Experimenting with the Automatic Assignment of Educational Standards to Digital Library Content

Anne R. Diekema
Center for Natural Language Processing
4-206 Center for Science and Technology
Syracuse, NY 13244-4500
1 (315) 443-5484
diekemar@syr.edu

Jiangping Chen
School of Library and Information Sciences
University of North Texas
P.O. Box 311068, Denton, TX 76203-1068
1 (940) 369-8393
jpchen@unt.edu

ABSTRACT
This paper describes exploratory research concerning the automatic assignment of educational standards to lesson plans. An information retrieval based solution was proposed, and the results of several experiments are discussed. Results suggest the optimal solution would be a recommender tool where catalogers receive suggestions from the system but humans make the final decision.

Categories and Subject Descriptors
H.3.7 [**Information Storage and Retrieval**]: Digital Libraries - *Standards*, - *Collection*.

General Terms
Algorithms, Experimentation.

Keywords
Natural Language Processing, Educational Standards, Metadata, Automatic Assignment, Experimentation, Evaluation.

1. INTRODUCTION
The No Child Left Behind Act of 2001 [1], has serious implications for the indexing and retrieval of educational materials because teachers are now required to use instructional materials that are aligned with relevant standards. To facilitate searching and retrieving of educational materials by state standards, they have to be indexed with standard information. Focus groups conducted by Devaul and Kelly [2] indicate that especially new teachers would like to be able to search by standards. However, while there is a strong movement to develop new educational resources that are in alignment with the competencies, there are vast repositories of educational resources suitable to address those competencies that have not originally been designed for that purpose. Manual classification or cataloging is knowledge-intensive and time-consuming work that requires the catalogers to possess the capability to correctly identify the category (i.e. standard) that a library material belongs to [3]. This research examines the feasibility of automatic assignment of standards to these vast repositories of un-mapped educational materials, specifically lesson plans as a gold standard for evaluation purposes was available.

JCDL'05, June 7–11, 2005, Denver, Colorado, USA.

2. AUTOMATIC ASSIGNMENT
As a research approach, we viewed the standard assignment task as essentially text categorization in which the educational resources are automatically assigned one or more of the identification numbers of the appropriate standards. Strategies considered for automatic classification include information retrieval and various statistical learning methods such as Bayes method, decision trees, K-nearest neighbor and support vector machine. [4]

We explored the feasibility of these different techniques and decided against using the statistical learning methods. The statistical methods derive rules or features for classifiers from pre-categorized data. These rules or features are then used for the categorization task. To be successful these methods require sufficient numbers of pre-categorized training materials for learning purposes. Unfortunately we only had 297 manually assigned science and technology standards for 573 different benchmark categories. With so many categories and so few manually tagged documents the statistical techniques were not feasible. Therefore we decided to pursue an information retrieval (IR) strategy. We treat each standard as a document, and each lesson plan or curriculum material as a query. A vector-space based information retrieval system indexes all the standards and then searches for the most relevant standards for each lesson plan. In matching the standards to the lesson plan, the standard assignment takes place.

3. INFORMATION RETRIEVAL FOR STANDARDS ASSIGNMENT

3.1 Creating Documents and Queries
The educational standards (our documents) we used in this research are the Compendium of Standards and Benchmarks [5]. From the Science Standards and Mathematics Standards, we automatically extracted standards at the Benchmark level. e.g. benchmark 8.1.1.1 is "Math; Problem Solving; Draws pictures to represent problems; pictoral respresentation." Since the benchmarks themselves tend to be extremely short, we (automatically) added information from the standard's higher level categories (for example algebra) and additional benchmark vocabulary to the benchmark document to create the "Expanded

Benchmarks". For example, our benchmark 8.1.1.1 gained the terms addition, subtraction, picture, representation, illustrate, illustration, depict, sketch, draw.

Lesson plan material was mined from the NSDL and Internet. Lesson plans (the queries) were then created by running a series of processes on the lesson plans (preprocessing, stemming, stopword removal, term filtering, NLP processing, term weighting, and term ranking). Using all or a selection of these processes allowed us to create a variety of queries for our experiments. For the experiments below, all queries were preprocessed, stemmed, stopword filtered, content filtered, NLP processed, weighted, and ranked.

3.2 Automatic Retrieval Evaluation

In order to test the system, standard assignment performance we put together a test collection for automatic evaluation. By using 297 math and science lesson plans that were manually tagged with benchmarks as a gold standard we could assess our system's assignment capabilities. Unlike traditional IR experiments where queries tend to have larger numbers of relevant documents (10-200), lesson plans in this evaluation only had a few relevant benchmarks assigned to it (75% <= 3). Since there are limited relevant documents and the assignment task requires great accuracy, these experiments merit a special evaluation metric: precision at rank 1.

4. EXPERIMENTS AND RESULTS

Table 1 displays the results of four different experiments. The experiments differ in the types of documents (either expanded or unexpanded benchmarks) and the number of query terms. Queries (lesson plans) with 20-25 terms were the most successful in retrieving relevant expanded documents (benchmarks). The system assigned the correct standards as the first pick 56% of the time. Although this precision score at rank 1 for the final experiment (0.5623) is very respectable from an information retrieval standpoint, especially when you consider how few relevant documents there are per benchmark. However this is not sufficient for automatic assignment and there is ample room for improvement.

Table 1. Automatic benchmark assignment experiments

Experiment ID*	N	Rel. in top 10	precision at rank 1	recall at rank 1
expanded, 15 terms	297	195	0.3468	0.1888
not expanded, 20 terms	297	222	0.3434	0.1915
expanded, 20 terms	297	262	0.5657	0.3152
expanded, 25 terms	297	265	0.5623	0.3158

There were a number of queries (lesson plans) that did not retrieve any relevant documents (benchmarks) so no automatic standard assignment could take place. In addition, there were certain queries (lesson plans) that only retrieved their standards at a rank other than the top 3.

5. ANALYSIS OF THE RESULTS

The main problem in trying to automatically map standards to lesson plans is the vocabulary gap between the two. In some extreme cases there is no overlap between the language in the lesson plan and the language in the standard assigned to that lesson plan. Standards and lesson plans have a different level of granularity. Standards are written in abstract or theoretical language while lesson plans are concrete examples of concepts presented in the standards.

Although our weighting scheme by definition assigns a level of term importance, it is often difficult to determine which terms in a lesson plan are important and which ones are not. Initially we had problems with terms describing materials to be used in class, such as markers, scissors, and even pumpkins. We remedied this by including a query filter to exclude irrelevant terms from the final query. We also expanded our general stopword list with terms that are unsuitable for retrieval purposes culled from the terminology of the lesson plans.

Finally we encountered problems with the training data (standards with manually assigned standards) used to fine-tune our system because not all standard assignments seemed appropriate.

6. CONCLUSIONS AND FUTURE RESEARCH

We successfully developed a middleware tool for the automatic assignment of content standards to benchmarks. However, the task of automatically mapping standards to lesson plans is inherently difficult due to the vocabulary gap between the two objects. Although our system assigns the correct standards as the first pick in 56% of the cases, there is much room for improvement. At this stage of the development we recommend the tool be used as a recommender-type system rather than a fully automatic assignment system. In a recommender capacity, the system could provide suggestions to catalogers as to what standards may be most appropriate for certain benchmarks, while the system can learn from the vetted assignments. This research is being continued at Center for Natural Language Processing.

7. ACKNOWLEDGMENTS

This work was funded through the NSF NSDL Program.

8. REFERENCES

[1] 107th Congress. *No Child Left Behind Act of 2001*. http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf. (H.R.1, Public Law No: 107-110, Section 1001 Statement of Purpose).

[2] Devaul, H. and Kelly, K. (2003). Searching by Educational Standards in DLESE: What does it mean and what do users want? http://nsdl.comm.nsdl.org/meeting/poster_docs/2003/752_dlese_edStandards.pdf

[3] Taylor, A. G. *Wynar's Introduction to Cataloging and Classification. 9th edition*. Libraries Unlimited, Englewood, CO., 2000.

[4] Yang, Y. An evaluation of statistical approaches to text categorization. Information Retrieval 1, 1 (1999), 69-90

[5] McREL Compendium of Standards and Benchmarks. http://www.mcrel.org/standards-benchmarks/.