Automatic Mapping of Resources to Content Standards

NSF/NSDL-2: DUE-0121543

Jiangping Chen, Anne Diekema, Elizabeth Liddy

Objectives of the Project

• Explore effective strategy for metadata generation for educational materials

• Develop middleware tool for the automatic assignment of content standards and benchmarks
Strategies Explored

• Automatic text categorization

   Problems
   – Sparse training data
   – Only 24% of Math and Science standards covered by more than 1 Lesson Plan
   – Lesson plans fit multiple categories

• Information retrieval

Methodology

• Information Retrieval for standard assignment

   – Standards as documents
   – Educational materials such as lesson plans as queries
   – Standard assignment as matching
   – Information retrieval experiments for evaluation
Standard Assignment as Retrieval Process

- Lesson Plan as Query
- Indexed Terms from Standards

Matching Standard to Lesson Plan

Procedures

- Document processing
  - Splitting, Document expansion, and NLP processing
- Query processing
  - Stemming, query weighting, query filtering, formulation
- Indexing & Retrieval
  - Phrases as well as words as index terms
  - CNLP search engine
Splitting

- Converting large Standard document into short passages
- Name each passage with name reflecting classification code of that particular benchmark.

Math

Problem Solving

1

1.1.1

Draws pictures to represent problems

pictorial representation
Math
Problem Solving
Draws pictures to represent problems
pictorial representation
</TEXT>
</DOC>

Document Expansion

- Expand the Standards with vocabulary from pre-assigned educational resources and other resources
- Boost Standard content
Math Problem Solving
Draws pictures to represent problems
pictorial representation

addition
subtraction
picture
representation
illustrate
illustration
depict
sketch
draw
Query Formulation

- Generating queries by extracting important terms from the lesson plans
- Stemming
- Stopword removal
- Query filtering
- Query weighting

Description: Students can develop strategies for learning multiplication facts by finding patterns on a multiplication table.

Goals: Students will develop strategies for learning multiplication facts.

Objectives: Student will:
- discover patterns on a multiplication table
- create strategies for factors 4-9

Materials:
- Overhead Projector
- Multiplication tables for each student

Procedure: ………
Automatic Evaluation

• Manual evaluation is costly
• Developed test collection for automatic evaluation
• Relevance judgments for 297 queries (lesson plans)
• Very few relevant documents per query (75% <= 3)
• Performance measure to fit the task
  – Precision/recall at n-th document level
  ( n = 1, 2, 3… 10)
Information Retrieval Experiments

- Different ways for document processing
  - Expanded vs. non expanded standards
  - NLP processed vs. original standards

- Different ways for query processing
  - Words vs. phrases
  - Query size

Retrieval Results on Math and Science Standards

<table>
<thead>
<tr>
<th>Experiment</th>
<th>N</th>
<th>Relevant standard in top 10</th>
<th>Precision at rank 1</th>
<th>Recall at rank 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanded, 15 terms</td>
<td>297</td>
<td>195</td>
<td>0.3468</td>
<td>0.1888</td>
</tr>
<tr>
<td>Not expanded 20 terms</td>
<td>297</td>
<td>222</td>
<td>0.3434</td>
<td>0.1915</td>
</tr>
<tr>
<td>Expanded, 20 terms</td>
<td>297</td>
<td>262</td>
<td>0.5657</td>
<td>0.3152</td>
</tr>
<tr>
<td>Expanded 25 terms</td>
<td>297</td>
<td>265</td>
<td>0.5623</td>
<td>0.3158</td>
</tr>
</tbody>
</table>
Findings

• Standard expansion increased performance
• Query filtering increased performance
• Phrase extraction did not help performance
• Unsuccessful attempts (need further investigation)
  – Relevance feedback: no effect on performance
  – Topic-ranking: negatively affects performance
  – Using metadata: no effect on performance

Failure Analysis

• Vocabulary gap between standards and lesson plans
  – standard contains abstract highly theoretical language, lesson plans contain specific language
• Query representation: missing important query terms:
  – challenge to extract an appropriate number of key terms from lesson plans, some of which may be quite long
• Incompleteness or error in the relevance judgment file
Vocabulary Gap – An Example

lesson plan A

: ....... Inform the students that they will have to find a place to live. Have several Newspapers ready for them to find a suitable place to fit their "families". They have to write a check to you for rent, deposit, etc. The students will also have to write out checks for all of their bills, extra spending, or anything extra that might pop up.

Standard 8.1.3.4.xml

Math
Problem Solving
Formulates a problem, determines information required to solve the problem, chooses methods for obtaining this information, and sets limits for acceptable solutions
problem solving
problem formulation
information
method
data collection
limit
solution
solving problem
analyze

Lesson plan A should match Standard 8.1.3.4.xml.
However, none of the language overlaps

Ongoing Research

• Manual Evaluation of Standard Assignment
Future Research

• Filling the gap
  – High level understanding of the lesson plans –
    develop summarization capability for
    educational resources.
  – Expanding the Standards employing appropriate
    knowledge
  – Expand matching to allow semantic alternatives
• Improving query representation
  – Template querying
• Machine learning with larger training set

http://www.cnlp.org

Center for Natural Language Processing