# Explaining and Combining Recommender Algorithms for Decentralized Architectures (ENCORE)

**Martin Svensson**
Swedish Institute of Computer Science
Box 1263, SE-164 29, Kista, SWEDEN
martins@sics.se

## 1 Project Summary

Collaborative filtering is a new method for retrieving information. Instead of basing the search on the content in the information units, a collaborative filter or recommender system uses databases of user ratings as a basis for predictions. From our user studies we see that collaborative filtering can be of great help to users when they search for information. In order to be truly useful collaborative filtering needs more research. Specifically, we have seen that it can be improved by exploring new ways of *explaining* to users why they get a particular recommendation. Secondly, for some application scenarios, collaborative filtering has to work in networks where there is no central server – thus requiring a *decentralized recommender* model. Finally, it is necessary to find ways of *combining* traditional content-based methods for information filtering with collaborative information to improve the filtering process and solve some of the problems with collaborative filtering.

ENCORE aims to tackle these three problem areas and package the resulting solutions in a general-purpose recommender platform. The platform will be applied in two very different applications to investigate its feasibility and integration qualities in different domains and for different tasks. Together with IFS AB (Industrial and Financial Systems AB) we will integrate the platform in their main product IFS Applications, which is a professional system for customer management and resource planning. The second application is a research prototype for position-based mobile information filtering. In both cases, the applications will be evaluated with end users.

## 2 Project Objectives and Relevance

The goal for a collaborative filter is to predict a user's preference regarding some set of titles: books, news articles, or perhaps radio channels. To do this, a collaborative filter use ratings from a database of users (either from all users or a subset of the users in the database). The filter can then be used in either of two ways: users can explicitly ask what the system believes they would think about a certain title, or users can ask the system to return a ranked list of titles the system thinks they like. Since collaborative filters try to suggest titles that a user would like, they are often called *recommender systems*.

Collaborative filtering is especially useful to filter out information that is not text, since it is difficult to do any content-based analysis on that type of data. For this reason, many e-commerce companies have adopted variants of collaborative filtering as a way of guiding users to the product that fits their need. The idea is that in the same way you get recommendations from your friends, you could get recommendations from a set of anonymous users also interested in, for example, the same books as you are. Companies such as amazon.com and CDNow use collaborative filtering successfully.

To date, collaborative filtering has not been used for finding more expensive items such as holiday travels [10]. One explanation for this is that a user can risk going to a movie, but is more reluctant to choose a holiday destination solely based on a recommendation from a set of anonymous peers. When collaborative filtering becomes more elaborate and integrated with other techniques for finding information we believe that it can be used to find other types of data. There are a number of issues that needs to be resolved before recommender systems becomes truly powerful tools for navigating and filtering out information, to name three:

**Decentralization**: Traditional recommender systems depend on a central server that keeps track of and matches user profiles. In networks where there are no such servers how can "matching" of similar users still be carried out?

**Extended filtering methods**: Collaborative filtering relies on the ability of matching user profiles and will work poorly when little information has been collected, e.g. the bootstrapping problem [22].

**Explanations**: A common problem found in recommender systems is how to convey a recommendation to a user. If a user understands how or why she gets a particular recommendation she can better judge the quality of the recommendation [23].

ENCORE will (1) investigate scenarios where there is no central repository of user profiles and how that will effect both content-based and collaborative filtering, (2) deliver methods for how to combine content-based and collaborative filtering in order to get more powerful filtering thus alleviating, for example, the bootstrapping problem, (3) tackle the problem of explaining recommendations to users, (4) make available a public platform for information filtering, hereafter named SIRE. The platform will encapsulate the novel techniques developed within the project. Finally, ENCORE moves away from the traditional domain for recommender systems and together with IFS AB integrates recommender functionality into corporate systems using the SIRE platform.

## 2.1 State of the Art

Collaborative filtering has its foundation in the broader research field of social navigation [24]. The model for a collaborative filter is simple: collect and store user profiles at one central location and when a recommendation for a user is needed match other users profiles with that user's profile and recommend items that the users has not yet seen. In Amazon.com, for example, collaborative filtering is used for recommending all sorts of products. At a technical level collaborative filtering could look slightly different:

**Memory-based Prediction**. Memory-based prediction uses a nearest-neighbour approach to find a subset of all users that have the most similar preference history as the active user [11]. The advantage of the memory-based scheme over other methods is that its structure is dynamic and immediately reacts to changes in the user database. The downside is that user profiles are matched against each other every time a prediction is needed, a process that can be slow and require large amounts of memory.

**Model-based Prediction**. In model-based collaborative filtering, the user database is run through an algorithm that builds a model of that data in terms of the probability that a user would like a given title depending on what she has done in the past. In terms of predictive accuracy the model-based schemas perform as well as memory-based prediction [2]. The model-based prediction does not suffer from the performance and memory bottlenecks found in memory-based prediction. However, once the model has been built it is difficult to update it without rebuilding it.

**Inverted file prediction.** To overcome the shortcomings with the memory-based model and still have a scalable solution as in model-based filters we have suggested an approach that is built around an inverted file structure [4] in the same way as most text retrieval systems are built. Data is inverted so instead of having user profiles containing items, each item contains all users that have voted on that particular item. From experiments we see that this structure is superior to the classical memory-based approach in terms of speed. Furthermore, initial findings indicate that not all votes are equally important. By removing the ratings for certain highly frequent items it is possible to improve the overall precision of the system.

**Previous work by project participants**

At the Social Computing group we have looked at several ways of doing information navigation, from pure social navigation (e.g. allowing people to directly communicate with each other or visualizing their traces in space [24]) to advanced techniques for content-based information extraction based on machine learning. As a result of this research, three platforms have emerged: the Social Navigator [24], an initial prototype of the SIRE platform [5] and a text retrieval system. SIRE has been used in three different systems: *Kalas*, *RIND* and *GeoNotes*. Kalas [12] is a social navigation system for finding recipes. It uses collaborative filtering for recommending recipes to users. RIND [6] uses recommendations to guide users in the process of configurating a PC. In systems such as RIND, which essentially uses content and collaborative filtering as a way to guide people, it is clear that the methods should be combined and not kept apart. GeoNotes [7] is a mobile system for populating the "real world" with virtual post-it notes, where users can filter or view only those notes that are selected from a collaborative filter.

IFS has explored the possibilities with various collaborative filtering techniques, for expert localization, as well as searching and filtering of information both in intranets and the Internet. Particularly, a working prototype of a future IFS Applications featuring expert locating has been shown at a customer conference.

From our work at the forefront of mobile, ubiquitous and social computing systems, we see that new technologies and usage patterns put new demands on collaborative filtering systems. It is crucial to investigate how and what can be done with collaborative filtering in mobile systems. There are certainly opportunities; the fact that users have a physical context ought to be information that can be used by the collaborative filter. It is also true that when users are moving through space they are not interested in one type of item. Depending on their physical location they want different sets of recommendations.

## 2.2 Scientific and technical objectives

The prime target for the ENCORE project is to create a general-purpose recommender system that works well in a variety of domains and settings – from centralized anonymous systems to decentralized systems. We want to create collaborative and content-based methods for filtering, access, and visualization of information that is driven by the user's needs.

ENCORE has the following scientific objectives:

- o Increase trust in recommender systems through developing new visualization and explanation models.
- o Use recommender systems to recommend people rather than items, turning the systems into expert locators.
- o Explore the possibilities of using recommender systems to create and sustain communities.
- o To investigate how a user's physical context can be used as input to a filtering system
- o Develop algorithms for combined content-based and collaborative filtering.

ENCORE has the following technical objectives:

- o Create a filtering platform that uses a decentralized architecture.
- o Create a filtering platform that combines collaborative and content based filtering.
- o Integrate the filtering platform into an existing large-scale information management system.

- Create three scenarios in which the research objectives are exemplified and implement these as prototypes.

## 2.3 Industrial relevance

There is an increasing demand on systems that filter and help people find the information they need. Collaborative filtering has been proposed as one method in this process. Many on-line stores use collaborative filtering to recommend products to people. There is also a need for social systems that in one way or the other let people get in contact with each other. Our partners operate in different domains (i.e. IFS and Volvo IT) but both see a potential gain in collaborative filtering.

The sheer amount of information available in today's business systems is huge, and the trend is that it increases at an even higher rate by each year. But still, most of the information needed by users is only in peoples' heads. To be able to both cope with a large information space and to utilize the vast source of knowledge that the users of our system represent there is a need for new approaches to navigation and information retrieval. Recommender systems promise a lot in both these areas as they can be used for both expertise location and collaborative information filtering. This has been recognized not only by IFS but also by their competitors (e.g. SAP with their NetWeaver initiative) and by industry analysts, e.g. [15].

Previous research in organisational settings, e.g. at Volvo, show that great benefits can be obtained from utilising social navigation in industry [18, 19]. The result from the Volvo studies suggest that in job-based organizations, employees cannot be encouraged to contribute to the success of external projects without returning tangible benefits also to their own organizational unit. We have to recognize that work increasingly is performed not by isolated workers but by cross-organisational project groups and that sustainable knowledge creation and business innovation depends on mixing input from a variety of competencies. The ambition to increase the level of trust and to facilitate the establishing of communities that ENCORE is aiming for is thus highly relevant to the industry.

The work at Volvo also showed that although being in an organisational context where content and activities can be expected to be work related, personal integrity and privacy are nonetheless important aspects to consider when leveraging interests. More research is therefore needed to understand the trade-off between giving up personal details and gaining tangible benefits. Industrial settings, as opposed to the web in general, are however better suited for solutions that build on personal behaviour and individually expressed preferences and the ENCORE project is likely to do well in collaboration with industrial partners.

## 2.4 Relevance for program

The ENCORE project is very much in line with current and future research. We see an emerging demand for collaborative filtering systems that are decentralized and work in ad-hoc and peer-to-peer networks. By combining social and content-based information we can create hybrid recommender systems that better support users in finding the information they need. In the near future we see a need for recommending people (instead of information), for building trust in recommendations and creating communities.

ENCORE fits well into the framework of VINNOVA's *NETPROG* as it (1) develops a filtering platform that will work in peer-to-peer networks and (2) is based on collaborative filtering techniques which allows a system adapt to individual users' need and preferences.

ENCORE is a joint project between SICS and IFS AB. The aim is to clearly show the potential benefit with collaborative filtering for the Swedish industry. We will, within the project, actively develop scenarios that are consistent with how IFS customers work on an everyday basis and use the SIRE platform to strengthen IFS competitiveness. At a research level the collaboration will force us to take into account how research can be transformed into usable technology for the industry.

# 3 Work packages

ENCORE is planned to start July 1, 2003 and run for three years. We are, however, already working with some of the issues in ENCORE since October 2002 with funding from SSF. This has allowed us to, for example, produce one licentiate thesis [5]. The funding also enables us to early on in the project have a working prototype of the SIRE platform. Furthermore, IFS will incorporate some aspects of the SIRE platform and develop the prototypes discussed in WP 1 before the project starts.

ENCORE has a total of 48 person months divided over 7 work packages.

## 3.1 WP 1: Scenario development and prototyping

**Coordinator:** IFS together with SICS

**Person months:** IFS 5, SICS 3

In ENCORE the research objectives and technical objectives will be exemplified in different scenarios that serve to illustrate how information filtering could be used in different contexts. Two scenarios are directly applicable to real-world working situations. To give us freedom to fully explore the research objectives in ENCORE a third scenario will be developed that has few technical limitations.

**Scenario 1:** Adam B. is a salesman at a global corporation selling complex products. He has got a lead indicating that a company is looking for a solution his own company can satisfy. By closer specifying the potential business in the customer relations management (CRM) system he finds out that they've had five previous contracts involving similar companies, where three have been profitable. He gets recommendations on how to approach the company and he starts to work out his strategy in the system. He also finds out that two other salespersons, which were involved in the previous contracts, are available for discussion. Together they agree on a strategy for Adam to approach them.

**Scenario 2:** Eric F is a field technician. While replacing a broken component on a machine Eric notices that another component shows some wear. He suspects that this will eventually lead to that the component breaks down, and needs to be replaced. He writes down a service request, which is immediately transferred to an issue tracking system. A short while later he receives an SMS telling him to call Greg H to discuss the need to replace it immediately.

**Scenario 3:** A person is visiting Dublin for the first time. Being in the hotel lobby, she is going through the maps and information on places to visit available at the front desk. Not knowing which to choose, she picks up her mobile phone and asks for recommendations on interesting places to visit. The phone connects to other users' phones in the lobby. The phone connects to those users and, by comparing public profiles, finds a number of users similar to her. From these users, a number of places are recommended. Being both surprised and happy, she walks out the lobby and sets her direction to the King's Theatre, to see a play written by one of her favourite playwrights.

**Approach**

The three scenarios will be described in more detail, specifically in terms of how they make use of the functionality provided by the SIRE platform. The scenarios will be implemented as concept demonstrators.

Either scenario 1 or 2 will be chosen for an in-depth field study. The procedure will be to study how a salesperson or field technician find and make use of information in their daily activities. The method we used in [9] will be applied. One or possibly two of IFS customers is chosen for the study.

**Deliverables**

D 1.1 Description and demonstrator of scenario 1 and 2

D 1.2 Description and demonstrator of scenario 3

## 3.2 WP 2: Decentralized filtering

**Coordinator:** SICS

**Person months:** SICS 9, IFS 2

Much research has been devoted to solutions for centralized recommender systems. What has been missing is a decentralized approach since not all domains are suitable for centralized recommender services:

- Domains where all computer power resides on the client such as ad-hoc networks where nodes constantly enter and leave the network [25].
- Domains where the users' privacy is of great importance. Collecting all user preferences at a single place might yield the data to unauthorized intruders or lead to unauthorized access and use by the service provider. In a decentralised network the user preferences could be spread among many peers, not all at one place, and the peers might co-operate to protect each other's privacy [8].
- Domains where the service availability is critical or where the users do not want to rely on a single service provider. In centralized recommender systems users are entirely dependent on the service provider for the service. If well designed, a decentralized recommender can continue to exist albeit the original provider vanishes.

There are two properties that affect the architecture of a decentralized recommender system: where information is stored and where recommendations are computed. Figure 1 depicts 9 possible setups between storage and computation. Each setup will affect what type of recommender system that is feasible (i.e., peer-to-peer, ad-hoc, decentralized, and centralized). Starting from a division between storage and computation we want to investigate how a decentralized recommender system might work and for what type of application scenario.

In decentralized recommender systems clients could come and go, meaning that not all users are accessible at any given point in time, which will also affect the accuracy of the system. In ad-hoc networks, where only a small number of users are connected, this can potentially lead to low quality recommendations. On the other hand, a person's context can outweigh the smaller set of users. It is often the case that people that are at the same location share a common interest (cf. being in the hotel lobby). In decentralized networks that are not based on physical location the same properties may apply. That is, there are possibilities in treating clients as "near" to each other. For example, in a peer-to-peer network where one client often accesses another client this could be an implicit indicator that they share some common interest, and thus, should be grouped together.

There are certainly technical problems that need to be solved with decentralized recommender systems. In centralized models, it is common to include both local and global properties of the information collection for placing a priori weights on items such as documents and users [2, 27]. This combination has a great impact for the accuracy of the system. In a decentralized setting the global properties are more difficult to extract, or even impossible. New ways of weighting information, perhaps based on sub-global properties and context parameters, is necessary to cover this gap. In terms of speed, decentralized collaborative filtering will not match a server-based approach, given that neighbourhood selection is based on exhaustive search. However, we have already shown that some pruning strategies can be applied to collaborative filtering [4] to reduce the search space.

**Approach**

There are four key issues that we want to investigate for decentralized recommender systems: application scenarios, accuracy, speed, and recall. Scenario building will be based on literature surveys and analysis of different network configurations (see Figure 1). To investigate the accuracy (or predictive power) of the decentralized model we will compare the recommendations with a traditional client-server solution. In terms of speed it is more complicated. We anticipate that it is necessary to simulate peer-to-peer networks and evaluate the recommender system by simulation. Several different networks are needed to cover most possibilities of how real networks could look like. Another possibility is to make a real world study of a large peer-to-peer network, something that is investigated in the MOSES project (see 4.1)

**Deliverables**

D 2.1 Theoretical paper on decentralized collaborative filtering

D 2.2 API specification

D 2.3 Final code

| | 1 client | | 2+ clients | | Server | |
|---|---|---|---|---|---|---|
| 1 client | N/A | N/A | Ad-hoc | P2P | Static | Client/Server |
| | Centralized | No Server | Decentralized | No Server | Decentralized | Server |
| 2+ clients | Ad-hoc | P2P | Ad-hoc | P2P | Static | P2P+C/S |
| | Decentralized | No Server | D+C | No Server | Decentralized | Server |
| Server | Static | Client/Server | Static | P2P+C/S | Static | Client/Server |
| | Decentralized | Server | Decentralized | Server | Centralized | Server |

Figure 1. Properties of recommender system architecture when splitting the storage of information (rows) from the computation of a recommendation (columns). Each cell defines a particular type of system.

## 3.3 WP 3: Combinations of information filtering methods

**Coordinator:** SICS

**Person months:** SICS 5, IFS 2

In collaborative filtering, there is no distinction between items that have the same set of votes or opinions. A tentative solution to this problem is to categorize items on the basis of their textual content. In text filtering, on the other hand, items may have similar content but it is hard to automatically separate items of low and high quality. Collaborative information such as votes, opinions and usage patterns is a possible starting point for categorizing items on the basis of their qualitative properties. To tackle these and related problems, we will investigate how collaborative and content-based algorithms and representations may be combined to improve filtering performance.

Different types of combinations have already been proposed in the research literature [3]: meta-learning methods, weighted combinations etc. The focus has been on how to combine existing algorithms rather than selecting well-founded algorithms to work on more sophisticated representations. Our approach is to use and extend state-of-the art representations of text content and collaborative information, and use robust machine learning algorithms.

**Representation**

Textual content is traditionally represented by word frequency vectors. There are several well-known problems with this simplistic representation. For example, it is not clear how to capture and treat synonyms and polysemic words.

State-of-the art techniques capture word usage patterns by utilizing vector space models that reduce the dimensionality of the problem space. One such example is Latent Semantic Analysis (LSA) [16], which transforms the representations of terms and documents matrix into vectors in a semantic space. The technique is based on an orthogonal transformations of a large sparse matrix, and is in many practical cases not feasible due to its high computational complexity. Recently, this matrix decomposition technique has been successfully applied to collaborative filtering [21].

Another, more recent, method is Random Indexing [13, 14], which relaxes the constraints on the transformation; instead of requiring orthogonal transformations it is based on near-orthogonal vectors, reducing the computational complexity considerably. The technique is very flexible with regard to managing new data. In contrast to methods such as LSA, Random Indexing can easily be applied to dynamical data sets where the information continuously changes and grows.

**Machine learning filtering algorithms**

As all information filtering algorithms stem from the broader field of machine learning, we will use and extend several known machine learning methods. Extending an algorithm in this sense is to incorporate background knowledge of the learning problem at hand, thus making the algorithm better informed, more accurate and hopefully more predictable. Background knowledge in our case will be information about word and concept relationships, and knowledge of collaborative usage patterns.

For the machine learning algorithms, we will focus on Kernel-based methods such as Support Vector Machines [26], local learning [1], and Bayesian techniques [17]. These methods are robust, fairly scalable and can incorporate background knowledge.

**Approach**

Research within collaborative filtering is starting to get some de facto standards and methods for evaluating recommender systems [2, 11]. They consist of standard metrics (e.g. ROC curves and Rank Score) and datasets (e.g. EachMovie and MSWeb). By using several existing data sets, we can compare combinations of content-based and collaborative methods with existing approaches to show how our approach will deliver better predictive accuracy and novelty detection.

**Deliverables**

D 3.1 API specification

D 3.2 Evaluation (Reported as paper)

D 3.3 Final code

D 3.4 Ph.D Thesis chapter (Rickard Cöster)

## 3.4 WP 4: Explaining and visualizing recommendations

**Coordinator:** SICS

**Person months:** SICS 7, IFS 2

Removing the author from a movie review will make that review much harder to judge. A recommender system is dependent on the ability to mediate recommendations to users in terms of their quality, relevance, and who affected it. As this is difficult it is often overlooked and the only information presented to the user is typically: "*Rate these items and you will then receive recommendations*". In a study by Herlocker [10], various explanation models were tried, and they found that it was extremely difficult to explain a recommendation; often the models were too complex to understand. The model that yielded the best result was the simplest: "*In XX per cent of the cases the system has been right*". In other words, they explicitly told the user how many times the recommender would rate her movies as she would rate them.

It could be argued that simple explanations in the form above is enough because people get some understanding of how well the system performs (and therefore can judge a recommendation). We argue that this is not the case. When a user does not get a more complete picture of why she gets a particular set of recommendations, she will be careful in acting upon them.

As the Volvo study [18] showed, employees asked for more personal details about their fellow employees, in order to initiate communication and collaboration. A deeper level of personal information would increase the sense of familiarity and thereby make it easier for organisational members to contact each other for information exchange, competence sharing, and building of communities. Another important aspect would be to make available more information about the signalling artefacts themselves. By making details such as the updating or visiting frequency available, the users would be able to derive the owners' level of engagement. This sort of information can hardly be expected to be shared freely on the Internet, but in an organisational context such as the intranet these data may already be available in form of homepages or employee records. So what other implications are there of presenting more information about the people who affected a recommendation?

First, instead of visualizing user profiles and how they are matched, a user can see what items and ratings other users have voted on. This is somewhat analogous to building a trust relationship for a movie reviewer, for whom you know the past ratings.

Second, we could use the recommender system as a way of creating communities. By community we mean "a group of people that share a common interest, either implicitly or explicitly". When using a recommender system to recommend and create communities for the user we: (1) increase the probability of creating communities since setting up the community and finding the members is done automatically, and (2) help people find people and communities that they would otherwise not find.

**Approach**

Before any implementation is done it is necessary to find explanation models that maximize the benefit for users. What type of information from other users is needed for people to gain better knowledge about them? A number of different design demonstrators will be created and evaluated on small sized user groups. We will in this stage focus on two aspects of the problem: the number of users to display and what information about users to visualize. The new visualization techniques are to be incorporated into the prototypes developed in WP 1.

**Deliverables**

D 4.1 Different visualizations models (in conjunction with WP 1)

D 4.2 User study

D 4.3 Software components for integration in SIRE

## 3.5 WP 5: Platform implementation

**Coordinator:** SICS

**Person months:** SICS 5

WP 2, WP 3 and WP 4 provide core functionality. In WP 5 we combine and package this functionality in the SIRE platform. Our aim is to produce a freely available filtering system that is suitable both for industry and academia and to make available public test suites that can be used to evaluate recommender system algorithms.

**Approach**
WP 5 will start by making slight modifications to the existing SIRE package to fit the additional functionality that is needed to support the extensions suggested in WP 2 – 4. As soon as new functionality becomes available from WP 2 – 4, it will be incorporated into SIRE. To a large extent this WP has a control function to ensure that all pieces fit together.

**Deliverables**
D 5.1 First SIRE platform with incorporated API specification from WP 2 – WP 4

D 5.2 Final SIRE platform

D 5.3 Publicly available data sets for evaluating recommender systems

D 5.4 Ph.D thesis chapter (Rickard Cöster)

## 3.6 WP 6: Integration

**Coordinator:** SICS together with IFS

**Person months:** SICS 3, IFS 1

When developing a general-purpose information filtering platform it is vital to ensure that it does what it promises. It has to be functionally sound and not cumbersome to use for developers. In the present case this is certainly true: not only are we targeting the research community, but we want to create SIRE in such way that it can be used by industry as well.

**Approach**
The scenarios developed in WP 1 will result in three prototypes (two that will be incorporated into IFS Applications and one developed as a stand-alone prototype). Thus, by ensuring that SIRE can be used both with IFS Applications and the research prototype we can conclude that SIRE is usable both for industry and the research community.

**Deliverables**
D 6.1 Enhanced IFS Applications

D 7.2 Research prototype

## 3.7 WP 7: User Evaluation

**Coordinator:** SICS

**Person months:** SICS 4

One objective with moving from the traditional domains for recommender systems to a domain with a completely different set of requirements and opportunities is to see if findings can be generalized across domains. Thus, once we have a working prototype for a business scenario some user evaluation is in place. In ENCORE one of the three scenarios developed will be selected for a user evaluation. In the HUMLE laboratory we have already conducted user studies on recommender functionality in the Kalas system [12, 23].

**Approach**
In WP 1 we will select a scenario and evaluate how users collaborate and use tools to find information. In this work package, a similar evaluation will be conducted with the enhanced version of IFS Applications. This enables us to do a comparative study of the scenario.

**Deliverables**
D 7.1 Conference/journal paper on user evaluation on selected prototype

# 4 Project participants

ENCORE is a joint project between the Swedish Institute of Computer Science (SICS) and IFS AB (Industrial and Financial Systems AB).

**The Social Computing group at SICS** is an interdisciplinary group. Our research range from pure users studies to the implementation of recommender systems. The group has successfully evaluated and enhanced a number of systems with social navigation, EFOL [23], Kalas [12], and GeoNotes [7], to name a few. The following researchers will participate in ENCORE:

Martin Svensson, Ph.D (Project Leader), completed his doctoral thesis in 2003 on social navigation (see appendix 1 for a complete CV). Martin has managed the SITI funded Kalas project and is currently managing the Social Computing theme and the MOSES project at SICS.

Rickard Cöster, Ph.Lic, is a Ph.D student in the Machine Learning Group at the Dept. of Computer and Systems Sciences, Stockholm University. His research is focused on information retrieval and filtering, especially on how to build scalable systems for personalized information access. He obtained a Ph.Lic degree in 2003 [5]. Rickard managed the VINNOVA funded project RIND.

Tomas Olsson is a Ph.D student in Computer Science at the Dept. of Information Technology, Uppsala University, within the Computing Science Division. His research is focused on decentralizing and bootstrapping recommender systems [20]. He will present his Tech.Lic. June 5, 2003.

Magnus Sahlgren is a Ph.D student in Computational Linguistics at the Dept. of Linguistics, Stockholm University, within the Graduate School of Language Technology, GSLT. His research is focused on using co-occurrence statistics to acquire semantic information, which is represented using vector-space models. In particular, Magnus studies the use of "brain-like" stochastic computing methods for building the semantic space.

Preben Hansen, Ph.Lic, investigates situation- context- and task-specific factors in information seeking and retrieval. This approach includes the perspective that IR activities are embedded within a broader information-seeking phase. Preben is currently focusing on and investigating the effects of collaborative information seeking and retrieval in information-intensive environments.

**IFS** is a Linköping based developer and supplier of component-based business applications for medium and large enterprises. The main product IFS Applications, which is based on web and portal technology, offers 60+ enterprise application components used in manufacturing, supply chain management, customer relationship management, financials, engineering, maintenance and human resource administration. IFS provides customers step-by-step evolution to the extended enterprise with e-business solutions that offer partner, customer, and supplier collaboration. From IFS the following people will participate:

Fredrik Eklund, Ph.Lic, is Principal Research Officer, with human-computer interaction as main responsibility. Before starting at IFS he worked at Nokia Home Communications as project manager for large early development projects of digital media convergence products. Nokia had previously acquired the consultancy company User Interface Design where Fredrik worked as project manager and developer. He came there from the Department of Computer Science at Linköping University after completing a Licentiate of Philosophy in Computer Science.

Pär Hammarström is Chief Research Officer, his main responsibility is to ensure IFS' long term competitiveness. With more than 10 years of experience within industrial research and development of large and complex standardized systems his work covers most aspects of a company's core businesses. He joined IFS 1996 and has had a number of positions including project manager, product director and program director. Before joining IFS, Pär worked at Saab Technologies in various positions including software engineer, system analyst and project manager.

## 4.1 Industrial partners

**Volvo Information Technology AB** (contact: Dick Stenmark) supplies global industrial companies with IT systems and solutions. In addition to infrastructure, operations and applications, Volvo IT also provides consulting services, training and support. Their customers come both from within and from outside the Volvo Group.

Dick Stenmark, Ph.D, has a Doctoral degree in informatics and a B.Sc in computer science, both from Göteborg University. Dick is a member of the Knowledge Management Hub at the Dept. of Informatics at Göteborg University and the Viktoria Institute and has been involved in various research activities since 1997. Dick is currently a member of the Content management group where he is responsible for the intranet search initiative.

## 4.2 Academic partners

**Mobile Services** (contact: Martin Svensson) aims to develop innovative mobile services and the corresponding supporting technology and platforms. The project focuses on creating a sustainable framework for user centred research on mobile applications. This should be done by a number of activities that aggregate knowledge about the unique properties of IT-use in mobile settings, and its consequence for theory/design approaches and methodology from ongoing research.

## 5 Schematic timetable

In figure 2 each work package is plotted in terms of the duration in ENCORE. Below each work package the accompanying deliverables are plotted.

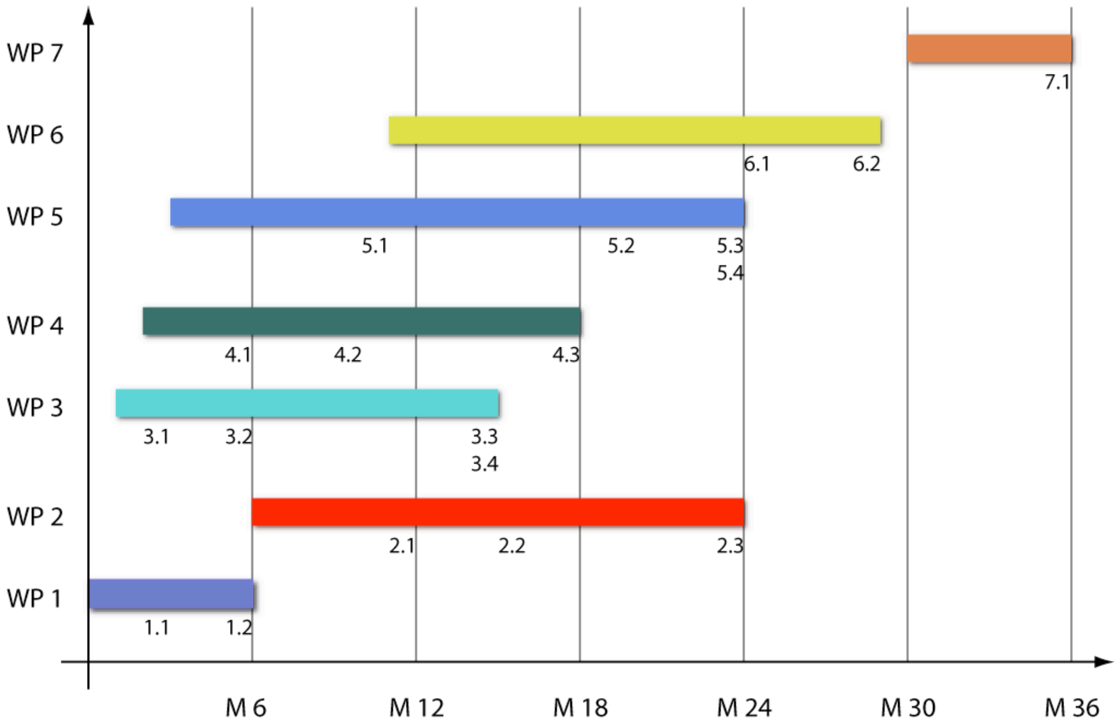


Figure 2. Time plan for work packages and deliverables

## 6 Budget

If ENCORE gets funded by VINNOVA we would like to start July 1, 2003 and end June 30, 2006. We have calculated for 1600 hours per year per person. The division between each project participant is as follows: Martin Svensson (40% at 720/h), Tomas Olsson (30% at 620 kr/h), Rickard Cöster (20% at 620 kr/h), and Preben Hansen (10% at 720/h). We intend to seek additional funds for Tomas Olsson. Rickard Cöster is also financed from Stockholm University and does not need any additional funding, Magnus Sahlgren is fully financed from the GSLT program, and Martin Svensson is partly financed from the MOSES project.

For IFS we have calculated for 1725 hours per year and person at a rate of 430kr/h. The division is as follows: 2003 5 months, 2004 4 months, and 2005 3 months.

| | 2003 | 2004 | 2005 | 2006 | SUM |
|---|---|---|---|---|---|
| VINNOVA | 536000 | 1072000 | 1072000 | 1072000 | 3752000 |
| IFS | 309063 | 247250 | 185437 | | 741750 |
| SUM | 845063 | 1319250 | 1257437 | 1072000 | 4493750 |

Table 1. Proposed budget

Explanation of Other Costs ("Övrigt") in the budget of the VINNOVA form is: These indirect costs cover SICS expenses for management, administration, mail, computer- and telecommunications, library, localities, and shared computer resources.

# 7 References


1. Bottou, L., and Vapnik, V. N. (1992) *Local learning algorithms*. Neural Computation, 4(6):888-900.

2. Breese, J., Heckerman, D., Kadie, C. (1998) *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, July.

3. Burke, R. (2002) Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction. 12(4), pages 331-370.

4. Cöster, R., Svensson, M. (2002) *Inverted File Search Algorithms for Collaborative Filtering.* International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 2002.

5. Cöster, R. (2002) *Learning and Scalability in Personalized Information Retrieval and Filtering*. Licentiate of Philosophy Thesis, Stockholm University, Department of Computer and System Sciences.

6. Cöster, R., Gustavsson, A., Olsson, T., Rudström, Å. (2002) *Enhancing web-based configuration with recommendations and cluster-based help*. Workshop on Recommendation and Personalisation in eCommerce, at 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, Malaga, Spain.

7. Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore E., Bylund M. (2001) *GeoNotes: Social and Navigational Aspects of Location-Based Information Systems*. In Abowd, Brumitt & Shafer (eds.) Ubicomp 2001: Ubiquitous Computing, International Conference Atlanta, Georgia, September 30 - October 2, Berlin: Springer, p. 2-17.

8. Foner, L. (1997) *Yenta: A Multi-Agent, Referral-Based Matchmaking System*. In Proceedings of The First International Conference on Autonomous Agents (Agents '97). pp 301-307. ACM Press.

9. Hansen, P. and Kalervo J. (2000) The Information Seeking and Retrieval process at the Swedish Patent- and Registration Office. Moving from Lab-based to real life work-task environment. Proceedings of the ACM-SIGIR 2000 Workshop on Patent Retrieval, Athens, Greece, July 28, 2000, pp. 43-53.

10. Herlocker, J., Konstan, J., Riedl, J. (2000) *Explaining Collaborative Filtering Recommendations*. In Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work, December 2-6.

11. Herlocker, J., Konstan, J., Borchers, A., Riedl, J. (1999) *An Algorithmic Framework for Performing Collaborative Filtering*. In Proceedings of the 1999 Conference on Research and Development in Information Retrieval. Aug. 1999.

12. Höök, K., and Svensson, M. (In Press) *Social Navigation of Food Recipes: Designing Kalas*. In A. Munro, K. Höök, and D. Benyon (eds.), *Social Navigation of Information Space (Second edition)*, Springer Verlag, In Press.

13. Kanerva, P., Kristoferson, J., and Holst, A. (2000) *Random indexing of text samples for Latent Semantic Analysis*. In L.R. Gleitman and A.K. Josh (eds.), Proc. 22nd Annual Conference of the Cognitive Science Society (Philadelphia), p. 1036. Mahwah, New Jersey: Erlbaum.

14. Karlgren, J. and Sahlgren, M. (2001) *From Words to Understanding*. In Uesaka, Y., Kanerva, P. & Asoh, H. (Eds.): Foundations of Real-World Intelligence, pp. 294-308, Stanford: CSLI Publications.

15. Kyte A. (2003) Business 2007: Drivers for Change, 17A, SPR4, 3/03, Gartner.

16. Landauer, T. K. and Dumais, S. T. (1997) *A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge*. Psychological Review, 104(2), 211-240.

17. Langley, P., Iba, W., and Thompson, K. (1992). *An analysis of Bayesian classifiers*. Proceedings of the Tenth National Conference on Artificial Intelligence (pp. 223-228). San Jose, CA: AAAI Press.

18. Lindgren, R., Stenmark, D. and Ljungberg, J. (2003). "Rethinking Competence Systems for Knowledge-based Organizations". European Journal of Information Systems, Vol. 12, Issue 1, pp. 18-29.

19. Lindgren, R. and Stenmark, D. (2002). "Designing Competence Systems: Towards Interest-Activated Technology". Scandinavian Journal of Information Systems , Vol. 14, pp. 19-35.

20. Olsson, T. (1998). *Decentralized Social Filtering based on Trust*. In: AAAI-98 Recommender Systems Workshop Papers. Technical Report WS-98-08. pp 84-88. AAAI Press.

21. Sarwar, B., Karypis, G., Konstan, J., Riedl J. (2000) *Application of dimensionality reduction in recommender systems-a case study*. In ACM WebKDD Workshop.

22. Sarwar, B., Konstan, J., Borchers, A., Herlocker, J., Miller, B., Riedl, J. (1998) *Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System*. In Proceedings of Computer Supported Cooperative Work, Seattle, WA.

23. Svensson, M., Höök, K., Laaksolahti, J., Waern A. (2001) *Social Navigation of Food Recipes*. In Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, Washington, USA, 341 – 348.

24. Svensson, M. (2003) *Defining, Designing and Evaluating Social Navigation*. Ph.D Thesis, University of Stockholm, Stockholm, Sweden.

25. Tveit, A. (2001). *Peer-to-peer based recommendations for mobile commerce*. In Proceedings of the first international workshop on Mobile commerce 2001. pp 26 - 29. ACM Press New York, NY, USA.

26. Vapnik, V. N. (1998) *Statistical Learning Theory*. Wiley.

27. Witten, I., Moffat, A., Bell, T. (1999) Managing Gigabytes: Compressing and Indexing Documents and Images (Second Edition)., Morgan Kaufmann Publishing.