Knowledge Maps for e-Learning

Jae Hwa Lee and Aviv Segev

Department of Knowledge Service Engineering, KAIST, Korea

Corresponding Author: aviv@kaist.edu

Abstract

Maps such as concept maps and knowledge maps are often used as learning materials. These maps have nodes and links, nodes as key concepts and links as relationships between key concepts. From a map, the user can recognize the important concepts and the relationships between them. To build concept or knowledge maps, domain experts are needed. Therefore, since these experts are hard to obtain, the cost of map creation is high. In this study, an attempt was made to automatically build a domain knowledge map for e-learning using text mining techniques. From a set of documents about a specific topic, keywords are extracted using the TF/IDF algorithm. A domain knowledge map (K-map) is based on ranking pairs of keywords according to the number of appearances in a sentence and the number of words in a sentence. K-map does not label links; instead K-map shows all sentences containing the two keywords placed at both ends of the relation chosen. Therefore, K-maps show promise as a tool for e-learning environments.

Keywords: Knowledge maps, e-learning, concept relation

1. Introduction

When people learn from textual material, they usually follow the order set by the author, as with reading books. Although this is the most common method of text-based learning, it is not efficient in the following three situations. First, in many cases, people have different levels of prior domain knowledge. However, when they learn from text material, they can only read it from the beginning to the end or use the table of contents to jump directly to a specific chapter. If a learner has a certain level of knowledge, then he can use the index to look for the information on a certain concept. However, the index usually contains hundreds of concepts listed in alphabetical order with no relational information between them. Second, in cases of learning under time pressure the learner can use the table of contents or index to identify the main parts. There is no other specific way to distinguish important information. Furthermore, if a person wants to learn about a domain from a web search, then the time limitation is more critical due to the huge number of documents on the web. He will read documents from top-ranked ones down and then will stop reading when the time is up. Third, if a document is complex or long, then a reader may find it difficult to recognize the important concepts and the relationships among them.

Concept maps or knowledge maps can be useful in these situations and thus can improve the e-learning experience. Key concepts and relationships can be recognized directly from the map, so learners can identify them with minimum effort. Time can also be saved when the amount of text is shortened.

However, the construction of concept maps and knowledge maps requires manual effort of domain experts. In this paper, a method for automatic generation of maps is proposed and an example of implementation with real-world data is presented.

2. Related Work

2.1 Concept Map

A concept map is a visuospatial representation of knowledge with text and graphical elements such as arrows, lines, ovals, and squares. It consists of nodes, containing a concept or item, and links connecting two nodes to each other and describing their relationship, where each node-link relation makes a proposition.

Concept maps are theoretically grounded in cognitive learning theory, in Ausubel's theories of assimilation and subsumption (Ausubel, 1968). Novak initially developed concept mapping tools to search for better ways to represent the learners' knowledge (Novak & Musonda, 1991).

Since then, concept maps have been used in various areas as assessment tools (McClure et al., 1999), cooperative learning tools (Stoyanova & Kommers, 2002), anxiety reduction tools (Czerniak, 1998), and tools that increase efficiency of search engines (Carvalho et al., 2001). Computer-based environments for concept mapping have been developed, such as the dynamic concept map proposed by Nesbit and Adesope (2005), in which audio presentation is synchronized with a node-link image. Cañas et al. (2005) developed CmapTools, tools that enable concept maps to be combined with multimedia resources.

2.2 Knowledge Map

Holley & Dansereau (1984) first developed knowledge maps, and Wiegmann et al. (1992) added types of links, which include type, characteristic, part, results in, leads to, and example. Knowledge maps are very similar to concepts maps, in terms of structures, goals, benefits, and areas of application; however, knowledge maps have specific types of links. A knowledge map is also a two-dimensional graphical display that presents information, but the method of relationship designation in knowledge-maps distinguishes the technique from other types of maps such as concept maps (Amer, 1994). O'Donnell et al. (2002) summarized findings on the field of knowledge maps over 12 years: students who use knowledge maps recall more central ideas than do students who use texts (Hall et al., 1992); students with low verbal ability or low prior knowledge benefit the most from the presentation of information in a knowledge map format with recall (Rewey et al., 1989); students who use knowledge maps as supports when interacting with peers in cooperative learning environments learn more effectively; and information is recalled better when presented in well-structured maps designed according to Gestalt principles than when presented in less well-structured maps (Wiegmann et al., 1992). Also, knowledge maps were found to help general purpose conceptualization processes (Gomez et al., 2000) and to help users reduce their anxiety and increase motivation (Hall & O'Donnell, 1996).

2.3 Topic Map

A topic map is an ISO standard for describing knowledge structures and associating them with their resources. A topic map contains basic concepts, such as Topics, Associations, and Occurrences (Pepper, 2000). A topic map also consists of topics and relationships between them. If the word 'topic' is generalized to 'word', topic maps are more or less similar to concept maps, except that topic maps are more focused on standards. Böhm et al. (2002) introduced a way to construct an initial set of topic maps or extend/optimize a given map using text mining technology. Dicheva & Dichev (2006) developed an environment for e-learning, called TM4L, where people use topic maps for learning.

2.4 Automatic Construction of Maps

In recent years interest in automatically building concept maps has grown. Chen and Xia (2009) reviewed researches about automatic construction of concept maps. They presented a traditional concept map construction method and auto-construction concept map methods. Chen et al. (2008) proposed a way to construct concept maps automatically from academic papers. They used author keywords as keywords after pre-processing and defined relations among them with four assumptions. Tseng et al. (2007) proposed a Two-Phase Concept Map Construction (TP-CMC) approach to automatically construct a concept map for a course from the learner's past test data. The first phase uses Fuzzy Set Theory to transform the numeric data into symbolic ones. The second phase uses multiple rule types to analyze the mined association rules and also uses a heuristic algorithm to remove redundancy and circularity when building concept maps. Relations in the concept map indicate learning paths. Urfat & Korhan (2009) introduced a way to extract key concepts and relations among learning concepts. To extract key concepts, n-gram, a sequence of characters that stands for a word, was used with a dictionary of technical terms. To extract relations (relevance) from sentences, three features were considered: word co-occurrences, logarithmic concept weights, and augmented normalized candidate learning concept frequency.

Previous work focused on automatic construction of maps and on construction of maps for learning based on user past learning experience. This paper proposes a method of automatic map generation from unexplored material for e-learning purposes and presents an implementation example with real-world data.

3. Model

3.1 Overview

K-map is a knowledge map that has nodes and links. Nodes are keywords that are considered important concepts for a specific domain, and links are relations between two keywords. The system uses several documents related to a certain domain to generate a K-map. Figure 1 is an example of the K-map tools screen shot automatically generated from a set of documents about John F. Kennedy. The map has 18 keywords and 30 relations.

Links have different thicknesses. Each link has a different score and thickness, when a higher score yields a thicker link. K-map tools serves as a K-map learning environment.

K-map has a hyperlink for each relation. If the user clicks a relation, he will see all the sentences that have the two keywords at both ends of the relation. For example, if a relation between ‘Kennedy’ and ‘president’ is chosen, Figure 2 appears. In Figure 2, all sentences that have the words ‘Kennedy’ and ‘president’ are displayed. The user can read some of the sentences and acquire domain knowledge from them. If the user chooses any sentence, he can get a direct access to the original document (Figure 3).

With K-map tools, the user can control the number of keywords and relations. Keywords and relations will appear and disappear according to their ranking and the limit that the user sets. This function assists in reducing Map shock (Dansereau et al., 1994) since the user can resize the map freely. There is a concept search window that helps the user search for a specific concept. When a relation has too many sentences, a user can be overwhelmed and lose interest. In this case, the words extracted from the set of sentences are believed to represent the relation. There are five extracted words for the relationship between ‘Kennedy’ and ‘Vietnam’ (Figure 4). If the user chooses one of those words, K-map Tools will show the sentences that have ‘Kennedy’, ‘Vietnam’, and the word he chose. In Figure 5, the word is ‘withdraw’. These words categorize the initial sentences.

With K-map Tools, as the user handles the K-map he accomplishes two tasks: he searches and he learns. In current search engines two steps occur in the searching and learning process. The first step is the user typing in the words he wants to know about; those words can be considered the topic. Then the search engine shows the user a list of the documents related to the query. The user goes over the list and at some point chooses a document and starts reading it. He repeats the same process until he learns enough about the topic. With current search engines, there may be inefficiencies – the user may not understand important concepts, may miss key concepts, or may waste his time trying to find a right document.

However, with K-map the user doesn’t have to find a right document, because the map already has extracted sentences, categorized by keyword pairs. He can recognize the key concepts and the strongly connected key concepts; he obtains a holistic view of the domain. Thus, K-map improves the searching and learning process.

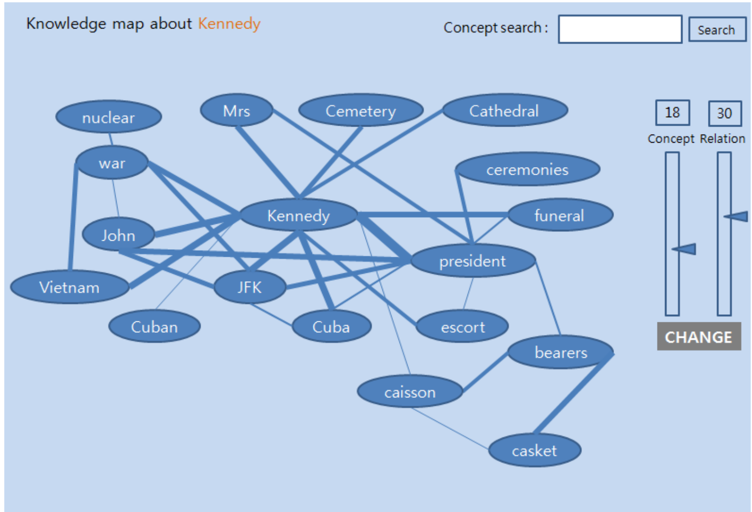

Figure 1. Screen Shot of K-map Tools - Knowledge Map of John F. Kennedy

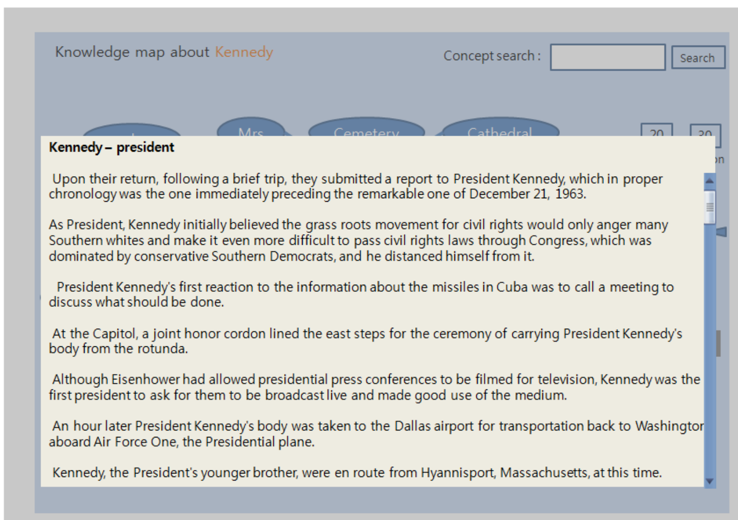

Figure 2. Screen Shot of Sentences Containing ‘Kennedy’ and ‘President’ in K-map Tools

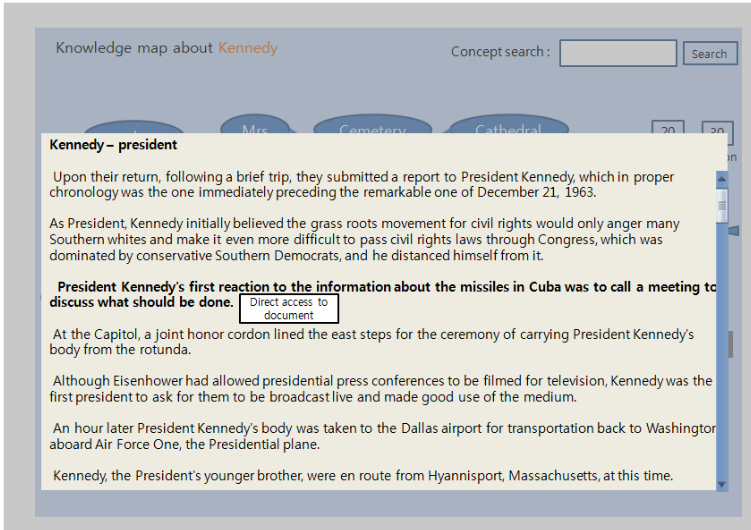

Figure 3. Screen Shot of K-map Tool Direct Access to a Document through a Sentence

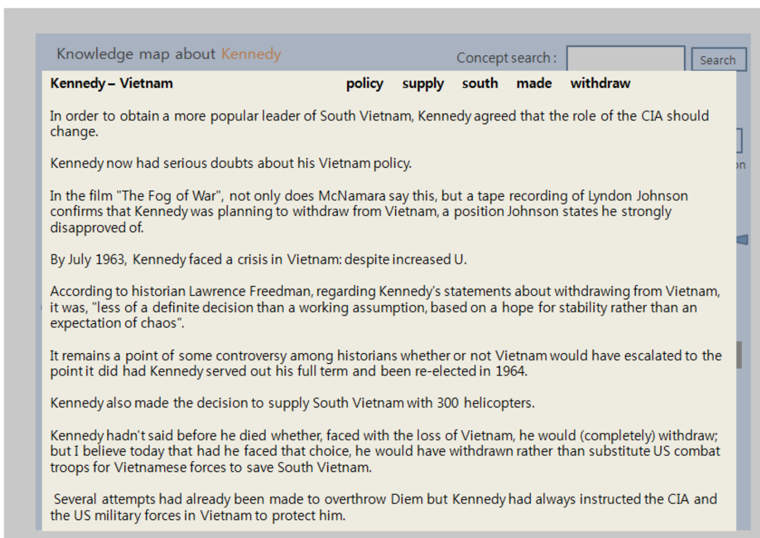

Figure 4. Screen Shot of K-map Tools Categorizing Sentences with More Words

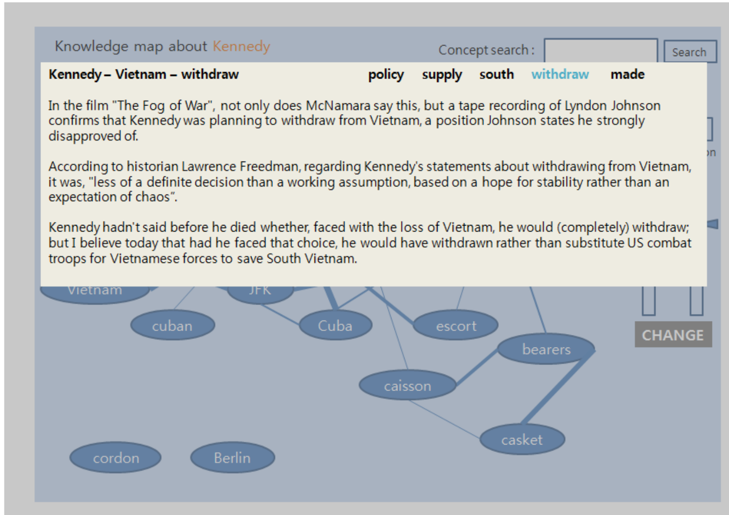

Figure 5. Screen Shot of K-map Tools Categorizing Sentences with Added Word 'Withdraw'

3.2 Model Structure

In this section, the processes of K-map construction are described. Model construction consists of three parts, which are keyword extraction, relation extraction, and relation labeling.

3.2.1 Keyword Extraction

In the first phase, keywords are extracted from a set of documents. The documents can be manually chosen or chosen from the user's query. Each term will receive a weight using the term weighting algorithm. Keywords will be selected from the top-ranked terms set by a limit given by the K-map user. Before running the term weighting algorithm, stop words, such as "and" and "the", are deleted from original sentences of a document. Next, a stemming process is conducted with those sentences that don't have stop words. The stemming process cuts the original word down to the root. For example, after the stemming process, the words "run", "running", and "runs" become the same word, which is "run", because all the words have the same meaning. To stem words, Porter's algorithm was used in the K-map system (Porter, 1980). Once these pre-processes are finished, we can use a term weighting algorithm to weight terms. The weighting algorithm is based on the Term Frequency / Inverse Document Frequency (TF/IDF) method (Salton & Buckley, 1988). The formula below is the weighting method using TF/IDF.

$$w_{ik} = \frac{tf_{ik}\log(N/n_k)}{\sqrt{\sum_{k=1}^{t}(tf_{ik})^2[\log(N/n_k)]^2}}$$

W_{ik} : weight of term k in document i

tf_{ik} : term frequency of term k in document i

N : total number of documents

n_k : number of documents that contain term k

After the weighting process, each term has its own weight. Since a word is weighted for each document using the TF/IDF algorithm, if the same word appears in other documents, its weight can then be different. In this case, the system takes the maximum value for its weight.

$$\mathbf{W_{MT} = Max(W_{D_i T})}$$

D_i = i-th document, i = 0,1,2…total number of the documents in K-map

W_T = Weight of term T in K-map

W_{DiT} = Weight of term T in Di

Once words are ordered by weights, keywords are selected from top-ranked words according to the limit that the user set. Unlike concept and knowledge maps, K-maps can have other parts of speech, not only nouns, as its nodes. We decided to take other parts of speech as well, because we thought that verbs also can have meaning if they are accorded a high weight.

3.2.2 Relation Extraction

Once the keywords of the K-map are decided, relations are defined. Chen et al. (2008) identified assumptions for defining relations. In this study, two factors are considered based on these assumptions. One factor is how frequently two keywords appear in a sentence together, and the other is how big of a role the two keywords play in a sentence. As the number of the words in a sentence increases, the weight decreases. In other words, the score of a relation in a shorter sentence is higher than the score of a relation in longer sentences. Each relation has its own score, and it is used to rank the relations.

$$\mathbf{R_{i,j} = \sum_{D_m} \sum_{S_n} \frac{2}{N_{D_m S_n}}}$$

i, j = keyword pair

$R_{i,j}$ = score of relation between word i and word j

m = 1,2,… total number of documents in a map

n = 1,2,… total number of sentences in document D_m

S_n = n-th sentence

D_m = m-th document

N_{DmSn} = total number of words in sentence S_n, document D_m

3.2.3 Relation Labeling

When the user chooses thicker lines, which are considered stronger connections, he can access many sentences at the same time. This can be overwhelming for the user, and therefore a way of decreasing the number of sentences is suggested: to extract words representing the relation and to use them in categorization. Initially, several sentences were extracted. Within the sentences, for each term its term frequency is calculated. TF/IDF was empirically found to be less reliable than TF. The most frequent words, not including the two keywords, can be considered representative words. Once the user chooses one of these representative words, he will see the sentences that have two keywords and all representative words. Representative words categorize the initial sentences. Categories might overlap one another. This phase is currently still under development and further research is needed.

Conclusion

K-map offers multiple benefits, especially when used in an e-learning platform. A user can see the key concepts in a domain and can identify which concepts are strongly related to others. As a user reads sentences, he can directly access a document he wants from a certain sentence; in other words, K-map can function as a search engine. By exploring the map, a user can learn about the domain at some level of knowledge without accessing original documents. As he explores a domain knowledge K-map, he can see the holistic/overall picture. In addition, since a user can choose relations based on keywords, he can selectively learn about the domain, which is hardly possible when he learns from text. Furthermore, according to initial experiments, K-map provides a mechanism with which to distinguish the more important sentences. Therefore, K-maps show promise as a tool for e-learning environments.

References

Amer , A. A. (1994). The effect of knowledge-map and underlining training on the reading comprehension of scientific texts. *English for Specific Purposes*, 13(1), 35-45.

Ausubel, D. (1968). *Educational psychology: A cognitive approach*. New York: Holt, Rinehart and Winston.

Böhm, K., Heyer, G., Quasthoff, U., Wolff, C., (2002). Topic Map Generation Using Text Mining. *Journal of Universal Computer Science*, 8(6), 623-633.

Cañas, A. J., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T. C., Lott, J., Carvajal, R. (2005). Concept maps: integrating knowledge and information visualization*" Lecture Notes in Computer Science*, 3426/2005, 181-184.

Carvalho, M. R., Hewett, R., Cañas, A. J. (2001). Enhancing web searches from concept map-based knowledge models. In *Proceedings of Multiconference on Systems, Cybernetics and Informatics*, 69-73.

Chen Y., Xia H., (2009). Research on the Auto-construction Methods of Concept Map, Proceedings of *International Conference on Intelligent Human-Machine Systems and Cybernetics*.

Chen, N. S., Kinshuk, Wei, C. W., Chen, H. J. (2008). Mining e-Learning domain concept map from academic articles, *Computers & Education*, 50(3), 1009-1021.

Czerniak, M. C. (1998). The effect of collaborative concept mapping on elementary preservice teachers' anxiety, efficacy, and achievement in physical science. *Journal of Science Teacher Education*, 9(4), 303-320.

Dansereau, D. F., Dees, S. M., Simpson, D. D. (1994). Cognitive Modularity: Implications for Counseling and the Representation of Personal Issues, *Journal of Counseling Psychology* 41(4), 513-523.

Dicheva, D., Dichev, C., (2006). TM4L: Creating and Browsing Educational Topic Maps. *British Journal of Educational Technology*, 37(3), 391-404.

Gomez, A. Morenoa, A, Pazosa, J., Sierra-Alonso, A. (2000). Knowledge maps: An essential technique for conceptualization. *Data & Knowledge Engineering*, 33,169-190.

Hall, R. H., Dansereau, D. F., Skaggs, L. P. (1992). Knowledge maps and the presentation of related information domains. *Journal of Experimental Education*. 61(1), 5–18.

Hall, R. H., O'Donnell, A. M., (1996). Cognitive and Affective Outcomes of Learning from Knowledge Maps. *Contemporary Educational Psychology*, 21(1), 94-101.

Holley, C. D., Dansereau D. F. (1984). *Spatial learning strategies: Techniques, applications, and related issues.* New York, NY: Academic Press.

McClure, J. R., Sonak, B., Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475-492.

Nesbit, J., Adesope, O. (2005). Dynamic concept maps. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 4323-4329.

Novak, J. D., Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal*, 28, 117-153.

O'Donnell, A. M., Dansereau D. F, Hall R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, 14(1), 71-86.

Pepper, S., (2000). The TAO of topic maps. *Proceedings of XML Europe*.

Porter, M.F. (1980). An Algorithm for Suffix stripping, *Automated Library and Information Systems* 14(3), 130-137

Rewey, K. L., Dansereau, D. F., Skaggs, L. P., Hall, R. H., Pitre, U. (1989). Effects of scripted cooperation and knowledge maps on the recall of technical material. *Journal of Educational Psychology*, 81, 604-609.

Salton, G., Buckley, C. (1988). Term-weighting approaches in automatic text retrieval, *Information processing & management*, 513-523.

Stoyanova, N., Kommers, P. (2002). Concept mapping as a medium of shared cognition in computer-supported collaborative problem solving. *Journal of Interactive Learning Research*, 13(1), 111-134.

Tseng, S. S., Sue, P. C., Su, J. M., Weng, J. F., Tsai, W. N. (2007). A new approach for constructing the concept map. *Computers & Education* 49(3), 691-707.

Urfat, N., Korhan, G. (2009). Relation extraction among learning concepts in intelligent tutoring systems, *Proceedings of Application of Information and Communication Technologies*.

Wiegmann, D. A., Dansereau, D. F., McCagg, E. C., Rewey, K. L., Pitre, U. (1992). Effects of knowledge map characteristics on information processing. *Contemporary Educational Psychology*, 17(2), 136-155.