RDF and Collections

CNI Presentation

Terry Noreault
OCLC Office of Research

Collections

- Collections will be increasingly important for effective access to electronic resources
- The “Web” encourages the decomposition of resources we previously would have thought were atomic (e.g. documents) making documents less useful as a unit of retrieval
- Collections are a mechanism for aggregating these objects

Defining Collections

- Identifying the context of resources, and how it relates to other resources, is important for access
  - e.g. Subject Classification
  - Yahoo is most popular search service
- Context can be defined in multiple ways
- Resources can have multiple contexts
  - e.g. can be in more than one collection at one time

Collection Tools

- Recognizing this, new tools are required to support the
  - Description...
  - Navigation...
  - Discovery...
  - Retrieval...
  - Management...
  - And analysis of these collections

Bookmarks as Collections

- Fine example of an “ad-hoc” collection
- Would like to say some additional “things” about the intellectual content of this collection
- More time spent on maintaining bookmarks than on describing web resources
- Next Generation: More structure, sharable, template for discovery of other things “like it”, etc.

Inadvertent Collections

- Collections that are discovered
  - Search results (by subject)
    - e.g. Yahoo
  - Search results (by relation)
    - e.g. Alexa
  - Search results (by explicit query)
    - e.g. AltaVista
Relations as Collections

RDF for Describing Resources:
- Web enables distributed publishing
- Uniform syntax, structure and semantics significantly improves interoperability and reuse of web resources
- RDF (Resource Description Framework) is a W3C initiative designed to support the description of web resources
- Collections are Web Resources

RDF Overview:
- W3C Initiative
- RDF is a metadata architecture
  - Uses XML (eXtensible Markup Language) as a transfer syntax
  - Allows for semantics to be defined by individual resource description communities
- Enables interoperability between applications that exchange metadata
- Targeted for many resource descriptive application areas (e.g. collection level description)

RDF for Collections:
- RDF provides a sound basis for associating properties with resources
- RDF provides a sound basis for defining the relationships between resources
- RDF can be used to define collections
  - Proposed for “site-maps”
  - Proposed for bookmarks

Bookmark Collection in RDF:
- `<RDF:RDF>`
- `<RDF:Description id="root" BM:Name = "Bookmarks for Eric Miller">`
- `<RDF:Description id="885608620RF.rdf" BM:Name="RDF">`
- `<BM:Child RDF:HREF="http://www.w3.org/XML/ns-sched.htm" BM:Name="Namespace schedule">`
</RDF:Description`
- `<RDF:Description id="885607908WWW.rdf" BM:Name="WWW">`
</RDF:Description`
- `<RDF:Description>`
</RDF:RDF>`
AMICO

- AMICO is an interesting collection for testing these collection tools
- Well-structured data
- Lots of interesting relationships (more than simple parent-child)
- Lots of interesting data
- Rich enough for discovering interesting sub-collections

AMICO Item in Collection

AMICO Related Resource

Conclusion

- Collections are increasingly important
  - Metadata to support collection level description is important
  - Tools to support collections therefore are also important
- AMICO is a fine set for testing the utility of these tools
- These are all active topics of research at OCLC