Combining Concept Mapping with CBR: Towards Experience-Based Support for Knowledge Modeling*

Alberto Cañas
Institute for Human and Machine Cognition
University of West Florida
40 South Alcaniz Street
Pensacola, FL 32501
acanas@ai.uwf.edu

David B. Leake and Ana Maguitman
Computer Science Department
Lindley Hall , Indiana University
150 S. Woodlawn Avenue
Bloomington, IN 47405
{leake,anmaguit}@cs.indiana.edu

Abstract

Knowledge management depends on effective methods for capturing knowledge in useful forms and making it available when needed. Electronic concept maps provide a promising representation for knowledge models that can be developed directly by the experts themselves, but the flexibility of concept mapping raises questions of how to support the knowledge modeling process and to standardize its results, in order to facilitate future examination and re-use. We describe ongoing research on the use of case-based reasoning methods to support the knowledge modeling process through proactive retrieval of relevant prior concept maps, in order to provide suggestions to aid the concept map generation process. The selection of relevant concept maps relies on an algorithm that combines textual and topological analysis. We describe the algorithm and present an example that illustrates concept suggestion procedures in the Mars exploration domain.

Introduction

The task of knowledge management is to capture useful knowledge and make it available in a usable form when it is needed in the future. Successful management of expert knowledge depends on the ability to elucidate the experts' understanding of a domain, to represent that understanding in a form that supports effective examination by others, and to make the encoded knowledge accessible when needed. A central question for both research and practice is how to capture and represent the needed knowledge. One approach is to develop carefully-crafted knowledge models in a structured and standardized form, which maximizes the usefulness of captured knowledge for automated processing but requires considerable involvement by knowledge engineers to mediate knowledge modeling. Another approach, at the other extreme, is to alleviate the knowledge acquisition burden by simply allowing experts to enter the knowledge they choose, as textual passages to be retained without further processing. This approach simplifies knowledge capture, but at the cost of usability—the resulting texts may be difficult for future users to understand and apply. This position paper proposes a middle approach, aimed at providing usable knowledge while controlling the knowledge acquisition burden: Exploiting AI methods to develop intelligent systems to support knowledge modeling, in order to empower domain experts to directly construct, navigate, share, and criticize rich knowledge models.

We are developing intelligent support tools to help experts represent their knowledge in a structured form, and to refine it in distributed collaboration with other experts. Our approach combines interactive tools for *concept mapping* (Novak and Gowin, 1984) with retrieval techniques from *case-based reasoning* (e.g., Kolodner, 1993, Leake, 1996, Watson 1997). In the combined approach, concept mapping provides methods for knowledge capture, representation, refinement, and examination; case-based reasoning techniques—taking advantage of the knowledge in the models themselves and contextual information gathered from the expert's navigation through them—provide mechanisms for storing and retrieving relevant prior concept maps for the expert to consider. This in turn provides the foundation for experience-based support for the expert's process of selecting important concepts and relationships to include. The goal is to provide scaffolding for experts building their own concept maps, consulting and critiquing prior concept maps, and linking their own concept maps to others'. The project aims to develop proactive support for knowledge access, comparison, and re-application, as well as automatic support for the development and standardization of concept map representations. This paper summarizes key issues and initial methods for this framework. The goal of this work is to support knowledge capture and sharing across time, through case-based reasoning, as well to support distributed knowledge sharing through access, integration,

* This research is supported in part by NASA under award No NCC 2-1035.

and comparison of concept maps, cases, and other forms of multimedia information across the Internet.

Background

Case-based reasoning is the process of learning and reasoning by capturing and reusing lessons from analogous prior experiences (Kolodner, 1993). The proficiency of case-based reasoners comes from having the right cases, being able to access them at the right times, and being able to apply them in the right ways. Because human experts frequently remember, gather, compare, and reason from specific examples, they often find CBR to be a natural method for supporting knowledge capture and sharing. As additional experiences or lessons are stored in the CBR system, they form a growing corporate memory to capture collective experience and make it available when needed in the future. Case-based reasoning is receiving considerable current attention in knowledge management and lessons learned systems (for a sampling of papers on this subject, see Aha et al. (1999) and Aha & Weber (2000)).

Case-based knowledge management systems often capture information in purely textual form. This facilitates knowledge capture, but may obscure the structure of the models being recorded, making it difficult to identify or compare key factors and relationships. Other systems use carefully-crafted structured representations, at the cost of requiring significant intervention and effort by knowledge engineers. We are investigating *concept maps* (Novak and Gowin, 1984) as a medium for knowledge models that are useful but also tractable for the experts themselves to build. Concept mapping is designed to tap into people's internal cognitive structures and externalize concepts and propositions. A concept map is a graphical display of concept names connected by directed arcs encoding propositions in the form of simplified sentences. When a concept map is generated in an electronic form, nodes in the concept map may also be associated with multimedia information to supplement and clarify its text, as illustrated in the sample concept map shown in Figure 1. Concept maps appear similar to semantic networks and conceptual graphs, but are not constrained by syntactic rules and have no associated semantics. They were developed as a

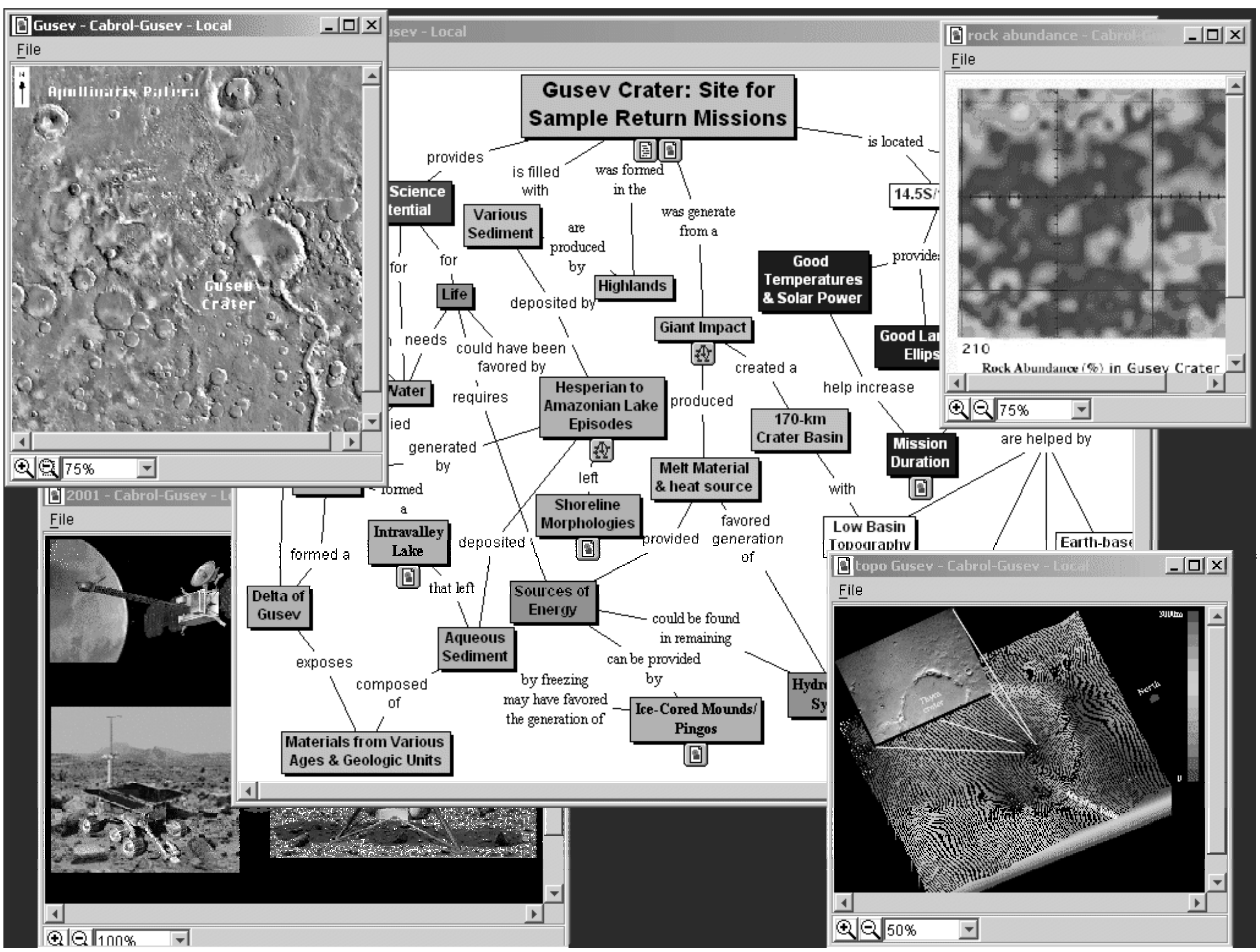

Figure 1 A concept map for the Mars exploration domain.

pedagogic device for use by humans "sketching out" concepts, rather than as a formal device for use by reasoning engines, and have received much use in educational settings for elucidating, sharing, and comparing knowledge.

Electronic concept maps provide an elegant representation of an expert's domain knowledge in a browsable, sharable form, easily understood by others, and the "informal" nature of concept map representations enables them to be generated by the experts themselves. This makes them a strong candidate method for direct entry, examination, and sharing of experts' knowledge. For example, concept mapping tools from the Cognition Institute at the University of West Florida (http://cmap.coginst.uwf.edu) have been used for applications such as the creation at NASA of a large-scale multimedia CD and web site on Mars (http://cmex.arc.nasa.gov/cmaps/Mars2000) (Figure 1 shows an example from this CD). These tools already provide the capability for distributed knowledge construction and access over the Internet, but they currently provide no automated support for retrieval of relevant prior concept maps or other intelligent support for the concept map generation process.

Towards Proactive Concept Map Retrieval

Relevant prior concept maps can be a valuable resource to the user who is capturing new knowledge, refining old conceptualizations, or seeking to better understand a domain. In our view, the effectiveness of concept map retrieval tools depends on their ability to anticipate which concept maps are relevant and automatically present them to users when needed. Simply providing users with a query facility is insufficient: As has previously been observed by ourselves and others, users may not use the query facility, in order to avoid the effort of querying, or may miss useful information by not querying at the right times.

Retrieval Issues and Approach

The success of proactive retrieval methods depends on the availability of contextual information (e.g., Budzik and Hammond, 2000), making context extraction a crucial issue. We are investigating the hypothesis that by monitoring the use of concept mapping tools and the user's navigation through existing maps, it is possible to gather a rich body of contextual information to guide retrievals. Our previous work has given promising results on using concept map information to focus retrieval in a domain-specific retrieval application (Cañas et al., 1999).

A central issue for concept map retrieval is how to recognize the similarity of related concept. Related concept maps can be represented in many different ways, requiring the similarity assessment/retrieval process to efficiently recognize the similarity between isomorphic concept maps, despite differences in their layouts. CBR research provides a wealth of approaches to build on for retrieving cases with structured representations. Because of the computational cost of matching structured representations, one promising method is to use a two-step process: inexpensive prefiltering to select likely candidates, followed by more subtle (and expensive) analysis of the selected cases (Gentner, Forbus and Law 1995). To summarize features for inexpensive initial matching we are investigating approaches based on Kleinberg's (1998) algorithms for topological analysis of graphs (previously applied to identify important nodes on the web), which efficiently infer features such as "hub nodes," or "centers of activity." These nodes can be computed at storage time for each concept map, to provide a weighted set of concepts to describe each map. These sets can then be matched against the current map, for fast filtering to find maps with similar important concepts.

Applying Topological Analysis to Concept Maps

Topological analysis can be applied to the analysis of concept maps to describe the relative arrangements of their concepts. The hypotheses underlying our use of this method are (1) that the topology of the concept map can convey useful information to determine the role of each concept in the whole map, and (2) that the topological roles of concepts in the map can be usefully summarized according to a small set of dimensions. Our characterization scheme describes concepts according to four node types:

- *Authorities* are concepts to which other concepts converge. They are the nodes that have the largest number of incoming links arriving from "hub nodes."
- *Hubs* (centers of activity) are the concepts that have the largest number of outgoing links ending at "authority nodes."
- *Upper Node*s generally correspond to those that appear at the top of the map when it is presented in a graphical representation. In general there is one main concept in each concept map specifying the main topic.
- *Lower Nodes* are generally the ones that appear at the bottom of the concept map in a graphical representation.

We associate to each concept four weights, *a-weight*, *h-weight*, *u-weight* and *l-weight* representing the degree to which the concept belongs to the categories mentioned above. Once these weights are computed, they remain static unless the topology of the concept map changes. Thus each concept's role in a concept map can be characterized using only its associated weights, and the roles concepts play in different maps can be compared by comparing their weights.

Cmap Concept Description Algorithm

We have developed an O (n^3) algorithm for characterizing concept map nodes, and are now testing it with promising results. This algorithm calculates weights as follows:

1. For each concept *c* in the set of concepts *CMap*, set *a-weight(c)*=1, *h-weight(c)*=1, *u-weight(c)*=1, and *l-weight(c)*=1.
2. Normalize weights such that
$$\sum_{\substack{c\in CMap \\ w\in\{a-weight,h-weight, \\ u-weight,l-weight\}}} w(c)^2 = 1$$
3. Compute
$$h-weight(p) = \sum_{(p,q)\in Links} a-weigh\,(q)$$
4. Normalize *h-weights* as described in step 2.
5. Compute
$$a-weight\,(p) = \sum_{(q,p)\in Links} h-weight\,(q)$$
6. Normalize *a-weights* as described in step 2.
7. Repeat steps 3 to 6 until a fixed point for the functions *a-weight* and *h-weight* is reached. This requires at most |*Cmap*| iterations.
8. Compute
$$u-weight(p) = \begin{cases} 1 \quad \text{if} \quad \neg\exists(q,p)\in\ Links \\ \sum_{(q,p)\in Links} u-weight(q)^2 \ \text{otherwise} \end{cases}$$
9. Normalize *u-weights* as described in step 2.
10. Repeat steps 8 and 9 until a fixed point for the function *u-weight* is reached.
11. Compute
$$l-weight\,(p) = \begin{cases} 1 \quad \text{if} \quad \neg\exists(p,q)\in\ Links \\ \sum_{(p,q)\in Links} l-weight\,(q)^2 \ \text{otherwise} \end{cases}$$
12. Normalize *l-weights* as described in step 2.
13. Repeat steps 11 and 12 until a fixed point for the function *l-weight* is reached.

This algorithm is based on the scheme presented by Kleinberg (1998), which associates weights to nodes in terms of their roles as authorities or hubs. However, it adds the calculation of two additional weights, *u-weights* and *l-weights*, which, as mentioned earlier, reflect the relative position of a concept in a graphical representation. These provide important information for comparing concept maps, because nodes higher in the concept map representation tend to be associated with the topic of the concept map.

Using the Concept Map Descriptions for Retrieval

Given the characterizations of the individual concepts in a map, we obtain the similarity degree between two concept maps m_1 and m_2 by comparing them as follows. First, we use simple keyword comparisons of node labels to calculate a similarity value used to determine how closely individual nodes in the two maps correspond to each other, by the following formula:
$$\mathrm{K}_w(m_1,m_2) = \sum_{(p,q)\in m_1\times m_2} |\, k_p \cap k_q \,| * w(p) * w(q)$$
where k_p and k_q represent the sets of keywords associated to the concepts *p* and *q* respectively.

Finally, the similarity metric *S* between entire concept maps is computed as follows:
$$S(m_1,m_2) = \sum_{\substack{w\in\{a-weight,h-weight, \\ u-weight,l-weight\}}} c_w * \mathrm{K}_w(m_1,m_2)$$

where the values associated to the c_w's determine in which weight categories we want to focus. For example, we can set $c_{a\text{-}weight}$=0.5, $c_{h\text{-}weight}$=0.5, $c_{u\text{-}weight}$=0, and $c_{l\text{-}weight}$=0 if we want to stress matches between concepts that have a higher rank as authorities or hubs.

Based on the described similarity metric the test system retrieves a set of maps similar to the target. As discussed in the following section, correspondences between individual nodes in similar maps suggest specific concepts relevant to those currently being edited by the user, enabling the user to suggest links from those concepts as possibilities for new links in the current map.

Applying Proactive Retrieval to Aid Generation of Sharable, High-Quality Concept Maps

The concept mapping process is intended to give maximal freedom to clarify and communicate the expert's potentially idiosyncratic understanding. However, this leaves the user with little guidance about how to build a concept map, increasing user effort and complicating later retrievals due to diverging representations for similar concepts. Automatic retrieval of relevant prior concept maps can help alleviate this problem, by presenting suggestions based on similar maps during concept map generation. We see this method as playing two main roles: Helping guide the user towards (1) possible factors to consider and (2) candidate terminology to use for those factors.

Using the previous retrieval techniques, we have been testing link and concept suggestion procedures in the Mars domain, using a body of over 150 concept maps on Mars as a sample case base. Given a concept map in progress,

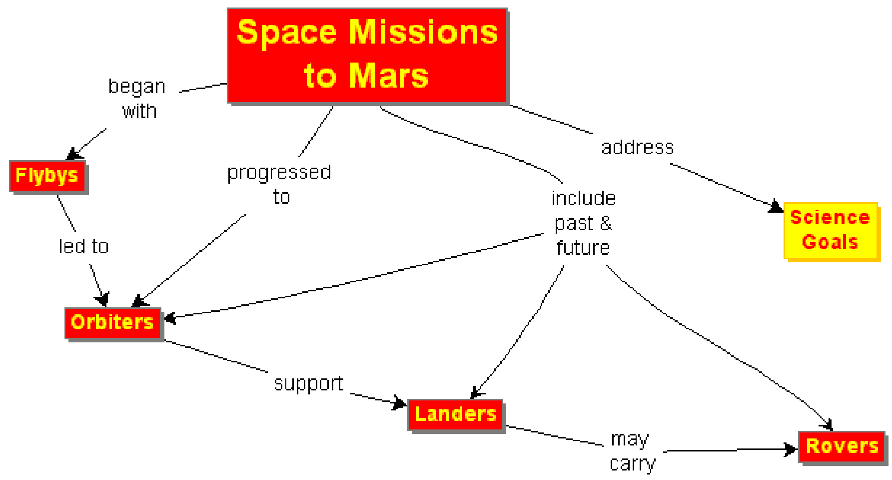

Figure 2 A partial concept map for the Mars domain.

our test system automatically retrieves similar prior concept maps, and suggests links and nodes from those maps for the user to consider adding to the current map. For example, for the sample map shown in Figure 2, and the active node "Space missions to Mars," the system suggests five additional links and the concepts that they point to in similar prior maps, to give ideas for possible additions. The links are "include → Russian and other non US missions," "are aiming toward → sample return," "can be launched every → 26 months," "will eventually lead to → human exploration," and "may include → airborne platforms." Each of these suggests types of elaborations that may be relevant to the development of the new concept map.

Conclusion

This paper describes ongoing research on applying case-based reasoning techniques to proactively retrieve relevant prior concept maps and provide suggestions during the knowledge modeling process, to support experts as they directly build, share, compare, and revise rich knowledge models represented as concept maps. Integrating concept mapping and CBR promises benefits in increasing the practicality of capturing rich knowledge, by helping to share knowledge relevant to the knowledge modeling process and suggesting concepts and links to consider. In addition, by facilitating access to relevant stored knowledge models, it can provide the opportunity to refine *prior* knowledge models in light of new lessons and perspectives.

References

Aha, D., Becerra-Fernandez, I., Maurer, F., and Muñoz-Avila, H. 1999. *Proceedings of the AAAI-99 Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning*. AAAI Press, Menlo Park, CA.

Aha, D., and Weber, R. 2000. *Proceedings of the AAAI-2000 Workshop on Intelligent Lessons Learned Systems*. AAAI Press, Menlo Park, CA.

Ausubel, D., Novak J., and Hanesian, H. 1978. *Educational psychology: A Cognitive View*. New York: Werbel & Peck, 1986.

Budzik, J., and Hammond, K. 2000. User interactions with everyday applications as context for just-in-time information access. In *Proceedings of the 2000 International Conference on Intelligent User Interfaces*, pp. 44-51.

Cañas, A., Leake, D., and Wilson, D. 1999. Managing, mapping, and manipulating conceptual knowledge. In *Proceedings of the AAAI-99 Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning*, pages 10-14, Menlo Park, 1999. AAAI Press.

Forbus, K., Gentner, D., and Law, K. 1995. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19:141-205.

Kleinberg, J. 1998. Authoritative Sources in a Hyperlinked Environment. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 668-677.

Kolodner, J. 1993. *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA.

Leake, D. 1996. *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. AAAI Press, Menlo Park, CA.

Novak, J. D., and Gowin, D. B. 1984. *Learning How to Learn*. Cambridge University Press, New York.

Watson, I. 1997. Applying Case-Based Reasoning: Techniques for Enterprise Systems. Morgan Kaufmann, San Francisco.