

# Element Matching in Concept Maps

Byron Marshall, Therani Madhusudan

University of Arizona, Department of Management Information Systems, Artificial Intelligence Lab  
McClelland Hall, The University of Arizona, Tucson, Arizona 85721-0108  
byronm@eller.arizona.edu

## ABSTRACT

Concept maps (CM) are informal, semantic, node-link conceptual graphs used to represent knowledge in a variety of applications. Algorithms that compare concept maps would be useful in supporting educational processes and in leveraging indexed digital collections of concept maps. Map comparison begins with element matching and faces computational challenges arising from vocabulary overlap, informality, and organizational variation. Our implementation of an adapted similarity flooding algorithm improves matching of CM knowledge elements over a simple string matching approach.

## Categories and Subject Descriptors

K.3.1 [Computer Uses in Education]: Computer-assisted instruction

## General Terms

Algorithms, Human Factors

## Keywords

Concept Mapping, Education, Knowledge Representation

## 1. INTRODUCTION

Concept maps (CM) are informal, semantic, node-link conceptual graphs used to represent knowledge in a variety of applications. Digital libraries containing CMs maps could represent textual resources semantically and capture multiple personal perspectives. Student concept mapping is a theoretically sound and empirically verified educational technique, but map evaluation is a time-consuming manual process. Algorithms that compare CMs would be useful in supporting educational processes and in leveraging indexed digital map collections. Student and instructor maps can be compared to provide instructional feedback, maps in a collection can be compared to provide context for user queries, and maps must be compared before they can be merged into combined knowledge structures. Map comparison begins by matching individual knowledge elements between map pairs. Knowledge elements include concept nodes, concept-link-concept propositions, hierarchical clusters, and other map sub-structures.

Concept mapping is an effective educational technique. [3] reviews evidence that CMs can be the basis of effective study and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL '04, June 7–11, 2004, Tucson, Arizona, USA.

Copyright 2004 ACM 1-58113-832-6/04/0006...\$5.00.

learning strategies: (1) in cooperative interactions, (2) as pre- and post-study aids, (3) as a substitute for traditional text, and (4) for updating and editing knowledge. Several computerized concept mapping systems have been deployed [1, 6].

Concept mapping was developed to support education based on a constructivist notion of meaningful learning [8]. In creating a concept map a learner identifies key concepts and relations. Meaningful learning processes exhibited in the organization of student concept maps include: (1) new concept learning in the node names, (2) subsumption in specific items connected to more general previous knowledge, (3) progressive differentiation in multiple hierarchical layers, and (4) integrative reconciliation in cross-links connecting separate branches of a hierarchical tree. At least 128 different methods of creating and scoring CMs have been proposed [9] but existing computerized CM-systems do not include algorithms to support student assessment. A system that counts levels of hierarchy, matches corresponding hierarchical clusters, and automatically identifies cross-links would be useful for concept map evaluation.

Specially annotated CMs called Attributed Concept Maps were used in [2] to evaluate student achievement. Students drew maps using a closed list of nodes and links. Overall structural similarity between “master” and student maps was shown to correlate with scores on other tests of student knowledge. Correlation was sharply reduced when students chose equally correct but different organizational structures. It was suggested that the system should be enhanced to match concept-link-concept propositions.

## 2. COMPUTATIONAL CHALLENGES

The GetSmart system combines concept mapping, curriculum, and digital library tools [6]. Use of the system in the fall of 2002 generated a collection of more than 1,400 concept maps, drawn by 114 students, on a variety of topics from two computer-science related graduate courses at Virginia Tech and the University of Arizona. Ongoing analysis of this map collection has helped us identify several key challenges facing computerized concept map evaluation systems including term variation, informality, and organizational variation.

We compared different labels used for the same concept in several sets of maps from the GetSmart collection. For example the information retrieval term *E Measure* was entered as *E-Measure*, *emeasure*, *E Measure*, *E Measrue*, and *E evaluation measure*. Allowing for minor spelling adjustments, we found that the most common word for a concept appeared in about 75% of the maps. About half of the remaining labels were quite similar and the rest were not very similar at all. Using a measure from [4] this equates to between 50% and 60% vocabulary overlap. Thus effective CM element matching must allow for term variation.

There is a dichotomy between a human user’s need for a flexible and forgiving (informal) CM system and the computer’s need for a (formal) system with strong semantics [5]. CMs are

characteristically informal; they are therefore difficult to manipulate but useful for knowledge acquisition [1]. Effective CM element matching routines cannot rely on rigid formalism.

Students use many different organizational structures when representing the same set of concepts. Based on a review of the maps in our extensive collection, we have identified several commonly occurring patterns. An effective CM element matching routine will need to be robust over these patterns.

- *Added/missing elements*: Some students insert extra nodes between related concepts.
- *Organizational variations*: The same clusters of nodes are often represented by one student as a major concept connected to a set of subsidiary concepts while another connects those same concepts in a linear fashion. Node clusters are also frequently moved to different positions in a hierarchical tree.
- *Cross-links*: Students frequently provide different cross-links between branches of hierarchical clusters. These variations may reflect personal differences related to integrative reconciliation.
- *Granularity and cardinality differences*: A single node in one map may correctly match to two or more nodes in another (a difference of cardinality) i.e. *b/b+ tree* to *btree* and *b+tree*. An example of a granularity variation occurs when a node-link-node structure should be correctly matched to a larger structure like node-link-node-link-node.

### 3. ELEMENT MATCHING

We implemented an element matching system appropriate for concept maps. It begins with a simple string matching routine that assigns a similarity value to nodes and propositions in one map to the nodes and propositions in another. The string matching algorithm addresses some of the term variation found in our concept map collection. We compared a master map to 30 topically-similar, student-drawn maps found in the GetSmart collection. Matches made by human experts were compared to the matches found by a simple string match algorithm. The algorithm correctly identified 88% of the node matches and 50% of the concept-link-concept matches. It is unlikely that this level of accuracy would be very helpful to an instructor in providing feedback to students. In addition, this approach does not allow for matches between structures with different levels of granularity.

To improve on the initial string match results we implemented a variation of the similarity flooding algorithm (SF) presented in [7]. SF leverages structural information in matching elements in models or schemas. The algorithm operates on the assumption that whenever two elements, one from Graph1 and one from Graph2 are found to be similar, the similarity score of adjacent elements should be increased. Over a number of iterations, the initial similarity of any two nodes propagates through the graphs. SF is appropriate for this task because it can be adapted to support matching of sub-structures with different granularities or cardinalities and it does not rely on strict formalism in the underlying graphs. Our implementation of this algorithm significantly improved on the string match results. Matches were correct for 94% of the nodes and 79% of the concept-link-concept propositions. A review of the results suggests that they can be improved if an instructor encodes key term indicators and a few contextually appropriate synonyms in the master map. No special annotations are applied to the student maps. The SF approach allows matching at different granularity levels although our current system only partially implements this capability.

## 4. CONCLUSIONS

A matching system that uses structural information in addition to node and link labels is promising for matching knowledge elements in collections of concept maps. Element matching can be used to support identification of hierarchical sub-structures and cross-links in student maps. Semi-automatic routines employing this kind of matching could be used to help educators efficiently provide feedback on student maps using a semi-automatic CM element matching and scoring system. The same kind of matching routines would also be useful in processing contextually rich CM-based queries and in managing collections of informal, personalized, semantic node-link graphs representing DL resources. This would enhance the use of digital libraries for educational processes built around concept mapping.

## 5. ACKNOWLEDGMENTS

This research was partially supported by NSF Award No. DUE-0121741, Program 7444. September, 2001 - August, 2003. We thank and recognize Hsinchun Chen, other members of the U of A's AI Lab as well as Ed Fox, Rao Shen, and Lillian Cassel who used and evaluated GetSmart, helping guide its development.

## 6. REFERENCES

- [1] A. J. Canas, D. B. Leake, and D. C. Wilson, "Managing, mapping and manipulating conceptual knowledge: exploring the synergies of knowledge management & case-based reasoning," presented at AAAI Workshop on Exploring Synergies of Knowledge Management and Case-based Reasoning, Menlo Park, CA, 1999.
- [2] S.-W. Chen, S. C. Lin, and K. E. Chang, "Attributed concept maps: fuzzy integration and fuzzy matching," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 31, pp. 842-852, 2001.
- [3] T. L. Chmielewski and D. F. Dansereau, "Enhancing the recall of text: Knowledge mapping training promotes implicit transfer," *Journal of Educational Psychology*, vol. 90, pp. 407-413, 1998.
- [4] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The Vocabulary Problem in Human-System Communication," *Communications of the ACM*, vol. 30, pp. 964--971, 1987.
- [5] R. Kremer, "Concept mapping: informal to formal," presented at Proceedings of the International Conference on Conceptual Structures, University of Maryland, 1994.
- [6] B. Marshall, Y. Zhang, H. Chen, A. Lally, R. Shen, E. Fox, and L. Cassel, "Convergence of Knowledge Management and E-Learning: the GetSmart Experience," presented at Joint Conference on Digital Libraries, Houston, Texas, 2003.
- [7] S. Melnik, H. Garcia-Molina, and E. Rahm, "Similarity flooding: a versatile graph matching algorithm and its application to schema matching," presented at Proceedings of the 18th International Conference on Data Engineering (ICDE '02), San Jose, Ca, 2002.
- [8] J. Novak and D. B. Gowin, *Learning How To Learn*. Cambridge, UK: Cambridge University Press, 1984.
- [9] J. A. Rye and P. A. Rubba, "Scoring Concept Maps: an Expert Map-Based Scheme for Relationships," *School and Science Mathematics*, vol. 102, pp. 33-44, 2002.