A Portrait of the Semantic Web in Action

Jeff Heflin and James Hendler
University of Maryland

2008. 11. 27

Kyusik Kim

Wireless Networks lab
Contents

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

❖ World Wide Web
   ■ Enormous quantity of data
     ▪ difficult to locate, organize, integrate the available information
   ■ We need to hand off portions of the tasks to machines

❖ Problems
   ■ Natural language processing is unsolved
   ■ Machines cannot understand web pages

❖ Solutions
   ■ Change current web to make more understandable by machines
   ■ development of semantically-enriched languages
     → Semantic Web languages

❖ Semantic Web
   ■ “The Semantic Web is a web of data, in some ways like a global database.” by Tim Berners Lee
Question about Semantic Web

- How it can be used in practice
- How are the semantic descriptions generated
- How are these descriptions discovered by agents
- How can the Semantic Web be queried
### HTML vs. XML vs. Semantic Web

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

- **Semantic Web languages**
  - must describe meaning in machine readable way
    - needs the ability to
      - specify a vocabulary
      - formally define vocabulary

- **Semantic Web vs Knowledge representation (KR) in AI**
  - Semantic Web languages are greatly influenced by KR
  - Difference
    - Semantic Web languages are based on existing standards
      - HTML, XML
    - integration with other web technologies is possible
Ontology

- **Semantic Web vocabulary can be formally specified using ontology**

- **Expressivity of language for building the Semantic Web**
  - First step for building the Semantic Web
  - Define class taxonomies, properties, relationship, axioms
  - Database for communication between web agents

- **Sharable**
  - different users can agree to use the same definition

- **Extensible**
  - different users can agree on definitions, add terms and definitions
Examples of Semantic Web Languages

- **RDF (Resource Description Framework)**
  - least expressive
  - based on a semantic network model
    - special links for defining category, property taxonomies, etc

- **SHOE (Simple HTML Ontology Extension)**
  - based on a frame system
  - allows Horn clause axioms
  - focuses on dealing with the problems of dynamic, distributed environment

- **OIL (Ontology Interchange Language)**
  - based on a frame system augmented with description logic

- **DAML (DARPA Agent Markup Language)**
  - combines the best features of RDF, SHOE, OIL
Producing Semantic Markup

- **Descriptions that use the terminology defined in ontologies**

- **To make content of web pages machine understandable**
  - the web pages must contain semantic markup

- **The first step to produce semantic markup**
  - consider the domain of the web pages
  - choose an appropriate ontology

- **Authoring tools**
  - semantic markup can be added using a simple text editor → error prone
  - allow markup to be created by making selections and filling in forms
  - Ex. Knowledge Annotator (in SHOE project)
Knowledge Annotator

- has an interface that displays instances, ontologies, claims
- user can add, edit, remove any of the objects
- provides various methods to view the knowledge in the document
  - the source of HTML, logical notation, etc.
- error checking ability
Generating Markup on a Large Scale

- Limitation using authoring tools to generate large amount of markup

- Automatically generating semantic markup
  - Mapping regular structure of web pages to semantic markup
  - Translation tools can be used (ex. Running SHOE)

- Running SHOE
  - Tool that helps users to extract SHOE markup from regularly structured web pages

Problem
- Distributed system
  - When markup authored by different people describes the same entity
- Common identifier (key) – URL
  - Different URL for same object
Running SHOE

<table>
<thead>
<tr>
<th>Id</th>
<th>Start</th>
<th>End</th>
<th>Type</th>
<th>Category</th>
<th>Name</th>
<th>Arg1</th>
<th>Arg2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Facultys</td>
<td>cs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>members</td>
<td>cs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashok K. Agrawal</td>
<td><a href="http://www.cs.umd.edu/dept/faculty.html">http://www.cs.umd.edu/dept/faculty.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>John Martin McNamee</td>
<td><a href="http://www.cs.umd.edu/CHCN/people/bar.html">http://www.cs.umd.edu/CHCN/people/bar.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>William A. Arbaugh</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Victor R. Basili</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Benjamin S. Bederson</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Samir Bhattcharjee</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Rajeev Kumar Banerji</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Sudeshan S. Chandrasekhar</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Rama Chellappa</td>
<td><a href="http://www.cs.umd.edu/faculty/chellaps.html">http://www.cs.umd.edu/faculty/chellaps.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Yehuda Chud</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Larry S. Davis</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Bonnie Dar</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Howard C. Elman</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Christine Faloutsos</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Molly Franklin</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>Michael J. Franklin</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>John D. Gannon</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
<tr>
<td>William Gassarch</td>
<td><a href="http://www.cs.umd.edu/adult-web2.html">http://www.cs.umd.edu/adult-web2.html</a></td>
<td>Faculty</td>
</tr>
</tbody>
</table>
Limitation of Running SHOE

- cannot extract information from some important resources

- **CiteSeer** ([http://citeseer.ist.psu.edu](http://citeseer.ist.psu.edu))
  - index of online CS publications
  - Problem using Running SHOE
    - Multistep process
      - issuing a query to one page, viewing a result page, selecting a result to get a page about a particular publication

- **Publication SHOE Maker (PSM)**
  - tool for extracting SHOE from CiteSeer
  - issues query to get publications
    - retrieves some fixed number of publication pages from results
Integrating Resources

❖ How to integrate information with semantic markup

❖ Crawling the web and storing the information in a central repository
  ■ Exposé
    ■ Web crawler for web pages with SHOE markup
    ■ graph traversal using cost function
    ■ parse each web page
    ■ stores SHOE category, relation statement, ontology information in Knowledge base (KB)

❖ Knowledge representation system for KB
  ■ appropriate knowledge representation system must be chosen
  ■ Example
    ■ Parka - High performance frame system
    ■ XSB - Deductive database
    ■ OKBC - Compliant knowledge bases
Querying the Semantic Web

- **General Purpose Query Tools**
  - users submit structured queries
  - users open documents by clicking on the URLs in the results
  - Ex) SHOE Search
**SHOE Search**

- **Users can browse the Semantic Web**
  - choose ontology
  - choose class of desire object
  - type in desired values for properties
  - issue a query and receive a set of results in a tabular form
  - web page will open by double-clicking on a URL

- **For the web pages that were not described by SHOE markup**
  - SHOE Search have a web search feature
    - translate the user’s query into a similar search engine query
    - submit to one of the popular search engines
Semantic Search

- Search engine using technologies described so far

Figure 4. Semantic Search Main Page
Summary

- How to build Semantic Web and use
  - Add markup to web pages
  - Web crawler discovers and stores in repository
  - Query
Conclusion

 Vision
  ■ Much easier to locate useful information on the Internet
  ■ Integration of diverse resources will be simplified

 Obstacle
  ■ Interoperability between independently developed ontologies
  ■ Determining who and what to trust
Thank you!!!!!