

# Using “Human-in-the-loop” in an Adaptive System: An Evaluation Study of the ConCall System

Master Thesis in Informatics

Charlotte Averman

[averman@sics.se, s94lotta@hobbe.informatik.gu.se]

SICS/HUMLE

Swedish Institute of Computer Science,  
The Human Computer Interaction and  
Language Engineering Laboratory

School of Economics and Commercial Law,  
Department of Informatics,  
Göteborg University

## Abstract

The rapid development of new information technology has brought forward the possibilities of efficient collaborative work. People can seek in vaster information spaces, and are the target of a tidal wave of information surging by every day through email, newsgroups, news sites, etc. The access to information is escalating and we feel overwhelmed. We call it information overload. One way of solving this problem would be to reduce the information that passes through each individual's sphere. The aim would be to create a system that could filter through the incoming information in an intelligent way so that we reduce the flow but at the same time get through *all* the information relevant and interesting to the user of such a system.

This master thesis presents an evaluative study of the ConCall system, where we take a look at how to use the best of both human and machine to solve this problem. ConCall is an adaptive system, implementing the EdInfo ideas which is to combine human expertise with machine intelligence in order to achieve a high quality of filtered information to its end users. ConCall was built up to be a call for paper and participation filtering system targeting researchers as users. The study of ConCall was an experimental evaluation aiming both to look at the functionality and utilities in ConCall and show that this concept works. The study was also one of the last steps in a bootstrapping circle with the intentions to be a steppingstone to the start of the next circle of development.

The study showed that a filtering system like this could be both useful and was desired as a help to sort through the continuous stream of incoming calls for papers and participation as an alternative to what most of the participants used today: unstructured streams of calls coming in through e-mail. The study also showed that recommendations are preferred having colleagues and friends as senders. Another interesting result concerned a dependency between the motivation in users and the filtering performance in means of precision and recall. Lessons learned from this study had to do both with setting up experimental situations and the difficulties of developing adaptive systems.

## Table of Contents

<b>ABSTRACT</b> .....	<b>1</b>
<b>TABLE OF CONTENTS</b> .....	<b>2</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>3</b>
<b>I. INTRODUCTION</b> .....	<b>4</b>
USING THE BEST OF BOTH HUMAN AND MACHINE .....	4
INTRODUCTION TO THE CONCALL SERVICE.....	5
DISPOSITION.....	5
PROBLEM DEFINITION .....	5
<b>II. BACKGROUND</b> .....	<b>6</b>
A TIDAL WAVE OF INFORMATION .....	6
INFORMATION FILTERING SYSTEMS .....	6
TWO INFORMATION FILTERING PARADIGMS.....	7
<i>The Content-Based Filtering Paradigm</i> .....	7
<i>The Social/Collaborative Filtering Paradigm</i> .....	7
THE APPROACH IN CONCALL.....	8
ADAPTIVE FILTERING.....	9
MEASURING THE EFFECTIVENESS OF AN INFORMATION FILTERING SYSTEM .....	9
BOOTSTRAPPING ADAPTIVITY .....	11
<i>Why Must Adaptivity be Bootstrapped?</i> .....	11
<i>How to Bootstrap Adaptivity</i> .....	11
<i>Bootstrapping the ConCall Service</i> .....	12
THE CONCALL SYSTEM.....	12
<i>Functionality, architecture and implementation</i> .....	12
<i>Technical Details</i> .....	13
<i>The Buzzwords and the Buzzwordlist</i> .....	14
<i>The User Profile</i> .....	14
<i>The Editor Role</i> .....	15
<b>III. STUDY METHOD</b> .....	<b>15</b>
<i>Study Aims</i> .....	15
<i>Subjects</i> .....	16
<i>Study Structure</i> .....	16
<i>The editor role; editing of calls and adding buzzwords</i> .....	16
<i>The Logs</i> .....	17
<i>Measures</i> .....	17
<b>IV. RESULTS</b> .....	<b>17</b>
COMMENTS AND EXPRESSED NEEDS FOR HANDLING CFPS .....	17
THREE MAJOR THEMES AND RECOMMENDATIONS .....	19
<i>Theme1 - Reminders:</i> .....	19
<i>Theme2 – Friends and Colleagues:</i> .....	19
<i>Theme3 – Filtering Performance:</i> .....	19
<i>About Recommendations</i> .....	21
<i>Perception about the information provider/broker</i> .....	22
OUTSIDE THE SCOPE OF THE STUDY – “SIDE RESULTS” .....	23
<i>Profile Handling and the Candidate Profile</i> .....	23
<i>Feedback on the GUI</i> .....	23
<i>The Wish for a Graphical Overview</i> .....	23
<b>V. DISCUSSION AND SUGGESTIONS FOR REDESIGN</b> .....	<b>24</b>
<i>Design Issues</i> .....	24
<i>User Profiles</i> .....	24

<i>The Experimental Situation</i> .....	25
<i>Trust and Privacy</i> .....	25
<i>Editor Aspects and Support</i> .....	25
<b>REFERENCES</b> .....	<b>26</b>
<b>APPENDIX I</b> .....	<b>27</b>
FRÅGEFORMULÄR (I): .....	27
FRÅGEFORMULÄR (II): .....	29
<b>APPENDIX II</b> .....	<b>32</b>
RESULTS FROM THE TEST STUDY.....	32
<i>Questionnaire I</i> .....	32
<i>Questionnaire II</i> .....	35

## **Acknowledgements**

This master thesis has been done at the HUMLE laboratory group at SICS (Swedish Institute of Computer Science) within the EdInfo project. I would like to thank Annika Waern, head of the HUMLE laboratory and my supervisor, for her percipience and way of directing me towards the “important” subjects and objects. I would also like to thank Åsa Rudström, Kia Höök, Mark Tierney and Jarmo Laaksohanti (HUMLE/SICS) for insightful comments and fruitful discussions. My thanks also go to Johan Redström at the Viktoria Institute, for his philosophical directions in writing scientific reports.

## I. Introduction

### ***Using the best of both human and machine***

The field of AI has a history reaching back to the birth of the “computer age”, starting in the 1960’s. Attempts have been made over and over again, in different forms and shapes and with different methods and perspectives, to try and create machine intelligence. The AI research community has tried everything from creating a virtual human brain to minimalistic and task specialised soft- and hardware. These efforts have been made in order to try and create new “workers” that could replace the human so that we as humans could be freed up to spend our time and efforts on tasks we either are better at or prefer to do. To enable people to delegate tasks to autonomous or semi-autonomous software entities is today a more and more appreciated feature in intelligent software and agent technology. Another aim has been to speed up certain processes and task accomplishments where the machine capability to compute and calculate supersede the human capability. A third aim that has come to mean more and more to us humans in later years, is the power of the machine to search for, find and filter the information we desire. This has become a more eminent the last few years.

Information is getting more widespread and is increasing exponentially in our environment. A tidal wave of information is flowing over us each day, through TV, radio, email, newsgroups, letters, conversations, etc. We have quickly reached the threshold of the human cognitive capability to deal with all this information. If we cannot deal with it, we cannot use it. We get exhausted and will not get the value of the information that is so precious to us today. The resent research which deal with this problem, *information overload*, has put great effort into and tried their best to make it more pleasant and intuitively for us to reach all this information. Though a greater concentration has been directed towards the question of how we could diminish the information wave and at the same time get the really interesting and relevant information through to the right person.

Though we come far in the AI area, several experiments and studies has shown that the combination of both human and machine intelligence used in solving this question has proven to be more fruitful than just relying on machine intelligence. The machine is used for its speed, calculation capability and efficiency to sort through large sets of data. While the human is used to evaluate and annotate information objects<sup>1</sup> and to define preferences. The human competence in biased evaluation and selection is prized highly in this context. A human can often account for more and more subtle variables that will play a role in the choice of information, than a machine. A machine might be able to theoretically do the same, but both the rules it bases its choice on and the computation time of the same, quickly become unrealistic.

---

<sup>1</sup> *Object* is used in this master thesis to represent information in all kinds of forms and shapes, from audio and video to pure ASCII text documents. *Object* is used in this interchangeably with the term item(s) and includes the information form *document(s)*.

### ***Introduction to the ConCall Service***

The ConCall (Waern et al., 1998) system is a call-for-paper/participation (CFP) filtering service, built on an agent-based architecture. The system is an actual implementation of the EdInfo concept, where the idea is to combine human expertise (an editor or information broker) with machine intelligence in order to achieve a high quality of the filtered information provided to the end user (Höök, Rudström and Waern, 1997). The editor’s role is to survey and conduct the information retrieval and to shape the rules for annotating CFPs. There is also a high user involvement, where the users of the system are a vital part for the shaping and creation of conformity between users and information brokers. (A more extensive presentation of ConCall is given under *Background/The ConCall System.*)

### ***Disposition***

In chapter I the problem and research question is described along with a basic introduction to the ConCall system. In chapter II the background to this master thesis is presented. Chapter III describes the study-outline and study aims along with a presentation of the methods used. Selected facts and findings from the study are presented in chapter IV including a discussion thereof. In chapter V a discussion is given alongside with some suggestions for redesign in future work and studies. This text ends with two appendices, Appendix I and II, where the questionnaires used and the raw data of from the study can be found.

### ***Problem definition***

This study came about with the intention to initialise an evolutionary and iterative development of a user adaptive system, namely ConCall. The aim has been in this first step to evaluate the service as a whole and to see whether the offered functionality is or is not sufficient. That is, sufficient in providing the users with both an intuitive way of reaching their information seeking goals and to support the collaboration and adaptation in ConCall.

There are also intentions to find possible extensions or modifications to future versions and re-implementations.

## II. Background

### *A Tidal Wave of Information*

With the new information technology a rapid development have brought forward the possibilities of efficient collaborative work. It has enabled people to communicate with people that they before would not have interacted with at all. People are both enabled to seek in vaster information spaces and the target of a tidal wave of information surging by every day through email, newsgroups, news sites, etc. Not to mention all those little clients we have on a desktop informing us of everything from a woman in Idaho bearing eight children to the latest report on the trial against Bill Clinton. The access to information is escalating at an exponential rate (info\_overload and JIT, med. rep. ?? andra refs?). We get overwhelmed. We call it information overload, which we see as a problem. One way of solving this problem would be to reduce the information that passes through each individual's sphere. A risk with this though is that we reduce this so that vital and interesting information does not come through, which would not be a desired effect. One of the solutions to this could be intelligent information filtering. To create a system that could filter through the incoming information in an intelligent way so that we reduce the flow but at the same time get through *all* the information relevant and interesting to the user of such a system. If this is not achieve or if the user of the system perceive the system of being incapable of providing sufficient information, the user will have to look to other sources or providers of information which would both be more time consuming and tiresome.

It would be wonderful if a system as described in the section above were just to be created, but more than just a little coding is needed for it to work. First of all would such a system build on some sort of personalised preferences matching the users needs, and to achieve that the users would need to directly or indirectly through her/his actions tell the filtering systems of his or her preferences. All well with that, but humans are not always as good as computers to explicitly state or express their needs nor do they always know exactly what they want in before hand. To ease this process it is therefor important to have an interface between the user and the system that help the user formulate his or her needs. Both in a way that is intuitive to the user and explicit enough for the system to be able to use that information in its filtering work.

### *Information Filtering Systems*

Douglas W. Oard (1997) shortly and concisely stated that: "The goal of an information filtering system is to enhance the user's ability to identify useful information" (Oard, 1997). He also argues for the fact that in combining machine and human abilities the user satisfaction could be raised. By using the best of both an interactive combination could achieve better results than a system purely based on a machine automated filtering or a humans manual search and filtering process. We could use the speed of the computers, give it some rule based adaptive instructions and try to create intelligence, but so far just such human-machine combinations have given the best results.

### **Information Filtering in relation to Information Retrieval**

Information filtering is related to Information retrieval though differs on a few points. (Belkin and Croft, 1992)

**One**, when dealing with information retrieval the information source is seen to be a rather static collection of for the most part documents, while in the case of information filtering the source is seen as a constant stream of information distributed by someone.

**Two**, when talking about users’ interests and goals in an information retrieval system the need or formulation thereof is more immediate, whilst in information filtering systems the user needs (or needs for a group of users) is more constant over time, and aim at more long-term goals and tasks.

**Three**, in an information retrieval system, the need is expressed through direct queries but in an information filtering system the needs are represented and expressed through a profile.

**Four**, when the comparison happens in an information retrieval system, there is also usually an interaction phase where the user accepts or decline recommended information or their representations. In an information filtering system, at the state of comparison, there is instead the automatic filtering process.

## ***Two Information Filtering Paradigms***

### **The Content-Based Filtering Paradigm**

In content-based filtering, each user is assumed to operate independently (Oard, 1997). There are no additional sources, like in the case of social/collaborative filtering, e.g. document annotations, other users preferences. Due to this there are only the content of the document available to create document representations from. Recommendations in systems with a content-based approach are based on what users have liked or disliked in past events and the task of rating each retrieved document is essential for future performance. Thus a pure content-based system relies on its performance on its users efforts of rating. With little imagination one can easily visualize this to grow to an enormous task for the users.

### **The Social/Collaborative Filtering Paradigm**

When using collaborative (social) filtering systems, the criteria for recommending a document of a specific information representation is based first of all on available personal profiles, i.e. other users’ preferences. These profiles are either manually set up and maintained by the user or automatically by the system. These profiles are then open for adaptation in different ways (See *Adaptive Filtering*).

The documents that are recommended initially through use of the available profiles are then subject to be annotated by the user. The annotations can be keywords from the text, the domain of the text, associated terms, judgements etc. An additional way to go about social/collaborative filtering and the collaborative effect it builds on, is to let the users add and delete keywords or terms from their profiles (showing likes and dislikes), where the effects of the changes made are disseminated to the users of the system.

The difference from content-based filtering is that instead of matching the contents of items to past preferred items, users’ are matched after similarities in their preferences. (Balabanović, Shoham, 1997) In systems using the collaborative approach a try is made to identify other users with similar preferences and recommend what these users have preferred. This approach demand less effort from the users, since it is not as dependant on a fairly large quantity of

ratings from one user to perform adequately. There are problems with this approach as well though. For instance, it implies that all members in the user group need to have some critical and basic *similarity* in order to get any recommendations at all. A user with interests deviating from that of his/her user group will get a low performance out of a system based on this approach.

In a social filtering system, several studies has shown that in order to get a good system, i.e. a system that has a high accuracy in its recommendations, there is a critical mass of users needed (Oard, 1997, p. 156).

Oard (1997) states another interesting factor for social filtering in his paper: the limitation put on the social filtering by user motivation. Since a lot of the social filtering systems implemented so far, includes the momentum of annotating documents (or other objects being filtered), there is a need for a fairly high motivation in the users of such a system. If there is no motivation to annotate, give feedback or any recommendations, there will subsequently not be any grounds to base the filtering on.

### **Recommender Systems**

In our everyday social interaction we seek and gather information that can help us in making decisions about everything from which car to buy to which video to rent in the video store. We seek recommendations. The recommendations we get we evaluate and grade based on our perception and degree of trust in the recommendation provider. A recommender system is supposed to ease and support this process. Most existing recommender systems use social (collaborative) filtering methods that base recommendations on other users' preferences.

A recommender system typically takes recommendations from its users or other contributors, as input. These recommendations are then gathered and distributed out to information seeking users, either through matching the users' need or representation thereof (e.g. through profiles), or through matching the user with a certain recommendation provider (as in the case when a user is seeking the opinion of a certain expert). A recommender system can be based on virtual communities<sup>2</sup> of users and the disseminated effect of those users' feedback, annotations and other ways of recommending items<sup>3</sup>.

### **Collaborative Filtering**

One often hears the term “collaborative filtering” used along side recommender systems. Though this term is seen as to be more specific, partly because a recommender system need not be based on direct collaboration between the recommendation provider and the receiving user. And also partly because the term “recommender systems” need not mean that any filtering is done as the term “collaborative filtering” more explicitly indicate.

### ***The approach in ConCall***

In ConCall today there is no supported interaction between users, nor do individual users have any support from or access to information about other users or their preferences or profiles. However, through conveying their preferences and feedback to the editor(s) (through adding their own buzzwords), convergence may arise and form an implicit information-flow back to

---

<sup>2</sup> Hill, W., Stead, L., Rosentstein, M. and Furnas, G., *Recommending and Evaluating choices in a Virtual Community of Use*. Available at: <http://www.apparent-wind.com/navigation.videos.html>.

<sup>3</sup> *Items* in this case could mean information in the form of documents, pictures, references, audio, video or other information representations.



the users. In the case of ConCall, document representations are manually derived from the document contents (the original CFP), through “human intervention”, thus are no demands made on ConCall to deduce document representations.

In the ConCall system users define and set-up their profiles initially from a given list of terms. The maintenance is the users own responsibility though the system give suggestions to alterations. The system is thus trying to give adapted feedback to such things as the user’s behavior, the change of content in the database, changes in the group of users’ preferences, convergence between editor and users, etc.

The users are given the opportunity to give feedback annotations through adding their own terms to their profiles. The profiles are subject to review of the editor(s) and the idea is that the editor(s) then will be able to get suggestions to new annotations, domains to cover and see how the terms are used and perceived by the users.

### ***Adaptive Filtering***

Adaptive filtering techniques are based on user profiles, where the profiles are put together from experiences of user behavior and evaluations of previous recommended objects. These profiles are then used as a basis for selecting and recommending newly received objects.

Observations about user behavior could be based on time spend reading a document, if the user saved the document for future references, if the user deleted the document, etc. Then these observations can be used as a base from which rules for filtering can be inferred. These rules could then be presented to the user to be approved or modified. This is an iterative process, where the system continuously readapts the user’s profile according to the observations, approved rules and modifications. Hence this filtering technique is called adaptive.

ConCall is adaptive in the way that the system give suggestions to alterations so that editors for instance adapt their way of annotating and users change their profiles. This is done through giving the editors access to the user profiles and the users are presented with a candidate profile indicating what the system “think” the user could be interested in. These suggestions could contain both annotations the users previously avoided to select or annotations new in the database.

### ***Measuring the Effectiveness of an Information Filtering System***

In order to present reproducible results from an information-filtering task, it is a necessity to postulate a few things about such a task (Oard, 1997, p. 154). Such assumptions could be that a user’s judgement of the relevance of a presented document is constant over time or that we are limited in our span of grades when judging a recommended document on its relevancy. As Oard (1997) argues, due to the fact that human judgement *do* vary significantly both over time and depending on who does the evaluation, the above fails to satisfy the fundamental concept of relevance on which it rests.

Nevertheless, a measure of effectiveness is needed to evaluate a filtering system. Within the information retrieval field, there are three such measurements commonly used when looking at the effectiveness of an information retrieval task. These are precision, recall and fallout.

Measuring precision, recall, and fallout as described by Rijsbergen (1979):

	Relevant	Non-Relevant	
Retrieved	$A \cap B$	$\bar{A} \cap B$	$B$
Not Retrieved	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	$\bar{B}$
	$A$	$\bar{A}$	$N$

(Where N is the number of items.)

$$\textbf{Precision} = \frac{|A \cap B|}{|B|}$$

$$\textbf{Recall} = \frac{|A \cap B|}{|A|}$$

$$\textbf{Fallout} = \frac{|A \cap B|}{|\bar{A}|}$$

**Precision** – the part of detected documents, that actually where relevant to the user.

**Recall** – the part of all the documents that were relevant to the user *and* where correctly classified as such by the system.

**Fallout** – the part of non-relevant documents, classified by the system as relevant.

The precision and recall measure the detection effectiveness, whilst fallout measure the rejection effectiveness. When measuring the detection effectiveness, no regard is made to the size of the document collection. While precision is less expensive to evaluate (only part of the collection need to be scored), both recall and fallout easily get out of hand when the collection grows bigger, since every document would have to be calculated. When dealing with a large document collection, recall and fallout are usually calculated only on a chosen sample of the collection.

Precision and recall are presented in values ranging from 0 to 1. The closer to 1 the values of precision and recall are the more optimal is the system that was measured.

Precision, recall, and fallout are used in areas that does not explicitly fall within the area of information retrieval but rather under different information filtering approaches. One example of that is (Robertson, 1981, p. 60).

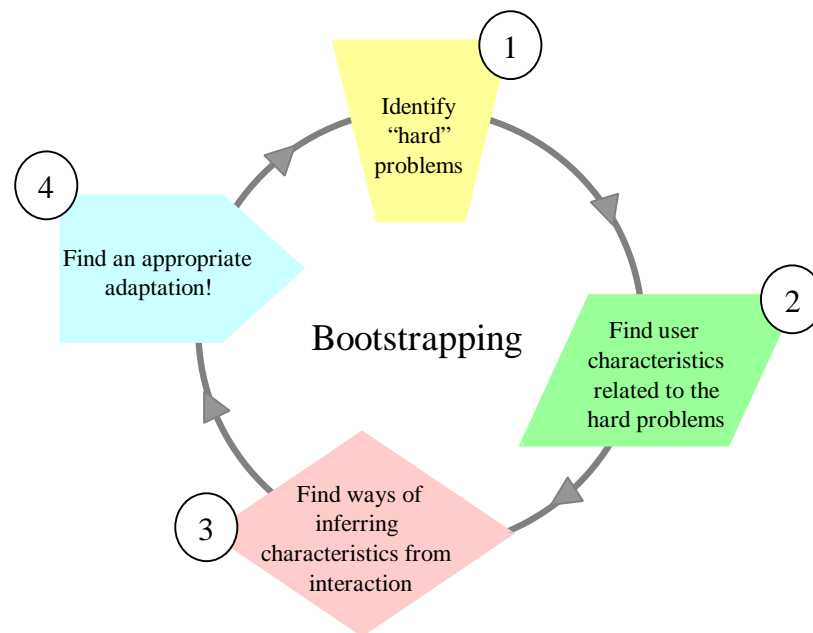
## Bootstrapping Adaptivity

Traditionally system development follow the three-phases: Analysis, Design and Evaluation. When developing Adaptive User Interfaces another strategy is taken. The development of the adaptivity is bootstrapped through an iterative process.

### Why Must Adaptivity be Bootstrapped?

There are essentially three motivations for bootstrapping Adaptive Systems. First of all, dealing with adaptive systems the user modelling rules need to be understood from what the users actually do with the system. Secondly, it follows that the user behaviour these rules are based on, will change once the systems starts to adapt. In addition, it is necessary to evaluate the adaptation design in itself.

In developing an adaptive system, the analysis of users’ tasks and needs is a necessary part of the process. The development of ConCall has taken the form depicted in figure 1 (as proposed by Höök, 1998). Where the test done and presented herein this thesis could be placed between step 4 and 1. Previous development steps are not discussed in this text.



**Figure 1** The development life-cycle used in the evolution of ConCall.

There is always a risk that (pre)-defined rules for adaptation ceases to be relevant once the system starts to adapt to user behaviour, in the cases of systems using implicit methods of inferring user characteristics.

### How to Bootstrap Adaptivity

One option would be to bootstrap adaptivity at design time in parallel with evaluation of the adaptivity design. Methods such as controlled studies or longer trials with ‘real’ users can be used. Bootstrapping can also be done when the system is installed and in use, using either: 1) fully automated methods, as in the case of recommender systems, or 2) semi-automatic

methods where the adaptivity is tweaked using logs and user feedback. (The ConCall system mostly use this approach)

### Bootstrapping the ConCall Service

The ConCall study was an evaluative study combined with gathering data for bootstrapping the adaptivity of the system. In the study data was in terms of logs and user profiles, that would later be used to tune the adaptation algorithms for user profiles, as well as provide future editors with a suitable collection of keywords to start annotating CFPs with.

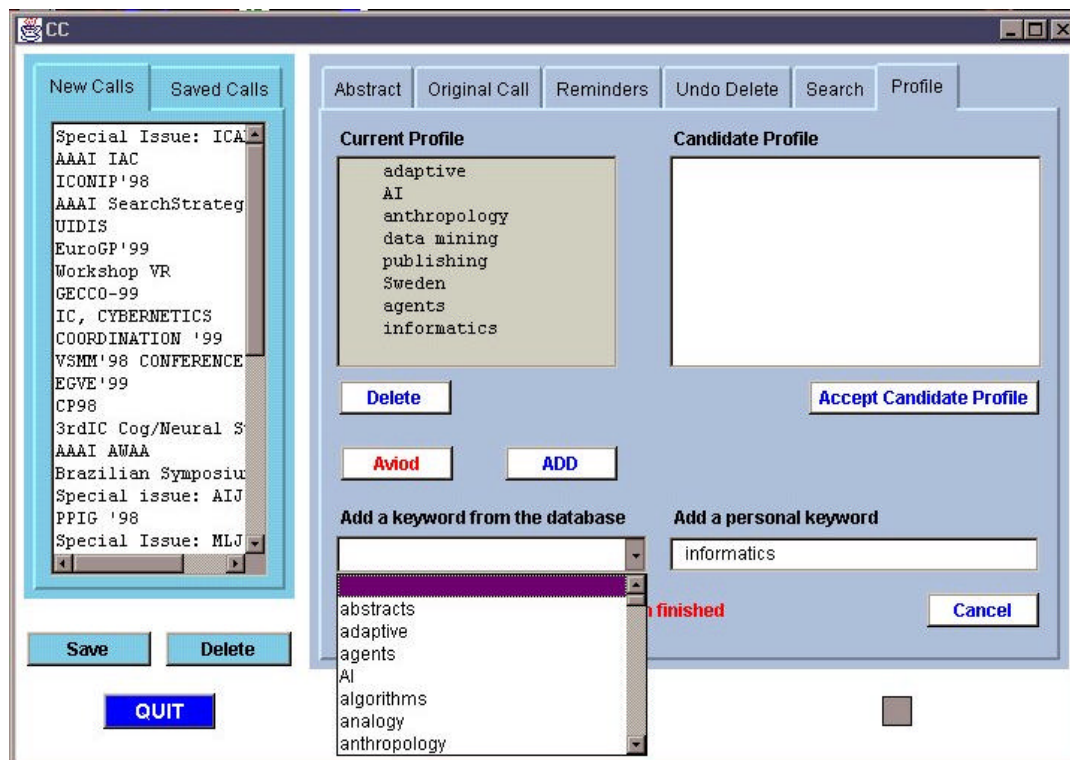
The system also logged all relevant user actions (save, delete, remind, looking at original call), as well as changes to user profiles. The intention was that this information would later be used in tuning adaptive functionality as well as comparing different algorithms for user adaptation.

The adaptation in ConCall is advisory, that is the users themselves set up their profiles manually and the system give continuously suggestions adapted after the users' behaviour. It is possible to tune ConCall entirely at run time since logs monitoring the actual user behaviour can be reviewed. This will have the effect of letting the adaptivity come out of an actual usage of the system, instead of having to infer rules for adaptivity from more constructed situations. Though a problem with run-time bootstrapping is that initially the advisory functionality will perform poorly and give bad advice. This in turn could (often does) lead to that users get initial bad results and thus will experience problems in trusting the system, which will carry on even though the system will eventually perform much better (see further Averman and Waern, 1999).

### ***The ConCall system***

#### Functionality, architecture and implementation

The ConCall service is an agent-based system and supports collection, filtering and browsing of calls of papers and participation (CFPs) (Waern et al., 1998 and 1999). The ConCall service enables the user to set up a personalised filter, the user profile. ConCall then use this profile to filter through a database containing CFPs and present the user with an individualised selection of calls. These calls are then open for the user to view, brows through and set up reminders for deadlines on. The profile is set up by the user her/himself through adding pre-defined “buzzwords” from a buzzword-list and by adding her/his own choice of words to her/his “current profile”. The user-added buzzwords do not immediately affect ConCall's choice of calls, but is instead meant to be a channel of communication to the editor (as feedback), so as to let the editor modify buzzwords and annotations to better suit the users' needs.



**Figure 2** The profile tab in the ConCall user interface, showing the buzzwordlist at the middle-bottom and the field for user added buzzwords at the bottom-right corner. (The user’s profile is shown in the text-area named “current profile”)

### Technical Details

The ConCall system is built up of a number of agents, each performing tasks within its area of specialisation. The following agents are at work within the ConCall service:

**The Personal Service Assistant (PSA)** – handles the interaction with the user and provides a central point of interaction between the user and the agents in the architecture.

**The User Profile Agent** – stores the preferences of the user. The user profile is based on information about the user’s actions received from the ConCall agent. The user can inspect and change the profile.

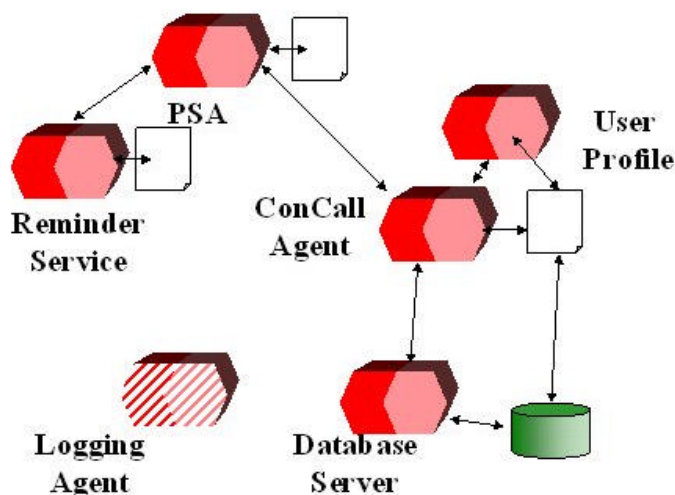
**The ConCall Agent** – does the filtering of calls for each user.

**The Reminder Agent** – provides a reminder service to the user also interacts with the other agents.

**The Database Agent** – handles transactions with a database that stores the conference calls.

**The Logging Agent** – is accessible from all other agents and enables agents to keep a record of events.

The agents communicate through KQML messages and the content is represented in Prolog terms. Users communicate with the agents by special-purpose user interfaces. The Personal Service Assistant and the Reminder Agent have their own user interfaces. The user communicates with the User Profile Agent, the ConCall Agent and the Database Agent through one shared applet. The agents in ConCall have one interface towards user and one towards other agents. (See Figure 3)



**Figure 3** The ConCall architecture.

### The Buzzwords and the Buzzwordlist

The word buzzword was chosen to represent the annotations for calls over the word ‘keyword’. This was partly done so not to confuse or draw any implications from the already established meaning of ‘keywords’.

The idea of using buzzwords was that the annotations should be signified by a buzz flavour. A concern for the profiling and use of annotations in ConCall were to create a structure as open as possible, with high potential of adaptation and to still keep it simple in maintenance. To achieve this, neither users nor editors were to be limited by any one specific or predefined ontology when setting up their profiles or choosing annotations. The buzzwords are then chosen not only after their quality of describing the topic of a CFP, but could also be words representing the conference’s geographical location, a committee name or other associated annotations.

### The User Profile

The user profile is set up and maintained by the individual user. The user is given the full list of buzzwords available. They then pick out buzzwords from this list that represents his/her needs and interests, and add them to the profile. The users are also given the option to add their own buzzwords. Words that they think is missing or better represent their interests. These user-added buzzwords do not affect the filtering during the same session but are rather a means of feedback to the editor; part of the indirect channel of communication between the users and the editors.

## The Editor Role

The editor’s role is that of an information broker. The editor is responsible for adding new information to the database. He/she is also responsible for reacting on the users’ behaviour, and to adapt the information in the way it is annotated and the use of buzzwords. The editor either seeks out new information or makes use of such things as mailing lists. The editor will need supportive tools to handle both the editing of calls and the maintenance of annotations and buzzwords. This support can come from tools that graphically display structures of grouped calls, indicate that other versions of the same buzzword is used, or that calls are outdated, etc.

## III. Study Method

### Study Aims

The test study done on ConCall had the following aims:

- To find out whether the buzzword structure is a sufficient way of communication between users and editors/information brokers.
- To gather information about which, of a number of possible extensions, is the most appropriate to deal with potential problems concerning the communication between editors and users.
- To collect personal profiles (logged automatically by the system) that made up the feedback to the editor to use for the following test study and to for use of tuning the user modelling algorithms.
- To evaluate ConCall as a service, and reconnaissance assumptions and expectations the reader might have about the system and the service in general.

It is important to emphasise that the design and layout were not any direct points of evaluation in this first test, though some data and observations were collected about this as well, reported as “side results”.

The following are questions that were written down in order to ease and structure the formulation of the test study. It also served as a direction in what to look for and observe during the test study.

- Do users/readers change their profiles once they are set up?
- Do users/readers ever type in their own buzzwords, or do they choose from the buzzwordlist or from the suggestions in the candidate profile?
- Do users need a more expressive way of formulating their needs (profiles)?
- Would users find it useful to review other users’ profiles?
- Would users like to have the possibility to categories their personal calls?
- Would users appreciate recommendations and if so from whom? [editors, friends/colleagues, special groups]
- Of how much importance would a reminder service be to the users?

## Subjects

There were 11 subjects in the experiment. All the participants had higher academical education, either master students, Ph.D. students or senior researchers. They all had extensive experience with computers and most of the subjects (eight out of eleven) read and handled CFPs on a regular basis. All of the subjects knew what a CFP was, and were familiar with conferences within their field.

## Study Structure

The test was conducted through three-step, individual sessions. The first step was a 10-minute part where the test participant were put before two interview questions, and then asked to fill out a questionnaire. Then ensued the actual running of and interacting with the system, taking approximately 20 to 30 minutes. Each participant logged in with her/his email address or something similar, and the logs corresponding to each participant was labelled with this login name.

The test-run was followed by a part where the participant got a paper stack of conference calls, containing all the conference calls within the database. The participant was asked to sort through the calls and indicate which, if any, calls they would have liked to have seen, i.e. any of interest regardless if they been presented to them during the hands-on testing of the system. The callid<sup>4</sup> number was written out on each call so that the monitor could relate to the information in the logs. The sorting through of calls is to see if any calls were missed when using the profile filtering in ConCall. We want to see if the profile filtering is doing its intended job, i.e. to accurately provide the reader/user with calls that are of interest to that particular reader/user. The data from we got out of this part were later used together with the logs to calculate precision and recall. The last part was a questionnaire with 10 questions evaluating post-reactions and possible extensions to the system were put to the user.

**Note:** Questionnaire I and II are attached and can be found in Appendix I.

## The editor role; editing of calls and adding buzzwords

The test was angled to look at the users and not the editors and therefore a secondary method of choosing buzzwords and annotations was used. The editor did neither have full-feathered support in means of tools nor pre-data of user preferences.

The buzzword-list that was used in the first test study was generated from uninformed annotations that in turn had been collected from researchers, that in turn had forwarded call for papers in e-mail format. The sender annotated the forwarded conference calls and the test study monitor was acting editor and administered the data collected and added the information to the database. The call for papers collected were kept intact and added to the database as original calls, though the e-mail headings like “from:”, “to:” and “subject:” was stripped of, as was more personal messages and annotations in the e-mails.

---

<sup>4</sup> All the CFPs in the database had an associated callid number. This number was not dependent on anything but the order the CFP was entered into the database. Neither users, editor nor the system were affected by the fact that a certain CFP had a certain callid.



## The Logs

By “remembering” user-induced events<sup>5</sup> the system generated the logs. Date, user id, the current buzzwords in the profile, the conference call name and the callid is also logged. These data together with the data collected at the sorting through of calls (described above) will serve as the input when assessing the filtering performance of the ConCall system. The personal profiles generated at the users’ first use of the ConCall system will be saved. These logs will also be the used tune the user modelling for future study.

## Measures

This study is using indicators such as precision<sup>1)</sup> and recall<sup>2)</sup> from information retrieval to assess the performance of an information filtering system. (The use and intentions of precision and recall are further described under *Background/Filtering- and Recommender Systems/Measuring Effectiveness in an Information Filtering System*)

### How Precision and Recall were Calculated in this Study

$${}^1) \text{ Precision} = \frac{X \text{ (CFPs shown that were of interest to the user)}}{Y \text{ (total of CFPs shown)}}$$

$${}^2) \text{ Recall} = \frac{X \text{ (CFPs shown that were of interest to the user)}}{Z \text{ (total of CFPs the user found interesting in the DB)}}$$

All other measures or indicators are of a qualitative nature. Scales have been used, ranged between 1 and 5, with the base unit of one. The scales have been given with an explanation of the approximate range of the scale, e.g. [1 – unimportant, ..., 5 – important].

## IV. Results

**Note:** Detailed summary of the results can be found in Appendix II.

### ***Comments and expressed needs for handling CFPs***

When talking to the test participants, there was a generally expressed need for a tool to filter and sort through incoming CFPs. Comments like the following expressed the participants’ current situation:

- “It is hard to keep up with all existing conferences.”

---

<sup>5</sup> I.e. actions performed such as delete, save, remind, and looking at abstracts.

- “I usually do not have time to handle all the CFPs that comes in through e-mail etc, and usually do not have the time to read all of them.”
- “The CFPs comes in unsorted, and I rather not go through a whole list of unsorted CFPs every time.”
- “Both interesting and uninteresting calls comes through with e-mail.”
- “Too much time goes to reading through calls from unknown senders, and one is forced to read through the whole call to be able to judge if it is relevant.”
- “The ones that comes in are often irrelevant.”
- “I want to get calls from someone that I can trust know that certain calls are relevant and of interest to me.”

One can notice a general frustration in these comments. There is a frustration not being able to handle all the incoming information. There is also a sense of frustration over the fact that it’s so hard to keep track of all the conferences there are and especially not being able to sort out or even acquire CFPs to the conferences relevant and/or vital to the user.

Needs expressed, was given through comments like the ones in the following list:

- “I have need for a way of handling calls that comes in.”
- “I could use some form of graphical overview, a sort of time-scale like a calendar that preferably stretches over a full year, in order to be able to see annually conference happenings.”
- “I would have a need for getting reminded of deadline-dates etc.”
- “I have a need to find out if a certain conference is of interest to me or relevant to me.”
- “I have a need to find out about deadlines etc, in order to be able to synchronise scheduled events and reminders and to get reminders some time before.”
- “I need a way to sort calls into categories like: papers, short papers, etc.”
- “I have a need to find out which conferences exist.”
- “To be able to filter our irrelevant calls.”
- “To get summaries and abstracts from CFPs, especially about unknown conferences, so one quickly can judge if this conference is of interest of not.”
- “To be able to filter out calls, there are a lot coming in at the same time.”
- “To get relevant calls.”

From these comments and other think-aloud comments, three major expressed needs could be identified: (1) a reminder service, (2) a filtering service, (3) to be able to sort CFPs into categories.

Researchers (test participants) with a little less experience of sorting through CFPs, found as it is today hard to figure out which conferences that really were of any interest to them, and therefor worth spending time on reading the whole CFP for, i.e. problems judging the relevance.

One question that repeatedly occurred in the interviews was the issue of trust. That is, does the service really do what you think it is doing, i.e. will it give me *all* the conference calls I want? If the user feel that she/he cannot trust the service then she/he would have to double-check the information, and the time and work that was supposed to be saved, by using the service, is lost. The trust issue really relates to two aspects in this case. One, being able to trust ones information broker. Two, being able to trust the profile filtering. The former is

somewhat more open to user influence i.e. a user can choose a/several information brokers or change editor if not satisfied. (Though not included in the test the idea is to have several editors or information brokers providing its users with information through ConCall.)

What did not come up though when talking about trust, was the trust of privacy that so often is discussed when dealing with systems using user modelling today (Höök, 1998). This could be due to the fact that it was not clear to the participants that their maintained profiles were going to be reviewed. Or even more likely, that the experimental situation lulled the participants into a sense of security, since this situation in a sense was perceived of being fictitious and therefor not a threat.

### ***Three Major Themes and Recommendations***

Three major themes were identified in the test study result: Theme1 – Reminders, Theme2 – Friends and Colleagues, Theme3 - Filtering Performance. In addition to these themes a discussion over extensions to the current version arose and a general discussion about recommendations is included. There were issues that came up that really was outside the test study scope, but is included as “side results”.

#### **Theme1 - Reminders:**

To be able to set reminders on up-coming calls were a service almost all of the test participants found to be something that could become essential for them. As a matter of fact, most of the participants wished they had a way of putting reminders on deadlines for CFPs already today. Nine out of eleven gave it a high priority grade, while two gave it one grade down and one ranked it to medium importance. There is clearly a trend towards having a memory-supportive function of some sort. The reminder function did not only get strong support because of its memory-support, but also because it could give the users an organisational support, through its potential to graphically and/or time-wisely plot out events. This would give the user a better over-view and a way of putting events in relation to each other.

#### **Theme2 – Friends and Colleagues:**

When the test participants were put before the question of whom they would prefer to get recommendations about conferences from, friends and colleagues stood out to be the preferred source. Having an editor or a person with “similar” interests giving recommendations were the other two alternatives. Nine out of the eleven test participants showed a high interest in getting information about friends and colleagues’ profiles.

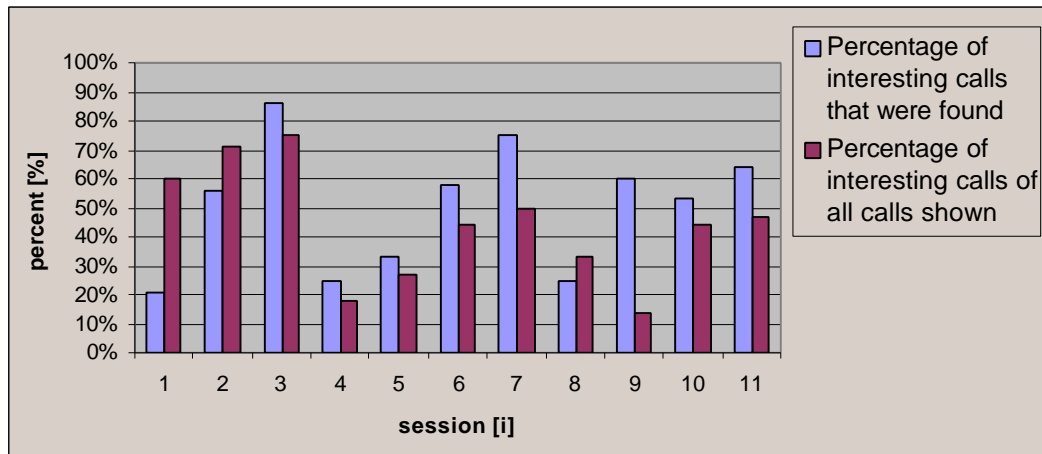
The users seem to prefer to get recommendations from someone they have a perception about. At least from someone they know something about, in order to base some form of opinion about the information (recommendation) provider.

#### **Theme3 – Filtering Performance:**

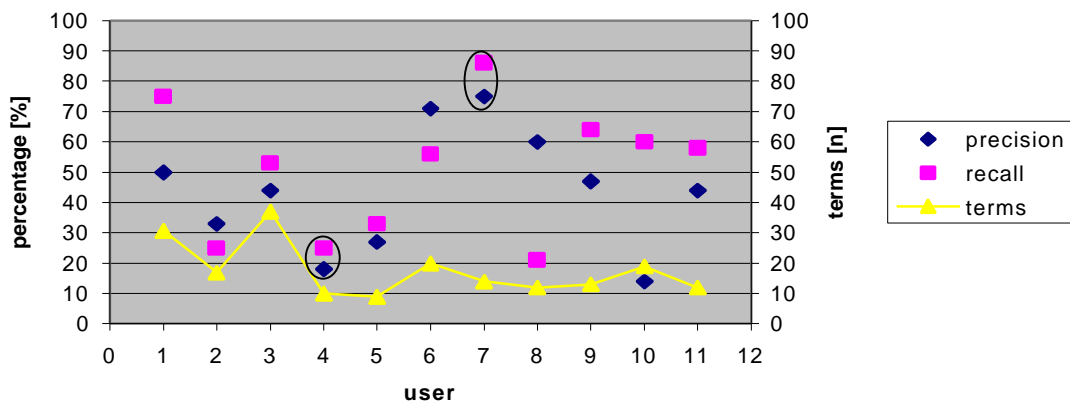
The filtering performance was measured by looking at the percentage of interesting calls that were found and at the percentage of interesting calls of all calls shown during the hands-on test-runs. Since the buzzword annotations in this first study were not tuned to real user needs,

the results were expected to be rather bad in terms of precision and recall. The interest lie instead in seeing if users were able to accomplish this task at all i.e., if some users were able to set up working filters, and if so, how they accomplished this task.

Another aspect is that the experimental situation in itself did not encourage users to tune their filters. Participants were encouraged to tune their filters until they were reasonably satisfied with their performance. Nevertheless, most users would just toy with the system until they had understood its functionality. Since they were not able to use the system outside of the study, there were little incitement for them to spend any effort in setting up a working filter.



**Figure 3:** ConCall filter performance: recall (left columns) and precision (right columns). Showing precision and recall in terms of users being able to set up their filters so that the interesting calls were retrieved.



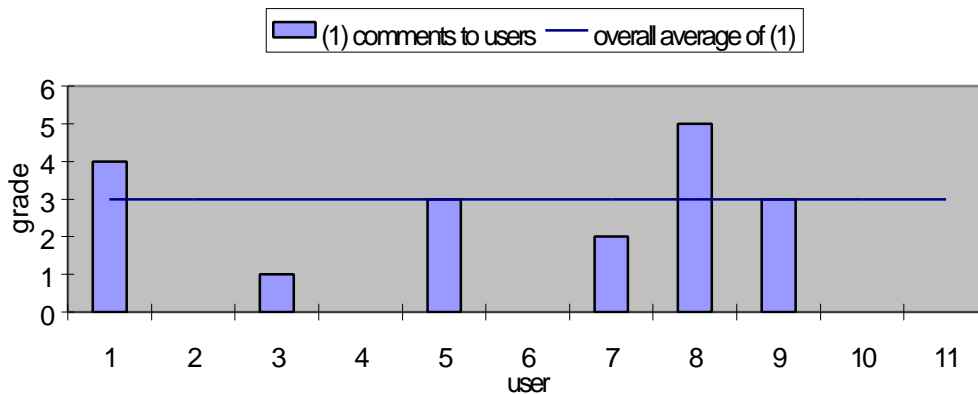
**Figure 4:** Precision, recall and number of terms entered into a profile.

As can be seen, on average both precision and recall was very low in this study. As mentioned earlier, this was expected. What is more interesting is if it is at all possible to achieve good performance with this type of filter. The graph (see figure 3) shows that one of the users, user 7, achieved fairly good results in both precision and recall. This user was highly motivated, and simulated as near a *real-life* situation this test could make possible. The same user was satisfied with the system and tuned the filter more than most others. User 4 on the other hand

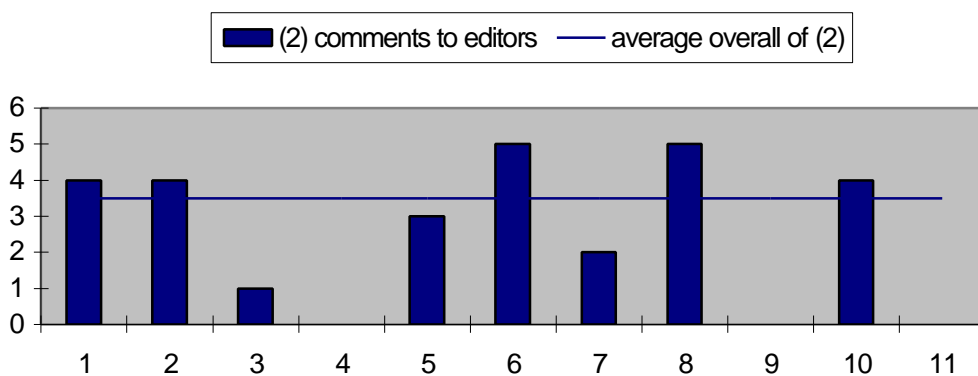
had both a low recall and a low precision. Though both user 7 and 4 had entered a fairly low amount of terms into their profiles (14 and 10 terms respectively). Both changed their profiles more than any of the other participants. This suggests that there is something more than the amount of terms in a users filter or corrections over time to their profile that influences the precision and recall for a set of users. The only observable difference between them was their degree of motivation for using the system. User 4 mostly flicked through the system and did not bother about being as sincere in using it. User 7 had more experience in his/her field when it came to handling calls for papers.

### About Recommendations

When asked if they could see any usefulness in being able to give comments about conferences and CFPs, just over half of the test participants showed only a moderate interest having other users as the target (See Figure 5, (1)). Eight out of eleven participants showed slightly more interest in giving comments as feedback to the editor or information broker (See Figure 6, (2)).

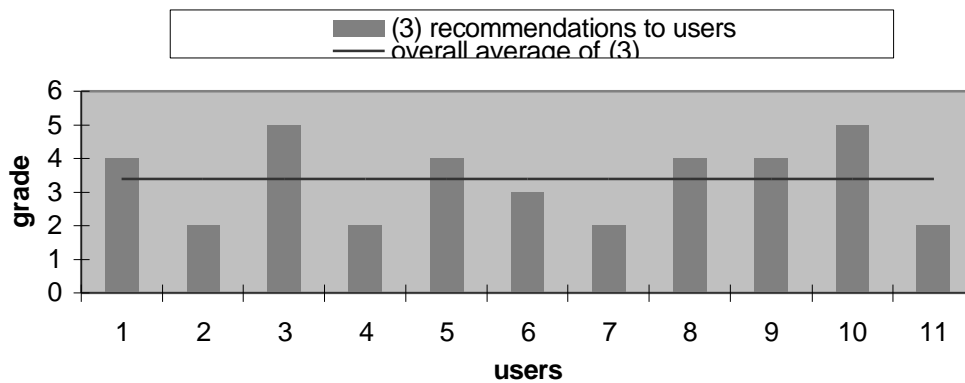


**Figure 5:** Users' interest of giving comments targeted at other users.



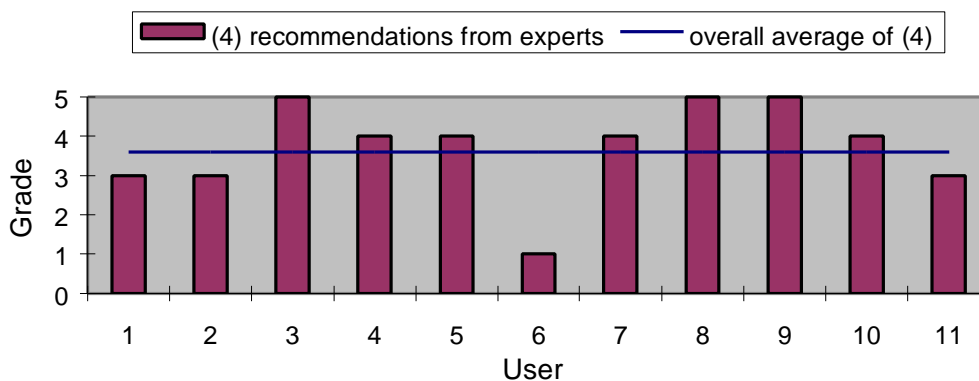
**Figure 6:** Users' interest in giving comment target at editors.

There was a slightly higher interest (seven out of eleven indicated a grade of 3+) in being able to give recommendations about conference calls to other users (See Figure 7, (3)).



**Figure 7:** Showing the users’ interest in being able to give recommendations to other users.

When asked how much it would mean to get information about how an *expert* ranked or felt about a conference, nine out of eleven indicated a high interest (See Figure 8, (4)).



**Figure 8:** Showing the users’ interest in getting recommendations from an expert/experts.

### Perception about the information provider/broker

An insight into what the information provider is about could be based on, among other things, personal experience (friends and colleagues) and professional respect (colleagues and experts). Pre-opinions about the provider of information, recommendations and comments could give an impression of security, in the sense of being able to trust the provider to be accurate and relevant. This could be because there is a ground to base a judgement on.

## ***Outside the Scope of the Study – “side results”***

### **Profile Handling and the Candidate Profile**

The majority (eight out of eleven) of the participants found setting up their own profiles a fairly easy task. The way that predefined buzzwords were available and the feel of building your own profile with provided *building blocks* was well received. The participants also appreciated the opportunity to add their own buzzwords. However, this feature was somewhat misunderstood by all the participants. They all thought that the words they added would affect the filtering within the same session, which was not the case since the user added buzzwords had another intent. The intent was that the user added buzzwords would serve as one of the feedback channels to the editor, giving suggestions to what the users thought were missing or needed rephrasing.

The candidate profile was received with mixed responses. The suggestions presented in the candidate profile sometimes agreed with the user’s area of interest, but it did also come back with suggestions that were of no relevance, or once or twice even with something contradicting the users interests (as they were represented in the user’s profile). Despite this the suggestions were welcome by the participants. The use of the candidate profile suggestions were quite low, which was mostly due to the fact that the participants was preoccupied with examining their results from the profile filtering and few went back and changed their profile at all.

### **Feedback on the GUI**

When using the ConCall service, a few points came up regarding the user interface. The interface is still in a demo version layout and has not been worked on as much. The next version will have a much more intuitive interface where the layout has been thought through much more. When developing the prototype, evaluated herein, the GUI had not been a priority.

Among a few comments about the GUI, the users thought it was unclear which “tab” did what and the profile tab was the one the users had most problems with. Since the test participants only used the system for a short period of time (20-30 minutes) no or little time was given for the users to really get to know the user interface.

One thing that got brought up was a wish for highlighting buzzwords. Both having highlights on the buzzwords that existed in the user’s profile and occurred in abstracts and original calls, and having indications in the candidate profile over which buzzwords that were already a part of the user’s profile.

### **The Wish for a Graphical Overview**

An interesting point that came up was the need for somehow getting a visual representation of reminders, deadlines, interesting conferences, etc. This has so far no support in ConCall except for the linking to a reminder service with a calendar design, and this service was not included in the test and will therefor not be evaluated as an existing and functional service here.

One idea was to create a time-line that visualised where in time the deadlines, etc. would lie in respect to each other. Another idea was to link reminders to a calendar that the user already

used, like the Netscape Calendar, and automatically get the reminders synchronised or added in to the calendar.

## V. Discussion and Suggestions for Redesign

This test had the aim of being a pre-study to evaluate the service and the functionality of the implemented system ConCall. The test aimed to show how the users of the system could interact with it and if it actually was beneficial to the users. I call it a pre-study since this is the first step of an evolutionary development and is meant to be a stepping stone for further evaluative tests and improvements. The following is a discussion and some suggestions for future work on and redesign of ConCall.

### Design Issues

When it came to profile handling either a better explanation or modification of the function and layout is needed. The interface gave way to several misunderstandings and some frustrated comments from the participants, though the profile set-up was found to offer a flexible way of handling one’s initial profile set-up as well as maintenance. There is however a need to redesign the interface, so to diminish the possibility of users just getting frustrated by the system and consequently stop using it or be less sincere in using it.

An interesting suggestion from a couple of the test participants was to add a highlighting feature to the interface. The participants wanted to have it as an option to be able to let the system highlight buzzwords present in the users’ profiles when they occurred in a text. Particularly at the abstract viewing interface but also when viewing the original calls. This could be a valuable feature to add to the systems interface, since it will make it easier for the users to judge a text’s relevancy.

Another function that would be most useful to the users and that one of the participants expressed a wish for, is a *undo* functionality in the system, so that mishaps can be corrected.

### User Profiles

The profiles gathered through the logs were intended to be used for tuning the user modelling in ConCall, but did not prove to be of much help. This due to the fact that the annotations used were suffering from a cardinal problem in how they were both gathered and used by the editor initially. The effect of having to initialise annotations in an unused collaborative or recommender system is not a unique problem for the ConCall study, but the lesson learned would be to at least consider alternative methods (Averman and Waern, 1999).

The study gave low figures in terms of precision and recall, which showed that the profiles obtained during the study actually were bad profiles. This poses a serious problem for using this data in bootstrapping adaptivity, as the profiles that the system *should* generate are still largely unknown. A suggestion could be to instead generate optimal profiles 'backwards', from the ratings of documents by users. Instead of using the profiles that users themselves constructed, a profile could be constructed for the user that got the highest “relevancy” in retrieved calls, and use these profiles as the optimal target for adaptive behaviour of the system.



## The Experimental Situation

The ConCall study produced an experimental situation that led to a somewhat unnatural situation. Unfortunately the experiment gave some undesirable effects due to the fact that the set-up of the experiment did not facilitate the participants with the situation of “normal” interaction. This in turn led to the fact that some of the participants were not encouraged to perform as sincerely as they would given normal usage. A suggestion for future studies of the ConCall system would be to take better care of creating a more undisturbed and less restricted situation when letting the participants use the system.

It is important for a system such as ConCall to have a fairly high motivation in both users and editors, since the recommender feature in ConCall will otherwise suffer from insincere recommendations and annotations which will lead to fake situations, not giving the desired results.

It is also important to take into account the fact that a system like ConCall would benefit (probably crucial), from having a critical mass of users. Therefore it would be advisable to try and get a larger group of participants in future studies, in order to give the system as optimal prerequisites as possible.

## Trust and Privacy

An important issue that has not got much attention in this report is the issue of trust and privacy. No real incidences indicated that it was a problem for the test participants that their profiles were used and reviewed by the editor. Though this is not to say that it will not arise as a problem in future versions or tests of ConCall. Also as discussed above in the results section, that this was not a concern for the participants could be due to the experimental situation. In future studies this might be something to look (out) for.

## Editor Aspects and Support

The editor had few or not tested tools for adding and editing of CFPs and annotations to the database, this in turn led to a somewhat primitive situation for the editor. The work of adding new calls were a time consuming task, which in an experimental situation as the ConCall study is acceptable but will not be in future studies in larger scale, where more than one editor is going to be used. It is important to support the editor so that the task will be as easy as possible. One can imagine that both a graphical overview of the database structure (if categorisation is to be implemented) and over CFPs and annotations could be useful. Another tool that could be most useful and should be implemented, is a good way of viewing the users’ profiles, since otherwise it would be impossible for the “loop” to close and no convergence or adaptivity would be possible.

## References

Averman, C., and Waern, A., (1999) The Hen and Egg problem of Bootstrapping Adaptive Services, Ensuring Usable Intelligent User Interfaces, i3 Spring Days Workshop, March 1999.

Balabanović, M., Shoham, Y., (1997) Content-Based, Collaborative Recommendation, Communications of the ACM, Vol. 40, No. 3, pp. 66-72.

Belkin, N.J., Croft, B., (1992) Information Filtering and Information Retrieval: Two Sides of the Same Coin?, Communications of the ACM, Vol. 35, No. 12, pp. 29-38.

Höök, K., (1998) Steps to Take Before Intelligent User Interfaces Becomes Real. *Interacting with Computers*.

Höök, K., Rudström, Å. And Wearn, A. (1997) Edited Adaptive Hypermedia: Combining Human and Machine Intelligence to Achieve Filtered Information. Available at <http://www.sics.se/~kia/papers/edinfo.html>

Oard, D.W., (1997) The State of the Art in Text Filtering, User Modeling and User-Adapted Interaction, 7, pp. 141-178.

Robertson, S.E., (1981) The methodology of information retrieval experiment, Information Retrieval Experiment in K. Sparck Jones, Ed. Chapt. 1, pp. 9 – 31, Butterworths.

Van Rijsbergen, C.J., (1979) Information Retrieval, Second Ed., ISBN 0-408-70929-4, Butterworths.

Waern, A., Averman, C., Tierney, M. and Rudström, Å.(1999) Information Services Based on User Profile Communication. in Proceedings of the seventh International Conference on User Modelling, Banff, Canada, forthcoming June 1999.

Waern, A., Tierney, M., Rudström, Å., Laaksolahti, J. and Mård, T. (1999) ConCall: Edited and Adaptive Information Filtering. Proceedings of the Intelligent User Interfaces Conference, Los Angeles, California.

Waern, A., Tierney, M., Rudström, Å. and Laaksolahti, J. (1998) *ConCall: An information service for researchers based on EdInfo*. SICS Technical Report, T98:04, ISSN 1100-3154.

## Appendix I

---

Oviktigt	Viktigt			
<input type="checkbox"/>	<input type="checkbox"/>			
<input type="checkbox"/>	<input type="checkbox"/>			
<input type="checkbox"/>	<input type="checkbox"/>			
<input type="checkbox"/>	<input type="checkbox"/>			
<input type="checkbox"/>	<input type="checkbox"/>			
1	2	3	4	5

### Frågeformulär (I):

Hur viktigt tror du att följande skulle vara att ha med?  
1. Att ha möjlighet att söka efter aktuella kallelser i den befintliga databasen?

2. Prenumerera på nya kallelser?

➤ Och i sådana fall hur tycker du att det borde fungera?

Genom till exempel:

- speciella konferenser
- kategorier
- en form av intresse-profiler

eller genom:

3. Hur viktigt skulle det vara för dig att kunna få rekommendationer om kallelser som "liknar" det du prenumererar på?

➤ Och i sådana fall, vem skulle du vilja få rekommendationer från?

- Av personer med liknande intressen som tyckte något var bra.
- Av kollegor eller vänner.
- Av en redaktör som förstått att det skulle passa dig.

4. Hur viktigt skulle det vara för dig att ha möjligheten att sortera kallelser i ett eget "bibliotek"?

5. Hur viktigt skulle det vara för dig att ha möjligheten att rekommendera andra om bra kallelser?

6. Hur viktigt skulle det vara med möjligheten att kunna ge kommentarer till kallelser, som feedback till:

• Andra läsare?

• Till redaktören?

7. Hur viktigt skulle det vara att ha en påminnelsefunktion för deadlines?

**Frågeformulär (II):**

---

1. Tror du att du skulle kunna ha användning av en service som ConCall?

Ja	Kanske	Nej	Vet Inte
----	--------	-----	----------

2. Vad tycker du om ConCall efter att ha använt systemet, på en skala 1 till 5?

1	2	3	4	5
---	---	---	---	---

(Där 1 är dåligt och 5 är bra)

3. Nämn åtminstone en funktion som du tyckte var bra.

4. Nämn åtminstone en funktion som du inte tyckte var bra eller kunde ha gjorts på ett bättre sätt.

5. Tyckte du det var lätt att använda ConCall, eller fann du några svårigheter i att använda det? (Där 1 är lätt och 5 är svårt)

1	2	3	4	5
---	---	---	---	---

Skala för fråga 6 till 11:

1 2 3 4 5

---

6. Tyckte du det var lätt eller svårt att sätta upp (och/eller konfigurera/ändra) din profil, (på en skala mellan 1 och 5, där 1 är lätt och 5 är svårt)?

7. Indikera hur viktig någon eller några av de följande skulle vara som hjälp för dig när du sätter upp din profil: (På en skala mellan 1 och 5, där 1 är oviktigt och 5 är viktigt)

➤ Information om andras profiler som liknar dina.

➤ Information om vänner eller kollegors profiler.

➤ Information om typiska nyckelord för en viss informationskälla eller ett visst område.

---

Vilken eller vilka av de följande funktionaliteterna skulle du ha velat sett i systemet och hur intressant skulle det vara att ha med dem (skala 1 till 5, där 1 är ointressant/oviktigt och 5 är intressant/önskvärt/viktigt)?

8. Möjligheten att kunna kategorisera buzzwords efter typ av information, t.ex. efter plats.

- Ett exempel skulle kunna vara:  
"Patty Maes" är ett viktigt ord om det förekommer i programkommitte-kategorin.

9. Möjligheten att kunna indikera att ett visst buzzword bara gäller vid en viss typ av konferens?

- Ett exempel skulle kunna vara:  
"Patty Maes" är ett viktigt ord om det handlar om en agentkonferens.

10. Möjligheten att kunna sortera dina profiler, d.v.s. att kunna ha olika profiler för olika intresse områden?

--	--	--	--	--

11. Att få information om hur viktig en expert/flera experter bedömer en konferens att vara (inom sitt område)?

## Appendix II

---

### Results from the Test Study

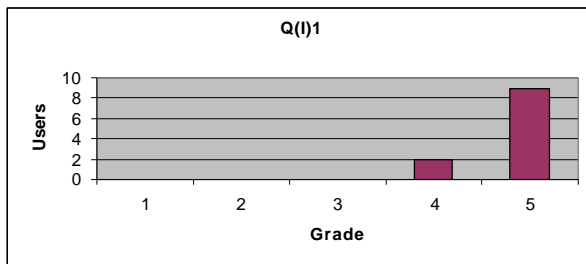
Even through the interview questions, from part I and II, come before the questionnaires I respectively questionnaire II they are presented after the results from the questionnaires in chapter V. **Results** under section *Comments and Expressed Needs for Handling CFPs*.

#### Questionnaire I

How important would any of the following be to you? [1 – no importance, ..., 5 – important]

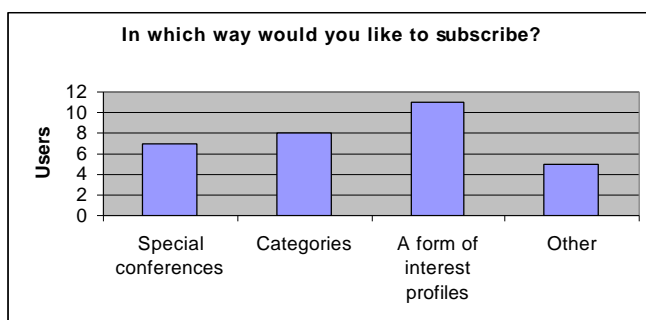
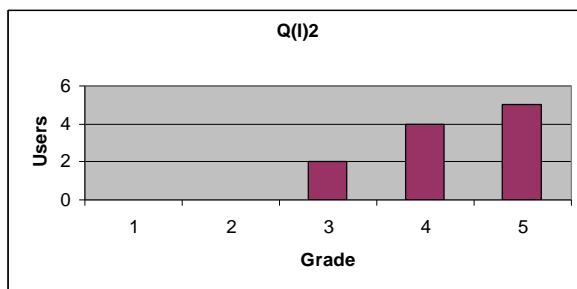
Q(I) 1: Have the possibility to search for calls in the database be for you?

$$\bar{X} \cong 4.8$$



Q(I) 2: Be able to subscribe to new calls?

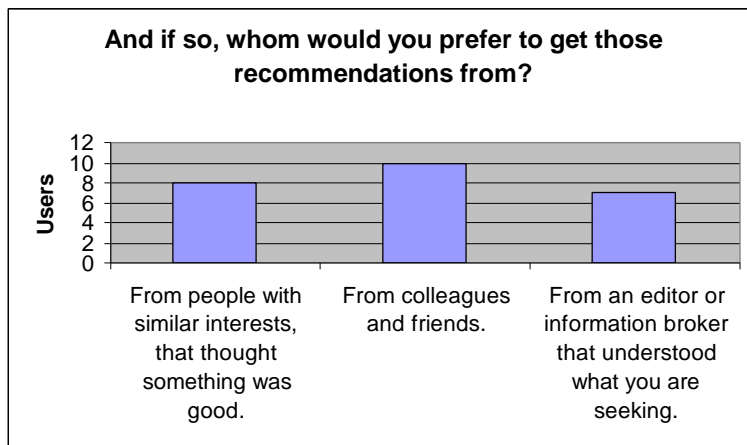
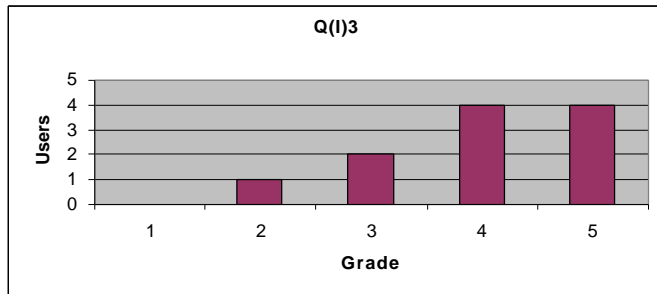
$$\bar{X} \cong 4.7$$





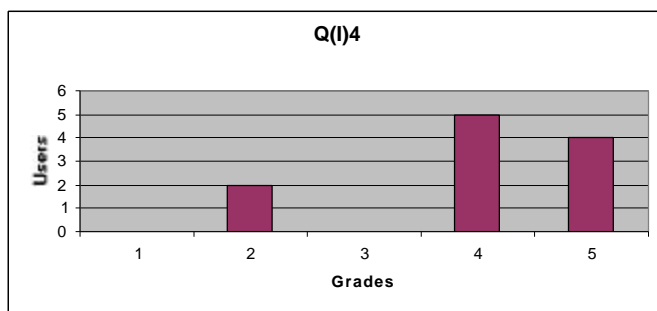
Q(I) 3: To get recommendations about calls that is "similar" to what you subscribe to?

$$\bar{X} \cong 4.7$$



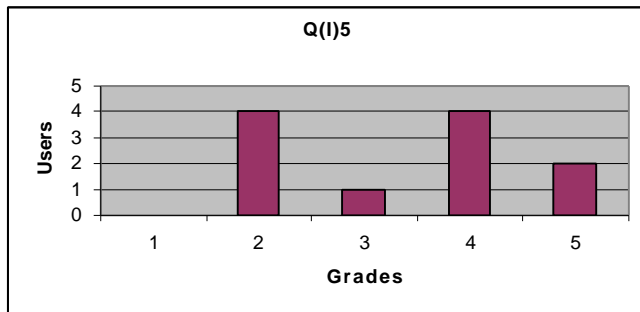
Q(I) 4: To be able to sort calls into you own "library"?

$$\bar{X} \cong 4.0$$



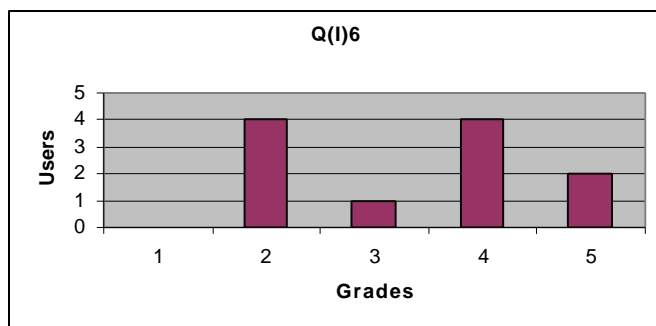
Q(I) 5: To be able to recommend other users/readers about calls you food interesting or good to know about?

$$\bar{X} \cong 3.4$$



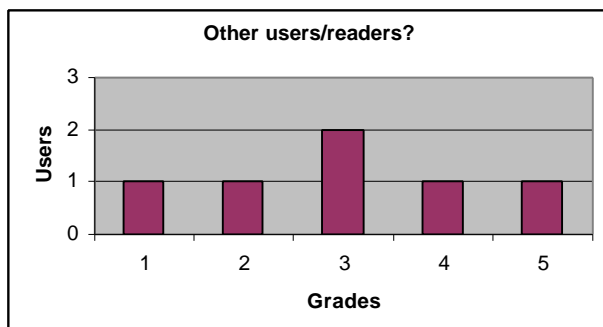
Q(I) 6: To be able to give comments on calls?

$$\bar{X} \cong 3.2$$



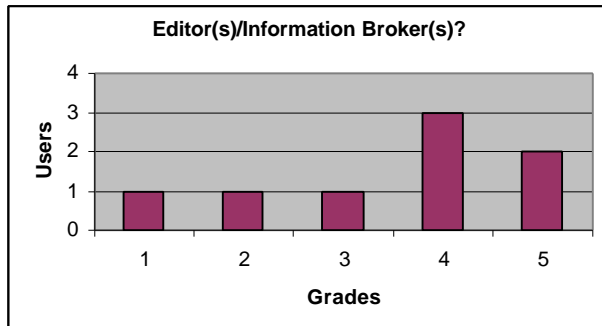
As feedback to:

$$\bar{X} \cong 3.0$$



Note: Six out of the eleven test participants answered this question.

$$\bar{X} \cong 3.5$$

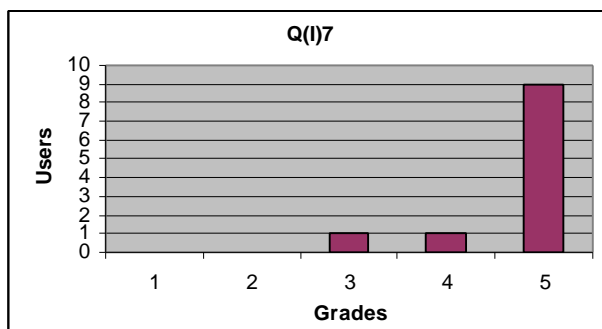


Note: Eight out of the eleven test participants answered this question.

---

Q(I) 7: To have access to a reminder service for deadlines?

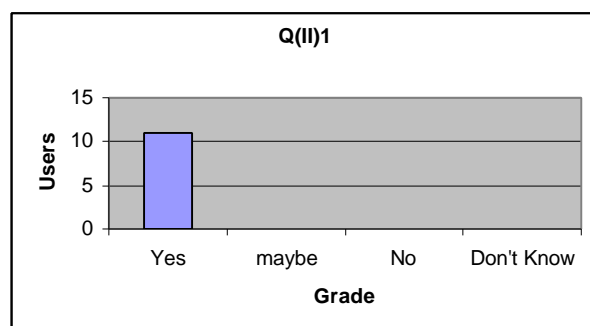
$$\bar{X} \cong 4.7$$



---

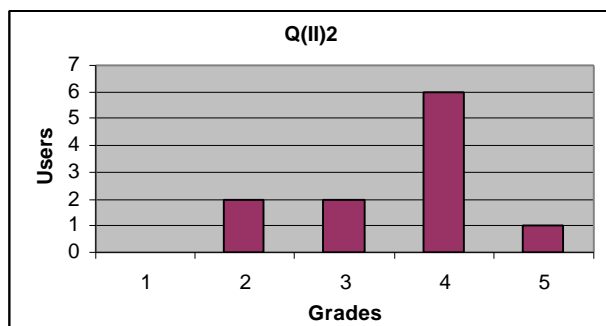
## Questionnaire II

Q(II) 1: Do you think you would have use for a service like ConCall?



Q(II) 2: What do you think about ConCall after using it? [1 – bad, ..., 5 – good]

$$\bar{X} \cong 3.2$$



---

Q(II) 3: Mention at least one function you thought was good.

- One "collected" (integrated) system for CFPs, where one can search for the information whenever one wants to. (compare to e-mail that always arrive at the wrong time and get lost!).
- To be able to create and build profiles and get calls filtered through it.
- To get suggestions on words to the profile.
- Buzzword-list is associated with every call, and the reminder function.
- The flexibility in handling the profile.
- The reminder service.
- Access to abstracts, original calls, the search function and that the system comes up with suggestions.
- The reminder function and the free-text-search.
- The reminder service and the filtering.
- That it prints out CFPs and that it searches out relevant conferences after my buzzwords. (A little bit of semantics wouldn't hurt though.)
- Probably time saving. The reminder function.

---

Q(II) 4: Mention at least one function you did not like or thought could have been implemented in a better way.

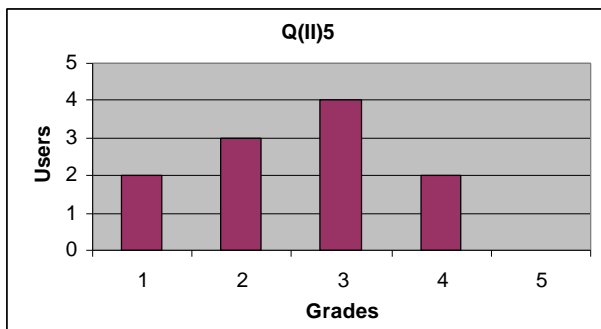
- The choice of buzzwords. Strange that my own added buzzwords did not affect the filtering.
- No direct manipulation, did not know if the system reacted, not intuitive when the update happened.
- I would have liked to see graded weights on the keywords and a better overview of the buzzword-list.
- When one is done with e.g. the "Reminders" and chooses new call from the "New Calls" list, one wants to get up the [now current] "Abstract" immediately.
- Save/Delete: Delete can be used to save calls to the next session (?) [By putting the calls in the trashcan ("Undo Delete")], but the problem with that is that the calls then get marked negatively, and thus interpreted by the system as being a negative action.

- If the “Abstract” is there to make it easier to judge if a call is worth spending time on reading, then the topics should be included and shown in the abstract, it will give more information.
- Highlighting of buzzwords in “Abstract” and “Original Call”.
- No extern linking of reminders. [No obvious linking]. Missed the possibility to list all the conferences.
- The Interface and the Buzzwords – too many and still did not reflect my “higher-level”-needs.
- The Interface needs some more work. The Reminder function – one had to save the conferences between sessions.
- The GUI. The interaction (sequence). The system rated fields as ”NOT interesting” despite my definition as interesting.

---

Q(II) 5: Did you think ConCall was easy to use, or did you find any difficulties in using the system? [1 – easy, ..., 5 – hard]

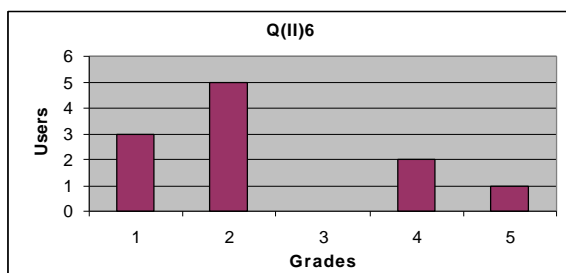
$$\bar{X} \cong 2.4$$



---

Q(II) 6: Did you find it easy or hard to set up (and/or configure/change) your profile? [1 – easy, ..., 5 – hard]

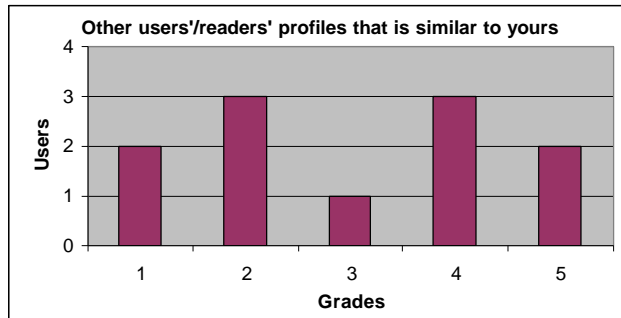
$$\bar{X} \cong 2.4$$



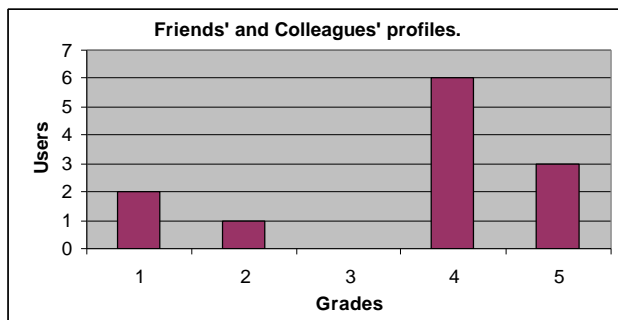
Q(II) 7: Indicate how important any of the following would be to you in helping you set up your profile? [1 – unimportant,..., 5 – important]

Information about:

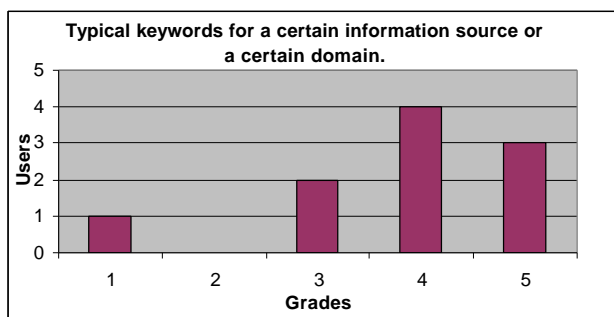
$$\bar{X} \cong 3.0$$



$$\bar{X} \cong 3.9$$



$$\bar{X} \cong 3.5$$



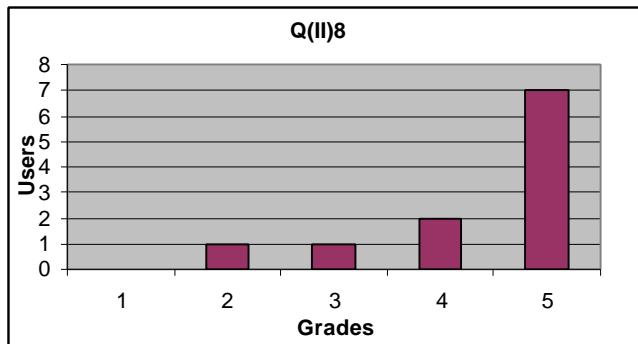
Which of the following functions would you have liked to see in the system and how interesting would they be to include in future versions?

[1 – uninteresting/unimportant, ..., 5 – interesting/important]

---

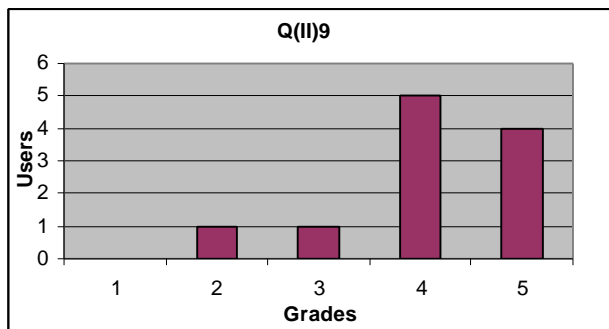
Q(II) 8: The possibility to be able to categorise buzzwords after type of information, e.g. after geographical location?

$$\bar{X} \cong 4.4$$



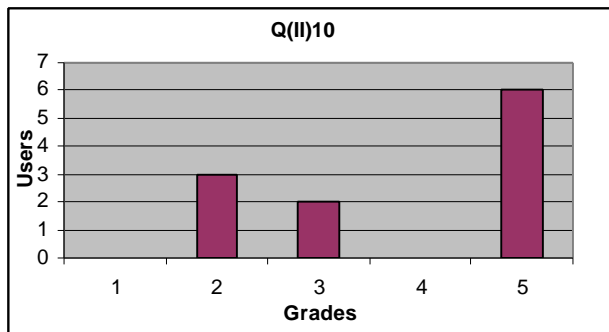
Q(II) 9: The possibility to be able to indicate that a certain buzzword only apply to certain conferences?

$$\bar{X} \cong 4.1$$



Q(II) 10: The possibility to be able to sort your own profiles, i.e. to be able to have different profiles for different interest domains?

$$\bar{X} \cong 3.8$$



---

Q(II) 11: To get information about what an expert/several experts think a certain conference?

$$\bar{X} \cong 3.6$$

