**Indian Journal of Science and Technology,** *Vol 10(23), DOI: 10.17485/ijst/2017/v10i23/115568, June 2017*
*ISSN (Print) : 0974-6846*
*ISSN (Online) : 0974-5645*

# Semantic Search Engine

**Shilpa S. Laddha[1] and Pradip M. Jawandhiya[2]**

[1]Government College of Engineering, Aurangabad – 431005, Maharashtra, India; kabrageca@gmail.com
[2]PL Institute of Technology and Management, Buldana – 443001, Maharashtra, India

## Abstract

**Background/Objectives**: Traditional keyword-based search engines retrieve web pages by matching the exact tokens or words in the user query with the tokens or words in the web documents. This approach has many drawbacks. Synonyms or terms similar to tokens or words in the user query are not taken into consideration to search web pages. The keyword based search engine gives equal importance to all keywords whereas when user will enter query, he may have different levels of importance for different keywords in his opinion. To get the correct relevant result, users may need to enter several synonyms on his own to get the desired information which may otherwise result into the omission of many valuable web pages. Another problem is of information overloading. The traditional keyword based search engines make it very tedious for end user to locate the really useful information from a huge list of search results. Existing web is dominated by keyword based Search Engines which does not provide an appropriate mechanism to classify and locate the relevant search results. This leads to wastage of precious time of end user if he does not know the key terms which are utilized to index preferred correct pages. To resolve the above mentioned issues that the users face, in this paper we have proposed search techniques to develop Ontology based semantic search engine. **Methods/Statistical Analysis**: Ontology based Semantic Search Engine for the tourism domain is developed that understands the meaning of the user query and relatively provides the direct, precise and relevant result. Not only the user entered keyword based pages would be returned but also the pages that are appropriate enough with the meaning of the user entered keyword were also be returned by using the Ontological synonym dataset developed by using WordNet. **Findings:** Firstly, the Ontology Synonym set is constructed using WordNet and then the ontology synonym set parser is used to map the user defined query with the query prototype. By comparing the Query Similarity for every prototype, the service/sub domain with maximum query similarity is identified and the respective service is invoked. Also if the similarity is 100% the extra keywords are also considered to provide the relevant and precise results to the end user. Meta-processor will provide meta information about the URL. **Application/ Improvements:** The proposed Ontology based semantic search engine in tourism domain is an enhanced model. This model can be used by search engines, tourism industry professional and customers in getting better results of the searches undertaken. After testing in real world, the improvements can be worked out.

**Keywords:** Keyword Based Search Engine, Ontology, Semantic Search Engine, Query Controller, Web Search

## 1. Introduction

The amount of the information available on the Web is very huge and significantly increasing at lightning speed. The main aim of information retrieval[1] systems is to provide direct, precise and relevant information to the end user. With the advent of the Semantic Web, it can be tremendously optimized if machines could "recognize" the information of web pages. The working of present keyword based information retrieval[2] techniques are almost based *simply* on the presence of user entered keywords in web documents. In other words it is restricted to word matching. But, the main disadvantage of this search approach is that it is based on the matching or occurrence of words or tokens, rather than the semantic of words or tokens. In other words, it is not considering the meaning of the user entered words. This results into

**Author for correspondence*

information overloading which leads to wastage of user precious time to locate the required relevant information.

The most extensively used algorithms used by the popular keyword based search engines are the PageRank algorithm and the HITS algorithm[3]. It would be possible to optimize keyword[4,5] based traditional search due to the extensive availability of machine recognizable content on the Semantic Web. It would be possible to provide the search results with high precision and recall[2], if machines could "recognize" the content of web pages.

Ontology based Semantic Search Engine[6] belongs to the class of information retrieval and semantic web. The evaluation of the web search tools and techniques need to be optimized[7]. The existence of large volume of information on the web as well as the increase in the number of new inexperienced users leads to the complication for information retrieval; however the current search engines provide a capable way to browse the web. But these tools do not take into account the semantics driven by the terms in the user query and web document.

This paper proposes a new semantic based approach using ontology for the evaluation of information retrieval systems; the objective is to provide the user with the direct, relevant and precise information as per the entered query. The results will be primarily not dependent on the keywords in the user entered query but it will also consider the related pages relevant to the user by taken into account the semantics of the query. The users will get only the relevant and direct information for what they are searching. The irrelevant results are omitted.

## 2. Conceptual Framework

Semantic Ontology based Web search intends to enhance the preciseness and the relevancy of search by understanding the search intent. Semantic search systems includes various concepts like natural language queries[8], contextual search, generalized and specialized queries, location, synonyms, intent, and variation of the words, conceptual matching to provide relevant search results.

The most dominant keyword based search engine namely Google include some elements of semantic search. Google's PageRank algorithm is used to determine the relevancy. Instead of using various ranking algorithms, semantic search uses unique approach to strive for getting most accurate search results. Generally semantic search is carried out based on Ontology–done with help of conclusions dependent on OWL concepts, relations and logics[9]. Ontology can be best described as a systematic description of part-of relationships and entity dependencies. Also explained as, one which consists of a hierarchical explanation of vital classes (or concepts) in a given domain, adjunct with the explanation of the properties (of instances) of each and every concept. Knowledge representation languages mainly consists of Web Ontology Language (OWL) and Description Logics (DL).Web Ontology Language (OWL) functions primarily authoring ontologies and Description Logics (DL) which can be used to represent the terminological knowledge of an application domain in a structured and well-understood formal way[9]. Nowadays description logic has become a basis of the Semantic Web[10] for its role in the design of ontologies. We proceed as follows: 1. the input is unstructured data; 2. Matching user query with the system prototype; and 3. the generation of result.

## 3. Proposed System

We propose Ontology based Semantic Search Engine for the tourism domain that understands the meaning of the user query and relatively provides the direct, precise and relevant result as shown in Figure 1. Not only the user entered keyword based pages would be returned but also the pages that are appropriate enough with the meaning of the user entered keyword were also be returned by using the Ontological synonym dataset developed by using WordNet. Ontological Synonym set is the set which is developed using WordNet but we need to manually remove some words not required for the tourism domain and also need to add some relevant words which are not given by WordNet. The proposed system is focused on the specific tourism domain to reduce the search space and to improve the search results, since all web pages are already classified by domains. The main modules of the proposed system are:

- Query Controller,
- Query Prototype,
- Query Similarity Mapper,
- State Parser,
- City-State Parser,
- Ontological Synset Parser,
- Distance Parser,
- Service Finder and Caller,
- Service Modules,
- Metaprocessor, and
- URL Generator.

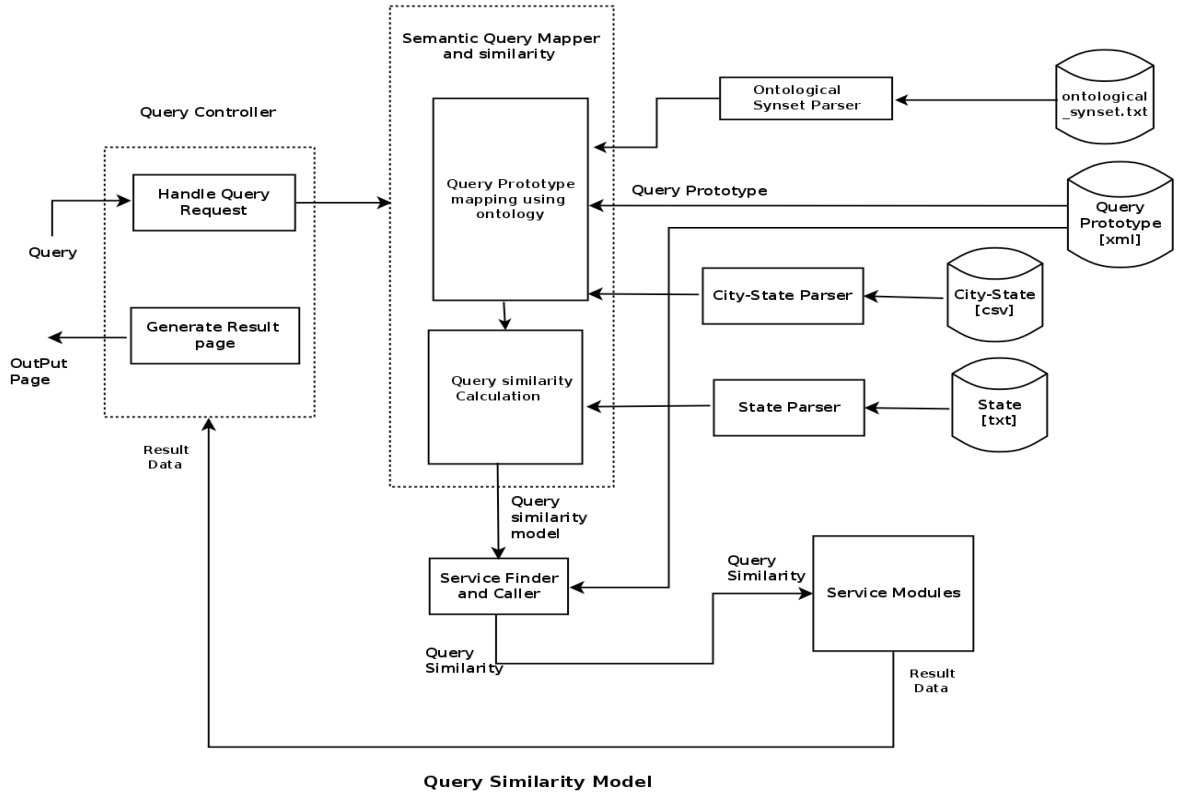


**Figure 1.** Proposed system.

## 3.1 Query Controller

This is the main module which provides the user interface through which the user will enter the query to search and the direct, precise and relevant results are provided to the user.

## 3.2 Query Prototype

In our proposed system we are concentrating on the tourism domain. So our domain is fixed. Our next target is to find out the sub-domain about which the query is asked by the user. To achieve this target, we will define different query prototypes. As we know that the different user may enter same query in many different ways but the pattern of writing the queries are limited so based on this concept, we identify all possible patterns for asking the query called Query prototypes with respect to that sub domain. Query Prototypes for the all sub domains need to be designed.

### *3.2.1 Query Similarity Mapper*

This module first categorizes the tokens from the user entered query into template tokens, keyword tokens and simple tokens. And then it generates the synonyms of template tokens using the Ontological Synonym parser used for mapping the defined query prototypes. Our proposed system is designed for the tourism domain. So it is expected that user will enter query related to tourism domain only. So the target of the Query similarity mapper is to identify the sub domain for which the user has entered the query.

For most of the queries of tourism domain, first thing is to identify the city name or state name or both. That is we need to check whether the user has entered the valid city name or state name and also we need to check whether the state or city name is written at right place or not. To solve this we need to maintain the data set of valid states and cities corresponding to state. Different parsers are designed like city-state parser, state parser, ontological synonym set parser etc.

The mapper maps the user entered query with the query prototypes using different available parsers and the similarity is calculated.

The prototype may match 100% or less. If it is matching 100% then the prototype with which it is matching

will return the sub domain and after identifying the sub domain the corresponding service will get invoked to provide the relevant result.

If it is not matching 100% then the system will check with which prototype it is matching maximum and it will be considered. And the prototype which matches maximum will return the sub domain which would be further used for invocation of the appropriate service to provide the desired relevant result.

Also there is the possibility that prototype is matching 100% but some extra keywords are present. So the system must consider these keywords while providing the result. The prototype with highest similarity provides the sub-domain of interest and accordingly the information is fetched and the precise and relevant results based on the additional keywords will be given to the user. In other word if omitting keywords 5 URL's are fetched. Then these fetched URL's were further refined considering the keywords and the relevant URL's would be provided to the user as result.

The similarity is calculated as follows:

Q = Input Query
Qp = Prototype of Query
similarity= S1 * S2 [0 - 1]
where,
S1 = N2/(N1 + N2)
S2 = N2/N3
N1 = No. of Keywords found in Q.
N2 = No. of tokens of Qp matched with Q.
N3 = No. of tokens of Qp

### 3.3 State Parser

State.txt file contain the list of all available states. So when the user will enter the query this state parser is used to check whether the state name entered by the user is valid or not.

### 3.4 City-State Parser

City-State.csv file contain the list of cities corresponding to state. So when the user will enter the query this City-State parser is used to check whether the city name entered by the user is valid or not.

### 3.5 Ontological Synonym Set Parser

OntologicalSynonymset.txt file contain the list of synonyms for the tourism domain. So when the user will enter the query this Ontological Synset Parser is used to match the prototype. The advantage of using the Ontological Synset is it will reduce the number of prototype list for the corresponding sub domain.

### 3.6 Distance Parser

The distance.txt files contain the distance between two cities which will be used to store the distance between two cities.

### 3.7 Service Finder and Caller

Based on the result of Query Similarity mapper the Service finder will find the sub domain and then invoke the appropriate service.

### 3.8 Service Modules

On invocation of the service module the results for the user entered query are fetched and provided to the user.

#### *3.8.1 Metaprocessor*

As the popular search engines provide the meta information and title of the fetched results Our proposed ontology based Semantic Search Engine also provide the meta information in the same fashion. Using page properties of web pages the meta information like title, status, basic web page information of the web pages are retrieved as background process and displayed to the user.

### 3.9 URL Generator

Based on the above processing, URL Generator will generate the actual URL and the result is provided to the user as shown in Figure 2.

## 4. Conclusion

In this paper, we have proposed to develop an Ontology based semantic search engine for tourism domain. Firstly, the Ontology Synonym set is constructed using WordNet and then the ontology synonym set parser is used to map the user defined query with the query prototype. By comparing the Query Similarity for every prototype, the service/sub domain with maximum query similarity is identified & the respective service is invoked. Also if the similarity is 100% the extra keywords are also considered to provide the relevant and precise results to the end user. Meta-processor will provide meta information about the URL.

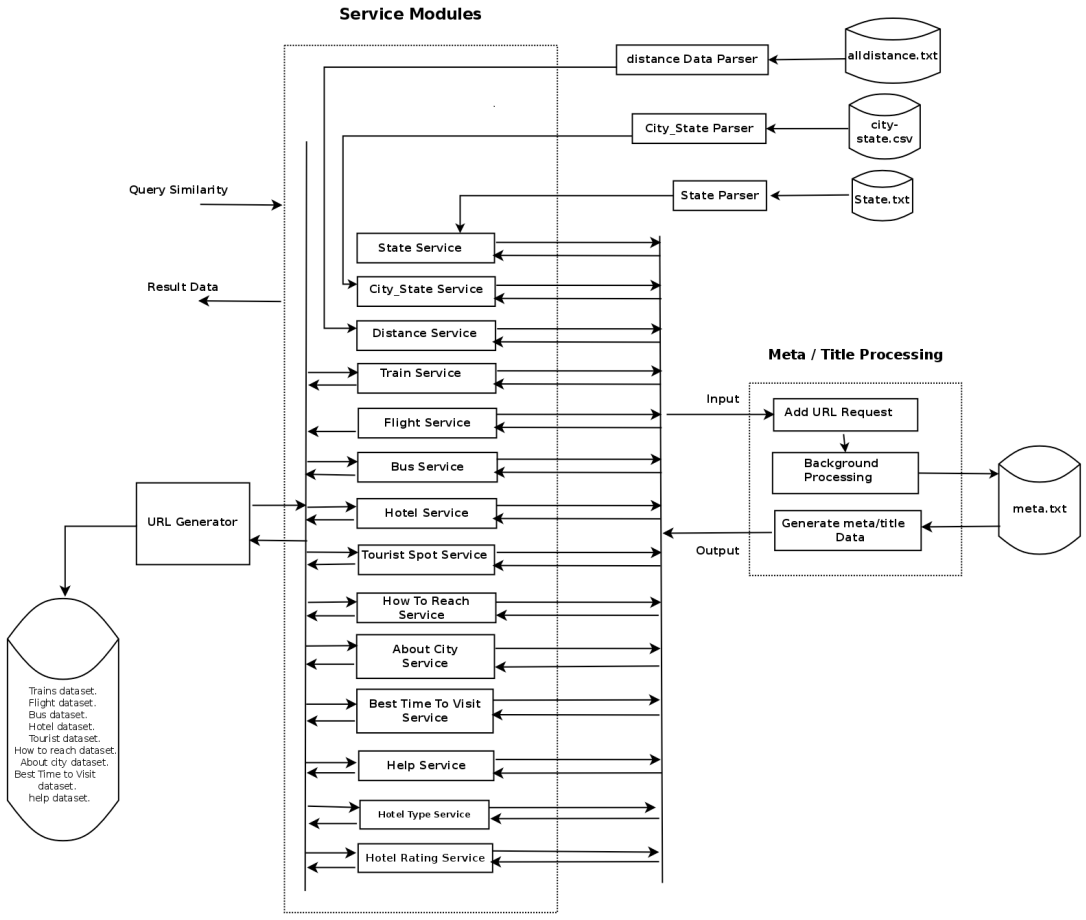


**Figure 2.** URL generator.

## 5. Acknowledgement

The authors wish to acknowledge the Government College of Engineering, Aurangabad, and Maharashtra for their support and technical help throughout this work.

## 6. References

1. Shah U, Finin T, Joshi A, Cost RS, Matfield J. Information Retrieval on the Semantic Web. In: CIKM '02: Proceedings of the Eleventh International Conference on Information and Knowledge Management; 2002. p. 461–68. Crossref
2. Mihalcea R, Moldovan D. Semantic Indexing using Wordnet Senses. In: Proceedings of the ACL-2000 Workshop on Recent Advances in Natural Language Processing and Information Retrieval, Association for Computational Linguistics; 2000. 11:35–45. Crossref
3. Shah D, Somaiya J, Nair S. Fuzzy Semantic Search Engine, International Journal of Computer Applications. 2014; 107(15):1–3. Crossref
4. Zhou Q, Wang C, Xiong M, Wang H, Yu Y. Spark Adapting Keyword Query to Semantic Search. In: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference; 2007. p. 694–707. Crossref
5. Wang H, Zhang K, Liu Q, Tran T, Yu Y. Q2 Semantic a Lightweight Keyword Interface to Semantic Search. In: ESWC; 2008. p. 584–98. Crossref
6. Celino I, Valle DV, Cerizza D, Turati A. Squiggle an Experience in Model-Driven Development of Real-World Semantic Search Engines. In: Luciano Baresi Piero Fraternali and Geert-Jan Houben. Editors ICWE volume 4607 of Lecture Notes in Computer Science; 2007. p. 485–90. Crossref
7. Uren V, Lei Y, Lopez V, Liu H, Motta E, Giordanino M. The Usability of Semantic Search Tools a Review, Knowledge Engineering Review. 2007; 22(4):361–77. Crossref

8. Shvaiko P, Euzenat J. Ontology Matching State of the Art and Future Challenges, IEEE Transactions on Knowledge and Data Engineering. 2013; 25(1):158–76. Crossref
9. Farrag TA, Saleh AI, Ali HA, Toward SWSs Discovery Mapping from WSDL to OWL-S Based on Ontology Search and Standardization Engine, IEEE Transactions on Knowledge and Data Engineering. 2013; 25(5):1135–47. Crossref
10. Davies J, Weeks R. Quizrdf Search Technology for the Semantic Web. Hawaii International Conference on System Sciences; 2004. p. 8. Crossref