Open Ontology Repository

Planning Meeting for
Ontology repositories:
approaches, technologies, collaboration

Ken Baclawski

June 15, 2009

OOR Charter

Promote the global use and sharing of ontologies by:

- − establishing a hosted registry-repository;

- − enabling and facilitating open, federated, collaborative ontology repositories, and

- − establishing best practices for expressing interoperable ontology and taxonomy work in registry-repositories.

http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository

Scope of the OOR

- Repository: *"An ontology repository is a facility where ontologies and related information artifacts can be stored, retrieved and managed"*
  - The persistent store
  - The registry
  - Value-added services

- Ontology:
  - all types of artifacts on the ontology spectrum
    - from folksonomies, terminologies, controlled vocabularies, taxonomies, thesauri, ... to data-schema, data-models ... to OWL ontologies ... and, axiomatized logical theories
    - from shared understanding ... to ontological commitments ... to the future of standards

- Open:
  - open access; compliance with open standards; open technology (with open source); open knowledge (open content); open collaboration (with transparent community process)
  - open to integration with "non-open" repositories via an open interface

OOR Activities

- Develop rationale and motivation
- Collect user needs and requirements
- Identify research challenges
- Seek funding
- Initiate design and implementation
- Maintain consistency with the rationale and motivation

Isn't the Semantic Web notion of distributed islands of semantics sufficient as a de facto repository?

- If you put it out there, will they come?
- If you build it better and put it out there, will they prefer yours?
- History does not show this laissez faire "field of dreams" is good reality
- The "clickable" web has been very successful in employing this strategy for html documents
- However the use and content of the semantic web has different characteristics that make it far less tolerant of the change and frequent errors which are commonplace on the clickable web.

Distinguishing characteristics of the Semantic Web

- Machines rather than humans are the primary consumers of content. Errors that a human may be able to diagnose and fix (such as a change in location of a document) are likely fatal for machine processing

- The use of owl:imports creates a strong transitive dependency between ontology documents; changes in any imported document (imported directly or through nested import) can cause the resulting import closure to be inconsistent or to change its meaning or computational characteristics significantly.

- Ontologies convey a precise meaning with an unambiguous machine interpretation. This means that, when using this content, careful selection and precise reference is critical.

OOR Value Added (1)

- The OOR is reliably available
- The OOR is persistent and sustainable, so you can be confident when committing to its use
- The OOR has information about when, why, and how an ontology has changed, so you can be aware of changes that may effect its usability
- You can find ontologies easily
- Ontologies are registered, so you know who built them
- Metadata provides the ontology purpose, KR language, user group, content subject area, etc.

OOR Value Added (2)

- The OOR includes mappings, so you can connect ontologies to other ontologies
- The OOR content has quality and value, as gauged by recognized criteria
- The OOR supports services, so that ontologies can map and be mapped, find and be found, can review/certify and be reviewed/certified, can hook your own services into and can use the services others have hooked in
- Ontologies can reuse or extend other ontologies, including common middle and upper ontologies
- The OOR can be easily extended

Top Requirements

- A well-maintained persistent store (with high availability and performance) where ontological work can be stored, shared and accessed

- Properly registering and "governing" ontologies, with provenance and versioning support, made available (logically) in one place so that they can be browsed, discovered, queried, analysed, validated and reused

- Allow ontologies to be "open" and unencumbered by IPR constraints, in terms of access and reuse

- Providing services across disparate ontological artifacts to support cross-domain interoperability, mapping, application and making inferences.

- Registering semantic services to support peer OORs

- OOR Use Case development is ongoing.

9

Research Challenges (1)

- Inadequacies of the "clickable web" as a basis for the semantic web
  - Need long-term maintenance rather than ontologies maintained in author's web site.
  - Varying levels of coverage
  - Intellectual property concerns
- Best practices
  - Policies and procedures
  - Provenance to enable trust

Research Challenges (2)

- Ontology metadata
  - Ontology Summit 2007 Dimensions
  - Ontology Metadata Vocabulary
  - eXtended Metadata Registry
- Interface ontologies
  - Internal APIs for core modules and plug-ins
  - External APIs, especially web services
  - Federation APIs, among OORs

Funding

- NSF DataNet proposal outcome:
  - Unanimous agreement that the OOR is important
  - Mismatch with the specific requirements of the program
- Current efforts:
  - Incorporate OOR into other proposals
  - Volunteers
  - Meet with program officers

Design and Implementation

- OOR Sandbox based on BioPortal

- OOR meetings on Fridays

- Joint Ontolog-OOR conference calls on
  - July 16 (Mark Musen)
  - July 30 (Natasha Noy)

Links

- Much of this presentation was derived from Ontology Summit 2008 Presentation
- Home page: OOR