

A Modest (*Metadata*) Proposal

Elizabeth D. Liddy

Geri Gay, Sarah Harwell, Tina Finneran

Center for Natural Language Processing

School of Information Studies

Syracuse University



Center for Natural Language Processing

Questioning of Assumptions

- Do we need metadata?
 - Why?
- If so, how much metadata do we need?
 - For what purposes?
- Which elements do we need?
 - For which digital library tasks?
- How do information-seekers utilize the metadata?
 - When browsing / searching / previewing?
- Can automatically generated metadata perform as well as manually assigned metadata?

Why Evaluation is Needed

- Only 1 study has empirically evaluated metadata's utility
 - 33% Recall and 17% Precision for 24 queries searched against the EPA website's metadata records (Quam, '01)
- Digital Libraries need guidance on appropriate set of metadata elements
 - For use by both humans and machines
- Will findings replicate past manual vs. automatic indexing experiments, which showed that:
 - Automatic performs better
 - Automatic has greater consistency
 - Automatic is preferred by professional searchers

Life-Cycle Model of Metadata Evaluation

1. Initial generation

- Methods
 - Manual
 - Automatic
- Costs
 - Time
 - Human Resources
 - Technology

2. Accessing DL resources

- Users' interaction
 - Browsing
 - Searching
- Relative contribution of each element

3. Search Effectiveness

- Precision
- Recall

Proposed Evaluation Methodology

1. Automatically metatag a DL collection that has already been manually meta-tagged
2. Solicit range of appropriate DL users
3. Have users qualitatively evaluate metatags
4. Conduct searching & browsing experiments
5. Monitor with eye-tracking & think-aloud protocols
6. Develop metrics of relative utility of each meta-data element (manual & automatic) for both tasks
7. Report results on query-by-query + average basis

Behavioral Experiments

- How do users seek/find meaningful resources?
 - What are users thinking, doing, looking at?
 - Which metadata elements are most important?
- Collect demographic & personality data
 - Does study replicate contribution of role in prior experiments?
- Eye-tracking & think-aloud protocols
 - Number & duration of fixations, scan paths, dilation, etc

Quality of Representation Experiments

- Two methods – qualitative vs. quantitative
- Two generation conditions – manual vs. automatic

	Qualitative	Quantitative
Manual	?	?
Automatic	?	?

Manual vs. Automatic Generation

- Qualitative Evaluation
 - Larger version of pilot study
 - Teachers, education students, & professors
 - *Does the metadata record usefully reflect the content of the resource?*
 - Expectation & Satisfaction conditions

Qualitative Pilot Study Results

<u>Generation Method</u>	<u>Expectation</u>	<u>Satisfaction</u>	<u>Combined</u>
Manual (average)	(153) 4.03	(571) 3.81	(724) 3.85
Automatic (average)	(139) 3.76	(532) 3.55	(671) 3.59
<hr/>			
Average Score Difference	.27	.26	.26

Manual vs. Automatic Generation

- Qualitative Evaluation
 - Larger version of pilot study
 - Teachers, education students, & professors
 - *Does the metadata record usefully reflect the content of the resource?*
 - Expectation & Satisfaction conditions
- Quantitative Evaluation
 - Run queries against manual and automatic records
 - Evaluate and compare for Recall and Precision

Possible Operational Outcomes

MetaData Generation



Desired Achievements

- Provide experimental results to guide Digital Library development
- Develop metrics of metadata quality & utility
- Inform HCI design
- Reduce the metadata generation bottleneck
 - Redo metadata standards to include only useful elements
 - Determine if automatic metadata generation can perform comparably to manually generated metadata