International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 506-509
e-ISSN:2278-621X

A STUDY ON THE IMPACT OF WORDNET FOR QUERY EXPANSION OVER GENERAL SEARCH AND VERTICAL SEARCH ENGINES

Mr. Ruban.S[1], Chaithra[2] & Chaithranjali[3]

Abstract: The role of General search Engines in searching relevant information from the web is inevitable. Apart from the General search engines that can be used to search any information in the web, there is also a category of search Engines that focuses only on a particular category of Web content which are normally called as Vertical Search Engines. They are also referred to as Topical Search Engines or specialty Search Engines. The vertical Search Engine area may be based on a domain, Topic or content genre. These search engines are also used by a vast majority of users who rarely spend time in formulating the query that will give better results while retrieving information. So it is up to the respective retrieval systems to define and use methods that will help to reformulate a query into a more responsive one which will help the search Engine to yield better results. This paper studies the impact of ontology over the query used in General search Engines and vertical search Engines.
Keywords: Information Retrieval, Query Expansion, Vertical Search Engine, Thesaurus.

1. INTRODUCTION

Information Retrieval is a domain that existed before the presence of World Wide Web, yet at any rate it was just after the approach of web crawler, Retrieval has advanced toward getting to be and indispensable bit of the web preparing Framework. An IR system is planned to find a scrap of Information that is sensible to customer information required which is imparted by a request. Ordinarily the IR system explores an enormous gathering of data which may have structure or not. IR structure is always used when the degree of collections gets more prominent and immeasurable size where the pioneer posting system can't work. With a regularly expanding number of data of different sorts, structure and nature, added to the web at reliably, it is hard to find some other course beside using the interest applications to find correlated things.

A large portion of the work that was done around there focus on the calculations which takes the data need of the client as the info and get some applicable archives as yield. At present, systems for look are absolutely in light of Keywords and no extent of seeking in view of ideas or significance of the substance .

The Information Retrieval space has moved from its center objectives of Text Indexing and Searching for applicable records in a store because of the impact of the web . The achievement of getting proper data to a question is impacted by the client work and furthermore by the thinking of the archives grasped by the Information Retrieval framework. Data Retrieval can be valuable for the improvement, usage and assessment of a web index. Data Retrieval models contribute towards the improvement of the Information recovery framework. Boolean model, Vector demonstrates and Probabilistic Information Retrieval shows are considered as the run of the Information Retrieval models over the time frame [1]. Throughout the years elective demonstrating ideal models for each kind of the established models have been proposed. Despite the fact that every one of these models have prompted the improvement of Information Retrieval frameworks, they have their own constraints and issues that they endure. Thus none of the Information Retrieval framework grew so far can be considered as an immaculate, errorless Information Retrieval framework.

In the next section we will be explaining about Vertical Search Engines, then the Literature review showing the role of Ontology in Query Expansion, the implementation of the system, results, discussion and then will be the conclusion.

2. VERTICAL SEARCH ENGINES

The part of General web crawlers in seeking pertinent data from the web is unavoidable. Aside from the broadly utilized web indexes there is likewise a class of web indexes that spotlights just on a specific classification of Web content which are typically called as Vertical Search Engines. They are additionally alluded to as Topical Search Engines or claim to fame Search Engines. The vertical Search Engine region might be founded on an area, Topic or substance kind. As of late these web search tools are additionally broadly utilized on the grounds that the importance of the outcomes that these web search tools deliver are more and particular than the thousand of results the other universally useful web indexes give. These web crawlers are additionally utilized by a dominant part of clients who once in a while invest energy in defining the question that will give better outcomes while recovering data. So it is up to the separate recovery frameworks to characterize and utilize techniques

[1] Assistant Professor, Department of MCA,AIMIT, St Aloysius College,Mangalore-575022,'karnataka,India
[2] Student , Department of MCA, AIMIT St Aloysius College,Mangalore-575022,'karnataka,India
[3] Student , Department of MCA,AIMIT St Aloysius College,Mangalore-575022,'karnataka,India

that will reformulate an inquiry into a more responsive one which will assist the web crawler with yielding better outcomes. For this investigation we utilized two search engines one general search engine called Bing and another is the vertical web indexes called as Trulia. Trulia is web look entry situated in United States that helps home purchasers, land experts and dealers with data identified with land industry.Their imaginative inquiry components finds the homes and related data that is significant for home purchasers.

3. LITERATURE REVIEW

Recent studies have suggested the usage of ontology for query reformulation. Word net is one of the commonly known ontology which is used is many studies conducted in this perspective. G. A. Miller developed Word net in Princeton University. [2] Voorhees [3] using word net for query reformulation in her studies concluded that it has a positive impact over the short queries.

In the work Probabilistic query expansion using query logs [4] the authors considered adding different words from the logs that are already available to the actual query as part of the reformulation procedure. Authors in the study based on improving weak ad-hoc queries [5] considered adding words from Wikipedia as part of query reformulation procedure. In the study titled Query Manipulation involving Multiple Information sources [6] the authors used the words taken from the web.

Query reformulation also witnessed the usage of ontology [7]. Ontology can be developed taking a particular domain into consideration which we refer as Domain dependent ontology or it also contains words which are very general, hence we call them as Domain independent ontology. Word net is an example of Domain Independent ontology. Some studies done in this context showed enhancement, there are also occurrences where the system performance has been degraded because of the ontology. In the study titled a re examination of query expansion using lexical resources [8] the author's experiment shows an improvement in the system results. This study also proves the point that; ontology when properly used will help to enhance the performance. In a study titled Query Expansion using Term distribution and Query Association [9] the authors studies the impact of expanding the query using the association and the distribution that exists between the words. This study also reveals a drastic improvement in the system performance.

Our proposed work in this study is to analyze the impact of word net over queries that are executed in the General search engines and vertical search engines.

4. MATERIALS AND METHODS

4.1. Query Expansion methodology:

Query Expansion procedure takes the actual query given by the user as the input and considers every keyword present as the actual seed terms. The Query Expansion is done using Ontology. In this study we used Word net which is a widely and commonly used in many studies. Domain dependent ontology can also be used for the study. It will be however useful only for the domain specific queries which pertain only for that domain.

The methodology involves adding more words from the ontology which is relevant to the actual seed terms, Instead of Automatic query reformulation we used Interactive query reformulation where the user picks the words to be added to the actual query. The reformulated query is later send to the Search Engine. In our study we used Bing and Trulia.

4.2. Selection of Queries

We chose couple of arbitrary questions, yet for our investigation since we focus on the effect of word net over the inquiries. Our examination does not arrange them into Transactional, navigational or Informational but rather simply take them as they seem to be. A few questions that we utilized as a part of the test think about are recorded beneath.

Table 1 : Experimental Queries

No	Experimental List
1.	Types of flats available in California
2.	Homes in New York
3.	3 bedroom house needed in Chicago
4.	Flats available in Westside, Atlanta
5.	Homes needed to keep pets in New Jersey

5. RESULT AND OUTCOME

The Study was planned and done in an overall duration of 3 months which involved collecting user information needs and executing them in different vertical search engines which were used for this study. Though we came across some Vertical search Engines, we did select Trulia and Bing a general search engine for our study and the code was written using JENA. The queries were analyzed in two different ways.

i : User Information need given directly to the General Search Engine and Vertical Search Engine.

ii: User Information need reformulated using Word net then given to the General and Vertical Search Engine.

The queries were given to the General or the vertical search Engine in the first case, and then the results were manually evaluated for relevance. Since it was a web environment, we considered the first 100 values and every individual query's precision was computed.. The details about the queries that were executed and their precision values are given below in the following tables Table2, Table3,

Queries	**Bing**	**Trulia**
Types of flats available in California	6	35
Homes in New York	77	49
3 bedroom house needed in Chicago	64	84
Flats available in Westside, Atlanta	82	28
Homes needed to keep pets in New Jersey	58	29

Table 2: Experimental queries without Query Expansion

Queries	**Bing wordnet**	**Trulia Wordnet**
Types and kinds of flats available in California	23	40
Homes and house in New York	82	23
3 bedroom or room house or house needed in Chicago	69	0
Flats or apartments available in Atlanta	87	0
Houses or homes needed to keep pets or animals in New Jersey	65	0

Table 3: Experimental queries with Query Expansion

6. DISCUSSION

The Analysis of the results derived from the experimental study on two search engines, i) General search engine and ii) vertical search Engines are plotted below.

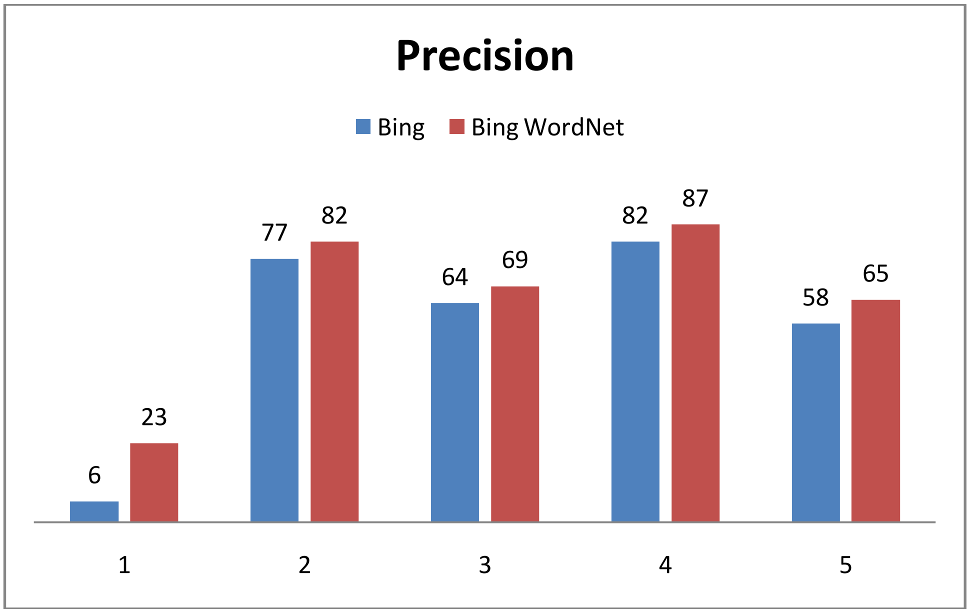

Figure 1 : Precision in Bing Vs Bing WordNet

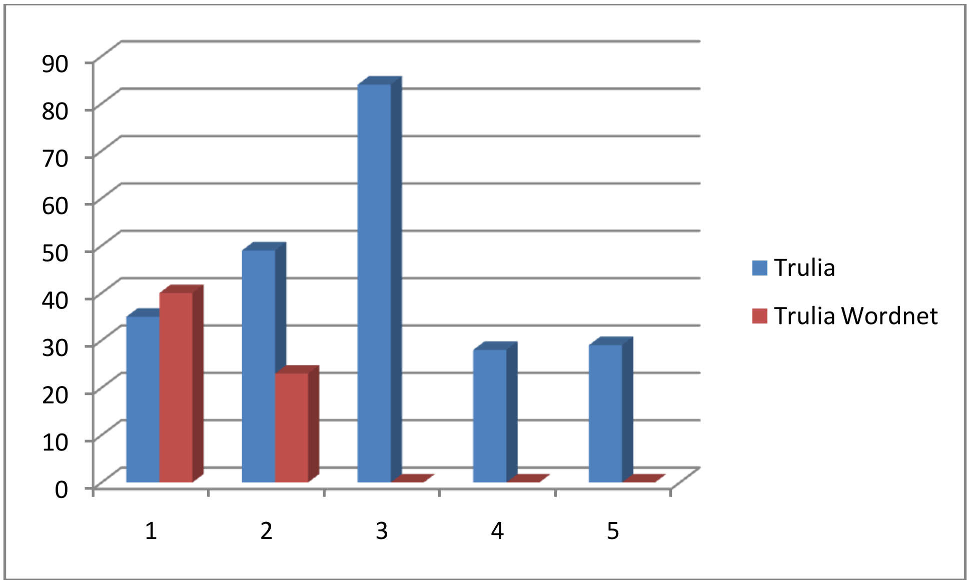

Figure 2 : Precision in Trulia Vs Trulia WordNet

All the above figures (Fig1, Fig2) that represent the comparison of queries executed before and after reformulation clearly indicates that, the query reformulation using the domain independent ontology has some improvement in General search engine whereas the domain independent ontology is not having a greater influence in the system performance or will have no influence in the system performance.

7. CONCLUSION

Though, it is proved that though query reformulation, 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 query reformulation, the query reformulation using the domain independent ontology has some improvement in General search engine whereas the domain independent ontology is not having a greater influence in the system performance or will have no influence in the system performance of the vertical search engine, because of the fact that the vertical search engines are much focussed on a particular domain or specialization.

8. REFERENCES:

[1] R.Baeza-Yates and B.Ribeiro-Neto, "Modern Information Retrieval in practice", 1st ed. Reading, MA: Addison-Wesley, 2009.

[2] G.A Miller, 1990, Special Issue, "Wordnet:An on-line lexical database", International journal of Lexicography, 3(4).

[3] [3] Ellen M. Voorhees, 1994, "Query Expansion using Lexical-semantic relations", proceedings of the 17th ACM-SIGIR Conference, pages 61-69.

[4] [4] Cui et al., (2002) "Probabilistic query expansion using query logs", Proceedings of the 11th international conference on World Wide Web (pp. 325–332). New York, NY, USA: ACM.

[5] [5] Chung et al., (2007). "Improving weak ad-hoc queries using Wikipedia as external corpus". Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 797–798). New York, NY, USA: ACM.

[6] [6] Croft, W. B. (2012). "Effective query formulation with multiple information sources". Proceedings of the fifth ACM international conference on Web search and data mining (pp. 443–452).

[7] [7] Bhogal, J., Macfarlane, A., & Smith, P. (2007, July). "A review of ontology based query expansion", Journal of Information Processing and Management. 43(4), 866–886.

[8] [8] Fang, H. (2008). "A re-examination of query expansion using lexical resources" Proceedings of ACL-08: HLT (pp. 139–147).

[9] [9] Pal, D., Mitra, M., & Datta, K. (2013). "Query expansion using term distribution and term association" CoRR, abs/1303.0667.