A103 Search Using Metadata

James A. Larson
Larson Technical Services
and Information Today, Inc
Metadata

• Metadata = data about data
• Used to locate requested data quickly
• Examples of metadata from real life
  – Table of contents at the front of a book
  – Index at the back of a book
  – Card catalog in a library
  – Directory of the day’s events posted in a hotel lobby
Standard XML-based metadata

- OWL (Web Ontology Language)
- RDF (Resource Data Format)
- `<meta>` and `<metadata>` tags in XML documents
- XML tags
The Big Picture of Voice and XML

Formulate request

Search metadata

Present results

OWL

RDF

<meta> and <metadata> tags

XML tags

Data

Data

Data
XML Tags

- Used to locate information of the desired type

<employee>
  <name> Smith </name>
  <salary> 22000 </salary>
  <hireDate> 2006-04-09 </hireDate>
</employee>

OWL (Web Ontology Language)
RDF (Resource Data Format)
<meta> and <metadata> tags in XML documents

XML tags
<meta> Tags

Used to
- Provide documentation
- Search <meta> tag contents instead of searching ALL of the words in a document

<meta name="keywords" content="tree, plant, perennial, trunk, branches" />

OWL (Web Ontology Language)
RDF (Resource Data Format)
<meta> and <metadata> tags in XML documents
XML tags
<metadata> Tags with “Dublin Core” Annotations

• Used to search and query

RDF (Resource Data Format)

OWL (Web Ontology Language)

<meta> and <metadata> tags in XML

XML tags
What Is RDF?

• RDF stands for Resource Description Framework
• RDF is a framework for describing resources on the web
• RDF provides a model for data, and a syntax so independent parties can exchange and use code
• RDF is designed to be read and understood by computers
• RDF is not designed for being displayed to people.
• RDF is written in XML.
• RDF is a part of the W3C’s Semantic Web Activity.
• RDF is a W3C Recommendation.
RDF (Resource Description Framework)

- Used primarily for link chasing during browsing

```xml
<Description about="some.uri/evergreen">
  <isA resource="some.uri/tree"/>
</Description>
<Description about="some.uri/pine">
  <isA resource = "some.uri/evergreen " />
</Description>
```

- RDF (Resource Data Format)
  - OWL (Web Ontology Language)
  - `<meta>` and `<metadata>` tags in XML documents
  - XML tags
What Is Ontology?

- Ontology is a description of things and their relationships.
What Is OWL?

- OWL stands for Web Ontology Language.
- OWL is built on top of RDF.
- OWL is for processing information on the web.
- OWL was designed to be interpreted by computers.
- OWL was not designed for being read by people.
- OWL is written in XML.
- OWL has three sublanguages.
- OWL is a web standard.

© 2008, Larson Technical Services
OWL Is Different from RDF

• OWL and RDF are much of the same thing, but OWL is a stronger language with greater machine interpretability than RDF.

• OWL comes with a larger vocabulary and stronger syntax than RDF.
OWL (Resource Description Framework)

- Used primarily for reasoning

\[
\text{pine isA evergreen} \land \text{evergreen isA tree} \rightarrow \text{Pine isA tree}
\]

- RDF (Resource Data Format)
- XML tags
- `<meta>` and `<metadata>` tags in XML documents

© 2008, Larson Technical Services
What Is DITA?

• Darwin Information Typing Architecture
• An XML-based architecture for authoring, producing, and delivering technical information.
Using DITA

• Search topics
• Combine and format topics
• Present to user
Where Does Metadata Come From?

• Reuse existing metadata
• Build it yourself
  – Manual development
  – Automatic extraction of metadata from documents

OWL (Web Ontology Language)
RDF (Resource Data Format)
<meta> and <metadata> tags in XML documents
XML tags
Summary

- Use metadata to reflect the content of your web pages
- Make metadata part of your searching strategy
- Review detailed language specifications

OWL (Web Ontology Language)
http://www.w3.org/2004/OWL/

RDF (Resource Data Format)
http://www.w3.org/RDF/

<meta> and <metadata> tags in XML documents
http://dublincore.org/

XML tags
http://www.w3.org/XML/
Questions