doi: 10.1111/j.1365-2729.2007.00230.x

Organizing learning materials through hierarchical topic maps: an illustration through Chinese herb medication

B.-J. Shih,* J.-L. Shih† & R.-L. Chen‡

*Department of MIS, Tajen University, Pingtung, Taiwan
†Department of Educational Technology, Tamkang University, Taipei, Taiwan
‡Tajen University, Pingtung, Taiwan

Abstract

This research aims to use hierarchical topic maps to compile digital learning material and to discuss its design and application possibilities. The system renders tremendous original assets and then embeds a self-organizing map (SOM) in the material database to produce topical learning materials, as in this case, an illustration through Chinese herb medication. It helps to demonstrate robust professional information as well as knowledge structures, and provides a customized and interactive learning dynamic to support both progressive and constructive learning styles.

The paper first gives a detailed procedural description of the material construction, explains how topic map techniques were applied, and observes the implications and potentials of the technology to education.

Both the technical and educational evaluations of using SOM topic maps in compilation of learning materials have resulted in positive feedback. SOM allows users to review the complete databank in structural hierarchical order, which provides comprehensive understanding of the entire set of learning materials, and also brings opportunities to users to discover knowledge related to their study area.

Keywords Chinese herb medication, digital learning material, self-organizing map, topic map.

Introduction

Finding the best way to effectively induce useful knowledge from the inexhaustible resource mine is a big challenge for most teachers. With potentially hundreds of attributes to review for a course, it is hard for the instructor to have a comprehensive view of the information embedded in the transcript (Dringus & Ellis 2005). Computer-based systems have great potential for delivering learning material (Masiello *et al*. 2005), which frees teachers from handling mechanical matters so they can practice far more humanized pedagogical thinking. However, information comes from different sources embedded with diverse formats in the form of metadata, making it troublesome for the computerized programming to create professional materials.

A topic map seems to be a good solution. Fisher *et al.* (2000) said that a topic map presents a group of assets in a network structure; the knowledge network is completed by interlinked conceptual nodes, and spread out in a framework. It not only involves sets of concepts but also the organization of concepts in terms of their interrelationships. Topic maps have long been used in fields such as computer science, electronic engineering, and other information-related fields, and have been

Accepted: 2 February 2007
Correspondence: Ju-Ling Shih, 151, Yin-Chuan Road, Tamsui, Taipei Hsien 251, Taiwan. E-mail: juling@mail.tku.edu.tw

applied to the manipulation and presentation of data in many ways. However, their application to education has seldom been touched upon. Therefore, designing appropriate tools for teaching and learning is a feasible approach to reduce the barriers teachers might encounter when adopting technology in their teaching (Marx *et al*. 1998; Putnam & Borko 2000).

It now seems to be a new trend in medical education to implement Information and Communication Technology (ICT) (Govindasamy 2001) to support student learning that we see pharmacy is an appropriate field to experiment upon. This paper focuses on the material compilation of topic map technology for the education of Chinese herb medication. A detailed procedural description of the material construction is provided, followed by pedagogical application suggestions observing the implications and potentials of the technology to education. In the literature review section, the concept of knowledge management is introduced. The educational characteristics of Chinese herb medication are analysed to provide advance understanding of the nature and learning factors of the subject, which is useful for the development of digital learning materials using topic map techniques. Then, we give a full view of our experience conducting the project 'Shannon Sutra: Digital Learning Content and Experimental Research on Chinese Herb Medication Safety' supported by the National Science Committee in Taiwan, and explain how topic map techniques were applied. Both the technical and educational evaluations of using self-organizing map (SOM) topic maps in compilation of learning materials produced resulted with positive feedback. Our conclusion brings some ideas and suggestions to the alternative possible integration of technology into learning in all fields.

Literature review

It is seen that a key factor of hypermedia-based learning is customizable cognitive style as it suffices users' information processing habits, representing individual user's typical modes of perceiving, thinking, remembering and problem solving (Lee *et al*. 2005). From the perspective of learning theories, cognitive psychologists recognize that knowledge has a basic structure which presents the interrelationships between concepts. Anderson (1980) distinguished knowledge into declarative and procedural types to identify their characteristics as abstract or practical functions. In understanding knowledge, Ausubel (1968) provided two strategies including progressive differentiation and integrative reconciliation for making meaningful learning focusing on the systematic methods to learning. The implication of these theories is twofold: (1) knowledge has an internal structure to sustain itself; and (2) systematic retrieval and understanding of knowledge is an effective method for learning.

In order to effectively use technology to assist the education process, helping learners to collect, process, digest and analyse the information, we introduce data/text mining, topic maps, vector space representation, and SOM technology, describing how the technology facilitates the processing of data, and how the dynamic map can achieve the pedagogical goal.

The Web currently contains more than 2.5 billion pages, consisting of at least 10 terabytes of textual information (Roussinov & Zhao 2003). This 'Information Overload' problem is even more emphasized with the growing amount of text data in electronic form and the availability of the information on the constantly growing World Wide Web (Mladenic & Grobelnik 2003). The ability to access a large amount of resources is good, but the ability to locate specific, relevant and accurate information is poor. When a user enters a keyword, such as 'Herb Mediation', into a search engine, the returned result is often a long list of web pages, many of which are irrelevant, moved, or abandoned (Smith & Ng 2003). It is virtually impossible for a single user to filter out quality information under such an overloading situation.

In certain organizations, knowledge objects can be found in new contexts, regardless of potential manually assigned categories in groupware-based organizational memories. Therefore, new contextual knowledge can be gathered (Nonaka & Takeuchi 1995; Klemke 2000). Researchers and practitioners from various domains consider knowledge management crucial to the adaptation and survival of organizations in the face of continuous environmental changes (Malhotra 1998). The widespread availability of data mining software has given practitioners a variety of new alternatives to traditional, statistical data analytic techniques. These alternatives include several techniques based on concepts from machine learning, pattern recognition and neural networks.

Specifically, growing interest in the usage of machine learning on text data is especially stimulated by the intensive growth of the World Wide Web that can be seen as a widely accessible, large, distributed source of text data (Mladenic & Grobelnik 2003). To effectively process, manage, share and apply the huge amount of data has become a major task of knowledge management.

Generally speaking, there are two approaches to organize and present knowledge: top-down and bottom-up. Top-down approaches consist of supervised learning. They require human experts to define knowledge ontology (Noy & McGuinness 2001) and taxonomy (Bruno & Richmond 2003). Accordingly, the machine classifies the knowledge by those predefined rules, based on co-occurrence and Artificial Neural Network (ANN). Inversely, bottom-up approaches are fully automatic and represent unsupervised learning, including SOM (Kohonen 2001) and knowledge maps. However, most applications took hybrid approaches, that is, a human expert (SME: subject material expert) to predefine ontology and taxonomy and then clustering and organizing the knowledge by using ANN or SOM. This is sometimes called reinforcement learning (Bruno & Richmond 2003) and will be used in this project.

The key concepts of topic maps are topics, occurrences of topics, and relationships between topics (topic associations). First, a topic is a construct that represents a real world subject, and, in this sense, a topic can be almost everything: a theme, a concept, a subject, a person, an entity, etc. (Smolnik & Erdmann 2003). By applying topic maps to large sets of heterogeneous information resources, reusable structured semantic link networks are created on top of those resources. Furthermore, the standard provides navigation paradigms to enable users to search knowledge structures (Rath & Pepper 1999).

In traditional Chinese medication practice, patients actively looked for related information besides doctor's advice. Such information would be mostly distributed around the World Wide Web, or in magazines or books. This project targets the integration of diverse information into a comprehensive whole by using text-mining technology. For example, users can input a keyword, 'Gin-Seng', to find recipes with similar functions or clearly figure out the differences between recipes. Most important of all, users can discover unexpected patterns of information; especially those not noted on the textbooks by unsupervised ANNs. With proper (1) visualization (Yang *et al.* 2003), (2) summarization and (3) clustering (categorization), that information can be re-organized as learning material for medical practitioners and patients.

Lin *et al.* (1991) proposed to use SOM to generate topic maps and earlier topic maps were single-level plane maps, where topics are listed on the map without flexibility. Thus, many researchers have devoted themselves to improve the visualization (Chen *et al.* 1998; Yang *et al.* 2003) and labeling of topics on the map (Rauber & Merkl 1999; Dittenbach *et al.* 2000). Some topics of various domains have been well developed such as patent maps (Yoon *et al.* 2002) and web-browsing maps (Yang *et al.* 2003) based on hyperlinks. There are three major deficits: (1) those maps are static snapshots of data, hence they will not provide real-time updates; (2) to keep the integrity and consistency of data, human effort is unavoidable; and (3) after users clicked and jumped out of the map, information overload persisted. The detailed technical information of how to use SOM to generate topic maps is included in Appendix A.

Methodology and experiment

In the Shannon project, the researchers compile a hierarchical topic map of Chinese herb medication as e-learning materials. The developing process of topic maps is illustrated in Fig 1.

Retrieval of documents

The process includes selecting the most appropriate documents using a suitable criteria and the document processing to the eventual generation of the topic maps. First, the system retrieves documents from various sources (web pages, textbooks, reports, etc.) and stores them in a data warehouse. To ensure the learning material is up-to-date and accurate, we programmed a meta-search smart agent to automatically retrieve documents from 65 professional websites – which was previously filtered by experts of National Chinese Herb Medication Committee to ensure their correctness and coherence with the subject. The documents are then processed to

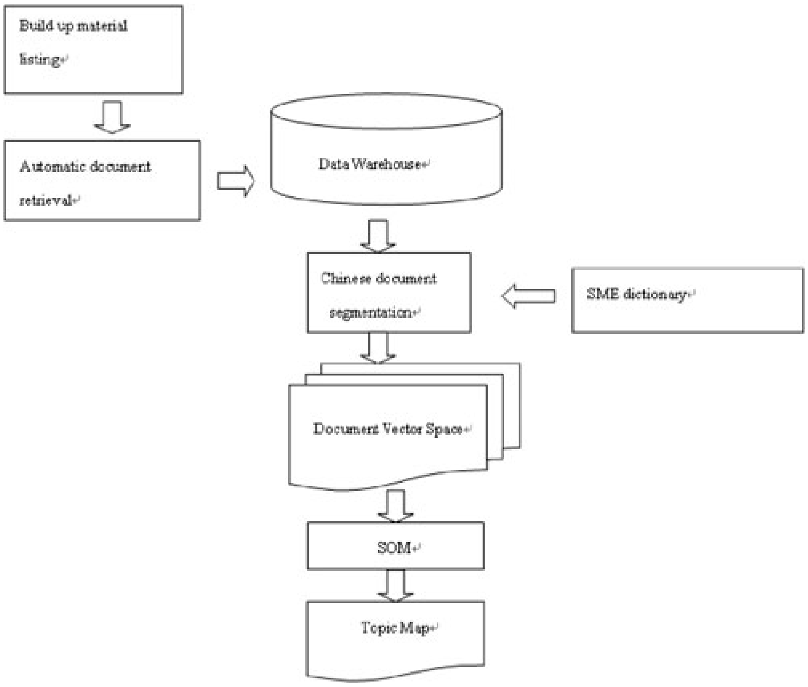

Fig 1 The development of topic map. SME, subject material expert; SOM, self-organizing map.

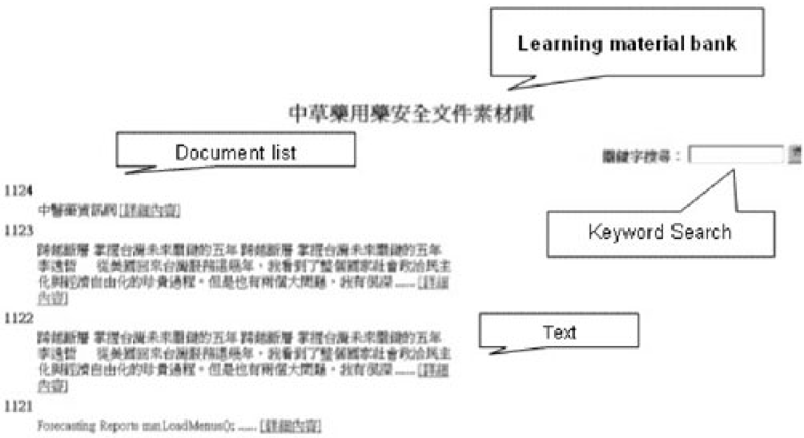

Fig 2 The learning material bank (English translation in notes).

be pure-text HTML (Hyper-Text Markup Language) and stored in the data warehouse for further queries (see Fig 2).

Document processing

The subject matter experts then built up a dictionary for our topic maps. The dictionary contains 100–1000 terms which are considered the most important concepts or materials to learn (see Fig 3, for this round, 625 terms are picked-up).

Next, word segmentation and vectorization of documents are done according to the dictionary. The number of documents is not fixed, and is normally between 10 000 and 100 000, because of frequent (daily, weekly, or monthly as needed) updates. The whole process could take more than 24 h on current single PC server.

After running the text-processing procedure, all the documents are clearly represented as vectors independently. The SOM engine automatically classifies the documents into several groups after training and then label those groups by topics according to their

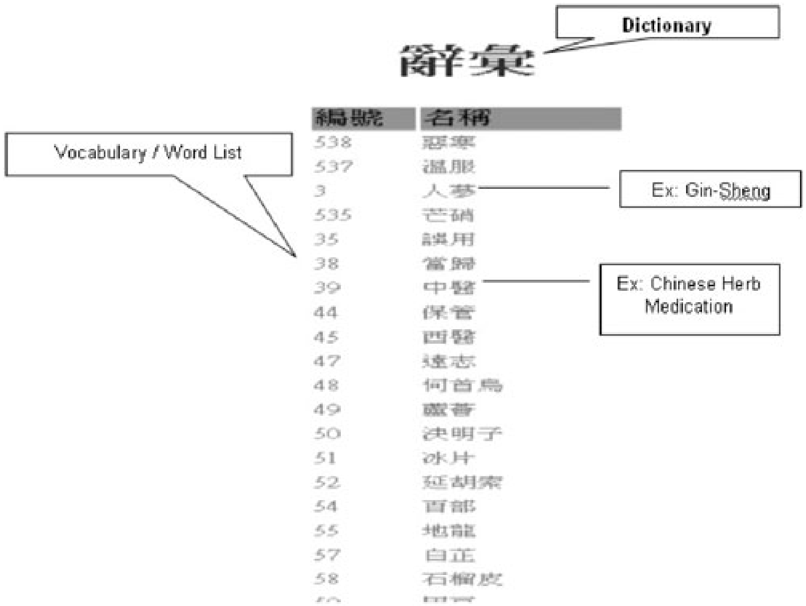

Fig 3 Dictionary for vector presentation (English translation in notes).

weight ranking. If the user is interested in certain topics, he or she can simply follow the hyperlinks to trace the topic into the next level. All the maps in the following levels are organized by similar algorithm.

The hierarchical topic map usually contains more than 10 000 professional documents which are divided into 6–12 groups. It provides different views from textbooks of Chinese herb medication. It is a full picture of domain knowledge for learners and researchers. In order to let users discover and learn interactively by their own traces, we transform the links between map levels into action queries. So the map user is actually querying the map engine in every link he or she clicks. The query results are automatically organized as topic maps by SOM iteratively as shown in Fig 4.

Educational evaluation

After the SOM topic maps are compiled, we follow up with learning usability evaluation to assess the result by implementing it into related course sessions, and test it within more specific context. The SOM system and materials were presented to 135 college freshmen in the Department of Pharmacy in Tajen University in Taiwan. A short questionnaire was distributed to all student participants after the trial for the result appraisals measured by the Likert scale from one to five. We gathered their general opinions on the functions of the topic map and its value in sustaining learning experiences. In addition, in order to reach a fuller understanding of the implicit learning effects, we coordinated a couple of focus groups consisting of both teachers and students to discuss the related use and issues of SOM. Their suggestions were collected to provide information for improving the design in future stages. The results are presented in Table 1.

Results

Topic map presentation

Once the topic map is structured, it presents a main page showing a simplified table of contents with multiple levels of categories embedded by LabelSOM (Rauber & Merkl 1999). Users start from the first level, and enter the second, and so forth in the same way going deeper into the matrix. It has the same searching mechanism as regular web browsers. Resulting topic maps are presented in XML form individually (see Figs 5 and 6).

The '11922 × 103' heading signifies the topic map resulting from 11 922 documents processed with 103 terms (dictionary words). The document distribution for each group is presented by visualization. Each cluster is labeled with a different colour and each little dot represents an individual document with its

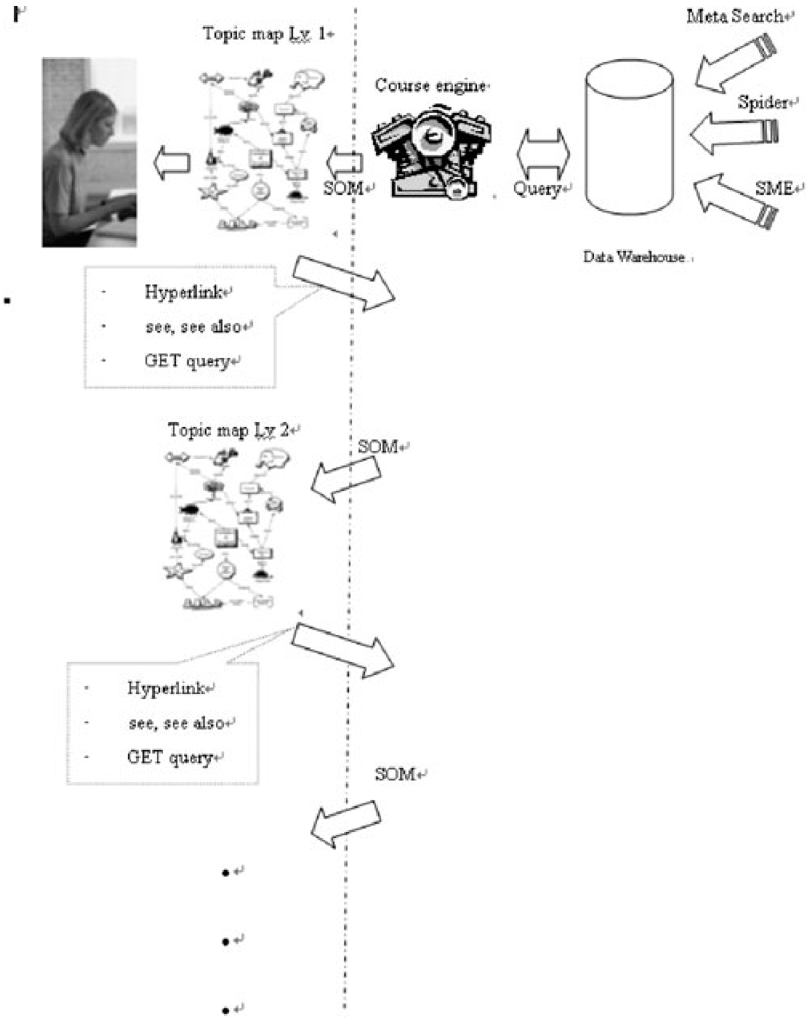

Fig 4 Multilevel topic maps. GET, parameter–passing through URL; SME, subject material expert; SOM, self-organizing map.

hyperlink. Users can trace down the full text by simply clicking on it.

In every map, we sorted the term attributes by their average distance ascent; in other words, the terms that are labelled first are terms 'mostly' representing the cluster. For example, the blue group in the 11922 × 103 map is labelled with 'fragile state', which means that group is mostly about topic 'fragile state'.

Technical evaluation

In the data mining world, the goodness of group splitting is often calculated by the Gini index (Berry & Linoff 2000), which represents the diversity of documents. The Gini index can range from 0 to 1; the higher the index, the better is the resulting grouping. (see Table 2). As is shown, the execution time increases with the dimension and number of documents. The distribution of resulting groups worsens as the document number increases, too. Limited by the current computation capability, very large scale SOM text-mining is virtually impossible.

For verifying the coherence of document grouping, the SME set up a gold standard, which consists of 500 documents divided into 12 clusters artificially. The experts reviewed the document and manually mark up three most important concepts (terms) embedded in each cluster. Thereafter, the SOM results are compared with the gold standard statistically to see the matching ratio, which is called 'hit rate' (see Table 2). What should be noted is, to set up the gold standard manually based on a large data set (such as the 11935 × 103 case we have implemented) is tedious and meaningless (if not impossible).

Table 1. Educational evaluation results.

	Mean	SD
1. The interface of topic maps presented in HTML is easy to understand and manage	3.74	0.52
2. The categories presented by topic maps are clear and easy to search	3.92	0.79
3. The topic maps can respond quickly to the inquiry	2.78	0.72
4. It is convenient to retrieve resources by SOM	3.06	0.51
5. The hierarchical presentation of topic maps is meaningful and corresponding to the knowledge structure of learning materials	3.14	0.53
6. The SOM provides comprehensive understanding toward learning materials	3.16	0.47
7. The resources compiled by SOM are related to course content	3.21	0.53
8. The content retrieved by SOM is rich and proper	3.66	0.77
9. The content of topic maps matches the learning objectives	3.19	0.48
10. The topic maps provide much learning resources outside of textbooks	3.56	0.78
11. The topic maps are interactive	4.58	0.62
12. The topic maps are easily customizable according to needs	2.52	0.75
13. The structure of SOM can be easily adopted as teaching and learning materials	2.56	0.79
14. The topic maps assist in discovery new knowledge	3.19	0.51

SOM, self-organizing map.

Educational evaluation

The SOM system analyses a large amount of data, uses expert-defined terms to categorize them, and then generates hierarchical materials for teaching as presented in Figs 5 and 6. From the educational perspective, the teachers can more effectively use the system in their education process, helping them to collect, process, digest and analyse information.

From the effective questionnaires done by 126 rookie students taking 'Introduction of Chinese herb' at the end of the second semester – week 16, Spring 2006, we investigated the usability and instructional value of the SOM topic maps, which include the presentation of material categories and the trouble-free search of materials; the convenience of data retrieval; the meaningfulness and usefulness of resources; the value of the system's assistance in new knowledge discovery; the usefulness of instructional needs; the appraisal of the overall conceptual presentation of large amounts of information; and the level of acceptance and comprehensive understanding of the learning materials, and so forth. The result shows that the questions related to the presentation and functions of topic maps are ranked the highest in the evaluation. Users consider topic maps to have good interaction (mean = 4.58), functions in analysing large amounts of data, and can shuffle them into clear categories (mean = 3.92), and the interface presented in HTML is easy to understand and manage (mean = 3.74). In terms of the content compiled by the SOM, users rank them between 3.10 and 3.20. They think that the hierarchical presentation of topic maps is meaningful and corresponding to the knowledge structure of learning materials, and the SOM provides comprehensive understanding toward learning materials. At the same time, the resources compiled by SOM are related to course content, matches the learning objectives, and can assist in the discovery new knowledge. However, questions related to the immediacy of use are rated below 3. To understand the reasons of low ratings and the implications of it, we conducted a focus group interview to gather more feedback.

Discussion

From the interviews, we find that different from other browsers or databases, topic maps have several features

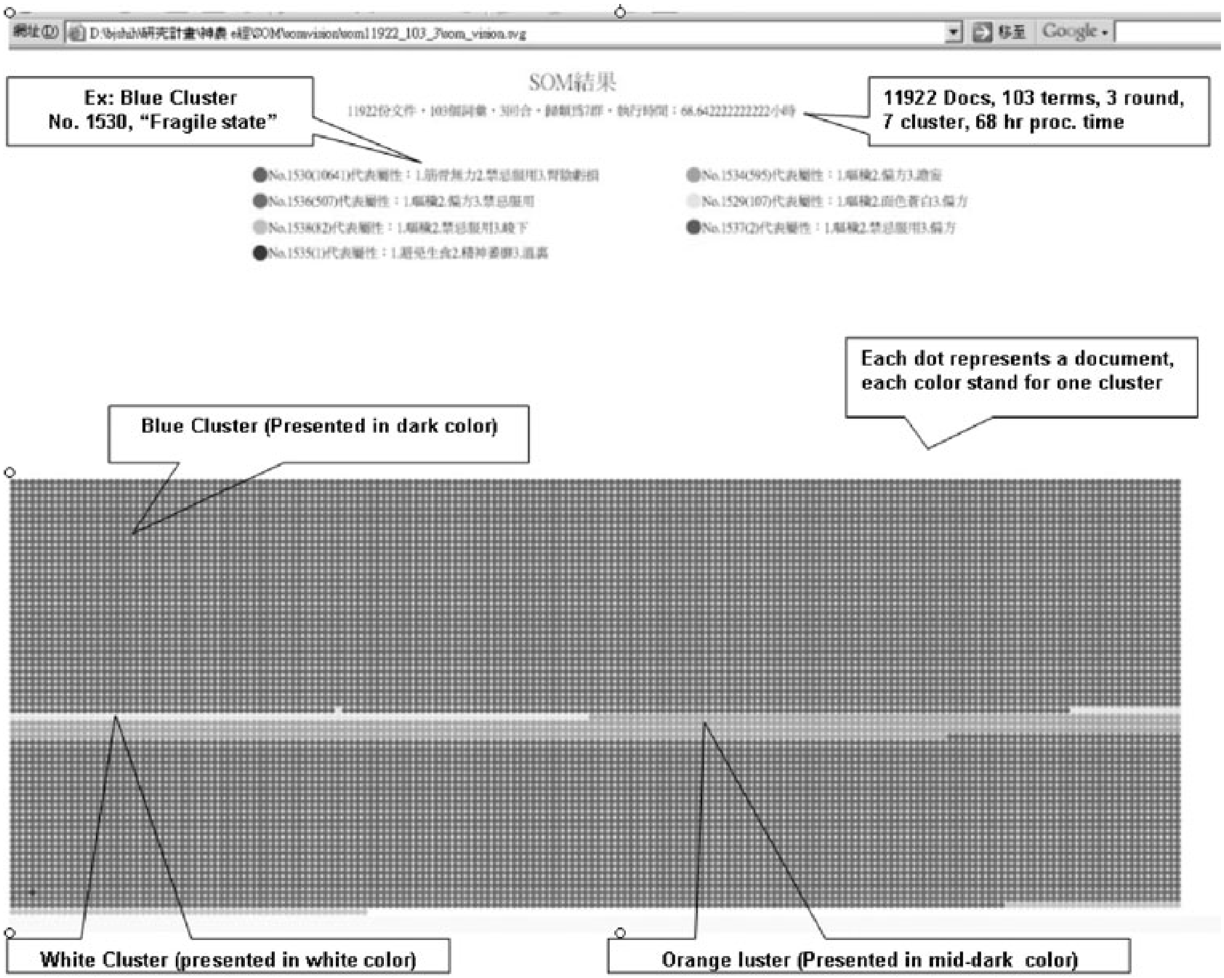

Fig 5 11922 × 103 result in map form.

that are unique. We find that the hierarchical topic maps can assist in the construction of Chinese herb medication digital learning materials in several ways. We analyse the interview discussions more in-depth from three perspectives.

First, from the content perspective, SOM provides rich professional resources, and generates continuously updated information. The topic map is oriented from the concept of a semantic network. It is most useful to sort documents that are rich in text. Although the interface is simple, the connotation of it is that it embeds a large selection of files of different sources and different formats. The documents are mostly complicated with professional depth. For learning in a professional field, having rich resources is the beginning of research and the key of learning. As the sources in a databank are constantly changing, the menu of topic map material is correspondingly dynamic. Users can always retrieve the newest information as they are clicking on the groups or terms. This function is as useful to teachers as to students when they are organizing teaching and learning materials with the newest information. In addition, SOM shows a comprehensible presentation of knowledge structure. For professional knowledge which has complex meanings embedded in it, knowledge structure provides the patterns and consists of conceptual nodes to help users grasp the gist of it. It offers a grid to support extensive exploration and research. It helps to formulate the implicit abstract concepts making it an explicit knowledge so that users can more feasibly conduct analysis with it.

From the educational perspective, SOM can support systematic learning as well as constructive learning. On

Fig 6 11922 × 103 result in XML form.

one hand, a multilevel topic map can assist the categorization of knowledge and present it with labels for each term. A topic map also shows the frequency each term appears in all documents and the relationship between terms, which can effectively guide users through systematic browsing and inquiry. With this function, SOM works more as a dynamic researching tool than static learning material. On the other hand, SOM sustains constructive learning. Although the functionality of a topic map is formulaic and systematic,

Table 2. Technical statistics of SOM results.

Document No. × Term No.	Execution time	Cluster No.	Gini index	Hit rate
11935 × 103	68.64 h, (execution distributed on 10 PC)	7	0.2	0.44
11744 × 84	44.06 h, (execution distributed on 10 PC)	5	0.238	0.56
500 × 84	245.76 h (executed on 1 PC)	12	0.622	0.87

SOM, self-organizing map.

it is also feasible for task-based learning. From the constructive point of view, learners need resources from multiple sources for the purpose of independent research. The mechanism can suffice the exploration of various learning styles, tendencies of interests and professional abilities. More importantly, this guidance is not provided by teachers working in the classrooms, but an autonomous system which is supported by a professional team with a wide array of resources. It turns learning into an information-guided dynamic. Therefore, this material helps users 'discover' new knowledge by presenting explicit and implicit knowledge so that they are able to see ideas and concepts that are most unexpected. This process matches the basic principles of constructivist learning.

From the usability perspective, SOM can carry out autonomous processing and presentation. Topic maps provide teachers and learners with an autonomous abstract environment. Even facing substantial documents, computers can replace human labour to efficiently process the tedious algorithm, but maintain high-level humanistic and professional analysis. Teachers can apply it on the teaching websites, and let the system compile thematic materials for them, reducing time spent on copying and pasting, coding and rewriting. It is customizable and interactive. The menu of a topic map interacts with users, opening up layers of information upon every selection and inquiry. Hence, the route taken by every user is different, and the system returns with different results upon every choice. The system has proper support for customized learning. Best of all, it is easily manipulated. Most search engines, websites and databases are designed to carry documents with different formats, content areas and inquiry methods. Users who are unfamiliar with each system can get lost in it, and every inquiry can take up much energy and time. A topic map presents materials with only a few themes on the menu. It is easy to use, and simple for users to grasp the key terms generated from the knowledge content. Users do not need to spend much time and energy to get a hold of the main theme of the massive resources. Even learners without prerequisite knowledge, and lack of sufficient subject understanding to reach into the depth of content, can still get on the system quickly, starting from the key terms, and advance progressively.

Conclusion

The goal of this research is to use text-mining techniques – SOM and topic map – to generate learning materials based on the subject knowledge. It is hoped that the course engine can automatically generate professional learning materials, and can offer student opportunities to take hold of large amounts of information and turn it into useful learning resources. From the observations in the primary stage, the grouping result of this hierarchical topic map can produce fairly meaningful learning materials on Chinese herb medication education.

The development of this system uses the trial-and-error method to adjust the values betweens factors including learning rate and initial values. There are no set rules or predetermined values for the system to use because the learning materials for each professional subject matter are different by nature. From the aspect of making learning materials, this system can generate reasonably satisfactory results for the subject matter expert so far. It is believed that more system 'training' can lead to the generation of more complete and trustworthy results.

This research is a cross-discipline approach between knowledge management, information technology, and educational design. The integration of these theories stimulates alternative research direction, which encourages more cooperative research between different realms. In the near future, there will be developments of topic maps in conjunction with various types of course modules, including digital learning platforms and user interface design. The prototype of learning content in the format of website and CD-ROM will be modeled. Following that will be a cognitive research report using action research principles. After the evaluation has been completed, the digital learning content can be launched for official educational use.

We hoped that this research provides an insight to an automation tool which can bring teaching and learning an alternative in searching and compiling materials with technology. We believed all teachers and students in higher education would see a chance of applying it and find it useful in their academic works.

Acknowledgements

This research was supported by National Science Committee in Taiwan under the category of National Technology Plan for Digital Learning, and as a part of the result of the project 'Shannon Sutra: Digital Learning Content and Experimental Research on Chinese Herb Medication Safety' (NSC-93-2524-S-127-001). Special thanks are owed to the research team in Digital Content and e-Learning Laboratory in Tajen University, Taiwan.

Appendix A

Vector space representation of documents

Documents based on natural language contain abundant descriptive material. However, they are difficult to retrieve by machines which are designed based on program language. Most text-mining approaches inherently rely on the algorithms representing content of the text mess ages through a vector space model, in which each mess age is represented by a vector (Roussinov & Zhao 2003). It takes terms in documents as attributes, and computes the attribute weight. This approach is more difficult for the Chinese language (Wong & Li 1998) because it does not contain 'stop-words' between terms. This problem has also been noticed in a more general domain of text technologies. It is traditionally known as the vocabulary problem and there has not been an effective solution to it. Since natural languages are very am biguous and diverse, solving this problem would require knowing semantic relationships between all possible words and phrases. This task is believed to be 'AI-complete', which means solving it would require solv ing all the other Artificial Intelligence (AI) tasks such as natural language understanding, common sense reasoning and logical thinking (Roussinov & Zhao 2003).

Word segmentation has to be done before any mining attempt. Among word segmentation algorithms, the dictionary approach (Chien 1997) is most intuitive and precise. With this method, SME has to predefine a dictionary containing words of interest. The program accordingly segments the paragraphs into term-sets afterward. It requires human effort and frequent updates. The quality of mining results is highly associated with the dictionary too. However, when the dimensionality of the domain contains a large number of attributes (variables), the size of the search space for model induction grows in a combinatorially explosive manner (Kiang & Kumar 2001). Approaches to this problem include methods to reduce the effective dimensionality of the problem and the use of prior knowledge to identify irrelevant variables. Normally, several hundred to a few thousand words will be retrieved from documents in the usual applications.

Salton (1989) developed vector space representation (sometimes called the IDF approach) to convert documents into vectors. It computes the term frequency and document frequency. Term frequency refers to the occurrence of term *i* in document *d*. Document frequency refers to the number of documents in the mining set containing term *i*. Term Frequency Inverse Document Frequency was then used as the weights of certain documents:

$$w_i(d) = tf_i(d) * \log\left(\frac{N}{df_i}\right)$$

$w_i(d)$: weight represent term i in document d.
$tf_i(d)$: the frequency of term i in document d.
N: the number of documents in mining set.
df_i: the number of documents containing term i.

The IDF rule values those terms with higher occurrence in fewer documents, which means they are more 'special' and 'meaningful' intuitively. We can then use the vector representation to perform the following classification and visualization.

SOM

Two approaches are commonly used for document classification. Some applications classify the documents according to their individual 'external' attributes – length, size, source, predefined subject categories, etc. (Fox *et al*. 1993; Shneiderman *et al*. 2000) or simply the frequency of certain terms occurring in documents (Veerasamy & Belkin 1996). The other approach is based on the similarity of documents and employs neural network-based machine-learning algorithms such as SOM (Kohonen 2001). A general *a priori* justification for preferring neural network-based approaches to statistical ones is that they do not require the invocation of assumptions about the underlying data generating mechanisms (e.g. the distributional assumption of multivariate normality that is invoked to justify the use of several multivariate statistical modeling procedures) (Kiang & Kumar 2001).

One of the most well-known data mining approaches is WEBSOM (Smith & Ng 2003). It includes groupings of addressable information objects (occurrences) around topics and relationships (associations) among topics (Liu & Hsu 2004). There are several advantages to using the SOM to cluster documents, rather than people, due to the objectivity of the process. In addition, the process is automatic (hence the name 'self-organizing') (Smith & Ng 2003). The common belief behind those approaches is that automated processing techniques can reduce the cognitive load of meeting par-

Begin

Set neighborhood parameters R, learning rate parameters (try and error)

Initialize weights (by random, Principle component analysis or simply pick up some inputs as seeds)

While

For each input factor $\mathrm{x}_k = [x_{k1}, x_{k2}, \ldots, x_{km}]^{\mathrm{T}}$

For each output node, compute the distance:

$$d_j \leftarrow \|\mathrm{x}_k - \mathrm{w}_j\|,\ j = 1, 2, \ldots, n$$

Find index j such that d_j is a minimum

For node j and its neighborhoods, updates according to

$$\varsigma \leftarrow \exp(-R^2 / 2\sigma^2)$$

$$\mathrm{w}_j \leftarrow \mathrm{w}_j + \varsigma \times \eta \times (\mathrm{x}_k - \mathrm{w}_j)$$

Reduce learning rate η and radius of R of neighborhood

Until (Convergence or maximum no. of iterations is exceeded)

End

Fig A1 Self-organizing map algorithm.

ticipants even if manual post-processing is still required (Roussinov & Zhao 2003).

The main function of a SOM network is to map the input data from an n-dimensional space to a lower dimensional (usually one- or two-dimensional) plot while maintaining the original topological relations (Kohonen 2001). It has been successfully applied as a classification tool to various problem domains, including speech recognition, image data compression, image or character recognition, robot control, and medical diagnosis. The SOM output format (used by WEBSOM) is very useful in conveying the relationship of the web documents in a user-friendly manner (Smith & Ng 2003). Several previous studies have explored automated detection and summarization of structures in meeting messages by identifying and listing the most important concepts, representing the messages with semantic maps, or clustering the messages into automatically created subsets of topics (Roussinov & Zhao 2003). These applications can help users to effectively define the subject within groups (Rasmussen 1992; Zamir & Etzioni 1999). The SOM algorithm is illustrated in Fig A1.

References

Anderson J.R. (1980) *Cognitive Psychology and Its Implications*. W.H. Freeman and Company, San Francisco, CA.

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

Berry M.A. & Linoff G.S. (2000) *Mastering Data Mining*. John Wiley & Sons, New York.

Bruno D. & Richmond H. (2003) The truth about taxonomies. *Information Management Journal* **37**, 44–51.

Chen H., Houston A.L., Sewell R.R. & Schatz B.R. (1998) Internet browsing and searching: user evaluation of category map and concept space techniques. *Journal of the American Society for Information Science* **49**, 582–603.

Chien L.F. (1997) PAT-Tree-Based keyword extraction for Chinese information retrieval. In *Proceedings of the 20 International ACM SIGIR Conference on Research and Development in Information Retrieval.* pp. 50–58. 27–31 July, Philadelphia, PA.

Dittenbach M., Merkl D. & Rauber A. (2000) The growing hierarchical self-organizing map. In *Proceedings of the Int'l Joint Conference on Neural Networks (IJCNN'2000)*, Vol. VI, pp. 15–19. IEEE Computer Society Press, Como, Italy.

Dringus L.P. & Ellis T. (2005) Using data mining as a strategy for assessing asynchronous discussion forums. *Computer and Education* **45**, 141–160.

Fisher K.M., Wandersee J.H. & Moody D. (2000) *Mapping Biology Knowledge*. Kluwer Academic, Boston, MA.

Fox E., Hix D.L., Nowell T., Brueni D.J. & Rao D. (1993) Users, user interfaces, and objects: envision, a digital library. *Journal of the American Society for Information Science* **44**, 480–491.

Govindasamy T. (2001) Successful implementation of e-learning; pedagogical considerations. *Internet and Higher Education* **4**, 287–299.

Kiang M.Y. & Kumar A. (2001) An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research* **12**, 177–194.

Klemke R. (2000) Context framework – an open approach to enhance organizational memory systems with context modeling techniques. In *Proceedings of the Third International Conference on Practical Aspects of Knowledge Management (PAKM2000).* 30–31 October, Basel, Switzerland. [WWW document.] URL http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34

Kohonen T. (2001) *Self-organizing Maps*, 3rd edn. Springer-Verlag, New York.

Lee C.H.-M., Cheng Y.-W., Rai S. & Depickere A. (2005) What affect student cognitive style in the development of hypermedia learning system? *Computer & Education* **45**, 1–19.

Lin X., Soergel D. & Marchionini G. (1991) A self-organizing semantic map for information retrieval. In *Proceeding of 14th the ACM SIGIR Conference Research and Development in Information Retrieval.* pp. 262–269. 13–16, October, Chicago, IL.

Liu D.-R. & Hsu C. (2004) Project-based knowledge maps: combining project mining and XML-enabled topic maps. *Internet Research* **14**, 254–266.

Malhotra Y. (1998) Deciphering the knowledge management hype. *Journal for Quality and Participation* **21**, 58–60.

Marx R.W., Blumenfeld P.C., Krajcik J.S. & Soloway E. (1998) New technologies for teacher professional development. *Teaching and Teacher Education* **14**, 33–52.

Masiello I., Ramberg R. & Lonka K. (2005) Attitudes to the application of a Web-based learning system in a micro biology course. *Computer and Education* **45**, 171–185.

Mladenic D. & Grobelnik M. (2003) Feature selection on hierarchy of web documents. *Decision Support Systems* **35**, 45–87.

Nonaka I. & Takeuchi H. (1995) *The Knowledge-Creating Company – How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, New York.

Noy N.F. & McGuinness D.L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology.* Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March.

Putnam R.T. & Borko H. (2000) What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher* **29**, 4–15.

Rasmussen E. (1992) Clustering algorithms. In *Information Retrieval Data Structures and Algorithms* (eds W.B. Frakes & R. Baeza-Yates), pp. 419–442. Prentice Hall, Englewood Cliffs, NJ.

Rath H.H. & Pepper S. (1999) Topic maps: introduction and allegro. In *Proceedings of the Markup Technologies 99* (ed. B.T. Usdin), December 7–9. Graphic Communications Association. Philadelphia, PA. [WWW document.] URL http://www.ontopia.net/topicmaps/materials/allegro.pdf

Rauber A. & Merkl D. (1999) Automatic labeling of self-organizing maps: making a treasure-map reveal its secrets. In *Proceedings of the 3rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD′99)* pp. 228–237. 26–28 April, Beijing, China.

Roussinov D. & Zhao J.L. (2003) Automatic discovery of similarity relationships through Web mining. *Decision Support Systems* **35**, 149–166.

Salton G. (1989) *Automatic Text Processing*. Addison-Wesley, Boston, MA.

Shneiderman B., Feldman D., Rose A. & Grau X.F. (2000) Visualizing digital library search results with categorical and hierarchical axes. In *Proceedings of 5th ACM Conference on ACM 2000 Digital Libraries*. pp. 57–65. 2–7 June, San Antonio, TX.

Smith K.A. & Ng A. (2003) Web page clustering using a self-organizing map of user navigation patters. *Decision Support Systems* **35**, 245–256.

Smolnik S. & Erdmann I. (2003) Visual navigation of distributed knowledge structures in groupware-based organizational memories. *Business Process Management Journal* **9**, 261–280.

Veerasamy A. & Belkin N.J. (1996) Evaluation of a tool for visualization of information retrieval results. In *Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR)*. pp. 85–92. 18–22 August, Zurich, Switzerland.

Wong K.F. & Li W.J. (1998) Intelligent Chinese information retrieval – why is it so difficult? In *Proceedings of the First Asia Digital Library Workshop.* pp. 47–56. 6–7 August, University of Hong Kong, Hong Kong.

Yang C.C., Chen H. & Hong K. (2003) Visualization of large category map for Internet browsing. *Decision Support Systems, Special Issue on Web Retrieval and Mining* **35**, 89–102.

Yoon B.-U., Yoon C.-B. & Park Y.-T. (2002) On the development and application of a self-organizing feature map-based patent map. *R&D Management* **32**, 291–300.

Zamir O. & Etzioni O. (1999) Grouper: a dynamic clustering interface to web search results. *Computer Networks* **31**, 1361–1374.