

# Expertise visualization: An implementation and study based on cognitive fit theory

Zan Huang<sup>a,\*</sup>, Hsinchun Chen<sup>b</sup>, Fei Guo<sup>b</sup>, Jennifer J. Xu<sup>c</sup>,  
Soushan Wu<sup>d</sup>, Wun-Hwa Chen<sup>e</sup>

<sup>a</sup> Department of Supply Chain and Information Systems, Smeal College of Business, The Pennsylvania State University,  
419 Business Building, University Park, PA 16802, United States

<sup>b</sup> Department of Management Information Systems, Eller College of Management, The University of Arizona,  
Tucson, AZ 85721, United States

<sup>c</sup> Department of Computer Information Systems, Bentley College, Waltham, MA 02452

<sup>d</sup> College of Management, Chang-Gung University, Taiwan

<sup>e</sup> Department of Business Administration, National Taiwan University, Taiwan

Received 30 July 2003; received in revised form 29 December 2005; accepted 18 January 2006

Available online 10 March 2006

## Abstract

Expertise management systems are being widely adopted in organizations to manage tacit knowledge. These systems have successfully applied many information technologies developed for document management to support collection, processing, and distribution of expertise information. In this paper, we report a study on the potential of applying visualization techniques to support more effective and efficient exploration of the expertise information space. We implemented two widely applied dimensionality reduction visualization techniques, the self-organizing map (SOM) and multidimensional scaling (MDS), to generate compact but distorted (due to the dimensionality reduction) map visualizations for an expertise data set. We tested cognitive fit theory in our context by comparing the SOM and MDS displays with a standard table display for five tasks selected from a low-level, domain-independent visual task taxonomy. The experimental results based on a survey data set of research expertise of the business school professors suggested that using both SOM and MDS visualizations is more efficient than using the table display for the *associate*, *compare*, *distinguish*, and *cluster* tasks, but not the *rank* task. Users generally achieved comparable effectiveness for all tasks using the tabular and map displays in our study.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Expertise management; Information visualization; Self-organizing map; Multidimensional scaling; Visualization evaluation; Cognitive fit theory

## 1. Introduction

The knowledge management community has defined two forms of knowledge: explicit and tacit. Explicit knowledge is knowledge that can be articulated in a formal language and transmitted among individuals. Tacit knowledge is personal knowledge

\* Corresponding author. Tel.: +1 814 863 1940.

E-mail addresses: [zanhuang@psu.edu](mailto:zanhuang@psu.edu) (Z. Huang),  
[hchen@eller.arizona.edu](mailto:hchen@eller.arizona.edu) (H. Chen), [guof@email.arizona.edu](mailto:guof@email.arizona.edu) (F. Guo),  
[jxu@bentley.edu](mailto:jxu@bentley.edu) (J.J. Xu), [swu@mail.cgu.edu.tw](mailto:swu@mail.cgu.edu.tw) (S. Wu),  
[andychen@ccms.ntu.edu.tw](mailto:andychen@ccms.ntu.edu.tw) (W.-H. Chen).

embedded in an individual's experience and is involved with intangible factors such as personal beliefs, perspectives, and values [32]. In most organizations, explicit and tacit knowledge resides in two major forms: written documents and expertise in employees' minds. These knowledge forms carry the intellectual assets of an organization that are critical to its competitiveness and long-term success. Knowledge management involves creation, organization, dissemination, and utilization of such intellectual assets to achieve organizational objectives. Document management and expertise management can be viewed as two subfields of knowledge management that deal respectively with explicit and tacit knowledge.

Information technologies are critical to facilitating the knowledge management practices of organizations. Some researchers even explicitly include information technologies into the definition of knowledge management: "it (knowledge management) embodies organizational processes that seek synergistic combination of data and information processing capacity of *information technologies*, and the creative and innovative *capacity of human beings*" [24]. Many well-established research fields, such as information retrieval, digital library, and natural language processing, have developed a rich set of technologies to support various aspects of document management, including document creation and collection, document searching and retrieval, document recommendation, and document visualization. Recently, many researchers and practitioners have started to apply similar technologies to manage the expertise in the minds of employees of an organization. Various systems have been implemented to perform automatic expertise database building, expertise retrieval, and expertise recommendation.

Despite the wide applications of many well-studied document management technologies in expertise management systems, few studies have explored the potential of adopting technologies developed in data and document visualization. The recent proliferation of visualization research has demonstrated the value of visualization techniques in providing more efficient and effective access to large amounts of data and documents. These techniques are expected to also help users explore an expertise space more efficiently and effectively. This study aims to demonstrate and evaluate the application of visualization techniques to support expertise management. In addition, we are also interested in understanding the effects of visualization on the efficiency and effectiveness of human subjects in performing different types of tasks for exploring the expertise space of an organization.

We focused on a basic form of expertise representation, in which experts are represented by a set of expertise fields. Due to the potential high dimensionality of such expertise data, we chose to examine two dimensionality reduction visualization techniques that have been widely applied in visualizing data and documents with inherent high dimensional characteristics: the Self-organizing Map (SOM) and Multidimensional Scaling (MDS). Both techniques generate a map-like representation in which the objects' positions embody inter-object similarities or dissimilarities derived from object attributes or other information sources. While being able to produce intuitive visual repetitions of the original high-dimensional data, these techniques inevitably introduced distortion of the original data during the dimensionality reduction process. Whether these distorted visual representations of the high-dimensional expertise data make sense and provide benefits to users of an expertise management system is unclear. In our study, we implemented SOM and MDS visualizations of an expertise space derived from the research expertise of business school professors in Taiwan. Two types of maps were generated, an expert map and an expertise field map. Both maps showed reasonable grouping of experts and expertise fields according to the actual expert and expertise similarities. The expertise field map generally was consistent with our understanding of the relationships among business research disciplines. To the best of our knowledge, this is the first study to implement and assess dimensionality-reduction visualization techniques for the expertise management domain.

Literature on graphical data representations suggests that effects of data representations on human information processing performance are contingent on the task types. We adopted a low-level, domain-independent visual task taxonomy used in "defeating" visual interface evaluation studies as the basis for our research. Based on cognitive fit theory [42], the accurate graphical data representations (e.g., bar charts for numerical values) should facilitate better cognitive fit for human information processing with spatial tasks and thus enhance performance. However, it is not clear to what extent the cognitive fit theory still applies with the distorted graphical representations of high-dimensional data such as the visualization results generated by the SOM and MDS techniques. We selected from the visual task taxonomy tasks of a spatial nature (based on Vessey's definition [42]) that are critical to expertise space exploration and designed specific formulations of these tasks in our context. We

conducted user experiments using the expert map of Taiwan business school research and empirically test the effects of visualization techniques on human subjects' task performance.

The remainder of the paper is organized as follows. Section 2 presents a literature review of research on use of information technologies for expertise management. We also review in this section visualization technologies developed for document management that can be applied to expertise data. Section 3 describes the expertise data representation used in the study as well as the SOM and MDS techniques for visualizing such expertise data. In Section 4, we review the literature on comparison between graphical and tabular data representations and present the expertise space exploration tasks for which we hypothesized that the SOM and MDS visualization would improve the performance of human subjects. We then present our testbed and visualization results in Section 5. Section 6 provides details of an experimental study for testing our hypotheses. We conclude the paper in Section 7 by summarizing our research contributions and pointing out future directions.

## 2. Literature review

In this section, we review applications of information technologies to expertise management. Although research in expertise management largely aligns with that in document management, we point out a lack of visualization research. We then present a review of visualization research in document management in an attempt to identify applications relevant to expertise management.

### 2.1. Information technologies for expertise management

Expertise management has recently attracted substantial academic and industry interest as a subfield of knowledge management that focuses on the management of the employees' expertise of an organization. Many similar concepts such as expertise capitalization/leveraging, skill mining, competence management, intellectual capital management, expertise network, knowledge sharing system and the like have been widely discussed by researchers and practitioners [47]. Many research prototype systems and practical large-scale systems have been built to support expertise management requirements, often adopting techniques developed for document management in research fields

such as information retrieval, digital library, and natural language processing.

Using concepts from document management, we categorize past research in expertise management into three classes: expertise collection, expertise retrieval, and expertise recommendation.

- *Expertise collection* refers to the process of gathering information regarding people's expertise, identifying those who have expertise, the extent of each expert's knowledge, etc. Document management technologies such as automatic indexing have been applied to transform expertise collection from a manual process as in early expertise management systems such as Microsoft's SPUD and Hewlett-Packard's CONNEX [10] into an automatic process to address the dynamic nature of expertise information. Yimam-Seid and Kobsa [47] provided an extensive review of automatic expertise collection systems. Examples are *Expert/Expert-Locator (EEL)* (also called "Bellcore Advisor") [37], which constructed an expertise index of a research group based on a representative collection of the technical documents produced by that group, and *ContactFinder* [20], which used an intelligent agent to monitor discussion boards and identify contacts in specific areas as experts.
- *Expertise retrieval* focuses on providing access to a repository of expertise information and is similar to document retrieval in digital libraries and web page retrieval in search engines. Initiated by the visionary system *HelpNet* [26], many systems have applied various retrieval models developed in information retrieval, ranging from the basic TF-IDF weighting to Latent Semantic Indexing (LSI), to retrieve expertise [37].
- *Expertise recommendation* focuses more on putting the expert-finding process into a collaborative environment and applying techniques such as collaborative filtering that originated in recommender systems [34] to achieve more effective expertise-request matching. Examples are *ReferralWeb* [18], which combined social network and collaborative filtering to recommend expertise, and *Expertise Recommender* [27], which extensively exploited detailed heuristics and social interactions to recommend sources of expertise in an organizational environment.

Across all three classes of information technology support for expertise management, we have observed parallel development with document management technologies. This is mainly because of the similar characteristics of document and expertise data. A key

feature that differentiates document data from other types of data, such as product specification data and sales data in a business context, is its unstructured nature. Additional processing such as keyword- or concept-based indexing is needed to bring some structure to the documents. With the vector space model, documents are typically represented by high-dimensional, and often sparse, vectors consisting of large numbers of elements representing the occurrence of a particular keyword or concept in the documents. Expertise data is similar to document data in that a person's expertise is also unstructured in nature and is difficult to characterize completely. Adopting a similar approach as document representation, researchers have used expertise-field indexing and similar high-dimensional sparse vectors to provide a computable representation of expertise data. The similarity between document and expertise data characteristics suggests that other technologies developed for document management could be applied to support expertise management as well. Document visualization is one of such technologies that promise potential benefits to expertise management systems.

## 2.2. Document visualization technologies

The information retrieval community has explored a variety of post-retrieval document visualization techniques as alternatives to ranked list presentation. As the Web becomes the largest document collection ever, much recent research has focused on visualizing Web search engine results. Mann [25] classified these research activities into three areas of interest: document level, web-site level and document-set level. The document-set level visualization is most relevant to our study because its techniques can be directly applied to visualize expertise information by treating experts as documents and their fields of expertise as keywords in the documents. We present a detailed review of these visualization techniques below.

Document-set level visualization provides a visual metaphor for the whole set of retrieved documents. Such visual representations of the retrieved documents are provided to help users access query results more efficiently and effectively. One type of set-level document visualization uses interactive scatter plots in different forms, which is also referred to as “dimensions and reference point systems” [29]. Visualization techniques of this type attempt to display additional information about the retrieved documents and to group documents that share similar characteristics. These characteristics may include the relationship

between the documents and the query terms [1], predefined document attributes such as size, date, source or popularity [12,30], and user-specified attributes such as predefined topics [31]. Morse and Lewis recently conducted a systematic evaluation of several techniques that visualize the relationship between retrieved documents and query keywords. Their study evaluated text, table, icon, graphical, and “spring” displays, and demonstrated the utility of visualization for supporting information retrieval activities.

A second category of techniques attempts to visualize inter-document similarities. This form of visualization is also referred to as “map systems” [29]. There are four major techniques for inter-document similarity visualization: document networks [38], physically based modeling techniques [5], document clustering [2,13] and geographic map metaphors [7,22]. Other interesting approaches to visualizing a document set include the use of a wall metaphor [23], Venn diagrams [36], and cone trees [35].

Besides search engine result visualization, some other research has explored visualizing large collections of documents to map knowledge domains. One example is ThemeSpace and Galaxies [46], the underlying techniques of which are multidimensional scaling and principle component analysis. Another example is WebSom [14], which utilizes the self-organizing map algorithm to visualize large numbers of documents. Such research aimed to use the map metaphor to visualize large document sets such as digital libraries, regulations and procedures, archived reports, patent collections, etc. Chen et al. [8] conducted experiments to compare the utility of a self-organizing-map-based visualization in supporting users browsing a large Internet information space with that of alphabetical and hierarchical organizations. Their results suggested that the map-based visualization provided better support for broad browsing tasks.

In this study, we were interested in applying the visualization technologies developed for document management to support expertise management and in understanding particular expertise space exploration tasks that could benefit from these visualization technologies. In the next section, we discuss visualization for expertise management and describe the specific expertise data representation and visualization techniques we focused on in this study.

## 3. Visualization for expertise management

We first review the limited previous research on visualization of expertise data and describe the specific

expertise data representation we chose to focus on. We describe the algorithmic details of two selected visualization techniques, the self-organizing map and multidimensional scaling.

### 3.1. Applying visualization techniques to expertise data

Very few studies have addressed the issue of visualization in the context of expertise management. Mockus and Herbsleb [28] presented a system named “Expertise Browser” in the context of collaborative software engineering for change management systems. They embedded some simple visualization elements such as the tree structure and other visual elements to present expert attributes. Zhu and Chen [49] developed the “Communication-Garden System” to visualize computer-mediated communication processes. The “people visualizer” component of this system uses a glyph-based flower representation to identify active participants (experts). Another related research stream deals with the visualization of a knowledge domain and is represented by Chen’s research on large-scale network visualization of papers and authors related to a scientific paradigm [6,9]. Such research on mapping knowledge domains provides examples of building connections between document and expertise visualization. However, the focus of these studies was not on expertise visualization.

In the context of large enterprises, where large volumes of expertise information exist, information overload could create challenges for many expertise management tasks. Applying well-developed visualization techniques for expertise management should improve users’ ability to explore an expertise space more efficiently.

Our research explored this idea by focusing on a simple form of expertise database, in which each expert is represented by a list of predefined expertise fields. Each expert can be represented as a vector, each element of which represents whether the expert possesses expertise in the associated predefined field. As described previously, this expertise representation shares the same high-dimensional and sparse characteristics as vector space representation of documents. We chose two dimensionality reduction techniques for visualizing document space commonly used in the literature, the self-organizing map and multidimensional scaling, to generate map metaphors to visualize the expertise space of an organization. We hypothesized that such a map-based view of the expertise space would help users perform expertise space exploration tasks in the same way a document map benefited document browsers. We

were also interested in learning the particular types of tasks that would benefit from map-based visualization of the expertise data. We present the algorithmic details of the two visualization techniques in Section 3.2 and leave the detailed discussion of the techniques’ effects on expertise space exploration tasks for later sections.

### 3.2. Visualization techniques: SOM and MDS

In this section, we describe the self-organizing map (SOM) and multidimensional scaling (MDS) algorithms adopted in our research. These two techniques can map a high-dimensional data set to a space with smaller dimensions. Both have been frequently applied to generate a two-dimensional or three-dimensional map display of multidimensional data. There have been many visualization projects and software aimed at producing three-dimensional and even higher-dimension visualization including some examples of document visualization we have reviewed previously. These higher-dimension visualizations have the advantage to preserve more accurate original high-dimensional data patterns and reducing the information loss during the projection process. However, there is likely to be a dimension overload problem as typical users are not able to access and manipulate high-dimension visualization intuitively and might not actually benefit from such visualization in terms of exploring the information space more effectively and efficiently [50]. Based on this consideration, in our study we have focused on investigating the impact of two-dimensional visualization that regular users can intuitively explore, thus restricting our SOM and MDS algorithms to produce two-dimensional expertise maps.

#### 3.2.1. Self-organizing map

Self-organizing Map (SOM) was first introduced by Kohonen [19] and has attracted substantial research interest in a wide range of applications. SOM is an unsupervised learning mechanism that clusters objects having multi-dimension attributes into a lower-dimension space, in which the distance between every pair of objects captures the multi-attribute similarity between them.

The input for the SOM algorithm is a set of objects with multiple features. Each object is represented as a vector of the feature values and is referred to as an input in the subsequent description. In the context of our research, each expert was an object and the fields of expertise were the features associated with an object.

Kohonen based the neural network on the associative neural properties of the brain [19]. This network

contains two layers of nodes: an input layer that represents the object features and a mapping layer in the shape of a two-dimensional grid that determines the positions of the objects. The mapping layer acts as a distribution layer to summarize general feature patterns in the collection of objects. Each node in the input layer corresponds to one of the features of an object. Each node in the mapping layer is connected to all input layer nodes with certain link weights. Thus, a mapping layer node can be also viewed as a feature vector with link weights as the feature values.

Each time a new input (the expertise profile of an expert) is presented, the algorithm calculates distance  $d_j$  between the input and each mapping node  $j$ , as shown in (1), where  $x_i(t)$  represents the value of feature  $i$  in the input presented at time  $t$ , and  $w_{ij}(t)$  represents the link weight between input node (feature)  $i$  and mapping node  $j$  at time  $t$ .

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2 \quad (1)$$

The node  $j^*$  with the minimum distance becomes the “winning” node and the link weights between a mapping layer node and the input nodes are updated according to its distance to this “winning” node. Many learning functions have been used in previous research. The one we used is shown in (2), where  $\eta(t)$  is the learning rate for updating link weights and  $R(t)$  is the radius factor of the neighborhood. These two factors decrease over time and converge to a specified minimum value. Intuitively, the mapping nodes closer to the winning node get larger updates, and over time, the amount and scope of the update decrease and the link weights of the mapping nodes tend to stabilize.

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \eta(t)e^{-\frac{\text{distance-to-winner}}{R(t)}}(x_i - w_{ij}(\text{old})) \quad (2)$$

This process of weight updating is performed for a specified number of iterations. At the final stage, each expert is assigned to a mapping layer node having the smallest distance to the expert’s profile, thus each expert obtains a position in the two-dimension map.

### 3.2.2. Multidimensional scaling

Multidimensional scaling (MDS) is a multivariate statistical technique that is often used to map high-dimensional numerical data to a spatial structure of lower dimensions. It has been extensively applied in psychology research as a psychometric method [39] and in marketing research in areas such as product positioning

and market segmentation [4]. It has also been applied to document visualization as mentioned previously. A classic MDS algorithm was given by Kruskal [21].

The input to MDS is a square, symmetric matrix indicating relationships among a set of objects. Such matrices are usually either similarity or dissimilarity matrices. In the context of our research, a similarity matrix was formed on the basis of similarity scores of expert pairs derived from the Jaccard’s similarity function [15] as shown in (3), where  $A$  and  $B$  represent the sets of expertise fields associated with experts  $a$  and  $b$ , respectively.

$$\text{Similarity score } (a, b) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

MDS attempts to find a set of vectors in the  $p$ -dimension space ( $p$  is much smaller than the dimension of the original matrix) such that the matrix of Euclidean distances among the vectors in the  $p$ -dimension space correspond as closely as possible to some function of the input relationship matrix. In our study  $p$  was set to 2 to generate two-dimension maps. A stress function is used to measure the degree of correspondence between the distance matrix implied by an MDS map and a specified function of the input relationship matrix. We used a general form of the stress function as defined in (4), where  $x_{ij}$  refers to the expertise similarity between experts  $i$  and  $j$ ,  $d_{ij}$  refers to the Euclidean distance between experts  $i$  and  $j$  in the MDS map,  $f$  is a specific function of the input data, and Scale refers to a constant scaling factor to keep stress value between 0 and 1. In our study, we were interested in directly visualizing the expert similarities in the map, thus  $f(x_{ij}) = x_{ij}$  in our study.

$$\text{Stress} = \frac{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (f(x_{ij}) - d_{ij})^2}}{\text{Scale}} \quad (4)$$

A sketch of the MDS algorithm is presented as follows:

- Step 1. Obtain an initial configuration of the experts’ coordinates in a two-dimension map;
- Step 2. Compute Euclidean distances among all pairs of experts, to form the distance matrix;
- Step 3. Evaluate the stress function in (4);
- Step 4. Adjust coordinates of each expert in the direction that best minimizes the stress value;
- Step 5. Repeat Steps 2, 3, and 4 for a specified number of iterations or until no decrease in stress is possible.

We adopted a standard implementation of the MDS algorithm, and implemented the Singular Value Decomposition (SVD) algorithm from [33] to obtain the initial map configuration.

#### 4. Testbed and implementation

We used an Internet survey data set regarding research interests of a group of researchers to form a simple expertise database. We used the research fields as descriptors and represented each researcher as a binary vector of these research fields, in which an element taking the value 1 indicates the researcher is interested in the corresponding field.

##### 4.1. Data set

The Internet survey data set contains research expertise of professors in the business and management fields in Taiwan. The survey was conducted by the National Science Council in Taiwan (equivalent to the National Science Foundation in the United States) and covered most business school professors in Taiwan. We believe this survey data set contains high-quality and comprehensive information about the expertise landscape of Taiwan business and management professors.

The data set contained 597 researchers, who had selected their research interests or expertise from a two-level hierarchy of research fields. In Fig. 1, we show the 11 first-level research fields and a sample of second-level research fields under the first-level field of Finance. The numbers in parentheses for the first-level research fields indicate the numbers of second-level research fields associated. There were 128 second-level research fields and 2865 combinations of researchers

and second-level fields in the data set. We used the second-level fields as expertise descriptors.

We applied the vector space model to form data representation for our visualization purpose. Each of the 597 researchers was represented by a binary vector with 128 elements, which corresponded to the research fields. The expertise similarity between two researchers was derived using vector similarity functions. We also had a dual representation for research fields. Similar to the researcher representation, each of the 128 research fields was represented by a binary vector with 597 elements, which corresponded to the researchers. With this dual representation, similarities among research fields depended on the number of overlapping researchers. Such similarities may reflect the common theoretical/analytical foundations or closely related application domains of the research fields, based on the assumption that researchers typically work on closely related research fields. With these two representations, we can adopt existing document visualization techniques to generate maps of expertise and expertise fields.

##### 4.2. Expertise map visualization results

We present in Fig. 2 the expertise map visualization results generated by the SOM and MDS algorithms. In these two map displays, 597 researchers were positioned in a 20 by 10 two-dimensional map. The labels within the map grids are researcher names. We manually grouped grids of researchers with common research fields into regions labeled by the corresponding research fields. The field labels and groupings only provide support for a high-level overview of the map. We were mainly interested in looking at the grid positions of individual researchers in this study. The visualization



Fig. 1. Research field hierarchy.





researchers assigned to the block) were presented. The similarity scores for the blocks with block distance of 1 (blocks B, D, F, and H) were typically larger than those with block distance of 2 (blocks A, C, G, and I). When looking at the distribution of expertise fields, we found that expertise fields in blocks B, D, E, F, and H were mainly “Internet and E-commerce Application” and “Internet and Database Technologies,” while blocks A, C, G, and I also included other fields such as “DSS/ES and Artificial Intelligence Applications,” “Human Resource Management,” and “Knowledge Management.” We also observed some inconsistencies between block distances and similarity scores. Block B is only 1 block away from block E but the similarity score was quite low (0.19), while block A was positioned 2 blocks away from block E despite a relatively high similarity score (0.33). The reason for such inconsistencies was that the positions of the experts were globally optimized to provide the best match with the original expertise data. In this example, blocks A and B were “misplaced” locally in order to achieve a better global match. For example, block A might have high similarity measures with the blocks to its upper-left direction. These inconsistencies are examples of distortions introduced during dimensionality reduction which made the SOM and MDS visualization results only approximate representations of the original expertise data.

In order to understand the two visualization algorithms’ abilities to preserve original similarity information given these distortions, we conducted a regression analysis to test the general correlation between map distances of the experts and their similarities based on expertise-field overlaps. The expert similarities were calculated using Jaccard’s similarity function. A Euclidean distance function was used to calculate the map distances of researcher pairs. We randomly sampled 100 pairs of researchers to conduct the regression analysis using the linear regression specification,  $S_{ij} = \alpha + \beta \cdot D_{ij} + \varepsilon$ , where  $S_{ij}$  and  $D_{ij}$ , respectively, represented the Jaccard’s similarity and map distance between experts  $i$  and  $j$ ,  $\alpha$  and  $\beta$  are the regression coefficients and  $\varepsilon$  is the disturbance term. Results on both the SOM and MDS maps showed significant correlations between the map distances and expert similarities ( $\beta = -0.0016$ ,  $R^2 = 0.145$ , and  $F$ -test  $p$ -value close to 0 for the SOM regression and  $\beta = -0.0027$ ,  $R^2 = 0.1092$ , and  $F$ -test  $p$ -value = 0.0007 for the MDS regression). These statistics showed that SOM and MDS both preserved a large portion of the expert similarity information (as reflected by the significant  $F$ -test results), although with considerable amount of distortions (as reflected by the

relatively small  $R^2$  values. These analysis results reveal the nature of the approximate graphic representations generated by the SOM and MDS algorithms. It is an interesting line of research to investigate the effect of these approximate graphical representations on human information processing ability within an expertise space.

#### 4.3. Expertise field map visualization result and analysis

Fig. 4 presents the expertise field map generated by the SOM algorithm. A similar map was also generated using the MDS algorithm. In this section, we focus on analyzing the SOM visualization result. Based on the SOM and MDS algorithms, the map distances between the research field pairs correspond to the field similarities derived from overlapping researchers. We observe from Fig. 4 that research fields having underlying similarities based on common theoretical/analytical foundations and application domains were grouped together.

The upper left corner of the map (A–C, II–IV) is occupied by business research fields that mainly employ quantitative methods: “3.2: Optimization Theory and Application” (A-II), “10.6: Managerial Economics and Economics Analysis” (B-II), “3.4: Multi-Objective Decision Analysis” (B-III), “10.1: Decision and Risk Analysis” (A-IV), “1.2: Investment” and “1.5: Risk Management” (C-III), and “1.6 Asset Valuation” (C-IV). Based on the notation introduced in Fig. 1, we found that these second-level research fields were from three first-level fields: (1) Finance, (3) Operations Research/Quantitative Methods, and (10) Decision Science and Managerial Economics. Our map visualization provided an alternative grouping of these second-level research fields based on closeness of the underlying analytical tools and methods.

Regions A-VI and B-VI are occupied by research fields using simulation approaches (e.g., “10.7: Dynamic System and Simulation” (A-VI), and “3.3: Dynamic Programming and Control” and “3.7: System Simulation” (B-VI)). These fields were positioned close to the group of fields using general quantitative methods, but formed a small cluster because of the emphasis on simulation tools. Research fields regarding system design and organization management start to appear in regions A-VII, B-VIII, and C-VIII in the upper right corner. A dense cluster of accounting research fields occupies Region C-VII, which demonstrates that large numbers of researchers in these fields had research across different subfields. The lower right corner (D–H, VII–VIII) is mainly occupied by research fields in the

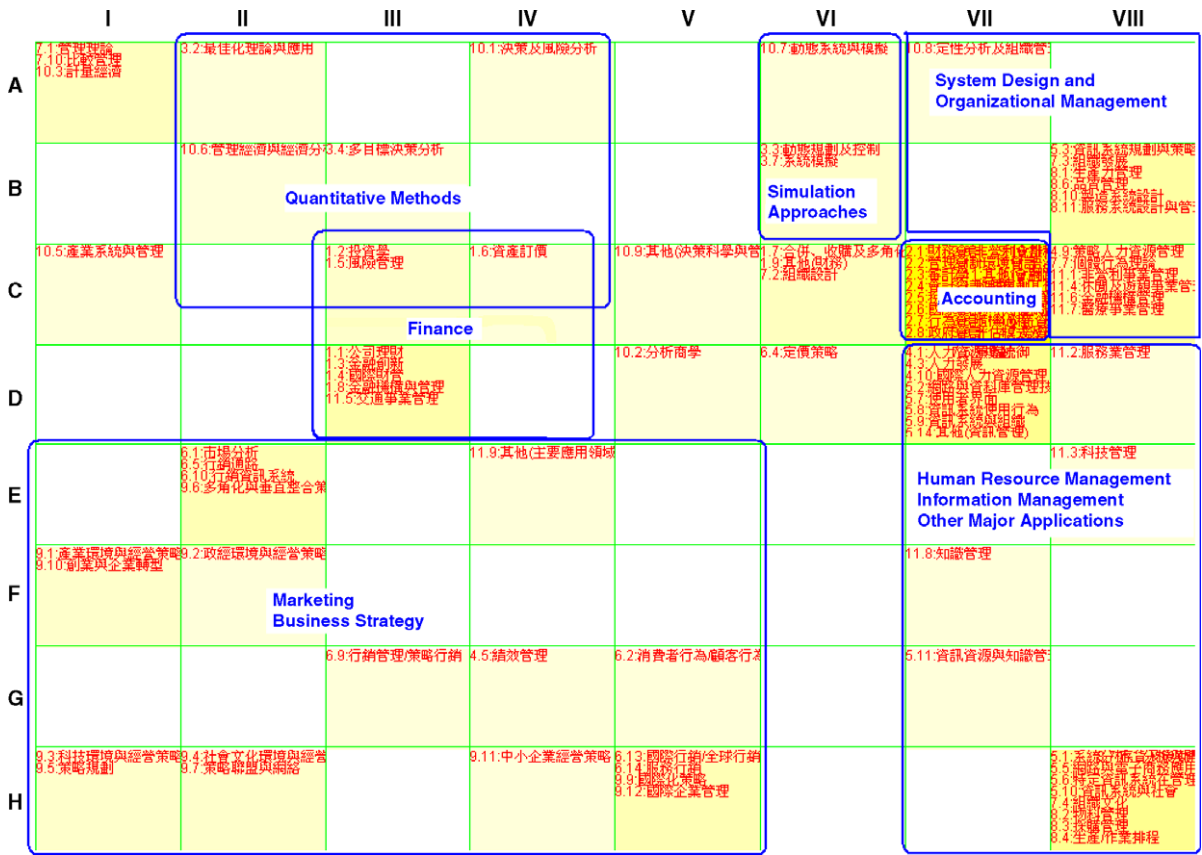


Fig. 4. A sample SOM map of fields of expertise.

following first-level fields: (4) Human Resource Management, (5) Information Management, and (11) Major Application and Others.

Several less quantitative finance fields were positioned in Region D-III, close to Regions C-III and C-IV, which are also finance fields but more quantitative. The remaining portion of the map is mainly occupied by the research fields “(6) Marketing” and “(9) Business Strategy.” The grouping indicates the intensive interactions between marketing study and business strategy development in current business practices.

The expertise field map generated by our visualization techniques revealed meaningful groupings of research fields based on experts’ co-occurrence patterns in multiple research fields. Although we can obtain such grouping of business research fields from other knowledge sources (e.g., expert knowledge on the business research areas), it is typically more difficult to find such knowledge sources for less well-known expertise domains, such as the expertise space of a particular organization. Our results demonstrate the potential of using a typical expertise data set and visualization algorithms to achieve automatic mapping

of the expertise field structure of a generic expertise space. Such visualization results have the potential to characterize the expertise landscape of an organization. Combining these expertise field visualization results with the expert map presented in the previous section may largely improve the efficiency and effectiveness of expertise searching and browsing in an organizational environment.

### 5. A study based on cognitive fit theory

Both the SOM and MDS techniques presented above construct graphical expertise data representations that use map metaphors to present the inter-expert similarities. These graphical representations have the potential to facilitate more effective and efficient exploration of the expertise space of an organization than less visual representations such as the tabular and list representations. The research literature suggests that effects of data representation on decision-making performance are dependent on the particular task types under investigation. In this section, we present an empirical study guided by the cognitive fit theory to understand the task-

level effectiveness of the SOM and MDS graphical representations of expertise compared to the traditional tabular representation.

In Section 5.1, we first review the research literature on cognitive fit theory that provided guidance for our study of the task-level effects of expertise visualization. We then present in Section 5.2 five expertise space exploration tasks customized from a generic visual task taxonomy on which the cognitive fit theory predicts superior performance with graphical representations. Section 5.3 presents the overall design of our study and the specific research hypotheses regarding effects of the display types on human subjects' task-level performances for expertise space exploration. In Section 5.4, we present an experimental study to verify the hypotheses and discuss the findings.

### 5.1. Cognitive fit theory and its application to expertise visualization

There has been a considerable amount of research on the effects of graphical and tabular representations of numerical data on decision-making performance. Overall performance comparisons of the two representations were largely inconsistent in early studies [11,16]. Later research suggested that the performances of the two representations were contingent on the particular information processing tasks. Vessey proposed the cognitive fit theory based on the information processing theory to explain the performances of the graphical and tabular representations under different types of tasks [42–44]. In her framework, the performance differences of graphical and tabular representations on different types of tasks were explained by how well the representations match the tasks. She characterized the graphical and tabular representations as presenting the same information in fundamentally different ways, with particular emphasis on the spatial and symbolic information, respectively. She also divided tasks into two types, spatial and symbolic, based on the type of information that facilitates their solutions. Performance on a task is enhanced when there is a cognitive fit (match) between the information emphasized in the representation type and the representation required by the task type. Vessey's cognitive fit theory successfully consolidated the controversial research results on performance comparison of graphical and tabular representations and provided valuable guidance to our study of the effects of the map-based visualization on expertise information exploration.

The literature on graphical versus tabular representations has focused on studying different representations

of numerical data. These studies typically have included basic data charts such as bar charts, line charts, and scatter plots that could present exactly the same information as tabular representations. For example, the relationship between two numerical values can be accurately represented by the length of the bars in a bar chart. In our study, however, the high-dimensional nature of the expertise data made the typical visual data representations not applicable. Techniques like SOM and MDS summarize the high-dimensional expertise data into inter-expert similarities that are represented by distances in a two-dimensional map. The transformation from a high-dimensional space to a two-dimensional map inevitably results in information loss and imperfect mapping. In other words, the SOM and MDS visualization results only provide a two-dimensional *approximation* of the original high-dimensional expertise data to maintain a large portion of the inter-expert similarity information. The existing literature does not provide direct answers regarding the effects of these approximate graphical representations of the high-dimensional data on tasks in comparison with other representations such as table and list representations. In this study, we were interested in testing the effects of these approximate (SOM and MDS) visualization of expertise data on spatial tasks for which the cognitive fit theory predicts better performance of graphical representations.

### 5.2. Expertise space exploration task design

A variety of tasks of different complexity levels can be performed on an expertise data set. Researchers have recognized that taxonomy of visual tasks is needed such that effects of graphical representations can be systematically interpreted within the taxonomy [17]. We adopted a well-defined taxonomy of low-level domain-independent visual tasks used in “de-featuring” visual interface evaluation studies [45,48]. Some examples of the tasks in such taxonomy include locate, identify, distinguish, categorize, cluster, rank, compare, associate, correlate, trace, and outline. These tasks are low-level tasks in that they are the elementary components of higher-level visualization tasks users may perform. They are also domain-independent, as they are applicable to visualization systems in any application domain. We believe testing the effect of expertise data visualization on these low-level tasks can provide a detailed and comprehensive understanding of the effect of visualization on expertise management tasks in general.

We selected low-level visual tasks from the task taxonomy that match Vessey's definition of spatial

tasks. Following the cognitive fit theory introduced in Section 5.1, graphical representations would facilitate users to perform better with the spatial tasks than the tabular representation because of the inherent match between the two. A main difference between our study and the previous studies is that the graphical representations we study are approximate representations of the original expertise data. We hypothesized that the cognitive fit theory can be extended to approximate graphical representations such as those generated by the SOM and MDS techniques. Thus we hypothesized that visual representations of expertise data generated by the SOM and MDS techniques would outperform symbolic representations such as the tabular representation for the spatial tasks.

Vessey’s definitions of “spatial” and “symbolic” tasks were extended from Umanath et al.’s intraset pattern and point value recall tasks [40,41]. Spatial tasks assess the problem area as a whole rather than as discrete data values. These tasks require making associations or perceiving relationships in the data. Symbolic tasks, on the other hand, involve extracting discrete data values. In our study, we selected five tasks from the visual task taxonomy that can be categorized into spatial tasks and customized them to expertise-space-exploration tasks. The definitions and formulations of these five visual tasks (associate, compare, distinguish, rank, and cluster) are presented in Table 1. The specific interpretations of these tasks in the context of expertise management are also presented in Table 1.

We use the *rank* task as an example to provide further discussions about the task selection and design. The rank task is defined as “finding the extremes (the best and the worst case)” in the visual task taxonomy. This is a spatial task according to Vessey’s definition, since it requires processing the data as a whole and perceiving relationships (comparison) between data

Table 2  
An example of table display of experts and fields of expertise

	Operations management	User interface design	Optimization theory	... Information systems
John Doe	✓		✓	
Jane Smith		✓		✓
...				
Tom Porter	✓			✓

objects. Because of the expertise data representation we adopted, it was not meaningful to find the extreme experts based on particular expertise fields. Thus we customized the formulation of this task for expertise space exploration as to be finding the expert having overlapping expertise fields with the greatest number of other experts presented. This task is important to expertise management because it deals with finding versatile experts (or experts in multiple areas) who can substitute many other experts for a variety of expertise in demand. Identifying such experts may have critical implications for business decisions such as job allocation and team formation, while a similar task formulation for document management could be less meaningful.

### 5.3. Research design and hypotheses

To evaluate the effects of the two visualization techniques, we used a table display as a benchmark. A typical table display of experts and their fields of expertise is presented in Table 2. To make the table display and the two map-based displays comparable, we also provided lists of expertise fields for each expert in addition to the table or map displays. Table 3 shows the corresponding lists of expertise fields for the table display in Table 2.

Table 1  
Visual task formulation

Task	Definition	Formulation	Interpretation
Associate	Form relationships between visual objects	Among the presented experts, which one is the most similar to the specified expert based on their research interests?	Identify substitutable or complementary experts
Compare	Make comparison among relationships between visual objects	Among the presented experts, which two experts are the most similar to each other based on their fields of expertise?	Identify substitutable or complementary experts
Distinguish	Distinguish visual objects by certain attribute(s)	Among the presented experts, which one is unlike any of the other experts based on the fields of expertise?	Identify experts with unusual expertise
Rank	Find the extremes (the best and the worst case)	Among the presented experts, which one has overlapping fields of expertise with most other experts?	Identify versatile experts (experts with expertise in multiple areas)
Cluster	Find the similarity among visual objects and form groups	Cluster the presented experts into 3 groups based on their fields of expertise.	Identify groups of experts with similar expertise

Table 3  
An example of lists of expertise fields

Expert	Expertise
John Doe	Operations management, optimization theory
Jane Smith	User interface design, ..., information systems
Tom Porter	Operations management, ..., information systems

When asked to perform certain visual tasks, the subjects were presented the lists of expertise fields of specified experts, together with one of the three display types: table, SOM map, or MDS map display. Under such a setting, subjects had the same amount of expert information (provided by the lists of expertise fields) but different assisting visual representations to complete the tasks. This allowed us to accurately measure the effects of the visual representations on different tasks.

The dependent variables in the study were the subject’s task performance, consisting of effectiveness and efficiency in completing the tasks, and display format preference. Effectiveness was measured by the percentage of correct answers and efficiency was measured by time to completion. The display format preference was revealed by the subjects at the end of the study. These dependent variables were adopted from the previous visualization evaluation studies (e.g., [45,48]) to assess the effects of data representations on information processing of human subjects. According to the cognitive fit theory, for special tasks that require subjects to form corresponding spatial mental representations, a better match would be achieved when the subjects are presented with the graphical data representation. This level of match between the problem and mental representation enhances the subjects’ information processing ability, which may result in greater effectiveness (more correct answers to questions associated with the tasks) or greater efficiency (shorter time to complete the tasks). As the third measure, the

revealed preference complements the effectiveness and efficiency measures by reflecting the potential mental burden imposed on the subjects when performing the tasks. For example, when comparable effectiveness and efficiency measures were achieved, a preference of the graphic representations by the subject may still indicate an effect of data representations on information processing because less mental burden might be involved. The three dependent variables together gives a comprehensive picture of the effect of display types on human subjects’ information processing when performing the expertise space exploration tasks. The independent variables of interest were display type and visual task type. Our overall research design is summarized in Fig. 5.

Based on the above discussion the two map-based displays, the SOM and MDS displays, should provide a better match with the mental representations required by the five spatial tasks, thus leading to more effective and efficient task performance and be preferred by the subjects. Due to the limitation of the data availability, in our study we conducted statistical tests only for the effectiveness and efficiency performance measures. Enough data points were collected by asking subjects to answer multiple questions of different task types. We also report the subjects’ revealed preferences on the three types of displays, but without statistical tests. We developed two sets of hypotheses regarding the effects of the SOM, MDS, and table displays on task performance measures.

**H1.** The two map displays will outperform the table display for all five tasks in terms of efficiency and effectiveness.

**H2.** There will be no significant differences between the performances of the two map displays for all five tasks.

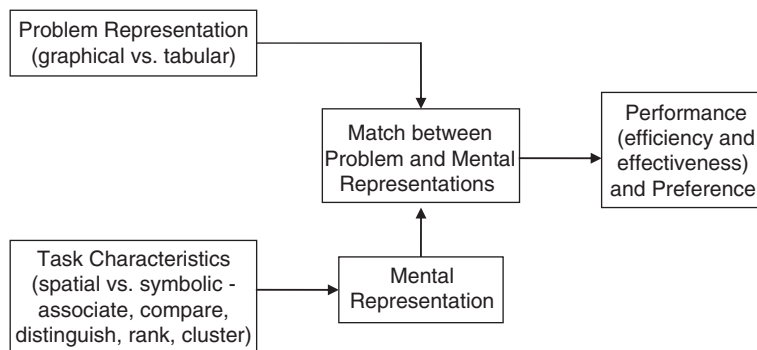


Fig. 5. Overall research design.

#### 5.4. An experimental study

##### 5.4.1. Experimental design and procedure

Six questions were designed for each of the five tasks presented in Section 4 following the formulation presented in Table 1. Each question involves a set of presented experts. The only difference among the six questions of each task was the different sets of experts presented to the subjects. The number of experts involved in these questions ranged from 6 to 14. Three of the six questions were designed to involve fewer experts (6–8) and the other three involved more experts (12–14). The number of experts presented generally determined the difficulty level of the questions. We designed the questions such that even questions across different tasks involved different sets of experts to avoid potential learning effects of subjects benefiting from their experiences from earlier questions. Each subject was required to answer all 30 questions (6 questions  $\times$  5 tasks). The correct answers to the questions were determined by Jaccard's similarity function as shown in (3). The questions were designed to have unambiguous answers and closed-form multiple choice questions were designed to reduce the complexity and subjectivity of interpreting the experimental results.

For example, a cluster task question takes the form "cluster the presented experts into 3 groups based on their fields of expertise." When designing such a question involving 6 experts, we selected 6 experts from the data set that fell into 3 unambiguous clusters based on their expertise-field-vector-based similarities computed by Jaccard's function. One of the three displays (table, SOM, and MDS displays) and the detailed listing of the expertise fields of the 6 experts were presented to the subject together with the correct answer and four incorrect ones as possible answers/choices.

A Web-based system was developed for question display and answer submission during the experiment. We randomized the assignment of display types and the order of presentation across all subjects for each visual task. The random assignment assured that each question was presented with all three display types to a similar number of subjects and that each subject answered 2 questions for each task using each of the three types of display types (2 questions  $\times$  3 displays  $\times$  5 tasks). The random assignment also crossed different difficulty levels of the questions, thus one of the two questions for each subject–task–display combination was easy and the other was difficult. This enabled us to obtain the information on overall performance of all subjects for each question using different display types, and

individual performance differences from using different display types for each of the tasks.

Twenty-eight graduate students in the business school participated the experiment. The subjects came from a variety of departments within the business school, including Management Information Systems, Accounting, Finance, Marketing, and Public Administration. Ten subjects stated that he/she was familiar with the concept of self-organizing map, Nine stated that they had introductory-level knowledge of multi-dimensional scaling, and eight stated were familiar with similarity functions. These also indicated that the degree of familiarity was at the introduction level.

A training session was provided to introduce the basic features of the map display and how maps can help in completing different tasks. The similarity function that determines the correct answers was also presented to the subjects. Using an example question, we demonstrated the benefit of using a map as an assisting tool in completing the questions. We also showed the distortion introduced by the dimension reduction techniques. The subjects were explicitly informed that they could not rely only on the maps to obtain the accurate answer to the questions and that they were expected to use the map as a supporting tool to evaluate the original expertise information listing. A bonus reward for subjects with the best performance (in addition to the participation compensation) was set before the experiment started, and the subjects were explicitly notified that the performance is measured by both the correctness and speed of completion. This setting was intended to encourage subjects to perform to their abilities in both effectiveness and efficiency. Based on the experimenters' observation, most of the subjects in the experiment did attempt to answer questions correctly and were considerably serious about the best performance reward.

After the subjects completed all 30 questions, a post-test questionnaire was used to collect preference information on the table display and map display (the information about whether the map is derived from SOM or MDS was hidden from the subjects to avoid potential biases).

##### 5.4.2. Experiment results and discussion

Two-sample *t*-tests were conducted to evaluate the differences in subjects' performance measures when presented the table, SOM map, and MDS map displays for each visual task. For accurate comparison, simple heuristics were used to remove the outliers for each display–question combination. The analysis results are summarized in Tables 4 and 5.

Table 4  
Efficiency and effective measures

Task type	Average efficiency (s)			Average effectiveness (percentage of correct answers)		
	Table	SOM	MDS	Table	SOM	MDS
Associate	64.85	49.69	50.40	81.25%	76.60%	87.76%
Compare	59.64	43.78	45.39	94.12%	88.00%	87.50%
Distinguish	54.16	45.31	45.37	50.00%	65.31%	52.94%
Rank	58.80	58.95	56.31	54.17%	57.14%	59.18%
Cluster	68.22	64.94	49.90	56.00%	62.26%	57.45%

Based on the results shown in Table 5, we drew the following conclusions:

- *Map displays generally outperformed the table display in achieving higher efficiency.* The general superior performance of the map displays was attributed to their succinct visual presentation of similarity information. Using the map displays, where inter-researcher similarities were captured in geographic distances in the map, subjects could quickly capture a large amount of similarity information visually. This enabled subjects to complete the tasks more efficiently than when using the table display. For example, when performing the *associate* task, a subject could identify a small number of candidate researchers who were positioned close to the specified researcher and check against the lists of expertise fields to determine the correct answer. When using the table display, a subject needed to first identify the research fields associated with the specified researcher and then to browse the entire table to determine the most similar researcher by evaluating the research field overlaps. For a more difficult task, the *cluster* task, the map displayed intuitive groupings of similar researchers, while forming the similar groups with the table display

required the subject to examine the research field distribution patterns in the table.

However, the benefits of the visualizations might be partly cancelled out by distortion introduced when reducing dimensionality. This led subjects to use the map displays as a supporting tool to identify candidate answers and then to check with the lists of expertise fields to determine the final answer. Using the map display in this way sometimes was as difficult as using the table display.

Specific task-level conclusions are summarized as follows.

- The *SOM* display was observed to achieve *higher efficiency* (shorter completion time) than table display for the *associate*, *compare*, and *distinguish* tasks (at 10% significance level).
- The *MDS* display was observed to achieve *higher efficiency* than table display for the *associate*, *compare*, *distinguish* and *cluster* tasks (at 10% significance level).

Note that all five tasks we have chosen in this study were the spatial tasks under Vessey's definition and were supposed to benefit from the visual presentations if such presentations are accurate reflection of the original data. The task-dependent effects of the *SOM* and *MDS* displays result jointly from the cognitive fit effect and the distortion effect. Our results show that the distortion introduced by dimensionality reduction visualization techniques such as *SOM* and *MDS* might be material for certain spatial tasks (such as the rank task) but not for others (such as the *associate*, *compare*, and *distinguish* tasks).

- There were generally *no significant efficiency differences between the SOM and MDS display types*, with one exception, the *cluster* task, for which *MDS* display achieved significantly higher efficiency (at 10% significance level). The results

Table 5  
Efficiency and effectiveness differences

Task type	Table (T) vs. SOM (S)		Table (T) vs. MDS (M)		SOM (S) vs. MDS (M)	
	Efficiency	Effectiveness	Efficiency	Effectiveness	Efficiency	Effectiveness
Associate	T < S ( $p=0.012$ )	No difference ( $p=0.292$ )	T < M ( $p=0.019$ )	No difference ( $p=0.191$ )	No difference ( $p=0.900$ )	No difference ( $p=0.158$ )
Compare	T < S ( $p=0.008$ )	No difference ( $p=0.144$ )	T < M ( $p=0.016$ )	No difference ( $p=0.131$ )	No difference ( $p=0.803$ )	No difference ( $p=0.940$ )
Distinguish	T < S ( $p=0.074$ )	T < S ( $p=0.065$ )	T < M ( $p=0.026$ )	No difference ( $p=0.386$ )	No difference ( $p=0.766$ )	No difference ( $p=0.212$ )
Rank	No difference ( $p=0.490$ )	No difference ( $p=0.385$ )	No difference ( $p=0.329$ )	No difference ( $p=0.311$ )	No difference ( $p=0.658$ )	No difference ( $p=0.840$ )
Cluster	No difference ( $p=0.348$ )	No difference ( $p=0.261$ )	T < M ( $p=0.001$ )	No difference ( $p=0.443$ )	S < M ( $p=0.065$ )	No difference ( $p=0.628$ )

generally confirmed our hypothesis that visualizations generated by the SOM and MDS algorithms had similar effects on exploration of the expertise space, although the two algorithms generated quite different visualizations and introduced different distortions.

- Generally *no significant differences were observed in effectiveness* of different display types. The one exception was the *distinguish* task, for which the SOM display achieved significantly higher effectiveness than the table display. Lack of differences in task effectiveness might have been due to the insignificant information loss during projection process of the SOM and MDS algorithms. However, there are several other alternative explanations. Under our experimental design, a detailed expertise list for each expert is always presented in addition to the table or map display. The subjects use the map or table displays as supporting tools and can always go back to the lists to get complete information. The effect of the distortion of introduced by the SOM and MDS algorithms might not be directly reflected in the subjects' effectiveness measures. The literature also indicates that it is possible that the subjects might have adjusted the time to spend on each question to maintain consistent level of accuracy/correctness of their answers [3]. The interpretation on the effectiveness measure alone is not clear, but the comparable effectiveness measures on all tasks assured that the efficiency improvements for the tasks discussed above were not at the price of incorrect answers to the questions and are indeed evidence of potential benefits of the distorted visualizations.
- The table display did not outperform the two map displays in either efficiency or effectiveness for any of the tasks. This was mainly due to the high dimensionality of the data. Subjects needed to perform comparisons on such large numbers of attributes that it might have been typically difficult to use the table display.

The subjective preference measure obtained from 26 subjects who completed the post-test questionnaire are summarized in Table 6. *Map display was preferred* to table display for the *compare*, *distinguish* and *cluster* tasks, which generally conform to the results from the objective measures. The subjects liked map display most for the *cluster* task, although only the MDS display achieved significantly higher efficiency for this task. Lower cognitive burden might have been the reason for users' preference for map displays in completing such tasks. Also interesting was that, for the *associate* task, subjects did not express a clear preference between map

Table 6  
Subject preferences

Task type	Preference		
	Map	Table	No preference
Associate	14	9	3
Compare	19	6	1
Distinguish	17	7	2
Rank	9	9	8
Cluster	24	0	2

display and table display, even though they achieved significantly higher efficiency using both SOM and MDS map displays.

## 6. Conclusions and future directions

In this study, we demonstrated the application of information visualization techniques to expertise management. We used a simple expertise data representation, and applied two commonly used dimensionality reduction visualization techniques, the SOM and MDS, to generate map displays that approximately captured the structure of an expertise space. Using a data set regarding the research expertise of business school professors in Taiwan, we generated expert maps and expertise field maps. Both types of maps showed reasonable grouping of experts and expertise fields. The expertise field map generally was consistent with our understanding of the relationships among business research disciplines. As a first study formally assessing the dimensionality-reduction visualization, our results show that such visualization techniques hold promise to be applied in organizational contexts to provide intuitive presentation of the structure of the organization-specific expertise space that is often not clear to the organizations and support more effective expertise management.

We also tested the applicability of cognitive fit theory to distorted visual displays of high-dimensional data by comparing visual representations of expertise data in SOM and MDS displays and a tabular display in the context of performing spatial tasks selected from a low-level, domain-independent visual task taxonomy. The results showed that the two visualization techniques achieved better efficiency in supporting the *associate*, *compare*, *distinguish*, and *cluster* tasks than the standard table display of expertise information. Most subjects in our study expressed preference for using a map display for the *compare*, *distinguish*, and *cluster* tasks. Our experimental results suggested that the benefit from the cognitive fit brought by SOM and MDS visualization outweighed the effects of distortions introduced during dimensionality reduction process for most spatial tasks.



However, for some other tasks, such as the *rank* task, the two effects may cancel one another and result in similar performances of graphical and tabular displays. The results of our study imply that organizations intend to implement visualization in their expertise management systems might want to analyze whether the frequently performed high-level expertise management tasks are composed of the elementary tasks that can benefit from visualization to make system implementation decisions.

One limitation of the research is that we did not focus on optimizing the SOM and MDS algorithm. More careful fine-tuning of the algorithm parameters such as the learning rate and neighborhood radius factor of the SOM algorithm and number of iterations in both algorithms might improve the performance of the two map displays. Another important future direction for this research is to extend the cognitive fit theory to include degree of data representation distortion among the explanatory variables in addition to task type and representation type. This would permit applying the theory to a wide range of visualization techniques that provide approximate representation of data and would guide more effective implementation of visualization support in expertise management and other types of decision support systems.

## Acknowledgements

This research is partly supported by NSF Digital Library Initiative-2, “High-performance Digital Library Systems: From Information Retrieval to Knowledge Management,” IIS-9817473, April 1999–March 2003. We would like to thank Anne Hsu for her help in initial data processing and discussion of research ideas.

## References

- [1] C. Ahlberg, B. Shneiderman, Visual information seeking: tight coupling of dynamic query filters with starfield displays, Proceedings of the ACM CHI'94 Conference on Human Factors in Computing Systems, 1994, pp. 313–317.
- [2] R.B. Allen, P. Obry, M. Littman, An interface for navigating clustered document sets returned by queries, Proceedings of the ACM Conference on Organizational Computing Systems, 1993, pp. 166–171.
- [3] J.R. Bettman, M. Zins, Information format and choice task in decision making, Journal of Consumer Research 6 (1979) 141–153.
- [4] J.D. Carroll, P.E. Green, Psychometric methods in marketing research: Part II. Multidimensional scaling, Journal of Marketing Research 35 (1995) 385–391.
- [5] M. Chalmers, P. Chitson, Bead: exploration on information visualization, Proceedings of the 15th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, 1992, pp. 330–337.
- [6] C. Chen, Visualizing scientific paradigm: an introduction, Journal of the American Society for Information Science and Technology 54 (5) (2003) 392–393.
- [7] H. Chen, C. Schuffels, R. Orwig, Internet categorization and search: a machine learning approach, Journal of Visual Communication and Image Representation, Special Issue on Digital Libraries 7 (1) (1996) 88–102.
- [8] H. Chen, A.L. Houston, R.R. Sewell, B.R. Schatz, Internet browsing and searching: user evaluation of category map and concept space techniques, Journal of the American Society for Information Science 49 (7) (1998) 582–603.
- [9] C. Chen, J. Kuljis, R.J. Paul, Visualizing latent domain knowledge, IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews 31 (4) (2001) 518–529.
- [10] T. Davenport, L. Prusak, Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, 1998.
- [11] G. Desanctis, Graphs as decision aids, Decision Sciences 15 (1984) 463–487.
- [12] M.A. Hearst, C. Karadi, Cat-a-cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy, Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1997, pp. 246–255.
- [13] M.A. Hearst, J.O. Pedersen, Reexamining the cluster hypothesis: scatter/gather on retrieval results, Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval, 1996, pp. 76–84.
- [14] T. Honkela, S. Kaski, K. Lagus, T. Kohonen, Websom—self-organizing maps of document collections, Proceedings of the Workshop on Self-Organizing Maps, 1997, pp. 310–315.
- [15] P. Jaccard, The distribution of flora in the alpine zone, New Phytologist 11 (1912) 37–50.
- [16] S. Jarvenpaa, G.W. Dickson, Graphics and managerial decision making: research based guidelines, Communications of the ACM 31 (6) (1988) 764–774.
- [17] S. Jarvenpaa, G.W. Dickson, G. Desanctis, Methodological issues in experimental IS research, MIS Quarterly 9 (2) (1985) 141–156.
- [18] H. Kautz, B. Selman, M. Shah, ReferralWeb: combining social networks and collaborative filtering, Communications of the ACM 40 (3) (1997) 63–65.
- [19] T. Kohonen, The self-organizing map, Proceedings of the IEEE 78 (9) (1990) 1464–1480.
- [20] B. Krulwich, C. Burkey, The ContactFinder agent: answering bulletin board questions with referrals, Proceedings of the 1996 National Conference on Artificial Intelligence (AAAI-96), 1996, pp. 10–15.
- [21] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika 29 (1964) 1–27.
- [22] X. Lin, D. Soergel, G. Marchionini, A self-organizing semantic map for information retrieval, Proceedings of the International ACM SIGIR Conference on R&D in Information Retrieval, 1991.
- [23] J.D. Mackinlay, G.G. Robertson, S.K. Card, The perspective wall: detail and context smoothly integrated, Proceedings of the ACM SIGCHI '91 Conference on Human Factors in Computing Systems, 1991, pp. 173–179.
- [24] Y. Malhotra, Tools@work: deciphering the knowledge management hype, Journal for Quality and Participation 21 (4) (1998) 58–60.

- [25] T.M. Mann, Visualization of WWW-search results, Proceedings of the International Workshop on Web-Based Information Visualization (WebVis'99), 1999, pp. 264–268.
- [26] M.E. Maron, S. Curry, P. Thompson, An inductive search system: theory, design and implementation, IEEE Transactions on Systems, Man and Cybernetics SMC-16 (1) (1986) 21–28.
- [27] D. McDonald, M. Ackerman, Expertise recommender: a flexible recommendation system and architecture, Proceedings of the ACM Conference on Computer Supported Cooperative Work (2000) 231–240.
- [28] A. Mockus, J.D. Herbsleb, Expertise browser: a quantitative approach to identifying expertise, Proceedings of the International Conference on Software Engineering, 2002.
- [29] E. Morse, M. Lewis, K.A. Olsen, Evaluating visualizations: using a taxonomic guide, International Journal of Human-Computer Studies 53 (2000) 637–662.
- [30] L.T. Nowell, R.K. France, D. Hix, L.S. Heath, E.A. Fox, Visualizing search results: some alternatives to query document similarity, Proceedings of the 19th Annual International ACM SIGIR Conference (1996) 67–75.
- [31] K. Olsen, R. Korfhage, K. Sochats, M. Spring, J. Williams, Visualization of a document collection: the VIBE system, Information Processing and Management 29 (1) (1993) 69–81.
- [32] M. Polanyi, The Tacit Dimension, Doubleday, NY, 1966.
- [33] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1986.
- [34] P. Resinick, H. Varian, Recommender systems, Communications of the ACM 40 (3) (1997) 56–58.
- [35] G.G. Robertson, J.D. Mackinlay, S.K. Card, Cone trees: animated 3D visualizations of hierarchical information, Proceedings of the Conference on Human Factors in Computing Systems (CHI'91), 1991, pp. 189–194.
- [36] A. Spoorri, A visual tool for information retrieval and management, Proceedings of the Conference on Information, Knowledge and Management (CIKM), 1993, pp. 11–20.
- [37] L.A. Steer, K.E. Lochbaum, An expert/expert locating system based on automatic representation of semantic structure, Proceedings of the Fourth IEEE Conference on Artificial Intelligence Applications, 1988, pp. 345–394.
- [38] R.H. Thompson, B.W. Croft, Support for browsing in an intelligent text retrieval system, Journal of Man Machine Studies 30 (6) (1989) 639–668.
- [39] W.S. Torgerson, Multidimensional scaling: i. Theory and method, Psychometrika 17 (1952) 401–419.
- [40] N.S. Umanath, R.W. Scammell, An experimental evaluation of the impact of data display format on recall performance, Communications of the ACM 31 (5) (1988) 562–570.
- [41] N.S. Umanath, R.W. Scammell, S.R. Das, An examination of two screen/report design variables in an information recall context, Decision Sciences 21 (1) (1990) 216–240.
- [42] I. Vessey, Cognitive fit: a theory-based analysis of the graphs versus tables literature, Decision Sciences 22 (2) (1991) 219–240.
- [43] I. Vessey, The effect of information presentation on decision making: a cost–benefit analysis, Information and Management 27 (1994) (1994) 103–117.
- [44] I. Vessey, D. Galletta, Cognitive fit: an empirical study of information acquisition, Information Systems Research 2 (1) (1991) 63–84.
- [45] S. Wehrend, C. Lewis, A problem-oriented classification of visualization techniques, Proceedings of the IEEE Visualization '90, 1990, pp. 139–143.
- [46] J.A. Wise, J.J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, V. Crow, Visualizing the non-visual: spatial analysis and interaction with information from text documents, Proceedings of the IEEE Information Visualization 95 (InfoViz'95), 1995, pp. 51–58.
- [47] D. Yimam-Seid, A. Kobsa, Expert finding systems for organizations: problem and domain analysis and the DEMOIR approach, Journal of Organizational Computing and Electronic Commerce 13 (1) (2003) 1–24.
- [48] M.X. Zhou, S.K. Feiner, Visual task characterization for automated visual discourse synthesis, Proceedings of the Conference on Human Factors in Computing Systems (CHI'98), 1998, pp. 392–399.
- [49] B. Zhu, H. Chen, Communication-Garden System: Visualizing a Computer-mediated Communication Process, submitted to ACM Transactions on Information Systems, 2002.
- [50] B. Zhu, H. Chen, Information visualization, in: B. Cronin (Ed.), Annual Review of Information Science and Technology, Information Today, Medford, NJ, 2005.



**Zan Huang** is an Assistant Professor of the Department of Supply Chain and Information Systems at the Pennsylvania State University. His research interests include recommender systems, data mining and text mining for bioinformatics and financial applications, knowledge management technologies, and experimental economics-related research for electronic markets. His articles have appeared in *ACM Transactions on Information Systems*, *Journal of the American Society for Information Science and Technology*, *Decision Support Systems*, *Journal of Nanoparticle Research* and other publications. He received the B.Eng. degree in Management Information Systems from Tsinghua University, Beijing, China and the Ph.D. degree in Management Information Systems from the University of Arizona.



**Hsinchun Chen** is McClelland Professor of Management Information Systems at the University of Arizona and Andersen Consulting Professor of the Year (1999). He received the B.S. degree from the National Chiao-Tung University in Taiwan, the MBA degree from SUNY Buffalo, and the Ph.D. degree in Information Systems from the New York University. He is author/editor of 10 books and more than 130 SCI journal articles covering intelligence analysis, biomedical informatics, data/text/web mining, digital library, knowledge management, and Web computing. His recent books include: “Medical Informatics: Knowledge Management and Data Mining in Biomedicine” and “Intelligence and Security Informatics for National Security: Information Sharing and Data Mining,” both published by Springer. He serves on the editorial board of *ACM Transactions on Information Systems*, *IEEE Transactions on Intelligent Transportation Systems*, *IEEE Transactions on Systems, Man, and Cybernetics*, *Journal of the American Society for Information Science and Technology*, and *Decision Support Systems*. Dr. Chen is a Scientific Counselor/Advisor of the National Library of Medicine (USA), Academia Sinica (Taiwan), and National Library of China (China), and has served as an advisor for major NSF, DOJ, NLM, and other international research

programs in digital library, digital government, medical informatics, and national security research. Dr. Chen is conference co-chair of ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2004 and has served as the conference/program co-chair for the past eight International Conferences of Asian Digital Libraries (ICADL), the premiere digital library meeting in Asia that he helped develop. Dr. Chen is also (founding) conference co-chair of the IEEE International Conferences on Intelligence and Security Informatics (ISI) 2003–2005.



**Fei Guo** is currently a doctoral student in Finance at the University of Arizona. She received her MS degree in Management Information Systems and Master of Accounting from the University of Arizona. Her research interests include mutual fund performance evaluation and institutional investor herding behaviors and have previously worked on research projects involving information visualization, user interface design, and visualization evaluation.



*tion Systems, and Annual Review of Information Science and Technology.*

**Jennifer J. Xu** is an assistant professor in the Computer Information Systems Department at Bentley College. She received her Ph.D. in Management Information Systems from the University of Arizona. Her research interests include data mining and knowledge discovery, knowledge management, social network analysis, human–computer interaction, and information visualization. Her papers have appeared in *Communications of the ACM, ACM Transactions on Information*



Pacific Journal of Finance, International Journal of Accounting and Information Systems, etc. He has also served as an ad hoc reviewer for academic journals in the fields of Accounting, Finance, and Information Management. Dr. Soushan Wu served as a chief editor of several journals in the fields of Accounting, Finance, and Information Management. Currently, he serves as the chief editor of Journal of Management (Taiwan).

**Soushan Wu** received his Ph.D. in Finance from the University of Florida in 1984. He is currently a Chair Professor and Dean of the College of Management, Chang-Gung University, Taiwan. He is also a visiting scholar at Clemson University, Hong Kong Polytechnic University. His research interests cover Management Science, Investment Science, Capital Markets, and Information Systems. He has published more than 70 articles in the areas of Finance, Financial Management, Asia-



**Wun-Hwa Chen** received his Ph.D. in Management Science from the State University of New York at Buffalo in 1989. He is currently a Professor of Operations Management in the Department of Business Administration at National Taiwan University. Prior to his current affiliation, he was an Assistant Professor in the Department of Management and Systems at Washington State University. His research and teaching interests lie in production planning and scheduling, AI/OR algorithmic design, data mining models for customer relationship management, and mathematical financial modeling. He has published articles in several journals including *Annals of Operational Research, IEEE Transactions on Robotics and Automation, European Journal of Operational Research, and Computers and Operations Research*. In addition, Dr. Chen has held the Chairmanship in Information Technology Advisory Committees of several government agencies in Taiwan and has extensive consulting experiences with many major corporations in Taiwan. He was also a visiting scholar at the University of New South Wales, Sydney, Australia, in 1995 and the University of Arizona in 2001.