Web Technology (Semantic Search System)

Monam Sharma
Department of Computer Science
Uttar Pradesh Technical University
Ghaziabad, India
monamsharma@gmail.com

Anu Varshney
Department of Computer Science
Uttar Pradesh Technical University
Ghaziabad, India
anu.varshney984@gmail.com

Pragya
Lecturer, Department of Information Technology
Uttar Pradesh Technical University
Ghaziabad, India
agrawal.pragya11@gmail.com

***Abstract*— A web search engine is a software code that is designed to search engine for information on the World Wide Web. Web search engine work by storing information about many web pages. When user submits a query, the generated information is often very large and inaccurate, that result in increased user perceived latency. In this paper, a novel approach of definition based search is being introduced that solves this problem. The proposed system searches and displays results based on themes, definitions and synonyms of the query keywords generally extracted from the web resources and stored in a separate definition repository. It extends the traditional keywords 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 spaces 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 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 varying from user to user. The semantic web is being developed based on the current web with refreshed framework information resources are described using logic based knowledge representation language. Search engine enable user to navigate through the web information content incrementally and interactively. Keyword based indexing in search engine is another factor towards irrelevancy of search result. They are unable to associate the query words with related fields of our daily world. We have magnitudes of words, out of which many possess more than one meaning. Semantic search uses semantic 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 short through a list of loosely related keyword results. Its aim to enable to 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 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. Most of the information towards resolving the above said issues exits on the web in *definition-based* or *dictionary-based* sites. This paper proposes a semantic search system that utilizes the existing web resources to carry out a definition based search and enhance the search quality.

II. RELATED WORK

The development of semantic web search system has been an emerging area of query research since the last few years. Query refinement and expansion has become an essential information retrieval approach that interactively recommends new terms related to a particular query. As keyword based queries are likelyto miss important results due to unstructured and semantically. Heterogeneous nature of the Web, Traditional Information Retrieval (IR) technology is based almost purely on the occurrence of words in documents. Search engines like Google [9], augment this in the context of the Web with information about the hyperlink structure of the Web. The availability of large amounts of structured, machine understandable information about a wide range of objects on the Semantic. A critical look at the available literature indicates that most of existing semantic and context-based search techniques suffers from a couple of following limitations:

- They are not able to capture the full set of synonyms for each type of query submitted by the users.
- Most of the techniques require complex analysis involving natural language processing and linguistic preprocessing to discover the context and semantics of query terms.
- The techniques utilizing external web resources require complex retrieval efforts and the resource integration.
- Many techniques do not recommend the users with related semantically meaningful queries, from which,

ISBN: 9788175157538

user can browse any one of his interest and moreover, results may not be represented in a user navigational form.

This paper contributes towards developing a search system for Semantic and context-based information retrieval, while at the same time, keeping in view the above limitations. Web offers some opportunities for improving on traditional search.

III. 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"***. Basically the system is divided in to two sub-systems

- Definition Repository Generation
- 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. 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.

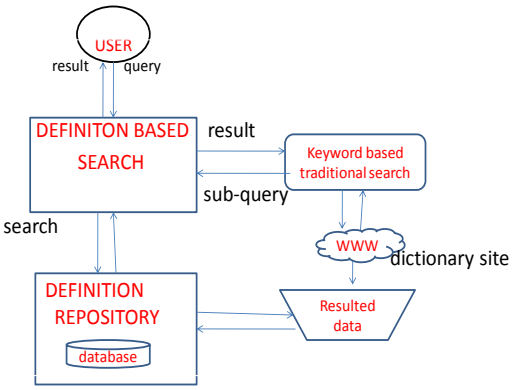

Figure 1. Architecture of Semantic System

A. *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".

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 word in dictionary based sites.

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.

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, annotate them with query term and store the results in definition repository.

B. *Definition Based Search*

When user submits a keyword based query to *Semantic Search System*, the Keywords are passed to "Definition Repository Generation "Subsystem to build the definition Repository as well as to the "Definition based Search" subsystem to respond to the various modules which are used in definition based query processing the user in the form of cluster of result URLs.

- Query Analyzer
- Definition Searches
- Query Transformer and Processor

1) *Query Analyzer:* Query analyzer is responsible to check whether definition based search is applicable to the query or not? Analyzer first examines the query to decide whether a query is topical or not? A topical is the query, which is generally framed by simple keywords.

ISBN: 9788175157538

2) Definition Searcher: The job of Definition Searcher is to check whether the topical query already exists in the system's Definition Repository. If it is there, that means some user has already queried it and as result of which its definitions are already stored in the database.

3) Query Transformer & Processor: Query transformer is responsible for making the retrieved definitions as a sequence of well-defined queries so that they could be searched individually as independent queries. The original query and the matched set of definitions will be combined to form a new set of sub-queries.

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.

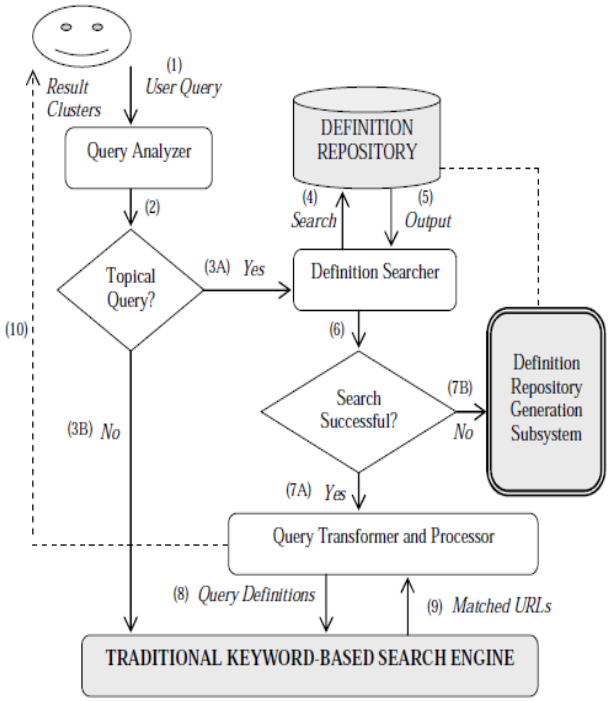

Figure 2. Definition Based Search Subsystem

TABLE I

TERM TABLE

Term_id	Term_Title	Term_Description

TABLE II

DEFINITION ANNONATION TABLE

Def_id	Def_title	Def_Description	Term_id

TABLE III

SCHEMA OF DEFINION REPOSITORY

Field	Description

III. ALOGRITHM

This algorithm is used to depict the local site searching in which first we find the local site form and then used to submit a query terms. Then, it fetch the response pages from the database i.e. definition repository and parse that page and their links.

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
```

IV. RESULT ANAYSIS

This figure is used to depict the result of Query Terms

ISBN: 9788175157538

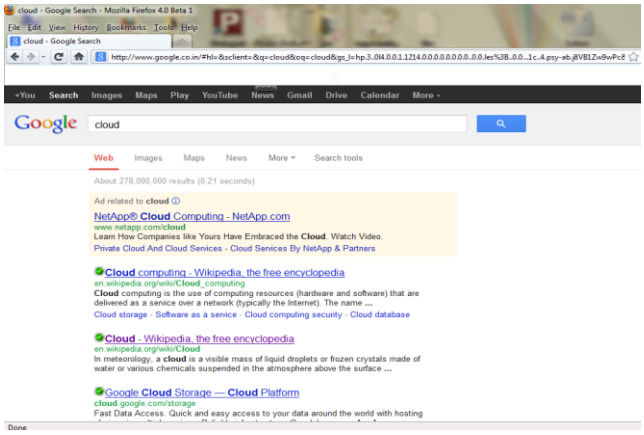

Figure 3. Result for Query "Cloud"

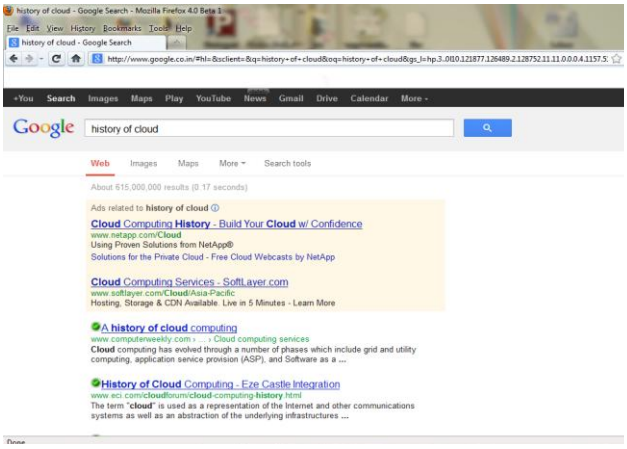

Figure 4. Results for Query "History of Cloud"

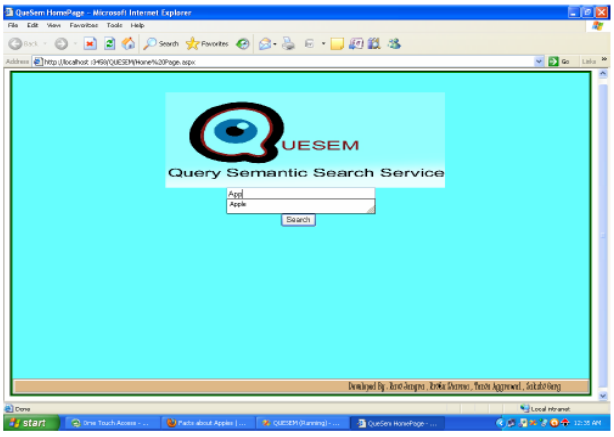

Figure5. Interface of Semantic System

VI CONCLUSIONS

A Semantic search system, 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: http://doi.ieeecomputersociety.org/10.1109/ICSC.2008.8

[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

[7]. Conference on Semantic Computing, Berkeley, CA, September 14-16, 2009.

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

[9]. Y. Li, R. Krishnamurthy, S. Vaithyanathan, and H. V.Jagadish, "Getting work done on the web: supporting

[10]. Transactional queries," In Proceedings of the 29th ACMSIGIR, Seattle, Washington, 2006, pp. 557- 564.

[11]. 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).

ISBN: 9788175157538