International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Optimized Content Retrieval Using Ontology Mapping

A. Leo Prabu, Dr. A. Valarmathi

Department Of MCA, Anna University (Bit Cmpus), Trichy, Tamil Nadu, India

ABSTRACT

Ontology similarity calculation and ontology mapping are important in information retrieval. This paper identifies the nature of semantic mismatches and essential elements that need to be taken into account for ontology mapping via the ranking learning algorithm. The similarity between two concepts that can be measured by comparing the difference between their corresponding ranking in terms of real numbers. Search optimization based on ontology provided by enhancing page to semantic –web on context aware pages.

Keywords : Semantic Mismatches, Ontology Mapping, Ranking, Semantic-Web, Context Aware.

I. INTRODUCTION

The Web Search involves searching through huge volume of web pages. Looking for the word that is requested by the browser. A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs). The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

Ontologies are the backbone technology for the Semantic Web and – more generally - for the management of formalized knowledge in the context of distributed systems. They provide machine-processable semantics of data and information sources that can be communicated between different agents (software and people). In Other words, information is made understandable for the computer, thus assisting people to search, extract, interpret and process information.

The problem of information retrieval is how to find the useful information for user's need from the mass of information. At present, the key reason that leads the low-quality of text information retrieval is lacking of semantics retrieval tools. This retrieval mode mainly based on word frequency analysis techniques. The disadvantage is that it spends a longer time on searching there are no suitable relevant categories for retrieval to the new emerging concepts. The essence of difficulties of the above-mentioned information retrieval is that the traditional information retrieval technologies are lack of knowledge processing capacity and understanding capacity. Ontology raise the accurate ratio and recall ratio of information retrieval. As the ontology has the ability to express concept semantics through the relationship between concepts. So, the reason for the ontology research so common. There isn't a standard communication of grammar or semantics between computer systems and people.

II. METHODS AND MATERIAL

A. The Optimized Content Retrival Using Ontology

MAPPING is an application developed using HTML and java script . It helps to relate the content between the user entered search and the web document provided by the user. It rates the documents provided as a related web page by the user and finally suggests the web page that is more related to the search provided by the user.

The URL provided by the user are analyzed and the mapping and reducing is performed to make the key ,

value pairs out of the content. The Words are made to be the Key and the Count is the value from the operation.

Then it is checked for relation in the document using the distribution of queried words in the document and the length between the instances of the words queried for. Finally, rating is produced for the page using the above process results.

Finally, The Suggestion is given to the user regarding the search results.

B. Scope And Objectives

- To get a search text or interested words from the user
- To get output URL's which the user feels to be a possible Resultset.
- To Create a system that analyze the URL's.
- Performs Map and Reduce operation on the content.
- Produces a Result set of Key, Value pairs with Key as Words and the Value as corresponding Count.
- Then Identify Distribution of interested words in the Content.
- Converting these into a rating for the File Content.
- Suggesting the better URL's which holds the interested words as well as better content semantically.

C. Literature Survey

1. **Finding and Ranking Knowledge on the Semantic Web ?**Li Ding, Rong Pan, Tim Finin, Anupam Joshi, Yun Peng, and Pranam Kolari at Proceedings of the 4th International Semantic Web Conference, Galway IE, November 2005, Springer-Verlag

Conventional web navigation and ranking models are not suitable for the Semantic Web for two main reasons:

(i) they do not differentiate SWDs from the overwhelming number of other web pages; and

(ii) they do not parse and use the internal structure of SWD and the external semantic links among SWDs. Hence, even Google, one of the best web search engines, can sometimes perform poorly in finding ontology. For example, the FOAF ontology (the most used one for describing a person) is not among the first ten results, when we search Google using the phrase "person ontology".

2. **A Survey on Semantic Web Search Engine** by G.Sudeepthi , G. Anuradha ,Prof. M.Surendra Prasad Babu IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012ISSN (Online): 1694-0814

Present World Wide Web is the global database that lacks the existence of a semantic structure and hence it becomes difficult for the machine to understand the information provided by the user in the form of search strings. As for results, the search engines return the ambiguous or partially ambiguous result data set. Semantic web is being developed to overcome the following main limitations of the current Web.

D. Limitations

- ✓ The web content lacks a proper structure regarding the representation of information.
- ✓ Ambiguity of information resulting poor interconnection of information.
- ✓ Unable to deal with enormous number of users and content ensuring trust at all levels.
- ✓ Incapability of machines to understand the provided information due to lack of a universal format.
- ✓ Automatic information transfer is lacking.

3. **International Journal for Advance Research In Engineering and Technology Wings to your Thoughts**….. Survey on Semantic Search Engine using Ontologies by Prof. Sumedh Pundkar, Kapil Baheti

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

4. The role of Foundational ontology for Domain ontology Engineering:

Giancarlo Guizzardi, NEMO Group, Federal University of Espírito Santo, Brazil

Ontologies are commonly used in computer science either as a reference model to support semantic interoperability,or as an artifact that should be efficiently represented to support tractable automated reasoning. Thisduality poses a tradeoff between expressivity and computational tractability that should be addressed in differentphases of an ontology engineering process. The inadequate choice of a modeling language, disregardingthe goal of each ontology engineering phase, can lead to serious problems in the deployment of the resultingmodel.

5. Web Semantics: Science, Services and Agentson the World Wide Web:

Petar Ristoski , Heiko Paulheim Data and Web Science Group, University of Mannheim, B6, 26, 68159 Mannheim, Germany

The tasks performed in that field are knowledge intensive and caoften benefit from using additional knowledge from various sources. Therefore, many approaches have beenproposed in this area that combine Semantic Web data with the data mining and knowledge discovery process. This survey article gives a comprehensive overview of those approaches in different stagesof the knowledge discovery process. As an example, we show how Linked Open Data can be used at various stages for building content-based recommender systems. The survey shows that, while there arenumerous interesting research works performed, the full potential of the Semantic Web and Linked OpenData for data mining and KDD is still to be unlocked

6. System Analysis

Existing System

The Existing System uses Keyword and the Time spent by a user on a web page.

Keyword is the word that is given by a user, for the web content made by the user. Such that any search about the keyword returns his web content in the search result or when the words used by the user is queried for in the search engine, the web page is provided to the browser.

Use of keywords for filtering by the search engine often leads to irrelevant search results compared to the one that the browser looks for. Instead of looking into content and context of browser query, only the keywords are considered to display the search result.

Time spent by the browser is used to calculate the page rank. The Time Spent by the browser is sometimes meaningless, as the browser may look for or search for particular text or portion or word in a page. Obviously, the browser may spend more time on the web page.

Using the page rank as the filter, the web content is provided to the browser based on the keywords or words just used in the web content.

System Architecture

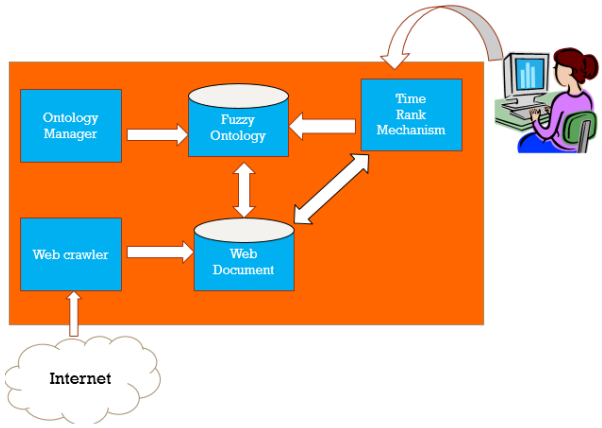


Figure 1. Existing system Architecture

III. RESULTS AND DISCUSSION

A. **Proposed System**

The Proposed System uses the content and the occurrence of instances and the relationship in the web content to produce a rating.

The rating provides insight to the browser about the web content's relation to the search query provided by the browser. From the rating, suggestion is given to the browser regarding the better content related to the search query provided by the browser from a set of results.

This requires less memory and acts as a filter .

Advantages:

- Real-time processing of data.
- It parses the html content to plain text.
- Then finds the Relationship between search query to the content of the resultset.
- The concern is provided to the user-end.
- Takes less memory.
- Easy to use.

B. System Architecture

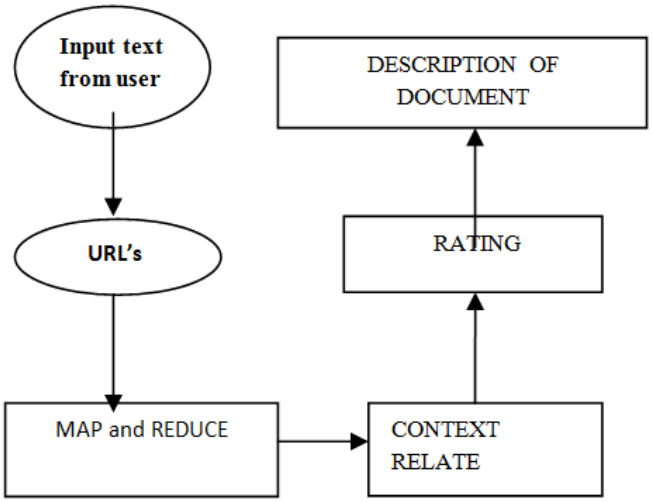


Figure 2. Proposed System Architecture

C. Module Description

MAP AND REDUCE

This module performs the operations of mapping and reducing the resultset web content based on the words as keys and the count of words as the value. It produces a list of key and values with words and frequency.

RELATIONSHIP FINDER

The Map and reduced resultset is used to perform the operation, finding relationship between the content and the search query text given. It tries to find the distribution of the words in the web content of each resultset web page for interested search query. The Relationship is produced as a mathematical metric.

Rating Generator

The relationship metric is used to produce a rating within the rating of 0-15 range. The 15 indicates the high relationship between the resultset and the text queried for by the browser. Whereas, the 0 rating indicates the resultset without any relation between the resultset and the text query.

Suggestion Maker

The Suggestion maker uses the rating to check the result sets with higher rating and suggests it to the browser. It displays the name of the webpage to the browser that is the best in the search resultset.

D. System Requirements

HARDWARE SPECIFICATION

- SYSTEM : INTEL PENTIUM 2.1 GHZ
- HARD DISK : 50 MB
- RAM : 1GB

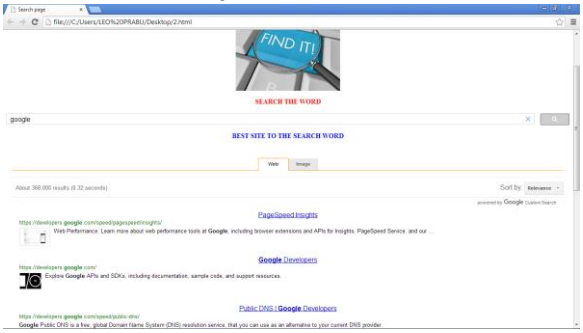
SOFTWARE SPECIFICATION

- OPERATING SYSTEM : WINDOWS 7
- FRONT END : HTML,Java script

E. Screenshot

Main Page:

Context Search By Relevant Context:



Context Search By Date:

F. System Testing

SOFTWARE TESTING

- Software testing is a process of executing a program or application with the intent of finding the software bugs.
- It can also be stated as the process of validating and verifying that a software program or application or product:
- Meets the business and technical requirements that guided it's design and development

SOFTWARE TESTING LEVELS

Testing levels are basically to identify missing areas and prevent overlap and repetition between the development life cycle phases. In software development life cycle models there are defined phases like requirement gathering and analysis, design, coding or implementation, testing and deployment. Each phase goes through the testing. Hence there are various levels of testing. The various levels of testing are:

1. Unit testing: It is basically done by the developers to make sure that their code is working fine and meet the user specifications. They test their piece of code which they have written like classes, functions, interfaces and procedures.
2. Component testing: It is also called as module testing. The basic difference between the unit testing and component testing is in unit testing the developers test their piece of code but in component testing the whole component is tested.

IV.CONCLUSION

In the Existing system, the resultsets are analysed once and not updated frequently considering the changes as the cached memory helps to retrieve the URL's without processing further, in the search results.

The memory required for processing goes to higher extent as the page ranking algorithm and database overhead are high in terms of saving resultsets for analysis in the search engines.

The proposed system performs the rating and suggestion in a faster rate and the time is saved for the browser in terms of irrelevance avoidance in the search.

V. REFERENCES

[1]. Semantic Web Portal Survey- the state of the art survey by Holger Lausen, Michael Stollberg, Rubén Lara Hernández,Ying Ding, Sung-Kook Han, Dieter Fensel

[2]. Semantic Web- An extensive Literature Review byRashmi Bakshi, Assistant Professor, VSIT,VIPS Abhishek Vijhani, MCA Student, JIMS

[3]. Semantic Integration: A Survey Of Ontology Based Approaches by Natalya F. Noy

[4]. The Use of Semantic Web Technologies for Decision Support - A Survey by Eva Blomqvist

[5]. Quality Assessment for Linked Data: A Survey

[6]. A Systematic Literature Review and Conceptual Framework by Amrapali Zaveri a;, Anisa Rula b, Andrea Maurino b, Ricardo Pietrobon c, Jens Lehmann a and Sören Auer