A Portrait of the Semantic Web in Action

Jeff Heflin and James Hendler
IEEE Intelligent Systems

December 6, 2010
Hyewon Lim
Outline

- Introduction
- Overview of Semantic Web Language
- Producing Semantic Markup
- Integrating Resources
- Querying the Semantic Web
- Conclusion
Introduction

- World Wide Web
  - Enormous quantity of data
  - Machines cannot understand the web pages
Creating a Semantic Web

- Change the Web to make it more understandable by machines
- The key lies in the development of semantically-enriched languages
  - RDF, SHOE, DAML

Semantic Web

- An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (T. Berners-Lee, 2001)
  - Web of Knowledge
  - Web of Data
Overview of Semantic Web Language

- A Semantic Web language
  - must describe meaning in a machine readable way
  - the ability to specify a vocabulary & to formally define the vocabulary
    - It can be used in automated reasoning

- Syntaxes of Semantic Web language
  - Based on existing standards such as HTML or XML
    - Integration with other Web technologies is possible

<table>
<thead>
<tr>
<th>HTML</th>
<th>XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;table&gt;</td>
<td>&lt;english&gt;</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>&lt;tr&gt;</td>
<td>&lt;name&gt;홍길동&lt;/name&gt;</td>
</tr>
<tr>
<td>&lt;td&gt;홍길동&lt;/td&gt;</td>
<td>&lt;score&gt;90&lt;/score&gt;</td>
</tr>
<tr>
<td>&lt;td&gt;90&lt;/td&gt;</td>
<td>&lt;grade&gt;A&lt;/grade&gt;</td>
</tr>
<tr>
<td>&lt;td&gt;A&lt;/td&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/tr&gt;</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>&lt;/english&gt;</td>
</tr>
<tr>
<td>&lt;/table&gt;</td>
<td></td>
</tr>
</tbody>
</table>
Overview of Semantic Web Language

- Web is decentralized
  - The language must allow for the definition of diverse, and potentially conflicting, vocabularies

- Rapid evolution of the Web
  - The language must allow for the vocabularies to evolve as human understanding of their use improves

- Immense size of the Web
  - Scalability
Overview of Semantic Web Languages

- A Semantic Web vocabulary can be formally specified using an ontology or schema

- Ontology
  - From the study of *being* or *existence*: “What is there?”
  - A data model that represents a set of concepts within a domain and the relationships between those concepts (in Wikipedia)
  - Enable computers to deal with semantics
  - Sharable and extensible
  - The potential richness of an ontology’s definitions is determined by the expressivity of the language
    - Concepts, properties, relationships, constraints
Overview of Semantic Web Languages

- Major differences between the Semantic Web languages
  - 1. RDF (Resource Description Framework) with RDF Schema
     - The least expressive
     - Based on a semantic network model
  - 2. SHOE (Simple HTML Ontology Extensions)
     - Based on a frame system
     - Allows Horn clause axioms
       - Can be used to define things not possible in RDF
     - Deal with the problems of a dynamic, distributed environment
  - 3. OIL (the Ontology Inference Language)
     - Based on a frame system augmented with description logic
  - 4. DAML (DARPA Agent Markup Language)
     - Combine the best features of RDF, SHOE, and OIL
Producing Semantic Markup

- Describing a set of web pages using a Semantic Web language

- To make content machine understandable
  - The web pages must contain semantic markup

- First step
  - Consider the domain of the pages
  - Choose an appropriate ontology

- Our running example will use the SHOE language
Knowledge Annotator

- Allow markup to be created by making selections and filling in forms
Producing Semantic Markup
- Generating Markup on a Large Scale

- Automatically generating semantic markup
  - Tedious to use an authoring tool to generate large scale markup

- Web pages have regular structure
  - Can be mapped to an ontology

* Figures from: An Automated Approach for Retrieving Hierarchical Data from HTML Tables
  Seung-Jin Lim, Yiu-Kai Ng (CIKM99)
Running SHOE

- Extract SHOE markup from the web pages
- Easy to regenerate new SHOE markup if the content of the page changes
Different people describes the same entity
- A fundamental problem in distributed system
- Use a common identifier (or key)

Irregularities in a page’s HTML
- Can often cause fields or records to be extracted improperly

Limitation
- Cannot extract information from important resources
  - CiteSeer
    - An index of online CS publications
    - Multi-step process to get a particular publication page
      - Publication SHOE Maker (PSM)
- Producing Semantic Markup
- Generating Markup on a Large Scale

- Other extraction tools
  - Include machine learning or natural language processing techniques
  - Transform a simple XML vocabulary using stylesheets

- If the author of a web page provides semantic markup, then the process can be even easier
  - E.g., keep Web data in database
Integrating Resources

- Crawl the Web and store the information in a central repository
  - Exposé
    - A web-crawler that searches for web pages with SHOE markup
    - Parse each web page
    - The cost function determines the order of the traversal
      - Based on the distance from the start node
        - Paths through non-SHOE pages are more expensive
        - Paths that stay within the same directory on the same server are cheaper

- Knowledge base
  - Determine the query capabilities of the system

- Backend knowledge representation system
  - Tradeoffs between query response time and the degree of inference
    - Parka API vs. XSB vs. database
Querying the Semantic Web

- SHOE Search tool

Translates the user’s query into a similar search engine query & submits it to the popular search engines
Querying the Semantic Web

- Some individuals may have multiple URLs
  - Partially alleviated if the language included the ability to specify identifier equivalence
    → absent from SHOE, but present in DAML

- Semantic Search
Conclusion

- Simple Semantic Web system

- In many case, simple extraction tools can generate accurate markup with minimal human effort
Conclusion

- Achieve the vision of the Semantic Web
  - Become much easier to locate useful information on the Internet
  - Integration of diverse resources will be simplified

- Obstacles to overcome
  - Need better schemes for ensuring interoperability between independently developed ontologies
  - Approaches for determining who and what to trust

- When user will add semantics to web pages easily, then the next Internet revolution will arise