A Graphical, Self-Organizing Approach to Classifying Electronic Meeting Output

Richard E. Orwig and Hsinchun Chen
MIS Department, Karl Eller Graduate School of Management, University of Arizona, McClelland Hall, Tucson, AZ 85721. E-mail: rorwig@bpa.arizona.edu; hchen@bpa.arizona.edu

Jay F. Nunamaker, Jr.
MIS Department, College of Business and Public Administration, University of Arizona, Tucson, AZ 85721. E-mail: nunamaker@bpa.arizona.edu

This article describes research in the application of a Kohonen Self-Organizing Map (SOM) to the problem of classification of electronic brainstorming output and an evaluation of the results. Electronic brainstorming is one of the most productive tools in the Electronic Meeting System called GroupSystems. A major step in group problem solving involves the classification of electronic brainstorming output into a manageable list of concepts, topics, or issues that can be further evaluated by the group. This step is problematic due to information overload and the cognitive demand of processing a large quantity of textual data. This research builds upon previous work in automating the meeting classification process using a Hopfield neural network. Evaluation of the Kohonen output comparing it with Hopfield and human expert output using the same set of data found that the Kohonen SOM performed as well as a human expert in representing term association in the meeting output and outperformed the Hopfield neural network algorithm. In addition, recall of consensus meeting concepts and topics using the Kohonen algorithm was equivalent to that of the human expert. However, precision of the Kohonen results was poor. The graphical representation of textual data produced by the Kohonen SOM suggests many opportunities for improving information organization of textual information. Increasing uses of electronic mail, computer-based bulletin board systems, and world-wide web services present unique challenges and opportunities for a system-aided classification approach. This research has shown that the Kohonen SOM may be used to automatically create "a picture that can represent a thousand (or more) words."

1. Introduction

This research used an artificial intelligence approach to understanding the problem of classification of concepts (topics) in an electronic meeting setting. Electronic meeting systems (EMSs) provide support for large groups interactively working on a single problem or collection of problems (Nunamaker, Dennis, Valacich, & Vogel, 1991; Vogel, Nunamaker, Martz, Grohowski, & McGoff, 1989). Large groups of people are thereby enabled to use a network of computers to discuss complex organizational problems electronically. These electronic discussions create large quantities of text in a very short period of time. A major stage in the group problem solving process involves classifying these large quantities of text into a manageable list or set of concepts/topics. Experience with this classification process has shown that meeting convergence is problematic for participants and meeting facilitators.

The prevailing EMS provides only clerical classification support for browsing the text and creating a list of topics for group members. However, it does not provide system support for managing or organizing the large volume of text that may be created as output from an electronic brainstorming session. The synthesis of electronic brainstorming comments is a classification problem. It is something that humans currently do well, but not willingly, a situation that suggests using an artificial intelligence approach to understanding how humans classify concepts and developing a system to test whether better classification support for groups can be provided.

This research is a continuation of research previously reported in Chen, Hsu, Orwig, Hoopes, & Nunamaker (1994) in which the underlying classification algorithm used was a Hopfield neural network. This research used a Kohonen Self-Organizing Map (SOM) to classify electronic meeting output, and reports the experiment that was performed to evaluate the classification output of a human, the Hopfield algorithm, and the Kohonen SOM.

Section 2 provides a description of an electronic meet-

Received July 18, 1995; revised January 29, 1996; accepted January 29, 1996.

ing system and describes in more detail the classification problem that exists in the group problem-solving process. Section 3 surveys the literature concerning classification. Section 4 describes this research—the application of the Kohonen SOM to the meeting output classification problem. Section 5 describes an experiment that evaluated the classification performed by the Kohonen SOM by comparing it with those of a human expert and a Hopfield neural network. Finally, Section 6 discusses conclusions and directions for future research.

2. The Problem: Classification in an Electronic Meeting System

EMSs are a subset of a general category of systems, called ''groupware.'' Ellis et al. define groupware as ''computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment'' (Ellis, Gibbs, & Rein, 1991). In addition to Electronic Meeting Systems, other types of groupware include electronic mail, group calendar systems, group project management, and multi-user document editing systems.

What distinguishes an EMS from other types of groupware is that the common goal (problem resolution) typically involves complex organizational problems, which benefit from maximum participation from appropriate organizational members. Examples of such problems include performance of organizational analyses through vision or mission statement construction or organizational model building, strategic plan development, organizational decision making, and project status reporting. Research and experience with EMSs have demonstrated that electronic support of group meetings can improve meeting productivity through greater participation by organizational membership and shorter time required for information gathering and for group problem resolution (Nunamaker, Dennis, Valacich, Vogel, & George, 1991; Post, 1992). In particular, groups using EMS have been found to generate more unique alternatives for creative tasks and higher quality of decisions related to intellective tasks than non-EMS-supported groups (George, Easton, Nunamaker, & Northcraft, 1990).

EMS experience has demonstrated a consistent pattern of use that may be described as a ''group problem solving'' process. This process is goal-directed insofar as specific types of information are identified for support of problem resolution. The EMS then is used to gather the information that is appropriate for the goal. We label the stages of the information gathering process as ''divergence,'' ''convergence,'' and ''consensus checking.'' Particular EMS tools are selected for each of these stages depending upon the characteristics of the particular goal or task before the group. Electronic brainstorming (EBS) is a particularly good divergent technique/tool for collecting information related to complex tasks in which maximum, unstructured, and anonymous participation is desired. Idea Organization (IO) techniques and tools may then be used to organize the information into a list of topics or concepts that address the complex problem (the convergent stage). Techniques and tools for Voting may then be used to test whether all meeting participants agree on the definitions of those topics or concepts or on the importance of particular topics with respect to the particular goal.

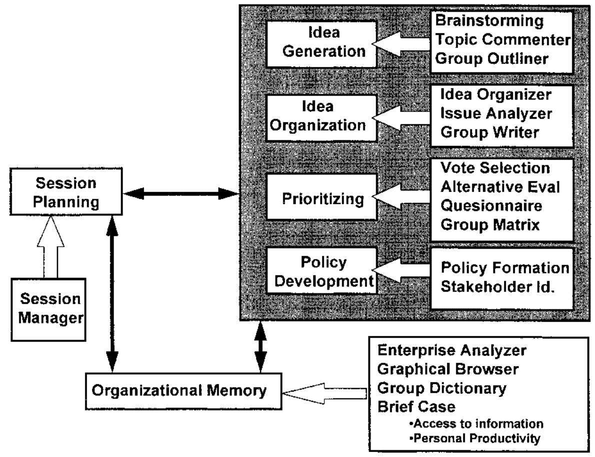

FIG. 1. Electronic meeting system research model.

2.1. Description of an EMS: GroupSystems

An EMS consists of hardware, specialized software, and facilitation methods and techniques for solving group problems. An EMS typically includes 8–30 networked personal computers or workstations, special software that allows people to enter comments and manipulate shared data, and support for electronic projection of the shared data as well as other audio/visual equipment that supports meetings. Software for general-purpose group problem solving consists of a set of programs that can be mixed and matched to suit the characteristics of the problem as well as the characteristics of the group of participants. GroupSystems is an specific example of such specialized software tools that was developed at the University of Arizona. A pioneering example of meeting software technology, GroupSystems has been installed at more than 80 universities and a total of more than 400 organizational sites in business, government, and university settings.

Figure 1 presents a diagram of the types of EMS activities that occur in an electronic meeting as well as a mapping of particular GroupSystems tools that support the group problem-solving activities. Four EMS activities diagrammed in the center of the figure are the main types of tools needed for group problem solving: Idea Generation, Idea Organization, Prioritizing, and Policy Development. Session planning software controls the overall meeting and is, itself, a set of software tools for creating an agenda, starting and stopping participant tools, and implementing system utilities such as saving data and printing reports. Facilitation is the activity of matching appropriate EMS hardware and software tools to the organizational problem to be solved. This requires collecting information from organizational members to gain an understanding of the characteristics of the problem, characteristics of the group, and the organizational context of the problem (Nunamaker et al., 1991). The facilitator is typically NOT an expert in the group problem-solving domain. Rather, the facilitator is responsible for understanding the group's problem well enough to select the appropriate EMS tools to be used in the appropriate circumstances and guide meeting participants' use of these tools during the meeting. Like an arbitrator, the facilitator is supposed to play an objective, non-biased role in helping the meeting participants solve the group problem using the tools of the EMS.

The benefits of group problem solving using an EMS come from three major EMS attributes: Anonymous participant input, parallel data entry, and electronic recording of all user input. Allowing anonymous input of information by participants often results in a freer exchange of ideas. Parallel data entry allows multiple participants to enter ideas or comments at the same time. Participants can enter information while ideas and thoughts are fresh in their minds. Electronic recording of all of the participant input supports the creation of reusable ''organizational memory.'' However, experience with GroupSystems has shown that the unique features of anonymity, parallel data entry, and electronic recording are less effective in supporting the convergent stage of group problem solving (Chen et al., 1994).

2.2. Convergence Problems in EMS

The objective of the Idea Organization activity in an EMS is to classify the participants' comments into a list of important topics that is appropriate for the desired outcome, goal, or subgoal. Difficulties in performing this idea organizing activity are due to various aspects of the output of the divergent process that precedes it: Large numbers of comments and different interpretations of the vocabulary within the comments. Other difficulties are related to the increased cognitive demand of the Idea Organization process itself as well as software interface problems with the current Idea Organization tools.

Information overload

Successful use of the electronic brainstorming tool has been found to result in an unanticipated information overload problem (Chen et al., 1994). In a typical meeting of 10–20 participants, several hundred EBS comments can be generated in less than an hour. The characteristics of parallel and anonymous data entry often result in hundreds of lines of text being generated in a short period of time, which makes it extremely difficult for participants to browse comments and consolidate ideas. Meeting participants are often impressed with the number of ideas they have generated, but become overwhelmed by the task of organizing the ideas into categories or topics.

Vocabulary problem

Much of the complexity in interpersonal communication is related to semantics. Connotations surrounding terms vary according to the perspectives of different individuals and even within the same individual in different problem settings. One aspect of the vocabulary problem is that meeting participants may have different definitions of the same terms. Another aspect is the converse: Multiple terms may have a single meaning for meeting participants. In prior research, Furnas, Landauer, Gomez, and Dumais (1983, 1987) found that in spontaneous word choice for objects in five domains, two people favored the same term with less than 20% probability. This vocabulary difference is even more problematic in the context of collaborative systems which involve a large number of participants (Chen, 1994).

Increased cognitive demand

EBS comments need to be consolidated and organized by meeting participants within a short period of time. Usually each participant has to browse and understand the ideas generated in the EBS comments, judge the merits of these ideas, merge similar ideas, eliminate redundant or irrelevant ideas, consult other members' opinions, and so on in less than an hour. Because the EBS process encourages creative, diverse, and uncensored ideas, many EBS comments are raw or unpolished and it may require special effort from meeting participants to synthesize them. Because the group topic organization process is sometimes frustrating and sub-optimal, the satisfaction level and productivity of the meeting participants may go down significantly, and/or some unique EBS ideas may be lost during the process.

EMS interface problems

Most EMSs are based on windows (Microsoft or X-Windows). However, a 30+ line computer screen is still a severe constraint for someone browsing hundreds of lines of text while trying to synthesize ideas into a list of issues. Several windows may be needed: A window for browsing the comments, a window for formulating a participant's local list of issues, a window for viewing the public list, and a window for attaching comments to the list items. However, management of these windows could be difficult for even computer experts, let alone the average person participating in a meeting.

The need for automated and ''intelligent'' support of the convergent process arises from dissatisfaction experienced by meeting participants, the inconsistency of the group topics generated, and the time constraint for meeting convergence. In light of these problems, this research was undertaken to build upon prior research in artificial intelligence and information science, and our extensive experience in EMS, in order to develop a more pro-active and ''intelligent'' EMS idea classification system.

3. Classification Literature Review

The classification of EMS ideas involves an understanding of the mapping of that idea text into a set of categories that model or represent that text. Thus, to develop an automated approach to classifying electronic brainstorming comments, we needed to review the textual analysis and cluster analysis literatures to review methods for mapping the text into an appropriate representation. For further details and a more in depth literature review, see Orwig (1995).

Textual analysis techniques aim to identify descriptors and develop an unambiguous internal representation for a document. In clustering, documents are analyzed and structured based on their degrees of similarity or relevance.

3.1. Textual Analysis

Automatic indexing in information science and natural language processing in artificial intelligence are two sets of techniques frequently used for textual analysis.

Automatic indexing

Automatic indexing is defined as the machine performance of the process of constructing document surrogates by assigning content identifiers to text items (Salton, 1989). Salton describes three theories for indexing terms to aid in discriminating among terms to better address term specificity: Term-frequency, term-discrimination, and probabilistic term weighting. Three blueprints for automatic indexing are summarized in Salton (1989). Two of these blueprints depend upon the existence of a thesaurus. Since electronic brainstorming text is generated in ''real time'' during a meeting and since most organizations do not have a documented ''organizational thesaurus,'' the third approach of word identification, stop wording, stemming, and term-phrase formation seems more appropriate for the EMS application.

Natural language processing

Natural language processing (NLP) aims to develop unambiguous internal representations for human languages. Identification of concepts contained within electronic brainstorming comments would seem to demonstrate some level of natural language understanding. NLP typically involves several levels of analysis, i.e., morphological analysis, syntactic analysis, semantic analysis, discourse integration, and pragmatic analysis. Morphological analysis separates the input text into terms and/or term phrases. Syntactic analysis parses the output of morphological analysis and applies a sentence structure or other grammatical structure to the input. Semantic analysis derives meaning from the input by mapping the terms into appropriate objects in a knowledge base and converts the syntactical structure into a structure consistent with that represented in the knowledge base (''frames'' or ''scripts,'' for example). Linking these structures together to demonstrate the linkages between the structures created in semantic analysis (e.g., link pronouns of one sentence with their antecedents of previous sentences) is the objective of discourse integration. Finally, actual interpretation of the input and determining what action to take in response to the input is the objective of pragmatic analysis.

Work in syntax and grammars of English text has been occurring for decades. Several techniques in particular have attracted special attention. Noam Chomsky (1965) is best known for his work in syntax theory. Context-free grammars attempt to identify the building blocks of language with syntax providing the rules of construction, which is limited to phrases and sentences. Woods (1972) describes the augmented transition network (ATN parser)—a system similar to a finite state machine of states and arcs, where the states of the network represent positions through a text passage and arcs represent possible words, word categories, or procedures which cause a transition to the next state. An ATN parser includes both context-free grammars and transformational grammars, both considered syntactic analysis. Semantic grammars, developed by Burton (1976), include semantic rules as well as syntactic functions in grammar rules. Rather than embed the semantics within the grammar rules, Fillmore (1968) creates semantic relations to structure the syntactic rules and called this a *case grammar.* Thirty years of work in syntax and semantics have demonstrated the complexity of processing the English language (or any human language). There is still much work to do to create semantic analysis output in a generalizable and domain-independent way.

The lexical analysis associated with NLP today still has far to go before generalizable text processing is achievable. The ''noise'' found in EBS comments (typographical errors, poor sentence structure, unique vocabulary, etc.) combined with unpredictable topics of discussion make the implementation of natural language understanding techniques impractical for concept categorization. As will be discussed below, the automatic indexing technique was later adopted in our research.

3.2. Cluster Analysis

Classification of EBS comments requires grouping (or clustering) similar concepts/terms as a category or topic, a process calling for *cluster analysis* techniques. Two approaches to cluster analysis exist: *The serial, statistical* *approach* and *the parallel, neural network approach.* In this section, we provide only a brief summary of the conventional statistical approach. However, because our ongoing research is based on a neural network algorithm, we provide a more detailed review of the newer parallel, neural network approach to document classification.

The serial, statistical approach

Rasmussen (1992) defines cluster analysis as ''a statistical technique used to generate a category structure which fits a set of observations. The groups which are formed should have a high degree of association between members of the same group and a low degree between members of different groups.'' He points out that cluster analysis is a technique for multivariate analysis that has application in many fields. A number of software packages support various methods and algorithms (SAS, SPSSX, BMDP, CLUSTAN, CLUSTAR/CLUSTID).

Automatic document classification involves determining a document representation structure and methods for determining similarities between documents. The hierarchical clustering of documents can be carried out either *divisively* or *agglomeratively* (Salton, 1989). Divisive clustering breaks one complete cluster down into smaller pieces. In agglomerative clustering ''individual item similarities are used as a starting point and a gluing operation collects similar items, or groups, into larger groups'' (Salton, 1989).

Stepp (1987) describes conceptual clustering as the new frontier in artificial intelligence that arose from Michalski's introduction of the concept in 1980. Algorithms for clustering involve co-occurrence of feature values, discovering conjunctive features among the attributes rather than variations in the value taken by a single attribute, and clumping concepts based upon most commonly occurring relations in the data. Two stages occur in conceptual clustering: An aggregation phase and a characterization phase. The aggregation phase produces the groupings and the characterization phase attempts to assign some meaningful interpretation of the groupings.

Using these techniques, classes of similar objects are basically found by doing pairwise comparisons among all of the data elements. These clustering algorithms are serial in nature in that pairwise comparisons are made one at a time and the classification structure is created in a serial order.

The parallel, neural network approach

A new approach to addressing clustering and classification problems is based on the connectionist approach, or neural network computing. Algorithms based upon neural networks are parallel in that multiple connections among the nodes allow for independent, parallel comparisons.

Neural networks are patterned after the biological ganglia and synapses of the nervous system. The concept is not new—McCulloch and Pitts suggested the description of a neuron as a *logical threshold limit* in 1943. The essential element of the neural network is the neuron. A typical neuron j receives a set of input signals from other connected neurons, x_i, each of which is multiplied by a synaptic weight factor of w_{ij}. All activation weights are then summed to produce the activation level for neuron j. Many neural network topologies and error correction (learning) algorithms have been developed since the early 1980s (Lippmann, 1987).

The adjustments of the weights of the nodes of the neural network enable the total network to ''learn'' in that a neural network's performance can be adjusted to fit a known set of data characteristics. In supervised learning, a set of training examples is presented, one by one, to the network. The network then calculates outputs based on its current input. The resulting output is then compared with a desired output for that particular input example. The network weights are then adjusted to reduce the error. In unsupervised learning, network models are first presented with an input vector from the set of possible network inputs. The network learning rule adjusts the weights so that input examples are grouped into classes based on their statistical properties (Dalton & Deshmane, 1991; Rumelhart, Hinton, & Williams, 1986).

Neural network computing recently has been applied to information science applications with interesting results. Among the early researchers, Belew (1989) developed a three-layer neural network of authors, index terms, and documents in a system called AIR. Relevance feedback from its users changed the representation of authors, index terms, and documents over time. Kwok (1989) used a modified Hebbian learning rule in a similar three-layer network of queries, index terms, and documents. Chen et al. adopted a variation of the Hopfield network for concept space generation and concept space traversal in a series of experiments (Chen & Lynch, 1992; Chen & Ng, 1995; Chen, Schatz, Yim, & Fye, 1995). Doszkocs, Reggia & Lin (1990) provide an excellent overview of connectionist models in information retrieval including artificial neural networks, spreading activation models, associative networks, and parallel distributed processing. Chen (1995) provides an up-to-date review of various machine learning techniques, neural networks, and genetic algorithms for information retrieval applications.

Several information science researchers have developed neural network clustering algorithms for information science applications. MacLeod and Robertson (1991) present a two-layer neural network and an algorithm that is used for document clustering. This algorithm adopts an unsupervised approach to clustering by way of partitioning. Inputs are binary-valued vectors representing documents. The algorithm uses two similarity measures in order to allow proof of algorithmic behavior, cluster stability, and characterization of final clusters. The algorithm is multi-pass in that documents are repeatedly read into the network until two consecutive passes produce identical classifications for each document (convergence). Only those clusters that are successful in classifying one or more documents during the last pass are active. Inactive clusters do not take part in any subsequent retrieval operations.

Chen et al. (1994) describe an algorithm for concept classification of electronic brainstorming comments that combines automatic indexing of the electronic brainstorming comments, term co-occurrence concept space generation, and a Hopfield neural network classifier (Hopfield, 1982). Results of experiments comparing the output of their algorithm with human experts and novices found that the Hopfield classifier performed as well as the novices but two human experts out-performed the novices and Hopfield classifier significantly. Reasons given for the performance differences centered on the ability of the experts to associate terms more appropriately.

In addition to these unique neural network based clustering algorithms for information science applications, prior research in neural networks has strongly suggested the Kohonen SOM as an ideal candidate for clustering of textual documents.

Kohonen (1989) bases his neural network on the associative neural properties of the brain. This network contains two layers of nodes—an input layer and a mapping (output) layer in the shape of a two-dimensional grid (Caudill, 1993; Hiotis, 1993). The input layer acts as a distribution layer. The number of nodes in the input layer is equal to the number of features or attributes associated with the input. Each node of the mapping layer also has the same number of features as there are input nodes. Thus, the input layer and each node of the mapping layer can be represented as a vector which contains the number of features of the input. The network is fully connected in that every mapping node is connected to every input node. The mapping nodes are initialized with random numbers. Each actual input gets compared with each node on the mapping grid. The ''winning'' mapping node is defined as that with the smallest Euclidean distance between the mapping node vector and the input vector. The input thus maps to a given mapping node. The value of the mapping node vector is then adjusted to reduce the Euclidean distance. In addition, all of the neighboring nodes of the winning node are adjusted proportionally. In this way, the multi-dimensional (in terms of features) input nodes get mapped to a two-dimension output grid. After all of the input is processed (usually after hundreds or thousands of repeated presentations), the result should be a spatial organization of the input data organized into clusters of similar (neighboring) regions. Many engineering and scientific applications which involve numeric data (e.g., image recognition, signal processing) have successfully adopted the SOM approach to parallel clustering (Kohonen, 1995).

Several recent studies adopted the SOM approach to textual analysis and classification. Ritter and Kohonen (1989) apply the Kohonen SOM to textual analysis in an attempt to detect the logical similarity between words from the statistics of their contexts. Their first approach represents context of terms as a set of attribute values that occur in conjunction with the words. The second approach defines context by the sequences in which words occur without consideration of any associated attributes. They analyze simple verbal statements consisting of nouns, verbs, and adverbs and suggest that ''such phrases or clauses involve some of the abstractions that appear in thinking, namely, the most common categories, into which the words are then automatically grouped'' (Ritter & Kohonen, 1989). They argue that a similar process may be at work in the brain. Typical self-organizing maps reflect metric distance relations between patterned representation vectors. Ritter and Kohonen associate this with lower levels of perception. To map symbols topographically one must present the symbol ''in due context, i.e., in conjunction with all or part of the attribute values of the item it encodes, or with other, correlating symbols'' (Ritter & Kohonen, 1989). The example they give uses 16 animal types (dove, hen, fox, tiger, cow, etc.) and 13 attributes (size, number of legs, hunts, swims, etc.). Input of these attributes for each of the animal types into the SOM produces an organization of the animals on a map where animals are spatially ordered into essential ''family relationships.''

Miikkulainen (1993) developed DISCERN (DIstributed SCript processing and Episodic memoRy Network) as his dissertation project. DISCERN is a prototype of a subsymbolic natural language processing system that shows how script instantiation and inferences can be learned from examples by a parallel distributed processing architecture that is based upon a Kohonen Self-Organizing Map. However, DISCERN does not implement the entire computational theory or even the full complexity of symbolic script-processing systems. Miikkulainen provides an interesting discussion concerning expansion of classification into a hierarchy of layers of abstraction whereby classification occurs between each level of abstraction.

In support of using Kohonen for textual document classification, Lin, Soergei and Marchionini (1991) used the Kohonen SOM for classifying documents for information retrieval. Documents are represented as vectors of binary values. Each coordinate of the vector represents a specific term or term phrase with the value set to ''1'' if the term or term phrase is found within the document and ''0'' otherwise. After several passes through the input file (a collection of similar documents), the Kohonen layer is trained. The resulting map provides an intuitively-appealing organization of the input data. The documents are classified according to their content and conceptual regions are formed and named on a two-dimensional grid.

4. Kohonen Classification for EBS Comments

Lin's work first demonstrated the feasibility of using the Kohonen algorithm for classification of textual documents. In addition, since associated terms were juxtaposed on the map, the algorithm provided potential for suggesting association of terms in a manner that might approach the experts in the study reported by Chen et al. (1994). In the context of electronic meetings, the graphical, associative output is appealing for use with typical electronic meeting participants (non-computer people). The algorithmic robustness and graphical output produced by the Kohonen SOM prompted us to experiment with it in the context of EBS classification. In our research, we adopted the human facilitators' output and the Hopfield neural network classification results from our prior study (Chen et al., 1994) as benchmarks for comparison.

4.1. A Kohonen Algorithm for Text Classification

The Kohonen algorithm for classifying EBS comments uses the same automatic indexing process as the Hopfield network classifier reported in Chen et al. (1994). Output of the automatic indexing process is converted into a form compatible with the Kohonen input format. Each comment of the EBS session is represented by a term (index) vector of 1 or 0. The number of 1s in each comment vector is equal to the number of terms in the comment and each vector position corresponds with one unique term. The SOM algorithm we adopted for EBS comment classification is summarized below:

1. Initialize the output (mapping) layer

Initialize the weights connecting the inputs to the mapping layer to small random values and initialize the neighborhood size.

2. Process input

Present an input vector and compute the difference between this input vector and each mapping node vector. The distance measure is based on a Euclidean function of:

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

where $x_i(t)$ is the input to node i at time t and $w_{ij}(t)$ is the weight from input node i to output node j at time t. N is the total number of input nodes. The node on the map with the minimum d_j is the ''winning'' node. The node is selected as the center of a neighborhood to be adjusted.

3. Adjust weights in a neighborhood

With the ''winning'' node identified (say $j*$), adjust the vector weights for the minimum node and all of the nodes in its defined neighborhood. The new weights are computed based on a simple error correction function of:

$$w_{ij*}(t+1) = w_{ij*}(t) + \eta(t)(x_i(t) - w_{ij}(t))$$

for $j*$ and all nodes in the neighborhood and where $0 \leq i \leq N-1$. The gain term, $\eta(t)$, which is between 0 and 1, decreases in time.

See Kohonen (1995), Lippmann (1987), Hiotis (1993), and Caudill (1993) for detailed algorithmic discussions and variations.

4.2. Kohonen Classification for EBS: An Example

A sample EBS output file, shown in Figure 2, is used as an example. Each numbered paragraph represents an EMS comment contributed by a meeting participant. Automatic indexing was performed on the EBS comment file in a manner identical with that performed with the Hopfield classifier (Chen et al., 1994). Output of automatic indexing contained index terms for the comments is shown in Figure 3 and a list of terms in decreasing order of frequency for the entire EBS file is represented in Figure 4. These two files were used as the Kohonen input file. For this sample session, we created an input file of 202 vectors, where 202 is the number of EBS comments contained in the EBS file. Each input vector consisted of 190 nodes, representing the top (in frequency) 190 terms in the EBS file.

We used a 20 by 10 grid map for our example based upon what would fit on an output screen. We used a hexagonal neighborhood area which considers six surrounding nodes to be a node's immediate neighborhood. Finally, we used the bubble adjustment method, which is simply an adjustment of the weights of neighboring nodes based upon the decreasing gain term. In the initial training phase, we ran the input with 1,000 iterations, a gain term adjustment of 0.05, and a neighborhood size of 10. In the fine-tuning phase we ran the input with 10,000 iterations, a smaller gain term adjustment of 0.02, and a smaller neighborhood size of 3. The training phase ended after 1,000 input vectors had been processed (recycling through the input file if necessary).

After training, the Kohonen SOM visualization consisted of running the same input file against the trained map and reporting the map grid location that is closest in Euclidean distance to each input. Each document (vector) and each term (represented as a unit vector) were thus mapped to a node and also a region (of the same nodes) on the map. Labeling each node on the map with the appropriate term resulted in the map shown in Figure 5.

The numbers in the map (Fig. 5) correspond with the documents that are classified into a particular concept region. Each document can only be assigned to one single region. Thus, it is interesting to note that ''Standard'' (lower right portion of the map) is an issue that is relevant to important information technology problems but each document that discussed standards also discussed other concepts and was mapped into one of the other areas.

Electronic Brainstorming Report
Session: technology Date: 11/18/1993 2:08pm

What are the most important information technology problems with respect to
Collaborative Systems to be solved over the next five years?

1.1 development of "standard" user interfaces for reomte access to info
and communication

1.2 Doesn't XYZ Software give us such a standard?

1.3 XYZ is one implementation, but it certainly is not yet a standard.
!!!! Maybe we need to have a standard that is NOT owned by a company which
has a history of initiating look and feel lawsuits when someone else copies their "standard" !!!!!!

1.4 XYZ is not a good standard even if they are the industry leader
at this time. XYZ is a terrible architecture.

2.1 Voice recognition is one of the promising technologies.

...

30.3 Understanding how we actually use the tools in a distributed mode.
We need to understand process.

30.4 Total integration between all applications.

30.5 how to measure the validity and reliability of groupware tools

30.6 Yes, total integration. Right now artifacts of the tools limit us
significantly in ways we need not be limited.

30.7 No matter what we do we will be limited in ways we need not be
limited. Thank God for unanswered prayers.

FIG. 2. Sample EBS session output file.

One would need to search for the keyword ''standard'' to obtain all of the documents that discussed this concept (or find them in the conceptual areas adjacent to ''Standard.'') Thus, the Kohonen algorithm seems to produce a conceptual map of the textual data with the one best place for storing each document on the conceptual map. From a document management perspective, the algorithm would seem to provide an organizing structure but one would need to use the keywords of a storage area to search for documents in order to get higher recall of documents related to the issues of that given storage area.

Analysis of several session outputs revealed that there seemed to be a general relationship between terms that were adjacent on the map. For example, the sample Kohonen EBS output revealed that ''Collaborative'' and ''Systems'' (top-right corner) were terms frequently used together as well as ''Distributed'' and ''Meetings'' (bottom-right corner) all on neighboring regions of the map (see Fig. 5: A Sample Kohonen EBS Classification Output). And interestingly, ''Video,'' ''Networks,'' and ''Hardware'' issues seemed to be clustered on the bottom-left corner of the SOM map. ''People'' and ''Training'' issues are adjacent, as well as ''(Lotus) Notes'' and ''Memory/Tool.''

The size of the region also seemed to bear some relationship to the importance of the topic. The largest areas were represented by terms that addressed the brainstorming question directly, ''What are the most important information *technology* problems with respect to *collaborative systems* to be solved over the next five years?'' ''Tools,'' ''Meeting,'' ''Technology,'' ''Collaborative,'' ''Systems,'' and ''Facilitator'' were among the largest regions on the map. Due to the competition among regions, some regions were formed without any comment having been assigned. On the other hand, some comments did not contain any relevant or specific ideas and thus were mapped to a ''miscellaneous'' area, e.g., 31 comments were mapped to a node at the bottom center of Figure

1.1 1 1 2 1 DEVELOPMENT
1.1 1 1 4 1 STANDARD
1.1 1 1 5 1 USER
1.1 1 1 5 2 USER INTERFACES
1.1 1 1 6 1 INTERFACES
1.1 1 1 8 1 REMOTE
1.1 1 1 11 1 INFO
1.1 1 1 13 1 COMMUNICATION
1.2 1 1 3 1 XYZ
...
30.7 1 1 3 1 MATTER
30.7 1 1 10 1 LIMITED
30.7 1 1 17 1 LIMITED
30.7 1 2 19 1 GOD
30.7 1 2 21 1 UNANSWERED
30.7 1 2 21 2 UNANSWERED PRAYERS
30.7 1 2 22 1 PRAYERS

FIG. 3. Sample EBS automatic indexing output.

5. These comments were mostly noise (i.e., typos) or completely unique ideas.

In summary, the Kohonen map seemed to provide a graphical representation of the collective discussion of the group of 21 people in approximately 6 minutes. Not only are the issues related to important information technology problems surfaced but the relatedness of the issues with respect to subject area and importance (size of the area) seem to be represented. This objective and quick analysis of the Kohonen algorithm would seem preferable to the potentially-biased and slow (+40 minutes) method involving a human facilitator. Further evaluation of the output was warranted.

5. Evaluation of Kohonen Classification

Several research questions arise in looking at the Kohonen classification results. How does the Kohonen network compare with a human facilitator in classifying EBS comments? Does the Kohonen neural network produce results better than those generated by the Hopfield classifier developed earlier? And does the graphical, associative nature of Kohonen classification help in producing more relevant classification?

5.1. Experiment Design

To address the above research questions, we recently designed and conducted a three-stage experiment. The first stage involved comparing the topic lists generated by a human expert facilitator, the Hopfield Algorithm, and the Kohonen SOM. This stage addressed SOM list quality in comparison with the output of the human expert and the Hopfield algorithm. To address the issue of term relevance, we pooled all of the terms in the three lists, presented them to several facilitator subjects in alphabetical order, and asked the subjects to rate the relevance of the terms to the EBS question. Thus, the second stage measured the degree of relevance of the terms contained within the topic lists. The third stage measured the relative degrees of association that existed between each of the terms on the three lists. This stage addressed the significance of the adjacent areas of the Kohonen map and whether the apparent high degree of association actually exists between the areas. If so, the Kohonen SOM might prove more useful in supporting group convergence by suggesting graphical conceptual clusters as regions in the map.

First stage: List comparison

EBS output from an actual electronic brainstorming session was used in our experiment (the same session shown in Fig. 2). The session was chosen because the subject domain was in ''collaborative systems,'' an area with which all our facilitator subjects were familiar. During the actual group electronic brainstorming process, an expert facilitator browsed the participants' comments and created a set of keywords to categorize the comments pertaining to the question: ''What are the most important information technology problems with respect to Collaborative Systems to be solved over the next five years?'' The 21 meeting participants were all researchers and practitioners in this area. The EBS session output file was then run through the Hopfield classifier (Chen et al., 1994) and the Kohonen algorithm described above. The facilitator took about 40 minutes to generate his topic list; while the Hopfield classifier and the Kohonen algorithm took 4

0 TECHNOLOGY
1 MEETINGS
2 COLLABORATIVE SYSTEMS
3 SUPPORT
4 TOOLS
5 COLLABORATIVE
6 SYSTEMS
7 MEETING
8 DISTRIBUTED
9 PEOPLE
10 ENVIRONMENTS
11 TECHNOLOGIES
12 ENVIRONMENT
13 INFORMATION
14 VIDEO
15 NETWORKS
16 ISSUES
17 NETWORK
18 LINEAR THREAD MEETING
19 HARDWARE
20 DISTRIBUTED MEETINGS
21 PROCESSES
22 CULTURAL
...
181 CHANGE
182 WORLD
183 COMMON SPOKEN LANGUAGE
184 DIFFERENCES
185 REMOTE COLLABORATIVE MEETINGS
186 WINDOWS
187 REMOTE SCREEN SHARING
188 MODE
189 RELIABILITY

FIG. 4. Ranked terms in a sample EBS file.

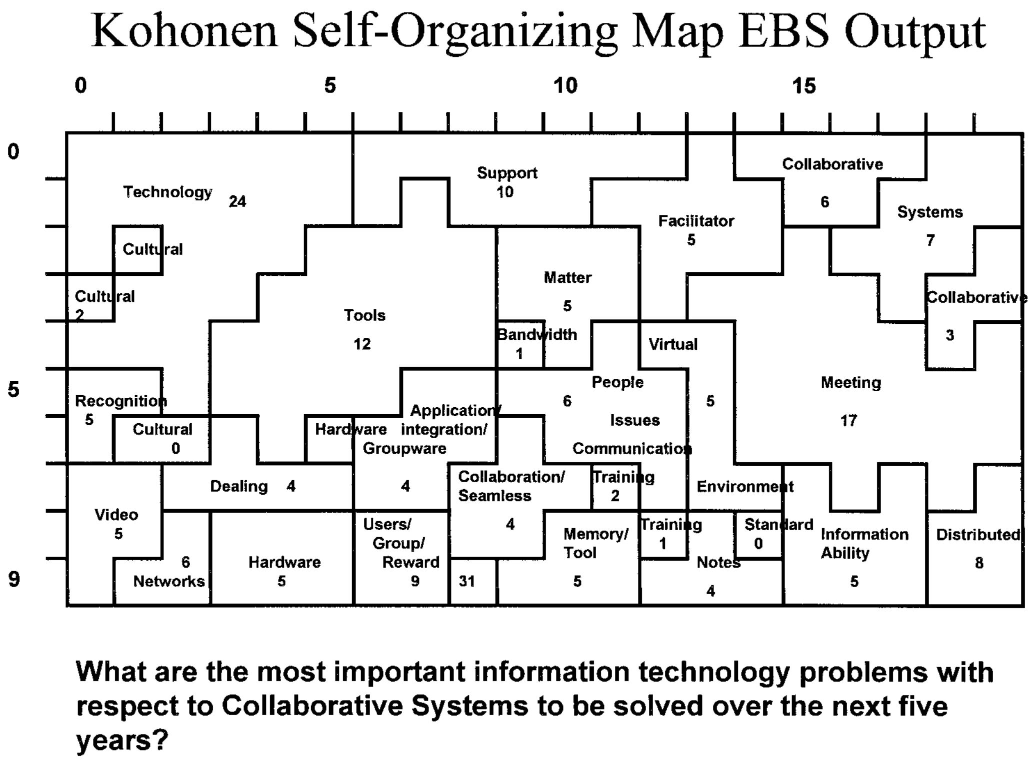

FIG. 5. A sample Kohonen EBS classification output.

minutes and 6 minutes, respectively. The topic lists created by the three methods are shown in Table 1.

Eight subjects of the experiment were given the complete text output of the electronic brainstorming session and the three lists—each on a separate sheet of paper. Subjects read through the comments (optionally creating their own lists). They then ranked the three lists from most appropriate to least. Criteria for evaluating the lists were the options of the facilitator-subjects but they were given the task as ''Rank the lists as most appropriate for addressing the EBS question.''

Subjects were then asked to ''correct'' each list by deleting inappropriate terms and adding terms that were missing. A past experiment had demonstrated that facilitators had a criteria of a correct list as a list of issues that covered all of the EBS discussion appropriate to the EBS question and ''balanced'' with respect to level of abstraction (Chen et al., 1994). Subjects could use the lists that they created themselves in order to make each list of collaborative technology problem issues more appropriate to the EBS question. The size of each corrected list as identified by the subjects are reflected in the Target rows of Table 2.

Subjects completed this stage in times that ranged from just under 1 hour to over 2 hours.

Second stage: Term relevance evaluation

The terms on the three lists (Facilitator, Hopfield, and Kohonen) were pooled into one alphabetical list of 83 terms. Combinations of terms forming individual topics (e.g., ''video/projection'') were split and treated as separate terms. Term phrases were retained (e.g., ''distance issues'' was treated as one term as was ''distance''). Subjects were not allowed to re-visit the three individual lists to determine the source of the terms.

Subjects were asked to rate the relevance of the terms with respect to whether they should be included within a topic list pertaining to the electronic brainstorming session. The ranking scale was: 0—irrelevant (too vague, too specific, or otherwise inappropriate); 1—somewhat relevant; 2—most relevant. The instructions plus the 83 terms were listed on a single sheet of paper. Subjects called out the number of the term plus a score of ''0,'' ''1,'' or ''2'' to the experimenter. Subjects completed this task in 10 to 15 minutes.

Third stage: Term association evaluation

The list of 83 terms was converted to a shorter list by combining terms which were singular and plural forms of the same term (e.g., ''application'' and ''applications'') and removing terms which did not have semantic content or were obviously too general (e.g., ''matter'' and ''issues''). Term phrases were also combined with individual terms (e.g., ''culture,'' ''cultural,'' and ''cultural differences'' into ''culture/cultural differences''). A list of 59 topics resulted. Subjects were given a one-page printout of the numbered 59 terms and asked to rate

TABLE 1. Three sample EBS topic lists.

Facilitator	Hopfield	Kohonen
1. video/projection	1. technology/dealing/cultural differences	1. technology
2. network/bandwidth	2. collaborative systems	2. tools
3. multimedia/hypertext/multi-media	3. meetings/distributed/human	3. meeting
4. group memory/project memory/repository	4. linear thread meeting	4. support
5. voice	5. ai/data/amounts	5. collaborative
6. culture/style	6. environments/virtual	6. facilitator
7. language	7. voice recognition	7. matter
8. standards	8. tools/culture	8. systems
9. distributed/distance issues/distance/ different place	9. people	9. people/issues/communication
10. facilitation	10. technologies	10. application/integration/groupware
11. research methodologies	11. information/ability	11. recognition
12. cost/money	12. video/desktop	12. cultural
13. team	13. networks	13. video
14. reward	14. ussues	14. hardware
15. integration	15. hardware	15. networks
16. social/societal/society	16. distributed meetings	16. dealing
17. performance	17. training	17. users/group/reward
18. virtual	18. groups	18. collaboration/seamless
19. education/train/learn/teach	19. applications	19. memory/tool
20. human/people/user/individual/interpersonal	20. distributed environment	20. training
		21. virtual
		22. environment
		23. notes
		24. information/ability
		25. distributed
		26. standard
		27. bandwidth

the associations between any term-pair (0: not associated; 1: somewhat associated; 2: strongly associated). This stage required 1,711 pairwise comparisons ($59 \times 58 \div 2$) and assumed that terms were symmetrically associated. Times to complete this third stage ranged from just under 1 hour to about 2 hours.

Eight subjects completed the three stages. Three of the subjects were experienced facilitators who each had facilitated more than 25 meetings over the past 2 years. Two subjects were less experienced but had been trained in facilitation and had each facilitated five to 10 meetings. The three remaining subjects had facilitated several sessions or assisted in many electronic meetings over a 2-year period. While relatively inexperienced in guiding groups through the convergent process, they were very much aware, as observers and participants, of the objectives of the meeting convergence process.

5.2. Experimental Results

List comparison: The Kohonen output was less precise than the facilitator output. However, the recall levels of the three lists were not statistically different

In overall list ranking, each subject selected the facilitator list as the best. For stage one, the relative quality of the three lists was measured by recall and precision, commonly used in information retrieval. Table 2 contains a summary of the results of list comparison. The ''Identified'' row pertains to the number of concepts identified by the human facilitator, Hopfield algorithm, and Kohonen SOM (i.e., the number of list items). The ''Target'' row contains the number of concepts that were contained in each list after the subject added missing concepts and deleted inappropriate concepts. The ''Relevant'' row contains the number of concepts in the original lists that remained in the corrected lists. ''Recall'' is the number of relevant concepts divided by the number of target concepts identified by each subject. ''Precision'' is the number of relevant concepts divided by the number of concepts originally identified in the ''Identified'' row.

The Stage 1 row of Table 3 contains the results of statistical analysis (using MINITAB; Ryan, Joiner, & Ryan, 1985) of the comparison of the recall levels of the three lists as judged by the subjects. Analysis of variance (ANOVA) among the recall scores of the eight subjects indicated that there were no differences among the three lists ($p = 0.362$) and *t*-tests also indicated no significant differences among the three lists (at the 10% significance level). However, analysis of variance and *t*-tests among the precision scores indicated that the facilitator list out-performed the Kohonen list ($p = 0.021$).

Except for some obvious noise generated by the two algorithms, the Kohonen list and the Hopfield net produced 81 and 74% recall, comparable to the 89% recall generated by the facilitator. Considering the time required of the facilitator to generate such a list manually and the cognitive demand and dissatisfaction involved in this meeting convergence process (described earlier), our Hopfield classifier and Kohonen

TABLE 2. List comparison results.

Subjects	Identified	Facilitator 20	Hopfield 20	Kohonon 27
1	Target	15	20	11
	Relevant	14	18	9
	Recall	93%	90%	82%
	Precision	70%	90%	33%
2	Target	32	32	32
	Relevant	14	10	11
	Recall	44%	31%	34%
	Precision	70%	50%	41%
3	Target	10	10	10
	Relevant	10	6	7
	Recall	100%	60%	70%
	Precision	50%	30%	26%
4	Target	22	22	20
	Relevant	20	14	18
	Recall	91%	64%	90%
	Precision	100%	70%	67%
5	Target	19	19	22
	Relevant	18	17	21
	Recall	100%	89%	95%
	Precision	90%	85%	78%
6	Target	17	16	14
	Relevant	16	13	13
	Recall	94%	81%	93%
	Precision	80%	65%	48%
7	Target	19	15	24
	Relevant	18	14	24
	Recall	95%	93%	100%
	Precision	90%	70%	89%
8	Target	22	17	19
	Relevant	20	14	16
	Recall	91%	82%	84%
	Precision	100%	70%	59%
	Average recall	89%	74%	81%
	Average precision	81%	66%	55%

SOM algorithm for EBS have demonstrated the potential and feasibility of such system-aided meeting facilitation approach.

Term relevance evaluation: Scores of term relevance for the Hopfield list and the Kohonen list were significantly worse than that of the facilitator list

With respect to relevance of the terms used in the three lists (Stage 2), the scores given each term by each subject were summed for the terms on each list and an average score per item was computed for each list. Each of the eight subject's average scores for each list was then analyzed statistically using ANOVA and *t*-test. The results are listed in the Stage 2 row of Table 3. Terms selected by the facilitator scored significantly better than those generated by either the Hopfield classifier or the Kohonen algorithm ($p = .042$ and $p = .037$, respectively). The two system-generated lists clearly were out-performed in

TABLE 3. Experimental statistical results.

	ANOVA	T-test: Facil vs. Hopfield	T-test: Facil vs. Kohonen	T-test: Hopfield vs. Kohonen
Stage 1				
List recall	0.362	0.16	0.460	0.50
List precision	0.046	0.12	0.021	0.30
Stage 2				
Term relevance	0.055	0.042	0.037	0.71
Stage 3				
Term association	0.000	0.000	0.64	0.0009

the term relevance evaluation because of their low precision levels.

Term association evaluation: The score of term association of the Kohonen list was comparable to that of the facilitator list and both of these were significantly better than that of the Hopfield list

In the term association evaluation, our hypothesis was that the Kohonen map would be better than the Hopfield list and comparable to that of the human facilitator due to its graphical display of term-association proximity on a map. Statistical analysis of the term association results (Table 3) shows a significant difference between the Hopfield list and both the facilitator and Kohonen lists ($p =$ 0.000 and 0.0009, respectively) and no difference between the facilitator list and the Kohonen list ($p = 0.64$), verifying this hypothesis.

In summary, the results from our evaluation were encouraging. The Kohonen SOM approach was shown to be comparable to human facilitation in generating high concept recall and rich term association. Its term association capability was also significantly superior to the Hopfield classifier we had developed earlier. However, the precision levels of the system-generated lists were inferior to that of the facilitator output. In light of the cognitive demand and the cumbersome nature of the meeting convergence process, we believe this research has shed light on a promising and intuitively appealing neural network based textual analysis and classification approach.

6. Conclusions and Future Directions

This research demonstrates how software can be designed to perform an intelligent classification of EBS output. In evaluation, we compared the software's performance with that of a human expert facilitator in solving the same complex problem as well as with results of previous work using a Hopfield neural network algorithm. Specific conclusions from this research include: Recall of the topic list generated by the Kohonen was not different from that of the human expert or of the output of the Hopfield algorithm. Topics produced by the Kohonen algorithm were less precise than those produced by the human facilitator but not different from those produced by the Hopfield algorithm. With respect to relevance of the terms to the electronic brainstorming question itself, both the Kohonen and Hopfield algorithms suffered in comparison with the facilitator's selection of terms. Finally, in term association, the Kohonen algorithm performed as well as the expert facilitator.

More work is needed to enhance the SOM-based user interface. Better real-time classification support should be provided in order to allow adjustment of the dimensions of the Kohonen map. Users will need to be able to browse and change the underlying comments/terms within map regions to apply better labels to the regions, to merge similar areas of the map into one conceptual region, or to place comments/terms in regions that they may feel are more appropriate. Algorithmic fine-tuning and testing of different actual EBS sessions will also be performed in the near future.

The graphical representation of the Kohonen SOM provides an intuitive abstraction that should also be useful for organizing large-scale information spaces such as digital libraries, internet homepages, and other business data collections (or organizational memory). These larger collections may require extending the Kohonen algorithm to provide multiple layers of maps to aid in concept-based organization and retrieval of relevant documents. Our ongoing semantic retrieval research involving the Illinois digital library project is investigating such a multi-layered SOM approach \cite{ChenChris95}.

7. Acknowledgment

This project was supported mainly by the following research contract and grants: U.S. Army Corp of Engineers Waterways Experimental Station Research Contract, DACA39-92-K-0042-P00003, 1992–1995 (J. F. Nunamaker); NSF/ARPA/NASA Digital Library Initiative, 1994–1998 (B. Schatz, H. Chen, et al., ''Building the Interspace: Digital Library Infrastructure for a University Engineering Community System''); NSF CISE Research Initiation Award, IRI-9211418, 1992–1994 (H. Chen, ''Building a Concept Space for an Electronic Community System''); and AT&T Foundation Special Purpose Grants in Science and Engineering, 1994–1995 (H. Chen). We would also like to thank the following people for their kind assistance in system development and evaluation: Doug Dean, Jim Lee, Jon Stapley, Boris Nevstrujev, Nina Katic, Jeff McKibben, Andrea Houston, Ken Walsh, Karen Ng, Dorbin Ng, Joanne Martinez, Chris Schuffels, Siunie Sutjaho, and Dr. Jerome Yen.

References

Belew, R. K. (1989). *Adaptive information retrieval.* Paper presented at the Twelfth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Cambridge, MA.

Burton, R. R. (1976). *Semantic grammar: A technique for efficient language understanding in limited domains.* Unpublished doctoral dissertation, Computer Science Department, University of California at Irvine.

Caudill, M. (1993). A little knowledge is a dangerous thing. *AI Expert, 8*(6), 16–22.

Chen, H. (1994). Collaborative systems: Solving the vocabulary problem. *IEEE Computer, 27(5),* 58–66.

Chen, H. (1995). Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms. *Journal of the American Society for Information Science, 46,* 194–216.

Chen, H., Hsu, P., Orwig, R., Hoopes, L., & Nunamaker, J. F., Jr. (1994). Automatic concept classification of text from electronic meetings. *Communications of the ACM, 37(10), 56–73.*

Chen, H., & Lynch, K. J. (1992). Automatic construction of networks of concepts characterizing document databases. *IEEE Transactions on Systems, Man and Cybernetics, 22(5),* 885–902.

Chen, H., & Ng, D. T. (1995). An algorithmic approach to concept exploration in a large knowledge network (automatic thesaurus consultation): Symbolic branch-and-bound vs. connectionist Hopfield net activation. *Journal of the American Society for Information Science, 46,* 348–369.

Chen, H., Schatz, B. R., Yim, T., & Fye, D. (1995). Automatic thesaurus generation for an electronic community system. *Journal of the American Society for Information Science, 46,* 175–193.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Dalton, J., & Deshmane, A. (1991). Artificial neural networks: An approach to increasing machine intelligence. *IEEE Potentials, 4(33),* 33–36.

Doszkocs, T. E., Reggia, J., & Lin, X. (1990). Connectionist models and information retrieval. *Annual Review of Information Science and Technology (ARIST), 25,* 209–260.

Ellis, C. A., Gibbs, S. J., & Rein, G. L. (1991). Groupware: some issues and experiences. *Communications of the ACM, 34(1),* 38–58.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, Rinehart, and Winston.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1983). Statistical semantics: Analysis of the potential performance of keyword information systems. *Bell System Technical Journal, 62(6),* 1753–1806.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM, 30(11),* 964–971.

George, J. F., Easton, G. K., Nunamaker, J. F., Jr., & Northcraft, G. B. (1990). A study of collaborative group work with and without computer-based support. *Information Systems Research, 1(4),* 394–415.

Hiotis, A. (1993). Inside a self-organizing map. *AI Expert, 8(4),* 38–43.

Hopfield, J. J. (1982). Neural network and physical systems with collective computational abilities. *Proceedings of the National Academy of Science, 79(4), 2554–2558.*

Kohonen, T. (1989). *Self-organization and associative memory.* Berlin: Springer-Verlag.

Kohonen, T. (1995). *Self-organization maps.* Berlin, Heidelberg: Springer-Verlag.

Kwok, K. L. (1989). *A neural network for probabilistic information retrieval.* Paper presented at the Twelfth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Cambridge, MA.

Lin, X., Soergei, D., & Marchionini, G. (1991). *A self-organizing semantic map for information retrieval.* Paper presented at the Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Chicago, IL.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine, 4,* 4–22.

MacLeod, K. J., & Robertson, W. (1991). A neural algorithm for document clustering. *Information Processing & Management, 27(4),* 337–346.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas imminent in neural nets. *Bulletin of Mathematical Biophysics, 5,* 115–137.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory.* Cambridge, MA: The MIT Press.

Nunamaker, J. F., Jr., Dennis, A. R., Valacich, J. S., & Vogel, D. R. (1991). Information technology for negotiating groups: Generating options for mutual gain. *Management Science, 37(10),* 1325–1346.

Nunamaker, J. F., Jr., Dennis, A. R., Valacich, J. S., Vogel, D. R., & George, J. F. (1991). Electronic meeting systems to support group work. *Communications of the ACM, 34(7),* 40–61.

Orwig, R. E. (1995). *A graphical, self-organizing approach to classifying electronic meeting output.* Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Post, B. (1992). *Building the business case for group support technology.* Paper presented at the 25th Hawaii International Conference on System Sciences, Maui, Hawaii.

Rasmussen, E. (1992). *Clustering algorithms.* Englewood Cliffs, NJ: Prentice Hall.

Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics, 61,* 241–254.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning internal representations by error propagation.* Cambridge, MA: The MIT Press.

Ryan, B. F., Joiner, B. L., & Ryan, T. A. (1985). *MINITAB Handbook, 2nd ed.* Boston: PWS-KENT Publishing Company.

Salton, G. (1989). *Automatic text processing.* Reading, MA: Addison-Wesley Publishing Company, Inc.

Stepp, R. E. (1987). *Concepts in conceptual clustering.* Paper presented at the Proceedings of the Tenth International Joint Conference on Artificial Intelligence, Milan, Italy.

Vogel, D. R., Nunamaker, J. F. J., Martz, W. B. J., Grohowski, R., & McGoff, C. (1989). Electronic meeting system experience at IBM. *Journal of Management Information Systems, 6(3),* 25–43.

Woods, W. A. (1972). An experimental parsing system for transition network grammars. In R. Rustin (Ed.), *Natural language processing.* New York: Algorithmics Press.