Semantic Web and Ontology Based for Information Interoperability

Dr. I Wayan Simri Wicaksana, S.Si, M.Eng
iwayan@staff.gunadarma.ac.id
Outline:
• Interoperability
• Peer to Peer
• Semantic Web
• Ontology
• Example Research Topic
Outline:

- Interoperability
- Peer to Peer
- Semantic Web
- Ontology
- Example Research Topic
Motivation of Web Interoperability

Internet and Web:

- Web is massive
- Web is distributed
- Web is dynamic
- Web is open world
Dimension of Interoperability

• **Distribution**
  • multi database in a company
  • multi database on enterprise-wide, impact of Internet
  • multidatabase, global information infrastructure

• **Autonomy**
  • design autonomy
  • communication autonomy
  • execution autonomy
  • association autonomy

• **Heterogeneity**
  • Information H
    • Semantic H – Semantic Interoperability
    • Structural/Schematic H – Structural I
    • Sytatic H – Sytatic I
  • System H
    • Information System H (Media, DBMS, etc)
    • Platform H (OS, HW)
## Interoperability Era

<table>
<thead>
<tr>
<th>Level of interoperation concern</th>
<th>System, data</th>
<th>System, data, information</th>
<th>System, data, information, knowledge, process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Interoperability Architecture</td>
<td>Multidatabases, federated databases</td>
<td>Federated information system, mediator</td>
<td>Mediator, information brokering</td>
</tr>
<tr>
<td>Triggering</td>
<td>LAN</td>
<td>Internet</td>
<td>Global Information Infrastructure</td>
</tr>
<tr>
<td>Software and information system architecture</td>
<td>Term Acc, , mainframe, minicomputer, CS (2 tier)</td>
<td>CS (3 tier)</td>
<td>Network, distributed, mobile</td>
</tr>
<tr>
<td>Type of data</td>
<td>Structured database and files</td>
<td>Structured and semi-structured</td>
<td>All forms digital format</td>
</tr>
<tr>
<td>Research Prototype</td>
<td>ADDS, DDTS, Interbase, Multibase, etc</td>
<td>GARLIC, HERMES, InfoHarness, TSIMMIS, etc</td>
<td>Knowledge Sharing Effort, Intelligent Integration of Information</td>
</tr>
<tr>
<td>Products</td>
<td>DataJoiner(IBM), OmniConnect (Sybase)</td>
<td>Harness(Bellcore), TIE(Tesserae)</td>
<td></td>
</tr>
</tbody>
</table>
Federated Architecture

US1
Federated Schema 1
ES 1
Filtering Processor
Component Schema 1
Transforming Processor
Local Schema 1
Component db1

US2
Federated Schema 1
ES 2
Filtering Processor
Component Schema 2
Transforming Processor
Local Schema 2
Component db2

User / external schema
Federated schema
Export schema
Component Schema /
Common data model
Local schema/
data model

Constructing processor
Data Warehouse

Diagram:
- USERS
- Query
- Meta Data
- Target Database
- Integrate
- Source Database
- Source Database
- Source Database
Mediation Architecture

Application

Query

Global View

Global Data Dictionary

Wrapper

Data Source

Wrapper

Text File
Broker and Mediator

a). Single Ontology Approach

b). Multiple Ontology Approach

c). Hybrid Ontology Approach

Shared Vocabulary
Outline:

- Interoperability
- Peer to Peer
- Semantic Web
- Ontology
- Example Research Topic
History of P2P

- Sun with JXTA project
- Microsoft with .Net project
- Nepster (2001) sharing music file MP3, trigger P2P become popular
# New and Not New in P2P

<table>
<thead>
<tr>
<th>Perspective</th>
<th>What is New</th>
<th>What is Not New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical/Evolutionary (Computing)</td>
<td>Computing on the edge of the Internet</td>
<td>Distributed scheduling</td>
</tr>
<tr>
<td>Cultural/Sociological (Content sharing/Services)</td>
<td>Direct sharing (privacy, anonymity)</td>
<td>Decentralization</td>
</tr>
<tr>
<td>Communication/Collaboration</td>
<td>Dealing with disconnection</td>
<td>Ad-hoc NW, disconnected operation</td>
</tr>
<tr>
<td>Architectural</td>
<td>Cost of ownership</td>
<td>P2P concept and application</td>
</tr>
<tr>
<td>Algorithms/Programming model</td>
<td>Particular algorithms</td>
<td>Distributed state algorithm in general</td>
</tr>
</tbody>
</table>
Deffinition of P2P

• The Intel P2P working group gives definition: “the sharing of computer resources and services by direct exchange between systems”

• Ross Lee Graham state P2P definition trough three key requirements: a) they have an operational computer of server quality; b) they have an addressing system independent of DNS; and c) the are able to cope with variable connectivity

• Clay Shirky of O'Reilly and Associate say “P2P is a class of applications that takes advantage of resources – storage, cycles, content, human presence – available at the edges of the Internet. Because accessing these decentralized resources means operating in an environment of unstable connectivity and unpredictable IP addresses, P2P nodes must operate outside the DNS system and have significant or total autonomy from central servers”.

• Kindberg’s definition is “P2P systems as those with independent lifetimes”.

Characteristics of P2P

• Main Characteristics:
  - Self-Organizing
  - Symmetric Communication
  - Decentralized

• Derivative Characteristics:
  - Autonomy
  - Anonymity, vs intellectual property right
  - Cost of Ownership
  - Scalability
  - Connectivity
  - Addressing
  - Joint and leave
# Pros & Contras P2P

<table>
<thead>
<tr>
<th></th>
<th><strong>Pros</strong></th>
<th><strong>Contras</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collaboration</strong></td>
<td>No agreements are needed before deploying server</td>
<td>Each collaboration is a new overhead</td>
</tr>
<tr>
<td><strong>Standards</strong></td>
<td>Plug-in capability is flexible and can track new standards</td>
<td>Need to support all variants</td>
</tr>
<tr>
<td><strong>Costs</strong></td>
<td>Running costs are low</td>
<td>Bu in costs are high</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>Limited only by extranet bandwidth and server CPU</td>
<td>May need high capacity server</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>You have total control</td>
<td>You are responsible for server access control, etc</td>
</tr>
</tbody>
</table>
## Illustration Application of P2P

<table>
<thead>
<tr>
<th>Type Of Activity</th>
<th>SCOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>consumer</td>
</tr>
<tr>
<td>@work</td>
<td>Collaboration, communication</td>
</tr>
<tr>
<td></td>
<td>enterprise</td>
</tr>
<tr>
<td></td>
<td>Distributed computing, storage, collaboration, communication</td>
</tr>
<tr>
<td></td>
<td>public</td>
</tr>
<tr>
<td></td>
<td>Communication, digital rights management</td>
</tr>
<tr>
<td>@play</td>
<td>games</td>
</tr>
<tr>
<td></td>
<td>Sponsored event</td>
</tr>
<tr>
<td></td>
<td>Digital media, digital experience</td>
</tr>
<tr>
<td>@rest</td>
<td>Music sharing</td>
</tr>
<tr>
<td></td>
<td>Content consumption</td>
</tr>
<tr>
<td></td>
<td>Instant messaging</td>
</tr>
</tbody>
</table>
Consideration Implementation of P2P

- Budget
- Ressource redevance to participants
- Trust
- Rate of system change
- Critically
Architecture of a Peer

Outgoing Events sent Via HTTP

Module Container
Hold dynamically loaded Modules and provides access to Services and the API

Express Module

GUI
Chat
Web agent (WAP)
Web agent (HTML)

Access Control

Buddy Manager (location, presence, grouping)

Event Service

Request Manager

Servelet Engine

Web Server

Core Component

Extension Module

Web Service Foundation
P2P Model

- Degree of Centralization
  - Pure P2P / Decentralized
  - Partially
  - Hybrid

- Network Structure
  - Structured
  - Loosely Structured
  - Unstructured

- Organize Network
  - Random Graph
  - Power Law Network
  - Publish / Subscribe Network

- Searching
  - Routed Response
  - Direct Response (with & with/o invitation)
  - Routed Response with metadata
  - Direct Metadata Response (with & with/o invitation)
  - Combination (Switch, Shift, Pipeline)
Degree of Centralization

Request from A

Pure P2P

SP : super-peer / super-node
P : peer / node

Sample Query

Partially P2P

Hybrid P2P
<table>
<thead>
<tr>
<th>Degree of Centralization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
</tr>
<tr>
<td><strong>Pure P2P</strong></td>
</tr>
<tr>
<td><strong>Partially</strong></td>
</tr>
<tr>
<td><strong>Hybrid</strong></td>
</tr>
</tbody>
</table>
Network Structure

<table>
<thead>
<tr>
<th>Pure Decentralized</th>
<th>Chord, CAN, Tapestry, Pastry</th>
<th>Freenet</th>
<th>Gnutella</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partially Decentralized</td>
<td></td>
<td></td>
<td>KaZaA</td>
</tr>
<tr>
<td>Hybrid Decentralized</td>
<td></td>
<td></td>
<td>Napster</td>
</tr>
</tbody>
</table>
Organize Network

- Random Graph, i.e. Gnutella with flooding
- Power-Law Network, like super-node model
- Publish / Subscribe Networks,
Improvement Matching Query

- RR switch to DR
- DR switch to RR
- DR shift to DR
- Pipe Line
Other Issues

- Security for P2P
- Interoperability between P2P models
- Application layer in P2P for information integration/interoperability/sharing
- Real implementation
Summary

- P2P as one of distributed system solution
- Characters of P2P are points of consideration in implementation of P2P
- P2P environment as media for information integration / interoperability / sharing
Outline:

- Interoperability
- Peer to Peer
- Semantic Web
- Ontology
- Example Research Topic
Definition

Jonathan West, www.ncl.ak.uk/sml/staff/west/sml12_week10b.htm, 2000 & Tim Berners Lee, W3 [22]

- Semantic: study of meaning
  - Sense and reference
  - Synonymy, hyponymy, incompatibility, antonymy complemenariry & proper, converseness, homonymy (homography & homophony)
  - Characteristic object, cognate object, characteristics action/state/process, characteristic quality, symptom and state.

- Semantic Web: information linked up to be easily processable by machines on global scale. Is web of data

- Semantic Web: mesh of information link up to be easily process by machine on global scale [23]

- Inference: derive new data from available data

- Semantic Web – ER model, RDF – model of ER [26]

- Semantic Web – relational database, record – RDF model, colom – RDF prototype type, record – value [26]
What & Where Semantics
A Sheth, 1999 [8]

What is semantics?
Vocabulary
Content
Structure

Where is semantics?
Ontology (domain specific)
Metadata
Data

relationships
content-descriptive
content-based (indices)
content-independent

abstract structure
domain specific metadata
Type of Meta Data

• **Metadata Media**
  – Type specific metadata, ie: texture, frequency
  – Processing-specific metadata
  – Content-specific metadata

• **Metadata classification**
  – Content-independent, ie: location, type of sensor
  – Content-dependet, ie: no or row,color
    • Direct content-based, ie: inverted tree, document vector
    • Content descriptive, ie: fragrance of floweer
      – Domain-independent, ie: C++, HTML
      – Domain-specific, ie: relief, population in GIS

• **Metadata** is machine understandable information about web resources or other (Tim Berners-Lee [24])
Definition Data Semantic
Amit Seth, Data Semantic What, Where and How [8]

• Meaning and use of data (Wood)
• A Connection from database tor the real world outside of the database (Rosenthal)
• Meaning of data and reflection of the real world (Mark)
• Set of mapping represent language to concept in the real world (Meersman)
• Anything about data has a conceivable practical consequence/in application (David B)
Data Semantic (1)
Amit Seth, Data Semantic What, Where and How [8]

• Where to get
  – People interact, communicate, interpretation and use of data
  – Semantic (software semantic) in data structure, know representation program, communication protocol and encoding
  – Semantic (procedural semantic), program procedure, social protocol, law
Data Semantic (2)
Amit Seth, Data Semantic What, Where and How [8]

• Representing
  – Semantic proximity – degree similarity/relationship (Kashyap&Sheth)
  – Reuse, Evolution, Merging and Modularity (Rosehl)

• On Use
  – Multidatabase manipulation
  – Interoperability
Contribute to Create Semantic Web

Jonathan West, www.ncl.ak.uk/sml/staff/west/sml12_week10b.htm, 2000

- Published some globally useful data in RDF
- Write an inference engine in the language of your choice
- Spread the word: do some education and outreach
- Help in the development of RDF Schema and/or DAML
- Contribute in representing in RDF
- Apply your own development background to the Semantic Web, give new angle
- Instead of using some proprietary system for your next application, consider making it a Semantic Web Project

Instead
Orthogonal Layers in The Semantic Web
Alexander Maedche, Steffen Staab, 2003

Semantic Web Representation Layers

Semantic Web Application

Semantic Web Management

Trust
Proof Interchange
Rule Language
OWL
RDF(S)
UML

Portal
Store
Web Services
CMS
Modify
Evolve
Versioning
Monitoring
Interfacing

Versioning
Modify
Access
Store
Portal
HTML - XML

- Confused in content presentation
- No explicit structure
- How schema when querying the Web (HTML)
- Not appropriate for data exchange
- XML shall be straightforwardly usable over the Internet.
- XML shall support a wide variety of applications.
- XML shall be compatible with SGML.
- It shall be easy to write programs which process XML documents.
- XML documents should be human-legible and reasonably clear.
- The XML design should be prepared quickly.
- The design of XML shall be formal and concise.
- XML documents shall be easy to create.
XML Stack

Standardized Applications
- XML 1.0
- Locators (URI)
- Unicode
- DTDs

Specific Applications
- Layout
  - XSL
  - CSS
- Hyperlinks
  - XLink
  - XPointer
- Metadata
  - RDF
  - RDFS
- API
  - DOM
  - SAX
- Schemas
  - XSD
  - Namespaces
- Queries
  - XPath
  - XQuery

Standardized Applications
- XHTML, SVG, SML, MathML

Specific Applications
- Tutorial – WISE 2002
XML Basic

- Defines a class of XML documents
- Defines datatypes, elements and attributes
- Documents described by an XML schema can be called an instance
- The schema language, considerably extends the capabilities of XML 1.0 document type definitions (DTD)

Simple tutorial: http://www.w3schools.com/
Web Evolution
Outline:

- Interoperability
- Peer to Peer
- Semantic Web
- Ontology
- Example Research Topic
Definitions Ontology
Ansuncion Gomez Perez, 1999 [10]

• An ontology defines the basic terms and relation comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary (Neches et al)

• An ontology is an explicit specification of a conceptualization (Gruber)

• An ontology is a formal (machine readable), explicit specification of a share (concensual knowledge) conceptualitation (Studer, et al)

• There are other 8 definitions
Track and State of The Art

- Research areas: database-collection metadata, application areas: web services, eLearning, PtP network, knowledge management and knwoledge protal (Sturder [16])
- Better interoperability and search; need knowledge representation, vocabulary, thesaurus; easy to use open source tools (James Handler [17])
- Useful standard for exchanging information; OWL, reasoning engines and services, clusters (Patrick Hayes [18])
- Moving from XML – XML/s – RDF – RDF/s – Web Onto (Tim Berners-Lee [42])
Ontology Spectrum

- Catalog/ID
- Thesauri
  - "narrower Term" relation
- Terms/Glossary
- Informal Is-a
- Formal Is-a
- Formal instance
- Frames (properties)
- Value Restrs
- General Logical Constraints
- Disjointness, Inverse, Part-of
IEEE Standard 1074-1995

1. Software life cycle model process
2. Project management process
3. Development – oriented processes
   1. Pre-development processes
   2. Development process
   3. Post-development process
4. Integral process
   - Planning
   - Control
   - Quality Management
   - Environment study
   - Feasibility study
   - Requirements
   - Design
   - Implementation
   - Installation
   - Operation
   - Support
   - Maintenance
   - Retirement
   - Evaluation
   - Documentation
   - Configuration Management
   - Training
Skeletal Methodology
Ushold and King, 1995

1. Identify Purpose and Scope

2. Building the ontology
   1. Ontology capture
   2. Ontology coding
   3. Integrating existing ontologies

3. Evaluation

4. Documentation

5. Guidelines for each phase

- Identify key concepts and relationships
- Produce unambiguous text definition
- Identify term to refer to such concept and relations

- Commit to a meta-ontology
- Choose a representation language
- Write the code

How and whether to reuse ontologies that already exist
Gruninger & Fox Methodology
Gomez Perez, 1998

Motivating Scenarios ➔ Informal Competency Question ➔ Formal Terminology ➔ Formal QC ➔ Formal Axioms ➔ Completeness Theorems

Identify intutively possible Application and solutions.

As an entailment of consistency problems with respect to the axioms in the ontology.

Conditions under which the solutions to the questions are complete.

Identify Queries:
- Answers: Axioms
- Questions: Terminology

Objects ➔ Constants
KIF ➔ Variable
Attributes ➔ Functions
Relations ➔ Predicates

Defined as a first-order sentence Using the predicates of the ontology.
• Specification of the application

• Preliminary design based on relevant top-level ontological categories

• Ontology refinement and structuring

Kactus Methodologies
Bernaras et alia, 1996
METHONTOLOGY
Gomez Perez, 1998
SENSUS Methodology
B Swartout, R Patil, 1997

1. Identify “seed” terms
2. Link seed terms to SENSUS by hand
3. Include nodes on the path to root
4. Add entire subtrees using the heuristic:
   If many nodes in a subtree are relevant, the other nodes in the subtree are relevant
On-To-Knowledge (OTK) Methodology

York Sure, Rudi Studer, Toward The Semantic Web, 2003

Feasibility Study
- Go/ no Go
- Select tool
- Focus domain
- Identify people

Ontology Kickoff
- Semi formal Description
- Task Anal / TM-1
- Know Anal / TM-2
- Agent Mod / AM-1

Refinement
- Target Ontology
- Requirement specification
- Anal knowledge resources
- Create semi formal description of ontology

Evaluation
- Knowledge elicitation with domain experts
- Formalize (target ontology)
- Check requirements
- Test in target applications
- Analyze usage patterns

Maintenance & Evaluation
- Manage organizational maintenance process (PIC)
- Evolution of ontologies

Target Ontology
- O-based Application
# Methodology Compliance With IEEE Standard

Lopez, 1999

<table>
<thead>
<tr>
<th>Management Processes</th>
<th>Development-oriented process</th>
<th>Integral Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-Develop</td>
<td>Development</td>
</tr>
<tr>
<td></td>
<td>Requirement</td>
<td>Design</td>
</tr>
<tr>
<td>Skeletal, Uschold &amp; King</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gruunger &amp; Fox</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Kactus, Bernaras</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Methontol</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Sensus</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>OTK, York</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Available

Partly Available

No Available
# Maturity of The Methodologies

Lopez, 1999

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Recommended Life Cycle</th>
<th>Compliance with IEEE Std</th>
<th>Recommended Techniques</th>
<th>Ontologies and application</th>
<th>Detail of the methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skeletal, Uschold &amp; King</td>
<td>X</td>
<td>O</td>
<td>not known</td>
<td>1 domain</td>
<td>very little</td>
</tr>
<tr>
<td>Gruunger &amp; Fox</td>
<td></td>
<td>O</td>
<td>not known</td>
<td>1 domain</td>
<td>little</td>
</tr>
<tr>
<td>Kactus, Bernaras</td>
<td>X</td>
<td>O</td>
<td>not known</td>
<td>1 domain</td>
<td>very little</td>
</tr>
<tr>
<td>Methontol</td>
<td>✓</td>
<td>O</td>
<td>O</td>
<td>several domain</td>
<td>a lot</td>
</tr>
<tr>
<td>Sensus</td>
<td>X</td>
<td>O</td>
<td>not known</td>
<td>several domain</td>
<td>medium</td>
</tr>
<tr>
<td>OTK, York</td>
<td></td>
<td>O</td>
<td>O</td>
<td>several domain</td>
<td>medium</td>
</tr>
</tbody>
</table>

Available

Partly available

No available
Outline:

● Interoperability
● Peer to Peer
● Semantic Web
● Ontology

→ Example Research Topic
Expressivity and Semantic Power

OWL

RDF

XML

Top Quadrant Survey

Enterprise Support

 Sources: Irene Polikoff, and Dean Allemang, Top Quadrant Technology Briefing: Semantic Integration, Strategis and Tools, Top Quadrant, 2003
Research Topic

• Interoperability
  – Doctoral: Multiple Ontology, Ontology Maintenance, Query Optimization, Query Rewriting, Matching
  – Master: Implementation and approach to be concept or algorithm level
  – Bachelor: Bring to coding and real implementation.

Example:
  • Wrapper for currency in business
  • Wrapper for adding cost in business
  • Wrapper for unstructured information
  • Collection & comparing airline ticket price from many sources
  • Create Ontology
Terimakasih

Thank you – Merci - Danke