Automatically-generated Concept Maps as a Learning Tool

Xia Lin

Abstract—Concept maps can be generated manually or automatically. It is important to recognize differences of the two types of concept maps. The automatically generated concept maps are dynamic, interactive, and full of associations between the terms on the maps and the underlying documents. Through a specific concept mapping system, Visual Concept Explorer (VCE), this paper discusses how automatically generated concept maps are different from manually generated concept maps and how different applications and learning opportunities might be created with the automatically generated concept maps. The paper presents several examples of learning strategies that take advantages of the automatically generated concept maps for concept learning and exploration.

Keywords—Concept maps, Dynamic concept representation, learning strategies, visual interface, Visual Concept Explorer.

I. INTRODUCTION

CONCEPT maps are conceptual graphs of nodes and links for concept organization and representation [1]. The construction of concept maps is typically done either manually or automatically. When the concept maps are created manually, the maps are most often used to represent how a user or users understand a topic or a domain of relation concepts. The construction of such concept maps is a process of learning and discovery where users add more concepts and links to the maps as they learn and understand more about their studying topics or domains. A good evaluation of such maps is to compare the maps that the user generated before and after a study (such as a course or a set of learning activities) or to compare student's maps with teacher's maps. When the maps are constructed automatically, the maps are generated from a body of documents related to the topics or domains that the concept maps represent. The construction process is focused on computational learning, i.e., using computer's learning algorithms to extract and learn about concepts and their relationships from the underlying documents and represent them in a graphical form. While there have been significant numbers of research on using manually constructed concept maps for teaching and learning [2-4], there is little research on applying automatically constructed concept maps for concept learning and exploration.

Xia Lin is with College of Information Science and Technology, Drexel University, Philadelphia, PA 19104 USA (215-895-2482; fax: 215-895-2494; e-mail: xlin@drexel.edu).

It is important to recognize differences between the manually and automatically constructed concept maps and develop different applications for them. In this article, we first describe a specific automatic concept map application and then explore several different learning strategies that might be applied to the use of automatically generated maps for learning and concept exploration.

II. VISUAL CONCEPT EXPLORER

Visual Concept Explorer (VCE) is a visualization prototype developed to explore potentials of concept mapping and exploration. Traditionally, in library and information science, the conceptual space of a domain is defined by subject thesauri or controlled vocabularies [5]. Typically, a group of domain experts selected relevant terms and built hierarchical term relationships and cross-references to provide a comprehensive coverage of the domain. Terms included in the controlled vocabularies are later used by professional indexers to index documents to provide a consistent "concept space" of the domain. The controlled vocabularies might be considered as a huge "concept map" that was built "top-down" without a corresponding graphical or visual form, whereas concept maps are "bottom-up" representing individual concepts and their relationships in visual forms. VCE is an attempt to take advantage of the traditional controlled vocabularies to create instant concept maps through visual mapping and learning algorithms. In this section, we describe the three major components of the VCE system, the knowledge base, the mapping algorithms, and the visual interfaces.

A. The knowledge base and the conceptual space

Implemented as a client-server application, VCE contains a Java-based Web service that interacts with a very large ontology, the Unified Medical Language System (UMLS) as the Knowledge Source and a live search engine, the free PUBMED search engine available on the Web [6].

UMLS is created and managed by the U.S. National Library of Medicine. It includes a very large vocabulary database, the metathesaurus, which contains information about biomedical and health related concepts, their various names, and the relationships. The metathesaus links concepts of many different medical thesauri and classifications such as Medical Subject Headings (MeSH), Thesaurus of Psychological Index terms, Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT), etc. Because of linking in the metathesaurus, concept relationships shown in one source of

vocabularies can be easily translated to concept relationships of another vocabulary.

The metathesaurus also include a table of co-occurrence counts of all the MeSH terms in the Library of Medicine's MEDLINE database. Such data are an excellent source for studying co-occurrence-based concept relationships. Co-occurrence analysis is one of the popular methods used for extracting semantic relationships of document collections. The underlying assumption of using co-occurrence analysis is that if two items occur often in the same documents, there must be some relationship between the two items. Many research has shown the co-occurrence term relationship can be used to approximate semantic relationships of concepts and create concept spaces [7-8]. Based on this assumption, VCE extends the "concept spaces" created in MeSH and UMLS through co-occurrence mapping of concepts.

VCE is built to be flexible to connect to different databases with thesaurus, often through a new customized web service. Currently, VCE can also be used with many different DIALOG databases with thesauri such as ERIC, PSYINFO, INSPEC, and ABI/INFORM.

B. The mapping algorithms

Mapping concept relationships to a visual graph and providing interactive functions for the user to explore and understand the visual representation of concept relationships is not a trivial task. It is essentially a task of learning the complex term relationship in a high-dimensional concept space and simplifying them to a lower dimensional visual space. Many dimensional reduction algorithms have been tested for such a task. Two of the most popular algorithms, which are described below, are adopted and implemented in VCE.

The first is an unsupervised learning algorithm of artificial neural networks, Kohonen's self-organizing feature map (SOM) [9]. As a dimension-reduction technique, SOM takes a set of input objects, (such as a matrix comprising N-dimensional vectors of co-occurrence counts), and maps them onto nodes of a two-dimensional (2D) grid. Relationships of the data in the high dimensional space are maintained as much as possible within the local neighborhood relationships of the 2D display. The algorithm groups several nodes into a display region when their responses to input data are similar. Adjacent regions indicate relatively strong co-occurrence relationships, and distant regions imply weaker co-occurrences. As the result, an orderly map display is generated to present high order statistics of terms and their relationships. SOM has been applied to map document and term relationships in many of the research projects [10].

The second mapping algorithm implemented is the Pathfinder Networks (PFNETs), originally created by cognitive psychologists [11]. PFNET is a computational procedure to devise a simplified network model for a set of inter-related data. Similar to Multidimensional Scaling techniques, PFNETs examine all the data relationships and create paths that mark the most efficient connections. As the result, the data and data relationships are represented as a graph of nodes and links that direct the user's attention to the best "associative paths," thus most similar to the manually generated concept maps.

C. The interface

VCE implements a visual interface to let users explore concept maps generated automatically. The earlier version of the interface was implemented in Java and the current version is implemented in Adobe FLASH. Fig. 1 shows an example of the interface for the term "cognition." The interactive functions of the interface will be described in later examples. One can also go to the web link: http://cluster.cis.drexel.edu/vce to test the interface interactively.
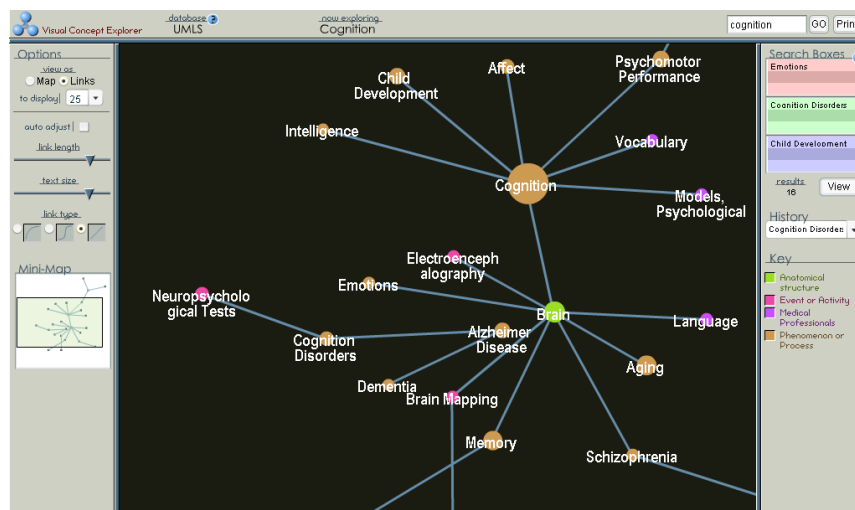

Fig. 1 An example of automatically generated concept maps

III. Dynamic concept representation

VCE provides a platform to experiment how to derive concept maps automatically based on term semantic relationships in documents. It represents a new approach to concept representation – dynamic concept representation [15]. When a concept can be presented dynamically upon user's request, how do we interpret it? What does the representation show us? Do we expect a "true" representation of the concept regardless of the mapping algorithms and the underlying document collections used?

There are many open research questions here. There are a series of assumptions that need to be tested empirically:

- The meaning of a concept can be understood through the concept's relationships with other related concepts.
- The meaning of a concept is dependent on the context and use in a document collection within a specific domain.
- A concept can be represented through the mapping of semantic relationships of related concepts.
- A concept can be represented through revealing how the concepts are used in a domain or in a document collection.

To a large extent, Gardenfors' work on conceptual spaces [12] has provided excellent discussions on these assumptions. Gardenfors argued that, in addition to the two most popular cognitive science' representation modeling methods, i.e. the symbolic approach and the associationism approach, the conceptual approach based on geometrical structures might be the most promising approach to modeling concepts and concept relationships.

There are also many other projects that have worked on building meanings or presenting concept structures through visual representations [13-14]. More discussion on dynamic concept representation can be seen in [15]. What we are focused on here is comparing the dynamic, and interactive concept maps with manually generated concept maps.

Table 1 is a summary of a comparison of their characteristics and features. In general, the manually created concept maps need significant time to create; the maps are often created for special purposes; and the maps are a result of meaningful learning. On the other hand, the automatically generated maps can be created or re-created instantly upon user's request; the maps can be scaled large or small as needed; the users can interact with the maps and request new mapping; and the maps are a result of a dimensional-reduction mapping based on how the concepts are used in documents. While the manual maps represent concept relationships in the way how the creator understands them, the automatic maps represent concepts within the context of their use in documents.

IV. LEARNING STRATEGIES WITH THE AUTOMATICALLY-GENERATED CONCEPT MAPS

Although one single, well crafted, and static concept map can provide very rich information for the user to study, digest, and understand concept relationships, the automatically generated concept maps provide much more options for the users to explore and understand concept maps. In this section, we use a number of examples to demonstrate several learning strategies that can be applied to dynamic concept maps but may not be easily applied to manually generated concept maps.

A. Learning through interactive exploration of concept maps and the underlying documents

For concept maps to be generated automatically, the system needs to be fully integrated with a knowledge base or document retrieval system. In the case of VCE, the interface serves as both a concept exploration tool on top of UMLS and a retrieval interface for PUBMED search engine. This tightly integration will allow users to get support directly from relevant documents when they explore relationships of the concepts as demonstrated in the following example.

When a user enters a term to VCE, in this case, "cognition," the system quickly generates a concept map for the matching concept based on the co-occurrence data and learning algorithms. The display shown in figure 1 is generated by the PFNET algorithm. Now the user can see how the "cognition" is linked to "brain," "intelligence," "vocabulary," "Models, psychological", etc., and the term "brain" is connected to "language," "emotion," and "electroence phalography," etc. The user can explore how the terms related to each other or select multiple concepts to retrieve documents. The user can also double-clicking on a concept to regenerate a new concept map focusing on the concept. In this example, the user starts with a query term "cognition", and selects "Emotions" and "cognitive disorders" and "child development" to construct a Boolean query. All the three terms are selected from the mapping of "Cognition" and its related terms.

This example shows that the concept maps can guide users in search query construction. Users will be able to clink on terms to construct complex queries and select multiple links to retrieve documents related the linking concepts. Ideally, when they read documents, they can highlight a term to see a new concept map. This feature has not been implemented but we have seen the advantage of the tight integration of document retrieval systems and concept maps.

B. Learning through multiple displays of the same concept

When the computer can generate a concept map instantly, how the user interacts with the concept maps will be very differently. In VCE, the users can zoom in or zoom out to see different part of the maps; they can request a series of related maps and compare their differences; they can request to see different styles of maps; and they can compare maps generated by different algorithms for the same concept. The last one is perhaps most significant as different maps expose different perspectives for users to understand the concept relationships. A series of empirical research have been done in comparing how the users interact and understand different

maps of the same concept, notably in two dissertations completed in our university [16 -17]. It is found that users will receive new information when they switch from one kind of maps to the other. In the case of VCE, users can use both Kohonen's map displays and PFNET network displays for searching and concept learning equally well. While novice users (newcomers to a domain) prefer the network displays, domain experts slightly prefer map displays as they can get better understanding of concept relationships on the maps.

As a learning strategy, the user should be allowed to select or switch from one display to other as they explore the concept maps. VCE allows the user to see simple text hierarchical displays, network displays, and map displays and all the displays provide similar interactive functions. Figure 2 is an example of the concept maps in the network display form and in the map display form.

C. Learning by mapping of the same concept in different domains

VCE can be connected to different document resources or search engines to do mapping in different domains. This provides a new way of concept exploration. For example, how do we understand the concept "visualization"? There are dictionary definitions, such as "to formal a mental image" or "to make visible" (http://www.dictionary.com). There are also definitions from thesauri, such as "act or power of forming mentally visual images of objects not present to the eye" (ERIC thesaurus, http://www.eric.ed.gov/ thesaurus/). We can also try to understand the concept of "visualization" from the psychology perspective, educational perspective, engineering perspective, and many others. To get a comprehensive understanding one might need to read hundreds or thousands of documents in order to summarize the different perspectives.

VCE can provide mappings from different perspectives. Figure 3 shows four concept maps of the concept "visualization" generated from four different document databases, INSPEC, PSYCINFO, ABI/INFORM, and ERIC. The map generated from INSPEC, a database of scientific and technical literature, displays concepts (and their relationships) mostly related to technical approach to visualization: the four main areas of related concepts include: Data visualization (data analysis and data mining), Engineering graphics (rendering, animation, and virtual reality), Image processing (image reconstruction, image segmentation, computerized tomography) and Programming (software tools, visual programming, object-oriented programming, and graphical user interface). These four areas represent a good summary of the technical approach to "visualization."

Similarly, the map generated from psycINFO, a database of psychology literature, shows that the concept of "Visualization" is related to "cognition," "brain," "education," and "technology," whereas the map generated from ABI/INFORM, a database of business literature, reveals four areas of "visualization" related to business: "decision-making," "information-management," "search engines", and

"CAD (technology)." Again, these are high-level summaries of how the concept "visualization" is mentioned in different domains of literature. Much more detailed concept relationships are shown in these maps. The user can choose different maps to explore the concept and the underlying documents interactively. All the four maps shown are generated by the Kohonen self-organizing map algorithm; the user can also select different mapping algorithms to display the maps in different formats.

V. FUTURE RESEARCH AND DEVELOPMENT

In this paper, we presented a visual concept mapping system to let users interact with the automatically generated concept maps. The focus of the discussion is how such concept maps are different from manually generated concept maps and how different applications and learning opportunities might be created once we have the automatically generated concept maps. It is important to recognize different potentials of automatically and manually generated concept maps. We presented some examples of new learning strategies that take advantages of the automatically generated concept maps. Much more research needs to be done in this area. The VCE system is currently being extended to make it easy for users to select mapping methods, view different styles of mapping results, access different collections for mapping, and view multiple concept maps on the same screen. More empirical studies will also need to be done. As the system is further improved, VCE will become a good vehicle to experiment and learn how to generate concept maps according to the user's needs and how to utilize the automatically generated concept maps for concept learning and exploration.

REFERENCES

[1] Novak, J. D. & Cañas, A. J. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition. Available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

[2] Anderson, J. R., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological Review, 111(4), 1036–1050.

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

[4] Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. Journal of Research in Science Teaching, 31(1), 91-101.

[5] Lancaster, F. W. (1986). Vocabulary Control for Information Retrieval (2nd ed.). Arlington, VA: Information Resources Press.

[6] UMLS (2007). Unified Medical Language System. http://www.nlm.nih.gov/research/umls/

[7] Chen, H. Schatz, B., Yim, T., & Fye, D. (1995). Automatic Thesaurus Generation for an Electronic Community System. Journal of the American Society for Information Science, 46(3), 175-193.

[8] Schatz, B.R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net. Science, 275(5298), 327-334.

[9] Kohonen, T. (1997). Self-Organizing Maps. Springer-Verlag, Berlin.

[10] Honkela, T.; Kaski, S.; Lagus, K., Kohonen, T. (1998). Self-organizing maps of very large document collections: Justification for the WEBSOM method. In Balderjahn, I., Mathar, R., and Schader, M., editors,

Classification, Data Analysis, and Data Highways, pages 245-252. Springer, Berlin.

[11] Schvaneveldt, R. W. (Ed.). (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

[12] Gardenfors, P. (2004). Conceptual spaces: The geometry of thought. MIT Press.

[13] Takayama, Y.; Flournoy, R.; Kaufmann, S. (1998). Information mapping: Concept-based information retrieval based on word associations. Unpublished technical report. Stanford University.

[14] Widdows, D. (2004). Geometry and meaning. CSLI Publications, Stanford University.

[15] Lin, X.; Aluker, S.; Zhu, W.; & Zhang, F. (2006). "Dynamic Concept Representation through a Visual Concept Explorer." Advances in Knowledge Organization, Vol. 10; Proceedings of the Ninth Conference of the International Society of Knowledge Organization (ISKO 2006, July 4 – 7, Vienna, Austria), pp. 367-374.

[16] Buzydlowski, J. (2002). A Comparison of Self-Organizing Maps to Pathfinder Networks for the Mapping of Author Co-Citation Analysis. (Unpublished Ph.D. dissertation, October, 2002, Drexel University).

[17] Williams, J. (2007). The Impact of Concept Map Visualizations on the Information Behavior, Perceptions of Performance, Learning and Use of Novices in the Information Retrieval Context. (Unpublished Ph.D. dissertation, May, 2007, Drexel University).
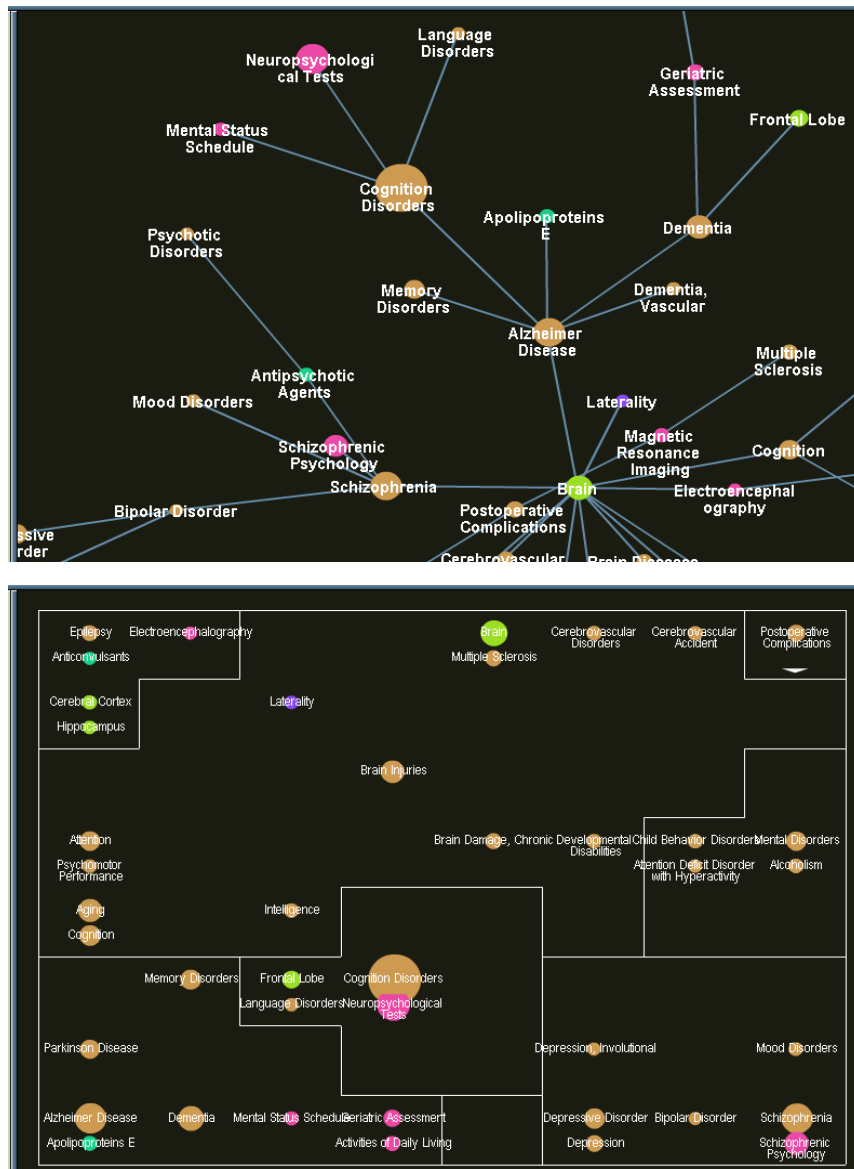
Fig. 2 Two different styles of concept maps for "Cognitive Disorders"

(a) The concept map of "Visualization" in INSPEC, a database of scientific and technical literature.

(b) The concept map of "Visualization" in psycINFO, a database of psychology and related behavior and social sciences.

(c) The concept map of "Visualization" in ABI/INFORM, a business-oriented literature database

(d) The concept map of "Visualization" in ERIC, a database of educate-related literature

Fig. 3 Mapping the same concept in different domains