Proceedings of KFIS Spring Conference 2007 Vol. 17, No. 1.

Natural Language Query Framework on the Semantic Web

Jin Sung Kim

School of Business Administration, Jeonju University
Hyoja-Dong 3-1200, Wansan-Ku, Jeonju, Jeonbuk, 560-759, Korea
Tel: +82-63-220-2932, Fax: +82-63-220-2787
E-mail: kimjs@jj.ac.kr

Abstract

This study proposes a Natural Language Query Framework (NLQF) on the semantic web to support the intelligent deduction at semantic level. A large number of former researches are focused on the knowledge representation on the semantic web. However, to revitalize the intelligent agent (IA)-based automated e-business contract with human customers, there is a need for semantic level approach to the web information. To enable accessing web information at semantic level, this paper discusses the pattern of complex natural language processing at first, and then the semantic web-based natural language inference in e-business environment. The NL-based approach could help the IAs on the web to communicate with customers and other IAs with more natural interface than traditional HTML-based web information. Therefore, our proposed NLQF will be used in semantic web-based intelligent e-business contracts between customers and IAs.

Key Words : Intelligent agent, Natural language, Web information systems, E-business, Semantic web

1. Introduction

During the development of web-based intelligent information gathering, the most important is the description of the information and/or knowledge resources [3]. Recently, web-based knowledge representation and gathering is increasingly moving to the semantic web, where the information is annotated with concepts from sharing ontology, thus the semantics of web information can be understood and consumed by web agents [1]. In addition, a great number of web information sources are increasing rapidly that are available on the semantic web. However, most of former researchers have focused on how to annotate the HTML information efficiently. As an effective annotation mechanism developed by former researchers, OWL is an important formatto store the current web information.

To construct an OWL-oriented web document, ontology and schema can be used to describe the semantics of the web resources and to make the content explicit. Then the use of schema and ontology for the explication of implicit knowledge is a possible approach to overcome the limitations came out from semantic heterogeneity.Therefore, with respect to the description of information resources on the semantic web, they can be used for the identification and association of semantically corresponding information concepts [2]. However, there is still remains a barrier which protect the IAs from intelligent inference by using information resources on the web.

In the paper, to overcome the limitations, we will discuss about the OWL and Natural Language (NL) oriented knowledge transformation and inference framework on the semantic web. The rest of the paper is organized as follows. Section 2 reviews the related works. In Section 3, we propose a natural language query framework (NLQF) for the semantic web. The conclusion to this research is in Section 4.

2. Research Background

2.1. Resource Description Framework

Resource Description Framework (RDF) organizes information in a Subject-Verb-Object (SVO) or *Resource-Property-Value* triples form, thus the RDF files can be processed semantically on the semantic web [1]. RDF is a general purpose knowledge representation framework and RDF-Schema (RDFS) is the RDF vocabulary description language to be used to provide a further description mechanism to define classes or groups of related resources and the relationships between the resources [3]. RDF developed by W3C is based on an underlying model with triples made of *resource*, *property*, and *value* [4].

- A *resource* is an entity accessible by an URI on the Semantic Web (e.g. an XML document). Resources are the elements described by RDF statements.
- A *property*defines a binary relation between resources and/or its atomic values. A property enables us to attach detailed information to resources, and provide descriptions for resources.
- A *value*can be either a character string or a resource. Reification of resources using property and values enables us to transform the tripe into a resource.

Some primitives are defined in RDF-Schema (RDFS) and there are several successors of RDFS such as DARPA Agent Markup Language (DAML), Ontology Inference Layer (OIL), DAML+OIL, and Web Ontology Language (OWL). OWL has three increasingly-expressive sublanguages OWL Lite, OWL DL (Description Logic) and OWL Full. These languages all follow the RDF structure [1].

When the ontology languages are ready, the ontology for different domains can be created. And based on the ontologies, the web information can be annotated for IAs to process. Most of the current techniques focus on the RDF and RDFS-based annotations of the information on the web.

2.2. Genetic Model for Ontology Organization

Li & Ling (2006) proposed a genetic model to organize ontologies on the semantic web. Their genetic model includes the four operators such as *Inheritance, Block, Atavism* and *Mutation*. In Figure 1, each ontology reuse the primitives of ontology languages e.g. RDF, RDFS and OWL based on the Inheritance operator. And lower level ontology reuse the concepts of higher level ontologies, e.g. Employee Ontology inherits Person Ontology. The concept home_phonecontact_number mutation of the contact_number

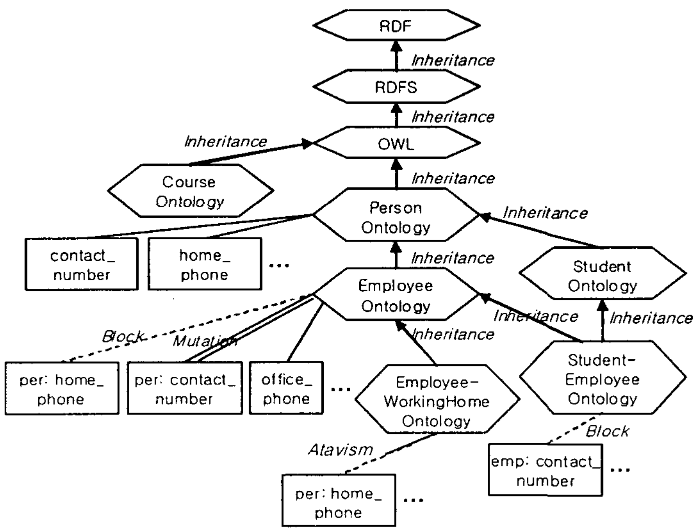

Fig. 1 Ontology hierarchy [1]

2.3. Conceptual Graph (CG)

The Conceptual Graph (CG) is a method of knowledge representation developed by Sowa [5] based on Charles Peirce's Existential Graphs and semantic networks of AI [7]. CG has a direct mapping to and from NLand a graphic notation designed for human readability. Therefore, CG express meaning in a form that is logically precise, humanly readable, and computationally tractable. Many popular graphic notations and structures ranging from type hierarchies to entity-relationship or state transition diagrams can be viewed as special cases of CGs [7]. On the CG the knowledge is divided into two parts: the *terminological knowledge* (*support*) which contains the "ground" vocabulary and the *assertional knowledge* which consist of a set of conceptual graphs built by means of the terminological knowledge [6]. The *support* provides the ground vocabulary used to build the domain knowledge base: the type of *concepts* used, the *instances*of these types, and the types of *relations* (*conceptual relationships*) linking the concepts. The concepts can be linked by means of relations and the support contains the set of *conceptual relation types*.

3. Research Methodology

The proposed mechanism has three steps such as Knowledge Manipulation, NL-Query, and Knowledge Inference. The first step Knowledge Manipulation has two sub steps Knowledge Extraction and Knowledge Storing. The second step NL-Query also has two sub steps NL Query Input and NL Representation. The third step Knowledge Inference was composed of Knowledge Retrieving and Knowledge Inference.

Where, Schema and Ontology are used to explicit the knowledge and mapping with other knowledge represented on the semantic web. The NL-KB is used to store and retrieve the knowledge extracted from the semantic web.

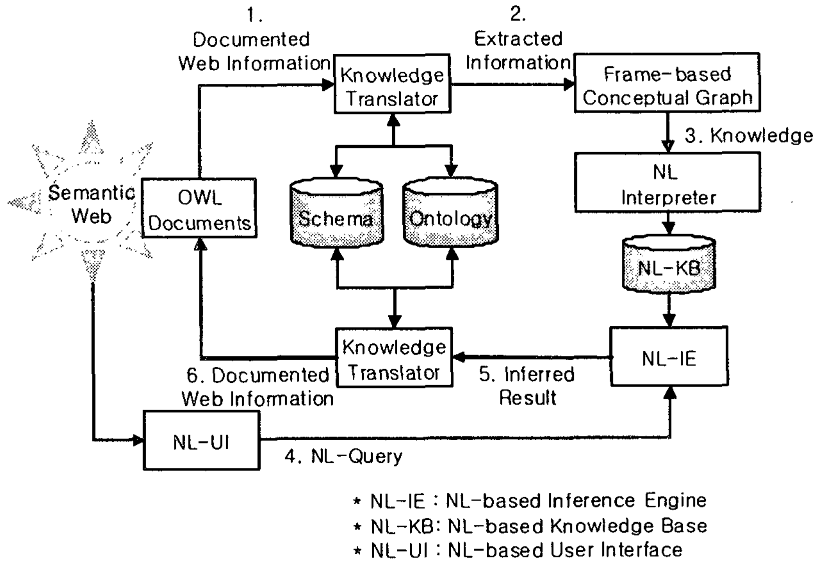

Fig. 2 Natural language query framework (NLQF) on the semantic web

3.1. Knowledge Manipulation

Knowledge Extraction:

Web information represented by OWL contains several kind of knowledge which is inferable. In this procedure, Knowledge Translator converts the OWL-based web information to Frame-based Conceptual Graph (FCG).

Knowledge Storing:

After the translation from Web Information to FCG, NL Interpreter will transform the knowledge on the FCG as a NL-based inferable knowledge. For example PROLOG+CG Artificial Intelligence (AI) language can be adapted directly to that procedure.

3.2. NL-Query

NL Query Input:

For a long time, in the field of AI, NL was used as an efficient tool to manage the complicate human knowledge manipulation and inference. At the same time, semantic network was used to represent the experts'network concept was adapted to the knowledge representation on the semantic web. All of knowledge represented by semantic network is inferable by using NL tools. The revised versions of NL tools are applicable to this procedure.

NL Representation:

The results inferred by NL-IE should be represented on the semantic web to help the user'understanding. Furthermore, the agents on the semantic web will comprehend the information and can react to that.

3.3. Knowledge Inference

Knowledge Retrieving:

To infer with the knowledge stored in NL-KB, knowledge retrieving procured is required. In this procedure, the knowledge interpreted by NL-IE is compared with the users

Knowledge Inference:

The capability of knowledge inference is rely on the NL-IE. Recently, PROLOG and LISP were revised as an Object-Oriented programming language. Therefore, FCG-based knowledge retrieving and inference are possible in this procedure.

4. Concluding Remarks

In this paper, the proposed mechanism uses three steps such as Knowledge Manipulation, NL-Query and Knowledge Inference to extract and infer the knowledge on the semantic web. For a valid inference on the semantic web, their schema and ontologies are adapted firstly. Then the extracted knowledge is converted to FCG which can manipulate experts'srequest which is composed by NL-based query interface NL-UI. However, the proposed NLQF has so many limitations to adapt in the real-world semantic web. Most important things are the successful combination of NL-UI and NL-IE with OWL-based semantic web. In the further research, the combination and detailed and efficient inference procedures should be developed.

References

[1] C. Li and T. W. Ling, From XML to Semantic Web,*Database Systems for Advanced Applications*, Berlin: Springer, *The 10th International Conference on Database Systems for Advanced Applications*(DASFAA 2005), LNCS 3453, pp. 582-587, Beijing, China, 17-20 April 2005.

[2] J. Gu, B. Hu, and Y. Zhou, Semantic Query Planning Mechanism on XML Based Web Information Systems,*The 7th International Conference on Web Information Systems Engineering* (WISE 2006), LNCS 4256, pp. 194-205, Wuhan, China, 2006.

[3] H. Yao and L. Etzkorn, L., Automated conversion between different knowledge representation formats, *Knowledge-Based Systems*, 19, pp. 404-412, 2006.

[4] O. Corby, R. Dieng, and C. Hébert, A conceptual graph model for W3C resource description framework, *Proceedings of the 8th International Conference on Conceptual Structures (ICCS '2000)*, Berlin, August 2000.

[5] J.F. Sowa, *Conceptual structures: Information processing in minds and machines*, Addison-Wesley, Reading, MA, 1984.

[6] J. Dibie-Barthélemy, O. Haemmerlé, and E. Salvat, A semantic validation of conceptual graphs, *Knowledge-Based Systems*, 19, pp. 489-510, 2006.

[7] A. Kayed and R.M. Colomb, Using BWW model to evaluate building ontologies in CGs formalism, *Information Systems*, 30, pp. 379-398, 2005.