A Framework for Ontology Based Repeated Rule Acquisition from Similar Web Sites

E.DEEPAK KUMAR, M.VARUN KUMAR, P.DEEPA SITE, VIT UNIVERSITY, INDIA-632014.
varunkumar.m@vit.ac.in

Abstract—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. There are some problems with extracting rules from text. First, which words of the Web page are rule components and which types of rule components are they? Second, how can we 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 situations under a couple of assumptions. We proposed two main steps of rule acquisition, which consists of rule component identification and rule composition with the identified rule components. In other words, we identify rule components such as variables and values in Web pages by using RuleToOnto in the first step, and we combine the variables to compose rules in the second step.

Index Terms—Ontology based Rule Acquisition, RuleToOnto, Rule Acquisition, Rule Generation

1. INTRODUCTION

The Semantic Web 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. Therefore, rule acquisition is the process of gaining knowledge or rules from similar websites. The rule acquisition is also an important issue, and the Web that implies inferential rules can be a major source of rule acquisition. We expect that it will be easier to acquire rules from a site by using similar rules of other sites in the same domain rather than starting from scratch.

Rule acquisition is as essential as ontology acquisition; still however rule acquisition is at a standstill a bottleneck in the operation of rule-based systems. This is time consuming and difficult, because it wants knowledge proficient as well as domain specialists, and in attendance are statement problems among them. Nevertheless, little bits rules have previously been implied in Web pages, and it is potential to acquire them from Web pages in the same manner as ontology learning. The upper part is the text in a Web page explaining the return policies of Barnes & Noble.com, which is one of the representative online bookstores. The lower part shows acquired rules from the text. We can see that the words with bold font in the upper part are used as the variables and values of the acquired rules in the lower part. 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. As we repeat the rule acquisition process across several sites, we can accumulate rules. However, as the size of rule base increases, it becomes hard to reuse rules. Therefore, we used an ontology named RuleToOnto, which symbolizes the information about rules including terms, rule component types, and rule structures.

2. RELATED WORK

The knowledge acquisition in current approaches can be grouped into two types, ontology acquisition and rule acquisition. There research is about an automatic knowledge acquisition procedure from the web that consists of unstructured texts. Acquiring knowledge is based on the type of ontology, and using ontology as a knowledge schema in the knowledge acquisition is more common than rule acquisition.
I. SEMANTIC SIMILARITY

The Semantic Web is an evolving development of the World Wide Web in which the Inferential rules are as essential to the Semantic Web applications as ontology. Therefore, rule acquisition is also an important issue, and the Web that implies inferential rules can be a major source of rule acquisition. We expect that it will be easier to acquire rules from a site by using similar rules of other sites in the same domain rather than starting from scratch. We proposed an automatic rule acquisition procedure using a rule ontology RuleToOnto, which represents information about the rule components and their structures. The rule acquisition procedure consists of the rule component identification step and the rule composition step. We developed A* algorithm for the rule composition and we performed experiments demonstrating that our ontology-based rule acquisition approach works in a real-world application.

II. ONTOLOGY BASED RULE ACQUISITION

This section, we recommend a process which routinely acquires rules through RuleToOnto. In step 1, RuleToOnto is produced from rules which are acquired in another site. In step 2, variables and values are routinely recognized as of the Web page by means of RuleToOnto and the first rule draft is caused. In step 3, rules are automatically invented by combining the identified variables and values. We developed A* algorithm for this purpose. Still, the generated rules may be unfinished. So, the knowledge engineer requires to filter the second rule draft to make it complete in step 4. Rule Ontology Generation RuleToOnto is domain specific knowledge that provides information about rule components and organizations. It is feasible to directly use the rules of the previous system as an alternative of the proposed ontology. Conversely, it requires a large space and extra processes to utilize information on rules, while RuleToOnto is a generalized compact set of information for rule acquisition. Thus, we use RuleToOnto instead of the rules themselves. While the rule component identification step needs variables, values, and the relationship between them, the rule composition step needs simplified rule structures.

III. RULE COMPONENT IDENTIFICATION

This article expanded RuleToOnto by adding synonyms of each term using WordNet. In the comparison between the terms of RuleToOnto and the terms of the Web page[29], if there is used semantic matching instead of simple string comparison. In order to find the semantic similarity between two terms used WordNet. Identified components are denoted in the format of variable instances with variable abbreviation and number. If there are, rules acquired from Amazon.com (in short Amazon), as shown in the upper-left part of Fig. 1, it can make an ontology which shows the variables and values used in the rules, such as that shown in the Fig. 1. By using the information, there can be identify rule components in a new site such as (in short BN)[20].

From the ontology, there are easily recognize that refund and days of the shipment of the Web page in the middle of Fig. 1. The goal of rule component identification is to extract variables and values by comparing parsed words of the given text with the variables and values of RuleToOnto.

For example, this paper can perceive that item is omitted from the Web page shown in Fig.1, because books, CDs, and VHS tapes are values of item in the ontology shown in Fig.1. In addition, it is possible to assign variables to corresponding values, because every value has its matching variable in the ontology.

This measure is calculated only when one term is a hyponym of the other term, and the path length is the path length between the two terms in the hyponym hierarchy. There are decided that two terms are semantically related when the measure is larger than 0.25[20].
3. PROPOSED WORK

We proposed an automatic rule acquisition procedure using rule ontology RuleToOnto, which represents information about the rule components and their structures. The rule acquisition procedure consists of the rule component identification step and the rule composition step. We proposed two main steps of rule acquisition, which consists of rule component identification and rule composition with the identified rule components. In other words, we identify rule components such as variables and values in Web pages by using RuleToOnto in the first step, and we combine the variables to compose rules in the second step.

The purpose of using ontology in our approach is to automate the rule acquisition procedure. The starting point of our approach is that it will be helpful for acquiring rules from a site, if we have similar rules acquired from other similar sites of the same domain. Rule ontology, which we use in this paper, includes the information about rules including terms, rule component types, and rule structures. RuleToOnto is domain specific knowledge that provides information about rule components and structures. It is possible to directly use the rules of the previous system instead of the proposed ontology. While the rule component identification step needs variables, values, and the relationship between them, the rule composition step requires generalized rule structures. Therefore, RuleToOnto represents the IF and THEN parts of each rule by connecting rules with variables with the IF and THEN relations, in addition to basic information about variables, values, and connections between variables and values. The RuleToOnto schema has three object properties HasValue, IF and THEN, and three classes, Variable, Value and Rule to generate the RDF graph. An instance of Variable should have at least one Value instance for its Has Value property, and Rule should have at least one Variable for each of the IF and THEN properties.

I. Extract Rule From Websites

First, select the particular domain and select the Input Rule webpage from the particular domain. Then acquire the rule from input webpage. From the webpage, the rule can automatically extract by using screening method and display the rule webpage. Get the source from the rule webpage and identify variables and values

II. Construct Ontology

After acquired the rules from webpage, this article can construct the ontology based on identified variables and values. In addition, get another similar rule webpage related to input webpage by using this ontology information, there can be identify rule component in new input site.

III. Rule Refinement

Once the rules are determined, the next step is to complete the rules by assigning variable and value pairs to IF or THEN. The identified rule instances can be converted to the variable value pairs by matching variables and values with identified values and the ontology. Assigning the pairs to IF or THEN is very simple. If the variable belongs to an IF part in the rule instance of RuleToOnto, this can be assign the pair to the IF.
part of the rule. Otherwise, if it belongs to a THEN part, this paper assign it to the THEN part. The rules automatically generated are not complete in most cases, so they need to refine them. The knowledge engineer checks the rules and modifies connectives and values.[20] The following rule is an example of the refined rule. The knowledge engineer changed the operator of days_of_shipment from “=” to “<=” and added the value full by referencing the ontology and the target Web page.

IF days_of_shipment <= 40
   AND return_status = “original condition”
   AND item = “book”
THEN
   Refund = “full”

4. Conclusion
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. We proposed an automatic rule acquisition procedure using ontology, named RuleToOnto, that includes information about the rule components and their structures. We started from the idea that it will be helpful to acquire rules from a site if we have similar rules acquired from other similar sites of the same domain. RuleToOnto is a generalized, condensed, and specifically rearranged version of the existing rules.

5. REFERENCES