International Journal Of Advance Research In Science And Engineering http://www.ijarse.com

IJARSE, Vol. No.2, Issue No.3, March, 2013 ISSN-2319-8354(E)

SEMANTIC SEARCH SYSTEM USING QUERY REFINEMENT

[1]Monam Sharma, [2]Anu Varshney, [3]Pragya Agarwal

UG, [1,2] Department of Computer Science Engineering,
[3]Asst. Prof., Department of Computer Science Engineering,
Raj Kumar Goel institute of Technology for Women, Ghaziabad,

Gautam Buddh Technical University, Lucknow, (India)

ABSTRACT

A web search engine is a software code that is designed to search for information on the world wide web. Web search engine work by storing information about many web pages. Semantic search is also a web search engine .Semantic search seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms. In this paper a semantic search service is being developed to help users gather relevant documents more efficiently unlike traditional web search engine. When a user enter a query into a search engine (typically by using a keyword).The engine examine its index and provides a listing of best-matching web pages according to its criteria. This search engine display result based on synonyms or multiple meanings of query that is extracted from web resource and stored in the database i.e. definition repository. It extends the traditional keyword based web search in order to provide semantic and context based search. The system works as a layer above the keyword based search engine to generate sub-queries based on different meanings of query keywords. This experiments show that this approach is efficient as it results in reducing the search space to a large extent and it support information overkill.

Keywords – Query Processing, Query Refinement, Search Engine, Semantic Search Technique, World Wide Web.

I INTRODUCTION

The Semantic web is an extension of of the current web in which information is given well defined meaning, better enabling computer and people to work in cooperation. The outcomes of search typically depend on submitted queries cannot be guaranteed queries. But the effectiveness of query cannot be guaranteed as they are vary from user to user. Semantic Search uses Semantics or the science of meaning in language to produce highly relevant search results. In most cases, the goal is to deliver the information queried by a user rather than have a user sort through a list of loosely related keyword results. The semantic web is being developed based on the current web with refreshed framework information resources are described using logic-based knowledge representation languages. Its aims to enable computers to automatically process information and to promote

reusability and interoperability across heterogeneous system. A search engine query is a request for information that is made using search engine. Every time a user puts a string of characters in a search engine and presses "Enter", a search engine query is made. The string of characters acts as keyword that the search engine uses to algorithmically match results with a query. The results are displayed on the search engine in order to significance most of the data in this engine is solved by definition based sites. Every search engine query adds to mass of analytical on the internet. The more data search engine collect, the more accurate the search results become-and that's a good thing for the internet users. The development of semantic web search system has been an emerging area of research since the last few years and many researchers have shown their interest in this particular field. Query refinement and expansion has become an essential information retrieval approach that interactively recommends new terms related to a particular query. This paper contributes towards developing a search service for semantic and context –based retrieval , while at the the same time , keeping in view the above limitation. The proposed technique has been made to take the advantage of dictionary based information or definition related terms available on the web to gather possible meanings and generate the query response accordingly. The result will be displayed in the form of clusters accordingly to the newly found sub-term. The section 2 explain about the semantic search system, the section 3 explain about the algorithm, section 4 give about the table and figure and section 5 give conclusion of the paper with a discussion of future work.

1I SEMANTIC SEARCH ENGINE

Semantic Search in search engine mean the search engine would provide relevant search results based on the intent and contextual meaning of the search term. Semantic search engine is used to search the word or meaning or we can say synonyms based on keyword based indexing or query based search engine that's why quesem is also called (QUESEM)/(QUERY SEMANTIC SEARCH ENGINE) . While quesem basically improve the search quality. Quesem is used to maintained a database that we called **" Definition Repository "**.

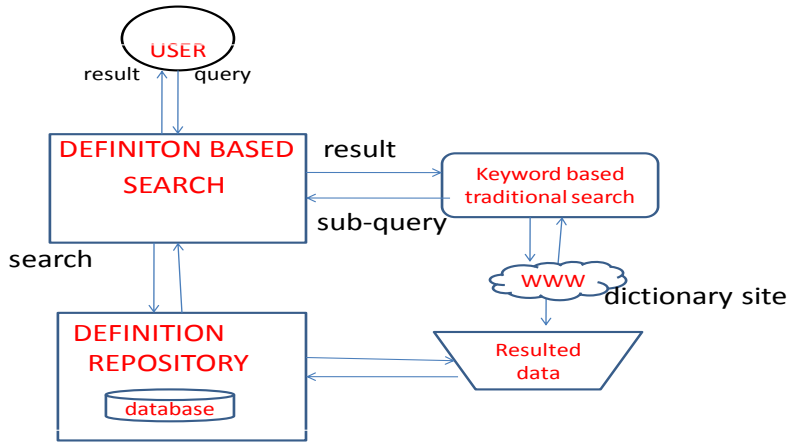

Fig:1 Architecture Of QUESEM

Basically the system is divided in to two sub-systems

1. **Definition Repository Generation**
2. **Definition-Based Search**

When a user enter a query then definition based search use to analyses the query then search that keyword based query in the definition repository database otherwise cal query based search engine .When a user enter a query then definition based search use to analyses the query then search that keyword based query in the definition repository database otherwise cal query based search engine

After that in definition repository tropical crawler use to perform crawling by using an algorithm i.e. local site searching afterwards definition generation and definition annotator use to perform its function. The web use to contain all the sites and dictionary based sites. After that sub queries are formed.

2.1 Definition Repository Generation

Definition based repository stores different meaning related word synonyms of keyword. It consists at least two relation like table 1 give the various description various field in the database i.e. definition repository. Fig 2 gives the state of the word. E.g. the term" **cloud"** consists various definitions like "cloud computing", "cloud atlas".

TERM TABLE

Term_id	**Term_Title**	**Term_Description**

DEFINITION ANNOTATION TABLE

Def_id	**Def_title**	**Def_Description**	**Term_id**

Schema of DEFINITION REPOSITORY

Field	Description
Term_id	An identity number to the term under consideration.
Term_Title	It contains actual query terms entered by the user.
Term_Description	It contains a limited length snippet to have a small description of the term.
Def_id	It represents the identity no.
Def_Title	It define actual definition of query
Def_Desccription	It give the small description of each definition.

Various field in DEFINITION REPOSITORY

Example

TERM TABLE

1	Cloud	It is visible mass of liquid droplet

DEFINITION ANNOTATION TABLE

1.1	Cloud computing	It's a company	1
1.2	Cloud atlas	Third book by British author David Mitchell	1
1.3	Cloud 9	It is a state of happiness	1

2.2.1 Information Resource:

In this all the dictionary based sites are present, which act as a input resource. It is a basically a rich store of semantic definition of term. A tropical is used to crawl each and every words in dictionary based sites.

2.2.2 Tropical crawler:

The basic aim of crawler is to estimate the related word and meaning of query. To start the crawl process ,some topic like "dictionary" is needed to be provided .a publically available search service (Google SOAP API) can be used to filter(e.g. dictionary ,definition etc) in the title of the home page of sites. The crawler will output a set S of sites related to dictionary based content ,which is stored in a local repository to be referred by DEFINITION GENERATOR.

2.2.3 Definition Generation and Annotation:

Generator simply extracts the *prefix* and *suffix* tokens present consecutively with query term from the pages. Annotator is used to combine with query term to give proper word.

"Extracts suffix and prefix query term from page, annanate them with query term and store the results in definition repository

Algorithm (local site searching)

Algorithm: LOCAL_SITE_SEARCHING (Initial query, S)

I/P= Initial query *q* and Set *S* of Sites stored in a repository
O/P= Unstructured document set *D* containing documents that are response pages against the initial query *q*.
// Start of Algorithm

```
Begin
    For (every site si∈ S)          // perform the local-site search
        Begin
            Step1: fetch the homepage of si
            Step2: find the local-site search form
            Step3: Submit the query terms (q)
            Step4: Fetch the response pages in local Repository D
            Step5: For (every response Page          //Find linked pages
                        Begin
                            Parse the Page;
                            Fetch all result links;
                            If synonyms link exist
                              Begin
                                  Fetch synonyms page and place in D
                               End
                        End
            End
    Return the fetched page set D
End
```

III FIGURE

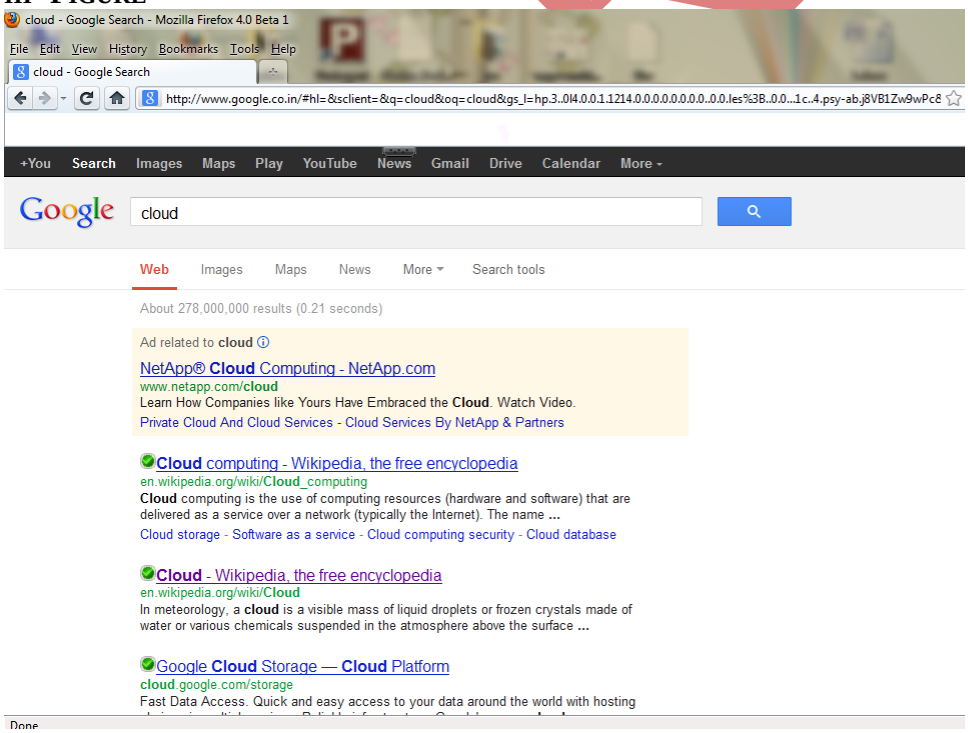

Figure 2: Results for query "Cloud"

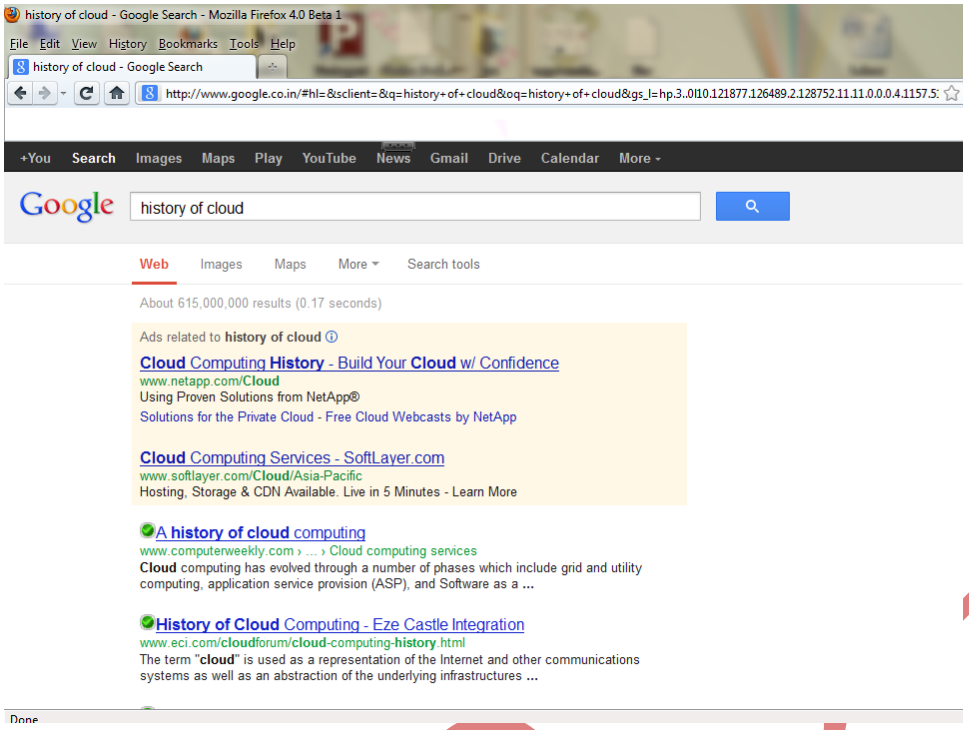

Figure 3: results for query "History of Cloud"

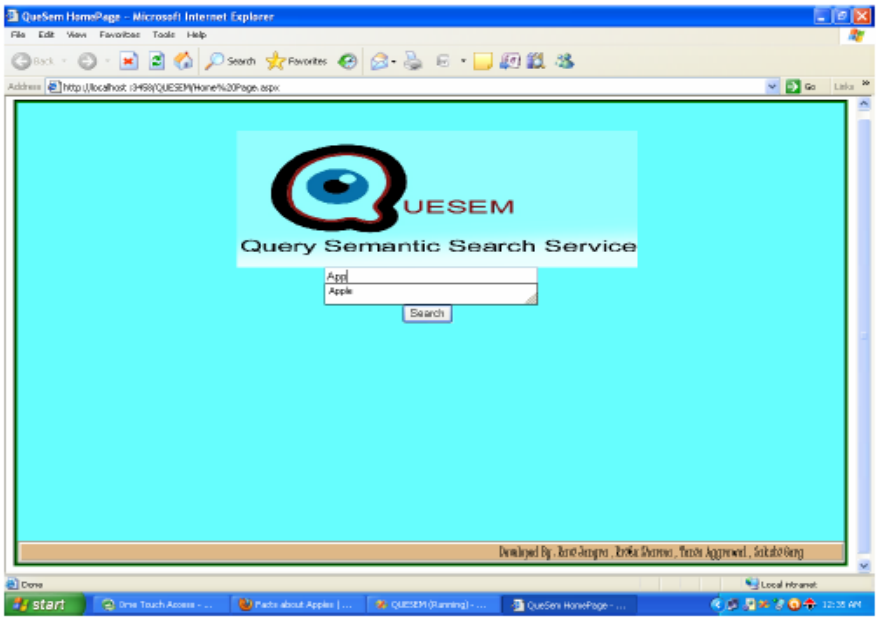

Figure 4: Interface of QUESEM

IV CONCLUSION

A Semantic search system, QUESEM, to address the "information overkill problem" has been developed. Its acts like a Meta Search Service which is made to utilize the existing web resources to automatically extract the synonym related to user queries and enhance the search efficiency. Assisted by information of definitions, proposed system is able to understand users' queries in a better way to perform more meaningful searches. The future research include enhancing the system towards serving different types of complex queries.

REFERENCES

[1] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., and Raghavan, S., "*Searching the Web*". ACM Transactions on Internet Technology. Aug. 2001, pp: 2-43.

[2] Aggarwal, C. C., Al-Garawi, F., and Yu, P. S., "*On the design of a learning crawler for topical resource discovery*".ACM Transactions on Information Systems. Vol. 19, No. 3,
Jul. 2001, pp: 286-309.

[3] Chakrabarti, S., Vandenberg, M., and Dom, B. "*Focused crawling: a new approach to topic-specific Web resource discovery"*. In Proceedings of the Eighth International Conference on World Wide Web, Toronto, Canada, 1999. pp: 1623-1640.

[4] Ziming Zhuang, Silviu Cucerzan, *"Exploiting Semantic Query Context to Improve Search Ranking*". IEEE International Conference on semantic Computing, 2008 DOI.

[5] Liu, Y. and Agah, A, "*Crawling and Extracting ProBeijing"*, China, August 17-19, 2009, Springer-Verlag, Berlin, Heidelberg, LNAI 5678, pp: 545-552.

[6] Liu, Y. and Agah, A, "*A Prototype Process-Based Search Engine"*. In Proceedings of the third IEEE International Conference on Semantic Computing, Berkeley, CA, September 14-16, 2009.

[7] OpenNLP, http://opennlp.sourceforge.net/cess "*Data from the Web*". In Proceedings of the 5th International Conference on Advanced Data Mining and Applications.

[8] Y. Li, R. Krishnamurthy, S. Vaithyanathan, and H. V.Jagadish, "*Getting work done on the web: supporting Transactional queries*," In Proceedings of the 29th ACM SIGIR, Seattle, Washington, 2006, pp. 557- 564.

[9] Giorgos Akrivas, Manolis Wallace, Giorgos Andreou, Giorgos Stamou and Stefanos Kollias, "*Context Sensitive Semantic Query Expansion".* Proceedings of the 2002 IEEE International Conference Artificial Intelligence System (ICAIS'02).