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

Determining the Impact of Domain Independent Ontology over Transactional Queries

S.Ruban[1], Dr.S.Behin Sam[2], Naveen Thomas[3], Arun Angamaly[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 systems can also be described by the scale at which they operates. In web search, the system should provide search over billions of related documents stored on different millions of computers. Distinctive issues are compulsory to gather documents for indexing, being able to build large systems that work efficiently at this large scale, and handling particular aspects of the web, such as the profiteering of hypertext and not being fooled by site providers change page content in an attempt to boost their search engine rankings, given the commercial priority of the web. The field of information retrieval helps the users in browsing or filtering document collected or different set of retrieved documents. Query handling plays an important role in Information Retrieval. To enhance the query it is important to understand the user need behind this. There are different classifications of queries that have been proposed over the given time of period. One such classification, categorizes queries into Navigational, informational and transactional. In this experimental study we analyze the impact of domain independent ontology when expanding the transactional queries.

KEYWORDS: Query, WordNet, Query Expansion, Transactional query, Information Retrieval.

I. INTRODUCTION

Information retrieval (IR) is finding documents of an unstructured nature (usually text) that satisfies an information need from large collections (usually stored in computers). The field of information retrieval also covers supporting users in browsing or filtering documents, clustering which is the task of coming up with a good grouping of the documents based on their contents.

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 given increased challenge overall. 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 carried forward as 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.

The term Information Retrieval means to a search that may cover all 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. IR deals with the representation, storage, organization and access to Information.

The Rest of the paper is as follows, in section 2 we will be briefly discussing about Retrieval process , in section 3 we will be discussing about the Query and query expansion in general, in section 4 we will be discussing about Semantic web and ontology, in section 5 we will discuss about the related work and the methodology we adopted to improve 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

query then in section 6 we will be presenting the results based on the evaluation we have done and finally the conclusion.

II. INFORMATION RETRIEVAL PROCESS

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

i) The representation of the content of the documents,
ii) The representation of the user's information need, and
iii) The comparative study of the two representations.

The processes are visualized in Figure 1.

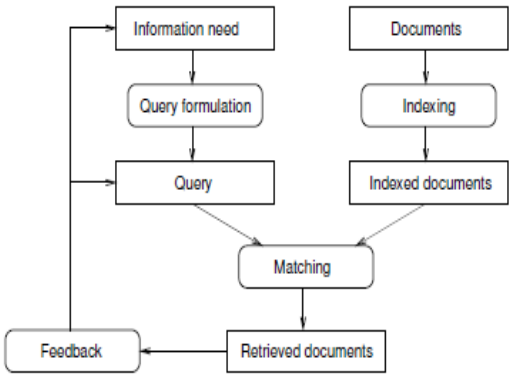

Figure 1: The Retrieval Process

Information Retrieval can be useful for the development, implementation and evaluation of a search engine. Indexing process is referred as representing the documents. The process of representing the information need of a user is referred as the query formulation process. The resulting representation is the query. The comparison of the query of the document representations is said to be the matching process. Retrieval strategy refers to information retrieval model. Retrieval strategies assign a measure of similarity between a query and a document. Any Information Retrieval system is held upon IR process. An Information Retrieval process model is a quadruple {D, Q, F, R (qi,dj) } where

i) D is a set composed of logical views (or representations) for the documents in the collection.
ii) Q is a set composed of logical views (or representations) for the user information needs. Such representations are called queries.
iii) F is a Framework for modeling document representations, queries, and their relationships.
iv) R(qi,dj) is a scaling function which includes as a real number with a query qi € Q and a document representation dj € D.

Such scaling defines an indexing among the documents with regard to the query qi. A model of information retrieval serves as an original document which is used to implement as an actual information retrieval system.

III. QUERY AND QUERY EXPANSION

Query :

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. A query can be simply a word or a more complex combination of operations involving several words. The most elementary query that can be formulated in a text retrieval system is a word. The output of word queries is the set of documents containing at least one of the words of the query.

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

A simple way to formulate user queries is using keywords in place of natural languages to match user's information needs. In Information Retrieval, the user's input queries usually are not detailed enough to allow fully satisfactory results to be returned. Query expansion can solve this problem.

Query Expansion :

Query Expansion is generally aimed to formulate 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 demonstrated great potential for significantly improving results given short queries. since then there have been lot of work that is carried on query expansion by the researchers in the IR community. In literature many query expansion approaches are proposed, and each of them have their own benefits and limitations.

IV. ONTOLOGY

The Artificial-Intelligence literatures have variety of definitions about ontology; most of these contradict one another. For the purposes of this guide ontology is a formal explicit description of concepts in a domain of discourse, properties of each concept describing different properties and features of the concept, and restrictions on slots. Ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base starts.

Classes are the focus of much ontology. Classes describe concepts in the domain. Every class can have subclasses that represent concepts that are more specific than the super class
In practical terms, ontology development includes:

- Defining classes in the ontology,
- Arranging the classes in a taxonomic (subclass–superclass) hierarchy,
- Defining slots and describing allowed values for these slots,
- Filling in the values for slots for instances.

We can then develop a knowledge base by defining one by one instance of these classes filling in particular slot value information and additional slot limitation.

In this research, we developed a framework to study the refinement of Transaction queries using the domain independent ontology wordnet. We picked up the user queries which were transactional in nature. We implemented in a program that automatically applied wordnet to the given transactional queries. We discuss later how this wordnet impacts the queries that are transactional in nature. We provide the experimental results based on the five mostly used search Engines across the world.

The next section presents the related research concerning the different query expansion approaches that used Wordnet for query refinement.

V. RELATED STUDIES

Cui et al [1] in their experiment suggested using the concepts from the query logs to be added as part of the query expansion process. Li et al [2] in their work titled improving ad-hoc queries suggested to use the concepts from Wikipedia for the query expansion process, Another group of researchers [3] in their work titled Effective query expansion with multiple information sources suggested using the query expansion terms from the web.

Ontologies are also studied to be used as part of query expansion [4]. There are two types of ontology, such as Query Independent Ontology and Query dependent ontology. Wordnet is a example of Query independent ontology. Though the previous experiments that used wordnetfor query expansion showed improvement in some cases and the performance was degraded in some cases, still handpicking and manually expanding the query would be helpful to overcome some flaws that may arise

When expanding using the automatically created
Thesauri.
But very recently Fang [5] in their work involving wordnet showed a increase in the performance of the system in

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

the axiomatic framework. This result gives a clear indication that wordnet used correctly can be helpful in refining the query thereby increasing the accuracy of the system.

Pal, D., Mitra, M., &Datta, K. [6] in their work involving wordnetto expand the query not based on the candidate terms generated from wordnet, but using the term distribution and term association. This work showed a great increase of performance of the system.

Carpineto, C., & Romano, G. [7] in their survey of automatic query expansion in information Retrieval lists out the different possible ways by which the query can be expanded and the new candidate terms can be added.

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 and Yahoo 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 Ontology 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.

+	**Transactional Queries**	**Google**	**Yahoo**
1.	Ola Cab Booking	38	39
2.	Udupi Properties	33	42
3.	MP4 Player Download	44	43
4.	Service Oriented Architecture By Thomas URl	40	48
5.	Ford Fiego Second	38	37
6.	Flickart	50	49
7.	Corporation Bank	44	40
8.	Taken Movie Download	30	34

Table 1: Transactional Queries with their precision values

In the second case, the queries were given to the Stanford parser, then they were given as input to the ontology i.e., Wordnet, the terms that were generated are called as synsets, the relevant terms are added to the original query hence the new refined query is formed, which is then passed to the Search API. The first 50 results were manually evaluated for relevance and the precision is calculated. The following table displays the original value that was got during the experiment when the queries were passed to the Search API.

Table 2 : Refined Transactional Queries with their precision values

	Refined Transactional Queries	**Google**	**Yahoo**
1.	Ola Cab Booking or Reservation	31	30
2.	Udupi Properties or Places Details	37	41
3.	MP4 Player Download or Participant Download	29	35
4.	Service Oriented Architecture By Thomas	40	29

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

	URl		
5.	Ford Fiego Second or 2nd Sales	30	33
6.	Flickart	50	49
7.	Corporation or Company Bank Online	40	28
8.	Taken Movie or Film Download	29	15

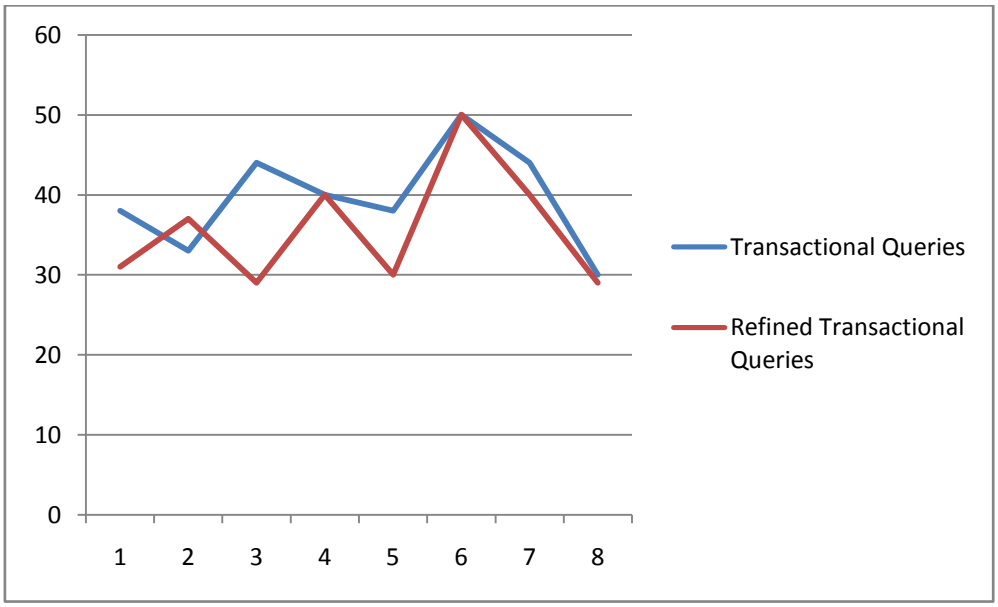

Fig 2 : Transactional Queries Vs Refined Queries in Google

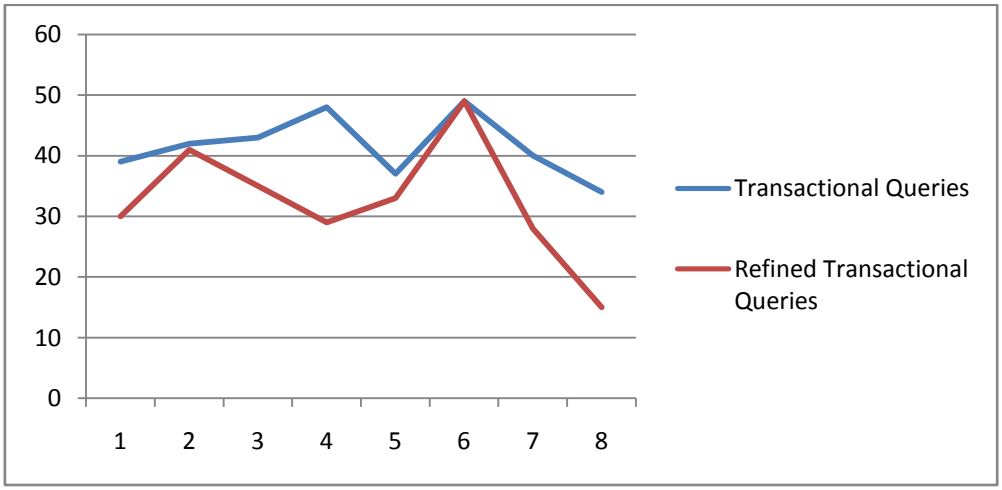

Fig 3 : Transactional Queries Vs Refined Queries in Yahoo

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

VII. CONCLUSION

In our study we have proved that in the case of Transactional queries, the query refinement using wordnet 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 for transactional purposes will never benefit from wordnet.

REFERENCES

1. Cui, H., Wen, J.-R., Nie, J.-Y., & Ma, W.-Y. (2002). Probabilistic query expansion using query logs.In *Proceedings of the 11th international conference on World Wide Web* (pp. 325–332). New York, NY, USA: ACM.
2. Li, Y., Luk, W. P. R., Ho, K. S. E., & Chung, F. L. K. (2007). Improving weak ad-hoc queries using wikipediaasexternal corpus.In *Proceedings of the 30th Annual International ACM SIGIRConference on Research and Development in Information Retrieval* (pp. 797–798). New York, NY, USA: ACM.
3. Bendersky, M., Metzler, D., & Croft, W. B. (2012). Effective query formulation with multiple information sources. In *Proceedings of the fifth ACM international conference on Web search and data mining*(pp. 443–452).
4. Bhogal, J., Macfarlane, A., & Smith, P. (2007, July). A review of ontology based query expansion. *Inf.Process. Manage.*, *43*(4), 866–886.
5. Fang, H. (2008). A re-examination of query expansion using lexical resources. In *Proceedings of ACL-08:HLT*(pp. 139–147).
6. Pal, D., Mitra, M., &Datta, K. (2013). Query expansion using term distribution and term association.*CoRR, abs/1303.0667*.
7. Carpineto, C., & Romano, G. (2012, January). A survey of automatic query expansion in informationretrieval.*ACM Comput.Surv.*,*44*(1), 1:1–1:50.