INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

ONTOLOGY BASED WEB SEARCH

RUTUJA. A. AJMIRE[1], PROF. A. V. DEORANKAR[2], DR. P. N. CHATUR[3]

1. M. Tech. Scholar, Department of Computer Science and Engineering, Government College of Engineering, Amravati, Maharashtra, India.
2. Associate Professor, Department of Computer Science and Engineering, Government College of Engineering, Amravati, Maharashtra, India.
3. Head of Department, Department of Computer Science and Engineering, Government College of Engineering, Amravati Maharashtra, India.

Abstract

Information Retrieval Mechanisms From The Web Become Tedious As The Amount Of Information Is Growing Dynamically Day By Day. Web Search Is Done Either Keyword Based Or Snippet Based. In This Paper, It Is Observed That The Users Rarely Have The Patience To Navigate For Content Beyond The First Five Web Result Pages. This Paper Shows The Common Problems Of Existing Search Engine And Proposed A New Approach Based On The Ontological Search Engine. The Aim Of This Paper Is To Improvethe Web Search Result Using Semantic Similarity Which Will Improve The Quality Of Search Engines.

INTRODUCTION:-

As the information on the Internet explodes astonishingly, search engines play a more and more important role. However, due to the diversity of web users' search requests; it is important for search engines to improve their keyword-based search techniques. Web search is a key technology of the Web. An extension of the current Web by standards and Technologies that help machines to understand the information on the Web, to support richer discovery, data integration, navigation, and automation of tasks. Very recent joint initiative of Google, Microsoft, and Yahoo to addmeaning to Web pages to aid search. The development of a new semantic search technology for the Web is called semantic search on the Web, is currently a very hottopic, both in Web-related companies and in academic research.

Information Retrieval (IR) is a domain that is interestedin the structure, analysis, organization, storage,search and discovery of information. The challengeof IR is to find in the large amount of available documents;those that best fit the user needs. The evaluation offers is to measure its performance regarding to the

user needs, for this purpose evaluation methods widely adopted in IR are based on models that provide basis for comparative evaluation of different system effectiveness by means of common resources. In this context, several questions arise regarding the improvement of the information retrieval process, and the manner in which returned results are evaluated.[1]

Current search engines use a keyword matching approach. Using high frequency of keyword matching, the search engine returns a result to the user. These search engines has following pitfalls:First problem is, in Keyword matching search engine cannot guarantee the selected candidates have high correlation with the user query, given the different positions and meanings of the keywords. Another problem is most of the time web users cannot expresses his intent of search accurately using keywords. And thereafter the result return by the search engine does not satisfy the users need. Using Keyword matching, the top rank result contains high frequency of keyword. If the one of the page is more related to users query but does not contain high frequency of keyword, then the page may

lose its ranking position. This will leads to the awkward situation: spammers try their best to pollute the web document corpuswith term spamming tricks such as repetition, dumping and weaving.

RELATED WORK

Ontology definition:-

Ontology is an explicit specification of a conceptualization". Based on this definition, ontologies are used in the IR field to represent shared and more or less formal domain descriptions in order to add a semantic layer to the IRS.[2]

WordNet is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The purpose is twofold: to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications. WordNet distinguishes between nouns, verbs, adjectives and adverbs because they follow different grammatical rules. It does not include prepositions, determiners etc. Every synset contains a group of synonymous words or collocations.[3]

Use of Ontology in search engine:-

The need for using ontology for information retrieval (IR) has been explored by some approaches to better answer users' queries. In the information filtering process, this aspectwill be the subject of the contributionthat we present in this paper. The idea is touse ontology to add the semantic similarity between the user intent and the web return result. This can be done byextracting the query terms and their semanticprojection using the WordNet ontology onthe set of returned documents. The result ofthis projection is used to extract concepts relatedto each term, thus building a semanticrelation which will be the base of the web page ranking.

Existing Approach:-

When a user submits a query the retrieve results returned by search engines:-

1.   Check the information content of each returnedpage.

2.   Project the user query on the linguistic resource,the WordNet ontology in our case.

3.   Measure the results relevance by calculatingthe relevance degree of each of them.

4.   Generate a semantic rank of results accordingto the calculated relevance based on their degreeof in formativeness.

5.   Assign a score to each search engine based onits position in the new ranking.[2]

Calculation of Semantic Similarity using WorldNet

For Calculating semantic similarity, ontology must be specified first. For example, WordNet1 is a very famous and widely used ontology. Based on WorldNet, researchers have already put forward some semantic similarity formulas. For example, Leacock and Chodorow propose the following formula to compute the semantic between two concepts:

$$sim(\pi_1, \pi_2) = -\log \frac{len\,(\pi_1, \pi_2)}{2 * max_{\pi \epsilon wordnet} depth\,(\pi)} \quad (1)$$

where*len*(π1; π2) is the length of the shortest path between concept π1 and π2 in WordNet and *depth*(_) is the length of the path from to the root.[5]

Theoretically Proposed Approach:-

This paper present an approach in which, first a query is fire on a middleware which usesa backend search engine like Google, yahoo msn.e.g Google will return a result which is based on keyword matching. For improving the Web search result, first snippets are fetched from the N top result return by the Google search engine. Stemming and part of speech are performed on the snippets and then using ontology like WordNet a semantic similarity is calculated. Based on this semantic similarity relevance is calculated. Relevance is the semantic similarity between the specified web document and keyword or a query keyword. This approach also consider the original rank of each page return by the Google search engine and assign the importance score to each web page.Using this relevance value and importance score of each web page, this approach computes a new score. Based on this new scoremiddleware search engine performs a re-ranking of the original search result.

The main stages of the algorithm are as follows:-

1.   First step is to perform a Pre-processing on retrieved informationi.e snippet from search engine.

2.   Calculating semantic similarity score between snippets and topics.

3.   Calculating the importance Score for each document.

4.   Balanced methodis used for ranking web pages from semantic similarity and importance score.

5.   Apply the new ranking method on the web pages.

CONCLUSION

The paper presents an approach for improving the web search result.Ontology similarity is unquestionable important for Semantic Web search engine.This paper tries to propose an ontology similarity basedapproach to measure similarity between the users query and web page. This paper makes an attempt to propose a solution for information retrieval to retrieve higher occurrence of the concepts, within the web pages dynamically, Which reduces the effort made by the user for searching the required concept, In which semantic similarity provides a relevance score for a web page and a new score is formed for re-ranking which is more relevance to the user query.

REFERENCE

1. http://www.cs.ox.ac.uk/people/thomas.lukasiewicz/ssw11.pdf

2. AbdelkrimBouramoul, Mohamed-KhireddineKholIadi and Bich-LiênDoa "An ontology-based approach for semantic ranking of the web search engines results"

3. http://wordnet.princeton.edu/

4. Andre Freitas, Edward Curry, Sean O'Riain," A Distributional Approach for Terminological Semantic Search on the Linked Data Web" *SAC'12* March 25-29, 2012

5. Wang, S. Jiang, Y. Zhang, and M. Wang, "Re-ranking search results using semantic similarity", in Proc.FSKD, 2011, pp.1047-1051.

6. Zhi Lu, TianyongHao, Liu Wenyin "Categorizing and ranking search engine's results by semantic similarity" international conference on Ubiquitous information management and communication *2009.*