International Journal of scientific research and management (IJSRM)
||Volume||2||Issue||5 ||Pages|| 871-875 ||2014||
Website: www.ijsrm.in ISSN (e): 2321-3418

Profile Based Web Search Using Rule Ontology

S.Amithapatchan[1], B.Sivananthan[2]

[1] PG Student,
Affiliated to Anna University Chennai, Dept. of Computer Science and Engineering
Gnanamani College of Engineering,
Namakkal, India.
Amithapshri05@gmail.com

[2] Asst. Prof Dept. of CSE
Dept. of Computer Science and Engineering
Gnanamani College of Engineering,
Namakkal, India
sivananthan.b@gmail.com

Abstract: *Web search engines help users find useful information on the World Wide Web (WWW). However, when the same query is submitted by different users, typical search engines return the same result regardless of who submitted the query. Generally, each user has different information needs for his/her query. Therefore, the search results should be adapted to users with different information needs. In this paper, we first propose a user profile based web search using rule acquisition through ontology. Rule is build based on user profile details, then the search result from engine is parsed based on this rule. Rule acquisition procedure using rule ontology RuleToOnto has been applied on the search result. The rule acquisition procedure consists of the rule component identification step and the rule composition step. And result is parsed and performed semantic matching and displayed to user according to the rule build up on user requirement. Only the information related to user profile will get displayed to the user and all other search results will be truncated.*

Keywords: Profile Based Web Search, Rule Ontology, Rule Acquisition, Semantic Matching.

1. Introduction

The World Wide Web is a diverse source of information for billions of Web users. This variety provides a significant challenge in enabling a user‘s access to information because large portions of the Web may fall outside of a particular user‘s overall interests, comprehension level, or comprehension within a particular domain. It has become increasingly difficult for users to find information on the WWW that satisfies their individual needs since information resources on the WWW continue to grow. Under these circumstances, Web search engines help users find useful information on the WWW. However, when the same query is submitted by different users, most search engines return the same results regardless of who submits the query. In general, each user has different information needs for his/her query. For example, for the query “Java,” some users may be interested in documents dealing with the programming language, “Java,” while other users may want documents related to “coffee.” Therefore, Web search results should adapt to users with different information needs.

Knowledge is an essential part of most Semantic Web applications and ontology, this is a formal explicit description of concepts or classes in a domain of discourse [9], is the most important part of the knowledge. However, ontology is not sufficient to represent inferential knowledge. This is because ontology-based reasoning has limitations compared with rule-based reasoning, even though ontology-based reasoning with description logic is a popular issue of the Semantic Web. Many attempts have been made at knowledge acquisition in order to obtain enough knowledge for Semantic Web applications. Ontology learning, which refers to extracting conceptual knowledge from several sources and building ontology from scratch, enriching, or adapting an existing ontology.

Rule acquisition is as essential as ontology acquisition, even though rule acquisition is still a bottleneck in the deployment of rule-based systems [8]. This is time consuming and laborious, because it requires knowledge experts as well as domain experts, and there are Communication problems between them. Let us suppose that, if they have to acquire rules from several sites of the same domain. The sites have similar Web pages explaining similar rules from each other. A comparison-shopping portal can be an example. The comparison of simple data such as book prices does not need rules, but delivery cost calculation with various options and applying free shipping rules and return policies needs rules.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

2. Related Works

2.1 Search Engine

A web search engine is designed to search information on World Wide Web. The World Wide Web (WWW) contains large amount data. Even in the day-to-day process we use search engine frequently. This is the reason for the increasing

popularity and necessity of search engine. Search engine crawls the web and parses the WebPages and the URL's, which contains the user query keywords matched with the content of that specific webpage, are given as a results to the user.

Search engine helps to find information stored on a computer system such as the World Wide Web, or a personal computer.' Search engines use regularly updated indexes to operate quickly and efficiently. Basically through search engine we can access WebPages efficiently and easily.

2.2 Hyperlink Based Personalized Web Search

The field of Web information retrieval focuses on hyperlink structures of the Web, for example with Web search engines such as Google and the CLEVER project [6]. To address several problems with these engines, i.e., (1) the weight for a Web page is merely defined, and (2) the relativity of contents among hyperlinked Web pages is not considered, we proposed several approaches to refining the TF-IDF scheme for Web pages using their hyperlinked neighboring pages [4, 5]. In personalized Web searches, the hyperlink structures of the Web are also becoming important. The use of personalized Page Rank to enable personalized Web searches was first proposed in [5], where it was suggested as a modification of the global Page Rank algorithm, which computes a universal notion of importance of a Web page. The computation of (personalized) Page Rank scores was not addressed beyond the original algorithm. Haveliwala [7], used personalized Page Rank scores to enable "topic sensitive" Web searches. Experiments in this work concluded that the use of personalized Page Rank scores can improve a Web search. However, no experiments based on a user's context such as browsing patterns, bookmarks, and so on were conducted. Therefore, it is not clear if search results obtained using this approach actually satisfies information needs that is different user by user.

2.3 Ontology Learning

The algorithm builds the taxonomy with linguistic analysis and identifies relevant candidates of classes and instances based on statistical analysis. The Ontologies are composed from automatically obtained taxonomies. Some approaches used somewhat different learning methods for identifying instances and relations. For example, WEBfiKB [10] used Bayesian and First Order Logic learning methods, and Sanchez and Moreno [11] suggested a knowledge acquisition technique that built ontologies with a multi agent system. TextOntoEx defined and used semantic patterns to identify not only simple taxonomic relations but also non taxonomic conceptual relations (e.g. causes, caused by, treat, contain, etc.).The approach using the Multiple Classification Ripple-down Rules (MCRDR) methodology, in ontology, learning is somewhat similar to this approach in its framework. They use a graph search algorithm instead of MCRDR to extract inference rules. In addition, they accumulate the rule ontology by repeating rule acquisition across different sites

2.4 Rule Acquisition

Learning by examples is a very different concept from rule acquisition from texts, which imply IF-THEN rules. Therefore, it is impossible to apply those methods in this paper problem, because their target is structured data while there target is unstructured text. Compared to rich studies of ontology learning, rule acquisition from the Web is not popular. Moreover, acquired rules are limited to a certain purpose and type [12], and are not general-purpose inference rules. Most significantly, studies about automatic rule acquisition from text are quite rare while there are some studies that discover rules from existing data.

Even though these can be separated by the Related Works section into ontology learning and rule acquisition, the extraction of rules is one of the research areas in ontology learning, because the inference rules could be an outcome of ontology learning. The term "inference rule" means the relationship between two phrases in entailment rule approaches. Moreover, the rules are generated with statistical methods by calculating frequencies and probabilities while the rules are directly generated from the Web in this approach.

The eXtensible Rule Markup Language (XRML) approach is a framework for extracting rules from texts and tables of Web pages. The core of the XRML framework is rule identification, in which a knowledge engineer identifies various rule components such as variables and values from the Web pages with a rule editor. The effectiveness of the rule acquisition procedure of the XRML approach depends on the rule identification step, which also depends on the large amount of manual work done by the knowledge engineer..

3. Basic Ideas of Ontology In Rule Acquisition

Ontologies capture the structure of the domain, i.e. conceptualization. This includes the model of the domain with possible restrictions. The conceptualization describes knowledge about the domain, not about the particular state of affairs in the domain. In other words, the conceptualization is not changing, or is changing very rarely. Ontology is then specification of this conceptualization - the conceptualization is specified by using particular modeling language and particular terms. Formal specification is required in order to be able to process ontologies and operate on ontologies automatically. Ontology describes a domain, while a knowledge base (based on ontology) describes particular state of affairs. Each knowledge based system or agent has its own knowledge base, and only what can be expressed using ontology can be stored and used in the knowledge base.

3.1 To Expanding an Ontology

To developing, Ontology includes,

- Defining classes in the ontology,
- Arranging the classes in a taxonomic (subclass–super class) hierarchy,
- Defining slots and describing allowed values for these slots,
- Filling in the values for slots for instances.

3.2 OntoLT

The OntoLT approach, is available as a plug-in for the widely used Protégé ontology development tool , which enables the definition of mapping rules with which concepts (Protégé classes) and attributes (Protégé slots) can be extracted automatically from linguistically annotated text collections. A number of mapping rules are included with the plug-in, but alternatively the user can define additional rules.

OntoLT provides a precondition language, with which the user can define mapping rules. Preconditions are implemented as XPATH expressions over the XML-based linguistic annotation. If all constraints are satisfied, the mapping rule activates one or more operators that describe in which way the ontology should be extended if a candidate is found.

3.3 Semi-Automatical Ontology Acquisition Method

The process of acquiring ontology can be divided into two stages: acquiring ontological structure and acquiring ontological instances. In the stage of acquiring ontological structure, it is necessary to capture information about database schema firstly, and then based on the information ontological structure can be constructed. Since the constructed ontological structure may not be ideal, the evaluation and refinement about it is needed.

SOAM, which consists of four steps.
Step1.Capture the information about relational database schema;
Step2.Acquire ontological structure according to the database schema information;
Step3 Refine the obtained ontological structure;
Step4.Acquire ontological instances based on refined ontological structure.

SOAM tries to balance the cooperation between user contributions and machine learning in order to ensure the quality of constructed ontology and improve the automatic degree of acquiring process.

3.4 Text2Onto

Text2Onto is a framework for ontology learning from textual resources. Three main features distinguish Text2Onto from there earlier framework TextToOnto as well as other state-of-the-art ontology learning frameworks. First, by representing the learned knowledge at a meta-level in the form of instantiated modeling primitives within a so-called Probabilistic Ontology Model (POM), they remain independent of a concrete target language while being able to translate the instantiated primitives into any knowledge representation formalism.

Second, user interaction is a core aspect of Text2Onto and the fact that the system calculates the condense for each learned object allows to design sophisticated visualizations of the POM. Third, by incorporating strategies for data-driven change discovery, it can avoid processing the whole corpus from scratch each time it changes, only selectively updating the POM according to the corpus changes instead. Besides increasing efficiency in this way, it also allows a user to trace the evolution of the ontology with respect to the changes in the underlying corpus.

4. Profile Based Web Search Using Rule Ontology – Proposed System

Same query is submitted by different users, typical search engines return the same result regardless of who submitted the query. Generally, each user has different information needs for his/her query. Therefore, the search results should be adapted to users with different information needs. In this paper, we first propose a user profile based web search using rule acquisition through ontology. Figure-1 shows system architecture of "Profile Based Web Search Using Rule Ontology".

4.1 Screening Method

An authorized user request is send to screening method. In screening method, user profile details are fetched from database. User profile details contain details about user area of interest and profession details. Hence the rule will be building based on this. Screening method collects the required information used to generate rule ontology.

4.2 Generate Ontology – Building Rule

Rule is build based on user search query and user profile details. Rule is built on XML format. The search query is send to Google search engine to search the content. All searches are performed based on user profile. User needs to provide text to be searched and the same will be searched and filtered using user profile details. And then rule is applied on the search result.

4.3 HTML Parsing and Stemming

4.3.1 HTML Parsing

Parsing HTML is an automated task, performed by HTML parsers. They have two main purposes:

HTML traversal: offer a interface for programmers to easily access and modify of the "HTML string code". Canonical example: DOM parsers.

HTML clean: to fix invalid HTML and to improve the layout and indent style of the resulting markup. Canonical example: HTML Tidy.

4.3.2 Stemming

Information Retrieval (IR) is essentially a matter of deciding which documents in a collection should be retrieved to satisfy a user's need for information. The user's information need is represented by a query or profile, and contains one or more search terms, plus perhaps some additional information such importance weights. Hence, the retrieval decision is made by comparing the terms of the query with the index terms (important words or phrases) appearing in the document itself. The decision may be binary (retrieve/reject), or it may involve estimating the degree of relevance that the document has to the query.

Unfortunately, the words that appear in documents and in queries often have many morphological variants. Thus, pairs of terms such as "computing" and "computation" will not be recognized as equivalent without some form of natural language processing (NLP).

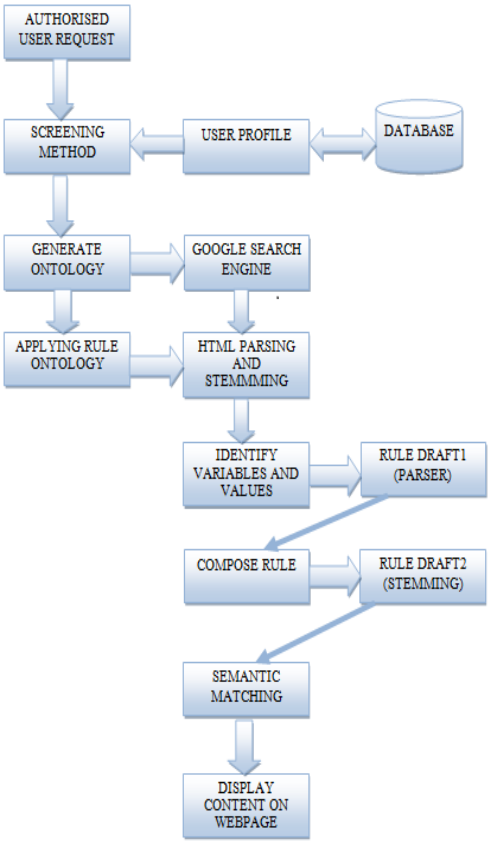

Fig 1: Authenticated User Module

4.4 Rule Acquisition through Ontology

In order to automatically acquire rules through ontology, we divided the rule acquisition procedure into two main steps in order to apply proper methods to each step. In the rule component identification step, we identify variables and values by using an ontology that describes frequently used variables and values in other rule bases. Subsequently, we compose rules from the identified rule components by using the rule structures of the ontology. The ontology helps to recommend feasible rules with variables.

4.4.1 Rule Component Identification

Fig. 2 shows how we can use ontology in rule component identification [2]. If we have rules acquired from Amazon. com (in short Amazon), as shown in the upper-left part of Fig. 2, we can make an ontology which shows the variables and values used in the rules, such as that shown in the upper-right part of Fig. 2. By using the information, we can identify rule components in a new site such as Barnes&Noble.com (in short BN). From the ontology, we can easily recognize that refund and days of the shipment of the Web page in the middle of Fig. 2 are variables and books, CDs, and VHS tapes are values [2]. The basic algorithm of identification is based on text matching between ontology and the text on a Web page [3]. Moreover, we can use information about omitted variables and the relations between the variables and values described in the ontology [3]. For example, we can perceive that item is omitted from the Web page shown in Fig. 2, because books, CDs, and VHS tapes are values of item in the ontology shown in Fig. 2. Also, it is possible to assign variables to corresponding values, because every value has its matching variable in the ontology.

4.4.2 Rule Composition from Identified Variables

The basic idea of rule composition is using patterns of rules in similar sites [2]. If BN uses similar regulations on refund policy to Amazon, the acquired rules will also be similar to the rules of Amazon in their patterns. The lower-right part of Fig. 2 shows a rule in the ontology generated from the rules of Amazon. It shows that the identified variables and values of the Web page can make a similar rule, as shown in the lower-left part of Fig. 2. The ontology plays a role of rule summary. The referenced rule of Amazon and the newly generated rule of BN are different from each other, but the ontology can connect them by summarizing the patterns of rules, as shown in Fig. 2.

In the previous study [3], a knowledge engineer could designate a target range for just one rule in the rule identification step so that the algorithm retrieved the most similar rule from the ontology by using a similarity measure. If there are ten rules in a Web page, the knowledge engineer should divide the area into ten ranges and repeat the rule selection step ten times. That is, there was no rule composition concept in the previous study. The objective of rule composition suggested in this paper is to retrieve a combination of similar rules for a given range and automatically assign variable instances to the rules. We expect that the burden on the knowledge engineer can be reduced compared to the previous study.

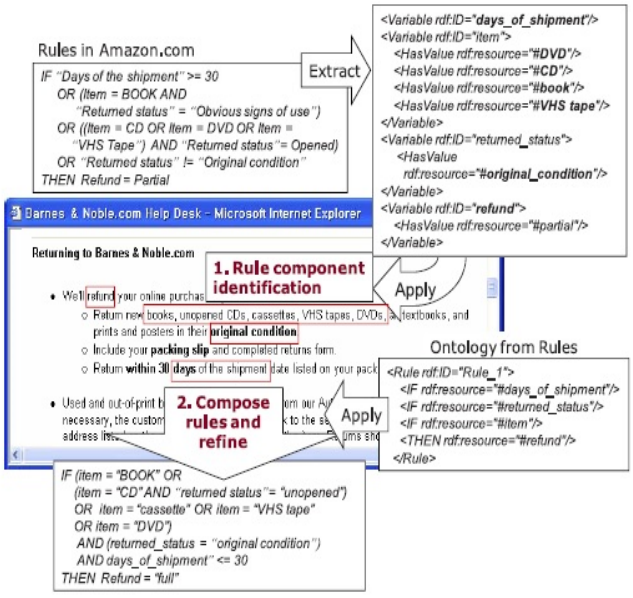

Fig 2: An example of rule acquisition through ontology

4.5 Semantic Matching

Semantic matching is a type of ontology Matching technique that relies on semantic information encoded in light weight ontologies to identify nodes that are semantically related.

Given any two graph-like structures, like classifications, database or XML schemas and ontologies, matching is an operator that identifies those nodes in the two structures which semantically correspond to one another. For example, applied to file systems it can identify that a folder labeled "car" is semantically equivalent to another folder "automobile" because they are synonyms in English. This information can be taken from a linguistic resource like WordNet.

Semantic matching represents a fundamental technique in many applications in areas such as resource discovery, data integration, data migration, query translation, peer to peer networks, agent communication, schema and ontology merging. It has been proposed as a valid solution to the semantic heterogeneity problem, namely managing the diversity in knowledge. Interoperability among people of different cultures and languages, having different viewpoints and using different terminology has always been a huge problem. Especially with the advent of the Web and the consequential information explosion, the problem seems to be emphasized. People face the concrete problems to retrieve, disambiguate and integrate information coming from a wide variety of sources.

5. Conclusion

In this paper, in order to provide each user with more relevant information, we proposed profile based web search using rule ontology to adapting search results according to each user's information need. Our approach is novel in that it allows each user to perform a fine-grained search, which is not performed in typical search engines, by capturing changes in each user's profile data. Rule is build based on user profile details, then the search result from engine is parsed based on this rule. The rule acquisition procedure consists of the rule component identification step and the rule composition step. And result is parsed and performed semantic matching and displayed to user according to the rule build up on user requirement. Only the information related to user profile will get displayed to the user and all other search results will be truncated. We believe that the technique proposed in this paper can be applied to situations where users require more relevant information to satisfy their information needs.

References

[1] Sangun Park and Juyoung Kang, "Using Rule Ontology in Repeated Rule Acquisition from Similar Web Sites," IEEE Transaction on Knowledge and Data Engineering Vol.24, NO.6, June2012.

[2] S. Park, J.K. Lee, and J. Kang, "A Framework for Ontology Based Rule Acquisition from Web," Proc. First Conf. Web Reasoning and Rule Systems, pp. 229-238, 2007.

[3] S. Park and J.K. Lee, "Rule Identification Using Ontology While Acquiring Rules from Web Pages," Int'l J. Human-Computer Studies, vol. 65, no. 7, pp. 644-658, 2007.

[4] K. Sugiyama, K. Hatano, M. Yoshikawa, and S. Uemura. A Method of Improving Feature Vector for Web Pages Reflecting the Contents of their Out-Linked Pages. In Proc. of the 13th International Conference on Database and Expert Systems Applications (DEXA2002), pages 891–901, 2002.

[5] K. Sugiyama, K. Hatano, M. Yoshikawa, and S. Uemura. Refinement of TF-IDF Schemes for Web Pages Using their Hyperlinked Neighboring Pages. In Proc. of the 14th ACM Conference on Hypertext and Hypermedia (HT '03), pages 198–207, 2003.

[6] IBM Almaden Research Center. Clever Searching. http://www.almaden.ibm.com/cs/k53/clever.html.

[7] T. H. Haveliwala. Topic-Sensitive PageRank. In Proc. of the 11th International World Wide Web Conference (WWW2002), pages 517–526, 2002.

[8] D. Richards, "Addressing the Ontology Acquisition Bottleneck Through Reverse Ontological Engineering," Knowledge and Information Systems, vol. 6, no. 4, pp. 402-427, 2004.

[9] T. Gruber, "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993.

[10] M. Craven, D. DiPasquo, D. Freitag, A.K. McCallum, T.M.Mitchell, K. Nigam, and S. Slattery, "Learning to Construct Knowledge Bases from the World Wide Web," Artificial Intelligence,vol. 118, no. 1/2, pp. 69-113, 2000.

[11] D. Sanchez and A. Moreno, "A Methodology for Knowledge Acquisition from the Web," Int'l J. Knowledge-Based and Intelligent Eng. Systems, vol. 10, no. 6, pp. 453-475, 2006.

[12] Y. Xu, J. Liu, and D. Ruan, "Rule Acquisition and Adjustment Based on Set-Valued Mapping," Information Sciences, vol. 157, no. 1/2, pp. 167-198, 2003.

Author Profile

S.Amithapatchan received the B.TECH degree in Information Technology from Kongu Engineering College, Affiliated to Anna University, Chennai, in 2006. He is working towards the M.E degree in Computer Science and Engineering from Gnanamani College of Engineering, Affiliated to Anna University, Chennai since September 2012.

B.Sivananthan received the M.E degree in Computer Science and Engineering from Gnanamani College of Technology, Affiliated to Anna University, Chennai. Received B.TECH degree in the Information Technology from Tamilnadu College of Engineering, Affiliated to Anna University, Chennai in 2008. Now working as Assistant Professor in Gnanamani College of Engineering, Affiliated to Anna University, Chennai Since June 2012. His research interest includes Cloud computing, Networking, VANET. He is a member of the ISTE.