Workflow-centric Information Distribution through Email

J. Leon Zhao
Department of MIS
University of Arizona
Tucson, AZ 85721
lzhao@bpa.arizona.edu

Akhil Kumar
College of Business, Box 419
University of Colorado
Boulder, CO 80309-0419
Akhil.Kumar@Colorado.Edu

Edward A. Stohr
Stern School of Business
New York University
New York, NY 10012, USA
estohr@stern.nyu.edu

Abstract

Organizations require ways to efficiently distribute information such as news releases, seminar announcements, and memos. While the machinery for information storage, manipulation, and retrieval exists, research dealing directly with its distribution in an organizational context is scarce. In this paper, we address this need by first examining the pros and cons of the conventional "mailing lists" approach and then proposing new workflow mechanisms that improve the efficiency and effectiveness of information distribution through email. The proposed approach is relevant to other information distribution approaches beyond e-mail. The main contributions of this study include: (1) offering a workflow perspective on organizational information distribution; (2) analysis of workflows in two new information distribution methods based on dynamic mailing lists and profile matching, respectively; and (3) proposing a new way of matching supply and demand of information that extends existing information filtering algorithms.

1. Introduction

In recent years, organizations have deployed workflow management systems (*WFMS*) to support routing of documents and tasks in electronic form, thereby enabling the automation of business processes across teams, functional departments, customers, and suppliers [7, 13, 30]. WFMS are natural repositories for organizational memory, especially with regard to business processes and logic [36], and are also well suited for targeted delivery of organizational information due to their ability to provide dynamic, as needed, connections between organizational members.

Our goal in this study is to develop new approaches to organizational information distribution from a workflow perspective. We use seminar announcements sent over email as an example of information distribution in our discussion. Seminars are a common means for sharing and enhancing organizational information, and seminar announcements are representative of other information distribution formats such as memos and news offerings [6]. We first examine the pros and cons of the conventional "mailing lists" approach and then propose new workflow mechanisms intended to improve the efficiency and effectiveness of information distribution.

Our approach is *workflow-centric* because we view information distribution as an organizational process and investigate process-oriented, efficient and flexible solutions for it. Moreover, these solutions are asynchronous

in that each step can be done independently of others as long as each step is performed in a permissible order with respect to the others.

Our research contributes to the literature in two ways:

(1) We propose a workflow-centric perspective to the information overload problem in the context of organizational information management. Within this perspective, we analyze three types of workflow in organizational information distribution: workflow with static mailing lists, workflow with dynamic mailing lists, and workflow with automatic profile matching.

(2) The two new mechanisms we propose, namely, dynamic mailing lists and automatic profile matching, promote a more accurate distribution of information by taking user preferences and behaviors into account. With the dynamic mailing lists approach, the user acts are logged and used to update the mailing lists. With the profile matching approach, we propose a matching algorithm that makes use of profiles of information objects and user profiles.

Our proposal aims to avoid information overload by reducing the transmission of e-mail to people to whom it is irrelevant. This is different from e-mail filtering methods [6] that provide tools for blocking irrelevant incoming e-mail. It has been shown that approaches to filtering "electronic junk" mail based on prioritizing or blocking incoming channels are not always effective [11]. A by-product of the information distribution system proposed in this paper is that it can provide useful inputs to management concerning the firm's information and knowledge resources.

At the outset it is helpful to clarify how our work relates to organizational knowledge management. Organizational knowledge is a form of collective competence based on the know-how of an organization's people and systems. Increasingly, an important part of this competence is system-based rather than people-based [5, 31, 32]. Research in knowledge management, organizational memory and organizational learning has focused on the development of models and mechanisms for the capture, storage, filtering, manipulation and distribution of knowledge in an organizational setting [1, 12, 25, 29, 35]. In this paper, the focus is primarily on the distribution aspect of knowledge, and hence the term *information* is perhaps more appropriate. With the explosion in the amount of information available worldwide, this problem is becoming increasingly important while good solutions are still lacking.

The remainder of the paper is structured as follows. Section 2 provides a brief literature overview. Section 3 presents preliminary concepts needed for the development of the workflow models and matching algorithms that are described in later sections. Section 4 develops three workflow process models: conventional mailing

lists, dynamic mailing lists, and automatic matching of information supply and demand. Section 5 delineates the data structures of user and message profiles. Section 6 gives a two-stage matching algorithm that extends the keyword-based information filtering methods found in the literature. Section 7 discusses potential contributions of the workflow approach to an organization's information base. Finally, Section 8 summarizes the paper and outlines future research directions.

2. Related Work

Our research lies at the intersection of workflow systems and information retrieval. In this section, we briefly describe some related ideas and research areas.

2.1. Ad Hoc Workflow Management Systems

Workflow Management Systems (WFMS) can be classified on a number of different criteria [7]. A common scheme is to divide them according to their transaction throughput capacity and the latitude they allow users in the choice of processing steps. *Production workflow systems* are designed for heavy transaction processing and enforce rigid controls, and strict routing rules. *Administrative workflow systems* help automate less intensive, less complex, and often more sporadic processes for tasks such as expense form processing. They are easier to implement but also enforce rigid routing schemes. *Ad hoc workflow systems* support the definition of new, unanticipated work flows such as those needed to support collaborative work [33]. Ad hoc WFMS are close in spirit to the systems proposed here. As in our proposals, ad hoc WFMS are often implemented as enhancements to basic e-mail transport mechanisms and they are often used to support information distribution. Our "dynamic mailing list" proposal may be classified as a hybrid between an administrative and an ad hoc WFMS; the processing steps are pre-defined but the receivers of messages are determined ad hoc.

2.2. Advanced Electronic Mailing Systems

This subsection reviews various enhancements to e-mail systems that have been proposed in the literature. The *Object Lens System* [17] was the first system to allow users to filter their e-mail. The *Coordinator system* [34] attempted to increase the meaningfulness of e-mail messages by requiring senders to classify the messages they sent in terms of a number of "speech acts." Motiwalla and Nunamaker [20] studied the use of knowledge-based e-mail systems as a tool for supporting managerial decisions. Several studies have been conducted on information filtering methods in the context of e-mail for document filtering [9] and for document sharing [6]. These studies focused on full text matching based on keywords with various weighting schemes. Motiwalla

[19] proposed an intelligent system for prioritizing e-mail received based on personal preferences (or profiles). In contrast to most of these approaches, our filtering methods are applied at the source of the messages rather than at their destinations.

2.3. Conceptual Clustering and Concept Space

A major approach for information retrieval is conceptual clustering, in which documents are classified based on the terms contained in the documents and queries are processed based on the terms specified by the user. The basic technique is the vector space model using simple measures of word frequency [18]. In the vector space model, documents are specified as vectors in the multidimensional space of keywords and are clustered according to the frequency of keywords appearing in them. The main advantage of the vector space model is that the formation of the conceptual clusters can be automated using computational procedures.

Similarly, a concept space approach has been proposed to create meaningful and understandable domain-specific networks of terms and weighted associations, which are used to represent the underlying information spaces, i.e., documents in different domain-specific databases [3]. The concept space approach consists of (1) acquiring complete and recent collections of documents as the sources of vocabularies, (2) automatically indexing all terms in the documents, (3) clustering the documents based on term frequency and document frequency, and (4) organizing documents based on multi-term associations.

The research on conceptual clustering and concept spaces indicates that techniques exist that are capable of automatically creating conceptual networks of millions of domain specific terms [3].

2.4. Knowledge Management Systems

Knowledge Management Systems are based on the idea that knowledge can be made explicit, stored in a database and distributed on demand to users [12]. A major problem in such systems is to maintain and communicate the relative quality and worth of the stored items of information. The *Annotate system* [8] addresses this problem by employing a simple document (knowledge item) rating scheme to attach meta-information, such as quality indicators, to documents. (In Annotate, ratings are assigned by the user.) Our objective in this paper is to advance this general idea by refining the knowledge distribution component of such a system. While we center our discussion on e-mail distribution systems, the ideas can be transferred to other modes of information distribution and knowledge sharing.

To summarize, our work in this paper complements previous work on workflow management systems, enhanced e-mail systems, and knowledge management systems. We combine features from each of these areas to design a real-time system involving user feedback and dynamic matching of messages with user needs.

3. Pros and Cons of Electronic Mailing Lists

In this section, we discuss some preliminary concepts including the characteristics of seminar announcements, types of mailing lists, and information supply and demand.

3.1. Example Seminar Announcement

Seminars are a common mode of organizational learning in most organizations. Attendance at a seminar may be mandatory in some cases, but is more often voluntary. Seminar announcements, such as that shown in Figure 1, provide information to the potential audience.

HKUST
Department of Information & Systems Management

Seminar: One Country, Two Systems -- Telecom Deregulation in Hong Kong and China

Dr. Yan Xu, Assistant Professor
Department of BUS, HKUST

Abstract

Telecommunications deregulation strategies are self-evidently different in Hong Kong and China due to differences in political and economic systems. This paper provides an overview of contemporary trends in telecommunications deregulation in these two territories. Comparisons will be drawn between the two systems in regard to ownership, foreign direct investment, respective regulatory frameworks and the government's perseverance in propelling deregulation.
Date: Friday, April 16, 1999
Time: 3:00 pm - 4:00 pm
Venue: Conference Room 4379, 4/F (Lift Nos. 17 - 18)
Remarks: For further inquiries, please contact Ms. Pancy Wong (x7923, e-mail: imstpancy). We would appreciate it if you could let us know if you intend to attend the seminar.

Fig. 1. An example of seminar announcement.

The example specifies that a seminar will be offered by Dr. Xu (the *presenter*) about telecommunication deregulation in China (the *topic*) on Friday, April 16, 1999 (*date*) at 2:45 pm (*time*). An *abstract* describes the seminar. The unique structure of seminar announcements makes it possible to convert them into machine-readable form so that they can be matched to the needs of receivers. The specific content of a particular announcement indicates the subset of personnel in the organization for whom the seminar may be of most

interest. (In this example, it is people interested in international telecommunications.) The seminar coordinator would like to estimate the number of people who are likely to attend the seminar in order to arrange the right room and prepare for the tea reception.

Seminar announcements are normally distributed to individuals on mailing lists. As we elaborate next, the method of announcing through mailing lists has several drawbacks. We therefore propose a workflow-centric approach that increases the efficiency and reduces the costs of information distribution.

3.2. Mailing Lists

Mailing lists are a basic tool for information distribution in many modern organizations [11, 19]. In general, an organization such as a university can have many mailing lists of different types:

- *Administrative lists*: This type of mailing list mirrors the organizational hierarchy and is used to distribute important messages that concern the personnel in various organizational units. In an educational context course lists are a common form of administrative list.

- *Information lists*: This type of mailing list is used to inform people of news items and events that are of interest to a general audience.

- *Interest-group lists*: This type of mailing list is designed to serve the needs of special groups of people with common interests.

There are also many cross-institutional mailing lists such as the "ISWORLD LISTSERV" mailing list, which is used by thousands of subscribers around the world. However, in this paper, we focus on intra-institutional mailing lists. The ideas underlying our approach should also be useful in cross-institutional mailing lists.

The management of mailing lists requires a lot of work and has therefore been automated to some extent. The Majordomo system is an example of such a mechanism [27]. As we discuss next, while Majordomo (or similar software) helps with the management of mailing lists, there are several drawbacks with the traditional approach to information distribution using mailing lists.

3.3. Information Overload Problems with Mailing Lists

While electronic mail has made the distribution of information much cheaper and faster, the side effect it creates is information overload. The ease with which information can be distributed electronically encourages overuse of e-mail in general, and of mailing lists in particular. A basic *uniformity assumption* is that everyone

subscribing to a mailing list has the same information needs. The adverse consequences of this assumption are evident when the size of the list increases and it is used beyond its original purposes. This leads to a number of problems including junk mail, irrelevant mail, unmet need and high costs of e-mail management:

- *Junk mail* is an all-too-familiar problem for e-mail users. For each junk mail message, one may have to spend a few seconds or sometimes even a few minutes, to identify and discard it.

- *Irrelevant mail* messages are somewhat more difficult to identify than junk mail messages. This category of mail may appear to be useful, but is not really relevant.

- *Unmet needs* occur if users are not informed of messages that are relevant to them. This can happen if senders adopt a conservative approach to the inclusion of people on mailing lists or minimize the number of mailing lists to which messages are sent.

- *Labor cost of using mailing lists* is a major hidden cost, considering everyone in the organization may spend fifteen minutes or more per day to browse, sort, delete and file e-mail messages.

The above issues call for a more accurate way of matching the supply and demand for e-mail based on the contents of the message and the interests of the user. A better messaging system will reduce the time spent on irrelevant messages and potentially increase the productivity of users. Moreover, a better mailing system can support information management objectives. Consider figure 1, which illustrates the use of mailing lists to distribute messages to users who have subscribed to various lists.
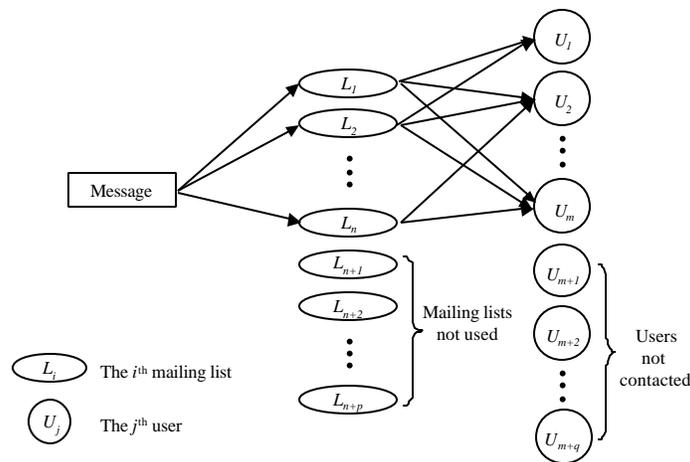

Figure 1. The use of mailing lists.

It is evident that the mailing list approach has several drawbacks:

- *Overload versus Starvation*. Two types of information distribution errors may occur, *information overload* and *information starvation*. The former occurs when users who are not interested in the message are sent the message, and the latter when users who are interested in the message are NOT sent the message. Information overload occurs when the message is sent to a mailing list that has few interested users, and information starvation occurs when the message is NOT sent to a mailing list that contains more than one interested user.

- *Feeback from users.* There is no reliable feedback mechanism for the message sender to track the user's responses. As a result, if the sender would like to follow up with the first message, s/he must send to the same mailing lists again. This contributes to information overload. This lack of feedback mechanism aggravates both the information overload and starvation problems since the sender has no way of learning the interests of message recipients.

- *User receptiveness/tolerance.* Users of mailing lists vary in their tolerance towards receiving messages. Some users are hungrier for information than others and exhibit higher tolerance to information overload in order to avoid information starvation. This type of user tends to subscribe to many mailing lists. On the other hand, other users are more conservative in subscribing to mailing lists to avoid information overload, and are thus more vulnerable to information starvation.

While the user has the option to choose which mailing lists to subscribe to, the information overload and starvation problems cannot be readily resolved because the mailing lists offer a crude and static way of satisfying unique user needs.

- Mailing lists require that the sender must first select the relevant list(s) to post her message to. Except for the case of administrative lists, a user can subscribe to or unsubscribe from the lists as he or she sees fit.

- Mailing lists provide a crude form of selecting receivers, since people on the same list may have varying needs and interests.

In this study, we examine the process of organizational information distribution via mailing lists and provide several new mechanisms to control the two types of information distribution errors, namely, information overload and starvation. We achieve this objective by:

- Adding a formal feedback mechanism to the process of organizational information distribution so that the information overload and starvation errors can be discovered systematically and corrected.

- Selecting the appropriate message recipients by matching the message profile with the user profiles to minimize information overload and starvation at the same time.

- Allowing users to specify their level of aggressiveness in receiving messages so that user variations can be taken into account in the message filtering. We achieve this goal by combining the use of mailing lists and user profiles as detailed in later sections.

- Developing a mechanism for learning the evolution of user interests so that the user profiles can be updated on a regular basis. The results of user evolution would be to modify the user profile and add the user to additional mailing lists.

3.4. Matching the Supply and Demand for Information

One way of looking at the information distribution problem is to consider it as a supply and demand problem. We could have the supplier, i.e., the advertiser of the seminar, provide a profile of the announcement, which would then be matched to the users' profiles. We call this the *supply and demand matching* problem. In short, mailing lists do not match supply and demand for information well and this results in irrelevant mail and unmet needs. On the other hand, mailing lists have the advantages of simplicity and of reflecting the informal and formal structure of the organization.

In the remainder of this paper, we consider a hybrid approach that combines automatic matching of supply and demand with the use of mailing lists. The mailing lists will be used as the initial filtering apparatus prior to matching the seminar and receiver profiles. While we focus on the distribution or "push" of information to users through e-mail mechanisms, we also assume that the firm has an Intranet bulletin board on which important notices are posted. This is a "pull" mechanism that allows interested users to gather information beyond what they might glean from the e-mail distribution system. As explained later, the bulletin board also allows us to gather additional information about users' interests that can be used to help improve the matching of information supply and demand.

4. Analysis of Three Workflow Processes

We investigate three workflow processes for organizational information distribution. The first type is workflow with static mailing lists, which represents the conventional approach for information distribution in most institutions. We propose two new workflow processes: workflow with dynamic mailing lists, and workflow with profile matching. The new workflow approaches improve static mailing lists by adding workflow mechanisms such as automatic logging of receiver acts, automatic feedback between senders and

receivers of seminar announcements, and automatic matching of supply and demand for information based on seminar and user profiles.

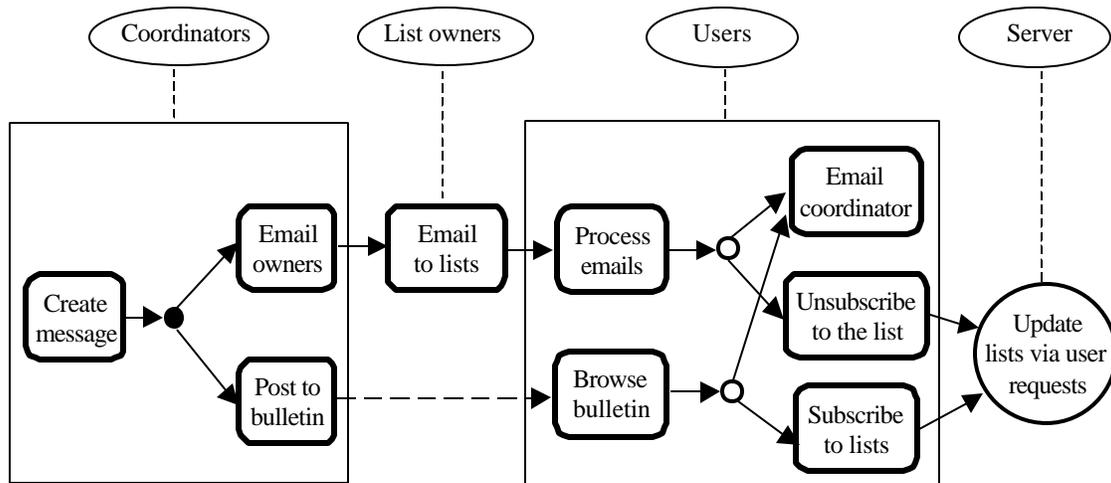## 4.1.   Workflow with Static Mailing Lists

Figure 2. Workflow with static mailing lists.

The notation used in the figure is based on the Information Control Network (*ICN*) approach [2]. The ovals denote human and system roles. The circle represents activities performed by the system, and the rounded rectangle represents human activities that may involve the use of software tools. The arrow indicates a sequence between two activities, and the dashed arrow indicates a sequence with unspecified delays. For instance, users may browse the bulletin board for a message only after the message has been posted by the coordinator; however, this sequence does not have to occur at all and has an uncertain delay even if it occurs. The black dot is an AND split, and the small circle is an OR split. We also extend the ICN notation by denoting human and system roles explicitly in the diagram and by enclosing the activities of a role in a rectangle for ease of identification.

In Figure 2, four roles are identified: C*oordinators* organize the seminars; *List owners* are authorized to send messages to the mailing lists; *Users* subscribe to the mailing lists; and, *the server* is the software program that enables the administration and use of the mailing lists. While a seminar coordinator and a list owner can sometimes be the same person, they are usually separate.

Workflow with static mailing lists includes the following main activities (identified by role):

- *Coordinator*

1) Create message for the seminar

2) Send message to mailing list owners

3) Post message to the electronic bulletin board

- *List Owner*

4) Send message to mailing list subscribers

- *User*

5) Browse, read, delete, file and reply to selected messages

6) E-mail to coordinators if additional information is needed

7) Unsubscribe to mailing lists when desired

8) Browse the electronic bulletin periodically for seminar announcements

9) May subscribe to mailing lists when interested in receiving messages regularly

- *Server*

10) Update mailing lists based on explicit user requests

The following points are relevant in the context of Figure 2:

- After the "process e-mail" activity, the default of doing nothing at all is also possible.

- To keep the figure simple, some other potential activities are not shown, such as interactions between list owners and coordinators and between coordinators and users.

- The static mailing lists approach involves little process automation except for the administration of mailing lists through a software tool such as Majordomo.

- There is no required feedback between the senders and potential attendees of the seminar. As a result, the seminar organizers (or the seminar coordinators) must guess the level of interest in the seminar. In reality, most seminar organizers send e-mail repeatedly to users to solicit responses, which further increases the information overload. This is a very typical annoyance and later we will suggest a solution for it that follows from our techniques.

In workflow with dynamic mailing lists as discussed next, we propose to improve the workflow effectiveness by logging user acts and enabling senders to track user responses indirectly. The system updates the mailing lists automatically based on user responses to the announcements.

4.2. Workflow with Dynamic Mailing Lists

A workflow with dynamic mailing lists is illustrated in Figure 3. In this approach, users "tag" each incoming e-mail message to indicate its relevance to them (c.f., the Annotate System mentioned earlier). Using this and other information, the server dynamically updates the mailing lists. The main activities are listed below. Activities depicted in bold font differ from those in the static mailing list system.
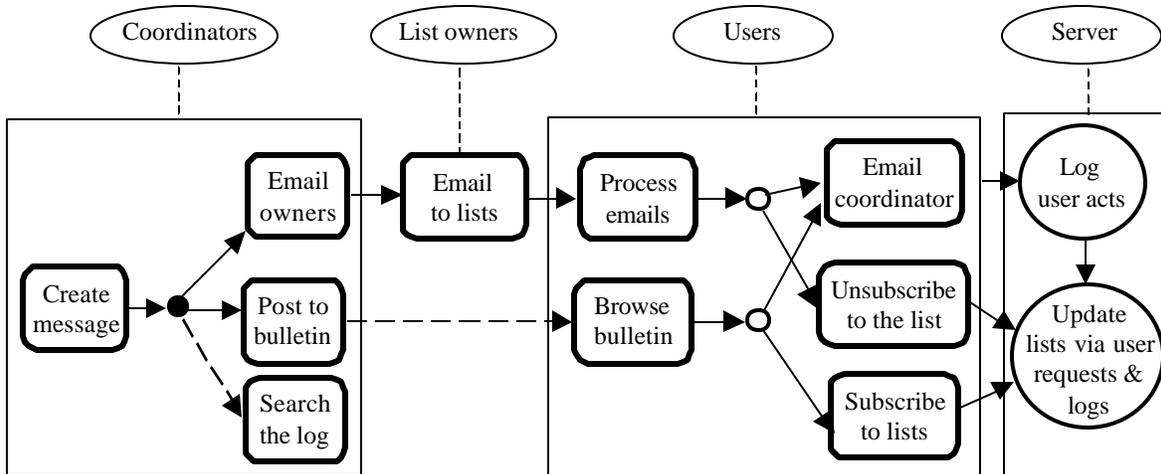

Figure 3. Workflow with dynamic mailing lists.

- *Coordinator*

1) Create message for the seminar

2) Send message to mailing list owners

3) Post message to the electronic bulletin

4) Search the log files to determine the number of potential attendees and their e-mail addresses for communicating any changes to the seminar contents or schedule

- *List owner*

5) Send message to mailing list subscribers

- *User*

6) Browse, read, delete, file, **tag** and/or reply in writing to incoming messages

7) E-mail to coordinators if additional information is needed

8) Unsubscribe from mailing lists when desired

9) Browse the electronic bulletin periodically for seminar announcements

10) May subscribe to mailing lists when interested in receiving messages regularly

- *Server*

11) Log user acts with respect to the e-mail messages and the electronic bulletin board

12) Update mailing lists based on user requests and **logs of user actions**

Note that in Figure 3, the arrow into "Log user acts" originates from the rectangle enclosing the users' activities to the server logging activity, indicating that logging may be done with respect to more than one activity by the users. This is a way to simplify the diagram.

Compared with the static mailing list approach, this workflow includes three new activities: logging of user actions and reactions to messages by the server, queries to the logs by coordinators, and update of the mailing lists by the server based on system log information as well as user requests. Brief explanations of these system functions and their implementation methods are given next:

- An automatic logging component is needed to capture user reactions to the seminar announcements (the tags) along with e-mail actions that might need to be taken. This could be done by adding buttons to the e-mail for users to identify their interests such as "irrelevant", "not interesting", "interesting, but will not be able to attend", "may attend", "will attend", "remind me one week prior", and "add to calendar". Clicking on one of these buttons could be a voluntary action or a mandatory preliminary to exiting the current message and moving on to the next message. Requiring the users to make a choice before exiting the message would ensure a complete response – at the cost of additional effort by the user and the possibility of a negative reaction to the system.

- A single data table, USER_LOG, with the following data elements: User ID, Message ID, Data Received, User Action Type, and User Comments, can be used to log user actions. User logs are very similar to workflow histories. Koksal, Arpinar and Dogac [14] have developed efficient data structures and maintenance algorithms for workflow history management in a database environment.

- Automatic update algorithms are needed to maintain mailing lists more dynamically based on user actions. Because accurately predicting user' desires is difficult, we favor an approach that prompts the user with optional updates in two circumstances:

  (1) When the user replies with "irrelevant" responses to a mailing list for a number of times in a row, the workflow system will give the user the option of dropping the mailing list or continuing it. The user

13

can then decide what to do. The default value for the threshold number can be initialized by the system operator and can later be modified by the user.

(2) When the user retrieves information about a seminar from the electronic bulletin, the workflow system will prompt the user with the option of adding his/her name to one or more of the mailing lists to which the seminar was announced. The user can then make the decision whether or not to join the list(s).

- Senders of seminar messages can access user logs to find out how many people are interested in the seminar, how many of them are planning to attend, and who is interested (and not interested) in the offered seminars. This information can be useful for planning and scheduling seminars and, perhaps, interacting directly with those who are likely to attend. This takes some of the guesswork out of seminar planning and obviates the need to send the same message repeatedly for fear of insufficient attendance.

- The query facilities envisaged for the user logs are normal database functions and can therefore be implemented in a straightforward fashion.

In summary, there are three major advantages to the dynamic mailing list approach. (1) The server now has a mechanism to update the mailing lists, (2) Coordinators can find out easily who is interested in attending the seminars and can also interact more precisely with potential attendees in case of changes to the seminar such as a change of venue and time, and (3) The system automatically gathers valuable information about the intellectual interests and tastes of the participants.

Obviously, there is a privacy issue involved when people have to share information about their plans with others. Experience in progressive companies like Sun Microsystems has shown that people are willing to even share their calendars with others, and this results in an increase in overall productivity for everybody [10]. On the other hand, in other situations, major privacy issues may be involved. In this paper, the authors are not in a position to pursue this matter in any detail.
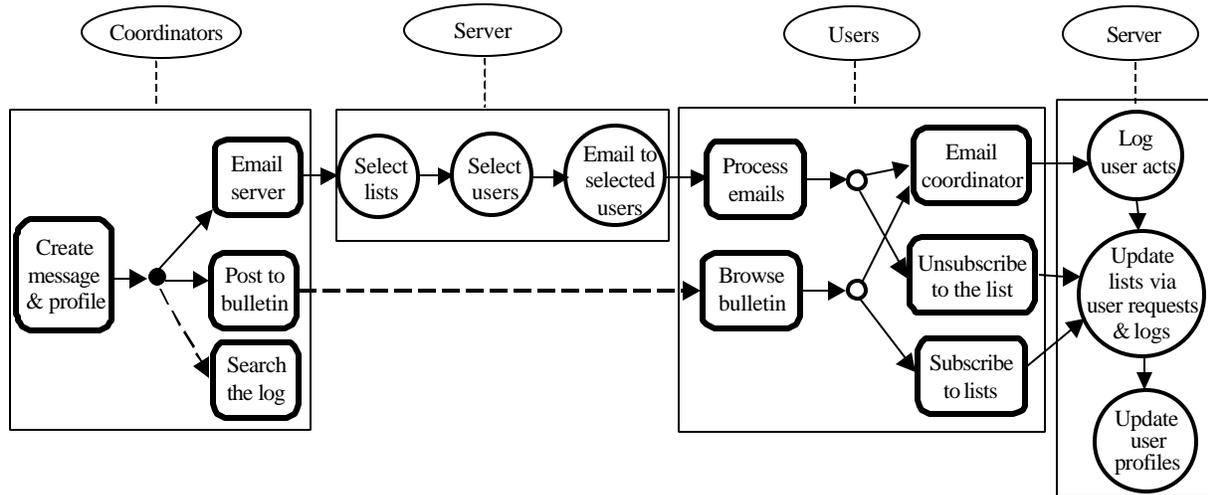
4.3.   Workflow with Profile Matching

Figure 4. Workflow with automatic profile matching

The workflow approach with dynamic mailing lists improves system efficiency by automatically logging user acts and utilizing the logged information to support mailing list maintenance and queries for potential attendees by coordinators.   However, certain mailing lists such as administrative mailing lists cannot be updated due to their mandatory nature.   Furthermore, dynamic mailing lists, though better matched to user needs than static mailing lists, may still cause information overload due to the uniformity assumption that all users in the same mailing list are uniformly interested in the same things.   Therefore, we propose another approach to automatically match information supply and demand.   The proposed workflow system, which automatically matches seminars with interested users, includes the following activities (elements that are new relative to the previous two approaches are shown in bold face):

- *Coordinator*

1) Create message for seminar announcement and **seminar profile.**

2) Send message to the seminar server **along with seminar profile.**

3) Post message to the electronic bulletin board.

4) Search the log files to determine the number of potential attendees and their e-mail addresses for communicating any changes to the seminar contents or schedule.

- *Server*

5) Select relevant mailing lists based on mailing list profiles using a list matching algorithm

15

6) Select users in the relevant mailing lists based on user profiles using a user matching algorithm

7) Send message to selected users

- *User*

8) Browse, read, delete, file, tag and/or reply in writing to incoming messages

9) E-mail to coordinators if additional information is needed

10) Unsubscribe to mailing lists when desired

11) Browse the electronic bulletin periodically for seminar announcements

12) May subscribe to mailing lists when interested in receiving messages regularly

- *Server*

13) Log user acts with respect to the e-mail messages and the electronic bulletin board

14) Modify the profiles and members of mailing lists periodically based on the user log

15) Update mailing lists based on specific user requests and user logs

This new workflow design adds automatic profile matching between seminars and mailing lists and between seminars and interested users. As a result, the role of mailing list owners is removed from the workflow since abusive uses of mailing lists are now controlled by the server's matching algorithm. The matching is done in two stages, the mailing list match and the user match. The advantage of the two stage matching method is to reduce computational cost by making use of the dynamic mailing lists. The main feature of the new workflow is that it attempts to send a message to only relevant mailing lists, and only to interested users within the relevant mailing lists.

Note that under the profile-based approach, the user and mailing list profiles must be created. This may be done either manually, or semi-automatically. We discuss algorithms for the initialization and maintenance of profiles in the next section along with details of the matching algorithms.

4.4. Summary of Techniques

Table 1 summarizes our various techniques and highlights their main features. This table shows that the three approaches differ markedly in terms of their features. Our proposed approach, automated profile matching, should improve productivity by targeting messages from senders to receivers in a more intelligent way thus saving the time of receivers in reading only what they are interested in and also saving the senders the trouble of selecting which lists are most appropriate for their missives. However, it depends critically on how

the user profiles are managed and matched with message profiles. Therefore, the rest of the paper focuses on this issue.

	Static mailing lists	Dynamic mailing lists	Profile matching
Basic approach	Users subscribe to existing lists	Users subscribe and reply to the received messages	System matches messages and users automatically
Implementation effort	Easy	Hard	Even harder
User involvement	Subscribe/Unsubscribe periodically to lists; reply to messages	Subscribe/unsubscribe to lists; tag and/or reply to messages	Tag and/or reply to messages. Define and update interests as keywords
User feedback frequency	Only occasional	Continuous	Continuous
List owners	Must maintain & modify lists regularly	Must maintain & modify lists regularly	No list owners required
Sender perspective	Must pick the appropriate list(s)	Must pick the appropriate list(s)	Does not pick a list; monitors message profile
Success factors	Well-targeted and focused lists	Users willing to give feedback	Well chosen keywords
Organizational Effectiveness	Low	Medium	High

Table 1: A comparison of the three approaches.

5. Management of User Profiles

In this section, we present the data structures and algorithms for creating and maintaining profiles for seminar announcement messages, users and user groups using the profile matching approach.

5.1. Organizational Concept Space

We propose a new data structure referred to as the *organizational concept space* (*OCS*). This data structure is used to store user interests in an organization and to support the efficient matching of message profiles with those of users and groups. The OCS categorizes all concepts (or keywords) into two classes, disciplines and topics. As we shall illustrate, the extended data structure of our concept space enables us to

process the matching of message with users more efficiently. The OCS also forms a part of the organization's essential information base as discussed in Section 7.

An organizational concept space includes two dimensions, discipline and topic, and is represented as a 2-D matrix $OCS\{D,T\}$, where,

D is the set of disciplines in the organization, $D = \{d_j\}$, $1 \leq j \leq J$;

T is the set of topics in the organization, $T = \{t_k\}$, $1 \leq k \leq K$.

An OCS contains the disciplines relevant to the organization and the relevant topics within each discipline. The OCS is updated when a new message or a user specifies concepts (topics or disciplines) that do not yet exist in the space. There are two steps to this update. First, the OCS is updated by adding the new topics and/or disciplines to the dimensions of the matrix. Second, the users will be asked to identify their interests in those new concepts. This step can be done either periodically or when user profiles are updated for other reasons.

5.2. User Profile

We now describe the data structure and the update algorithm for the user profile. Building a user profile lies in the area of user modeling (see for instance, [23, 26, 28]).

5.2.1. Contents of User Profile

A user profile includes the following data:

- User unit: department to which the user belongs.

- Personal information: name and e-mail address of user.

- Interests information: disciplines (U-Disc) and topics (U-Topics) of interest to the user.

- User receptiveness to e-mail (U-Attitude): aggressive, moderate, or conservative

These data elements can be defined in the data structure UP:

 UP(UID, U-Dept, U-Name, U-E-mail, U-Disc, U-Topics, U-Attitude).

Because individuals may be expected to have widely differing attitudes towards e-mail, we can allow a certain amount of customization. We do this by allowing users to select their own values for the U-Attitude variable - *aggressive, moderate*, or *conservative*. Aggressive users will desire more e-mail messages than either moderate

or conservative users. How user receptiveness can be used to modify the distribution of e-mail messages will be covered in a later section.

5.2.2. Initializing and Updating the User Profile

We assume that there is sufficient off-line information to be used in the initialization of user profiles. This is a reasonable assumption in information intensive organizations such as universities because there are usually ample documents in the organization, such as personal vitae, annual personnel reports, web pages, and staff profiles. Consequently, it is reasonable to assume that the user profiles above can be created without too much difficulty. The most challenging part of user profile initialization has to do with two attributes, the disciplines (U-Disc) and topics (U-Topics) in which the user is interested. Although the aggressiveness of the user in receiving messages of seminar announcements needs to be determined as well, its value does not change greatly over time. Next, we concentrate on how to update the set of disciplines and topics of interest to the user.

We also assume that the user will interact with the information distribution system in two ways, *pull* and *push*. First, the user can visit periodically the web-based seminar lists and register his/her interest by interacting with the web page in some simple and convenient manner. This is a pull event. Second, the user will receive email messages and respond to the messages with standard feedback mechanisms such as those mentioned previously. These are push events.

In summary, user profiles can be updated dynamically in response to at least three different events:

(1) When the user completes his or her annual research report and personal research interests, the system can update the user profile using the newly available information.

(2) When the user responds with "interesting, but cannot attend," "may attend," or "will attend," the system can prompt the user with the disciplines and topics contained in the seminar profile and ask the user to consider selecting new disciplines and topics to add to her user profile.

(3) Finally, when the user browses the electronic bulletin and registers interest in some seminars, the system can also update the user profile by interacting with the user. How best to perform the updating process is a subject for future research.

The user profile will be updated with respect to both pull and push types of events as follows:

- *Profile enhancement:* When a user pulls a message from the bulletin board that contains discipline or topic items not contained in the user profile, the system will extract the disciplines and topics that may represent new interests of the user. We call these keywords *the positive set of interest items*. The user will be asked

to approve or modify the new interest items. If the user approves any new profile additions, the user profile will be enhanced. Furthermore, if the new items do not exist in the organizational concept space, the latter will be extended as well. We present an algorithm for profile enhancement next.

Algorithm User-Profile Enhancement:

Given a user, u, the user profile $UP<UID=u>$, and the pulled message profile $\{D^M, T^M\}$, where D^M is the set of discipline keywords and T^M is the set of topic keywords for the message.

Step 1: Compare the message profile with the user profile.

Step 2: Determine the new interest items from the pulled message by comparing the message profile with the existing user profile.

Step 3: Ask the user to review the new interest items (new disciplines and topics); if an update of his/her user profile is desired by the user, continue. Otherwise, stop.

Step 4: Update the user profile $UP<UID=u>$ and extend the OCS as necessary.

- *Profile reduction:* When a user replies to a message with a tag indicating a low degree of interest, there might be a need to trim the user's profile. The user profile should be reduced when a user repeatedly rejects messages from a user group to which he or she belongs. The system can attempt to identify the reasons for the uninteresting messages by extracting the common keywords in the rejected messages. We call these keywords *the negative set of interest items*. The user will be asked to approve the profile reduction because the rejection of messages could be for reasons other than a change of interest. We present an algorithm for profile reduction next.

Algorithm User-Profile Reduction:

Given user u, the user profile $UP<u>$, a set of e-mail messages that have not been of interest to the user, and the message profiles $\{S^M, \{D^M, T^M\}\}$, where S^M is the set of rejected or low-interest messages, and $\{D^M, T^M\}$ represents all sets of discipline and topic keywords for the message set.

Step 1: Analyze the message profiles to determine the common keywords

Step 2: Determine the set of discipline (L-Disc) and topic (L-Topics) keywords that may have caused the loss of user interest.

Step 3: Ask the user to review the set of possibly uninteresting disciplines and topics.

Step 4: Update the user profile $UP<UID=u>$ by removing the uninteresting disciplines and topics.

5.3. Mailing List Profile

5.3.1. Basic Data Structure

In the profile-based approach, mailing lists are dynamically maintained by the system and used both to facilitate the mail distribution process and to automatically collect information on groups of users with like interests. A mailing list profile should include the following data:

- List affiliation: Name of department with which the list is affiliated.

- Type of mailing list: for example, administrative, information, discussion, or class list.

- Collective interests: disciplines (L-Disc) and topics (L-Topics) of interest

These data elements can be structured as follows:

MLP(MID, L-Dept, ML-Type, L-Disc, L-Topics)

where MID is the mailing list identifier.

5.3.2. Mailing List Profile Update

The mailing list profiles are derived based on the profiles of users associated with the mailing list. When the user profiles are updated, the mailing list profiles should be updated accordingly. The algorithm is given next.

Algorithm Mailing List Profile Update:
    Given: user u and the set of discipline and topic keywords.
    If the set of keywords is a positive set, add the new keywords to the list profile,
    Else, if it is a negative set, then remove the keywords from the list profile when user u is the only one who
        is interested in them.

5.3.3. Inheritance of User Profile from a List Profile

In an earlier subsection, we described how individual profiles may be established and modified. It is also possible to set up default initial profiles for users by inheriting them from the profiles of the mailing list(s). If a user belongs to multiple lists, then s/he would inherit from all of them.

5.4. Seminar Profile

We continue our explanation of the profile-based approach to information distribution using the seminar announcement example. However, similar data structures and algorithms would apply to other information distribution situations.

5.4.1. Basic Data Structure

A seminar profile (or message profile) includes the following data:

- Sponsorship: Name of the department coordinating the seminar

- Presenter: Name, position and affiliation of the presenter

- Seminar description: Title, abstract, related research disciplines (R-Disc), related research topics (R-Topics)

- Logistics: Date, time, venue, contact person for the seminar.

These data elements can be structured into a table:

SP(SID, C-Dept, P-Name, P-Posi, P-Affi, S-Title, S-Abst, R-Disc, R-Topics, S-Logi)

SID is the seminar identifier. Note that some of the fields such as research disciplines and topics (R-Disc and R-Topics) may be multi-valued.

5.4.2. Creation of a Seminar Profile

We assume that the organizational concept space has been set up. The message profile can be created by extracting those discipline and topic phrases from the message text that exist in the OCS. This extraction process can be done by a keywords analysis of the message text with respect to the OCS. The message sender is then asked to modify and approve the profile. An algorithm is given below for this purpose.

Algorithm Seminar Profile Creation:

Given a seminar announcement and the OCS.

Step 1: Extract all discipline and topic phrases from the seminar message by comparing it to the OCS

Step 2: Rank the phrases according to the frequency of their appearances in the announcement

Step 3: Ask the message creator to approve the discipline and topic keywords

Step 4: Ask the message creator to add additional keywords if any.

6. Profile-Based Matching of Information Supply and Demand

We propose a two-stage matching algorithm. The first stage matches a seminar profile with the mailing list profiles, and the second stage matches the seminar profile with the profiles of users in the mailing lists who qualified during the first stage.

6.1. A Two-Stage Matching Algorithm

The overall matching strategy is as follows. Given a seminar announcement and a set of mailing lists, first screen the mailing lists, and then match individual users. Initial screening of the mailing lists reduces system computation costs, especially when there may be hundreds or thousands of users of the mailing system. Matching user profiles in the second stage allows for individualized distribution decisions and reduces the information overload of the users.

6.1.1. Mailing List Screening

To screen a mailing list, we use a simple matching algorithm that compares the disciplines and topics listed in the seminar announcement with those contained in the mailing list. The disciplines and topics of interest associated with each mailing list are the superset of the interests of each user in the mailing list.

Given a seminar message s, and all mailing lists, m $\hat{I}$ M, the following algorithm is applied to generate the set, M^S, *of mailing lists for use in the second stage of the matching process*:

> Let S^{disc} be the set of disciplines and S^{topic} be the set of topics that the seminar covers, which are contained in the seminar profile *SP<SID=s>*.
>
> FOREACH mailing list m $\hat{I}$ M DO
>
> > Let T^{disc} be the set of disciplines and T^{topic} be the set of topics that the mailing list m concerns
> >
> > IF (S^{disc} Overlap T^{disc})THEN
> >
> > > IF (S^{topic} Overlap T^{topic})THEN
> > >
> > > > Insert m into M^S
> > >
> > > ENDIF
> >
> > ENDIF
>
> ENDFOR

The function *Overlap* is a matching operator that returns a *True* value if there is one or more common elements in the designated *input* sets. The same function is applied to the matching of both disciplines and topics between the seminar and mailing list profiles. If there is a match, the mailing list is inserted into the

relevant mailing lists set M^S for seminar s. Note that a mailing list is considered relevant only when there are matches for both disciplines and topics.

The *Overlap* function is implemented by the following algorithm: Given two sets, X and Y, where X and Y contain elements from the same semantic domain, perform the following steps of a Merge-Sort algorithm (Kruse, 1987):

(1) Replace the contents of X and Y with the corresponding integer codes in the organization's manual.

(2) Sort X and Y.

(3) Merge the two sorted lists X and Y into a single sorted list XY.

(4) IF XY contains any duplicate, THEN return True ELSE Return False.

This overlap algorithm can also return the number of duplicates, which can serve as a measure of the degree of overlap between X and Y.

6.1.2. User Matching

To match users to the seminar announcement, we apply a matching algorithm similar to the one used for the screening of mailing lists. Given a seminar message s, the following matching algorithm is applied:

Let S^{disc} be the set of disciplines and S^{topic} be the set of topics that the seminar covers, as listed in the seminar profile $SP<SID=s>$.

FOREACH mailing list m $\hat{I}$ M^S, DO

    FORALL users u in m, DO

        Let T^{disc} be the set of disciplines and T^{topic} be the set of topics that is of interest to user u

        IF (S^{disc} Overlap T^{disc})THEN

          IF (S^{topic} Overlap T^{topic})THEN

            Insert u into U

          ENDIF

        ENDIF

      ENDFOR

    ENDFOR

The users in the user set U are considered to be potential attendees of the seminar.

6.2. Customizing to Individual Users' Receptiveness

As explained earlier, the U-Attitude parameter of each user's profile has the possible values, *aggressive, moderate* or *conservative*. Using this parameter we can devise different matching algorithms for different users. For instance, we can send an "aggressive user" all messages considered relevant to all the mailing lists to which he or she belongs. For "moderate users", we can lower the bar by sending a message when there is a

match of discipline without a matching of topics. Finally, the full two-stage matching algorithm can be used for "conservative users" as a default.

6.3.   An Example of Two-Stage Matching

We now present an example of the two-stage matching of seminar and users, which illustrates the concepts with a very simple data set based on the seminar announcement in Figure 1.

Let the seminar profile be:

SP("S00001", "BUS", "Dr. Yan Xu", "Assistant Professor", "BUS, HKUST", " One Country, Two Systems -- Contrasting Approaches to Telecommunications Deregulation in Hong Kong and China", "Abstract: …", "information technology, telecommunication policy, political economics, applied economics", "deregulation, political systems, foreign investments, China economic reform").

Next, take two departmental mailing lists from the School of Business (BUS) and Applied Economics (Economics). The records in the MLP table are:

- MLP(ML0001, "BUS", "admin", "information technology, information systems, operations management, applied statistics", "database management, supply chain management, telecommunication policy, workflow management, financial information systems").

- MLP(ML0002, "Economics", "admin", "economics, political economics", "econometrics, political systems, telecommunication policy").

Further, the records in the UP table for four user profiles, two each from BUS and Economics, are as follows:

- UP("U0003", "BUS", "Ted Clark", "ted-clark@ust.hk", "information technology, telecommunication policy", "supply chain management, online auction, virtual organizations, inter-organizational systems", "aggressive").

- UP("U0005", "BUS", "James Kwok", "james-kwok@ust.hk", "information systems, information engineering", "image processing, image databases, web information systems", "moderate").

- UP("U0007", "Economics", "Mary Jones", "mary-jones@ust.hk", "econometrics, applied economics", "matrix theory, linear programming, economic forecasting", "conservative").

- UP("U0009", "Economics", "Leonard Chang", "leonard-chang@ust.hk", "political economics, applied economics", "China economic reform, political systems", "moderate").

In the first stage, the seminar profile S00001 is compared to the two mailing lists ML0001 and ML0002. Intuitively, seminar S00001 matches both lists because both contain the discipline "information technology". Similarly, a match is found between disciplines in S00001 and ML0002 due to a common discipline "political economics".

In the second stage, user U0003 matches the seminar on the discipline "information technology" and "telecommunication policy", but there is no common topic between them. Nevertheless, since the user has an "aggressive" receiving attitude, he would still receive this message. The next user U0005 does not achieve a match with the seminar in either the disciplines or the topics, and does not receive the message.

Under the Economics mailing list, user U0007 matches the seminar on discipline "applied economics," but not on any topic. Therefore, since this user selected a "conservative" receiving attitude the message is not sent to her. Finally, user U0009 matches the seminar in both disciplines "applied economics" and "political economics" and in topics "China economic reform" and "political systems". Hence, he would receive the message.

This simple example, by eliminating two out of four potential recipients, illustrates how the two-stage matching is performed and indicates intuitively that the matching process can be potentially effective. However, the efficiency of the matching algorithm will need to be determined further through experiments and theoretical analyses in a subsequent study.

7. Discussion

7.1. Advantages of the Matching Algorithm

Most information filtering algorithms are focused on full text search using keywords [6, 9]. Many algorithms rely on the automatic extraction of keywords from the texts [13]. User profiles under full text filtering are also based on keywords. Because of the enormous number of keywords in general contexts, it is inappropriate to ask the user to initialize and update the keywords that are of interest to her. As a result, the usual approach is to extract keywords from those documents that are considered interesting by the user [6]. Furthermore, due to the large number of keywords in the user profile and the texts being filtered, certain weighting schemes are needed to take different keywords into account with varying relative importance. Note

that the concept of mailing list profiles is not found in the literature and therefore is an innovative feature of our approach.

Our matching algorithm is innovative in two respects. First, it uses only certain types of keywords, namely, words and phrases about disciplines and topics. Second, it uses disciplines and topics in the matching process in two steps, i.e., it first matches disciplines and then matches topics. Furthermore, due to the small number of disciplines and topics of interest to a typical organization, we assume the existence of a manual or thesaurus that codes the known disciplines and topics with due attention to synonyms. This assumption is realistic since such information is often available in documents such as annual research reports and personnel resumes. In summary, the limited scope of our matching tasks enables a simple yet effective matching via the proposed two-stage process. A comparison of the efficiency and effectiveness of our approach and conventional filtering algorithms is an interesting topic that we intend to pursue in a separate study.

Finally, another advantage of the matching algorithm is that it can be tailored to reduce clutter and repetitive email. The matching algorithm can return a match coefficient that is a measure of the degree of match between the user's interests and the seminar. The first email message about a seminar would be sent to everybody above a certain threshold. For subsequent reminder email, the threshold would be increased. Thus only those individuals who have a very high match or have explicitly indicated an interest in the seminar will receive reminders for it.

7.2. Contributions to the Organization's Knowledge Repository

The organization can gain important "self knowledge" as a by-product of the workflow-based information distribution system described in this paper. In particular, the proposed data structures provide important management information:

1) The *Organizational Concept Space (OCS)* contains a dynamically changing list of the disciplines and topics of current interest to the organization. If this list is augmented by frequency counts for each keyword, the OCS can provide information on the perceived relative importance of each topic. Comparing the development of the OCS at different points in time can help the organization understand its changing intellectual focus. If this focus remains relatively static, for example, it could imply that the organization is losing its innovative edge.

2) The *User Profiles (UP)* contain information on the interests of each user. This information can be integrated with the company's Human Resources database to provide a dynamically changing perspective on the skill sets of employees. Static analysis of the User Profiles can help the organization gauge its

relative depth (number of people) in different intellectual areas. Analysis of changes in the profiles over time can help the company understand its evolution towards a knowledge-based organization.

3) The *Mailing List Profiles* (*MLP*) contain current information on the clusters of topics of interest to users in the organization as well as the identity of each list's sponsoring organization. Association of users with each MLP (using the two-stage algorithm explained in Section 6) can help the organization determine how its employees are clustered in informal interest groups and provide a basis for team formation and work assignments.

4) The *Seminar Profiles (SP)* record an important component of the organization's learning program together with the sponsoring organizations.

In summary, the system logs and data structures, produced and maintained by the proposed workflow-based information distribution system, can automatically contribute to the organization's knowledge base. The workflow system logs can be analyzed to produce new insights for management. The data structures can be integrated with the organization's knowledge base to provide current information concerning who is interested in what within the organization. The use of the proposed information distribution system to augment and update organizational knowledge repositories will be left for future research.

8. Conclusions

This paper described a workflow-centric approach to the targeted delivery of information. Our objective was to support the communication component of information management systems. We concentrated on electronic messaging as the main distribution mechanism. However, the issues raised and the general approach are relevant in other situations where the objective is to distribute information or share encoded knowledge.

We proposed two new workflow processes that improve the accuracy of information distribution by (a) logging user responses to the information distribution process and allowing the sender to query the log file, and (b) automatically matching seminars and users. The three resulting workflow processes differ in their complexity and the level of automation.

Our main contribution is the introduction of a workflow perspective into the domain of information filtering and delivery. We showed in this paper that it is possible to integrate workflow techniques with information filtering technique to achieve more accurate information distribution. The ability to capture user responses through logging gives rise to novel approaches to matching the supply and demand for information, and in particular, to combating the information overload problem.

The proposed workflow-based techniques provide *a mechanism* to help individuals and organizations achieve a balance between the alternative extremes of information overload and information starvation. However, there are many important issues that are not addressed in this paper. These include: user incentives to use the system, privacy concerns, limitations on synergistic discovery of new topics of interest to a user, and the inevitable trade-off between human and system-based mechanisms for deciding which information is or is not interesting. These issues can only be fully resolved through actual implementation of a pilot system.

Our study is only a first step towards the development of more dynamic communication mechanisms for organizational information distribution. In future research, we foresee several possible directions as follows:

- Exploring ways of receiving implicit feedback from users [23] and incorporating it into the user profile on a continuous basis.

- Applying the concepts and framework in other, non-academic environments. It is likely that the specifics of the information management process in non-academic institutions may be significantly different from academia and therefore lead to new challenges even though the basic issues remain the same.

- Evaluating the proposed workflow approaches in terms of quantitative parameters such as reduction in email clutter, overall user satisfaction, lost messages (because of filtering mistakes) and improvement in communication efficiency.

- Studying algorithms for "mining" the collected data further to reorganize the various lists dynamically based on usage patterns. The log information can be used to split certain lists or merge other lists based on usage.

- Investigating the contributions of the proposed workflow approach to the more general problem of developing and maintaining an organizational knowledge repository.

- It will also be interesting to relate ideas of information overload to the concept of information richness. The information richness theory has been studied first in its classical, positivist form by Daft and Lengel [4], and subsequently, Ngwenyama and Lee [21] have given it a critical social theory perspective and emphasized the importance of organizational context in communications.

References

1. Akscyn, R.M. et al., "KMS: a distributed hypermedia system for managing knowledge in organizations." CACM, v31, n7, 1988, p820-35.

2. Blumenthal, R. and Nutt, G., "Supporting Unstructured Workflow Activities in the Bramble ICN System," in Proceedings of the 1995 ACM Conference on Organizational Computing Systems (COOCS'95), pp 130-137, Milpitas, CA.

3. Chen, H. et al., "A parallel computing approach to creating engineering concept spaces for semantic retrieval: the Illinois digital library initiative project", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 18, No. 8, August 1996. Pp. 771-782.

4. Daft, R.L. and Lengel, R.H., "Organizational information requirements, media richness and structural design," Management Science (32:5), 1986, 554-571.

5. Davenport, T. H., and Prusak, L. *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1998.

6. Foltz, P.W. and Dumais, S.T. "Personalized information delivery: an analysis of information filtering methods." *Communications of the ACM*, vol.35, no.12, Dec. 1992, pp.51-60.

7. Georgakopoulos, D. et al., "An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure", Distributed And Parallel Databases, V3, N2, 119-153, 1995.

8. Ginsberg, M. and Kambil, A., "Annotate! A Web-based Knowledge Management Support System For Document Collections." Working Paper #IS-98-19, Stern School of Business, New York University.

9. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. 1992. "Using collaborative filtering to weave an information tapestry." *Communications of the ACM*, 35, 61-70.

10. Grudin, J. and Palen, L., 1997, "Emerging groupware successes in major corporations: Studies of adoption and adaptation," In *Proceedings WWCA97*.

11. Hall, R.J. "How to avoid unwanted e-mail." *Communications of the ACM*, March 1998, vol.41, (no.3):88-95.

12. Holsapple, C. W. and Joshi, K. D., "Description and analysis of existing knowledge management frameworks", HICSS'99.

13. Kindo, T.; Yoshida, H.; Morimoto, T.; Watanabe, T. "Adaptive personal information filtering system that organizes personal profiles automatically." IJCAI 97, Nagoya, Japan, 23-29 Aug. 1997, p. 716-21 vol.1.

14. Koksal, P., Arpinar, S., Dogac, A., "Workflow History Management", *Sigmod Record*, Volume 27, Number 1, March 1998.

15. Kruse, R.L. *Data Structure and Program Design*, Prentice Hall, Inc., 1987.

16. Kumar, A. and Zhao, J. L. "Dynamic Routing and Operational Controls in Workflow Management Systems", *Management Science*, Feb. 1999, vol.45, (no.2):253-72.

17. Lai, K-Y., T.W. Malone, and K.-C. Yu. "Object Lens: a 'spreadsheet' for cooperative work." *ACM Transactions on Office Information Systems*, Oct. 1989, vol.6, (no.4):332-53.

18. Munoz, A. "Compound key word generation from document databases using a hierarchical clustering ART model", *Intelligent Data Analysis*, Vol. 1, Number 1, January 1997.

19. Motiwalla, L.F. "An intelligent agent for prioritizing e-mail messages." *Information Resources Management Journal*, Spring 1995, vol.8, (no.2):16-24.

20. Motiwalla, L.F. and Nunamaker, J.F., Jr. 1992. MAIL-MAN: a knowledge-based mail assistant for managers. *Journal of Organizational Computing*, 2, 131-54.

21. Ngwenyama, O.K and Lee, A.S., "Communication richness in electronic mail: Critical social theory and the contextuality of meaning," MIS Quarterly, Volume 21, No. 2, 145-167, June 1997.

22. Oard, D.W., "A Conceptual framework for text filtering," University of Maryland Technical Report, CS-TR-3643, May 1996.

23. Oard, D.W. and Kim, J.,"Implicit Feedback for Recommender Systems," Proceedings of the AAAI Workshop on Recommender Systems, Madison, WI, July, 1998.

24. Orlikowski, W., "Learning from Notes: Organizational Issues in Groupware Implementation," *Proceedings of the CSCW Conference*, 1992.

25. Paradice, D.B. and Courtney, J.F., "Organizational knowledge management." *Information Resources Mgmt Journal* vol.2, no.3 (summer 1989) p1-13.

26. Rich, E.A., "User modeling via stereotypes," *Cognitive Science*, 3:329-354, 1979.

27. Schwartz, A. *Managing Mailing Lists*, O'Reilly & Associates, 1998.

28. Stadnyk, I. and Kass, R., "Modeling Users' interests in information filters," Communications of the ACM, Vol. 35, No. 12, December 1992.

29. Stein, E. W. and Zwass, V. "Actualizing organizational memory with information systems." *Information Systems Research,* 1995, v6n2, Jun p. 85-117.

30. Stohr, E. A. and Zhao, J. L. "The Expanding Mission of Workflow Technology." *Document World*, Oct.-Nov. 1998, vol.3, (no.5):21-6.

31. Swanson, E.B. "The new organizational knowledge and its systems foundations." HICSS96, p140-6, vol.3.

32. Tuomi, I. "The communicative view on organizational memory: power and ambiguity in knowledge creation systems." *HICSS96*, p147-55 vol.3.

33. Voorhoeve, M. and Van der Aalst, W. "Ad-hoc workflow: problems and solutions." *Proceedings of the 8th International Workshop on Database and Expert Systems Applications*, Toulouse, France, 1-2 Sept. 1997, p. 36-40.

34. Winograd, T. and Flores, F. *Understanding Computers and Cognition: A New Foundation for Design*, 1986. Norwood, NJ: Ablex Publishing.

35. Zack, M. H., "Managing codified knowledge", Sloan Management Review, Summer 1999, pp. 45-58.

36. Zhao, J. L. "Knowledge management and organizational learning in workflow systems." *Proceedings of AIS 98*, Baltimore, Aug. 14-16, 1998.