Challenges for Information Systems Towards The Semantic Web

Artem Chebotko, Shiyong Lu, Farshad Fotouhi
Department of Computer Science
Wayne State University, Detroit, MI 48202
{artem, shiyong, fotouhi}@cs.wayne.edu

The Semantic Web [1, 2, 3] is the next generation Web, in which information is structured with well-defined semantics, enabling better cooperation of machine and human effort. The Semantic Web is not a replacement, but an extension of the current Web.

The Extensible Markup Language (XML) [4] will play an important role as a syntactic foundation for the Semantic Web. XML provides a set of constructs for representing data on the Web, but XML only supports limited semantics via its nested structure. To illustrate the limitations of XML, consider the two XML documents shown in Figure 1. Suppose we would like to retrieve all the *persons* from these two documents. To return the complete results, the query processor needs to understand that each "student" or "professor" is also a "person". In addition, to integrate the search results from both documents, the query processor also needs to understand that "familyName" in document 1 is equivalent to "lastName" in document 2, but "date" in document 1 is different from document 2 — while "date" in document 1 means a date of birth, "date" in document 2 means the date when a student or a professor enters a university. These semantic relationships between different elements are not supported by XML, and current XML query languages such as XQuery and XPath do not support the correct processing of the above query.

```
<person id="1">
      <firstName>Emily</firstName>
      <familyName>Lu</familyName>
      <date>2001-01-15</date>
</person>
<person id="2" >
      <firstName>Shiyong</firstName>
      <familyName>Lu</familyName>
      <date>1972-08-14</date>
</person>
```

```
<student id="3">
      <firstName>Artem</firstName>
      <lastName>Chebotko</lastName>
      <date>Fall 2003</date>
      <gpa>3.0</gpa>
</student>
<professor id="4">
      <firstName>Farshad</firstName>
      <lastName>Fotouhi</lastName>
       <date>Fall 2003</date>
</professor>
```

Figure 1: XML document 1 and 2

To overcome the limitations of XML, ontologies [5] will be used to represent various concepts and the relationships between these concepts for a domain. For example, one can represent that a *"student" is subClassOf "person"*, and *"familyName" equivalentProperty "lastName"* in an ontology and use the ontology to guide the query processor to process the above query properly.

What challenges does the Semantic Web pose to information systems? In this article, we share our perspective to this question based on our experience with building the Semantic Web for endangered languages [6]. We focus our discussion on the following challenges.

The challenge of the development of a domain ontology. Ontologies now play an important role in enabling the Semantic Web. Semantic Web communities develop ontologies in their domains, which involves many experts in the same domain and each of them might have his own perspective (a social challenge). This requires a collaborative ontology development environment (a technical challenge) that will enable version control, proposal and release control, and coordination and collaboration support. The development of such an environment is a challenge. Most of the ontology development tools today, like Protégé-2000 [7], are personal ontology editors and they lack these functionalities. To the best of our knowledge, the only tool that has the collaboration support is OntoEdit [8], but it needs further improvements, such as rights- and user- management layers.

The challenge of ontology mapping, alignment and merging. Over the past few years, people have come to the consensus that even in one domain, it is very difficult to enforce a single ontology to be used for each data source. Instead, people should have the full freedom to use their own proprietary ontology to annotate their data source, and then, if they are willing, provide additional mapping to map them to a standard (central) domain ontology to support data interoperability and queries across data sets. This mapping is a challenge because there might exist heterogeneities between ontologies: syntactic, schematic and semantic. The mapping process might include not only ontology alignment to make ontologies coherent, but also ontology merging to add new terms in a central ontology. For further information on available algorithms and tools in this research area, please refer to [9, 10].

The challenge of annotation management. One challenge of annotation management is the integration of a domain ontology and an annotation tool. For each domain, an annotation tool should be customized based on a given ontology and the requirements of the users. For example, linguists require the support of International Phonetic Alphabet in an annotation tool and this requirement is not specified in the domain ontology. A second challenge is the development of a universal but customizable annotation tool for several domains since it requires an open architecture and the support of common annotation functionalities needed for various domains. A third challenge is the development of a shared annotation capability, where the support of concurrency control is needed. Shared data annotation is especially efficient when users simultaneously annotate data segments that do not intersect. For example, one person annotates the gestures of a speaker, while another person adds linguistic characteristics of the speech. Two different ontologies can be used for this purpose and each person may be familiar with only one of them. User rights play an important role in client co-work coordination. One of the projects in the shared annotation direction is Annotea [11], which implements a client-server architecture, where server stores annotated data and manages its access. Finally, the consistency maintenance between annotations and the corresponding ontology is also a challenge. When ontology evolves, the annotations might need to change accordingly.

The challenge of ontology-based information retrieval. Annotated data is not useful if one cannot search through it. One of the promises of the Semantic Web is high precision. Search engines now should exploit available semantics and ontology reasoning to return not only precise results, but also specify meaningful relationships between them. New opportunities also require new approaches to query refinement and user interface tactics [12]. But the major challenge is searching across data sets annotated using different ontologies. Generally, there can be several ontologies for one domain since each domain can be modeled by several ontologies or a domain may require the usage of several ontologies. As a result, not only ontology mapping is required, but also user query mapping will be needed.

Conclusions and outlook

In this article, we discussed some challenges that we experienced working on the Semantic Web for linguistics. These are (1) the challenge of the development of a domain ontology; (2) the challenge of ontology mapping, alignment and merging; (3) the challenge of annotation management; and (4) the challenge of ontology-based information retrieval. We conclude our article with a comment from Tim Berners-Lee on the state-of-the-art of the Semantic Web: “It's not unlike the early days of the Web, when once people saw how it worked, they understood its power. We're entering that phase now, where people can see the beginnings of the Semantic Web at work.”

References

1. T. Berners-Lee. “Semantic Web Road Map”. IW3C Design Issues. Cambridge, MA: W3C. Available at: http://www.w3.org/DesignIssues/Semantic.html. 1998.
2. T. Berners-Lee, James Hendler and Ora Lassila. “The Semantic Web”. Scientific American, May 2001.
3. S. Lu, M. Dong and F. Fotouhi, "The Semantic Web: Opportunities and Challenges for Next-Generation Web Applications". International Journal of Information Research, 7(4), 2002.
4. T. Bray, J. Paoli, C. Sperberg-McQueen, E. Maler, and F. Yergeau, “Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation 04 February 2004”. 2004.
5. T. Gruber, "A translation approach to portable ontology specifications." Knowledge Acquisition, 6(2), 199-221. 1993.

6. S. Lu, D. Liu, F. Fotouhi, M. Dong, R. Reynolds, A. Aristar, M. Ratliff, G. Nathan, J. Tan, and R. Powell, "Language Engineering For The Semantic Web: A Digital Library For Endangered Languages", International Journal of Information Research, 2004.
7. Protégé-2000. Protégé Project, http://protege.stanford.edu/index.html. 2004.
8. Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, D. Wenke, "OntoEdit: Collaborative Ontology Development for the Semantic Web". ISWC 2002.
9. D. McGuinness, R. Fikes, J. Rice, and S. Wilder, "An environment for merging and testing large ontologies". KR-2000.
10. N. Noy and M. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment". AAAI 2000.
11. M. Koivunen, R. Swick, E. Prud'hommeaux. "Annotea Shared bookmarks". KCAP 2003.
12. E. García and M. Sicilia, "User Interface Tactics in Ontology-Based Information Seeking". PsychNology Journal, 2003.