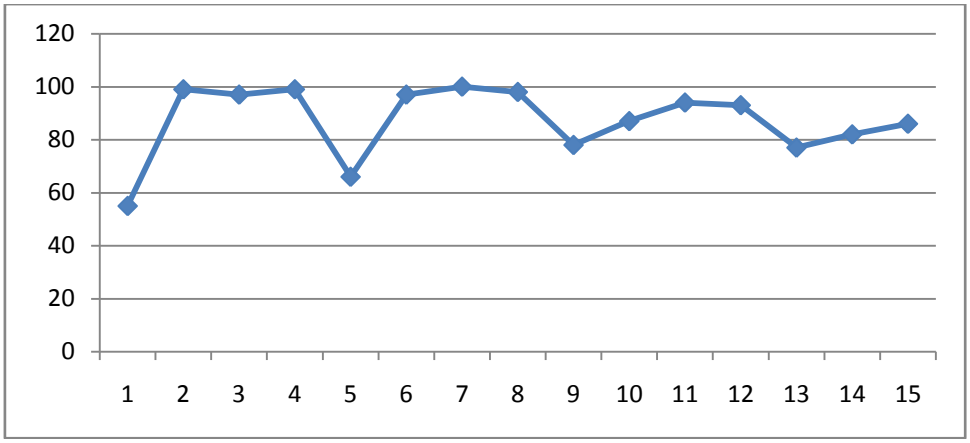

Fig. 1: Navigational Queries with their precision values

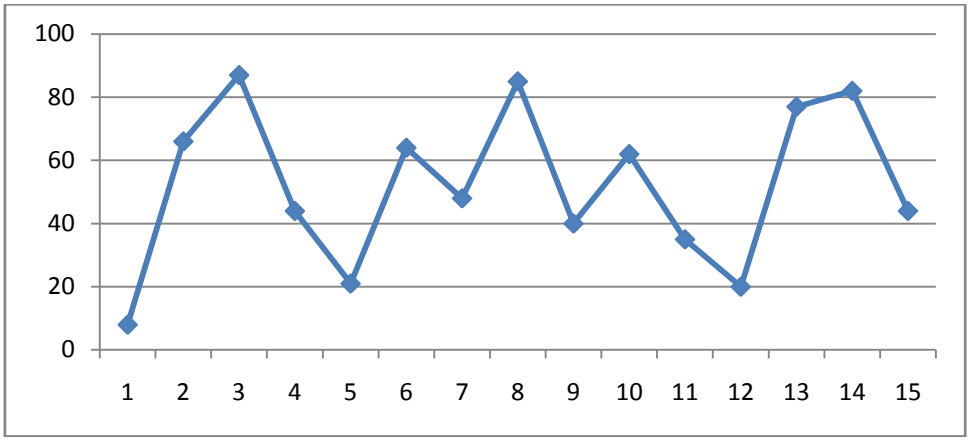

Fig. 2: Refined Navigational Queries with their precision values

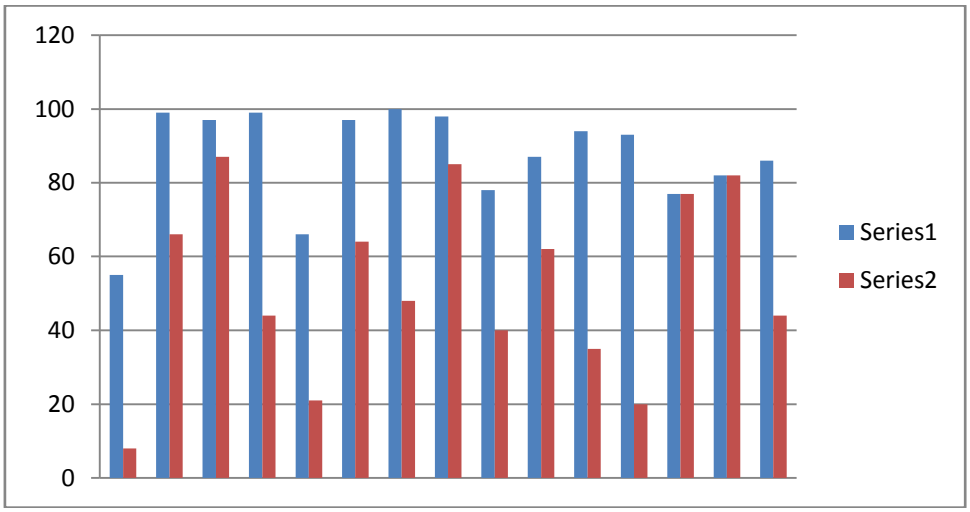

Fig. 3: Navigational Queries Vs Refined Navigational Queries with their precision values

Since there is a drastic decrease in the precision value of the navigational queries when we use ontology to it, It is hereby concluded that Any Query which is of navigational in nature will never be improved using Query expansion using ontology. The following table displays the group statistics. Query 1 represents the set of Navigational queries and Query 2 represents the set of refined Navigational queries.

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 244-248

Table 3: Group Statistics

	NavQuery	N	Mean	Std. Deviation	Std. Error Mean
Result	Query1	15	87.20	13.455	3.474
	Query2	15	52.20	25.038	6.465

Conclusion:

If the search engine performance has to improve, it is very necessary to understand the user intent behind the user information need that is given for searching. Though, it is proved that though query expansion a query can be made into a more responsive representation which will retrieve more relevant results. But in our study we have proved that in the case of Navigational queries, it will decrease the performance of the system. Though the results we have obtained may vary based on the timing of execution of the queries, but based on our experimental results we conclude that any query that is aimed at using a search engine to navigate to a particular site or location will never benefit from domain independent ontology based query expansion.

REFERENCES

Baeza-Yates, R., L. Calder´on-Benavides and C. Gonz´alez, 2006. In The intention behind Web queries (pp. 98–109). Paper presented at the string processing and information retrieval (SPIRE 2006), 11–13 October, Glasgow, Scotland.

Bernard, J. Jansen, Danielle L. Booth, Amanda Spink, 2008. Determining the informational, navigational and transactional intent of web queries, Information processing and Management., 44: 1251-1266.

Broder, A., 2002. A taxonomy of Web search. SIGIR Forum, 36(2): 3-10.

Carmel, E., S. Crawford and H. Chen, 1992. In Browsing in hypertext: A cognitive study (pp. 865–884). Paper presented at the IEEE transactions on systems, man and cybernetics, 5–10 October, Chicago IL.

Dai, H.K., Z. Nie, L. Wang, L. Zhao, J.-R. Wen and Y. Li, 2006. In Detecting online commercial intention (OCI) (pp. 829–837). Paper presented at the World Wide Web conference (WWW2006), 23–26 May, Edinburgh, Scotland.

Jansen, B.J., A. Spink and J. Pedersen, 2005. Trend analysis of AltaVista Web searching. Journal of the American Society for Information Science and Technology, 56(6): 559–570.

Lee, U., Z. Liu and J. Cho, 2005. In Automatic identification of user goals in Web search (pp. 391–401). Paper presented at the World Wide Web conference, 10–14 May, Chiba, Japan.

Marchionini, G., 1995. Information seeking in electronic environments. Cambridge: Cambridge University Press.

Rose, D.E., and D. Levinson, 2004. In Understanding user goals in Web search (pp. 13–19). Paper presented at the World Wide Web conference (WWW 2004), 17–22 May, New York, NY, USA.