Concept Maps: Theory, Methodology, Technology
Proc. of the First Int. Conference on Concept Mapping
A. J. Cañas, J. D. Novak, F. M. González, Eds.
Pamplona, Spain 2004

THE HIGH COST OF KNOWLEDGE RECOVERY

Robert R. Hoffman, Paul J. Feltovich, & Lynne Roetzer
Institute for Human and Machine Cognition, Pensacola FL
David W. Eccles, Florida State University, Tallahassee FL
Email: rhoffman@ihmc.us   www.ihmc.us

Abstract. Knowledge recovery is the process of culling meaningful propositions from documents and crafting them in the form of Novakian Concept Maps, thereby making the material more useful and usable. This report discusses our first major effort at knowledge recovery. Expert knowledge about terrain exists in the form of traditional documents discussing landforms, soils, rock types, etc. (Hoffman, 1983; Mintzer & Messmore, 1984). These were recast as Concept Maps in ROCK-TA ("Representation Of Conceptual Knowledge-Terrain Analysis"), which is a very large knowledge model, consisting of over 150 Concept Maps, containing over 3,000 propositions, and dozens of multimedia resources (text pieces, aerial photos, maps, etc.). The knowledge recovery effort was very costly in terms of time and effort, suggesting that knowledge-based organizations should attempt to make knowledge capture an on-going aspect of work, rather than finding themselves in the situation of needing to recapture knowledge. A second aspect of the work reported here involved the need to support navigation among the many Concept Maps in ROCK-TA.

1 Introduction

Terrain analysis is a formal systematic process in which aerial photographs, topographic maps, and other data sources are used to understand terrain and then apply that understanding, especially for engineering and land use (e.g., engineering properties of soils and rocks, waterways management, land-use planning, etc.). Expertise at terrain analysis is achieved only after years, if not decades of experience at interpreting aerial photographs (Hoffman, 1985; Hoffman & Pike, 1995). Therefore, the capture of terrain analysis expertise is of great potential value, for knowledge sharing (e.g., training) as well as knowledge preservation.

Expert-level knowledge about terrain and terrain analysis exists in the form of traditional hardcopy documents. As such, the information is hard to access and is of limited usefulness and usability. It therefore seemed an appropriate topic for a study in the process of knowledge recovery, in which the information is recast in the form of Concept Maps.

2 Method: Creating the Knowledge Model

2.1 Materials

The Terrain Analysis Data Base (TADB) (Hoffman, 1983) is a corpus of about 1500 property and association statements about terrain. These were derived by documentation analysis and structured interviews with expert terrain analysts at the US Army Corps of Engineers Engineering Topographic Laboratory. The TADB involved representing domain concepts and propositions as hierarchical lists, for two reasons: 1). The domain itself is largely one involving hierarchical classification, and 2). The TADB entries were ordered because they were intended to be an intermediate representation for possible applications in knowledge bases for rule-based expert systems. Thus, the TADB entries were of a form readily transformable into Concept Maps. An example of a TADB entry is:

- Landforms
    - Fluvial Landforms
        - Alluvial Fans
            - Fan-shaped deposits
            - Gentle slope
            - Rugged upstream boundaries
            - Box-, v-gullies with steep gradients
            - Usually in arid regions

2.2 Procedure

The first step was to transform the TADB entries into propositions. In most cases, this involved the need to modify the wording of the TADB statements, most of which had an implicit propositional structure. Thus, for example, the Concept Map about Alluvial Fans included the following propositions:

Alluvial fans have morphology
Morphology includes fan-shaped deposits
Morphology includes gentle slope
Morphology includes rugged upstream boundaries
Alluvial fans have drainage pattern
Drainage pattern is box- and V-shaped gullies
Gullies have steep gradients

As another example, some high-level information about Deserts was represented in the TADB as:

- Deserts
  - Sand and gravel soils
  - Usually arid climate
  - Desert varnish tones
  - Occasional silt soils
  - Usually with dunes

This same information was represented as propositions and then represented as a Concept Map that appears in Figure 1, below. Note that this preserves the semi-heirarchical structuring of the TADB: Higher order concepts are nearer the top of the map, and lower order concepts are nearer the bottom of the map.

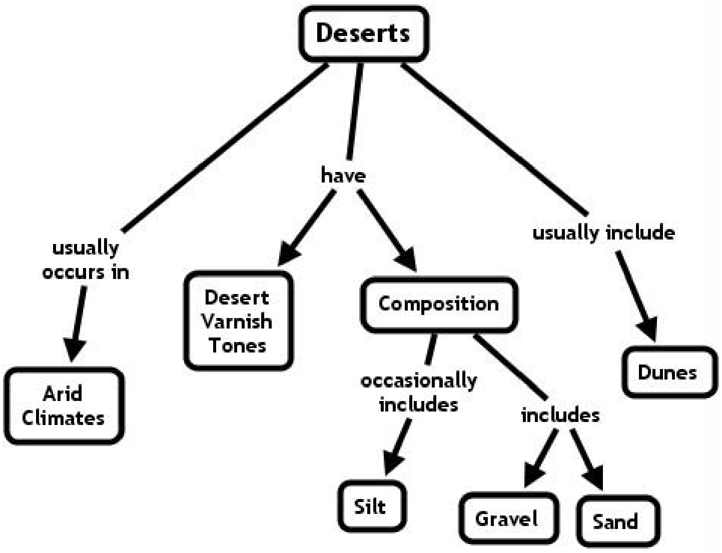

Figure 1. An example Concept Map representing propositions from the TADB.

Once the draft Concept Maps had been created, a series of structured interviews was held. The first and fourth authors of the present Report served as facilitators. The first author served as the local expert and also served as the Mapper. Working together, they reviewed the TADB and compared it to the draft Concept Maps. Refinements were made to the Concept Maps. A triple-check was conducted to insure that all of the statements in the TADB were included as propositions in the Concept Maps.

Another important resource for our work was the "Terrain Analysis Procedural Guide FOR Surface Configuration" (Mintzer & Messmore, 1984). This includes a great many photo interpretation keys for landforms of all types, including topographic maps and aerial photographs. These interpretation keys and aerial photos were used as multimedia resources for the final Concept Maps.

3 Results

The knowledge model consisted of 150 Concept Maps, containing a total of 3,341 concepts, 1,634 link labels, and 3,352 propositions. The Concept Maps contained an average of about 22 propositions. It took a total of between 187.5 person-hours and 225 person-hours to draft all of the 150 Concept Maps. This equates to an average of 75-90 person minutes to create each of the draft Concept Maps. (Detailed notes of time and effort were not made during the initial drafting of the first dozen Concept Maps. The lower and upper bounds we present here were based on estimates for that first wave of map drafting, but detailed notes of time/effort that were taken in all of the subsequent work.) It took 33 person-hours to refine the draft Concept Maps. This means that the process of recovering the TADB information in the form of propositions in Concept Maps took a total of between 220.5 and 258 person-hours.

Given that the final set of Concept Maps contained 3,352 propositions, the ROCK-TA Concept Maps were drafted and refined at a rate of between 0.25 and 0.22 propositions per task minute. This result falls considerably short of the benchmark for effective knowledge elicitation (Hoffman, Shadbolt, Burton, & Klein, 1995). The yield of 0.2 propositions per task minute is roughly equivalent to that for knowledge elicitation via protocol analysis, which is believed to be the *least* efficient method for eliciting expert knowledge (see Hoffman et al., 1995). For comparison purposes, Concept Mapping knowledge elicitation (without any knowledge recovery step) has a yield of about 2 propositions per task minute (seer Hoffman et al., 2000).

4 Implication of the Results

The meaning of this effort calculation—a yield of about 0.2 propositions per task minute—should be considered in light of the fact that the participating local expert who had helped create the TADB happened to also be a proficient Concept-Mapper. No doubt, the refining of the Concept Maps would have taken even longer had this not been the case, and the yield of propositions even less than 0.2 per minute. Furthermore, our effort leveraged the fact that the TABD entries were already of a form readily transformable into Concept Maps. Were this not the case, that is, if the TADB had been in the form of traditional text, then the knowledge recovery process would have taken much longer, arguably perhaps as much as an order of magnitude longer.

In light of these considerations, the moral of our results is clear for any organization that in confronting issues involving the loss of expertise (see Hoffman & Hanes, 2003): *Knowledge recovery is costly.* Any knowledge-based organization is better off capturing knowledge in a usable and useful form as a part of the organization's on-going knowledge management program, rather than finding itself in a position of losing expertise because of retirement, or the "wasting" of knowledge because it is archived in older forms of media (i.e., hardcopy text) that are not easily compatible with newer hypermedia forms and formats.

Having created ROCK-TA, we realized that the 150 Concept Maps would benefit from having multiple support tools for navigation. This was a focus of a second effort.

5 Creating the Navigator

Concept Map knowledge models depend on a method to support navigation. This has involved appending CmapTools icons to concept nodes (e.g., Hoffman, et al., 2000). Given the size of ROCK-TA, it was decided that it would be prudent to exercise a known principle of cognitive science: In information search and perception, redundancy is helpful. This idea was manifested in our adoption of redundant methods for navigating among the ROCK Concept Maps, and leveraged the fact that the ROCK-TA Concept Maps were arranged hierarchically in accordance with the hierarchical arrangement of the TADB.

The ROCK-TA Navigator included a number of mechanisms, described in Table 1.

The top concept node in each Concept Map, which bears the title of the Concept Map (i.e., its main topic), had appended to it a Cmap resource icon that would take the viewer up one level in the hierarchy.
A high-level Concept Map was created, called "The Representation of Conceptual Knowledge in Terrain Analysis." This Concept Map provides some important high-level explanatory concepts ("landforms," "climate," "soils," etc.) that form a general description of the major areas of knowledge in terrain analysis. This Concept Map serves as the user's gateway into the knowledge model. From it, the user can go to the Navigator or to the Top Maps.
A "Map of Maps" was created. A Maps of Map has nodes for all of the Concept Maps, and shows how each Concept Map is related to the others in the overall hierarchical scheme.
"Top Maps" were created to cluster Concept Maps into meaningful groups. Thus, for example, all of the Concept Maps about various types of sand dunes were linked together in a Top Map about Dunes. Like all Concept Maps, Top Maps display higher-order concepts (in this case, nodes depicting other concept Maps) nearer the top, and lower-order concepts nearer the bottom. The user can open a Top Map and use to see the overall relations between higher order concepts and use it to navigate between the subsumed Concept Maps, because double clicking a node opens the relevant Concept Map. Third, where appropriate, Top Maps were themselves organized under Subsuming Top Maps. For instance, the Dunes Top Map was subsumed under the Aeolian Landforms Top Map, which was itself subsumed under the Landforms Top Map.
"Maps of Top Maps" were made to show how all of the Top Maps relate to one another.
The top concept node in each Concept Map was hyperlinked to the "The Representation of Conceptual Knowledge in Terrain Analysis" Concept Map.
A "ROCK-TA" concept node was added in the upper left corner of each Concept Map. From this node, one would be able to navigate to the Map of Top Maps or the Map of Maps.
A Navigator "Cmap Piece" was added at the left-hand side of every Concept Map. This Cmap Piece showed how the given Concept Map fit into its subordinating hierarchy of Top Maps. Using the Navigator Cmap Piece, users can navigate up to superordinate Top Maps.
Once all the Navigator Cmaps and Top Maps had been created, all of the Concept Maps were hyperlinked. For example, the map "Terminal Moraine" describes key concepts associated with the moraine deposited by glaciers, one of which is the concept of "depressions." A separate Concept Map, called "Depressions" describes depressions in greater detail.

Table 1. Mechanisms of the ROCK-TA Navigator

The goals for including these multiple redundant navigational tools were: a). To allow the user to get from anywhere in the knowledge model to anywhere else, in two clicks at most, and b). To allow the user to always be able to see, know, and keep track of where s/he is in the knowledge model. Figure 2 presents an example of a subsuming Top Map. In ROCK-TA, all Top Maps have a colored background. The Top Map in Figure 2 example points to a number of subsumed Top Maps (colorized). The uncolored nodes refer to Concept Maps.

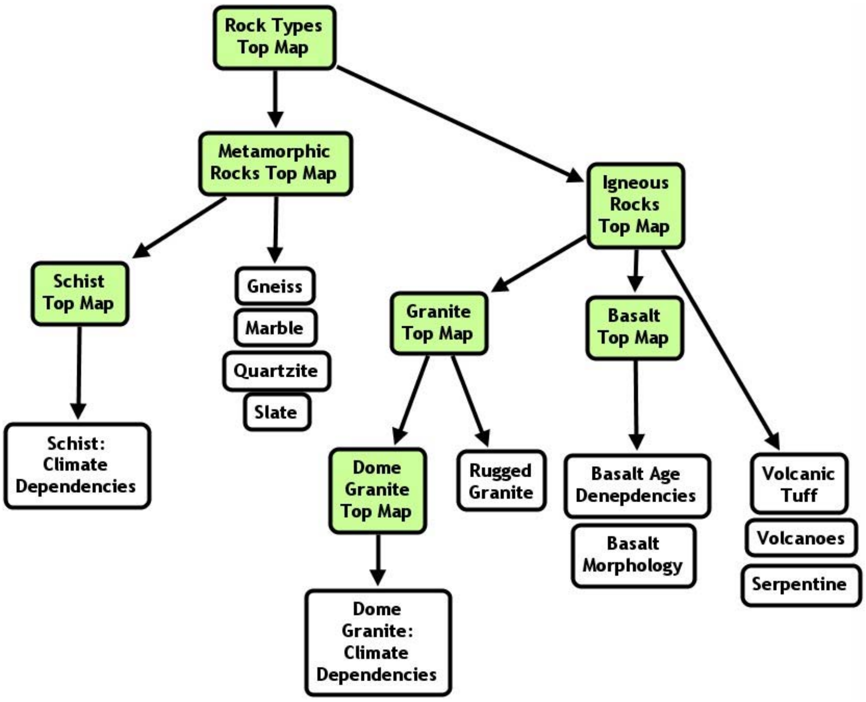

Figure 2. An Example Top Top Map in ROCK-TA.

Figure 3 presents an example of a Top Map that contains its Navigator Cmap Piece, on the left-hand side of the view. In such a Cmap Piece, the lowest node represents the Top Map that is immediately up in the hierarchy, and each node above that represents the next levels up in the hierarchy. The subordinating relationships of the nodes in the Navigator are indicated by arrows. Clicking on any of Cmap resource icons appended to the nodes in the Navigator opens the relevant Top Map. In Figure 3, the concept "Deserts" is subsumed by the higher order concept "Aeolian Landforms."

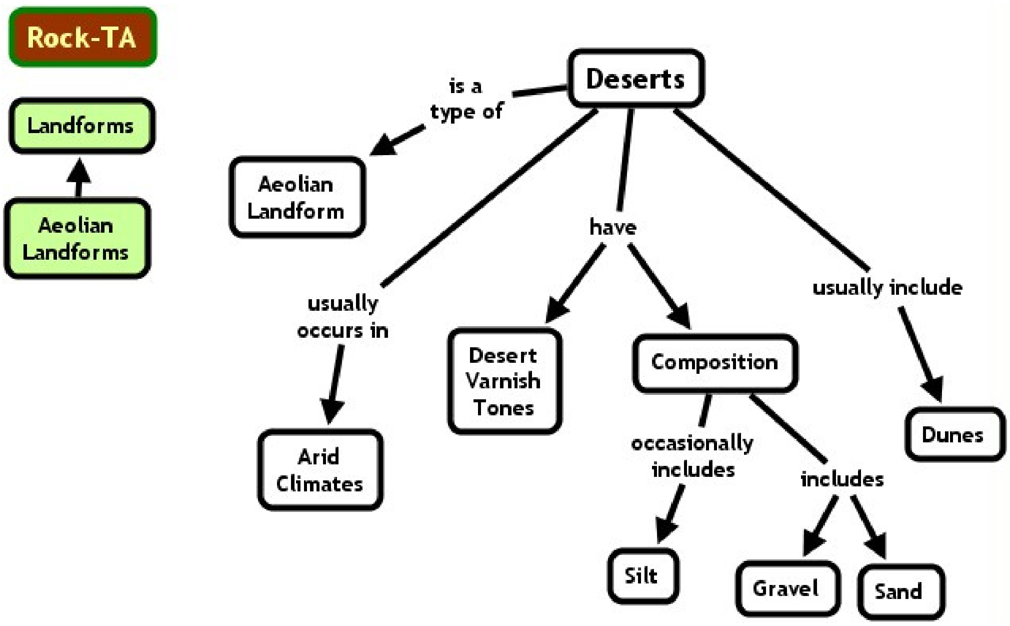

Figure 3. An example of a Concept Map that includes its Navigator.

A rule of thumb in Human-Centered Computing (HCC) is that if a display or interface system requires navigation, and especially if the navigation system has to be explained, then the system will create make-work for the user, no matter how “user friendly” it is intended to be. It is important to keep in mind, however, that HCC regards the human-machine as a *system*. ROCK-TA is intended for use by individuals who are already accustomed to working with CmapTools knowledge models. In other words, the intended users belong to a new culture for knowledge sharing. Thus, for the human-machine system that we envision, explanation of the Navigator should not be an issue: It should explain itself.

6 General Discussion

It took over 366 person-hours to create ROCK-TA. (This includes the time taken to append dozens of various multimedia resources, such as example aerial photographs.) As we have explained, the knowledge recovery step was time-consuming and relatively inefficient compared to other knowledge elicitation procedures (see Hoffman, et al., 1995). On the other hand, the result is a knowledge model that includes all of the propositions from the Terrain Analysis Date Base. This single compilation, transferable to CD, contains the collective experience, skill, and guidance from of some of the world's leading terrain analysts, most of whom retired not long after they created the documents and other sources upon which we relied.

In its current form, ROCK-TA might be most useful as a teaching tool or learning aid in courses on terrain analysis. It might also be a useful resource for professionals to have on hand. ROCK-TA contains generic information about terrain. “Daughters of ROCK" might be specifically tailored for local needs, such as the representation of information about terrain in particular regions. We see such daughters of ROCK-TA as relying on ROCK-TA for their backbone of domain knowledge, but going into more detail for more specific regions and purposes.

7 Summary

A neglected aspect of knowledge management and the preservation of expertise is “knowledge recovery.” Expert-level knowledge in many domains has already been elicited and preserved, but exists in the form of traditional text documents. This makes the material difficult to access and difficult to understand and use. Knowledge recovery is the process of culling meaningful propositions from documents and crafting them in the form of Novakian Concept Maps, thereby making the material more useful and usable. This report discusses our first major effort at knowledge recovery. Expert knowledge about terrain exists in the form of traditional documents discussing landforms, soils, rock types, etc. (Hoffman, 1983; Mintzer & Messmore, 1984). Previous IHMC research (e.g., Hoffman, Coffey, & Ford, 2000) had demonstrated that Concept Mapping could be used as a knowledge elicitation procedure and could result in a model of expert domain knowledge that captures relations among myriad elements of knowledge (Novak, 1998). Details of the process of creating ROCK-TA ("Representation Of Conceptual Knowledge-Terrain Analysis") are provided in the present Report. ROCK-TA is a very large knowledge model, consisting of over 150 Concept Maps, containing over 3,000 propositions, and dozens of multimedia resources (text pieces, aerial photos, maps, etc.). ROCK-TA is intended as both an instructional tool and a performance support tool for practitioners who have to engage in terrain analysis. A major finding from this effort concerns the process of knowledge recovery. Our participant expert at terrain analysis was involved in the creation of the original source documents on terrain analysis, and also was a proficient Concept Mapper and participated in the creation of the Concept Maps. This should have made the knowledge recovery process as efficient as possible. In fact, the effort was very costly in terms of time and effort. This result clearly suggests that knowledge-based organizations should attempt to make knowledge capture an on-going aspect of work, rather than finding themselves in the situation of needing to recapture knowledge. A second aspect of the work reported here involved the need to support navigation among the many Concept Maps in ROCK-TA. Our approach involved the notion of a “Cmap Piece.” A Cmap Piece is a small Concept Map that appears within the window of each main Concept Map and allows the viewer to always be aware of where they are in the knowledge model. A demonstration of ROCK-TA will show how the Navigator works.

8 Acknowledgement

We prepared this paper through participation in the Advanced Decision Architectures Collaborative Technology Alliance, sponsored by the US Army Research Laboratory under cooperative agreement DAAD19-01-2-0009.

9 References

Hoffman, R. R. (1983). "Methodological preliminaries to the development of an expert system for aerial photo interpretation." Report, Engineering Topographic Laboratories, US Army corps of Engineers, Ft. Belvoir, VA.

Hoffman, R. R. (1985). "Symbolic image interpretation techniques." Report, Engineering Topographic Laboratories, US Army Corps of Engineers, Ft. Belvoir, VA.

Hoffman, R. R., Coffey, J. W., & Ford, K. M. (2000). "A Case Study in the Research Paradigm of Human-Centered Computing: Local Expertise in Weather Forecasting." Report on the Contract, "Human-Centered System Prototype," National Technology Alliance.

Hoffman, R. R., & Hanes, L. F. (2003/July-August). The boiled frog problem. *IEEE Intelligent Systems*, pp. 68-71.

Hoffman, R. R., & Pike, R. J. (1995). On the specification of the information available for the perception and description of the natural terrain. In P. Hancock, J. Flach, J. Caird, & K. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems* (pp. 285-323). Mahwah, NJ: Erlbaum.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. A. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, *62*, 129-158.

Mintzer, O., & Messmore, J. A. (1984). "Terrain analysis procedural guide for surface configuration." Technical report ETL-0352. Engineer Topographic Laboratories.

Novak, J. D. (1998). *Learning, creating, and using knowledge*. Mahwah, NJ: Erlbaum.