Human IT
Tidskrift för studier av IT
ur ett humanvetenskapligt perspektiv

Are You Busy, Cool, or just Curious?

—CAFE: A Model with Three Different States of Mind for a User to Manage Information in Electronic Mail

by Juha Takkinen and Nahid Shahmehri

Laboratory for Intelligent Information Systems (IISLAB), Department of Computer and Information Science, Linköpings universitet, Sweden

---

Abstract

The design and implementation of a conceptual model, CAFE (a Categorization Assistant For E-mail), is described. The model supports the organization, searching, and retrieval of information in e-mail. Three modes are available for satisfying the users' needs in various situations: the Busy mode for intermittent use at times of high stress, the Cool mode for continuous use at the computer, and the Curious mode for sporadic use when exploring and (re-)organizing messages when more time is at hand. In the implementation each mode required using a different technique. The Busy mode uses the text-based Naive Bayesian algorithm, the Cool mode uses standard e-mail filtering rules, and the Curious mode uses a combination of clustering techniques known as Scatter/Gather. The design of the model is motivated partly by cognitive science theories (employed partly in a case study of categorization on the computer screen), and partly by a survey of e-mail clients. The model is related to information seeking theories in electronic environments.

Table of Contents

---

1. Introduction

Electronic mail (e-mail) is used both at home and at work, and important e-mail messages are increasingly often being mixed with less important messages in the evergrowing flow of information between users. In addition to this, people tend to collect and store information for later use, for personal business and, typically, for supporting decision-making [7][23]. Moreover, in e-mail and computer conferencing systems, such as KOM [21] and netnews (Usenet News), the storing of information is easy, while the search for—and the retrieval of—it often is more difficult. Also, it is easy to quickly disseminate information to many recipients at the same time. This asymmetry is characteristic of the electronic messaging systems being used today.

The e-mail user is rapidly finding herself in dire need of some kind of help in structuring and getting a better overview of the information contained in her e-mail messages. Furthermore, she is in need of retrieving the information in better ways. Typically, there is a lack of explicit semantic clustering of (or linkages between) relevant information. Moreover, conventional search techniques using keywords (either full text or index-based) [41] have their limits. A system with support for classifying the information would help the recipient in her task of reading and selecting relevant messages and avoiding "junk mail" or other messages of low interest. The system will have to consider both the static storage of messages and the dynamic flow of incoming messages. Finally, to make it possible for the user to satisfy her information needs, the system must allow the user to search for messages by entering queries—and examine the retrieved messages—interactively, and with a response time of only a couple of seconds.

In this paper we describe a conceptual model for the information management in e-mail. We look for inspiration in two places: cognitive science theories for categorization, and available techniques for retrieving and displaying e-mail messages, and organizing them on the computer screen. We concentrate our efforts mainly on textual information because e-mail is (still) a mostly textual medium. In section 2.1 we describe the principles for categorization, which is the basic principle behind information management. The findings from a case study of e-mail categorization and a survey of the filtering, organization, and visualization capabilities of some currently available clients are summarized in section 2.2. We then construct our conceptual model CAFE in section 3 and present a prototype implementing the model in section 4. We conclude our work and give some directions for future work in sections 5 and 6, respectively.

(Back to Table of Contents)

2. Background

2.1 Cognitive science theories for information categorization

Categorization of information is studied in both cognitive science and information retrieval and filtering (IRIF). According to psychologists there are two general and basic principles for creating categories [27]:

- The principle of cognitive economy states that the function of categories is to provide maximum information with the least cognitive effort. The maximum information with least cognitive effort is achieved if categories map the perceived world structure as closely as possible.
- The principle of perceived world structure states that the attributes, or features, that an individual will perceive in the world, and thus use for categorization of stimuli, are determined by the needs of the individuals. Moreover, these needs change over time and with the physical and social environment [27].

Since the perceived world is different for each individual, the categories are indeed personal to the individuals using them.

Similarity plays a central role in placing different items into a single category. The similarity of the items in a category varies, but to a certain degree—people want to minimize within-category variability of similarities between items while maximizing between-category variability [29]. The within-category similarity can be approximated by measuring the (subjective) “distance” between the least and the most typical item in a category [3]. Similarly, the between-category similarity can be approximated with the "distance" between the most typical items located in two different categories (see fig. 1—encircled items represent the most typical item in each category, while bold-typed items represent the least typical item in each category). However, similarity is really "in the eye of the beholder" and does not alone explain categorization, since no constraints are provided on what is to count as a feature or an attribute [34].

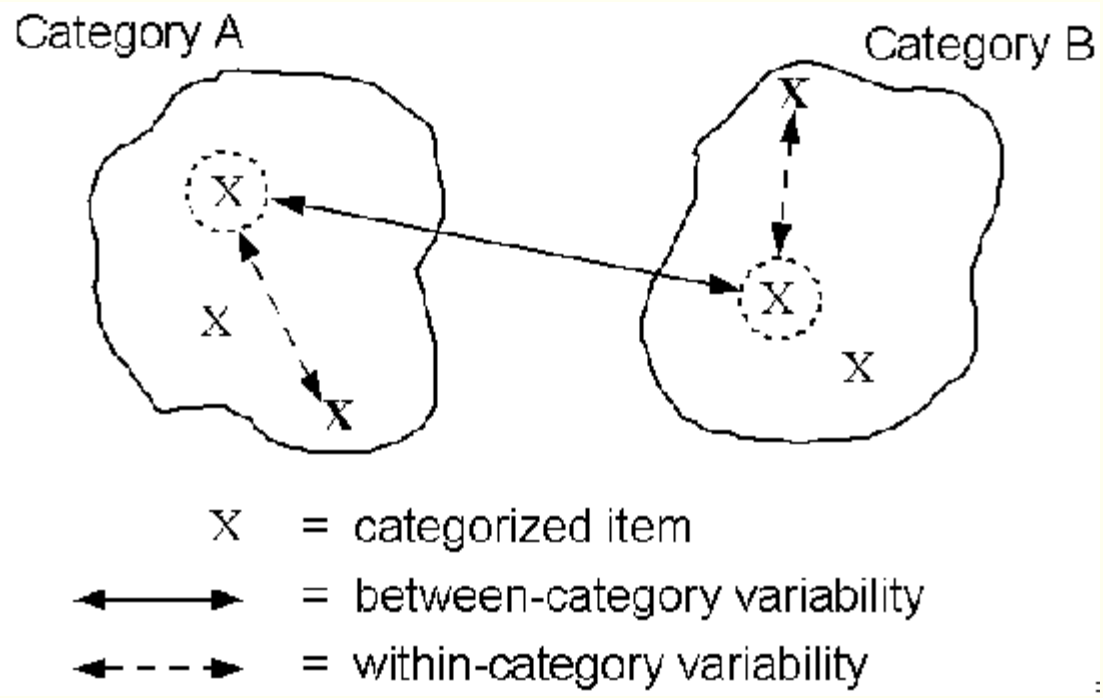

Figure 1

Categories and personal knowledge structures are of central interest to cognitive psychology researchers. The cognitive psychologists’ models of categorization and the human memory can provide useful clues for making the retrieval of information easier and more intuitive [35] (p. 178). Through the history, different theories for how categories are structured and created by humans have evolved [8]. Three examples are the classical view, the probabilistic view, and the theory-dependent view. The first two define categories solely based on the features or attributes that the items put in the categories have [27][34]. Of these, the classical view, first presented by Plato, describes categories as structured around features that define all of the items in each category. The probabilistic view, on the other hand, describes categories as either organized

around a prototype (or best example), or represented by all the individual instances that constitute it. Group membership can be graded. Two variants of the probabilistic view are available. The first variant is called the prototype view [27] and the second variant is called the exemplar view [8]. From an information-processing viewpoint [26] the classical and the probabilistic views are bottom-up: perception of similarity and attributes decide group membership. What counts as a feature or attribute is the main concern here. According to the theory-dependent view (Medin 1989, in [34]), theories determine our perception of what counts as an attribute or a feature. In this view, categories are based on knowledge and world theories (theories that humans use in categorization tasks).

In other words, people's individual theories determine the attributes or the features that will be important for a category. From an information-processing viewpoint the theory-dependent view is top-down: categories are defined by theories that determine perception—categories are put in a larger context. It may seem as though the theory-dependent would explain all the other models: for example, the classical view would be a theory that children develop. The research on categorization in cognitive science has progressed from the classical to the probabilistic view and from the idea that concepts are organized around similarity to the idea that concepts are organized around theories (Medin 1989, in [34]). Two examples of using the above mentioned theories for categorization in the IRIF area are neural networks (for example, [18]) and fuzzy sets [25]—the latter is, by the way, an attempt to use Rosch's prototype view [27] for modeling categories.

2.2 State-of-the-art of users' message management in e-mail clients

The user

Studies have shown a wide level of diversity in the way people use their e-mail clients and also a wide range of tasks for which they are used [14][31]. In a case study (The case study was a continuation and expansion of a previous preliminary study of people's categorization of text (e-mail messages and proverbs) on pieces of paper on a table [3, 24, 32]. For details about the set-up for the case study, see [33].) we examined how people create structures on the computer screen and how the structures evolve when increasingly more messages are sorted into them. Also, we examined how different representations of categories on the screen influence the development of structures and the retrieval of messages. The Subject line of the messages was extensively used for naming the categories, which is a result similar to an investigation made by the IntFilter Project at Stockholm University [12] (p. 26). The subjects were heavily reorganizing their structure for the categorization of the "junk mail" that was presented to the subjects. The type of messages influenced more the number of categories than the number of messages. Finally, there seems to be a need for a flexible way of changing the view of categories (folders), depending on the task (searching, sorting, etc.) that is to be performed. For a more detailed description of the results and discussion of the case study, see [33].

The e-mail clients

A survey of e-mail clients available for Internet-style e-mail (e-mail using SMTP and POP3/IMAP protocols) revealed a great uniformity of available functions [33]. Filtering functions for handling incoming messages are common, as are the use of folders for storing messages and two-paned or three-paned displays (fig. 3 in section 4) for presenting messages on the screen.

The most basic information management offered to the user in the e-mail client consists of the following functions: incoming messages are put (automatically by the delivery system) in an inbox and, typically, outgoing messages into an outbox, the user can read, print, compose, and send messages, and she can create folders (mailboxes) and manually file messages into the folders for permanent storage.

Typically, messages can be sorted into folders by way of a drag-and-drop interface that lets the user move around messages among folders with greater ease. The folders can be created according to an organization principle of the user's own devising and often in a hierarchy. Other functions or features commonly available in the e-mail client are the following:

- there is a folder list, a message summary, and a one-message preview window
- the filtering system looks at the text in the From and Subject lines of a message and, depending on the filter rules, moves messages into folders
- the messages can be searched for words
- the messages can be addressed through the use of aliases and addresses can be stored in an address book.

Most up-to-date clients offer a whole system of filters or rules that the user can use for automatically performing actions on (route, print, and otherwise process) incoming messages. Some clients even provide programming tools— powerful scripting languages—that can be used to build applications or trigger elaborate processes based on incoming e-mail [37][45]. Many times, however, these tools are hard to use, even at a basic level, e.g., Ishmail's patterns for rules [42]. The search functions vary from simple searching of words in message headers in one folder to advanced Boolean searching in all folders at the same time—cf. Exmh [38]. Below are two examples of e-mail clients, representing two different approaches:

- BeyondMail [37] is a commercial product. It is part of an integrated environment called groupware, which also includes bulletin boards, group schedules, and document flow, but it is also available as a standalone application with a lot of usable functions for organization of e-mail.
- Exmh [38] is a freely available and highly customizable program, with a multitude of user-definable functions for filtering, organization, and getting an overview of e-mail.

Most commonly, the vendors of the commercially available e-mail clients in our survey make the assumption that both sender and recipient of e-mail use the same product, i.e., the vendor's product. This makes it of course easier to incorporate handling of, e.g., priorities of messages (Urgent, Regular, etc.) and forms for special types of messages (meeting, phone message, etc.)—cf. BeyondMail [37]. These vendor-specific features can be of valuable use when creating a personalized structure of message categories. They can make the structure more meaningful and flexible to the individual user. Furthermore, sorting the received messages into categories according to priority coding or type of message helps making the messages more retrievable and viewable in new ways. However, few e-mail clients fully support this functionality without relying on vendor-specific features.

(Back to Table of Contents)

3. A Categorization Assistant For E-mail (CAFE)

The asymmetry in e-mail (see section 1) is both necessary and unavoidable. The sender does not want to manually classify a message, since it would mean more work. This is sometimes called the "senders' burden" approach [17]. Introducing a common, standardized classification scheme for messages would be costly (in time and money), as well as impractical. Each and every e-mail user should have to use the same kind of software for classifying and recognizing messages. Furthermore, the classification system would most certainly be difficult to maintain. Managing the software would be practically impossible, considering the wide variety of e-mail clients available [33]. Moreover, power-hungry supervisors, for example, could classify messages as being of high priority when they are not [30] (p. 75). The burden of categorization of messages should be put on the recipient's side instead. Hence, our aim is to aid the recipient in the classification,

organization, and getting an overview of her set of messages.

We want to make it possible for the recipient of messages to use different methods when looking at the information in her e-mail. In other words, using one technique to take care of all cases of message handling is not what we want to do. Furthermore, the current state of mind of the recipient is important: for example, the time available and the current information need are two components of the state information. It should be possible for the user to explicitly tell the e-mail client what her current state is. According to the principle of perceived world structure (see section 2.1), a computerized system for text categorization should be flexible in its management of the text and its representation of the user. By this we mean that text should be possible to classify in different ways, according to the needs of the user. This flexibility requires domain knowledge that changes over time. The knowledge about texts and users is usually modeled as a combination of the document representation and the (explicitly or implicitly defined) profiles of the user in the system. An example of a categorization system with these features is given by [13].

Our conceptual model for a Categorization Assistant For E-mail (CAFE) makes use of three different modes for specifying the user's state. CAFE is thus designed to support different strategies for reading, sorting, and searching messages. Both analytical and browsing strategies are supported, which, generally speaking, are central for overcoming the information problem [16] (pp. 7–8) and alleviating the user's "anomalous state of knowledge" (ASK) [1]. The conceptual model is shown in fig. 2. The modes are:

- The Busy mode is designed to be used intermittently, for locating important messages among the latest messages in the message storage. The user is typically in a situation when she has little time for reading new and unseen messages. The user is presented with a prioritized list of messages, grouped into the categories (folders) Important, 2nd Class, and Junk [6].
- The Cool mode is the default mode designed to be used continuously. It operates on the incoming message stream. The Cool mode is used in situations when the user can read messages little by little during her session at the computer. The user's own categories are used for storing the messages.
- The Curious mode is designed to be used sporadically (typically once a day), in situations when the user has time to spare. The mode is employed when the user wants to locate, organize, or reorganize previously stored messages. It supports the analysis of a larger collection of messages, typically messages from a mailing list, in all or a subset of the folders in the message storage. The user is presented with groupings of messages where she interactively can select categories to "zoom in on" and investigate further.

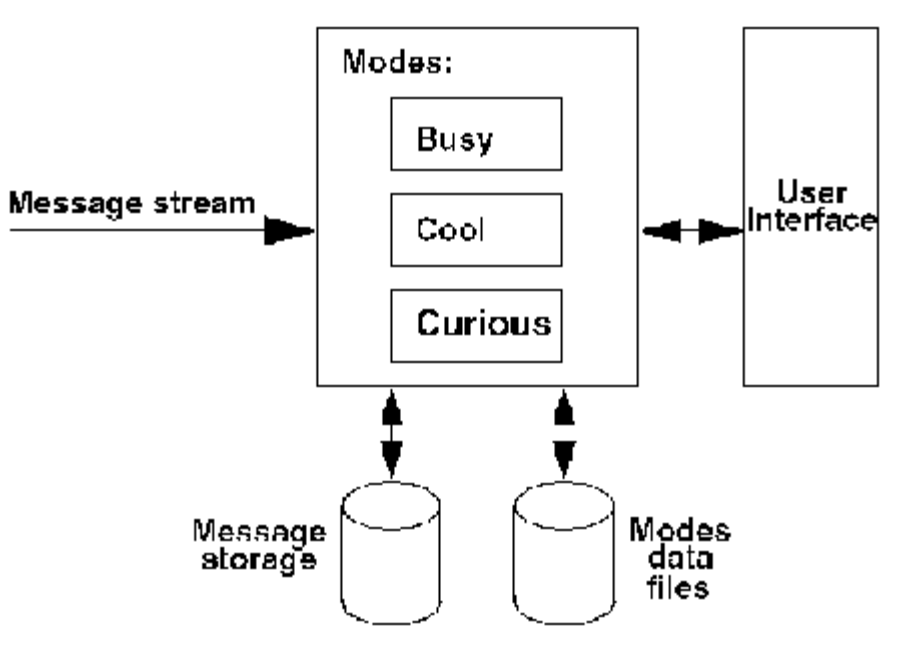

Figure 2

The user is allowed to select from the three representations (modes) according to her current personal style, experience, and information problem. This approach with using alternative representations is argued for by [16] (p. 140). The main argument is that cognitive science offers a variety of theories about how humans categorize and represent information and knowledge (see section 2.1). The need for flexibility in the representation of categories was also implied by the results of our case study of categorization in e-mail [33]. Moreover, the use and usage of e-mail in general [14] have been of great concern in the design of CAFE. A general design for a strategy to use in any system for accessing information is to use general queries and probes to identify a neighbourhood of interest, and then browse and filter [16] (p. 181). This is especially supported in the Curious mode in CAFE. The Curious mode and the other modes can be characterized by their different ways of viewing the information in e-mail. Messages already read and stored represent a collection that is static in its nature. New and unseen messages lying in the inbox or in folders form a semi-dynamic collection of messages, i.e., their state is likely to change in the near future. The incoming messages, finally, form a dynamic collection (a stream) of messages waiting to be classified and acted upon by the user or the system. In other words, we get the following characteristics of the different modes:

- in the Busy mode, we have a semi-dynamic or static deposit of messages (new and unread) on which dynamic, automatically created queries are applied
- in the Cool mode, we have a dynamic stream of messages and a set of static, user-defined queries that are applied to it
- in the Curious mode, we have a static message storage on which dynamic, interactively created queries are applied.

Our aim has been to use simple techniques and metrics, whose function and behaviour can be easily understood by the user—at least intuitively. A prototype of the conceptual model is presented in the next section.

(Back to Table of Contents)

4. The prototype of CAFE

The implementation of CAFE is based on the e-mail client called Exmh [38]. Exmh was

originally conceived with the assumption that the user would want to customize it—four ways of customization are available, depending on the desired extent [22]. Moreover, users are allowed to alter and make additions to the source code of Exmh, something which is a major bonus when developing an e-mail client. Exmh has been used as a basis for the development of different extensions by many users [5][40]. Finally, our implementation makes use of known algorithms and techniques in IRIF. Each mode in the implementation of CAFE uses a different information retrieval (IR) or text categorization technique. In this regard, the modes are described in more detail below.

The Busy mode

The Busy mode is illustrated by fig. 3, where the user is currently browsing the folder containing important messages. The contents of the menu under the Mode button are also shown in the figure. The folder display in the top pane of the window contains the folders used in the Busy mode:

- the three main folders Important, 2nd Class, and Junk, representing important messages, second class messages, and junk messages, respectively.
- the standard folders inbox, outbox, draft, and ToDo, representing incoming messages, outgoing messages, half- completed messages, and messages to be acted upon, respectively.

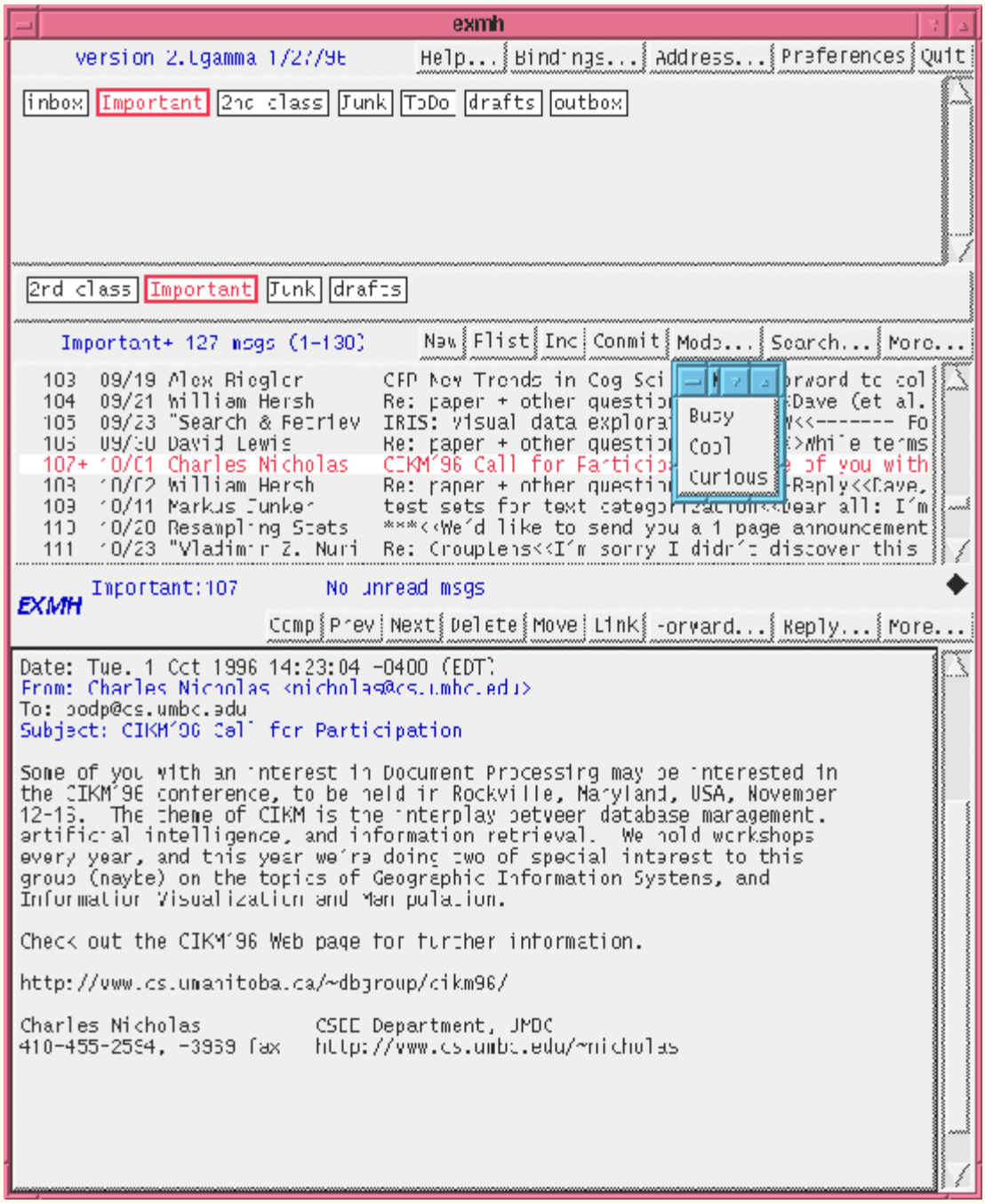

Figure 3

The text-based Naive Bayesian learning algorithm is used for filtering messages into the three main folders of the Busy mode. The algorithm uses Bayes' Theorem from probability theory. This algorithm makes the computations for training and classification simple, and it also performs rather well in practical applications of classification of text documents—see, for example [20]. It is employed via ifile, a filtering program developed by Jason Rennie at Carnegie Mellon University [40]. Messages are prioritized in ifile by giving the words on Subject and From lines higher weights in the computations.

Messages can be refiled by the user, either moving wrongly categorized messages into their right folders (folders available in the Busy mode) or saving messages for later action in the ToDo folder. The learning algorithm updates its parameters accordingly when messages are refiled.

Changing to or from the Busy mode changes the folder display. However, the standard folders (and the Junk folder) are used in all modes and remain the same. The messages in the three main

folders of the Busy mode are automatically moved to the user-defined folders when the user switches to the Cool mode, using the standard filtering rules of the Cool mode. Messages already in the Junk folder are not moved, however.

The Cool mode **In the Cool mode, the folder display shows the user-defined folders, which are used as targets for the standard user- defined rules that filter incoming messages. Messages that have not been filtered by the rules are left in the inbox and can be moved manually to their right folders by the user.**

In the current implementation, the filter rules are defined by the user in a separate filter file, one rule per line, using a text editor. Note that the categories (folders) in the Cool mode are created by the user and separate from those used in the Busy mode (see above).

The Curious mode **The Curious mode uses its own window for the display and selection of groupings of messages (fig. 4). Each grouping is shown in a scroll window of its own. A summary of each grouping is displayed in the header of each scroll window, consisting of the grouping number, the number of messages in it, and the ten most common words in the grouping. To use the Curious mode, the user typically selects a set of folders when she is in the Cool mode (the folders of the Busy mode can also be employed). The selection is done via a combination of keys that is consistent with the way Exmh is used. Thereafter, the user changes the mode to the Curious mode via the Menu button in Exmh (fig. 3), opening a separate window on the screen.**

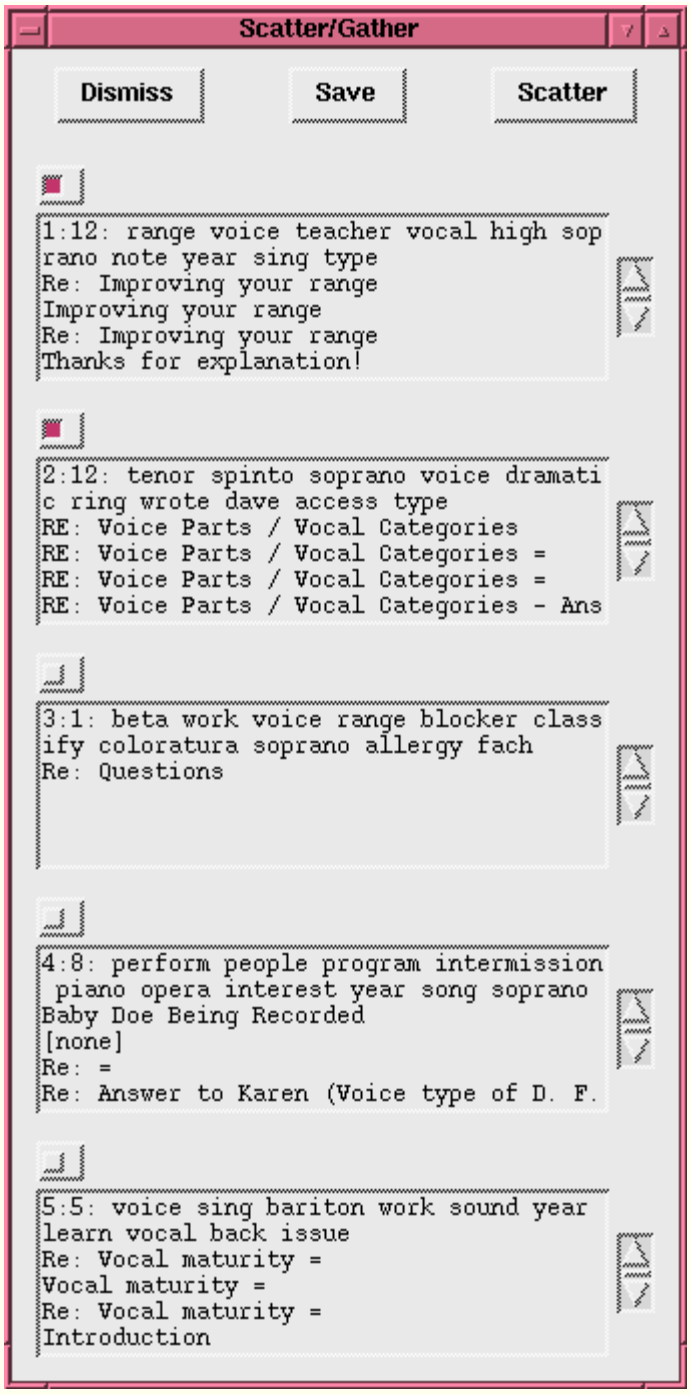

Figure 4

The messages are grouped into new categories based on groupings (clusters) that are created by a variant of the Scatter/Gather algorithm [4]. The algorithm uses a non-hierarchical partitioning strategy to cluster n documents into k groups. A strategy called Buckshot [4] is used to find initial centres for the clusters. Buckshot is non-deterministic, i.e., different (random) centres are output each time the same document set is given. The centres are used as starting points in the clustering algorithm that is employed to organize a set of documents into a given number of topic-coherent groups. We use Ward's method, a hierarchical agglomerative clustering method [10]. It uses the minimum variance measure to calculate "closeness" between points (documents). Though it is sensitive to outliers (documents far from the cluster centres), Ward's method produces compact groups of well distributed size and is deemed as appropriate for our domain. The input to the clustering algorithm are a pairwise similarity measure and the number of desired clusters. We use Dice's coefficient, since the documents are short and execution time is critical [10][28]. The

number of desired clusters can be set by the user via the Preferences window in Exmh (the default is 5). The assignment of documents to cluster centres is only done twice, since the assignment process makes its greatest gains in the first few steps [4]. The second time, new cluster centres are computed using the m most central documents in each group. We use the 70 % of the documents that are “closest” according to the minimum variance measure used in Ward’s method. Since the Scatter/Gather algorithm is interactive, Buckshot is therefore optimized for speed rather than accuracy (i.e., the rate of misclassification).

4.1 A worked example

Suppose the user has just arrived at her computer and starts her e-mail client (typically by clicking on an icon). Furthermore, suppose she is in a hurry, so she wants to see all important messages among all unseen and new messages. Thus, she changes the mode to Busy (the Cool mode is the default when the e-mail client is started) by selecting the mode from the menu under the Mode button (fig. 3). Now, the important messages are made available in a separate folder named Important (fig. 3). After doing some quick reading the user refiles a couple of messages into the ToDo folder, some other messages into the Junk folder, and another couple of messages into the 2nd Class folder. The user then exits the e-mail client, since she has skimmed through her new and unseen e-mail and is in a hurry to other places. Note that the filter rules of the Cool mode continue to work in the background and sort incoming messages into the user-defined folders available in the Cool mode.

Suppose the user comes back, now with more time on her hand. Let us say that she is interested in examining the messages from a music mailing list called VOCALIST [44] that she has stored in the folder with the same name. The messages have previously been routed to the folder by the user-defined rules in the Cool mode. The first action that she takes is to mark the VOCALIST folder—she could also have continued to select other folders by using the same marking procedure. She then changes to the Curious mode via the menu under the Mode button (fig. 3).

A separate window for the Curious mode appears, with a message asking the user to wait while the system creates groupings out of the selected folder (or folders) of messages. After a while, the result is shown (fig. 4) with five groupings of the messages from the mailing list. A summary of each grouping is displayed in the header of each scroll window, consisting of the grouping number, the number of messages in it, and the ten most common words in the grouping. Let us say that the user is especially interested in "voice types". She selects the groups with summaries containing the words "voice" and "type" (the first two groups in fig. 4) by clicking on the button in the header of the scroll windows. She then clicks the Scatter button to see new groupings of the newly selected messages. In this way, the user iteratively refines the search for interesting messages. When the user has satisfied her information needs, she has the option to save the groupings as new folders, before she quits the Curious mode by dismissing the window.

(Back to Table of Contents)

5. Discussion and conclusion

Our conceptual model, a Categorization Assistant For E-mail (CAFE), consists of three modes: the Busy mode, the Cool mode, and the Curious mode. Each mode treats the messages in different ways. Each mode is also used in a different situation, depending on the user’s "state of mind" and the amount of time that she has available. The messages can be viewed as either a continuous stream of messages or a stored collection of messages.

With CAFE, the filtering functions of the e-mail client can be personalized. That is, the sorting of messages into folders (categories) can be done in more than one way. The Cool mode gives the

user full control of simple filtering rules. Typically, the messages are sorted into categories that are topic-oriented or sender-oriented, i.e., based on the Subject or From lines of messages. More advanced rules can be derived via the machine learning algorithm in the Busy mode. The algorithm complements the filtering rules in the Cool mode. With the Scatter/Gather algorithm in the Curious mode the user can first seek broadly relevant information and then browse to reach the goal. Here, the user can make queries that she even cannot state, simply by selecting groups instead of individual queries. Apart from the explorative possibilities, a certain level of serendipity can also be achieved via the Curious mode.

As Marchionini [16] (p. 44) points out, the cost of flexible representations of information is in the various mechanisms for controlling the different representations. The mechanisms—usually paging, scrolling, and jumping— require the user to develop new strategies for manipulating the physical structure of the information, e.g., the length of a message or multiple windows on the screen [16]. In the implementation of CAFE, in this regard, we have not introduced any new mechanisms not available in the e-mail client Exmh before. For example, the folders are still represented in the same way, i.e., as collections of browsable message summaries in scroll windows.

Our experience suggests that in general, for the user to be able to formulate her information need, a successful implementation should make it possible for the user to use her experience and expertise.

Browsing is a central strategy in accessing information. In a terminology borrowed from Marchionini [16] this strategy can be supported using either probes, filters, or templates. In our prototype:

- the “probes” are represented by the different search functions, such as Scatter/Gather in the Curious mode (Furtthermore, Exmh has Glimspe [39]as a built-in search engine.)
- the “filters” are represented by the filtering rules in the Cool mode
- the “templates” are represented by the predefined folders in the Busy mode. (In addition, Exmh uses, among other things, the *components* file for creating templates [22].

The implementation of the hierarchical clustering algorithm (Ward’s method) in the Curious mode is currently too slow. Also, two documents with the same content, but written in different languages, are not treated as similar documents, since similarity in our implementation is based on keywords, which is a drawback of the simple techniques chosen. Furthermore, large clusters should be split into two clusters.

It is clear that the capability to manage heavy e-mail load is rapidly moving from a an extra feature, to something that is absolutely mandatory. Partly by examining individuals’ categorization processes and organization of messages on the computer screen, and partly by investigating the state-of-the-art of e-mail clients we were able to extract a number of interesting concepts and ideas for both an interface and a new conceptual model for handling e-mail messages.

Concludingly, the locus of control is still close to the user in CAFE, who gets a handful of new and usable possibilities of handling her e-mail. Furthermore, we alleviate some of the cognitive demand on the user in refining her "anomalous state of knowledge". Finally, the different modes ameliorate the possibilities to personalize the information management in e-mail.

(Back to Table of Contents)

6. Future work

The conceptual model can be extended in several ways: more personalized modes resembling user

profiles [19] can be added, the data in address book, calendar, and other "add-ons" associated with the e-mail client can also be included in the model. One example of an implementation that uses data available in the electronic environment of a user in the process of filtering her e-mail is described in [17]. Examples of add-ons are addressing through aliases, adding message signatures, supporting "advanced" text formatting, and spell checking. Information in other domains, such as netnews messages and personal document collections could also be managed. Fleming and Kilgour [9] have described an approach to restructuring the domain of e-mail, deriving message prototypes (templates) directly from users' formal or informal message structures. Incorporating these ideas, which relate to visual programming, can make the conceptual model even more flexible. For example, this could make searches based on message structures [15] such as "review form" and "meeting announcement" possible.

There are some aspects of the Busy mode that raise some questions and make it interesting to concentrate on in our continued research. One aspect is the ToDo folder that we introduced in the implementation of the model (see section 4). Some central questions are:

- What kind of actions or tasks does the ToDo folder imply?
- How can we support the user in the management of these tasks?
- If a message in the ToDo folder is redistributed, how can we aid the user in the redistribution of (or suggestion of recipients for) the message (and is this feasible to do or even needed)?
- Is a restructuring of the application domain (e-mail) needed? If so, how do we do it, based on the scenarios that we use and the tasks that we find?

Some support for decision-making will probably be considered here.

Extracting the so-called vendor-specific features (see section 2.2) and making them widely available to all Internet e-mail clients would make it easier to incorporate successful strategies for handling information in e-mail. Platform- independent implementations can be achieved by using, for example, Java [43] and MIME [2].

The Curious mode can be applied to the results of a search with Glimpse [39] in Exmh and thus enabling the user to view the search results in another way [11]. An important part is the definition and handling of the rules in the Cool mode (see section 4), which really should be done via a special user interface, as shown in, for example, [31].

Exmh is used by other persons in our department, which opens up the possibility to make an evaluation of CAFE in a real environment. However, the execution of the algorithms in the prototype must first be optimized before an evaluation of the prototype with real users can be done. Some of the optimizations concern the language that the algorithms are implemented in (Perl [36]). Changing the language completely can definitely make for for substantial efficiency savings. Furthermore, the initial cluster centres in the Scatter/Gather algorithm might be selected based on how dissimilar the clusters are, e.g., similarity measure less than 0.05, instead of a random selection.

Finally, we are considering making the prototype available on the Internet for Exmh users.

(Back to Table of Contents)

The Authors

Juha Takkinen has a licentiate degree in computer science and is a doctoral student in the Laboratory for Intelligent Information Systems (IISLAB) in the Department of Computer and Information Science (IDA) at Linköpings universitet, Linköping. He is also a teacher in

undergraduate courses covering IT, computer networks, and document management. He is currently writing his doctoral dissertation, which concerns the identification and delegation of tasks in electronic mail. For more information, please visit his web pages at http://www.ida.liu.se/~juhta/

Nahid Shahmehri is professor and the leader of the Laboratory for Intelligent Information Systems (IISLAB) in the Department of Computer and Information Science at Linköpings universitet, Linköping. For more information, please see her web pages at http://www.ida.liu.se/labs/iislab/people/nahsh/

References

[1] **Belkin**, N. J. (1980), "Anomalous States of Knowledge as a Basis for Information Retrieval," in Canadian Journal of Information Science, 5, pp. 133–143.
[2] **Borenstein**, N. S. & Freed, N. (1993), "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies". RFC 1521, SEPTEMBER 23, 1993. ftp://nic.nordu.net/rfc/rfc1521.txt
[3] **Cañas**, A. J., Safayeni, F. R., & Conrath, D. W. (1985), A Conceptual Model and Experiment on How People Classify and Retrieve Documents. Dept. of Management Sciences, University of Waterloo, Ontario, Canada, April 30, 9 pp.
[4] **Cutting**, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992), "Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections," in N. Belkin, P. Ingwersen, & A. M. Pejtersen (Eds.) SIGIR '92 Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press, pp. 318–329. ISBN 0-89791-523-2.
[5] **Danvind**, P. & Mattsson, M. (1996), Computational Mail. Master Thesis. Centre for Distance-spanning Technology, University of Luleå, Sweden. http://www.cdt.luth.se/%7emattias/ex-jobb/report/thesis.ps
[6] **Eberts**, R. (1993), "Postmaster: Trainable Neural Net-Based Agents for Sorting E-Mail Messages," in 2nd Industrial Engineering Research Conference Proceedings 1993, Los Angeles, pp. 534–538.
[7] **Edenius**, M. (1997), E-post – ett modernt dilemma. In Swedish. Stockholm: Nerenius & Santérus Förlag. ISBN 91-648-0124- 1.
[8] **Eysenck**, M. W. (ed.) (1990), The Blackwell dictionary of cognitive psychology. Basil Blackwell Ltd, Oxford. ISBN 0-631- 15682-8.
[9] **Fleming**, S. T. & Kilgour, A. C. (1994), "Electronic Mail: Case Study in Task-oriented Restructuring of Application Domain," in IEE Proceedings: Computers and Digital. Techniques, Vol. 141, No. 2, March 1994, pp. 65–71.
[10] **Frakes**, W. B. & Baeza-Yates, R. (1992), Information Retrieval: Data structures and Algorithms. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 504 pp. ISBN 0-13-463837-9.
[11] **Hearst**, M. A. & Pedersen, J. O. (1996), "Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results," in Frei, H-P, Harman, D., Schäuble, P., & Wilkinson, R. (Eds.) SIGIR '96 Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zürich, August 18–22, 1996. ACM Press, pp. 76–84. ISBN 0-89791-792-8. http://www.parc.xerox.com/istl/projects/ia/papers/sg-sigir96/sigir96.html
[12] **Kilander**, F., Fåhræus, E., & Palme, J. (1997), Intelligent Information Filtering. The IntFilter Project, Dept. of Computer and Systems Science, Stockholm University, Feb. 17, 1997. http://www.dsv.su.se/~fk/if_Doc/juni96/ifrpt.ps.Z
[13] **Liddy**, E. D., Paik, W., & Yu, E. S. (1994), "Text Categorization for Multiple Users Based on Semantic Features from a Machine-Readable Dictionary," in ACM Transaction on Information Systems, Vol. 12, No. 3, 1994, pp. 278–295.
[14] **Mackay**, W. (1988), "Diversity in the Use of Electronic Mail: A Preliminary Inquiry," in

ACM Transactions on Office Information Systems, Vol. 6, No. 4, October 1988, pp. 380–397.
[15] **Malone**, T. W., Grant, K. R., Lai, K-Y, Rao, R., & Rosenblitt, D. (1987), "Semistructured Messages Are Surprisingly Useful for Computer-Supported Coordination," in ACM Transactions on Office Information Systems, Vol. 5, No. 2, April 1987, pp. 115–131.
[16] **Marchionini**, G. (1995), Information Seeking in Electronic Environments. Cambridge University Press, 224 pp. ISBN 0-521- 44372-5.
[17] **Marx**, M. & Schmandt, C. (1996), "CLUES: Dynamic Personalized Message Filtering," in M. S. Ackerman (Ed.) Proceedings of the ACM 1996 Conference on Computer Supported Cooperative Work (CSCW 1996), pp. 113–121. ISBN 0-89791- 765-0.
[18] **McElligott**, M. & Sorensen, H. (1994), "An Evolutionary Connectionist Approach to Personal Information Filtering," in Irish Neural Networks Conference '94, University College Dublin, September 12–13, 1994.
[19] **Meadow**, C. T. (1992), Text Information Retrieval Systems. San Diego, California: Academic Press, Inc., 302 pp. ISBN 0-12- 487410-X.
[20] **Moulinier**, I. (1996), "A Framework for Comparing Text Categorization Approaches," in AAAI Spring Symposium on Machine Learning in Information Access, Stanford, March 25-27, 1996. http://www.parc.xerox.com/istl/projects/mlia/papers/moulinier.ps
[21] **Palme**, J. (1995), Electronic Mail. Boston: Artech House, 267 pp.
[22] **Peek**, J. (1995), MH & xmh: Email for Users & Programmers, 3rd Edition. O'Reilly & Associates, Inc., 738 pp. ISBN 1- 56592-093-7.
[23] **Rapp**, B. (1993), "Informationshantering på individ- och organisationsnivå," in Ingelstam, L. & Sturesson, L. (Eds.) Brus över landet, pp. 117–141. In Swedish. Carlssons Bokförlag. ISBN 91 7798 689 X.
[24] **Raymond**, D. R., Cañas, A. J., Tompa, F. W., & Safayeni, F. R (1989), "Measuring the Effectiveness of Personal Database Structures," in International Journal of Man-Machine Studies, No. 31, Sep. 1989, pp. 237–256. http://daisy.uwaterloo.ca/~fwtompa/.papers/ijmms.ps
[25] **Rocha**, L. M. (1994), "Cognitive Categorization Revisited: Extending Interval Fuzzy Sets as Simulation Tools for Concept Combination," in Proceedings of the 1994 International Conference of NAFIPS/IFIS/NASA. IEEE Press, pp. 400–404. http://ssie.binghamton.edu/~rocha/n94_abs.htm

[26] **Rosch**, E. (1994), "Categorization," in V. S. Ramachandran (Ed.) Encyclopedia of Human Behavior, pp. 513–523. ISBN 0- 12-226920-9.
[27] **Rosch**, E (1978), "Principles of Categorization," in Cognition and Categorization, E. Rosch, B. B. Lloyd (Eds.), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 27–48. ISBN 0-470-28377-6.
[28] **Salton**, G. & McGill, M. J. (1983), An Introduction to Modern Information Retrieval. New York: McGraw-Hill, 448 pp. ISBN 0-07-054484-0.
[29] **Smith**, E. (1990), "Categorization," in Osherson, D. N. & Smith, E. (eds.) An Invitation to Cognitive Science, Vol. 3, Thinking. MIT Press, pp. 33–53. ISBN 0-262-15037-9.
[30] **Sproull**, L. & Kiesler, S. (1991), Connections: New Ways of Working in the Networked Organization. Second edition. MIT Press: Massachusetts, 212 pp. ISBN 0-262-19306-X.
[31] **Takkinen**, J. (1994), CASUAR – en prototyp av ett användargränssnit med filtrering och automatik för hantering av elektronisk post. In Swedish. M.Sc. Thesis, Linköping University, Sweden, 126 pp. LiTH-IDA-Ex-9445. http://www.ida.liu.se/~juhta/publications/publications.html
[32] **Takkinen**, J. (1995), "An Adaptive Approach to Text Categorization and Understanding," in Conference Proceedings 1995, 5th Annual IDA Conference on Computer and Information Science, November 22, 1995, Department of Computer and Information Science, Linköping University, pp. 39–42.
[33] **Takkinen**, J. (1997), CAFE: Towards a conceptual model for information management in electronic mail. Licentiate Thesis No. 640, Dept. of Computer and Information Science. ftp://ftp.ida.liu.se/pub/publications/lic/1997/0640/
[34] **Tijsseling**, A. G. (1994), A Hybrid Framework for Categorization. Master Thesis, Dept. of Cognitive Artificial Intelligence, Faculty of Philosophy, Utrecht University. http://www.soton.ac.uk/~coglab/coglab/Thesis/

[35] **Vickery**, B. & Vickery, A. (1992), Information Science in Theory and Practice. London: Bowker-Saur. ISBN 0-408-10684-0.
[36] **Wall**, L., Christiansen, T., & Schwartz, R. L. (1996), Programming Perl. O'Reilly & Associates, Inc. ISBN 1-56592-149-6.

World-Wide Web URLs

[37] **BeyondMail**. http://www.coordinate.com/
[38] **Exmh**. http://www.smli.com/~bwelch/exmh/index.html
[39] **Glimpse**. http://glimpse.cs.arizona.edu/
[40] **The Official ifile Web Site**. http://www.cs.cmu.edu/%7ejr6b/ifile/
[41] **Intelligent Instruments for Information Management**. Project description, Dept. of Informatics, Göteborg University. http://www.adb.gu.se/~janl/III.eng.html
[42] **Ishmail User's Guide**. HAL Computer Systems, Inc., Campbell, California, USA, October 1995. http://www.ishmail.com/
[43] **Java Computing**. http://www.sun.com/java/
[44] **VOCALIST**. http://lists.oulu.fi/vocalist/
[45] **Z-Mail**. http://www.netmanage.com/

Åter till Human IT 2/1998