Intelligent Information Sharing Systems

Thomas W. Malone
Kenneth Grant
Franklin Turbak
Stephen A.Brobst
MIT

Michael D. Cohen
University of Michigan

90s: 86-025

December 1986

CISR WP# 147
Sloan WP# 1850-86

A slightly revised version of this paper appears in
Communications of the ACM, 1987, 30.

Management in the 1990s
Sloan School of Management
Massachusetts Institute of Technology

Management in the 1990s

Management in the 1990s is an industry and governmental agency supported research program. Its aim is to develop a better understanding of the managerial issues of the 1990s and how to deal most effectively with them, particularly as these issues revolve around anticipated advances in Information Technology.

Assisting the work of the Sloan School scholars with financial support and as working partners in research are:

American Express Company
BellSouth Corporation
British Petroleum Company, p.l.c.
CIGNA Corporation
Digital Equipment Corporation
Eastman Kodak Company
Ernst & Young
General Motors Corporation
International Computers Ltd.
MCI Communications Corporation
United States Army
United States Internal Revenue Service

The conclusions or opinions expressed in this paper are those of the author(s) and do not necessarily reflect the opinion of the Massachussetts Institute of Technology, the Management in the 1990s Research Program, or its sponsoring organizations.

Acknowledgements

This research was supported in part, by Citibank, N.A., by the Management in the 1990s Program at the Sloan School of Management, MIT, by an equipment grant from the Xerox Corporation, by the Center for Information Systems Research, MIT, and by National Science Foundation Grant No. SES-8213169. Portions of this article appeared previously in [Bro86] and [Mal86a].

Much recent work on the Information Lens system has been done by Kum-Yew Lai, Ramana Rao, and David Rosenblitt. Kevin Crowston implemented and maintains the mail gateway between our system and the outside network world, Chee-Seng Chow helped implement the first version of the central mail sorter, and Constance Perin collaborated in some of the early studies of information sharing in organizations. We gratefully acknowledge all these contributions.

Summary

As it becomes both technically and economically feasible to send electronic messages and other documents to large numbers of possible recipients, the problem of deciding who should receive a particular piece of information will become increasingly important. This paper focuses on the application of techniques from artificial intelligence, user interface design, and organizational science to help people share interesting and relevant information without being inundated by the potentially vast amount of less useful information.

Much of the work that has been done on computer-mediated communication systems, such as electronic mail, computer conferencing, and electronic bulletin boards, has focused on technical capabilities and standards for transporting and storing messages. In this paper, we propose to shift our focus to a more general problem that we will refer to as the information sharing problem. The information sharing problem involves disseminating information such that the information reaches those people to whom it is valuable and yet does not interfere with the information processing of people who will find no value in its contents.

We argue that this problem will become increasingly important with advances in communication technology and we will suggest three fundamental approaches to its solution. Our characterization of these approaches will be illustrated with (1) examples from a series of informal studies that we have conducted on how people share information in organizations, and (2) a description of an intelligent information sharing system, called the Information Lens, that we have developed.

It is already a common experience in mature computer-based messaging communities for people to feel flooded with large quantities of electronic "junk mail" [Den82; Pal84; Wil84; Hil85], and the increasing availability of inexpensive communication technology has the potential to overwhelm people with even more messages that are of little or no value to them. In current systems, people often adopt crude methods--such as removing themselves entirely from distribution lists that are of occasional interest--in order to avoid being inundated with information. At the same time, it is also a common experience for people to be ignorant of facts that would facilitate their work and that are known elsewhere in their organization. The system we will describe uses techniques from artificial intelligence and user interface design to help solve both these problems: it helps people filter, sort, and prioritize messages that are already addressed to them, and it also helps them find useful messages they would not otherwise have received.

We believe that the most important use of systems like this will not be to just reduce the flow of "junk mail", but instead to dramatically increase the amount of useful information that can be exchanged electronically without leading to information overload. Many previous communication technologies (e.g., printing press, broadcast television, photocopiers) have increased the amount of information

exchanged but their usefulness has been, to some extent, bounded by the limits of information overload (e.g., [Poo83]). Computer technology, unlike these earlier communication technologies, can not only increase the amount, speed, and distance of information flow, it can also increase the selectivity with which the information is disseminated.

Viewing the problem of information sharing in this general way suggests that much of the work that has been done separately on information retrieval systems (such as bibliographic search) and on database management systems is potentially relevant to the design of computer-mediated communication systems. Our discussion will show how concepts from both these fields are incorporated in the Information Lens system and how information sharing systems like this can help integrate, from a user's point of view, communications with both other kinds of systems. One of the key ideas behind the system is that many of the unsolved problems of natural language understanding can be avoided in intelligent information sharing systems by using semi-structured templates (or frames) for different types of messages. These templates are used by the senders of messages to facilitate composing messages in the first place. Then, the same templates are used by the receivers of messages to facilitate constructing a set of rules for filtering and categorizing messages of different types.

BACKGROUND

The most common approach to the information sharing problem in current electronic messaging environments is the establishment of some form of distribution list or conferencing structure based on interest groups (e.g., see [Hil78]). In such a scenario, users associate themselves with centralized distribution lists or conference topics which are used to disperse messages pertaining to particular subjects. In most cases, people may associate themselves with as many or as few distribution lists as desired. Senders can then address messages to these distribution lists or conference topics without having to know the specific individuals who will receive the messages.

Although distribution lists and conferences are topical in nature, their focus is generally much broader than the interests of a single person. Furthermore, since these methods of disseminating information are often targeted for relatively large audiences, it is usually impossible for the content of all information distributed to be of interest to all recipients. The users implicitly specify a general area of interest by associating themselves with particular distribution lists or conferences, but they

are ultimately responsible for deciding on a per message basis which information is of interest within this context.

Another method of dealing with the information sharing problem is to use keyword filtering and other techniques developed for large text retrieval systems (e.g., [Sal83]). Such techniques have rarely been applied to computer-mediated communication systems (see [Sta83], for an exception), and we believe they hold substantial promise. Traditional text retrieval systems, however, have their own limitations (e.g., see [Bla85], [Sal86]). We will see below how information sharing systems that support the composition of documents as well as their retrieval allow much more sophisticated retrieval possibilities.

Definition of "filtering." Before proceeding, it is important to clarify our use of the term "filtering." Even though the term has a literal connotation of leaving things out, we use it here in a more general sense that includes selecting things from a larger set of possibilities. As noted above, this positive kind of filtering (i.e., selection) may be much more important in information sharing systems than the negative kind of filtering (i.e., removal).

Studies of Information Sharing in Organizations

To help understand the kinds of automated aids for information sharing that would be desirable, we conducted several preliminary studies of how various kinds of information are shared in organizations. We performed relatively unstructured interviews of 19 people about their information filtering experiences, needs, and desires. We also conducted more structured interviews of over 30 people that focused on different kinds of information filtering environments: (1) processing the contents of their in-box, (2) reading and dealing with electronic mail, (3) scanning electronic bulletin boards, and (4) examining a table of contents as a basis for deciding which articles to read in a journal. In each case, we asked the subjects to explain in detail why they made the filtering decisions they did. Then we used these "expert protocols" to identify the general processes and specific kinds of knowledge and decision rules that people used for extracting valuable information from a large pool of available sources.

One valuable outcome of these organizational studies was the refinement of our notions of three different approaches that could be incorporated in an automated message filtering system. We will refer to these techniques as the *cognitive, social,* and *economic* approaches to information filtering. In

the following sections these approaches will be described and illustrated with examples from our studies.

Cognitive filtering

The cognitive filtering approach relies on characterizing the contents of a message and the information needs of potential message recipients and then using these representations to intelligently match messages to receivers. Distribution lists and simple keyword matching are rudimentary forms of this approach.

When making use of the cognitive approach, people in our studies often identified the content of a message by matching against certain keywords or phrases in its text. From this assessment of the content of a message, a decision is made regarding its applicability to current information needs. Throughout our organizational studies, examples of the cognitive approach were mentioned frequently by the people with whom we worked. Some examples of people using this filtering technique are given below.[1]

- *That one is important...anything to do with new product lines is a high priority item.* - engineer in a large company, electronic mail experiment.
- *That seminar on [pattern recognition] would have been of interest to me if it wasn't on a Tuesday afternoon.* - research staff member, bulletin board study.
- *I would do an exclusive scan looking for telecommunications articles except for voice mail. In [other subjects] I would do the opposite - I would search for a word that would trigger a hit.*- corporate executive, table of contents study.
- *That article sounds too OR-ish, so I wouldn't be interested.*- research administrator, table of contents study [describing an article about operations research ("OR")].

Notice that these decisions are often based on much more sophisticated heuristics than a simple keyword search. Complex combinations of both positive and negative keyword filtering techniques were used. People also used other kinds of information (such as the day of the week on which a meeting occurs) that are not easily recognized using keyword methods, and they used characteristics of items (such as "OR-ish") that may not occur in the text at all. The Information Lens system is designed to support a variety of these cognitive filtering techniques.

Social Filtering

The social approach to information filtering focuses on supporting the personal and organizational interrelationships of individuals in a community. This approach complements the cognitive approach by judging the potential value of a message based not just on its topic, but also on the characteristics of its sender.

One important dimension used for filtering by participants in our study was the organizational relationship between the sender and the receiver. For example, one participant said: *This message is from my supervisor. That means it's probably of importance to me.* The use of organizational relationships for filtering can often be quite subtle. In one case, a person focused on a particular message in his electronic mailbox because it was sent by his group leader's secretary: *I know that she would not have sent me anything, so the message must be on [my group leader's] behalf.* These organizational relationships often interact in complex ways with personal relationships as well: *"I would have to respond to a memo based on several requirements...if it came from James Long[2] [chairman of the board]...I would react quickly - that would be an alert. If it came from Sarah Rowe...who is next on the hierarchy...that would be another alert flag. If it came from John Ryan...who is my direct manager...that would be a real immediate alert flag - because...I deal with him [personally] - I work with him, assisting him on matters."* We will see below how this kind of filtering can be used in the Information Lens system.

Economic Filtering

The economic filtering approach relies on various kinds of cost-benefit assessments and explicit or implicit pricing mechanisms. In our studies, we found that persons often had to make a cost versus value decision to determine whether or not to process a particular piece of information. The length of a message was one of the primary factors used by recipients to estimate its cost. For example, one person described his filtering mechanism as follows: *I am extremely tight on time this week...I don't have the spare moments to read any messages with more than a screenful of text unless they are extremely important.*

A more subtle economic filtering technique employed by some of the subjects in our studies was to use the cost of the message to its sender as a consideration in evaluating the potential information gain from a message. For example, a secretary to the president of a small company explains her filtering

process as follows: *He [the president] is too busy to look at all the mail that comes into his office. I try to weed out anything that looks like a form letter or has a bulk mailing label on it.* If the number of recipients of some piece of information is very high (*i.e.*, form letters and bulk mailings), then the cost per person of formulating the information will be fairly small to the sender. However, the optimal piece of information to a reader is often one that is customized to his needs--at great expense to the sender of the information. This is a good explanation for the fact that people almost always read electronic mail addressed exclusively to them. The information content in these messages has presumably been chosen especially for the particular needs and interests of its recipient.

There is, however, an important tradeoff that takes place between quality and personalization. Even though published articles have a large number of recipients, the lack of personalization is often implicitly outweighed by higher quality. In fact, publications with a larger audience seem to have more appeal, at least in part, because they can afford to publish higher quality information. By spreading the cost over a larger number of subscribers, a large volume publication can afford to absorb the higher cost of providing extremely well formulated and carefully selected information to its readers. Insuring this high level of quality is, of course, one of the primary goals of an editor. The Information Lens currently includes the potential to filter on criteria such as whether the message is addressed to the receiver individually or to a distribution list. We will see below how much more elaborate economic filtering mechanisms, such as pricing schemes, could also be used.

THE INFORMATION LENS SYSTEM

In order to explore possibilities for these different approaches to information sharing, we have developed a prototype information sharing system called the Information Lens. The current version of this system emphasizes the cognitive approach to information filtering and exploits, in a simple way, techniques from artificial intelligence such as frames, production rules, and inheritance. Our discussion below will suggest how more elaborate kinds of processing and other approaches to filtering could also be included in the same general framework.

Key ideas

There are five key ideas that, together, form the basis of the Information Lens system. Though some of these ideas are empirically testable hypotheses, we treat them here as premises for our system design. We will list and briefly describe these ideas here. In the next sections, we will describe in more detail how the Lens system uses them:

(1) *A rich set of semi-structured message types (or frames) can form the basis for an intelligent information sharing system.* For example, meeting announcements can be structured as templates that include fields for "date", "time", "place", "organizer", and "topic", as well as any additional unstructured information. There are three reasons why this idea is important:

(a) *Semi-structured messages enable computers to automatically process a much wider range of information than would otherwise be possible.* By letting people compose messages that already have much of their essential information structured in fields, we eliminate the need for any kind of automatic parsing or understanding of free text messages while still representing enough information to allow quite sophisticated rules to process the messages.

(b) *Much of the processing people already do with the information they receive reflects a set of semi-structured message types.* In our informal studies we found that people often described their filtering heuristics according to categories of documents being filtered (e.g., *This is a brochure advertising a seminar. I usually throw these away unless the title intrigues me or unless it looks like a brochure I could use as a model for the ones I write.* -- research administrator, in-box study).

(c) *Even if no automatic processing of messages were involved, providing a set of semi-structured message templates to the authors of messages would often be helpful.* Two of the people in our informal interviews mentioned simple examples of this phenomenon: one remarked about how helpful it would be if any memo requesting some kind of action included, in a prominent place, the deadline by which the action needed to be taken; a second commented about how wonderful it would be if all the meeting invitations he received included a field about why he was supposed to be there. We will see below how message templates can be provided in a flexible way that encourages, but does not require, their use.

(2) *Sets of production rules (that may include multiple levels of reasoning, not just Boolean selection criteria) can be used to conveniently specify automatic processing for these messages.*

(3) *The use of semi-structured message types and automatic rules for processing them can be greatly simplified by a consistent set of display-oriented editors for composing messages, constructing rules, and defining new message templates.*

(4) The definition and use of semi-structured messages and processing rules are simplified if the message types are arranged in a frame inheritance lattice.

(5) The initial introduction and later evolution of a group communication system can be much easier if the process can occur as a series of small changes, each of which has the following properties: (a) individual users can continue to use their existing system with no change if they so desire, (b) individual users who make small changes receive some immediate benefit, and (c) groups of users who adopt the changes receive additional benefits beyond the individual benefits.

System overview

In order to provide a natural integration of this system with the capabilities that people already use, our system is built on top of an existing electronic mail system. Users can continue to send and receive their mail as usual, including using centrally maintained distribution lists and manually classifying messages into folders. In addition, the Lens system provides four important optional capabilities: (1) People can use structured message templates to help them compose their messages; (2) Receivers can specify rules to automatically filter and classify messages arriving in their mailbox; (3) Senders can include as an addressee of a message, in addition to specific individuals or distribution lists, a special mailbox (currently named "Anyone") to indicate that the sender is willing to have this message automatically redistributed to anyone else who might be interested; and (4) Receivers can specify rules that find and show messages addressed to "Anyone" that the receiver would not otherwise have seen.

By gradually adding new message types and new rules, users can continually increase the helpfulness of the system without ever being dependent on its ability to perfectly filter all messages.

System architecture. The Lens system was written in the Interlisp-D programming environment using Loops, an object-oriented extension of Lisp. The system runs on Xerox 1108 and 1109 processors connected by an Ethernet. We use parts of the Lafite mail system and the XNS network protocols already provided in that environment. The message construction aids and the individual filtering rules all operate on the users' personal workstations.

As Figure 1 illustrates, messages that include "Anyone" as an addressee will be delivered by the existing mail server directly to the explicit addressees as well as to an automatic mail sorter that runs on a workstation and periodically retrieves messages from the special mailbox. This automatic mail sorter may then, in turn, send the message to several additional recipients whose rules selected it.

Implementation status. The Information Lens system as described in this paper currently exists in prototype form. As of this writing, the system has been in regular use by about five members of our research group for several months, and planning for larger scale tests is underway.

Messages

The Lens system is based on a set of semi-structured messages. For each message type, the system includes a template with a number of fields or slots for holding information. Associated with each field are several properties, including the default value of the field, a list of likely alternative values for the field, and an explanation of why the field is part of the template.

Figures 2 and 3 show a sample of the highly graphical interaction through which users can construct messages using these templates (see [Tou82] for a similar approach to constructing database retrieval queries). After selecting a field of a message by pointing with a mouse, the user can point with the mouse again to see the field's default value, an explanation of the field's purpose, or a list of likely alternatives for filling in the field. If the user selects one of these alternatives, that value is automatically inserted in the message text. The user can also edit any fields directly at any time using the built-in, display-oriented text editor. For example, the user can add as much free text as desired in the text field of the message.

By providing a wealth of domain-specific knowledge about the default and alternative values for particular types of messages, the system can make the construction of some messages much easier. For example, Figure 3 shows how some message templates, such as a regular weekly meeting announcement, may have default values already filled in for most of their fields and require only a few keystrokes or mouse clicks to complete and send off.

Users who do not want to take advantage of these message construction aids can simply select the most general message type (*message*) and use the text editor to fill in the standard fields (To, From, and Subject) just as they would have done in the previous mail system. We expect, however, that the added convenience provided to the senders by semi-structured templates will be a significant

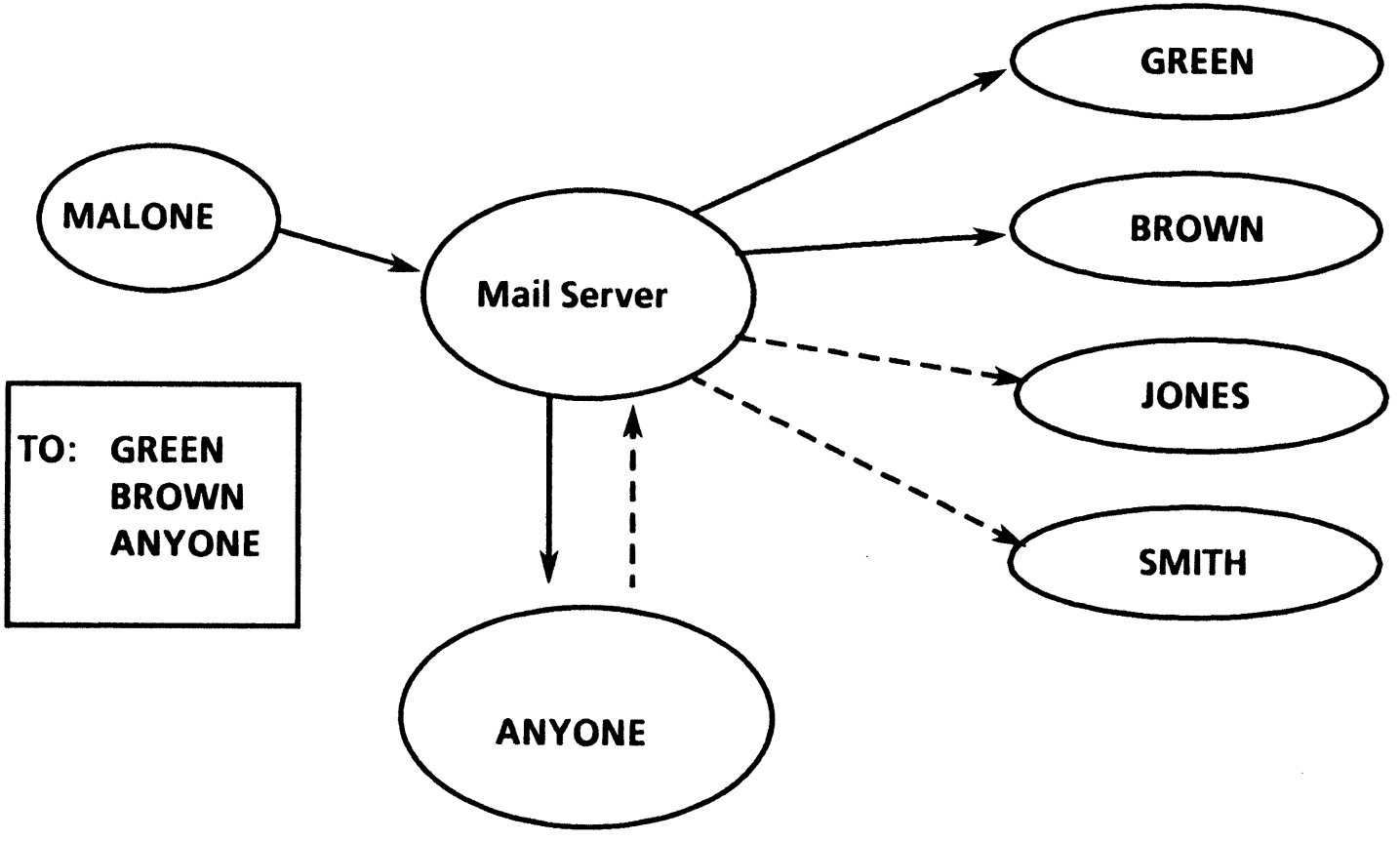

Figure 1

The Lens system includes components in the users' workstations and in a central server (called "Anyone"). Messages that include "Anyone" as an addressee are automatically distributed to all receivers whose interest profiles select the messages as well as to the other explicit addressees.

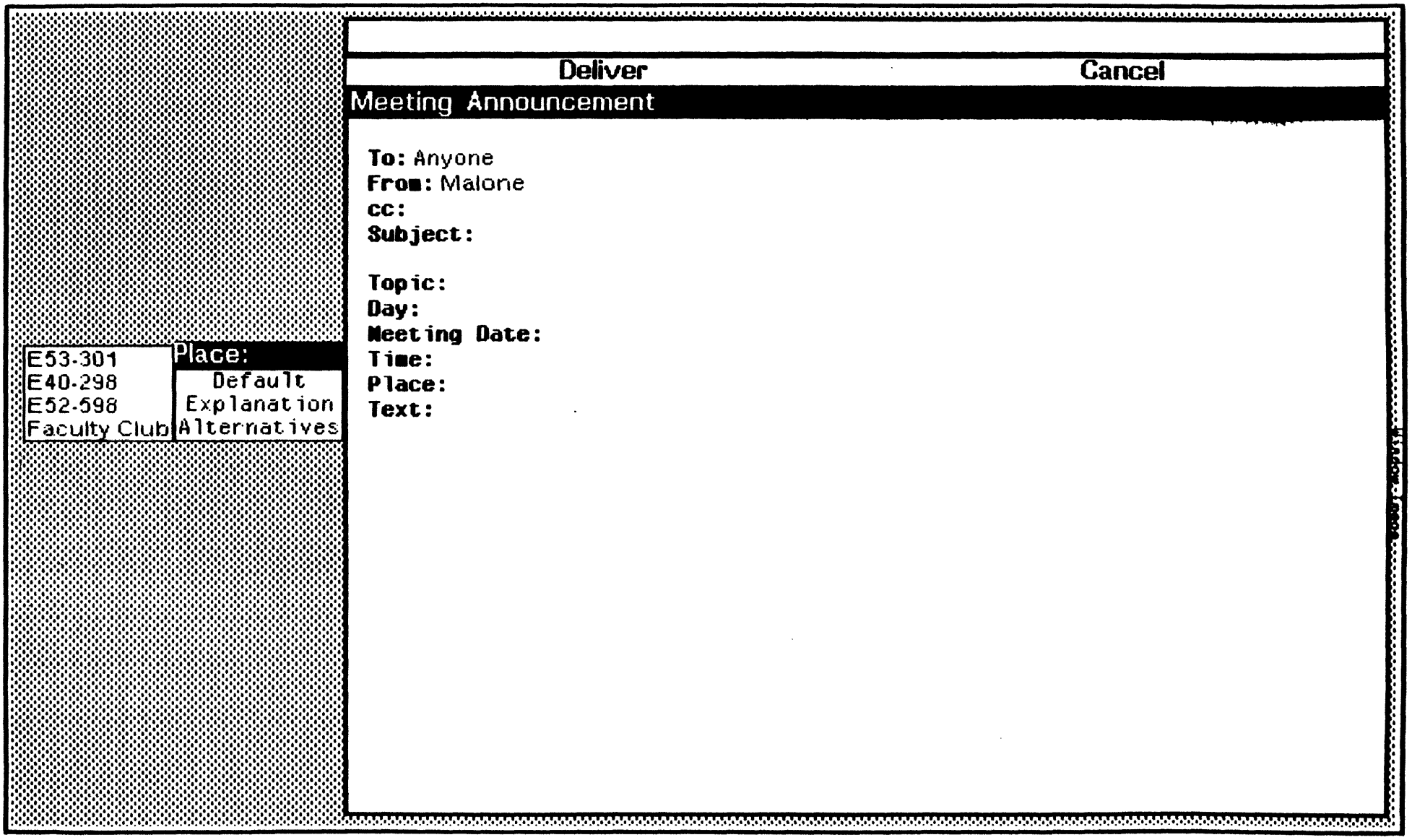

Figure 2

Messages are composed with a display-oriented editor and templates that have pop-up menus associated with the template fields.

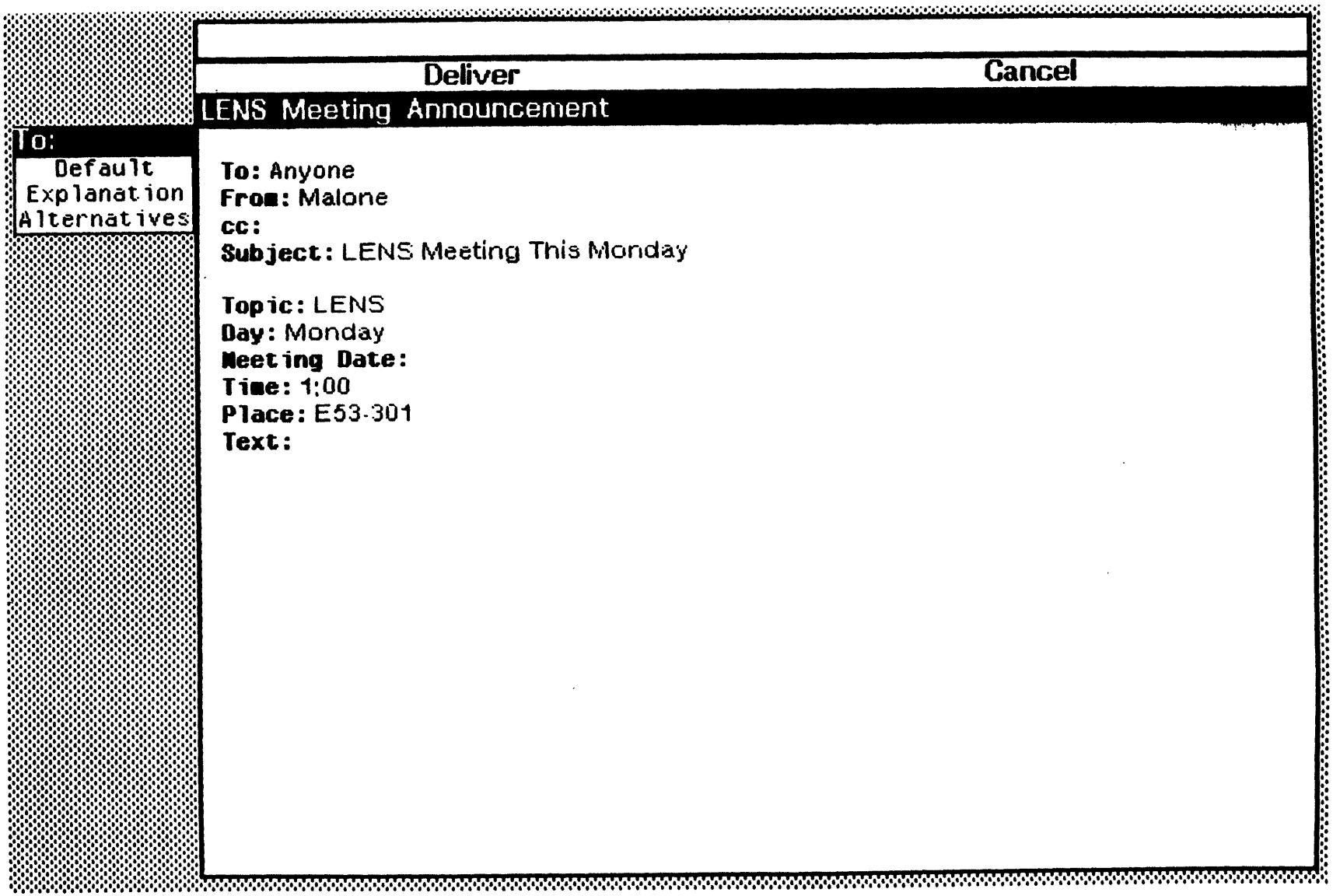

Figure 3

Some templates already have a number of default values filled in for different fields.

incentive for senders to use templates in constructing some of their messages. This, in turn, will greatly increase the amount of information receivers can use in constructing automatic processing rules for incoming messages.

Direct manipulation. Both the message editor and the rule editor (described below) embody the features that Shneiderman [Shn82] uses to characterize "direct manipulation" interfaces: "(1) continuous representation of the object of interest, (2) physical actions or labelled button presses instead of complex syntax, and (3) rapid incremental reversible operations whose impact on the object of interest is immediately visible." We believe that these features help explain the simplicity of using these editors.

Message types

To further simplify the construction and use of message templates, the templates are arranged in a network so that all subtypes of a given template *inherit* the field names and property values (e.g., defaults, explanations, and alternatives) from the *parent* template. Any subtype may, in turn, add new fields or override any of the property values inherited from the parent (e.g., see [Fik85]). For example, the *seminar announcement* template adds a field for *speaker* that is not present in its parent template *meeting announcement.* The *LENS project meeting announcement* (Figure 3) adds a number of default values that are not present in its parent. The inheritance network eliminates the need to continually re-enter redundant information when constructing new templates that resemble old ones, and it provides a natural way of organizing templates, thus making it easier for senders to select the right template.

The message type lattice is made visible to the user through the message type browser. Figure 4 shows this lattice browser for our sample network of message types. Users select a template to use in constructing a new message by clicking with the mouse on the desired message type in this browser. By clicking with a different mouse button, users can view or modify the rules (see below) associated with a particular message type. Like the other message type characteristics, these rules are inherited by the subtypes of a message template. Thus, for example with the network shown in Figure 4, the rules for processing "notices" and "messages" would be applied to incoming "meeting announcements" as well as the rules specifically designed for meeting announcements.

Group definition of message types. The network shown in Figure 4 includes some message types that we believe will be useful in almost all organizations (e.g., meeting announcements) and some that are important only in our environment (e.g., LENS project meeting announcement). Different groups can develop detailed structures to represent the information of specific concern to them. For example, a

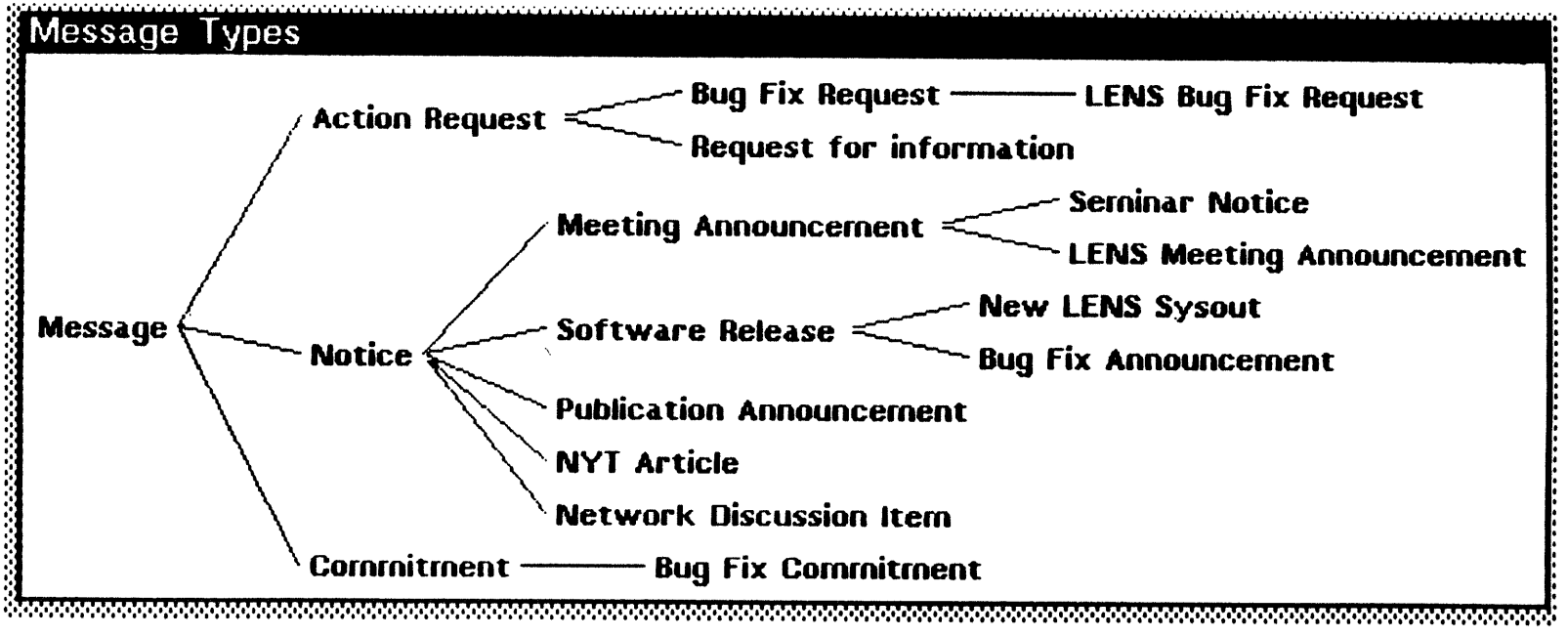

Figure 4

The message templates are arranged in a network with more general types at the "top" (shown at the left) and more specific types at the "bottom" (shown at the right).

product design team might have an elaborate network of message types describing different aspects of the product (*e.g.*, market size estimates, response time estimates, alternative power supply vendors). Then, for instance, marketing specialists who believe that the critical factors determining potential market size for the product are cost and response time can devote most of their attention to the messages concerning these factors and ignore all the rest of the technical specifications for the product.

We are developing another display-oriented editor, like the message editor shown in Figures 2 and 3, for creating and modifying the template definitions themselves. We expect that in some (e.g., rarely used) regions of the network anyone should be able to use this "template editor" to modify an existing message type or define a new one, while in other regions, only specifically designated people should have access to this capability. In the current version of the system people can use a simple version of this editor to personalize the *default, explanation,* and *alternatives* properties of the fields in existing message types.

A given user's *personal profile* consists of these customizations of the message types, together with a set of personal rules for processing messages (see next section), and a set of hierarchically arranged folders in which messages can be stored.

Message purpose. One characteristic of message classes that is critical in formulating filtering rules is the *purpose* of the message. We expect that an important part of the frame inheritance network for message types will be a taxonomy of the various *communication acts* that a message might embody (e.g., [Sea75; Ked84; Win86]). For example, messages whose purpose is to *request information* should be routed to people who know about the topic of the message, while messages whose purpose is to *provide information* should be routed to people who are interested in the topic of the message.

Rules

The Lens environment allows users to build rules for finding, filtering, and sorting messages. Rules consist of a test and an action; if a message satisfies the test, then the action specified by the rule is performed on the message.

Figure 5 shows an example of the display-oriented editor used to construct rules. This editor uses rule templates that are based on the same message types as those used for message construction. We expect that this template-based graphical rule construction will be much easier for inexperienced computer users than more conventional rule or query languages.

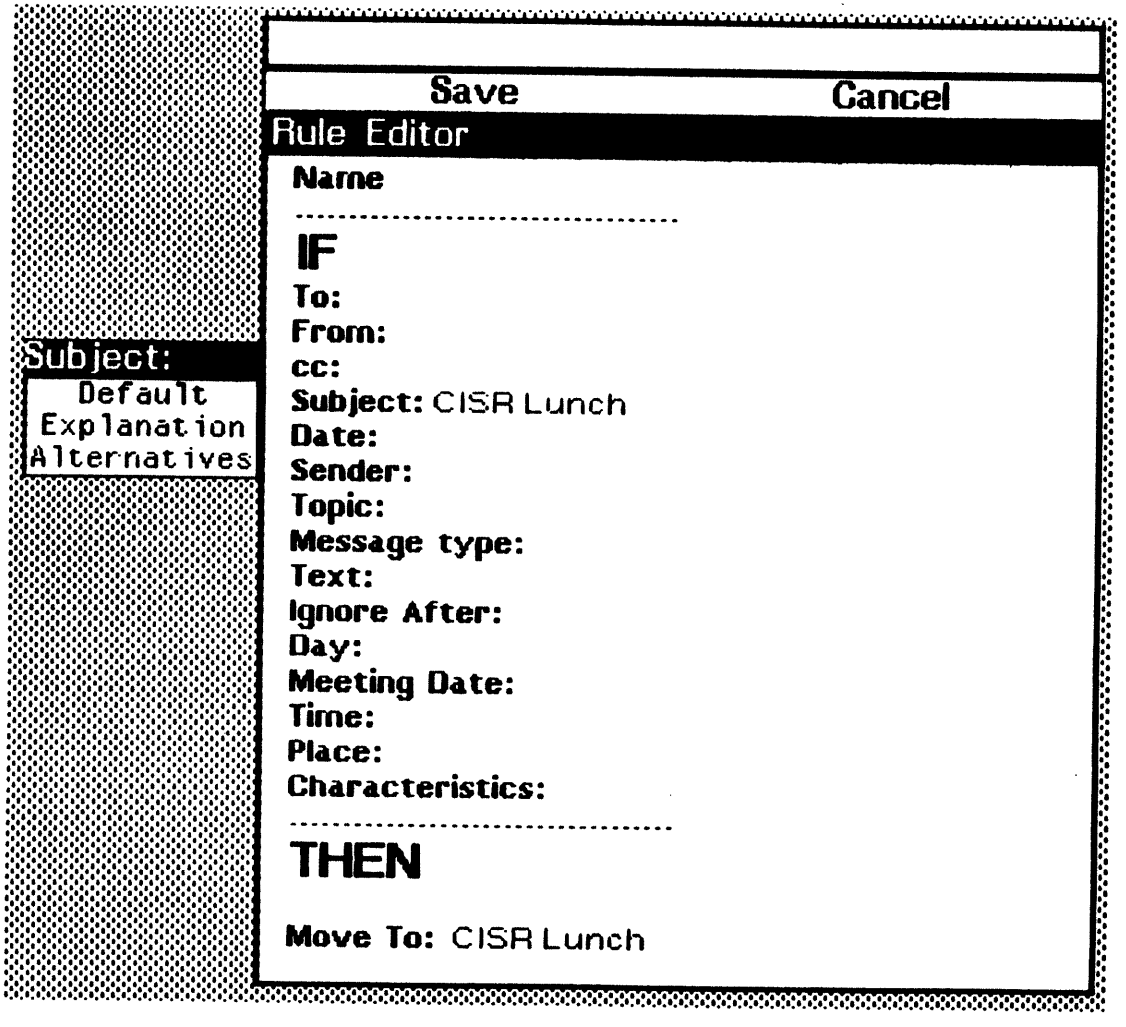

Figure 5

Rules for processing messages are composed using the same kind of editor and the same templates as those used for composing messages in the first place.

Constructing the "IF" part of a rule involves filling in selection specifications for the different fields of the message. The simplest kind of selection specification is a string that should appear somewhere in the specified field. More complex specifications for a field can be constructed by combining strings with *and*, *or*, *not*, and parentheses (i.e., arbitrary Boolean combinations are possible within a field). If specifications appear in more than one field, then all specifications must be satisfied at once for the rule to succeed (i.e., specifications in different fields are implicitly *and*-ed). As in the message editor, the default, alternatives, and explanation menus are available in the rule editor; here they facilitate the construction of selection specifications. To specify the action ("THEN" part) of a rule, users select the word "THEN" on the rule template and then choose an action from the menu of possibilities that appears. Typical actions classify messages in specific folders (Figure 6a) or delete messages (Figure 6b). In order to retain maximum flexibility, it is also possible to have arbitrary Lisp functions called as either the IF or the THEN part of a rule.

Rules can also be used to find messages of interest that are addressed to "Anyone" but which a particular user would not otherwise have seen. Figure 6c shows an example of such a rule. The *show* option for the action of a rule indicates that the central mail sorter should forward messages that satisfy the rule test to the person who constructed the rule.

Message characteristics. In our informal studies of people filtering information, we observed many instances of what could be modeled as a kind of multi-level processing: in the first phase, items were classified (e.g., "This message is from someone I don't know..." or "This article sounds too OR-ish. . .") and then, in the second phase, some action was taken (e.g., ". . .so I will throw it away"). A production rule formalism like the one we are using is, of course, well suited for representing this kind of multi-level reasoning. We capture a simple form of it by including in every rule template, in addition to the fields associated with the message type, a field for *characteristics*. Certain rules set the characteristics of messages, based on other field values, and then other rules can test messages for these characteristics. For example, Figures 6d and 6e show how we can construct a single rule that determines whether a message is from a VIP and then test for this characteristic in other rules. This kind of abstraction mechanism has obvious advantages over a mechanism that requires repeating the specifications of a VIP in all the rules that need to test for this characteristic. Although specifications in different fields are implicitly "and"-ed, the characteristics mechanism also makes it possible to construct tests that include any combination of features in any combination of fields (i.e., arbitrary Boolean combinations between fields).

Sample rules

(a)
IF Message type: Action request
Action deadline: Today, Tomorrow
THEN *Move* to: Urgent

(b)
IF Message type: Meeting announcement
Day: Not Tuesday
THEN *Delete*

(c)
IF Message type: Request for information
Subject: AI, Lisp
THEN *Show*

(d)
IF From: Silk, Siegel
THEN *Set Characteristic:* VIP

(e)
IF Message type: Action request
Characteristic: VIP
THEN *Move to* : Urgent

Figure 6

Rules can move messages to folders, delete messages, set "characteristics" to be used by other rules, or select messages addressed to "Anyone" to be shown to a given user.

Group use of message types. Individuals who begin using this system before most other people do can get some immediate benefit from constructing rules using only the fields present in all messages (To, From, Subject, Date). Groups of individuals who begin to use a set of common message types can get much greater benefits from constructing more sophisticated rules for dealing with more specialized message types. For example, a general rule might try to recognize "bug reports" based on the word "bug" in the subject field, but this would be a very fallible test. A community that uses a common template for bug reports can construct rules that deal only with messages the senders classify as bug reports. These rules can use specialized information present in the template such as the system in which the bug occurred, the urgency of the request for repair, and so forth. From the viewpoint of organization theory, we know that "internal codes" are among the most important productive assets of an organization (Mar58; Arr74). In effect, the Lens system provides a medium in which this collective language of an organization can be defined and redefined.

Future Directions

Connections with external information sources

One of the desirable aspects of the system architecture we have described is its versatility in dealing with external information sources. As indicated above, messages that are sent to users of the Lens system from people who do not use the system are simply repesented by Lens as messages of type "message." When Lens users send messages to others outside the system, all the fields in the template that are not part of the standard message header are sent as part of the text field.

It is also possible to do more intelligent translation into and out of certain message types. For instance, we currently receive daily on-line transmission of selected articles from the New York Times (via the system developed by [Giff85]). When the Anyone server receives these messages, it parses out the fields already present in the wire service feed (e.g., "title", "subject", "category", and "priority") and formats these messages as Lens templates with the same fields. Users of our system are then able to construct elaborate sets of production rules to select the articles they wish to see and sort them into categories.

One of the most interesting possibilities for such systems occurs in the formation of computer-based markets (e.g., see [Mal86b]). For example, techniques like those described here can be used to screen advertising messages and product descriptions according to precisely the criteria that are important to a given buyer. Air travellers, for instance, might specify rules with which their own "automated buyers' agents" could compare a wide range of possible flights and select the one that best matches

that particular traveller's preferences. The preferences might include decision rules for trading off between factors such as cost, convenient arrival and departure times, window seats vs. aisle seats, minimum number of stops, and so forth. A fairly simple set of such rules could, in many cases, do a better job of matching particular travellers' preferences than all but the most conscientious and knowledgeable human travel agents.

Natural language processing and information retrieval techniques

It is easy to imagine even more sophisticated facilities in this framework that could use whatever natural language understanding capabilities are available to parse unstructured documents into the templates used by Lens. The fields extracted in this way could then be used for automatic filtering or other processing after which human readers could look at the full text of selected articles to do more accurate processing themselves. We have already seen how the Lens facilities can be used without any automatic natural language understanding capabilities. However, as natural language parsers become more powerful and accurate, rules like those specified in Lens will become more useful for processing a much wider range of documents.

It is also important to realize that this framework can easily incorporate many existing techniques for information retrieval in unstructured text (e.g., see [Sal83]). For example, rules for documents that include free-text fields could specify adjacency conditions for keywords in the text fields, and other techniques such as term truncation and addition of synonyms could also clearly be used (e.g., [Sal86]).

Extended knowledge base

Our system currently includes lists of topics that are provided as alternatives for the topic field in different kinds of messages. One possible extension to our system is to generalize these topic lists into a *topic network* (similar to the network of message types that already exists) that represents topic-subtopic relationships. Then, for instance, receivers who indicated an interest in the topic "artificial intelligence" could automatically receive messages about subtopics of this topic such as "frames", "production rules", and so forth.

Another extension we would like to explore involves adding a *knowledge-base server* which will keep copies of public messages. Just as in a computer conferencing system, users will then be able to retrieve these messages at a later time even if they did not receive the messages originally. Significant power can also be added to the system by viewing field values as database objects rather than mere text strings. Rules can then access information from these *embedded objects* in deciding

how to handle messages. For example, rules could check the job title or organizational position of the message sender if these characteristics were stored in a central database.

Forms processing

One of the attractive features about the general system architecture we are using is that it is relatively easy to gradually add more and more kinds of knowledge (see [Mal86c]). For example, we would like to experiment with having the system automatically reply or otherwise respond to certain kinds of messages. Meeting announcements that meet certain criteria, for instance, can be automatically added to a calendar database. More elaborate forms processing systems can be built on top of the capabilities we already have. For instance, an accounting clerk might have rules that automatically processed travel expense reports that met certain criteria for being "routine" and saved all "non-routine" forms for the personal attention of the clerk. This kind of "fail soft" approach to forms processing is much more flexible than many traditional approaches to this problem.

Other approaches to information filtering

The Information Lens system relies primarily on the *cognitive filtering* approach. This approach relies on characterizing the information contents of a message and the information needs of potential message recipients. Even though the system depends on its human users to encode and interpret this information, the success of the system still relies on the ability of the message templates and rule templates to represent the information being communicated. The other two approaches to information filtering, *social* and *economic* filtering, suggest ways of facilitating information sharing that do not require the system to represent the content of the messages at all.

Social filtering

We have already seen how the Information Lens can include rules that process messages based on characteristics of the sender of the message. Another very widely used method of social filtering relies not just on the characteristics of the author of a message but also on the references and recommendations of other people. Perhaps the best example of this phenomenon is the everyday experience of most researchers who attempt to "keep up with the literature" in their field. It is our experience that almost no one actually scans all the journals that are potentially relevant to their interests. Instead, they read the articles which their colleagues most often refer to and recommend.

A rudimentary kind of support for this process already exists in most electronic messaging systems in the form of facilities for forwarding messages to other people who may find them interesting. A more elaborate kind of support would allow receivers to create lists of people whose opinions they value on various topics. When these people read messages, they may choose to "endorse" them. Messages can then be prioritized for a given receiver based on the number of endorsements they receive from people on that receiver's endorsement list. An obvious extension would allow assigning different weights to different endorsers and letting endorsers indicate the strength of their endorsements. For high volume topics, this process can be formalized still further by having some people perform the role of an "editor" who selects messages on the basis of both relevance and quality.

Economic filtering

The economic approach focuses on how to provide both positive and negative incentives to control the quantity and quality of information flows. The widespread use of distribution lists and electronic bulletin boards facilitates the dissemination of information to an extremely large audience with very little effort. The danger associated with this is that the very low marginal cost of sending additional copies of messages [Cra82] may encourage the proliferation of electronic messages that have little informational value to the majority of their recipients. It is the goal of the economic filtering mechanism to minimize this phenomena, while encouraging people to disseminate valuable information they might otherwise have neglected to send.

One feature of most current message systems that is highlighted by an economic perspective is that the cost of sending multiple copies of messages is borne primarily by the receivers who must process them. The economic approach to information filtering implies that it might sometimes be in the best interest of an electronic mail community to shift these incentives so that senders of messages are charged in a way that reflects some of the costs to receivers [Den82]. For example, the senders of unsolicited messages can be charged (in actual money or some point system) in proportion to the value of the time people will spend reading the messages, that is, more for long messages, more for messages sent to many people, and possibly even more for messages to highly paid recipients. In an even more extreme use of this idea, people who receive undesirable "junk mail" can indicate that fact and the sender will then be penalized by an additional surcharge. This scheme thus provides incentives to both reduce "junk mail" and to make messages more concise.

A somewhat subtler way to reduce "junk mail" is suggested by the observation above about how receivers use the cost of a message to its sender (*e.g.*, bulk rate versus first class postage) as a consideration in filtering. An electronic messaging system can use this approach by letting senders

spend limited resources to signal receivers that a message deserves more than the usual priority. Some receivers might then have "asking prices" on their mailboxes that screen out all messages below a certain cost [Den82]. Other receivers might construct rules that use the cost to the sender as one of a number of factors in filtering and prioritizing messages.

In any system where senders pay some of the costs of communication (including current systems where the cost is mostly time) there is a disincentive for people with potentially valuable information to disseminate it. To remedy this deficiency, the recipients of valuable information can sometimes compensate the sender. For example, [Tur83] describes a system that supports an internal "free market" for information and services within an organization. Certain kinds of information are highly desirable and employees who know (or have skills to obtain) this information can sell the information or "contract out" some of their time to high bidders anywhere in the organization.

Combined Approaches

It is important to realize that the most useful information sharing systems are likely to be combinations of the approaches we have described. For example, the pricing schemes are likely to work much more effectively when the topic structuring mechanisms described earlier are used to indicate which topics receivers are willing to pay for and which they would have to be paid to read about. One problem with this approach, however, is that senders might try to receive extra payments by classifying their messages into highly valued topic areas--for which the messages were actually irrelevant. This problem can be ameliorated by using social or economic filtering mechanisms. For example, "editors" can be paid for filtering information on certain topics. Editors who filter well will be able to charge more for their services and those who don't will find their "subscribers" dwindling away.

Potential problems with systems of this type

Almost any powerful technology that has the potential to benefit people also has potentials for misuse or unintended negative consequences. The system we have described is intended to help avoid some potential negative consequences of computer-mediated communication systems (e.g., information overload for individuals) and at the same time to take advantage of some even greater potential benefits (e.g., selective sharing of much more information in organizations as a whole). In order to use a new technology wisely, it is important to try to anticipate and encourage beneficial uses and to anticipate and avoid possible negative consequences. Since much of this paper has been devoted to

describing potential benefits from systems of this type, in this section, we will briefly describe a few potentials for misuse and some possible remedies.

Excessive filtering

Some people, on hearing descriptions of this system, worry that it might be used to decrease the flow of information in an organization. For instance, people might use it to filter out messages personally addressed to them and thus become less responsive to information from other people in their organization. While this is, in fact, a possible use of the capabilities we have described, we believe it is an unlikely one. The system leaves completely up to each user the decisions about how cautious or how reckless to be in specifying rules for automatic deletion of the messages they receive. There are already many social forces at work in organizations that affect how responsive people are to each other's communications, and in many cases, these forces would strongly discourage people from automatically deleting messages addressed to them personally. A much more likely scenario, we believe, is that people will use the capabilities of the system to sort and prioritize messages addressed to them personally and will use automatic deletion primarily for non-personal messages addressed to large numbers of people via distribution lists, conference topics, or bulletin boards.

In this case, of course, the ability of receivers to filter out "public" messages that are unlikely to be interesting to them increases the usefulness of the public communication channel in two ways: (1) receivers are more likely to attend to communication channels whose "richness" (i.e., probability of being interesting) is greater, and (2) senders are likely to send out more information if they are not worried about incurring the displeasure of many uninterested receivers whose mailboxes would be cluttered.

Imperfect finding

Another concern occasionally expressed about systems like this is that people may have difficulty knowing what they want and don't want to see until they have seen it. How, for instance, can you find out that another group in your organization is doing something of great interest to you if they use keywords for describing it that are unfamiliar to you? Here, of course, the relevant comparison should be, not an omniscient and perfect system, but the plausible alternatives that are available. No system, including the one described here, can do a perfect job of finding all and only the information in which a given user is interested. We believe, however, that capabilities like those we have described increase the likelihood that people will find useful information they would not otherwise have encountered.

One simple mechanism for helping people find messages they don't know they want is to give them the option of seeing some number of randomly chosen messages each day. (These messages should, of course, be chosen from the "public" messages addressed to "Anyone" not from private messages between individuals.) Some of the random messages may, in fact, be of interest and may lead their recipients to establish filters that select other similar messages in the future. A slightly more sophisticated version of this approach is to have each user's rules assign a "probable interest value" to all messages. Techniques used for document ranking (such as term weighting) could be helpful for this purpose ([Nor77] [Sal83]). The system could then show a user all the messages above some "interest threshold" and a sample of other messages that are below that threshold but are randomly selected in a way that favors messages of higher probable interest.

Excessive processing loads

In the prototype version of the Lens system, there is only one "Anyone" server for all the users of the system. Clearly, when systems like this are used on a larger scale, such a single server could easily become overloaded. It is a straightforward matter, however, to have multiple "Anyone" servers spread throughout an organization, each one, for example, serving a different group, department, or division. Each of these servers can, in turn, have rules that determine when to forward messages they receive on to other "Anyone" servers elsewhere in the organization.

Privacy concerns

Many important issues of privacy and security are raised by any computer-mediated communication system that carries personally or organizationally sensitive information. These issues are, of course, important in systems like the one we have described, but they are not unique here. For instance, it is already common in electronic mail systems to restrict the audience for certain messages by addressing the messages only to specific individuals or to distribution lists whose membership is restricted. The Information Lens system uses the underlying mail system in this way and adds one more level of "public" information (i.e., messages addressed to "Anyone"). We are also currently implementing a simple extension to the system that allows messages to be addressed to "Anyone-in-<distribution list name>". A message addressed in this way can be received only by people in a specific distribution list whose rules select it.

There are also some intriguing new possibilities raised by intelligent information sharing systems that are not present in all computer-mediated communication systems. For example, the rules about

how people filter, select, and prioritize their messages represent a new kind of potentially sensitive information that is stored in the system. Would employees, for instance, want their supervisors to know that they had filters selecting notices about job opportunities in other parts of the company? It is not clear, however, that people's rules should always be kept completely confidential. Sometimes, for instance, people may want others to know that they are interested in certain topics to encourage the formation of interest groups. There may also be times when it is desirable to tell the senders of messages addressed to "Anyone" how many people's rules actually selected the messages, without revealing the names of the recipients. Similarly, there may be times when it is desirable to display the numbers (but not the names) of people interested in different topics. Devices like this could thus provide a new kind of non-intrusive and (in some cases) non-objectionable method for conducting instant "opinion surveys" or "market research". Clearly, careful thought is needed about when and how these possibilities are desirable, but we think the possibilites are quite intriguing.

Conflicts of interest

Most of the capabilities for information sharing that are included in the current Information Lens system can be expected to work best in communities where people share goals and where there are not strong conflicts of interest about whether certain kinds of information are worthy of attention. When there are such conflicts, for example, when an "advertiser" wants you to pay attention to something that you will in fact regard as "junk mail", then filtering capabilities like those we have described can sometimes be defeated. For instance, someone who wants many people to read a particular message can indicate that the message is about a topic that is widely interesting when, in fact, the message is not about that topic at all. It is, of course, possible to evolve filters to combat such maneuvers (e.g., "delete all messages from X, regardless of the topic indicated"), but this kind of "game" can continue to escalate with each side adopting more and more subtle techniques to filter out (or filter in) the messages. We believe that situations involving conflicts of interest like this are probably better handled by the social and economic approaches to information filtering discussed above.

RELATED WORK

There are several other previous approaches to structuring information sharing in electronic communities that have been used much less widely than distribution lists, conference topics, and keyword retrieval methods. These include: (1) using associative links between textual items to

represent relationships such as references to earlier (or later) documents on similar topics, replies to previous messages, or examples of general concepts (e.g., [Eng68 ; Tri83]), and (2) representing and using detailed knowledge about specific tasks such as calendar management or project management (e.g., [Sat85]). Our system is, in some sense, at an intermediate level between these two approaches. It includes more knowledge about specific domains than simple associative links, but it can be used for communicating about any domain, even those for which it has no specific knowledge. A few systems (e.g., [McC85]) have used AI techniques such as production rules to reason about the contents of messages based on the presence or absence of keywords in unstructured text.

We have not focused here on facilitating the kind of real-time information sharing that occurs in face-to-face meetings (e.g., [Ste85; Sar84]) or teleconferencing (e.g., [Joh84]). We believe, however, that the aids we described could be useful in some real-time meetings (especially those involving very many people), and that these aids could eliminate the need for some meetings altogether.

CONCLUSIONS

As we move into an information age in which it is both technically and economically feasible for almost anyone to communicate large volumes of information almost instantaneously across long distances to many receivers, the users of computer-mediated communication systems are experiencing a desparate need for support in their information sharing task. To help solve this problem, we have identified in this paper three fundamental approaches to information filtering: cognitive, social, and economic. We have also described an intelligent system for supporting information sharing in organizations. The system appears to derive its power and simplicity from an unusual combination of ideas drawn from artificial intelligence, information retrieval, and user interface design.

We believe this system provides a prototypical example of how computer systems can be designed to include not only good *user interfaces* for supporting the problem solving of individuals, but also good *organizational interfaces* [Mal85] for supporting the problem solving of groups of people. For example, the system conforms to an important principle of organizational design: it can be superimposed on current procedures with little change, and will steadily increase in usefulness as more members of an organization employ its features.

Footnotes

[1]Italicized typeface is used to represent remarks of our subjects closely paraphrased from interview notes while quotations are reserved for verbatim transcriptions from tape recorded interviews.

[2]The names of all those persons referred to by the interviewees have been changed.

References

[Arr74] Arrow, K. *Limits of Organization.* Norton, New York, 1974.

[Bla85] Blair, D. and Maron, M. An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. *Communications of the ACM 28* ,3 (1985), 289-299.

[Bro86] Brobst, S. A., Malone, T. W., Grant, K. R. & Cohen, M. D. Toward intelligent message routing systems. In R Uhlig (Ed.), *Computer messate systems - 85: Proceedings of the Second International Symposium on Computer Message Systems.* North Hollanld, Amsterdam, 1986.

[Cra82] Crawford, A. Corporate Electronic Mail - A Communication Intensive Application of Information Technology. *MIS Quarterly,* (1982) , 1-14.

[Den82] Denning, P. Eectronic Junk. *Communications of the ACM 23,* 3, (1982) , 163-165.

[Eng68] Englebart, D. C. and English, W. K. Research center for augmenting human intellect. *Proceedings of Fall Joint Computing Congress,* AFIPS Press, December 1968, pp. 395-410.

[Fik85] Fikes, R. and Kehler, T. The role of frame-based representation in reasoning. *Communications of the ACM* 28, *7* (1985), 904.

[Gif 85] Gifford, D. K., Baldwin, R. W., Berlin, S. T., & Lucassen, J. T. An architecture for large scale information systems. *Proceedings of the 10th ACM Symposium on Operating Systems Principles* (Orcas Island, Washington, December 1985), pp. 161-170.

[Hil78] Hiltz, S. R. and Turoff, M. *The network nation: human communication via computer.* Addison-Wesley, Reading, Mass., 1978.

[Hil85] Hiltz, S. R. and Turoff, M. Structuring computer-mediated communication systems to avoid information overload. *Communications of the ACM* 28, 7 (1985), 680-689.

[Joh84] Johansen, R. *Teleconferencing and beyond: Communications in the office of the future.* McGraw-Hill, New York, 1984.

[Ked84] Kedzierski, B. Communication and management support in system development environments. *Proceedings of the Conference on Human Factors in Computer Systems* (Gaithersburg, MD, March 15-17, 1982).

[Mal85] Malone, T. Designing Organizational Interfaces. *Proceedings of the CHI '85 Conference on Human Factors in Computing Systems* (San Francisco, California, 1985).

[Mal86a] Malone, T. W., Grant, K. R., & Turbak, F. A. The Information Lens: An intelligent system for information sharing in organizations. *Proceedings of the CHI '86 Conference on Human Factors in Computing Systems* (Sponsored by ACM/SIGCHI) (Boston, Mass., April 1986).

[Mal86b] Malone, T. W., Yates, J., & Benjamin, R. I. Electronic markets and electronic hierarchies: Effects of information technology on market structures and corporate strategies. *Proceedings of the International Conference on Information Systems* (San Diego, California, December, 1986).

[Mal86c] Malone, T. W., Grant, K. R., Lai, K. Y., Rao, R, & Rosenblitt, D. Semi-structured messages as a basis for computer-supported coordination. *Proceedings of the Conference on Computer-Supported Cooperative Work*, Austin, Texas, December 3-5, 1986.

[Mar58] March, J. G. & Simon, H. A. *Organizations*. Wiley, New York, 1958.

[McC85] McCune, B. P., Tong, R.M., Dean, J.S., & Shapiro, D.G. RUBRIC: A system for rule-based information retrieval. *IEEE Transactions on Software Engineering*, Vol. SE-11,939-944, 1985.

[Nor77] Noreault, M. K., and McGill, M. J. Automatic ranked output from Boolean searches in SIRE. *Journal of the ASIS*, 1977, *28*, 333-339.

[Pal84] Palme, J. You Have 134 Unread Mail Do You Want to Read Them Now? IFIP Conference on Computer Based Message Services. Nottingham University, 1984.

[Poo83] Pool, I. Tracking the flow of information. *Science 221* (1983), 609-613.

[Sal83] Salton, G. and McGill, M. *Introduction to Modern Information Retrieval.* McGraw-Hill, New York, 1983.

[Sal86] Salton, G. Another Look at Automatic Text-Retrieval Systems. *Commmunications of the ACM*, Vol. 29, No. 7, July 1986.

[Sea75] Searle, J. A Taxonomy of Illocutionary Acts. In K. Gunderson (Ed.), *Minnesota Studies in the Philosophy of Language.* University of Minnesota Press, Minnesota, 1975.

[Sar84] Sarin, S.K. and Greif, I. Interactive on-line conferences. *Proceedings ACM-SIGOA Conference on Office Information Systems*, June 1984.

[Sat85] Sathi, A., Fox, M.S. and Greenberg, M. "Representation of Activity Knowledge for Project Management". *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7, No. 5 (1985), 531-552.

[Shn83] Shneiderman, B. Direct manipulation: A step beyond programming languages." *IEEE Computer 16* (1983), 57-69.

[Sta83] Stallman, R. "ZMail Manual". M.I.T. Artificial Intelligence Laboratory, 1983.

[Ste85] Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S. and Suchman, L. Beyond the chalkboard: using computers to support collaboration and problem solving in meetings. *Communications of the ACM*, in press.

[Tou82] Tou, F. N., Williams, M. D., Fikes, R. E., Henderson, D. A., & Malone, T. W. RABBIT: An intelligent database assistant. *Proceedings of the National Conference of the American Association for Artificial Intelligence* (P ittsburgh, Pennsylvania, August 18-20, 1982).

[Tur83] Turoff, M. "Information, Value and the Internal Marketplace." New Jersey Institute of Technology, Newark, New Jersey, 1983.

[Tri83] Trigg, R. *Network approach to text handling for the online scientific community.* Ph.D. thesis, Department of Computer Science, University of Maryland, CSTR-1346, November 1983.

[Wil84] Wilson, P., Maude, T., Marshall, C., and Heaton, N. The Active Mailbox - your on-line Secretary. IFIP Conference on Computer Based Message Services. Nottingham University, 1984.

[Win86] Winograd, T. & Flores, F. , *Understanding Computers and Cognition: A New Foundation for Design* . Ablex Publishing Corp., Norwood, New Jersey, 1986.