A Study on Semantic Web Construction Platform

Li Man, Du Xiaoyong, Wang Shan
School of Information, Renmin University of China, Beijing, 100872
liman1@ruc.edu.cn

Abstract The Semantic Web proposed by Tim Berners-Lee is the extension of the current Web. Although the architecture of the Semantic Web has been proposed, the method of constructing the Semantic Web is being explored and there is not an integrated platform for it. The paper proposes a practical methodology of evolution based on prototype by analyzing the characteristics of the Semantic Web construction, gives a framework of Semantic Web construction platform based on the methodology and introduces the main modules in detail. The methodology and platform has been applied in the Semantic Web project in the Renmin University of China.

Keywords Semantic Web, Ontology, Semantic Annotation

1. Introduction

The Semantic Web[1] proposed by Tim Berners-Lee has been regarded as the next generation of current Web, which aims to add semantics and better structure to the information available on the Web. According to the description of Tim Berners-Lee, the Semantic Web is the extension of the current Web.

Because the Web is open and dynamic, there exist many challenges for reconstructing the current Web to the Semantic Web. For example, how to construct the large-scale ontologies easily for the Semantic Web? How to annotate the numerous resources in the Web automatically? How to maintain related resources with the evolution of ontology? … There are not good resolutions for these problems. In addition, until now, the related tools for the Semantic Web can only support some function required by the Semantic Web construction. For example, OntoEdit[2], Protégé[3] and KAON[4], etc, are ontology editors. CREAM[5], Annotea[6] and MnM[7], etc, are ontology annotators. FCA-Merge[8] and PROMPT[9], etc, are ontology merger tools. Although they can be used to fulfill some tasks for the Semantic Web construction, the ideal Semantic Web platform should not be the simple combination of several existing tools, but an integrated platform supported by a kind of development methodology.

Although the architecture of the Semantic Web has been proposed, the methodology for developing the Semantic Web is still in research and there is no mature one to guide the practice. So this paper discusses the methodology of constructing the Semantic Web according to our experience in Semantic Web project in the Renmin University of China and proposes a practical methodology of evolution based on prototype. Then the paper gives a framework of Semantic Web construction platform based on the methodology and introduces the main modules. Finally, the paper presents an application case.

This paper is organized as follows. Section 2 discusses the method of constructing the Semantic Web. Section 3 proposes a related framework of Semantic Web construction platform and introduces the main modules in detail. Section 4 draws a conclusion.

2. Methodology for Constructing the Semantic Web
Construction of the Semantic Web is a huge project, which needs methodology as guidance. However, until now there is no mature methodology for it. Most existing methodologies only consider the ontology development method, such as skeletal methodology\textsuperscript{[10]}, enterprise modeling methodology\textsuperscript{[11]} and METHONTOLOGY\textsuperscript{[12]}, etc. They are used to guide the ontology development instead of the Semantic Web construction.

It is agreed that the current Web can be developed by common users freely, while the construction of Semantic Web depends on domain experts. Furthermore it is difficult to determine what is a well-constructed Semantic Web. So it is not practical for several domain experts to construct Semantic Web without the guidance of methodology. So we propose the methodology of evolution based on prototype for constructing the Semantic Web. In the methodology, “prototype” means that the Semantic Web should be built from some existing ontologies and few resources annotated according to the ontologies, because it is not possible to find a comprehensive ontology for resource annotation. The “evolution” means that the construction is a repeated process, because the Semantic Web may not be constructed well only once and it will evolve continuously.

In the methodology, evolution of the Semantic Web is a key. Because the main elements of the Semantic Web are ontologies and annotated resources based on existing ontologies, the construction of the Semantic Web comprises ontology development and resources annotation. During the process of building the subject-oriented Semantic Web, we find that the development of ontologies and annotation of resources should not be two separated processes. They are cyclic and can be promoted with each other. On the one hand, with the enrichment of the knowledge in ontology, the resources should be annotated repeatedly. On the other hand, during annotating the resources, new knowledge may be found and added to the ontology. The two processes interact and repeat. With the circulation, ontologies are constructed more reasonably and resources are annotated better, that is to say, the Semantic Web is built gradually. In conclusion, we believe that the evolution of Semantic Web will be fulfilled through the interaction of ontology development and resources annotation, which is shown in Fig.1.

![Fig.1. Methodology of Evolution Based on Prototype](image)

The methodology of evolution based on prototype emphasizes the characteristics of evolution during constructing the Semantic Web and gives the evolution approach of Semantic Web, which is practical.

3. A Semantic Web Construction Platform

Facing to the abundant Web resources, it is impossible to construct the Semantic Web by hand. However, the Semantic Web construction will involve the comprehension for natural language. It is not practical to build it fully automatically. So it is necessary to develop a related tool to aid developers to construct the Semantic Web semi-automatically.

This paper proposes the framework of Semantic Web construction platform based on the methodology of evolution based on prototype, which is shown in Fig.2.
The function of the Semantic Web construction platform in Fig.2 is to aid developers to reconstruct the current Web to the Semantic Web with the guidance of methodology of evolution based on prototype. The platform comprises three main modules: ontology development module, semantic annotation module and evolution management module. The ontology development module is used to construct ontology. The semantic annotation module produces the annotated resources based on the constructed ontology and the suggestion for enriching ontology. The evolution management module is responsible for the evolution management of ontologies and resources. The framework integrates the ontology development module and semantic annotation module based on the evolution management module, which meets the requirement of the methodology of evolution based on prototype.

Although some ontology development tools and semantic annotation tools have been developed, we find that they cannot be used directly in our platform and there are some limitations in them. So in our platform, we redesign and redevelop the related modules, which will be introduced in the following.

(1) Semantic Annotation Module

The existing annotation tools, such as CREAM, Annotea and MnM, etc only consider how to annotate resources according to ontology easily, but not support the evolution of ontology. Therefore they are not fit for our framework of the Semantic Web construction platform.

In our platform, the semantic annotation module includes the following sub-modules:

- Document Viewer. The document viewer visualizes the document contents, including the original documents and annotated documents. The document viewer support various format, such as HTML, XML, RDF, OWL, etc.
- Ontology Viewer. The ontology viewer can show the existing ontology for the use of resources annotation and show the relations between the annotated contents and corresponding ontological objects.
- Annotation Generator. The annotation generator supports annotating the resources according to existing ontology by hand or automatically.
- Knowledge Generator. The knowledge generator gives some advices for ontology development by applying information extraction technique to documents. It also supports selecting pieces of documents and aligning it with parts of the ontology.

Besides the above sub-modules, the semantic annotation module includes the crawler used to gather the related resources from the Web, and inference engine, which reasons on crawled and newly created instances and on the ontology. The inference engine also serves the ontology viewer and knowledge generator to allows user to query or modify the existing classes, instances and properties.

(2) Ontology Development Module

The success of the Semantic Web depends strongly on the proliferation of ontologies, which will involve
collaborative efforts of multiple developers. Therefore the module must support collaborative ontology development process.

However, collaborative construction of ontologies is a complicated task. Existing tools cannot meet the requirement well. For example, OntoEdit, Protégé and KAON only provide an environment for multi-user concurrent operations, but not provide ideal mechanism for negotiation and collaboration of multi-user.

So we proposed the role-based collaborative development method for ontology construction[13] and develop the related tool named CODE. Different from existing ontology development tools, CODE allows different kinds of developers, no matter whether they are domain experts or not, to develop ontology cooperatively. In CODE, multi-version mechanism is adopted for concurrency control and different ontology views are provided for different developers to avoid interference. The goal of CODE is to provide an open environment for more users to participate in the development of ontology, in order that the constructed ontology may be accepted by most of persons.

(3) Evolution Management Module

The evolution management module include two main sub-modules.

− Document Management. Two kinds of resources should be distinguished. They are annotated documents and original documents. It is simple to manage the original resources, while it is complicated for the annotated resources, because they may not be annotated once for all. They may need to be annotated repeatedly with the evolution of ontology. In this case one must avoid redundancy of annotation efforts. The goal of the document management in our framework will be the semi-automatic maintenance of annotated resources. When ontology changes, pattern matching may propose revision suggestions for the old annotations, and then the user may decide to ignore, modify or even delete the old annotations.

− Ontology Management. Domain changes, adaptations to different tasks, or changes in the conceptualization require modifications of the ontology. The evolution of ontologies causes operability problems, which will hamper their effective reuse. So we adopt multi-version mechanism to resolve those problems. The mechanism will make the relations between different revisions of ontologies explicitly for the semantic intercommunion among ontologies.

(4) Implementation

We have developed a Semantic Web construction platform according to the methodology of evolution based on prototype. Fig.3 is the screen snapshot of the platform.

![Fig.3. Semantic Web Construction Platform](image)

In Fig.3, the platform includes two main work areas: ontology development area and resources annotation area. The ontology development area is used to construct ontology, in which the constructed ontology is shown in tree shape. The resources annotation area is used to browse and annotate the documents, where the annotated content in the documents is shown in hyperlink. In our platform, if a hyperlink in the annotated documents is clicked, the related documents will be listed in the related resources area and the
corresponding term in the ontology will be indicated in the ontology development area. For example, the arrow from resources annotation area to ontology development area indicates the platform can show the annotated term’s position in ontology automatically and user can browse the detailed information about the term handily in the ontology development area indicates. If some content in the documents need to be added to ontology, then users can add it in the ontology development area directly.

4. Conclusion

In this paper we discuss the challenges and technologies in constructing the Semantic Web. We propose the methodology of evolution based on prototype to guide the Semantic Web construction and develop a related platform to aid developers’ work.

The methodology and platform has been applied in the Semantic Web project in the Renmin University of China, which is supported by the National Natural Science Foundation of China. The goal of the project is to explore the construction method of the Semantic Web and build a demonstrative Subject-oriented Semantic Web on economics and law. However, our existing work is an attempt to build the Semantic Web and we have a long way to go for researching, testing and improving our method and platform.

Acknowledges

The work was supported by the National Natural Science Foundation of China under Grant No. 60496325. Thanks to group members He Hu, Yan Wang, Yiyu Zhao and Wenjuan Wu et al for their works to the system realization and the Semantic Web construction.

References

Ontology Construction for Semantic Web: A Role-based Collaborative Development Method.