

Multidimensional scaling for group memory visualization ¹

Michael J. McQuaid ^{*}, Thian-Huat Ong ², Hsinchun Chen ³, Jay F. Nunamaker Jr. ⁴

MIS Department, Karl Eller Graduate School of Management, University of Arizona, McClelland Hall, Room 430, Tucson, AZ 85721, USA

Abstract

We describe an attempt to overcome information overload through information visualization — in a particular domain, group memory. A brief review of information visualization is followed by a brief description of our methodology. We discuss our system, which uses multidimensional scaling (MDS) to visualize relationships between documents, and which we tested on 60 subjects, mostly students. We found three important (and statistically significant) differences between task performance on an MDS-generated display and on a randomly generated display. With some qualifications, we conclude that MDS speeds up and improves the quality of manual classification of documents and that the MDS display agrees with subject perceptions of which documents are similar and should be displayed together. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Visualization; Multidimensional scaling; Organizational memory; Group memory; Collaborative computing; Group support systems

1. Introduction

Information overload frustrates productive Internet use more and more every day. Searching and browsing waste time and yield unsatisfying results. Users feel that desired documents are tantalizingly near, yet buried beneath an avalanche of irrelevant, time-wasting reading matter. The Internet is like a library where new books arrive at a staggering pace, are shelved randomly, and where uncoordinated masses of tireless librarians stand ready, if given a

slip of paper with a word scribbled on it, to race through the crowded aisles, retrieving 0 or 10,000 books containing the scribbled word, but nothing in between. It is a mess.

How can we make sense of the massive information overload that worsens every day? We can exploit our own vision system. Half of the human brain's neurons are associated with vision [13], suggesting not merely a ready audience for information visualization, but highly capable partners. Recent advances in computer graphics technology have begun to exploit the cognitive opportunity enough to identify visualization as a field of research. In the 9 years since the first paper naming it as a field, a flood of research has taken the visualization label. In this brief, productive period, numerous developments have transpired that could be applied to business, but few applications explicitly meant for business have appeared in the visualization literature.

^{*} Corresponding author. E-mail: mmcquaid@bpa.arizona.edu

¹ For color versions of the figures, please see the URL, <http://bpa.arizona.edu/~mmcquaid/dssfigs.pdf>, which will be maintained until August 2001, whereupon a forwarding URL will replace it.

² E-mail: tong@bpa.arizona.edu.

³ E-mail: hchen@bpa.arizona.edu.

⁴ E-mail: jnunamaker@bpa.arizona.edu.

The overall goal of our research is to reduce the problem of information overload, specifically in browsing and searching document collections, through information visualization. Although the broad domain of our work encompasses the Internet, we have chosen to narrow our focus, to start with, by working with the group memory generated by electronic meeting systems (EMS) [15]. Each meeting mediated by an EMS, like GroupSystems (groupware developed at the University of Arizona MIS Department, and marketed by Ventana), generates a log that can be used by future team members and others with a stake in the team's work. These logs are underutilized in practice. Our immediate goal is to produce and experiment with a series of maps/snapshots of meetings.

Fig. 1 shows just such a map. A stake holder could flip through this set of maps/snapshots to find relevant meetings and only inspect interesting meetings. Further, only the interesting portions of the relevant meetings need to be viewed, since clicking on a region of the map brings up only the text of the meeting related to that region.

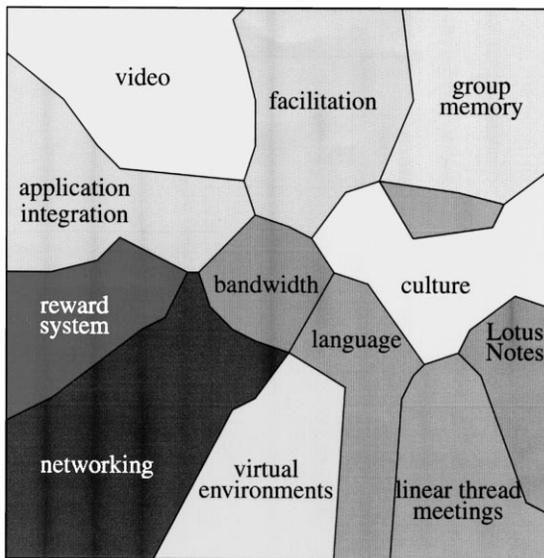


Fig. 1. An interactive map/snapshot of an electronic meeting with the topic, "Brainstorming: What are the most important information technology problems with respect to collaborative systems to be solved over the next 5 years?"

Previous experiments at the University of Arizona have shown that a neural network technique, based on Kohonen's Self-Organizing Feature Map (SOM), offers a viable alternative to traditional information retrieval methods for effective browsing [4]. The SOM system constructs a map (different from the map in Fig. 1) of terms or documents automatically, showing relationships that would not be apparent in keyword search or browsing through indexed lists. Major limitations remain to be overcome with this technique, opening the way for other techniques to replace or be synthesized with the SOM.

This paper describes an alternative technique from statistics, multidimensional scaling (MDS). MDS cannot directly replace SOM, since it only determines distance between objects. Two important properties of the SOM are that it applies (1) boundaries to regions containing objects, as well as (2) text labels for the bounded region. Our goal in the present research is to demonstrate the validity of MDS as a tool to locate objects relative to each other to improve user execution of a classification task. Success here would give us a foundation on which to use MDS as a visualization tool.

The remainder of the paper includes the following sections. In Section 2, we give a brief review of the visualization literature, focusing on a definition of visualization, and the history and taxonomy of the visualization literature. In Section 3, we describe the systems development methodology, which specifies an interplay between the development of a computer-based system and three research activities (experimentation, observation, and theory building) that system development contributes to. In Sections 3.2, 3.3 and 3.4, we describe our MDS system. In Section 3.5, we propose hypotheses to be tested with the working prototype of our MDS system. In Section 4, we give results of the test. In Section 5, we give conclusions based on the test results and present qualifications and future directions.

2. Visualization

2.1. Definition of visualization

The term *visualization* was coined in a seminal 1987 paper [13] advocating a federally funded initia-

tive to support a research stream, visualization in scientific computing, already recognized as distinct from computer graphics, but unnamed for several years.

Broad claims were made for visualization in the definition given in Ref. [13]. Visualization was said to be a method for seeing the unseen, for transforming the symbolic into the geometric, for generating images and interpreting images, and more. Visualization was claimed to unify diverse fields. While it may have been necessary to subsume so many disciplines and offer so much in a funding initiative, people working in visualization need a more practical definition.

Such a definition is suggested in Ref. [5], where some of the discussions and citations might lead to the following definition (not directly stated in the article):

Visualization is transformation and analysis to aid in the formation of a mental picture of symbolic data. Such a picture is simple, persistent, and complete.

The three qualities (simplicity, persistence, and completeness) help to assess the success of a given visualization. They distill some principles promulgated by William Playfair in his *Commercial and Political Atlas*, London, 1786, the first European example of time-series graphics [17]. Simplicity measures the degree to which the visualization is immediately understandable; persistence measures the propensity for the visualization to linger in the mind of the beholder; and completeness measures the extent to which all relevant information in the data is depicted.

One difficulty with such a definition is whether it blurs the distinction between visualization and other fields of research. For instance, in the capstone address to the IEEE Conference on Visualization 1994, Bela Julesz identifies the central difference between human and machine vision (focal attention). Should this be classified as research in vision or visualization? What about user interface studies that mix kinesthetic and visual information?

Since this paper simply explores a new area with a view to application to a domain of interest, we will not pursue a comprehensive definition. We will be

satisfied by a working definition, accepting any inconsistencies that do not hobble the stated purpose.

2.2. Motivation for visualization

That visualization is a burgeoning area of research cannot be disputed. Although this can be, in part, attributed to opportunities afforded by improvements in computing and, in part, to our vision-oriented brains, there is more — some problems simply need visualization, e.g., Anscombe [1] has shown four very different datasets that can be portrayed by precisely the same linear model. Visualizing the datasets demonstrates four contrasting patterns. A numerical view of this data misses meaningful distinctions without investigating residuals, yet the visual representation speaks simply, memorably, and comprehensively.

Tufte [20,21] offers a magisterial catalog of phenomena best depicted visually, as well as prescriptions for the best visual depictions. The above-cited text chronicles some very persuasive and even dramatic examples of the importance of visualization of information, ranging from a chart instrumental in freeing a crime boss to a data map used to halt a cholera outbreak that had claimed 500 lives. Compelling reasons to study visualization abound.

2.3. History

While information visualization has been documented in Asia for at least 850 years, the work done in the computer era has acknowledged only a European tradition, beginning soon after the invention of the printing press. Within this tradition may be found all the major techniques applied in the computer era. These techniques are documented in the above-cited sources and include scatter plots, time-series charts, contour plots, vector field maps, color maps, and icon plots.

Currently, The IEEE Visualization Conference groups studies under the following categories, also typically seen in federal funding documentation: flow visualization techniques and algorithms; volume visualization techniques and algorithms; surface rendering, often grouped under volume visualization; visualization applications (often with strong domain orientations in fields such as medicine, meteorology,

or geology); human factors/user interfaces for visualization and animation. Our work focuses on the visualization application area.

3. Research methodology

3.1. Systems development approach

We describe an example of the systems development approach to research, described in Ref. [14]. For the present research, we have built a tool and performed an experiment with its use, with the eventual goal of developing theories related to information overload, the overall challenge unifying our work.

The overall domain of the research is the problem of information overload, particularly as it confounds *browsing* [12]. Although scalability is a goal, this first experiment focuses on browsing small collections of EMS logs. The contribution is to construct a map to aid browsing for classification and to evaluate the map.

The UAMIS AI Group prior research in this area utilizes category maps built using Kohonen's SOM technique [4,3]. Protocol analysis has been used to evaluate this technique, which is an implementation of a category map approach, where materials representing similar concepts are grouped. In this case, the material may consist of terms (single words and term phrases) and documents, from which terms have been extracted.

For the prior research, the approach has been to:

1. collect documents;
2. index documents, using automatic indexing or noun phrasing;
3. perform a co-occurrence analysis of the terms contained in the documents; and
4. build a category map using SOM for browsing.

The present research mainly addresses and involves step 4. We use MDS to develop a map to aid users in a classification task.

3.2. Technique: MDS

MDS provides a way to “map” higher dimensional relationships in data to a lower dimensionality [6]. Traditionally employed on small data sets in the

psychometric literature, MDS has recently been applied to molecular chemistry, occasioning research [19] into the scalability of MDS.

Can MDS help in the present situation? Two well-known features of MDS may apply to this situation.

First, the use of an information-loss measure may improve precision, since it should more accurately map “distances” between terms than any procedure without such a measure.

Second, continuously — rather than discretely — valued distances may enable greater control by novices, since a radius may be meaningfully specified on a two-dimensional display of coordinates returned by MDS, while such a measure would merely act as a convenience without precise meaning in the grid display of the SOM.

3.3. MDS application

One way to apply MDS is to take a set of p -component vectors and to approximate them by a set of two- or three-component vectors. Fig. 2 shows four objects, A – D , described as four four-component vectors, $\langle 0,1,2,3 \rangle \dots \langle 3,2,2,0 \rangle$. MDS approximates them by four two-component vectors, which can then be displayed in a Cartesian coordinate system. Note that there is some information loss in the course of reducing the dimensionality. In this example, we can easily place A , B , and C , but when we come to D , we find that in two dimensions, there is no way to maintain all three relationships with the other points. MDS tries to minimize the divergence from a “truthful” representation of the relationships.

Fig. 2 shows a set of vectors that can be interpreted as a “dissimilarity” matrix, a kind of distance matrix. Objects have zero distance (dissimilarity) from themselves, and object B is equidistant between objects A and C , but more distant from D . D

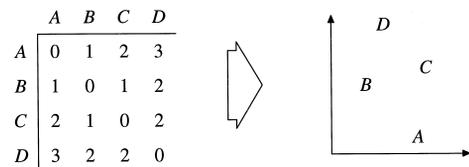


Fig. 2. MDS gives a lower-dimensional approximation of a higher-dimensional description.

is equidistant between B and C , but more distant from A . Portraying all these relationships simultaneously, while minimizing information loss, is the work of MDS in this application.

3.4. MDS process

To do MDS in the context described above, we perform several steps, beginning with the matrix of dissimilarities between objects (in this case, the objects are documents in an electronic meeting database, each document containing terms).

(1) Obtain an initial configuration using singular value decomposition (SVD). This approach has the advantage that the amount of improvement possible in this step serves as a guide to the difficulty of the problem and the degree to which one can expect pleasing results from MDS.

(2) Measure the information loss between the original matrix and the initial configuration using the STRESS metric of Kruskal [10].

(3) Find a new configuration with smaller information loss than the initial configuration using an isotonic regression algorithm to obtain fitted distances and a conjugate gradient descent algorithm to optimize.

(4) Repeat until a threshold information loss is reached or until a threshold number of iterations are performed.

Forsythe et al. [7] describe an easily implemented SVD algorithm. The original description is described by Wilkinson [22], but is not directly adaptable for coding. Press et al. [16] call SVD the only routine in their book that “you should just accept as a black box” so we have, simply using their algorithm and taking the coordinates for the first q eigenvectors, $q = \{2,3\}$, as initial coordinates.

MDS stands out among multivariate descriptive techniques due to its measure of information loss, used at each iteration to improve the configuration of points representing the location of objects in some low-dimensional space. The most influential measure is one proposed by Kruskal [10], known as STRESS or, more commonly, stress. Originally, an acronym for standardized residual error sum of squares, stress has gradually been accepted as such an appropriate word for the concept it portrays that it later spawned a variant strain that is not an acronym at all. In

Kruskal’s original formula, the stress of a configuration was termed S , and was defined as:

$$S = \sqrt{\frac{\sum (d_{ij} - \hat{d}_{ij})^2}{\sum d_{ij}^2}},$$

where the d_{ij} are interpoint distances in the present configuration, and the \hat{d}_{ij} are monotonically related to the values in the original dissimilarity matrix δ_{ij} so that $\hat{d}_{ij} \leq \hat{d}_{i'j'}$ whenever $\delta_{ij} < \delta_{i'j'}$.

A MDS research review [11] claimed that Kruskal’s measure remained popular over 20 years later, and that the many refinements offered since its introduction were small refinements. For the present research, we have chosen to use the original 1964 formula.

The stress measure should improve the precision of results over what is available with the neural net approach. The computational burden inherent in using it will remain to be addressed, but recent work in optimizing MDS [19] suggests that it can be done.

We then obtain fitted distances using isotonic regression. Grotzinger and Witzgall [9] show an $O(n)$ algorithm for isotonic regression (isotonic regression was proposed as an alternative name to monotonic regression in 1972 by Barlow et al. [2], who mention on p. 56 that monotonic could be interpreted as order-reversing). The authors have verified the optimality of their results by the satisfaction of Kuhn–Tucker conditions for the solution generated by their algorithm.

Having obtained fitted distances, we are now in a position to use them, making stress the objective function and trying different configurations of points. We have used a conjugate gradient descent algorithm [16] to obtain the new configuration.

This algorithm is an alternative to steepest descent, which requires a 90° turn at each step, whereas 90° only very occasionally turns out to be a good direction — a direction conjugate to the previous direction, though, makes the best use of the gradient information. Using this method drastically decreases search time compared to steepest descent.

3.5. Experimental hypotheses

After implementing the system described in Sections 3.2, 3.3 and 3.4, we evaluated it. We began by

recognizing that the input to MDS was based on keyword co-occurrence, so we assumed that a completely keyword-based benchmark was appropriate, and asked what we should see if MDS produced worthwhile maps in this context.

Before we began, we generated a benchmark map, shown in Fig. 3, using keywords only as a guide. This is the best categorization we can provide utilizing keywords directly. We validated this map with the help of an expert, one of the expert contributors to the original meeting on the future of collaborative

technology, and we discuss this expert's input in Section 5.

The map represents documents in the same category by dots of the same color. We see in Fig. 3 that dots of the same color are typically close together, signifying that the map tends to cluster similar documents. After developing this map, we began to form hypotheses.

First, we conjectured that if MDS properly clustered similar documents together, users would substantially agree on the positions, so we decided to

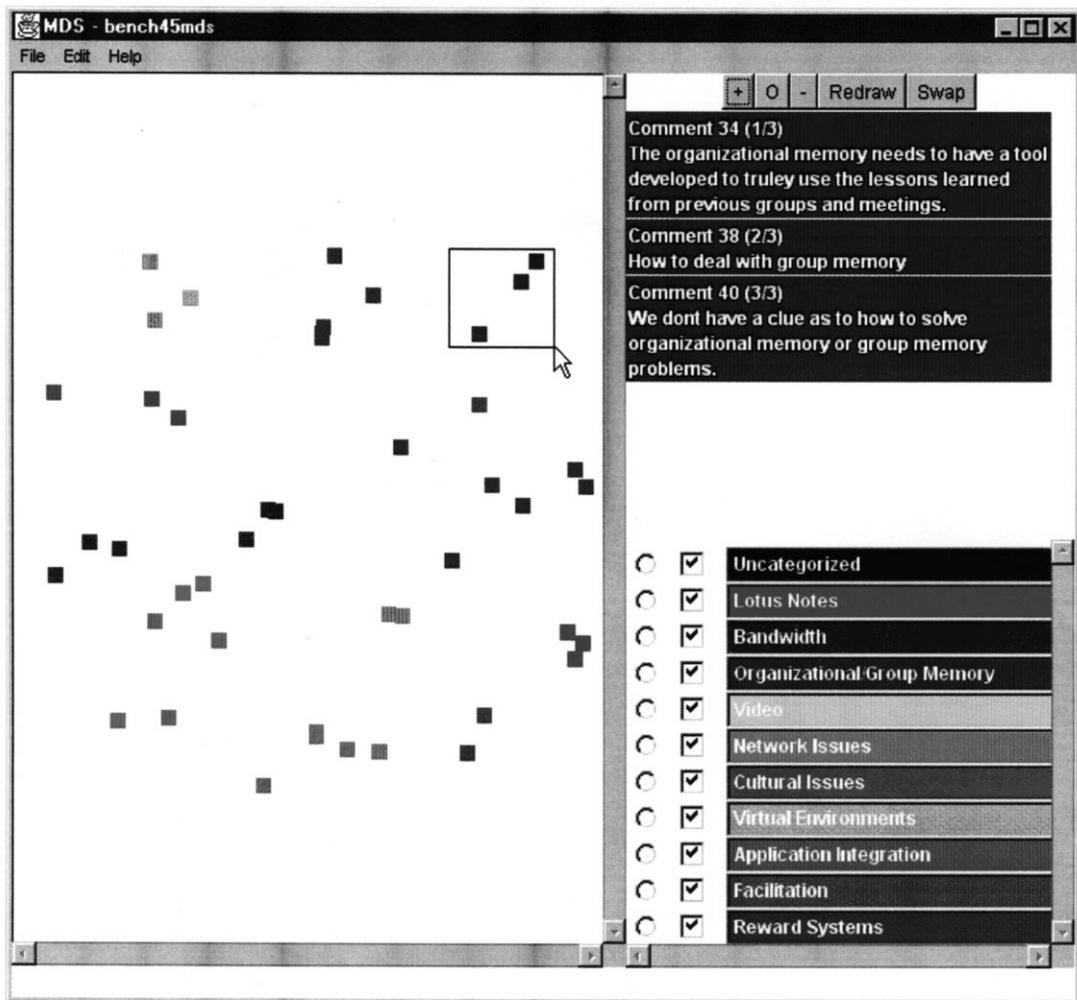


Fig. 3. Keyword benchmark, showing how a domain expert classified the material.

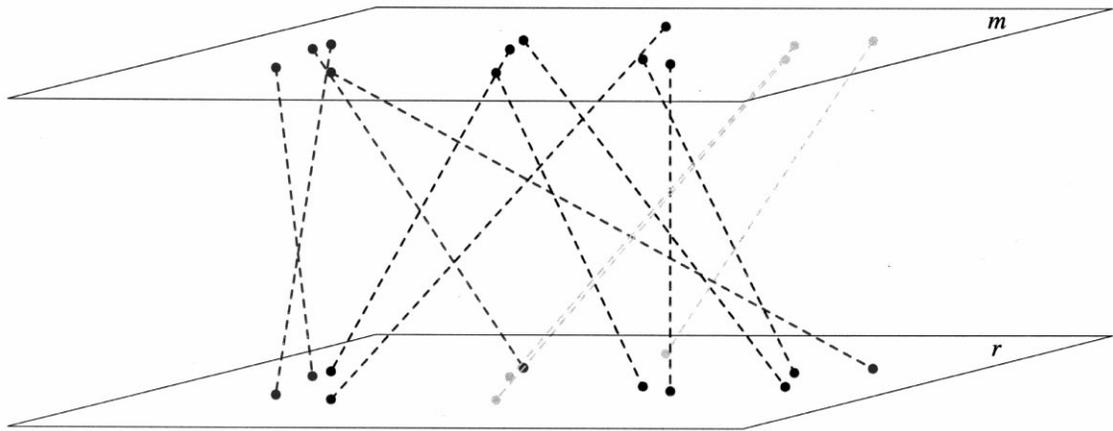


Fig. 4. Mapping m space to r space.

use a categorization task, where users would manually categorize documents into categories, with or without the aid of MDS.

Since some of the work of categorization has been done by grouping similar documents together, it should take subjects less time to perform categorization with the aid of MDS.

H₁: $\bar{t}_m < \bar{t}_r$ (The average time for subjects using MDS to categorize is less than the average time for subjects using a random display.)

The results of the categorization should reflect the fact, that MDS did a good job of grouping similar documents together. So we should be able to measure some function of the size of the clusters formed by this categorization and find that they are smaller, on average, for the subjects using our display than for subjects using a random display. Letting m and r be estimators of functions of the size of individual clusters, we should be able to show that m is on average smaller than r .

H₂: $\bar{m} < \bar{r}$ (The average size of clusters formed by subjects using MDS is smaller than for clusters formed by subjects using a random display.)

The MDS display should offer more than the potential for *faster* categorization. It should also make *better* categorization possible. To illustrate, consider a hypothetical subject. Instead of showing

the MDS display to this subject, we scramble the association of documents to coordinates as in Fig. 4 and show this subject the r display, or random display, instead of the m display, or MDS display.

Note that exactly the same documents and coordinates are used, but that we scramble the association between them when switching from the top map to the bottom map in Fig. 4. For the bottom map, each document is assigned to some other random document's dot.

The resulting display, far from showing small clusters of similar colors, is completely chaotic when shown with the benchmark color codes. Users do not see the benchmark color codes, but assign their own. What we expect is that their coloring of the randomly assigned map will be even more chaotic in that they will not be able to find or remember

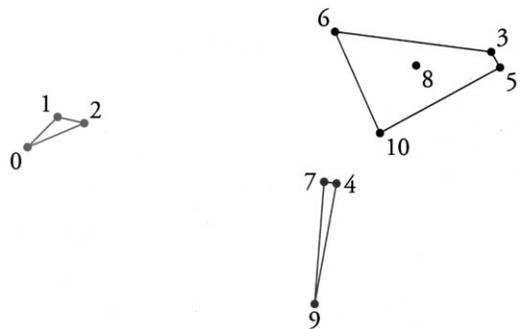


Fig. 5. Subject m_k (portion), a correct categorization on MDS coordinates.

documents in similar categories and will include some random, unrelated documents in each category.

Fig. 5 shows a portion of the keyword benchmark display in m space; i.e., the documents depicted possess keywords such that red (documents 0, 1 and 2) is the “reward category”, blue (documents 3, 5, 6, 8 and 10) is the “culture category”, and green (documents 4, 7 and 9) is the “language category”. We only show a few documents so we can easily compare these dots in different mappings (Figs. 5–8). In Figs. 7 and 8, only the reward (0, 1, 2 and 3), and culture (4, 5, 6, 7, 8, 9 and 10) categories are used.

The clusters in Fig. 5 are small and easy to pick out. A user of our map browser is very likely to drag a rectangle over each of the clusters, see the corresponding documents in another window, see that they are related, and assign categorizations like what we see in Fig. 5.

Fig. 6 shows the same documents as they are scrambled in the random mapping. Bear in mind that the entire random map has the same 45 dots as the MDS-based map, but this scrambled subset has some different dots. Fig. 6 also shows the benchmark colors, but we see that because documents of the same category are scattered, no meaningful small clusters emerge.

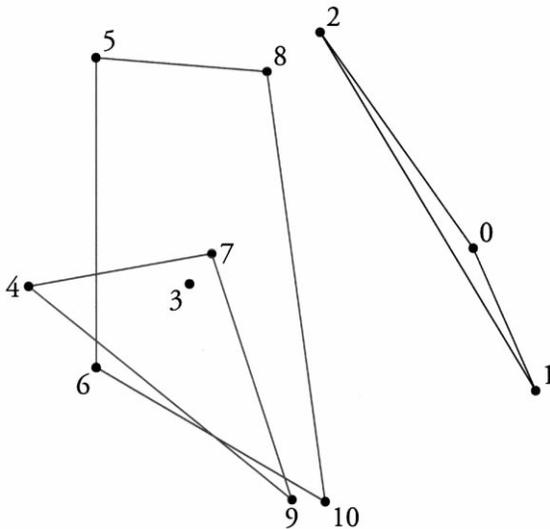


Fig. 6. Subject r_k (portion), a correct categorization on random coordinates.

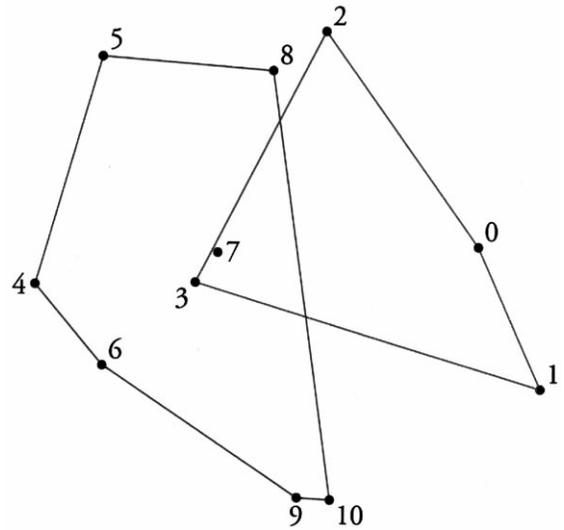


Fig. 7. Subject r_{13} (portion), an erroneous categorization on random coordinates.

The perimeters of the like-colored clusters are much larger than in the previous display and chaotically overlap clusters of other colors. We conjecture that because these similar documents are “all over the map”, it will be difficult for users to find every “language” document when building the language category, every “culture” document when building the culture category, and so on.

So we expect to find quite a few results such as what we later found with subject r_{13} , depicted in Fig. 7. This subject was so confused by the random display as to be unable to appreciate the actual categories. Bewildered by the randomly generated map, this subject simply placed all these documents into two categories, neither of which is consistent with the benchmark categorization. Yet, like most of the subjects viewing the randomly generated display, this subject found many documents that were properly assigned to the same category. How can we tell?

We measure a function (again the perimeter) of the size of the cluster as it would appear if we displayed the same erroneous coloring (categorization) the subject (r_{13}) assigned, but with the dots displayed as the MDS coordinates (Fig. 8). Inevitably, we found that these artificially generated maps were better (in the sense of tighter clusters)

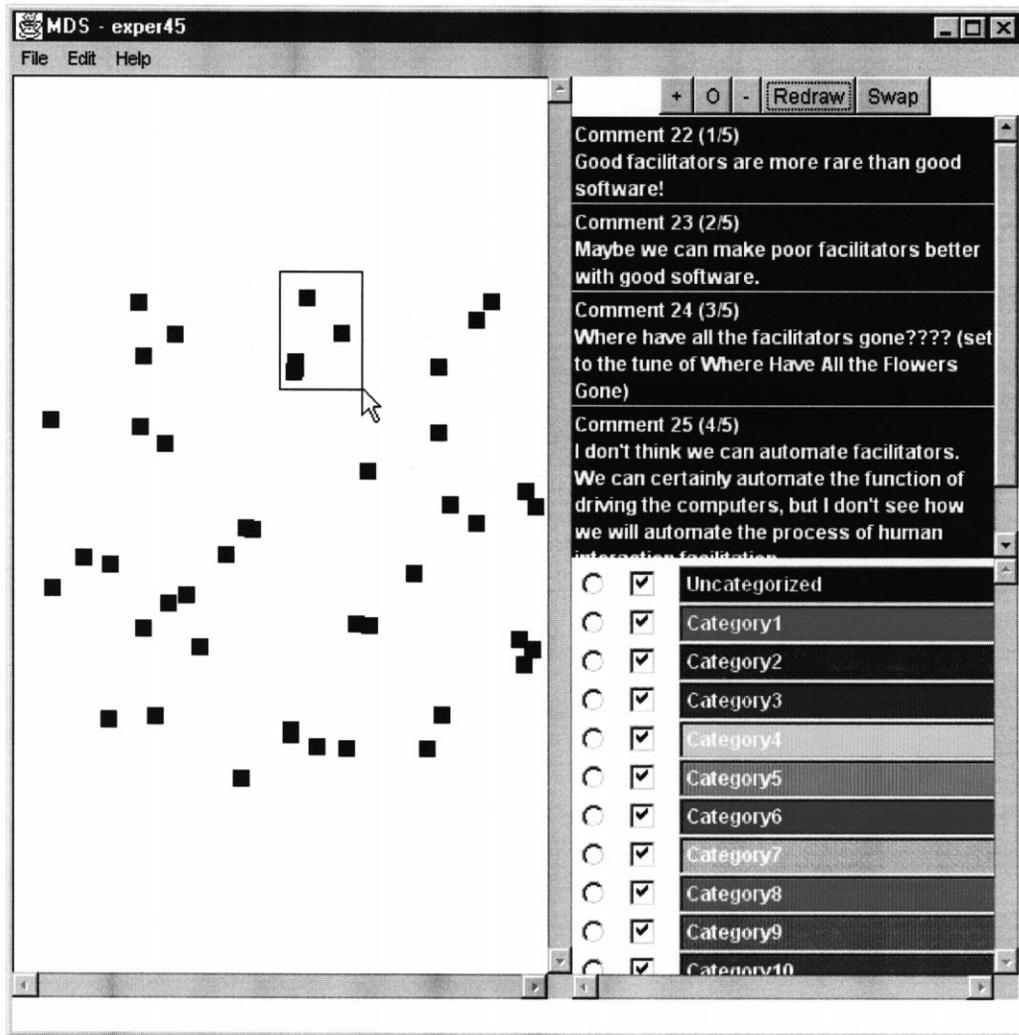


Fig. 8. Subject p_{13} (portion), an erroneous categorization on the MDS coordinates.

than the random maps, but not as good as those created by subjects using the MDS-based display.

Evidence of this can be seen by comparing Figs. 7 and 8. Not only are the clusters generated by the subject viewing a random display larger than those of m users, but they would *still* be larger, even if we put the dots as close together as they would be on the corresponding MDS display, as in Fig. 8.

If a subject viewing the r display assigns most of the documents to the same clusters as the keyword benchmark, but loses track of a few, then the average size of those clusters, when assigned to the m coordinates,

will be larger than the average clusters as assigned by those subjects viewing m space. We refer to the projection of r points into m space as p space. This space shows a categorization done “blindfolded” then displayed on an MDS generated map.

H₃: $\bar{m} < \bar{p}$ (The average size of clusters formed by subjects using MDS is less than the average size of clusters formed by subjects using a random display, but are then displayed using the MDS-generated coordinates.)

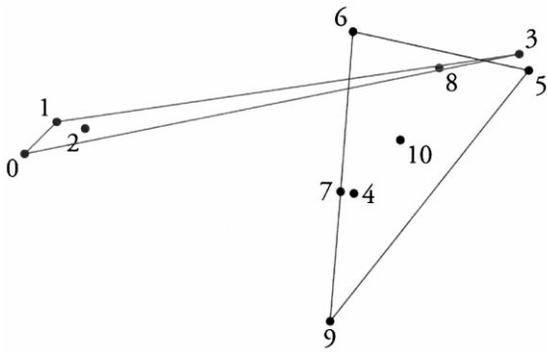


Fig. 9. Experimental interface, selecting one cluster, prior to any categorization.

Satisfaction of this hypothesis is evidence of the following. MDS visualization helps subjects see connections between documents that would be missed in a random ordering of documents.

4. Results

Sixty subjects, mostly students of MIS and ECE at the University of Arizona, were tested using the experimental interface shown in Fig. 9. The subjects were presented with a screen full of clumps of dots. A menu allowed them to color dots according to categories, as well as to type a name in each of the selected categories.

Subjects could see the content of a single document by selecting the associated dot or by dragging a group of dots, as shown in Fig. 9. Selected dot(s) could then be assigned a color, and the label for that color modified. Fig. 3 shows an almost completed experiment, in which all documents have been colored and all colors labeled with a category name.

Forty-five documents were used. Samples of two categories may be seen in Figs. 3 and 9. All the documents were generated in a meeting about the question captioned in Fig. 1.

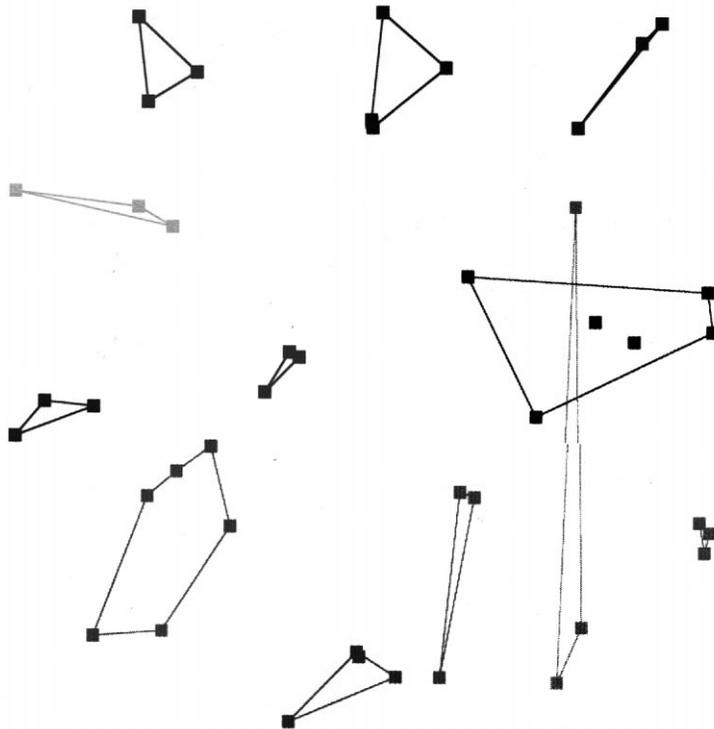


Fig. 10. The categories in the MDS display.

The subjects were divided into two groups, with 30 subjects randomly assigned to each group. (Because some of the subjects did not understand the instructions and gave nonresponsive results, some additional subjects were introduced to bring the total of usable subjects to 60.)

One group of subjects (the *m* group) was presented with a display generated by MDS. The other group (*r* group) was presented with a randomly generated display. A third group (the *p* group) was created by projecting the points from the randomly generated space into the MDS space. A typical comparison between the *m* group and *r* group is shown in Figs. 10 and 11, where we see plainly that the clusters specified by the *m* group are drastically smaller than those specified by the *r* group.

We selected the perimeters of the convex hulls of the clusters, in other words, the least enclosing convex polygons of the clusters, as measures of cluster size, and wrote a program to calculate the average convex hull perimeters for each group based on Ref. [8]. Given sets of points, the program returns the average perimeters of the convex hulls of the sets of points (Figs. 12 and 13).

4.1. $H_1: \bar{t}_m < \bar{t}_r$

Using *S+*, we tested the differences in timing between the *m* group and the *r* group. A *t*-test

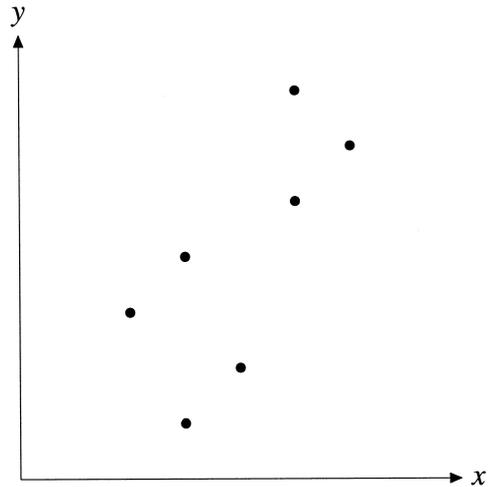


Fig. 12. A set of points.

shows that the difference is statistically significant at any measurable level.

An interesting point here is that two of the subjects were identical twins, both pursuing the same degree program. Although we randomly placed subjects into groups, and the twins happened to fall into different groups, we would not have been able to resist separating them had chance thrown them together. Their times were 21 min (*m*) and 28 min (*r*), respectively (both within a minute of the means for

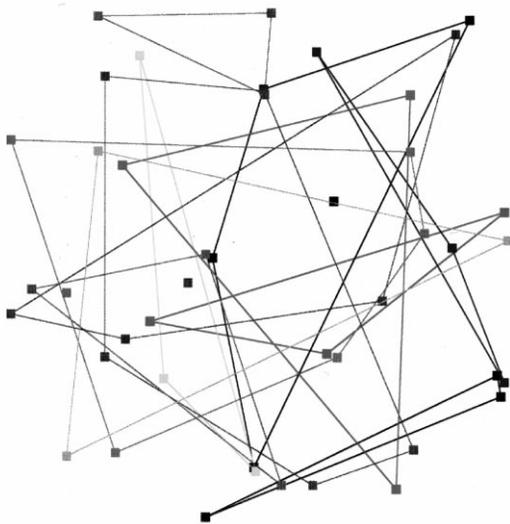


Fig. 11. The same categories as in the MDS display, but shown in the random display.

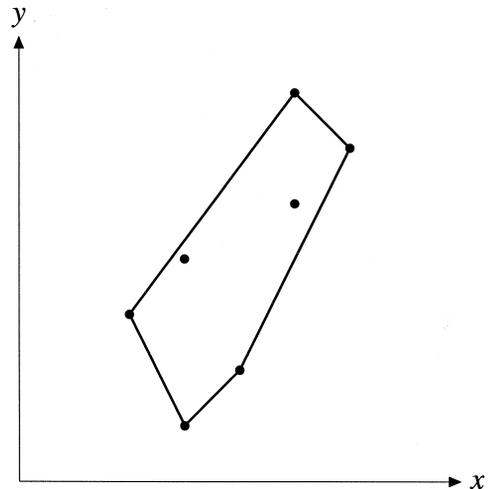


Fig. 13. The convex hull of the set of points shown in the previous figure.

their respective groups). The MDS display appears to cause an effect on their categorization.

\bar{m} subjects take significantly less time than \bar{r} subjects

> *t.test* (time[class = "m"], time[class = "r"])

Standard two-sample *t*-test

Data: time[class = "m"] and time[class = "r"]

$t = -2.541$, $df = 58$, $p\text{-value} = 0.0138$

Alternative hypothesis: true difference in means is not equal to 0

95% confidence interval: -12.846415 , -1.525109

Sample estimates: mean of $x = 21.62069$, mean of $y = 28.80645$

4.2. $H_2: \bar{m} < \bar{r}$

Using *S+*, we tested differences between the average sizes of the convex hulls of the clusters generated by the *m* group and by the *r* group. For this test, each subject is taken to have a single average convex hull, the average of the convex hulls of the clusters identified by that subject. There are 60 data points in this test data set, one for each subject in the *m* group and one for each subject in the *r* group. The output shows a striking difference between the average sizes, with a *p*-value close to zero. Additionally, the variability for the *m* group is less than for the *r* group.

\bar{m} is significantly smaller than \bar{r}

> *t.test* (avghullsize[class = "m"], avghullsize[class = "r"])

Standard two-sample *t*-test

Data: avghullsize[class = "m"] and avghullsize[class = "r"]

$t = -8.3818$, $df = 58$, $p\text{-value} = 0$

Alternative hypothesis: true difference in means is not equal to 0

95% confidence interval: -1.5389749 , -0.9456143

Sample estimates: mean of $x = 0.8694124$, mean of $y = 2.111707$

4.3. $H_3: \bar{m} < \bar{p}$

We compared the *m* group to the *p* group in the same way, with similarly strong results, shown below.

\bar{m} is significantly smaller than \bar{p} .

> *t.test* (avghullsize[class = "m"], avghullsize[class = "p"])

Standard two-sample *t*-test

Data: avghullsize[class = "m"] and avghullsize[class = "p"]

$t = -4.6342$, $df = 58$, $p\text{-value} = 0$

Alternative hypothesis: true difference in means is not equal to 0

95% confidence interval: -0.9774029 , -0.3877367

Sample estimates: mean of $x = 0.8694124$, mean of $y = 1.551982$

An important concern is the relationship between the number of clusters and cluster size. If the *r* group were simply grouping documents into a smaller number of clusters, false support for H_3 would be provided, since only 45 documents are available and the aggregation of documents into fewer clusters would necessarily lead to larger clusters.

We used *S+* to test for a difference between average numbers of clusters generated by the *m* group (11.3) and the *r* group (10.5). The result is that there is not a statistically significant difference between them at the $\alpha = 0.10$ level, as shown below.

r subjects do not use significantly fewer classes than *m* subjects at level $\alpha = 0.10$.

ttestcats.txt > *t.test*(cats[class = "m"], cats[class = "r"])

Standard two-sample *t*-test

Data: cats[class = "m"] and cats[class = "r"]

$t = 1.5238$, $df = 58$, $p\text{-value} = 0.133$

Alternative hypothesis: true difference in means is not equal to 0

95% confidence interval: -0.2509154 , 1.8509154

Sample estimates: mean of $x = 11.3$, mean of $y = 10.5$

Since the *p* group is simply a projection of the results of the *r* group into a different coordinate system, the number of clusters remains the same, and the corresponding *t*-test for the difference between the number of clusters is identical.

5. Conclusions and qualifications

These results confirm the hypotheses described in Section 3.5 as follows.

- Subjects using the MDS display need less time than other subjects, confirming H_1 .
- Cluster sizes generated by the m group are significantly smaller than those generated by the r group, confirming H_2 .
- Cluster sizes generated by the m group are significantly smaller than those generated by the p group, confirming H_3 .
- The average number of clusters generated by the two groups is not significantly different, eliminating the number of clusters as a possible source of the above results.

An important qualification about the above concerns the use of keywords to generate the MDS display. We assume that keywords offer an acceptable representation of the original documents. The map generated by a user employing only keywords as a guide has an average convex hull size very close to the average for the m group. Protocol analysis reveals that many subjects in both groups were consciously using keywords, especially when the subject matter was unfamiliar to the subject and the subject had no articulated construct for categorization.

An expert user (one of the expert participants in the original session) generated a map with similar average convex hull size to the m group, but with a notable difference. Protocol analysis reveals that this subject had a sophisticated construct due to extensive research into the topics, and grouped two clusters together that would ordinarily be kept separate. This subject lauded the system afterward for putting together most of the appropriate documents, calling it a major improvement over current methods.

Nevertheless, the expert subject highlights the need for a more sophisticated representation of the documents, and a future direction is to use the noun phrasing technique being developed in the University of Arizona MIS AI Laboratory [18].

Another direction is to use one-dimensional MDS to automatically generate categorizations. This can be done by scaling the results to one dimension and looking for patterns in the distances between points. These patterns allow us to break the points into a user-specified number of neighborhoods.

The third direction is, having demonstrated the validity of the approach, to attempt to scale it up to larger collections. The combination of this direction with the direction mentioned above gives us the

necessary tools to compare this system to the SOM mentioned Section 1.

Acknowledgements

Dr. Richard Orwig helped formulate experimental design. Dorbin Ng provided the output of his prior research (co-occurrence analysis and automatic indexing). Duffy Gillman contributed to the interface design. Numerous volunteers generously gave their time to perform the experiment. This project is sponsored in part by NSF/DARPA/NASA Digital Library Initiative, IRI-9411318, 1994–1998 (B. Schatz, H. Chen, et al., “Building the Interspace: Digital Library Infrastructure for a University Engineering Community”).

References

- [1] F.J. Anscombe, Graphs in statistical analysis, *American Statistician* 27 (1973) 17–21.
- [2] R. Barlow, D. Bartholomew, J. Bremner, H. Brunk, *Statistical Inference Under Order Restrictions*, Wiley, London, 1972.
- [3] H. Chen, A.L. Houston, R.R. Sewell, B.R. Schatz, Internet browsing and searching: user evaluations of category map and concept space techniques, *Journal of the American Society for Information Science* 49 (7) (1998) 582–603.
- [4] H. Chen, J.F. Nunamaker Jr., R. Orwig, O. Titkova, Information visualization for collaborative computing, *IEEE Computer* 31 (8) (1998) 75–82.
- [5] B.M. Collins, Data visualization — has it all been seen before? In: *Animation and Scientific Visualization Tools and Applications*, Chap. 1, Academic Press, London, 1993, pp. 3–28.
- [6] T.F. Cox, M.A.A. Cox, *Multidimensional Scaling*, Chapman & Hall, London, 1994.
- [7] G.E. Forsythe, M.A. Malcolm, C.B. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
- [8] R.L. Graham, An efficient algorithm for determining the convex hull of a finite planar set, *Information Processing Letters* 1 (1972) 132–133.
- [9] S.J. Grotzinger, C. Witzgall, Projections onto order simplexes, *Applied Mathematics and Optimization* 12 (1984) 247–270.
- [10] J.B. Kruskal, Nonmetric multidimensional scaling: a numerical method, *Psychometrika* 29 (2) (1964) 115–129.
- [11] W.J. Krzanowski, *Principles of Multivariate Analysis: A User’s Perspective*, Clarendon Press, Oxford, 1988.
- [12] G. Marchionini, B. Shneiderman, Finding facts vs. browsing knowledge in hypertext systems, *IEEE Computer* (1988) 70–79.

- [13] B.H. McCormick, T.A. DeFanti, M.D. Brown, Visualization in scientific computing — a synopsis, *Computer Graphics* 21 (6) (1987) 1–10.
- [14] J.F. Nunamaker Jr., M. Chen, T.D.M. Purdin, Systems development in information systems research, *Journal of Management Information Systems* 7 (3) (1990) 89–106.
- [15] J.F. Nunamaker Jr., A.R. Dennis, J.S. Valacich, D.R. Vogel, J.F. George, Electronic meeting systems to support group work, *Communications of the ACM* 34 (1991) 40–61.
- [16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, 2nd edn., Cambridge Univ. Press, Cambridge, 1992.
- [17] L. Tilling, Early experimental graphs, *British Journal for the History of Science* 8 (1) (1975) 193–213.
- [18] K.M. Tolle, Improving concept extraction from text using noun phrasing tools, University of Arizona, 1996.
- [19] M.W. Trosset, Optimization problems associated with multi-dimensional scaling, TR93-13, Department of Computational and Applied Mathematics, Rice University, 1993.
- [20] E.R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, 1983.
- [21] E.R. Tufte, *Envisioning Information*, Graphics Press, Cheshire, CT, 1990.
- [22] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

Michael J. McQuaid is a fourth year doctoral student in MIS at the University of Arizona, where he is a member of both the Artificial Intelligence Group and the Center for the Management of Information. His research interests include visualization and collaboration. He holds an MS in Operations Management and an MBA in Quantitative Analysis from the University of Wisconsin-Madison, and a BFA from the Experimental Theater Wing at New York University.

Thian-Huat Ong is a third year doctoral student in MIS at the University of Arizona, where he is a member of the Artificial Intelligence Group. His research interests include neural networks, visualization, and multilingual information retrieval. He has received a BS in Industrial Engineering and an MS in Mathematics from the University of Nebraska-Lincoln.

Hsinchun Chen is McClelland Professor of MIS and Andersen Professor of MIS at the University of Arizona, where he directs the UA/MIS Artificial Intelligence Group. Professor Chen is a Visiting Senior Research Scientist at NCSA. His articles have appeared in *Communications of the ACM*, *IEEE Computer*, *Journal of the American Society for Information Science*, *IEEE Expert* and many other publications. He has won numerous awards including Research Initiation (NSF), Best Paper (HICSS 1992), and an AT&T Foundation Award in Science and Engineering in 1994 and 1995. Professor Chen is a PI of the Illinois Digital Library Initiative project, and has received grant awards from the NSF, DARPA, NASA, NIH, NCSA, and NIJ. He is guest editor for special issues of *IEEE Computer* and the *Journal of the American Society for Information Science*.

Jay F. Nunamaker is Regents Professor of MIS, Computer Science, and Communications and Soldwedel Professor of MIS at the University of Arizona, where he founded the MIS Department, and founded and directs the Center for the Management of Information. His articles have appeared in *Management Science*, *Communications of the ACM*, *MIS Quarterly*, *Journal of the Academy of Management*, and many other publications. He has won many awards including Educator of the Year (DPMA), Professor of the Year (Andersen Consulting), PC Magazine Editor's Choice (for GroupSystems Software), and many best paper awards. He has edited many journals and currently serves on the editorial boards of six journals.