CAIS Paper: Ontology-based Indexing Technologies in Information Retrieval: Building a Topic Map (ISO 13250) for a Mathematics Education Database

Fei Shu (McGill University), and Charles-Antoine Julie (McGill University)

Abstract: This paper describes a project that has created a Topic Map search tool for a mathematics educational database containing articles from the journal *For the Learning of Mathematics*. The resulting website enables users to retrieve research articles based on a variety of topics such as mathematics classification, research methods, educational objectives, in addition to traditional bibliographic information.
Résumé:

1. Problem

Keyword-based search (e.g., Google) is the most popular tool for web searching and information retrieval; however, keyword-based searching also produces many irrelevant results because keywords can have multiple meanings or they often inadequately express users' intent (Michael Raj and Rajesh 2011). In order to find accurate information, users waste time browsing long lists of often-irrelevant results. In addition, keyword-based search does not provide adequate support to users but strictly returns the documents whose vocabulary matches the keyword terms (Wilson, Schraefel, and White 2009); therefore, it fails to recognize relevant documents that do not match the query (Mann 2008).

2. Ontology and Topic Maps

Some previous research tries to improve information searching by optimizing information organization. Ontology-based indexing technologies are a promising approach (Patkar 2011). By establishing an ontology on the basis of semantic relationships between concepts, the improved information organization may improve users' searching performance (Jimeno-Yepes, Berlanga-Llavori, and Rebholz-Schuhmann 2010, Yi 2008). Building ontologies can be done using Topic Maps, which is an international standard (ISO13250) for knowledge representation and exchange. According to Pepper (2000), Topic Maps represent information concepts and their relationships using the following elements:

- Topics: represent any concept, from people, countries, and organizations to software modules, individual files, and events;
- Associations: representing relationships between topics;
- Occurrences: representing information resources relevant to a particular topic.

Compared with the keyword searching, a Topic Map aims to reduce the gap between desired results (i.e., more relevant information) and the large result sets containing much irrelevant information returned by traditional keyword search tools (Kwong and Ng 2003, Yi 2008). A topic map represents knowledge through topics, associations between topics, and occurrences, which are locators pointing to information resources related to the given topics (Pepper 2000). A topic map can facilitate information discovery and retrieval

because it groups results by topics that users can visually navigate without inputting keywords (Garshol 2004). Topic Map users can retrieve information associated with a topic and discover information associated with previously unknown but related topics (Melgar Estrada 2011). Assuming a topic map application that is integrated with a keyword search engine, a student can filter the returned articles by topic, choose a topic, navigate the topic map, access other related topics via associations, and retrieve information without conducting additional searches.

This paper describes the design of a Topic Map search tool for a mathematics educational database containing articles from the journal *For the Learning of Mathematics*. The project aims to enable novice and expert users to retrieve scholarly articles based on a variety of topics such as mathematics classification, research methods, and educational objectives, in addition to traditional bibliographic information.

3. Previous Works

Topic Maps are used as an ontology-based search tool in various domains. Based on measures of recall and search time, an experimental study using 40 participants preforming information retrieval tasks showed that a Topic Maps-based ontology information retrieval system had a significant and positive effect on both recall and search time (Yi 2008). Topic Maps have been used to generate learning materials for Chinese herb medication (Shih, Shih, and Chen 2007), and they have been applied to language teaching (Urbaniak and Venkatesh 2012) and aboriginal language preservation (Pelczer et al. 2012) to represent grammatical and task-based structures.

Topic Maps are also used in E-learning applications. For example, they support the location of appropriate learning resources for a specific student in a given context to ensure effective learning (Kolås 2006), and Dicheva and Dichev (2006) describe the TM4L environment that enables the creation, maintenance and use of ontology-aware online learning repositories based on the Topic Map standard. Widhalm and Mueck (2003) suggest that Topic Maps may be used instead of SCORM/LOM for the characterization of E-learning resources, and Lalingkar and Ramani (2010) developed a Study Facilitation System (SFS) based on Topic Maps.

4. Methodology

This project requires a domain appropriate subject structure (i.e., topics and their relationships), the assignment of these topics to a collection of articles from the domain, and the development and implementation of a live Web site that allows browsing and searching of the collection based on the subject structure. The pilot test was also conducted by interviewing 5 users regarding their experience and satisfaction using the novel Topic Map search tool.

4.1 Creating a Subject Structure

Traditional academic databases (e.g. ERIC) allow users to search and group results by subject heading or descriptor. However, these subject headings and descriptors are pre-designated based on the articles' content, meaning that users cannot modify them according to their specific needs (e.g. neither the subject "Discrete Mathematics" nor the subject "Combinatorics" exists in ERIC). In this case, users would like to group the searching results not only by the subject discussed in the content but also by metadata not

currently available such as the research method or the educational objective applied by the studies. Therefore, a new subject structure was created based on users' requirement. The Topic Map standard was chosen as an appropriate tool for the creation of this subject structure.

Users' needs and requirements were collected by consulting subject-matter experts and conducting a needs assessment. The Mathematics Subject Classification (MSC) 2010 served as the basis for the subject structure topic design, which was modified according to users' requests. For example, new topics were defined (e.g., Academic Level, Research Method, Educational Objective) as well as their associations. Table 1 shows examples of topics, association and occurrence types created for the collection:

Topic	
Article	Academic Level
Author	Pre-School
Topic (Mathematics Classification)	Elementary
Discrete Mathematics	Secondary
Applied Mathematics	College
Mathematics Theory & Philosophy	University
History of Mathematics	General
Pure Mathematics	Article Type
Algebra	Opinions & Discussion
Arithmetic	Research Paper
Calculus	Peer Review
Combinatorics	Project Report
Function	Literature Review
Geometry	Educational Objective
Infinity	Teaching
Mathematics Analysis	Curriculum
Numbers	Pedagogy
Topology	Educator
Trigonometry	Learning
Research Methods	Learning Objective
Quantitative Research	Learning Environment
Qualitative Research	Student
Mixed Methods Design	History of Education
Other Methods	Instruction
Association	**Occurrence**
Write/Written by	Volume No.
Discuss/Discussed by	Issue No.
Applied to	Publish Time
Related to	Page
Classified as	Content
Use/Used by	

Table 1 List of Topics, Associations and Occurrences

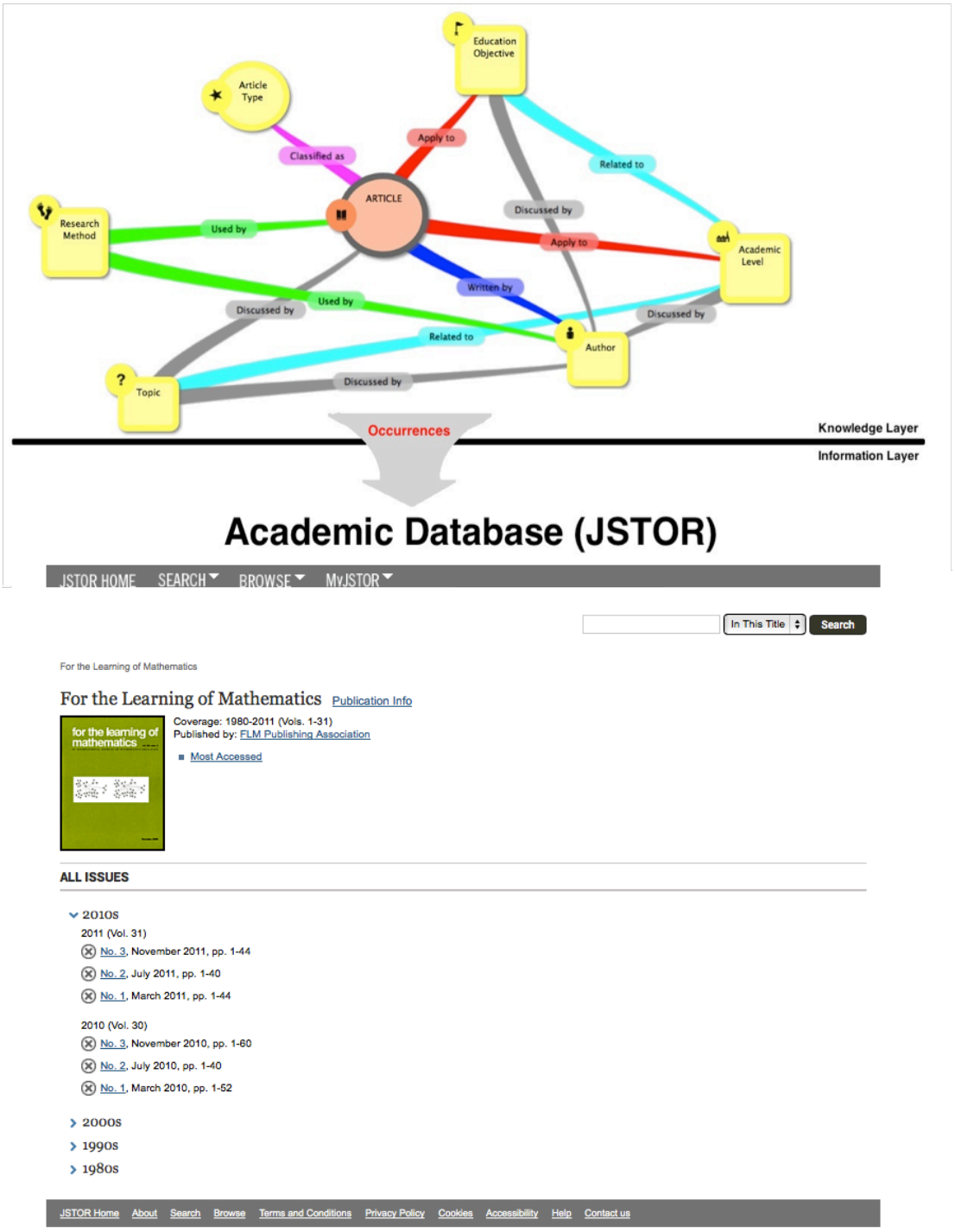

Figure 1 The 2-layer Topic Map model

Figure 1 shows that Topic Maps consist of two layers: the information layer contains a set of information resources (i.e., the database), and the knowledge layer contains a knowledge map consisting of topics and their associations. These two layers are connected by occurrences or individual documents. The 2-layer model separates the topics from the resources that demonstrate these topics, and it establishes the relationships between topics without accessing the database of resources (Venkatesh 2007). Therefore, search results can be efficiently filtered or grouped by topic and explored using the relationships between topics. Compared to traditional searching

applications that strictly retrieve documents matching the user specified keywords, the Topic Map search tool allows users to navigate from topic to topic without entering additional keywords.

4.2 Web Site Development

Ontopia was selected as a Web site development framework. It contains a set of tools whose engine stores and maintains the topic maps. Ontopia also includes Ontopoly and Omnigator, both used to design and display the Topic Maps. The current version provides access to 108 articles; this is a small collection used for the initial prototype and preliminary user testing.

As shown in Figure 2, Omnigator allows users to begin a search by visually exploring topics without entering keywords. When a user chooses a topic, results (e.g., articles) related to this topic are returned by the database. Since the results are grouped by topic, users can find the results related to the same topic quickly and accurately while wasting less time filtering irrelevant results. When users find a relevant article, they can either retrieve the article using an external link, or consult the article's metadata stored with the Topic Map. Users can also inspect related topics to retrieve articles assigned to other topics without restarting the search process. For example, as shown in Figure 2, assuming the user has found an article entitled "*Which Operations? Certainly not Division!*" related topics are listed in the associations (e.g., Applied to: elementary).

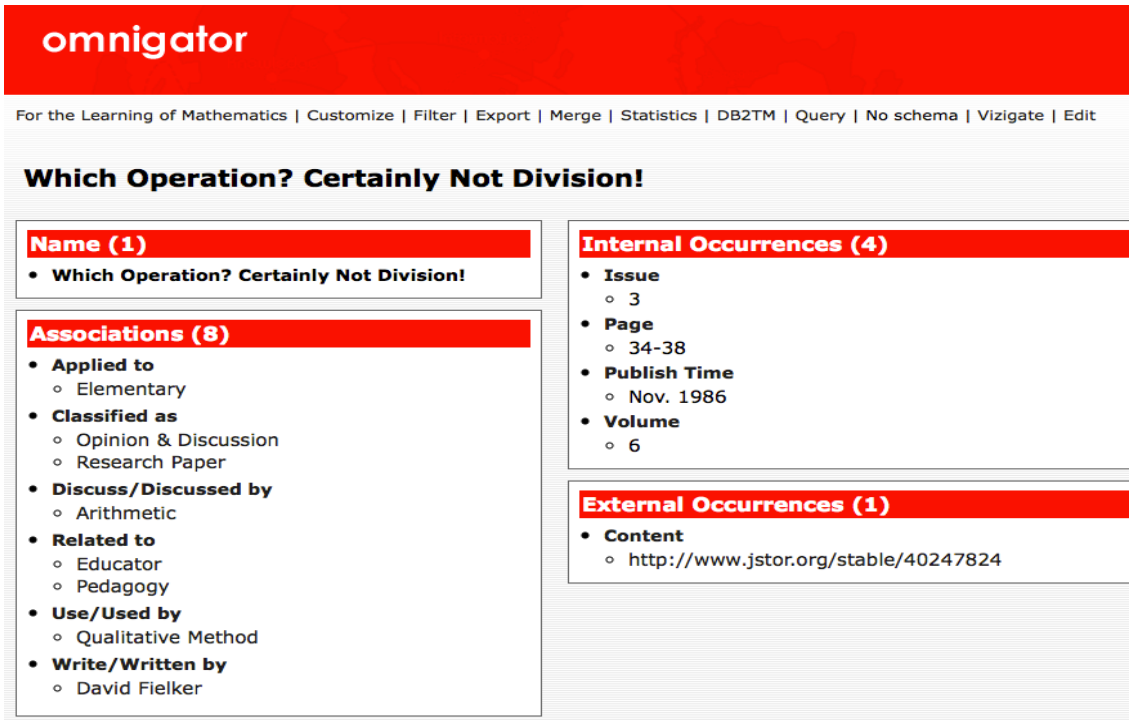

Figure 2 Screenshot of the Omnigator – Article

If the user desires to read more articles concerning Arithmetic, she may click on "Arithmetic" under its "Discuss/Discussed by" association and the system returns all articles assigned to "Arithmetic" as shown in Figure 3.

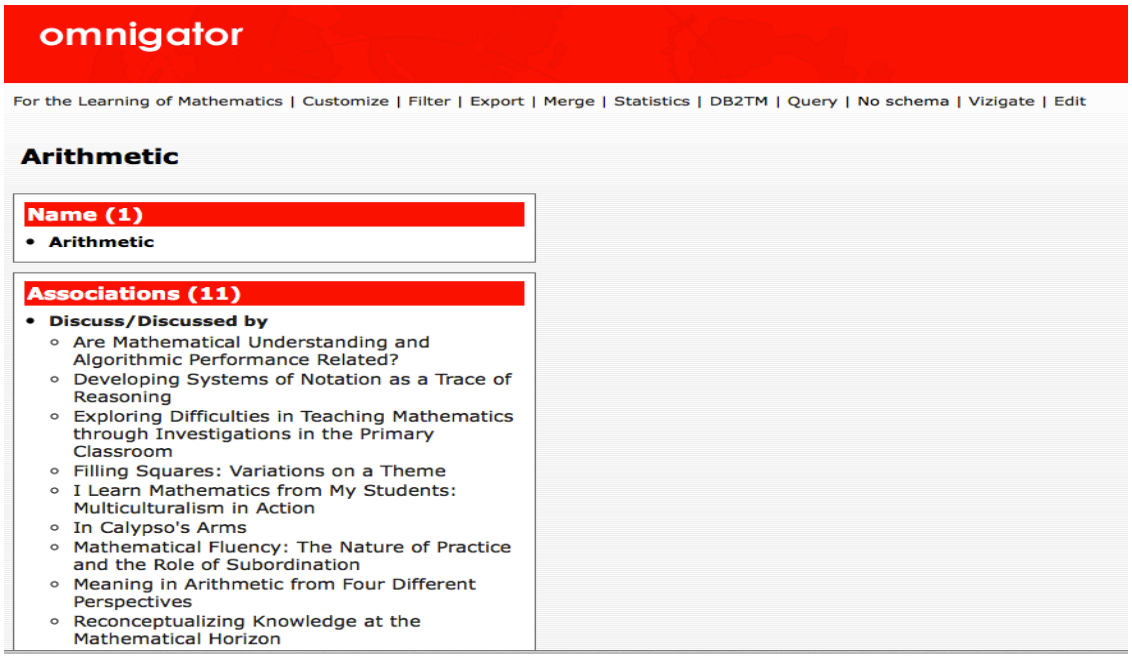

Figure 3 Screenshot of the Omnigator – Arithmetic

4.3 Pilot Test

A pilot test was conducted to collect users' experience with the system. This was not a task-based evaluation but a user-centered evaluation placing emphasis on users' satisfaction. Since the system is designed for retrieving scholarly mathematics articles, most users are graduate students or scholars in mathematics or information science. According to this user population, two participants were selected from School of Information Studies at McGill University and three participants were selected from Department of Mathematics at Concordia University. The participants were selected by means of a purposive sample using the snowball recruiting method to identify the participants. Participants were recommended by friends and informed the purpose of my research and the method of data collection.

Participants were not required to complete any specific tasks or queries but to share their experience in using this new Topic Maps search tool. They were asked to read a document that described the purpose of the test and brief usage instructions for Omnigator. They were asked to use Omnigator to find mathematics articles of interest. To ensure an adequate familiarization with the system, they were required to spend at least two hours using the system in one week. Afterwards, they would share their stories and express their feelings concerning the system by answering four open-end questions:

- Do you like this novel Topic Maps search tool (Ontopia) in general?
- Does Ontopia provide satisfactory search results in terms of accuracy as compared to searching directly in the academic database?
- Does Ontopia provide a satisfactory performance in terms of search time compared to searching directly in the academic database?
- Are you satisfied with Ontopia in terms of its layout, interface and functionality?

5. Results

All five participants stated they liked the idea of searching for articles by topics using the novel Topic Maps search tool (Ontopia). They believed that Ontopia had positive effects on search accuracy because it filtered the articles that did not concern their chosen topic

even when these articles matched the keyword search term(s). All five participants declared that Ontopia saved time because they could navigate from topic to topic, or from article to article, without typing keywords or accessing the database of articles. Four out of five participants were satisfied with Ontopia in terms of its layout, interface and functionality. One participant complained that he could not search through the full-text of the articles since Ontopia offers a full text search restricted to metadata. This could be a weakness of the system since a Topic Map integrated with full-text keyword searching may improve users' searching performance by offering both search techniques in conjunction.

6. Conclusion

This ongoing project illustrates the potential of Topic Map search tools, and ontologies in general, to retrieve articles from an academic database, and improve our understanding of the possible interactions between users and Topic Maps. A Topic Map integrated with a traditional keyword search engine may facilitate searching by grouping the search results by topics, and improve users' searching performance by allowing them to navigate from topic to topic without inputting keywords; thus, allowing users to retrieve more accurate information in less time.

REFERENCE

Dicheva, Darina , and Christo Dichev. 2006. "TM4L: Creating and Browsing Educational Topic Maps." British Journal of Educational Technology no. 37 (3):391-404.

Garshol, Lars. 2004. "Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all." Journal of Information Science no. 30 (4):378-391.

Jimeno-Yepes, Antonio, Rafael Berlanga-Llavori, and Dietrich Rebholz-Schuhmann. 2010. "Ontology Refinement for Improved Information Retrieval." Information Processing and Management no. 46 (4):426-435.

Kolås, Line. 2006. Topic maps in e-learning: An ontology ensuring an active student role as producer. Paper read at World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Kwong, Linus W., and Yiu-Kai Ng. 2003. "Performing Binary-Categorization on Multiple-Record Web Documents Using Information Retrieval Models and Application Ontologies." World Wide Web: Internet and Web Information Systems no. 6 (3):281-303.

Lalingkar, Aparna, and Srinivasan Ramani. 2010. A Topic Map-based System for Identifying Relevant Learning Objects. Paper read at World Conference on Educational Multimedia, Hypermedia and Telecommunications.

Mann, T. 2008. "Will Google's keyword searching eliminate the need for LC cataloging and Classification?" Journal of Library Metadata no. 8 (2):159-168.

Melgar Estrada, Liliana M. 2011. "Topic Maps from a Knowledge Organization Perspective." Knowledge Organization no. 38 (1):43-61.

Michael Raj, T. F., and K Rajesh. 2011. "A Comparitive study: FLOWS and PSL model in Selecting the Ontologies for Dynamic Web service Selection in Semantic Web Environment." International Journal on Computer Science & Engineering.

Patkar, Vivek. 2011. "A Passage to Ontology Tool for Information Organisation in the Digital Age." DESIDOC Journal of Library & Information Technology no. 31 (2):90-102.

Pelczer, Ildiko, Michael Conor Cook, Vivek Venkatesh, Elizabeth Gatbonton, and Norman Segalowitz. 2012. Building an Ontology for Inuttitut Language Samples: An Application of Topic Maps in Canadian Aboriginal Language Preservation. Paper read at World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Pepper, Steve. 2000. The TAO of topic maps. Paper read at Proceedings of XML Europe.

Shih, B. J. , J. L. Shih, and R. L. Chen. 2007. "Organizing learning materials through hierarchical topic maps: an illustration through Chinese herb medication." JCAL Journal of Computer Assisted Learning no. 23 (6):477-490.

Urbaniak, Kathryn, and Vivek Venkatesh. 2012. Building an Ontology for Student Models of English As Second Language Essays: A Synthesis of Literature and Results of

Development of Topic Map Indexes. Paper read at World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Venkatesh, Vivek. 2007. Theoretical and Practical Implications of Using Topic Map Technologies in E-Learning Applications. Paper read at World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Widhalm, Richard, and Thomas Mueck. 2003. Using Topic Maps for eLearning. Paper read at World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Wilson, Max L., M. C. Schraefel, and Ryen W. White. 2009. "Evaluating Advanced Search Interfaces Using Established Information-Seeking Models." Journal of the American Society for Information Science and Technology no. 60 (7):1407-1422.

Yi, M. 2008. "Information organization and retrieval using a topic maps-based ontology: Results of a task-based evaluation." J. Am. Soc. Inf. Sci. Technol. Journal of the American Society for Information Science and Technology no. 59 (12):1898-1911.