Overview of the Semantic Web

Jian-hua Yeh (葉建華)

真理大學資訊科學系助理教授

au4290@email.au.edu.tw

真理大學
Aletheia University

Lecture Outline

1. Today's Web
2. The Semantic Web Impact
3. Semantic Web Technologies
4. A Layered Approach

Today's Web

- Most of today's Web content is suitable for human consumption
  - Even Web content that is generated automatically from databases is usually presented without the original structural information found in databases
- Typical Web uses today people's
  - seeking and making use of information, searching for and getting in touch with other people, reviewing catalogs of online stores and ordering products by filling out forms

Keyword-Based Search Engines

- Current Web activities are not particularly well supported by software tools
  - Except for keyword-based search engines (e.g. Google, AltaVista, Yahoo)
- The Web would not have been the huge success it was, were it not for search engines

Problems of Keyword-Based Search Engines

- High recall, low precision.
- Low or no recall
- Results are highly sensitive to vocabulary
- Results are single Web pages
- Human involvement is necessary to interpret and combine results
- Results of Web searches are not readily accessible by other software tools

The Key Problem of Today's Web

- The meaning of Web content is not machine-accessible: lack of semantics
- It is simply difficult to distinguish the meaning between these two sentences:

    I am a professor of computer science.

    I am a professor of computer science, you may think. Well, . . .

The Semantic Web Approach

- Represent Web content in a form that is more easily machine-processable.
- Use intelligent techniques to take advantage of these representations.
- The Semantic Web will gradually evolve out of the existing Web, it is not a competition to the current WWW

Lecture Outline

1. Today’s Web
2. The Semantic Web Impact
3. Semantic Web Technologies
4. A Layered Approach

The Semantic Web Impact – Knowledge Management

- Knowledge management concerns itself with acquiring, accessing, and maintaining knowledge within an organization
- Key activity of large businesses: internal knowledge as an intellectual asset
- It is particularly important for international, geographically dispersed organizations
- Most information is currently available in a weakly structured form (e.g. text, audio, video)

Limitations of Current Knowledge Management Technologies

- *Searching information*
  - *Keyword-based search engines*
- *Extracting information*
  - human involvement necessary for browsing, retrieving, interpreting, combining
- *Maintaining information*
  - inconsistencies in terminology, outdated information.
- *Viewing information*
  - Impossible to define views on Web knowledge

Semantic Web Enabled Knowledge Management

- Knowledge will be organized in conceptual spaces according to its meaning.
- Automated tools for maintenance and knowledge discovery
- Semantic query answering
- Query answering over several documents
- Defining who may view certain parts of information (even parts of documents) will be possible.

The Semantic Web Impact – B2C Electronic Commmerce

- A typical scenario: user visits one or several online shops, browses their offers, selects and orders products.
- Ideally humans would visit all, or all major online stores; but too time consuming
- Shopbots are a useful tool

Limitations of Shopbots

- They rely on wrappers: extensive programming required
- Wrappers need to be reprogrammed when an online store changes its outfit
- Wrappers extract information based on textual analysis
  - Error-prone
  - Limited information extracted

Semantic Web Enabled B2C Electronic Commerce

- Software agents that can interpret the product information and the terms of service.
  - Pricing and product information, delivery and privacy policies will be interpreted and compared to the user requirements.
- Information about the reputation of shops
- Sophisticated shopping agents will be able to conduct automated negotiations

The Semantic Web Impact – B2B Electronic Commerce

- Greatest economic promise
- Currently relies mostly on EDI
  - Isolated technology, understood only by experts
  - Difficult to program and maintain, error-prone
  - Each B2B communication requires separate programming
- Web appears to be perfect infrastructure
  - But B2B not well supported by Web standards

Semantic Web Enabled B2B Electronic Commerce

- Businesses enter partnerships without much overhead
- Differences in terminology will be resolved using standard abstract domain models
- Data will be interchanged using translation services.
- Auctioning, negotiations, and drafting contracts will be carried out automatically (or semi-automatically) by software agents

Lecture Outline

1. Today's Web
2. The Semantic Web Impact
3. Semantic Web Technologies
4. A Layered Approach

Semantic Web Technologies

- Explicit Metadata
- Ontologies
- Logic and Inference
- Agents

On HTML

- Web content is currently formatted for human readers rather than programs
- HTML is the predominant language in which Web pages are written (directly or using tools)
- Vocabulary describes presentation

An HTML Example

<h1>Agilitas Physiotherapy Centre</h1>

Welcome to the home page of the Agilitas Physiotherapy Centre. Do you feel pain? Have you had an injury? Let our staff Lisa Davenport, Kelly Townsend (our lovely secretary) and Steve Matthews take care of your body and soul.

<h2>Consultation hours</h2>

Mon 11am - 7pm

Tue 11am - 7pm

Wed 3pm - 7pm

Thu 11am - 7pm

Fri 11am - 3pm<p>

But note that we do not offer consultation during the weeks of the State Of Origin games.

Problems with HTML

- Humans have no problem with this
- Machines (software agents) do:
  - How distinguish therapists from the secretary,
  - How determine exact consultation hours
  - They would have to follow the link to the State Of Origin games to find when they take place.

A Better Representation

```xml
<company>
    <treatmentOffered>Physiotherapy</treatmentOffered>
    <companyName>Agilitas Physiotherapy Centre</companyName>
    <staff>
        <therapist>Lisa Davenport</therapist>
        <therapist>Steve Matthews</therapist>
        <secretary>Kelly Townsend</secretary>
    </staff>
</company>
```

Explicit Metadata

- This representation is far more easily processable by machines
- Metadata: data about data
  - Metadata capture part of the *meaning of data*
- Semantic Web does not rely on text-based manipulation, but rather on machine-processable metadata

Ontologies

The term ontology originates from philosophy

- The study of the nature of existence

Different meaning from computer science

- An ontology is an explicit and formal specification of a conceptualization

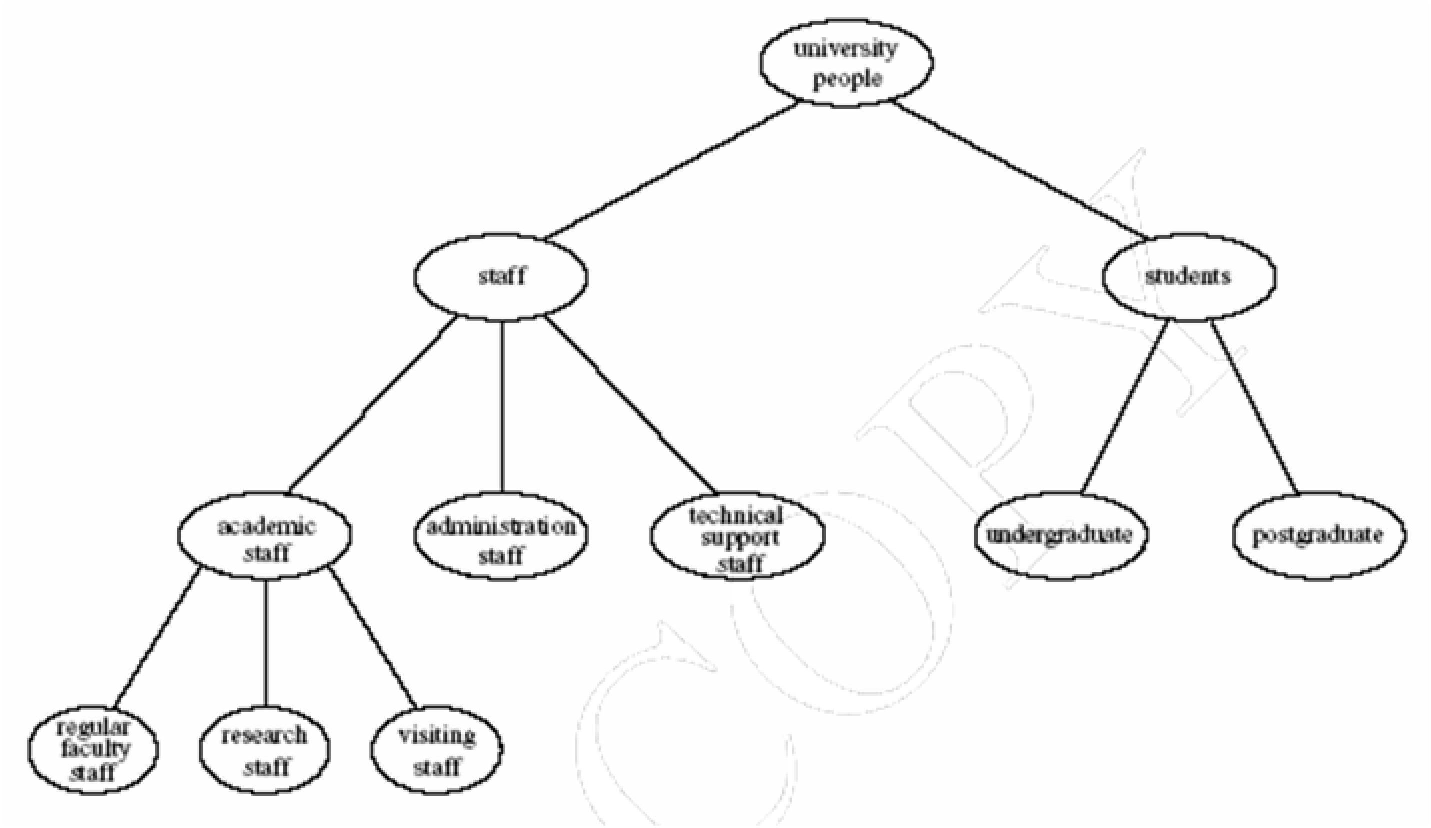
Typical Components of Ontologies

- Terms denote important concepts (classes of objects) of the domain
  - e.g. professors, staff, students, courses, departments
- Relationships between these terms: typically class hierarchies
  - a class C to be a subclass of another class C' if every object in C is also included in C'
  - e.g. all professors are staff members

Further Components of Ontologies

- Properties:
  - e.g. X teaches Y
- Value restrictions
  - e.g. only faculty members can teach courses
- Disjointness statements
  - e.g. faculty and general staff are disjoint
- Logical relationships between objects
  - e.g. every department must include at least 10 faculty

Example of a Class Hierarchy



The Role of Ontologies on the Web

- Ontologies provide a shared understanding of a domain: semantic interoperability
  - overcome differences in terminology
  - mappings between ontologies
- Ontologies are useful for the organization and navigation of Web sites

The Role of Ontologies in Web Search

- Ontologies are useful for improving the accuracy of Web searches
  - search engines can look for pages that refer to a precise concept in an ontology
- Web searches can exploit generalization/ specialization information
  - If a query fails to find any relevant documents, the search engine may suggest to the user a more general query.
  - If too many answers are retrieved, the search engine may suggest to the user some specializations.

Web Ontology Languages

RDF Schema

- RDF is a data model for objects and relations between them
- RDF Schema is a vocabulary description language
- Describes properties and classes of RDF resources
- Provides semantics for generalization hierarchies of properties and classes

Web Ontology Languages (2)

OWL

- A richer ontology language
- relations between classes
  - e.g., disjointness
- cardinality
  - e.g. “exactly one”
- richer typing of properties
- characteristics of properties (e.g., symmetry)

Logic and Inference

- Logic is the discipline that studies the principles of reasoning
- Formal languages for expressing knowledge
- Well-understood formal semantics
  - Declarative knowledge: we describe what holds without caring about how it can be deduced
- Automated reasoners can deduce (infer) conclusions from the given knowledge

An Inference Example

prof(X) $\rightarrow$ faculty(X)

faculty(X) $\rightarrow$ staff(X)

prof(michael)

We can deduce the following conclusions:

faculty(michael)

staff(michael)

prof(X) $\rightarrow$ staff(X)

Logic versus Ontologies

- The previous example involves knowledge typically found in ontologies
  - Logic can be used to uncover ontological knowledge that is implicitly given
  - It can also help uncover unexpected relationships and inconsistencies
- Logic is more general than ontologies
  - It can also be used by intelligent agents for making decisions and selecting courses of action

Tradeoff between Expressive Power and Computational Complexity

- The more expressive a logic is, the more computationally expensive it becomes to draw conclusions
  - Drawing certain conclusions may become impossible if non-computability barriers are encountered.
- Our previous examples involved rules "*If conditions, then conclusion*," and only finitely many objects
  - This subset of logic is tractable and is supported by efficient reasoning tools

Inference and Explanations

- Explanations: the series of inference steps can be retraced
- They increase users' confidence in Semantic Web agents: "Oh yeah?" button
- Activities between agents: create or validate proofs

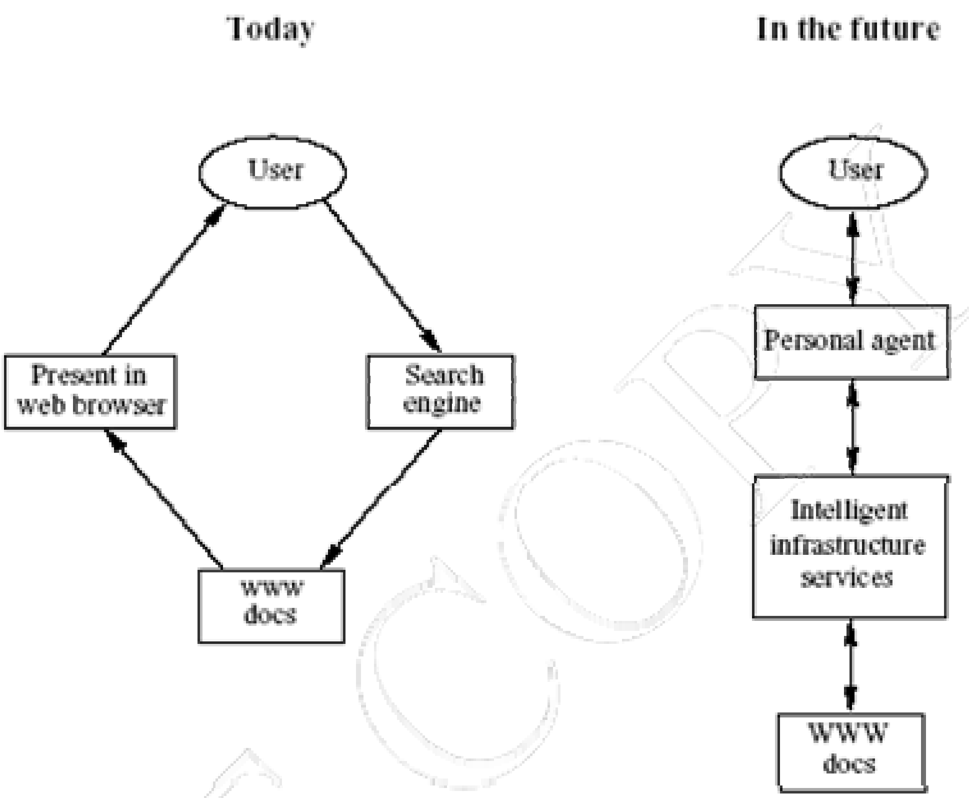
Typical Explanation Procedure

- Facts will typically be traced to some Web addresses
  - The trust of the Web address will be verifiable by agents
- Rules may be a part of a shared commerce ontology or the policy of the online shop

Software Agents

- Software agents work autonomously and proactively
  - They evolved out of object oriented and compontent-based programming
- A personal agent on the Semantic Web will:
  - receive some tasks and preferences from the person
  - seek information from Web sources, communicate with other agents
  - compare information about user requirements and preferences, make certain choices
  - give answers to the user

Intelligent Personal Agents



Semantic Web Agent Technologies

- Metadata
  - Identify and extract information from Web sources
- Ontologies
  - Web searches, interpret retrieved information
  - Communicate with other agents
- Logic
  - Process retrieved information, draw conclusions

Semantic Web Agent Technologies (2)

- Further technologies (orthogonal to the Semantic Web technologies)
  - Agent communication languages
  - Formal representation of beliefs, desires, and intentions of agents
  - Creation and maintenance of user models.

Lecture Outline

1. Today's Web
2. The Semantic Web Impact
3. Semantic Web Technologies
4. A Layered Approach

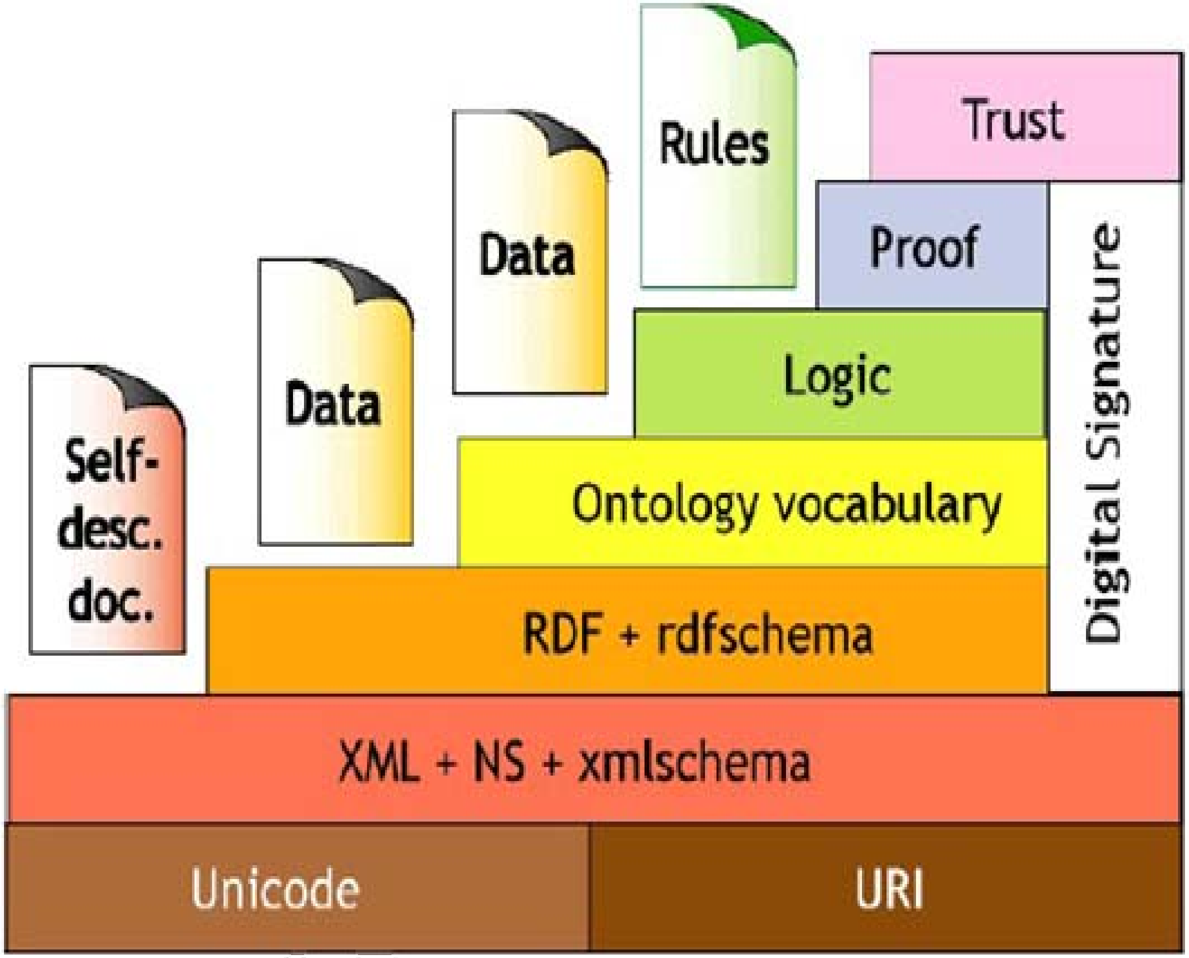
A Layered Approach

- The development of the Semantic Web proceeds in steps
  - Each step building a layer on top of another

Principles:

- Downward compatibility
- Upward partial understanding

The Semantic Web Layer Tower



Semantic Web Layers

- XML layer
  - Syntactic basis
- RDF layer
  - RDF basic data model for facts
  - RDF Schema simple ontology language
- Ontology layer
  - More expressive languages than RDF Schema
  - Current Web standard: OWL

Semantic Web Layers (2)

- Logic layer
  - enhance ontology languages further
  - application-specific declarative knowledge
- Proof layer
  - Proof generation, exchange, validation
- Trust layer
  - Digital signatures
  - recommendations, rating agencies ….