Considering Collaborative Filtering as Groupware: Experiences and Lessons Learned

Christopher Lueg
AI-Lab, Department of Computer Science
University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
lueg@ifi.unizh.ch

CEUR-WS.org/Vol-13/paper16.pdf

Abstract

Large information repositories, such as the World Wide Web or Usenet News, offer an amazing wealth of information. Searching for valuable information is an effort that permanently occurs in organizations and that contributes to the organizational memory. Active collaborative filtering builds on the exploitation of the results of this effort by encouraging the members of an organization to share information with others. To this extent, the members are provided with a tool that allows for the distribution of information with little effort. However, it is unclear which strategy is best to bootstrap or to enhance a self-supporting information dissemination activity. In this paper, we report on experiences we made when trying to bootstrap an active collaborative filtering activity in a medium-size organization. Our experiences suggest to view collaborative filtering as groupware in order to explain certain difficulties and to learn from the experiences with groupware.

1 Introduction

Large information repositories, such as the World Wide Web (WWW) or the global conferencing system Usenet News, offer an amazing wealth of information potentially suited to answer almost every question. Unfortunately, the information out there is not knowledge *per se* but the information has to be interpreted and to be used by humans in order to become knowledge [2, 10]. To this end, the relevant information has to be located first. Consequently, ways to locate useful information in the information jungle have become an increasingly hot topic.

So far, various tools have been developed to enable individual users to actively search information repositories for the information required. MetaCrawler[1], AltaVista[2], HotBot[3], and DejaNews[4] (the latter is specialized on Usenet News) are prominent examples for so-called search engines helping users to locate information in the web information jungle. Most search engines utilize proven information-retrieval techniques to deliver all information matching the user's query. However, to this extend, search engines require the user to explicitly describe the desired information, i.e, the user has to be aware of his or her interests. Also, the user must be willing to spend a significant amount of time on searching for the information.

A shift in perspective from individuals to organizations reveals that turning information into knowledge is a permanently occurring, distributed effort (e.g., [2]). Members of organizations often search information repositories in order to satisfy their personal information needs. The expertise needed to find relevant information in the information jungle is an important part of the organizational memory and ensures a constant flow of new knowledge into the organization. However, although all this contributes to the organizational memory, the effort is mostly geared towards

The copyright of this paper belongs to the paper's authors. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage.

Proc. of the 2nd Int. Conf. on Practical Aspects of Knowledge Management (PAKM98)
Basel, Switzerland, 29-30 Oct. 1998, (U. Reimer ed.)

http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-13/

[1] http://www.metacrawler.com/
[2] http://www.altavista.net/
[3] http://www.hotbot.com/
[4] http://www.dejanews.com/

individual information needs.

An open question is how an existing information sharing culture could be supported and enhanced. Also, it is unclear how to bootstrap a self-supporting information dissemination process. The issue is not merely a technical one in terms of adequate tools to disseminate information. Since it is crucial to *encourage* individuals to disseminate information and to share information with others, the issue is to bootstrap and to establish an "information sharing culture".

In this paper, we report on experiences and lessons learned when trying to bootstrap a self-supporting information dissemination process. We proceed as follows: first, we briefly summerize the active collaborative filtering idea. Then, we describe an experiment in bootstrapping an active collaborative filtering process within a medium-size organization. To this extent, we describe the dissemination tool being used in the experiment and discuss some of the assumptions underlying the development and the introduction of the tool. Then, we discuss the results of the experiment and contrast our assumptions with the results of an informal inquiry conducted after the experiment. Finally, we discuss lessons learned by interpreting active collaborative filtering as groupware and by comparing our experiences when trying to introduce an active collaborative filtering tool with experiences made by others when trying to introduce groupware applications as reported in the literature.

2 Active Collaborative Filtering

Members of organizations often search information repositories in order to satisfy their individual information needs. The idea of "active collaborative filtering" is to exploit the results of these individual efforts to locate interesting information in the information jungle. Such results are especially valuable since the information has already been found to meet certain quality criteria. Active collaborative filtering builds on encouraging the members of an organization to share interesting information with others. The term itself has been introduced at the 1995 CHI conference to describe a collaborative filtering application at Lotus Corporation [11]. "Active" as opposed to "passive" stresses that there has to be an intent on the part of the person who located this particular information (or the location) to share with others.

In order to bootstrap an active collaborative filtering process, participants are supplied with a tool allowing them to easily distribute "pointers" (see below) to others. The Maltz and Ehrlich tool [11] is implemented as an augmentation to the commercial groupware application Lotus Notes. The tool adds an additional button to the Lotus Notes SmartIcons bar. Clicking the button prepares a "pointer" to the Notes document the user is actually browsing. Such a pointer is a document reference which is automatically augmented with some contextual information, such as the title of the document, its creation date, and the database name. In addition, the user may add comments to the pointer and may send the pointer to private/group databases, distribution lists, email, or information digests [11].

The tool has been successfully introduced at Lotus Corporation in 1994 and is still used by a few people [1]. In the context of this paper, the experiences made with the introduction of the tool at Lotus Corporation are the most interesting thing since these experiences are quite different from ours (see below).

3 A Java-Based Approach to Active Collaborative Filtering

Encouraged by the success of the Notes filtering application, we have developed the Collaborative Recommender Agent CORA [8] which is an asynchronous distributed system for active collaborative filtering of World Wide Web documents [9]. Similar to the Maltz and Ehrlich dissemination tool for pointers to Lotus Notes documents, CORA enables users to recommend URLs (i.e., the locations of webpages) by issuing a single mouse click. However, CORA is not based on the proprietary Lotus Notes application but on the widely available Java/Javascript framework. CORA users are provided with a personal recommendation agent that monitors the user's web browsing behavior. The agent recognizes the webpage the user is actually browsing and displays the URL in a special recommendation window (see figure 1). If the user decides to recommend the URL to some user-defined group of recipients, the URL can be sent off by clicking the corresponding icon. Besides supporting the sending of URLs, the agent also supports the users in reviewing URLs he or she received from others. Recommended URLs are listed in the same window that is used for sending recommendations (see figure 1). A single mouse click on a recommended URL opens an additional browser window and displays the corresponding webpage.

In addition, CORA's interface provides an indicator for "use wear". The idea of "computational wear" has been investigated by Hill et al. [7] in order to visualize the interaction with computational objects, e.g., the usage of documents, with the help of attribute-mapped scroll-bars. These scroll-bars can be used to visualize "edit wear" data, such as how often which sections of a document have been modified or which author wrote which part of a document. Similarly, "read wear" data indicates which parts of a document have been mostly read. As with CORA, "use wear"

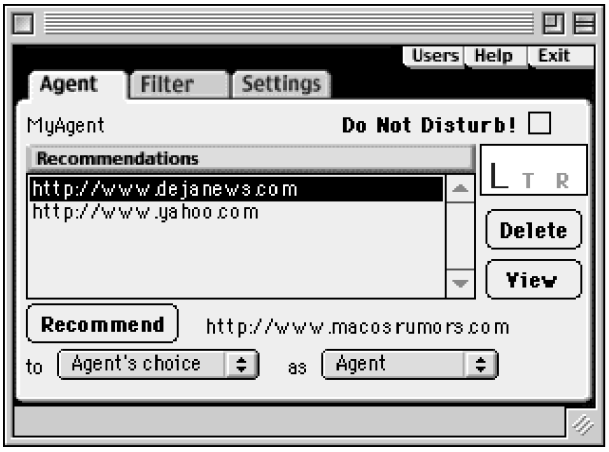

Figure 1: WWW interface of the active collaborative filtering tool CORA (explanations in the text)

is used to visualize if and -to a certain extent- how a recommended webpage has been used by the recipients of the recommendation. The size of the letters `L`, `T`, and `R` symbolizes the amount of links followed from the recommended webpage as starting-point, the time spent viewing the page, and a special recommendation factor (see figure 1) which is computed on the basis of the past usage of recommendations.

The development of CORA was based on the assumption of three basic requirements. First, a successful active collaborative filtering system (i.e., a system that is actually *used*) has to be easy-to-use. A long history of unused and therefore useless tools indicates that users are perhaps willing to inspect a new product but they are usually not willing to spend a lot of time to find out about how the product works. Moreover, experiences with collaborative filtering suggest that the effort required to recommend information should be as low as possible. CORA enables users of the WWW to send and view recommendations by issuing single mouse clicks.

In addition, sending a recommendation should not require the user to mentally switch between various application contexts as it is necessary if the user utilizes an email tool to send a reference to a web document. With CORA, all sending and receiving activities occur within the same browser context.

Last but not least, using CORA should not require any exotic hardware or software. This requirement was motivated by the demands of realistic testbeds that are commonly heterogeneous computing environments in these days. CORA has been implemented as a platform-independent Java/Javascript framework and can be used with most current browsers.

However, besides imposing stringent technical requirements, we assumed that we were able to raise interest in (active) collaborative filtering. Moreover, we assumed that providing a comfortable tool to easily disseminate webpage recommendations would encourage users to share URLs with others. A few enthusiastic users could be sufficient to bootstrap a self-supporting information dissemination process.

4 An Experiment in Active Collaborative Filtering

In order to verify our assumptions, we tried to establish the usage of CORA at the local computer science department. In terms of human and computational resources, the department is a well-suited testbed. It comprises approximately 100 researchers with scientific backgrounds reaching from biology and psychology to computer science and engineering. The department's computational equipment consisting of Sun workstations, PCs, and Macintoshs, is a typical example of a heterogeneous computing environment.

In particular, we assumed that the WWW is frequently used at the department so that participating in active collaborative filtering would be beneficial for the members of the department. In addition, we assumed that we were able to bootstrap a self-supporting information dissemination process at the department.

CORA has been announced on a mailing-list comprising all research assistants and several alumni of the department. The invitation email included a short introduction of the person responsible for the email, a brief summary of active collaborative filtering, and a reference to CORA's homepage which is located on the department's webserver. The aim of the email was both to announce CORA's homepage and to raise interest in (active) collaborative filtering.

Several weeks after CORA's introduction, the logfiles of the webserver showed a considerable hit rate on CORA's homepage. The homepage had been accessed from 55 different machines; 32 accesses originated from within the department, 4 from further places within the university, and some additional from outside the university. The total number of slightly more that 30 accesses might be interpreted as if we reached approximately 30% of the recipients and as if we raised considerable interest in CORA.

However, despite generating a lot of hits on CORA's webpage, the introduction of CORA did not raise particular interest in terms of usage or feedback. We clearly failed to bootstrap an active collaborative filtering process.

5 Investigations

In order to shed light on the course of the experiment we conducted an informal survey on the basis of the same mailing list as being used for the CORA announcement. We achieved a response rate of almost

50% (49 responses of approximately 100 recipients) which is considerably higher than the hit rate of 30% on the CORA homepage. The difference might be influenced to the fact that the introductory email has been sent by a student writing his diploma thesis at the department while the questionnaire has been sent by one of the department's research assistants.

The experiment was based on the assumption that the WWW is being heavily used at the department and that most browsers are technically sufficient for using CORA, i.e., the browsers are Java/Javascript-enabled. The results of the survey show that the assumptions concerning the WWW usage were correct since most participants use the WWW every day (see table 1).

Frequency	Count	Percent
Daily	44	90%
Weekly	4	8%
Monthly	0	0%
Less	1	2%

Table 1: WWW usage at the department

Also, the assumption about the browsers used at the department was reasonable. Actual browser versions that support Java/Javascript are widely used at the department (see table 2). Indeed, only 8% (4 out of 49) of the participants claimed that technical reasons kept them from inspecting CORA although the announcement even mentioned the 4.x version of the Communicator as a requirement for using CORA.

Browser	Count	Percent
Netscape Navigator	1	2%
Netscape Navigator 3.x	9	18%
Netscape Communicator	1	2%
Netscape Communicator 3.x	1	2%
Microsoft Explorer	3	6%
Microsoft Explorer 3.x	4	8%
Lynx	5	10%
Netscape Navigator 4.x	11	22%
Netscape Communicator 4.x	22	45%
Microsoft Explorer 4.x	7	14%

Table 2: Browsers used at the department (several participants named more than one product). Browsers in the lower part are suitable for using CORA.

In addition, the survey revealed that there is already a flow of recommendations at the department and that so-called "information sources" [11] exist (see table 3). All of these frequently (i.e., daily) recommendations sending persons claimed to use up-to-date browser versions that are suitable for CORA.

Recommendation rate	Count	Percent
Daily	8	16%
Weekly	12	25%
Monthly	19	39%
Less or none	10	20%

Table 3: Existing flow of recommendations

Put in a nutshell, the technical requirement of a Java/Javascript enabled browser seems to be an acceptable technical requirement although it actually prevents some persons from using CORA. Also, the asymmetric "information consumption" situation at the department is somehow similar to the situation observed at Lotus Corporation (as reported in [11]).

Unfortunately, the results of the experiment do not provide much information on the extent to which CORA itself has met the expectations of its potential users in technical terms. In particular, we could not get insights into the benefit of providing users with "use wear" in an active collaborative filtering context.

Interest	Count	Percent
Announcement missed	15	31%
Interested but lack of time	23	47%
Not interested in collab. filt.	9	18%
Not interested in agents	3	6%
Purpose and benefit unclear	3	6%
Additional information	25	51%
No additional information	20	41%

Table 4: Expressions of interest

Concerning the experiment itself, we assumed that we could raise enough interest in active collaborative filtering to bootstrap a limited, self-supporting filtering process on the basis of CORA. Furthermore, we assumed that announcing the tool on the internal mailinglist would be sufficient. The survey indicates, to the contrary, that the email announcement failed to get the attention of a large part of the participants (see table 4). Also, the overall interest in "collaborative filtering" was moderate. The interest is equally distributed among the participants. Asked whether they would like further information on collaborative filtering, 51% of the participants answered in the positive. However, 18% of the participants stated that they were actually not interested in collaborative filtering at all. Also, 41% answered in the negative to the question whether they were interested in further information on collaborative filtering. These results should also be

interpreted carefully since the questionnaire has been sent by a member of the department and the answers have not been anonymized.

The results of both the experiment and the informal survey conducted after the experiment indicate that important reasons for the lack of acceptance of CORA can be found in the startup phase of the experiment. It seems as if it is not so much the technology but the idea that did not make its way. This assumption is supported by the statement of some participants that they did not associate sharing URLs with collaborative filtering and that they did not see any good reason for sending and commenting on URLs. Others did not see the benefit of using the tool. This clearly suggests that further active collaborative filtering activities require a detailed planning of announcements and experiments. Besides using different communication channels, such as email announcements and announcements written on paper, a combination of the announcements with introductory talks and/or demo sessions should be taken into consideration. A related issue is that we probably failed to convey the active collaborative filtering idea. Thus, the potential benefit of participating in active collaborative filtering has to be outlined in more detail.

6 Lessons Learned: Collaborative Filtering and Groupware

When investigating our problems with establishing (active) collaborative filtering, we detected striking similarities with problems that are well-known in the context of groupware [4]:

- Lack of upper management support
- Need for user acceptance
- Benefit depends on usage
- "Cold-start" problem

Typical groupware applications in this context are project management applications, meeting support systems, co-authoring applications, workflow systems, and group calendars [5].

First, collaborative filtering and groupware are typically not an issue of particular concern for upper management. It is therefore unlikely that upper management will become involved in promoting a collaborative filtering application. This is quite different to the introduction of large organizational information systems where the use may be promoted and supported by the upper management [5].

The lack of upper management support implies that the success of a collaborative filtering application or groupware application depends on the acceptance of the application. The use of an application has to emerge bottom-up. Examples from the literature, e.g., the active filtering tool [11] or electronic calendars [6], show that this indeed happens. The Knowledge Pump [3] approach, however, includes an additional market system to provide more benefit for those supplying information.

Then, the benefit of collaborative filtering and groupware becomes effective *only* if a corresponding application is used by several persons. In addition, the potential benefit is probably not obvious at first sight. This is an issue of introducing the application in the right way to raise interest.

Finally, collaborative filtering is similar with groupware in that both have to overcome the so-called cold-start problem. As long as an application is not used by several people, the *benefit* does not pay. Experiences at Lotus Corporation [11] have shown that most active collaborative filtering participants will passively wait for information provided by others (or "the system") rather than actively distributing information. This means that some volunteers have to provide information without directly gaining benefit. Experiences with traditional collaborative filtering systems, such as [12], show that the cold-start problem is a major obstacle on the way to a working system. However, contrary to providing ratings for documents, many people already exhibit an information dissemination behavior that forms the basis of active collaborative filtering.

However, we also detected striking differences between active collaborative filtering and groupware:

- Target group
- Amount of active support needed

The target group of groupware is usually defined by the structure of an organization and the tasks allocated to its members. To the contrary, the target group of an interest-based information sharing tool are by definition those who are willing to share interests. From an organizational perspective, this means that such an application is well-suited to cross the traditional borders of workgroups and to enhance the flow of information among groups that are not directly interacting. However, experiences with Intranets show that it is not only the availability of technology that establishes the sharing of information. Informal workplace communication is an important prerequisite in order to know about the particular interests of others.

Also, active collaborative filtering has shown to provide benefit to its users even if only a few users participate actively while most users consume information passively [11]. In this respect, it resembles the global conferencing system Usenet News with its significant asymmetry between a few active participants (those

who actually post articles) and many passive participants (those who read articles). The degree of active support needed to provide benefit also distinguishes active collaborative filtering from traditional rating-based collaborative filtering. The large amount of ratings needed by the latter turned out to be a serious obstacle on the way to working filtering system that actually provides benefit to its users [12]. Groupware, such as meeting scheduling or calendar management, to the contrary, depends on being actively used by most members an organization.

7 Conclusions

In this paper, we have discussed an active collaborative filtering experiment and have shed light on related issues, such as communication problems, motivational problems, and the benefit question, i.e., who does the work and who gets the benefit. Our experiences show striking similarities between the problem to bootstrap active collaborative filtering and to introduce groupware. Viewing active collaborative filtering as a kind of groupware application enables us to profit from the large body of literature on the experiences with successes and failures of groupware. This is especially interesting since reports on experiences with collaborative filtering systems tend to focus on the benefit question only without considering the related social and organizational issues.

Acknowledgements

The author would like to thank Rolf Pfeifer for a stimulating research environment, and Jie Mei, Reinhard Riedl, and Ralf Salomon for their help. Special thanks go to Christoph Landolt for the Java/Javascript implementation of the collaborative filtering agent CORA.

References

[1] Kate Ehrlich. Personal communication, 1998.

[2] Kate Ehrlich and Debra Cash. Turning information into knowledge: Information finding as collaborative activity. In John L. Schnase, John J. Leggett, Richard K. Furuta, and Ted Metcalfe, editors, *Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries*, June 1994.

[3] Natalie S. Glance, Damián Arregui, and Manfred Dardenne. Knowledge pump: Supporting the flow and use of knowledge. In U. Borghoff and R. Pareschi, editors, *Information technology for management*. Springer-Verlag, 1998.

[4] Jonathan Grudin. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the International Conference on Computer Supported Collaborative Work (CSCW'88)*, pages 85–93, September 1988.

[5] Jonathan Grudin. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1):92–105, 1994.

[6] Jonathan Grudin and Leysia Palen. Why groupware succeeds: Discretion or mandate? In *Proceedings of the Fourth European Conference on Computer Supported Collaborative Work (ECSCW'95)*, pages 263–278, September 1995.

[7] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. Edit wear and read wear. In *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'92)*, pages 3–8. ACM Press, May 1992.

[8] Christoph Landolt. Autonome Agenten als Werkzeug für aktives gemeinschaftliches Filtern von Dokumenten. Diploma thesis (in German) IFI U 2227, Department of Computer Science, University of Zurich, Switzerland, February 1998.

[9] Christopher Lueg and Christoph Landolt. A java-based approach to active collaborative filtering. In Clare-Marie Karat, Arnold Lund, Joëlle Coutaz, and John Karat, editors, *Summary of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*, pages 319–320. ACM Press, April 1998.

[10] Christopher Lueg and Rolf Pfeifer. Cognition, situatedness, and situated design. In Jonathan P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska, editors, *Proceedings of the Second International Conference on Cognitive Technology*, pages 124–135. IEEE Computer Society, August 1997.

[11] David Maltz and Kate Ehrlich. Pointing the way: Active collaborative filtering. In Irvin R. Katz, Robert Mack, Linn Marks, Mary Beth Rosson, and Jakob Nielsen, editors, *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, pages 202–209. ACM Press, May 1995.

[12] Bradley N. Miller, John T. Riedl, and Joseph A. Konstan. Experiments with GroupLens: Making Usenet useful again. In *Proceedings of the 1997 USENIX Winter Technical Conference*, January 1997.