VOL. 10, NO. 7, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

# AN EXPERIMENTAL ANALYSIS AND IMPLEMENTATION OF ONTOLOGY BASED QUERY EXPANSION

S. Ruban[1] and S. Behin Sam[2]
[1]Department of Computer Science, Bharathiar University, Coimbatore, India
[2]Government Arts College, Chengalpattu, India
Email: rub2kin@yahoo.com

**ABSTRACT**

Search engines history goes back to the field of Information Retrieval, that has gone through a tremendous change over the period of time. Though search Engines have experienced many enhancements in the last few years, the query processing Techniques that the Information Retrieval (IR) Technology relies on are still based on Keywords. It is difficult to formulate queries which are well designed for retrieval purpose. Query Expansion can solve this problem. Query Expansion is normally aimed to form a query into one that is more responsive for Information Retrieval. Though there are many approaches aimed at query expansion, ontology based query expansion has been found to have more advantages compared to the traditional ones. This paper compares the performance of the Traditional query processing methodology with the Domain Independent Ontology based query expansion Methodology.

**Keywords:** information retrieval, semantic web, ontology, retrieval process, query expansion.

## INTRODUCTION

Search engines history goes back to the field of Information Retrieval, that has gone through a tremendous change over the period of time. Though search Engines have experienced many enhancements in the last few years, the query processing Techniques that the Information Retrieval (IR) Technology relies on are still based on Keywords. It is difficult to formulate queries which are well designed for retrieval purpose.

An Information Retrieval system helps to find information that will satisfy the user information need expressed using the user's query[1]. It deals with the recovery of documents from a collection, for a given user information need which is 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 has 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.

Information Retrieval has grown from its initial purpose of indexing text and looking for relevant documents in a collection due to the advent of World Wide Web. The ability to retrieve relevant information is affected by the way the user query is handled, as well as the way by which the documents are viewed by the retrieval system. The retrieval system helps the users to locate the information they are seeking for. It will never return the information that the user seeks for but it helps him to identify the documents that may contain such relevant information that he may be seeking for. The documents that satisfy the information need laid down by the user are called as relevant. If there is any retrieval system that is completely perfect, then it should retrieve all relevant documents, but there is no such system that is capable of retrieving only relevant documents because the user's information need that is expressed as a query is not complete most of the time. Since most of the users are novice user's this cannot be avoided completely.

Though Search Engines have experienced many enhancements in the last few years, the query processing techniques that the Information Retrieval (IR) Technology relies on are still based on keywords and hence the capabilities to predict the conceptualizations in user query is very limited [2].

A brief summary about the information retrieval process is described in the next section which will be followed by the short description about query expansion, later some of the related work in this area, followed by the methodologies adopted, then the experimental results and finally the conclusion.

## RETRIEVAL PROCESS

Any Information Retrieval system is supported by the Retrieval process which involves three basic processes, which are as follows:

- The representation of the document contents.
- The representation of the user's information need.
- The comparison between the above two.

The components are displayed in Figure-1. Information Retrieval can be helpful for the development, implementation and evaluation of a search engine. The document representation of the documents is usually done using the indexing process. Forming a query is the way of representing the information need specified by the user in the specified place in the retrieval system. Matching process involves the similarity comparison between the query and the document representation.

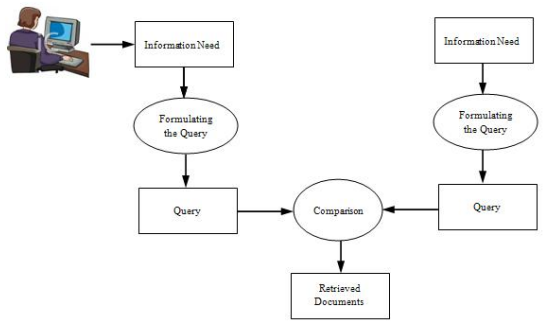


**Figure-1.** Information retrieval process.

The concept of Relevance [3] between the terms in the query and the document representation plays a vital role in Information Retrieval process. Hence representing the query terms and the document will significantly impact the Retrieval.

Retrieval strategy normally refers the information retrieval framework [4] Retrieval strategies assign a measure of similarity between a query and a document. Any Information Retrieval system is based on Information Retrieval process.

## QUERY EXPANSION

The query expansion process involves the following sub stages and each of them is explained below.

### a) Formulation of the query

A query is the formulation of a user Information need. Basically it can be considered as a composition of keywords that the user is looking for and the documents containing those keywords are being searched for by the search application. It can have a single word or many words based on some combination[5]. The most basic query that can be formed in a basic retrieval system is a word. The expected result of these queries will be the set of documents that may contain atleast one of the words that is available in the set of documents.

A very basic method of forming queries is to use keywords in the place of the natural language which will be equivalent to the information need expressed by the user [8]. In Information Retrieval, the user's input queries usually are not detailed enough to allow fully satisfactory results to be returned. Since forming queries with the correct words are hard for many users, query expansion techniques are widely used to enhance the accuracy of the Information Retrieval system and hence can solve this problem.

### b) An optimized representation of the user information need

Query Expansion is generally aimed to formulate a user query (user information need) into one that is more responsive for Information Retrieval. Earlier findings [6] have showed that expanding the query in detailed queries had little improvement, but it also demonstrated great improvement or significantly improved given short queries. Since then there has been lot of experiments that is carried on query expansion by the researchers in the IR community.

### c) Ontology usage to enhance and represent the User Information Need

With the advent of Semantic web, Ontology has gathered more momentum and is being used as a means to formulate the user query, thus using it to improve semantic capability. Ontology can be used to share information between systems. Though there are many definitions for ontology, it 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 said as a knowledge base which is machine understandable expressing the knowledge on a particular subject of a domain. The extensional knowledge and the intensional knowledge of a domain constitute a domain ontology. Ontologies also provides richer relationships between terms. It is these rich relationships that enable the expression of domain-specific knowledge. In our context, the ontology can be seen as a set of terms and relations between them, denoting the concepts that are domain specific.

## RELATED WORK

In Literature many query expansion approaches are proposed and each of them have their own benefits and limitations. With query expansion, the user is guided to formulate queries which enable useful results to be obtained [5].

For instance Relevance Feedback [9] offers many benefits for searching small databases. Relevance feedback would not be helpful if the user information need is misspelled, or if the query is based or concerned about a general concept.

Rinaldi A.M [7] in his work also depicted the advantages of using Ontology towards Information Retrieval.

Retrieval Feedback [10] adds terms from the top relevant documents to the query. This approach has shown improvement in many Information Retrieval tasks.. But later Aronson and Rindflesch[11] proved that Ontology based Query Expansion is a more effective and favourable method than Relevance Feedback. Since then lot of work is being done on using ontology to expand the queries thus aiming towards the optimized representation of the user information need.

There were some work carried out by some researchers where they tried using Ontologies for query expansion, but the attempts have not been been very successful. Word Net has been a popular general ontology used in the area of query expansion.

Navigli and Velardi [12] use sense information and ontologies for query expansion. Their contribution also involves the impact in the retrieval performance when expanding with synonyms etc.

Baziz *et al*. [13] state ontology based Information Retrieval is promising, in increasing the quality of

responses since document semantics are captured. Here the WordNet concepts are extracted and then globally disambiguated with reference to document terms to produce the optimum semantic network.

Nilsson, K. *et al* [14] also used ontology in his work where he showed improvement in the retrieval performance.

## METHODS ADOPTED

The Following diagram, Figure-2 represents the proposed methodology that we have adopted for query expansion. It consists of two different components i) Query formulation and the other one ii) Query Expansion.

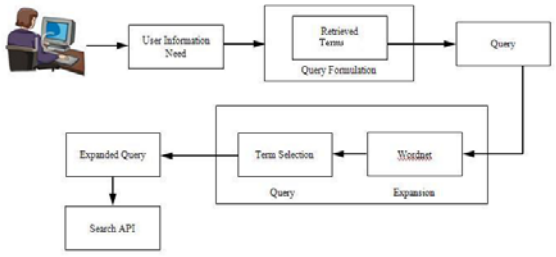


**Figure-2.** Proposed methodology for query expansion.

In this section the details about the methodologies that we have adopted to evaluate the performance of the ontology based query expansion method is explained. The query given by the user is parsed by the parser, in our case we used Stanford parser. The parsing is done to analyze the query syntactically which determines the part of speech of every word in the query, then the output is passed to the ontology. In the first case we used the Google API, ie the query was given to the search API directly, whereas in the next case the same query was refined further by adding terms from the ontology. We had used Jena API to do this. The refined queries were further passed to the search interface. In our case we used Google search API to achieve this.

## EXPERIMENTAL RESULTS

For our experimental evaluation, we selected 10 random queries but domain specific queries with related to sports, and evaluated separately under two circumstances and the first 100 retrieved links were verified for relevance. The experiment was conducted during the time interval of 3 months starting from May 2014 to July 2014.

- Queries directly given to the Google API.
- Queries that went through Domain independent Ontology based Query Expansion(Using Wordnet API)

Our Experiments reveal that the queries that were refined using the Domain Independent ontology gave more accurate results than that of the query that was given directly to the Search API. So we conclude that performance of any search engine will increase by using Ontology based Query Expansion.

**Table-1.** Queries Vs refined queries using independent ontology.

| Sample queries | Refined queries using domain independent ontology (Wordnet API) |
|---|---|
| Players of judo | Players of judo or Participants |
| Types of sports | Types of sports or Athletics |
| Karate fighter | Karate fighter or combatant |
| Type of Individual sports | Type or Types of Individual sports |
| Kinds of team sports | Kinds of team sports or sport |
| Who are competitors | Who are competitors or challengers |
| Who are winners of relay race | Who are the winners of relay race or achiever |
| Football team | Football team or squad |
| Winner or race walking | Winner of race walking victor |
| Type of motor sports | Types of motor sports or character |

The following graphs represent the comparison between the queries processed in the traditional way and queries processed using the general ontology.

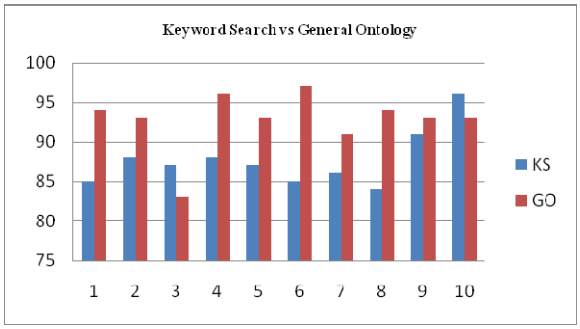


**Figure-3.** Queries Vs refined queries using independent ontology.

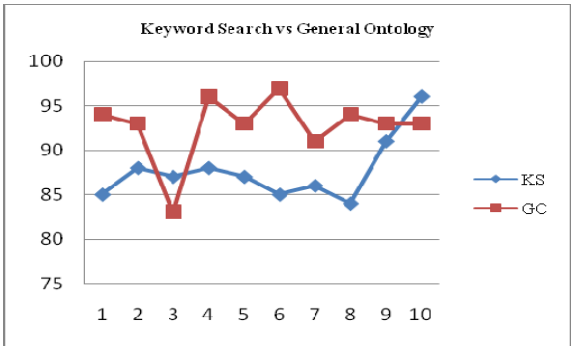


**Figure-4.** Queries Vs refined queries using independent ontology.

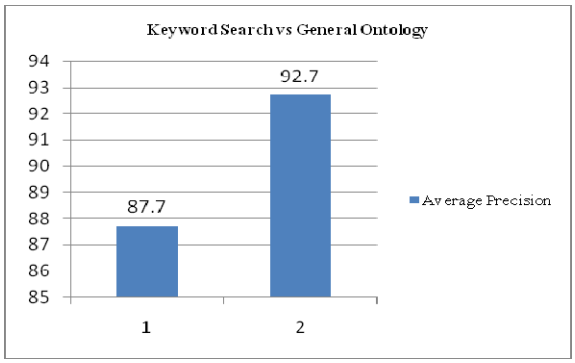


**Figure-5.** Average precision.

## CONCLUSIONS

The calculations done in this experiment may vary according to the domain and time of execution. The figure mentioned above depicts the average precision value of the Traditional way of executing the query, refining the query using Domain independent ontology. The average values of the result are shown above. From this we can infer that the higher value of average precision is for the system using query expansion based on Domain independent ontology. This system offers a better performance related to the accuracy in retrieving the results than the generic search engines. Though the Ontology based Query expansion has advantages over the classic way of handling the query, coverage of information about the domain, in the ontology holds the key. We plan to take further this work by developing a sports ontology and will compare the performance of Query Refinement using the Query Independent ontology and the Query Refinement using the Domain specific ontology.

## REFERENCES

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

[2] Jiewn Wu, IhabIlyas and Grant Weddell. 2011. "A Study of Ontology-based Query Expansion," Technical Report CS-2011-04, Cheriton School of Computer Science, University of Waterloo.

[3] Mark Sanderson and W. Bruce Croft. 2012. "The History of Information Retrieval Research", Proceedings of the IEEE, Vol. 100, May 13th 2012, doi:10.1109/JPROC.2012.2189916.

[4] Miriam Fernandez *et al.* 2011. "Semantically enhanced Information Retrieval: An ontology-based approach", Journal of Web Semantics: Science, Services and Agents on the World Wide Web. doi:10.1016/j.websem.2010.11.003.

[5] J. Bhogal, A. Macfarlane, and P. Smith. 2007. "A review of ontology based query expansion", Information Processing and Management, 43(4):866-886. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2006.09.003.

[6] Voorhees E. 1994. "Query expansion using lexical-semantic relations", In Proceedings of the 17th annual international ACM SIGIR conference on research and development in information retrieval table of contents. pp. 61–69.

[7] Rinaldi A. M. 2009. "An ontology-driven Approach for Semantic Information Retrieval on the web" ACM Trans. Internet Technol, 9, 3, Article 10, July. Doi:10.1145/15552291.1552293.

[8] David A. Grossman and Ophir Frieder. Information Retrieval Algorithms and Heuristics. Second Edition, Springer International Edition, ISBN 978-81-8128-917-9.

[9] Gerard Salton and Chris Buckley. 1990. "Improving Retrieval performance by relevance feedback", Journal of the American society for Information science, JASIS. Vol. 41, No. 4, pp. 288-297.

[10] Padmini Srinivasan. 1996. "Retrieval Feedback in MEDLINE", Journal of the American Medical Informatics Association, Vol. 3, pp. 157-167. doi: 10.1136/jamia.1996.96236284.

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

[12] Navigli, R., &Velardi, P. " An analysis of ontology-based query expansion strategies workshop on adaptive text extraction and mining (ATEM 2003)". In 14th European conference on machine learning (ECML 2003), September 22–26.

[13] Baziz, M. *et al.* 2005. "Conceptual indexing based on document content representation information context: nature, impact, and role. In: 5th International conference on conceptions of library and information sciences, CoLIS June 4–8 p. 171.

[14] Nilsson, K. *et al.* 2005. "SUiS – cross-language ontology-driven information retrieval in a restricted domain". In: Proceedings of the 15th NODALIDA conference.