International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2016, pp. 181-187
e-ISSN:2278-621X

A HYBRID APPROACH TO ENHANCE THE RELEVANCE OF SEARCH

Ruban S[1], Mohammed Arshad[2] and Shawn Merwin D'souza[3]

Abstract- Information Retrieval is a way of accessing related information from a collection of data. It can be based on text or content.The area of Information Retrieval has gone through lot of changes in the last two decades. None of us can deny the impact of search Engines in our day-to-day life. There has been lot of improvements in the way search engines work and process the information, but still there is a lot of challenges that has to be addressed. One such concern is about forming queries. Any Information retrieval system's performance depends on how a search query is given, how the information is stored and indexed.Any change in one component will have an impact in the performance of the system, which means if we can improve the quality of search query given or make use of efficient indexing and storage method, the efficiency of Information retrieval system increases. Inthis paper,we propose a new hybrid method of information retrievaltechnique thatexpands the query based on Ontology and Thesaurus. It uses an interactive method to refine query where the user entered search query is refined by adding more relevant words to its seed words. Our experiment proves that this hybrid method provides more accurate results than the traditional searching mechanism that is used.

Keywords – Information Retrieval, Ontology, Query Expansion, Retrieval process

I. INTRODUCTION

The area of Information Retrieval has gone through lot of changes in the last two decades. None of us can deny the impact of search Engines in our day-to-day life. There has been lot of improvements in the way search engines work and process the information, but still there isa lot of challenges that has to be addressed. One such concern is about forming queries, which is nothing, but the user information need specified in the query interface.Information Retrieval (IR) deals with the recovery of documents from a collection for a given user information need expressed with a Query. With enormous data emerging on the web, the process of searching and managing massive scale content have become increasingly challenging. This have led to the development of the IR models that seem to have an upper hand over the other with respect to performance, specifying the query, arranging the documents with regard to relevance and many other factors. The use of ontologies to overcome the limitations of the traditional keyword-based search has been put forward as one of the motivations of the semantic web. Since then, a lot of work began with the aim of getting the web to be a place that will facilitate a more meaningful search.

[1] *AIMIT, St Aloysius College, Mangalore.*
[2] *AIMIT, St Aloysius College, Mangalore.*
[3] *AIMIT, St Aloysius College, Mangalore.*

The term Information Retrieval refers to a search that may cover any form of information: structured data, text, videos, images, sound, musical scores, DNA sequences etc. For many years database system existed to search structured data and information retrieval meant the search of documents. Information Retrieval deals with the representation, storage, organization and access to Information.

The area of information Retrieval has grown well beyond its primary goals of indexing text and searching for useful documents in a collection due to the impact of World Wide Web. The effective retrieval of relevant information is directly affected both by the user task and by the logical view of the documents adopted by the retrieval system. The system assists users in finding the information that they need. A query is the formulation of a user Information need. In its simplest form, a query is composed of keywords and the documents containing such keywords are searched for. Query expansion techniques are applied with the aim of reformulating a seed query and improve retrieval. Query expansion technology appends related words to query and overcomes word mismatch and improves retrieval. So that through using query expansion, documents which are on the same matter can be found even if they do not contain the original query terms. Query Expansion is generally aimed to formulate a user query into one that is more responsive for Information retrieval.

Modifying the user query helps in improving information retrieval.There are many approaches to expand query where information retrieval system automatically refines initial query and user don't have any control over refining and expanding query. We propose an approach where we use WordNetOntology and another Thesaurus API(words.bighugelabs.com) to suggest related words and then user can choose from the list of provided words which will be then used in queryexpansion. We then compare our system's performance with the traditional retrieval systems using queries from TREC dataset.

The rest of the paper isorganized as follows.Section II talks about the process of Query Expansion and the related work is elaborated in section III. The proposed framework is discussed in section IV followed by the Experimental results and discussion and finally the concluding remarks.

II. Query Expansion

User written Query is not always efficient in retrieving information and one possible solution to overcome this is through Query expansion. It is also referred to as Query Refinement or Query Reformulation. Query Refinement is the process of transforming a user information need into another query that is more relevant for retrieving information. The Query expansion process comprises of the following stages:

A. Query Formation

User's information necessity is expressed as query. Query consists of keywords and these keywords are searched in the documents. Query can have more than one word and the result of this Query will be a set of documents which consists of this word. The most basic query consists of keywords which can help to retrieve information. Since most of the web users are naïve users, many a time the original user query is not efficient to retrieve relevant documents, so we include query expansion techniques to enhance the accuracy of Information Retrieval System.

B.Enhanced Query

Query Expansion aims to make Information Retrieval more efficient. From the earlier work done in this aspect it was found that query expansion meaningfully improves short queries and helps

in retrieving more relevant documents. Researchers are trying to improve Information Retrieval by Query Expansion.

C.Use of ontology for query expansion

Ontology can be considered as a vocabulary of terms or concepts that give a complete coverage about a specific domain or area of interest. It can be considered as a knowledge based on which the machine can understand and express the knowledge on a particular subject of a domain. Ontologies provide richer relationships between terms. In our context, the ontology can be seen as a set of terms, synonyms and relations between them which can also be used in Query Expansion.

III. RELATED WORK

There are various approaches quoted in the literature, which is explained in the related work section. Recently Researchers have started to use ontology to refine the query. Ontologies have been categorized as Domain independent ontology and Domain dependent ontology. One of the widely used domain independent ontology is WordNet which was developed by G.A. Miller in Princeton University [1].WordNet is a large lexical database for the English language. It groups English words into sets of synonyms called synsets.WordNet has been a popular general ontology used in the area of query expansion. WordNet can be considered as the combination of dictionary and thesaurus. WordNet's structure makes it a useful tool for computational linguistics and natural language processing. Usage of WordNet Ontology can be dated back to as early as 1990'sThough there has been lot of study on using the WordNet for Information Retrieval; Ellen M. Voorhees [2] in her study using WordNet for TREC Collection concluded that Less well developed queries can be significantly improved by refinement. One of the widely used query refinement method is called as Relevance Feedback techniques [3] which was proposed by Salton and Buckley, in which the terms featuring prominently in documents marked relevant by the user are automatically added to the query.

Later Srinivasan came up with a Retrieval Feedback technique [4] that adds terms from the top relevant documents to the query. This technique has shown considerable improvement in many retrieval tasks. Query logs was used as a means of query expansion by Hangs et al [5]. Later Huang et al [6] proposed a query expansion algorithm of pseudo relevance feedback based on matrix-weighted association rule mining.

However, in the year 2001 Aronson [7] proved that query refinement that is based on ontology is much more efficient than the other methods that were available. Using ontology for query expansion goes back to 1994 where Voorhees [2] attempted using the Domain independent ontology WordNet for query expansion. Since then there has been some works done in this area. The word sense information and the ontology was used for query expansion by Navigli and Velardi [8]. They succeeded in using ontology to extract the semantic domain of a word and then the query is expanded further using co-occurring words. Further Query refinement techniques based on domain and geographical ontology was studied by Fu,G et al [9]. The Domain ontology was modeled after tourism which consists of some non-spatial terms such as "near" whereas the geographical ontology consists of some spatial terms such as place names. A domain specific ontology based on Stockholm University Information systems(SUiS) was developed by Nilsson et al[10].

An interactive query expansion methodology was proposed by Ruban et al [11]. The work was done with a domain ontology and it shows a small improvement in the precision. In the current information retrieval method, the query used by the searchers is not relevant to how information

or documents has been indexed. A possible solution for this is Query Expansion where the user can select words which have similar meaning or have some relation with the query terms.

IV. PROPOSED FRAMEWORK

In our work we integrated WordNet Ontology and BigHugeLabsAPI(Thesaurus) for Query Expansion. Initially the user enters the query, which then undergoes stemming and stop word removal process and then this query is passed to WordNet Ontology and BigHugeLabs API which suggests similar words or synonyms or any related words to the user which the user can then select. Using this word, the Original Query entered by user is then expanded and passed to Google API and results are obtained.

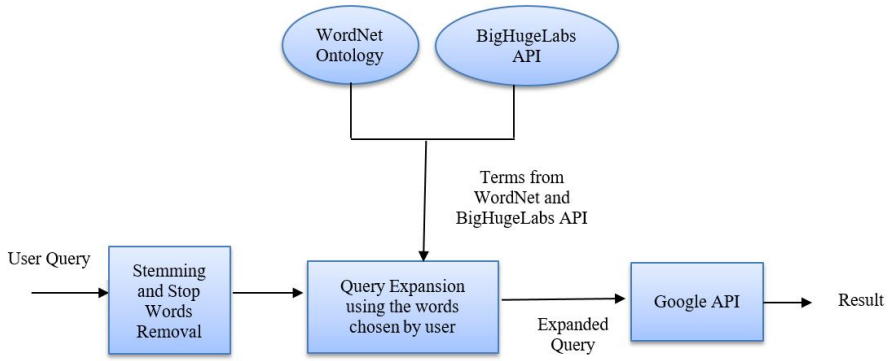

Figure 1. The Hybrid model

V. EXPERIMENTAL RESULTS AND DISCUSSION

In our experiment, we passed the original user Query to our system and expandedit. For example,suppose the user enters the query as "altitude sickness" this original query is passed to the WordNetOntology and BigHugeLabs API and all the relevant terms are shown to the user - "altitude"," height", "elevation"," EL"," ALT", "illness", "malady", "sickness", "nausea". In the suggested terms that are displayed,the user can select a term. The selected term is then added to the query and the query is expanded, suppose the user has selected the word "illness", then the refined query becomes as "altitude sickness or illness". Then we passed the Expanded Query to relevant search engines. In our work we used the Google Search Engine and Bing Search Engine. We also passed the Original Query to these search engines (i.e., Google Search Engine and Bing Search Engine) Then we compared the results. We checked the Efficiency of the new system by comparing the relevance of the search results.

The following table illustrates the list of queries that has been used in our experiment. The queries were selected from the web query track of the well-known TREC Data Set and is used for our Evaluation. We calculated the precision for our evaluation. The first hundred results were checked for relevance and was done manually.

Table 1 : Queries and refined queries with their precision values

Query No	Original Query	Expanded Query	Precision Rate for	
			Google	Bing

			Original Query	Expanded Query	Original Query	Expanded Query
1	identifying spider bites	identifying spider bites or spider	72	78	73	76
2	history of orcas island	history of orcas island or orca	61	73	59	70
3	hip roof	hip roof or roof	99	94	85	80
4	the american revolutionary	the american revolutionary or patriot	98	99	97	98
5	sun tzu	sun tzu or sun zi	96	96	95	95
6	benefits of running	benefits of running or benefit	77	80	87	88
7	folk remedies sore throat	folk remedies sore throat or tribe	85	87	83	86
8	balding cure	balding cure or remedy	94	98	95	96
9	altitude sickness	altitude sickness or height	96	99	96	97
10	how to tie a windsor knot	how to tie a windsor knot or tie	95	96	93	98
11	view my internet history	view my internet history or see	72	81	68	73
12	fidel castro	fidel castro or fidel castro ruz	98	98	97	98
13	benefits of yoga	benefits of yoga or benefit	95	98	96	99
14	norway spruce	norway spruce or spruce	93	98	97	96
15	hayrides in pa	hayrides in pa or pennsylvania	90	92	88	94

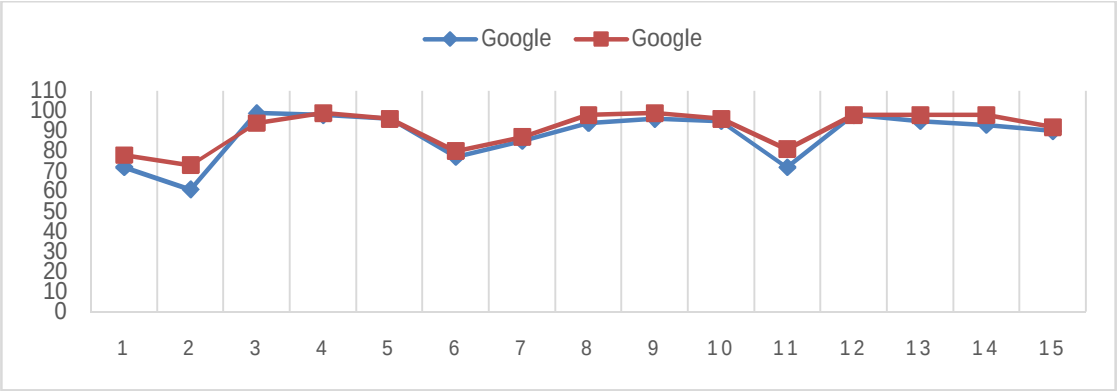

Figure 2. The above figure give a pictorial representation of our experimental results in google.

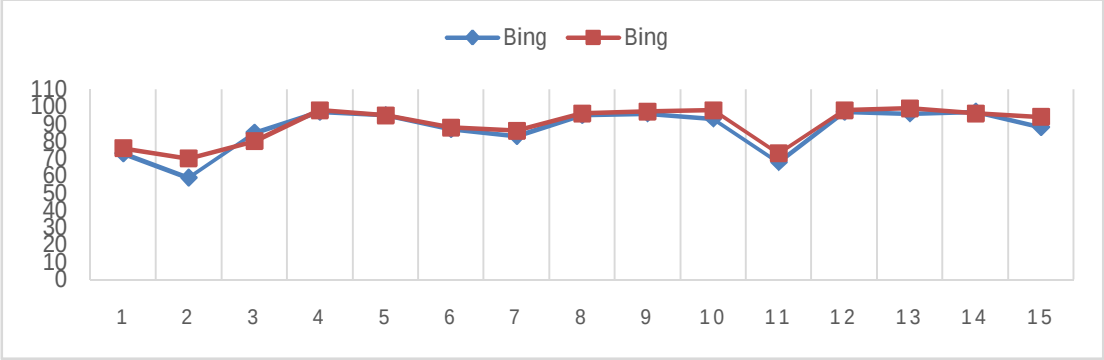

Figure 3. The above figure give a pictorial representation of our experimental results in bing.

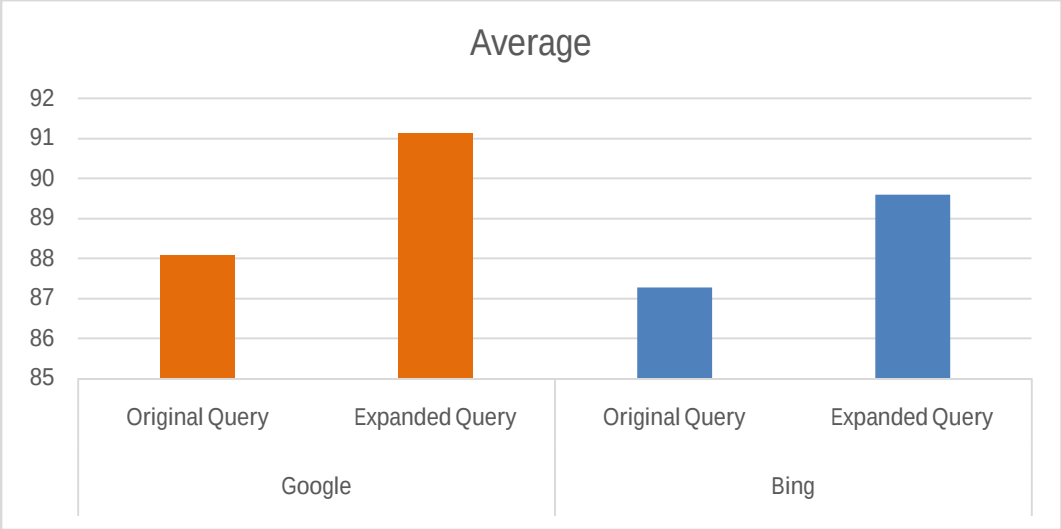

Figure 4. The above figure give a pictorial representation of average result our experiment..

VI.CONCLUSION

The Experimental results may vary according to the queries given and the time of execution of our queries. Our system compares the efficiency of the Information Retrieval system based on the Precision Rate of original user query and expanded query. From the experimental results,we can infer that the precision rate for expanded query generated by our model is more than the original query given by the user. The Precision Rate differs for different terms selected for Query Expansion by user. The system where user interactively selects terms for query expansion is more efficient than the traditional system where the user query is directly fed into the search engine.

REFERENCES

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

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

[3] Salton G & Buckley C (1990). Improving Retrieval performance by relevance feedback, Journal of the American society for Information Science.

[4] PadminiSrinivasan, "Retrieval Feedback in MEDLINE", Journal of the American Medical Informatics Association, 3(2):157-167, 1996c, doi: 10.1136/jamia.1996.96236284

[5] C.Hang, W. Ji-Rong and N. Jian-Yun, "Probabilistic query expansion using query logs", proceedings of the eleventh international conference on World wide Web(2002)

[6] M.Huang, x.Yan and S.Zhang, "Query expansion of pseudo Relevance feedback based on matrix-weighted association rules mining", Journal of Software , 20(7):1854-1865 (2009)

[7] Alan R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program." Proceedings of AMIA, Annual Symposium, pages 17-21, 2001.

[8] R.Navigli and P.Velardi, "An analysis of ontology-based query expansion strategies", workshop on Adaptive Text Extraction and Mining(2003).

[9] Lin Fu,Dion Hoe-LianGoh and Schubert shouo-Boon Foo, "Evaluating the effectiveness of a collaborative querying environment", proceedings of the 8th international conference on Asian digital libraries [2005].

[10] K.Nilsson, H.Hjelm and H.Oxhammar, "SuiS – cross-language ontology-driven information retrieval in a restricted domain", Proceedings of the 15th NODALIDA conference(2005).

[11] S.Ruban, Vanitha T, Behin Sam S, "Design and implementation of an interactive query Expansion methodology for information retrieval",vol. 11, no. 2, January 2016 Issn 1819-6608, ARPN Journal of Engineering and Applied Sciences.