ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Role of Domain Specific Ontology in Processing Transactional Queries

S.Ruban[1], Dr.S.Behin Sam[2], Shins Sebu Kurian[3], Ajit Lukose[3]

Assistant Professor, Dept. of MCA, AIMIT, St. Aloysius College (Autonomous), Mangalore, Karnataka, India[1]

Assistant Professor, Dept. of Computer Science, Rajeswari Vedachalam College, Tamil Nadu, India [2]

Student, Dept. of MCA, AIMIT, St. Aloysius College (Autonomous), Mangalore, Karnataka, India[3]

ABSTRACT: Information retrieval (IR) is the methodology of extraction of information resources related to a need of information from a set of information resources. Searches can be either metadata or content-based indexing. Information retrieval is used in various fields such as universities and public libraries, to provide access to books, journals and other documents. Search engines are the most visible Information Retrieval applications. The first process of information retrieval begins when a user enters a query into the system. Queries are statements that are used in order to retrieve required information from large amount of data. A query is not used to uniquely identify a single object in the collection; it can match several objects, perhaps with different degrees of relevancy. Any Information retrieval system has a query processing stage, indexing stage, matching stage and the ranking stage. Any change in any part of this system will have an overall impact in the system performance. In this experimental study we have studied the impact of domain specific ontology in our case we use a plants ontology that was developed, over transactional queries.

KEYWORDS: Information Retrieval, Ontology, Transactional queries, query processing.

I. INTRODUCTION

The search engines that we use such as google, yahoo or Ask have become a part and parcel of our day to day life. These search based applications roots goes back to the field of Information Retrieval. Traditionally the scope of information retrieval was in the field of libraries to search books etc. But the impact of web has brought a great change in the field of Retrieval. Information retrieval (IR) is the process of extracting information resources related to a need of information from a collection of information resources. Searches can be either metadata or content-based indexing. Information retrieval is used in various fields such as universities and public libraries, to provide access to books, journals and other documents. Search engines are the most visible Information Retrieval applications. The first process of information retrieval begins when a user enters a query into the system. Query is another word for question. In fact, outside of the computing terminology, the words "query" and "question" can be used vice versa. For example, if you need extra information from someone, you might tell, "I have a query for you." In computing, queries are also used to extract information. However, computer queries are sent to a computer system and are processed by a software program/code rather than a person himself. One type of query, which most people perform multiple times a day, is a search query. Each time when you search for something using a search engine, you perform a search query. As soon as you press Enter key, the keywords are sent to the search engine and are processed using an algorithm that retrieves relevant results from the search index. The results of your query appear on a search engine results page, or SERP. In the next section we will be explaining about query expansion, followed by semantic web and ontology, related work, materials, methods used, Results, Conclusion followed by References.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

II. QUERY EXPANSION

Query Expansion is generally aimed to develop a user query into one that is more responsive for Information retrieval. Earlier findings have showed that though query expansion had limited retrieval improvement on detailed queries, it determine great potential for significantly improving results given short queries. Since then there have been lot of work that is going on in query expansion by the researchers in the IR community. In literature many query expansion approaches are proposed and each of them has their own benefits and limitations.

III. SEMANTIC WEB AND ONTOLOGY

In the early of 21st century, Berners-Lee, Hendler and Lassila published an article entitled "The semantic web: A new web form content that is meaningful to computers will unleash a revolution of new possibilities". Since then lot of work began, with the aim of getting the web to be a place that will facilitate a more meaningful search. Some researchers have opted to build a hybrid system where Retrieval process will be more efficiently done taking the advantage of semantic web. Ontologies play a vital role in semantic web and it facilitates a semantic search systems. The word ontology is derived from the two Greek words 'on' and 'logos'. Ontology is a formal, explicit specification of a shared conceptualization.

- "Conceptualization" means an abstract model of phenomena in the world by having identified the related concepts of those phenomena.
- "Explicit" refers to the types of concepts used, and the constraints on their use are explicitly defined.
- "Formal" refers that the ontology should be machine readable.
- "Shared" reflects that ontology should capture consensual knowledge accepted by the communities

Ontology is more than taxonomy. Ontology's include richer relationships between terms. It is the rich relationships that enable the expression of domain-specific knowledge. This is a key distinction.

Advantages:

The advantage of ontology over traditional approach is that it represents real world information in a manner that is machine processable. The reason ontologies are becoming famous is largely due to what they promise: a shared and mutual understanding of a domain that can be communicated among application systems and people. Ontologies offer the following benefits:

- They aid in the communication between humans. Here, an unambiguous but informal ontology might be sufficient.
- They gain interoperability among computer systems by translating between different modeling methods, paradigms, software tools and languages. Here, the ontology can be used as an interchange format.
- They revise the process and quality of engineering the software systems.

IV. RELATED WORK

Some of the early works that were focusing on queries are being listed below. Since understanding the intention behind the user is taken into consideration for most of the classification of queries. The query classification which is mentioned in the literature is given below.

In the year 1992 some researchers did a study on browsing [1] and based on the queries they classified browsing into three types such as a) Browsing oriented towards search which is a process of finding relevant information for a particular task b) Browsing oriented towards review, which is the way of looking for information and c) Browsing towards scan, which is towards finding information without involving review.

Later in the year 1995 [2] in the work on Information seeking the researchers classified browsing patterns into directed, semi directed and undirected browsing.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Broder at IBM Research [3] proposed the widely known classification based on user intent as navigational web queries, Informational web queries and Transactional web queries. His classification was however based on some search logs that he collected when he was working with AltaVista Corporation and later was supported by the surveys. His experiments revealed that based on query log analysis 30% are transactional web queries, whereas 48% are informational and 20% are navigational web queries. Out of the user survey he conducted he reveals that an estimate of 39% is informational whereas an estimate of 36% is transactional and 24.5% are navigational.

Some studies that were conducted with AltaVista [4] states that there is an increased use of search engines for navigation. Because of the hypermedia capability of the web, it provides a very unique way of browsing where we use search engines to move to another site of interest or of need. Another classification of queries was made by [5] based on the query user used to search, the results which the user viewed. They sub categorized as informational queries, navigational queries and then resource queries.

The studies that were quoted above were based on query logs from the Search Engines and many of the classifications were manually done. The work based on the automatic identification of user goals [6] classified the queries into informational and navigational queries. They did this experiment with 50 queries collected from a group of students in a university. They could experience a success rate of over 50%.

The work on detecting commercial intention [7] studied whether the user information need given by user has any commercial motivation or not. They later revealed that 38% of the queries that were taken for the study have commercial intention. Another work [8] on the intention behind the web queries used supervised and unsupervised learning to divide the queries as informational or not informational or ambiguous. They declared that they were able to achieve 50% of precision.

The work on determining the user intent of queries [8] defined and presented three levels of hierarchical classification of user intent for searching web. In the next section we will be explaining about the material and methods in which we will be talking about the research objectives specifically with respect to this experimental study, which will be followed by Results, then Discussion and finally the conclusion, followed by the References.

V. MATERIALS AND METHODS

In this chapter let me narrate the materials and methods that we have used as part of this experimental study. The research objectives that were formed as part of our experimental study are listed out below.

Research Objectives:

1. Develop a Query Expansion methodology to enhance the query.

For research objective one we analyzed prior work in this area with an analysis of numerous actual query expansion approaches that has been developed and proposed over the period of time. It is very difficult to compare results across the different studies that have happened. But we would like to use the query Expansion that used ontology.

2. Identify the Transactional Queries

For Research objective two, we followed the classification that was laid down by Andrei Broder. This study was later redefined by Bernard J.Jansen, Danielle L.Booth and Amanda Spink where they did an automatic classification of queries that were used by the users. Based on the user intent, the queries are divided into navigational, informational and transactional we identified and selected few queries which are transactional in nature, i.e., the user intent was to reach a particular site of interest to perform some transactions.

3. Implement the Query Expansion methodology for the transactional queries.
For this research objective, we took the transactional queries that we have identified in the previous research objective and passed through query expansion method identified in the first research objective. The query is passed through the

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Stanford parser, where the initial parsing is done. Later it is passed to the Query expansion phase where the identified keywords of the query are taken into consideration and all its relevant terms which are called as synsets are displayed. Once the user selects the relevant synsets, they are added to the original query hence forms the refined query which is then passed to the Search API.

4. Evaluate the performance of the expanded queries using the comparative study with any Search API.

The evaluation of any Information Retrieval system is based on the precision and recall values. Since we evaluate our system directly in the web resource, it is impossible to calculate the correct the recall value because the environment in which we test is a dynamic and fluid environment. Hence we will be calculating the precision value for every individual query then we together calculate the Average Precision value.

VI. RESULTS AND DISCUSSION

The Experimental evaluation was conducted in the span of 2 months from the month of March 2015 to April

2015. The above queries were executed in two ways. We choose to use Google's Search API for our experiment and overall coding was done using JENA a Java based Framework.

Case i: Transactional Queries directly given to the Search API.

Case ii: Transactional Queries that has been refined using Domain specific Ontology (developed in protégé) which is then given to the Search API.

In the first case the queries were directly given to the search API, the first 50 results were manually evaluated for relevance, and the precision was calculated for every query.

Table 1: Transactional Queries with their precision values

Sl no	Transactional Queries	Google	Bing
1.	Variety of rice	32	25
2.	Type of silk plant	21	18
3.	Palm tree	28	38
4.	Verbal description of grassland Plants	16	16
5.	Kinds of plants	29	33
6.	Types of garden	35	32
7.	Popular crop by farmers	19	15
8.	Description about plants	36	34
9.	Potatoes used for cooking	38	37
10.	Name of almond trees	38	33

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

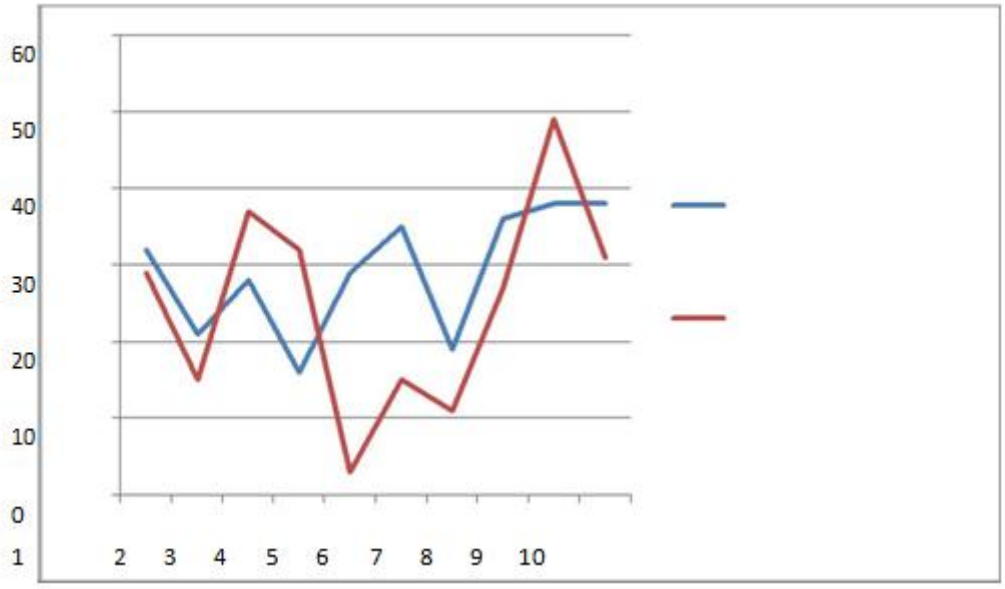
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Table 2: Refined Transactional Queries with their precision values

Sl no	Transactional Queries	Google	Bing
1.	Kinds or variety of rice	29	30
2.	Types or kinds of silk plant	15	11
3.	Palm tree or plants	37	51
4.	Description or Verbal description of grassland plants	32	18
5.	Variety or Kinds of plants	3	13
6.	Types or kinds of garden	15	14
7.	Popular or famous crop by farmers	11	2
8.	Description or explanation about Plants	27	30
9.	Potatoes or sweet potatoes used for Cooking	49	36
10.	Name or list of almond trees	31	22



Transactional Queries without Refinement in Google

Transactional queries with refinement in Google

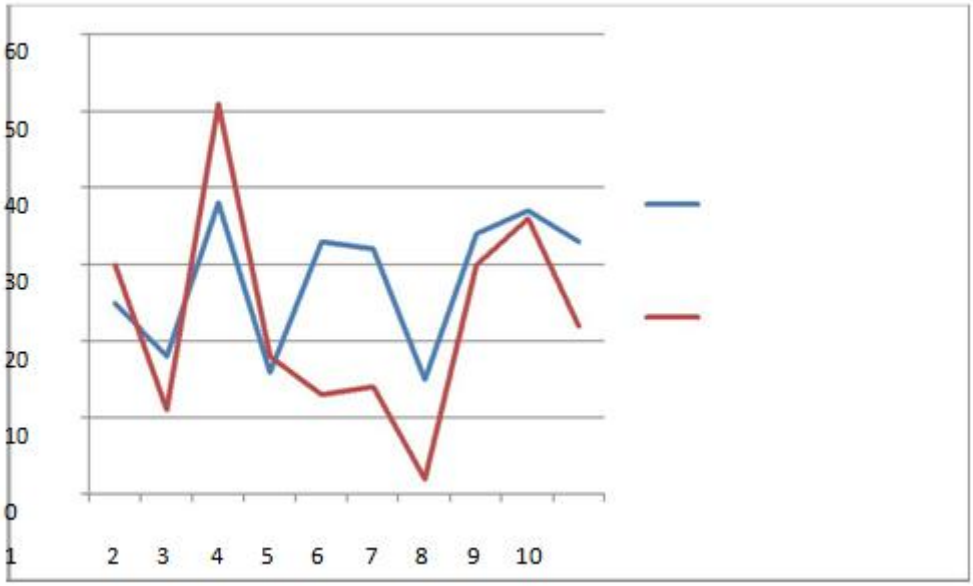
Figure 1: Transactional queries in Google

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015



Transactional Queries without refinement in Bing

Transactional queries with refinement in Bing

Fig 2: Transactional queries in Bing

V. CONCLUSION

It is always said that query expansion is a process to make a query into a more responsive representation which will retrieve more relevant results. In our study we have proved that in the case of transactional queries, it will decrease the performance of the system. Though the results we have obtained may vary based on the timing of execution of the queries, but based on our experimental results we conclude that any query that is aimed at using a search engine to navigate to a particular site or location to do some transactions such as downloading a file or executing a transaction will never benefit from domain specific ontology based query expansion in our case plants ontology.

REFERENCES

1. Carmel, E., Crawford, S., & Chen, H. (1992). In Browsing in hypertext: A cognitive study (pp. 865– 884). Paper presented at the IEEE transactions on systems, man and cybernetics, 5–10 October, Chicago IL.
2. Marchionini, G. (1995). Information seeking in electronic environments. Cambridge: Cambridge University Press.
3. Broder, A. (2002). A taxonomy of Web search. SIGIR Forum, 36(2), 3–10.
4. Jansen, B. J., Spink, A., & Pedersen, J. (2005). Trend analysis of AltaVista Web searching. Journal of the American Society for Information Science and Technology, 56(6), 559–570
5. Rose, D. E., & Levinson, D. (2004). In Understanding user goals in Web search (pp. 13–19). Paper presented at the World Wide Web conference (WWW 2004), 17–22 May, New York, NY, USA.
6. Lee, U., Liu, Z., & Cho, J. (2005). In Automatic identification of user goals in Web search (pp. 391– 401). Paper presented at the World Wide Web conference, 10–14 May, Chiba, Japan.
7. Dai, H. K., Nie, Z., Wang, L., Zhao, L., Wen, J. -R., & Li, Y. (2006). In detecting online commercial intention (OCI) (pp. 829–837). Paper presented at the World Wide Web conference (WWW2006), 23– 26 May, Edinburgh, Scotland.
8. Baeza-Yates, R., Calder´on-Benavides, L., & Gonz´alez, C. (2006). In The intention behind Web queries (pp. 98–109). Paper presented at the string processing and information retrieval (SPIRE 2006), 11–13 October, Glasgow, Scotland.
9. Bernard J. Jansen, Danielle L. Booth, Amanda Spink. (2008). Determining the informational, navigational and transactional intent of web queries, Information processing and Management 44(2008) 1251 – 1266.