Volume 3, Issue III, March 2015
ISSN 2320-6802

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

Survey on Semantic Search Engine using Ontologies

Prof. Sumedh Pundkar[1], Kapil Baheti[2]

[1]Head of Department of Computer Engineering
Usha Mittal Institute of Technology, SNDT University
sumedhpundkar@gmail.com
[2]Mukesh Patel School of Technology Management and Engineering, N.M.I.M.S.
JVPD Scheme, Vile Parle, Mumbai, India
kapilrbaheti@gmail.com

Abstract*: The search engines today are primarily focused on searching the web for user query on the basis of keywords. By using only the user entered key words, search engines are restricting themselves to deliver the results based on keywords only. The synonyms or similar words of same meaning are not taken into consideration. Also, all the keywords are treated equally. In this paper, a Search Engine will be described to help users by providing them satisfactory results which include synonyms of the entered keywords along with providing feature of personalization. Also, the proposed search engine will provide the best results and not the ones which are not authenticated, thereby increasing the user satisfaction. This is an ontology based search engine which will create ontologies by measuring the similarity between words. The similarity will also be assigned a numeric value.*
Keywords: *< ranking, search engine, searching algorithm, semantic search >.*

1. INTRODUCTION

All over the world, people use search engines for some or the other work. Searching the web has become the part of our daily life. This includes everything from searching a food recipe to searching the latest trends in different technologies. Though, searching the internet and user queries have increased but the satisfaction level of the users is still not up to the mark. Users still struggle to get the appropriate information on the internet. Getting the most accurate result for the searched query is a difficult task. Adding to the problem of the user, the number of results returned by the search engine are very large. It is practically impossible to go through all the links and get the answer. The basic problems of the users include:

- Displaying the results which are not relevant
- Large number of results making difficult for the user to browse
- Fetching the results which are not authorized
- User is unaware of the logic used to fetch the results for the query making it difficult for user to analyze the results
- Low Precision
- Low Recall

These problems can be observed on any search engine.
For example, if the search query is technical related to programming, then the top results are some blogging website. This makes the result less trust worthy for the user. Also, sometimes, user is not aware of the exact term needed to search. Thus, if exact keyword is not matching then the result may not be very accurate.
Search engines must not restrict themselves to keyword match only. The semantics of the words must also be taken into consideration. The logic should be fuzzy.
This paper compares existing systems providing such features and also proposes a search engine which will connect the information on existing web page with background ontological knowledge.

2. LITERATURE REVIEW

A semantic web search engine implementation needs to deal with the following aspects: developing a fuzzy ontology, natural language processing and crawler. In the paper "Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search", the author discusses about constructing a two-layered fuzzy ontology to organize terms that are elicited from WordNet[1]. WordNet is a large lexical database of English, built by Princeton University. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms, each expressing a distinct concept. In the two-layered fuzzy ontology, the first layer forms a domain hierarchy. Each domain contains a term lattice in the second layer.
Another technique that can be used to enhance semantic search is Natural Language Processing. The author "The HWS hybrid web search" discusses the use of an agent to process the natural language questions [2]. The agent employs LR (L means that the parser reads input text in one direction without backing up; that direction is typically Left to right within each line, and top to bottom across the lines of the full input file. The R means that the parser produces a reversed rightmost derivation) algorithm to complete the grammar parse for a given question. A given question is parsed to create a grammar tree which is then submitted to slot-filling. In contrast to the slot-filling of some nature languages, the agent employs slot-filling with grammar structure. Thus, in addition to pattern match, a given question can be processed based on grammar parser. The agent then employs Brill's part-of-speech tagger to analyse the words in a given question. The agent deletes certain

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

frequent words, acquired from the widely used WordNet, such as punctuation, preposition, article, or conjunction. It treats the rest of the question as keywords. In the meantime, the agent also employs WordNet to identify phrases as keywords [2]. Another algorithm is RK (Runge Kutta) algorithm to find the words that are closely related to the entered keywords and their synonyms [2].

An important pre-processing step to searching is crawling. Crawler is a program that searches a World Wide Web typically in order to create an index of data. In [3], the authors begin the discussion of the first component required for building the index, and thus for retrieving the raw RDF documents from the Web: that is, the crawler [3]. The crawler starts with a set of seed URIs, retrieves the content of URIs, parses and writes content to disk in the form of quads, and recursively extracts new URIs for crawling. In this paper, the author discusses the architecture and implementation of the Semantic Web Search Engine (SWSE). Following traditional search engine architecture, SWSE consists of crawling, data enhancing, indexing and a user interface for search, browsing and retrieval of information; unlike traditional search engines, SWSE operates over RDF Web data – loosely also known as Linked Data – which implies unique challenges for the system design, architecture, algorithms, implementation and user interface.

3. PROPOSED SCHEME

Search Engine that understands the meaning of the user query and relatively reasons him with the appropriate result is proposed. Not only the user entered keyword based pages would be returned but also the pages that is appropriate enough with the meaning of the user entered keyword. User will be provided with facility to mark pages which would be displayed first the next time user enters the related term.

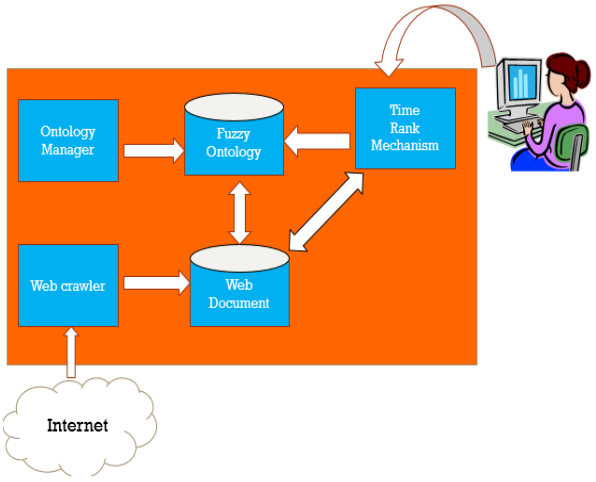

Fig. 2.5: Proposed System

Query Evaluator:

The Query Evaluator reduces each Semantic Web search query in an online step to a sequence of standard Web search queries on standard Web and annotation pages, which are then processed by a standard Web Search Engine, assuming standard Web and annotation pages are appropriately indexed. This block filters out the keywords from the user entered phrases and generates the synonyms to it. The Query Evaluator also collects the results and re-transforms them into a single answer which is returned to the user.

The search engine block takes the keywords from the query evaluator and checks it in the web document for the relevant pages which is returned to the inference system. Also, the annotations are used and algorithms are applied to generate the result.

Inference System:

Using background ontology Inference system adds all properties that can be deduced / induced from the ontology and returned to the web documents for other relevant pages.

Time Rank Mechanism:

A Time Rank Mechanism for ranking the pages which user searches can be implemented. This is a simple mechanism which ranks the pages based on the amount of time user has stayed on it previously. Higher the time, higher would be the rank of the page.

4. COMPARISION

The following conclusions can be inferred:

Search Engine	Crawler based	Ontology based	Semantic Web based
Metadata	Yes	No	Yes
Crawlers	Yes	No	Yes
Heuristics	No	Yes	Yes
Canonicalization	Yes	No	Yes
NLP efficiency	Medium	Medium	High

5. CONCLUSIONS AND FUTURE WORK

Semantic search on the Web, where standard Web pages are combined with background ontologies, on top of standard Web search engines and ontological inference technologies.

There is a formal model behind this approach. Generalized PageRank technique is used. Technique for processing semantic search queries [5] for the Web, consisting of an offline ontological inference step and

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

an online reduction to standard Web search queries. Implementation in desktop search along with very promising experimental results is expected. This search engine with time rank algorithm will be implemented. This mechanism will not only rank the pages based on its importance, but also on the basis of time user spends on each page.

REFERENCES

[1] Lien-Fu Lai, Chao-Chin Wu, Pei-Ying Lin, "Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search" in International Conference of fuzzy systems, June 2011

[2] Lixin Han, Guihai Chen "The HWS hybrid web search", in Information and Software Technology, pp687-695, September 2010

[3] E. Y. Chang "Confucius Proc.CIKM 2009, pp.3-3, 2009.

[4] P.T. Chang, K.C. Hung, K.P. Lin, and C.H. and its Intelligent Disciples." In Chang, a Comparison of Discrete Algorithms for Fuzzy Weighted Average, IEEE Transactions on Fuzzy Systems, pp.663-675, Oct. 2006.

[5]Aidan Hogan , Andreas Harth , Jürgen Umbrich , Sheila Kinsella , Axel Polleres , Stefan Decker, "Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine", in Web Semantics: Science, Services and Agents on the World Wide Web pp365-401, December 2011