VOL. 11, NO. 2, JANUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

DESIGN AND IMPLEMENTATION OF AN INTERACTIVE QUERY EXPANSION METHODOLOGY FOR INFORMATION RETRIEVAL

S. Ruban[1], Vanitha T[2]. and S. Behin Sam[3]
[1]Department of Computer Science, Bharathiar University, Coimbatore, India
[2]Department of MCA, AIMIT, St Aloysius College, Mangalore, India
[3]Department of Computer Science, RajeswariVedachalam Government Arts College, Chengalpattu, India
E-Mail: rub2kin@yahoo.com

ABSTRACT

Information Retrieval is a field of computer science that has seen a tremendous change in the past two decades. After the advent of World Wide Web, access to information became handy because of search engines. An Information Retrieval system consists of the following components i) Query interface where the use poses the information need ii) Index file which is created by the indexing process and the iii) Matching process which finds out the relevant documents from the available index file. The performance of the Information Retrieval system always depends on the above three components. So the improvement in one component can have a significant impact in the overall retrieval performance. In many of the query expansion approaches that are proposed in the literature, the system automatically selects the terms which will be then, be added to the initial query, whereas the user has no control in the query expansion process. Here we propose na interactive query expansion approach where the user gets the help from the system to select the terms that has to be added to the initial query. Our experiments conducted in this regard has also revealed that all the terms generated by the system may not be relevant for expansion but however the user decides which terms to be added and which may not be considered. In this experiment we intend to use a domain Ontology for interactive query expansion and we compare the performance of our system with the traditional one.

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

INTRODUCTION

Information Retrieval is a field of computer science that has seen a tremendous change in the past two decades. After the advent of World Wide Web, access to information became handy because of search engines. An Information Retrieval system consists of the following components i) Query interface where the use poses the information need ii) Index file which is created by the indexing process and the iii) Matching process which finds out the relevant documents from the available index file. The performance of the Information Retrieval system always depends on the above three components. So the improvement in one component can have a significant impact in the overall retrieval performance. In this context many studies have been done to analyze the way how query has been represented and handled, and it is been found that the way information need is presented has a significant impact in the retrieval performance. Hence Query expansion or query refinement or Query Improvement has become a vibrant field of Research in the domain IR.

One of the important reason behind query expansion is the word mismatch which means that concepts or keywords of user queries are often different from the words of the resource documents although these words have similar meanings [1], another reason being the vagueness of natural language where synonyms, homographs and inflection of words can all fool algorithms which see search terms only as a sequence of characters [2]. Query Expansion generally thought of as a recall-based technique is aimed to automatically formulate a user query into one that is more amenable for Information retrieval [3].Earlier research in this area has already showed that query expansion had limited retrieval improvement on detailed or complete queries, but it demonstrated great potential for significantly improving results for brief and short queries [4].

In many of the query expansion approaches that are proposed in the literature, the system automatically selects the terms which will then, be added to the initial query, whereas the user has no control in the query expansion process. Here we propose a interactive query expansion approach where the user gets the help from the system to select the terms that has to be added to the initial query. Our experiments conducted in this regard has also revealed that all the terms generated by the system may not be relevant for expansion, but however the user decides which terms to be added and which may not be considered. Since we intend to use Ontology for query expansion, we also have this notion that the user who gives the query is also familiar with the professional domain knowledge on which the query is based. Our experiment proves this point as well. Through human-machine interaction, undesired terms will not be added to the query string [5].In interactive query expansion users are displayed a list of terms suggested by the system after entering the initial query. Our approach is on the belief that not every suggested term is good for query refinement. Through human-machine interaction, undesired terms will not be added to the query string. Our overall experimental aim is to study the use of ontology to support interactive information retrieval. The specific objectives are to:

- Evaluate the retrieval effectiveness of ontology for selecting relevant documents in a retrieved document set.
- Evaluate the retrieval effectiveness of ontology to assist users in selecting candidate terms for interactive query expansion.
- Assess how ontology can bridge the gap between the query space and document space using interactive query expansion.

We will be covering the Existing work in the next section then our experimental methodology and its detailed design and description. Our experimental results and analysis will then be presented. We will be concluding then with our overall findings.

RELATED WORK

Literature Query expansion are studied in different ways for instance query expansion methods are classified into Automatic Query Expansion (AQE) methods and Interactive Query Expansion Methods (IQE) [6] Earlier studies reveal that many of the automatic query expansion methods rely on the Relevance Feedback techniques [7] 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 [8] that adds terms from the top relevant documents to the query. This technique has shown considerable improvement in many retrieval tasks. Query logs were used as a means of query expansion by Hangs *et al* [9]. Later Huang *et al* [10] proposed a query expansion algorithm of pseudo relevance feedback based on matrix-weighted association rule mining.

However in the year 2001 Aronson [11] 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 [12] 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[13]. 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* [14]. 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*[15].

But many of these methods are Automatic query expansion methods, but we focus on a Interactive query expansion method where after the user gives the query, Initially the relevant terms from the Domain independent ontology WordNet is displayed to the user, and the option is given to the user to select more terms which he thinks is more relevant, once the refinement is done then the Refined query is passed to the Domain dependent ontology say in our case the plants ontology that we have developed, the IR system then lists the different options that is available from the Ontology and user picks up the terms that he considers as relevant. After concatenating the terms both from the Domain Independent ontology and the Domain dependent ontology, the refined query is then passed on to the search API to search the web. Our experiment reveals that this optimized representation of the user information need will result in retrieving more relevant results than giving the queries directly.

EXPERIMENT METHODOLOGY

Framework of our Proposed System

The framework of our Interactive Information Retrieval system is given below. It consists of two important modules i) Query Recommendation module and ii) Query modification module. They both are responsible for the Query refinement process that forms the core process of this system. Given an initial query, we first pass the query to the Query Recommendation module which will recommend some relevant terms from the Query independent ontology to the user. After the user, selects the terms which are more preferable for refinement, the system will refine the query accordingly.

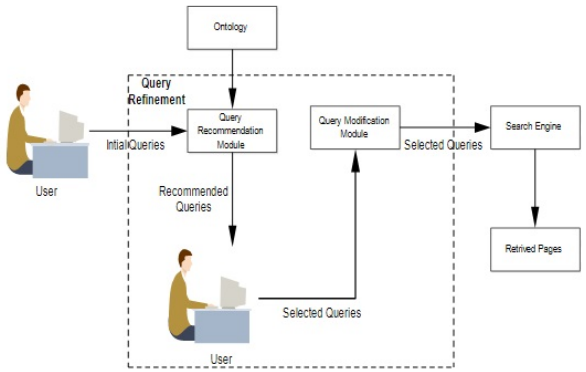

Figure-1. Framework of interactive information retrieval system.

Query Refinement

Query Refinement forms the core process of this System. It consists of Query Recommendation module and the Query Modification module. The following figure depicts the Query Refinement process.

As shown in the following figure, initial query is passed to the Domain Independent ontology in our case, we use wordnet. For eg suppose the user wants to know about “Palm Plants” and he gives the word in the user Interface, this initial query is passed to the Wordnet and all the relevant terms are shown to the user such as “Palm ”, Thenar”, “Palm tree”,”decoration”,”laurel wealth”, “medal”, “medallion”, “ribbon”, “handle”, “tree”, “tree diagram”, “Sir Herbert Beerbohm Tree”, “comer”,

"Shoetree". Among the above terms that is displayed , user makes a decision to select some terms, based on the interaction the terms that are selected are also added to the query now, suppose the user has selected the word "palm plants", then the refined query becomes as "Palm plants or palm trees".In our proposed framework the refined query is now passed to the domain specific ontology that is developed as part of this work, i.e the plants ontology, it displays two more options from the developed ontology say "Edible palms" and "palmstubs". Suppose the user after interaction with the system decides to select "Edible palms", then the refined new query will be "palm plants or palm trees or Edible palms". Instead of the search API processing the query as "Palm plants" the Search API will now process the query as "Palm plants or palm tress or Edible plants".

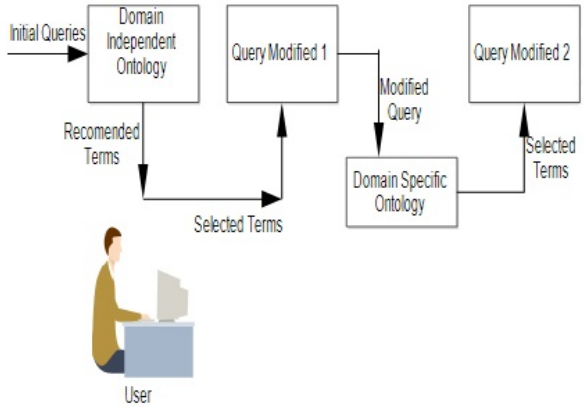

Figure-2. Query refinement process.

The greatest advantage about this approach is that the user has all the liberty to pick the terms that will be added to the query. Though it may little time before the query results be obtained, it is worth because the user information need well represented is better than the user information need that lacks the essential terms.

ONTOLOGY CONSTRUCTION

Ontology is defined as a formal, explicit specification of a shared conceptualization [16]. After the advent of semantic web, ontology has gained momentum and popularity among the research community. Ontologies may well capture domain knowledge if the ontology engineers have a solid understanding of the domain, because the quality of domain ontologies relies heavily on a correct representation of the domain knowledge, either by way of manual construction or by some automated tools [3]. The process of ontology construction and the steps involved in it are explained below.

Deciding the Scope of the Ontology

This phase helps to understand the usage of the ontology and the purpose behind creating it. Before creating the ontology the purpose should be clear and concise. In our experiment we use ontology to provide more relevant terms to refine the query that is coined by the user. The relevant terms are displayed to the user and he selects the terms that he thinks will give more relevant results.

Information Collection for Domain Ontology

There are two types of ontology domain independent ontology and Domain specific ontology. We have built a domain specific ontology which contains information about Domesticated Plants. The information was collected from domain experts, books on domesticated plants and some relevant websites. This phase is a precondition for building the domain specific ontology.

Class Definition

Currently there are three class design methods. A top-down development process which starts with the classification of the most general concepts in the domain and consequent specialization of the concepts, a bottom-up development process starts with the classification of the most specific classes, and grouping of these classes into more general concepts and a combination development process is a blend of the top-down and bottom-up approaches.[16] In this paper, a top-down approach was applied to construct plants ontology.

Defining the Attributes

The classes alone will not provide enough information to answer the competency questions. Once we have defined some of the classes, we must describe the internal structure of concepts.[16]. Initially, it is important to get a comprehensive list of terms between concepts they represent, relations among the terms, or any property that the concepts may have. These terms contain object properties and data type properties. All the sub-classes inherit the properties of the classes.

Instance Creation

The final phase is to create individual instances of the class. Defining an individual instance of a class requires 1) choosing a class, 2) creating an individual instance of that class, and 3) filling in the slot values. [16]

EXPERIMENTAL EVALUATION

Normally in our web search, many users pay attention only for the first few results. Therefore in this study we consider the top n results that are retrieved, and we checked them manually for relevance. We use precision to evaluate the performance of this system and n is set to 100.

$$\text{Precision} = \frac{\text{Relevant documents in the top 100 retrieved}}{100}$$

First we compared the performance of the system where user query is directly taken for processing without user interaction and the other one where the user query is

refined using the Query recommendation and Query modification module.

EXPERIMENTAL RESULTS

For our experimental evaluation, we selected 15 random queries but domain specific queries with related to plants, 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 September 2014 to December 2014

Case i. Queries directly given to the Search API.

Caseii. Queries that went through User Interaction.

The following table lists out the 15 different queries that were used for the experimental study. These queries were randomly generated and were used to compare the performance of the system when it is traditionally executed and when the query is enhanced using interactive query expansion

Table-1. Random queries used in the experimental study.

Ssl no	Sample Queries	Search API	Refined Queries	With Enhanced Query
		Precision		Precision
11	Palm Trees	0.86	Palm Plants or Palm Trees or Edible plants	0.89
12	Usage of herbs	0.82	Uses or Usage of herbs or herbal tea	0.81
23	Kinds of Fruit Trees	0.77	Variety or Kinds of Fruit Trees or Nut Trees	0.88
34	Most popular crop grown by farmers	0.77	Craw or Most popular crop grown by farmers or shrubs	0.88
55	Examples of ornamental Trees	0.89	Good example or Examples of ornamental Trees or Palms	0.93
66	Types of garden Pants	0.88	Type or Types of garden Pants or House Plants	0.96
67	Pesticides used for crops	0.94	Veggie or Pesticides used for crops or coffee	0.99
88	uses for Medicinal Plants	0.73	Usage or uses for Medicinal Plants or Shurbs	0.79
99	List of Cooking Plants	0.90	Name or List of Cooking Plants or Cereals	0.96
110	Fast growing vegetables	0.88	Rise or Fast growing vegetables or Vegetables oil	0.86
111	Some of Medicinal Plant	0.65	Variety or Some of Medicinal Plant or climber	0.71
112	Types of Grasses	0.96	Type or Types of Grasses or cyperus	0.95
113	Types of Hybrid Fruits	0.85	Type or Types of Hybrid Fruits or Citrus Hybrid	0.96
114	Which plants are used for vines	0.83	Plant or Which plants are used for vines or Hedera	0.85
115	vines plants example	0.85	illustration or vines plants example or Mandevilla	0.86

The following figures give a pictorial representation of the difference between the traditional way of processing a query and enhancing query using Interactive Query Expansion.

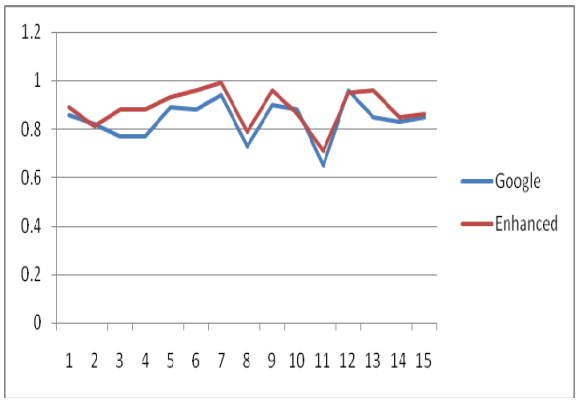

Figure-3. Search API vs after interactive query expansion.

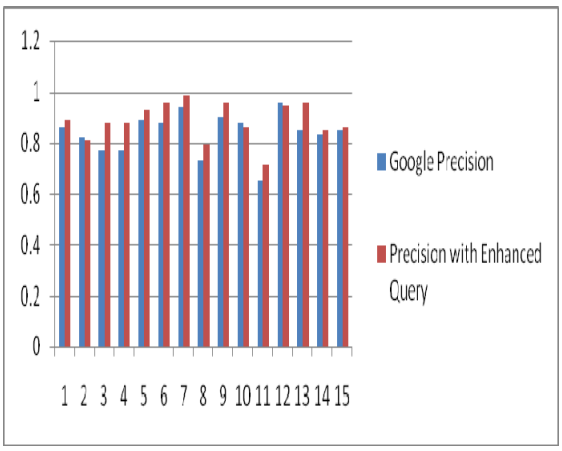

Figure-4. Precision of search API vs after interactive query expansion.

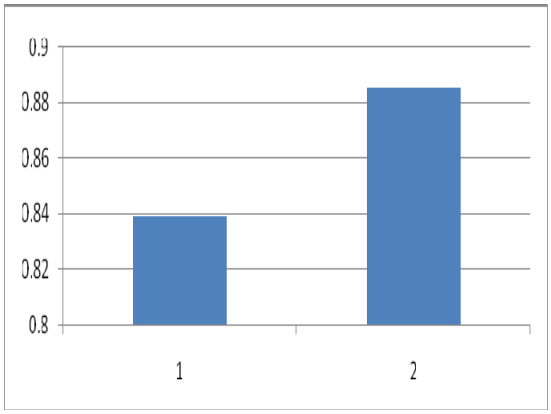

Figure-5. Average precision.

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 user interaction. 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 User Interaction. This system offers a better performance related to the accuracy in retrieving the results than the generic search engines. The performance of the system also changes depending upon the options selected by every user.

REFERENCES

[1] Design and Implementation of Ontology-Based Query Expansion for Information Retrieval, Fang Wu, Guoshi Wu and Xiangling Fu.

[2] Improving information retrieval effectiveness by using domain knowledge stored in ontologies, Gabor Nagypal.

[3] A study of Ontology-based Query Expansion, Jiewen Wu, Ihab Ilyas , Grant Weddel.

[4] Query Expansion using lexical-semantic relations Ellen M.Voorhees.

[5] Interactive Query Expansion Based on Fuzzy Association Thesaurus for Web Information Retrieval Hahn-Ming Lee.

[6] Query Expansion, Efthimiadis, E.N(1996), Annual Review of information science and Technology.

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

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

[9] 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).

[10] 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).

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

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

[13] E.M Voorhees, "Query Expansion using lexical-semantic relations", proceedings of the 17th annual international ACM SIGIR conference on Research and development in information Retrieval(1994).

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

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

[16] 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).

[17] Guber, T. (1993). A translation approach to portable ontology specification. Knowledge Acquisition, 199-220.

[18] McGuinness, N. F. (2000). What is an ontology and why we need it. Retrieved from http://protege.stanford.edu/: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html