Ollscoil na Gaillimhe
University of Galway

Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Semantic Social Collaborative Filtering with FOAFRealm
Author(s)	Kruk, Sebastian Ryszard; Decker, Stefan
Publication Date	2005
Publication Information	Sebastian Ryszard Kruk, Stefan Decker "Semantic Social Collaborative Filtering with FOAFRealm", Proceedings of the Semantic Desktop Workshop, in conjunction with ISWC 2005, 2005.
Item record	http://hdl.handle.net/10379/501

Downloaded 2024-04-25T12:38:57Z

Some rights reserved. For more information, please see the item record link above.

Semantic Social Collaborative Filtering with FOAFRealm

Sebastian Ryszard Kruk, Stefan Decker

Digital Enterprise Research Institute, Galway, Ireland*
<firstname.lastname>@deri.org, http://www.deri.org

Abstract. The most popular collaborative filtering implementations require either a critical mass of referenced resources and a lot of active users. Other solutions are based on finding a referral with an expertise on the given domain of discourse.
In this article we present the semantic social collaborative filtering solution to information retrieval. We describe how the concept of users' managed collections can be exploited to provide collaborative filtering system based on social network maintained by the users themselves.
We present FOAFRealm, a user profile management system based on the social networking and the FOAF metadata. FOAFRealm enables distributed collaboration between parties in the semantic social collaborative filtering way.

1 Introduction

The contemporary Internet contains a lot of information. In the unorganised structure of the Web all the information that we are looking for seems to be always just behind the corner. Though, still beyond our scope. And when we fail to find that information, it turns to be useless. Search engines and online catalogues tend to return a lot of resources as an answer to our queries. Very often some of results are unrelated to given queries. No wonder, we end up asking our friends and acquaintances for interesting references on the exact topic. Collaborative filtering is an idea of automating the process of asking around when looking for the information on the Internet[1].

Since early implementations of collaborative filtering, like introduced in [2], a number of methods have been developed for the "collaborative filtering" and "social filtering" [2–4]

Contributions The paper makes the following contribution to the field of collaborative filtering and user profile management systems:

* This material is based upon works supported by the Science Foundation Ireland under Grant No. SFI/02/CE1/I131. Authors thank all members of the JeromeDL (cf. http://www.jeromedl.org/) and the FOAFRealm (cf. http://www.foafrealm.org/) working groups for fruitful discussions on this document.

- We introduce a new approach to collaborative filtering - the semantic social collaborative filtering that covers both active and passive types and solves additionally some privacy/security issues.
- The reference implementation library (FOAFRealm), can be embedded into of web applications, providing additionally unified, distributed users management system based on FOAF.
- Our solution introduces goals like: distributed user profile management, privacy of the profile information, security of the provided knowledge, utilisation of social networks.

Outline of the paper The next section describes the architecture of the semantic social collaborative filtering in the context of other similar solutions. In section 3 we present the evaluation of the underlying model of social interactions in the semantic social collaborative filtering. We describe in section 4, the FOAFRealm system that implements a distributed user profile management system and delivers semantic social collaborative filtering features. Later, we discuss the relations between social collaborative filtering and digital library systems.

2 Semantic Social Collaborative Filtering

The semantic social collaborative filtering presented in this article is based on two concepts: distributed collections and annotations of resources. Each user classifies only a small subset of the knowledge, based on the level of expertise he/she has on the specific topic. This knowledge is later shared across the social network.

2.1 How does Social Collaboration Work

The problem that there is a trade-off between accuracy and scalability is often found in search engine applications. The information gathered in online collections is very precise, as the human factor is involved in the indexing process. But since the Internet is growing so fast, the process of creating the catalogue does not scale. On the other hand, search engines do the indexing work without involving the human activity. And results of queries are not always satisfiable.

A social network is a set of people or group of people, with some pattern of interactions or "ties" between them[5–8]. A social network is modeled by a digraph where the nodes represent individuals, and a directed edge between nodes indicates direct relationship between two individuals.

It is possible to construct a subgraph, on top of a social network, that represents flow of expertise in the certain domain. The idea of the semantic social collaborative filtering is based on this observation. Each person in the social network gathers the interesting information in collections he/she has created. Collections maintained by other people can be easily linked into own collections created by the user. As we show later (see section 3) the information disseminated through the collections linking across the social network corresponds to the expertise level on particular subject in the social network.

Distributed collections. The information is gathered in colletions by a number of people. Each of them handles specific domains of discourse within the colletions information space he/she has created. The quality of the information gathered across the colletions can be satisfied by approving the expertise in given domain of discourse.

Each user maintains his own collections (private bookshelf [9]) and renders them accessible to his/her friends [10]. We can assume that some of topics are better explored by some people. Each collection has a quality level assigned to it, based on the expertise the owner has on the related topic. Each user is also aware of the expertise level of other people on given topics.

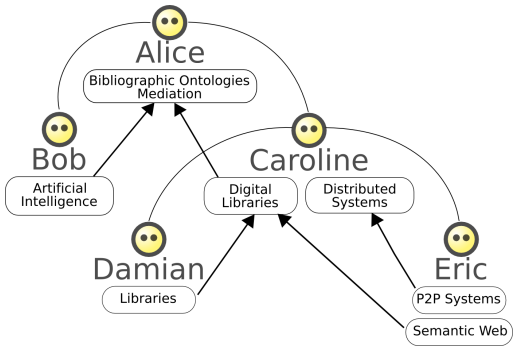

Fig. 1. The scenario of a simple semantic social collaborative filtering model

Resources annotations. Apart from managing collections by providing the categorisation description of resources, the semantic social collaborating filtering utilises comments and annotations provided by the users. The annotations are represented as fora with some additional semantic content. Annotations can be used by other people as a shorthand to quickly explore: (1) the content or meaning of the resource; (2) the context of resources; (3) the general opinion of other users.

2.2 Semantic Social Collaborative Filtering Scenario

In our example scenario, Alice writes a thesis on "Mediation in Bibliographic Ontologies". She registers to the digital library run by the University. She discovers that some of her friends are already registered to the library as well. With features based on online communities, she connects her profile to her friends profiles. Later on, Alice starts to gather the information required for her thesis topic. She keeps links to resources she has found in collections managed by the online bookmarks system. Soon she discovers that resources that she has bookmarked do not cover the topic of the thesis at satisfiable level. The following sections describe different algorithms Alice uses to find the desired information with the help of the semantic social collaborative filtering.

Simple Social Collaborative Filtering To find the desired information Alice sings up to the university digital library. The system used by the library is based on the simple semantic social collaborative filtering implementation (see Fig. 1).

Alice uses the searching features provided by the digital library web application (see Fig. 2(a)) to find interesting resources.

We introduce a solution to the problem stated in previous section (see 2.2) – a simple semantic social collaborative filtering model (see Fig. 2(b)). Each collection is categorised by the owner. Collaborative filtering feature in the digital library lists all the collections, within the given range of friendship neighbourhood, with topics related to the ones defined by Alice. Each collection has a quality level assigned to it. The quality of the collection corresponds to the expertise level of the owner on related topic. The expertise level can be computed with PageRank algorithm applied to graphs of collections inclusions and social network. Both graphs represents the rank value each person and each collection receives from other people. The rank values are assigned directly (by people to people) and indirectly (by including someone's collection to own collection).

Alice finds out that one of her friends, Caroline, gathers the information about digital libraries and her expertise level on that topic is very high. Though her direct friend Bob is interested in Artificial Intelligence, she finally decides to link resources provided by Eric, who has a highly ranked "Semantic Web" collection. From now on, Alice takes the advantage of the information gathered by Caroline and Eric in their collections.

Secured Semantic Social Collaborative Filtering Alice is still looking for more information required for her thesis. She decides to register in an open, hererogenous digital library. Some people protect their collections with access control restrictions (see Fig.3). The restrictions applied on the collection are based on maximal distance and minimal trust level between two people in the social network graph. Apart from defining friendship relations, users express the quality (trust level) of every outgoing social connection.

Since not all information should be accessible by everyone, some of it need to be protected from people from the outside of the given community. This is why access control lists (ACL) have been introduced (see Fig.4(a)). In the semantic social collaborative filtering environment based on ACL each collection has its own ACL, that defines the maximal distace and minimal friendship quantisation level from the specific person[1] to the person willing to access that collection. Only when this is satisfied the user can access and include this collection in his/her collections.

Alice wants to make use of the knowledge provided by Damian. But the algorithm for retrieving a list of collections in the secured environment (see Fig.4(b)) omitted some of collections. With ACL applied Alice is out of the range defined in Damian's ACL constrains. The collection managed by Damian is not presented to her.

[1] please note that it does not have to be an owner of the collection, thought the owner is the one that manages ACL

```
procedure ListCollections_SM(p,t) : collections[]
 for p' ∈ P with PeerDistance(p,p') < knowsRange
  C' ← C' ⋃ PeerCollection(p')
 end for
 sort C' according to FinalRanking_SM
end procedure
```

(a) Algorithm retrieving list of collections

P is a set of peers
C is a set of collections
$FoafKnows$ is a set of directed connections between peers
$G_{peers}(P, FoafKnows)$ is a digraph of friendship relations
T is a lattice of categorisation topics
We assume that each collection $c \in C$ has exactly one owner $p \in P$.

`PeerCollection`: $P \rightarrow 2^C$ – returns all collections owned by the peer
`OwnedBy`: $C \rightarrow P$ – returns the owner of the collection
`Expertise`: $(P, C) \rightarrow [0, 1]$ – computes the quality of the collection based on the peer's expertise on related topic
`Categorisation`: $C \rightarrow T$ – returns the list of topic describing collection
`PeerDistance`: $(P, P) \rightarrow \mathbf{N}$ – computes distance between two peers in the social network graph using Dijkstra algorithm
`Similarity`: $(T, T) \rightarrow [0, 1]$ – computes similarity level between two topics
`FinalRanking`$_{SM}$: (`PeerDistance`,`Similarity`,`Expertise`) $\rightarrow [0, 1]$ – computes ranking value for a collection based on distance to the owner, similarity level and quality measure

knowsRange – defines a maximal distance between two people when traversing the graph of friendship relations.

(b) Definition of model

Fig. 2. The simple model of social collaborative filtering

2.3 The Benefits

The main bottleneck of existing passive collaborative filtering systems is the process of gathering users' preferences[4]. A reliable system requires a very large number of people to express their opinion about a large number of topics. This requires from users to either fill out a survey or perform some activities (like e.g. buying a product, reading a book) over a certain time.

Active collaborative filtering solutions depends on maintaining the social network by users themselves. Outdated information on list of friends can mislead the person in his quest for an answer.

Backward Referral Chaining Maintaining a list of friends, posting a question and gathering the answers may be time consuming. That is why the social collaborative filtering (a new approach to active collaborative filtering) tends to

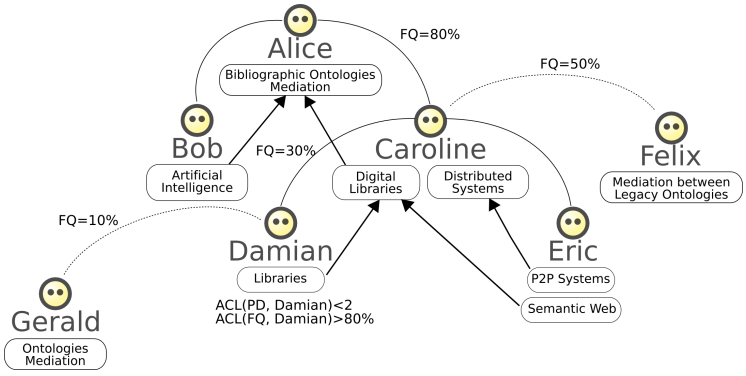

Fig. 3. The scenario in the secured semantic social collaborative filtering model

ease some hardships by introducing the concept of *backward referral chaining*, reusing existing classification schemata and extrapolating user's profile information with interests of his friends.

Usually, a user is not aware of the whole social network. To gather the knowledge outside of his direct friendship neighbourhood the user has to rely on references provided by his friends. Because the expert in the specific domain can be quite distant from the user, in terms of relationship links, the access to the answer provided dependents on the path to an expert. As it has been introduced in 2.2, an expert can restrict the access to some parts of information by applying access control lists.

The referral chaining[11] has two strong dependencies: *accuracy* of finding the right path to an expert, and *responsiveness* factor of the found expert. The backward referral chaining introduced in the social collaborative filtering inverses the process of finding an expert. The answers provided by different people (including experts) are being assembled into hierarchical knowledge base. Users link into their collections, information provided by some other people. In many cases, the expertise of the latter, on given topic is higher.

Connection to the established classification schemata. In social collaborative filtering each person can create own categories according to the local understanding of the world. The definition of the category might be hard to understand to other peers because of the use of ambiguous descriptions or an native language.

We propose to apply additional semantically reach description based on existing thesauri or classification ontologies, like WordNet[12] or Dewey Decimal Classification[13, 9]. This description can help to understand the meaning of the category both to people and machines. The latter can then utilise this knowledge in e.g. recommending related categories created by other peers.

ACL_{PD} is an access control constrains, defining maximal distance D (in number of 'hops') from user P
ACL_{FQ} is an access control constrains, defining minimal $FriendshipQuantization$ value (calculated across the graph) from user P

`DistanceACL`: $(C) \rightarrow 2^{ACL_{PD}}$ – defines a list of allowed maximal distances to the user
`QuantizationACL`: $(C) \rightarrow 2^{ACL_{FQ}}$ – defines a list of allowed minimal $FriendshipQuantization$ values
`Peer`: $(ACL) \rightarrow P$ – returns a peer from which the computation of ACL dinstance/level is do be performed
`Distance`: $(ACL_{PD}) \rightarrow \mathbf{N}$ – returns the maximal distance defined in ACL
`Quantisation`: $(ACL_{FQ}) \rightarrow [0,1]$ – returns the minimal $FriendshipQuantization$ level

(a) Definition of model

procedure ListCollections$_{ACL}$(p,t) : collections[]
$c_p \leftarrow$ `PeerCollection`(p)
for $p' \in P$ **with** `PeerDistance`$(p,p') < knowsRange$

for $c' \in$ `PeerCollection`(p')
with $\forall acl_{PD} \in$ `DistanceACL`(c') `PeerDistance`(`Peer`$(acl_{PD}), p) <$ `Distance`(acl_{PD})
with $\forall acl_{FQ} \in$ `QuantizationACL`(c') `FriendshipQuantization`(`Peer`$(acl_{PD}), p) >$ `Quantisation`(acl_{PD})
with `CollectionDistance`$(c_p,c') < inclusionRange$
$C' \leftarrow C' \bigcup \{c'\}$
end for
for $c \in C$ **with** `FriendshipQuantization`$(p,$`OwnedBy`$(c)) > quantisationLevel$
if $\forall acl_{PD} \in$ `DistanceACL`(c') `PeerDistance`(`Peer`$(acl_{PD}), p) <$ `Distance`(acl_{PD})
and $\forall acl_{FQ} \in$ `QuantizationACL`(c') **and** $\exists c' \in C'$ `CollectionDistance`$(c_p,c') +$ `CollectionDistance`$(c',c) < inclusionRange$
then $C'' \leftarrow C'' \bigcup \{c\}$
end for
sort $(C' \bigcup C'')$ according to `FinalRanking`$_{CI}$
end procedure

(b) Algorithm retrieving list of collections

Fig. 4. The secure model of social collaborative filtering

Extrapolated user's profile. When information about user's activities (personal bookshelf, resources' annotations) is gathered for a longer time it can be re-used during the search process. The query expansion process[9] takes into account semantically rich descriptions of users' preferences reflecting their activities. The result set becomes more user oriented than with a generic search.

New users registered to the system very often suffer from lack of rich profile information. This may have a strong influence on the quality of search results. To overcome this problem then social collaborative filtering paradigm introduces the concept of *an extrapolated user's profile*. The profile of the new user can

be represented with some probability depending on trust level (see 2.4) as a combination of profiles of his/her friends.

2.4 Security and Privacy Issues

Collaborative filtering implementations suffer in most cases from very weak security features or frequent privacy abuse. The information about the user in passive collaborative filtering systems is very often gathered without his knowledge. In the active collaborative filtering the user very often has no chance to protect himself from gathering information about him.

To implement the security and privacy features the concept of digraph of interpersonal connections have been utilised. Each user defines a list of his friends and states the level of trust to each of them. The user can then define the maximal distance and minimal trust level required from the person which wants to view information gathered in specific category.

As all the information about the user is provided by himself and he/she manages the access control lists for each piece of information, the privacy of the user is preserved.

3 Evaluation of Semantic Social Collaborative Filtering

Semantic social collaborative filtering utilises existing social networks instead of creating artificial connections between people. That is why on the contrary to other collaboration filtering solution, there is no need to evaluate an algorithm for creating a social network, as the social network is given explicitly.

On the other hand, since the semantic social collaborative filtering is based on friendship connections, the actual similarities of interests between connected users might differ. That is why, the evaluation of this collaborative filtering approach should prove that the dissemination of knowledge is possible within graph of semantically annotated friendship connections.

Simulation model In this section we present the implementation of the simple semantic social collaborative filtering model. We prove that average level of expertise in the subgraph of social network is almost maximal within 6 degrees of separation.

The definition of a simulation model has been based on similar ideas defined in Refferal Web project[11]. The main difference between social semantic collaborative filtering and the Refferal Web is that in the Refferal Web project, the process of finding an expert on certain topic is performed manually by the user. In semantic social collaborative filtering, semantical annotation on the knowledge provided in the social network is used to automate the process of finding the high quality of information. The simulation model itself might be similar to the one presented in [11], so we just need prove that it is possible to find an expert on the given maximal degree of separation.

Underlying assumptions. In the model of social network for the semantic social collaborative filtering, each user manages collections with information on selected topic. The different users represent different expertise on the given topic. We assume that:

- The quality of the information provided by a user on a certain collection is proportional to the expertise level of the user on the topic of collection.
- It is possible to find a user with a high expertise on given topic within the network of social connections.

According to simple social collaborative filtering model (see Fig. 2(b)) the simulation environment is modeled by a set of users and a set of collections managed by those users. There is exactly one user that owns each collection. On the base of the user's expertise on related topic the quality of the collection is defined. Each user has a predefined set of other users he knows (this relation should not be considered as implicitly symmetric).

Although according to the Small World Phenomena[14, 15] the distribution of the degree of the friendship connections is power-law based (Zipf's distribution, see Eq. 1) we have decided to perform second set of experiments where the degree of friendship connections is a bell-curve shaped (normal random variable see Eq. 2).

$$O(i) = \frac{n}{i^{\theta} H_{\theta}(V)}, H_{\theta}(V) = \sum_{i=1}^{V} \frac{1}{i^{\theta}} \tag{1}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

The distribution of expertise on a certain topic within the social network can be based on the Lotka's Law[16], stating that *the number of authors making n contributions is about* $\frac{1}{n^a}$ *of those making one contribution, where a is often nearly 2*. Since the expertise on a certain topic is proportional to the number of high quality of publications, the probability of the level of expertise (the level of expertise over the number of users that have one) is Zipfian shaped as well. Each collection has a quality value assigned to it that represents the expertise the owner has on the related topic.

In order to make sure that there would be at least one absolute (`Expertise`(T) = 1) expert in each topic T, we have normalised the associated expertise values dividing each but the value of the highest expertise in each topic.

The list of topics used to describe content of collection has been based on Dewey Decimal Classification[13]. This simplifies the computability of the model in the sense of comparison similarity between topics. Each category has a three-digit number (100 - 999) associated. Categories are structured as a three level tree. Although in the real world implementation categories are described additionally with WordNet words vectors, DDC seems to be enough for the modeling purpose.

```
procedure AverageMaximalExpertise(R) : Ē_max(R)
 for p' ∈ P with
  select t ∈ T_p find c that
   t =Categorisation(c)
   PeerDistance(p, Owner(c)) < R
   e =Expertize(Owner(c), c) is maximal
  AverageMaximalExpertise + = e/N_P
end procedure
```

Fig. 5. Algorithm calculating average maximal expertise in the semantic social collaborative filtering model in the given range

Definition of the experiment. During the experiment each user ($p \in P$, $\texttt{sizeOf}(P) = N_P$) tries to find in the social network within a given range R, the collection that provides the information on the topic $t \in T_p$. The topic is randomly selected from the list of topics associated to collections owned by the user. The average value of the highest expertise $\bar{E}_{max}(R)$ level found within given range is computed (see Fig.5).

We have performed four experiments. Each time the social network model consited of $N_P = \mathbf{1000}$ users.

Each user in our social collaborative filtering environment had only one collection associated. This simplification is correct since during the experiment we are looking only for collections with exactly the same topic as selected. So collections associated with each topic creates a subgraph that is independent of the actual number of collections owned by each user.

The expertise level for each collection has been randomly selected according to power law distribution. In the first two experiments the degree of friendship connections has been randomly selected according to normal distribution (μ =**25**, σ =**12.5**). In the last two experiments the power law distribution (θ =**1.9**) has been applied. During each experiment average maximal expertise values $\bar{E}_{max}(R)$ has been calculated for maximal degree of separation $R \in [1, 8]$.

Results of simulation. Table 1 presents results of all four experiments.

Table 1. Results of the experiment - average maximal expertise $\bar{E}_{max}(R)$

R	σ **= 12,5F (Bell)**		θ **= 1,9F (Zipf)**	
1	0,07072	0,06427	0,01793	0,01595
2	0,69098	0,69192	0,10557	0,09042
3	0,96399	0,96183	0,33044	0,29836
4	0,96796	0,96782	0,62892	0,61653
5	0,96796	0,96782	0,82896	0,82980
6	0,96796	0,96782	0,90953	0,91751

It is interesting that even for the power law based distribution user is able to find information with almost the highest possible quality within **6** degrees of separation (see Fig.6).

Conclusions on results of simulation Following experiments by Kauth[11] we have constructed similar social collaborative filtering model. The results revealed that each user is able to find (on average) the best quality of information provided by other users within the subgraph of social network bounded by **6** degrees of separation. These experimental results proved that the constructed social network model corresponds to the small world phenomena[14]. Hence, the assumptions underlying the social collaborative filtering has been fullfiled. It is possible to find an expert (with an average expertize level above 90%) within the small social network neighbourhood.

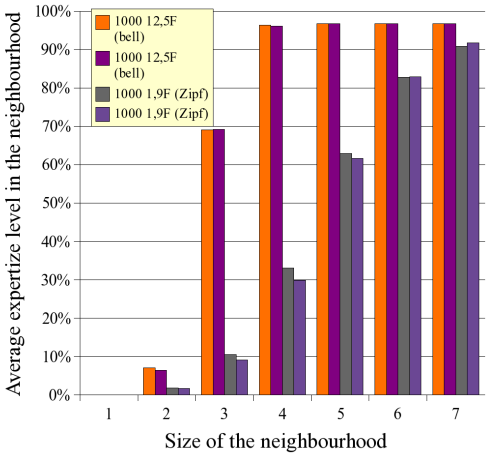

Fig. 6. Average expertise level in the neighbourhood of the given size

4 FOAFRealm - the Reference Implementation of Semantic Social Collaborative Filtering

The FOAFRealm is a library for distributed users management based on the FOAF vocabulary. It enables users to control their profile information, as the information can be accessed in the open FOAF format. Users can sign-in automatically across the P2P network (called D-FOAF[2]) of FOAFRealm enabled systems[17].

FOAFRealm provides a basic implementation of the semantic social collaborative filtering concept. The knowledge (annotations and private collections) can be shared among registered users. Security constraints can be applied to each piece of information separately.

The current implementation of FOAFRealm consists of four layers:

[2] D-FOAF project: http://d-foaf.foafrealm.org/

- The distributed communication layer providing access to highly scalable HyperCuP[3] P2P infrastructure to communicate and share the information with other FOAFRealm implementations..
- FOAF and collaborative filtering ontology management. It wraps the actual RDF storage being used from the upper layers providing simple access to the semantic information. The Dijkstra algorithm for calculating distance and friendship quantisation is implemented in that layer.
- Implementation of the `org.apache.catalina.{Realm,Valve}` interfaces to easily plug-in the FOAFRealm in to Tomcat-based web applications. It provides authentication features including autologin based on Cookies.
- A set of Java classes, Tagfiles and JSP files plus list of guidelines that can be used while developing user interface in own web applications.

The library has been successfully deployed as a user management system in JeromeDL - e-Library with Semantics[4]. It is used to handle private bookshelves of readers, and provides additional semantical annotations to the resources. The concept of extrapolated user profile has been adapted in the semantically enhanced search engine in JeromeDL. So that even new users to the system can benefit from the full-fledged semantic search process.

The FOAFRealm system has also become a part of MarcOnt Initiative[5] collaboration portal for ontologies management based on negotiations. The portal will utilise social networks based features of FOAFRealm to:

- *isolate outside world from the ontology management community.* The registered users will be allowed to take part of the ontology management process when they will be defined as a friend of at least on of the community members.
- *differentiate evaluations of ontology changes suggestions provided by different members of the community.* We will explore if evaluations provided by close friends of the person that posted the suggestion should be ranked lower than evaluations provided by people with higher degree of separation from the suggestion owner.

5 Related work

Collaborative filtering. The most popular types of the collaborative filtering systems are Active Collaborative Filtering and Passive Collaborative Filtering. The distinction between those is based on the activeness of the user that receives information. With passive collaborative filtering, the information about the user, such as: mailing-lists posts, links on home pages, citations in publications and co-authors of articles, is utilised. Since the user does not actively take part in maintaining his network of friends, he has no direct impact on information he/she receives.

[3] Lightweight HyperCuP Implementation project: http://www.hypercup.org/
[4] JeromeDL - e-Library with Semantics: http://www.jeromedl.org/
[5] MarcOnt Initiative: http://www.marcont.org/

Active collaborative filtering implements two models of information retrieval: *user pull model* - where a user generates a query to the network of other users, and *user push model* - where the answers on previously stated questions or information filters, are feed to the user.

Though by shifting from central (a search engine) to a distributed method of recommendation the problem tends to be more manageable, particular collaboration filtering implementations suffer various difficulties: **(1)** "heterophilous diffusion" (exchange information across different socio-economic groups) is neglected in favour to "homophilous diffusion" (exchange of information within socio-economic groups); **(2)** security and privacy issues are weakly supported; **(3)** meaning (semantics) of shared concepts are lost; **(4)** when the network of friends is created automatically by harvesting various databases with advance algorithms: the "critical mass" of registered users is required to provide satisfiable level of correlation to user's interests; it is impossible to create a digraph of social connection from most of commonly used sources; privacy of individuals is violated; monopolies are supported[18] because a service provider has to gather a lot of information to become accurate ("critical mass"); **(5)** when the user actively uses fora or mailing-lists: (i) there is no guarantee that there will be an answer to the posted question, or that the answer will be through; (ii) there might be no expert on the specific field of discourse in the "direct friendship neighbourhood" of the user; **(6)** some systems requires from users to answer long questionnaires [4] in order to find similarities in users' interests.

Hybrid filtering[19], the combination of content filtering and social filtering, is used to maximise precision with a recall still above specified limit.

Active collaborative filtering solutions concentrate on utilising the existing social connections provided explicitly. One of the approaches [10] is build on the the common practise where people tell their friends or colleagues of interesting documents. Users collect bookmarks on the interesting World Wide Web pages that they have found. [20] describes a social collaborative filtering system where users have direct impact on filtering process. The changes in the users interests are exploited to provide thorough relevance feedback to the system..

To format and distribute collections of bookmarks several simple system have been developed. With Simon system [21] users can create "subject spaces" which are lists of hypertext links to the WWW pages with annotations on them. One of other possible solutions is to find a personal referral that can answer the given query. The network of relationships can also help in exploring the hidden web, the part of the Internet that is not indexed by search engines [11], as some of the information is deliberately not accessible outside the intranets [10, 22].

Online social communities are the underlying key concept of the semantic social collaborative filtering presented in this article. In the last few years this field has been widely explored by several implementations.

Some of them, like Orkut[6] provides forum-like channels of dissemination of knowledge, where community members can ask questions to their friends

[6] Orkut online community portal: http://www.orkut.com/

or other members of specific thematic group. In the Semantic Web field the FOAF (a vocabulary for RDF [23]) format has been introduced to describe the interpersonal connections.

User Profile Management The existing implementations of user profile management lack: (i) fine granularity of security constraints; (ii) scalability; (iii) openness/privacy. Both of which play important roles in semantic social collaborative filtering.

One of the features that is becoming more and more important in social P2P environment is single-sign-on[24]. Each time a user uses a new web system, he would rather not provide all the same information about himself over and over again. Solutions like Microsoft Passport[7] or Sxip[8] provide such features.

6 Future Work and Conclusions

The semantic social collaborative filtering presented in the article opens new possibilities of exchanging and managing knowledge. Users can share their bookmarks (collections and their content) with their friends. Everyone can organise the knowledge by gathering collections that other people are maintaining. Since collections can be linked it is possible to find more relevant information in categories provided by some distant people. Annotations are also a key part of the semantic social collaborative filtering. Together with private collections (private bookshelves) they are utilised in the semantically enhanced information retrieval in systems like digital libraries.

FOAFRealm is a reference implementation of the semantic social collaborative filtering. It refers to social networks and open standards like FOAF. FOAFRealm provides support for J2EE based web applications for quick extending their features with user management and social collaborative filtering. Since the social network is represented as a digraph, FOAFRealm utilises informations about distance between two people and the trust level, to provide the security and privacy features.

Current implementation of FOAFRealm, D-FOAF, provides a distributed user profile management system and hence, the social semantic collaborative filtering across different systems. The future step, DigiMe, will deliver this features to mobile devices and will explore the ad-hoc social networks paradigm.

References

1. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM **35** (1992) 61–70

[7] Microsoft Passport: http://www.passport.net/
[8] SXIP - Passport/Liberty done right: http://www.sxip.com/

2. Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, ACM (1994) 175–186
3. Breese, J.S., Heckerman, D., Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. (1995) 43–52
4. Shardanand, U., Maes, P.: Social Information Filtering: Algorithms for Automating "Word of Mouth". In: Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems. Volume 1. (1995) 210–217
5. Trevor, J., Hilbert, D.M., Billsus, D., Vaughan, J., Tran, Q.T.: Contextual Contact Retrieval (2002)
6. Fukui, H.O.: SocialPathFinder: Computer Supported Exploration of Social Networks on WWW (2003)
7. Newman, M., Watts, D., Strogatz, S.: Random graph models of social networks. In: Proc. Natl. Acad. Sci., to appear. (2002)
8. Hoadley, C., Pea, R.: Finding the ties that bind: Tools in support of a knowledge-building community (2002)
9. Kruk, S.R., Decker, S., Zieborak, L.: JeromeDL - Reconnecting Digital Libraries and the Semantic Web. In: DEXA'2005. (2005)
10. Maltz, D., Ehrlich, K.: Pointing the way: active collaborative filtering. In: Proceedings of the Conference on Computer-Human Interaction. (1995) 202–209
11. Kautz, H.A., Selman, B., Shah, M.A.: The Hidden Web. AI Magazine **18** (1997) 27–36
12. Fellbaum, C.: WordNet An Electronic Lexical Database (1998)
13. Dewey, M.: A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library - Dewey Decimal Classification. guternberg.net, http://www.gutenberg.net/catalog/world/ readfile?fk_files=59063 (2004)
14. Milgram, S.: The small world problem. Psychology Today **67** (1967)
15. Barabasi, A.L.: Linked: The new science od Networks. Cambridge Perseus Press (2002)
16. Lotka, A.J.: The Frequency Distribution of Scientific Productivity. Journal of the Washington Academy of Sciences **16** (1926) 317–323
17. Grzonkowski, S., Gzella, A., Krawczyk, H., Kruk, S.R., Moyano, F.J.M.R., Woroniecki, T.: D-FOAF - Security Aspects in Distributed User Managment System. (In: TEHOSS'2005)
18. Techniques, R.P.: Privacy-Preserving Collaborative Filtering using (2003)
19. Basu, C., Hirsh, H., Cohen, W.W.: Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In: AAAI/IAAI. (1998) 714–720
20. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive web search based on user profile constructed without any effort from users. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, ACM Press (2004) 675–684
21. Simons, J.: Using a Semantic User Model to Filter the "World Wide Web" Proactively. (1995) 455–456
22. Kautz, H., Selman, B., Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering. Communications of the ACM **40** (1997) 63–65
23. http://www.w3.org/TR/rdf: Resource Description Framework (RDF). W3C Recommendation (1999)
24. Pearlman, L., Welch, V., Foster, I., Kesselman, C., Tuecke, S.: A Community Authorization Service for Group Collaboration (2002)