I J C T A, 9(7), 2016, pp. 3103-3108 ISSN: 0974-5572
© International Science Press

A Study on The Impact of Wordnet for Query Reformulation over Vertical Search Engines

Ruban S.* and Behin Sam**

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 vertical search Engines.

Keywords: Vertical Search Engine, Query, Query Expansion, Ontology.

1. INTRODUCTION

Information Retrieval is an area that did exist before the advent of World Wide Web, but however it was only after the advent of Web search engine, Retrieval has become and integral part of the web computing Framework. An IR system is designed to find a piece of Information that is suitable to a user information need which is expressed by a query. Normally the IR system looks into a vast collection of data which may have structure or not. IR system is always used when the size of collections gets bigger and unimaginable size where the pioneer cataloguing system cannot work. With more and more data of different types, structure and nature, added to the web at every second, it is impossible to find any other way except using the search applications to find relevant items.

Most of the work that was done in this area concentrate on the algorithms which takes the information need of the user as the input and get some relevant documents as output. Currently, techniques for search are purely based on Keywords and no scope of searching based on concepts or meaning of the contents.

The Information Retrieval domain has moved from its core goals of Text Indexing and Searching for relevant documents in a repository due to the influence of the web [2]. The success of getting appropriate information to a query is influenced by the user function and also by the reasoning of the documents embraced by the Information Retrieval system. Information Retrieval can be useful for the development, implementation and evaluation of a search engine. Information Retrieval models contribute towards the development of the Information retrieval system. Boolean model, Vector model and Probabilistic Information Retrieval model are considered as the typical Information Retrieval models over the period of time [1]. Over the years alternative modeling paradigms for each type of the classical models have been proposed. Though all these models have led to the development of Information Retrieval systems, they have their

* Asst Professor, St Aloysius College and Research Scholar, Dept. of Computer Science, Bharathiar University.

** Asst Professor, Rajeswari Vedachalam College, Chengalpattu, Tamilnadu, India.

own limitations and issues that they suffer. Hence none of the Information Retrieval system developed so far can be considered as a perfect, errorless Information Retrieval system.

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 role of General search engines in searching relevant information from the web is inevitable. Apart from the widely used search engines 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. In recent days these search engines are also widely used because the relevancy of the results that these search engines produce are more and specific than the thousand of results the other general purpose search engines give. 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. For this study we used two popular vertical search Engines based on US called as Trulia and Zillow. Trulia is web search portal based in United States that helps home buyers, real estate professionals and sellers with information related to real estate industry.Their innovative search mechanisms helps to locate the homes and related information that is valuable for home buyers. Zillow is an online real estate Database Company. Zillow has data on 110 million homes across the United States. It gives basic information on houses, their locations, features, pricesetc. Zillow’s search engine allows people to input keyword and other specifications like number of bedrooms, minimum and maximum price and others.

Different query reformulation approaches that used Ontology for query reconstruction are explained in the next section.

3. LITERATURE REVIEW

Query reformulation is a task of transforming the original user query into another query by attaching relevant words to the actual query placed by the user. The actual query is transformed into another query with the intention of retrieving more relevant information which may not be possible otherwise. Many studies that were done in this regard were looking into the different aspects from where the words can be picked from, to be added as part of query reformulation.Some of the studies that are listed below focus on the user intention which also helps in reformulating the query.

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 vertical search engines.

4. MATERIALS AND METHODS

4.1. Query Reformulation methodology

Query Reformulation procedure takes the actual query given by the user as the input and considers every keyword present as the actual seed terms. The Query reformulation 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 are 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 Trulia and Zillow.

4.2. Selection of Queries

We selected few random queries, but for our study since we concentrate on the impact of word net over the queries. Our analysis does not categorize them into Transactional, navigational or Informational but just take them as they are. Some queries that we used in the experimental study are listed below.

5. RESULTS AND OUTCOMES

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.

Table 1
List of 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
6.	Types and kinds of flats available in California
7.	Homes and house in New York
8.	3 bedroom or room house or house needed in Chicago
9.	Flats or apartments available in Atlanta
10.	Houses or homes needed to keep pets or animals in New Jersey

Though we came across some Vertical search Engines, we did select Trulia and Zillow 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 Vertical Search Engine.

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

The queries were given to 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 Table 2, Table 3,

6. DISCUSSION

The Analysis of the results derived from the experimental study on two different vertical search Engines are plotted below.

Table 2
Actual Queries and their precision Values

Sl no	*Queries*	*Trulia*	*Zillow*
1.	Types of flats available in California	35	11
2.	Homes in New York	49	90
3.	3 bedroom house needed in Chicago	84	37
4.	Flats available in Westside, Atlanta	28	84
5.	Homes needed to keep pets in New Jersey	29	80

Table 3
Reformulated Queries and their precision values

Sl no	*Queries*	*Trulia*	*Zillow*
1.	Types and kinds of flats available in California	40	0
2.	Homes and house in New York	23	99
3.	3 bedroom or room house or house needed in Chicago	0	3
4.	Flats or apartments available in Atlanta	0	0
5.	Houses or homes needed to keep pets or animals in New Jersey	0	0

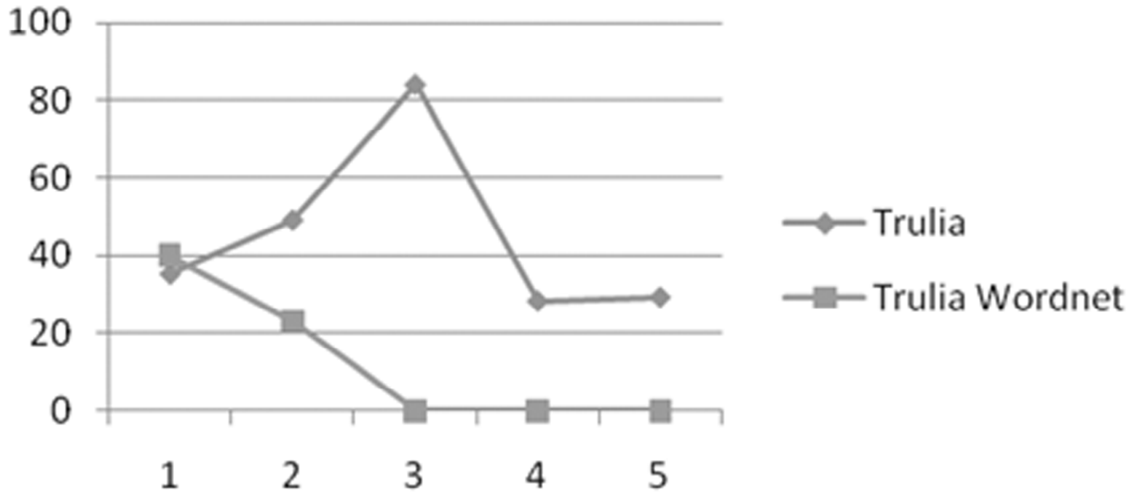

Figure 1: Trulia Precision values for queries

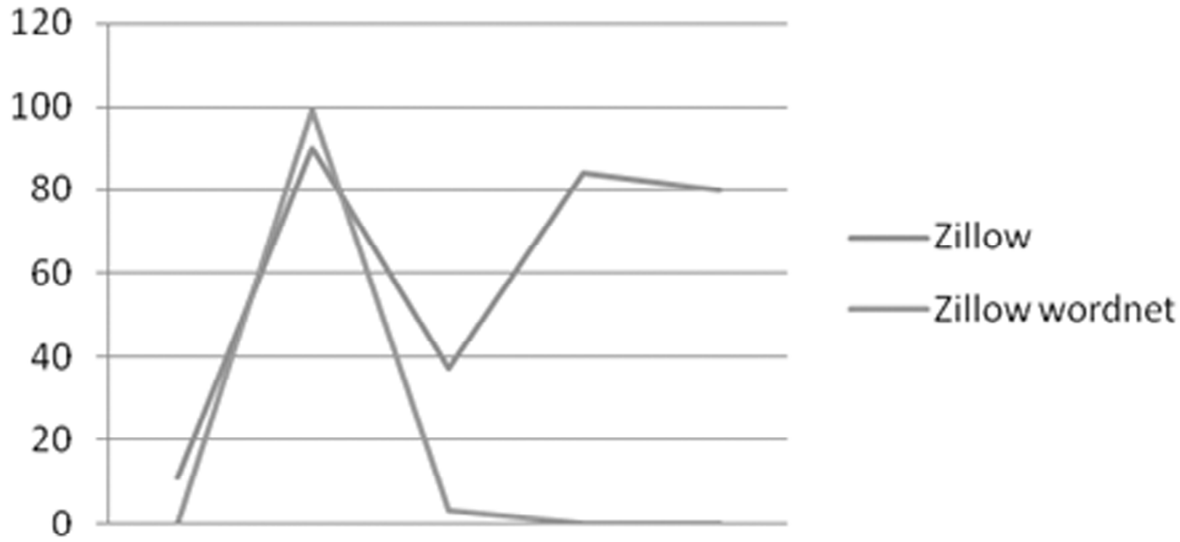

Figure 2: Zillow Precision values for queries

All the above figures (Fig. 1, Fig. 2) that represents the comparison of queries executed before and after reformulation clearly indicates that, the query reformulation using the domain independent ontology is not having a greater influence in the system performance or will have no influence in the system performance. Though the focus of our study was not to compare the performance the one search engine to another. But we observed, that the performance of these entire search engines towards executing the queries were similar.

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 queries, it is not having a greater influence in the system performance or will have no influence in the performance of the system.. The results we got was based on the time of execution of the queries, so on the basis of our experimental study results we conclude that the queries that are used in Vertical Search Engines may not benefit from using Wordnet for query reformulation because of the fact that the vertical search engines are much focussed on a particular domain or specialization.

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

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

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

[10] Carpineto, C., & Romano, G. (2012, January). “A survey of automatic query expansion in information retrieval”. ACM Comput Survey., 44(1), 1: 1-1:50.