A Workflow-centric Study of Organizational Knowledge Distribution

J. Leon Zhao
Department of MIS
University of Arizona
Tuscon, AZ 85721
Tzhao@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
Esthor@stern.nyu.edu

October 1999

A Workflow-centric Study of Organizational Knowledge Distribution

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 mechanisms to efficiently distribute knowledge such as news releases, seminar announcements, and memos. While the machinery for information storage, manipulation, and retrieval exists, research dealing directly with knowledge 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 knowledge distribution. The main contributions of this study include: (1) a workflow perspective on organizational knowledge distribution, (2) workflow analysis of two new knowledge distribution methods based on dynamic mailing lists and profile matching, respectively, and (3) a new way of matching knowledge supply and demand that extends existing information filtering algorithms.

1. Introduction

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 (Swanson, 1996; Tuomi, 1996). Research in knowledge management, organizational memory, and organizational learning has focused on the development of models and mechanisms for the capture, storage, and delivery of knowledge in an organizational setting (Akscyn et al., 1988; Paradice and Courtney, 1989; Stein and Zwass, 1995). With the explosion in the amount of information available worldwide, there is an urgent need for good solutions for this problem.

In recent years, workflow management systems (WFMS) have been deployed in organizations (Georgakopoulos et al., 1995; Stohr and Zhao, 1998; Kumar and Zhao, 1999). WFMS support the routing of documents and tasks in electronic form, thereby enabling the automation of business processes across teams, functional departments, customers, and suppliers. WFMS are natural repositories for organizational memory, especially with regard to business processes and logic (Zhao, 1998). They are also well suited for distributing organizational knowledge due to their ability to provide dynamic, as needed, connections between organizational members.

Our goal is to develop new approaches to organizational knowledge distribution by adopting a workflow perspective. We use seminar announcements sent over email as an example of knowledge distribution in our discussion. Seminars are a common means for sharing and enhancing organizational knowledge and seminar announcements are representative of other knowledge distribution formats such as memos and news offerings (Foltz and Dumas, 1992). 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 knowledge distribution.

Our approach is *workflow-centric* because we view knowledge distribution as an organizational process and investigate process-oriented solutions for it. This helps us design efficient and flexible solutions. Moreover, these solutions are asynchronous in that each step can be done independently of others as long as it is done in a permissible order with respect to the others.

Our research contributes to the literature in several ways:

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

2. The two new methods we propose, namely, dynamic mailing lists and automatic profile matching, promote a more accurate distribution of knowledge by taking user behaviors into account. In workflow with dynamic mailing lists, the user acts are logged and used to update the mailing lists. In the workflow with profile matching approach, we propose an information matching algorithm that makes use of profiles of knowledge objects and user profiles. The user profiles are continuously updated based on an analysis of user activities.

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 (Foltz and Dumais, 1992) that provide tools for blocking irrelevant incoming e-mails. Unfortunately, approaches to filtering "electronic junk" mails based on prioritizing or blocking incoming channels are not always effective (Hall, 1998).

The remainder of the paper is structured as follows. Section 2 presents preliminary concepts needed for the development of the workflow models and matching algorithms that are described in later sections. Section 3 develops three workflow process models: conventional mailing lists, dynamic mailing lists, and automatic matching of knowledge supply and demand. Section 4 delineates a two-stage matching algorithm that extends the keyword-based information filtering methods found in the literature. In section 5, we discuss several important issues underlying organizational knowledge distribution and relate our work to previous studies that proposed knowledge distribution mechanisms. Finally, Section 6 outlines the main contributions of the paper and points to future research directions.

2. Preliminaries

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

2.1. 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 have been 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.) Because of the unknown response, the seminar coordinator would like to know 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 through electronic mail. As we elaborate next, the method of announcing through mailing lists has several major drawbacks. We therefore propose a workflow-centric approach that increases the efficiency and reduces the costs of knowledge distribution.

2.2. Mailing Lists

Mailing lists are a basic tool for knowledge distribution in many modern organizations (Motiwalla, 1995; Hall, 1998). 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.
- *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-organizational 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 (Schwartz, 1998). 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 knowledge distribution using mailing lists.

2.3. Information Overload Problems with Mailing Lists

While electronic mail has made the distribution of knowledge 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 assumption is that everyone subscribing to a mailing list has the same information need. We call this the *uniformity assumption*. The adverse consequences of this assumption are more serious when the size of the list increases and when the list is used beyond its original purposes. This leads to a number of problems including junk mail, irrelevant mail, unmet needs, 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 could reduce the time spent on irrelevant messages and potentially increase the productivity of users.

The problems with the existing approach to using mailing lists for knowledge distribution stem from their static nature:

- 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.

Although mailing lists are cheap in terms of machine cost, the generation of unnecessary messages results in high human costs.

2.4. Matching the Supply and Demand for Knowledge

One way of looking at the knowledge 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 knowledge well and this results in irrelevant mails and unmet needs. On the other hand, mailing lists also have the advantages of simplicity and of reflecting the informal and formal structure of the organization.

In 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 seminar and receiver profiles.

3. Analysis of Three Workflow Processes

We investigate three workflow processes for organizational knowledge distribution. The first type is workflow with static mailing lists, which represents the conventional approach for knowledge 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 knowledge based on seminar and user profiles.

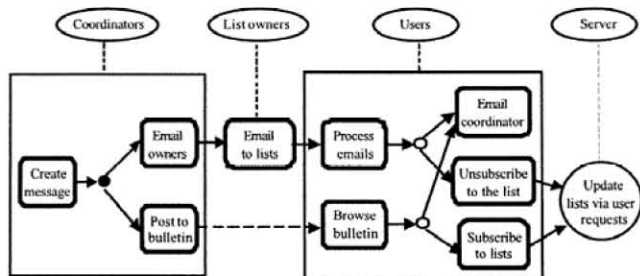### 3.1. Workflow with Static Mailing Lists

Figure 2. Workflow with static mailing lists.

The notation used in the figure is based on the Information Control Network (*ICN*) approach (Blumenthal and Nutt, 1995). 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: *Coordinators* 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 administrators
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 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 notes are relevant in the context of Figure 2:

- After the "process e-mails" 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-mails 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.

3.2. Workflow with Dynamic Mailing Lists

Workflow with dynamic mailing lists is illustrated in Figure 3. The main activities are listed below. Activities depicted in bold font differ from those in the static mailing list system.
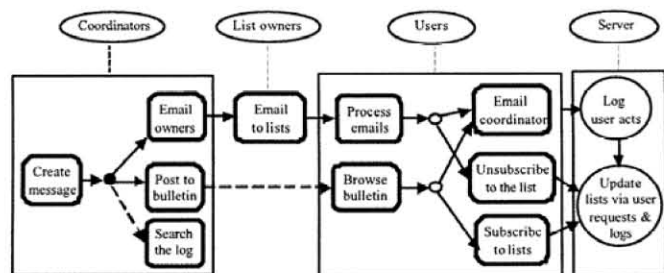

Figure 3. Workflow with dynamic mailing lists.

- *Coordinator*
1) Create message for the seminar
2) Send message to mailing list administrators
3) Post message to the electronic bulletin
4) **Search the log files for the number of potential attendees and their e-mail addresses useful 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, and reply to messages
7) E-mail to coordinators if additional information is needed
8) Unsubscribe to 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**
12) Update mailing lists based on user requests and **logs of user actions**

Note that in Figure 3, the arrow originates from the rectangle enclosing the users' activities to the server logging activity, indicating that the logging may occur for more than one activity by the users. This is a way to simplify the diagram.

This new workflow is characterized by the logging of user actions with respect to both e-mail messages and the electronic bulletin. The results are two-fold: (1) The server now has a mechanism to update the mailing lists, and (2) Coordinators can find out easily who is interested in attending the seminars. Furthermore, a coordinator can also interact more precisely with potential attendees in case of changes to the seminar such as a change of venue and time. Compared with the workflow with static mailing lists, this workflow includes three new activities: logging of user acts 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 explained next:

- An automatic logging component is needed to capture user actions with respect to seminar announcements. This could be done by adding buttons to the e-mails 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".
- 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 (1998) 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 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 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.

Note, however, that administrative lists (as defined in Section 2) cannot, by definition, be updated. In other words, dynamic mailing lists are not effective for administrative lists. Additional mechanisms are required to improve accuracy in the matching of the supply and demand for knowledge as described next.

It should also be noted that there is a privacy issue involved when people have to share information about their plans with others. However, we believe that this is not a serious problem within an organization because the information is not of a personal nature. Moreover, 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 (Grudin and Palen, 1997).

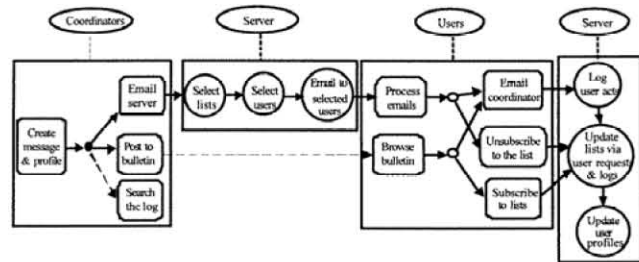### 3.3. Workflow with Automatic Matching

Figure 4. Workflow with automatic 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 the mandatory nature of administration. 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 knowledge 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 design 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) Query the log files for the number of potential attendees and their e-mail addresses to communicate any changes to the seminar.
- *Server*
5) **Select relevant mailing list based on mailing list profiles using a list matching algorithm**
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, and reply to 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 and categorize user acts
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 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 main feature of the new workflow is that it attempts to send a message to only relevant mailing lists, and to only interested users within the relevant mailing lists.

Note that under the profile-based approach, the user and 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. Profile-Based Matching of Knowledge 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.

4.1. Contents and Data Structure of Profiles

Profiles for seminars, mailing lists, and users must be created. The information for the seminar profile is in the seminar announcement and user profiles can be obtained from the individual users or constructed from the personnel profiles available in most knowledge-oriented organizations such as universities and research laboratories. The mailing list profiles are obtained by combining the user profiles of the list. The profiles may be updated periodically.

4.1.1. Seminar Profile

A seminar 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, related research 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. We denote an instance of a seminar profile by $s_i(a_1, a_2, a_3, a_4, a_5, a_6, a_8, a_9, a_{10})$, or simply $s_i<a_j \mid j=1 \; to \; 10>$, where the attributes are ordered in the sequence in which they appear in the SP table.

4.1.2. Mailing List Profile

A mailing list profile should include the following data:
- List affiliation: Name of department with which the list is affiliated.
- Type of mailing list: administrative, information, or discussion
- Collective interests: disciplines and topics of interest

These data elements can be structured into a table: **MLP(MID, L-Dept, ML-Type, I-Disc, I-Topics)**

MID is the mailing list identifier. For ease of representation, we denote an instance of a mailing list profile by $m_i(b_1, b_2, b_3, b_4, b_5)$, or simply $m_i<b_j \mid j=1 \; to \; 5>$, where attributes are ordered in the sequence in which they appear in the MLP table.

4.1.3. 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 and topics of interest to the user.
- User receiving attitude: aggressive, moderate, or conservative

These data elements can be defined in the UP table: **UP(UID, U-Dept, U-Name, U-E-mail, U-Disc, U-Topics, U-Attitude).**

We denote an instance of a user profile by $u_i(c_1, c_2, c_3, c_4, c_5, c_6, c_7)$, or simply $u_i<c_j \mid j=1 \; to \; 7>$, where the attributes are ordered in sequence in which they appear in the UP table.

4.2. A Two-Stage Matching Algorithm

The disciplines and topics within each discipline can be extracted from annual research reports of the institution and updated using the seminar profiles found in seminar announcements. 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 mailing lists can reduce computation costs while matching user profiles reduces information overload.

4.2.1. 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 profile s, and all mailing lists, M, the following algorithm is applied:

```
FOREACH mailing list m_j ∈ M DO
    IF (s<a_8> Overlap m_j<b_4>) THEN
        IF (s<a_9> Overlap m_j<b_5>) THEN
            Insert m_j into M^S
        ENDIF
    ENDIF
ENDFOR
```

Note that $s<a_8>$ and $s<a_9>$ are the relevant seminar disciplines and topics, respectively. Similarly, $m_j<b_4>$ and $m_j<b_5>$ are the interesting disciplines and topics, respectively, from the mailing list. The function *Overlap* is a matching operator that returns True if there is one or more common elements in the designated s *and* m_j 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.

4.2.2. User Matching

For user matching, we apply a matching algorithm similar to the one used for the screening of mailing lists. Given a seminar profile s, the following matching algorithm is applied:

```
FOREACH mailing list m ∈ M^S, DO
    FORALL users u_j in m DO
        IF (s_i<a_8> Overlap u_j<c_5>) THEN
            IF (s_i<a_9> Overlap u_j<c_6>) THEN
                Insert u_j into U
            ENDIF
```
```
        ENDIF
    ENDFOR
ENDFOR
```

The users in the user set U are considered to be potential attendees of the seminar. Note that $s<a_8>$ and $s<a_9>$ are the relevant seminar disciplines and topics, respectively. Similarly, $u_j<c_5>$ and $u_j<c_6>$ are the interesting disciplines and topics, respectively, of the user profile.

4.3. Customizing to Individual Users

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 individual "receiving attitude" - *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 original matching algorithm can be used for "conservative users."

4.4. 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.

Say, the seminar profile is

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", "Martin Jones", "martin-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. Since the user has an "aggressive" receiving attitude, we consider that a match is achieved. User U0005 does not achieve a match with the seminar in either the disciplines or the topics.

Under the Economics mailing list, user U0007 matches the seminar on discipline "applied economics," but not on any topic. Because the user selected a "conservative" receiving attitude, we declare that no match is achieved. On the other hand, user U0009 matches the seminar in both disciplines "applied economics" and "political economics" and in topics "China economic reform" and "political systems".

The example illustrates how the two-stage matching is done 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.

4.5. Discussion of the Matching Algorithm

Most information filtering algorithms are focused on full text search using keywords (Foltz and Dumais, 1992; Goldberg et al., 1992). Many algorithms rely on the automatic extraction of keywords from the texts (Kindo et al., 1997). 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 (Foltz and Dumais, 1992). 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 dictionaries are often available in documents such as annual research reports and personnel resumes. In summary, the limited scope of our matching tasks enables a simpler 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. The first email would be sent to everybody above a certain threshold. For subsequent reminder emails, 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.

4.6. Update of User Profiles

User profiles can be updated in response to 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.

(1) 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 her to consider selecting new disciplines and topics to add to her user profile.

(2) 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.

5. Related Work

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

5.1. Ad Hoc Workflow Management Systems

Workflow Management Systems (WFMS) can be classified on a number of different criteria (Georgakopoulos et al., 1995). A common scheme is to divide them according to 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 (Voorhoeve and Van der Aalst, 1997). 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 knowledge 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.

5.2. Advanced Electronic Mailing Systems

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 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.

6. Conclusions

This paper described a workflow-centric approach to the distribution of knowledge. Our objective was to support the communication component of knowledge 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 share encoded knowledge.

We proposed two new workflow processes that improve the accuracy of knowledge distribution by (a)