An Automatic Approach to Extracting Requirement Dependencies based on Semantic Web

Hui Guan
Shenyang University of Chemical Technology
Shenyang, Liaoning, China
h.guan@syuct.edu.cn

Guorong Cai
Shenyang University of Chemical Technology
Shenyang, Liaoning, China
1456177729@qq.com

Chuan Zhao*
Beijing technology and Business University
Beijing, China
zhaochuan2006tom@126.com

Abstract—For a complex system with a large number of requirement texts, manual analysis of the dependencies between the requirement texts is very time-consuming and requires heavy workload. This paper proposes an automatic approach to extract requirement dependencies based on semantic web. Firstly, requirements text is formalized as a set of keywords by segmenting the requirement text and making part-of-speech tagging. Secondly, the requirement dependencies are automatically obtained by constructing the semantic web ontology and defining inference rules. Finally, the results of requirement dependence extraction and expert extraction are compared and analyzed. The results show that the proposed method has high feasibility and accuracy which can save a lot of human resources and cost.

Keywords: requirement change; semantic web; requirement dependency; ontology; dependence extraction

I. INTRODUCTION

How change is handled is one of the most difficult and important parts of requirements engineering. As software projects become more complex and requirements texts become larger, it becomes difficult and error-prone to manually analyze the dependencies between requirements and the impact of requirements change. Under this background, this paper provides requirement dependencies that contain more semantic information, and uses semantic knowledge to realize automatic extraction of requirement dependencies.

The advantage of Semantic Web is that can be directly processed and operated by a computer after the information resources are formalized. It can pre-build the vast semantic web of a certain domain. On the basis of understanding the data relationships, it can obtain semantic dependency information between requirements through certain logical reasoning. The entire architecture of Semantic Web is divided into seven layers from bottom to top: Unicode and URI layer, XML layer, RDF layer, Ontology layer, logical reasoning layer, result verification layer and trust evaluation layer. In this paper, ontology and inference layer are used to realize automatic inference of requirement dependencies relationship, so as to save the corresponding human resources. Ontology is an abstract lexicon with semantic dependencies, which clearly defines the semantic dependency relationship between words and enables the computer to understand the semantic information of a certain field behind it. And reasoning layer contains a series of automatic requirement dependencies inference rules which can extract dependencies fact in the ontology semantic lexicon according to the requirements input. Therefore, automated reasoning of requirement dependencies can be achieved to save cost of the manpower analysis in project development.

In this paper, we propose a method to automatically extract requirement dependencies from a large number of requirement texts by combining semantic web technology. Based on this method, the requirements are pre-processed into noun and verb keyword sets by word segmentation tool HanLP (http://hanlp.com) and part of speech tagging. Then, ontology semantic lexicon is constructed by processing information resources in a certain domain, and semantic annotation is made for data with potential semantic relationship in ontology to enhance semantic dependency. By introducing inference mechanism for inference machine, the rule knowledge in the rule base is matched according to the user's requirement text and the dependency facts in the ontology semantic lexicon are extracted automatically which then are fed back to the user interface.

Section 2 of this paper briefly introduces the relevant research work; Section 3 describes the specific steps of requirement dependency extraction through Semantic Web technology; Section 4 makes a comparative analysis of the proposed method and the results extracted by experts to prove the feasibility of the proposed method through the specific experimental data. Section 5 concludes with a summary.

II. RELATED WORK

In terms of automatic requirements modeling development tools, Han et al. [1] combined the visually Unified Modeling Language (UML) with time automation, but this method focused on the implementation stage of the software and did not provide software-adaptive requirements modeling and formal testing. Dhikra Kchaou et al. [2] proposed a method of using use cases with structured
documents to represent requirements. Based on the analysis of the similarity between element in the use case text and element in the corresponding sentence, semantic trace-ability can be improved. Requirements were represented with a well-formed use-case template and designs model using class and sequence diagrams. Li [3] established a measurement model of conceptually semantic correlation based on word frequency, and established an associatively semantic query network based on the measurement results. Experiments were used to verify the rationality of automatic modeling.

In the field of relation extraction, there have been many targeted research results, which have been applied in a variety of different fields. Tang [4] proposed an entity relationship extraction model, which realized the extraction of data set relations based on the deep learning method of multiple instances and multiple labels. Rong [5] proposed a multi-channel convolutionally neural network model for entity relationship extraction of multi-channel convolutionally neural network. Other scholars such as Li [6] and Tian [7] have applied relation extraction to biomedical and event causality. At present, no method has been used to automatically obtain the dependencies between requirements.

Requirement dependency plays particularly important role in requirement analysis system. Pohl [8] summarized requirements described in several literature in the field of requirements engineering and divided requirement dependency into 18 types, which is relatively comprehensive at present. Luo [9] proposed the concept of requirement cluster and the method of dividing requirement cluster based on the relationship between requirements. According to the definition of UML, the requirement relation is defined as seven atomic relations: call relation, interrupt relation, arouse relation, modification relation, notification relation, control relation and resource control dependency.

In the application field of Semantic Web, Hu [10] proposed a semantic retrieval method of literature based on domain ontology technology. The purpose of semantic retrieval is to extend the user's query conditions semantically on the basis of accurate understanding of retrieval conditions. At the same time, the literature domain ontology can provide more accurate and comprehensive retrieval service for literature retrieval. Yan [11] proposed a knowledge reasoning method supported by semantic rough fuzzy ontology to realize knowledge reasoning for inaccurate knowledge. Achim Rettinger et al. [12] presented a new approach to processing Semantic Web knowledge bases that rely on statistical reasoning about their standard representations. Minu Rajasekar Indra et al. [13] used fuzzy rules to establish logical implication between domain specific ontology and entity. There is currently no way to use semantic web technologies for automatic reasoning about dependencies between requirements.

III. REQUIREMENT DEPENDENCY EXTRACTION

This paper proposes an automatic requirement dependency extraction method based on semantic web technology. The main steps of this method are shown in Figure 1. The method is mainly divided into two stages. The first stage is to formalize the requirements in which the original word segmentation tool HanLP is improved by adding words in specific fields. Then the requirement text is segmented into independent nouns and verbs by using part of speech tagging with HanLP and then combined into a set of keywords. The second stage is to automatically extract the dependencies by using Semantic Web technology. In this paper, requirement dependency relations is set up by using semantic annotation for the potential semantic relations of data in the ontology. A relational lexicon is added with domain-specific vocabularies and a series of requirement dependencies inference rules are introduced and incorporated into reasoning machine to extract requirement dependencies automatically.

Figure 1. Framework of requirement dependency extraction

A. Formalization of requirement

Definition 1. The requirements document consists of several requirements, and each requirement is represented by the following triples:

\[ Ri = \{R.content, R.kwset, R.relation\} \]

Based on the above definition, \( R.content \) represents the content of requirement text description, \( R.kwset \) represents the keywords set extracted after adding constraints in the requirement, and \( R.relation \) represents the dependency relationship between requirements.

Definition 2. Keywords set refers to the set composed of verbs and nouns after keyword extraction of requirements:

\[ R.kwset = \{v_1, v_2, \ldots, v_n, n_1, \ldots, n_i\} \]

Requirement processing steps are shown in Figure 2. Due to the importance of nouns and verbs in retaining semantics, the nouns and verbs in each requirement text are extracted and formalized as a set of keywords by segmenting the requirement text and making part-of-speech tagging. Table 1 shows a requirement document case given by [9] about the composition evaluation system of elementary and middle school students. Formalized requirement expression is shown in Table 2.

Figure 2. Details of the requirements processing step.
B. Definition of requirement dependency

In our previous research, we have defined the following kinds of requirement dependencies and made automatic extraction of them. The specific definition is as follows:

Similar: If \( R_1 \) and \( R_2 \) have the same requirements, they have a similar relationship.

Include: If the requirement \( R_2 \) is part of \( R_1 \), then \( R_1 \) and \( R_2 \) are include relationship.

Arouse: If the requirement \( R_2 \) needs to be realized after \( R_1 \), then \( R_1 \) and \( R_2 \) have an arousing relationship.

Notification: If \( R_1 \) allows \( R_2 \) to start to be realized after implementation, then \( R_1 \) and \( R_2 \) have a notification relationship.

Call: If \( R_1 \) needs to implement \( R_2 \) in the implementation process, that is, \( R_1 \) is implemented before \( R_2 \), but \( R_2 \) is implemented before \( R_1 \), then \( R_1 \) and \( R_2 \) have a calling relationship.

Conflict: If \( R_1 \) and \( R_2 \) cannot be realized at the same time, then \( R_1 \) and \( R_2 \) have a conflict relationship.

C. Framework of semantic web technologies

The framework of semantic web technology is shown in Figure 3. Firstly, the information resources need to be pre-processed in a specific domain. And then semantic dependencies are added for every pair of word in the ontology lexicon (fact knowledge base) using artificial semantic annotation. After the requirements are formalized, a pair of keyword sets of requirements text are imported into this system. And then the inference machine matches the dependency extraction rule in the rule knowledge base according to the inference request. If this succeeds, the facts (dependency relationship) in the ontology semantic lexicon are extracted and fed back to the user. If this fails, the reasoning ends.

(1) Annotating semantics

Ontology semantic network diagram is shown in Figure 4, which represents the semantic relationships between entity classes for a particular domain. In this paper, the ontology semantic network graph is constructed, and semantic dependency annotations are added to the data with potential semantic relations in the graph to strengthen the semantic connection between texts. At the same time, manual annotation is used to add semantic dependencies relationship for every pair of words in the relational lexicon with domain-specific vocabularies that corresponds to the ontology semantic network diagram.

![Figure 3: The framework of semantic web technology](image-url)
(2) Construction of ontology

Ontology constructed in this paper mainly includes fact knowledge base and rule knowledge base. The fact knowledge base is equivalent to a conceptual semantic lexicon with semantic dependencies relationship between words. Among them, ontology diagram corresponds to semantic lexicon. The semantic lexicon is shown in the Table 3. And rule knowledge base is a set of rules for dependency inference. Figure 5 is part of ontology diagram.

**TABLE III.** THE SEMANTIC LEXICON

<table>
<thead>
<tr>
<th>Dependency types</th>
<th>A lexicon with semantic dependency relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>arouse</td>
<td>[register, login, logout, record][choose, generate, send][log in, logout, record][log out, record][revise students’ homework, get a good score, get a top score]</td>
</tr>
<tr>
<td>conflict</td>
<td>[delete, add, supplement][delete, add][again, never again][determine, don’t determine][low score, medium, top score][estimate, determine][estimate, determine, affirm]</td>
</tr>
<tr>
<td>call</td>
<td>[choose, determine][judge, filtrate][modify information, inform whether modification was successful][finish homework, revise homework]</td>
</tr>
<tr>
<td>notification</td>
<td>[print, feedback][provide, use, send, inform][provide, deliver, answer question, finish the project]</td>
</tr>
<tr>
<td>similar</td>
<td>[need, hope][assist, help][good, top]</td>
</tr>
</tbody>
</table>

Figure 5. Part of ontology diagram.

**TABLE IV.** REQUIREMENT DEPENDENCIES ANALYSIS WITH SIMILARITY BELOW THE THRESHOLD.

<table>
<thead>
<tr>
<th>Requirement description</th>
<th>Similarity</th>
<th>Dependency relationship</th>
<th>Erroneous judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: The TA can help students answer questions in the system</td>
<td>16.67%</td>
<td>no</td>
<td>Notificati on relationship</td>
</tr>
<tr>
<td>B: Students completing course assignments will be corrected directly in the system</td>
<td>15.43%</td>
<td>no</td>
<td>Call relationship</td>
</tr>
<tr>
<td>A: “The teacher asks the students to finish their homework carefully and in time” B: “Students write homework and are corrected by the TA in the system”</td>
<td>22.36%</td>
<td>no</td>
<td>Call relationship</td>
</tr>
<tr>
<td>A: The instructor chooses to evaluate the composition in a particular grade B: “Instructor determines and revise s comments”</td>
<td>25.82%</td>
<td>no</td>
<td>Arouse relationship</td>
</tr>
<tr>
<td>A: The system shall provide information transmission equipment B: “The system should allow students to retrieve information about students and instructors”</td>
<td>25%</td>
<td>no</td>
<td>Arouse relationship</td>
</tr>
</tbody>
</table>

(3) Introduction of inference mechanism

a. Calculation of similarity.

First of all, for making requirements A and B correlated, the rule should meet an evaluation index of minimum similarity. Similarity is calculated by the frequency of keyword in all keyword sets of any two sentences, and the word frequency is expressed as a vector to calculate the cosine value. In this paper, requirement is represented by n-dimensional vector after keyword extraction, so the similarity formula for calculating requirement A and B is as follows:
\[ \text{similar}(A,B) = \cos \theta = \frac{A \cdot B}{|A| |B|} \]  

(3-1)

According to the data comparison of several tests (as shown in the table 4), it is found that there is no dependency relationship between most requirements when the similarity is less than 0.3. If a value less than 0.3 is retained, the result will be wrong if the dependency is judged through the ontology and rules. Therefore, the threshold value of similarity is set as 0.3 in this paper.

b. Dependency extraction rules

In this paper, production rules are used for dependency extraction reasoning. The specific inference rules are as follows:

1. Greater similarity (A, B, 0.3)
2. Have intersection (An, Bn)
3. Exist in ontology (An, Bn)
4. Have intersection (Av, Bv)

Interpretation of inference rules is listed as following:

- A or B is a requirement text.
- An or Bn represents a set of nouns keyword of requirement A or B.
- Av or Bv represents a set of verbs keyword of requirement A or B.
- Greater similarity (A, B, 0.3) represents that similarity of requirement A and B is greater than 0.3.
- Have intersection (An, Bn) represents a pair of nouns set of two requirements where it exists an intersection of nouns.
- Have intersection (Av, Bv) describes that a pair of verbs set whether it exists in ontology semantic lexicon.
- Dependency (A, B) is dependency of two requirements.

In the inference rules, dependencies relationship can be inferred by any of the below combination rules.

- Greater similarity (A, B, 0.3) && Have intersection (An, Bn) && Exist in ontology (Av, Bv) \rightarrow Dependency (A, B).

In order to prove the correctness and reliability of requirement dependency inference, this paper uses two requirement texts to achieve concrete reasoning rules. \( \text{R} \) represents instructor choose to revise comment, \( \text{B} \) represents instructor determine revise comment. \( \text{An} \) represents (instructor, comment), \( \text{Bn} \) represents (instructor, comment), \( \text{Av} \) represents (choose, revise), \( \text{Bv} \) is (determine, revise). Greater similarity (A, B, 0.3) is greater than 0.3, returning true value. Have intersection (An, Bn) exists intersection (instructor, comment), returning true value. In addition, exist in ontology (Av, Bv) exists in ontology semantic lexicon, so exist in ontology (Av, Bv) returns true value. Finally, dependency (A, B) is matched to the call of dependency type in the ontology semantic lexicon.

IV. VERIFICATION

The experimental results are mainly used to analyze and compare requirement dependency extraction. This paper further confirms the feasibility of the proposed method by comparing it with the method in [9]. This paper adopts three evaluation criteria for the comparison of the two methods: accuracy rate \( P \) and recall rate \( R \) and measure value \( F_1 \), among them: \( P \) is used to represent the percentage of number of dependencies that are labeled correctly as a certain dependency type in extraction results, \( R \) represents the percentage of number of dependencies that are labeled correctly as a certain dependency type in extraction results of test data set, \( F_1 \) indicates that the accuracy of the measurement experiment. In Equations 4-1 and 4-2, \( \text{count} \) represents the number of requirement pairs correctly labeled as a given relationship type in the extraction results, \( \text{ResultNum} \) represents the total number of requirement pairs with all relationship types in the results, and \( \text{TestNum} \) represents the total number of requirement pairs with all relationship types in the test data set. The experimental results are shown in Table 5.

<table>
<thead>
<tr>
<th>Dependent</th>
<th>Automatic extraction results</th>
<th>Manual extraction results</th>
<th>P(%)</th>
<th>R(%)</th>
<th>F1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notification</td>
<td>(R18, R19)</td>
<td>(R18, R19)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Arouse</td>
<td>(R1, R2), (R1, R3), (R1, R4), (R2, R4), (R3, R4), (R13, R16)</td>
<td>(R1, R4), (R2, R4), (R3, R4), (R13, R16)</td>
<td>71.43</td>
<td>100</td>
<td>83.33</td>
</tr>
<tr>
<td>Call</td>
<td>(R10, R11), (R12, R12)</td>
<td>(R11, R10), (R11, R12)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

\[ p = \frac{\text{count}}{\text{ResultNum}} \]  

(4-1)

\[ R = \frac{\text{count}}{\text{TestNum}} \]  

(4-2)

\[ F1 = \frac{2 \times P \times R}{P + R} \]  

(4-3)

In order to more intuitively represent the dependency relationship between requirement texts, Figure 6 shows the comparison diagram of the extraction results of dependency relationship between this paper and [9].

According to the indicators in Table 5 and the compared data from the extracted results, the efficiency and performance percentages in all aspects are relatively high after the addition of Semantic Web ontology and the definition of inference rules, which proves the feasibility and reliability of the method in this paper to a certain extent.
This paper proposes a method to automatically extract requirement dependencies based on Semantic Web technology. Based on this method, after formal processing of requirements, ontology semantic network graph and ontology semantic lexicon are established, rule knowledge base is introduced. Finally, dependency extraction of requirement text is automatically carried out. The innovation and main work of this paper is to enable the computer to understand and reason the requirement dependency type by using semantic Web technology, and to provide the semantic dependency basis for the analysis of requirement change in the future. At the same time, it also saves a lot of human resources and time costs.

REFERENCES


