Bibliographic Ontologies: CiTO, the Citation Typing Ontology

Kate Lauber
S604 Metadata & Semantics
December 11, 2009
Bibliographic Ontologies

- MARC21 (MAchine Readable Cataloging) precedes today’s bibliographic ontologies
  - Fields, tags, and indicators encode elements of a resource to make a sharable record
- Metadata description schemas
  - MODS (Metadata Object Description Schema), MADS (Metadata Authority Description Schema) are expressed in XML
  - Dublin Core is expressed XML or RDF
Bibliographic Ontologies con’t

• FRBR (Functional Requirements for Bibliographic Records)
  • Conceptual model for representing resources, expressed in XML
• MarcOnt
  • Ontology that borrows from MARC, Dublin Core and BibTex (XML, RDF, OWL)
• BIBO (The Bibliographic Ontology)
  • Provides semantics for describing citations and references (RDF)
How are ontologies used to add semantics to existing library and other bibliographic data?

MARC—researchers explored its extensibility as an ontology, incorporating authority data through FRBR

SIMILE Project at MIT has tools to convert MARC to MODS, then MODS to RDF; BibTex to RDF
• National Library of Sweden developed RDF wrapper to expose MARC records to the Semantic Web
  • Dublin Core for bibliographic data
  • FOAF (Friend of a Friend) for authority data
  • SKOS (Simple Knowledge Organization System) for controlled vocabularies
  • FRBR to link between records
• Library of Congress now uses SKOS to represent authority records
RDA & the Semantic Web

- RDA (Resource Description and Access), the new cataloging standard, has worked with Dublin Core to create metadata standards that are interoperable with the Semantic Web.
- Parts of RDA have been developed as an RDF vocabulary.
Martha Yee (2009) questions the rush to expose bibliographic data to the Semantic Web

Concerns about valuing machine-readable data rather than human end-users

- RDF is expressed as a tree—not useful for library catalog users
- Use of XSLT to clearly display RDF data to library users
CiTO

- Bibliographic ontology that describes citations in scholarly papers
- Potential for expressing nuances helpful for evaluating a scholar’s work (for tenure, etc.)
- Enables representation of *how* a scholar cites another’s work—does she agree or disagree? Critique the work or use it for background information?
- CiTO will reach its full potential in a fully Open Access environment
CiTO

- Citation characterization
  - Object properties `disagreesWith`, `usesDataFrom`, etc.
- Citation frequency
  - Object properties `inTextCitationFrequency`, etc.
- Characterization of the cited works themselves (FRBR)
  - Classes `Work`, `Expression`, `Manifestation`
  - Subclasses `ResearchPaper`, `BookReview`, etc.

http://purl.org/net/cito/
CiTO

- There are other bibliographic ontologies that represent citation data
- CiTO’s creator says it has more granularity
  - SWAP (Scholarly Works Application Profile)
  - BIBO (The Bibliographic Ontology)
  - SWAN (Scientific Discourse Relationships Ontology)
- What does a paper that uses CiTO look like?
  - http://dx.doi.org/10.1371/journal.pntd.0000228.x001
CiTO in Protégé

- Using CiTO, I modeled five scholarly papers in Protégé.
- Most citations were very simple relationships and used object properties `obtainsBackgroundFrom` or `obtainsSupportFrom`.
- Since I selected papers about the Semantic Web, there were many overlaps in authorship that could be represented using `sharesAuthorsWith`. 
Conclusion

- The scholarly article that has been semantically enhanced with CiTO, raw data, maps, taxonomies, etc., becomes a different kind of scholarship
- It is interactive and stimulates conversation
  - Access to data, nuances of relationship between sources
- Open Access materials help
- Feasible for scholars to use CiTO for all of their own work—decentralized