



Evaluating Metadata from Different Perspectives

Anne R. Diekema


Center for Natural Language Processing

Syracuse University

Metadata Tools for Digital Resource Repositories

JCDL Workshop, June 15, 2006

Today's Talk

A photograph showing a group of diverse children of various ethnicities and ages gathered around a globe in a classroom. They appear to be engaged in a learning activity, with some pointing at the globe. The background is slightly blurred, focusing attention on the children and the globe.

- Automatic metadata generation
- CNLP's metadata projects
- Evaluation Perspectives
- Conclusion

Automatic Metadata Generation

A photograph showing a group of diverse children in a classroom. They are gathered around a globe, looking at it with interest. The children are of various ethnicities and ages, and the scene is brightly lit, suggesting a positive learning environment.

- The generation of metadata through use of extraction or assignment software programs
- Many ways to evaluate this type of metadata and the systems generating this type of metadata

CNLP metadata projects

- MetaTest
 - evaluation of metadata from generation to use
 - MetaExtract
 - various metadata elements
- CASAA
 - Computer Assisted Standard Assignment and Alignment
 - CAT
 - standards metadata only

MetaTest Project



1. *Metadata quality study*

- How well does the metadata describe the resource?

2. *Information retrieval experiment*

- How well does automatic metadata perform compared to manual metadata and full text in an information retrieval system?

3. *Metadata user study*

- How, and in what ways, does metadata contribute to users' browsing in a digital library?

CASAA Project

- NSDL-funded service project
- Assists library developers in assigning content standards to learning resources
- Will provide a crosswalk between the different state and national standards
- Improves the ability of teachers to locate resources in the NSDL that will support their standards-based instruction





Evaluation perspectives: system versus user evaluation

- Creators of metadata (system development)
 - Metadata assignment quality
 - Metadata assignment coverage
 - Performance of metadata in search
 - Performance of learning for assignment
- Users of metadata
 - Difference between manual and automatic
 - Use of metadata in search
 - Use of metadata in browsing

Evaluation perspectives: metadata aspects



- Metadata assignment quality
- Metadata assignment coverage
- Metadata in searching or browsing
- Machine Learning and metadata assignment

Metadata Quality and Coverage

– Evaluation Methodology

- Create test collection by manually assigning metadata to selected resources
- Automatically tag the same collection
- Compare assignments
- Evaluation can be automated
- Inconsistencies not necessarily mistakes
- Agreement between human catalogers often lacking

Metadata Quality Evaluations 1



- Test collection: Math & Science lesson plans
 - Manual metadata: GEM digital library
 - Automatic metadata: CNLP's MetaExtract
- Subjects: 30 STEM educators reviewed metadata + document pairs
 - Unaware if were automatic or manual records
 - Judged quality of each element present as well as overall quality
 - *Very Poorly, Poorly, Well, Very Well, Unsure*
 - If (*Very*) *Poorly*, asked to comment on what was wrong

Quality of Metadata Assignment

- Results

- Of the 8 elements for which there was sufficient items with manual assignment to make comparisons
 - **Statistically insignificant** differences between manual and automatic on six elements
 - Title and Keyword were **barely significantly** better in the manually assigned records

Metadata Quality Evaluations 2



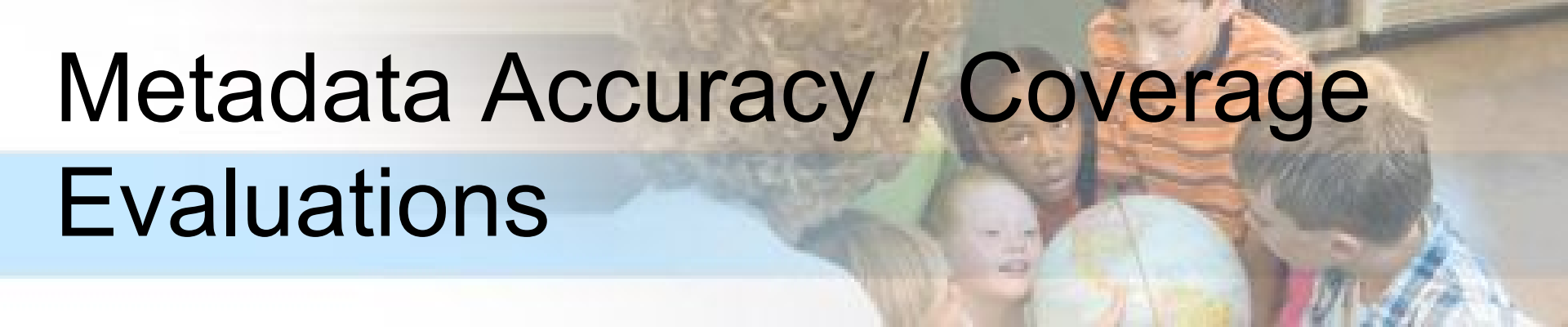
- Information Retrieval evaluation for internal system testing of quality and coverage
- Test collection with 403 lesson plans
- Automatic evaluation
- Recall (coverage measurement)
- Precision (accuracy measurement)

Metadata Coverage Evaluations 1

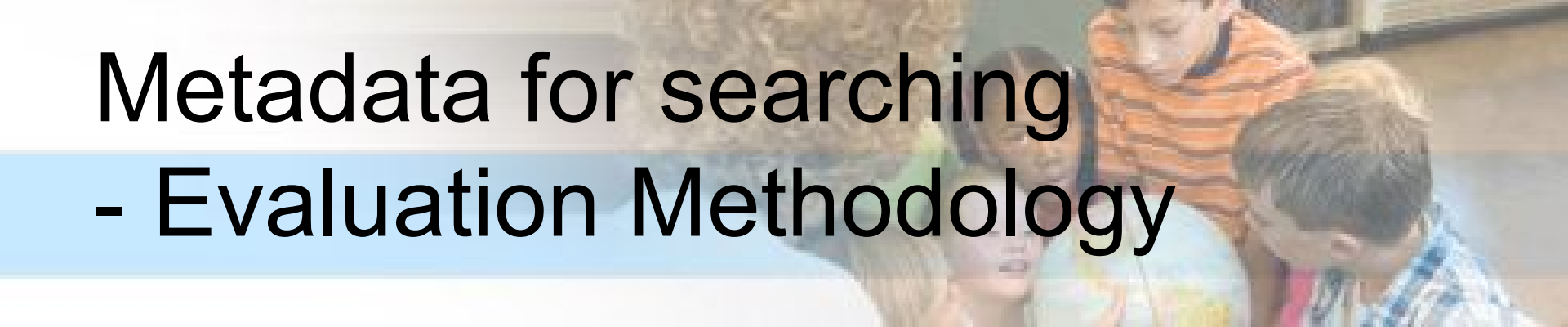
A photograph of several children in a classroom setting, gathered around a globe. They appear to be engaged in a learning activity, with some pointing at the globe. The background is slightly blurred, showing other students and classroom elements.

- Significantly poorer coverage by manual vs. automatic assignment method
- Only 11 of the 23 possible elements were assigned frequently enough to be analyzed and compared
- Most frequently assigned elements
 - Title
 - Duration
 - Keyword
 - Essential Resources
 - Description
 - Pedagogy Group / Method
 - Grade
 - Pedagogy Process
 - Relations
 - Pedagogy Assessment

Metadata Accuracy / Coverage Evaluations



- Shodor
 - On average, 3 out of 10 of CAT's suggestions were the same as the manual assignments
 - Additional 1-3 of the standards were also good
 - Reason why user is in the loop
 - Should **improve** over time as the system learns
- TeachEngineering
 - Similar results with fewer additional suggestions



Metadata for searching

- Evaluation Methodology

- Information Retrieval experiments
- Test Collection with lesson plans
- 20 Information Needs (queries) developed by certified teachers who also judged relevance
 - Varied on age-level, topic, and specificity
- Three different retrieval runs (trials)
 - fielded metadata search
 - free-text retrieval on metadata
 - full-text retrieval on entire lesson plan

IR Experimental Design

Trial I: Fielded retrieval on metadata (automatic & manual)

Automatic
Metadata

Title: Black Hole Lesson
Grade: 9-11
Keywords:
galaxy,black_hole,
set, model, object,
dark matter

Manual
Metadata

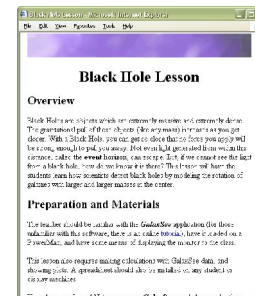
Title: GalaxSee: Black Hole Lesson
Grade: 9, 10, 11, 12
Keywords: Interactive teaching, simulations, Universe, GalaxSee software

Trial II: Free-text retrieval on metadata (automatic & manual)

Black Hole Lesson,
Grade: 9-11,
galaxy,black_hole,
set, model, object,
dark matter,...

GalaxSee: Black Hole Lesson, Grade: 9, 10, 11, 12, Interactive teaching, simulations, Universe, GalaxSee software,...

Trial III: Free-text retrieval on lesson plan



Sample Information needs



- **Abstract Type of Information Need:**
 - *I want to teach sixth graders about ecosystems, namely the factors that affect their stability.*
- **Specific Type of Information Need:**
 - *I would like students to do an experiment that shows how acid rain effects the environment.*

Metadata for searching

- Results

- Search precision better for auto assigned metadata
 - Performs as well or better than manually assigned metadata
 - Performs comparably to full text
- Precision of searches on manually assigned metadata varies widely
 - Strongly influenced by instances when not present

A photograph of four children, two girls and two boys, gathered around a computer monitor. They appear to be engaged in a collaborative activity, possibly a learning exercise or a game. The children are looking intently at the screen, which is partially visible. The background is slightly blurred, suggesting an indoor setting like a classroom or a library.

Metadata for browsing - Evaluation Methodology

- Eye-tracking experiments at Cornell
- User wears an eye-tracking device while browsing or searching STEM resources with and without metadata
- Eye fixations (stops) and saccades (gaze paths) are recorded

One Search Scanpath

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit Discuss

Address \\Testing\metatest\silicon_MDEDESC2.htm

Format	text/html
Language	en

Identifier <http://chemistry.beloit.edu/Chip/index.html>

Rights Copyright 2000 by the Trustees of Beloit College and the Regents of the University of California. Published through exclusive license with John Wiley Sons

The Chip Firing Game and Related Models

In this paper, we use a chip firing game as a model in physics, economics, computer science, and biology. We show that the stationary configuration space of such a system started in any configuration is finite. The lattice structure of the configuration space of a chip firing game is of great interest since it implies convergence (at least) if the configuration space is finite. We show that there is another kind of convergence: all the configurations reachable from two given configurations are reachable from a unique first configuration. Moreover, the chip firing game is a very general model, and we show how known models can be encoded as chip firing games, how some results about them can be deduced from this paper. Finally, we define a new model, which is a generalization of the chip firing game, and about which many interesting questions arise.

Subject	Pattern Formation and Solitons Adaptation and Self-Organizing Systems Cellular Automata and Lattice Gases	Creator	Phan, H. D. Latapy, M.
		Format	text
		Type	Text
		Date	2000-08-02

Identifier <http://arXiv.org/abs/nlin/0008003>

HyperTransport Technology: Simplifying System Design

HyperTransport technology is an architecture that allows data transfer between chips in excess of ten gigabytes per second. This white paper from Advanced Micro Devices (AMD), released in October 2002, considers the potential of HyperTransport in designing an input/output bus for a microprocessor. The technology, while accommodating high speed demands, requires relatively little power, making it ideal for a wide range of

Metadata for browsing

- Results 1

- When Description element is available
 - Users almost exclusively spend their time in the Description, Title, and Subject
 - Ignore all other elements
- When Description element is NOT available
 - Eye-fixations are evenly distributed across most of the metadata tags
 - Peaks in Title, Subject, Rights, and Creator fields

Metadata for browsing

- Results 2

- After viewing metadata, users were significantly faster and better at reformulating searches
- Benefited most from seeing Title and Description metadata elements
- Significantly more use of metadata in searching by subjects who were familiar with metadata






Machine Learning Metadata Assignment

– Evaluation Methodology

- Work done by DLESE Program Center (Holly Devaul)
 - Manually assign standards to set of 30-35 records
 - Submit resource URLs to system for comparative results analysis (using learning function)
 - Study proportion of suggested standards that match the human-assigned ones
 - Study nature of the matches and mismatches
- Ongoing research, no results yet



Machine Learning Metadata Alignment

– Evaluation Methodology

- Multi-label text classification approach to align standards metadata
- Collection: A2A+McRel compendix (611 labels)
- Minimum training size constraint resulted in 133 labels
- The training to test ratio was set at 0.75 (30 train, 10 test)

Alignment Metadata Evaluation



Feature Type	Precision	Recall	F-Measure
leaf	58.3	42.3	47.0
leaf + hierarchy	58.0	43.3	47.6
vocabulary	92.9	84.5	87.8
leaf + hierarchy + voc	81.5	76.7	78.4

Conclusion

- Many possible evaluations
- Pragmatic approach – what do you want to find out? What type of evaluation needed?
- Mostly driving system development
- Automatic evaluation sometimes possible
- Users needed in any case
- Use of domain experts to create test collection





This material is based upon work supported by the National Science Foundation under Grant No. 0435339. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.