HOW TO CREATE A HUMANE INFORMATION FLOW

Daniel Pargman, Jussi Karlgren, Ann Lantz, Olle Palmgren, Kristina Höök

Depts of Psychology and Computer and Systems Sciences, S-106 91 Stockholm University, Sweden

intfilter@sics.se

ABSTRACT

This paper describes some of the starting points of a project on intelligent filters for group communication systems at Stockholm University. Group communication systems cause an imbalance between sender and recipient of electronic information. The aim of our project is to give recipients more control than they typically have today by giving users better and more varied information on the material they receive, and tools to utilize it. We have tested some of our hypotheses by building a prototype system to work with Usenet News.

KEYWORDS

Information filtering, information overload, Conference systems, Electronic mail, Usenet News, NetNews

INTRODUCTION -GROUP COMMUNICATION SYSTEMS

Group communication systems such as electronic mail and conference systems are spreading increasingly fast as they are adopted by corporations and government agencies. They are effective means for fast coordination, dissemination and retrieval of information. As with all new media that replace or supersede older media they have specific advantages, but with their increased usage also some unforeseen *dis*advantages [1]. An obvious disadvantages is that of information overload.

INFORMATION OVERLOAD

Information overload in group communication systems stems from the radically altered relationship between the costs of the sender of information and those of the recipient in comparison with other media. Electronic group communication systems shift virtually all power and initiative from the recipient to the sender. Once the sender has created the information, the cost (in both money and effort) of sending it to many rather than to few recipients is very low. If this information is sent to people who have not requested it, or have no need for it, it is they who have to bear the costs of disposing of it.

If a conference system is used and controlled centrally by an organization, information overload can (1) be viewed as a social problem that can be addressed with social conventions [1], [2], *or* (2) it can be addressed by reorganizing the use of the conference system. If on the other hand the system has no central control, as is the case with the Usenet News system, filter systems are an alternative. By filter systems we mean software tools that help the user to structure and control an information flow by providing help in the entire spectrum from decision support to autonomous decision making. The aim is to give the recipient more control: to redress the imbalance between sender and recipient caused by characteristics of group communication systems.

FILTERING AND THE NEED FOR FEEDBACK

With ill-defined problems where no objective general optimal solution exists, one can but hope to find a solution that is good enough in regards to the effort given to reach it and its closeness to a perceived optimal solution. Filtering an information stream is a good example of such a problem. It combines the very real problem of information overload with the vague solution that is captured in the assertion that ”it is impossible to seek valuable information without encountering valueless information” [1, p.170].

Filtering is cognitively demanding for users and if it is to be done with a decision support system, and especially an autonomous decision making system, users need to be persuaded that the system is reliable and predictable. Above all, any system must have users feel that it gives them control and does not hide vital information from them.

Confidence can only be won through feedback. In filtering software, designed to trim an information flow, feedback mechanisms are the key to (1) build up confidence and delegate responsibility to the system and to (2) design a flexible system that can cope with the changing needs of users. In addition to providing the tools to customize an information flow, measures and mechanisms to help monitor the performance of the system has to be provided to help users evaluate the results of their actions. Changing or even ephemeral user interests exacerbate the problem. It is therefore essential that filter systems should be built with a high degree of support for change.

In systems that are built with ambitions to match human cognitive capabilities, many desirable characteristics can be summarized as perspicuity and control. This we intend to mean transparency and editability of system knowledge, as well as transparency, undoability, and predictability of system reasoning mechanisms.

THE INTFILTER PROJECT

The aim of the IntFilter project is to build ”intelligent” filters that will work with multiple knowledge sources. We conducted a pilot study of experienced Usenet News users [3]. We found that message selection currently was based only on key-words in the header or body of the messages and that sharing information about their own usage to aid other users is a sensitive question to users. Most would, but wanted to retain control of this information.

A prototype system, the GHOSTS system [3] has been built to filter messages in Usenet News. The GHOSTS system has basic functionality and should be seen as a framework for further extensions in the project. It is built with object-oriented methods to provide maximum modularity, reusability, portability and opportunity for extension. The filtering engine is for example independent of the rules it works with or the types of messages it works on. That is why it easily has been extended to also process ordinary electronic mail.

The GHOSTS system is at present a quite simple rule-based filtering system. It has rules that trigger on regular expressions in the header or body of a message [4]. The filter evaluates and classifies messages and performs actions according to the results. In Usenet News it stratifies messages into three categories in terms of interestingness and performs auxiliary actions (save, copy, print, forward etc.) as defined by the rules. The rules are constructed in a frame-based graphical environment. The rule-base can easily be inspected by the user and some feedback on the effectiveness of the rules can also be presented.

The GHOSTS system also incorporates a newsgroup browser. The browser is a specialized tool for Usenet News which graphically displays the newsgroup hierarchy and allows for certain actions on the newsgroups. The simple but surprisingly useful graphical interface makes the several thousand meetings of the Usenet News hierarchy manageable and encourages exploration. We are evaluating the GHOSTS system with experienced Usenet News users.

The next stage - multiple knowledge sources

The basic functionality of a filter for Usenet News is twofold: firstly the user will want to subscribe to interesting newsgroups, and secondly the user will want to read only interesting submissions in these newsgroups. In both cases several knowledge sources can be tapped for selecting interesting reading.

Newsgroups partition nicely into categories following several criteria: by name, by number and activity of participants, by internal dialogue structure, by the charter and the distribution of the newsgroup, by the nature of the newsgroup hierarchy, and by the characteristics of the messages.

There is also a large array of knowledge sources and criteria for message categorization. Messages can be categorized not only by keywords for probable content, but, when tagged by efficient surface syntactic processors used in the project [5], also by genre, and, tentatively, by quality. Information available to the system is taken from the header fields and the text itself, and includes simple textual data: word occurrence in text or header fields, average word length, average sentence length, part-of-speech information, and other easily obtainable data, as well as data typical of newsgroup messages rather than texts in general: dialog or comment structure that the message belongs in, length of the message, proportion quoted material, length of signature field. Which parameters to use is based partly on statistical evaluation as well as the aforementioned group discussion studies of experienced users. What the text genres of USENET News conferences is a research question, and we have initiated a study for this area. Lastly, we intend to experiment using data extracted from other users' profiles in the form of automatically emerging *recommendations* for newsgroups or messages [6].

Intelligence

The project will extend to investigating the utility of using learning mechanisms to observe user behaviour and suggest new rules and also to learn to weight the multiple knowledge sources available. However, the first step in the project is simply to give users more and better information. Any intelligent learning mechanisms must result in user editable and verifiable knowledge of the same sort that the users themselves have placed in the model, and any intelligent behaviour must be phased in incrementally, subject to continuos verification and editing by the user [4], [7].

REFERENCES

1 Lee Sproull and Sara Kiesler. 1991. *Connections - new ways to work in the networked organization*, MIT press.

2 Starr Roxanne Hiltz and Murray Turoff. 1985. ”Structuring computer-mediated communication systems to avoid information overload”, *CACM*, 28:7, pp. 680-689.

3 The IntFilter project. 1994. Working papers, Department of Computer and Systems Sciences, Stockholm University.

4 Thomas W. Malone, Kenneth R. Grant, Franklin A. Turbak, Stephen A. Brobst and Michael D. Cohen. 1987. ”Intelligent information-sharing systems”, *CACM* 30:5, pp. 390-402.

5 Douglass Cutting and Jussi Karlgren, 1994. “Identifying Text Types Through Simple Metrics”, submitted to SIGIR-94, Dublin.

6 Jussi Karlgren, “An Algebra for Recommendations” SYSLAB working paper.

7 Robyn Kozierok and Pattie Maes. 1993. “A Learning Interface Agent for Scheduling Meetings”, *ACM SIGCHI International Workshop on Intelligent User Interfaces,* ACM .