A Semantic Web Ontology for Research Community

Oct. 25, 2006

In-Su Kang, Hanmin Jung, Seungwoo Lee, Pyung Kim, Heekwan Koo, Mikyoung Lee, Namang Kuh, Won-Kyung Sung

Korea Institute of Science and Technology Information
Contents

■ Introduction
  ■ Semantic web & ontology

■ Previous works
  ■ Ontology development methodology
  ■ Ontology for research community

■ OFK ontology (for research community)
  ■ Ontology schema
  ■ Ontology instance
  ■ Instance management by URI-server
  ■ Instance representation

■ Construction of OFK ontology

■ Conclusion
Introduction - background

- Current web
  - Human-oriented, syntactic
    - Only understood by persons
    - Disallow automatic processing

- Semantic web [Berners-Lee et al., 2001]
  - Machine-oriented, semantic
    - Semantic tags assigned to information units
  - Prerequisite
    - Ontology
      - concept hierarchy for semantic tagging
    - Inference engine
      - ontology validation & implicit knowledge extraction
Introduction - ontology (1/2)

- **Shared, formal, explicit conceptualization** [Gruber’93, Borst’97]
  - for concepts and relationships b/w concepts

- **Two dimensions**
  - Schema level: \([Paper] \rightarrow [hasAuthor] \rightarrow [Person]\)
  - Instance level: ‘A Relational Model of …’ \(\rightarrow [hasAuthor] \rightarrow ‘Codd E.F.’\)
Introduction - ontology (2/2)

- **Relationship b/w concepts**
  - Object property
    - Relationship b/w concepts
    - Similar to relationship b/w entities in RDB
  - Datatype property
    - Relationship b/w concept and literal
    - Similar to relationship b/w an entity and an attribute in RDB

```
Paper
   hasAuthor
   isWith
   Thing
   subClassOf
   Organization
      isOwnedBy
   Person
      hasAuthor
   hasKorTitle
   hasEngTitle
   hasAcceptanceDate
   Datatype property
   subClassOf
   hasTotalPages
   Integer
   String
   String
   Date
```

Copyright © 2004-2006, KISTI
Previous works - ontology development methodology

- **Focused on schema modeling**

- **Need for instance modeling**
  - Identification system
    - (e.g.) SSN for persons, DOI for contents
  - Identity resolution
    - Synonymy problem
      - (e.g.) John R. Smith vs. John Richard Smith
    - Homonymy problem
      - (e.g.) John R. Smith in Harvard Univ. vs. John R. Smith at MIT
Previous works – ontology for research community

- Research area ontologies
  - KA2 ontology [Benjamins, 1999]
    - Modeling knowledge acquisition community (researchers, topics, etc.)
    - The first ontology for research area
  - SWRC (Semantic Web Resource Community) ontology [Sure et al., 2005]
    - KA2-based
    - Applied to creating social networks of researchers
  - AKT reference ontology [http://www.aktors.org/publications/ontology/]
    - Inferring top-level researchers / organizations / researcher’s cluster

- Summary
  - Includes Person, Organization, Project, Publication in common
  - Does not address identity resolution for instance modeling (except AKT case)
    - (e.g.) ambiguity of same-name authors
OFK ontology - overview

■ Motivation
  ■ Support researchers over full life-cycle of research activity
  ■ OFK: OntoFrame-K®(ver. 2006)
    □ Ontology framework for Knowledge/Korean/KISTI

■ Design principles
  ■ Schema-level
    □ Language-independent
    □ Scenario-oriented
      □ Do not include unnecessary elements from the viewpoint of application
    □ Ockham’s razor
      □ Do not represent properties derivable from rules

  ■ Instance-level
    □ Separate management of instances
      □ Through URI-server
        □ Instance storing
        □ Instance identity management
        □ Integrity check
OFK ontology – schema (1/2)

Core classes

- Person, Organization, Project, Outcomes (Paper, Patent, Report), Publication (Journal, Proceedings), Topic, CreatorsInformation, Location

Object property

- Outcomes – hasCreatorsInformation – CreatorsInformation
- Outcomes – hasOriginatedProject – Project
- Outcomes – hasPublication – Publication
- Outcomes – hasTopic – Topic
- CreatorsInformation – hasCreator – Person
- CreatorsInformation – hasOrganizationOfCreator – Organization
- Project – hasOrganizationOfFundingProject – Organization
- Project – hasOrganizationOfPerformingProject – Organization
- Organization – hasLocation – Location
- Person – hasOrganizationOfPerson - Organization
OFK ontology – schema (2/2)

- **CreatorsInformation**
  - Info. of a creator at the time when his/her outcome was written
    - Order of creator
    - Person corresponding to a creator
    - Organization of a creator
      - Different from organization of person
OFK ontology – instance

Identification system

Outcomes
- KOI (Knowledge Object Identifier)
  - Proceedings paper: ‘KISTI1.PCD.0001234’
  - Journal paper: ‘KISTI1.JNL.0000123’
  - Patent: ‘KISTI1.PTN.0000012’
  - Report: ‘KISTI1.RPT.0012345’

Person
- National Science & Tech. Personnel Identification system
  - 10-digit unique number
  - (e.g.) Clinton: ‘7010862430’

Organization
- Organization code compiled by Korea Research Foundation
  - 6-alphanumeric code
  - (e.g.) Seoul National Univ.: ‘114800’
  - (e.g.) Korea Institute of Science and Tech. Info.: ‘9R9048’
OFK ontology – instance management

Title: A Storage Structure for Nested Relations Using Signatures
Author: Hwan-Seung Yong, Sukho Lee
Publication: Journal of Systems Architecture
Volume(Issue)/Page: 43(5) / 245-250
Year: 1997

Sukho Lee → PER_4410022529
Jnl. of Sys. Arch. → PUB_SOJ000574

Register to URI-sever
(referential integrity check)
(duplicate check)

<table>
<thead>
<tr>
<th>URI Server Data</th>
<th>Type</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>PER_4410022529</td>
<td>Person</td>
<td>Name: Sukho Lee, Organization: ORG_114800</td>
</tr>
<tr>
<td>PER_0000012345</td>
<td>Person</td>
<td>Name: Hwan-Seung Yong, Organization: ORG_133600</td>
</tr>
</tbody>
</table>
| OBJ_KISTI.JNL.0000001 | Outcomes | Title: A Storage Structure for Nested Relations Using Signatures
| ORG_114800      | Organization | Name: Seoul National University |
| ORG_9R9048      | Organization | Name: Ewha Womans University |
| PUB:SOJ000574   | Publication | Name: Journal of Systems Architecture |
| TOP_030213      | Topic | Topic name: database |
Title: A Storage Structure for Nested Relations Using Signatures
Author: Hwan-Seung Yong, Sukho Lee
Publication: Journal of Systems Architecture
Volume(Issue)/Page: 43(5) / 245-250
Year: 1997

URI Server Data

<table>
<thead>
<tr>
<th>URI</th>
<th>Type</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>PER_4410022529</td>
<td>Person</td>
<td>Name: Sukho Lee, Organization: ORG_114800</td>
</tr>
<tr>
<td>PER_0000012345</td>
<td>Person</td>
<td>Name: Hwan-Seung Yong, Organization: ORG_133600</td>
</tr>
<tr>
<td>ORG_114800</td>
<td>Organization</td>
<td>Name: Seoul National University</td>
</tr>
<tr>
<td>ORG_9R9048</td>
<td>Organization</td>
<td>Name: Ewha Womans University</td>
</tr>
<tr>
<td>PUB_SOJ000574</td>
<td>Publication</td>
<td>Name: Journal of Systems Architecture</td>
</tr>
<tr>
<td>TOP_030213</td>
<td>Topic</td>
<td>Topic name: database</td>
</tr>
</tbody>
</table>
Construction of OFK ontology - schema

Statistics

- # of classes: 21
- # of properties: 64
  - # of datatype properties: 46
  - # of object properties: 18
- # of rules: 22
  - # of object properties derived by rules: 14

Ontology creation

- Ontology editing tool
  - Protégé 3.1.1
- Ontology description language
  - W3C OWL DL (http://www.w3c.org/)
Construction of OFK ontology - instance

Target bibliographic data
- Proceedings of major conferences/workshops/symposiums
  - Held at Korea during 2002 through 2006
  - # of papers: 12,016

# of RDF (Resource Description Framework) triples

<table>
<thead>
<tr>
<th>Type</th>
<th># of RDF Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>11,390</td>
</tr>
<tr>
<td>Outcomes (Paper)</td>
<td>12,016</td>
</tr>
<tr>
<td>Organization</td>
<td>12,586</td>
</tr>
<tr>
<td>Publication (Proceedings)</td>
<td>449</td>
</tr>
<tr>
<td>Topics (thesaurus-based)</td>
<td>31,719</td>
</tr>
<tr>
<td>Location</td>
<td>28,741</td>
</tr>
<tr>
<td>Others (Project, Department, etc.)</td>
<td>8,578</td>
</tr>
<tr>
<td>Instance relationship</td>
<td>1,553,575</td>
</tr>
<tr>
<td>Total</td>
<td>1,659,054</td>
</tr>
</tbody>
</table>
Conclusion

- **Sharing of ontology construction experience**
  - OFK ontology for research area

- **Need for ontology instance management**
  - Proposal of URI-server as a separate instance store
    - Instance storing
    - Instance identity management
    - Integrity check
      - Data integrity
      - Referential integrity
      - Duplicate detection
Thank you!

dbaisk@kisti.re.kr
swlee@kisti.re.kr
## Two Sides of Semantic Web

<table>
<thead>
<tr>
<th></th>
<th>Open domain</th>
<th>Closed domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assumption</strong></td>
<td>Open-world assumption</td>
<td>Closed-world assumption</td>
</tr>
<tr>
<td><strong>Type of data</strong></td>
<td>General Web contents</td>
<td>Legacy or security data</td>
</tr>
<tr>
<td><strong>Control of Instance integrity</strong></td>
<td>Don’t care</td>
<td>Highly needed</td>
</tr>
</tbody>
</table>