

EDUCATION BASED SEMANTIC SEARCH ENGINE USING WEB SERVICES

¹Asst. Prof. Parminder Singh, ²Er.Sandeep Kaur
¹SCIT, Lovely Professional University Phagwara
²Department of Computer Science & Engineering, LPU
¹ps_parminder@rediffmail.com
²Sandeepkaur_lpu@live.com

Abstract— Search Engines available in the current scenario provides the content in number of fields whether it is related to science, technology, research or education but there is a scope of improvement which can lead to a search engine for a specific filed such as search engine specifically for Education purpose. A search engine hardly provides essential educational related content relevant to the student's search query. Due to the text and their relationship of the request made by student is not analyzed up to full extent in generic search engines. So here the need for an education level optimized search engine arises. The area of education based optimized search engine has a great scope and it is not taken into consideration by any educational institute. The proposed topic focuses on university level search engine by using web services. Analyze query both syntactically and semantically as it will be used in SIEU (Semantic Information Extraction in University Domain).In SIEU web services related to students material not mentioned in the paper what they exactly used and location based mapping is also not there, it only uses Google search API. Web services like query for bus rental services, query for railway reservation, query regarding future scope of research papers, query for department exam papers etc. Location based mapping will be an additional feature to be covered in this paper. In future work will be done on location based mapping to implement it dynamically.

Keywords— Semantic web, University web services, ontology, WordNet API, Parser, Mapping system

INTRODUCTION

Introduction about modules used in the proposed system are:

A. Semantic Web

The term “Semantic Web” refers to Web of linked data. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. Semantic web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF and may more as recommending by W3C [15].

B. Parser

Meaning of Parser comes from Latin *pars* (*ōrātiōnis*), meaning lexical category (noun,verb).Parsing formally known as syntactic analysis which is the process of analyzing a text, made of a sequence of tokens (for example, words), to determine its grammatical structure with respect to a given formal grammar [12]. Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages.

Stanford Natural Language Processing Group develops statistical parser.

These parsers like natural language parser and probabilistic parsers. A natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences. The Stanford dependencies provide a representation of grammatical relations between words in a sentence [11].

C. WordNet API

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of synsets (a set of one or more synonyms), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. WordNet's structure makes it a useful tool for natural language processing. Synonyms means words that denote the same concept and are interchangeable in many contexts are grouped into unordered sets (synsets).

WordNet.Net was originally created by Malcolm Crowe, and it was known as a C# library for WordNet. After WordNet 2.1 was released, Troy released his version of Malcolm's library as an LGPL library known as WordNet.Net (with permission from Princeton and Malcolm Crowe, and in consultation with the Free Software Foundation), which was updated to work with the WordNet 2.1 database [14].

D. Ontology Construction

Ontology represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about entities. Ontology yield shared vocabulary which models a domain with the definition of objects and/or concepts and their properties and relations. Ontology components includes: Individuals, Classes, Attributes, Relations, Functions, events etc. A domain ontology (or domain-specific ontology) models a specific domain, which represents part of the world. Particular meanings of terms applied to that domain are provided by domain ontology. For example, the word *web* has many different meanings. Ontology about the domain of computer meanings is "web sites" or "web services" [13].

E. University Web Services

University Web Services (UWS) includes web services related to students query. Web Services like query for bus rental services, query for railway reservation, query regarding future scope of research papers, query for department exam papers, query for borrowing books from seniors.

F. Mapping System

Mapping system is a kind of system that can be designed in a static way by using university map. In this paper LPU map will be designed to get locations in an optimized way. It will be useful for fresher students for example, new student have problem where 32-block will be located, where admission block located, where campus café located, where Photostat machines located in blocks.

REVIEW OF LITERATURE

As we know that semantic information retrieval has become the core part of any search engine. Sanjay Kumar Malik ET. al (2010), Developing a University Ontology in Education Domain Using Protégé for Semantic Web [8] deals with constructing ontology. As we know by reading this paper that Ontology plays an important role in information exchange, use and re-use knowledge, shared and common understanding of a domain that can be communicated between people and across application systems which is the goal of semantic web. Objective of ontology is that the data capture by it can be machine understandable data on the current human-readable web. Ontology development is the most fundamental step which is the objective of this paper and been illustrated with an education domain example.

Joel Booth, Barbara Di Eugenio, Isabel F. Cruz, Ori Wolfson (2009), Query Sentences as Semantic (Sub) Networks [2] deals with describing process for representing queries in natural language as semantic networks. This means syntactic analysis of query is done by parsing query using

Stanford parser. But this paper does not make use the WorldNet gloss information. This algorithm mainly consist of parsing, domain model, candidate set generation, candidate selection, query target identification etc.

Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini (2009), A Relation-Based Page Rank Algorithm for Semantic Web Search Engines [1] they propose a relation-based page rank algorithm to be used in conjunction with semantic Web search engines that simply relies on information that could be extracted from user queries and on annotated resources. Relevance is measured as the probability that a retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition. In order to rank results, most of the existing solutions need to work on the whole annotated knowledge base. John Garofalakis, Panagiotis Kappos, and Dimitris Mourtouloukos (1999), Web Site Optimization Using Page Popularity [3] a link-editing algorithm based on relative page popularity can automatically revise a Web site's page structure to create a substantially more effective information scheme. One easy way to estimate a page's popularity is to count the accesses to this page based

exclusively on a given log file. Here considered only cases where the objective is to make it easier for a user to find the requested data: the faster the access, the better the organization of the Web server. But for commercial material the best organization may be one that achieves the highest AA (number of absolute accesses per page) for those pages with a commercial interest.

Swathi Rajasurya , Tamizhamudhu Muralidharan , Sandhiya Devi,Dr.S.Swamynathan ,Semantic Information Retrieval Using Ontology in University Domain [10] the objective here is to design, develop and implement a semantic search engine- SIEU (Semantic Information Extraction in University Domain) confined to the university domain.SIEU uses ontology as a knowledge base for the information retrieval process. The level of accuracy will be enhanced since the query is analyzed semantically. The system will be of great use to the developers and researchers who work on web. The Google results are reranked and optimized for providing the relevant links.

Mourad Ouzzani and Athman Bouguettaya (2004), Efficient Access to Web Services [5] For Web services to expand across the Internet, users need to be able to efficiently access and share Web services. Queries can be evaluated with different web services operations. Because efficiency plays a central role in these evaluations, so propose is to optimized query model based on aggregating the quality of Web service (QoWS) parameters of different Web services. The model adjusts QoWS through a dynamic rating Scheme and multilevel matching in which the rating provides an assessment of Web services' behavior. Multilevel matching allows the expansion of the solution Space by enabling similar and partial answers.

Sean W. M. Siqueira, Maria Helena L. B. Braz, Rubens Nascimento Melo (2003), Web Technology for Education

and Training [9] deals with a generic architecture for educational and training systems from the software development point of view. This architecture is discussed according to web technology, enabling a better understanding of the involved technological aspects of educational and training systems. The web-based education and training architecture enables a better understanding of the involved technological aspects of educational and training systems, allowing the identification of possible components that may be object of a standardization process and thus facilitates elearning software development and use.

Rupa R, M M M K Varma(2012), Ontology Mining for Personalized Web Information Gathering [6] it is challenging to use semantic relations of “kind-of”, “part-of”, and “related-to” and synthesize commonsense and expert knowledge in a single computational model. In this, a personalized ontology model is proposed attempting to answer this challenge. A two-dimensional (Exhaustively and Specificity) method is also presented to quantitatively analyze these semantic relations in a single framework. The model extracts world knowledge from the LCSH system and discovers expert knowledge from a large volume of specified information in the library catalogue.

S.Antoinette Aroul Jeyanthi, Dr.Latha Parthiban(2012), Ontology based Data Mining Approach – A Survey[7] It provide survey of ontology based data mining approaches in the past few years. The data mining process comprises of a sequence of steps ranging from data cleaning, data selection and transformation, to pattern evaluation and visualization. One of the important challenges in data mining is to extract interesting knowledge which is useful for expert users. The use of prior knowledge may significantly enhance the discovery of interesting patterns by considering the interestingness according to expert beliefs. Ontology’s are used to communicate domain knowledge and for discovering patterns.

Khurram Naim Shamsi, Zafar Iqbal Khan (2012), Development of an E-Learning System Incorporating Semantic Web [4] the Semantic Web-Based model for our e-learning system taking into accounts the learning environment at Saudi Arabian universities. The proposed system is mainly based on ontology-based descriptions of content, context and structure of the learning materials. It further provides flexible and personalized access to these learning materials. The framework has been validated by an interview based qualitative method.

PRESENT WORK

Present work consist of scope, objectives ,methodology as follows:

A. Scope of Study

- Students belonging to different departments can retrieve university, institute, and college’s related information in an easier way.

- Queries related to department papers for fresher students can be solved in an efficient way.
- Books taking by freshers from their senior’s service also provided by this proposed system.
- Location based mapping i.e. query related to location can be solved in a static and optimized way.
- Search related to university related information (news, announcements, events etc) can be done in an optimized way.
- Research papers future scope and their guide’s information also available in this proposed system which helps to use their time efficiently.
- Results can be provided with confidentiality and accuracy.
- Results also filter and ranked to get optimized results only not irrelevant results.

B. Problem Definition

There are many good search engines that are widely used for searching and retrieving user’ query in an optimized way. Due to overload of information, continuously failed to locate optimized result. In education field, data will be in gigabytes or terabytes related to universities, institutes, colleges, schools etc. in their systems. There is a problem, illustrated below:

Firstly, some of the challenges for the Semantic Web include vastness, vagueness,uncertainty, inconsistency.

Secondly, web services related to content are not provided by generic search engines automatically.

Thirdly, location based query not get an optimized result in a semantic search i.e. not exact location can be identified in an optimized way.

Fourthly, Google search API used by SIEU has some problems like there is a risk because Google reserves the rights to turn it off at any time. They will also limit the number of searches we perform per day (100 queries per day).

To overcome these problems proposed an optimized search engine which provides relevant and optimized result with confidentiality. The semantics of query is analyzed with following procedure:

- Student query is initially analyzed by parsing.
 - Check whether it is location based query or text based query.
 - If location based query:
 - ✓ Mapping system to map the location what student wants.
 - If text based:
 - ✓ Finding the domain words by using word net API which will provide the relevant result.

- ✓ Take description from ontology construction.
- ✓ Checking which web service student want in their query if not then skip this step.
- ✓ This optimized result then extracts relevant links using extraction.
- ✓ At last filtering the result then ranked it and show to user with confidentiality.

C. Objectives of Study

Research Question: The aim of the proposed system is to design and implement an education based search that provides optimized results related to student query.

It gives more appropriate information related to university, institutes, colleges etc i.e. education based things only because a lot of search engine that are highly optimized for giving good results according to user query. But they failed to give an apt results related to student query under university domain. System removed irrelevant search by filtering it and ranking the results after filtering with confidentiality. Web services and location based mapping will be used to get a complete system.

D. Research Methodology

In this paper action research will be done. Action research is a process in which participants examine their own educational practice systematically and carefully, using the techniques of research. Action research started with a particular problem study based and find out the solution for that problem which gives better results than previous study of that data. Action research is intervention researches which has a primary purpose to identifying the research based problem then find the best possible solution to solve it or the experimental hypothesis to be tested.

Architecture designed for problem is as follows:

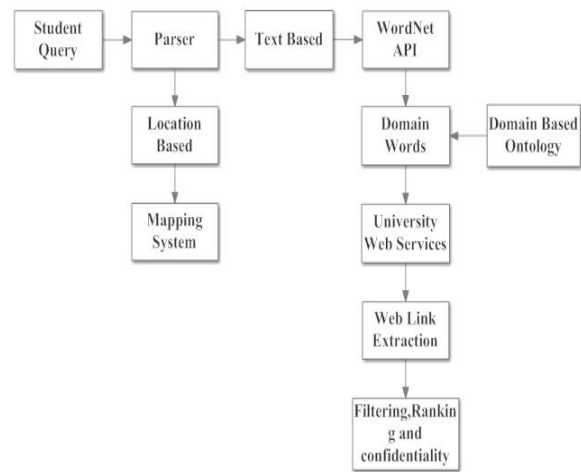


FIGURE 1-ARCHITECTURE OF PROPOSED SYSTEM

- First user query will be made on search engine.
- Parsing the query to get good results.
- Checker will decide whether query is text based or location based.
- If location based then mapping system will give the optimized result.
- If text based then using Word Net API which is known as database of English.
- When domain words related data will be fetching from ontology after getting domain words from WordNet API.
- Provide essential university related web services for data.
- Extracting links related to data student want.
- Filtering the links for an optimized search then ranked it with confidentiality.

FUTURE WORK

In this paper we propose an architecture using semantic web technology and university related web services. The future work related to this paper is that we implement it in an efficient and optimized way and compare their results with generic search engines. So that our proposed system will be beneficial for students and their work.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my advisor Mr. Parminder Singh for the continuous support of my thesis study, for his patience, motivation, enthusiasm, and immense Knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my M.Tech study.

REFERENCES

1. Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini," *A Relation-Based PageRank Algorithm for Semantic Web Search Engines*", 2009, IEEE Transactions on Knowledge and Data Engineering, Vol.21, no.1, pp: 123-136.
2. Joel Booth, Barbara Di Eugenio, Isabel F. Cruz, Ouri Wolfson," *Query Sentences as Semantic (Sub) Networks*", 2009, IEEE International Conference on Semantic Computing, Chicago, USA, pp: 89-92.
3. John Garofalakis, Panagiotis Kappos, and Dimitris Mourloukos," *Web Site Optimization Using Page Popularity*", JULY • AUGUST 1999, IEEE InternetComputing, University of Patras, Greece.
4. Khurram Naim Shamsi, Zafar Iqbal Khan," *DEVELOPMENT OF AN ELEARNING SYSTEM INCORPORATING SEMANTIC WEB*", International Journal of Research in Computer Science eISSN 2249-8265 Volume 2 Issue 5 (2012) pp. 11-14 www.ijorcs.org, A Unit of White Globe Publications doi:10.7815/ijorcs.25.2012.042.
5. Mourad Ouzzani and Athman Bouguettaya,"*Efficient Access to Web Services*", MARCH • APRIL 2004, Published by the IEEE Computer Society.
6. Rupa R, M M M K Varma," *Ontology Mining for Personalized Web Information Gathering*", July- August 2012,vol 3,No.4, International Journal of Advanced Research in Computer Science.
7. S.Antoinette Aroul Jeyanthi, Dr.Latha Parthiban," *Ontology based Data Mining Approach – A Survey*", July- August 2012, vol 3, No.4, International Journal of Advanced Research in Computer Science.
8. Sanjay Kumar Malik ET. al., "*Developing a University Ontology in Education Domain using Protégé for Semantic Web*", 2010, International Journal of Engineering Science and Technology, Vol.2, no.9, pp: 4673-4681.
9. Sean W. M. Siqueira, Maria Helena L. B. Braz, Rubens Nascimento Melo," *Web Technology for Education and Training*", 2003, Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03).
10. Swathi Rajasurya , Tamizhamudhu Muralidharan , Sandhiya Devi,Dr.S.Swamynathan," *SEMANTIC INFORMATION RETRIEVAL USING ONTOLOGY IN UNIVERSITY DOMAIN*", Department of Information and Technology, College of Engineering,Guindy,Anna University,Chennai-25.
11. *Stanford Parser:A Statistical Group* . Retrieved november 23, 2012, from stanford.edu.
12. *Parsing*. Retrieved November 23, 2012, from wikipedia.org:http://en.wikipedia.org/wiki/Parsing
13. *Ontology (information science)*. Retrieved november 23, 2012, from wikipedia.org:http://en.wikipedia.org/wiki/Ontology_%28information_science%29#Domain_ontologies_and_upper_ontologies
14. Troy Simpson, T. D. (2010, Feb 8). *WordNet-based semantic similarity measurement*. Retrieved November 23, 2012, from code project.com:http://www.codeproject.com/Articles/11835/WordNet-based-semantic-similarity-measurement
15. *Semantic web*. Retrieved november 23, 2012, from W3C.org:http://www.w3.org/standards/semanticweb/