Semantic Web and Knowledge Management

Ching-Long Yeh 葉慶隆
Department of Computer Science and Engineering
Tatung University
Taipei, Taiwan
Email: chingyeh@cse.ttu.edu.tw
http://www.cse.ttu.edu.tw/chingyeh
Web Technology Evolution
Web Technology

- HTML
- XML
- Electronic Business
- Semantic Web
HTTP (Hyper Text Transport Protocol)
HTML (Hyper Text Markup Language)
URL (Uniform Resource Locator)

- Human-to-machine interaction
- Information explosion
Interaction Using XML

XML (eXtensible Markup Language)
SOAP (Simple Object Access Protocol)

- Machine-to-machine interaction
- Service automation
Web Service Architecture

- Service registry
  - Service descriptions
  - Find WSDL, UDDI
  - Publish WSDL, UDDI
  - Bind
  - Service requester
  - Service provider
  - Services
  - Service descriptions
Electronic Business

- Application-to-Application
- Business Process Automation
- RosettaNet
- ebXML

Company A

- ERP
- Business processes, business documents
- Transport, routing, packaging

Company B

- ERP
- Business processes, business documents
- Transport, routing, packaging
The Semantic Web is a vision:

the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.

See “W3C Semantic Web Activity,” by Marja-Riitta Koivunen, for more descriptions.
Semantic Web

• The Semantic Web is the new generation of the World Wide Web, based on the semantic network knowledge representation formalism, which enables packaging information in the form of object-attribute-value statements, so called triplets.
• By assuming that terms used in these statements are based on the formally specified meaning (for the community of interest), i.e. ontologies, these triplets can be semantically processed by machine agents.
The Semantic Web Layered Architecture

Tim Berners-Lee: “Axioms, Architecture and Aspirations”
W3C all-working group plenary Meeting
28 February 2001
(http://www.w3.org/2001/Talks/0228-tbl/slide5-0.html)
The Big Picture of SW
An Architecture of Semantic Community Web Portals

Storing
Knowledge Warehouse

Inferencing
Inference Engine

Providing
Metadata Wrapper Fact Editor

Accessing
Navigation & Querying

Structuring
Ontology

(From: http://www9.org/w9cdrom/134/134.html)
A KM Framework over SW

(From: http://www2002.org/CDROM/poster/130.pdf)
Remainder

• Basics of Ontology
• Semantic Web Languages
• Semantic Web Portal
• RDF Annotation
• An RDF Store
• Discovery Services
• Related Work
• Conclusions
Basics of Ontology
Why Develop an Ontology

• The development of ontologies has been moving from the realm of Artificial-Intelligence laboratories to the desktops of domain experts.

• Ontologies have become common on the World-Wide Web.
  – Taxonomies and categories on web sites, e.g., Yahoo!, amazon.com

• W3C
  – RDF, OWL

• Standardized ontologies
  – UMLS(Unified Medical Language System)
  – UNSPSC(United Nations Standard Products and Services Code)
Ontology Basics

• Reasons why developing an ontology:
  – To share common understanding of the structure of information among people or software agents
  – To enable reuse of domain knowledge
  – To make domain assumptions explicit
  – To separate domain knowledge from the operational knowledge
  – To analyze domain knowledge
Ontologies

• Here we will be restricting our sense of ontologies to those we see emerging on the web.
• One widely cited definition of an ontology is Gruber’s [Gruber 1993] “A specification of a conceptualization”. 
What Is in an Ontology?

• An **ontology** is a formal explicit description of
  – concepts in a domain of discourse (**classes** (sometimes called **concepts**)),
  – properties of each concept describing various features and attributes of the concept (**slots** (sometimes called **roles** or **properties**)), and
  – restrictions on slots (**facets** (sometimes called **role restrictions**)).

• An ontology together with a set of individual instances of classes constitutes a **knowledge base**.
The Web’s Growing Needs

- The next generation of the web aims at pages for machine or programs consumption.
- The markup languages aimed at marking up content and services instead of just presentation information
  - XML, RDF, RDFS, DAML, etc. are becoming more accepted as users and application developers see the need for more understanding of what is available from web pages.
Ontology Spectrum

What is an Ontology?

Catalog/ID
Thesauri
"narrower term" relation
Terms/glossary
Informal
is-a
Formal
is-a
Frames
(proPERTIES)
General
Logical
constraints
Value
Restrs.

Disjointness,
Inverse, part-of…

Simple Ontologies and Their Uses

1. They provide a controlled vocabulary.
2. A simple taxonomy may be used for site organization and navigation support.
3. Taxonomies may be used to support expectation setting.
4. Taxonomies may be used as “umbrella” structures from which to extend content.
5. Taxonomies may provide browsing support.
6. Taxonomies may be used to provide search support.
7. Taxonomies may be used to sense disambiguation support.
Structured Ontologies and Their Uses

1. They can be used for simple kinds of consistency checking.
2. Ontologies may be used to provide completion.
3. Ontologies may be able to provide interoperability support.
4. Ontologies may be used to support validation and verification testing of data (and schemas).
5. Ontologies containing markup information may encode entire test suites.
6. Ontologies can provide the foundation for configuration support.
7. Ontologies can support structured, comparative, and customized search.
8. Ontologies may be used to exploit generalization/specialization information.
Ontology Acquisition

- One methodology for obtaining ontologies is to begin with an industry standard ontology and then modify or extend it.
- Another methodology is to semi-automatically generate a starting point for an ontology.
A Simple Knowledge-Engineering Methodology

Step 1: Determine the domain and scope of the ontology.
   – What, why, who, competency
Step 2: Consider reusing existing ontologies.
Step 3: Enumerate important terms in the ontology.
Step 4: Define the classes and the class hierarchy.
Step 5: Define the properties of classes—slots.
Step 6: Define the facets of the slots.
Step 7: Create instances.
Uses Cases of Web Ontologies

- Web portal
- Multimedia collections
- Corporate web site management
- Design documentation
- Agents and services
- Ubiquitous computing
Semantic Web Languages
What is XML?

- Extensible Markup Language
- A Syntax for Documents
- A Meta-Markup Language
- A Structural and Semantic Language, not a Formatting Language
- Not just for Web pages
XML Standards

- DTD
- Namespace
- Schema
- DOM
- CSS, XSL-T, XSL-FO
- XLink
- XPointer
XML Protocol: SOAP

1. SOAP client
2. HTTP
3. SOAP server
4. Remote object
5. HTTP
6. Response document

Request document
RDF M&S

• RDF (Resource Description Framework)
  – Beyond Machine readable to *Machine understandable*
• RDF consists of two parts
  – RDF Model (a set of triples)
  – RDF Syntax (different XML serialization syntaxes)
• RDF Schema for definition of Vocabularies (simple Ontologies) for RDF (and in RDF)
RDF Data Model

• Resources
  – A resource is a thing you talk about (can reference)
  – Resources have URI’s
  – RDF definitions are themselves Resources (linkage, see requirement 1)

• Properties
  – slots, define relationships to other resources or atomic values

• Statements
  – “Resource has Property with Value”
  – (Values can be resources or atomic XML data)

• Similar to Frame Systems
A Simple Example

• **Statement**
  - “Ora Lassila is the creator of the resource http://www.w3.org/Home/Lassila”

• **Structure**
  - Resource (subject) http://www.w3.org/Home/Lassila
  - Property (predicate) http://www.schema.org/#Creator
  - Value (object) “Ora Lassila”

• **Directed graph**
Another Example

• To add properties to Creator, point through an intermediate Resource.
Example: Bag

- The students in course 6.001 are Amy, Tim, John, Mary, and Sue
Example: Alternative

- The source code for X11 may be found at ftp.x.org, ftp.cs.purdue.edu, or ftp.eu.net
RDF Schema (RDFS)

• RDF just defines the data model
• Need for definition of vocabularies for the data model - an Ontology Language!
• The RDF Schema mechanism provides a basic type system for use in RDF models.
• The RDF schema specification language is less expressive, but much simpler to implement, than full predicate calculus languages such as CycL and KIF.
Most Important Modeling Primitives

• Core Classes
  – Root-Class rdfs:Resource
  – MetaClass rdfs:Class
  – Literals rdfs:Literal

• rdfs:subclassOf-property

• Inherited from RDF: properties (slots)
  • rdfs:domain & rdfs:range
  • rdfs:label, rdfs:comment, etc.

• Inherited from RDF: InstanceOf (rdf:type)
OWL provides three increasingly expressive sublanguages: OWL Lite, OWL DL, and OWL Full.
OWL
W3C Web Ontology Language

OWL Lite language constructs

**RDF Schema Features:**
- Class
- rdf:Property
- rdfs:subClassOf
- rdfs:subPropertyOf
- rdfs:domain
- rdfs:range
- Individual

**(In)Equality:**
- equivalentClass
- equivalentProperty
- sameAs
- differentFrom
- allDifferent

**Property Characteristics:**
- inverseOf
- TransitiveProperty
- SymmetricProperty
- FunctionalProperty
- InverseFunctionalProperty

**Property Type Restrictions:**
- allValuesFrom
- someValuesFrom

**Restricted Cardinality:**
- minCardinality (only 0 or 1)
- maxCardinality (only 0 or 1)
- cardinality (only 0 or 1)

**Header Information:**
- ontology
- imports
DAML-S

• Users and software agents should be able to discover, invoke, compose, and monitor Web resources offering particular services and having particular properties.

• As part of the DARPA Agent Markup Language program, we have begun to develop an ontology of services, called DAML-S.
Top Level of the Service Ontology

Resource provides Service
ServiceGrounding supports Service
Service presents ServiceProfile
Service describes by ServiceModel

(provided by)
(supports)
(presents)
(described by)
Process Modeling Ontology

ServiceModel

ProcessModel

Process

CompositeProcess

Sequence

Split

RepeatUntil

subPropertiesOf (parameter)
- input
- output
- participant

precondition

effect

parameter
Semantic Web Portal
Web Portals

• A web portal is a web site that provides information content on a common topic.
  – General portals, e.g., Yahoo, Excite, Netscape, Lycos, CNET, MSN, and AOL.com
  – Specialized portal e.g., gardeners.com, semanticweb.org

• Making valuable information to be found
  – directory service,
  – search facility
  – news, e-mail,
  – community forum
The Big Picture of SW

Ontology Articulation Toolkit

Ontology Construction Tool

Web-Page Annotation Tool

Ontologies

Annotated Web-Pages

Agents

Metadata Repository

Community Portal

Inference Engine

End User

(http://semanticweb.org/about.html#bigpicture)
Ontology-Based Web Portals

• Ontology represents
  – common knowledge and interests sharing within their community

• Tasks that ontology can be used to support a portal
  – Accessing a portal
    • Conceptual search and navigation
    • Inference capabilities
  – Providing information
    • Methods and tools accounting for the diversity of information sources
Technical architecture of An ontology-based portal
An Annotation Editor
We implemented a prototypical inference engine, which is able to translate RDF and answer F-logic and predicate logic queries about them. As an implementation platform, the Java language was chosen, because it is easy integratable with other components available on the web—for example, a Java servlet RDF Query Server could easily be created, and an HTTP-based API defined for query, update and result transactions.
We implemented a prototypical inference engine, which is able to translate RDF and answer F-logic and predicate logic queries about them. As an implementation platform the Java language was chosen, because it is easy integratable with other components available on the web – for example, a Java servlet RDF Query Server could easily be created, and an HTTP-based API defined for query, update and result transactions.
We implemented a prototypical inference engine, which is able to translate RDF and answer F-logic and predicate logic queries about them. As an implementation platform the Java language was chosen, because it is easy integratable with other components available on the web - for example, a Java servlet RDF Query Server could easily be created, and an HTTP-based API defined for query, update and result transactions.
Translation of RDF to Frame-based Representation
範例 (以下 RDF 是利用 Protée 2000 產生)

```xml
<?xml version='1.0' encoding='ISO-8859-1'?>

<rdf:RDF xmlns:rdf="&rdf;"
    xmlns:ka2new="&ka2new;"
    xmlns:rdfs="&rdfs;">

    <ka2new:Lecturer
        rdf:about="&ka2new;ka2new_00025"
        ka2new:address="ttu 602"
        ka2new:email="hhchen@cse.ttu.edu.tw"
        ka2new:firstname="hh"
        ka2new:lastname="chen"
        ka2new:name="hhchen"
        ka2new:phone="3295"
        rdfs:label="ka2new_00025"/>

</rdf:RDF>
```
範例（以下RDF係利用protage 2000 產生）

```xml
<?xml version='1.0' encoding='ISO-8859-1'?>
<rdf:RDF xmlns:rdf="&rdf;"
xmlns:ka2new="&ka2new;"
xmlns:rdfs="&rdfs;">
  <ka2new:Lecturer
    rdf:about="&ka2new;ka2new_00025"
    ka2new:address="tzu 602"
    ka2new:email="hchen@coe.ttu.edu.tw"
    ka2new:firstname="hh"
    ka2new:lastname="chen"
    ka2new:name="hchen"
    ka2new:phone="3295"
    rdfs:label="ka2new_00025">  
    輸出結果為 instance ka2new_00025 is a 'Lecturer'
    label is 'ka2new_00025'
    and phone is '3295'
    and name is 'hchen'
    and lastname is 'chen'
    and firstname is 'hh'
    and email is 'hchen@coe'.
  </ka2new:Lecturer>
</rdf:RDF>
```

KSL文件
Conceptual Search and Semantic Navigation
範例

frame food
default cost is "Number" and
default name is "String".

frame fruit is a food
default time is "Time".

frame fish is a food
default size is "String".

將Flex Ontology 載入系統
語法說明

簡單範例
run[Ka lavoro Person](C_Out_L_Out) 
give_AV[Ka lavoro Person][last-name-c][C_Out_L_Out] 
give_AV[Ka lavoro Person][In-Char][C_Out_L_Out] 
give_AV[Ka lavoro Person][last-name-Yu-Yang][C_Out_L_Out] 
give_AV[Ka lavoro Person][last-name-a-k-Char][C_Out_L_Out].

你所要查詢的內容
run[Ka lavoro Person](C_Out_L_Out)

Internal Query Language
根据输入的条件类别到RDF Triple Store找出资料
請選擇你要查詢的屬性
- number
- cp
- midname
- lastname
- isRelationWith
- firstname

請輸入查詢的值
- chen

搜尋結果：
你所找到的資料是屬於類別Lecturer的Instance

- address => ttu 502
- email => hhchen@cse.tu.edu.tw
- firstname => hh
- lastname => chen
- name => hhchen
- phone => 3295
- label => ka2new_00025

從Flex Data Store 找到資料
瀏覽學生的詳細資料
本網站主要是我們對讀者網的 研究心得，內容包含我
們的工作資料、相關文件的連結、出版資料 和常見問
題。在工作資料方面有我們所開設文件後的心得筆
記，使用 常見的標註工具後的感想，對於相關資料的
收集與分析和相關的系統架橋。相關文件方面，我們
彙整了一些標準的連結，提供我們方便查詢以及研
究。
Related Work
KA2
An Ontology-Based Community Web Portal
KA2

• Knowledge Annotation Initiative of the Knowledge Acquisition Community

• The basic scenario
  – WWW documents of the KS community were annotated according to the schema of an ontology.
  – The annotations enable intelligent access to these documents and infer implicit knowledge from explicitly stated facts and rules from the ontology.
The KA2 Ontology

Person-ontology

Class hierarchy
Person
  Employee
    Academic-Staff
      Lecturer
      Researcher
    Administrative-Staff
      Secretary
      Technical-Staff
  Student
    PhD-Student

Relations
Address, Affiliation, Cooperates-With,
Editor-Of, Email, First-Name, Has-Publication,
Head-Of-Group, Head-Of-Project, Last-Name, Member-Of-
Organization, Member-Of-Program
Committee, Member-Of-Research-Group,
Middle-Initial, Organizer-Of-Chair-Of,
Person-Name, Photo, Research-Interest,
Secretary-Of, Studies-At, Supervises,
Supervisor, Works-At-Project

Publication-ontology

Class hierarchy
On-Line-Publication
Publication
  Article
    Article-In-Book
    Conference-Paper
    Journal-Article
    Technical-Report
    Workshop-Paper
  Book
    Journal
      IEEE-Expert
      IJHCS
      Special-Issue

Relations
Abstract, Book-Editor, Conference-Proceedings-Title,
Contains-Article-In-Book, Contains-Article-In-Journal,
Describes-Project, First-Page, Has-Author,
Has-Publisher, In-Book, In-Journal,
In-Organization, In-Workshop, Journal-Editor,
Journal-Number, Journal-Publisher, Journal-Year,
Last-Page, On-Line-Version, ...
Accessing the Community Web Portal

• Query capability
  – In F-Logic mechanism
• Navigating capability
  – As the easy-to-use front-end of the query mechanism
KA² Portal - Microsoft Internet Explorer

Persons
Projects
Organizations
Publications
Research Topics
Events
About the System

knowledge acquisition community
KB Navigate
Search - Personalization - Feedback

Documents

find documents

Results:

DOCUMENTS
An Explication of Function
BPROW3
Intelligent Control at the Knowledge Level
The Ontology of Tasks and Methods
Representing Function as Effect: Assigning Functions to Objects in Context and out
Representing Functional Requirements and User-System Interactions
Separability Hypothesis,
Providing Information

- Integrating various syntactic and semantic formats based on the common ontology
- Three different modes of information provision are supported
  - Metadata-based information
  - Wrapper-based information
  - Fact-based information
Development of Web Portals

1. Requirement Elicitation
2. Web Site Design
   - Ontology Engineering
   - Terminology
   - Rule Development
   - Query Formulation
3. Creating Dynamic Webpages
4. Evaluation
5. Providing facts (user)
6. Maintain (Editor)
Topic Map
What Are Topic Maps?

• The GPS of the information universe

• A Topic Map is a network of
  - topics
  - associations (between topics)
  - occurrences (of topics)
The TAO of Topic Maps

- The basic building blocks are
  - Topics: e.g. “Puccini”, “Lucca”, “Tosca”
  - Associations: e.g. “Puccini was born in Lucca”
  - Occurrences: e.g. “http://www.opera.net/puccini/bio.html is a biography of Puccini”

- Each of these constructs can be typed
  - Topic types: “composer”, “city”, “opera”
  - Association types: “born in”, “composed by”
  - Occurrence types: “biography”, “street map”, “synopsis”

- All such types are also topics (within the same topic map)
  - “Puccini” is a topic of type “composer” … and “composer” is also a topic
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/" xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic id="chingyeh">
    <instanceOf>
      <topicRef xlink:href="#teacher"/>
    </instanceOf>
    <occurrence>
      <instanceOf>
        <topicRef xlink:href="#intro"/>
      </instanceOf>
      <resourceData>葉慶隆是大同大學資工系的老師</resourceData>
    </occurrence>
  </topic>
  + <occurrence>
  + <topic id="teacher"/>
  + <topic id="cjc"/>
  + <topic id="student"/>
  + <topic id="website"/>
  + <topic id="intro"/>
  - <association id="advisor-gradstud-association">
    <instanceOf>
      <member>
        <member>
          <roleSpec>
            <topicRef xlink:href="#gradstud"/>
          </roleSpec>
          <topicRef xlink:href="#cjc"/>
        </member>
      </member>
    </association>
  + <topic id="advrel"/>
  + <topic id="advisor"/>
  + <topic id="gradstud"/>
</topicMap>
葉基隆是大同大學資工系的老師

http://www.cse.ttu.edu.tw/chingye/index.html

陳佳揚 (受指導研究生)

簡介

網站

http://www.cse.ttu.edu.tw/chingye/index.html
Future Work

• Enhancing RDF stores
  – RDF to OODB (ZODB)
  – RDF to RDB
  – XTM (Topic Map)
• Intelligent Q&A
  – Plan-based content organization
• Ontology integration
  – Using ebXML Core Components
  – Using OASIS PSI
• Knowledge management study
  – strategy,
  – methodology
Summary

• Semantic Web portals
  – Machine-understandable information
  – RDF store
  – Accessing information
    • Navigation and query
  – Providing information
    • Annotation, wrapper, fact editing
  – Enabling automatic processing by software agents

• Much effort to be paid in content provision