AN AUTOMATIC CLASSIFICATION APPROACH TO IDEA ORGANIZATION IN GROUP SUPPORT SYSTEMS

by

Ming Yuan

A Dissertation Submitted to the Faculty of the

COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

2008

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Ming Yuan

entitled An Automatic Classification Approach to Idea Organization in Group Support Systems

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 05/27/2008
Jay F. Nunamaker, Jr.

____ Date: 05/27/2008
J. Leon Zhao

____ Date: 05/27/2008
Robert Briggs

____ Date: 05/27/2008
Don Fallis

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 05/27/2008
Dissertation Director: Jay F. Nunamaker, Jr.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Ming Yuan

ACKNOWLEDGMENTS

First of all, I would like to thank all my committee members, Dr. Jay F. Nunamaker, Jr., Dr. Robert Briggs, Dr. Don Fallis, Dr. Mohan Tanniru, and Dr. J. Leon Zhao for their guidance throughout my doctoral study at the University of Arizona. I believe that what I have learned from them will significantly benefit my future career. Special thanks to Jay for his generous support and supervision in completion of this dissertation. Without his vision and guidance, this study would not have been possible.

Special thanks to Dr. Robert Briggs, who have consistently and generously helped me in my research from my preliminary exam to my dissertation defense. Despite the short notice of my research, Bob took countless hours to provide insights and detailed suggestions on how to improve my dissertation study. His insights, expertise, and guidance are invaluable to me and will continue to inspire and guide me in my future career.

I would like to thank my good colleagues and friends in the University of Arizona. Special thanks to Chris Diller, who provided me with much-needed help in contributing to my experiments as an expert facilitator. His smile and generosity prove to be one of the best memories I have for my doctoral years.

DEDICATION

This dissertation is dedicated to my family for their constant support and encouragement throughout my graduate study in the U.S. First and foremost, I am grateful for the constant support, understanding, patience, and trust of my wife, Juan Wu. She has always been there to provide me with her unwavering support during the hard times.

I would also like to dedicate this dissertation to my parents, Xuanze Yuan and Baozhu Li, and my brother, Bo Yuan, who always care for me and support me in the pursuit of my Ph.D.

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

With the rapid advance of information technologies, human beings increasingly rely on computers to accumulate, process, and make use of data. Knowledge discovery techniques have been proposed to automatically search large volumes of data for patterns.

Group Support Systems (GSS) play an important role in streamlining group activities and improving group outcomes. Various attempts have been made to help automate several tasks in group activities under GSS environment. One of the approaches is to apply automatic approach to idea organization task in GSS.

This research designed and tested a system using singular value decomposition (SVD) techniques to automate the idea organization task in GSS. Specifically, this research was conducted to examine how the idea organization task, typically regarded as the most labor-intensive and cognitively demanding in group problems solving, can be automated using a system enabled by singular value decomposition techniques. For the purpose of evaluation, we compared the performance of our automated approach using SVD algorithm against that of human subjects. Two separate experiments were conducted to evaluate the performance of the automatic approach on two essential components of an idea organization; namely, generation of topic categories and placement of relevant comments into their respective categories.

The general conclusion that can be made from this research is that idea organization in group problem solving can be facilitated both efficiently and effectively with the SVD-enabled system that can automatically generate topic categories and place relevant

comments into their respective topic categories. Therefore, our automatic approach may provide a useful and promising tool for the idea organization task in GSS.

CHAPTER 1: INTRODUCTION

This chapter overviews group support systems and text mining and shows that this research is motivated by the need for improved cognitive support for the idea organization task in group support systems. Contribution of this research is summarized and the organization of the dissertation is presented.

1.1 Group Support Systems

Collaboration has played an important role in contributing to the success of many organizations (Mintzberg 1983). Many of the tasks faced by modern organizations are complex ones that can only be accomplished by pooling expertise and resources from many people. With collaboration playing an integral role in business environment, one of the major technological advances in the early 1980s is the development of group support systems to assist group collaboration, which can be seen as representing a shift in focus from the single user supported by decision support systems to a group of participants. Group support systems assist large groups to interactively work on a single problem or collection of problems (Vogel, Nunamaker et al. 1989; Nunamaker, Dennis et al. 1991). Under GSS environment, a number of groups of participants use a network of computers either through a local area network of terminals or through Internet to discuss and solve various kinds of organizational problems. GSS provide various tools to support group collaboration. For example, brainstorming tools allow participants to contribute ideas electronically in parallel without waiting in turn whereas idea organization tools allow participants to classify brainstorming comments into a manageable list of topics or topics.

1.2 Text Mining

Technological advances in the field of information technology have changed businesses around the world and given rise to new industries around controlling and providing information. That is the reason why today's world is often referred to as "information age", in which global economy shifts its focus away from the production of physical goods and toward the manipulation of information. The rapid development of information technologies has led to development of numerous new software applications, which in turn have produced vast amounts of digitized data and information. It is reported that past three decades have witnessed more information produced than in the previous five thousand years of human history (Trout 1997).

While people in the past could well manage to manually process limited information generated or rely on traditional data management and analysis technologies, the dramatic increase of digitized information may prove challenging for them to process information in an effective and efficient manner. It is no longer sufficient or even feasible to process a large amount of information manually or even with traditional technologies. The more information generated using new information technologies, the more difficult it is for people to understand it. Understanding typically involves the processing of the raw data into underlying knowledge. One important characteristic of knowledge is that all knowledge has a structure, which is essential to understanding. Structure organizes disparate information into a form, which is easy to digest and understand. For example, all newspapers group information into different categories such as politics, sport, business. Knowledge discovery is to "extract implicit, previously unknown, and potentially useful information from data" (Frawley, Piatetsky-Shapiro et al. 1991).

With rapid growth of the volume of electronic data currently available, automated knowledge extraction techniques are increasingly needed to process the huge amounts of data stored in the information systems. As one part of knowledge discovery, data mining is referred to as techniques applied in "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (Fayyad, Piatetsky-Shapiro et al. 1996). While data mining techniques are designed to operate on data in structured formats such as those stored in relational databases, text mining techniques are concerned with various tasks such as extraction of information implicitly contained in collections of textual documents such as text from the literature repository, thus extending the more traditional data mining approach to unstructured textual data.

1.3 Motivation

One of the most important effects of GSS is that GSS enable group participants to generate more brainstorming ideas than face-to-face meetings. Two factors might explain this increase in brainstorming contributions: parallel communication and anonymity. Unlike the take-turns approach in face-to-face meetings, GSS allow participants to brainstorm ideas via parallel communication anonymously. Specifically, parallel communication allows participants to simultaneously input their comments electronically into the system without waiting in turn whereas anonymity encourages participants to express their ideas freely without showing their identities. As a result, unlike face-to-face meetings, group support systems often generate large quantities of text in digital formats in a very short period of time during brainstorming sessions. While GSS greatly improve creative idea generation in the divergent task of brainstorming, the overwhelming amount of brainstorming comments exert tremendous pressure on group participants who need to complete the idea organization task.

As an important step in a collaborative process in GSS, idea organization is about moving from having many ideas to having fewer ideas worthy of further discussion and having more understanding of the relationships among ideas. This step of idea organization is important in the collaborative process due to its ability to reduce cognitive load on group participants. Instead of working on raw brainstorming comments, group participants can work on organized key ideas. While there may be different ways to organize ideas, this study focuses on one type of idea organization consisting of two parts: 1) consolidation of brainstorming comments into a manageable number of key topics; 2) placement of comments into their relevant topics. This type of idea organization constitutes as an important way of idea organization because it has been frequently used in different types of group sessions such as in system analysis and strategic planning. Another example is the group session results used as experimental data later in this study that were generated using this way of idea organization.

Idea organization is a much harder task cognitively than the divergent task of brainstorming in that people have to read and understand others' points and then reach consensus in categorizing these ideas into conceptually important themes. The overwhelming amount of comments generated by brainstorming sessions in GSS makes it very difficult for participants to generate topics and place comments into relevant topic categories. Therefore, the increase in creative ideas during brainstorming may tend to be offset by the information overload and cognitive demand experienced by both facilitator and group participants in the idea organization task.

Currently, the idea organization task is still performed manually by group participants and facilitators in GSS. Therefore, what GSS need is some form of cognitive support for group participants to help them make sense of the large quantity of brainstorming comments data. With the advancement of text mining research, it is possible to take advantage of the existing electronic textual data and reduce the work overload for group participants by applying text mining techniques to automatically perform the idea organization task in GSS. Automating idea organization can potentially reduce information overload, lessen cognitive demand, shorten idea organization time, and provide context to topics instantly.

This research is motivated by the question of how to automate idea organization task in GSS with quality comparable to humans.

1.4 Contributions

The objective of this research is to address the problems associated with the idea organization task in GSS by designing a tool using singular value decomposition to automatically identify important topic categories from all brainstorming comments and place relevant comments into their respective topic categories with quality comparable to human subjects and at much faster speed.

The major contributions of this research include an IT artifact: the SVD-enabled automatic idea organization tool, which is empirically tested to be able to automate idea organization efficiently and effectively with results comparable to human subjects. With the aid of this tool, group participants can reduce the information overload and cognitive demand problems, and further refine the topics and comments assignments for subsequent GSS stages.

1.5 Outline of the Dissertation

The rest of the dissertation is structured as follows: Chapter two describes the group support systems and explains in detail the problems that exist in the idea organization task under GSS environment. Chapter three reviews the related literature on automatic idea organization. Chapter four introduces the research question and presents the system development methodology that guides this research. Chapter five describes the application of singular value decomposition approach to idea organization. Chapter six describes two experiments that compare the system-generated output with manually generated results, and reports the results of the experiments. Chapter seven discusses the results of experiments, offering interpretation and insights into the results. This research is concluded with chapter eight, which describes conclusions, contributions, limitations and future directions of this research.

CHAPTER 2: GROUP SUPPORT SYSTEMS AND PROBLEMS WITH IDEA ORGANIZATION IN GSS

This chapter describes group support systems, groupware grid and collaboration processes. Problems with the idea organization task in GSS are discussed along with the need for automatic idea organization.

2.1 Group Support Systems

Collaboration has frequently taken place in the form of project teams or groups in today's business organizations. These teams and groups constitute as an important and almost universal organizational structure. During group collaboration, team participants are able to share information, generate ideas, make decisions and review the effects of decisions (Phillips and Phillips 1993). In contrast, individuals working alone may have to rely on their own abilities to tackle problems without help from others. With skills and abilities of a group of people harnessed, a complex organizational problem or an important opportunity can be addressed collectively in a better way than individually because collective knowledge, skill, and expertise of the group are greater than that of any individual (Maier and Hoffman 1960; Martz, Vogel et al. 1992). This is particularly true for problems that require multiple skill sets.

With collaboration playing an integral role in the business environment, group support systems have been created to assist group collaboration. Group support systems are a combination of hardware, software and facilitation support used to engage in intellectual collaborative work (Jessup and Valacich 1993). The CMI group of the

University of Arizona has been one of the pioneers in GSS field with its commercially available GSS software called GroupSystem and a web-based version ThinkTank-2.0. GSS help focus and structure a team's deliberation while minimizing cognitive costs of communication and information access and reducing distraction among team participants (Davison 2000). Instead of sitting around a table and taking turns to talk in conventional meetings, group participants in GSS environment are connected together using networked computers and interact with each other via keyboard and monitor screens.

Compared with face-to-face group collaborations, GSS tools provide many distinct characteristics including parallel communication, anonymity, group memory, and a structured pattern of discussion (DeSanctis and Gallupe 1987; Dennis, George et al. 1988). Benefits associated with the use of GSS are both tangible and intangible (De Vreede, Vogel et al. 2003). Among the tangible benefits are reduced time and resources required for the meeting, and increased amount of brainstorming ideas of higher quality as compared with face-to-face interactions without technology assistance (Dennis, Valacich et al. 1990; Grohowski, McGoff et al. 1990; Gallupe, Bastianutti et al. 1991; Gallupe, Dennis et al. 1992; Nunamaker, Briggs et al. 1996). Past research has also shown that GSS can lead to increased productivity, increased information flow between group participants, more objective evaluation of information, and synergy within the group (Nunamaker, Jr. et al. 1991; Adkins, Burgoon et al. 2003). Among the intangible benefits are increased level of group cohesiveness, improved problem definition, and stronger commitment from the group to the solution (Nunamaker, Briggs et al. 1996).

In the following, we will first discuss hardware components, followed by software tools and conclude with facilitation support in GSS.

2.1.1 Hardware Components in GSS

A GSS-supported meeting often requires hardware including networked personal computers or workstations and other peripheral equipment such as a projected screen for shared data and various audio and video equipments to enhance the communication channels among group participants.

A trend has been taking place that is moving from traditional client and server type of GSS to web-based GSS over the past several years. In a traditional GSS environment such as GroupSystems, a dedicated facility is often needed for electronic meetings, which requires investment of expensive resources. An example of such facility may include a dedicated electronic meeting room filled with GSS-installed client workstations and one or two dedicated computers that a meeting facilitator uses for controlling GSS installed in each participant's client machines. One or two projection screens are often set up in the front of the room for projection of the facilitator's screen when necessary. In contrast, in a web-based GSS like ThinkTank, there is no need for a dedicated facility including dedicated rooms or dedicated computers with installed GSS software. Everyone can take part in a GSS session anywhere as long as the participant has a computer with internet access. Unlike traditional client and server type of GSS, web-based GSS software is not required to be installed in computers of each participant. All these unique features make web-based GSS a popular choice for distributed collaborations.

2.1.2 Software Tools in GSS

GSS consist of a suite of software tools designed to connect group participants together through a network to support problem formulation and solution in group meetings (DeSanctis and Gallupe 1987). These software tools in GSS serve a variety of functions, which may help quickly generate lots of ideas to solve problems or find opportunities, distill those ideas to the very best, clarify exactly what is meant, organize the ideas, evaluate and prioritize them, build consensus among the team, and eventually generate deliverables that help the team take action. By mixing different tools together, GSS can be utilized to suit the characteristics of the problem as well as the characteristics of the group of problem-solving participants. Table 2.1 shows a suite of tools included in ThinkTank, a web-based GSS, to support one or more group activities such as generating, organizing and evaluating ideas.

Categorizer's Brainstorming Mode	Entering comments simultaneously
Categorizer 's Organize mode	Making lists of topics with underlying comments. Lists can be organized in definable categories
Rank/Order/Vote	Ranking priorities and evaluating ideas using various voting techniques
Alternative Analysis	Rating, comparing, and analyzing alternatives, as well as developing consensus concerning specific issues

Table 2.1 Modules in ThinkTank

Brainstorming is the default mode in ThinkTank's categorizer tool. It is similar to an instant messaging application where participants are to type ideas into the submission panel on the bottom of the screen and press 'enter'. This tool enables participants to interact and exchange ideas electronically via typing into keyboard instead of talking in conventional meetings. All participants can submit their ideas through the submission panel simultaneously and anonymously. All the comments submitted by participants can be seen by other participants on the screen. Individual participants can read others' contributions and make comments on those ideas. Originally, comments submitted by other participants appear in bold letters until being clicked to mark them as being read already. Functions that can be performed on submitted comments include edition, deletion, copy, etc.

The 'organize' mode embedded in categorizer tool allows participants to use the submission panel to create buckets or categories. Group participants can work with facilitators to read through all the comments generated from the brainstorming session and then work together to generate a manageable list of possible topics that sum up the majority of the comments. The 'organize' mode in categorizer tool also lets participants deposit ideas into one or more 'buckets' to organize them by dragging and dropping a comment into its appropriate category. In addition, categorizer allows participants to indent and out-dent ideas to display inter-relationships and connections. These tools are made available to organize the ideas generated from the brainstorming session.

ThinkTank's rank/order/vote tool enables participants to rank priorities, preferences, goals, or any other items. Rank/order/vote is a method in which participants rank ideas in order of preference. Winners are determined by giving each candidate a certain number of points corresponding to the position in which it is ranked. Ballot items can be dragged and dropped into their preferred order before votes are cast. In addition, ThinkTank's alternative analysis tool is a multi-criteria voting tool that offers a diverse range of capabilities. It allows teams to rate, compare, and analyze alternatives, as well as develop consensus concerning specific issues. Criteria for voting can be any one of six distinct types from a sliding scale to a simple yes or no. The session leader decides how many criteria to use and which voting methods to apply to each of the criteria. After the leader has set up the ballot, participants simply mark each item with their votes. Both these tools can be used to evaluate options or prioritize to determine what needs to get done and in what order.

ThinkTank's alternative analysis tool also allows group participants to spot voting results in disagreement more easily by using a statistical measure called standard deviation and the vote spread. The 'average' column appears in green if participants are in consensus and in red if they disagree. The identification of an area of disagreement allows participants to focus discussion on it and eventually may resolve the conflicts. The colors on the matrix allow participants to see instantly the items that should be discussed. These additional functions provided by alternative analysis tool can be used to build consensus among participants.

2.1.3 Facilitation Support in GSS

Another important component of GSS is facilitation support, which can be defined as a group of activities that a facilitator carries out before, during, and after a meeting in order to help a group during the decision making process (Bostrom, Anson et al. 1993). Facilitation support has been in existence long before the emergence of GSS. The availability of GSS tools present facilitators with additional responsibilities that require them to direct group participants on what GSS tools to use and when to use them (Dickson 1993). Facilitators may differ in terms of their expertise, experience and level of involvement with the group process (Aakhus, Adkins et al. 1997). In situations without GSS support, human facilitation has been shown to improve group performance (Reagan-Cirincione 1992). With respect to GSS-supported environment, previous research has shown the importance of facilitation in impacting productivity and outcome of the group (George, Dennis et al. 1992; Valacich, Dennis et al. 1994; Anson, Bostrom et al. 1995; Griffith, Fuller et al. 1998; Briggs, De Vreede et al. 2001)**.** Facilitator can support the group's social and cognitive processes, thus allowing participants to focus more on substantive issues in the decision making process and achieve better outcomes (Schuman 1996; Khalifa 2002).

There are three kinds of distinct process support provided by facilitators (Nunamaker, Jr. et al. 1991): a chauffeured style, a supported style and an interactive style. These different styles can be combined with each other at any stages during a collaborative process. The chauffeured style allows group participants to give directions to one person on what features or software tools to use. With the chauffeured style, group participants do not have access to GSS so they do not need to be familiar with the GSS tools. They verbally discuss the issues and then the facilitator is to enter information by managing tools in GSS. The supported style allows each group participant to have access to computers and all group participants can communicate using GSS tools. That is, all group participants have access to GSS tools rather than just the facilitator as in the chauffeured style. The meeting process proceeds using a mixture of verbal and electronic interactions. The role of facilitator is about providing technical support and teaching group participants how to use GSS tools. The interactive style also provides each group participant with computers and GSS tools. However, with the interactive style, all communications are conducted using GSS tools and almost no verbal communications are used. Similarly, facilitator also needs to provide technical support to make sure group participants use GSS tools appropriately.

2.2 GSS and Groupware Grid

Groupware can be defined 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 et al. 1991). To better understand functions of GSS, it is important to put GSS within the framework of groupware so as to understand their contributions to collaboration. In fact, GSS only represent a subset of groupware. The following describes the groupware grid and the mapping of GSS to the groupware grid to get a better understanding of GSS functions.

2.2.1. Groupware Grid

Groupware grid can be used as a theory-based heuristic model to map functionalities of a groupware technology to organizational work (Nunamaker et al.,2001). Table 2.2 shows that groupware grid consists of vertical and horizontal dimensions.

	Communication Support	Deliberation Support	Information Access Support
Concerted Work Level			
Coordinated Work Level			
Individual Work Level			

Table 2.2 Groupware Grid

Vertically, there are three levels of collaborative effort that can be made more effective with collaborative technologies. At the individual work level, group participants make individual efforts that require no coordination. Group productivity equals to the sum of individual efforts. Technologies such as word processors and spreadsheets may be used effectively to support individual work. At the coordinated work level, group work requires careful coordination among participants despite the fact that they still make individual efforts. Productivity depends on the level of individual effort and on the coordination among those efforts. Technologies such as E-mail and workflow automation may support coordinated efforts. At the concerted work level, group participants must make a continuous concerted effort in synchrony with other participants. Often times the performance of any one participant influences the performance of other participants. Tools that may be used to enhance concerted efforts include electronic brainstorming tools, group outlining tools, and idea categorizers in GSS.

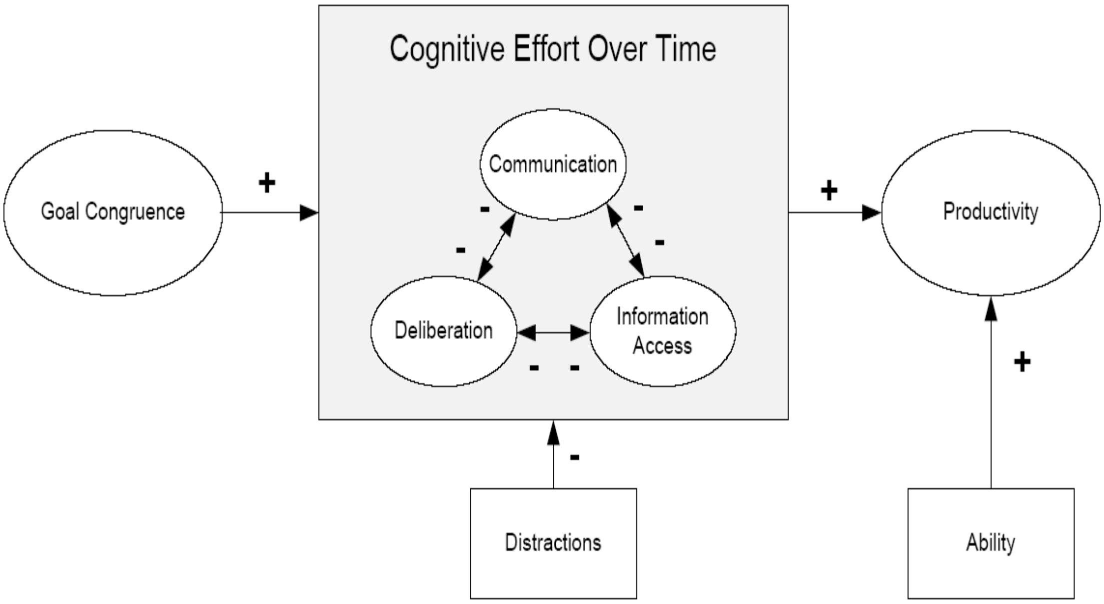

Figure 2.1 Team Theory of Group Productivity

The horizontal dimension of groupware grid comes from the team theory of group productivity, previously known as focus theory (Briggs 1994; Briggs, Reinig et al. 1997). Team theory, as shown in figure 2.1 presents a causal model for the productivity of a group in making a joint cognitive effort toward an objective. This theory posits that one of the key limits on group productivity is human attention resources. According to the team theory, team participants divide their attention into three cognitive processes: communication, deliberation and information access.

Communication refers to the process that the sender sends stimuli to the receivers who then assign meaning to the stimuli. Communication occupies attention resources in terms of choosing words, behaviors, images, and artifacts, and presenting them through a medium to other team participants. Under a GSS environment, communication is conducted in the form of textual messages, limiting the diversity and richness of the stimuli (Dennis and Valacich 1999). Deliberation refers to the cognitive tasks that are conducted in order to achieve the common goal. Deliberation utilizes attention resources via forming intentions toward accomplishing the goal and conducting problem-solving activities which consist of understanding the problem, developing and evaluating alternatives, selecting and planning a course of action, etc. Information access refers to the process of obtaining information while incurring cognitive costs associated with receiving and processing the information. Information access deals with attention demands of finding, storing, processing, and retrieving the information that the group participants need to support their deliberation. These three processes interfere within each other and thus limit group productivity.

To sum up, the horizontal axis of groupware grid addresses the potential for technology to reduce cognitive costs of joint effort. Groups may become less productive if the attention demands for communication, deliberation, or information access become too high. For example, as the size of the group grows, the amount of information communicated may exert a large cognitive burden on the participants, potentially hampering the collaborative process. The GSS process must minimize these cognitive costs so that the increased information flow enabled by GSS does not negatively affect the overall productivity of the group. That is, the GSS must improve productivity to the degree that the attention costs are reduced.

2.2.2 Mapping GSS to Groupware Grid

One of the benefits of using groupware grid is that individual groupware tools or entire groupware systems like group support systems can all be mapped into the cells of groupware grid. The three levels of group support on the vertical axis show that group support tools must strive to provide support on the concerted level. The three kinds of attention resources on the horizontal axis demonstrate the need for tools that can efficiently and effectively leverage the attention resources of participants within the group.

Table 2.3 shows that group support systems provide support at the concerted work level for all three attention resources: communication, deliberation, and information access (Nunamaker, Jr.; Briggs et al. 1996).

	Productivity Processes		
	Communication	**Deliberation**	**Info Access**
Concerted Level	Anonymity, Parallel Contribution	Structured and Focused Processes	Session Transcripts

Table 2.3 Classification of Benefits from Group Support Systems

For communication, GSS enable anonymity and parallel input to improve communication at the concerted level. For deliberation, there are various tools embedded in GSS that provide deliberation support in concerted level. For example, the brainstorming tool enables group participants to diverge from familiar thinking patterns by preventing them from thinking narrowly. Idea organizer also allows group participants to focus on a limited number of key points. Similarly, electronic voting and multi-criteria evaluation tools also help improve deliberation support. For information access, GSS improve information access at the concerted level by providing access to the information and by providing session transcripts to group participants.

To sum up, GSS are used to focus and structure group deliberation while reducing cognitive costs of communication and information access among teams working collaboratively towards a goal.

2.3 Collaboration Process in GSS

A GSS-enabled collaboration often follows a sequence of activities related to decision making. In order to understand the important steps during collaboration, a review of different problem-solving processes is followed:

Since the objective of collaboration is often to make decisions with input from multiple decision makers, collaboration often follows the general steps of a decision making process. Herbert Simon divided decision-making into three general steps including intelligence, design, and choice (Simon 1965). Specifically, intelligence step refers to the exchange of relevant information; design step involves the development of decision alternatives or available options; choice step allows selection of the best option or alternative.

As shown in table 2.4, Briggs et al. identified five basic patterns of collaboration in group decision making, which further refine Simon's three steps of general decision making (Briggs, Vreede et al. 2003).

Diverge	move from having fewer to having more ideas
Converge	move from having many ideas to a focus on and understanding of fewer ideas deemed worthy of

	more attention
Organize	move from less to more understanding of relationships among idea comments.
Evaluate	Move from less to more understanding of consequences for choices toward attaining group goals
Build Consensus (align goals)	Move from less to more agreement among a group of stakeholders; achieving more congruence between individual and group goals.

Table 2.4 Patterns of Collaboration

'Diverge', often called brainstorming, is to move from having fewer to having more ideas (Briggs, Vreede et al. 2003). Basically, brainstorming activity allows group participants to generate as many ideas or potential solutions as possible without regard to the feasibility of those ideas at this stage (Couger 1995). The objective of brainstorming activity is to develop new or innovative ways of looking at things and avoid established patterns of thinking. Therefore, brainstorming is a very useful activity that helps develop highly creative solutions to a problem by pooling the experience of all team participants into play during problem solving. The output from this divergent stage is typically a large number of unorganized textual idea comments.

'Converge' and 'organize' are often called idea organization. 'Converge' is to move from a state of having many unrefined ideas to a state of having a reduced set of ideas that are regarded as worthy of further attention. 'Organize' is to increase the understanding of the relationships among ideas. All brainstorming ideas are synthesized into specific topics that may provide potential solutions to the problem at hand. The key of this step is to generate a manageable set of topic categories, which serves as potential solutions and represents the majority of the themes reflected in all brainstorming ideas.

The output from 'converge' and 'organize' is a manageable list of key topics with ideas assigned to appropriate topics.

'Evaluate' is to improve understanding of possible consequences of all choices. After idea organization, group participants are presented with a manageable list of potential options. The task is to decide on the best possible solutions to the problem at hand based on a set of pre-determined criteria. The output from evaluation is a ranked list of best solutions available.

'Build consensus' is to improve the degree of agreement among group participants. Consensus is a strategy ensuring that everyone in the group plays a role in the decision making. It may not necessarily mean unanimity, but should represent extensive support among group participants. Consensus building provides workable solutions by removing possible conflicts, which consume efforts of the group and undermine progress towards the goal. With conflicts resolved through consensus, the group can create a better and long-lasting solution. However, it may not be easy for the group to achieve a consensus because evaluations from the previous step may reveal a large disagreement on certain issues. Therefore, evaluation and consensus building may be iterative, which means that the group may start over again until a certain level of consensus is reached.

In this research, a simplified four-step process is adopted. Step 1 – Brainstorming: group participants conduct brainstorming to identify all possible issues related to a challenge or problem. Step 2 – Idea Organization: group participants consolidate all issues into a manageable list of general topics of issues that are important to solving the problem at hand, and classify relevant issues into their appropriate topic categories. Step

3 – Evaluation: group participants evaluate the topic categories that are potential solutions to the problem so as to narrow down to a few best solutions. Step 4 – Consensus Building: group participants reach consensus on determining the relative importance of the general topics in solving the problem by polling group participants on a set of criteria. We refer to these steps as: brainstorming, idea organization, evaluation and consensus building.

This research mainly focuses on idea organization, which is discussed in detail in the next section.

2.4 Idea Organization

Idea organization mainly consists of two components including topic generation (or topic classification) and placement of relevant comments into their respective topics. Topic generation is to arrive at a list of topics that represents the content of brainstorming comments. Comment placement is to link comments to each topic in the topic list so that group participants can trace topics back to their original comments from which the topic list is derived. In a way, the comments that are linked to topics can serve as validation of the topics.

Table 2.5 shows a sample list of brainstorming comments from a real group session attended by college students using GroupSystems. The brainstorming question presented in the session is: “What can we do to make our student union center a better place?” With this brainstorming question in mind, group participants used the brainstorming tool provided by GroupSystems to generate as many comments as they could.

C1:	Clearly more technology related
C2:	But how do we foster viewpoint neutrality when student affairs overall is incredibly left-leaning in a usually unabashed way?
C3:	viewpoint neutrality - safe for all opinions, all sides, all viewpoints can be safely discussed
C4:	Much more technology based, greater demand for wireless
C5	I hope that viewpoint neutral does not mean that strong opinions are not expressed in the union, but rather that the union is a place where these opinions are shared in a civil manner, and that the union itself does not take a politically partisan position on issues
C6:	With less money we will be forced to find more efficiencies - with technology growth it's likely that more core function operate out of the unions i.e., financial aid, career development, academic advising - space won't be a concern due to technology
C7	Technology will mandate that the way we reach out to students and keep them engaged in new ways. Many reasons why students come to our facilities are no longer applicable
C8	Hopefully union will continue to be safe place/viewpoint neutral as our political climate changes
C9	Very flexible in order to stay with the many changing attitudes and needs of the student population-and will stay central to University community

Table 2.5 A Sample List of Brainstorming Comments

Table 2.6 shows an example of topics with relevant comments attached under each topic. This list of topics and associated comments is generated during idea organization after group participants browse through the sample of brainstorming comments.

The list contains two major topics: ‘technology’ and ‘viewpoint neutral’, which are closely related to solving the brainstorming question. There is an additional category called “other topics”, which lists ideas that do not belong to the identified two topics.

In addition to two major topics generated, the original comments from which the topics are derived are also linked to the appropriate topics. This comment placement (or comment linkage/assignment) provides context support and explanation to the semantics of the category labels ascribed to the two topics identified by group participants. Each of the topics contains either a single term or a term phrase as labels. Topic one is 'technology', which is a single term identified to provide one potential solution to making student union a better place. Topic two is 'viewpoint neutral', which is a term phrase identified as one of the ways that can make student union center a better place. The comments attached to these two topics provide rich context as to what these topics mean.

Topic 1: Technology	
C1:	Clearly more technology related
C4:	Much more technology based, greater demand for wireless
C6:	With less money we will be forced to find more efficiencies - with technology growth it's likely that more core function operate out of the unions i.e., financial aid, career development, academic advising - space won't be a concern due to technology
C7	Technology will mandate that the way we reach out to students and keep them engaged in new ways. Many reasons why students come to our facilities are no longer applicable
Topic 2: Viewpoint Neutral	
C2:	But how do we foster viewpoint neutrality when student affairs overall is incredibly left-leaning in a usually unabashed way?
C3:	viewpoint neutrality – safe for all opinions, all sides, all viewpoints can be safely discussed
C5	I hope that viewpoint neutral does not mean that strong opinions are not expressed in the union, but rather that the union is a place where these opinions are shared in a civil manner, and that the union itself does not take a politically partisan position on issues
C8	Hopefully union will continue to be safe place/viewpoint neutral as our political climate changes
Other Topics:	
C9	Very flexible in order to stay with the many changing attitudes and needs of the student population-and will stay central to University community

Table 2.6 A Sample List of Topics Generated and Related Comments

2.4.1 Problems with Idea Organization

Idea organization refers to moving from a state of having many ideas to a state of having only important topic categories drawn from all ideas. Group members browse the list of ideas and come up with preliminary suggestions of possible topic categories. During the process, topics that are redundant or similar are merged into one topic and topics that are irrelevant to the current goal are eliminated. Often times an appropriate level of abstraction has to be reached. Topics that are too general are dropped and topics that are too specific may be merged into a more general topic. When eventually a manageable list of topics is generated, group members typically link all the topics with original comments that support or explain the topic. This step may be used to demonstrate the validity of the topics generated and make group participants better trace the topics back to the original comments used to generate them.

While brainstorming is a divergent task, idea organization is a convergent task. As a convergent task, idea organization is more daunting because it requires participants to sort through all comments generated in the brainstorming session and come up with a list of important topics capturing major underlying themes of comments. Specifically, most difficulties related to idea organization arise from the information overload problems caused by the high output of the divergent activity of brainstorming and the high cognitive demand associated with the task of idea organization itself. The adverse effect of information overload and cognitive demand can be observed from the sharp drop in satisfaction level of group participants during idea organization in group sessions supported by GSS.

In the following, a detailed examination of information overload and cognitive demand problems is provided along with a discussion on the fluctuation of satisfaction levels during the course of electronic meetings supported by GSS.

2.4.1.1 Information Overload

The root cause of the information overload problem can be attributed to the divergent task adequately supported by brainstorming tools in GSS. The divergence stage in GSS is used to generate as many ideas as possible to tackle a specific problem. However, one of lessons learned over the years about brainstorming in GSS is that group participants in GSS environment produce many more contributions in the divergent task than do people in unsupported meetings (Nunamaker, Jr.; Briggs et al. 1996). Thus, electronic brainstorming tools in GSS are in effect a double-edged sword. On one hand, electronic brainstorming tools in GSS successfully enable creativity in generating ideas; on the other hand, there is an overwhelming amount of information generated, thus leading to information overload problem during idea organization.

Two key benefits of GSS may explain the information overload problem: parallel contribution and anonymity (Nunamaker, Dennis et al. 1991).

With parallel communication in GSS tools, all participants can contribute simultaneously to one shared repository while they can also instantaneously view contributions of any one user through their own computer screens. The distinct feature of parallel communication in GSS frees participants from the limitations of take-turns approach in conventional face-to-face meetings in three ways as identified by previous research: (1) participants who have ideas to contribute may have to wait, but may forget or suppress them because they may not be relevant later; (2) while waiting to contribute an idea, participants may focus on remembering that idea instead of generating new ones;

(3) participants may focus on listening to others' ideas rather than contributing their own (Jablin and Seibold 1978; Diehl and Stroebe 1987). With parallel communication, participants do not have to take turns to contribute their ideas, thus effectively inputting as many ideas as they have without the above-mentioned three adverse effects as seen in conventional face-to-face meetings.

Anonymity helps remove contribution barriers related to personalities, thus leading to more idea contributions. In conventional face-to-face interactions, every idea contribution can be traced directly to its contributors; with this in mind, some participants, especially those with low status, may tend to withhold ideas for fear of negative evaluation or for intended or unintended pressure to conform (Hackman and Kaplan 1974; Diehl and Stroebe 1987). With anonymity, GSS in effect encourage more contributions from participants, who can freely express their ideas. Therefore, with GSS removing barriers of idea contribution, participants tend to generate more ideas in GSS brainstorming sessions than conventional face-to-face environment.

To sum up, the characteristics of parallel input and anonymity associated with electronic brainstorming in GSS help create a situation where hundreds of lines of text in a short period of time can be generated. Typically, for a brainstorming session of 10-20 participants, comments can easily reach one hundred and sometimes may even number in several hundred comments in less than an hour. The large amount of textual comments makes it extremely difficult for group participants to browse and consolidate into important topic categories.

2.4.1.2 Cognitive Demand

In addition to information overload, huge cognitive demand on group participants is clearly present during idea organization. As a convergent task, idea organization is more demanding cognitively than the divergent task of brainstorming. In brainstorming, group participants express their thoughts through their own words, which is more natural for individual participants. However, considerable cognitive demand is required in order to complete an idea organization task.

An important part of idea organization is to identify ideas conducive to solving the brainstorming question. In the beginning of an electronic meeting in GSS, a brainstorming question is presented to all group participants as a way to define the scope of meeting. Every comment submitted should be relevant to achieving the goal set out in the brainstorming question. However, anonymity enabled by the GSS brainstorming tools may sometimes lead to comments irrelevant to the brainstorming question. Therefore, group participants are required to read and understand each comment, identify relevant ones and filter out those irrelevant ones from the brainstorming data.

Redundant or similar ideas are to be merged into a single general topic. It is inevitable that many comments submitted in brainstorming data share the same underlying theme or topic. A major responsibility during idea organization is to consolidate these synonymous ideas into a common topic, thus making it easier for the subsequent stage to conduct rank and order activities. In addition, brainstorming comments may contain different levels of abstraction ranging from very general issues to very specific issues. Comments that are too general need to be discarded and those that are too specific need to be merged.

Group participants also need to establish shared meaning for the topics in order to make sure that they understand the topics in similar ways to avoid confusion. However, it is not always easy to establish shared meaning because semantics of terms used during discussion may differ depending on the perspectives of different individuals, and may also differ in different contexts even for the same individual. Sometimes, the same meaning may be expressed using different terms. For example, Furnas et al. found that to choose word for objects in five domains, two people chose the same term with less than 20% probability (Furnas, Landauer et al. 1987). Therefore, the problems with semantics may lead to a situation where group participants may have different understanding of brainstorming comments.

In addition, comments from brainstorming tasks are unprocessed raw ideas, which may contain many grammatical errors and misspellings. Therefore, group participants have to be extra careful in order to understand the comments. Furthermore, like other tasks in group sessions supported by GSS, the idea organization task often has time constraints to meet. Therefore, with an overwhelming amount of comments to process, group participants may often experience high pressure, which further adds to an already high cognitive demand.

To sum up, idea organization requires considerable attention and cognitive resources from group participants.

2.4.1.3 Drop in Satisfaction Level during Idea Organization

The adverse effect of information overload and cognitive demand in idea organization can be clearly observed by the sharp drop of satisfaction level for group participants at the beginning of idea organization. Figure 2.2 shows changes in participants' satisfaction level over the duration of four types of activities (Chen, Hsu et al. 1994). In this figure, the horizontal line describes a typical problem solving sequence whereas the vertical axis represents the satisfaction level of the group participants.

The typical group problem solving sequence conducted under GSS environment is depicted in the horizontal line as consisting of four stages; namely, idea generation, idea organization, prioritizing and policy development.

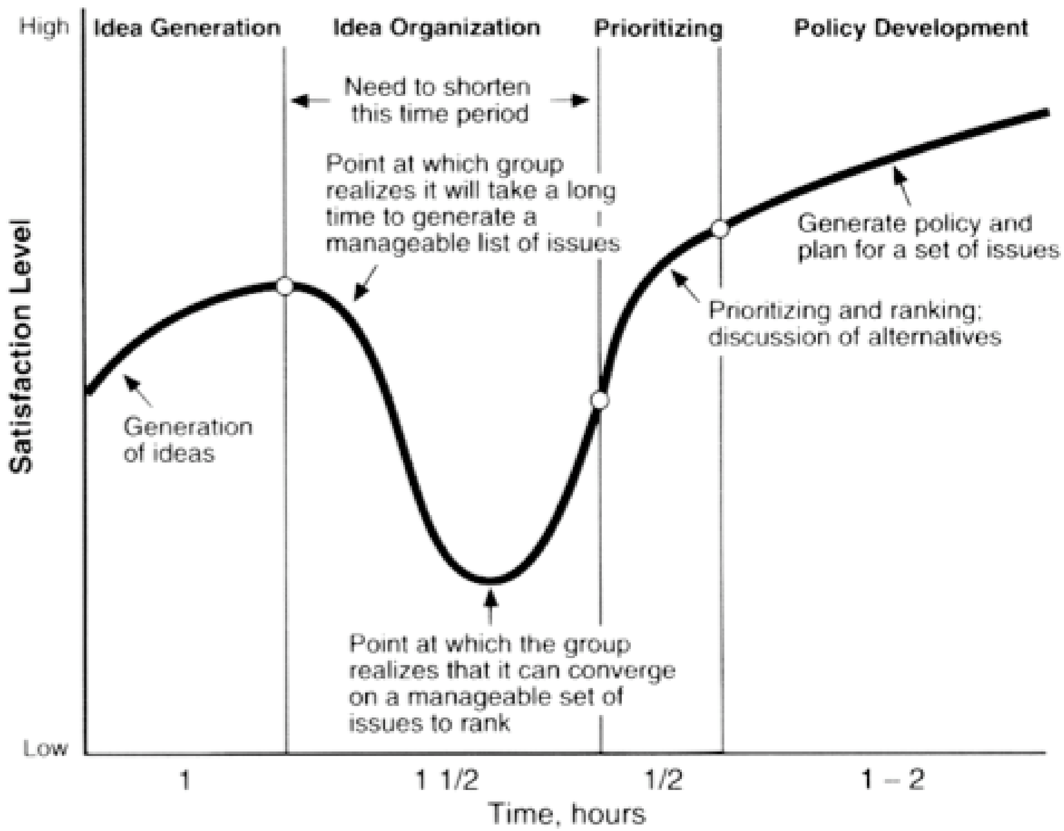

Figure 2.2 Satisfaction Levels over Time for the Collaboration Stages

The first stage in the figure is idea generation, which takes about an hour to complete. Although the actual time spent may vary from session to session, the relative duration compared with other stages in the figure may hold true in most sessions. It is observed that the satisfaction level increases over time during idea generation. That is, group participants feel increasingly satisfied because they feel that their efforts are well spent when more and more ideas are generated over time. The high output of brainstorming ideas is mainly due to process gains such as parallel input and anonymity provided by GSS.

In contrast, idea organization, the second activity, takes one and a half times longer than idea generation. This increase in time length is accompanied by a sharp drop in satisfaction level immediately after the completion of idea generation and just at the beginning of idea organization. Group participants realize that they have to read, understand and organize all brainstorming comments into topic categories. Faced with information overload and cognitive demand problems as explained previously, group participants have to take a long time working extremely hard to converge and organize ideas into key topics. Unlike the divergent task fully supported by various brainstorming tools in GSS, idea organization has very limited support provided by GSS, which leads to decreased satisfaction level among group participants who have to tackle information overload and cognitive demand problems on their own. The satisfaction level keeps dropping while group participants begin to come up with topic suggestions, consolidate duplicate topics, merge topics that are too specific and discard topics that are too general.

As a manageable list of important topics takes shape, the satisfaction level starts to increase over time.

The third stage in the figure is prioritization, which allows group participants to rank, order and vote the solutions to the brainstorming question based on the manageable list of topics obtained from idea organization. The figure shows that the prioritization activity takes half as long as the brainstorming activity and one third of the time required for idea organization activity. The satisfaction level increases over time during this stage because group participants could easily rank, order and vote their best solutions with adequate support for prioritization provided by various tools in GSS.

Policy development, the last stage in the figure, allows group participants to formulate a feasible policy based on the evaluations of possible solutions obtained from prioritization stage. Policy development ranges from one to two hours in duration depending on the context of the problem. Satisfaction level increases over time and reaches the highest point during this stage.

In conclusion, the sharp drop in satisfaction level as depicted in figure 2.2 draws attention to the inadequate cognitive support in idea organization stage. In contrast to the rise in satisfaction level in other stages, idea organization has seen a sharp decline in satisfaction level as group participants struggle to tackle an overwhelming amount of textual data in the absence of cognitive support in GSS. Though much of previous research has been focused on divergent activities, there has been limited research conducted on idea organization (Briggs, Vreede et al. 2003). As seen from the idea organization stage in figure 2.2, one crucial way to reverse the drop in satisfaction level is to have a manageable list of topics during idea organization to show to group participants rather than have them face an overwhelming amount of comments. Importantly, figure 2.2 also demonstrates the need to shorten the time period during idea organization.

2.4.2 Facilitation Support to Help Idea Organization

With the cognitive complexity of idea organization activity and the information overload problem, several group facilitation techniques can be employed to facilitate idea organization during collaboration. A variety of ways of facilitating idea organization tasks have been introduced in previous research (Orwig, Chen et al. 1997).

Given that much of the problem associated with idea organization is due to the large volume of brainstorming comments generated, one possible technique is for the facilitator to reduce the production of brainstorming comments to a manageable amount. For example, facilitators may try to limit the use of brainstorming tool in GSS. In order to do that, the facilitator may set a manageable threshold for the number of brainstorming ideas allowed in each session depending on the nature of the problem and the number of group participants. When the threshold is reached, the facilitator may stop the brainstorming function manually. Alternatively, depending on the nature of the problem, the facilitator may choose to avoid using the brainstorming function completely by opting for more structured tools such as topic commenter or group outliner in GroupSystems. However, previous research has shown that parallel input and anonymity features of electronic brainstorming can lead to increased creativity among group participants. Therefore, the drawback for not utilizing parallel input and anonymity of electronic brainstorming to the fullest extent may be the decrease in the number of creative ideas generated.

Facilitation support can also be used to reduce the information overload. One possibility is to ask group participants themselves to suggest which topics to be included. Group participants make the suggestions based on their contributions and their understanding of others' contributions. In essence, this method is easy to follow and reduces information overload by dividing the workload among group participants. However, this approach may not be able to involve all participants in the group. Some participants may be very active to suggest topics while others may not. Therefore, the topic list generated may be biased and not representative of all ideas generated. In addition, participants may not read all comments so their understanding of comments is limited.

Alternatively, the facilitator could divide up the work evenly among all group participants by assigning each group participant an equal number of comments. This divide and conquer approach can make the best use of human resources without relying on any particular person. Each participant can then generate a topic list for his or her own set of comments. Then topic lists may be collected from all participants and presented to all participants for a final round of idea organization. However, topic lists collected from all participants may still be very large if each participant generates around 5 topics. For a 20-participant group, an average of 5 topics each could lead to 100 topics for the final round, which are still a large amount of information to process. Furthermore, the division of work may inevitably lead to more chances of redundancies among topics, which may take group participants more time to merge redundant ideas.

Similarly, for idea organization, the facilitator may ask each group participant to propose one topic, which may add up to a list of topics. Then it is up to group participants to evaluate each of these topics to determine the appropriateness with respect to solving the problem at hand. However, since each participant is to suggest one topic, it is highly possible that some important topics might be overlooked while there might be more chances of redundant topics because it is prone for group participants to suggest the obvious topics.

Another possible approach is for the facilitator to devise topics while group participants are busy with brainstorming activity. Time could be saved this way because the facilitator is using the time during brainstorming to come up with a topic list. However, this approach leaves the work of topic generation solely to the facilitator, who could otherwise perform other important facilitation tasks such as monitoring the brainstorming activity. Furthermore, since comments are submitted very fast via electronic brainstorming, it is very difficult for the facilitator to catch up with all the ideas generated. If the facilitator is not expert on the domain of the discussion, he or she may have to spend extra time digesting and understanding the ideas to come up with a topic list and the quality of the list generated may not be limited.

Despite the usefulness of the above-mentioned various facilitation support to improve idea organization, it is known that experienced human facilitators may not be always available (Nunamaker, Briggs et al. 1996; Niederman 1999; Wong 2003). In some cases, human facilitators may be too involved in participation, which may lead to failures (De Vreede, Davison et al. 2003). Therefore, there should be other ways that can be used to mitigate the adverse effect of information overload and cognitive demand, which lead us to automatic idea organization.

2.4.3 Need for Automatic Idea Organization

As explained in the previous section, GSS provide various software tools that can be employed to support individual steps in a collaboration process. For example, GroupSystems contain tools such as electronic brainstorming, topic commenter and group outliner to support divergent activities. For idea organization, tools like idea organizer, issue analyzer and group writer can be used. While divergent tools prove very useful in divergent activities, GSS do not have effective tools available to help group participants reduce information overload and cognitive demand associated with idea organization in GSS. Current tools associated with idea organization rely exclusively on participants to manually classify all comments into topic categories with little or no support to help participants reduce the high cognitive demand.

Although facilitation techniques can be used to reduce the adverse effect of information overload and cognitive demand problems, it should be noted that these facilitation techniques themselves are insufficient. Therefore, automatic idea organization is needed to overcome the following problems associated with idea organization in GSS. As idea organization consists of two components: topic generation and placement of relevant comments into their respective topics. Automatic idea organization must achieve automation in both components; that is, on one hand, the system generates topic categories automatically; on the other hand, the system automatically places relevant comments into their respective topic categories. There are various benefits that can be achieved with automatic idea organization as discussed in the following.

2.4.3.1 Reducing Information Overload

Information overload problem is mainly caused by the effective brainstorming tool in GSS. The brainstorming tool in GSS encourages group participants to generate as many ideas as possible via methods such as parallel contribution and anonymity. Therefore, during idea organization, group participants have to browse through a large volume of textual comments data, which are quite overwhelming for them.

Information overload problem can be completely solved with automatic idea organization because group participants do not have to browse through the long list of comments, which are directly fed into the automatic idea organization system and results can be produced very quickly.

2.4.3.2 Reducing Cognitive Demand

Various tasks that require constant attention and cognitive power need to be performed in idea organization. For example, redundant or similar ideas are to be merged into a single general topic; topics that are too general need to be discarded and those that are too specific need to be merged. All of these efforts prove cognitively very demanding for group participants.

Cognitive demand can be reduced, if not eliminated, with automatic idea organization because group participants do not have to use their attention resources in processing cognitive tasks. Instead, the automatic idea organization tool takes as input all the brainstorming comments and intelligently processes the data and produces the output of idea organization automatically. With the computer-generated output, group participants can then further refine the results, a task which demands much less cognitively of participants than the task of completing idea organization from scratch.

2.4.3.3 Shortening Idea Organization Time

There is a trade-off existing between the value of information and the time spent searching it (Pirolli and Card 1995). In the case of idea organization, the trade-off means that group participants have to spend a lot of time in order to achieve results with good quality. Figure 2.2 shows that the time spent for idea organization under GSS environment takes one and a half times longer than brainstorming time. Each of the two components of idea organization takes a long time to finish. In addition, it is often impossible to manually explore all relevant textual comments exhaustively during idea organization.

With automatic idea organization, the computer system can process large quantities of textual data much faster than humans can read through. As a result, time spent on idea organization will be in minutes or even in seconds, which are significantly shorter than the average one and a half hour as shown in Figure 2.2.

2.4.3.4 Providing Context to Topics Instantly

During a GSS-enable group session, group participants manually place relevant comments into their respective topic categories after the topic list is finalized, thus providing meaning and context to the topics generated. However, when the topics are generated, the context or meaning of the topics are not immediately clear because they are mostly captured in participants' own minds as they come up with the topics. Therefore, there may be discrepancy between the topics and the context provided after the topics are generated.

Automatic idea organization can instantaneously place comments into their respective topics immediately after topics are generated. What group participants will see are not only a topic list but also a series of comments associated with each topic. With context provided to each generated topic, automatic idea organization helps clarify the meaning of topics generated and makes it easier for group participants to understand the generated topics.

2.4.3.5 Eliminating Bias

In order to identify appropriate topics, group participants typically browse brainstorming comments data in a serial manner. Therefore, the topics generated may be biased towards comments near the beginning of the comments list. In addition, the contribution of topics may not be evenly spread among group participants with some participants very active and others not. Thus, the topic list may be biased towards those active participants.

With automatic idea organization, all comments are processed in the same manner without bias as seen in manual idea organization.

To sum up, the problems surrounding idea organization such as information overload and cognitive demand make it crucial to automatically classify a large amount of raw brainstorming comments into meaningful topic categories and place relevant comments into their respective topic categories. Automatic idea organization helps relieve group participants from the labor-intensive job of manually processing comments.

In fact, automatic idea organization can be mapped into groupware grid to show its usefulness. As explained in previous sections, groupware grid is used to show each cell with examples of the kind of support available for a particular process at a particular level of work. Table 2.7 shows that an automatic idea organization tool can be considered to be useful in improving information access during a concerted effort (Nunamaker, Jr.; Briggs et al. 1996). That is, an automatic idea organization tool is necessary if we want to improve information access at the concerted level so as to offset the huge costs of manually acquiring, storing, processing, and retrieving information. In addition, given that a person's attention resources are limited, time and cognitive power saved by an automatic idea organization can also be used towards future stages of GSS to achieve more productivity improvements.

	Productivity Processes		
	Communication	**Deliberation**	**Info Access**
Concerted Level	Anonymity, Parallel Contribution	Structured and Focused Processes	Session Transcripts ***Automatic Topic Classification***

Table 2.7 Classification of Benefits from Group Support Systems

Therefore, the objective of this research is to develop an automatic idea organization tool and evaluate its performance. In the next chapter, a literature review is performed to determine the level of understanding in terms of automatic idea organization in previous research.

CHAPTER 3: RELATED LITERATURE

Brainstorming comments are essentially textual documents. Each individual comment can be considered as a separate document. One important characteristic of idea organization is that there is no prior information on what the topic categories are. Therefore, automatic idea organization can be regarded as an unsupervised document clustering problem and thus calls for adopting document clustering techniques to automate idea organization with textual input from brainstorming comments. Specifically, the automation process requires the completion of the two important components of idea organization: automatic generation of topic categories with labels showing the content and automatic placement of relevant comments into their respective topic categories.

There have been several limited studies on automatic idea organization using neural networks like Hopfield network and Kohonen's network. These techniques to automatically generate topic categories have been drawn from a larger body of literature closely related to document clustering. Therefore, in order to apply appropriate technique to automate idea organization from electronic brainstorming data, it is necessary to first conduct a review of literature on document clustering. In addition, before any clustering techniques can be applied, a proper way of preprocessing, feature selection and document representation has to be determined.

This chapter discusses related literature on preprocessing, feature selection and document representation, followed by related literature on document clustering. Finally, several related studies on automatic idea organization are discussed.

3.1 Preprocessing

The objective of preprocessing is to convert the raw textual documents into a form suitable for further document clustering algorithms. A number of text processing techniques have to be taken to automatically preprocess textual documents as follows:

- Word Identification: Word identification is used to extract terms from the text of a document. Terms represent the semantic content of a document. Thus, the extracted terms from all documents can then be used to index the entire document collection. The simplest way to extract terms is to make the algorithm recognize white spaces as term delimiters and certain punctuation marks such as "?" as sentence delimiters. More sophisticated techniques are often needed to deal with special characters such as digits and punctuation marks that do not contribute to the meaning of textual documents (Baeza-Yates and Ribeiro-Neto 1999).
- Stop Words Removal: Stop words, also known as function words, are terms that are too general and have no identifying meaning. Examples of stop words are "the", "this", "I", etc. Since these stop words often occur frequently in most documents, they have very limited discrimination power and thus are not very useful in textual processing. Therefore, it is often necessary to have stop words removed. The application of stop words removal can often lead to a reduction of 40% or more in the size of an inverted index (Baeza-Yates and Ribeiro-Neto 1999).
- Stemming: A stem is a portion of a word that remains after removing its affixes including suffixes and prefixes. When stemming is performed, different

grammatical formats of a word can be reduced to a single base form. For example, ‘listened’ and ‘listening’ can be reduced to the word ‘listen’. There are a number of stemming algorithms available for English language, of which Porter stemmer is the most commonly used (Porter. 1980).

3.2 Feature Selection and Document Representation

With preprocessing completed, the next step is to determine what features to use in the final representation of documents. There are two major reasons for feature selection. One reason is to limit the number of features so as to increase time efficiency of the algorithm. The other reason is to help remove noise from the text so that a higher accuracy of result can be achieved. A variety of feature selection methods have been proposed in the literature that includes simple frequency threshold algorithm and more complex information theoretic algorithms (Tao Liu 2003).

Document representation describes the number of clusters, the number of available documents, and the number, type and scale of the features available to the clustering algorithm (Jain, Murty et al. 1999). The most commonly used method for document representation is the vector space model (Salton 1989). In this model, each document is represented by a multidimensional feature vector with length equal to the number of unique document attributes in the entire document collection. Each component of the vector refers to a particular term feature associated with that document. A weight is associated with each component of the feature vector, which indicates the importance of the particular attribute in representing the document. That is, the weight represents the degree of the relationship between the term feature and the document that carries the term.

Past research has proposed a variety of ways to measure the degree of relationship between a term feature and a document, often called term weighting. The simplest method is binary weighting, in which the weight can take the form of a binary representation consisting of either 0 or 1 depending on whether the term attribute occurs in the document or not. A more advanced method is term frequency (tf), which indicates the frequency of occurrence of an attribute in the document. Term frequency is more accurate than binary weighting by showing the strength of the relationship between the term and the document. Another improved term weighting method is the term frequency inverse document frequency (tf-idf), which is to give each term a weight based on the frequency of occurrence of the attribute in the entire document collection rather than just one document. The rationale is to balance the local and global term occurrences in the documents by deemphasizing terms that appear frequently in many documents because they may have limited discrimination power.

3.3 Related Research on Document Clustering

A cluster can be defined as "a set of entities which are alike and entities from different clusters are not alike" (Everitt 1974). As one major kind of data mining techniques, data clustering is the unsupervised partitioning of a data set into clusters of similar objects. The result of clustering is that the data in each cluster share some common features based on some defined distance measures. One essential characteristic of generated clusters is that each cluster has objects that are similar to each other and very different from objects in other clusters (Rasmussen 1992).

The idea of clustering originally comes from statistics and only applies to numerical data. Increasingly, other types of data such as text and multimedia are used in data clustering. The objective of document clustering is to summarize and organize text documents by aggregating similar documents into clusters optionally with a high-level summary of each cluster. Good document clustering produces clusters where documents within one cluster are more similar to each other than documents in other clusters. That is, every cluster has one topic shared by all documents in that cluster and topics in different clusters are different from each other. With automatic document clustering, there is no need to browse through the entire document collection to understand each document.

Document clustering research is very relevant to automatic idea organization. Similar to document clustering, automatic idea organization deals with text-based brainstorming comments. Each comment generated in the brainstorming session can be thus treated as a document. Also similar to document clustering, idea organization is aimed to organize comments into meaningful topic categories with thematically similar documents grouped under one topic category. Therefore, it is the document clustering algorithm that should be regarded as the foundation upon which our automatic idea organization system is built.

It should be noted that there is a difference between clustering and classification from the perspective of data mining. Classification is a supervised process where a collection of pre-classified training data is given in advance and the task is to classify a new data item into its appropriate class. The training data provided is typically used to learn the description of classes. The knowledge learned through training is applied towards classifying new data. In contrast, clustering is an unsupervised process where a collection of unlabeled data is partitioned into distinct clusters. Unlike in supervised classification, there is no prior information about the data in unsupervised clustering.

Past research on document clustering has focused on developing various algorithms to cluster traditional documents. Increasingly, these algorithms have been applied to web documents clustering. In the following, we discuss several approaches to the problem of document clustering.

3.3.1 Classic Document Clustering Methods

According to clustering methods used, document clustering approaches can be roughly classified into the following categories: hierarchical, partitional, graph-based, and neural network-based. One important characteristic of all these methods is that all these methods obtain clusters of documents first based on numerical comparisons and then often generate cluster labels by selecting the most significant terms with each cluster. All these algorithms require vector space model and some forms of distance measures to be used to calculate similarity between objects to be clustered. Past research has identified a variety of distance measures (Anderberg 1973; Jain and Dubes 1988). Some of the most commonly used measures include cosine coefficient, jaccard coefficient and dice coefficient, of which cosine coefficient is the most frequently used measure that computes the cosine of the angle between the two feature vectors. More on the similarity measures can be found in (Willett 1988; Strehl 2000).

3.3.1.1 Hierarchical Clustering

Hierarchical clustering generates a nested series of partitions based on a criterion for merging or splitting clusters using similarity measures. The similarity between each pair of documents is stored in a similarity matrix. Clusters can be either merged using agglomerative methods or divided using divisive methods. Hierarchical clustering produces a tree-like structure called dendrogram with a top cluster including all the documents in the collection and many bottom clusters with one document in each of them. Clusters are derived by choosing an appropriate level of the dendrogram. The dendrogram is particularly useful when retrieval is performed on a clustered set of documents because dendrogram provides the paths through which the retrieval can be conducted (Rasmussen 1992).

Agglomerative hierarchical clustering (AHC) algorithms are the most common methods used for hierarchical clustering. In order to compute the similarity between two clusters, the hierarchical agglomerative clustering algorithms may use the following methods:

- Single link: single link algorithms compute the similarity between two clusters as the similarity between the two most similar documents with one in each cluster. Example of single link algorithms include Rijsbergen's algorithm (Van Rijsbergen 1979), SLINK (Sibson 1973), Minimal Spanning Tree (Rasmussen 1992) and Voorhees's algorithm (Voorhees 1986).

- Complete link: complete link algorithms compute the similarity between two clusters as the similarity between the least similar documents with one in each cluster. Examples of complete link algorithms include CLINK algorithm (Defays 1977) and Voorhees's algorithm (Voorhees 1986).
- Group Average: group average algorithms compute similarity based on all similarities between documents, thus avoiding the pitfalls of the single-link and complete-link criteria, which equate cluster similarity with the similarity of a single pair of documents. Examples of group average algorithms include Voohee's group average algorithm (Voorhees 1986) and Steinbach et al.'s algorithm (Steinbach, G. Karypis et al. 2000).
- Centroid/Median Methods: As each cluster is generated, it is represented by the group centroid or median. During clustering, the two clusters that have the most similar mean centroid or media are merged. The median method is different from the centroid method in that the median method is not weighted proportionately to the size of the cluster.
- Ward's Method: this method is different from all the methods mentioned above because it uses an analysis of variance approach to compute the distance between clusters. For any two clusters that can be formed at each step, this method attempts to minimize the increase in the total within-group error sum of squares based on the distance between the cluster centroids. Examples are the reciprocal-nearest neighbor algorithm (Murtagh 1983) and its modified version for document clustering (El-Hamdouchi and Willett 1986).

3.3.1.2 Partitional Clustering

Partitional clustering, also known as non-hierarchical clustering, directly partitions the data into a predefined number of disjoint clusters. A major characteristic of partitional clustering is the use of a global criterion function. The objective of the criterion function is to optimize different aspects of intra-cluster similarity, inter-cluster dissimilarity and their combinations. Typically the global criteria involve minimizing some measure of dissimilarity in the samples within each cluster, while maximizing the dissimilarity of different clusters. Partitional clustering algorithms consist of iterative or reallocation methods and single pass methods.

One of the most commonly used partitional clustering algorithms is k-means, which is an iterative clustering algorithm in which clusters are built around k central points called centroids. The algorithm starts with selecting k centroids and computing the cosine similarity measure between each document in the collection and the centroid. Each document is then assigned to the cluster with the nearest centroid. With all documents assigned to clusters, the algorithm recalculates the new cluster centroid and the process continues iteratively until some criterion is met. Spherical k-means is a variation of the k-means algorithm in which all centroids are normalized to have unit length. Examples of different variations of the k-means include ISODATA (Jain, Murty et al. 1999) and bisecting k-means (Steinbach, G. Karypis et al. 2000). Buckshot-fractionation algorithm is another partitional clustering algorithm used in the Scatter/Gather system to cluster web search results by D. Cutting (Cutting, Karger et al. 1992).The objective of buckshot-fractionation algorithm is to find the initial cluster centers and then obtain the clusters by assigning each document to its nearest center. The single pass method proposed by Rasmussen is another example of partitional clustering algorithm in which one pass of data forms the clusters without iteration (Rasmussen 1992).

These partitional clustering algorithms are simple to implement and very low in terms of computational complexity. However, the result of clustering depends on many parameters such as the predefined number of clusters, the selection of initial cluster centroids.

3.3.1.3 Graph-based Clustering

When graph is used to cluster documents, nodes are used to represent documents and edges to represent the relationship between them. Each edge has a weight indicating the strength between two documents. As the main method used in graph-based cluster algorithms, graph partitioning forms clusters by removing edges from the graph in order to minimize the weights of the edges that are cut. By cutting edges that have the minimum sum of weights, the algorithm minimizes the similarity between documents in different clusters. The result of cutting edges is to make the weights of the edges in the same cluster greater than the weights of the edges in different clusters, and thus related documents are placed in the same clusters.

There are a variety of graph-based ways to cluster documents that may vary according to the way how the graph is constructed and the graph partitioning algorithms. For example, Chameleon adopted k-nearest neighbor approach for a graph-based representation of documents (Karypis, Han et al. 1999). With the k-nearest neighbor approach, an edge is constructed between two documents if either of the documents is among the k most similar documents of the other document. Then a graph partitioning algorithm, hMETIS, is applied to find the clusters. Finally, a hierarchical agglomerative clustering algorithm and a dynamic model are used to further cluster the results to determine the similarity between two clusters. Given that a hierarchical agglomerative cluster is used in addition to the graph-based method, Chameleon's algorithm is a hybrid document clustering algorithm. Another example of graph-based document clustering is Association Rule Hypergraph Partitioning (ARHP) (Boley, Gini et al. 1999). ARHP algorithm is based on hypergraph, which is a variation of a graph where each hyperedge can connect more than two nodes. As a result, a frequent item set can be formed with hyperedges connecting a set of nodes together, representing the relationship between two or more documents. Typically, a frequent item set includes documents that share common terms. An association rule discovery algorithm is used to determine the frequency item set in the document collection and to give weight to the hyperedge. Then a hypergraph partitioning algorithm is applied to partition the hypergraph to get the clusters. Other graph-based document clustering approaches include Dhillon's algorithm that applies iterative bipartite graph partitioning (Dhillon 2001).

There are both advantages and disadvantages for graph-based methods. The advantages are that these approaches can represent the structure of the data and work effectively in high dimensional space. However, the graph constructed takes a lot of memory to store.

3.3.1.4 Neural Network Clustering

Neural networks are inspired by nervous systems consisting of closely interconnected sets of neurons. One neuron may be limited in terms of its simple structure, but dense networks of interconnected neurons could perform complex tasks such as pattern recognition and classification. Therefore, neural networks are created to imitate the type of nonlinear learning that takes place in the networks of neurons found in nature.

One of the major features of neural networks is that a neural network can "learn" to the extent that its performance can be adjusted to fit a known set of data characteristics. In unsupervised learning like document clustering, inputs data are fed into the input nodes of a neural network, which then uses its learning rule to adjust the weights so that inputs data are clustered based on their statistical properties (Rumelbart, Hinton et al. 1986; Dalton and Deshmane 1991).

The Kohonen's network represents a type of self-organizing maps (SOM) and is a widely used unsupervised neural network model (Kohonen 1989). The goal of self-organizing maps is to convert a complex high-dimensional input signal into a simpler low-dimensional map (Haykin 1990). Therefore, SOM can be used to uncover underlying hidden patterns among documents and thus are suitable for cluster analysis. Previously SOM has been used for classifying textual documents for information retrieval (Lin, Soergei et al. 1991). Another example of neural network clustering is the hierarchical feature map, which uses a hierarchical organization of more than one self-organizing maps for clustering (Merkl 1998).

Specifically, SOM divides the output nodes into clusters of nodes such that nodes close to each other are more similar to each other than to nodes that are farther away. Unlike most neural networks, SOM has no hidden layer, but one input layer and one output layer in the shape of a two-dimensional grid (Caudill 1993; Hiotis 1993). The input layer consists of input nodes representing a collection of documents whereas the output layer has output nodes corresponding to clusters. Each node of the output layer also has the same number of features as the input nodes. The network is fully connected in the sense that every node in the input layer is connected to every node in the output layer. The data from the input layer is passed directly to the output layer. Each connection between input and output nodes has a weight which is initialized randomly to a value between zero and one. Adjusting these weights is very important for the learning process. Each node on the output layer is also associated with a 'neighborhood', which consists of neurons surrounding it. Self-organizing maps are based on competitive learning where the output nodes compete among themselves to be the winning node so as to be activated by a particular input node. The result of SOM is a 2-dimenstional space where the input documents data are organized into clusters of similarity.

3.3.2 Phrase-based Methods

The above-mentioned classic document clustering methods put main focus on generating clusters of documents based on comparison of numerical measures. These methods often describe each generated cluster of documents with its most frequent terms; for example, the two most frequent terms from each cluster of documents may be used as labels for that cluster. Instead of relying on frequencies of individual terms, phrase-based methods can describe clusters with phrases that preserve the order of words in documents. These phrases are often more intuitive and descriptive of the cluster contents than a combination of single terms. Phrase-based methods assume that the topic category of a document is closely correlated with its most frequent phrases. Therefore, finding clusters of documents can be converted to finding groups of documents that share a high percentage of frequent phrases, which can then be used to describe the cluster

One prominent example of phrase-based methods is suffix tree clustering, which was originally developed to cluster web search results (Zamir and Etzioni 1998). Specifically, the algorithm first constructs a tree structure which represents shared suffixes between all documents. These shared suffixes are used to identify base clusters of documents, which are then combined into final clusters. Another study done by Hammouda and Kamel proposed document index graph to encode word order information and define similarity based on matches in word order (Hammouda and Kamel. 2004). The algorithm represents each word as a vertex in a directed multigraph. Each vertex keeps track of a table of documents in which the word is present and information on how a sentence forms along a sequence of edges. Phrases are formed by following the directed edges in the graph. In addition to phrases used to capture word order information, a weighted average method is adopted in both studies to combine term frequency and word order similarities. The results of both studies showed that word order information along with word frequency can improve clustering accuracy. However, since both studies require the construction of trees or graphs to capture word order information in phrases, they are more expensive than classic document clustering methods in terms of efficiency.

3.3.3 Dimension Reduction Methods

Various kinds of matrix factorization techniques are often called dimension reduction methods, which have been used to reduce the dimension of the term-document vectors in information retrieval systems. These matrix factorization techniques such as principal component analysis (PCA), singular value decomposition (SVD), and non-negative matrix factorization (NMF) have proven efficient in information retrieval tasks because they can reduce the number of dimensions used to represent documents by multiple orders of magnitude with manageable error.

One of the most important dimension reduction methods used in document clustering is singular value decomposition, which is the mathematical logic underlying Latent Semantic Indexing (LSI) introduced by G. Golub and W. Kahan to calculate the singular values, pseudo-inverse and rank of a matrix (Golub and Kahan 1965; Berry, Dumais et al. 1995). Singular value decomposition is performed on the original matrix A by breaking it up into three components in the form of the following: $A \approx U_r \times S_r \times V_r^T$, where U and V are orthogonal matrices that are referred to as the left and right singular vectors of A respectively. S is a diagonal matrix that contains the singular values of matrix A. Subscript r represents the rank of A. With the singular values sorted in decreasing order, the truncated SVD can be used to project the original matrix A onto a k-dimensional space in the form of $A_k = U_k \times S_k \times V_k^T$, which can be considered as the rank k approximation of A. In the new approximated A_k, S_k is a diagonal matrix that contains the k singular values along the diagonal line with higher value of k representing the closer approximation of A_k to the original matrix A.

More recently, dimension reduction methods such as SVD have been not only used to reduce dimensions in information retrieval, but also utilized directly as clustering methods on their own. These methods use decomposed orthogonal vectors to create clusters and use terms or phrases to generate cluster labels. For example, the Semantic Online Hierarchical Clustering (SHOC) is an algorithm that used singular value decomposition to generate clusters and use phrases to as labels for generated clusters (Zhang and Dong. 2004). Similarly, Lingo is a description-first approach that used singular value decomposition algorithms to cluster web search results (Osinski, Stefanowski et al. 2004).

3.4 Related Research on Automated Idea Organization

There has been limited research on automating idea organization from brainstorming comments in GSS (Chen, Hsu et al. 1994; Chen, Schuffels et al. 1996; Orwig, Chen et al. 1997). These studies have applied neural network algorithms including Hopfield network and Kohonen's self organizing maps described earlier.

3.4.1 Hopfield Network

A Hopfield network algorithm has been used for topic classification of electronic brainstorming comments (Chen, Hsu et al. 1994). The focus of Hopfield network is to cluster topics rather than documents. The clusters of topics serve as textual descriptors for the underlying topic categories in the comments.

The algorithm consists of three stages: automatic indexing of the electronic brainstorming comments, topic space generation and classification using the Hopfield neural network. The objective of topic space generation is to construct a matrix of term descriptors with their weighted relationship. Then the Hopfield network activation procedure is used to find clusters of relevant descriptions of the topic space through their weighted connections. During the process, each term in the topic space matrix is regarded as a neuron and the asymmetric weight between any two terms is treated as the unidirectional and weighted link between neurons. An information loss analysis is performed to reduce the number of terms used in Hopfield network analysis. The Hopfield network algorithm then uses each remaining term as an input pattern, activates its neighbors, and combines weights from all associated neighbors. The algorithm continues this process repeatedly until the output pattern converges. The result produces terms that are semantically relevant to the input term. This process is repeated one by one for all the terms based on their decreasing occurrence frequencies. When all terms are processed through the Hopfield network, related clusters of topic are identified for the electronic brainstorming comments. Experiments compared the output of their algorithm with human experts and novices and found that the Hopfield algorithm performed as well as the novices but two human experts outperformed the novices and Hopfield classifier significantly.

3.4.2 Kohonen's Self-organizing Map

Kohonen's SOM has also been applied to classify important topics for electronic brainstorming comments (Chen, Schuffels et al. 1996; Orwig, Chen et al. 1997). During the process, automatic indexing is first performed to represent each comment. The result is transformed into a form compatible with the Kohonen input. Each vector in the transformed input file represents a comment with each vector position representing a term. The value of each vector value is '1' if the term that it represents is contained within the comment and '0' otherwise. The Kohonen algorithm consists of four steps: map initialization, map training, quantization error computation and map visualization.

Each comment is treated as an input vector with its most frequently occurring terms as coordinates. A coordinate is set to 1 if the comment has the corresponding term and 0 if there is no such term. A two dimensional map is used as the output. Each node in the input layer is equivalent to a comment. The connections between the input layer and the output layer represent the weights. The Kohonen algorithm first initializes the output layer. The weights connecting the inputs to the output layer are initialized to small random values while neighborhood type and size are also initialized. Then each document in the input file is fed into the network, followed by the computation of the distance between this input vector and each output node. The node on the output grid with the minimum difference is the winning node, which is used as the center of the neighborhood. With the winning node identified, the weights for the minimum node and all of the nodes in its defined neighborhood are then adjusted. After the updates, the input vector becomes more similar to the neighborhood of the winning node. After repeated input of all documents to the network with each document presented at least five times, a winning term is assigned to each output node by choosing the one corresponding to the largest weight. Neighboring nodes that contain the same winning terms are merged to form a topic region. Finally, each document is submitted as an input to the trained network again in order to be assigned to a particular topic region in the map. The result of the self-organizing map algorithm is a 2-dimensional map with relevant comments assigned to regions of important topics. In addition, the map presents the topic regions in such a way that related topics are co-located on the output map.

CHAPTER 4: REMAINING PROBLEMS AND RESEARCH METHODOLOGY

Given that automatic idea organization is needed for GSS, it is important to find out an appropriate technique to automate idea organization with quality comparable to human group participants. This chapter describes remaining problems and the research methodology used in this research.

4.1 Remaining Problems

Literature review has shown that there have been very limited studies on automatic idea organization. The only available studies have exclusively focused on applying neural networks including Hopfield network and Kohonen's SOM. Hopfield net (Chen, Hsu et al. 1994; Chen, Schuffels et al. 1996; Orwig, Chen et al. 1997). However, one of the major drawbacks with SOM and Hopfield network algorithms in processing brainstorming comments is that they tend to be very slow due to high computational complexity. When processing brainstorming comments data, the training data often have to be presented in several passes and the network can only be gradually improved after multiple adjustments to its weights. That is, these neural network algorithms require frequent iteration, which may require relatively long time. For example, when SOM was applied to categorizing the output from brainstorming comments, a coffee break of 20-30 minutes has to be taken for the slow SOM algorithm to produce a result (Orwig 1995). To reduce time, information loss analysis are performed to reduce the number of terms used in iteration, which result in terms with frequencies over four being excluded from the analysis. Therefore, the quality of the results may be potentially reduced with the exclusion of useful terms from the final analysis. In some cases, it is possible that the number of remaining terms after information loss filtering is still large and thus makes iterations time-consuming and resource-intensive. In addition, these studies focused primarily on generating and evaluating topic categories. There has been a relative lack of empirical analysis on evaluating the quality of comment placement, which measures whether all the comments under a topic category is appropriate to that category and whether there are other relevant comments not listed under the assigned category.

All these studies on automatic idea organization using neural network algorithm were conducted in mid 1990s. With the recent advances in document clustering, especially in web search results clustering, dimension reduction algorithms such as singular value decomposition technique have proven very useful in clustering search results (Osinski, Stefanowski et al. 2004; Zhang and Dong. 2004).

However, despite the success of singular value decomposition in clustering web search results, it is net yet known whether it can be applied to automatic idea organization. What is also important is that previous research on clustering with singular value decomposition adopts a merge-then-cluster approach, which compares a known cluster structure to the results of clustering the same set of documents algorithmically. This kind of evaluation focuses only on clustering quality, which measures whether the documents within a cluster consists only of documents from a single original category. Instead, automatic idea organization in GSS is concerned with two different objectives. The first and most important objective is to find out whether the topic categories are appropriate topics representative of a majority of brainstorming comments. The second objective is to find out whether all the comments under a topic category are appropriate to that category and whether there are other relevant comments not listed under the assigned category. These two objectives are different from the web results clustering which is only concerned about clustering quality. The output of neural network algorithms such as Hopfield network and SOM does not always produce desired results when compared with human subjects. As a result, it is not known how SVD approach compared with human subjects in conducting idea organization. Since neural network algorithms have been known to be very time consuming, it is not yet known how efficient SVD can be when applied to automating idea organization.

In conclusion, given the limitations of neural networks approach and the potential of SVD approach in automating generation of topic categories and placement of comments into their respective categories in a manner that might approach human performance, we decided to apply the SVD approach to automating idea organization and evaluate the results of automated approach with those generated manually by human subjects.

4.2 Research Question

The research question is "Can a singular value decomposition approach be used to automate idea organization from electronic brainstorming comments in a manner that equals to or exceeds the performance of humans?" Since automatic idea organization consists of two components: generation of topic categories and placement of comments into their respective topic categories, we divide the research questions into two sub questions:

1. Can a singular value decomposition approach be used to automate generation of topic categories in a manner that equals to or exceeds the performance of humans?
2. Can a singular value decomposition approach be used to automate placement of relevant comments into their respective topic categories in a manner that equals to or exceeds the performance of humans?

4.3 Potential Issues with Automatic Idea Organization

Previous literature on document clustering was primarily focused on application areas such as documents or web search results. However, GSS is a unique application domain, which has several unique characteristics that need to be taken into consideration when we use singular value decomposition techniques to automate idea organization (Chen, Hsu et al. 1994).

4.3.1 Noisy Input

Unlike textual data in web pages, news stories or other published sources used in information retrieval research, brainstorming comments in GSS sessions are more prone to errors. Despite the availability of word editors in most GSS, group participants may tend to contribute their ideas as fast as they type them rather than proofread them before submission. One possible reason may be that since output mechanisms in GSS resemble those commonly seen in chat rooms and instant messengers, group participants may follow the styles they adopt in those applications and thus rely less upon spell checkers to reduce mistakes in their idea contributions. Time constraints in brainstorming may also force participants to type quickly without paying much attention to spelling and grammars. As a result, more typos, incomplete sentences and grammatical errors may be made. All of these features related to textual data may pose difficulties in designing a tool for automatic idea organization.

4.3.2 Domain Independent

GSS can host any kind of meetings no matter what domain the meeting is involved in. Meetings hosted in GSS can cover a variety of subject issues, ranging from a distributed software requirements meeting for software engineers to a problem-solving session for school administrators. Given the domain-independence of GSS, terms and terminologies used in different meeting sessions may vary greatly. Therefore, this characteristic precludes the use of domain-specific natural language processing techniques commonly used in information retrieval fields.

4.3.3 Time Constraint

Time efficiency is also an important requirement for idea organization in GSS. A meeting in GSS environment typically follows a preset agenda with time allocated to each stage of the meeting. Automatic idea organization should be done as fast as possible because group participants may still need to further refine the list generated by the system. This time efficiency requirement may make unfavorable the algorithms that may require a long time to process, especially those algorithms like neural network techniques that may need many iterations.

4.4 Solution – A SVD-enabled System for Automatic Idea Organization

We need to find an appropriate solution to automatic idea organization that can cope with the unique characteristics in GSS. Our system for automatic idea organization is designed and implemented by adopting the following techniques covering preprocessing, term weight assignment and feature selection, and singular value decomposition.

4.4.1 Preprocessing

Preprocessing is an important step to extract content from brainstorming comments data. The domain-independence makes it difficult to utilize natural language processing techniques. Automatic indexing method has proven to be a simple, fast and domain-independent approach to representing textual content (Salton 1989). Therefore, we adopted automatic indexing method to preprocess brainstorming comments data into meaningful words.

4.4.2 Feature Selection and Document Representation

Since each term may vary in terms of its importance in representing a comment, it is important to represent each term with appropriate weight. Of all the term weighting techniques, term frequency and inverse document frequency have proven to be very useful (Salton 1989). Term frequency allows more weights to be assigned to a term if that term occurs more frequently in a comment. Inverse document frequency gives fewer weights to terms that occur more frequently across different comments and thus specific terms get higher weights than general terms. Therefore, we followed Salton's vector space model to associate each term with a weight indicating its relative importance. In addition, we limited the number of terms used in vector space model by setting up a term frequency threshold. In this way, it can help remove noise from our brainstorming comments for higher accuracy of topic generation and at the same time increase time efficiency of the algorithm

4.4.3 Singular Value Decomposition

The most important question in our automatic approach is how to derive topic categories from brainstorming comments. We decided to apply the singular value decomposition approach for this purpose. SVD is an algebraic method of matrix decomposition for discovering the orthogonal basis of the original term-document matrix. This basis consists of orthogonal vectors that hypothetically correspond to topics present in the original term-document matrix. Therefore, a reduced SVD-decomposed term document matrix can be used to identify abstract topics. In this research, we applied SVD to automate idea organization in GSS by reducing the original number of comments vector dimensions into a smaller number of dimensions representing the underlying topic categories in comments data. We then used the generated topic labels as key words to search for relevant comments that can be placed into their respective topic categories.

4.5 Research Methodology

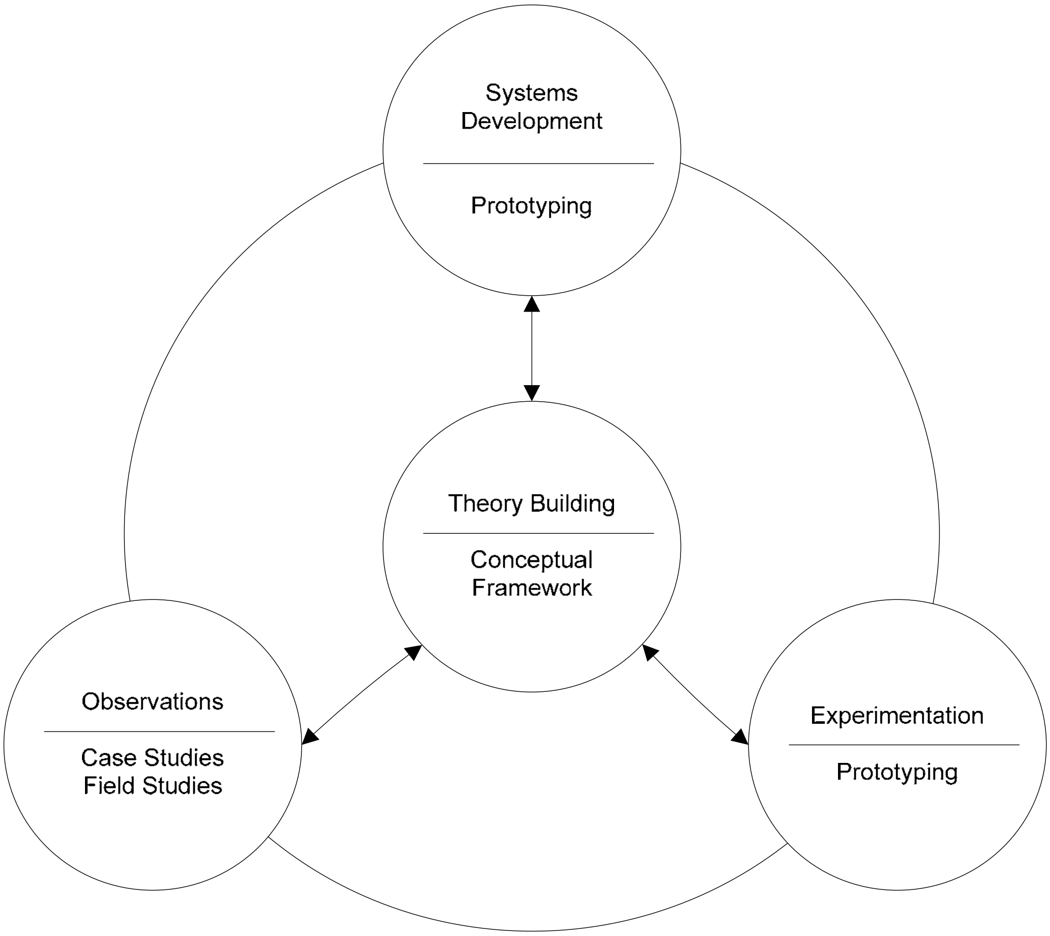

Figure 4.1 Systems Development Multi-methodological Research Approach

To address the research question, we decided to follow the system development methodology (Nunamaker, Jr. et al. 1991), as shown in figure 4.1. System development methodology is a multi-methodological research approach with focus on investigating information systems phenomena by building and evaluating an IT artifact. The system development approach consists of four research strategies including theory building, experimentation, observation, and systems development. Of these four research methodologies, no one is sufficient by itself to be the preeminent methodology. Instead, the use of multiple methodologies in the research can complement each other. Given the complexity of the group support systems, it is necessary to use an integrated multi-methodological research methodology.

Following system development methodology, this research started by selecting and designing appropriate algorithms, then proceeded to system prototyping and finally empirically evaluated our prototype system.

4.5.1 Theory Building

Theory building is used to develop new ideas, topics, conceptual frameworks, new methods or models. The theories generated are important for formulating research hypotheses, guiding the design of experiments and conducting observation.

4.5.2 Experimentation

Experimentation involves the use of research strategies including laboratory and field experiments to investigate research hypotheses based on theory. Experimentation is guided by theories and facilitated by system development. The outcomes may be used to refine theories and improve systems. Experimentations also close the gap between theory building and observation in terms of validating the underlying theories and dealing with issues such as acceptance and technology transfer.

4.5.3 Observation

Observation is used to investigate the phenomenon of interest in order to have a good understanding of what is involved in a research domain. Observation can be accomplished using various unobtrusive research methodologies including case studies, field studies and sample surveys. The result of observation may be useful for researchers to design hypotheses to be tested through experimentation or to arrive at generalizations for later investigations.

4.5.4 System Development

System development explores and synthesizes various technologies to produce the system, which is an IT artifact central to the system development process. It is the hub of research that forms an integrated research program by interacting with other research methodologies. This IT artifact that results from systems development serves as the important bridge connecting technological research to the social research. Therefore, in the multi-methodological approach, system development plays a crucial role in understanding the development and use of information systems. During the system development process, various difficulties may be encountered, which can in turn be used to modify the theories.

Figure 4.2 shows that there are five different stage during systems development including topic design, construction of the architecture of the system, prototyping, product development, and technology transfer (Nunamaker, Jr. et al. 1991).

- Construct a conceptual framework. A conceptual framework is the starting point for the system development. Research questions are stated and research objectives are outlined. A framework is provided to guide the entire research process.

- Develop system architecture. System architecture is a blue print for the entire systems building process. During this stage, components are defined and relationships among system components are specified. System requirements and functionalities are defined so that the design and implementation of the system are based on the requirements collected.
- Analyze and design the system. During this step, researchers analyze the system requirements to understand the studied domain, apply relevant scientific and technical knowledge and propose various alternative solutions to the research problem. The results are design specifications used as a blueprint for the implementation of the system.
- Build the system. IS researchers conduct research by building a prototype system. With the prototype system, the feasibility of the design and the usability of the research project can be demonstrated. While implementing the prototype, researchers can gain insights into the topics, the frameworks and the chosen design alternatives.
- Experiment, observe, and evaluate the system. With a system prototype built, researchers then use the system to conduct experiments, observation, and evaluation of the system to test its performance and usability and observe its impacts on individuals, groups, or organizations. The experience gained

through experiments, observation and evaluation can be used to further refine the system or formulate a new theory to explain the observed phenomena.

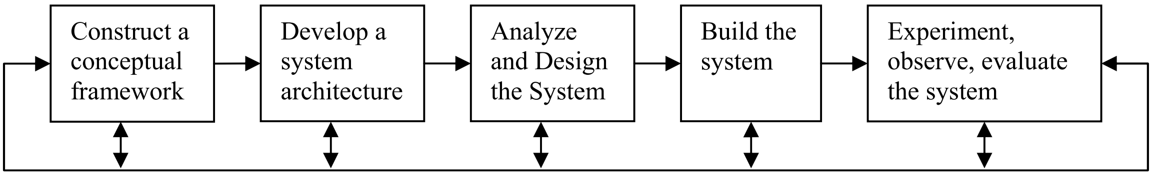

Figure 4.2 System Development Stages

CHAPTER 5: A SINGULAR VALUE DECOMPOSITION APPROACH TO AUTOMATIC IDEA ORGANIZATION

In this chapter, we discuss the details on how singular value decomposition can be applied to automatic idea organization in GSS.

We based our analysis on a set of brainstorming comments data collected and combined from 2 group sessions. Both of these group sessions were conducted in GroupSystems and attended by college students to answer the brainstorming question: "What can we do to make our student union center a better place in the university?"

The basic information unit in the brainstorming comments data is a textual comment, submitted by group participants as their contribution to solving the brainstorming question. A comment is composed of short paragraphs. Given that the two group sessions address the same brainstorming question, we are able to combine the comments from two sessions into one single data set. There are a total of 138 comments in the combined list of brainstorming data with 72 comments from one session and 66 comments from the other. Figure 5.1 shows a sample of brainstorming comments from the combined list along with the brainstorming question listed at the top. Each comment is assigned a number ranging from 1 to 138.

Some of the characteristics displayed by this brainstorming data set include the occurrence of typos (e.g., “fiexible” in comment 2; “communities” in comment 3), special abbreviations (e.g., “rec” in comment 75; “high tech” in comment 13), simple word or phrases (e.g., “fiscal smart” in comment 15; “versatile” in comment 121), incomplete sentences (e.g., “More efficient to operate due to technology” in comment 1, “either new construction or renovated with high tech features” in comment 13). Of these characteristics, typos occur most frequently with occurrences in 34 comments out of the 138 comments. Some comments are vague, e.g., "will extend well beyond the borders of the traditional building" in comment 97 whereas other comments are very specific, e.g., "Clearly more technology related" in comment 98. The shortest comment is a one-word comment "versatile" in comment 121 and the longest comment is one paragraph of more than 60 words in comment 67.

Brainstorming Question: What can we do to make our student union center a better place in the university?	
1	More efficient to operate due to technology
2	The Union must be diverse and fiexible in it;'s array of conveniences and services
3	As mini communties (and community centers) spring up on campus, some sense of larger community is lost. There will still be place for the traditional union, but the challenge will be to pull people together in one location vs. the segmentation that decentralized communities bring
……	
13	either new construction or renovated with high tech features
14	But how do we foster viewpoint neutrality when student affairs overall is incredibly left-leaning in a usually unabashed way?
15	fiscal smart
……	
75	Will more likely to be combined with other campus facilities such as Rec/Sports Centers, Housing and Academic/Classroom facilities.
……	
120	will develop students to be citizens
121	versatile
122	will still rent facilities for meetings, receptions, dinners
……	

Figure 5.1 Sample Comments of Combined Brainstorming Data in GSS

5.1 Preprocessing

We treated each comment in the brainstorming data as a separate document in our analysis. We selected automatic indexing to preprocess brainstorming data.

5.1.1 Word Identification

The objective in this step is to identify all meaningful words in each comment. For each comment passed as input into our system, we removed all characters that have no meaning, which include non-letter characters like punctuations or '#', '@', etc. At the same time, we also removed terms with one character letter and also overly long terms that exceed 20 characters. Terms are identified by using special characters such as white space characters as word delimiters and special punctuation marks such as "." and "?" as sentence delimiters. Numerical data formats such as the comment number associated with each comment were ignored during preprocessing. The result of word identification is a cleaned representation of terms.

5.1.2 Stop-Wording

Stop-wording is a step to identify and remove stop words that occur frequently but carry no substantial meanings such as "the" and "that" that can affect the quality of topic descriptions. Stop wording is often carried out by implementing a stop word list, which is simply a list of stop words in English. We developed a special "stop word" list by complementing common non-semantic stop words with special additions that are typical of brainstorming comments. For example, the new stop word list includes abbreviated phrases such as "we've, I'm, it's, etc" and words that often occur in group discussions but carry no substantial meanings in generating topic categories such as "agree, disagree, exactly, etc".

5.1.3 Stemming

Stemming is used to reduce the impact of syntax by replacing words with their respective stems. A stem is a part of a word that remains after removing its affixes including prefixes and suffixes. Stemming is often performed by using a stemming algorithm called stemmer to reduce different forms of a word to its base form. Given that Porter stemmer has been one of the most commonly used stemming algorithm (Porter. 1980), we decided to use Porter's algorithm to find out word stems for all terms whenever possible.

5.1.4 Phrase Extraction

Previous research on phrase-based methods as described in the literature review section showed the importance of using phrases in providing descriptions for clusters in document clustering. Therefore, it is important that phrases are also incorporated in the singular valued decomposition approach for generating topic categories.

Phrases are recurring ordered sequences of terms. Unlike single words, they maintain the original order of terms and do not cross sentence boundaries. Since phrases are much easier to understand than single words, they are often more appropriate to represent topic categories than single words.

We based our definition of frequent phrases on "complete phrases" introduced by Zhang and Dong (Zhang and Dong. 2004). The use of complete phrases allows the generation of longest possible frequent phrases. A complete phrase is both right-complete and left-complete. Informally, a complete phrase is a phrase that cannot be altered by preceding or trailing it with additional elements, because at least one of these elements is different from the rest. An algorithm was proposed by Zhang and Dong to identify complete phrases and their frequencies by using a data structured called "suffix array" (Zhang and Dong. 2004). We adopted the algorithm and generated a total of 50 phrases that occurred at least twice in comments.

5.2 Feature Selection and Vector Space Model

The brainstorming data revealed the prevalence of typos with around 25% of comments containing one or more typos. For feature selection, it is useful to set a threshold to filter out misspelled words. At the same time, it is also important to keep as many words in the comments as possible to accurately represent each comment. Therefore, given that most misspellings only occurred once, we chose to set term frequency threshold at 2, indicating that the vector space model was built from terms that occurred more than once in the brainstorming comments.

We constructed a vector space model by building term vectors for all comments. Specifically, for each comment, we calculated the term frequency, tf_{ij}, representing the number of times term j occurs in comment i and the document frequency, df_j, representing the number of comments in which word j occurs. Next, based on df_j, we computed its inverse, idf_j, by dividing the total number of comments by dfj. We then computed w_{ij}, the weight of term j in comment i, based on the commonly used tf-idf with the following formula: $w_{ij} = tf_{ij} * \log(idf_j)$. The weight w_{ij} measures the frequency of term j in comment i with a factor that discounts its importance when term j appears in almost all comments. Previously, Chen et al. adopted document frequency instead of inverse document frequency to capture the consensus topics in the comments (Chen, Hsu et al. 1994). However, after experimenting with document frequency, we concluded that the derived topics were too general; that is, content classification did not achieve the desired topic categorization, but rather generated categories that were too general to qualify as a single topic.

```
The t=14 terms               The c=9 comments

T1 : Technology tf=5         C1: Clearly more technology related
T2 : Union      tf=5         C2: But how do we foster viewpoint neutrality when student affairs overall
T3 : Viewpoint  tf=5             incredibly left-leaning in a usually unabashed way?
T4 : More       tf=4         C3: viewpoint neutrality - safe for all opinions, all sides, all viewpoint
T5 : Neutral    tf=4             can be safely discussed
T6 : Students   tf=4         C4: Much more technology based, greater demand for wireless
T7 : Opinions   tf=3         C5: I hope that viewpoint neutral does not mean that strong opinions are n
T8 : Safe       tf=3             expressed in the union, but rather that the union is a place where the
T9 : Way        tf=3             opinions are shared in a civil manner, and that the union itself does
T10: Changing   tf=2             take a politically partisan position on issues
T11: Hopefully  tf=2         C6: With less money we will be forced to find more efficiencies - with
T12: Place      tf=2             technology growth it's likely that more core function operate out of t
T13: Political  tf=2             unions i.e., financial aid, career development, academic advising - sp
T14: Stay       tf=2             won't be a concern due to technology
                             C7: Technology will mandate that the way we reach out to students and keep
                                 them engaged in new ways. Many reasons why students come to our
                                 facilities are no longer applicable
The p=2 phrases              C8: Hopefully union will continue to be safe place/viewpoint neutral as ou
                                 political climate changes
P1: Viewpoint Neutral tf=4   C9: Very flexible in order to stay with the many changing attitudes and ne
P2: More Technology   tf=2       of the student population-and will stay central to University communit
```

Figure 5.2 An Example of Terms and Phrases Extraction on a Reduced Data Set of 9 Comments

For illustration purpose, we obtained a reduced data set containing 9 comments from the original brainstorming data mentioned earlier and used this reduced data set as an example in the following in order to help explain the procedure in detail. Figure 5.2 shows this reduced brainstorming data set containing 9 comments with 14 terms and 2 phrases extracted. Figure 5.3 shows the vector space model drawn from the reduced data set of 9 comments.

$$
A = \begin{array}{cc}
 & \begin{array}{ccccccccc} C1 & C2 & C3 & C4 & C5 & C6 & C7 & C8 & C9 \end{array} \\
\begin{array}{c} T1 \\ T2 \\ T3 \\ T4 \\ T5 \\ T6 \\ T7 \\ T8 \\ T9 \\ T10 \\ T11 \\ T12 \\ T13 \\ T14 \end{array} &
\begin{bmatrix}
0.59 & 0.00 & 0.00 & 0.59 & 0.00 & 0.55 & 0.21 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.62 & 0.37 & 0.00 & 0.30 & 0.00 \\
0.00 & 0.37 & 0.42 & 0.00 & 0.15 & 0.00 & 0.00 & 0.22 & 0.00 \\
0.81 & 0.00 & 0.00 & 0.81 & 0.00 & 0.75 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.37 & 0.21 & 0.00 & 0.15 & 0.00 & 0.00 & 0.22 & 0.00 \\
0.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.58 & 0.00 & 0.23 \\
0.00 & 0.00 & 0.39 & 0.00 & 0.57 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.79 & 0.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 \\
0.00 & 0.69 & 0.00 & 0.00 & 0.00 & 0.00 & 0.79 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.32 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.28 & 0.00 & 0.00 & 0.40 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.28 & 0.00 & 0.00 & 0.40 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.28 & 0.00 & 0.00 & 0.40 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.92
\end{bmatrix}
\end{array}
$$

Figure 5.3 An Example of Vector Space Model for a Reduced Data Set of 9 Comments

5.3 Singular Value Decomposition Approach to Automatic Idea Organization

Our system applied singular value decomposition in steps described as follows to achieve automatic topic generation and automatic comment placement, which are the two major components of idea organization.

5.3.1 Automatic Topic Generation

After the term-document matrix was constructed, we applied SVD method to first reduce the dimension of the original matrix. That is, we decomposed the original term-document matrix to find the orthogonal basis of the matrix. Any rectangular matrix A, such as a $t \times d$ term-document matrix, can be decomposed into a product of three sub-matrices including U, S and V such that $A \approx U_r \times S_r \times V_r^T$. U is a $t \times t$ orthogonal matrix whose column vectors are often called left singular vectors of A whereas V is a $d \times d$ orthogonal matrix whose column vectors are called right singular vectors of A. S is a diagonal matrix of singular values, which are ordered decreasingly along its diagonal. Subscript r represents the rank of A, which is the number of its non-zero singular values. Importantly, the r columns of U form an orthogonal basis for the term space of the original matrix A. The orthogonal vectors in U are uncorrelated and different from each other and can thus be perceived as vector representations of the abstract topic categories drawn from brainstorming comments.

However, the number of r orthogonal basis vectors is often very large, often close to the number of comments in the input data. Therefore, an important approximation step must be conducted to minimize the number of r orthogonal basis vectors. Since the singular values in S are ordered decreasingly according to their significance to the process of reconstruction of the original matrix A, a truncated SVD that keeps only the first k columns of S can be used to project the original matrix A onto a k-dimensional space in the form of $A_k = U_k \times S_k \times V_k^T$, which can be considered as the rank k approximation of A. In the new approximated A_k, S_k is a diagonal matrix that contains the k singular values along the diagonal line with higher value of k representing the closer approximation of A_k to the original matrix A. Therefore, a reduced U that keeps only the first k columns of U can be used as the most important categories of topics because it captures the major associational structure of the data and throws out much of the noise. We used this reduced model U to generate topic categories from our comments data.

One of the benefits of using SVD is that the value of k can be determined by using Frobenius Norm, which calculates the percentage distance between the original term-document matrix A and its k-rank approximation A_k. Specifically, if we determine that a certain percentage of the original matrix A is to be retained by its k-rank approximation A_k, we can then use this percentage to calculate the value of k (Berry, Drmac et al. 1999). We want the value of k to be large enough to fit all the real structure in the data, but small enough so that we do not overly fit the sampling error or unimportant details. At the same time, we also need to consider the manageable number of topic categories that is suitable for group participants to make decisions on. According to previous research on idea organization in group settings, the desired number of topic categories that group participants can manage should be seven plus or minus one or two (Miller 1956). Therefore, after running many experiments, we chose to set the threshold value of the percentage at 0.52, which means that 52% of original structure is retained.

For illustration, figure 5.4 shows that for the reduced brainstorming data set of 9 comments, the value of k is set at 2, which means the first 2 vectors in U separated by vertical lines are retained vectors for topic categories.

$$
U = \left\{ \begin{array}{cc||ccccccc}
0.59 & -0.09 & 0.07 & 0.01 & 0.03 & -0.03 & -0.26 & 0.07 & -0.74 \\
0.17 & 0.26 & -0.30 & -0.01 & -0.50 & 0.25 & -0.03 & -0.71 & -0.00 \\
0.05 & 0.36 & -0.06 & 0.11 & 0.27 & 0.06 & 0.51 & -0.03 & -0.17 \\
0.77 & -0.19 & -0.04 & -0.01 & 0.09 & -0.02 & 0.21 & 0.06 & 0.55 \\
0.04 & 0.30 & -0.01 & 0.07 & 0.08 & -0.04 & 0.61 & 0.03 & -0.21 \\
0.07 & 0.32 & 0.48 & -0.09 & -0.06 & 0.03 & -0.11 & -0.00 & 0.22 \\
0.05 & 0.28 & -0.25 & 0.09 & -0.01 & 0.69 & -0.18 & 0.48 & 0.08 \\
0.04 & 0.37 & -0.28 & 0.08 & 0.68 & -0.15 & -0.42 & -0.25 & 0.11 \\
0.10 & 0.41 & 0.63 & 0.19 & -0.12 & -0.04 & -0.16 & -0.01 & 0.04 \\
0.02 & 0.17 & -0.09 & -0.35 & -0.01 & -0.43 & -0.04 & -0.03 & 0.06 \\
0.04 & 0.22 & -0.20 & -0.04 & -0.24 & -0.25 & -0.05 & 0.26 & 0.01 \\
0.04 & 0.22 & -0.20 & -0.04 & -0.24 & -0.25 & -0.05 & 0.26 & 0.01 \\
0.04 & 0.22 & -0.20 & -0.04 & -0.24 & -0.25 & -0.05 & 0.26 & 0.01 \\
0.01 & 0.09 & 0.09 & -0.89 & 0.12 & 0.24 & 0.05 & 0.01 & -0.07
\end{array} \right\}
$$

Figure 5.4 An Example of Matrix U for a Reduced Data Set of 9 Comments

Next, we treated the generated phrases as if they were comments and built a new matrix P of size $t \times (p + t)$ where t is the number of terms and p is the number of frequent phrases. In this new matrix P, single terms were represented as row vectors and frequent phrases and single words represented as column vectors. We used the same tf-idf weighting schemes to construct matrix P as we used earlier for term-document vector space model. For illustration, figure 5.5 shows an example of P matrix calculated on 2 phrases and 14 terms extracted from 9 comments.

Next, we computed which phrase or single terms can be used to label the two topic vectors. Since topic vectors in the reduced U were expressed in the column space of the original term-document, they share the same space as the newly constructed frequent phrases and terms vectors in matrix P. We normalized the column vectors of both U and P so that we can use cosine similarity measure to compute how close phrases and single terms vectors in P are to topic vectors in the reduced U. That is, we constructed a new matrix M = $U_k^T \times P$ with phrases and single words represented as columns and abstract topics represented as rows. Thus, for each row, the highest similarity score indicates the phrase or single term that best approximates the corresponding topic vector. We selected the phrases or single terms that have the highest similarity score to represent the generated topic categories in GSS. As a result, we completed generating topic categories from brainstorming comments. For illustration, figure 5.6 shows an example of matrix M, which multiples matrix U_k^T by matrix P. The product takes 2 phrases and 14 terms as column vectors and 2 reduced U vectors as row vectors, and shows the similarity between label candidates and topics. Figure 5.6 indicates that P2 “more technology” and P1 “viewpoint neural” have the highest score respectively for U1 and U2 topic vectors and are thus used to represent the 2 topics in the reduced brainstorming data of 9 comments.

$$P = \left\{ \begin{array}{lcccccccccccccccc}
 & P1 & P2 & T1 & T2 & T3 & T4 & T5 & T6 & T7 & T8 & T9 & T10 & T11 & T12 & T13 & T14 \\
T1 & 0.00 & 0.59 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T2 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T3 & 0.71 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T4 & 0.00 & 0.81 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T5 & 0.71 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T6 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T7 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T8 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T9 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T10 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
T11 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\
T12 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\
T13 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\
T14 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00
\end{array} \right\}$$

Figure 5.5 An Example of Matrix P Calculated on 2 Phrases and 14 Terms Extracted from a Reduced Data Set of 9 Comments

$$M = U_K^T P = \left\{ \begin{array}{lrrrrrrrrrrrrrrrr} & \text{P1} & \text{P2} & \text{T1} & \text{T2} & \text{T3} & \text{T4} & \text{T5} & \text{T6} & \text{T7} & \text{T8} & \text{T9} & \text{T10} & \text{T11} & \text{T12} & \text{T13} & \text{T14} \\ \text{U1} & 0.06 & 0.97 & 0.59 & 0.17 & 0.05 & 0.77 & 0.04 & 0.07 & 0.05 & 0.04 & 0.10 & 0.02 & 0.04 & 0.04 & 0.04 & 0.01 \\ \text{U2} & 0.47 & -0.20 & -0.09 & 0.26 & 0.36 & -0.19 & 0.30 & 0.32 & 0.28 & 0.37 & 0.41 & 0.17 & 0.22 & 0.22 & 0.22 & 0.09 \end{array} \right\}$$

Figure 5.6 An Example of Matrix C for a Reduced Data Set of 9 Comments

In figure 5.6, two phrases have the highest score respectively for two topic vectors and are thus selected as labels for these two topic vectors. In fact, what figure 5.6 shows is a simplified version which does not distinguish between single words and phrases. Given that phrases are better choices for describing topic categories, our system is designed to give more weights to phrases in determining choices to describe each topic vector. Specifically, for each topic, our system first selects two potential labels using both a single word and a phrase that have the highest score in all single words and all phrases respectively, and then finalize the choice by comparing the weight for the selected single term and phrase. During comparison, our system multiples a weight factor “w” to the phrase’s score to increase its weight. Repeated experiments showed that it is appropriate to set the weight factor “w” at 1.7 if the phrase contains the single word and at 1.2 otherwise. The difference in the value of weight factors is due to the fact that it is much more appropriate to choose the phrase instead of the single word if the phrase contains the single word itself. Figure 5.7 shows that from the combined 138 brainstorming comments data, our system generated 8 topic categories, of which there are seven phrases and one single word.

1. Central Location
2. Community Center
3. Programs and Services

4. Technology
5. Space for Students
6. Academic Affairs
7. Viewpoint Neutral
8. Commercial Vendor

Figure 5.7 System-Generated Topic Categories for 138 Comments

5.3.2 Automatic Comment Placement

After topic categories are generated, the next step is to place relevant comments into their respective topic categories. The comment placement process follows a standard information retrieval task by using the generated topic labels as queries against the term-document matrix generated earlier. If the similarity between the topic label and a comment exceeds a certain placement threshold, the comment is considered relevant to the topic and then placed directly under the topic category. The comment placement threshold ranges from 0.0 to 1.0. Ideally, the placement threshold value should not be too low so that it may lead to more comments placed but with lower degree of accuracy. Similarly, the value should not be too high either because it may lead to very few comments considered "relevant". Experiments showed that comments are most appropriately placed under respective categories when the threshold value is set to 0.125. While it is possible to allow overlapping of clusters by placing a comment into multiple topics when it reaches or exceeds the threshold value, we ran the algorithm with and without overlapping and compared the results. We decided against overlapping because the system places a comment into the topic that has the highest score of all topics that exceed the threshold value, and thus allows for the most appropriate placement.

For illustration, we again used the sample of 9 comments as an example. Figure 5.8 shows matrix Q, which is composed of topic vectors for P1 and P2 extracted from the first two columns of matrix P.

$$
Q = \left\{ \begin{array}{lll}
 & \text{P1} & \text{P2} \\
\text{T1} & 0.00 & 0.59 \\
\text{T2} & 0.00 & 0.00 \\
\text{T3} & 0.71 & 0.00 \\
\text{T4} & 0.00 & 0.81 \\
\text{T5} & 0.71 & 0.00 \\
\text{T6} & 0.00 & 0.00 \\
\text{T7} & 0.00 & 0.00 \\
\text{T8} & 0.00 & 0.00 \\
\text{T9} & 0.00 & 0.00 \\
\text{T10} & 0.00 & 0.00 \\
\text{T11} & 0.00 & 0.00 \\
\text{T12} & 0.00 & 0.00 \\
\text{T13} & 0.00 & 0.00 \\
\text{T14} & 0.00 & 0.00
\end{array} \right\}
$$

Figure 5.8 Topic Vectors P1 and P2

For illustration, figure 5.9 shows an example of matrix C, which multiples matrix Q^T by matrix A. The product takes 9 comments as column vectors and 2 topics as row vectors and shows the similarity between comments and topics. Each topic label is treated as a query to search for the best comment that contains the topic label. Specifically in this example, for each comment, two scores are generated which represent the similarity between the comment and both topics. The highest of the two scores is then compared with the threshold value of 0.125. If the highest score equals to or exceeds the threshold value of 0.125, the comment is then placed under the corresponding topic. Otherwise the comment is placed into the category called “Other Topics”.

$$C = Q^T A = \left\{ \begin{array}{cccccccccc} & D1 & D2 & D3 & D4 & D5 & D6 & D7 & D8 & D9 \\ P1 & 0.00 & 0.52 & 0.45 & 0.00 & 0.22 & 0.00 & 0.00 & 0.31 & 0.00 \\ P2 & 1.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.93 & 0.13 & 0.00 & 0.00 \end{array} \right\}$$

Figure 5.9 Similarity Matrix C with Topics as Rows and Comments as Columns

Topic 1: More Technology (P2)

D1: Clearly more technology related [1.00]
D4: Much more technology based, greater demand for wireless [1.00]
D6: With less money we will be forced to find more efficiencies – with technology growth it's likely that more core function operate out of the unions i.e., financial aid, career development, academic advising - space won't be a concern due to technology [0.93]
D7: Technology will mandate that the way we reach out to students and keep them engaged in new ways. Many reasons why students come to our facilities are no longer applicable [0.13]

Topic 2: Viewpoint Neutral (P1)

D2: But how do we foster viewpoint neutrality when student affairs overall is incredibly left-leaning in a usually unabashed way? [0.52]
D3: viewpoint neutrality - safe for all opinions, all sides, all viewpoints can be safely discussed [0.45]
D5: I hope that viewpoint neutral does not mean that strong opinions are not expressed in the union, but rather that the union is a place where these opinions are shared in a civil manner, and that the union itself does not take a politically partisan position on issues [0.22]
D8: Hopefully union will continue to be safe place/viewpoint neutral as our political climate changes [0.31]

Other Topics:

D9: Very flexible in order to stay with the many changing attitudes and needs of the student population-and will stay central to University community

Figure 5.10 An Example of Results from Comment Placement

For illustration, figure 5.10 shows the final result of automatic idea organization. For topic category generation, there are two topics generated and displayed; namely, “More Technology” and “Viewpoint Neutral”. For comment placement, there are four comments displayed under “Topic 1: More Technology” including D1, D4, D6 and D7 with similarity scores of 1.00, 1.00, 0.93 and 0.13, respectively. Similarly, there are also four comments displayed under “Topic 2: Viewpoint Neutral” including D2, D3, D5 and D8 with similarity scores of 0.52, 0.45, 0.22, 0.31, respectively. All comments listed under either topic 1 or topic 2 have similarity scores that exceed the threshold value of 0.125. D9 is listed under “Other Topics” because its similarity scores for both topics are both 0.00, well below the threshold value 0f 0.125.

Figure 5.11 is the final output of our system, which shows all 8 topics and a sample of comment placement results produced by our system. The system list has each topic name followed by a series of comments indicating that the comments listed under the category are ones closely relevant to the respective category.

Programs and Services
76 Programming base for the campus. Programming will ocur everywhere on campus and not restricted to the union.
88 The philosphy of student programming should be examined. What do students really want? Do they want to be programmed to at all?
………………..
Technology
105 In many ways they will be where technology leads us
132 As technology becomes more mainstream and a greater part of our lives, will students accept not having access to this technology?
………………..
Space for Students
112 More space for student organizations and student government
57 space is being created in several academic buildings and lounge space is being built into athletic and academic buildings-difficult to compete for the students' time
………………..
Community Center
34 It will still be the community center of campus. However, it might be both brick and morter and virtual 3 As mini communties (and community centers) spring up on campus,
………………..
Academic Affairs

91	we will be more aligned with academic affairs
92	It will need to be a place where tangible partnerships between student affairs and academic affairs occurs on a regular basis.
………………..	
Central Location	
70	The central location to help with diversity because it should be the home for all students.
111	The large central located unions may not be able to serve the campus effective where smaller but several at locations around the campus may be more effective
………………..	
Commercial Vendor	
66	Commercial vendor interface with self operated board plans is essential for future growth.
119	More commercial enterprises outsourced
………………..	
Viewpoint Neutral	
14	But how do we foster viewpoint neutrality when student affairs overall is incredibly left-leaning in a usually unabashed way?
79	viewpoint neutrality - safe for all opinions, all sides, all viewpoints can be safely discussed
………………..	
(OTHER)	
103	The union will continue to be in competition for student disposable income, as well as student fee dollars.
73	the types of services the students want to see will change as a result of trends that we need to stay in tune with
………………..	

Figure 5.11 System's List of Comment Placement

CHAPTER 6: SYSTEM EVALUATION

This chapter describes the process of using two experiments to evaluate our system tool in terms of topic generation and comment placement, followed by analysis of the experimental results. With the system now in place for automatic idea organization from brainstorming comments, there are two primary evaluation questions that need to be answered using experiments:

1. How does the list of topic categories generated by the system compare with the quality of human lists?
2. How does the automatic comments placement produced by the system compare with the comment placement done by humans?

To address these important evaluation questions concerning the performance of the automatic idea organization approach, we designed and conducted two experiments which compared the system's output on topic generation and comments placement with that generated by human subjects.

6.1 Brainstorming Comments List

The brainstorming comments list used in the experiment are the same data set used as input data fed into our system. The list contains brainstorming output from two electronic brainstorming sessions in real electronic meetings using GroupSystems. The two sessions consist of 72 and 66 textual comments, respectively. The topic of discussion used in both sessions is: "What can we do to make our student union center a better place in the university?" Given that both sessions address the same brainstorming question, we combined the data from two sessions into a single data set consisting of 138 comments in total. In addition, we randomized the order of the 138 comments and used the randomized list as our final list of 138 comments. In the final list, each comment is preceded with a number which ranges from 1 to 138. The list of 138 comments was presented to each human subject during all stages for experiment 1 and 2.

6.2 Human Subjects Pool

The majority of human subjects were recruited from an introductory management information systems class in the Eller College of Management at the University of Arizona. We randomly selected and invited 65 senior undergraduate students out of 356 students across two sections of the class. Of the 65 email invitations, 40 students accepted the invitation. The participating students were given course credits for their participation in the laboratory experiment.

In addition to these senior undergraduate students, we also invited 6 expert facilitators who have at least three years of experience in facilitating GSS session. Four out of the 6 experts accepted our invitation and participated in our study.

In total, there are 44 human subjects involved in different stages of the two experiments. Of the 44 human subjects, 40 are senior undergraduates with the rest being expert facilitators. The participating human subjects completed their tasks in different stages of experiment 1 and 2.

6.3 Experiment 1: Evaluating System's Topic List

Typically after the electronic brainstorming during a GSS session, the topic categories with key words as labels are generated by group participants with help from facilitators to categorize the comments into topics addressing the brainstorming question. The first experiment examines the performance of automatic topic generation as compared with undergraduate subjects, who are treated as domain experts. Evaluation is conducted by human subjects including both undergraduate domain experts and expert facilitators. There are three separate stages involved in experiment 1; namely, manual generation of topic categories, evaluation of topic generation by undergraduate subjects and evaluation of topic generation by expert facilitators. The three stages vary in terms of tasks, subjects and procedures.

6.3.1 Stage 1: Manual Generation of Topic Categories

The first stage of experiment 1 is designed to allow domain experts to manually generate topic categories from brainstorming comments in an idea organization task. This stage produced 9 lists of manually generated topic categories, with which automatically generated topic categories were later compared in stage 2 and stage 3.

6.3.1.1 Subjects

The brainstorming question is: what can we do to make our student union center a better place in the university? Clearly, the subject matter under discussion involves a student union center at a university. Even though college undergraduates may not be experienced in guiding groups through the convergent process during an electronic meeting, they are ideal domain experts who are familiar with the topic under discussion. Out of the pool of 40 undergraduate subjects, we randomly selected 9 students to participate in this stage.

6.3.1.2 Procedure

When all 9 student subjects arrived at the experimental laboratory, the laboratory facilitator handed out an instruction sheet to each participant to ensure consistent delivery of instructions, see Appendix A. The instruction sheet described the background and the objective of the experiment, step-by-step direction and a detailed example of how to complete the task. After reading the instruction, each subject was then provided with the list of 138 comments along with a form for them to write down a list of topic categories, see Appendix B. Human subjects were asked to read through the list of 138 comments and identify their own topic category list with a maximum of 8 topic categories using either phrase or single key words as labels. A maximum of 8 topic categories were allowed in order to reduce cognitive workload of human subjects and to make it consistent with the list of 8 topic categories automatically generated. All human subjects were told to take as long time as they wish and were required to record the time they spent for completing the manual generation of topic categories. After completion of topic generation, the subjects were thanked and dismissed from the room.

6.3.1.3 Data Collection

The elapsed time and the generated topics were collected. With 9 human subjects, this stage of manual topic generation produced 9 lists with each human subject producing one list. The 9 lists were used in the second and third stage to be compared with the list of automatically generated topic categories.

6.3.2 Stage 2: Evaluation of Topic Generation by Undergraduate Subjects

The second stage of experiment 1 allows undergraduate subjects to evaluate the 10 lists one by one in order to measure the performance of the automatically generated topic list as compared with the 9 manually generated topic lists. This stage produced 10 revised topic lists along with 10 lists of ranked orders in terms of topic quality.

6.3.2.1 Subjects

Similar to the first stage, stage 2 also uses undergraduates as human subjects because these undergraduate subjects are familiar with the topic under discussion. However, in this stage, we exclude those undergraduates who have already participated in the first stage of experiment 1. That is, our study chose a group of human subjects for the evaluation task different from those who participated in stage 1. Out of the pool of the remaining 31 undergraduate subjects, we randomly selected 10 students to participate in this stage.

6.3.2.2 Procedure

When all 10 student subjects arrived at the experimental laboratory, the laboratory facilitator handed out an instruction sheet to each participant to ensure consistent delivery of instructions, see Appendix C. The instruction sheet described the background and the objective of the experiment, step-by-step direction and a detailed example of how to complete the task. After reading the instruction, each subject was then provided with the list of 138 comments. All human subjects were also given a form with 10 lists of topic categories, see Appendix D. They were told about the process how the 10 existing categories were derived from the 138 comments and that the all lists were generated by human subjects who had gone through the entire process. They were not told that the list 10 was a list generated automatically by the system. Then all the subjects were asked to browse the list of 138 comments. Based on their understanding of all the comments, they were asked to delete categories that do not represent important topics for the 138 comments and add missing categories. If rewording or editing of a topic category is needed, it was treated as a deletion followed by an addition. Human subjects were told to keep the number of final topics for each list at a maximum of 8 topics. Finally, after all the 10 topic category lists were corrected, they were asked to rank the 10 lists from the most appropriate to the least appropriate with respect to identifying solutions that could address the brainstorming question. All human subjects were told to take as long time as they wish. After completion of topic generation, the subjects were thanked and dismissed from the room.

6.3.2.3 Data Collection

For each list, we recorded deleted topics and added topics. These added new topics and deleted topics for each list were then used later in the analysis stage to generate quality measures in terms of topic recall and topic precision. In addition, information on ranked order for 10 lists was collected.

6.3.3 Stage 3: Evaluation of Topic Generation by Expert Facilitators

The third stage of experiment 1 allows expert facilitators to evaluate the 10 lists one by one. Two objectives can be accomplished in this stage: on one hand, similar to stage 2, the performance of the automatically generated topic list can be compared with the 9 manually generated topic lists; on the other hand, another comparison can be made to find out if there is any difference between domain experts and expert facilitators in terms of evaluation results. This stage produced 4 revised topic lists along with 4 lists of ranked orders in terms of topic quality.

6.3.3.1 Subjects

Different from the second stage, stage 3 used expert facilitators who have at least three years of GSS facilitation experience as human subjects. The experience with facilitating electronic meeting could help these facilitators develop criteria on which to develop a good list of topic categories. There are a total of 4 expert facilitators who agreed to participate in the study.

6.3.3.2 Procedure

The entire procedure is the same as the stage 2 in this experiment.

6.3.3.3 Data Collection

Data collection is the same as the stage 2 in this experiment.

6.4 Experiment 2: Evaluating System's Comment Placement List

Typically after the topic categories are generated during a GSS session, group participants browse a list of comments and place comments to their respective topic categories by dragging and dropping comments into their respective categories in GSS. The objective of this experiment is to address the quality of comment placement list by comparing the output of the human subjects and that of the automatic system. Evaluation is conducted by human subjects who are undergraduate domain experts. There are two separate stages involved in experiment 2; namely, manual comment placement and evaluation of automatic comment placement. The two stages vary in terms of tasks, subjects and procedures.

6.4.1 Stage 1: Manual Comment Placement

The first stage of experiment 2 is designed to allow domain experts to manually generate topic categories in an idea organization task. This stage produced 5 lists of manually generated comment placement, with which the automatically generated comment placement list was later compared in stage 2.

6.4.1.1 Subjects

Placing relevant comments into their respective topic categories requires human subjects to be familiar with the subject matter under discussion, which involves a student union center at a university. Similar to the first stage of experiment 1, we decided to use undergraduates because they are ideal domain experts who are familiar with the topic under discussion. Out of the remaining 21 undergraduate subjects, we randomly selected 5 students to participate in this stage.

6.4.1.2 Procedure

When all the 5 student subjects arrived at the experimental laboratory, the laboratory facilitator handed out an instruction sheet to each participant to ensure consistent delivery of instructions, see Appendix E. The instruction sheet described the background and the objective of the experiment, step-by-step direction and a detailed example of how to complete the task. After reading the instruction, each subject was then provided with the list of 138 comments. A form with 8 system-generated topic categories was also presented to each subject. Under each category, a large space was available for subjects to write down comment number which they think should belong to that category, see Appendix F. Human subjects were asked to read through the list of 138 comments and identify all relevant comments for each of the 8 topic categories. To make it consistent with the automatically generated comment placement and also reduce cognitive load for subjects, we asked human subjects to limit one comment to only one topic category. That is, no overlapping of topics was allowed. For comments that subjects did not believe should be placed into any of the 8 topic categories, they were left untouched without being placed into any topic categories. All human subjects were told to take as long time as they wish and were required to record the time they spent for completing the manual generation of comment placement. After completion of comment placement, the subjects were thanked and dismissed from the room.

6.4.1.3 Data Collection

Finally, the elapsed time and the generated comment placement lists were collected. With 5 human subjects, this stage of manual generation of comment placement produced 5 lists with each human subject producing one list. The 5 lists were then used in the second stage to be compared with the automatically generated comment placement list.

6.4.2 Stage 2: Evaluation of Automatic Comment Placement

The second stage of experiment 2 allowed undergraduate subjects to evaluate the 6 lists one by one in order to measure the performance of the automatically generated comment placement list as compared with the 5 manually generated comment placement lists. This stage produced 6 revised comment placement lists together with 6 lists of ranked orders in terms of the comment placement quality.

6.4.2.1 Subjects

Similar to the first stage, stage 2 also used undergraduates as human subjects because these undergraduate subjects were familiar with the topic under discussion. However, in this stage, we excluded the 5 undergraduates who have previously participated in stage 1 of experiment 2. That is, our study chose a different group of human subjects for the evaluation task. We asked the remaining 16 students to participate in this stage.

6.4.2.2 Procedure

When all the 16 student subjects arrived at the experimental laboratory, the laboratory facilitator handed out an instruction sheet to each participant to ensure consistent delivery of instructions, see Appendix G. The instruction sheet described the background and the objective of the experiment, step-by-step direction and a detailed example of how to complete the task. After reading the task, each subject was then provided with the list of 138 comments. All human subjects were also given 6 different lists of comment placement. They were told about the process how relevant comments were placed under their respective topics and that all lists were generated by human subjects who had gone through the entire process. They were not told that there was a list generated automatically by the system. All subjects were asked to delete irrelevant comments from topics and add missing comments to topics. For comments that do not belong to any of the 8 system-generated topic categories, subjects were told to move them to “Other Topics”. Finally, after all the 6 comment placement lists were corrected, human subjects were asked to rank the 6 lists from the most appropriate to the least appropriate with respect to associating relevant comments to their respective categories, see Appendix H. All human subjects were told to take as long time as they wish. After completion of topic generation, the subjects were thanked and dismissed from the room.

6.4.2.3 Data Collection

For each topic within each list, we recorded the deleted and added comments. These added and deleted comments for each topic of 6 lists were then used later in the analysis stage to generate quality measures in terms of placement recall and placement precision. In addition, information on ranked orders for 6 comment placement lists was collected for all 16 human subjects.

6.5 Experimental Results

In the following, we report the statistical analysis and results from the two experiments, including results related to topic category generation and related to comment placement.

6.5.1 Experiment 1 – Results

Table 6.1 contains 10 lists of topic categories generated from the 138 comments data. 9 lists of topics ranging from list 1 to list 9 were created by human subjects in stage 1 of the first experiment. The system list, i.e., list 10, contains the topic categories generated automatically by the system using the SVD techniques. From these lists, we could see that list 7 contains 5 topics, the smallest number of topic categories of the 10 lists. List 1, 2 and 3 all have 6 topic categories while list 8 has 7 topic categories. The remaining lists, i.e., list 4, 6, 9 and the system list, all have 8 topic categories generated. Table 6.1 also shows the time spent for each of the lists. System list finishes the entire idea organization task for 0.67 second whereas the time spent by humans for topic generation task alone ranges from 20 minutes to 41 minutes.

List 1: 28 minutes	List 2: 20 minutes
1. one-stop place	1. one stop food and services
2. increase technology	2. high quality but personal
3. leadership/mentor laboratories	3. efficient and effective at meeting students needs
4. commercialization	4. accessible
5. social events	5. cater for student meeting, activities, growth
6. adaptable	6. technologically advanced
List 3: 25 minutes	**List 4: 24 minutes**
1. one-stop shop/centralization	1. investment
2. up-to-date/new technology	2. entrepreneurship
3. commercialization	3. accessibility
4. more programming/services	4. diversity
5. maintain values (learning and development)	5. communication
6. re-evaluate staffing	6. technology
	7. security
	8. openness
List 5: 33 minutes	**List 6: 35 minutes**
1. one-stop shop	1. community
2. commercialization	2. technology
3. technology	3. Commercial
4. viewpoint neutrality	4. Adaptable
5. adaptable	5. one-stop shop
6. decentralization	6. Efficient
	7. extra-curricular
	8. Irrelevant
List 7: 31 minutes	**List 8: 37 minutes**
1. one stop shop	1. technologically advance
2. a campus living room	2. adaptable
3. union funding	3. commercialization
4. service variety	4. diverse community/cultural center
5. change with the times	5. Unify student services
	6. student development center

	7. "face of the university"
List 9: 41 minutes	List 10 – System: 0.67 second (including comment placement)
1. greater demand for wireless	1. central location
2. transportation of goods to students	2. community center
3. facility for citizenship	3. programs and services
4. 24 hour service with assistance	4. technology
5. more rooms to study	5. space for students
6. more active with students teachers	6. academic affairs
7. construction for more space	7. viewpoint neutral
8. more career showcase at various hours	8. commercial vendor

Table 6.1 Manually Generated Topic Lists and System-generated Topic List

6.5.1.1 Results from Topic Quality Evaluation by Undergraduate Subjects

Table 6.2 shows the results of the list evaluation by 10 undergraduate subjects in stage 2 of the first experiment. The names of the 10 lists are listed on the top. The 10 subjects that give evaluations are listed on the left. For each subject, there are 6 rows associated. The first row indicates the rank assigned to each of the 10. The second, third and fourth row refer to the "target", "identified" and "relevant", respectively. The "target" is the total number of items that remain in each list after the subject completes adding missing topics and deleting inappropriate topics. The "identified" is the number of items that the system or the human subject identifies (i.e., the number of listed topic items). The "relevant" shows the intersection of the "target" and "identified" rows, which are the number of topics in the original list that still remain in the corrected list.

The use of "target", "identified" and "relevant" allows the analysis of topic quality by the standard recall and precision techniques from information retrieval literature. In this experiment, we defined topic recall and topic precision, as shown on the fifth and sixth row for each individual human subject. Topic recall is generated by dividing the number of relevant items by the number of target items for each list, representing the percentage of relevant categories captured in the revised list. Topic precision is computed by dividing the number of relevant items by the number of identified, representing the percentage of relevant topics in the original list. Values of these two measures are between 0 and 1. Higher values of topic recall and topic precision are desirable.

Table 6.2 indicates that the list that was ranked the best most frequently is the system list, which was ranked the best by 4 out of 10 subjects whereas the list that received the second highest number of the best rank is list 3, which was ranked the best by 2 out of 10 subjects. List 2, 4, 6 and 7 were all ranked the best by only 1 out of 10 subjects. List 1, 5, 8 and 9 are the only lists that were not ranked the best by any of the 10 subjects. In addition, system list was ranked the second best by 2 out of 10 subjects, which was on par with list 3 and list 6, both of which were also ranked the second best by 2 out of 10 subjects. List 2, list 4, list 5 and list 7 were all ranked the second best by only one out of 10 subjects. On average, our system list was ranked the best out of the 10 lists.

As further corroboration of the quality of our system list, subjects again rated system list as the best in terms of both average topic recall and average topic precision, which are 0.95 and 0.9 respectively. The second best list is list 3, which has the second best topic recall at 0.91 and the second best topic precision at 0.88. The worst list for topic recall is list 5 at 0.77 whereas the worst list for topic precision is list 6 at 0.76.

		LISTS									
SUBJECTS		List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10 system

Subject 1	Rank	7	4	3	1	9	6	10	5	8	2
	Target	6	6	6	8	7	8	6	7	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	5	6	7	5	7	4	7	6	8
	Recall	0.83	0.83	1.00	0.88	0.71	0.88	0.67	1.00	0.75	1.00
	Precision	0.83	0.83	1.00	0.88	0.83	0.88	0.80	1.00	0.75	1.00
Subject 2	Rank	10	9	8	2	6	4	5	3	7	1
	Target	6	4	5	6	6	7	5	8	7	6
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	4	5	6	6	6	4	7	6	6
	Recall	0.83	1.00	1.00	1.00	1.00	0.86	0.80	0.88	0.86	1.00
	Precision	0.83	0.67	0.83	0.75	1.00	0.75	0.80	1.00	0.75	0.75
Subject 3	Rank	5	3	8	9	2	4	10	6	7	1
	Target	8	5	7	7	6	7	6	6	7	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	6	5	5	5	4	7	4	5	5	8
	Recall	0.75	1.00	0.71	0.71	0.67	1.00	0.67	0.83	0.71	1.00
	Precision	1.00	0.83	0.83	0.63	0.67	0.88	0.80	0.71	0.63	1.00
Subject 4	Rank	10	2	4	6	3	9	8	5	7	1
	Target	5	6	7	8	6	7	6	7	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	5	6	7	6	5	5	6	8	8
	Recall	1.00	0.83	0.86	0.88	1.00	0.71	0.83	0.86	1.00	1.00
	Precision	0.83	0.83	1.00	0.88	1.00	0.63	1.00	0.86	1.00	1.00
Subject 5	Rank	6	9	7	10	8	1	4	5	3	2
	Target	7	6	5	8	6	8	6	5	4	7
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	5	5	7	4	7	4	3	2	7
	Recall	0.71	0.83	1.00	0.88	0.67	0.88	0.67	0.60	0.50	1.00
	Precision	0.83	0.83	0.83	0.88	0.67	0.88	0.80	0.43	0.25	0.88
Subject 6	Rank	4	3	7	6	8	10	2	5	9	1
	Target	6	6	6	7	6	8	4	6	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	4	5	7	4	7	4	5	6	8
	Recall	0.83	0.67	0.83	1.00	0.67	0.88	1.00	0.83	0.75	1.00
	Precision	0.83	0.67	0.83	0.88	0.67	0.88	0.80	0.71	0.75	1.00
Subject 7	Rank	6	7	2	10	5	8	1	3	9	4
	Target	7	5	6	8	7	8	4	6	7	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	6	4	5	5	5	6	4	6	7	7
	Recall	0.86	0.80	0.83	0.63	0.71	0.75	1.00	1.00	1.00	0.88
	Precision	1.00	0.67	0.83	0.63	0.83	0.75	0.80	0.86	0.88	0.88
Subject 8	Rank	8	1	2	4	9	7	10	3	6	5
	Target	7	6	6	7	7	7	5	7	7	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	6	6	6	6	5	7	5	7	7	7
	Recall	0.86	1.00	1.00	0.86	0.71	1.00	1.00	1.00	1.00	0.88

	Precision	1.00	1.00	1.00	0.75	0.83	0.88	1.00	1.00	0.88	0.88
Subject 9	Rank	9	5	1	4	3	2	8	7	6	10
	Target	6	6	6	8	7	8	5	7	8	7
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	4	4	5	7	5	7	4	6	6	5
	Recall	0.67	0.67	0.83	0.88	0.71	0.88	0.80	0.86	0.75	0.71
	Precision	0.67	0.67	0.83	0.88	0.83	0.88	0.80	0.86	0.75	0.63
Subject 10	Rank	8	9	1	3	10	2	4	5	6	7
	Target	8	5	5	8	6	7	4	7	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	4	5	6	5	7	3	7	8	8
	Recall	0.63	0.80	1.00	0.75	0.83	1.00	0.75	1.00	1.00	1.00
	Precision	0.83	0.67	0.83	0.75	0.83	0.88	0.60	1.00	1.00	1.00

Average Recall	0.80	0.84	0.91	0.84	0.77	0.88	0.82	0.89	0.83	0.95
Average Precision	0.87	0.77	0.88	0.79	0.82	0.83	0.82	0.84	0.76	0.90

Table 6.2 Topic Recall and Topic Precision for 10 Lists

For topic recall, statistical analysis of the results showed that based on 5% significance level with P at 0.05, the F-test (analysis of variance, ANOVA) found that there was no significant difference among the 10 lists with respect to topic recall (P: 0.07) even though the P value of 0.07 is very close to the significance level of 0.05. Table 6.3 contains a summary of the ANOVA analysis data for topic recall. In addition to the ANOVA results, table 6.3 also includes descriptive statistics for all 10 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	.7970	.10863	.03435	.7193	.8747	.63	1.00
2	10	.8430	.12383	.03916	.7544	.9316	.67	1.00
3	10	.9060	.10648	.03367	.8298	.9822	.71	1.00
4	10	.8470	.11861	.03751	.7622	.9318	.63	1.00
5	10	.7680	.13071	.04133	.6745	.8615	.67	1.00
6	10	.8840	.09958	.03149	.8128	.9552	.71	1.00
7	10	.8190	.13763	.04352	.7205	.9175	.67	1.00

8	10	.8860	.12536	.03964	.7963	.9757	.60	1.00
9	10	.8320	.16963	.05364	.7107	.9533	.50	1.00
10	10	.9470	.09707	.03070	.8776	1.0164	.71	1.00
Total	100	.8529	.12829	.01283	.8274	.8784	.50	1.00

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.258	9	.029	1.880	.065
Within Groups	1.372	90	.015		
Total	1.629	99			

Table 6.3 ANOVA and Descriptive Statistics for Topic Recall

Pairwise two-sample t-tests were performed with results showing that our system list significantly outperformed list 1, 2, 4, 5, and 7 in topic recall with P levels at 0.00, 0.05, 0.05, 0.00, 0.03, respectively. However, our system's recall level was not statistically different from the topic recall levels of list 3, 6, 8 and 9 with P levels at 0.38, 0.17, 0.24 and 0.08, respectively. Table 6.4 contains a summary of the two-sample t-tests between our system list and all the 9 lists generated by human subjects.

	t-test for Equality of Means						
	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	-3.256	17.777	0.004	-0.15	0.046	-0.247	-0.053
List 2 Vs. System	-2.09	17.029	0.052	-0.104	0.05	-0.209	0.001
List 3 Vs. System	-0.9	17.848	0.38	-0.041	0.046	-0.137	0.055
List 4 Vs. System	-2.063	17.323	0.054	-0.1	0.048	-0.202	0.002
List 5 Vs. System	-3.477	16.612	0.003	-0.179	0.051	-0.288	-0.07
List 6 Vs. System	-1.433	17.988	0.169	-0.063	0.044	-0.155	0.029
List 7 Vs. System	-2.403	16.178	0.029	-0.128	0.053	-0.241	-0.015
List 8 Vs. System	-1.217	16.939	0.24	-0.061	0.05	-0.167	0.045
List 9 Vs. System	-1.861	14.324	0.083	-0.115	0.062	-0.247	0.017

Table 6.4 Two Sample T-test (Topic Recall)

For topic precision comparison, statistical analysis of the F-test (analysis of variance ANOVA) also revealed that based on the 5% significance level with P at 0.05, there was no significant difference among the 10 lists with respect to topic precision (P: 0.29). Table 6.5 contains a summary of the ANOVA analysis data for topic precision. In addition to the ANOVA results, table 6.5 also includes descriptive statistics for all 10 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values for each of the 10 lists.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	.8650	.10544	.03334	.7896	.9404	.67	1.00
2	10	.7670	.11412	.03609	.6854	.8486	.67	1.00
3	10	.8810	.08212	.02597	.8223	.9397	.83	1.00
4	10	.7910	.10354	.03274	.7169	.8651	.63	.88
5	10	.8160	.12140	.03839	.7292	.9028	.67	1.00
6	10	.8290	.08837	.02795	.7658	.8922	.63	.88
7	10	.8200	.11353	.03590	.7388	.9012	.60	1.00
8	10	.8430	.18421	.05825	.7112	.9748	.43	1.00
9	10	.7640	.21634	.06841	.6092	.9188	.25	1.00
10	10	.9020	.12761	.04035	.8107	.9933	.63	1.00
Total	100	.8278	.13331	.01333	.8013	.8543	.25	1.00

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.193	9	.021	1.230	.287
Within Groups	1.567	90	.017		
Total	1.759	99			

Table 6.5 ANOVA and Descriptive Statistics for Topic Precision

Pairwise two-sample t-tests were performed with results showing that our system list significantly outperformed list 2 and 4 in topic precision with P levels at 0.02 and 0.05, respectively. However, our system's recall level was not statistically different from the topic precision levels of list 1, 3, 5, 6, 7, 8 and 9 with P levels at 0.49, 0.67, 0.14, 0.16, 0.15, 0.42 and 0.1, respectively. Table 6.6 contains a summary of the two-sample t-tests between our system list and all the 9 lists generated by human subjects.

	t-test for Equality of Means

	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	-0.707	17.382	0.489	-0.037	0.05235	-0.14726	0.07326
List 2 Vs. System	-2.494	17.78	0.023	-0.135	0.054	-0.249	-0.021
List 3 Vs. System	-0.438	15.363	0.668	-0.021	0.048	-0.123	0.081
List 4 Vs. System	-2.136	17.267	0.047	-0.111	0.052	-0.221	-0.001
List 5 Vs. System	-1.544	17.955	0.14	-0.086	0.056	-0.203	0.031
List 6 Vs. System	-1.487	16.018	0.156	-0.073	0.049	-0.177	0.031
List 7 Vs. System	-1.518	17.759	0.147	-0.082	0.054	-0.196	0.032
List 8 Vs. System	-0.833	16.021	0.417	-0.059	0.071	-0.209	0.091
List 9 Vs. System	-1.737	14.586	0.103	-0.138	0.079	-0.308	0.032

Table 6.6 Two Sample T-test (Topic Precision)

We can see from the results that when evaluated by undergraduate subjects, the system's topic list may be at least comparable to, if not better than, the human-generated topic lists.

6.5.1.2 Results from Topic Quality Evaluation by Expert Facilitator Subjects

Table 6.7 shows the results of the list evaluation in stage 3 of the first experiment by 4 expert facilitator subjects. The names of the 10 lists are listed on the top. The 4 subjects that give evaluations are listed on the left. For each subject, there are 6 rows associated. The first row indicates the rank assigned to each of the 10 lists. The second, third, fourth, fifth and sixth row refer to the "target", "identified", "relevant", "recall" and "precision", respectively. The definitions for each of these five terms are the same as in stage 2 of the first experiment. The only difference here is that the evaluation was given by 4 expert facilitators instead of 10 undergraduate subjects.

Table 6.7 indicates that as far as the rank is concerned, list 9 is the most controversial list because it was ranked as the best by 2 expert facilitators and the worst by the other two expert facilitators. Our system list was ranked the second worst twice but at the same time was ranked the second and third place by the other two expert facilitators. On average, our system list was ranked in the middle of the 10 lists.

Further analysis revealed that expert facilitators rated system list as the third best in terms of both average topic recall and average topic precision, which are 0.81 and 0.81, respectively. The best list is list 6, which has both the best topic recall at 0.94 and the best topic precision at 0.84. The worst list is list 9, which received both the lowest topic recall at 0.28 and the lowest topic precision at 0.28.

		LISTS									
SUBJECTS		List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10 system
Subject 1	Rank	6	8	9	10	4	5	7	3	1	2
	Target	7	6	7	8	8	8	8	8	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	4	4	4	6	6	5	6	1	6
	Recall	0.71	0.67	0.57	0.50	0.75	0.75	0.63	0.75	0.13	0.75
	Precision	0.83	0.67	0.67	0.50	1.00	0.75	1.00	0.86	0.13	0.75
Subject 2	Rank	2	1	3	5	4	8	6	7	10	9
	Target	8	7	6	8	6	5	7	6	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	5	6	6	6	5	5	5	6	0	7
	Recall	0.63	0.86	1.00	0.75	0.83	1.00	0.71	1.00	0.00	0.88
	Precision	0.83	1.00	1.00	0.75	0.83	0.63	1.00	0.86	0.00	0.88
Subject 3	Rank	10	6	4	5	3	2	7	8	1	9
	Target	6	6	6	8	6	8	5	7	8	8
	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	6	5	6	8	6	8	5	7	8	8
	Recall	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Precision	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Subject 4	Rank	7	9	1	8	6	2	5	4	10	3
	Target	6	5	4	5	5	8	4	7	6	8

	Identified	6	6	6	8	6	8	5	7	8	8
	Relevant	2	1	4	4	2	8	0	1	0	5
	Recall	0.33	0.20	1.00	0.80	0.40	1.00	0.00	0.14	0.00	0.63
	Precision	0.33	0.17	0.67	0.50	0.33	1.00	0.00	0.14	0.00	0.63

Average Recall	0.67	0.64	0.89	0.76	0.75	0.94	0.58	0.72	0.28	0.81
Average Precision	0.75	0.67	0.83	0.69	0.79	0.84	0.75	0.71	0.28	0.81

Table 6.7 Topic Recall and Topic Precision for 10 Lists by Expert Facilitators

For topic recall, statistical analysis of the results showed that based on 5% significance level with P at 0.05, the F-test (analysis of variance, ANOVA) found that consistent with the evaluation given by undergraduate subjects, there was no significant difference among the 10 lists with respect to topic recall (P: 0.21). Table 6.8 contains a summary of the ANOVA analysis data for topic recall. In addition to the ANOVA results, table 6.8 also includes descriptive statistics for all 10 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	4	.6675	.27548	.13774	.2291	1.1059	.33	1.00
2.00	4	.6400	.30496	.15248	.1547	1.1253	.20	.86
3.00	4	.8925	.21500	.10750	.5504	1.2346	.57	1.00
4.00	4	.7625	.20565	.10282	.4353	1.0897	.50	1.00
5.00	4	.7450	.25252	.12626	.3432	1.1468	.40	1.00
6.00	4	.9375	.12500	.06250	.7386	1.1364	.75	1.00
7.00	4	.5850	.42115	.21057	-.0851	1.2551	.00	1.00
8.00	4	.7225	.40582	.20291	.0767	1.3683	.14	1.00
9.00	4	.2825	.48224	.24112	-.4849	1.0499	.00	1.00
10.00	4	.8150	.16010	.08005	.5602	1.0698	.63	1.00
Total	40	.7050	.32151	.05084	.6022	.8078	.00	1.00

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.220	9	.136	1.447	.213
Within Groups	2.811	30	.094		
Total	4.031	39			

Table 6.8 ANOVA and Descriptive Statistics for Topic Recall by Expert Facilitators

Pairwise two-sample t-tests were performed with results showing that our system's topic recall level was not statistically different from all other lists generated by human subjects with all values of P much greater than 0.05. Table 6.9 contains a summary of the two-sample t-tests between our system list and the 9 lists generated by human subjects.

	t-test for Equality of Means						
	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	-0.926	4.819	0.399	-0.148	0.159	-0.562	0.267
List 2 Vs. System	-1.016	4.537	0.361	-0.175	0.172	-0.632	0.282
List 3 Vs. System	0.578	5.545	0.586	0.078	0.134	-0.257	0.412
List 4 Vs. System	-0.403	5.660	0.702	-0.053	0.130	-0.376	0.271
List 5 Vs. System	-0.468	5.076	0.659	-0.070	0.149	-0.453	0.313
List 6 Vs. System	1.206	5.667	0.276	0.123	0.102	-0.130	0.375
List 7 Vs. System	-1.021	3.849	0.367	-0.230	0.225	-0.865	0.405
List 8 Vs. System	-0.424	3.912	0.694	-0.092	0.218	-0.704	0.519
List 9 Vs. System	-2.096	3.653	0.111	-0.533	0.254	-1.265	0.200

Table 6.9 Two Sample T-test (Topic Recall) by Expert Facilitators

For topic precision comparison, statistical analysis of the F-test (analysis of variance ANOVA) based on the 5% significance level with P at 0.05 also revealed that, consistent with the evaluation given by undergraduate subjects, there was no significant difference among the 10 lists with respect to topic precision (P: 0.49). Table 6.10 contains a summary of the ANOVA analysis data for topic precision. In addition to the ANOVA results, table 6.10 also includes descriptive statistics for all 10 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values for each of the 10 lists.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	4	.7475	.28964	.14482	.2866	1.2084	.33	1.00
2.00	4	.6675	.35799	.17900	.0979	1.2371	.17	1.00
3.00	4	.8350	.19053	.09526	.5318	1.1382	.67	1.00
4.00	4	.6875	.23936	.11968	.3066	1.0684	.50	1.00
5.00	4	.7900	.31696	.15848	.2856	1.2944	.33	1.00
6.00	4	.8450	.18556	.09278	.5497	1.1403	.63	1.00
7.00	4	.7500	.50000	.25000	-.0456	1.5456	.00	1.00
8.00	4	.7150	.38897	.19449	.0961	1.3339	.14	1.00
9.00	4	.2825	.48224	.24112	-.4849	1.0499	.00	1.00
10.00	4	.8150	.16010	.08005	.5602	1.0698	.63	1.00
Total	40	.7135	.33039	.05224	.6078	.8192	.00	1.00

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.957	9	.106	.967	.486
Within Groups	3.300	30	.110		
Total	4.257	39			

Table 6.10 ANOVA and Descriptive Statistics for Topic Precision by Expert Facilitators

Pairwise two-sample t-tests were performed with results showing that our system's topic precision level was not statistically different from all other lists generated by human subjects with all values of P much greater than 0.05. Table 6.11 contains a summary of the two-sample t-tests between our system list and all the 9 lists generated by human subjects.

	t-test for Equality of Means						
	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	-0.408	4.677	0.701	-0.067	0.165	-0.502	0.367
List 2 Vs. System	-0.752	4.154	0.492	-0.148	0.196	-0.684	0.389
List 3 Vs. System	0.161	5.827	0.878	0.020	0.124	-0.287	0.327
List 4 Vs. System	-0.886	5.237	0.415	-0.128	0.144	-0.493	0.238
List 5 Vs. System	-0.141	4.437	0.894	-0.025	0.178	-0.499	0.449
List 6 Vs. System	0.245	5.874	0.815	0.030	0.123	-0.271	0.331

List 7 Vs. System	-0.248	3.609	0.818	-0.065	0.263	-0.826	0.696
List 8 Vs. System	-0.475	3.988	0.659	-0.100	0.210	-0.685	0.485
List 9 Vs. System	-2.096	3.653	0.111	-0.533	0.254	-1.265	0.200

Table 6.11 Two Sample T-test (Topic Precision) by Expert Facilitators

6.5.2 Experiment 2 – Results

The second experiment consists of two stages. Stage 1 asked 5 undergraduate subjects to place relevant comments into their respective categories. As a result, a total of 5 comment placement lists was generated by human subjects. Our idea organization system produced a comment placement list automatically. These six lists altogether were given to 16 undergraduates for evaluation in stage 2 of the second experiment.

Table 6.12 shows the time spent for each of the lists. System list finishes the entire idea organization task for 0.67 second whereas the time spent by humans for comment placement task alone ranges from 21 minutes to 32 minutes.

List	1	2	3	4 (system)	5	6
Time	27 minutes	21 minutes	27 minutes	0.67 second including topic generation	25 minutes	32 minutes

Table 6.12 Time Spent for Each Comment Placement List

Table 6.13 shows the results of the comment placement list evaluation by 16 human subjects in stage 2 of the second experiment. The names of the 6 lists including the system list (i.e., list 4) and the manual list (i.e., list 1, 2, 3, 5, 6) are listed on the top. The 16 human subjects that give evaluations to the 6 lists are listed on the left. Similar to the evaluation of topic categories, there are 6 rows associated for each human subject. The first row indicates the rank assigned to each of the 6 lists. The second, third and fourth row refer to the “target”, “identified” and “relevant”, respectively.

SUBJECTS		LISTS List 1	List 2	List 3	List 4 System	List 5	List 6
Subject 1	Rank	5	1	4	3	6	2
	Target	123	123	125	126	126	125
	Identified	138	99	101	82	106	106
	Relevant	38	40	58	51	51	62
	Recall	0.31	0.33	0.46	0.40	0.40	0.50
	Precision	0.28	0.40	0.57	0.62	0.48	0.58
Subject 2	Rank	3	6	2	1	4	5
	target	132	125	122	124	131	128
	identified	138	99	101	82	106	106
	relevant	66	33	58	53	55	68
	recall	0.50	0.26	0.48	0.43	0.42	0.53
	precision	0.48	0.33	0.57	0.65	0.52	0.64
Subject 3	Rank	6	3	4	5	2	1
	target	128	108	102	109	111	113
	identified	138	99	101	82	106	106
	relevant	76	61	90	77	70	88
	recall	0.59	0.56	0.88	0.71	0.63	0.78
	precision	0.55	0.62	0.89	0.94	0.66	0.83
Subject 4	Rank	3	2	4	6	1	5
	target	125	122	122	122	125	120
	identified	138	99	101	82	106	106
	relevant	80	38	56	48	52	57
	recall	0.64	0.31	0.46	0.39	0.42	0.48
	precision	0.58	0.38	0.55	0.59	0.49	0.54
Subject 5	Rank	6	4	1	3	2	5
	target	122	100	95	87	97	95
	identified	138	99	101	82	106	106
	relevant	100	71	80	67	76	75
	recall	0.82	0.71	0.84	0.77	0.78	0.79
	precision	0.72	0.72	0.79	0.82	0.72	0.71
Subject 6	Rank	6	3	5	2	4	1
	target	133	109	108	97	113	123
	identified	138	99	101	82	106	106
	relevant	121	93	95	78	101	104
	recall	0.91	0.85	0.88	0.80	0.89	0.85
	precision	0.88	0.94	0.94	0.95	0.95	0.98
Subject 7	Rank	5	6	2	3	4	1
	target	125	125	125	125	124	126
	identified	138	99	101	82	106	106
	relevant	52	48	68	57	56	70
	recall	0.42	0.38	0.54	0.46	0.45	0.56
	precision	0.38	0.48	0.67	0.70	0.53	0.66
Subject 8	Rank	6	5	1	2	4	3

	target	145	128	119	111	123	123
	identified	138	99	101	82	106	106
	relevant	96	89	99	80	87	99
	recall	0.66	0.70	0.83	0.72	0.71	0.80
	precision	0.70	0.90	0.98	0.98	0.82	0.93
Subject 9	Rank	1	5	2	6	4	3
	target	143	117	116	102	117	120
	identified	138	99	101	82	106	106
	relevant	122	92	96	75	97	102
	recall	0.85	0.79	0.83	0.74	0.83	0.85
	precision	0.88	0.93	0.95	0.91	0.92	0.96
Subject 10	Rank	3	5	6	4	2	1
	target	135	103	101	86	103	102
	identified	138	99	101	82	106	106
	relevant	125	88	90	74	99	100
	recall	0.93	0.85	0.89	0.86	0.96	0.98
	precision	0.91	0.89	0.89	0.90	0.93	0.94
Subject 11	Rank	6	3	4	5	2	1
	target	138	119	109	94	114	114
	identified	138	99	101	82	106	106
	relevant	117	92	101	79	106	106
	recall	0.85	0.77	0.93	0.84	0.93	0.93
	precision	0.85	0.93	1.00	0.96	1.00	1.00
Subject 12	Rank	5	2	6	1	3	4
	target	134	106	103	89	111	112
	identified	138	99	101	82	106	106
	relevant	107	79	83	65	86	85
	recall	0.80	0.75	0.81	0.73	0.77	0.76
	precision	0.78	0.80	0.82	0.79	0.81	0.80
Subject 13	Rank	5	2	6	1	3	4
	target	138	108	111	91	106	105
	identified	138	99	101	82	106	106
	relevant	87	70	84	70	90	94
	recall	0.63	0.65	0.76	0.77	0.85	0.90
	precision	0.63	0.71	0.83	0.85	0.85	0.89
Subject 14	Rank	5	6	3	4	2	1
	target	147	118	106	93	114	113
	identified	138	99	101	82	106	106
	relevant	127	89	95	81	106	104
	recall	0.86	0.75	0.90	0.87	0.93	0.92
	precision	0.92	0.90	0.94	0.99	1.00	0.98
Subject 15	Rank	5	6	2	4	3	1
	target	135	106	109	95	108	111
	identified	138	99	101	82	106	106
	relevant	114	84	98	79	96	100
	recall	0.84	0.79	0.90	0.83	0.89	0.90
	precision	0.83	0.85	0.97	0.96	0.91	0.94

Subject 16	Rank	1	6	2	3	5	4
	target	136	103	99	81	107	104
	identified	138	99	101	82	106	106
	relevant	128	97	97	77	99	99
	recall	0.94	0.94	0.98	0.95	0.93	0.95
	precision	0.93	0.98	0.96	0.94	0.93	0.93

Average Recall	0.72	0.65	0.77	0.70	0.74	0.78
Average Precision	0.70	0.73	0.83	0.85	0.78	0.83

Table 6.13 Placement Recall and Placement Precision for 6 Lists

The definitions of “target”, “identified” and “relevant” in comment placement list evaluation are a little different from those in topic evaluation. For each list, the “target” shown in table 6.13 is the sum of individual “target” value for each of the 8 topic categories within the list. Each topic has its own “target”, which refers to the number of comments that remain under that category after the human subject completes adding missing comments and deleting irrelevant comments. Similarly, “identified” shown in table 6.13 is the sum of individual “identified” value for each of the 8 topics within the list. Each topic category also has its own “identified”, which is the total number of comments that are listed under the associated category before the human subject starts deleting and adding comments to that category. “Relevant” shown in table 6.13 is obtained by adding up individual “relevant” values for each of the 8 topics. Each topic category has its own “relevant” value, which is the intersection of its own “target” and “identified”; that is, “relevant” for each topic refers to the number of comments in the topic of the original list that remain in the topic of the corrected list.

The use of “target”, “identified” and “relevant” allows the analysis of quality of comment placement by standard recall and precision techniques from information retrieval literature. In this experiment, we defined placement recall and placement precision, which are shown on the fifth and sixth row for each individual human subject in table 6.13. Placement recall is generated by dividing the number of relevant items by the number of target items for each subject, representing the percentage of relevant comments placed in the corrected list. Placement precision is computed by dividing the number of relevant items by the number of identified, representing the percentage of relevant comments in the original list. Values of these two measures are between 0 and 1. Higher values of placement recall and placement precision are desirable.

In terms of the ranks given to the 6 lists, table 6.13 indicates that list 6 was rated the best by 7 human subjects. Our system list was ranked the best by 3 human subjects. List 1 and 2 were rated the best twice whereas list 3 and 5 were rated the best only once.

In terms of average placement recall, our system's list is the second worst (recall = 0.70). The list that has the highest level of average comment placement recall is list 6 (recall = 0.78). However, in terms of average placement precision, our system list has the highest value of 0.85, which is followed by list 6 (precision = 0.83), list 3 (precision = 0.83), list 5 (precision = 0.78), list 2 (precision = 0.73), and list 1 (precision = 0.70).

With respect to placement recall, statistical analysis of the results showed that the F-test (analysis of variance, ANOVA) found that there was no significant difference among the 6 lists based on the 5% significance level with P at 0.05 (P=0.42). Table 6.14 contains a summary of the ANOVA analysis data for comment placement recall. In addition to the ANOVA results, table 6.14 also includes descriptive statistics for all 6 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	16	.7219	.19222	.04806	.6194	.8243	.31	.94
2	16	.6494	.21524	.05381	.5347	.7641	.26	.94
3	16	.7731	.17980	.04495	.6773	.8689	.46	.98
4	16	.7044	.18122	.04530	.6078	.8009	.39	.95
5	16	.7369	.20684	.05171	.6267	.8471	.40	.96
6	16	.7800	.16927	.04232	.6898	.8702	.48	.98
Total	96	.7276	.19150	.01954	.6888	.7664	.26	.98

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.186	5	.037	1.013	.415
Within Groups	3.298	90	.037		
Total	3.484	95			

Table 6.14 ANOVA and Descriptive Statistics for Placement Recall

Pairwise two-sample t-tests were performed with results showing that our system's placement recall level was not statistically different from the placement recall levels of all other lists including list 1, 2, 3 5 and 6 with P levels at 0.80, 0.44, 0.29, 0.64, 0.23, respectively. Table 6.15 contains a summary of the two-sample t-tests for placement recall between our system list and other 5 lists generated by undergraduate subjects.

	t-test for Equality of Means						
	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	0.265	29.896	0.793	0.018	0.066	-0.117	0.152
List 2 Vs. System	-0.782	29.154	0.441	-0.055	0.07	-0.199	0.089
List 3 Vs. System	1.077	29.998	0.29	0.069	0.064	-0.062	0.199
List 5 Vs. System	0.473	29.49	0.64	0.033	0.069	-0.108	0.173
List 6 Vs. System	1.22	29.862	0.232	0.076	0.062	-0.051	0.202

Table 6.15 Two Sample T-test (Placement Recall)

For placement precision comparison, statistical analysis of the F-test (analysis of variance ANOVA) revealed that based on the 5% significance level with P at 0.05, there was no significant difference among the 6 lists with respect to placement precision (P: 0.15). Table 6.16 contains a summary of the ANOVA analysis data for placement precision. In addition to the ANOVA results, table 6.16 also includes descriptive statistics for all 6 lists including mean, standard deviation, standard error, 95% confidence interval, minimum and maximum values.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	16	.7063	.20409	.05102	.5975	.8150	.28	.93
2	16	.7350	.22441	.05610	.6154	.8546	.33	.98
3	16	.8325	.15801	.03950	.7483	.9167	.55	1.00
4	16	.8469	.13671	.03418	.7740	.9197	.59	.99
5	16	.7825	.18916	.04729	.6817	.8833	.48	1.00
6	16	.8319	.15609	.03902	.7487	.9150	.54	1.00
Total	96	.7892	.18374	.01875	.7519	.8264	.28	1.00

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.270	5	.054	1.656	.154
Within Groups	2.937	90	.033		
Total	3.207	95			

Table 6.16 ANOVA and Descriptive Statistics for Placement Precision

Pairwise two-sample t-tests were performed with results showing that our system list significantly outperformed list 1 on placement precision with P levels at 0.03. However, our system's placement precision level was not statistically different from placement precision levels of list 2, 3, 5, and 6 with P levels at 0.10, 0.79, 0.28, and 0.77, respectively. Table 6.17 contains a summary of the two-sample t-tests for placement precision between our system list and other 5 lists generated by human subjects.

	t-test for Equality of Means						
	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
List 1 Vs. System	-2.29	26.205	0.03	-0.141	0.061	-0.267	-0.014
List 2 Vs. System	-1.703	24.786	0.101	-0.112	0.066	-0.247	0.023
List 3 Vs. System	-0.275	29.392	0.785	-0.014	0.052	-0.121	0.092
List 5 Vs. System	-1.103	27.311	0.28	-0.064	0.058	-0.184	0.055
List 6 Vs. System	-0.289	29.488	0.774	-0.015	0.052	-0.121	0.091

Table 6.17 Two Sample T-test (Placement Precision)

We can see from the results that in terms of placement precision, the system's comment placement list is at least comparable, if not better than the manually generated placement lists by human subjects. For placement recall comparison, system's list was also proven to be at least comparable to all the human-generated comment placement lists.

CHAPTER 7: DISCUSSION

The chapter contains a discussion of the experimental results that were presented in Chapter 6. The discussion follows the structure of the two experiments, with each experiment discussed separately. Interpretations and insights are given based upon the evidence obtained from experiments.

7.1 Experiment 1

This section discusses the major findings from the study about quality of the topics generated by the system. The goal of experiment 1 is to examine whether the system-generated topic list is comparable to human-generated topic lists. Three stages of the first experiment were conducted separately consisting of manual generation of at most 8 topic categories by 9 undergraduate subjects, evaluation of topic lists by 10 different undergraduate subjects, and evaluation of topic lists by 4 expert facilitator subjects.

The manual generation of topics took between 20 and 41 minutes, considerably longer than 0.67 second (including topic placement) clocked by the system. This indicates that the system has tremendous edge over manual processing in terms of time efficiency. In addition, our system compares favorably to previous studies. For example, one of the previous studies using Hopfield algorithm produced a topic list in 7 minutes for around 300 comments (Chen, Hsu et al. 1994).

The results given by undergraduate subjects also suggest that topic recall of system-generated topic list was rated the highest of all the lists on average and also significantly higher than that of five human lists. Ideally, the set of topics generated would cover all topics present in the input collection of brainstorming comments. The higher the topic recall, the higher the capability to generate diverse topics that correspond to the ideal topics. The topic recall results show that the topic list generated by the system is significantly closer to the list of ideal topics than the five of the nine human-generated lists of topics, demonstrating that the system has the capability to generate a diverse set of important topics as compared with human subjects. The results are not surprising and are in line with the fact that the decomposed orthogonal vectors are uncorrelated and thus can be treated as distinct topic vectors in our SVD-enabled system. These topic vectors are the basic building blocks upon which our eight topic labels are generated.

The results given by undergraduate subjects also reveal that topic precision of system-generation topic list was rated the highest of all the lists on average and significantly higher than that of two human lists. Ideally, the set of topics generated would all be right topics without any irrelevant ones. The higher of topic precision, the higher the capability to generate accurate topics. The topic precision results show that the list of 8 topics generated by the system are significantly more accurate than two human lists of topics, demonstrating that the system has the capability to generate a more accurate list of important topics than human subjects. The results are also in support of the underlying technical properties of our SVD-enabled system in that the system tries to use the smallest number of topic vectors to reconstruct the largest percentage of the original term-document matrix. The result is that the system chooses the top eight orthogonal vectors that maximize the percentage of the original matrix that can be captured to 52%. Therefore, these eight topics are the most representative of all the brainstorming comments. In addition, the high precision of the system may also be attributed to how the topic labels are generated. The system gave more weights to phrases than single terms that result in 7 topics represented by phrases. These phrases are much easier to understand and more precise in meaning than single terms when used to represent the content of topics.

One of the major reasons for using two groups of human subjects to evaluate topic quality is to find out whether the same lists are evaluated differently by expert facilitators as compared with undergraduate subjects. While undergraduates may be domain experts familiar with the subject matter under discussion, expert facilitators have more experience in facilitating groups to generate topics. Expert facilitator subjects further corroborate the results given by undergraduate subject to a certain degree by rating both topic recall and topic precision as the third highest on average. The ANOVA results show that regardless of who the human subjects are, there is no significant difference among all the lists including human-generated lists and system-generated list. Given the limited number of human subjects, no substantial claims could be made about this statistical result. Undergraduate subjects rated the system-generated topic list more favorably than expert facilitators in topic recall and topic precision. That is, for undergraduate subjects, topic recall and topic precision of system-generated topic list were both rated higher than those of several human-generated topic lists. However, for expert facilitator subjects, no significant difference was detected on topic recall and topic precision between system-generated topic list and any one of the human-generated lists. The relatively low ratings given to system-generated list by expert facilitators may be partially attributed to the limited number of expert facilitators in the experiment. There are only 4 expert facilitators participating in the experiment as compared with 10 undergraduate subjects. Given the limited number of expert facilitators, it is relatively hard to detect a pattern on the quality of a topic list. For example, out of the 4 expert facilitator subjects, two subjects rated system's list as the second worst whereas the other two subjects rated as the second and third best. In addition, list 9 was ranked as the best by 2 expert facilitators but as the worst by the other two expert facilitators. In comparison, out of the 10 undergraduate subjects, six rated system's topic list as the best and the second best whereas two subjects rates system's topic list as worst and fourth worst.

7.2 Experiment 2

This section discusses the major findings from the study about quality of the comment placement list generated by the system. The goal of experiment 2 is to examine how the system-generated comment placement list is compared with human-generated comment placement lists. Two stages of the second experiment were conducted separately consisting of manual generation of comment placement lists by 5 undergraduate subjects, evaluation of comment placement lists by 16 different undergraduate subjects.

The manual generation of topics took between 21 and 32 minutes, considerably longer than 0.67 second (including topic generation) clocked by the system. This indicates that the system has tremendous edge over manual processing in terms of time efficiency. Previous studies did not reveal time spent for comment placement per se.

The results given by undergraduate subjects suggest that system's comment placement list was rated the second best in terms of overall ranks given. However, there was quite a difference in terms of placement recall and placement precision as explored in the following.

The results given by undergraduate subjects also suggest that placement recall of system-generated topic list was rated the second worst of all the 6 lists on average although there is not significant difference in placement recall between the system-generated list and the human-generated comment placement lists. Ideally, the set of all relevant comments would be listed under their respective topics. The higher the placement recall, the higher the percentage of relevant comments to be placed under their respective topics. The placement recall results show that the comment placement list generated by the system allows relatively fewer relevant comments to be placed into their respective topics though that difference is not significant between the system's comment placement list and any of the five human-generated lists of topics.

In contrast to placement recall, the results given by undergraduate subjects suggest that placement precision of system-generated comment placement list was rated the best of all the 6 lists on average and significantly higher than that of one human-generated comment placement list. However, there is no significant difference in placement recall between the system-generated comment placement list and the human-generated comment placement lists. Ideally, the set of comments listed under topics would all be relevant to their respective topics. The higher the placement precision, the higher the percentage of comments that is relevant to their respective topics. The placement precision results show that the comment placement list generated by the system allows topics to have significantly more relevant comments than one of the comment placement lists generated by human subjects.

The results of low placement recall and high placement precision may be explained by the relatively high placement threshold value we set in the system. Since the comment placement is based on the traditional information retrieval techniques, there is always a tradeoff between recall and precision. The purpose of comment placement in idea organization is to provide a context to and validate the content of the topics generated. Therefore, one of the design decisions is to increase the precision value. The results correspond well to the design decisions we made in the system. Given the limitation of the key word search, there are several cases where comments sharing the same topic are not assigned due to the difference in key words. For example, “technology” is generated as a topic label. By using “technology” as a key word, relevant comments that contain abbreviations such as “high tech” but not “technology” are assigned a smaller weight, and thus may end up not being placed to the “technology” topic.

CHAPTER 8: CONCLUSIONS, CONTRIBUTIONS, LIMITATIONS AND FUTURE DIRECTION

This chapter concludes the research by presenting a summary of our automatic approach to idea organization in GSS as well as a summary of the experimental results. We further explore implications and limitations of this research and discuss possible future direction of this line of research.

This research is aimed at complementing GSS with automatic idea organization support. Group support systems are often equipped with various important software tools that can be employed to support individual steps in a collaboration process. While divergent tools prove very useful in divergent activities, GSS do not have effective tools available to help group participants reduce information overload and cognitive demand associated with idea organization in GSS. Current tools associated with idea organization rely exclusively on participants to manually classify all comments into topic categories with little or no support to help participants reduce the high cognitive demand.

This research continues the line of research on providing cognitive support by automating idea organization task in GSS. To help achieve the goal of automating idea organization task in GSS, this research designs and tests a SVD-enabled system to automate the often problematic idea organization task in GSS. Specifically, this research is conducted to examine how the idea organization task, typically regarded as the most labor-intensive, time-consuming and cognitively demanding in group problems solving, can be automated using a system enabled by singular value decomposition algorithm, and how the automatic idea organization approach is compared with a manual approach done by humans in terms of performance. For the purpose of evaluation, we compare the performance of our automated approach using SVD algorithm against that of human subjects. Two separate experiments are conducted which deal with evaluating the performance of the automatic approach on two essential components of an idea organization; namely, generation of topic categories and placement of relevant comments into their respective categories.

We can conclude from this research that important progress can be made in providing cognitive support to idea organization task in GSS with an automatic approach. In particular, the output of our research, an SVD-enabled idea organization system tool, is proven to be able to automatically accomplish idea organization task in terms of creating topic categories and placing relevant comments into their respective topics with performance comparable to human performance levels. Further conclusions are discussed below. This research contributes to the solution of managing the complexities involved with cognitively processing large volume of noisy textual brainstorming data in idea organization by automatically suggesting a list of topic categories with each category followed by a list of relevant comments.

8.1 Conclusions

This research demonstrates how an automatic approach with SVD algorithms can be used to perform an intelligent idea organization task including generation of topic categories and placement of relevant comments into their respective topics from brainstorming comments data produced in GSS sessions. Two separate experiments are conducted to evaluate the performance of our automatic idea organization system. The brainstorming comments data used in both experiments are collected from two group sessions in GSS organized to tackle the same brainstorming question of finding ways to improve a student union center.

The results of this research demonstrate that our automatic idea organization approach can be achieved both efficiently and effectively.

In terms of efficiency, the system completes the generation of topic categories and the placement of relevant comments into their respective categories in no less than 1 second. Comparatively, the time spent by individual for finishing both the task of topic category generation and comments placement ranges from 40 minutes to more than one hour.

However, efficiency alone cannot make automatic idea organization feasible unless a satisfactory level of performance is reached. With respect to effectiveness, we measure the system's performance in terms of recall and precision in two separate experiments covering both topic category generation and comment placement.

To evaluate the system's performance in terms of topics generated, we first have the system automatically generate 8 topics. Then 9 undergraduates subjects work as domain experts to manually produce their own lists of topics. With the output from the system and the 9 undergraduate subjects, the measure of performance that we use is to compare the system's list of topics with the 9 lists of topics generated separately by 9 undergraduate subjects. The comparison is conducted by two groups of human subjects.

The first group consists of 10 undergraduate subjects different from those who generate the topic lists earlier. Specific conclusions from this experiment include: In term of general perception of performance, system's topic list is ranked the best by 4 out of 10 human subjects, more frequently than any other lists. System's topic list has the highest average topic recall although statistically topic recall of the system is not different from that of other 9 lists. More promisingly, topic recall of the system significantly outperforms that of other 5 lists out of a total of 9 human lists. Similarly, system's topic list has the highest average topic precision although statistically topic precision of the system is not different from that of the other 9 lists. Topic precision by the system also significantly outperforms that of other 2 lists.

The second group consists of 4 expert facilitators. Specific conclusions drawn from this experiment include: In terms of general perception of performance, system's topic list is ranked in the second and the third best out of 10 lists by 2 of expert facilitators respectively, but ranked the second worst by the other two expert facilitators. System's topic list has the third highest average topic recall although statistically topic recall of the system is not different from that of other 9 lists. A pairwise comparison of topic recall of the system with that of other 9 lists does not reveal any significant difference. Similarly, system's topic precision has the third highest average topic precision although statistically topic precision of the system is not different from that of other 9 lists. A pairwise comparison of topic precision of the system with that of other 9 lists also does not indicate any significant difference.

To evaluate the system's performance in terms of comment placement, we first have the system automatically place relevant comments into the 8 topics earlier generated by the system. Then 5 undergraduates subjects work as domain experts to manually place what they believe relevant comments into their respective topics out of the 8 system-generated topics. With the output from the system and the 5 undergraduate subjects, the measure of performance that we use is comparing the system's list of comment placement with the 5 lists of comment placement generated separately by 5 undergraduate subjects. The comparison is conducted by a group of 16 undergraduate subjects, who are different from the 5 subjects who earlier generate the comment placement. Specific conclusions from this experiment include: in terms of general perception of performance, system's comment placement list is ranked the best out of 6 lists by 3 human subjects, which is a fairly good evaluation though not as good as that of list 6, which is ranked the best by 7 human subjects. System's comment placement list has the second worst average placement recall although statistically placement recall of the system is not different from that of other 5 lists. A pairwise comparison of placement recall of the system with that of other 5 lists does not reveal any significant difference. Interestingly, system's placement precision has the highest average placement precision although statistically placement precision of the system is not significantly different from that of other 5 lists. A pairwise comparison of placement precision of the system with that of other 5 lists shows that the system's comment placement list outperforms one human list out of a total of 5 human lists.

To sum up, the research demonstrates that regardless of who the human subjects are, there is no significant difference among all the lists including human-generated lists and system-generated list in terms of topic recall and topic precision. Given the limited number of human subjects, no substantial claims could be made about this statistical result. Particularly, when evaluated by undergraduate subjects, the topic recall and topic precision of the system even significantly outperforms several human lists. When evaluated by expert facilitators, the relatively moderate performance of system in terms of topic recall and topic precision may be because there are only 4 expert facilitators available as compared with 10 undergraduate subjects. Again, regardless of what groups of human subjects are, there is no significant difference among all the lists including human-generated lists and system-generated list in terms of placement recall and placement precision. Given the limited number of human subjects, no substantial claims could be made about this statistical result.

8.2 Contributions

The principal contribution of this research is the creation of an IT artifact: an SVD-enabled automatic idea organization tool that can facilitate idea organization in group problem solving by automatically generating topic categories and placing comments into their respective topic categories. This is in line with the objective of design science in that our research extends the boundaries of human and organizational capabilities by creating new and innovative artifacts (Hevner, March et al. 2004).

The success of future GSS may well depend on partially or completely automating individual tasks to free group participants from various complex tasks such as idea organization in GSS environment. The proposed system greatly reduces the complexity of consolidating large volume of brainstorming output into manageable and meaningful topics attached with relevant comments. Therefore, our automatic approach may provide a useful and promising direction for future GSS research.

As demonstrated by a series of experiments, the system accomplishes idea organization both efficiently and effectively. As a result, it is seen that the efficiency and effectiveness of the system could provide a useful contribution to the group collaboration literature in terms of applying automatic approaches to specific tasks within a group collaboration process.

This system can act as an auxiliary intelligent tool to automate idea organization in the process of group problem solving, a useful addition to the repertoire of tools in GSS. With the assistance provided by our system, group participants will find it much easier to understand, internalize and adapt a system-suggested topic list validated with relevant comments attached than to start from scratch in organizing the raw brainstorming comments. This new dimension to traditional GSS meetings provides a useful contribution to the group problem solving process that can be applied automatically in a consistent way.

8.3 Limitations and Future Direction

This research study is subject to several limitations and future study needs to be conducted to overcome these limitations.

The experiments used in this study reveal several limitations. The participants in the experiments were sampled from both student and expert facilitator populations. The results from these participants may not generalize well to other populations for several reasons. The students possess limited experience and knowledge regarding collaborative work in a professional setting. Given the complexity of the tasks involved in the experiments, expert facilitators and undergraduate subjects may not be motivated enough to induce high levels of cognitive effort in the experimental tasks. In addition, the study used a small sample size, limiting the power of the findings. The number of subjects used ranges from 4 expert facilitators used in stage 3 of experiment 1 to 16 undergraduate subjects used in stage 2 of experiment 2. The small sample size may lead to the limited statistical power in the results of the experiments. Therefore, future studies need to have a more representative sample from subjects and recruit more subjects to increase the statistical power.

The configurations set in the system also have some limitations. For example, the results from the second experiment demonstrate that the system at current configurations is much better at placement precision than placement recall. Since the very purpose of comment placement is to validate the content of the topic, it is preferable to have better placement precision than placement recall. However, it remains to be seen whether more placement recall can be achieved without sacrificing placement precision. In the future, it may be worthwhile to tweak the system's configuration in terms of the threshold value for comment placement in order to achieve the best possible trade off between placement recall and placement precision. In addition, the experiment is designed not to allow overlapping of topics for any comments; that is, one comment can only be placed into one topic instead of multiple topics. While reducing the cognitive load of human subjects, not allowing overlapping topics may not reflect the reality of idea organization where one comment may be placed into multiple topics. Therefore, it may be worthwhile to simulate the real world situation in future experiments to allow for overlapping of topics.

System-generated output has its limitations in terms of helping group participants better understand the relationships among the concepts. On one hand, the output from the system can potentially serve as a starting point to help group participants get a better understanding of relationships among the concepts than to organize ideas from scratch. For example, the system-produced output provides a consolidated set of key topics so that group participants could get a general understanding of the relationships among them and how the brainstorming question are related to these key topics. With relevant comments placed topics, group participants could immediately see the linkage between the topics and underlying comments. In addition, browsing and understanding the relationships among the concepts from brainstorming comments exerts considerable cognitive demand from users. Therefore, the efforts and cognitive power saved from the initial round of manual idea organization could be better spent on refining the system-produced output and thus achieve a better understanding of the relationships among the concepts. On the other hand, while system-generated results have the potential to help group participants better understand the relationships among concepts, it is also the case that the act of starting from scratch to organize and discuss around group organizing could invoke deeper learning about relationships among concepts. Therefore, future studies may help untangle the impact of system-generated results as compared with manual processing on the understanding of relationships among concepts.

The system does not have a user-friendly user interface. We are planning in the future to develop an interface to our automatic idea organization tool and have results displayed visually in front of all group participants. In that way, participants can visually see the result and make further refinements by revising topic labels and adjusting the placement of comments into topic categories. In addition, rather than use our system as a separate tool, we are also planning to embed our system into the commercial product of GroupSystems or ThinkTank2.0. It is expected that the graphical display and integration of our system to GroupSystems for user interactions will further enhance the usefulness and performance of GSS, thus continuing its wide-spread adoption among collaborative participants.

APPENDIX A EXPERIMENT 1 – STAGE 1: INSTRUCTION – MANUAL TOPIC GENERATION

Idea Organization Experiment

Background

Brainstorming session in group support systems (GSS) is used to generate as many ideas as possible to tackle a specific problem. Idea comments generated from a brainstorming task in GSS need to be organized into general topic categories before users can rank order them and decide on final options to solve the problem. This involves two steps

- Generate categories which summarize important ideas out of all comments;
- Assign relevant comments to their respective categories.

These two steps of organizing idea comments into general topic categories is called *idea organization* in GSS.

Objective of the Experiment

Given a set of brainstorming output, please generate a maximum of 8 topic categories that can capture important ideas from the brainstorming output.

Instruction:

What you will see is a list of 138 comments generated from a brainstorming session where participants were college students who collaborated in an electronic meeting environment using GroupSystems. The question posed to them is – What can we do to make our student union center a better place in the university? All participants brainstormed and generated a list of 138 comments which stated how they would make the student union center a better place in the university.

The goal of the categorization process was to derive the lists of topic categories for use in rank order vote. The voting issues will be "In what order should we adopt these ways to make thee student union center a better place?" Assume that you are a facilitator and will lead a group through the voting. Think about the criteria that you will use to evaluate the lists. As a facilitator, what makes a good list?

Read through the list of comments generated from the brainstorming session. Then complete the following two tasks

1. Please read the comments carefully and come up with a maximum of 8 categories (Please do NOT exceed 8 categories) to summarize important ideas out of all the 138 comments. Please use single words or phrases in the comments for topic category names
2. After you are finished with generating topic categories, please record the time you spend for this task

Here is an example of what you should do.

Example Question: What are some of the ways that you can think of to solve the parking problems in the university?

Brainstorming Comments Data:

1. I guess we need to build more garages, as simple as that
2. use existing space more efficiently
3. get more garages
4. we should make efficient use of existing space
5. I don't know. There are no solutions

Your Response:

- Topic categories generated:
    - Topic 1: garages;
    - Topic 2: efficient use of existing space
- Time spent: 1 hour

APPENDIX B EXPERIMENT 1 – STAGE 1: FORM – MANUAL TOPIC GENERATION

Idea Organization Experiment

Before you begin, please first record your start time.
Start Time: ____

Topic Categories Generated (A Maximum of 8 Categories allowed, You do not have to generate all 8 categories)

Topic Category 1: ____________________________________

Topic Category 2: ____________________________________

Topic Category 3: ____________________________________

Topic Category 4: ____________________________________

Topic Category 5: ____________________________________

Topic Category 6: ____________________________________

Topic Category 7: ____________________________________

Topic Category 8: ____________________________________

Please record your finish time.
Finish Time: ____

APPENDIX C EXPERIMENT 1 – STAGE 2 & 3: INSTRUCTION – EVALUATION OF TOPIC CATEGORIES

Idea Organization Experiment

Background
Brainstorming session in group support systems (GSS) is used to generate as many ideas as possible to tackle a specific problem. Idea comments generated from a brainstorming task in GSS need to be organized into general topic categories before users can rank order them and decide on final options to solve the problem. This involves two steps

- Generate categories which summarize the majority of comments;
- Assign relevant comments to their respective categories.

These two steps of organizing idea comments into general topic categories is called *idea organization* in GSS.

Objective of the Experiment
Given a set of brainstorming output and 10 lists of topic categories generated from the 138 comments, objectively evaluate the quality of these lists of topic categories.

Instruction:
What you will see is a list of 138 comments generated from a brainstorming session where participants were college students who collaborated in an electronic meeting environment using GroupSystems. The question posed to them is – What can we do to make our student union center a better place in the university? All participants brainstormed and generated a list of 138 comments which stated how they would make the student union center a better place in the university.

What you will also see is 10 lists of topic categories generated to summarize the important ideas from the 138 comments. Any individual topic category list consists of up to 8 topic categories. These topic categories will be used subsequently in the original meeting for rank order vote. The voting issues will be "In what order should we adopt these ways to make thee student union center a better place in the university?" When deriving topic categories, participants think about the criteria that they will use to evaluate the lists.

Here is an example of how a list of topic categories is generated

Example Question: What are some of the ways that you can think of to solve the parking problems in the university?

Brainstorming Comments Data:

1. I guess we need to build more garages, as simple as that
2. use existing space more efficiently

3. get more garages
4. we should make efficient use of existing space
5. I don't know. There are no solutions

Result

- Topic categories generated:
    - Topic 1: garages;
    - Topic 2: efficient use of existing space

Here is what you should do in this particular experiment. In this experiment, you are asked to evaluate each of the 10 lists of topic categories already generated to summarize the important ideas from the 138 comments.

You will read through the 138 brainstorming comments, and alter these lists by adding, deleting or editing specific topic categories in the list to make them better. Think about the criteria that you will use to evaluate the lists. Browse through the printout of the brainstorming session. Feel free to take notes if you wish. After you are finished of correcting all 10 lists, please rank order these 10 lists to in terms of their quality.

You are asked to alter each list (adding/deleting topic categories) to make them better. Please treat each line item on each list as a topic category. The following are three activities that you should perform for each of following 10 topic category lists. Please make sure that after altering each list, the resulting list should have a maximum of 8 topic categories.

1. Delete inappropriate topic categories, add missing topic categories
    - Delete: Please *cross out* categories that you believe that are inappropriate and not representing important topic categories for the 138 comments.
    - Add: Please *add* new category to the original list if you think there are important topic categories not shown in the topic category list.
    - Edit: An *edit* action (cross out one topic category and change it to another topic category) is treated as a deletion followed by addition.
2. Put the following lists in a rank order with 1 being the best list of topic categories.

Here is an example of what you should do in this experiment
Example Question: What are some of the ways that you can think of to solve the parking problems in the university?

Brainstorming Comments Data:

6. I guess we need to build more garages, as simple as that
7. use existing space more efficiently
8. get more garages

9. we should make efficient use of existing space
10. change the class times to spread out the traffic
11. we can arrange for the classes in different buildings and in different times so that parking space is not that crowded.
12. I don't know. There are no solutions

Topic Category List 1:

- More garages
- Parking space

Topic Category list 2:

- Garages
- arrangement

Your Response:

Topic Category List 1:	New Topic Categories (if any):
• More garages • Parking space •	Change class times ________________ ________________ ________________ ________________

Topic Category List 2:	New Topic Categories (if any):
• garages ~~• arrangement~~	Parking space Change class times ________________ ________________ ________________ ________________ ________________ ________________ ________________

Please rank the above list in terms of its topic category quality with 1 being best and 10 being the worst

• 1: List No.____2_____ • 2: List No.____1_____	

APPENDIX D EXPERIMENT 1 – STAGE 2 & 3: FORM – EVALUATION OF TOPIC CATEGORIES

Idea Organization Experiment

The following are three activities that you should perform for each of following 10 topic category lists.

- Delete: Please <u>*cross out*</u> categories that you believe that are not representing important topic categories for the 138 comments.
- Add: Please <u>*add*</u> new category to the original list if you think there are important topic categories not shown in the topic category list.
- Edit: An *edit* action is treated as a deletion followed by addition.

After altering each of the 10 topic category lists, please remember to rank the all 10 lists in terms of its topic category quality with 1 being best and 10 being the worst.

<u>IMPORTANT: Please note that the final list should contain a Maximum of 8 categories.</u>

<u>Topic Category List 1:</u>	New Topic Categories (if any):
• one-stop place • increase technology • leadership/mentor laboratories • commercialization • social events • adaptable	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

<u>Topic Category List 2:</u>	New Topic Categories (if any):
• one stop food and services • high quality but personal • efficient and effective at meeting students needs • accessible • cater for student meeting, activities, growth • technologically advanced	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

<u>Topic Category List 3:</u>	New Topic Categories (if any):
• one-stop shop/centralization • up-to-date/new technology • commercialization • more programming/services • maintain values (learning and development)	________________ ________________ ________________ ________________ ________________ ________________

• re-evaluate staffing	________________ ________________

Topic Category List 4:	New Topic Categories (if any):
• investment • entrepreneurship • accessibility • diversity • communication • technology • security • openness	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 5:	New Topic Categories (if any):
• one-stop shop • commercialization • technology • viewpoint neutrality • adaptable • decentralization	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 6:	New Topic Categories (if any):
• community • technology • commercial • adaptable • one-stop shop • efficient • extra-curricular • irrelevant	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 7:	New Topic Categories (if any):
• one stop shop • a campus living room • union funding • service variety • change with the times	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 8:	New Topic Categories (if any):
• technologically advance • adaptable • commercialization • diverse community/cultural center • unify student services • student development center • "face of the university"	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 9:	New Topic Categories (if any):
• greater demand for wireless • transportation of goods to students • facility for citizenship • 24 hour service with assistance • more rooms to study • more active with students teachers • construction for more space • more career showcase at various hours	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Topic Category List 10:	New Topic Categories (if any):
• Central Location • Community Center • Programs and Services • Technology • Space for Students • Academic Affairs • Viewpoint Neutral • Commercial Vendor	________________ ________________ ________________ ________________ ________________ ________________ ________________ ________________

Please rank the above list in terms of its topic category quality with 1 being best and 10 being the worst

• 1: List No.________ • 2: List No.________ • 3: List No.________ • 4: List No.________ • 5: List No.________	• 6: List No.__________ • 7: List No.__________ • 8: List No.__________ • 9: List No.__________ • 10:List No.__________

APPENDIX E EXPERIMENT 2 – STAGE 1: INSTRUCTION – MANUAL COMMENT PLACEMENT

Idea Organization Experiment

Background

Brainstorming session in group support systems (GSS) is used to generate as many ideas as possible to tackle a specific problem. Idea comments generated from a brainstorming task in GSS need to be organized into general topic categories before users can rank order them and decide on final options to solve the problem. This involves two steps

- Generate categories which summarize the majority of comments;
- Assign relevant comments to their respective categories.

These two steps of organizing idea comments into general topic categories is called *idea organization* in GSS.

Objective of the Experiment

Given a set of brainstorming output and 8 existing categories, please assign relevant comments to their respective categories.

Instruction:

What you will see is a list of 138 comments generated from a brainstorming session where participants were college students who collaborated in an electronic meeting environment using GroupSystems. The question posed to them is – What can we do to make our student union center a better place in the university? All participants brainstormed and generated a list of 138 comments which stated how they would make the student union center a better place in the university.

What you will also see is a separate list of 8 existing categories. These 8 categories are generated to summarize the important ideas out of 138 comments.

Read through the list of comments generated from the brainstorming session. Please complete the following:

- For each category, identify relevant comments and assign these relevant comments (comment number) to their respective categories. One comment can only be placed under one topic. Please note that for comments that do not belong to any of the existing 8 categories, do NOT put them in any categories.
- After you are finished with assigning comments to their respective categories, please record the time you spend for this task

Here is an example of what you should do.

Example Question: What are some of the ways that you can think of to solve the parking problems in the university?

Brainstorming Comments Data:

1. I guess we need to build more garages, as simple as that
2. use existing space more efficiently
3. get more garages
4. we should make efficient use of existing space
5. I don't know. There are no solutions

Topic Categories

- Topic 1: more garages;
- Topic 2: efficient use of existing space

Your Response:

- Assignments Made:
  - Topic 1: more garages
    - comment 1,
    - comment 3
  - Topic 2: efficient use of existing space
    - comment 2,
    - comment 4
- Time spent: 1 hour

APPENDIX F EXPERIMENT 2 – STAGE 1: FORM – MANUAL COMMENT PLACEMENT

Idea Organization Experiment

The following are the 8 existing topic categories:

Topic 1: Community Center Topic 2: Programs and Services Topic 3: Technology Topic 4: Space for Students	Topic 5: Academic Affairs Topic 6: Viewpoint Neutral Topic 7: Commercial Vendor Topic 8: Central Location

Instruction: Please put the comment number of each relevant comment under its respective topic category. For comments that do not belong to any of the existing 8 categories, do NOT put them in any categories.

Before you begin, please first record your start time

Start Time: ___________________________________

Topic Category 1: Community Center

Topic Category 2: Programs and Services

Topic Category 3: Technology

Topic Category 4: Space for Students

Topic Category 5: Academic Affairs

Topic Category 6: Viewpoint Neutral

Topic Category 7: Commercial Vendor

Topic Category 8: Central Location

Please record your finish time.
Finish Time: ____

APPENDIX G EXPERIMENT 2 – STAGE 2: INSTRUCTION – EVALUATION OF COMMENT PLACEMENT

Idea Organization Experiment

Background

Brainstorming session in group support systems (GSS) is used to generate as many ideas as possible to tackle a specific problem. Idea comments generated from a brainstorming task in GSS need to be organized into general topic categories before users can rank order them and decide on final options to solve the problem. This involves two steps

- Generate categories which summarize the majority of comments;
- Assign relevant comments to their respective categories.

These two steps of organizing idea comments into general topic categories is called *idea organization* in GSS.

Objective of the Experiment

Given a set of 138 brainstorming comments output, 8 existing topic categories generated from the 138 comments to summarize important ideas, and 6 comments assignment lists that are obtained by assigning relevant comments to their respective topic, please objectively evaluate the quality of the 6 comments assignment lists.

Instruction:

What you will see is a list of 138 comments generated from a brainstorming session where participants were college students who collaborated in an electronic meeting environment using GroupSystems. The question posed to them is – What can we do to make our student union center a better place in the university? All participants brainstormed and generated a list of 138 comments which stated how they would make the student union center a better place in the university.

What you will also see on a separate page are 8 existing categories generated to summarize the important ideas out of 138 comments.

What you will also see are 6 different comments assignment lists where comments are placed under their supposedly appropriate categories out of the 8 existing categories by different individuals. Within each list, in addition to the 8 existing categories, there is another special category "other categories", which is located at the end of each comment assignment list. "other categories" is used to place comments that do not belong to any of the 8 existing categories.

For each of the 6 comments assignment lists, please make revisions to make sure that each comment is placed into its correct categories. Specifically, you are asked to perform the following activities.

First, evaluate comments under each of the 8 categories and do the following

- If you feel that a comment is irrelevant to the category under which it is placed, please cross it out and put its ***comment number*** under the category which you think are relevant category. In addition, if you think that comment does not belong to any of the 8 categories, place its ***comment number*** under "other categories”, which is located at the end of each comment assignment list. During editing, comments can be added to or delete from a particular category. One comment can only be placed under one topic.

Second, evaluate comments under the “Other categories”, which is located at the end of each comment assignment.

- If you think that any of the comments under “other categories” should belong to one of the 8 existing categories, please cross it out under “other categories” and place its ***comment number*** under the category which you think is relevant.

<u>Here is an example of what you should do in this experiment</u>
Example Question: What are some of the ways that you can think of to solve the parking problems in the university?

Brainstorming Comments Data:

13. I guess we need to build more garages, as simple as that
14. use existing space more efficiently
15. get more garages
16. we should make efficient use of existing space
17. change the class times to spread out the traffic
18. we can arrange for the classes in different buildings and in different times so that parking space is not that crowded.
19. I don’t know. There are no solutions

Existing 3 categories:

- Topic 1: More garages
- Topic 2:Efficient use of existing space
- Topic 3: Make arrangement for classes

<table>
<tr>
<td>Placement List 1:
o Topic 1: more garages
▪ comment 1,
▪ comment 4
o Topic 2: efficient use of existing space
▪ comment 2,
▪ comment 3
o Topic 3: Make arranges for</td>
<td>Placement List 2:
o Topic 1: more garages
▪ comment 1,
▪ comment 3
o Topic 2: efficient use of existing space
▪ comment 2,
▪ comment 4
o Topic 3: Make arranges for</td>
</tr>
</table>

classes ▪ Comment 5 o Other Categories ▪ Comment 6 ▪ Comment 7	classes ▪ Comment 5 o Other Categories ▪ Comment 6 ▪ Comment 7

<u>Your response</u>

<u>Placement List 1</u> o Topic 1: more garages ▪ comment 1, ▪ ~~comment 4~~ <u>new comments</u>: • comment 3 o Topic 2: efficient use of existing space ▪ comment 2, ▪ ~~comment 3~~ <u>new comments</u>: • comment 4 o Topic 3: Make arranges for classes ▪ Comment 5 <u>new comments</u>: • Comment 6 o Other Categories ▪ ~~Comment 6~~ ▪ Comment 7	<u>Placement List 2</u>: o Topic 1: more garages ▪ comment 1, ▪ comment 3 o Topic 2: efficient use of existing space ▪ comment 2, ▪ comment 4 o Topic 3: Make arranges for classes ▪ Comment 5 <u>new comments</u>: • Comment 6 o Other Categories ▪ ~~Comment 6~~ ▪ Comment 7

Please rank the all 2 comments assignment lists in terms of its comment assignment quality with 1 being best and 2 being the worst

- 1: Placement List No.____2____
- 2: Placement List No.____1____

APPENDIX H EXPERIMENT 2 – STAGE 2: FORM – EVALUATION OF COMMENT PLACEMENT

Idea Organization Experiment

First, please evaluate comments under each of the 8 categories and do the following

- If you feel that a comment is irrelevant to the category under which it is placed, please cross it out and put its ***comment number*** under the category which you think are relevant category. In addition, if you think that comment does not belong to any of the 8 categories, place its ***comment number*** under "other categories", which is located at the end of each comment assignment list. During editing, comments can be added to or delete from a particular category.

Second, please evaluate comments under the "Other categories", which is located at the end of each comment assignment.

- If you think that any of the comments under "other categories" should belong to one of the 8 existing categories, please cross it out under "other categories" and place its ***comment number*** under the category which you think is relevant.

Third, Please rank the all 6 comments assignment lists in terms of its comment assignment quality with 1 being best and 6 being the worst

- 1: Placement List No._________
- 2: Placement List No._________
- 3: Placement List No._________
- 4: Placement List No._________
- 5: Placement List No._________
- 6: Placement List No._________

REFERENCES

Aakhus, M., M. Adkins, et al. (1997). Layers of learning: Facilitation in the distributed classroom. The Thirtieth Hawaii International Conference on Systems Sciences.

Adkins, M., M. Burgoon, et al. (2003). "Using group support systems for strategic planning with the United States Air Force." Decision Support Systems 34(3): 315-337.

Anderberg, M. R. (1973). Cluster analysis for applications, Academic Press.

Anson, R., R. Bostrom, et al. (1995). "An experiment assessing group support system and facilitator effects on meeting outcomes." Management Science 41(2): 189-208.

Baeza-Yates, R. and B. Ribeiro-Neto (1999). Modern Information Retrieval, ACM Press.

Berry, M. W., Z. Drmac, et al. (1999). "Matrices, Vector Spaces, and Information Retrieval." SIAM Rev 41(335-362).

Berry, M. W., S. T. Dumais, et al. (1995). "Using Linear Algebra for Intelligent Information Retrieval." SIAM Rev.: 537-595.

Boley, D., M. Gini, et al. (1999). "Partitioning-based clustering for web document categorization." Decision Support Systems 27(3): 329-341.

Bostrom, R., R. Anson, et al. (1993). Group Facilitation and Group Support Systems. Group Support Systems: New Perspectives. L. M. J. a. J. S. Valacich. New York, N.Y., MacMillan: 146-168.

Briggs, R. O. (1994). The team theory of group productivity and its application to the development and testing of electronic group support technology. Unpublished doctoral dissertation, University of Arizona.

Briggs, R. O., G. J. De Vreede, et al. (2001). ThinkLets: achieving predictable, repeatable patterns of group interaction with group support systems (GSS). The 34th Annual Hawaii International Conference on System Sciences.

Briggs, R. O., B. A. Reinig, et al. (1997). Quality as a function of quantity in electronic brainstorming. the Thirtieth Hawaii International Conference on Systems Sciences.

Briggs, R. O., G. J. d. Vreede, et al. (2003). "Collaboration Engineering with ThinkLets to Pursue Sustained Success with Group Support Systems." Journal of Management Information Systems 19(4): 31-63.

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

Chen, H., P. Hsu, et al. (1994). "Automatic Concept Classification of Text from Electronic Meetings." Communications of the ACM 37(10).

Chen, H., C. Schuffels, et al. (1996). "Internet categorization and search: a self-organizing approach." Journal of Visual Communication and Image Representation 7(1): 88-102.

Couger, J. D. (1995). Creative Problem Solving and Opportunity Finding. Danvers, MA, Boyd & Fraser Publishing Company.

Cutting, D. R., D. R. Karger, et al. (1992). "Scatter/gather: A cluster-based approach to browsing large document collections." the Fifteenth Annual International ACM Conference on Research and Development in Information Retrieval: 318-329.

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

Davison, R. M., and Briggs, R.O. (2000). "GSS for presentation support." Communications of the ACM 43(9): 91-97.

De Vreede, G. J. (1997). Support for collaborative design. Thirtieth Annual Hawaii International Conference on System Sciences.

De Vreede, G. J., R. M. Davison, et al. (2003). "How a Silver Bullet May Lose its Shine." Communications of the ACM 46(8): 96-101.

De Vreede, G. J., D. R. Vogel, et al. (2003). Fifteen years of in-situ GSS use: A comparison across time and national boundaries. the Thirty-Sixth Hawaiian International Conference on System Sciences, Alamitos, CA, IEEE Computer Society Press.

Defays, D. (1977). "An efficient algorithm for the complete link method." The Computer Journal 20(364-366).

Dennis, A., J. F. George, et al. (1988). "Information technology to support group electronic meetings." MIS Quarterly 12(4): 591-624.

Dennis, A. R. and J. S. Valacich (1999). "Rethinking media richness: towards a theory of media synchronicity." the 32nd Annual Hawaii International Conference on System Sciences.

Dennis, A. R., J. S. Valacich, et al. (1990). "An experimental investigation of the effects of group size in an electronic meeting environment." IEEE Transactions on Systems, Man and Cybernetics 20(5): 1049-1057.

Dennis, A. R. N., J.F.,Jr.; Vogel, D.R. (1990-91). "A comparison of laboratory and field research in the study of electronic meeting systems." Journal of Management Information Systems 7(3): 107-135.

DeSanctis, G. and R. B. Gallupe (1987). "A foundationfor the study of group decision support systems." Management Science 33(5): 589-609.

Dhillon, I. S. (2001). Co-clustering documents and words using Bipartite Spectral Graph Partitioning. UT CS Technical Report TR2001-05 20,.

Dickson, G. W., Partridge, J.-E. L. and Robinson, L. H. (1993). "Exploring Modes of Facilitative Support for GDSS Technology." MIS Quarterly 17(2): 173-194.

Diehl, M. and W. Stroebe (1987). "Productivity loss in brainstorming groups: toward the solution of a riddle." J. Personality and Social Psychology 53(3): 497-509.

El-Hamdouchi, A. and P. Willett (1986). Hierarchic document clustering using Ward's method. the Ninth International Conference on Research and Development in Information Retrieval, Washington, ACM.

Ellis, C. A., S. J. Gibbs, et al. (1991). "Groupware: some issues and experiences." Communications of the ACM 34(1 January): 38-58.

Everitt, B. S. (1974). Cluster Analysis. New York, John Wiley & Sons, Inc.

Fayyad, U. M., G. Piatetsky-Shapiro, et al. (1996). Advances in Knowledge Discovery and Data Mining. Menlo Park, CA, AAAI Press/MIT Press.

Fjermestad, J. H., and Hiltz, S. R. (1997). "Experimental studies of group decision support systems: an assessment of variables studied and methodology,." Proceedings of the Thirtieth Hawaii International Conference on System Sciences II: 49-53.

Fjermestad, J. H., and Hiltz, S. R. (1998-99). "An assessment of experimental research of group support systems: methodology and results." Journal of Management Information Systems 15(3): 7-149.

Frawley, W. J., G. Piatetsky-Shapiro, et al. (1991). Knoweldge Discovery in Databases: An Overview, AAAI Press.

Furnas, G. W., T. K. Landauer, et al. (1987). "The Vocabulary Problem in Human-System Communication." Communications of the ACM 30(11): 964-971.

Gallupe, R. B., L. M. Bastianutti, et al. (1991). "Unblocking brainstorms." Journal of Applied Psychology 76(1): 137-142.

Gallupe, R. B., A. R. Dennis, et al. (1992). "Electronic brainstorming and group size." Academy of Management Journal, 35(2): 350-369.

George, J. F., A. R. Dennis, et al. (1992). "An experimental investigation of facilitation in an EMS decision room." Group Decision and Negotiation 1: 57-70.

Golub, G. and W. Kahan (1965). "Calculating the singular values and pseudo-inverse of a matrix.". Soc. Indust. Appl. Math. Ser. B Numer. Anal. 2: 205-224.

Griffith, T. L., M. A. Fuller, et al. (1998). "Facilitator Influence in group support systems: Intended and unintended effects." Information Systems Research 9(1): 20-36.

Grohowski, R. B., C. McGoff, et al. (1990). "Implementing electronic meeting systems at IBM: Lessons learned and success factors." MIS Quarterly 14(4): 369-383.

Hackman, J. R. and R. E. Kaplan (1974). "Interventions into the group process: an approach to improving the effectiveness of groups." Decision Science 5: 459-480.

Hammouda, K. M. and M. S. Kamel. (2004). "Efficient phrase-based document indexing for web document clustering." IEEE Transactions on knowledge and data engineering 16(10): 1279-1296.

Haykin, S. (1990). Neural Networks: A Comprehensive Foundation. Upper Saddle River, NJ, Prentice Hall.

Hevner, A. R., S. T. March, et al. (2004). "Design Science in Information Systems Research." MIS Quarterly 28(1): 75-105.

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

Jablin, F. M. and D. R. Seibold (1978). "Implications for problem solving groups of empirical research on 'brain-storming': a critical review of the literature." The Southern States Speech Communication J. 43: 327-356.

Jain, A. K. and R. C. Dubes (1988). Algorithms for Clustering Data. Upper Saddle River, NJ., Prentice-Hall, Inc.

Jain, A. K., M. N. Murty, et al. (1999). "Data clustering: a review." ACM Computing Surveys 31(2).

Jessup, L. M. and J. S. Valacich (1993). On the study of group support systems: An introduction to group support system research and development. Group Support Systems: New Perspectives. L. M. Jessup and J. S. Valacich. New York, NY, Macmillan Publishing Company.

Karypis, G., E. H. Han, et al. (1999). "CHAMELEON:A Hierarchical Clustering Algorithm Using Dynamic Modelling." IEEE Computer 32(8): 68- 75.

Khalifa, M., Davison, R. and Kwok, R. C.-W. (2002). "The Effects of Process and Content Facilitation Restrictiveness on GSS-mediated Collaborative Learning." Group Decision and Negotiation 11(5): 345-361.

Kohonen, T. (1989). Self-Organization and Associative Memory. Berlin, Springer-Verlag.

Koniger, P. and K. Jonowitz (1995). "Drowning in information, but thirsty for knowledge." International Journal For Information Management 15(1): 5-16.

Lin, X., D. Soergei, et al. (1991). A self-organizing semantic map for information retrieval. the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Chicago, IL USA.

Maier, N. R. F. and L. R. Hoffman (1960). "Quality of first and second solutions in group problem solving." Journal of Applied Psychology 44: 278-283.

Martz, W. B., D. R. Vogel, et al. (1992). "Electronic meeting systems: Research in the field." Decision Support Systems 8: 141-158.

Merkl, D. (1998). Text Data Mining. A handbook of natural language processing: techniques and applications for the processing of language as text. R. Dale, Moisl, H., Somers, H. New York, Marcel Dekker.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information." Psychological Review 63(1): 81-97

Mintzberg, H. (1983). The Adhocracy. Structures in five, Prentice-Hall International ed.,**:** 253-281.

Murtagh, F. (1983). "A survey of recent advances in hierarchical clustering algorithms." The Computer Journal 26: 354-359.

Niederman, F. a. V., R. J. (1999). "The Effects of Facilitator Characteristics on Meeting Preparation,Set Up, and Implementation." Small Group Research 30(3): 330-360.

Nunamaker, J., A. Dennis, et al. (1991). "Electronic Meeting Systems To Support Group Work." Communications of the ACM 34(July): 40-61.

Nunamaker, J. F., A. R. Dennis, et al. (1991). "Information technology for negotiating groups generating options for mutual gain." Management Science 37(10): 1325-1346.

Nunamaker, J. F., Jr. , R. O. Briggs, et al. (1996). "Lessons from a dozen years of group support systems research: A discussion of lab and field." Journal of Management Information Systems 13(3): 163-207.

Nunamaker, J. F., C. Jr., M., et al. (1991). "Systems development in information systems research." Journal of Management Information Systems 7(3): 89-106.

Nunamaker, J. F., D. Jr., A. R., et al. (1991). "Electronic meeting systems to support group work." Communications of the ACM 34(7): 40-61.

Nunamaker, J. F., R. O. Jr.; Briggs, et al. (1996). "Lessons from a decade of group support systems research." Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences 3: 418 - 427.

Orwig, R. E. (1995). A graphical, self-organizing approach to classifying electronic meeting output. Dept. of MIS, University of Arizona.

Orwig, R. E., H. Chen, et al. (1997). "A graphical, self-organizing approach to classifying electronic meeting output." Journal of the American Society for Information Science 48(2): 157-170.

Osinski, S., J. Stefanowski, et al. (2004). Lingo: Search results clustering algorithm based on singular value decomposition. Intelligent Information Systems. M. A. Klopotek, S. T. Wierzchon and K. Trojanowski. Springer, Advances in Soft Computing: 359-368.

Phillips, L. D. and M. C. Phillips (1993). "Facilitated work groups: Theory and practice." Journal of the Operations Research Society 44: 533-549.

Pirolli, P. and S. Card (1995). Information foraging in information access environments. Conference on Human Factors in Computing Systems.

Porter., M. F. (1980). "An algorithm for suffix stripping." Program 14(3): 130-137.

Rasmussen, E. (1992). Clustering Algorithms. Information Retrieval. W. B. Frakes and R. Baeza-Yates. New Jersey, Prentice Hall PTR.

Reagan-Cirincione, P. (1992). Combining Group Facilitation, Decision Modeling, and Information Technology to Improve the Accuracy of Group Judgment. Twenty-Fifth

Annual Hawaii International Conference on System Sciences (HICSS-25), koala, Hawaii.

Rumelbart, D. E., G. E. Hinton, et al. (1986). Learning internal representations by error propagation, MIT Press.

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

Schuman, S. P. (1996). "What to Look For in a Group Facilitator." Quality Progress 29(6): 69-72.

Sibson, R. (1973). "SLINK: an optimally eficient algorithm for the single link cluster method." The Computer Journal 16: 30-34.

Simon, H. A. (1965). The Shape of Automation for Men and Management New York, Harper & Row.

Steinbach, M., G. G. Karypis, et al. (2000). "A Comparison of Document Clustering Techniques." KDD Workshop on Text Mining.

Strehl, A., Joydeep, G., Mooney, R. (2000). Impact of Similarity Measures on Webpage Clustering. 17th National Conference on Artificial Intelligence: Workshop of Artificial Intelligence for Web Search,.

Tao Liu, S. L., Zheng Chen, and Wei-Ying Ma. (2003). An Evaluation on Feature Selection for Text Clustering. Twentieth International Conference on Machine Learning, Washington, D.C.

Trout, J. (1997). The New Positioning: The Latest on the World's #1 Business Strategy, McGraw-Hill.

Valacich, J. S., A. R. Dennis, et al. (1994). "Idea generation in computer-based groups: A new ending to an old story." Organizational Behavior and Human Decision Processes 57(3): 448-467.

Van Rijsbergen, C. J. (1979). Information Retrieval. Butterworths.

Vogel, D. R., J. F. Nunamaker, et al. (1989). "Electronic Meeting System Experience at IBM." Journal of Management Information Systems 6(3): 25-43.

Voorhees, E. M. (1986). "Implementing agglomerative hierarchic clustering algorithms for use in document retrieval." Information Processing & Management 22: 465-476.

Willett, P. R. (1988). "Recent Trends in Hierarchic document Clustering: a critical review." Information & Management 24(5): 577-597.

Witten, I. H. and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. San Francisco, Morgan Kaufmann.

Wong, Z. a. A., M. (2003). "Automated Facilitation of Electronic Meetings." Information and Management 41(2): 125-134.

Zamir, O. and O. Etzioni (1998). Web document clustering: a feasibility demonstration. 21st annual international ACM SIGIR conference on research and development in information retrieval, New York, NY, ACM Press.

Zhang, D. and Y. Dong. (2004). Semantic, Hierarchical, Online Clustering of Web Search Results. the 6th Asia Pacific Web Conference, Hangzhou, China.