Concept Maps: Theory, Methodology, Technology
Proc. of the First Int. Conference on Concept Mapping
Pamplona, Spain 2004

MINING THE WEB TO SUGGEST CONCEPTS DURING CONCEPT MAP CONSTRUCTION

Alberto J. Cañas, Marco Carvalho, Marco Arguedas, Institute for Human & Machine Cognition, USA
David B Leake, Ana Maguitman, Thomas Reichherzer, Indiana University, USA
Email: {acanas, mcarvalho, marguedas}@ihmc.us, www.ihmc.us
Email: {leake, anmaguit, treichhe}@cs.indiana.edu, www.cs.indiana.edu

Abstract. The most challenging aspect of constructing a concept map is not coming up with the list of concepts to include, but linking the concepts into meaningful propositions to create a coherent structure that reflects the person's understanding of a domain. We present an algorithm which, during the process of concept mapping, takes the partially constructed map as input to mine the Web, and presents to the user a list of suggested concepts that are relevant to the map under construction. We previously reported that testing an initial implementation of the algorithm with a set of users during a concept-mapping workshop seemed to support its viability. In this paper, we present the results of an experiment in which users rank the relevance of concepts suggested by the algorithm during a concept map construction session, using an implementation of the algorithm in the CmapTools software.

1 Introduction

Concept mapping is a process of meaning-making. It implies taking a list of concepts – concepts being perceived regularities in events or objects, or records of events or objects, designated by a label (Novak & Gowin, 1984) – and organizing them in a graphical representation in which pairs of concepts and their linking phrases form propositions. Hence, key to the construction of a concept map is the set of concepts on which it is based. In educational settings, teachers often prompt students by providing an initial set of concepts that they should include in their map.

Coming up with the list of concepts to include in a concept map is primarily an issue of retrieving from long-term memory. In fact, rote learners are particularly good at listing concepts for a domain. It is the process of linking the concepts to create meaningful propositions within the structure of a concept map that is the most difficult task. Jonassen (2000) emphasized that nontrivial effort may be required to choose a linking phrase that represents the relationship between two concepts, not only because of the large number of possibilities, but also because of the need to set that relationship in the context in which the pair of concepts is presented. However, the need to identify relevant concepts may require effort that distracts from the task of creating meaningful propositions.

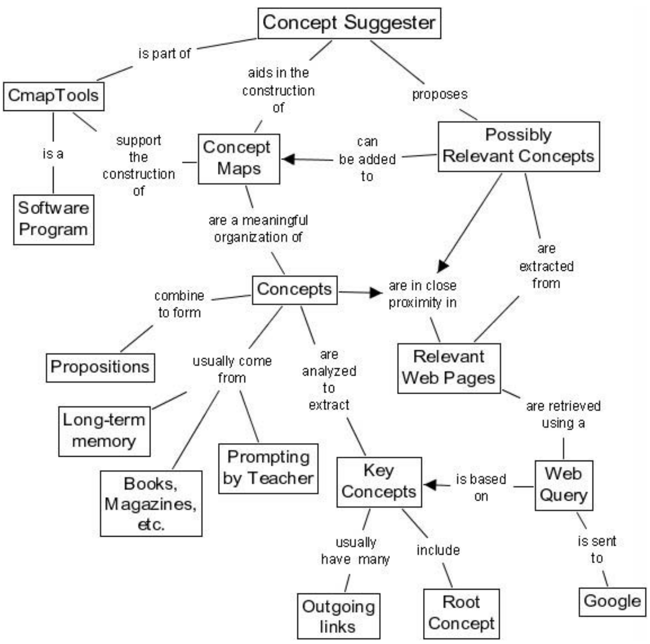

Figure 1: A concept map on the *concept suggester* module.

Often, while constructing a concept map, users – whether elementary school students, scientists, or other professionals – pause and wonder what additional concepts they should include in their map. It is not that they do not know more about the domain they are modeling, it is that they cannot “remember” what other concepts are relevant.

At the Institute for Human and Machine Cognition (IHMC) we have developed CmapTools (Cañas et al., 2004), a widely used software program that supports both the construction of concept maps and the annotation of the maps with additional material such as images, diagrams, video clips, and other such resources. It provides the capability to store and access concept maps on multiple servers and supports knowledge sharing across geographically distant sites.

In this paper we report on the usability of a *concept suggester* module, part of version 4 of CmapTools, which automatically analyzes a concept map under construction, extracts information from it, proactively searches on the World Wide Web (WWW) for concepts that may be relevant to the context of the map, and presents the user with a list of concepts as suggestions for possible inclusion in the map. Previously, we reported preliminary results (Cañas, Carvalho, & Arguedas, 2002) that lead to the development of the module and its inclusion in the software. This *concept suggester* module is part of a larger effort to aid users in the construction of concept maps. Leake et al. (2002) described a module that suggests prior concept maps and associated resources that the user can compare and possibly include as part of the concept map being constructed. Leake et al. (2004) presented a general framework on how to search the WWW using a concept map to generate the query.

This paper begins with a short description of concept mapping. It then presents an example of the use of the *suggester* followed by the algorithm used to extract relevant concepts from the WWW. Finally, results from an experiment in which users ranked the suggested concepts are presented and discussed. Figure 1 shows a concept map summarizing the purpose and function of the *concept suggester*.

2 Concept Maps and Concept Mapping

Concept maps are tools for organizing, representing, and sharing knowledge. Specifically, concept maps, developed by (Novak & Gowin, 1984), have been designed to tap into a person’s cognitive structure and externalize concepts and propositions. A concept map is a two-dimensional representation of a set of concepts constructed so that the interrelationships among them are evident (see Figure 1). The vertical axis expresses a hierarchical framework for the concepts. More general, inclusive concepts are found at the highest levels, with progressively more specific, less inclusive concepts arranged below them. These maps emphasize the most general concepts by linking them to supporting ideas with propositions.

Concept maps are assimilation theory’s major methodological tool. Ausubel’s (Ausubel, 1968) assimilation theory belongs to the family of theories contributing to a constructivist model of human representational processes. Ausubel posited that meaningful learning involves the assimilation of new concepts and propositions into existing cognitive structures. This assimilation of new meaning leads to progressive differentiation and reintegration of cognitive structures. He explicated various forms of meaningful (as opposed to rote) learning involving the assimilation of new information. Ausubel assumed that meaningful learning requires that the learner’s cognitive framework contain relevant anchoring ideas to which new material can be related. Indeed, he argued that the most important single factor influencing learning is what the learner already knows. Ascertain this and teach accordingly.

Meaningful learning results when learners make a conscious effort to relate new knowledge to be learned with relevant knowledge they already possess. In contrast, rote learning results when learners memorize the new information and make little or no effort to relate and integrate this with their prior knowledge. Information learned by rote is notoriously soon forgotten, and there is little chance for the application of this knowledge in new problem solving contexts (Novak, 1998).

A growing body of research indicates that the use of concept maps can facilitate meaningful learning. During concept map construction, meaning-making occurs as the learner makes an effort to link the concepts to form propositions. The structure of these propositions in a map is a reflection of his/her understanding of the domain. Therefore, the meaning-making process involves both coming up with the list of concepts to include in a map and establishing the relationship between concepts. A rote learner may very well come up with the same list of concepts as a meaningful learner, but may not be able to establish explicitly the relationship between the concepts in the form of propositions. On the other hand, providing a meaningful learner with a richer set of concepts on which to build his/her map can help the learner construct a more complete representation of his/her understanding of the topic.

3 CmapTools and the Concept Suggester

Software programs like CmapTools make it easier for users to construct and share their knowledge models based on concept maps. In CmapTools we have extended the use of concept maps to serve as the browsing interface to a domain of knowledge, and have provided a tool that allows users to construct, organize, navigate, criticize, and share knowledge models. The software is widely used all over the world, by users who range from elementary school children, to professors creating content for distance learning courses, to NASA scientists. Applications of the tools range from supporting collaborative knowledge construction by students from different countries (Cañas et al., 2001), to just-in-time training (Coffey et al., 2003), to a large multimedia knowledge model about Mars at NASA (Briggs et al., 2004).

This broad range of users and applications has provided us with extensive feedback on the process of concept map construction. Taking advantage of this information, we have enhanced the tools with additional features to proactively aid the users in the construction of their knowledge models. Within this effort, we propose that unobtrusively presenting to the user a list of concepts that seem to be relevant within the context of the concept map being constructed, would allow users to concentrate on the meaning-making process of linking the concepts to form propositions and structuring the map, while freeing them from the effort of "remembering" which concepts are missing.

Consider the concept map about concept maps constructed by Joe Novak shown in Figure 2. Which other relevant concepts should be added to this map? The pane on the right shows what the *concept suggester* lists as possible concepts to add. (Figure 3 shows a larger version of this list.) These terms appear automatically as a user constructs the maps, and are updated when the program determines that the map has undergone significant changes. We have highlighted the term "Novak" in the list of suggestions. By analyzing the content of the map and searching through the Web, the *suggester* has determined that the term "Novak" is relevant to this map. Obviously, a relevant addition to the map would be to include that "concept maps were invented by Joe Novak," and the user may have forgotten to add it, although s/he was aware of it. (In this particular case, the authors are certain that Joe Novak was too modest to include it.) Other relevant concepts suggested in the list include "assessment," which is one of the main applications of concept maps, "study," "student," "science," and "understand (understanding)." Within the list, some suggestions may not make sense, for example, "reference"

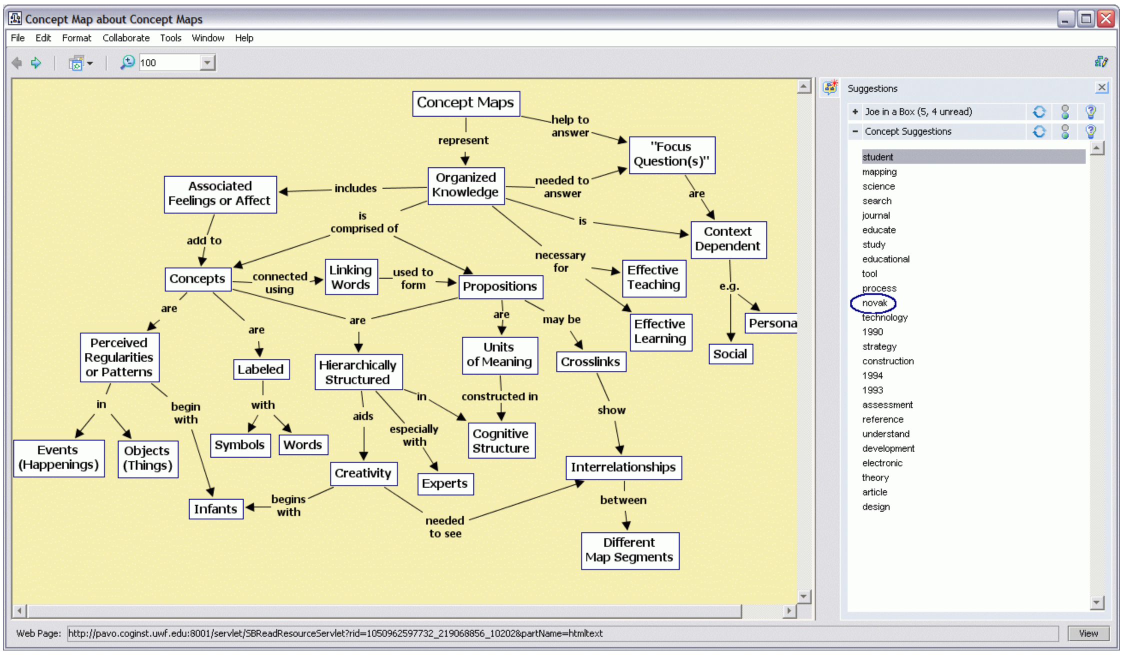

Figure 2: Concept map about concept maps by Joe Novak, and a list of suggested concepts from the CmapTools *concept suggester*.

and "article." The user can double-click on the suggestions to display Web pages that include the selected term within the context determined by the concept map.

We are aware that the *concept suggester* will not be able to come up with a list of terms that are all relevant to the map, given that all it is doing is mining the Web without any real understanding of the map itself. However, we believe that even if only two or three relevant concepts in the list trigger ideas in the user, it will result in an improved concept map.

4 The Concept Suggester

To find and suggest relevant concepts, we take advantage of various key characteristics of concept maps:

1. Concept maps have structure: By definition, more general concepts are presented at the top with more specific concepts at the bottom. Therefore, different weights can be assigned to the concepts in the partially constructed map according to their relative vertical position. Other structural information, for example, the number of ingoing and outgoing links of a concept, may provide additional information regarding a concept's role in the map. (Leake, Maguitman, & Reichherzer, 2004 presented experimental support for the cognitive importance of such factors.)

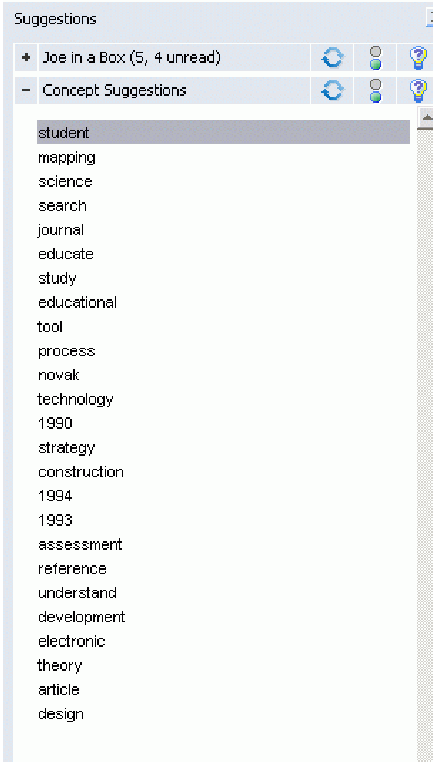

Figure 3: List of concepts suggested by CmapTools for the concept map about concept maps in Figure 2.

2. Concept maps are based on propositions: If two concepts form a proposition, the Web search for relevant documents may take into account that they are related, by seeking documents in which the two concepts appear related due to their proximity to each other in the text.

3. Concept maps have a context: A concept map is a representation of a person's understanding of a particular domain of knowledge. As such, all concepts and linking phrases are to be interpreted within that context, and the concept finder can take advantage of it.

As the user proceeds in the construction of the concept map, the program automatically reviews the changes as they are made and determines when it is appropriate to update the list of suggested concepts. The process of preparing a list of concepts consists of the following steps:

a) Analyzing the partial concept map to prepare a relevant query to use in searching the Web;
b) Retrieving relevant documents from the Web;
c) Extracting the relevant concepts from the retrieved Web pages;
d) Presenting the concepts to the user.

In this section, we describe the implementation of steps a) through c) of this algorithm in the CmapTools program. The system has undergone initial evaluation, and the results are described later in this paper.

4.1 Analyzing the Partial Concept Map

This phase consists of extracting from the concept map a limited set of words that represents the concept map's context, to be used as a query for our metasearch engine.

In traditional information retrieval, word frequency analysis is used to extract keywords from text. This approach, however, would not be effective in a concept map. The concise nature of the map would distort the frequency of words and, furthermore, because in a good map concepts are not repeated, all terms would most likely have the same frequency.

Our approach is to perform a graphical analysis of the partial map to identify the key concepts that play an important role in the context. Specifically, we try to identify concepts that refer to the focus question characterizing the topic of the map and concepts that are authority nodes.

Ideally, concepts consist of a single word, or a small set of words. In practice, though, during the process of building a map it is common to find concepts that consist of a large number of words, or even small phrases. For each concept, we try to identify the most relevant words by removing all stop words. If the result is still three words or longer, or if the result is an empty concept, it is discarded for the rest of the process.

At any stage of development, the root node of the map is usually a good representation of the overall topic of the map, or the focus question. The root node is assumed as an important concept, and its label is included as part of our query as long as it consists of less than three words once the stop words are removed.

Authority nodes are those with the highest number of outgoing links to other nodes. We assume that the outgoing links are indicative of further elaboration of these concepts, and therefore a gauge of their relevance in the context of the map. The algorithm selects, among all the nondiscarded concepts, those with the largest number of outgoing links. If more than one concept has the same (largest) number of outgoing links, they are all included in the query.

The process then consists of scanning the concept map to locate the root concept and the authority node(s). The overall number of concepts retrieved is dependent on the size of the map. Large maps might have many authority nodes, which would result in a larger number of key concepts. Given the restriction on concepts having less than three words, the process could yield an empty query, in which case the *suggester* cannot proceed. The query is constructed from the resulting concepts in no particular order.

4.2 Retrieving Relevant Documents

We use the query constructed from the key concepts in the previous step to retrieve and rank Web pages and build our collection of documents for the concept mining.

We have developed a metasearch engine, based primarily on Google (Brin & Lawrence, 1998), in order to retrieve an initial set of documents from the public Internet. The meta-search engine returns a small set of 10 to 20 URLs, depending on the query.

With the documents retrieved, parsed, filtered for stop words, and indexed, we proceed to the next phase, the actual mining for relevant concepts.

4.3 Extracting Relevant Concepts

Our current approach to extracting relevant concepts is simple: search all retrieved documents for all nondiscarded concepts from the map. Each time a concept is found in a document, all the neighboring words (excluding stop words) are saved in a temporary table. A word is considered a neighbor if it is part of same sentence as the concept in the text and is within a specific distance (currently three words) from the concept. In the current implementation, all neighboring words have an equivalent weight and are potential candidates for suggestion.

The result of searching for all of the map’s concepts in all the documents is a large collection of terms that are neighbors of the map’s concepts in the text. We now proceed to rank these terms using frequency analysis to obtain an ordered list of candidates for suggestion. This list is then part-of-speech (POS) tagged and sorted by frequency. The final set of concepts that is suggested to the user includes 15 concepts composed by the first five top ranking nouns in the sorted term list, followed by the top five verbs and the top five remaining terms. The *suggester* will automatically adjust these proportions as a function of the size of the subset of terms to display to the user.

5 Experimental Procedure

The experimental evaluation was based on human subjects ranking the relevance of proactively suggested concepts during the construction of a concept map. The goal was to determine the proportion of users that would find the suggestions useful during the map construction process.

Each participant received written and verbal instructions at the beginning of the experiment. The instructions described the goal of the experiment, the criteria for ranking suggestions, and details on how to use the *concept suggester* and assign relevance scores to the suggestions. The participants were also provided with a topic for a concept map and a version of CmapTools that was modified specifically for this experiment.

A single topic was proposed for all subjects. The participants were asked to build a single concept map about "Computers" with at least 20 linked concepts. During the map-building process, the *concept suggester* was running and proactively listing suggested concepts. At different times during the construction of the map, the participant was prompted to rank the displayed list of suggestions. At this time, s/he stopped the construction of the map and specified the relevance of each suggestion. Each suggested term was ranked on a scale from 1 to 5, where 1 meant "low relevance" and 5 "high relevance." Participants were instructed to assign higher relevance to suggested concepts that would satisfy the following:

a) The concept (or some variation of it) could be added to the map, or could be used as a replacement for a concept in the map in order to improve or extend its breadth (the user did not have to add the term to the map to rank it highly; we were interested in determining whether the user found the concept relevant).
b) The concept triggers a search for information that leads to the extension or improvement of the map.

On average, each participant was prompted at least five times for suggestions. At each instance, all the rankings were saved, together with the list of suggestions and the state of the map used to search the Web for suggestions. There were no constraints on time allotted for the construction of the map or for the ranking of suggestions.

After scoring all the suggested terms, the participant continued with the construction of the map. When the map reached 20 concepts, the experiment concluded. Figure 4 shows an illustration of the client application prompting the user for ranking.

During the experiment, the participants were allowed to request new concept suggestions at any time. In addition, CmapTools proactively generated new suggestions for users when significant changes in the concept map were detected. This is the normal behavior of the *concept suggester* within CmapTools. In both cases, however, at the appropriate time the user was prompted for ranking.

Fourteen participants volunteered for the experiment. The data from one of the participants had to be discarded because partly through the experiment it was determined that the instructions had been misunderstood

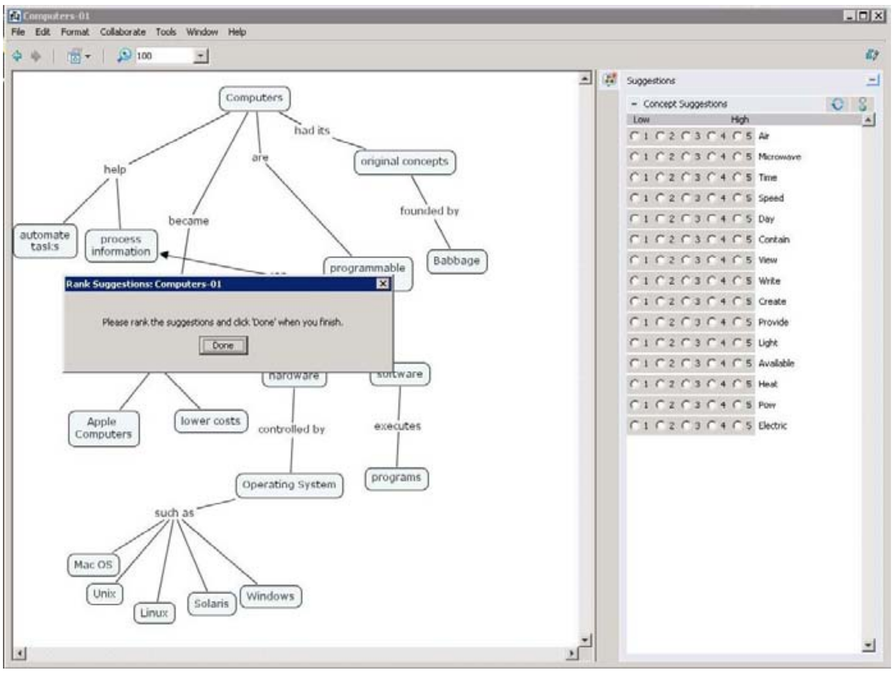

Figure 4: Prompting the user to rank the suggestions.

and the participant was ranking the suggested terms with respect to the selected concept in the map, not with respect to the whole map. All participants were familiar with the CmapTools application and with concept mapping although none of them had previous experience using the *concept suggester* during the construction of a map. The great majority of the participants (12 out of the 13 analyzed) had a computer science background, and all were familiar with the topic chosen for the map.

When the map reached 20 linked concepts, the subject was notified by the application that the experiment was over.

6 Experimental Results

We were interested in determining whether users reported that a *reasonable* number of suggested terms were *relevant*. To do this, we first established that a term was *relevant* if it was ranked 3 or above by the subject. Next, we determined a reasonable minimum number of relevant concepts at each stage. To do this, we considered that: (a) even a single relevant concept could have an important impact on the map construction, and (b) given the nature of the algorithm, in which suggested terms are mined from the Web, we could not expect most of the suggested terms to be relevant. We considered at least three or four relevant terms to be a reasonable criterion for a useful *suggester*. This number is subjective, of course.

Figure 5 shows the result of the experiment based on these two parameters. The chart on the left displays the percentage of subjects that ranked at least 4 concepts (out of the 15 concepts suggested) as relevant, for each of five stages in time while building the concept map. As can be seen, 85% (11 out of 13 subjects) considered at least 4 of the concepts suggested to be relevant when the map construction was at its early stages. This number decreased to 55% by the fifth ranking, however, as the construction of the map evolved. The chart on the right shows the percentage of subjects that ranked at least 3 concepts as relevant. In this case, the percentage decreased from 100% to 73%.

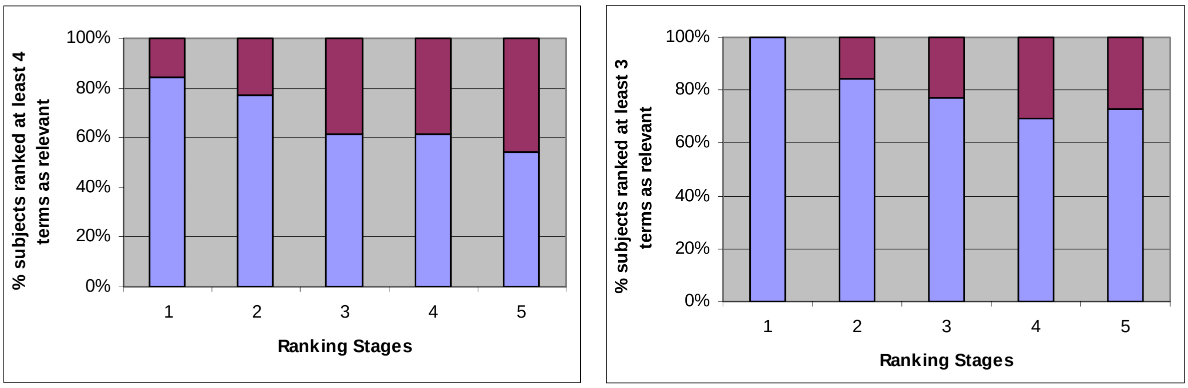

Figure 5: Percentage of subjects that ranked at least four (left) and at least three (right) suggested concepts as relevant in each of the ranking stages.

7 Discussion

The results show that the *suggester´s* algorithm is quite effective at the early stages of the construction of the concept map. Up to the third stage of ranking, the percentage of subjects that ranked at least four of the suggested terms as relevant was above 62%. Moreover, the percentage that ranked at least three as relevant was 77% or above. These numbers suggest that the algorithm is effective. However, in both cases we observe that as the concept map grows, the algorithm is less effective at finding relevant terms. The algorithm tries to determine "what the map is about." When the concept map is small, it is relatively easy to select the key concepts. As the map grows, the key concepts become spread throughout the map, making it more difficult for the algorithm to determine what the complete map is about. Also, as the map grows, the users tend to work on a piece or section of the map at a time. Consequently, for the concepts to be deemed relevant, they should be selected within the context of that piece of the map, not based on the map as a whole. Therefore, we propose that the algorithm be adjusted to take into account the dynamics of map construction.

8 Summary

Concept map construction is a meaning-making process in which listing the concepts to be included in the map is a less central task than selecting the appropriate linking phrase to form propositions. We have found, however, that users often struggle to "remember" new concepts to add to their maps when they should be concentrating their efforts on how to link those concepts to the map. We have implemented in the CmapTools software a proactive *concept suggester* module which, during map construction, analyzes the concept map, creates a relevant query to search the Web for documents related to the map, extracts relevant concepts from the retrieved Web pages, and presents the results as suggestions to the user. This module searches for new suggestions whenever it determines that the map has undergone significant changes. Tests of this module with a group of users suggest that it is effective in presenting relevant concepts to the users. This effectiveness diminishes as the map grows, however, which implies that the algorithm should be revised to take into account that in larger maps, users are most likely working on a piece of the map, and the suggested concepts should be determined by the context of that piece.

Acknowledgements

This work was partially funded by NASA's Intelligent Systems Program under awards NCC 2-1216 and 2-1297 and by the U.S. Navy's Chief of Naval Education and Training.

References

Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston.

Briggs, G., Shamma, D., Cañas, A. J., Carff, R., Scargle, J., & Novak, J. D. (2004). Concept Maps Applied to Mars Exploration Public Outreach. In A. J. Cañas, J. D. Novak & F. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the 1st International Conference on Concept Mapping*. Pamplona, Spain: Universidad Pública de Navarra.

Brin, S., & Lawrence, P. (1998). *The Anatomy of a Large-Scel Hyptertextual Web Search.* Paper presented at the 7th WWW Conference.

Cañas, A. J., Carvalho, M., & Arguedas, M. (2002). *Mining the Web to Suggest Concepts during Concept Mapping: Preliminary Results.* Paper presented at the XIII SBIE: Simpósio Brasileiro de Informática Educativa, Porto Alegre, Brasil.

Cañas, A. J., Ford, K. M., Novak, J. D., Hayes, P., Reichherzer, T., & Niranjan, S. (2001). Online Concept Maps: Enhancing Collaborative Learning by Using Technology with Concept Maps. *The Science Teacher, 68*(4), 49-51.

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M., & Carvajal, R. (2004). CmapTools: A Knowledge Modeling and Sharing Environment. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the 1st International Conference on Concept Mapping*. Pamplona, Spain: Universidad Pública de Navarra.

Coffey, J. W., Cañas, A. J., Reichherzer, T., Hill, G., Suri, N., Carff, R., Mitrovich, T., & Eberle, D. (2003). Knowledge Modeling and the Creation of El-Tech: A Performance Support System for Electronic Technicians. *Expert Systems with Applications, 25*(4).

Jonassen, D. (2000). *Computers as Mindtools for Schools* (2nd ed.). Columbus OH: Merrill.

Leake, D., Maguitman, A., Reichherzer, T., Cañas, A. J., Carvalho, M., Arguedas, M., & Eskridge, T. (2004). Googling from a Concept Map: Towards Automatic Concept-Map-Based Query Formation. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping*. Pamplona: Universidad Pública de Navarra.

Leake, D. B., Maguitman, A., & Cañas, A. J. (2002). *Assessing Conceptual Similarity to Support Concept Mapping.* Paper presented at the Proceedings of FLAIRS: Fifteenth Florida Artificial Intelligence Research Symposium, Pensacola.

Leake, D. B., Maguitman, A., & Reichherzer, T. (2004). *Understanding Knowledge Models: Modeling Assessment of Concept Importance in Concept Maps.* Paper presented at the Twenty-Sixth Annual Conference of the Cognitie Science Society.

Novak, J. D. (1998). Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations. Mahwah, NJ: Lawrence Erlbaum Associates.

Novak, J. D., & Gowin, D. B. (1984). *Learning How to Learn*. New York: Cambridge University Press.