Establishing the Value of Socially Created Metadata for Image Indexing

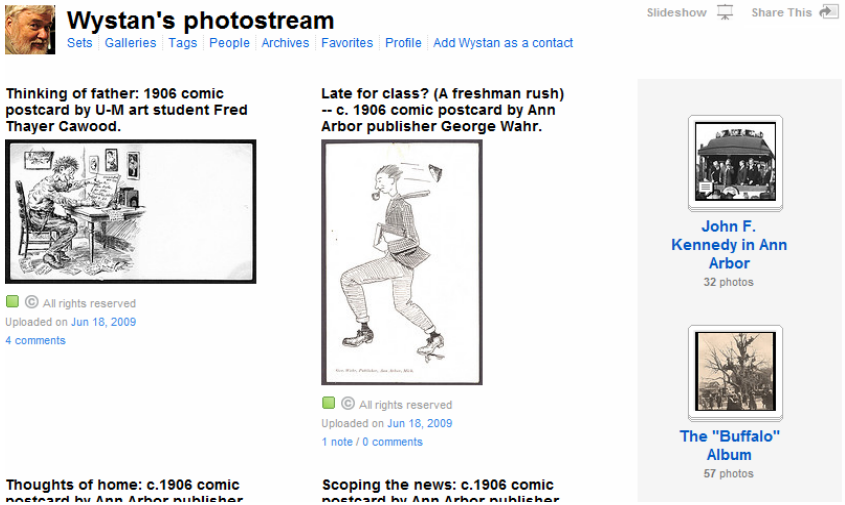
Besiki Stvilia, Corinne Jörgensen,
Shuheng Wu

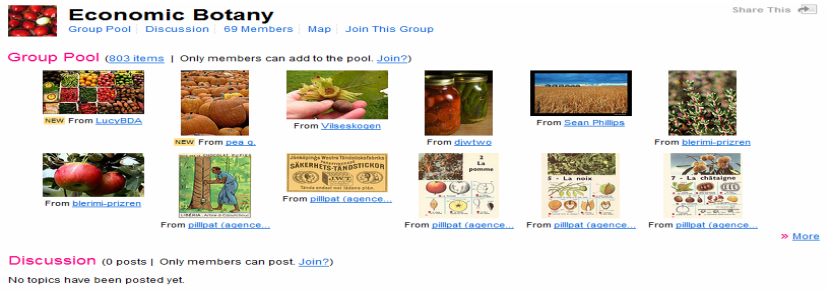
School of Library and Information Studies
Florida State University
ALISE, January 2011

The study was partially supported by
an OCLC/ALISE Research Grant, 2010

Networked collections, many subjects

- Many roles for users:
  - Providers
  - Indexers
  - Annotators
  - Aggregators
  - Quality control agents





- Need support for:
  - Multiple user activities
  - Not just search/retrieve

2010 Study: Value Added by “Tags”?

- What are the relationships among socially-generated metadata, controlled vocabularies, and user characteristics?
  - Can “tags” add value to controlled vocabularies?
  - Can useful relationships be extracted from metadata sources?
  - Are metadata concepts medium-specific?
  - Can combining collection- and item-level metadata results in a better quality folksonomy?
  - Are there relationships between user demographics and their valuations of social terms?
  - What are the demographic characteristics of productive and quality contributors of index terms?

Overall approach taken

- Extract concepts, terms, and relations (thesaurus elements) from Flickr and Wikipedia metadata
- Integrate these with metadata from controlled vocabularies
  - Thesaurus for Graphic Materials (TGM)
  - Library of Congress Subject Headings (LCSH)
- Explore if value is added to the controlled vocabularies

Methodology

- Guided by an earlier proposed metadata value measurement model (Stvilia & Gasser, 2008)
- Adapted the experimental design used by Jörgensen (1998) and Chen et al. (1995)
- A combination of methods, including controlled experiments, auto-ethnographies (“diaries”), semi-structured interviews, content and statistical analysis
- Evaluating added value of social metadata along two facets: *perceived* and *objective*

Experiment Procedure

- 35 participants – students and staff @ CCI
- Experimental System: modified Steve Tagger
- Experiment Tasks
  - Describe each photograph spontaneously by assigning tags
  - Rate pre-assigned index terms on usefulness to description task on a five-level Likert scale
  - Write queries to locate the photographs with a search engine

Data

- 10 photographs selected from the set of 7,192 photographs of the LoC Flickr photostream
  - on different subjects
  - with moderate number of tags to allow for task completion

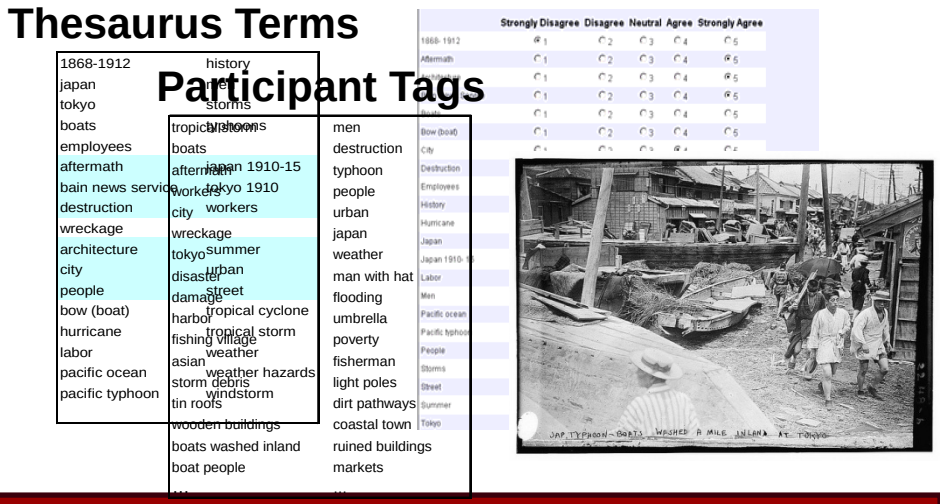
- Index terms selected from TGM, LCSH and Flickr tags, supplemented with related terms from the LoC Flickr folksonomy, Flickr relatedTags API, and English Wikipedia
- Sets of terms used by participants in description and search tasks

Descriptive Statistics

- Undergraduate students - 46%
- Doctoral students - 28%
- Master's students - 20%
- Master's degree holders - 6%
- Age ranged from 19 to 59 years old
- Female - 43%
- Male - 57%
- Tagging experience - 34%
- Intermediary knowledge of indexing - 11%
- Non-native speakers of English - 29%
- Median number of tags used in description task – 5
- Median number of query terms - 4

Facet One – Perceived Value

- Ratings of social terms were significantly higher than the baseline rating score – 3 (‘neutral’)
- But significantly lower than TGM & LCSH term ratings (mean=3.7; median=4)



Facet Two – Degree of Coverage

- Degree of added coverage provided by the social terms
  - Median percentage of added coverage provided by the social metadata was above 100%
    - 127% for description task terms (i.e., tags)
    - 108% for search task terms (i.e., query terms)

$$addedCoverage = \frac{\#\,of\ \_Tags\ \_Covered\ \_by\ \_Social\ \_Index\ \_Terms}{\#\,of\ \_Tags\ \_Covered\ \_by\ \_TGM\ \&\ LCSH\ \_Terms} \times 100$$

Facet Two – Degree of Coverage

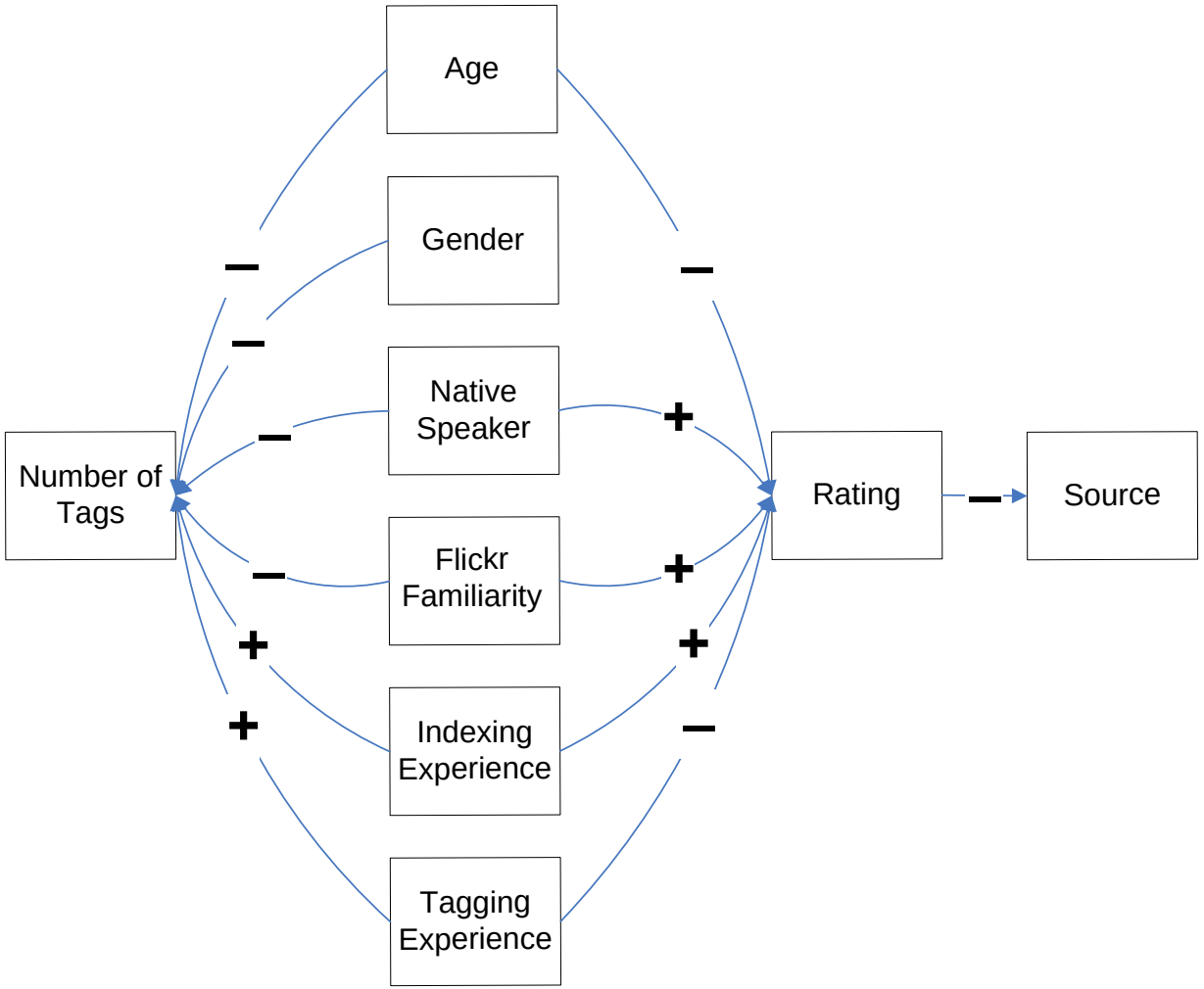
- The median set overlaps with the complete set of pre-assigned index terms, and with TGM & LCSH terms alone were:
  - for Tag terms from Description Task: 0.1 vs. 0.05
  - for Query terms from Search Task: 0.19 vs. 0.12

- The median set overlap of query terms from Search Task with tag terms from Description Task was even higher – 0.23

The degrees of the overlap were calculated as Dice coefficients

$$c = \frac{2|A \cap B|}{|A| + |B|}$$

Demographics, Ratings, & Productivity



Discussion

- Social terms provide an added value to the controlled vocabularies and the activity of image indexing in general
  - Social terms were perceived mostly useful
  - Median rating for the social terms was significantly higher than the baseline – the neutral rating.
  - Addition of social terms provided twice higher coverage of participant terms on average than controlled vocabulary terms alone

Discussion

- The sets of descriptors assigned by participants provided best coverage of query terms
    - The overlap between query terms and descriptors used by participants was the highest – 0.23

Discussion

- The TGM and LCSH captured most important and preferred terms
  - Participants valued controlled vocabulary terms higher than the social terms
  - The value of folksonomies are in extending and enhancing expert created KOS by providing additional descriptors and access points, not in substituting KOS.

Discussion

- The relationships between participant demographics and term ratings
    - The participants with greater tagging experience evaluated terms less favorably than the participants with indexing experience

Discussion

- The relationships between participant demographics and the number of terms used in description task
    - Positive interaction with indexing and tagging experiences
    - Older, male, native speakers, and participants with higher Flickr familiarity assigned lower numbers of tags

Next Steps & Limitations

- Next Steps
  - Complete the analysis of the exit interviews and search diaries to gain an additional insight into the user's value structure for index terms and image seeking behavior
  - Investigate the relationship between user demographics and term quality

- Limitations
  - Replicating the experiments with a more representative and diverse sample of participants would be desirable

Thanks!