# A Semantic Query Transformation Approach Based on Ontology for Search Engine

SAJENDRA KUMAR[1], RAM KUMAR RANA[2], PAWAN SINGH[3]
[1, 2, 3]Department of Computer Science,
IIMT Institute of Engineering &Technology,
Meerut, India
[1]saajvoice@gmail.com,[2] ram_shobhit@rediff.com,[3]pawansingh3@yahoo.com

***Abstract* -- These days we are using some popular web search engines for information retrieval in all areas, such engine are as Google, Yahoo!, and Live Search, etc. to obtain initial helpful information. Which information we retrieved via search engine may not be relevant to the search target in the search engine user's mind. When user not found relevant information he has to shortlist the results. These search engines use traditional search service based on "static keywords", which require the users to type in the exact keywords. This approach clearly puts the users in a critical situation of guessing the exact keyword. The users may want to define their search by using attributes of the search target. But the relevancy of results in most cases may not be satisfactory and the users may not be patient enough to browse through complete list of pages to get a relevant result. The reason behind this is the search engines performs search based on the syntax not on semantics. But they seemed to be less efficient to understand the relationship between the keywords which had an adverse effect on the results it produced. Semantic search engines – only solution to this; which returns concepts not documents according to user query matching. In This paper we proposed a semantic query interface which creates a semantic query according the user input query and study of current semantic search engine techniques for semantic search.**

***Keywords: Semantic search, Query transformation, Semantic Web, WWW.***

## I. INTRODUCTION

The web could be a huge distributed and linked mass of the many resources that are poorly unstructured and unorganized. Web Search engines are information retrieval systems designed to go looking for information stored within the websites. Today there's an amazing increase within the resources obtainable on the online. Even supposing search engines give a listing of websites in an exceedingly fraction of milliseconds, most of them are irrelevant to the context. Secondly, the order of webpage's listed out based mostly on some metrics and ranking algorithms creates any confusion to the user. The explanation behind the dearth of quality results attributes to the inefficient processing of keywords while not understanding the means. This can be a lot of obvious when the search phrase is comprised of a collection of keywords.

The basic plan of semantic web is to counterpoint this time web with machine cognitive information regarding the semantics of data content. Semantic looking out understands the relation between the keywords and therefore the contexts during which they're used and returns results in keeping with the that means of the question. As a primary requisite, the resources ought to be semantically annotated in terms of ontology. Ontology could be a formal, specific specification of a shared conceptualization. Owning to it does perform of describing conceptual that means and relationships among completely different ideas, domain ontology will type a pleasant conceptual hierarchy and supply glorious support for logic reasoning. However annotating every and each document obtainable on the online could be a mightier task and highly impractical. Thanks to this reason our work concentrates on annotating solely the domains by taking the inputs from a domain skilled. This conjointly reduces ambiguity which could happen if annotation is completed by somebody with very little information regarding the domain. Automatic text categorization may be employed in assigning resources to at least one or a lot of such semantically annotated domains. The incompatibility between the question languages to question the semantic annotation and text creates one more downside within the development of semantic search engines. This could be resolved by hiding semantic annotations from the documents and keeping it separately. This helps us to question the annotation using SPARQL and therefore the text using SQL separately.

### *A. A Semantic Search Engine*

WWW is that the biggest revolution that went on to the technology. It still retains it pride because it serves and helps human kind indeed in many ways. Search engines are information retrieval systems designed to look

for information stored within the web content. And these search engines incorporates a crucial half in success of web and currently it's an inevitable a part of one‘s life. the internet could be a huge distributed and linked mass of the many resources that are poorly unstructured and unorganized. It’s forever a surprise for everybody how search engines retrieve a large list of web content in a very fraction of seconds. This result connects man to the resources unfold worldwide despite of the geographical boundaries. However the relevancy of leads to most cases might not be satisfactory and also the users might not hold back enough to flick thru complete list of pages to induce a relevant result.

The reason behind this can be the search engines performs search based mostly on the syntax not on semantics. The keyword based mostly search engines fails to grasp and analyze the context during which keywords are used. A thing worsens when the search phrase could be a combination of keywords. The standard of the results degrades with irrelevant results of documents that contains solely the part of search phrase leaving the that means aside. A semantic search –Tim Berner Lee‟s unrealized dream - resolves this issue by analyzing the contexts and also the relationships between the key words and therefore manufacturing a “high quality” and “less quantity” results [15]. The essential plan of semantic web is to counterpoint this net with machine cognitive info regarding the semantics of knowledge content.

## II. BACKGROUND

The semantic search engine needs semantically annotated resources as its necessity. Firstly it's a mightier task to semantically annotate innumerable resource accessible on the WWW. and also the range of documents added to internet rises in an exponential manner per year. Since nobody owns the web and there's no rules to stay the web going it's troublesome to trace whether or not documents - each added and existing – are semantically annotated. the WWW contains several helpful resources which may be used as plug-ins for different systems. WordNet [7] – a lexical electronic database developed at Princeton University contains a network of connected words that are classified into completely different contexts and tagged with correct sense [1]. This database conjointly provides out choices to different systems to form use of their vast database.

Ontology could be a set of assertions specifying the ideas concerned within the domain. By using ontology to prepare text documents, data are often represented in an exceedingly approach that facilitates understanding between computers and humans. This conjointly helps to make a hierarchical illustration composed of super-concepts and relevant sub-concepts and therefore makes data a lot of organized and straightforward to subsume [2].

Automatic text categorization that categorizes electronic resources into specific classes is currently a part of intensive analysis. A unique algorithm presented by Maciej Janik and Krys Kochut ontologically classifies the domain entities that are interconnected by relationships furthermore as instances. This algorithm works well with semantically annotated documents and features a satisfactory accuracy in classifying a well outlined semantic schema of entities [3].

SemSearch [4] is an ontology primarily based search engine that features a Google like question interface that prompts the user to enter the keyword and its corresponding concept. It then builds an inspiration relation graph with weighted arcs between the connected entities that later decides the resultant set. This search engines performance degrades when the user fails to relate between the keywords and ideas. Conjointly the removal of less weighted arcs from the graph would possibly chop off some relevant data that provides unsatisfactory results.

The semantic enlargement search model Sem-Exp-M [5] expands the question keyword semantically via semantic enlargement reasoning algorithm. The expanded question set is then used to retrieve results from Semantic Index repository that could be a vast resource of semantically annotated documents. Such retrieved documents have a better precision ratio and relevancy compared to different search models. Optimizing the time taken for Semantic annotation could be a highly suggested for this search model.

The translation of SPARQL queries to SQL queries [6] preserving the semantics is extremely useful since most of the RDF stores use RDBMS as its backend. This helps to bridge the gap between the semantic internet and databases.

## III. PROPOSED ARCHITECTURE

The model proposed here makes use of the prevailing resources like lexical database for English language, semantically annotated documents of various domains, RDF stores stored in RDBMS, Indexers and crawlers employed by the search engines.

The search model provides a semantic query interface through 3 sections; we pass keyword via these sections and make a semantic query. To understand the various contexts of the keywords we will create use of

lexical database of English language. Once the contexts are analyzed we will proceed with understanding the link between ideas, finding the synonyms and therefore expand the keyword set with that when searched can provide give quality results these sections are discussed below.

***The First section*** finds out the various contexts within which the keyword is used. In Figure 1, a lexical database of English language is employed because the context extraction tool for this section. A word in English language will seem in numerous contexts counting on how it's used. This section helps the search engine to precisely decipher what the user needs. It additionally helps to get rid of any ambiguity of the user relating to the domain of the keyword.

The contexts extracted are presented the user in order that he will choose the domain of the question keyword. Currently the search engine overhead is reduced drastically by confining the search to that individual domain. The synonyms of the domain selected are retrieved and stored in a very specialised knowledge structure for the later use.

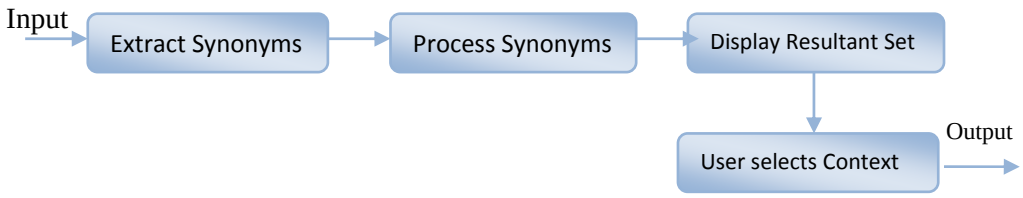


Figure 1: First Section

***In the second section***, the search is focused solely in the user specified domain. As a necessity for this section, the domains ought to be semantically annotated and also the resources ought to be classified into it totally different domains. This approach helps this model in numerous ways that. In Figure 2, Firstly, if documents are individually annotated the incompatibility between the question languages raises a difficulty which might be sorted if we tend to leave the documents because it is and annotate the domains. Secondly, the semantic annotation demands high information of the domains which can need the help of a site professional. it's not possible to hunt help of the domain professional to annotate each single resource on the net classified to be during this domain. An another excuse is since nobody owns web and there's no regulatory body to manage the functioning of internet it should not be potential to trace whether or not every document added to web is semantically annotated or to see whether or not semantic annotation is correctly done. The exponential rise in range of resources added to the net makes it still a lot of impractical.

The domains selected by the user are searched so as to retrieve connected entities. The algorithms work on sure metrics based mostly on the relationships between the ideas. This set of results is displayed once more to the user from that the user will choose rather more specific space. The keyword so selected by the user along with its synonyms can be added to the on top of said specialised knowledge structure. This second section is iterated until the user is completed with refining the search.

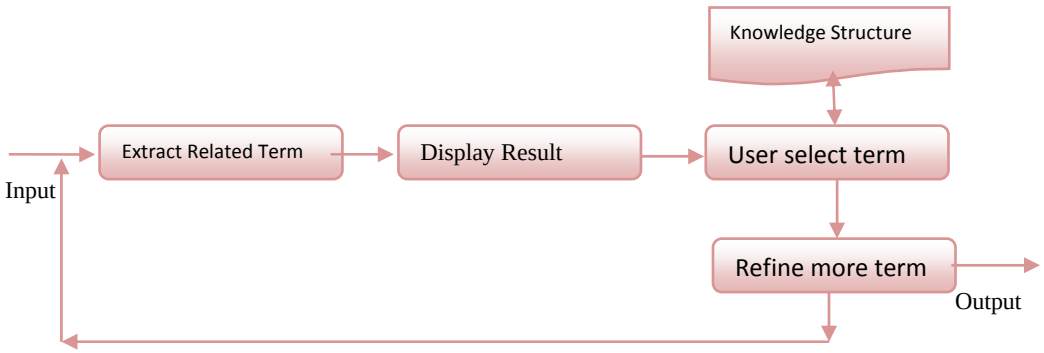


Figure 2: Second Section

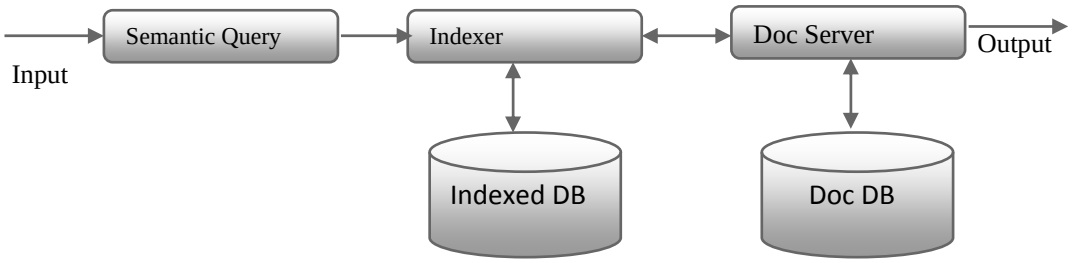


Figure 3: Third Section

***The third section*** starts processing by taking the knowledge structure as input and creates a semantic query as shown in figure 3. This knowledge structure contains the key words and its synonyms that are combined using AND and OR relations to create a group of search phrases. This processed search phrases are sent to the indexer of the search engine that retrieves the results from net servers and document servers. After processing the keyword from first section to last section this query now become a semantic query. The processed architecture is shown in figure 5.

For example in figure 4: user pass the following keywords- Spider man

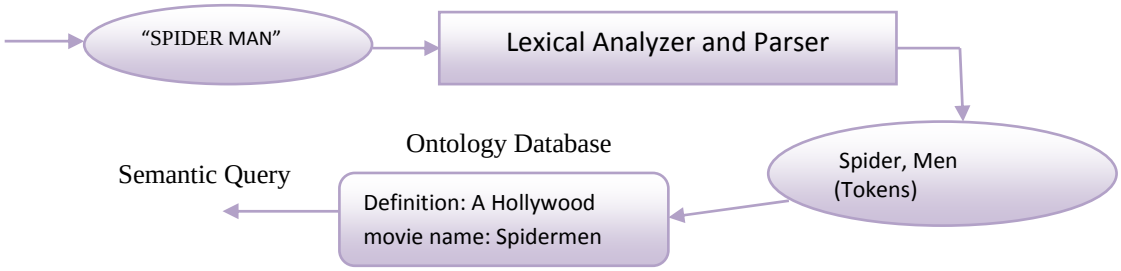


Figure 4: Example

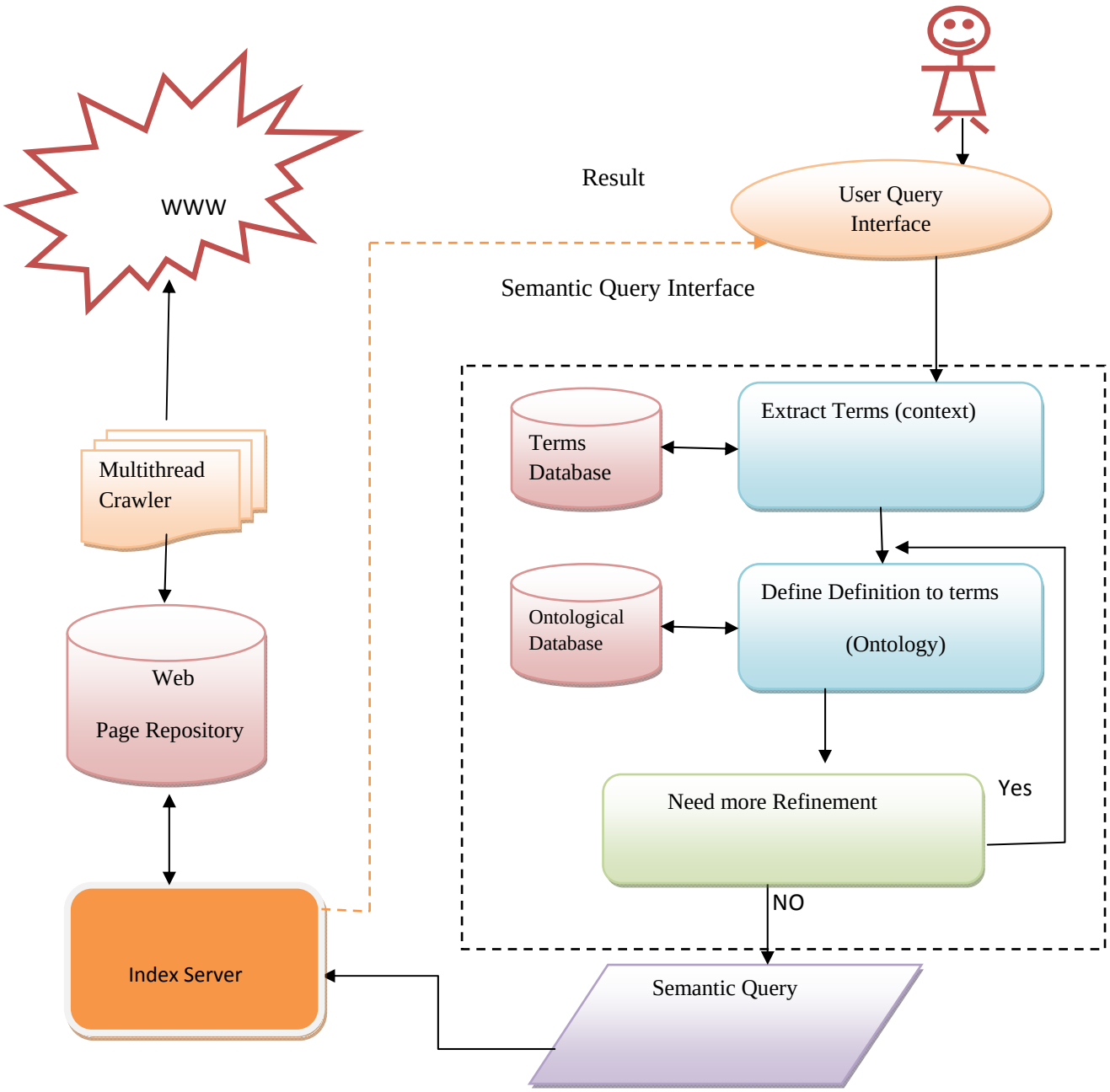


Figure 5: Proposed Architecture for Search Engine by using Semantic Query Interface based on ontology.

## IV. CONCLUSION AND FUTURE SCOPE OF WORK

This paper proposes A new way of carrying out semantic search by making use of existing resources. This method can provide better, accurate and most relevant result with compare to traditional search engine results, but search time may be large because semantic search engines transform simple query into the semantic query which pass through the different phases. The future work includes implement this search and make a scalable semantic search engine.

## V. ACKNOWLEDGEMENT

The author sincerely thanks Associate Prof. Amit kumar, Dr. Sohan Garg, Mr. Vijay Maheshwari and Ms. Nidhi Taygi for their timely, invaluable and indispensable guidance and consequently encouragement shown towards the work group.

## REFERENCES


[1] Artem Chebotko, Shiyong Lu, and Farshad Fotouhi, *Semantics Preserving SPARQL-to-SQL Translation*, Data & Knowledge Engineering 2009.

[2] Chris Sherman, *Google Power: Unleash the full potential of Google*, Mc Graw Hill Osborne 2005.

[3] Fritz Schneider, Nancy Blachman, Eric Fredricksen, *How to do everything with Google*, Mc Graw Hill Osborne 2004.

[4] Guobing Zou Bofeng Zhang Yanglan Gan Jianwen Zhang, *An Ontology-based Methodology for Semantic Expansion Search*, Fifth International Conference on Fuzzy Systems and Knowledge Discovery 2008.

[5] Hakia Team, *hakia Semantic Search Technology making sense of the worlds information*. White paper Jan 2010.

[6] John Hebeler, Matthew Fisher, Ryan Blace, Andrew Perez – Lopez, *Semantic Web Programming*, Wiley India 2009.

[7] Karin K. Breitman, Marco Antonio Casanova, Walt Truszkowski, *Semantic Web: Concepts, Technologies and Applications*, Springer Verlag Berlin Heidelberg London 2007.

[8] Lizhen Li, Zhifeng Dong, Keming Xie, *Ontology of general concept for Semantic Searching*, Second International Conference on Computer Modeling and Simulation 2010.

[9] Luiz André Barroso, Jeffrey Dean, Urs Hölzle, *WEB SEARCH FOR A LANET: THE GOOGLE CLUSTER ARCHITECTURE*, IEEE Computer Society 2003.

[10] Maciej Janik and Krys Kochut, *Training-less Ontology-based Text Categorization*, European Conference on Information Retrieval (ECIR) March 30 2008.

[11] Ronen Feldman, James Sanger, *The text mining handbook: Advanced approach in analyzing unstructured data*, Cambridge University Press USA 2007.

[12] Siegfried Handschuh, Steffen Staab, *Annotation for the semantic web*, IOS Press Amsterdam 2004.

[13] Steffen Staab, Rudi Studer, *Handbook on ontologies*, Second Edition Springer Verlag Berlin Heidelberg London 2007.

[14] Thomas B. Passin, *Explores Guide to the Semantic Web*, Manning Publications Co. GreenWich 2004.

[15] Tim Finin, Yun Peng, R. Scott Cost, Joel Sachs, Anupam Joshi, Pavan Reddivari, Rong Pan, Vishal Doshi, Li Ding, *Swoogle: A Search and Metadata Engine for the Semantic Web*, Proceedings of the thirteenth ACM international conference on Information and knowledge management 2004.

[16] Yuangui Lei, Victoria Uren, and Enrico Motta, *SemSearch: A Search Engine for the Semantic Web*: IEEE Transactions on knowledge and Data engineering, VOL.19, NO. 2, FEBRUARY 2007.


## AUTHORS PROFILE

**Sajendra Kumar** is student of M.Tech. in Department of Computer Sc. and Information Technology, Shobhit University, Meerut. He received his B.Tech. Degree in Computer Science and Engg. in 2008. Currently he is working in Department of Computer Sc. and information technology in IIMT IET, Meerut as a Assistant Professor. His areas of interests are Search Engines, Semantic Web Crawlers and Mobile Ad-hoc Network.

**Ram Kumar Rana** pursuing M.Tech. in Department of Computer Sc. and Information Technology from Shobhit University, Meerut. He received his B.Tech. Degree in Computer Science and Engg. in 2008 from ABES Engg. Collage affiliated to UPTU. Currently he is working in Department of Computer Sc. in IIMT IET, Meerut as a Assistant Professor. His areas of interests are Search Engines, Web Crawlers and Semantic Web.

**Pawan Singh** received the B.E degree in Computer Science Engineering. From CCS University in 2002 and M.Tech in Information Technology in 2009 from GGSIP University Delhi. Presently, he is working as Associate Professor in Computer Sc. department in IIMT IET, Meerut. His areas of interests are Web Mining, Software Engg. and Crawlers.