Metadata:
The Organization of Information

LIB 1201

Monday, October 15, 2012

Why must information be organized?

What is metadata?

- information about information
- a surrogate for the real thing
- can be embedded in an object
- can be stored separately and linked via database
- or both, for example: books

What does metadata do?

- describes resources for the purpose of organizing them

- makes resources findable

- supports archiving and preservation

- NISO 2004, Understanding Metadata (pdf)

Two ways metadata is used:

- Taxonomies

- Folksonomies

Taxonomy

- A hierarchical classification system
- Library catalog:
    - Helps you find information on shelves
    - Helps librarians with cooperative cataloging
    - Helps control for variation in words used to describe similar topics
- Controlled vocabularies
    - Subject headings, subject terms, thesaurus

Folksonomy

- For example, tagging our blog posts
- User-generated metadata
- "Flat" non-hierarchical system
- Flexible, individual, social
- Used in blogs, wikis, other social media
  - "Like" or "favorite" is a familiar example
- Can help create community of users around shared knowledge and interests

Why are we talking about metadata in this class?

Activity

- Pick two blog posts of yours from anytime during the semester
  - If these posts are missing tags, add them now
- Search here http://id.loc.gov/ and assign at least 2 subject headings to your 2 blog posts; tag your posts with these subject headings
- Or se the Library of Congress website at http://www.loc.gov/catdir/cpso/lcco/ to browse subjects & subject headings
- Share your results with the class

Metadata for non-text media

Challenges

- For text, computer programs can pull out metadata
  - for example, search engines search keywords
- For images, videos, multimedia machine-generated metadata is trickier
  - Some added during the creation process
  - Filenames can also contain metadata
  - Human-generated metadata is time-intensive

Process

- New technologies can assign machine-generated metadata to non-text media
  - Content-based image retrieval (pixel recognition)
  - Voice recognition
  - Optical character recognition (OCR) for text scanned as images
- Users can assign metadata during the publishing or storage process

For Wednesday, October 17

Discussion: Search mechanics; what is a database; how a search engine works

Reading: Badke Ch. 3, Liddy, Leibman

Assignment: One reading response blog post