

SICS Technical Report  
T94/09

ISRN : SICS-T—94/09-SE  
ISSN : 1100-3154

## **The glass box user model for filtering**

by

Jussi Karlgren, Kristina Höök, Ann Lantz,  
Jacob Palme and Daniel Pargman

July 1994

Swedish Institute of Computer Science  
Box 1263, S-164 29 KISTA, SWEDEN

---

T94:09

ISRN: SICS-T--94/05--SE

ISSN: 1100-3154

## **The glass box user model for filtering**

Jussi Karlgren, Kristina Höök, Ann Lantz, Jacob Palme, Daniel Pargman

Departments of Computer and Systems Sciences, Computational Linguistics, and  
Psychology, Stockholm University

Swedish Institute of Computer Science Box 1263, S-164 28 Kista, Stockholm

Sweden

`intfilter@sics.se`

July 1994

### **Abstract**

The first requirement on an interactive system in a domain such as information filtering is to be an interface to knowledge, rather than just a knowledgeable interface. We borrow the computation instruction metaphor of a system as “a black box in a glass box” as a means to conceptualize the problem of giving a user control over the actions of an interactive system. The application domain we work in is that of information filtering. In the “black box”, we hide complex knowledge of the domain objects such as facts and assumptions about text genre identification, while the “glass box”, which is what the user sees, only shows the neat top level knowledge of the domain conceptual categories such as e.g. categorization rules.

**Keywords:** Information filtering, user modelling, interface design, Usenet News



# The glass box user model for filtering

Jussi Karlgren, Kristina Höök, Ann Lantz, Jacob Palme, Daniel Pargman

Departments of Computer and Systems Sciences, Computational Linguistics, and  
Psychology, Stockholm University

Swedish Institute of Computer Science Box 1263, S-164 28 Kista, Stockholm

Sweden

`intfilter@sics.se`

## 1. Introduction

Filtering software is recognised as one way of helping a user to manage large information flows - a problem that is becoming increasingly relevant in the emerging information society. The design of such filters involves two problems: one is finding good criteria and methods for filtering, the other is designing conceptually acceptable interaction methods for the user of the filter. Both of these problems are addressed by the IntFilter project at Stockholm University and some of the findings are presented in this report.

The first problem is harder than might be expected at first. When studying experienced users using the Usenet News conference system, we found that very few clues are in fact given to the user on which to base decisions about whether a message or a news group is interesting or not. Experienced users basically filter on the title of the message, the news group it appears in, and in some cases by author. This is not because these criteria are sufficient: users would filter on other factors were they easily accessible [Pargman, 1994].

The second problem, designing the interaction, became obvious to us when we started designing a prototype system, GHOSTS, in which users can write personal sets of rules to be used as their personal filter, [Pargman, 1994].

We decided early on that it is important that the user have full *control* over the filter. This is why simple rules seemed appealing, as they are easy to display, conceptually reasonably easy to grasp, and easily editable by the user. The problem we encountered was that the data the rules need to use are very complex. Since users in general are disinclined to spend any substantial amount of time learning how to use interactive systems, the rules cannot be too complex. This is why we propose and discuss a "glass box" user model, where certain parts are hidden from the user in "the black box".

## 2. Filtering

New electronic media such as electronic mail and conference systems spread fast as they are adopted by corporations, government agencies, and individuals. They are effective and inexpensive means for fast co-ordination, dissemination, and retrieval of information.

With increased usage, the new media have also shown some disadvantages. The most obvious of the disadvantages is that of information overload or information overflow: simply, the receipt of more information than is desired. An explanation to why electronic media are especially susceptible to information overflow stems from the skewed relationship between the costs of the sender of information, whom the new media benefit to a great extent, and those of the recipient, whose power and initiative have been reduced. Once the sender has created information, the cost in both money and effort of sending it to many rather than to few recipients is very low. When 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. On average, the time to write an electronic message is four minutes, and the time to read it half a minute: ten spurious recipients thus expend more time with the message than did the sender. And adding a dozen or so spurious recipients is not unusual in electronic messaging systems.

From the user's point of view, it is very likely that the problem is framed as a problem of insufficient quality. A user will notice that although Usenet News is an interesting medium with occasional nuggets of interesting information, most of the material is irrelevant, uninteresting, and of low quality. This, it turned out in study II [Lantz, 1993], was actually more of a problem in the conference system domain than was information overload. With a lowered threshold of publication, the initial filter which exerts a certain quality control at the source by demanding senders to invest greatly in the dissemination of their message is rendered moot.

If a conference system is used and controlled centrally by an organisation or used by a small group of people, problems such as these can better be addressed through social conventions [Hiltz and Turoff, 1985; Sproull and Kiesler, 1991], or by reorganising and formalising the use of the conference system and by demanding a certain competence or level of experience of its members. On the other hand, if the system has no central control and a large number of participants, as is the case with Usenet News, filter systems are a viable alternative, where filter systems are to be understood as any software tool that helps the user to structure, control, and select information from an information flow in a manageable way.

This paper describes the work of the IntFilter project.<sup>1</sup> Our work is directed towards conferencing systems, not at mail systems (as have others, e.g. [Malone et al 1987]). Although similar in many ways, conferencing systems involve some additional aspects: the information flow is likely to be less relevant on average than a mail flow, and there will be the additional problem of selecting conference groups (newsgroups in Usenet News). Large conferencing systems also are used as information bases by users, so that

---

<sup>1</sup> The IntFilter project is an on-going project at the departments of Computer and Systems Sciences, Computational Linguistics, and Psychology of Stockholm University, in cooperation with SICS (Swedish Institute of Computer Science). It is funded by NUTEK, the Swedish Board for Technical Development.

the filtering task also involves more information retrieval than does the typical mail organisation task.

The design of filtering software is not trivial. As outlined above, there are two ends to the problem, or two (not unrelated) research problems: one is finding good criteria and methods for filtering and the other is designing a habitable interaction between the filter and the user. This paper will address the latter as its main point, but we will first, in some detail, give an overview of our principles for the former.

### 3. Studies of Filtering in Usenet News

Finding good criteria and methods for filtering is non-trivial. We have performed a series of studies with experienced users of Usenet News in order to determine where automatic filtering would aid the user, and where users can and should be in control of filtering. Below we report on three studies of experienced users of Usenet News carried out within the IntFilter project.

Before elaborating further on these studies, we give a short introduction to Usenet News in order to explain where the problem lies.

#### 3.1 Usenet News

Usenet News can be accessed with any of a large number of interface systems (“newsreaders”) and most or all of them provide the user with extremely few clues on which to base decisions about whether a message or newsgroup is interesting or not.

Naturally, the first step of filtering in this type of hierarchically ordered system is to decide what newsgroups to subscribe to. This information - a user model of sorts - is saved in a file (the .newsrsrc file) which the newsreader accesses at every sessions.

Usenet News involves several million users<sup>2</sup>, carries many thousands of different newsgroups and grows very fast (10-20 % per month). Each newsgroup can have tens or even hundreds of thousands of subscribers<sup>3</sup>. The average traffic in Usenet News is 30-35 000 messages or 55-75 megabytes of data every day<sup>4</sup>. During the course of a week, several dozen new newsgroups can be created of which most belong to the alt, or “alternative” hierarchy, where 46%<sup>5</sup> of all Usenet News traffic by volume is transmitted.

Newsgroup: comp.ai.nat-lang		Articles: 123 of 4258/11
a.Manuela Boros	14	robust parsing
b.miles@minster	116	>robust parsing
c.Carole Hafner	24	>>robust parsing
d.Carole Hafner	24	>>robust parsing
e,Shona Douglas	45	Position in NLP at Edinburgh
f,Graeme Hirst	92	Research positions on Computational Linguistics
g,rshain@mitre.org	12	Autmatic Generated Abstracts
h,J.J. Paijmans	15	>Autmatic Generated Abstracts

<sup>2</sup> 3 million users according to Hauben (1993). 4,2 million users in November 1993 as estimated by the DEC Network Systems Laboratory after sample measures.

<sup>3</sup> November 1993 statistics as measured by the DEC Network Systems Laboratory, to be found in the newsgroup news.lists.

<sup>4</sup> November 1993 statistics by UUNET communications (the lower figure) and the DEC Network Systems Laboratory (the higher figure), from the newsgroup news.lists.

<sup>5</sup> January 1994 statistics by UUNET communications, from the newsgroup news.lists.

```

i,Francesc Ribas 20 Request for spanish corpora
j,Ben B. Nystuen 13 Request for CD-ROM Corpora
k,Tony Davis 31 information retrieval systems
l,Ted Dunning 18 >request for ASCII-Dictionarys
m,Raj Dash 22 >>request for ASCII-Dictionarys
n,Gabriel.Climesc 282 CFP: ConTI'94, Timisoara, Romania
o,Ted Pedersen 24 unification vs. constraint based grammar
p,Alex Franz 16 >unification vs. constraint based grammar
q,Henry Thompson 29 >unification vs. constraint based grammar
r,rshain@mitre.org 12 Representing Mandarin in ASCII
s,kfc@gmv.ifas 16 >>Representing Mandarin in ASCII
t,M Covington 8 >Representing Mandarin in ASCII
u,Mark Van Haren 16 >>>Representing Mandarin in ASCII
v,Shawn O'Donnell 15 >Representing Mandarin in ASCII
w,Brian Wilkins 33 -
x,kfc@gmv.ifas 23 >Representing Mandarin in ASCII
y,Carlos McEvelly 44 >Representing Mandarin in ASCII
z,Stefan Svenberg 19 multi-lingual generation
0,Tim Dugan 4 wanted:public NLP software written in C/C++

-- 18:44 -- SELECT -- help:? -----Top 28%-----<level 2>--

```

Figure 1: A typical research-oriented newsgroup, and a typical newsreader interface.

### 3.2 Study I: How Users Do

The first study, I, was an explorative interview study with the aim to investigate the users strategies for reading contributions in electronical conference systems. The six users were academics working in different areas (bio-medicine, computer science, computer technique and psychology) [Pargman, 1994].

We found that users used Usenet News for various reasons, and consequently, used various methods to select interesting information. The tree main purposes for using Usenet News are: Goal-oriented usage to search for specific information and to solve immediate work-related problems, work-related usage to keep informed about a certain area of interest and to keep up with colleges in a field of research or development and leisure usage and recreation. It is primarily work-related usage that is of interest to us. The criteria is that Usenet News should be a useful tool, i.e. more useful than other information channels. Leisure usage can be more sensitive to examine, more random and dynamic in its nature and does not feel as pressing for us to examine. In goal-oriented usage, Usenet News is used as a traditional database in search of solutions for immediate problems, with the extra option of posing a question to “search for information in the brains of the meeting participants”.

For most users, Usenet News was their first experience with a large constant flow of information. This novelty of the media in itself render old methods that have been used successfully with other media (radio, television, newspapers etc.) of little value. Different strategies have to be made up to cope with the new situation. In a study by Wendy Mackay, [Mackay, 1988] e-mail users had different strategies to cope with their situation and she divided them into "prioritisers" and "archivers".

It seems many users do not learn to use News from anybody, but have learned to use News by themselves. A number of different news readers are used, but most users had not tried or heard very much about news readers other than the one they use, and the one they use is often the first one they tried. It sometimes seemed to be a pure coincidence that a particular news reader was used.

The differences in strategies between novice and expert Usenet news users are large. What the novice first of all need is an introduction to the concept of conferencing systems and knowledge of Usenet News' structure. The expert on the other hand knows where to

find interesting information and needs help to filter and order messages in these newsgroups to discern between interesting and uninteresting messages. Inbetween these groups are the experience novices who begin to understand Usenet News and asks himself where information about a certain subject is located.

The study was too small to draw any conclusions about different strategies to read News as the behavior varied much between the users and we felt it was premature to try discern any trends from such a small study.

### **3.3 Study II: Experienced users**

Study two, II, was a pilot study with the aim to investigate how experienced users of Usenet News select their information in the system and how they react towards assisting the system. Four subjects were interviewed and they also demonstrated how they selected their information in the system, both in meetings they were familiar with and in meetings they never entered before, but that still were relevant to the subject's work. During this demonstration the subjects were instructed to think-aloud [Lantz, 1993].

We found that experienced users mostly filter on information in the header: mainly keywords they see in the title, and partly through browsing through the text, where keywords and general layout are used for selection. In an unfamiliar news group, the selection process takes longer, apparently since the title keywords are unfamiliar. In this pilot interview, none of the subjects complained about overflow *per se*, but instead stressed the low quality of contributions.

The information given to the recipient is impoverished as compared to the non-textual surface information implicit in the appearance of messages arriving through the paper mail. Envelope form, quality, size, thickness and other such factors that are used for initial filtering are completely lacking in the electronic mailbox and the incoming flow can be much larger. One aim of the project is to identify information, including, but not limited to, what is found in the header fields, that can be used to filter on.

We also asked subjects what their attitudes would be to assist the system by providing feedback about their reading habits or preferences for public use. This seemed to be an acceptable way to work, but the subjects stressed the need to retain control over what information the system would have access to.

In addition, our conclusion was that, from the viewpoint of the user, we cannot separate the filtering functionality from the problem of providing the user with an *overview* and *information seeking tools* for the entire system. We need not only a filter, but an interface which provides structure to the entire information flow. Below, we will outline the design of GHOSTS, our prototype system.

### **3.4 Study III: What We Get From Users**

We aim to evaluate texts along several parameters, or after membership in several categories: what they are for a specific domain is an empirical question. Every domain has its own patterns of variation, and we are investigating Usenet News texts for factors of this sort. We need to have users categorise and evaluate texts in order to provide us with information about relevance on typical variation of texts along parameters.

In study three, III, the goal was to find criteria for filtering information. Here we asked experienced users of Usenet News to rate and comment a set of contributions from meetings they were familiar with. Three subjects were interviewed and were instructed to think aloud during the session of rating and commenting the contributions read [Lantz, 1994].



As a generalised problem, filtering is a question of categorisation, or evaluation, depending on the slant of one's method. In the case of Usenet News, the problem is categorisation. The crudest description is that texts are categorised into categories "interesting" and "uninteresting", or sorted along a scale from "interesting" to "uninteresting".

The results from our study III show that the three subjects in their ratings what to read now, what to read later and what to not read, did use several criteria that could be useful for filtering. These criteria can be divided into two groups. The first group includes criteria that can be used in pattern recognition, i.e. what the contribution looks like, for example: the length, if it includes natural language or programming language. This pattern recognition is automated and does not require too much cognitive effort from the subject.

The second group is non-automatic and demands that the user reads the text in order to be able to decide how to handle the specific contribution. The non-automatic group is requiring cognitive work load. The user has to read the text or parts of the text to find out if the text is: interesting or not; geographically relevant; correct compared to the subject line of the contribution, etc.

Since pattern recognition does not contribute to the cognitive load as much as the scanning and checking of message text, we conclude that there are two roads we should travel: one is to strengthen the pattern of messages to allow users to improve their own filtering skills and the other is to help the user to analyse the text and create meaningful measures and categories of the text upon which automatic or semi-automatic filtering can be based.

## **4. Methods for filtering**

### **4.1 Topical filtering**

The first step in any message filtering task is to try to recognize the subject matter of the message at hand. For this there are standard information retrieval metrics, see for instance the textbook by [Salton and McGill, 1983].

In addition, in conference systems, there is a broad categorization by topic in that messages usually occur in conferences or newsgroups, and in discussion threads within them. This is compounded by the information found in the "Subject:" and possibly, the "Keywords:" fields. This is what our prototype system makes use of now; *topical filtering*.

But this is not enough especially keeping in mind what really contributes to the users cognitive load, namely the scanning and reading of the text in the message. There are numerous parameters of variation that a user could make use of, if they only were displayed in a comprehensible manner.

### **4.2 Dialogue structure**

The *dialogue structure* wherein messages are placed is useful for categorizing messages. A message which is a comment on a comment will probably not be as pertinent to the topic alluded to in the subject line as the first original message was. Also, a selection of messages from within a conversation, where messages refer to each other, will not be as useful as either the entire conversation or none of it. Similarly, a message which is *cross-posted* to a number of groups usually has low relevance for any one of the addressed groups.

### 4.3 Genre identification

Due to new advances in efficient surface syntactic processing we are able to “tag” large amounts of text very rapidly, marking up words according to grammatical category and identifying syntactic constituents, [Koskenniemi, 1983; Karlsson, 1990; Tapanainen and Voutilainen, 1993]. We will be able to use these data for several purposes: for regular information retrieval with keyword and phrase identification, as well as for *genre identification*. [Biber, 1990; Karlgren and Cutting, 1994]. What the relevant genres are in the medium of Usenet News conferences is a research question, and we are currently pursuing a text study with newsgroup material, using parameters such as proportion of nouns, average sentence length, average heaviness of noun phrases and other well and less well established textual data.

### 4.4 Other methods

Newsgroup messages, in addition to typical general textual data such as the ones mentioned above, have *medium specific textual characteristics*. Length of the message, proportion of quoted material, length of signature field, and other characteristics may all be useful for categorization. Other users’ explicit or implicit *recommendations* may be useful as well, after the earlier and more robust data have been used to make the larger distinctions. Data extracted from other users’ profiles is being analyzed for informational relevance, [Karlgren, 1991; Goldberg et al, 1992].

Not only messages, but the newsgroups themselves can be partitioned in several typical *newsgroup types*, by data on the flow and type of messages transmitted.

Whatever data we choose to use, we will have a problem trying to explain to the user what the system does. All the data sources are on a low level of granularity – whereas the problem of the user is not. This is a question of interacting with the knowledge in the filter.

## 5. Interaction with a Filter

Coming back to the second problem outlined in the introduction, designing the interaction well, is common to all projects that have worked in the domain of intelligent filtering [Malone et al, 1987; Maes and Kozierok, 1993]. Naturally, the interaction with a filter is subject to all the usual caveats of interface design: situation-, user-, task-, and domain-dependence, and in this case, the underlying domain and task is of special interest to us.

The task is to give the recipient control. Today, the domain objects, the messages, do not give enough information for this to be easy: all the user gets is the information in the header fields. As outlined in the previous section, we have a grab bag of tentatively useful methods to extract information from the texts and the discourse itself, most of which on their own will not be particularly useful, but in combination with each other may yield reliable results. The problem is that the level of abstraction is too low to be presented to the user.

Our prototype system, GHOSTS, is rule based. GHOSTS is built on top of a freely available and widely spread newsreader: **nn**. GHOSTS is designed to interact with the user only through its interfaces to the system knowledge: any actions GHOSTS takes are performed through the **nn** interface. GHOSTS does not require that the messages or the text material be more structured than the newsreader itself requires. In GHOSTS, the user writes rules that customize the interaction. At present, the rules can address data on standard restricted text retrieval metrics like word occurrence in specified fields or in the text. In order to improve the filtering precision, and to be able to extend it to tentative filtering

on quality as well as genre and style, we realize we need a better interface to the data that the system has access to. This has led us to borrow the metaphor of *the black box in a glass box*.

### 5.1 User Models - A Black Box in a Glass Box

The rule set and the subscription list (the .newsrsrc file) together form a user model. We stress the fact that the model must be editable and comprehensible, to give the user explicit power and integrity over the interaction. The knowledge must be separable from the system itself, and the users must feel that the user model is entirely under their control. The drawback is, of course, that a user-model based system will demand some sophistication of the user. This is all right, as long as the user's conceptual model as regards the domain can stay simple. The problem is that the rules are to use complex data.

The solution for us has been to design GHOSTS so that the complexities of text analysis can be hidden from the user. We call this, borrowing a metaphor which has been used in discussing the teaching of computation, *a black box in a glass box*. [Du Boulay, O'Shea, and Monk, 1981]. By the black box we mean the interface backdrop, behind which we hide unpalatably complex and confusing facts and assumptions about text genre identification, for instance, from the user.

The glass box is where we present the categorization rules, and let the user make use of them. With an intractable and dynamic domain but with a the user community that has a set of relatively stable conceptual categories, we can thus make use of the categories without binding ourselves for specific algorithms. The algorithms in themselves can then be improved along the way. An example of what can be hidden in the black box are metrics such as what the number of agentless passives is in a message (see Figure 2). That metric, weighed together with other similar ones – number of past-participial

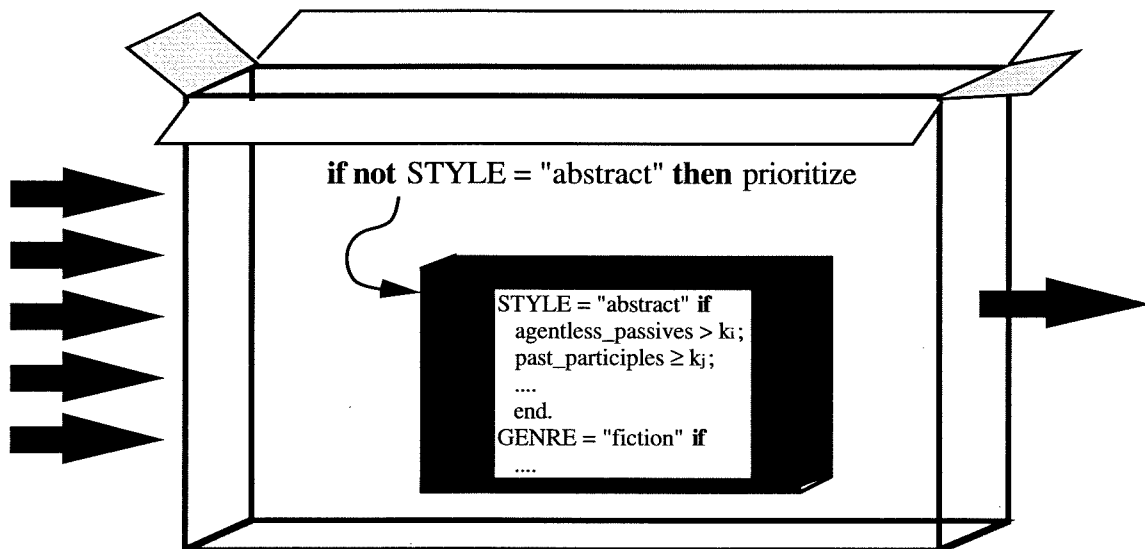


Figure 2: A glass (or openable cardboard) box, with a forbidding black box inside.

clauses, for instance – can in the glass box be represented as a judgment of whether the text is of an abstract or non-abstract style, [Biber, 1990].

## **5.2 Organisational structure and domain dependency**

It remains yet to be seen which metrics will be most useful on the glass box level. Our assumption is that this will be closely related to how the conference system is used. These measures can be more specific and domain-dependent the more closely coupled its use is to an organizational structure and its purpose. If, on the other hand, the system is used for leisure, more general measures imported from other similar situations can be used.

## **5.3 Intelligence and Adaptivity in the Black Box**

Some of the same motivations as for the black box can be advanced (and indeed, have been advanced!) for adaptive interfaces. When the domain is too complex in detail, but understandable in terms of conceptual categories, automatic learning algorithms of different types seem to be useful. The project will extend to investigating the utility of using machine learning mechanisms to learn to weight the multiple knowledge sources available for message analysis automatically, and to learn new categories automatically. However, the first step in the project is simply empowering users. 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 verification and later editing by the user, much as in the system Maes and Kozierek have presented, [Maes and Kozierek, 1993].

We know from other attempts to use adaptive techniques, e.g. in the design of knowledge based systems, that control and visibility are a real problem when a system starts to act on its own [Berry and Broadbent, 86; Wærn et al. 93]. In the worst case, an intelligent interface that was intended to aid users in their tasks, may instead turn into an unpredictable, uncontrollable and obtrusive front-end to the system that the user is trying to use. It is necessary to give users some kind of view of the internal reasoning of an intelligent system. It does not necessarily have to be the complete picture, but again, rather a glass box that the black box resides in.

## **6. Interface to Knowledge**

Finally, we should point out that even if we have emphasised the design and use of a filter, we know that the ideal filtering system is more than just a filter. As noted above, we realized from interviews with users that it is hard for many users to distinguish conceptually between the processes of filtering information, presenting information and information retrieval. To the users, the ideal filtering system is a tool that also helps them to navigate and search in information space. This suggests that the ideal filter should be an integrated part of a system that also supports these functions.

The complexity of coping with Usenet News has been addressed by the GHOSTS system by explicit and editable knowledge representations. For example, to manage newsgroup subscription, and to structure the entire information flow in general, GHOSTS includes a newsgroup browser (see Figure 3). The only complexity the user needs to watch out for here is the unmanageable size of the hierarchy, and a simple graphical tool suffices to handle it.

The browser displays the conference hierarchy and allows for certain actions on the conferences. The simple but surprisingly useful graphical interface makes the several thousand meetings of the Usenet News hierarchy manageable and encourages exploration.

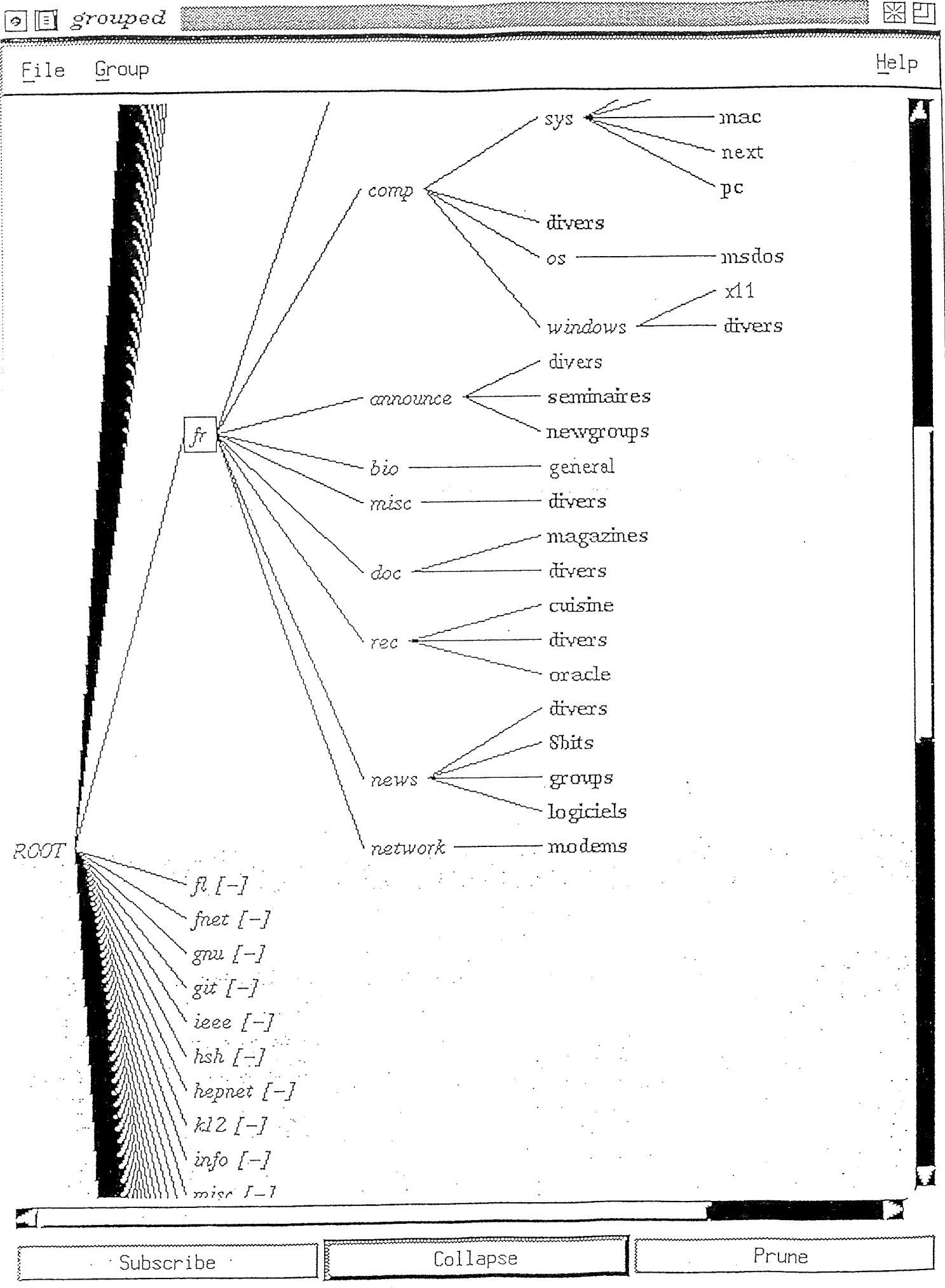


Figure 3: The newsgroup browser creates a manageable interaction with the newsgroup hierarchy.

## 7. Summary

Through studying experienced users of Usenet News, we have found that pattern recognition used by users to do filtering today, does not necessarily provide to a high cognitive load. Instead, the scanning, reading and analysing of the actual text in the message, is what contributes most to the cognitive load. Therefore, our work has concentrated on finding methods for automatically analysing text in order to provide new and good measures upon which filtering can happen.

We have presented a model for filtering based on simple rules, where the data that the rules use are partly hidden by a “black box” in a “glass box”. These rules may in the future be learned automatically. Hiding the data is especially necessary in view of the fact that we intend to augment the filter functionality by genre, style, and other elusive parameters of variation, most of which only can be monitored through a large number of relatively sophisticated parameters.

We have emphasised the need to allow the user to inspect and alter the filter to stay in control of its actions.

Finally, we have pointed out that the actual filter must only be one part of a system solution where information is not only filtered but also structured and presented in such a manner that the user feels in control of the information flow.

## 8. References

- Douglas Biber. 1989. “A Typology of English Texts”. *Linguistics* 27:3-43.
- Berry, D. C. and Broadbent, D. E. 1986. “Expert systems and the Man Machine Interface: part two: The user interface”. *Expert Systems: The international journal of Knowledge Engineering*, 4.
- Benedict du Boulay, Tim O’Shea, and John Monk. 1981. “The Black Box Inside the Glass Box: Presenting Computing Concepts to Novices”, *International Journal of Man-Machine Studies*, 14:237-249.
- David Goldberg, David Nichols, Brian M. Oki and Douglas Terry. 1992. “Using collaborative filtering to weave an information tapestry”. *Communications of the ACM* 35:12:61-70.
- Hauben, R. 1993. “The evolution of Usenet News: the poor man’s Arpanet” Speech presented at MACUL (Michigan Association of Computer Users in Learning) in March 1993, transcript posted in the newsgroup comp.society on March 28, 1993.
- Starr Roxanne Hiltz and Murray Turoff. 1985. “Structuring computer-mediated communication systems to avoid information overload”. *Communications of the ACM*, 28:7:680-689.
- Jussi Karlgren. 1991. “An Algebra for Recommendations” SYSLAB working paper. Stockholm: Department of Computer and Systems Sciences, University of Stockholm.
- Jussi Karlgren and Douglass Cutting. 1994. “Identifying Text Types Through Simple Metrics”, submission to COLING-94, Kyoto.
- Fred Karlsson. 1990. “Constraint Grammar for Parsing Running Text”. In *Proceedings of COLING -90*, Karlgren (ed.), Helsinki.
- Kimmo Koskenniemi. 1983. *Two-level Morphology: a General Computational Model for Word-Form Recognition and Production*. Publication no. 11, Dept. General Linguistics, University of Helsinki.
- Robyn Kozierok and Pattie Maes. 1993. “A Learning Interface Agent for Scheduling Meetings”. *ACM SIGCHI International Workshop on Intelligent User Interfaces*. Orlando: ACM.

- Ann Lantz. 1994. "How do experienced users of Usenet News select their information?". *IntFilter project document*. available from DSV, Stockholm University, Sweden.
- Ann Lantz. (in preparation). 1994. "Finding criteria for filtering by studying experienced users". *IntFilter project document*. available from DSV, Stockholm University, Sweden.
- Robert M Losee, Jr. 1989. "Minimizing information overload: the ranking of electronic messages". *J. of Information Science* **15**:179-189.
- Mackay, W.E. 1988. "Diversity in the use of electronic mail: a preliminary inquiry". *ACM transactions on office information systems*, **6**:4:380-397.
- Thomas W. Malone, Kenneth R. Grant, Franklin A. Turbak, Stephen A. Brobst and Michael D. Cohen. 1987. "Intelligent information-sharing systems". *Communications of the ACM* **30**:5:390-402.
- Jacob Palme. 1981. "Experiences with the use of the COM Computerised Conferencing System", Technical report available via gopher and ftp from dsv.su.se at Dept. Comp & Sys. Sci, Stockholm University, Stockholm.
- Daniel Pargman. (in preparation). 1994. *Creating a Humane Information Flow*, Master's thesis at Uppsala University, Uppsala.
- Lee Sproull and Sara Kiesler. 1991. *Connections - new ways to work in the networked organisation*. MIT press.
- Atro Voutilainen and Pasi Tapanainen. 1993. "Ambiguity Resolution in a Reductionistic Parser", *6th Conference of the EACL*, Utrecht, Holland.
- Annika Wærn, Peter Holm, Kristina Höök, and Rune Gustavsson. 1993. "The commonKADS Communication Model". KADS-II Report, M3/TR/SICS/006/2.0, Swedish Institute of Computer Science.