Available Online at http://www.journalajst.com

ASIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ISSN: 0976-3376

Asian Journal of Science and Technology
Vol. 6, Issue 04, pp. 1278-1281, April, 2015

RESEARCH ARTICLE

RAMOLOGY: RULE ACQUISITION METHOD USING ONTOLOGY

[*,1]Vaishali Tyagi, [2]Shilpi Saxena, and [3]Divya Gupta

Department of Computer Science and Engineering, SRM University, India

ARTICLE INFO

Article History:
Received 12th January, 2015
Received in revised form
01st February, 2015
Accepted 24th March, 2015
Published online 30th April, 2015

Key words:
Ontology-based Rule Acquisition,
Semantic Web, BFS,
Rule Extraction.

ABSTRACT

Extracting information helps the users to salvage their needed information from the bulk of knowledge present on Internet. There is variety of information extraction techniques to explore our information from the internet but they do not resume the desired unique information. Ontology techniques can be used for extracting creative information but they have some issues. Rule acquisition is an important issue because the rules are essential part of semantic web applications. So, the proposed work states an algorithmic approach RAMOLOGY to form the rule draft by retrieving structure of websites and further utilizing ontology. It is also an underlying assumption that the rules which are implemented on web can be a major source for acquisition of rules and it is easier to acquire rules from websites of same domain.

INTRODUCTION

The Semantic Web is capable of answering search requests by employing natural language processing systems capable of converting samples of human language into more formal representations that are easier for computer programs to manipulate. Our current search engines rely on a less rigid representation of natural language where latent semantic indexing and co-occurrence matrices are utilized to find patterns. Latent semantic analysis is a technique of natural language processing in which relationships between a set of documents and terms they contain are analyzed. It produces a set of concepts related to the documents and terms. While this has been, to date, surprisingly effective, this analysis is limited because the data it attempts to understand has been constructed by inconsistent means. Natural inconsistencies between people and cultures create a body of work in which the language is simply not rigorous enough to allow computer to discern a significant amount of meaning. The Semantic Web, which is the key component of Web 2.0 and Web 3.0, is an evolving development of the World Wide Web in which the semantics of information and services on the Web are being defined. This is enabling the Web to understand and satisfy the requests of people and machines to use the Web content. Knowledge is an essential part of most Semantic Web applications and ontology, which is a formal explicit description of concepts or classes in a domain of discourse, is the most important part of the knowledge. The foundation of Semantic Web is comprised of a set of design principles and enabling technologies. Some examples of these include Resource Description Framework (RDF) and the Web Ontology Language (OWL) (Smith *et al.,* 2004), all of which are intended to provide a formal description of concepts, terms and relationships within a particular field of knowledge, or domain. Ontology learning, which refers to extracting conceptual knowledge from several sources and building ontology from scratch, enriching, or adapting an existing ontology, is one of the attempts at knowledge acquisition. Most ontology learning approaches acquire knowledge from the Web, because it offers a large amount of valuable information for every possible domain. However, ontology is not sufficient to represent inferential knowledge. This is because ontology based reasoning has limitations. Rule acquisition is as essential as ontology acquisition, even though rule acquisition is still a bottleneck in the deployment of rule-based systems (Richards *et al.,* 2004), (Tang *et al.,* 1994). This is time consuming and laborious, because it requires knowledge experts as well as domain experts, and there are communication problems between them. However, sometimes rules have already been implied in Web pages, and it is possible to acquire them from Web pages in the same manner as ontology learning. That is, most of the rule components already exist in Web pages. It means that we can acquire rules more easily by using an automatic rule acquisition method rather than the old method with domain experts and knowledge experts. However, there are some problems with extracting rules from text. First, which words of the Web page are rule components, which types of rule components are they, how can they be extracted? For example, item is a variable and book is a value in Fig. 1. Second, how to compose rules with the rule components. There are numerous possible combinations of making rules. Our idea for solving these problems is using rules of similar sites in limited

Corresponding author: Vaishali Tyagi,
Department of Computer Science and Engineering, SRM University, India.

situations under a couple of assumptions. Let's suppose that rules are to be acquired 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. Therefore, the portal should acquire rules about delivery options, shipping rules, and return policies from shopping.

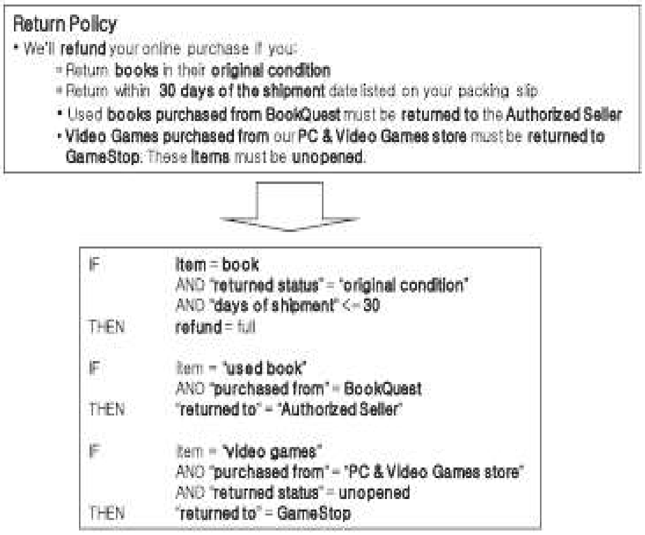

Fig.1. Example of Rule Acquisition

Related Works

Basically, our research is about an automatic knowledge acquisition procedure from the Web that consists of unstructured texts. Knowledge acquisition in current approaches can be grouped into two types: ontology acquisition and rule acquisition.

Ontology Tools

Aiming to eliminate one of the major obstacles to wider acceptance and deployment of ontology based software solutions, the Semantic Web community has devoted a significant attention to development of tools that would facilitate and scaffold the ontology development task (Marek *et al.,* 2012).

Following three categories are there for the available tool support:

1. Tools for handcrafting ontology.
2. Tools supporting reuse of existing ontology.
3. Tools for (semi)automatic development of ontology.

Tools for Handcrafting Ontology

Manual creation of ontology using a specialized ontology editor is still the predominant approach in ontology development. Its main drawback is the fact that the majority of ontology editors are suitable only for experts in the field of ontology development.

Tools Supporting Reuse of Existing Ontology

Supporting tools are needed to facilitate the searching process and evaluation of the retrieved ontology.

Tools Supporting (Semi) Automatic Ontology Development

These tools aim at reducing the human intervention to supervision of the development process and refinement of the results (Cimanio *et al.,* 2006), (Zouaq and Nkambou 2010). They offer users with suggestions for ontology elements, typically based on the analysis of existing domain documents.

Theoritical Background

Ontology

Ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts.

Use of Ontology

Ontology is used as a form of knowledge representation about the world or some part of it. As ontology defines the concepts and relationships within a domain, it provides a standardized vocabulary for that domain and the relationships between those concepts. Ontology is now central to many applications such as scientific knowledge portals, information management and integration systems, electronic commerce, and semantic web services.

Types of Ontologies

Upper Ontology

Concepts supporting development of an ontology (meta-ontology).

Domain Ontology

Concepts relevant to a particular topic or area of interest, for example, information technology or computer languages, or particular branches of science.

Interface Ontology

Concepts relevant to the juncture of two disciplines.

Process Ontology

Inputs, outputs, constraints, sequencing information, involved in business or engineering processes.

Ontology Learning

The main idea of ontology learning is very similar to ours. The approach using the Multiple Classification Ripple-Down Rules (MCRDR) methodology (Valencia-Garcia *et al.,* 2008) in ontology learning is somewhat similar to our approach in its framework. They extract concepts and relations by using Part of Speech (POS) tagging and inference with the MCRDR system with rules.

Rule Acquisition

The previous approaches to automatic rule acquisition before the Semantic Web era, mostly involved machine learning research based on pattern classification and learning by examples but learning by examples is a totally different concept from rule acquisition from texts. 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 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. The extraction of rules is one of the research areas in ontology learning, because inference rules could be an outcome of ontology learning.

Best-First Search

Best-first search is a graph-based search algorithm which is widely used as a problem solving technique in the field of artificial intelligence. In this, the search space can be represented as a series of nodes connected by paths. It is applicable to a discrete optimization problem in which it is assumed that the state space is represented in the form of a tree. Best-first search estimates the promise of node n with a heuristic evaluation function f (n). The order of traversal of the search graph is determined by the evaluation function. Moreover, the search typically stops as soon as a complete solution is found.

Problem Description

Problem Definition

Acquisition of rules through ontology till now is applied to some constraint areas with similar domains but performance varies from one area to another, as it depends on the nature of each area and its respective domain**.**

Problem Formulation

RAMOLOGY can be used for rule acquisition to extend its application in different areas having similar domains**,** by extracting rules using tools for ontology extraction and algorithms for rule acquisition.

Methodology Used

Rule acquisition is a multi-phase process, as shown in Fig 2. The two main phases of rule acquisition are rule component identification and rule composition.

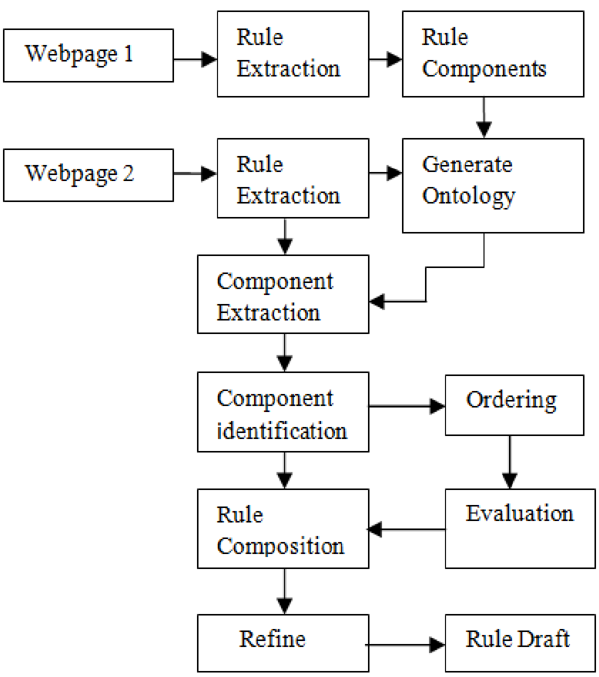

Fig. 2. Rule Acquisition Procedure

Rule Component Identification

In the rule component identification phase, variables and values are identified by using an ontology that describes frequently used variables and values in other rule bases as shown in Fig. 3. The goal of rule component identification is to elicit variables and values by comparing parsed words of the given text with the variables and values of given web page.

Steps involved in process of Rule Component Identification

The process of rule component identification consists of following steps:

1. Rules from both web pages are acquired with the synonyms of each term using Word Net (Miller *et al.,* 1995).
2. The stemming algorithm (20) is applied to the acquired rules in order to normalize the terms for the next step.
3. Parse and stem the Web page for rule acquisition.
4. Compare the terms of both Web pages using semantic matching.

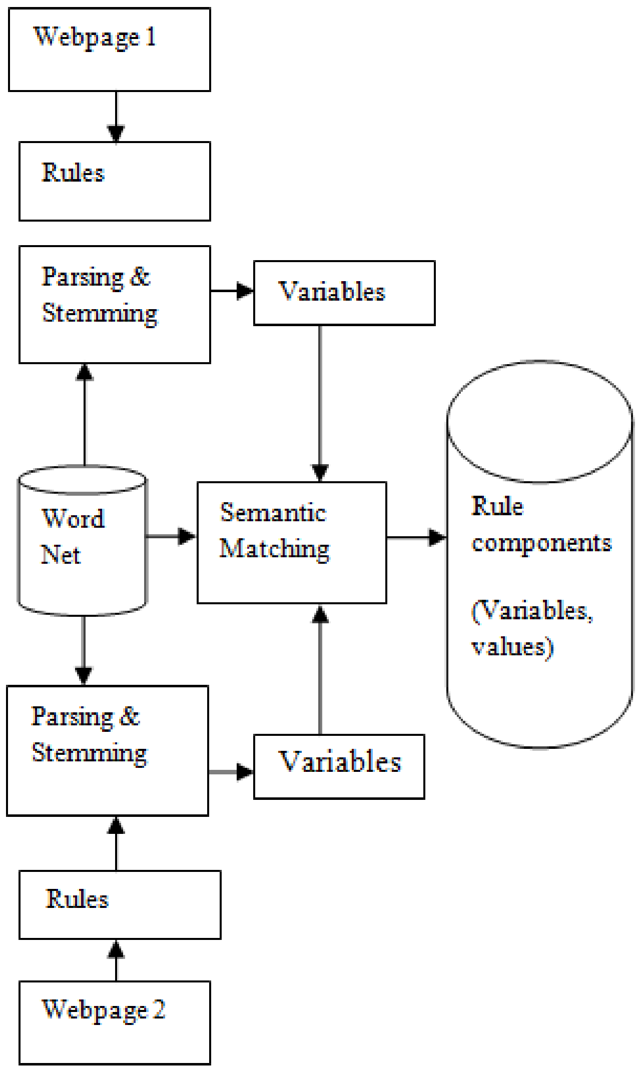

Fig. 3. parsing and stemming Component Identification Process

Rule Composition

In rule composition phase, rules from the identified rule components have been composed by using the rule structures of the ontology. The objective is to combine identified variable instances into rules.

Steps involved in Rule Composition Process

Rule Composition process consist of following steps:

1. In the first step of rule composition that is the preparation step, appropriate rules are selected by comparing the identified variable instances with the variables of the rule components which are identified in identification step as shown in Fig. 4.
2. Second step is to choose the candidate rules and the input of this step is the variable instances of rule components that is an output of the rule component identification step.
3. The third step is rule ordering, which generates Rule Order from the variable instances and the candidate rules.
4. The next step is variable ordering, which generates Total Order with the Rule Order and Variable Order that is calculated in this step.
5. Finally, best-first search algorithm is applied in the last step, to compose the rules that generate the final rule draft, as output.

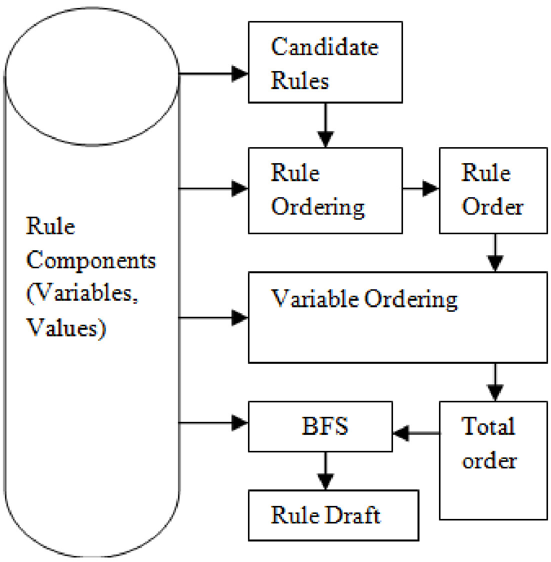

Fig.4. Rule Composition Process

Conclusion and Future Work

Inferential rules are as essential to the Semantic Web applications as ontology. However, rule acquisition research is relatively unpopular while there are many works on ontology learning. An automatic rule acquisition procedure using ontology is proposed in this paper named as RAMOLOGY that includes information about the rule components and their structures. This work started from the idea that it will be helpful to acquire rules from a site if there are similar rules acquired from other similar sites of the same domain. "RAMOLOGY" is a generalized, condensed, and specifically rearranged version of the existing rules. The rule acquisition procedure consists of the rule component identification step and the rule composition step. Subsequently stemming is performed with semantic similarity in the former step. Experimental results can show the potential of this approach, even though the experiments are very limited in terms of the domain setting. There are several challenging research issues that must be addressed in order to meet the ultimate goal of our research. A screening method to select exact parts that contain rules from Web pages can be developed. And this approach is needed to be applied into various domains, in order to extend its use and to see the performance rate and to check whether it varies or not while moving from web-pages of one domain to another, depending upon their respective features.

REFERENCES

Cimanio, P., Volker, J. and Studer, R. 2006 ."Ontologies on Demand ? – A Description of the State-of-the-Art, Applications, Challanges and Trends for Ontology Learning from Text Information," Wis-senschaft und Peaxis, vol. 57, nos. 6/7, pp. 315-320.

Sanchez, D. and A. Moreno, 2006. "A Methodology for Knowledge Acquisition from the Web," Int'l J. Knowlege based and Intelligent Eng. Systems, Vol.10, No.6.

Delin Jing, Hongji Yang, and Yingchun Tian "Abstraction Based Domain Ontology Extraction for Idea Creation".

Marek Hatala, Dragan Gasevoic, Melody Siadaty, J. Jovanoic, and Carlo Tornial, 2012. "Ontology Extraction Tools: An Empirical Study with Educators," IEEE Transactions on Learning Technologies, Vol.5, No.3, July-Sep..

Miller, G.A. 1995. "Word Net a Lexical Database for English," Comm. ACM, vol. 38, no. 11, pp. 39-41.

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

Sangun Park and Juyoung Kang, 2012."Using Rule Ontology in Repeated Rule Acquisition from Similar Web Sites," IEEE Transactions on Knowledge and Data Engineering,Vol.24,No.6,June.

Smith, M. K., Welty, C. and D. McGuinness, 2004. "OWL Web Ontology Language Guide, http://www.w3c.org/TR/owl-guide/.

Tang, Y.Y., Yan, C.D. and C.Y. Suen, 1994. "Document Processing for Automatic Knowledge Acquisition," IEEE Trans. Knowledge and Data Eng. vol. 6, no. 1, pp. 3-21, Feb.

Valencia-Garcia, R., Fernandex-Breis, J.T., Ruiz-Martinez, J.M., Garcia-Sanchez, F. and R. Martinez-Bejar, 2008 ."A Knowledge Acquisition Methodology to Ontology Construction for Information Retrieval from Medical Documents," Expert Systems, vol. 25, no. 3, pp. 314-334.

Zouaq, A. and R. Nkambou, 2010. "A survey of Domain Ontology Engineering: Methods and Tools," Advances in Intelligent Tutoring Systems, Nkambou, Bourdeau, and MMizoguchi, eds., pp. 103-119, Springer.

****