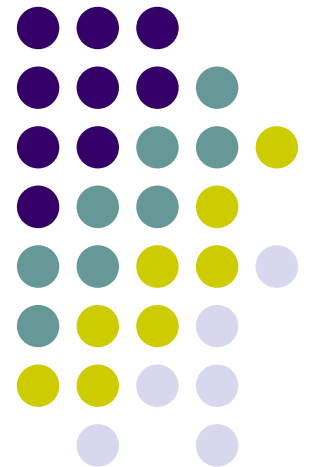# Cushman Exposed!

Exploiting Controlled
Vocabularies to Enhance
Browsing and Searching of an
Online Photograph Collection

Michelle Dalmau and Jenn Riley
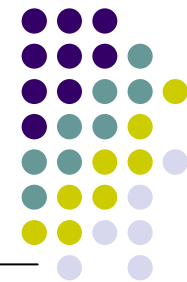Indiana University Digital Library Program

Overview

- Background Information

- Metadata

- Browse and Search Functionality
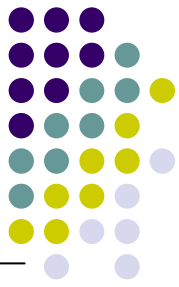
Controlled Vocabularies for Increased Access

- Why do we use controlled vocabularies (CV)? To increase access!

- But do we really increase access when we don't utilize the CV's structure in browsing and searching? No!

- So then, how can we utilize the CV structure to promote access?

Looking Back

- U.S. Steel Gary Works Photograph Collection
  - ~2,200 Images
  - Archival descriptions
  - Assigned subject terms from CV
- Subject field search requires referencing the A-Z list of subjects
- Usability studies revealed not using the CV's syndetic structure impacts searching

The Cushman Collection

- Funded with an Institute of Museum & Library Services (IMLS) grant

- ~15,000 color slides taken between 1938-1969

- Cushman provided a significant amount of description

- Additional metadata to enhance genre, subject and geographic access

Metadata for Image Collections

- Advantages to "free-text" descriptions:
    - Preserve photographer's notations
    - Resembles the user's language
- Advantages to CV descriptors:
    - More access points
    - Structure
    - Collocation
    - Disambiguation
    - Interoperability
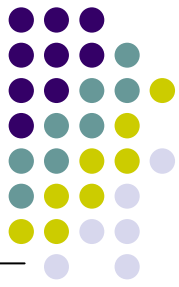
Metadata for the Cushman Collection

- Cushman's description
  - Dates
  - Location
  - Names
- TGM I – LC Thesaurus for Graphic Materials: Subject Terms
- TGM II - LC Thesaurus for Graphic Materials: Genre & Physical Characteristics
- TGN – Getty Thesaurus of Geographic Names

TGM I: Subject Terms Strengths and Weaknesses

- Strengths include:
    - Pre-defined relationships between concepts
    - Some lead-in vocabulary
- Weaknesses include:
    - Syndetic relationship lacking for new terms
    - Language not user-friendly
    - Not enough lead-in vocabulary
    - Form and number of top-level categories not useful for a browse structure

Browsing Image Collections

- Research shows:
  - Browsing is exploratory (Bawden)
  - Guided, flexible browsing in context works (Flamenco and SI Art Image Browser projects)
- Our Usability studies show:
  - Structure is important
  - Contents should be easily exposed
  - Flexible and combinatorial browsing is desired
  - Browsing cultivates searching

Searching Image Collections

- Research shows:
  - Using thesaurus structure helps searching (Greenberg)
    - Automatic expansion of synonyms and narrower terms
    - User-initiated expansion of broader and related terms
- Our Usability studies show:
  - Referencing an A-Z list with no lead-in terms for searching is NOT helpful at all
  - Concerns about word choice
  - Iterative reformulation of queries in context is desired
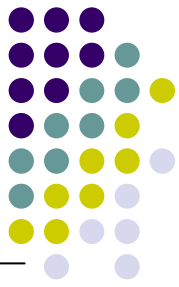
Cushman Specifications: Browsing
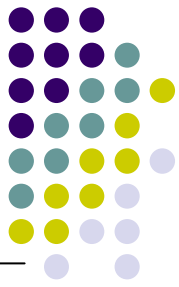
- Date
- Genre
- Subjects (hierarchical)
  - Access via broadest term or actual descriptors with ability to move up and down (pending user studies)
  - Related terms for lateral movement may also be supported (pending user studies)
- Location (hierarchical)
- Combination of categories

Cushman Specifications: Searching

- [Integrated search](#) against BOTH "free-text" descriptions and thesaurus

- Mapping from lead-in vocabulary

- Retrieval of all records with narrower terms

- [User-initiated](#) broadening and narrowing

- Related term access

Browsing Implementation

- To be implemented
  - Hierarchical (subjects & location)
  - Combinatorial
- Under investigation:
  - How do we preserve the structure, when only 1/3 of TGM I terms are used, to effectively browse the collection?
  - Grouping by broadest terms or assigned descriptors (transportation facilities or bridges)?
  - How to present results comprehensibly when narrower terms are expanded automatically?

Searching Implementation

- ## To be implemented
  - Integrated search against BOTH "free-text" descriptions and thesaurus
  - Mapping from lead-in vocabulary
  - Retrieval of all records with narrower terms
- ## Under investigation
  - User-initiated broadening and narrowing
  - Related term access

Looking Forward

- Strive to make our collections truly accessible even if only incrementally

- Discovery of new technologies to help facilitate an integrated and user-initiated search is underway (Oracle 9.i Thesaurus Support)

- Ongoing user studies to address integrated browse and search needs for Cushman and future image collections

- Defining functionality for future image repository for all IU image collections

References

- Bawden, D. (1993). Browsing; theory and practice. *Perspective in information management*, 3 (1): 71-85.
- Choi, Youngok and Rasmussen, Edie M. (2003). Searching for Images: The Analysis of Users' Queries for Image Retrieval in American History. *Journal of the American Society for Information Science and Technology*, 54 (6): 498-511.
- Greenberg, J. (2001). Optimal QE Processing Methods with Semantically Encoded Structured Thesauri Terminology. *Journal of the American Society for Information Science and Technology*, 52 (6): 487-498.
- University of California, Berkeley: Flamenco Project -- **http://bailando.sims.berkeley.edu/flamenco.html**
- University of Michigan: SI Art Image Browser -- **http://www.si.umich.edu/Art_History/**