Master Thesis

Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation

by
Kai Bielenberg and Marc Zacher
submitted to the Program of Digital Media, Universität Bremen
in partial fulfillment of the requirements for the degree of
Master of Science in Digital Media

16. Aug. 2005

Thesis Advisor
Prof. Dr. Michael Koch, Technische Universität München

Chair
Prof. Dr. Herbert Kubicek, Universität Bremen

digitalmedia

Master Thesis

Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation

by
Kai Bielenberg and Marc Zacher
submitted to the Program of Digital Media, Universität Bremen
in partial fulfillment of the requirements for the degree of
Master of Science in Digital Media
16. Aug. 2005

Thesis Advisor
Prof. Dr. Michael Koch, Technische Universität München

Chair
Prof. Dr. Herbert Kubicek, Universität Bremen

We hereby declare that this thesis was written by ourselves. No other sources than those specified have been used.
All parts of this thesis have been developed together. Kai Bielenberg has written the Chapters 2.1, 2.3, 3.1, 3.3.2, 3.3.3, 4.1.1, 4.2.1 and 4.3, Marc Zacher wrote the Chapters 2.2, 2.4, 3.2, 3.3.1, 3.3.4, 4.1.2 and 4.2.2. All other chapters, as well as the practical implementation of "GROOP.US" have been realized collaboratively.

Wir erklären hiermit, dass wir diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben.
Alle Teile dieser Arbeit wurden gemeinsam erarbeitet. Kai Bielenberg hat die Kapitel 2.1, 2.3, 3.1, 3.3.2, 3.3.3, 4.1.1, 4.2.1 und 4.3 geschrieben, Marc Zacher die Kapitel 2.2, 2.4, 3.2, 3.3.1, 3.3.4, 4.1.2 und 4.2.2. Alle nicht aufgeführten Kapitel, sowie die praktische Implementierung „GROOP.US" wurde gemeinsam angefertigt.

Kai Bielenberg Marc Zacher

Abstract

Currently, weblogs and social web services attract more and more attention, encouraging information publishing and collaborative annotation by a broad mass. Shared metadata creation - especially in the form of adding tags - enhances structured access to information. Moreover, usage of tags may give hints about people with similar interests and similar ways of thinking and speaking.
The goal of this thesis is to explore the potential of tags to construct social networks and to identify groups with shared contexts that assure a common understanding of tags and related resources. These shared contexts are the basis for deriving recommendations to realize social navigation.
Founded on theories from social network analysis, the field of social navigation and concepts like folksonomies and transactive memories we will propose a framework for social navigation based on tagging. We will apply this framework in the implementation of a prototype called *GROOP.US* to show practicability of the chosen approach.

Das Interesse an Weblogs und sogenannten Social Web Services nimmt gegenwärtig mehr und mehr zu. Diese Applikationen ermöglichen der breiten Masse komfortabel Informationen zu veröffentlichen und gemeinsam zu annotieren. Gemeinsame Metadatenerstellung, vor allem in Form einfacher Tags, erleichtert einen strukturierten Zugriff auf Information für alle Beteiligten. Darüber hinaus kann die Auswahl von Tags Aufschluss über gemeinsame Interessen, gemeinsame Denkweisen und einen gemeinsamen Sprachgebrauch geben.
Ziel dieser Arbeit ist das Potenzial von Tags für die Erstellung sozialer Netzwerke und darüber hinaus die Identifikation von Gruppen mit einem gemeinsamen Kontext zu erforschen, der ein einheitliches Verständnis für Tags und zugehörige Ressourcen sicherstellt. Auf Grundlage dieser gemeinsamen Kontexte sollen anschließend Empfehlungen für die Verwirklichung von Social Navigation abgeleitet werden.
Basierend auf Theorien der Analyse sozialer Netzwerke, dem wissenschaftlichen Bereich Social Navigation und Konzepten wie Folksonomies und Transactive Memories werden wir ein Framework für Social Navigation basierend auf Tagging vorschlagen. Das Framework wird dann prototypisch in unserer Anwendung GROOP.US umgesetzt um das Framework zu analysieren.

Table of Contents

1. Introduction

Being up to date has always been a challenge in any knowledge domain. Filtering important information from the data available involves having access to many sources, tracking cross-references, knowing important authors and editors and having a network of people who give hints and information about information you have not yet retrieved.
Today's technical networks - and in particular the Word Wide Web - are both helpful and a threat to this process. Helpful because they potentially facilitate information exchange and knowledge sharing. A threat because of the higher amount of - partly unreflected - data, as it is even more difficult to track down the signal from noise.

A common approach to face "information overflow" in the World Wide Web is using search engines. Search engines try to index the content of as many web pages as possible. A person looking for information specifies keywords or phrases that she thinks signify the topic she is interested in. But instead of analyzing the semantic meaning of a search request, most often search engines simply perform a syntax based pattern matching between the search string and all indexed resources ignoring any semantics.
Automatic evaluation of a resource's semantic meaning is exceptionally difficult. Thus, reliable semantics have to be provided manually. People have to assign additional information about resources, like descriptions or categories. This information is called metadata. Examples for metadata based information retrieval are web directories like Yahoo (www.yahoo.com), shopping and auction systems like amazon.com (www.amazon.com) or ebay (www.ebay.com) or community portals like dooyoo (www.dooyoo.de).

Information filtering based on metadata seems more promising than mere syntax matching. One reason why Google (www.google.com) dominates the search engine market is that they do not only rely on pure syntax comparison but additionally "exploit metadata about the structure of the World Wide Web" [Doctorow2001]. Google counts the number of links which point to a web page assuming that frequently linked pages are more important than seldom linked ones. Evaluating information like links and references is one option to gather implicit metadata about the relevance of a resource and relations between resources. But how to engage people to not just create implicit metadata like links but to state explicit information about a resource?

With the ongoing success of personal web journals - also called weblogs - there is a tendency towards widespread annotation of information. People create their own content on weblogs. Being highly interlinked, weblogs connect friends and people with similar interests in social networks. Within such a network, people comment on other people's written thoughts and thus providing direct feedback on a shared topic of interest.
Based on the principles of weblogging - sharing, interlinking and commenting on information - new web services such as bookmark managers, shared photo collections or personalized online radios emerged. They all have in common that people share their content (weblinks, pictures, music playlists) with others allowing them to add metadata like comments, ratings and keywords (also referred to as tags or labels). Software that "supports, extends, or derives added value from human social behaviour" ([Coates2005]) can be defined as social software.
What is worth noting in this development is the fact that not only a few experts are revaluating information with metadata but that a broad mass of former passive recipients is now actively participating in information annotation - or more abstract: in a collaborative knowledge sharing

process. This process happens within the social context of a web service or a network of connected weblogs. People add metadata not only for their individual advantages (like categorization for eased find-ability for later retrieval) but to share value with friends and acquaintances to benefit in return from others' efforts. Metadata creation does not only become widespread, but also a social activity.

1.1 Our Focus

Metadata creation on a broad level is a precondition for semantic based information retrieval. Weblogs and contemporary web services foster metadata creation, but it has to be kept in mind that this metadata reflects only the subjective attitude of its creator. To estimate the value of metadata the context of its creation always has to be considered. Otherwise metadata may be ambiguous, it may contradict other statements, it even may be wrong. The question is how individual metadata can be valuable for a broader audience while minimizing these problems. Social networks could be one solution. If people with a shared understanding about a topic are brought together in a group, the probability is high that subjective information of an individual is understood by other group members. A shared understanding may be based on a joint vocabulary, a similar level of knowledge and a shared topic of interest.

Currently social software and weblogs are becoming important sources for studying social network phenomena. For example contributions on weblogs give hints about the topics their authors are interested in. Links to related articles or people with similar interests may be shown. Comments from others indicate who is actively interested in the discussed topic. People may link to other related resources and discussions - probably on their own weblog. To sum it up, a social network among authors, commentators and readers evolves.
But relations within such an online network often stay vague and are only obvious to the people that constantly follow all discussions. A lot remains unsaid or implies information from other not explicitly mentioned contexts. The interesting question is how such implicit social networks - groups of people with shared interests, opinions or activities - can be identified? How can information in such networks be made available for a larger audience? And can similar but unrelated social networks be brought together to foster information exchange?

The central aspect of our thesis is the identification of groups in social software to foster shared knowledge management. The data we will rely on originates from external web resources - like weblogs, bookmark or image sharing services - and is aggregated on the basis of standard exchange formats.
In contrast to group identification in the area of social network analysis we do not have precise information - like tie strengths or density (cf. 2.3) - about relations and interactions within the network we are analyzing. Instead, we are solely relying on evaluating subjective metadata stated by every participant herself. The information we will concentrate on are tags for categorizing web resources. This metadata does not only provide semantic information about a resource, but even more important gives clues about people acting in similar ways. That is, adding similar metadata to the same resources. Or: talking about the same topics with similar words. Bringing together people that act and talk similarly corresponds to the idea of merging people's individual contexts to a shared group context. Such shared group contexts provide the basis for recommendation of resources - one aspect of social navigation.

The whole concept can be summed up in the title of this thesis: "Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation".

1.2 Related Work

Related work will be referred to along the way of elaborating the scientific basis of our approach (cf. Chapter 2). Nevertheless, as an introduction some related projects in the areas of social networks, collaborative filtering and tagging shall be presented. We start with work that partly influenced our approach. Then, a short glance on social web services our work is inspired and based on will be followed by a presentation of two projects also relying on social web services.

One of the most widely respected projects in the area of collaborative filtering is GroupLens and its related scientific work in the field (cf. [Resnick1994], [Konstan1997]). The application enhances existing usenet clients. Entries can be rated which is recorded in a user profile. Recommendations are based on correlation of user profiles.

Kanfer et al. 1997 discuss how to use email archives to create a network for social navigation. Their "Know-Who" email Agent takes a natural language query as input and identifies people in the social network who are likely to help in the subject. For this reason a user's inbox is divided into boxes for each sender. Important keywords of a query are compared to the cumulated content of each sender's box. For the best matches contact data of the sender is proposed. If there is no match, the search is conducted in the inboxes of the user's contacts. The final distance of an adequate person from the user is returned as confidence level (cf. [Kanfer1997].

A very similar approach is proposed by Golbeck and Hendler [Golbeck2004a]. They discuss to use machine readable information about social networks such as FOAF (cf. [Brickley2005]). An algorithm is proposed that calculates a reputation value for other people in the network based on transitivity (cf. Chapter 2.3.1).
The technique was applied to a "TrustMail" client that predicts trust values for senders of incoming mails in general or related to a certain topic. This allows to prioritize mails in the inbox.

A third project dealing with email is "Social Network Fragments" [Viégas2004]. The authors criticize that email and chat contain a lot of meaningful interaction which is not leveraged for navigation purposes. The application shows a social network that is derived from properties of emails (sender, recipients, time). Changes can be tracked over time. Strength of relations in a certain time span is made up by the format of the addressing a person (directly, through lists, in blind copy field). In addition, the predominant role the communication took part in is shown. It is either derived from different sender addresses (home/work) or explicitly stated.
The results are viewable as a network diagram that changes over time. It allows to reflect past email communication.

Reconstruction of social networks is also the basis for Kautz's et al. (1997) ReferralWeb. While also including mail archives to the scope of applications they mainly rely on web documents. Co-occurrences of names in web documents are identified to construct a social network graph. This network complements a search engine by limiting search results to areas of the network such as documents close to node X or anything within the range of 4 nodes from the investigator (cf. [Kautz1997])

Baier et al. (2004) propose global identity management to track a person's interaction on the web and correlate people accordingly. They propose a client called "CoInternet" that is designed for social navigation. The user may choose different profiles for surfing depending on tasks or situations. Websites may be annotated and rated. User profiles are compared to predict the relevance of unvisited websites. CoInternet allows to reindex Google search results this way (cf. [Baier2004]).

Social Web Services

The following projects are less scientific. Yet they are zeitgeisty applications that inspired this work and are frequently mentioned in the next chapters. They all feature social collaboration through tags.

del.icio.us (del.icio.us) is a social bookmark manager. URLs are annotated with title and description and - more important - a list of keywords deemed helpful for personal archival. The archive can be queried by single tags or combinations.
The system is social because it allows to watch other peoples bookmarks. For each URL in the system it can be tracked how often, by whom and under which keywords it has been archived. Results from a query to the system may be exported by standardized formats (e.g. RSS and Atom) or through a documented API for reuse. Several applications arose that use del.icio.us data for visualization, evaluation, search, etc. (see below).

flickr (flickr) utilizes tagging for photos. Additionally, several annotation methods are available. Shared pictures are viewable by everyone or defined groups. Tags and annotations may also be assigned by authorized members (friends). By this, they can note their spontaneous associations when watching the pictures (just as watching real world photo album).
Photos from your own archive or from all members are accessible by tags so there is no need for folders or albums. Reuse is possible by standardized formats and an API as well. This way flickr may be used as a huge picture archive where people benefit from each other. Interesting applications like "flickrgraph" arose which show networks of flickr members and their pictures.

A third and last service that shall be mentioned here is technorati (technorati.com). It is a search engine for weblogs. Weblog applications ping technorati (and other similar services) whenever an entry is published.
Technorati indexes the content of each post to allow research on certain topics in the blogosphere. Moreover, links in the posts as well as trackbacks (explicit pointer to related posts) are evaluated. This way people can check on technorati how often an entry was referenced and by whom. Finally, technorati introduced a syntax to define tags in XHTML. Such tags are also indexed so that any weblog entry is accessible by tags or a combination of tags.
Like the services mentioned before technorati delivers any results in standardized export formats.

Projects based on Social Web Services

Finally we want to mention two projects that rely on social web services as well.

"Ultra Gleeper" is a recommender system that uses input resources of social web service accounts and claims weblogs as a starting point to crawl web pages within four links of each resource. Potentially, any crawled document is a recommendation. They are evaluated similarly to Google's page rank (cf. [Doctorow2001]) taking into account any incoming and outgoing links. Resources with many outgoing links are considered valuable. Many incoming links from other valuable pages in turn show that a page is well-regarded (cf. [Richardson2005]).

A search engine based on tagged URLs is proposed by Michal (2005). "CollaborativeRank" searches del.icio.us resources by tags that are provided just as a search term for an ordinary search.
The search engine records searches performed and values del.icio.us users who early provided frequently used tag combinations for a particular item. The higher valued a user is the higher will her bookmarks be ranked in the search results. This greater influence shall be an incentive for useful suggestions of tags. In addition to a search interface a list of the top 500 de.icio.us users regarding their importance in the system is viewable (cf. Michal2005).

1.3 Outline

This thesis is structured into five chapters. Following this introduction Chapter 2 provides the theoretical background for further elaborations. Specifically, Section 2.1 defines the term metadata with an emphasis on collaborative categorization of resources. Concluding that metadata has to be interpreted in the context it was created in, Section 2.2 examines individual and shared context. Section 2.3 relates shared contexts to groups in social networks and introduces the concept of foci. Finally, in Section 2.4 the field of social navigation is introduced which leverages findings from social network evaluation for helpful visualization of a social environment and computation of recommendations.
Chapter 3 relates our theoretical findings to a concrete proposal of a framework for social navigation based on tagging resources. After introducing a metaphor describing the framework in Section 3.1 we will identify the main steps of the framework in Section 3.2. Section 3.3 describes each of the four steps - structuring of individual clusters, mapping individual clusters to foci, building a social network and calculation recommendations and feedback - in detail. We conclude requirements for an implementation of the framework in Section 3.4.
The practical implementation of a prototype realizing the framework is subject of Chapter 4. We will illustrate the design of its architecture and interfaces in Section 4.1, followed by implementation details and a documentation of the functionality and interfaces in Section 4.2. Initial results of an evaluation of the prototype will be presented in the final Section 4.3.
Chapter 5 is a conclusion of our work with a final outlook.

2. Foundations

In this chapter we will discuss fundamental concepts regarding our topic: groups in social software for social navigation. We will start looking at metadata creation in Section 2.1 and more specifically at collaborative categorization of resources with freely chosen keywords which we will introduce as a ”folksonomy”. These folksonomies do not only offer ways for meaningful information filtering and retrieval, but also for integrating individual backgrounds - or contexts - to shared understandings about a topic.

We will discuss the importance of contexts in Section 2.2 in detail, defining individual and shared contexts and their correlations.

In Section 2.3 we will have a closer look at preconditions for creating shared contexts: groups in social networks. After introducing foundations of network analysis we will explain relations within and between groups, based on the concept of foci.

Finally, Section 2.4 introduces the field of social navigation. The main purposes is to derive recommendations by transferring actions of the individual to people with the same shared contexts. Additionally, concepts to visualize processes within social networks are examined to foster an understanding of the social environment the interaction is happening in.

2.1 Metadata

Metadata is information about physical or digital resources (like books, documents, images, etc.). It is often distinguished between descriptive, structural and administrative metadata [NISO2004]. Descriptive metadata describes resources to foster discovery and identification. Examples are title, author, abstract and keywords. Structural metadata relates resources, for example to combine chapters to a book. Administrative metadata supports the management of resources by providing information like creation date, file type or who is allowed to access a resource. In short: metadata is data about data.

2.1.1 Metadata Creation

Some metadata can be generated by machines automatically. Examples are creation date and information about file formats. Feature extraction techniques like shape or color detection for images and videos or automated text categorization may even derive basic low level semantic information. But quite often the results are insufficient and high level semantics (like abstracts or summaries) cannot be generated reliably today.
Semantic metadata of higher quality has to be created manually. Mathes (2004) differentiates between professional created, author created and user created metadata [Mathes2004].
Librarians create metadata on a professional basis by following strict rules and categorization schemes, for example the Library of Congress Classification Scheme. Producing high quality metadata is a non-trivial task only applicable by well-trained people.
A different approach is author created metadata. The Dublin Core Metadata Initiative provides guidelines for a standardized and common understandable annotation of resources for content

creators [Hillmann2003]. But there are several issues with author created Metadata, like a lack of control of quality and correctness and an additional workload for creators.
Thirdly, user created metadata is based on implicit and explicit statements about resources by their recipients. Examples for implicit statements are reading or purchasing habits (see 2.4.4.1), explicit statements are ratings, classifications or comments.

2.1.2 Comparing Taxonomies and Folksonomies

In the following we will focus on user created metadata. From an economic point of view user created metadata is the only realistic option for large-scale environments like the World Wide Web. But instead of looking at metadata on the whole we will concentrate on one specific aspect: categorization of resources. Categorization metadata is of special interest because it is not only descriptive - a category is a semantic description - but also structural relating similar categorized resources to one another.

The "traditional" way of categorizing resources is usually based on a predefined vocabulary of categories which is often hierarchically structured in a taxonomy. Large and consistent taxonomies have to be maintained and controlled by experts.

A completely different approach is the tagging of resources with freely chosen keywords, often called tags. If this categorization process is done collaboratively by different users, it is often referred to as "folksonomy", as coined by Thomas Vander Wal [Sterling2005]. Other commonly used terms are "grassroots classification" [Mathes2004], "ethnoclassification" [Merholz2004] or "social classification" [Hammond2005].
The number of people participating in the categorization of one resource defines the dimension of a folksonomy. In narrow folksonomies one resource is only tagged by few people. This is shown by the photo sharing service called flickr.com (cf. 1.2) where a photo can only be tagged by its creator and her friends.
Broad folksonomies are characterized by many people tagging one resource. For example the bookmark manager del.icio.us (cf. 1.2) allows anybody to tag any web resource. Broad folksonomies are particularly interesting. The distribution of different tags of a frequently tagged resource shows a power law curve [VanderWal2005]. High spikes show popular tags, the "long tail" of the curve shows infrequently used tags, see Figure 2.1. We will refer to both phenomena in later paragraphs.

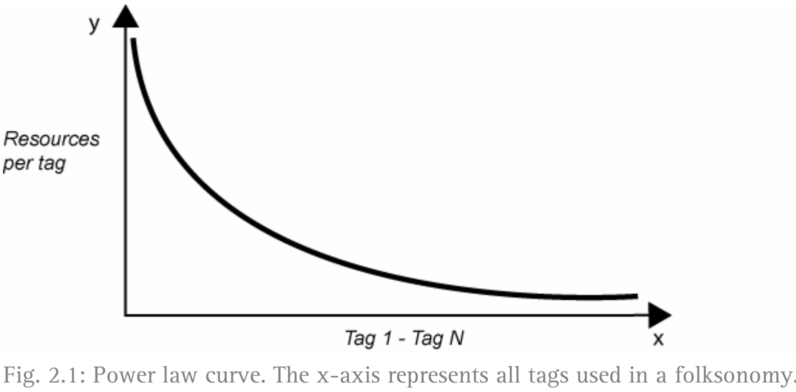

Fig. 2.1: Power law curve. The x-axis represents all tags used in a folksonomy. On the y-axis the number of resources tagged with the tag on the x-axis are shown. Only a few tags are used for many resources. Quite a few tags are only used for a small number of resources.

Comparing the taxonomy and folksonomy approach which concept is preferable for non-professional collaborative categorization of web resources? Following Smith (2005) this question can be addressed from different perspectives: retrieval, quality, usability, authority, scalability and social implications.

Retrieval of resources through categories is a subjective process depending on a person's background and her intentions. A strict taxonomy forces a person "to view the world in potentially unfamiliar ways" [Merholz2004], namely with the eyes of the taxonomy's creator. Nevertheless a taxonomy's hierarchical structure offers support for searching and browsing. In this respect folksonomies offer only marginal assistance. It is more a kind of serendipitous form of browsing [Rosenfeld2005].
When categorizing a resource, taxonomies have to face many drawbacks compared to folksonomies. Even well thought-out taxonomies miss categories or a desired category is not found at the expected place in the hierarchy. A result may be an inflating use of categories like "miscellaneous" or "other", which are only marginal descriptive.
Folksonomies allow a person to choose a keyword freely for describing a resource in a way that makes sense to her. Instead of enforcing a controlled vocabulary a "vocabulary of users" [Mathes2004] evolves. This may lead to categories a professional taxonomy designer never would have thought of (like the tags "cameraphone", see www.flickr.com/photos/tags/cameraphone or "whatsinyourbag", see http://www.flickr.com/photos/tags/whatsinyourbag).
The concept of using personal tags for accessing resources shows strong correspondence to what Wegner (1987) calls a person's "individual memory system" in which personal labels are used for storing and accessing information (cf. 2.2.2 and [Wegner1987], p.186). In this sense individual tags can be seen as a direct representation of a person's labels to external resources. But the connection between tag and resource is probably not meaningful for other people.

The ambiguity of individual tags or labels is directly related to the **quality** of categories. Tags may be interpreted differently from the creator's original intention outside of the creator's background. Reasons could be cultural differences, different knowledge domains or differing degrees of knowledge. Or as Lawley (2005) puts it: "Describing things well is hard, and often context-specific" [Lawley2005].
Furthermore tags can be completely personal (like "toread") or only meaningful for people knowing the creator's background. Unlike taxonomies there is no hierarchy that gives information about superior categories, for example to distinguish synonyms (think of the super category "fruits" or "computer" for the category "apple"). And finally tags may simply be wrong. An often mentioned example is the tag "archeology" for a resource about dinosaurs.
Taxonomies are most often maintained by professionals to ensure correctness, accuracy and unambiguity with the costs of excluding any information that cannot be represented by a predetermined vocabulary. The fixed vocabulary of taxonomies cuts of the long tail of folksonomies [Porter2005] by limiting the way of expressing oneself to the common categories of the given taxonomy.

The possibility of choosing a free text description for tagging a resource has also advantages from a **usability** perspective. People do not have to browse through long keyword lists to look for matching categories or sub-categories. The common way to tag a resource is to type all keywords separated by space into one input field without any constraints like considering predefined categories.
Tagging resources with keywords is a common practice, not only in folksonomies. The difference to previously existing systems is that after assigning a tag to a resource other resources labelled with

the same tag are displayed - this way implementing a tight feedback loop [Udell2004]. If your resource does not fit in a cluster thematically you will probably change the assigned tags. Mathes calls this process "asymmetrical communication between users through metadata" [Mathes2004]. He states that "groups of users do not have to agree on a hierarchy of tags or detailed taxonomy, they only need to agree, in a general sense, on the "meaning" of a tag enough to label similar material [...] for there to be cooperation and shared value" [Mathes2004].
The ongoing success of folksonomies originates largely from the easy creation process of tags and the direct feedback of related items.

Creation of tags relates to the issue of **authority**: who controls the categorization process? According to Boyd (2005) the question "comes down to benevolent dictator vs. crowd behavior" [Boyd2005]. The structure of a taxonomy is enforced by few but with the goal to achieve the best result for all later users. In folksonomies the individual is self responsible for creating categories for a resource but probably without having in mind all other groups of potential beneficiaries. The tight feedback loop may support the mediation of meaning of a tags, but there is a risk that popular tags will dominate less frequently used tags. Or more abstract: popular ideas and concepts will dominate the less popular ones. Without any precautions majorities may suppress minorities. Boyd calls this the homogenization of the crowd [Boyd2005].
When talking about giving authority to general public, one also has to talk about possibilities of systematic misuse. Popular tags could be exploited to deliver spam to people who look for specific tagged resources, as shown by Shirky [Shirky2005a]. Furthermore people could tag offensive material without hints to its potential insulting character, as discussed by Powers [Powers2005].

Problems like spam and offensive material will increase in larger folksonomies. But these are not the only issues where **scalability** has to be taken into account. Folksonomies seem to attract more and more people. One reason is the opportunity of information discovery based on a shared understanding of a meaning of a tag. But this meaning grows within the community and is not explicitly available. Is this implicit negotiation process still possible in large communities? More precisely: Are tags on their own sufficient for information retrieval in large-scale environments? Or like Shelly Powers asks: "can cheap semantics scale" [Powers2005]?
To answer the question a look at the semantics of taxonomies is helpful. In a professional taxonomy semantic meaning of a classification entry should be well defined. The relation between different entries - like an apple is a kind of fruit - can be formalized in the form of an ontology. Ideally these relations are written down in a machine understandable way, allowing semantic driven information requests. Explicit semantics make concepts behind categories visible. Not only for a small community with a shared understanding about a topic, but for a wide range of recipients - as long as these concepts are correct on a global level and do not change over time (a cat is a mammal and will be a mammal tomorrow). But people may have differing concepts for the same thing and concepts are quite often a subject of change (think of the concept of a state, like "East Germany"). Accordingly taxonomies have to be changed continuously and - even more problematic - respect differing concepts of groups and individuals.
In contrast, folksonomies are dynamic, new tags and concepts can be introduced by anyone. Shared value arises if others adopt newly introduced tags. But this presumes that the implicit concept behind a new tag is understood by its adaptors (the tag "whatsinyourbag" is an example for this. Someone took a picture with the contents of her bag and tagged it with "whatsinyourbag". Many others followed, see www.flickr.com/photos/tags/whatsinyourbag). But how to ensure that an implicit concept is understood by many? One approach could be to use folksonomies as a basis to evolve more formal taxonomies by evaluating popularity and relations of tags but without solving the problem of respecting differing concepts of different groups.

A different and in our opinion more realistic approach would be to strengthen the community structure that is the foundation of a folksonomy.

Hence it is necessary to look at the **social implications** of folksonomies. Tags form communities. Or as Weinberger (2005) states: "Find people who tag items the same way as you do and you've now found a social group based not around shared interests but around shared ways of thinking and shared ways of speaking" [Weinberger2005]. The challenge is to support and mediate the integration of individual contexts into group contexts. Once groups with a shared understanding are formed formal semantics are not necessarily needed, because "all of us can describe things easily understood by ourselves or our immediate social groups" [Powers2005].

2.1.3 Summary

At the moment manually created metadata seems to be the only realistic solution to get high quality descriptive information about resources in wide-scaled networks. With increasing information on the World Wide Web metadata cannot be exclusively created by professionals without enormous costs. Consequently, information recipients have to participate in information annotation, or as Shirky (2005b) states: "the mass amateurization of cataloging is a forced move" [Shirky2005b].
Focussing on the categorization of information two options are to mention: hierarchical structuring on the basis of a taxonomy or collaborative tagging in the form of a folksonomy.

Taxonomies offer support for hierarchical information browsing and state explicitly - in conjunction with ontologies - the semantic meaning of a category. On the other hand they are expensive to maintain, inflexible for changes and most important difficult to apply and limiting for many people.

Folksonomies are easy to use - adding freely chosen keywords can be done by anyone. Not only related information, but also people with a shared understanding can be identified. Not surprisingly folksonomies become more and more popular and are also the choice for our further elaborations. But it has to be kept in mind that tags may be interpreted differently outside of their creator's context. Developing concepts for identifying shared contexts within social networks is the challenge we address in this thesis. But before analyzing social networks in detail we have to examine the term "context".

2.2 Context

We argued previously that information has to be interpreted in respect to the context it is shaped in. But what is associated with the notion of context? Referring to contemporary discussions regarding context and ubiquitous computing in HCI two major directions can be identified: a positivist and a phenomenological approach (cf. [Dourish2004], [Chalmers1999]).
In the sense of positivist theories "context is a form of information" ([Dourish2004], p.4) that can be identified and stored independently from any activity that happens "within" the context. That means context is defined by a number of characteristics of the environment in which activities take place. These characteristics can be captured and stored in parallel to encoding the specific activity.
In contrast, the phenomenological view describes context as the relation between activities and

environment. Context cannot be seen on its own, rather something is ”contextually relevant to some particular activity” ([Dourish2004], p. 5) or not. Context is dynamic and varies according to the particular action and the particular setting. ”Context arises from the activity” ([Dourish2004], p. 5), it is based on the interaction between people and their environment.
In contrast to the positivist perspective context cannot be seen independently from an individual’s activity. Context is not part of the setting, it is an outcome of people’s actions. ”In other words, context and content (or activity) cannot be separated” ([Dourish2004], p. 6).
From a phenomenological view a shared context has to be mutually identified based on interaction between different parties referring to their cultural and common-sense understandings of the world [Dourish2004].

We will subscribe to the inseparable view of context and content, hence taking up the phenomenological approach. We will provide three examples from the scientific disciplines of semiotics, psychology and sociology to describe context from different perspectives. The semiotical approach will focus on the understanding of information (or signs) in individual contexts. The examples from psychology and sociology will concentrate on the interaction between people to foster a shared context.

2.2.1 Semiotics

Referring to Eco (1976) ”semiotics is concerned with everything that can be taken as a sign” ([Eco1976] p. 7). This can be words, images, sounds, gestures, objects - anything as long as someone interprets it as standing for something. Or as Pierce (1931-74) stated: ”Nothing is a sign unless it is interpreted as a sign” ([Pierce1931-74], p.2306). In this paragraph we will concentrate on a further definition of signs, following the model of Pierce. The main motivation is to get an understanding about the process of interpreting signs in the context of the individual.

Pierce describes a sign with a triadic model. He calls the representation of a sign the *representamen*. A representamen ”is something which stands to somebody for something in some respect or capacity” ([Pierce1931-74], p. 2228). This interpretation in someone’s mind - the sense made of the sign - is named the *interpretant*. Finally, the sign refers to an *object*. Pierce calls the interaction between representamen, interpretant and object ”semiosis” or sign process. Note that the interpretant created in someone’s mind may serve as a representamen in a new sign process - looking up a word in a dictionary may lead to a synonym which, again, can be found in the dictionary. In the sense of Pierce, Eco refers to this as ”unlimited semiosis”.

The triadic model of representamen, interpretant and object was later represented in a triangular diagram, called ”the semiotic triangle” [Nöth2000]. Figure 2.2 shows additionally the currently widespread terminology, namely sign vehicle (representamen), sense (interpretant) and referent (object). The given example refers to the approach of tagging resources, whereas the tag ”Folksonomy” is a sign vehicle. The meaning or sense of this sign vehicle is ”one or more resources about the collaborative way of categorizing resources with freely chosen keywords”. The represented object may be a single document about ”Folksonomies” or may even refer to the whole class of documents related to this topic.

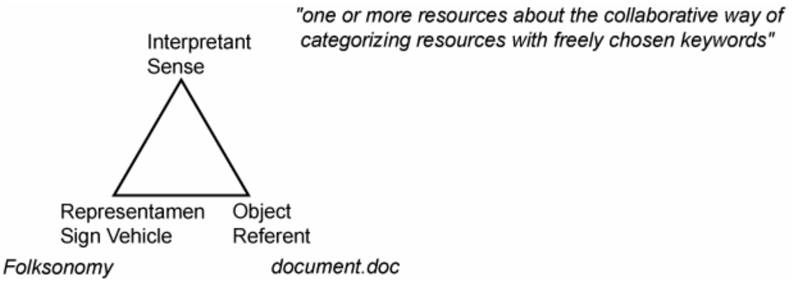

Figure 2.2: the semiotic triangle, cf. [Nöth2000]

The interpretation of a sign (semiosis) takes place on three levels ([Morris1938], [Nadin1988], [Chandler2002]): syntax, semantic and pragmatic (Fig. 2.3). Syntax describes the relation between signs and how signs are recognized. Semantic is the relation between representamen and object - understanding the intended meaning of a sign. The relation between the sign (on the whole) and the user (or interpreter) is called pragmatic. It describes the subjective interpretation process of a sign through an individual in terms of relevance, agreement, etc.

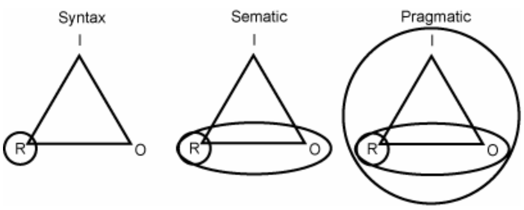

Figure 2.3: syntax, semantic and pragmatic, cf. [Nadin1988]

In this sense, an individual context refers to the notion of pragmatic. Figure 2.4 depicts differing individual contexts. Two persons use two different sign vehicles (the terms "Folksonomy" and "Ethnoclassification") for describing the same thing. Consequently, the two persons will not understand each other when talking about the same object, as long as there is no agreement (for example in the sense of defining a synonym) on the terms used. Problems will also arise if people use the same sign vehicle to refer to different meanings. A shared context, which can be defined as a shared understanding about a sign for an object, can only be achieved on the pragmatic level. That is, the relation between sign vehicle, meaning and object is interpreted similar by all persons of one context.

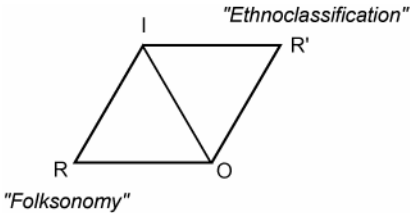

Figure 2.4: differing vocabulary

2.2.2 Psychology - Transactive Memories

Psychology addresses behavior, mind and thought of the individual. It can give us a clue about a person's very own context. However, individual contexts are influenced by one's social environment. This interplay of individuals and its implications on information processing in groups were examined by Wegner (1987) in his theory of "Transactive Memories". Wegner proposes that social interaction is crucial for developing something like a shared memory [Wegner1987].

Wegner starts arguing by describing how humans memorize things in three steps. At first information enters memory in an encoding stage. It persists in memory in a storage stage and is retrieved from memory in a retrieval stage. He stresses that one or more labels are developed in the process of encoding and storing of memory items that are used in the retrieval process. Furthermore, Wegner differentiates internal and external memory. It is not effective for people to store everything in their internal memory. Instead much information is stored externally and can be accessed by an internal label. Therefore, the retrieval process actually includes memory items, corresponding labels *and* a location - that is the internal memory or external resources like a book or a computer system. Moreover, other people may serve as location of an individual's external memory. In this case one person knows labels to retrieve information from another person's memory. Wegner coined the notion of "Transactive Memories" for this, which he defines as "individual memory systems taken together with the communication that takes place between those individuals" ([Wegner1987], p.186).
Transactive Memory Systems preferably occur in tightly knit networks and groups like married couples or workgroups because a high level of communication and shared experience is required to agree on who stores what and how something is retrieved. The latter is done by group labels which are "common denominators" ([Wegner1987], p. 195) that are developed in the process of "transactive encoding" ([Wegner1987], p. 195): similarly to the encoding stage in an individual's information processing, information that approaches a group (for example in a group meeting) is discussed and members agree on a location of storage. This happens either explicitly - as a result of the discussion - or implicitly based on experience of each other's expertise. Keywords developed and used in this process, which may have no meaning for somebody not involved in the process, serve as labels [Wegner1987].

Of course, there is a high benefit for the individual from a Transactive Memory System. A person is able to store and use much more information than her own capacity would allow. Furthermore, groups that realize such a system are more powerful and may reach their goals more efficiently. But as already mentioned, Wegner stresses that communication and deep knowledge about each other's domains of expertise is required for a Transactive Memory System to arise and work properly, as referred to by Kanfer. "Members of a group need time to learn about each other's knowledge by communication and observation of who talks to whom" [Kanfer1997]. Consequently, groups with a grown information infrastructure are superior to new established groups with a similar but only individually organized memory.
Accordingly, a Transactive Memory System is not applicable from outside, neither by selecting specific people for a group nor by forcing an information structure on a group. "[...] It is a property of a group" ([Wegner1987], p. 191) and when a group splits up its Transactive Memory System will be lost.

Concluding this example, we can infer that an individual context is made up by a person's own memory system. That is, the processes of encoding, storing and retrieving items and especially labels are highly individual. This corresponds to our hypothesis that assigning metadata in terms of keywords is effective when people can choose them freely to perfectly match their individual context and therefore experience direct value in the retrieval stage.
But keywords are supposed to be valuable for other people as well. This is not provided by individual labels. Instead, they may have no meaning to others at all. According to Wegner, whenever people want to organize their information together a shared concept of storage and shared labels has to be developed. A social context both influences this procedure by providing the environment for communication and observation and is in return influenced resulting in a

Transactive Memory System. We will refer to a social context that realizes a Transactive Memory System as a shared context later.

But how can such a shared context be developed? The formation of shared labels was mentioned more than once as a crucial factor. A last question that shall be addressed here is whether publication of individual labels in a shared context may boost the creation of a Transactive Memory System. Wegner discusses this in a comparison of Transactive Memory Systems to computer networks (see [Wegner1995]) as a part of "directory updating" which are processes whereby "people learn what others are likely to know about" [Wegner1995]. Directories in Wegner's metaphor ideally provide a full index to the memories of all other individuals. As this is not possible in reality he discusses strategies to update directories based on observations and communications in a social context as proposed in his original theory. The question can be raised whether publication of individual keywords as an index to externally stored resources - or memories - in a computer and network based system may be helpful to a faster clarification of each other's expertise.

2.2.3 Sociology

We learned that social interaction, which is the basic issue of sociology, is required to establish a Transactive Memory System. Sociologists like Harper (1999) are interested in implications a given social context (that is any formation of human beings and their interaction: kinship, workplace, societies, etc.) has on information processing. Precisely, Harper investigated how an organizational context in terms of motivation and relevance determines accessing, judgement and further processing of information (cf. [Harper1999]).
Based on fieldwork, attendance and survey of economists whose task is to evaluate the economic situation in countries their company is active in, he came up with two major findings. First, information is evaluated regarding its relevance for the context in question only. In an example Harper describes a desk officer who regularly reads an official party newspaper of a totalitarian state. The officer judges the validity of information by its origin, saying that facts from the paper are not objective and therefore worthless for his company. However, it is still valuable as it helps him - again in favor of his company - to judge the overall contemporary mood in the country. Another finding Harper points out is the relevance of information for *all* parties involved. He found out that an economist did not consider an analysis of the macro-economic situation in a country performed by another institute because the authorities of that country did not take the information on board [Harper1999]. Though the analysis was objectively relevant its relevance for the given social context was diminished because "what mattered was whether the information was in sight of the authorities, not whether it ought to be or was naturally relevant" ([Harper1999], p. 348).

Concluding Harper's findings, a person has to judge relevance of a piece of information in light of her individual context which is influenced by a given social context (such as a company). Moreover, an individual must consider whether other parties of a given social context include information into their field of relevance to judge its overall relevance ([Harper1999], p. 349). For a shared context to emerge it is a precondition that such overall relevance of each piece of information is conceived by all individuals involved.

The requirements identified by Harper can easily be transferred to information processing on the World Wide Web. Whenever information is retrieved literate people will ask questions about it like "where is this coming from?", "who reviewed the article?", "what is the expertise of the reviewer?",

etc. By finding out the social context of information they determine how to judge a resource in light of their individual context, that is their motivation or their goal. Furthermore, people often try to find out the relevance of a piece of information for a community (for example by using Google which calculates relevance for the community of *all* web users, [Doctorow2001]) thereby assessing whether some information is already accepted by a relevant community and therefore may be dealt with differently than with a new or less considered resource ([Harper1999], p. 349).
The outcome of this is a basic requirement for web applications that support information processing on the web: they have to convey the social context of a piece of information. Crucial is that people with a similar motivation or task (such as workgroups, communities of interest) know each others experience so that a shared context can be established. Therefore technology must ensure "that when information is used by any one individual, this can be indicated to other parties in the relevant institutional process" [Harper1999, p.353]

2.2.4 Context - Conclusion

People interpret information based on the background of their individual contexts. We examined the process of formation of a person's individual context from the perspectives of semiotics and psychology. From a semiotical point of view individual context refers to the process of semiosis - the interpretation of signs in a person's mind.
Wegner stressed that information is stored and retrieved from internal and external memory in a highly individual manner which allows us to explain context from a psychologist's viewpoint. According to Wegner people use personal labels to access information stored internally in their memory or in external resources. We will refer to this individual memory system as individual context.

Interaction between a person and its environment - particularly communication between people - determines whether information processing is effective and reliable for both the individual and the group.
We could observe that under certain circumstances positive effects in groups arise. Examples are Wegner's Transactive Memory Systems - describing the development of shared labels to offer information of individual memory systems to groups - or Harper's research about common judgement of relevance of information. In our ongoing work we will refer to such positively loaded contexts as shared contexts.

2.3 Social Networks

Social networks emerge when people interact. Social relations within these networks may be based on friendship, kinship, ethnical background or shared activities at work or in leisure time. Financial aids, emotional support, supportive services or companionship are only a few options to show how the individual may benefit from social networks [Wellmann1990].
Traditionally, sociologists analyze community structures in areas like work, school, neighborhood or in respect to social affiliation. In the last few years a new focus for social network analysis arose: online communities in computer networks (see for example [Wellmann1996]).

In the following, we will look at social network analysis. Referring to Kanfer (2000) social network analysis "provides a quantitative approach to analyzing social actors and patterns of relations among them" ([Kanfer2000], p.26). We will start with a quantitative description of social networks, trying to identify basic properties and recurrent patterns based on statistical phenomena. Taking these quantitative and abstract elaborations as the grounding, we will afterwards focus on a qualitative description of social network theories related to real world observations.

2.3.1 Patterns in Social Networks

In social network analysis, networks are often described as graphs, where each member of the network is represented by a node. Relations between members are shown by edges or ties. A relation between two persons is asymmetric if only one person relates to the other. If both persons have a mutual relation it is called symmetric [Feld1997]. If symmetric and asymmetric relations are distinguished the resulting graph is a directed one.
Referring to online communities an example for an asymmetric relation is when one person is reading somebody else's weblog without leaving any hints of her presence. The relation becomes symmetric if the reader comments articles to start a discussion with the weblog's author.
Based on the observation of ties between individuals social networks can be characterized in terms of several properties or patterns like density, centrality ([Freeman1977], [Borgatti1997], [Palau2004]), community structure [Girvan2002], bridges and transitivity ([Granovetter1982], [Girvan2002]).

The largest possible number of ties in a network is (K-1) * K, where K is the number of all of the network's members (i.e. its size, [Borgatti1997, [Palau2004]). The **density** of a network is the proportion of all possible ties compared to those actually present. In networks with low density only few connections between its members exist, in contrast to high density networks where many members are connected to each other.

Network **centrality** can be measured in various ways. Examples are degree centrality, closeness centrality or betweenness centrality ([Freeman1977], [Borgatti1997], [Palau2004]). *Degree centrality* is defined as the number of ties a member of a social network has. In directed graphs degree centrality can be refined in in-degree and out-degree. The in-degree counts the number of other people relating to a member. A high in-degree value of a person is often interpreted as high prestige, because the person's opinion is taken into account by many. The out-degree reflects the number of ties a person has to other persons. Having a high number of relations to others offers more alternatives for support and reduces dependency. Furthermore, the opportunities to actively influence others are higher.
Closeness centrality does not only take into account the direct neighbors of a person. Instead, all shortest paths (called geodesics) to all other members are considered. The closeness centrality is inversely proportional to the sum of all shortest paths from one person to all others.
Betweenness centrality counts all shortest paths that go through a specific node of a graph in relation to all alternative paths for each of these shortest paths. High betweenness centrality of a person shows a high dependency of many on this node to connect to others.

Summing up, the different measurements of network centrality allows to analyze the distribution of power within a social network. The more central the position of a member is, the more other people may be influenced by this person or have to rely on her.

Many social networks show unequally distributed ties. In this case some nodes may be more central than others. Another explanation is the formation of groups within a network. A group is a number of tightly knitted nodes with only a few loose connections to other groups. The degree of group formation within social networks is called **community structure** [Girvan2002]. The process of detecting community structure is also often referred to as clustering. Clustering approaches can be based on identifying strongly related groups of nodes within networks. A different approach is based on the detection of least central ties that lie ”between” groups. Girvan refers to it as ”edge betweenness” [Grivan2002], others call a connecting tie between groups a **bridge** [Granovetter1982]. A more detailed overview of clustering algorithms is given in [Jain1999].

Network **transitivity** describes the phenomenon that two nodes connected to a third one are also quite often related to each other [Girvan2002]. Or as Granovetter states: ”[transitivity is] the tendency of one's friends' friends to be one's friends as well” ([Granovetter1982], p. 218). Transitivity is found in very dense networks.

The properties and patterns described allow statistical analysis and comparison of social networks. In addition, they provide the basis for developing more profound theories to explain social phenomena in communities.

2.3.2 Social Network Theories

Social network analysis shows patterns among social relationships, like transitivity, centrality or network bridges. But these patterns have to be interpreted in a larger social context. The following paragraphs offer deeper theoretical interpretation and consolidation of these patterns with one main goal: to explain groups in social networks. This includes relations of individuals within groups and relations between single group members of different groups. The most important concepts we will rely on are Granovetter's theory of the strength of weak ties [Granovetter1982] and Feld's focus theory ([Feld1981],[Feld1997]) described in detail in the next sections.

2.3.2.1 Refining Ties

Before we talk about groups the definition of ties has to be refined. Ties are not only symmetric or asymmetric, they can also be characterized according to their ”strength”. Granovetter distinguishes between weak and strong ties [Granovetter1982]. He describes weak ties informally as acquaintances and strong ties as close friends. Based on empirical data Wellman (1990) defines the strength of a tie more detailed with three characteristics: socially close intimacy, voluntary interaction and participation of both members in multiple social contexts [Wellman1990]. We will adopt this view of a tie's strength, abbreviating the three characteristics with the terms ”friendship”, ”interaction” and ”multiplexity” (an explanation for the term will follow in the next paragraphs). Furthermore we consider a tie between two persons as weak if the sum of all three characteristics is lower than a specific threshold, otherwise the tie is strong.
Considering weak and strong ties the concept of transitivity has to be narrowed. Transitivity can only be expected from strong ties. If two persons are only weakly connected with a third, one cannot safely assume that the two know each other [Granovetter1982].

2.3.2.2 Relations within Groups: Defining Foci

We have already mentioned groups in social networks. Analyzing a social network, a community structure (see 2.3.1) can be identified: strongly related nodes with high density and only few connections to other clusters of nodes. But how can this pattern be theoretically supported?
Feld's focus theory concentrates on the explanation of social circles within social networks [Feld1981]. For this reason Feld introduces the term "focus". He defines a focus as "a social, psychological, legal or physical entity around which joint activities are organized (e.g. workplaces, voluntary organizations, hangouts, families, etc.)" ([Feld1981], p. 1016). We complement this definition with any kind of joint activity around a topic in online networks.
A focus brings people together actively (like a sports club) or may act as a passive constraint (like neighborhoods). In both cases foci support interaction between people. Persons who mutually interact within a focus become symmetrically tied and form a cluster. Such a cluster is called a group. Foci foster the formation of groups, because the probability of interaction between two persons around a joint activity is much higher than between two individuals without a shared focus [Feld1981]. Taking together various different and loosely related foci (with one or more groups within a focus) shows a "community structure"-like pattern.

Feld states that "a social context can be seen as consisting of a number of different foci and individuals, where each individual is related to some foci and not to others" ([Feld1981], p. 1016). Consequently a person may be part of several groups.
A person's social context cannot easily be expressed in one single network graph without loosing information about the individual strength of a tie within a single focus. The reason for this is that if two persons share more than one focus together (this is called "multiplexity" [Feld1981], p. 1025) the strength of their relation will always be the average of all their shared ties. Hence, we will evaluate ties not only in the context of one overall social network, but additionally in the context of foci.

Relations within a focus are considered to show a high degree of transitivity. If two persons are connected to a third one and share a focus with this person, the probability is high that they both are also tied together. It gets even higher if all three share more than one focus. Another characteristic of densely tied foci is how constraining they are. The more time a focus consumes and the more frequent and intensive interactions between members of a focus happen, the more ties within this focus are developed.
Summing up, transitivity directly relates to the number and the constrictions of foci [Feld1981].

If one person is a member of various different foci it can be assumed that many of her ties do not share the same foci. In this case the person might try to relate the unconnected acquaintances by organizing her activities in a new shared focus. This tendency can primarily be observed when foci are highly constraining - thus demand high efforts from the individual - but still similar enough to be joined (cf. [Feld1981], [Donath2004]).

2.3.2.3 Relations Between Groups: Weak Ties and Bridges

Although transitivity is considered to be high within foci not all ties have to be strong. Especially in less constraining foci, ties may only be weakly connected. Of special interest are weak ties between persons of two different foci. These weak ties "bridge social distance" ([Granovetter1982],

p.209). They become connections between otherwise unlinked social circles. Bridging weak ties allow information flow over large distances of a social network. Without weak ties one had to rely on information from one's own intimate friends - which is more likely to be already known. In this respect people with bridging weak ties - also called local bridges - are in an advanced position regarding information retrieval. For this reason Granovetter entitled his theory of weak ties the "strength of weak ties" [Granovetter1982]. In terms of the previously mentioned patterns of social networks people with local bridges have a high betweenness centrality.
A local bridge will retain its bridging character as long as the time and effort invested in this tie remains small. This is the case as long as the tie is weak. Accordingly people may be able to manage many weak ties. If the time investment gets too high - and the tie transforms from a weak to a strong relation - the bridging character vanishes, the two formerly bridged foci may merge. This effect may not be desirable. Weak ties allow diverse cultural groups to interact, but without loosing their cultural identity in a process of homogenization.

Just to clarify: one could get the impression that weak ties should be favored over strong ties. This is not necessarily the case. Strong ties are important for immediate help. They are easily available and provide strong assistance. Wellman showed that strong ties provide broader support than weak ties, including emotional aid, supportive services and companionship [Wellman1990]. Furthermore Granovetter quotes Weimann: "weak ties provide the bridges over which innovations cross boundaries of social groups; the decision making, however, is influenced mainly by the strong-ties network in each group" ([Granovetter1982], p.219).
Friedkin (1980) refers to the integrative character of regular transmissions between strong ties as microintegration [Friedkin1980]. Respectively, bridging weak ties that connect different social groups support macrointegration.
Summing up, strong ties are responsible for active support and communication within foci, whereas weak ties connect isolated foci with each other to ensure network-wide information flow.

2.3.3 Conclusion

In this chapter we introduced theories explaining the formation of groups, in particular the interaction between members of one group and members of different groups. Groups consist of strongly tied clusters of people with only a few weak ties to other groups. Thus, evaluating the strength of ties is essential for group identification. As defined before the strength of a tie is dependent on the number and kind of interaction between two people, the social context (or the foci) they share and the intimacy of their friendship.
According to Feld's focus theory a focus fosters the formation of groups as it supports interaction between people around a shared activity, topic or constraint. The more time consuming a focus is the more probable are strong ties.
Basically, data about social networks may be incomplete, accordingly not all present ties may be retrieved. Absent ties within a focus may be proposed basing on phenomena like transitivity.

Interaction between members of separate groups, especially of groups from different foci, is described in Granovetter's theory of "weak ties". Weak ties between foci - or local bridges - support communication between unrelated social circles. Thus, people acting as local bridges have an integrative role. Identifying and supporting the creation of local bridges is of special interest for our work.

2.4 Social Navigation

Human beings are social by nature. Consequently, we use social information - information about our social context - we gather through interaction and observation for finding information ([Dieberger2000], [Dourish1993], [Höök2003]). Relying on the possibility to ask passengers for the way instead of taking a map with you, following the flock when leaving a plane at the airport to find the baggage return, asking other people for a recommendation for a dentist - there are plenty of other examples from real life that show how much we rely on hints from our social context when navigating through and towards information [Höök2003].
All these hints make up a field of relevance for a particular information (cf. [Harper1999] and chapter 2.2.3). However, when information is retrieved from computer systems and through networks most applications neglect this valuable information. On the one hand social information from real life is not transferred to the medium, on the other hand applications do not provide means of creating and observing clues from other people or for other people respectively to evaluate relevance. Finding concepts and techniques to address this problem is the issue of social navigation.
For our purposes we will define the meaning and scope of social navigation. Based on an overview of tendencies, assessment factors and applications in the field we will identify two major areas, social awareness and recommender functionality which we consider most important for our work. They will be introduced more in-depth.

Social navigation was first introduced by Dourish and Chalmers (1994) who took "moving towards a cluster of other people, or selecting objects because others have been examining them" ([Dourish1994], p.1) as examples. Dieberger et al. (2000) give a detailed description of the field. They include direct traces left for other people such as bookmarks in their definition ([Dieberger2000], [Höök2003]). Höök et. al. (2003) conclude, "the concept of social navigation has broadened to include a large family of methods, artefacts and techniques that capture some aspects of navigation" ([Höök2003], p.4). Thus, it touches many areas of computer science, including Collaborative Virtual Environments (CVEs), Computer-Supported Cooperative Work (CSCW), Intelligent User Interfaces (IUIs) and Information Retrieval ([Dieberger1999], [Höök2003]). It is an interdisciplinary field that leverages findings and theories from social psychology, anthropology, sociology, artificial intelligence, social network analysis and more. No matter their viewpoint, researchers of social navigation consider people to be inside information space rather than outside of computer systems. Hence, their major goal is not to bridge the gap between humans and machine but to give people means to navigate information space just as they navigate in real world [Höök2003].

2.4.1 Direct and Indirect Social Navigation

Basically two types of social navigation are distinguished ([Dieberger1999] and [Höök2003]). Direct social navigation refers to synchronous interaction such as (textual) chatting or seeing other people in a virtual world. It is characterized by a direct intention of the sender to more or less guide known people. Indirect social navigation is provided through any information that is left "as a by-product of our activities" [Dieberger1999]. A list of authors a customer has ordered books from as well as a topic in a forum that is marked as "hot" because many people have looked at it are examples for indirect social navigation. Communication is rather asynchronous and people who guide each other often remain anonymous.

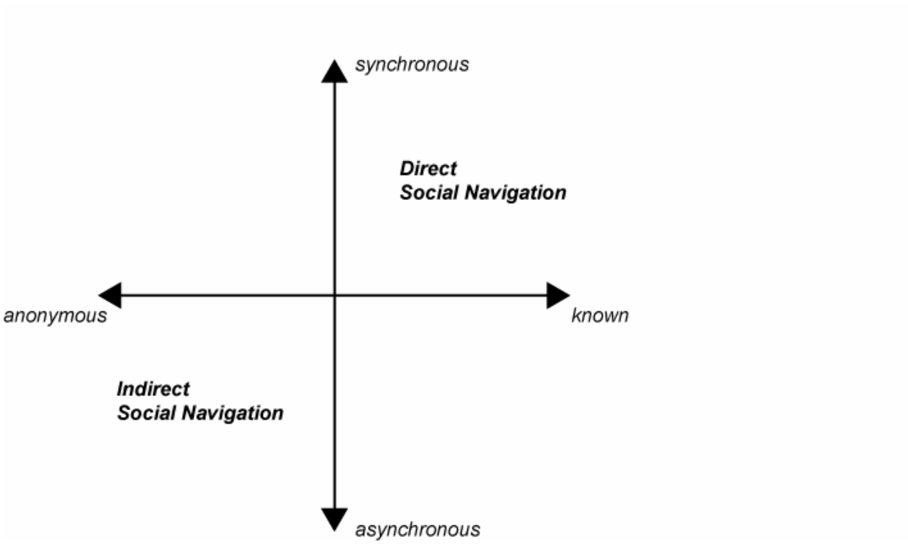

Figure 2.5: Characteristics of direct and indirect social navigation

Direct and indirect social navigation may not be seen as absolute characteristics of social navigation systems. They rather describe the degree of anonymity and the prevailing type of communication in a system or a part of a system. Figure 2.5 shows that the more synchronous communication is used and the better known the users are, the more direct the social navigation. The diagram shows that hybrid forms are possible as well. Furthermore, direct and indirect social navigation might be used complementary in a system.

2.4.2 Development of Social Navigation Research and Applications

Social navigation could be observed on the internet from the very beginning. In order to structure the inevitably unstructured World Wide Web people created private homepages and published tips and tricks as well as bookmark collections for themselves and their friends, colleagues, relatives and like-minded people. Pointers to interesting information were exchanged via email and newsgroups.
Since then, a shift towards automated recommender functionality as a key element of social navigation systems occurred in applications and in research. These recommender systems ”support choices based on the experience of other people” ([Dieberger1999], p.295f.). This is a shift from rather direct towards more indirect social navigation. In an often cited example from the World Wide Web, Amazon.com’s online shop, anonymous traces left by other customers are evaluated to calculate recommendations for further products (”people who bought this item also bought this item”). Other applications evaluate direct social navigation within email ([Kanfer1997], cf. Chapter 1.2) or newsgroups ([Resnick1994], cf. 1.2) to generate recommendations of email contacts for certain problems or interesting news articles.
Lately, systems have been developed that consider holistic approaches, that is to include both direct and indirect social navigation. Examples for such systems are ”Kalas”, a grocery store where buying food is the outcome of social interaction around food recipes resulting in recommendations for further recipes [Svensson2003] and ”Babble” [Erickson2003] which visualizes chat situations and enables retrieval of archived chats.

In order to assess the quality of social navigation within such systems it is important to define criteria. Dieberger et al. (2000) and Höök(2003a) propose to look at effects social navigation would ideally have when it is supported by a system. These effects are *filtering*, *quality*, *social affordance* and *usage reshapes functionality and structure*.

History enriched environments as well as calculated recommendations allow for **filtering** of relevant information from a huge amount that is available. To ensure quality of recommendations feedback mechanisms for evaluation are required.

We already discussed in chapter 2.2.3 that in some cases relevance of a piece - or the objective **quality** - of information cannot be judged by an individual alone but it is determined by usage of other people like the validation by experts or how often it has been cited by people involved in a field of research.

Social affordance refers to the property of an application to convey social interaction of fellow users and potentials for social interaction.
Being aware of others and their response to one's own actions has two important effects. On the one hand, it makes an application more alive allowing to feel more comfortable, getting support by others and feel encouraged to try functions that other people are using and one would not have touched otherwise. On the other hand, people can get an impression of appropriate behavior within a system and are particularly guide through this awareness.

One example for real world social navigation are paths that appear as a shortcut in a forest or meadow because people consider a certain way through it useful. These paths are an important clue for navigation for people who are not familiar with a location. Once a path is not used any more because a better shortcut was found, it vanishes and another path emerges somewhere else. Similarly, in information spaces action of each individual may shape the structure that is used by everybody to navigate. Browsing through recommendations instead of categories or by search terms is an example how **usage reshapes functionality and structure**.
In order to make people effectively shape a system it is important that they get an idea of the modalities how to influence it. The response of my own action (I leave footprints on the meadow), interaction (I tell others about my new shortcut) and other's actions (more footprints, finally the grass vanishes on the path) must be conveyed.

From these effects two main areas of interest can be identified (though distinctions are hard to make because there is overlapping). *Filtering* and *quality* are rather related to recommender functionality while *social affordance* and *usage reshapes functionality and structure* are more important for social awareness and interfaces of social navigation systems. Our distinction reflects the "shifting perspectives" identified by Höök et al. (2003). Information is not decontextualized anymore but enriched with social information and provided accordingly. This heavily relies on recommender functionality. Secondly, Höök et al. see a shift from space to place in how people perceive their information systems. While a space is designed and constructed by architects and designers having in mind requirements and opportunities for social interaction people will finally make the space a place through their usage and behavior. Or as Dieberger (1999) puts it: "we do live in space but we act in place" ([Dieberger1999], p.299). Here the quality of interfaces (or a space to stay in the metaphor) and the resulting social awareness decides how well people conceive opportunities for valuable interaction.

For our ongoing work we want to take a closer look at social awareness and recommender functionality. Interesting approaches and results from research and application shall provide a basis for discussion of our own approach to integrate these main aspects of social navigation.

2.4.3 Social Awareness

The internet is used by millions of people every day. When navigating through information of the World Wide Web, however, we do not meet many of them. Donath (1995) points out "first, wandering about the Web is a solitary pursuit: one is unaware of the presence of the many fellow explorers. Second, one cannot communicate directly to another person online within the rich context of the Web" [Donath1995, p. 3].

Some researchers explored ways to create social awareness by visualizing social interaction in information systems (including [Minar1999], ([Viégas2004] cf. 1.2), [Svensson2003] and [Erickson2003]). While all of them foster social affordance, some rely more on spatial metaphors such as "Visualizing the Crowds at a Website" [Minar1999] which shows how present and past visitors have moved around a website to indicate popular pages and paths through information space. Others rely more on visualization of artefacts left by others such as "Kalas" ([Svensson2003], see 2.4.2).
Indication of how *usage reshapes functionality and structure* - the other quality of social awareness - can be found in the examples as well, most often in new ways of navigation. Erickson and Kellogg (2003) developed an interesting approach called "social translucence" which mainly deals with this issue. We will present this approach which eventually lead to development of the "Babble" system [Erickson2003, see 2.4.2] more in-depth.

2.4.3.1 Social Translucence

Erickson and Kellogg (2003) coined the term "translucent" for displaying people and their actions within an information system. Instead of being transparent and therefore without privacy, people voluntarily admit some constraints which are vital to the process they are involved in. Basically, everybody is aware of how transparent she is to other users. In effect everybody knows that one is accountable for her action. Consequently, who and how many are watching determines what we say and do. Erickson and Kellogg clarify this valuable trade-off between privacy and visibility by an example [Erickson2003]: The authors of a forthcoming book discussed its structure in a special manner. For each chapter a physical location in a room was defined. The author's subchapters as paper copies could be moved around and chapters could be renamed. The fact that moving a subchapter to another chapter could be observed by other authors (who gathered around a specific chapter) forced the mover to explain or discuss her intention. Another constraint resulting from the physical distance between each chapter was that nobody could follow every action and listen to every discussion. This distracted authors from defending their own chapters because it was not possible to participate in the overall structuring process otherwise.
Two major conclusions have been drawn from this example. First, in offline face-to-face situation people are "aware of the existence and nature of the constraints" ([Erickson2003], p.21) which leads to norms of behavior such as asking or explaining yourself when you move a chapter. Secondly, each participant was aware of the other participants' awareness of constraints. In the example everybody knew that it is impossible to listen to every discussion. Therefore being insulted because somebody moved away your chapter while you were at the other end of the room was not

considered an appropriate behavior. [Erickson2003]
Erickson and Kellog argue that online interaction in contemporary systems misses many of the features observed offline. In order to prevent people from being "socially blind" [Erickson2003] social translucent systems will provide for.

- **Visibility**, that is visualization of as much information about participants and their actions as required.
- **Awareness** of each other's visibility and its consequences and
- **Accountability**, which is a consequence of the shared awareness of constraints because everybody knows that inappropriate behavior might be sanctioned.

Shirky (2003) illustrates that constraints are sometimes very subtle and therefore even harder to visualize. Groups develop patterns of behavior like "identification and vilification of external enemies" and "religious veneration" [Shirky2003]. If somebody is new to a newsgroup and discredits something that is held religious she will be identified as a common enemy and punished by a flamewar (numerous insulting messages). Such information about the group has to be made visible as well so that novices are aware of it and adjust their behavior if they still want to participate.

2.4.3.2 Social Proxies

In order to realize Social Translucent systems Erickson and Kellog propose a concept for visualization called a "social proxy" [Erickson2003]. It is a minimalist visualization of social situations online. Participants and settings are pictured in meaningfully shaped and colored figures. The relative position of each participant to the setting has a meaning as well. Fig. 2.6 shows a social proxy for a chat situation.

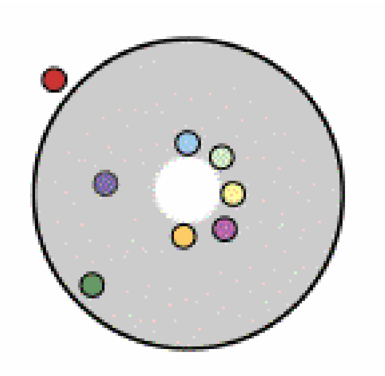

Figure 2.6: A social proxy visualizing a chat situation, cf. [Erickson2003]

Everybody participating is visualized as a dot within a circle (the chat room). The more active somebody is in the conversation, the closer her dot moves to the center. By this, you get an intuitive impression of who is only reading and who is really engaged in a discussion.

Social proxies have two more important features. First, they do not allow personal views to ensure a shared awareness of visibility only. Moreover, social proxies are always viewed in third person view so that every response to your own behavior and that of others can be observed to be able to pick up the social norms as quickly as possible [Erickson2003].

The chat proxy was leveraged in the ”Babble” system [Erickson2003]. The authors report that the social proxy was considered useful in giving the chat room a sense of place and awareness of who is really participating. Particularly the function of people (dots) moving closer to the center when they start typing was considered useful in order to instantly judge the response on a comment [Erickson2003].

The example shows that visualization is an effective means of creating social awareness. Still there is social information that is hard to convey as Shirky states (see above). It is grown from the interaction within the group and as a result of the identities of its participants. In this sense it is a property of the group. Visualizing a group’s emerging information space is also an example of how ”usage reshapes functionality and structure”. Shirky (2003) proposes to include reputation mechanisms and barriers to entry in online groups in order to protect a group’s reshaped functionality and structure against mischief and newbies. These mechanisms would allow the latter for example to become aware of the social norms of a group [Shirky2003]. Transferring this to social awareness, the question is how properties of groups and changes of these properties may be mediated by means of visualization.

2.4.4 Recommender Functionality

The challenge of recommender systems is to propose information to people that they individually consider most valuable. In order to identify this information a technique called collaborative filtering (CF) is used. Konstan and Riedl (2002) describe collaborative filtering as ”taking advantage of patterns of agreement and tastes between users” ([Konstan2002], p.44). Crucial is consistency of similarities over time so that predictions can be made. Thus collaborative filtering relies on the assumption that ”users who agreed with each other in the past are likely to agree with each other in the future” ([Konstan2002], p.44).
In order to discuss capabilities of collaborative filtering for our work we will describe the evolution towards its predominant form today, automated collaborative filtering. Afterwards some important issues for effectiveness of recommendations will be discussed, namely implicit measures of interest, situational recommendations, user control and transparency of the recommendation process (cf. [Konstan2002]).

2.4.4.1 From Manual to Automated Collaborative Filtering

Similar to early social navigation system described above, early forms of collaborative filtering can be seen as a reaction of the internet users to the vast quantity of information. Any communication technique (email, newsgroups, World Wide Web) is leveraged to publish evaluation of information. Everybody who has access to this information can search or browse it to find recommendations. As this is an active process this form of collaborative filtering is called ”pull-active” [Konstan2002]. For smaller communities pull-active collaborative filtering can be very effective because people are likely to know each other. Therefore they know what to expect and browse or formulate search queries accordingly. Moreover, the relevance of results can be judged better when the recommenders are known to the recipient. To implement pull-active collaborative filtering in large communities there has to be some basic structure and reputation mechanisms. Product

recommender websites such as ciao.com (www.ciao.com) and dooyoo.com (www.dooyoo.com) use pull-active collaborative filtering for huge communities by providing an edited structure for recommendations as well as information about the recommenders.
Another type of collaborative filtering could be observed in mailing lists since the early days of the internet. People send evaluations or - in the case of social navigation - pointers to a known community. Konstan and Riedl (2002) call this "push-active" collaborative filtering. It is a precondition that the recommender knows the recipients' taste. Therefore push-active CF only works in smaller communities. To apply push-active collaborative successfully filtering to large communities would require the recommender to specify potentially interested recipients by criteria such as category. The recommendation is then added to that category and interested people will retrieve it later in a pull-active process. Konstan and Riedl call this a hybrid form of pull-active and push-active collaborative filtering [Konstan2002].

As illustrated, both early forms of collaborative filtering are well suited for small communities while significance of recommendations diminishes with increasing number and heterogeneity of recommenders and recommendations. Not only do people want to know where many people have gone or what information was consumed most but what people like themselves have done. To include this requirement into the collaborative filtering process in either way - pull or push - the recommender or the recipient had to perform an enormous effort to judge whether a recommendation for or from a stranger is suitable. Whenever things become too expensive for people to perform, computers come into play. Automated collaborative filtering stores opinions of many people in a database. Based on the assumption stated in the beginning that people who shared opinions in the past are likely to share them in the future, similarities between people are calculated by special algorithms. Then, information from like-minded people (also called neighbors) is recommended.

2.4.4.2 Approaches to Automated Collaborative Filtering

At this point we want to give an overview of different techniques of automated collaborative filtering. Algorithms will be introduced only in terms of their coarse functionality. A more detailed view will follow for the approach we will take in our later work.
Nearest-neighbor algorithms determine the distance of neighbors on the basis of the history of all users in a system. The weighted average of opinions about a special product within the neighborhood is then used to decide whether an item is recommended or not and how likely the recipient will like the recommendation (expressed in a ranking of recommendations). Nearest-Neighbor algorithms are accurate because they are able to incorporate new information quickly but slow because similarity calculation is based on a user to user comparison which is harmful to performance in large systems ([Sarwar2001], [Konstan2002]).
Nearest-neighbor is an example of a *memory-based* collaborative filtering algorithm. It uses statistical analysis which is applied to the whole user data each time. In contrast, *model-based* approaches such as Bayesian networks derive models of the decision processes for finding recommendations. Bayesian networks build such a model where each node represents an item and contains a decision tree which represents different user information. Traversal of the tree leads to recommendations. Bayesian networks are fast, less data is produced but it takes longer to build because the algorithm is learning based on training sets ([Konstan2002], [Sarwar2001], [Breese1999]).
Clustering techniques build a model by combining people with similar opinions in groups. In the ongoing recommendation process predictions are made on the basis of the average opinions of

other members of a cluster. An individual may be a member of several clusters. In this case an average of similarity from all clusters is used. Participation in each cluster is considered to weight their influence. Clustering techniques are very fast once the clusters are established. However, recommendations are not as personal as those resulting from the approaches presented above especially for those who are located in the periphery of a cluster. Sometimes clustering is a first step towards more accurate similarity calculation. Clusters are built and a nearest-neighbor algorithm is applied afterwards resulting in less user to user comparisons and therefore better performance ([Konstan2002], [Sarwar2001]).
There are several more approaches including horting and dimensionality reduction. The former is able to explore transitive relationships between individuals so that it is not required that a person involved in the recommendation process has rated the item at stake. Dimensionality reduction mainly addresses cases where not enough recommenders are available. It proposes to combine, for example, several movie genres to a broader, more popular genre to transfer opinions from one genre to the other. (cf. [Konstan2002], [Sarwar2001], [Wolf1999]).

2.4.4.3 Automated Collaborative in Huge Systems

Automated collaborative filtering techniques are successfully applied to many systems but have two important shortcomings regarding sparsity and scalability ([Sarwar2001], [Linden2003]). Sparsity of opinions - also referred to as cold-start or first-rater problem - occurs in huge systems such as the amazon.com online-shop (www.amazon.com) with millions of items in their databases. In such systems there are few people who have rated even 1% of the items available [Sarwar2001]. Consequently, accuracy of recommendations is probably low as it may be hard to find neighbors who swim in the same part of the ocean. Even if there are neighbors with a significant overlapping they are only able to recommend items in the amount of a tiny fraction of the range of products. The scalability issue comes up when too many people use a system. The approaches described here aim to make recommendations based on user-to-user comparison. Especially in the nearest-neighbor approach this is very costly in terms of performance.
To face the sparsity and scalability issues Sarwar et al. (2001) and Linden et al. (2003) discuss an item-to-item approach where products or pieces of information that are likely to be rated together are the starting point. For each individual similar items from those already consumed or rated are recommended. Amazon.com uses item-to-item collaborative filtering in their online shops [Linden2003]. As our approach will particularly rely on similarity between users we will not further discuss item-to-item collaborative filtering.
Another method to prevent the cold-start problem is using filter bots. These programs analyze characteristics of items in order to perform ratings and make recommendations. The effect is twofold. People feel more encouraged to rate items when they first use the system because obviously others have done so before. Moreover they enjoy getting recommendations and are willing to give something back to the community.

2.4.4.4 Implicit Measurement of Interest

Another strategy to avoid sparsity of ratings is to interpret them from usage of the system. If somebody reads an article for an appropriate amount of time, it can be inferred that the recommendation was appreciated. In an online shop actions like visiting a product page very often or buying a product may indicate that a customer likes an item.
Implicit ratings are potentially noisy because actions may be misinterpreted. A customer might buy

a product for somebody else or an article might have been opened for reading but the reader went away to do something else. However, implicit measures of interest are far more available than explicit ratings. The latter may not be provided by those people who do not invest time in rating anymore once a recommender system works for them (free-rider problem, cf. [Konstan2002]), p. 58-59). Moreover, evaluation of user action is more honest than subjective user rating. Finally, people are not distracted from their task. Implicit interest measure is leveraged effectively by an online radio called last.fm (www.last.fm). Whenever they like a track they play it to the end. When they do not want to listen to a song at the moment they just skip the track, thus keeping it in their profile. This way the actual procedure of listening to music in a player is also the evaluation procedure. Explicit rating is only done if they never want to hear the song again, thus banning it.

2.4.4.5 Situational recommendations

We mentioned that automated collaborative filtering is based on significance of a person's history for her current and future taste. But such consistency over time is not always given. Selection of movies and music for example is dependent on the recipient's mood and her current activity (leisure time, workplace). The challenge to collaborative filtering is then to react to such changes in a person's context appropriately.
A solution to this is the categorization of items according to common situations or moods, such as a "comedy" movie genre or music for "driving home". But syntactical description and perception of what is appropriate in certain situations is a highly subjective matter. Some like rock on their way home, others classical music.

Other approaches include temporal weighting of ratings and identification of people who are in a similar situation regarding retrieval of information or a task [Konstan2002]. At any rate it must be considered that environment (mood, time of day) and activity (working, cycling, lying in bed) have to be taken into account to detect recurring situations - or contexts - for appropriate recommendations. We have discussed this phenomenological approach in Chapter 2.2.

2.4.4.6 User Control and Transparency of Recommendations

Here we address the issue of the individual's influence on the recommendation process. We illustrated that many collaborative filtering systems rely either entirely on manual recommendations or are completely automated. A hybrid approach would allow a recipient of recommendations to influence her neighbors and how opinions of recommenders are combined and weighted. In order to be able to make conscious decisions the recommendation process has to be understood. Therefore a CF system should convey who is involved and how opinions are combined to come to recommendations (cf. [Konstan2002]).
Transparency of the recommendation process also fosters confidence in the recommender system. Imagine you get a recommendation for a holiday resort. You will probably make sure that it really matches your interests before you travel there. In order to assess a recommendation, information about quantity and quality of the recommendation process has to be displayed. Examples for such information include [Konstan2002]:

- Data the system has collected about the target user
- Size of the neighborhood (people with similar tastes)
- Proximity to the neighbors in terms of agreement and consistency over time

- Number of neighbors that have rated for the item in question
- Consistency of the rating among those neighbors

Some recommender system allow to specify criteria that have to be matched in order to reduce recommendations and evaluate significance [Konstan2002].

2.4.5 Summary

We defined social navigation as field of research and application that leverages findings and methods from various areas - in our case semiotics, psychology, sociology and network analysis - to enhance navigation based on social interaction.
After looking at the development of social navigation applications and criteria for assessing its quality we identified two major characteristics for effective social navigation, social awareness and recommender functionality. **Social awareness** includes any direct and indirect social navigation technique that supports people in perceiving the social environment in terms of opportunities and constraints. The importance of constraints and a shared awareness of the social environment in general was stressed in the social translucence concept. This way, an application allows people to conceive the social context of an information space - thus considering it as a place.
Social awareness is more related to interfaces though results from collaborative filtering are required to create history enriched environments as well.

Recommender functionality in contrast is concerned with the underlying system to obtain the highest possible value from social interaction in form of recommendations. We described the evolution of recommender systems from pull-active and push-active collaborative filtering to automated collaborative filtering, that is from rather direct to indirect social navigation. Then, algorithms for automated collaborative filtering were introduced and issues like sparsity of opinions, measurement of interest by activity instead of rating (as a remedy for the cold-start problem) and situational - or context sensitive - recommendations have been discussed. Finally, user control and transparency of recommendations were introduced as means to foster trust in recommender systems. Control and transparency are provided by feedback mechanisms and interfaces with high social awareness of the environment a recommendation was made in. In other words: the better social awareness in a recommender systems is, the better the quality of recommendations in terms of trust of the recipient. Social awareness in return is dependent on the quality of the facts it is based on which are determined by the collaborative filtering algorithm. Thus, social awareness and recommender functionality turn out to be mutually influencing.

2.5 Conclusion

Economic metadata creation for semantic based content retrieval in large scaled environments like the World Wide Web can only be achieved by widespread participation of "ordinary" content recipients. Folksonomies realize such a collaborative way of categorizing content with freely chosen keywords - called tags. But although applicable by everyone, the meaning of a tag is quite often only unambiguous in the context of its creator.

For this reason we examined the term context more closely, distinguishing between individual, shared and social context. The concept of **individual contexts** refers to a person's individual

process of storing, accessing and interpreting information. If a group of people develops and mediates a shared process of storing and interpreting information, we talk about a **shared context** or **group context**. Shared contexts are based on experiences about people's previous interactions. Social environments with frequent and intensive interactions - like groups - foster the creation of a shared context.
The development of shared contexts happens within constraints of the **social context** of people. The sum of all constraints for social interaction of an individual, such as neighborhood, job, interests, nationality, etc. makes up her social context. Accordingly, formation of a shared context is influenced by constraints of the social context it is located in.

Folksonomies ease the identification of social contexts as the basis for evolving shared contexts. We assume that people who assign the same tags to a particular resource have a common interest in one topic - the social context - and are likely to share ways of thinking and speaking, which we call a shared context.
Besides folksonomies we rely on different concepts to identify shared contexts or - more explicit - groups: social network analysis and theories.
Quantitative network analysis is used to identify patterns within existing and well-defined social networks. Social network theories try to provide explanations for these phenomena.
Our approach - identifying groups in social networks on the basis of tagging - relies on incomplete and limited information about ties between the people involved. Hence, quantitative network analysis is not sufficient. Additionally, we will rely on social network theories to propose groups and relations within a presumed social network.
A central aspect for this approach is the concept of focus: a joint activity around a topic as one part of a person's social context. Foci - like thematically related online resources - relate people with shared interests to each other. A number of strongly related persons within such a focus with only a few loose connections to other clusters of people is called a group.
Relations within groups can be proposed analyzing how constraining the associated focus is. The more time and effort people invest in a focus the more probable is the assumption that they have ties to other members of the focus. Phenomena like tie strength and transitivity can be considered to evaluate the network structure within a focus.
Relations between different people of different foci are called local bridges. These bridging weak ties allow information exchange between different social circles or foci. One interesting question is whether bridges can be proposed between people of different foci to enable information flow across borders of unconnected but possibly related foci.

After identifying a social network structure the network can be utilized for social navigation. We concentrated on two important aspects: social awareness and recommender functionality. Social software (i.e. software that implements social navigation) supports social awareness by allowing its users to conceive their social environment. This requires visualization of interaction processes of members in a social network, thus validating and strengthening relations within the network.
Creating social awareness will positively affect recommender functionality in a social navigation system. If relevance of a recommendation and the context it was made in is perceived, then a recipient is able to evaluate the quality of a recommendation. Thus, feedback to the system is of higher quality and the algorithm can generate recommendations more precisely in the future. Consequently, the overall trust in the system will rise.

After drawing a detailed picture of theories it is based on, we will now introduce our approach - identification of groups in social software for social navigation on the basis of tagging - step by step and associate it with relevant theoretical foundations elaborated here.

3. A Framework for Social Navigation

In this chapter we will envision a framework to establish a social network based on collaborative tagging of web resources. Within this social network groups with shared understandings about a topic shall arise. Social relations within and between these groups are utilized to filter resources and to derive recommendations. Our assumptions will be founded on concepts from the prior chapter.
To ease the understanding of the framework's different steps we start with introducing a metaphor comparing our approach of social network creation through collaborative tagging with organizing an information exchange congress. The metaphor is presented in a scenario-like narrative. Afterwards we will identify key aspects of the metaphor and relate them to the different steps of our framework. Finally, we will again abstract from the metaphor to describe each step in detail.

3.1 A Metaphor: Information Exchange Congress

The information exchange congress was brought to life to mediate knowledge transfer between people who are working on or are interested in similar topics and problems without knowing each other. The congress should offer possibilities to communicate with people from different knowledge domains or disciplines, diverse social and cultural backgrounds and varying experiences and focal points. Everybody - workers, scientists, artists, students, etc. - is welcome to participate.

The congress is divided into several workshops, each related to a specific domain of interest. Instead of predetermining the number of workshops and topics for discussion, they are dynamically formed and based on related interests of the congress' participants. For this reason each participant previously handed in a number of papers and works she had done, a list of especially interesting articles and topics she had read before and information about the people she is collaborating with. The congress organization then not only decided which workshops on which topics to form and who is participating in which workshop, but also who is going to sit next to whom. Of course everybody is free to change places, leave a workshop to participate in another or attend several workshops. A big variety of workshops was formed, for example about "modern architecture" or "social network theories".

The congress takes place in a big building with every workshop in a single room. Everybody within a room can see each other but intensive communication is only possible with your direct neighbors. During a workshop people read and comment articles and information they get from other participants. Of course it is often impossible to read all suggested articles. So when forwarding recommendations to workshop members, the organization team evaluates which recommendation could be of interest for each individual. Their decision is constituted on similar reading and commenting habits, but even more important on the feedback they get for each forwarded recommendation. Accordingly, each participant not only submits her recommendations to the organization team but also submits how suitable previously received recommendations were. Based on this information a prioritized list of recommendations is constantly revised for each participant.
Of course it is always visible from whom information originates. So, if two people constantly recommend each other interesting articles they may decide to move chairs to sit together to talk directly. If more people are involved in this mutual recommendation process, they may decide to

group around a table. After a while the whole seating arrangement may have changed, showing a number of different groups within a workshop.
Some people may only participate infrequently, some may leave the workshop for a while to listen to other discussions in other workshops. Some may even never come back because they are better off in another workshop. People who participate in various workshops may contribute valuable information from different perspectives picked up from other knowledge domains.
Whenever people may discover that discussions within two different workshops are related they might tell people from both workshops and propose to form a new joint workshop.

Once or twice a day there is a break where people gather in the huge cafeteria. Here you can meet your friends and colleagues, people you have dealt with a lot during the day and people you have met in many different workshops to exchange your experience and the latest information.

3.2 Identifying Major Functions of the Framework from the Metaphor

Of course, we are not developing a software to organize congresses. Instead, we propose a system to solve the problem of exchanging information among people - about their work, findings and research from their personal archives - shown by the example of an information exchange congress.
The congress stands for a computer system that brings people together. It analyzes information about reading habits provided by each participant to determine interests and information to share in each area of interest for each participant. How this is done is the first major challenge of our envisioned system - **structuring of individual resources** (see Fig. 3.1, part 1) based on tags a person has applied to her resources.

Once the interests of participants are known similarly interested people are joined in appropriate workshops. Workshops are our metaphor for foci, parts of social contexts that are made up by constraints (in our case an area of interest) where information exchange and communication takes part. **Mapping individuals' interests and resources to foci** (see Fig. 3.1, part 2) is the second major function we have to address for our approach.

In our congress metaphor people are not just signed up for different workshops but are also assigned a place where to sit in each workshop. The (preferred) location of an individual within a workshop - or focus - is our metaphor for the social network that arises as a result of similarity between people and the degree of valuable interaction. **Proposing a social network** (see Fig. 3.1, part 3) as a starting point for interaction in the beginning and constantly updating it as a response to interaction observed is the third big challenge for our system. The network may span over various foci, as a result of people being in several foci or because of foci independent relations, such as friends or colleagues.

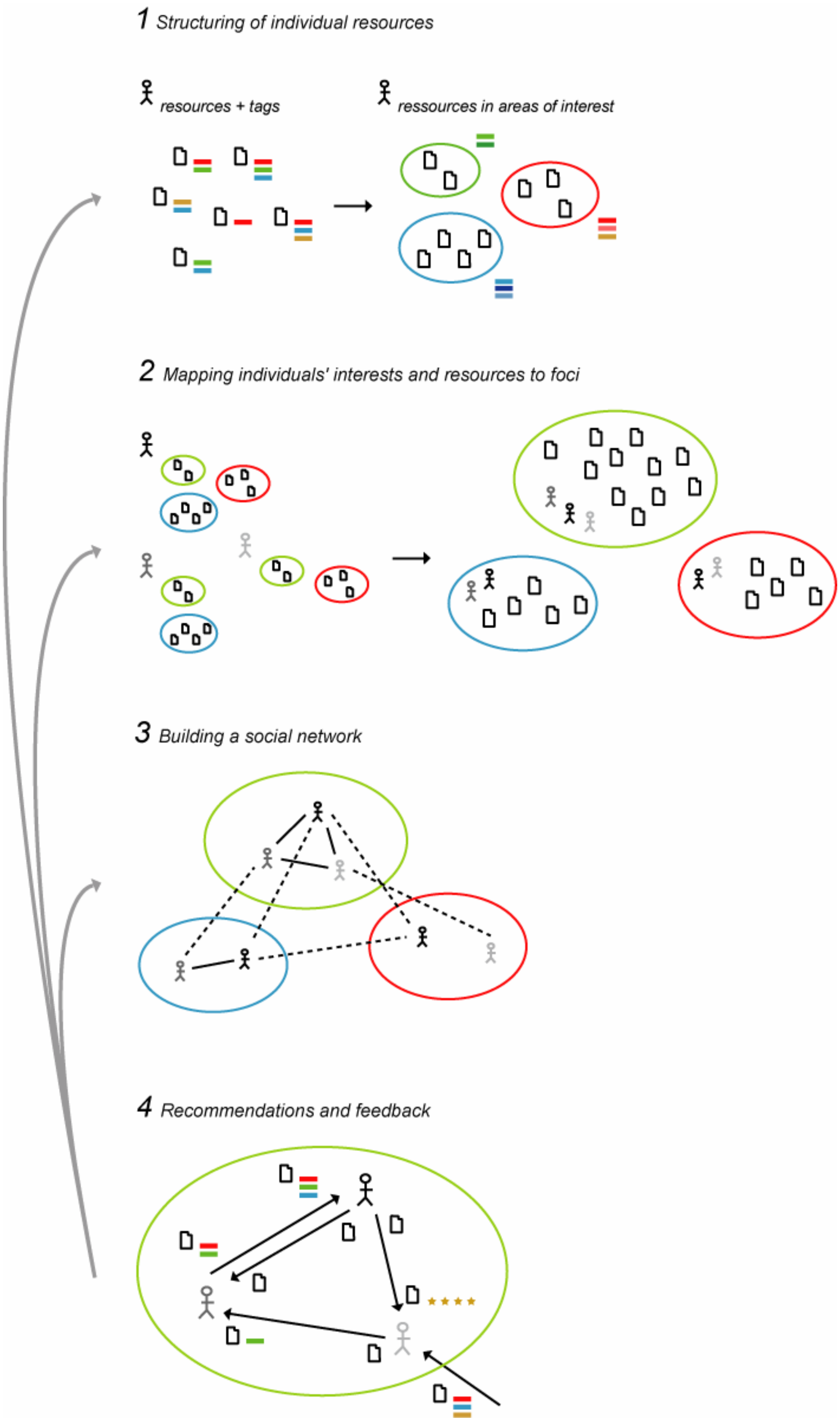

Figure 3.1: Overview of the framework with its four main functions

The organization team of our metaphorical workshops stands for the fourth major function in our envisioned system that operates **recommendations and feedback** (see Fig. 3.1, part 4). As proposed in the metaphor, it helps people within a focus to find the information that is most valuable for them based on the social network. Challenges include definition of rules for forwarding recommendations, prioritization of received items and evaluation of feedback. Feedback may result in a reorganization of the social network.

Figure 3.1 shows the four major functions as a sequence. However, the whole process is iteratively, results from interaction in part four are considered in steps one through two: the structure of an individual's resources will change due to newly acquired resources, shifting interests and changing tagging behavior. Consequently, her interests might end up in a different focus in step two. The social network is updated on basis of results from both, reformation of foci (step two) and degree and quality of interaction from the recommendations and feedback.

3.2.1 Interface Issues

The metaphorical information exchange congress does not only contain the main functionality of our approach but also some issues for interfaces. We already stressed - and will do so in the ongoing deeper discussion - the importance of interfaces for social navigation, especially for social awareness (cf. chapter 2.4.3). In the congress two levels of social awareness can be identified. On the one hand, we have the workshops showing constraints and relations within and across workshops. In our system people shall be aware of these constraints and relations, when they explore their foci, get and evaluate recommendations. On the other hand, we have the cafeteria where people meet during the breaks. It symbolizes an environment and a social network that is not directly connected to foci but a summary of membership and interaction in multiple foci as well as relations that do not stem from the system (such as predefined friends). How social awareness might be created and how it can support the process will be discussed in Chapter 3.3.5.

There are further important issues for interface design in each step. For example the structuring process of individual resources which shall be understood and performed by the user (Chapter 3.3.1). Issues like these will be raised in the discussion of each step. We will not discuss each of them in detail and propose interfaces for each. Important issues will be revisited in Chapter 3.3.5 for further elaboration.

3.3 Development of a Framework

In the previous section we have identified the main steps of our framework. Next, we will describe each step in detail, relating our findings from Chapter 2 to a concrete approach.

3.3.1 Structuring of Individual Resources

In the first step of our framework people's areas of interest shall be identified. Like in our metaphor evaluation of an individual's resources is the basis for building workshops - or foci - of common interests where social interaction will take place. In our envisioned system material that is handed in by people comes in form of digital resources with a URL and an arbitrary number of tags for each resource (note that for the sake of simplicity we will assume that each resource is always tagged with one or more tags). Other information, such as personal data including interests and profession, full-text indexing of documents and other metadata than labels could be evaluated as well. We will not consider these for our approach as they are not as easy to obtain as tags (cf. Chapter 2.1) and because we are particularly interested in the potential of tags as basis for information exchange. However, we consider personal data, especially information about relations to other people (such as FOAF, [Brickley2005]) as a valuable complement to our approach.

The goal of structuring individual resources is twofold. On the one hand it allows for **grasping** the structure of your personal archive. Consequently, validation and potentially adjustment is possible. On the other hand knowing about your individual areas of interest that are connected with similar areas of others helps to comprehend the process of building foci, leading to a higher **transparency**. In the following we will describe these goals more in-depth and propose how to reach them. Afterwards, some issues regarding tagging will be discussed.

3.3.1.1 Grasping and Editing Personal Archives

Tagging is usually done along the way of editing and consuming digital resources. It is performed in various environments and applications such as bookmarking, archiving and weblogging in a more or less strategic manner. A tagging strategy might include attaching many tags (also referred to as tag sets) to each resource in principle, that might result in more cross references and therefore easier information access and possible reorganization to piles of related items. Moreover, consistency of terms is part of a strategy. One example for this is using consistent tags for resources that belong together from a subjective point of view in addition to other tags to denote relations in personal categories (such as adding the tag "university" to any item that is somehow related to your university activities). Consistency also means to prevent ambiguity by using consistent tags in terms of orthography or number and avoidance of synonyms (or usage of additional unambiguous tags).
Thus, a good tagging strategy will lead to a highly interlinked and subjectively meaningful related archive. However, when we introduced folksonomies and their advantages we underlined that tagging is an easy process because people do not have to take care of strategic decisions like selecting appropriate categories from one or more taxonomies (cf. Chapter 2.1.2). The question is then, how can an application support adding appropriate tags without making it more complicate? We also mentioned the importance of a tight feedback loop earlier. In contemporary systems such as del.icio.us (del.icio.us, cf. 1.2), a person's effort of adding tags is valued in turn by showing related tags and statistical information on tags (e.g. overall number of a particular tag) used by herself or all users. However, we think that this is not enough information to draw an overall picture of the quality (as described above) of your tagged archive. Hence, we propose to make relations visible and therefore a direct part of the feedback loop. Our hypothesis is that this additional information will enable people to reflect on their tagged resources to evaluate if their strategy corresponds to the way the resources are structured. If the structure does not meet their expectations, they have to adjust their tagging behavior.

3.3.1.2 Clustering to Support Information Processing

To show relations we will cluster resources and tags provided by a person using a hierarchical clustering algorithm that groups resources based on similarity of their tag sets (we will not describe hierarchical clustering in detail, which is not the scope of this paper. For a good overview see [Jain1999]). Basis for clustering is a similarity value (S) for each pair of resources (AB) which is determined by counting the number of correlated tags (CT) and weighting the result compared to all tags that do not match of both resources (which is the overall number of tags (OT) minus the number of correlated tags (CT)):

$$S_{AB} = CT_{AB} / (OT_{AB} - CT_{AB})$$

Each similarity value is filled in a NxN distance matrix (N = number of all resources a user has) and used as the input for the clustering process. A first run of the hierarchical clustering algorithm shows how far the difference between two clusters has been when merging them. High values show big distortions resulting in a union of two rather dissimilar clusters. By comparing the inter-cluster distance of two succeeding clustering steps one can estimate the optimal number of clusters.

Figure 3.2: Tag clouds from an early version of our prototype application for individual clustering. Font-size is related to the importance of tags for the cluster.

The outcome of the final clustering is a number of clusters of resources with lists of weighted tags, also referred to as tag clouds [Speroni2005]. Tag clouds contain all tags that are attached to resources of a cluster. For every tag the number of occurrences is known. The list of resources is optionally ordered by their tag set's congruence with the middle of a cluster. Tags may occur in

several clusters.
The number of resources a person has increases constantly, for example through recommendations or news feeds, accordingly the clustering has to be repeated frequently.

The examples in Figure 3.2 and 3.3 show clusters of related resources which are described by one or more important tags. The owner of these clusters gets a quick impression of the structure of her archive or individual context (cf. Chapter 2.2.2). Wegner (1987) notes that such ”memory about memory, or metamemory” ([Wegner1987], p. 187) of an individual is crucial to use her memory capacities effectively ([Wegner1987], p.187). According to Wegner, metamemory contains valuable information for all steps of personal information processing (cf. Chapter 2.2.2). Tag clouds as feedback allow to revise tagging strategies for better encoding (because you know your interests and the tags you usually use for them) and consequently storage (containing more and stronger relation information). Finally, browsing clusters fosters retrieval for both, the owner of an archive as well as visitors. Thus, effective communication about each others archives is possible by looking at each others tag clouds and their contents in order to clarify ”who knows what” and ”what are her labels”.

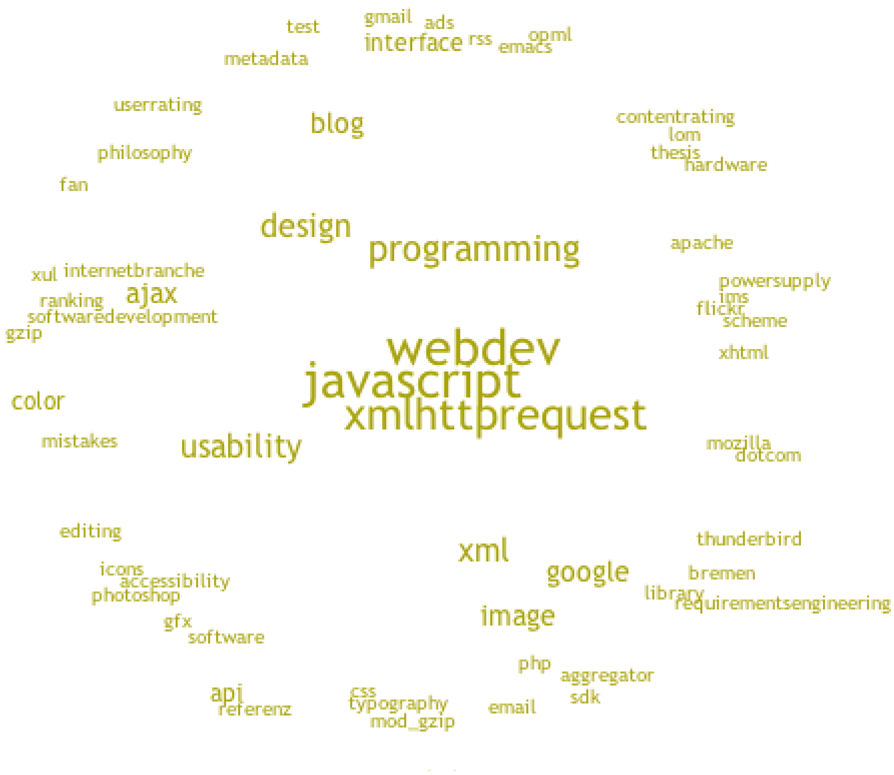

Figure 3.3: A more experimental presentation of clusters from our prototype. Font-size and position from center to orbits shows the importance of a tag.

In a transactive memory system, effective information processing within groups (cf. 2.2.2) only occurs as a result of communication about each others existing individual memory systems. As just explained, tag clouds foster both, the quality of individual memory systems as well as their communication. Accordingly, tag clouds are helpful for transactive memory systems to arise and for their quality. In our envisioned system they are not only helpful for people but also for the computer system because individual clusters are automatically matched to foci for further communication and collaboration. The quality of the input - that is clusters of highly related resources and many representative tags - will then decide about the quality of foci (cf. following Chapter 3.3.2).

3.3.1.3 Transparency of Foci Building Process

We discussed the importance of transparency for influence on collaborative filtering systems and for trust in results they deliver (Chapter 2.4.4.6). Structuring individual resources is the first step in our framework that has the ultimate goal to filter tagged resources collaboratively. Regarding transparency the goal of the first step is to make the process of building foci (cf. Chapter 3.3.2) comprehensible. By comparing of individual clusters with foci people can evaluate their position in a focus and see which personal clusters have matched foci at all.
If resources in a cluster are appreciated by an individual even though her position in a focus is not central, she might decide to adopt important tags from this focus. Our system will also evaluate the position in a focus for building a social network and make recommendations accordingly (cf. Chapters 3.3.3 through 3.3.4).
Knowing about personal clusters that have not matched any foci might encourage to rethink tags that have been used. If desired, other tags may be found by looking at the tags others have used for resources from the cluster in question or by browsing foci to find those with similar resources and adopt tags from these.

A glance on context in sociology taught us that relevance of information is evaluated in light of your social context (Chapter 2.2.3). Judging relevance of information coming from a focus by taking into account how well you fit into it (cf. prior paragraph) is an example for this.
Clusters that did not match any foci contain tags or resources that are not in sight of others, thus having no overall relevance. As discussed in Chapter 2.2, no shared context will be reached for this cluster and different tags are required as proposed above.

3.3.1.4 Known Issues

We argued that the structuring of individual resources is an important part of our framework and how we are going to approach it. However there are issues related to tagging that require special treatment in this step.

First, the resulting clusters are not necessarily areas of interest of an individual. Our clustering algorithm performs only syntactic comparisons. Thus, unless somebody does not intentionally add tags to each resource that relate it to personal interests, clusters will contain resources that are related, but not necessarily considered to be part of an area of interest by its owner. An example for this might be a number of resources that all have been labelled ”to_read”. These will probably be joined to one cluster which is not necessarily an area of interest but probably a collection of different areas. Depending on the other tags that were assigned to these resources they might also match clusters that rather reflect areas of interest like biology, university or the like. However, our current clustering approach does not add resources to different clusters.
Basically, it does not matter that clusters do not reflect areas of interests. We are not going to insert them into a global category of interests later. Instead, related resources with subjective representative labels are offered to other people. If others use the same tags in combination and

everybody agrees on tags and associated resources, a shared context emerges (cf. Chapter 2.2 and Chapters 3.3.2 through 3.3.4). An example might be a focus with a tag "++" for very good resources from any domain, because some people consider the tag "++" appropriate. As expressed by Wegner (1987) labels in such a shared context or transactive memory system may not be meaningful for people who are not involved (cf. Chapter 2.2.2).
In order to raise the probability of proposing foci that lead to shared contexts it would be interesting - as mentioned before - to offer resources in different contexts (that is adding them to several clusters, if applicable). Fuzzy clustering algorithms would approach these problems but we will not go further into detail (cf. [Jain1999]).

Another issue originates from the relations inherent to the tags of a person's archive. As discussed above, people who use tagging are not necessarily aware of its potential in terms of relations and consistency. Here, we expect a clash of highly structured resources and "loosely" tagged collections. If lots of information about relations is given by well considered tags, the algorithm can be adjusted to cluster items only with a high similarity. To derive any clusters from loosely tagged archives in contrast, this similarity threshold must be lowered. While this is technically applicable, the question is, whether this makes sense with regard to the second step of our framework: in the process of building foci clusters with a big variance of relation of tags would be mixed up.
There are two approaches to solve this problem. On the one hand we could add mechanisms to manipulate individual clusters. Once they are created users could move resources, add or remove tags and create new clusters based on information they get from the interface about the quality of their clusters. Behind the scenes this would mean that tags have to be attached and removed to resources accordingly. This consequence has to be transparent to the user to ensure that she learns that clusters are reflecting the tag sets of her archived resources. Whenever cluster are manipulated they have to be recalculated in response. This will probably not be possible in real-time.
On the other hand we can address this issue in the process of building foci. The quality of clusters in terms of cross references between its resources could be evaluated and taken into account when clustering. Following, only clusters of high quality could be joined to a focus. Clusters of bad quality should not be considered at all. The module for building foci could also hand back those clusters which were not matched to others because of bad quality to a manipulation interface for individual clusters as described. Clusters that need validation would be marked accordingly.
For our approach we will not consider manual cluster manipulation though it might be a valuable complement for our framework. As we suggested, it is quite complex and contains performance issues. Still we will consider density of clusters. After clustering individual resources those with few cross references will be marked as bad and will not be considered for the process of building foci.

3.3.1.5 Summary

We argued that clustering of individual resources is not just a technical precondition for building foci but that results allow to grasp your archive - or individual context - of tagged resources and enhance its structure. Further, we argued that the result is a better metamemory which allows for more effective retrieval and communication of your own archive for both, human and machine. Consequently, preconditions for a transactive memory system or a shared context to arise in later steps of our framework are enhanced.

Being able to trace how your clusters are related to foci supports transparency of collaborative filtering, because it allows to judge the relevance of focus and consequently the relevance of recommendations originating from it (cf. Chapter 3.3.4).

Finally, we pointed out critical issues related to tagging that need special care when introducing our envisioned system to a broad mass. However, for a first prototype to show our concept some limitations are acceptable while concentrating on the core functionality.

3.3.2 Mapping Individuals' Interests and Resources to Foci

The second step of our framework describes the process of joining interests and resources of individuals - as defined in the first step - to foci. According to Feld (1981) a focus emerges around joint activities or constraints and fosters interaction between a focus' participants (cf. 2.3.2.2). Our notion of foci refers to Feld, but relies on a slightly different starting point. We assume - as described by the concept of folksonomies (cf. 2.1.2) - that people labelling the same resources with similar tags are not only interested in a shared topic but also have a mutual understandable way of expression. Bringing these people together in one context is our understanding of a focus.

In the following we will start with describing the process of mapping individual areas of interests to foci. Thereby we will concentrate on propositions about significance and similarity of individual interests. Next, we will specify properties of foci with the goal to introduce means for their evaluation.

3.3.2.1 Building Foci

The information for building foci results from the previous step of the framework. Instead of using a person's complete archive of tagged resources we will limit the input data to her individual interests, also referred to as personal clusters. This does not only reduce the number of items to compare drastically but is also the precondition for showing relations between personal clusters and foci, as postulated in Section 3.3.1.3.
Although the previous step should provide personal clusters with a high significance of related tags we have to ensure that loosely related clusters do not falsify the foci building process. Having a high number of resources with unrelated tags in a personal cluster would lead to imprecise and broad foci.
The relevance of a personal cluster can be measured by its density. A high density points out a high number of similar tagged resources, represented by a high number of tags used for more than one resource. A special case is the usage of one or only a few tags for the majority of resources within a personal cluster. This form of "hierarchical" tagging - using one tag as a kind of relating super category - is also seen as highly relevant.

The actual foci building relies on comparing similarities - also referred to as distances - of personal clusters. We start with calculating a NxN distance matrix between all personal clusters (N) of every users. The distance between two personal clusters is calculated analogously to the distance of single resources (cf. 3.3.1.2), based on the correspondence of tags of both clusters. The tags of one cluster are all tags assigned to resources within this cluster. But in contrast to the comparison of tags of resources only the main tags of personal clusters are considered. Main tags are all tags that are more often used within a cluster than the averagely used tag. The goal is to restrict

comparisons to a few tags that signify the topic of a personal cluster. Additionally, main tags are rated according to their occurrence in relation to the overall number of resources in a cluster. The result is a number of main tags ranked in relevance for each personal cluster.
The distance of two individual clusters is calculated on the basis of similar main tags. If a tag is found in both clusters the similarity of both clusters according to this tag (S_{tag}) is defined by the difference of both tag ranks within the specific cluster. Two similar ranks result in high similarity, the higher the rank and the more similar both ranks are the higher the resulting similarity is. In detail, we divide the higher tag rank (R1) by the difference of both tags (R1-R2, rounded to one decimal place) normalized by a factor F.

$$S_{tag} = R1/(R1-R2) / F$$

All similarity values of matching tags are added to form the overall similarity (OS). The overall similarity of two clusters has to be weighed against the number of mismatching tags. For this reason we sum up all tag ranks of main tags that were present in only one of the clusters, leading to an overall dissimilarity value OD. Both - similarity and dissimilarity value - are normalized to a value between 0 and 1 by dividing through the overall number of all main tags (OT). Subtracting the normalized values from each other gives the distance (D) ranging from -1 to 1:

$$D = OS/OT - OD/OT$$

The distance is transformed to a value between 0 and 1 as required by the clustering algorithm.

After calculating distances of all personal clusters the most similar personal clusters are merged to foci. This is done in a clustering process using a hierarchical clustering algorithm (cf. 3.3.1.2).
The clustering has to be redone frequently to reflect changes in personal clusters. The risk - especially in the beginning - is that drastic changes of personal clusters, for example caused by many new participants, will lead to heavy varying foci after each new clustering. We expect some time to pass until a critical mass of participants and accordingly a significant amount of tagged resources - the basis for personal clusters - leads to a number of stable foci. Until then, people may experience that their personal clusters are matched to different foci from one day to another which of course is not a good precondition for transparency as described in Section 3.3.1.3. After establishing a solid basis of larger foci consisting of many similar personal clusters transitions should only happen slowly over time, for example if many people adopt new tags or change their tagging habits.
To lower the risk of varying foci in the beginning the initial foci building process has to be based on a large data set from many different users. Hence, we will collect suitable data from contemporary web services to minimize cold start problems.

In our system a focus is defined by the number of people that are related to it, the number of resources in that focus and the number of tags assigned to these resources. The mapping of people and resources to foci happens through personal clusters.
Overall, we expect stable foci to follow a power law distribution of tags (cf. 2.1.2) with only a few important main tags originating from many different personal clusters and a long tail of infrequently used tags.
To judge the stability or maturity of a focus we will describe its properties in detail in the next chapter.

3.3.2.2 Describing Foci

A focus can be described by various properties: size, breadth and height, activity or age. Taking together these parameters define a focus' maturity.

The **size** of a focus is defined by the number of members. Its **breadth** describes how many different subtopics are present. Narrow foci have only very few strongly related main tags - indicating one dominating topic in a focus. A focus with only one dominating main tag has a folder-like character (cf. Fig. 3.4a). Broader foci show not one dominating tag but a number of tag sets of similar relevance, which are still similar enough to form one focus. Figure 3.4b gives an example showing the tags ”folksonomy”, ”tags” and ”del.icio.us” with similar strong relevance. If a focus gets to broad, the interrelation between its topics may get too loose.
The number of resources in a focus in relation to the number of main tags defines its **height**. Broad foci are supposed to be less high than narrow ones.

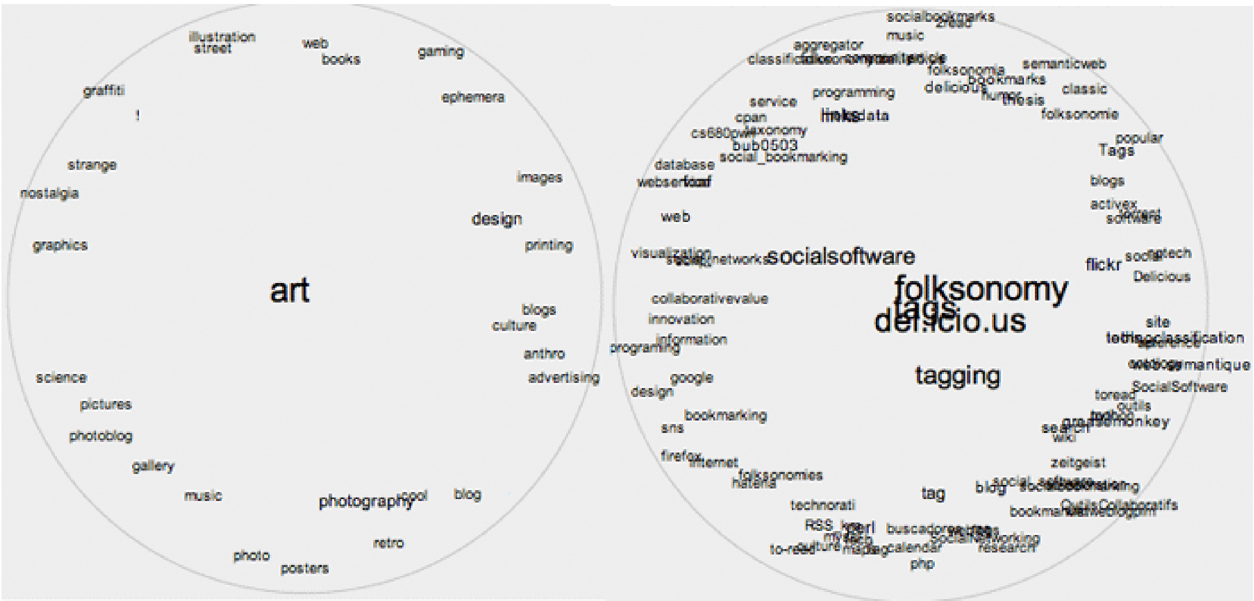

Figure 3.4: a) Narrow focus with one main tag. b) Broader focus with several main tags.

The **activity** of a focus describes the average of resources that are submitted by each of its members per day. The overall activity is the middle of all daily activity values. Personal activity is the number of resources a member submits to its focus in relation to all submitted resources. Finally, the **maturity** of a focus is the relation of the sum of all previously mentioned properties to the **age** of a focus.

Referring to Feld's focus theory it is especially interesting how constraining - that means how time consuming (cf. 2.3.2.2) - a focus is. To evaluate how constraining a focus is from the perspective of one of its members we have to determine how much ”time” this person spends within this particular focus in comparison to all other foci she is also part of. We measure the number of resources she submits to one focus compared to the overall number of resources she submitted to all foci. These personal constraint values allow us to rank all foci of a person according to their relevance. We will use this information in the step of individual structuring of resources (cf. 3.3.1) to organize the user interface and in later steps to rank recommendations (cf. 3.3.4).

3.3.2.3 Summary

A focus combines similar areas of interest of different people into one context. Attributes like maturity and how constraining a focus is allow statements about the quality of a focus. Accordingly, we may ignore foci with only loosely related resources, few members or low age. But we cannot directly influence these values, because they are a result of people's tagging habits. What we can do is give direct feedback how tagging influences the foci building process. This can be shown by visualizing the relation between personal clusters and foci (cf. 3.3.1.3) and by offering support for tagging resources (cf. 4.1.2).

Foci offer opportunities to relate people around shared topics of interest. But in contrast to foci in real life situations - sports clubs, neighborhood, work, etc. - the probability of direct interaction between two people does not coercively increase in an online focus because there are no physical constraints that foster communication. Instead, ways of interaction and communication have to be mediated and visualized by computer systems to create awareness of social relations. We consider a focus as the basis for establishing such a differentiated social network.

3.3.3 Proposing a Social Network

Our framework's third step describes the process of proposing a social network on the basis of previously defined foci. We will base our approach on social network analysis and social network theories introduced in Section 2.3.

Social network analysis offers quantitative methods to describe patterns in social networks - like density or group structures. The precondition for social network analysis is data about interaction between people: examples are frequency of mail exchange, newsgroup communication or - to analyze offline relations - inquiries about neighborly contacts, job recommendations or helpful services. In our case information about interaction is rare. In fact, we will initially propose a social network solely based on indirect interaction: shared ways of tagging combined in a focus.
As described in the previous section we assume that people within such a focus are interested in shared topics. But compared to real life situations you do not "see" people reading the same online resources - as long as computer systems do not support means of social awareness. Our aim is - equivalently to foci in real world environments - to propose and visualize relations between people frequently acting similar in a shared focus. Based on interaction - for example the exchange of recommendations - these relations have to be validated.

We will proceed in three steps, starting with looking at relations within foci including the weighting of ties and identification of groups. Next, relations between foci are discussed with an emphasis on analyzing information flow across borders of foci. Finally, we combine information about relations in and between different foci to describe a foci spanning social network.

3.3.3.1 Relations within Foci

Within our Social Navigation Framework people in a newly set up focus have no known direct relations at all. Our aim is to propose ties between people, weight these ties and merge strongly connected people to groups.

Proposing Ties

The first step to relate people relies on concepts from collaborative filtering and folksonomies. The available information is limited to the foci a person is part of and the resources he or she has imported so far. These resources are related to foci through personal clusters according to their tags.
Collaborative filtering is based on similarities of users (cf. 2.4.4), in our case a person's reading and tagging history. Similar tagging and reading habits indicate near neighbors. For near neighbors we propose ties. Note that this relation is not based on any personal interaction and therefore has to be validated in later steps. We will refer to this in the process of weighting tie strength.
Regarding the concept of folksonomies similar usage of tags is of special relevance, because it signifies people with shared ways of thinking (cf. 2.1). Accordingly, the initial tie strength calculation of two people is based on the number of tags two people share in a focus. This number is divided through the overall number of tags they contributed to the focus (through their individual clusters)

The number of people within a focus is already significantly smaller than in the whole social network, thus reducing person-to-person comparisons to a fair amount.
At this point the initial tie strength is only of marginal significance. For further refinements of the network structure personal interaction has to be considered, the main influence for weighing ties.

Tie Strength

We defined tie strength between two people as the weighted sum of level of interaction, strength of friendship and the number of social circles or foci they share (referred to as multiplexity, cf. 2.3.2). Level of friendship (F) and the overall number of shared foci (M) stay the same for all foci two people share. In contrast the intensity of interaction (I) may vary in different foci. The interaction value is different for each perspective of both related persons. Consequently, ties are considered asymmetric (cf. 2.3.1), which means that tie strength differs for each direction of a tie.
As we consider interaction as the most important factor for tie strength it constitutes 50% of that value, whereas friendship and multiplexity are both weighted 25%. Accordingly tie strength of two people (A,B) in a focus (TS) is defined as following:

$$TS_{AB}=I_{AB}*0.5+K_{AB}, \text{ where } K_{AB}=F*0.25+M*0.25$$

The strength of friendship to a specific person can be set manually by each user of the system based on a given selection of attributes like "contact", "acquaintance","colleague" etc. (as proposed by the XHTML Friendship Network, cf. [Celik2005]). In addition to manual input, data may be taken from existing resources like blogrolls or FOAF files. All these attributes result in a numerical weight. Like interaction the strength of a friendship value may be different for both directions of a tie. Person A may regard person B as her best friend while person B considers person A only as one acquaintance among others. Still, the overall friendship value used in the above tie strength

calculation is the average of both directions. From our point of view friendship is a kind of mutual agreement. If someone considers somebody else as a close friend, but this assumption is not based on mutuality, the overall friendship value is the middle of both points of view.
The number of shared foci shows how many different interests two people have in common. The more interests they share the more likely is interaction. The number of shared foci is divided through the overall number of foci to get a normalized value between 0 and 1.
The level of interaction is significant for the tie strength of two people in a focus. We distinguish between direct and indirect interaction. In our sense indirect interaction refers to similar reading and especially tagging habits, as explained in the previous section and summed up with the term similarity. We will consider indirect interaction only as a starting point to propose initial ties because direct interaction is seen as much more important. Accordingly relations proposed on the basis of similarity are significantly weaker than ties based on direct interaction.
Evaluation of recommendations from other people is considered as direct interaction. A person receives recommendations from people he or she is strongly connected to within a focus. But although the source of the recommended article is a known person (the one who first added the article to the system), the recommendation itself is performed by the system (the recommendation process is described in detail in Section 3.3.4). For this reason automatic recommendations are no personal interaction only the reaction to a recommendation belongs to this category. Details of the feedback process are described in Section 3.3.4.2.
High interaction values have to be verified steadily. If no interaction takes place within a specific period of time the interaction value decreases until a certain minimal threshold is reached. Otherwise people would continuously get recommendations from all people they interacted with in the past, leading to a constantly growing inbox. Of course automatic reduction of interaction values has to be seen in context of the overall time a person is spending to read her recommendations. If someone has only little time to follow her recommendations or is absent for a while no reduction takes place. Only if a significant amount of time is spent within the system a recalculation is performed.

Group Identification

Until now the network structure within a focus shows people with strong or weak relations. A cluster of many strong ties with only a few weak ties to other people is considered a group (cf. 2.3.2.2). We can apply the same clustering techniques described in Section 3.3.1.2 to identify these groups. We start with building a N x N distance matrix of all people within a focus. The distance between two people is based on the tie strength between both. Ties between two people are normally asymmetric because they may have differing friendship values (see previous paragraph) and a varying degree of interaction (one reads more of the other's recommendations than the other way round). To ease calculations we will define the distance of two people as the average of their tie strengths in both directions. We will substitute this inaccuracy in the process of deriving recommendations (cf. 3.3.3). Based on the distance matrix a hierarchical clustering is performed. Groups will be utilized in the subsequent recommendation process.

Additionally, the clustering process offers a lot of valuable information. The density of groups shows how well-connected members within a group are. Dense groups have many strong ties, interaction in such a group is supposed to be high. We may consider a dense group as highly constraining similar to highly constraining foci (cf. 3.3.2.2.).
Furthermore, clustering reveals information about most central person of a group (in the sense of degree centrality, cf. 2.3.1). High degree centrality means broader access to information and more

possibilities to influence others by sending recommendations. We will consider this information to propose relations between unrelated social circles, called bridges later (cf. 3.3.3.2).

Especially in the beginning not all probable ties between members of a group are present, mainly because most ties are only proposed on the basis of similarity and not on direct interaction, as mentioned in the previous paragraphs. Section 2.3 introduced the concept of transitivity that allows the proposal of a tie between two people which are both strongly connected to a third one. Groups show a high number of strong ties, therefore creation of ties based on transitivity can be used to complete the network structure.

Still, there may be some groups within a focus that are not connected to other groups. To foster information flow in the whole focus connections between isolated groups should be introduced. People in a focus should at least share interests in a common topic independent of being in the same group. Therefore, it seems reasonable to propose ties between all people in a focus. Accordingly, we will forward possibly interesting recommendations between unrelated groups within a focus. The goal is to initialize interactions (rating, archiving) as a starting point for further refinement of ties. Details of the recommendation process will be discussed in Section 3.3.4.

3.3.3.2 Relations between Foci

For now our social network consists of a number of foci. Each focus contains several groups, each group in return several - mostly strong - connected people. Some members of different groups are weakly tied. But what about ties between different foci? Ties between foci are important to foster information exchange across otherwise closed social circles (cf. 2.3.2.3). The idea is not only to relate unknown people to each other but to connect different areas of interest and perspectives with the goal to support exchange of thoughts between separated knowledge domains.
In real-life situations people are often weakly connected to people in other foci without being member of these foci. Ties between people connecting unrelated foci are called local bridges.

Regarding our evolving social network we can basically distinguish between two different approaches bridging unrelated foci. First, we may evaluate existing relations between people to propose new bridges. For example, if two people have several foci in common and additionally are good friends - referred to as multiplexity and friendship value (see previous paragraphs) - we might propose a bridge between two foci they do not share conjointly to see if they share more interests than the system is aware of. Based on recommendations over such a new bridge further ties across borders of foci may evolve (cf. 3.3.4.1).
We stated in Section 2.3.2.3 that bridges between foci are weak ties because they only connect loosely related people. The given example shows that in our network a bridge may also evolve from strong ties, due to the fact that high friendship values can be set manually and may be based on communication and real life relations which are not known to our system (thus, in its original sense the term local bridges is not appropriate).

Second, we may have no known personal connections between different foci at all. In this case a mechanisms to propose bridges between unrelated members of different foci is needed.
A naive approach would be to create random bridges between different members of unrelated foci from time to time. These bridges would be cut off again if no positive feedback is given.
More promising are bridges between persons in at least reasonably similar foci. These persons are not similar enough to share a focus but have at least something in common, for example they have

used a few important similar tags or showed common interest in a couple of resources. Special candidates for local bridges are the most active people within a focus, their recommendations should reflect what is significant for this focus. Moreover, if they receive recommendations from a different focus and rate it positively a wide distribution is guaranteed. But the high engagement of persons may lead to strong limitations of topics of his or her own focus without the interest or openness to new influences. Therefore, it might be beneficial to direct recommendations from other foci to less engaged people. These people could bring in new ideas to the core of their own focus to support innovation flow from the borders, as described in Section 2.3.2.3.

3.3.3.3 The Overall Social Network

Most likely people will be part of many different foci. It will be a common case that two people share more than one focus. Figure 3.5 depicts such a social network from the point of view of one member of the network (marked red and bold). This person is a member of two different groups in two foci. The given numbers show tie strength values for directed ties of this person. Local bridges are marked with dotted lines.

Instead of looking at the relation in each focus one could look at the social network on the whole. The strength of a relation would then depend on all relations of a pair of people within and across all foci. The friendship and multiplexity values stay the same for all foci and consequently for the whole network. What changes from one focus to another is the interaction value. But instead of simply summing up all interaction values of all foci the values should be weighted according to the time a person is spending in a focus. If a focus is more constraining for a person, it is weighted stronger in the overall network. Additionally, all bridges between two persons from two different foci are included in the calculation. Again, the interaction value of the bridging tie is taken and weighted against the constraint value of the focus the bridge originates from.

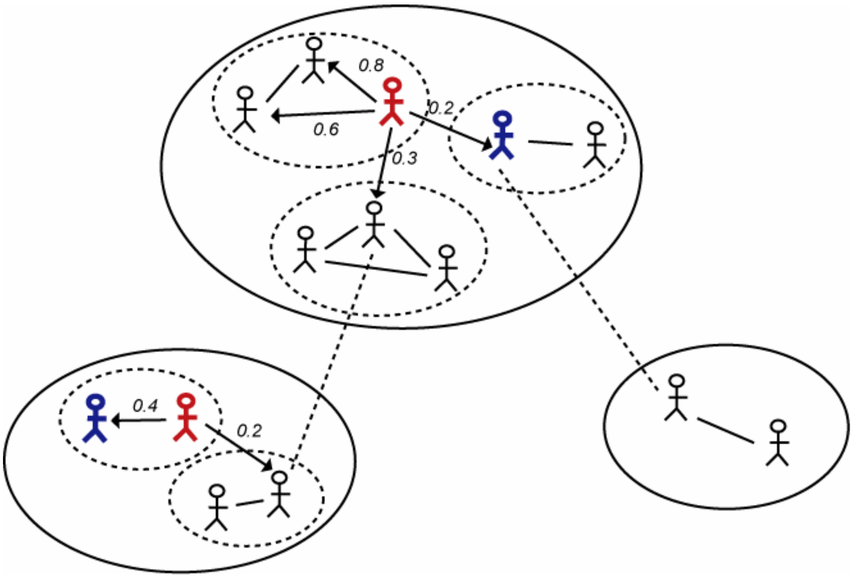

Figure 3.5: The social network with foci

The resulting overall tie strength may differ for both directions of a tie caused by varying interaction values. To ease further calculations we will average the overall tie strength of both ties between two people.

Figure 3.6 shows the overall social network of the previous example. In the previous graph the red bold person shares two foci with the blue marked person. In the overall network both tie strength values are merged to one single tie to that person.

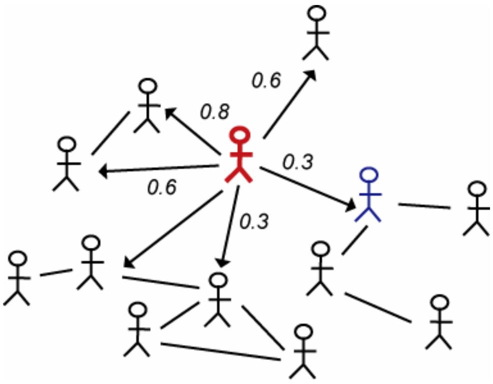

Figure 3.6: The final social network

The overall social network allows use to evaluate relations without considering foci, for example to show all relations a person has or to rank recommendations from a focus the receiver is not part of (more to that in Section 3.3.4). Furthermore, the overall network is considered for proposing local bridges between friends with unrelated areas of interests, as shown in Section 3.3.4.1.

3.3.3.4 Summary

Establishing a social network based on foci is the key concept of our framework and the precondition for deriving recommendations in the next step. In the beginning, we started with creating weak ties between similar people within a focus using collaborative filtering techniques. These ties are proposed by the system and have to be validated through affirmative interaction by the focus' participants. Besides friendship and the number of foci two people share the degree of interaction is the main aspect to refine tie strength.

Based on information about tie strength groups of strongly related member of a focus can be identified. Members of a group are detected and connections between unrelated groups proposed. Again, all steps support the recommendation process described later.

Not only connections between groups but also between foci can be proposed. The goal is to prevent occurrences of isolated foci to support unhindered information flow across the whole network.

Finally, we propose a different presentation of the arising social network. By merging the averages of all tie strengths two persons have in all existing foci an overall tie strength between those two can be calculated. Repeating this for all ties in the network an overall social network is created. The overall network allows focus independent evaluation of people's relations.
In the next step we will utilize our concepts for distributing recommendations.

3.3.4 Recommendations and Feedback

Coming back to our metaphor, the congress will start once workshops are established and seating is arranged within workshops. We have reached this point in our framework as well: foci have been established in step one and two and a social network was proposed in step three that contains information about relations within and between foci.
Though any efforts undertaken so far are related to the multidisciplinary field of social navigation (cf. Chapter 2.4), it is especially this last step that realizes navigation towards resources based on other people's experience - the ultimate goal of social navigation. Precisely, this means taking advantage of the network structure to receive recommendations. Therefore, a collaborative filtering system is initiated that makes recommendations based on constraints of social relations in and across multiple foci, comparable to the seating arrangements and workshops in our metaphor.

Our approach - making recommendations based on foci - allows for situational recommendations as postulated in Chapter 2.4.4.5. Individual clusters reflect individual contexts which are mapped to foci. People will visit a focus that contains the particular individual context they currently are dealing with to get adequate recommendations.

Referring to the classification discussed in Chapter 2.4.4.1 our collaborative filtering system can be defined as automated. Yet it has elements of push-active and pull-active collaborative filtering approaches: recipients' personal clusters define the areas of interests they expect to get recommendations for. Recommenders determine potential recipients by adding tags respectively. Having such additional knowledge about recommended or expected resources distinguishes our approach from a common automated collaborative filtering as described in Chapter 2.4.4.
It is this additional information that supports the finding of meaningful relations based on relative few items (as shown in steps one through four) whereas collaborative filtering based on statistical evaluation of viewing items requires larger samples and might therefore suffer a sparsity or first-rater problem (cf. Chapter 2.4.4.3).

Most of the structure our system calculated so far is based on the assumption that social contexts can be derived from similar reading and tagging habits that might eventually lead to shared contexts. In Chapter 2 we elaborated that communication is crucial to develop shared contexts. Thus, to validate our assumptions, we will introduce a predefined way of communication that consists of recommendations and feedback to validate foci and the social network we proposed. Whenever feedback is positive we will strengthen relations and consequently foci.
The communication is asynchronous, thus, a rather indirect form of social navigation. However, interaction between people within the network is visible, so navigation is not completely indirect (cf. Chapter 2.4.1). Yet we think that synchronous means of communication could be a valuable complement to our approach later.
Recommendations are not only made within groups but are also suggested for separated groups within foci and between foci. The goal is to validate correctness of these separations. We suggested local bridges earlier (cf. Chapter 2.3.2). We will make use of such bridges to send recommendations to otherwise unrelated people.

In the following, we will explain the basic recommendation process between people. For reasons of clearness we will consider only recommendations *within* groups first. Then, recommending over local bridges will be explained. Finally, a feedback mechanism will be described.

3.3.4.1 Recommendation Engine

Recommendations between people happen on various levels: among close friends within a focus - tied together in groups - between groups in a focus and between people of different foci. Basically, recommendations are forwarded to everybody in a focus. The relevance of a recommendation is defined on the recipient's side in consideration of the tie strength to its sender. Ties to group members are strong ties (cf. Chapter 3.3.3). Strong ties and especially recommendations within the same group result in a higher relevance. A ranked list of recommendations is maintained for every member of a focus.
For now, we will only look at these group recommendations. Communication across the borders of groups and foci will be discussed below.

Starting point are any resources of the sender of a recommendation that are assigned to the focus the interaction takes place in. They are either resources and tags that led to the creation of that focus in the first place (cf. Chapter 3.3.1 and 3.3.2), resources arriving in feeds (standardized exchange formats, cf. Chapter 4.2.1) that already have been tagged in external systems and fit to the focus in question or resources that have been archived and tagged within our proposed system. The latter is ideal because people have in mind their personal clusters and foci they are part of when tagging.

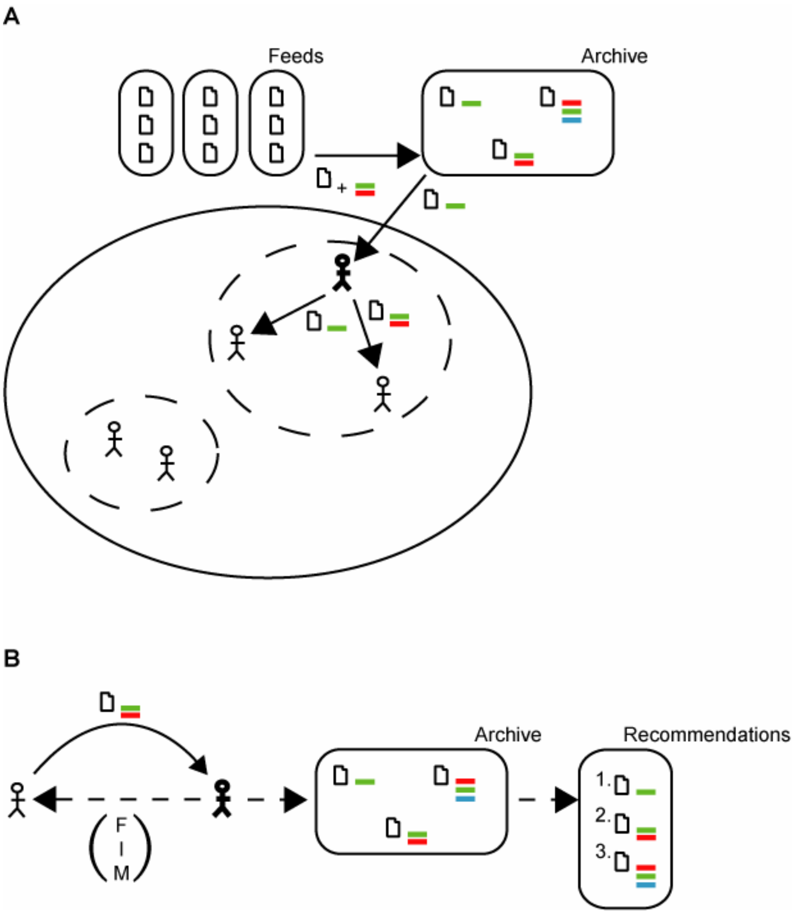

Figure 3.7: (A) Tagged resources of a sender (marked bold) are distributed in a group within a focus. (B) Recipient (marked bold) evaluates tie strength and similarity with archived tags to determine ranking in list of recommendations.

Any of these resources are forwarded to all members of a focus. If not yet received by a group member, a resource is evaluated on basis of the relation to the sender and regarding the resource's tag set's similarity to the recipient's other resources. The goal is to prioritize any incoming resource for insertion into her recommendation list for this focus.
The relation to the sender is defined by an asymmetric tie of a specific strength (consisting of friendship, interaction, multiplexity, cf. Chapter 3.3.3) where the interaction value from the recipient to the sender is taken into account instead of a mutual interaction value. We do this because we are mainly interested in the amount of positively valued interaction from the recipient's point of view. Feedback will affect this value (see below).

Yet, another possible approach would include the interaction value directed from the sender to the recipient but only weigh it 20% (versus 80% for the interaction value from the viewpoint of the recipient). This would ensure that tie strength is considered as a mutual feature. Anyway, mutual interaction values in each cluster are precondition for a high auto-generated friendship value (cf. chapter 3.3.3.1) that is included in the calculation as well.

For comparing the recommendation with the recipient's personal archive, we use the same procedure for comparing tag sets as in the process of clustering of individual resources (cf. Chapter 3.3.1.2). After comparing the item in question to each resource of the archive we will have an average similarity value between 0 and 1.

Overall similarity to a recipient's other resources will be weighted 30% whereas tie strength counts 70%. We place emphasis on information about personal relationship and experience opposite to mere syntactical comparison.

After all, the averaged and weighted evaluation value for the recommendation is between 0 and 1. Any recommendation coming from a group member is prioritized by adding 1 in the end to make sure that group recommendations are higher in the list of recommendations for the current focus than those coming from local bridges (see below). We do this because we aim at boosting group contexts.
Finally, the resource is added to the recommendation list of a context. Figure 3.7 sums up the recommendation process.

In addition to recommendation lists for foci we will maintain an overall list for each person that contains recommended items from all foci. Items are prioritized by weighting them with the constraint value of the focus they originate from.

Popular items will be recommended to an individual several times as a consequence of any resource being forwarded to all group members. In this case, a resource will show up in a recommendation list several times. The higher an item is prioritized in the recommendation engine the higher it will be listed. In order to have every item in the list only once all recommendations prioritized lower are taken from the list and shown with the top rated item. Consequently, items recommended by several people are listed under the name of the strongest tied recommender. The question is to whom feedback shall be given. To the originator of an item (author or importer into the system)? Or the closest neighbor who recommended it? Everybody in the recommendation chain? We think that probably feedback should be given also to the predecessors of a recommender in alleviated form (such as weighting according to tie strength). We will not adopt this for our framework to keep things simple. Yet there is already a mechanism that regulates who is addressed with feedback which is quite fair: temporal consumption. People more or less work through articles every day. If they read a lot, they will likely receive a recommendation by one of the early adopters of an item thus fostering weak ties in case of positive feedback. In this case their effort is valued by extension of their network. If people are rather lazy and only consume the topmost articles they have to wait until close friends forward a recommendation coming from other areas of the network. In this case the recipients horizon in terms of relations to various people will not be broadened. Of course temporal consumption may not always be interpreted this way (take a recipient who is going to holidays for weeks) but we think it is still significant.

Recommendations Over Local Bridges

We raised the issue of local bridges earlier. To make use of potential local bridges we propose two new paths for recommendations.
At first, any member of a focus may be considered a weak tie for everyone else - even though direct interaction has not taken place yet - since a common interest was identified. Thus, recommendations are sent to all members of a focus in addition to group members to ensure that potential *valuable* weak ties can be identified (cf. 2.3.2.3). Recommendations from weak ties are prioritized lower by a recipient because the tie strength is lower. Still weak ties can lead to positively valued recommendations between different groups. Checking for such bridging weak ties within foci is important to avoid unnecessary separation of groups. Moreover, it has a useful side-effect. In case of sparsity of recommendations in a group there are still recommendations from the "surrounding" focus. Especially small groups and foci that are likely to face a sparsity problem will benefit (cf. Chapter 2.4.4).

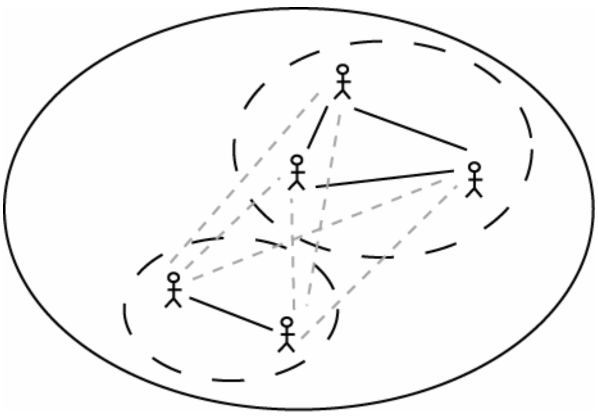

Figure 3.8: Weak ties as local bridges. A focus with two groups (dashed ellipses). Members of a group have strong ties, members of the focus weak ties (dashed and light).

Secondly, as mentioned in Chapter 3.3.3.3 our overall network can be used to identify local bridges. Irrespective of foci, recommendations are made to all neighbors on the network to whom a strong tie exists. We described in detail in chapter 3.3.3.3 that strong ties are the result of interaction values to friends in various foci (the averaged and weighted interaction value), strong friendship which was manually declared (or is a result of a high interaction value) and a strong correlation of multiple interests (multiplexity value). Consequently, neighbors on the overall network are not necessarily group members or even in the same focus (in case of a friendship that is not topic related or coexistence in many foci). Those connections may serve as local bridges between groups and foci even though they are strong ties (cf. Chapter 3.3.3.2).
Bridges between foci foster flow of innovation among otherwise closed social circles. Regarding our way of creating foci - based on syntactical comparison of tagged resources - it is moreover a possibility for people who have not used similar tags in the first place to communicate and develop

a shared context eventually (cf. Chapter 2.2 and Chapter 3.3.3). Finally, bridging foci is another way of facing sparsity of recommendations. People in foci with few activity will have items from other foci on top of their list. Of course, this might lead to an adoption of a low activity focus by a similar one with high activity.

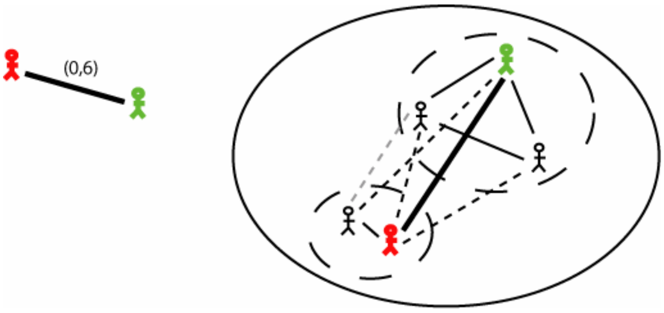

Figure 3.9: A strong tie bridging two groups of a focus.

Figure 3.9 shows a local bridge between two groups of a focus. A strong tie is given for the red bold person and the green bold person as a result of the average of both directed ties (displayed in left part of the image, cf. also Chapter 3.3.3.3). Recommendations are sent on strong and weak ties through the network of the focus (displayed to the right). The bridging tie that was added (bold line) allows for boosting already existent weak ties (cf. prior paragraph) by transitivity (black dashed lines). A recommendation for example that is sent from the red person to the green may be archived by her and forwarded to a strong tied person in her focus consequently. Our recommendation engine will prioritize the forwarded item higher than those made by way of the weak tie between the red person and a member of the green person's group. The higher an item in the recommendation list is the higher the chance that it will be noticed and feedback will be given is. Feedback is not only given to the sender (green) of the recommendation but also to the originator (red). This "indirect" feedback (see 3.3.4.2) then defines the weak tie between recipient and originator. The interaction value of this tie will be lowered or raised so it might become a strong tie after a while.
Once feedback for a recommendation over a bridge is given (bold line) the new tie strength in that direction is stored and not taken from the overall network anymore. This is how we make sure that inappropriate bridges loose their impact and people are not spammed. The same goes for bridges between foci.

Bridging of foci is illustrated in Figure 3.10. Again, we have a bidirectional strong tie between the red and the green person in the overall network (upper left). In the first situation (A) the tie between two persons of different foci is caused by high interaction in (multiple) other foci and friendship. By recommending over this tie we evaluate whether there is also a common interest in foci the two persons do not share yet. Moreover, this connection serves as a bridge for emergence of new weak ties between the two foci in question. For example, the red person receives an item from the green one, archives it and accordingly recommends it to the other group members. This item would not have reached this person without the bridge. By means of indirect feedback (see below) to sender and originator we can establish a weak tie that will serve as new local bridge between the foci. This tie would be directed from the receiver to the originator (green person) so that recommendations can be sent from the green person to the receiver. Positively valued recommendations from the receiver to the red person and on to the Green would establish a weak tie in the opposite direction. The illustration shows all possible weak ties (grey dotted lines).

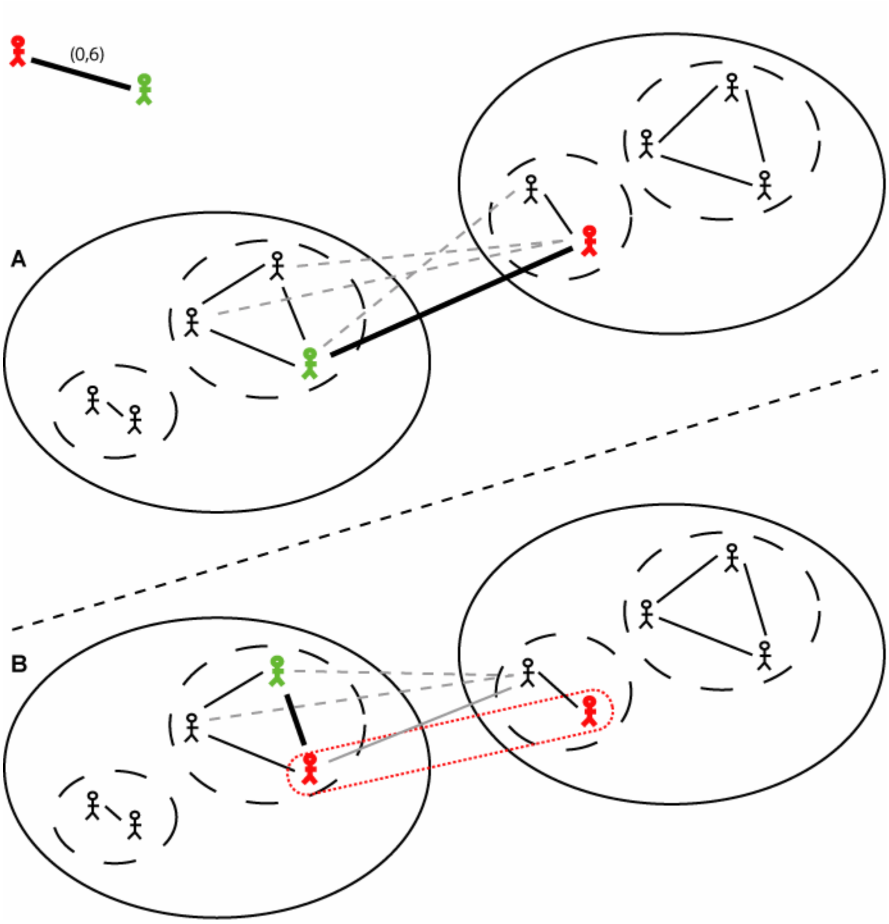

Figure 3.10: Local bridges between foci based on strong ties from the overall network made up by two people (A) or a single person (B).

Situation B shows how a single person (Red) can bridge foci. In the overall network Red has strong ties to her group members from all foci. This is how recommendations can cross the borders of foci. If feedback is given, weak ties can be formed as the illustration shows (light lines). Note, that a recommendation that is passed from Green to Red will be forwarded to Red's group member in the other focus based on the averaged tie strength from the overall network instead of the tie strength that respects interaction in the focus (cf. Chapter 3.3.3.3). To store feedback from this recommendation we have to introduce a new strong tie between the red person in the left focus to the recipient in the right focus (light line) so interaction across foci will not affect ties made up from interaction within foci.

The approach presented - deriving local bridges from the overall network - could be complemented by a statistical approach. We could identify for example less constraining foci or similar foci and propose bridges within or between. Bridging ties could be established between people who are very engaged in their groups, those that are near the border and therefore closer to another group or focus or randomly selected people.

3.3.4.2 Feedback

Feedback to recommendations allows people to influence the collaborative filtering system as discussed in Chapter 2.4.4.6. It will enable the system to validate whether a focus is considered appropriate by a person by measuring if any recommendations are positively valued. Furthermore,

shared contexts are developed or detected respectively based on positive feedback within groups. Shared contexts are leveraged because those recommendations from neighbors with highest interaction values are preferred in the future.
We discussed in Chapter 2.4.4.6 that transparency of the recommendation process fosters quality of feedback. In our case the recipient knows about the thematic and social environment of the recommendation (the focus), her relation to that environment (personal area of interest) and about the relation to the sender from the social network.

Feedback completes communication we enforce in our system (see above). It influences the interaction value directed to a sender for a certain cluster. However, sender and recipient are only slightly aware that they are communicating directly. As described, recommendations are generated by the system. Thus, a sender might not know that she is sending recommendations. Feedback does not address the sender directly either. Basically there are two ways to make communication transparent. *Implicit* hints will be delivered by the system behavior. Every sender is also recipient of recommendations and will therefore recognize that recommendations are received from other group members because the sender of a recommendation is always known to the recipient. Moreover, the network structure of the group and highly ranked recommendations from certain group members will reveal that there has been positive interaction. Additionally, the interface of our envisioned system will contain *explicit* hints. An interactive map of a group and its focus can reveal information flow and changes in the structure of the network within a group in general. For each item a custom view on the network could show its penetration, that is how much positive feedback was given.
The issues addressed here affect social awareness in the system. We will come back to this at the end of this chapter.

Indirect Feedback

Our discussion of local bridges above yielded that not only feedback to the sender is important but also to the originator of a recommendation. Only this way new ties that originate from transitivity can be defined (in terms of interaction values). How direct feedback (to the sender) and indirect (to his predecessor) feedback differ will be explained in the following introduction to our feedback mechanism.

A Feedback System

In our elaboration of a social network we pointed out that feedback will basically mean that a recommendation is *accepted*, *declined* or *ignored* (that is, there is no feedback, cf. Chapter 3.3.3).

We are able to detect the type of feedback based on indicator actions such as **reading, archiving** and **rating**. The former two are forms of implicit interest measure and therefore easier to obtain (cf. Chapter 2.4.4.4). However, measuring reading may be problematic because it has to be distinguished from skimming on the one hand and unintentional long intervals (because attention was drawn off) on the other hand. Reading frequency, that is how often a resource was revisited, would be a good indicator for interest in a resource.
We consider reading time as too noisy for the reasons we mentioned. Reading frequency in contrast will be one of our indicators even though it complicates the feedback loop because not all feedback data will be available with the first feedback. For this reason we are going to neglect it in our prototype.

Archiving is another example of implicit interest measure because it is not performed solely for the purpose of giving feedback but also archiving it in a meaningful way. For the interpretation of this indicator action it is interesting to evaluate the tags that are used. Basically, a recipient might adopt the tags from the proposed resource or remove and add tags. Adaptation may be interpreted as agreement. Removing tags shows disagreement depending on how many tags are removed in relation to the overall number of tags. New tags might fit to the focus the recommendation is made in or not. Tags that fit into the current cluster are considered as sign of agreement. The resource might even be tagged with completely different tags so it ends up in another area of interest of the recipient. In this case the recommendation was appreciated but not considered to be labelled correctly. Feedback will not be positive because it is not feedback for the resource alone but for the resource in combination with its tags.
Rating is an explicit form of feedback. Ratings are performed on a scale predetermined by a system. For our system we define a simple scale from ”Not interesting” to ”Interesting” and ”Very interesting”. This scale could be represented with icons of a hand showing thumbs down to thumbs up. Often a more differentiated scale is used. But the bigger the mental scope for interpretation (e.g. what does four stars on a scale from one to 5 stars mean”) the more difficult it is to rate and the less significant or comparable ratings are. For our purposes it is most important to evaluate if a person accepts a recommendation (that is, will not rate ”not interesting”).
The following table sums up all of our indicator actions and how they will affect the value of the feedback:

Action	Value	Description
Archiving	+1	Tags adopted
	+1 -N	Tags erased: $N_{tagsErased}$ / $N_{tagsInTagSet}$
	+N	Tags added: $N_{tagsAddedInSenderArchive}$ / $N_{tagsAdded}$
	+0	Tagged differently, resource ends up in another focus
Rating	-1	Rated "Not interesting"
	+0.5	Rated "Interesting"
	+1	Rated "Very interesting"
Reading frequency	+N	+ 0.1 for each visit until 10 visits reached, update after each visit

For archiving with adopting all tags the feedback value is 1. If tags are removed, a relation value of removed tags to the overall number of tags in the tag set is subtracted from 1. Whenever labels are added the relation of new tags to the number of tags that are in the recommenders personal clusters of the focus in question is taken because we are interested whether sender and recipient use the same ”language”. For both, adding and removing tags, values will be in the interval between 0 and 1. The final value will be averaged.
Finally, if tags are edited in a way that makes the item end up in another cluster the feedback value will be 0 (see prior paragraph).

Reading frequency will optionally increase the interaction value later. For each visit 0.1 is added until 10 revisits are reached, so the range is between 0 and 1.

If an item is rated ”Not interesting” the feedback value will be -1. ”Interesting” is an expression of agreement so a moderate value of 0.5 is taken. Rating ”Very interesting” is the strongest positive feedback and will be valued +1.

The result in each of the three sections Archiving, Rating and Reading frequency is a value between 0 and 1. While archiving with one or more tags is mandatory in case of positive feedback (because we only want to have tagged resources in the archive) people are not forced to archive *and* rate resources. Thus, for a first feedback of a recommendation we will get either only a value for archiving or a value for both. In case of the latter we will weigh both 40%. Otherwise archiving will be weighted 80%. If negative feedback is given, that is rating is "not interesting" we only have a rating value which will be weighted 80%. Reading frequency always counts 20%. To be able to update the overall influence on an interaction value each time an item is revisited, any results from an interaction have to be stored.
The final value is added or subtracted from the current interaction value of the tie strength in direction to the sender.

For indirect feedback, that is feedback to a recommender's predecessor in case of local bridges, evaluations of archiving is done pertaining to the tags of the predecessor because it influences the interaction value to her. Furthermore, the final feedback value will only be weighted 50% in contrast to direct feedback because the flow of information would not have been possible without the strong tie to the mediating person.

Feedback Interface

We just mentioned that archiving and rating are optional. Our feedback interface will basically consist of two sections, one for **archiving and optional rating** and another for **negative rating.**
For archival the recommender's tags are displayed in a text input field for adaptation or editing (see Fig. 3.11). The rating scale is reduced to "Interesting" and "Very interesting". "Not interesting" would contradict the action of archival which is an expression of agreement.
Refusing a recommendation is possible by selecting the negative rating option (see Fig. 3.11).

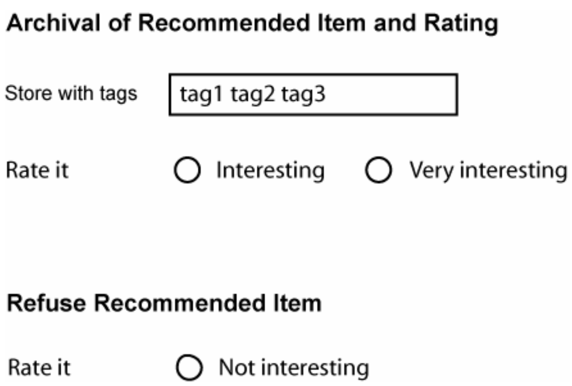

Figure 3.11: Basic concept of a feedback interface

3.3.4.3 Summary

We presented a recommendation engine which is incentive for interaction in terms of feedback in our framework. It is important to note that filtering of recommendations which are sent on weak and strong ties within foci happens on the recipient's side. This forwarding of recommendations to

all people in a focus fosters interconnectedness within a social context while interaction in groups is boosted for development of shared contexts. Filtering happens as prioritization in a list of recommendations for a focus respecting experience of interaction (part of tie strength) as well as concordance with the individual context, that is the tags of the individual cluster that matched the foci the recommendation is made in.
Identification and utilization of local bridges was described leveraging the overall network developed in step three (Chapter 3.3.3). We showed how strong ties, developed from acquaintance in other foci or outside of the system, can support interlinking of groups and foci to validate separation and to allow flow of innovation among otherwise unrelated circles on weak ties.

Finally, the feedback mechanism was introduced which is the heart of our framework because it implements communication we postulated in Chapter 2 as a precondition for a shared context. By giving implicit and explicit feedback in terms of tagging received recommendations and rating, we redefine tie strength within foci which affects the social network and is used in turn for further prioritization of recommendations. Thus, items of a certain neighbor might become very popular in an individual's list or loose their impact because feedback is rather negative.
We also introduced indirect feedback to establish and evaluate weak ties that originate from local bridges.

Together recommendations and feedback as well as display of their influence on the social network (cf. forthcoming Chapter 3.3.5) are the core elements of social navigation in our framework. Yet, they are not just results but will further shape the social environment as well as resulting navigational hints.

3.3.5 Social Awareness: Basic Requirements for Interface Design

We emphasized in Chapter 2.4.3 that social navigation systems like ours are intrinsically tied to an adequate interface that supports people by creating social awareness which in turn helps the individual to navigate more effectively. The concepts of social awareness or social translucence (Chapter 2.4.3) do not propose concrete solutions for interface design but rather guidelines to identify and use social elements that are fruitful to the overall goal of the application.
Both concepts - social awareness and social translucence - differentiate between the main questions one has to ask which are **"How can I convey social components of my application to make people act social?"** (social affordance and visibility, cf. 2.4.3) and **"How can I make people experience that they are acting social and how will it shape the environment for further interaction?"** (usage reshapes functionality and structure, awareness and accountability, cf. 2.4.3). Concluding, there basically has to be an incentive for using social navigation capabilities as well as feedback to the people and the social environment (the information space becoming a place, cf. Chapter 2.4.2).

In order to develop requirements or innovations for interface design both questions have to be answered to convey social capabilities to the user and to find ways of showing effects of social navigation. We will not perform a complete analysis here but confine ourselves to a single yet important social function of our application: Within groups people communicate with and observe other group members communicating. According to our theoretical considerations (2.2) and to Wegner's transactive memory system in particular (2.2.2) this is crucial for developing a shared context.

According to question one we have to expose this function to the people. Thus, whenever a focus or a group is visited it shall be transparent that items are recommended by certain other people and response to recommendations is possible in terms of feedback. Moreover, the system shall convey to the user that the same is happening between other members of a group or focus. These are two basic requirements for interface design for foci.
In order to valuate social interaction we have to answer question two. Valuation or feedback to communication results in a changed relationship to a particular person in terms of tie strength. The system shall present the (constantly updated) social network of a person's group members according to communication that happened. It should also reveal whether recent communication brought people closer together or degraded their relationship. Having a constantly changing social network within the group as response to acting social is also the answer to the second part of the question - how the environment for further interaction will be shaped. People experience their influence on the formation of foci and groups as well as the network structure within. Furthermore they will get different recommendations. As proposed in the concept of social proxies (2.4.3.2) the social network should be shown from a third person perspective where everybody watches the same situation and concludes that she is visible to others. A shared awareness of being visible to each other is created that fosters fruitful phenomena from real world groups including striving for reputation and acting responsible. We also expect the shaped environment to influence decisions made when giving feedback. For example, tags of an important member who has a high betweeness centrality (cf. Chapter 2.3.1) are more likely to be adopted than those of an unimportant.

We will apply the requirements identified here in the design process of a focus interface in Chapter 4.1.2.

3.4 Summary: Conclusion of Requirements for a Prototype

In this chapter we proposed a framework for social navigation based on tagging. Basically, we distinguished four steps: structuring of individual resources on the basis of tagging to identify areas of interests, mapping these individuals' interests and resources to foci, proposing a social network based on foci and last utilizing the social network to derive recommendations and feedback to verify the network. We described each step in detail, introducing a variety of problems and possible solutions. Concluding, we want to summarize the core functionality of each step, in order to point out what we intend to realize in a prototype (implementing the framework) that is presented in the next chapter. The result is a set of basic requirements for each step.

Requirements Engineering

The requirements we derive here are mostly functional. They are on a quite abstract level and would have to be redefined in further iterations of the software development process. Even though we will not do any further project management - these requirements shall just give an overview to the reader - we are going to consider elements of requirements engineering as proposed by Rupp (2002). We will state the requirements in a standardized way (syntactical pattern) to raise comprehensibility (cf. [Rupp2002], p. 469). Moreover we use standardized verbs to indicate the level of prioritization and obligation for each requirement (cf. [Rupp2002], p. 165). We will mainly use ”shall” for obligatory requirements and ”should” for optional or requested requirements. However, this is not binding for implementation of the prototype which only realizes our envisioned system in part. We will neglect even obligatory features which are necessary to complete the framework we propose. References for every requirement are added to ensure traceability and to show some (yet incomplete) interrelations.

3.4.1 Structuring Individual Resources

In the process of structuring individual resources the core functionality is the clustering of resources by similarity of their tag sets (Requirement 1.1, see table below). Other functions which will be summed up in the following support the two main goals we identified in this step: to enable people to grasp their own archive and to make the process of building foci transparent. First, counting the number of each tag in a tag cloud (R 1.2) allows to present it according to its importance. Ordering resources of a cluster by congruence of their tag sets to the most important tags of the cluster (R 1.3) shows which resources are most representative.
Assigning resources to more than one cluster if affiliation to an individual cluster is not unambiguous (R 1.4) could be useful in order to better match areas of interest of a person (even though there might be redundant clusters then) and to raise the probability of matching individual clusters to foci. Even though we expect this feature to deliver better results, we will not implement it in our prototype in order to be able to give a more complete overview of our framework.
We also proposed to approach the problem of too few structure within people’s archives by adjusting the number of resulting clusters according to the degree of cross-references contained in their whole archive (R 1.5). Another problem that results from this individual treatment is a lack of comparability of clusters of different people. This issue was discussed in Chapter 3.1. A solution was proposed in Chapter 3.2 (see following chapter). A function for providing individual clusters with higher density is manipulation by their owner (R 1.6) including identification of bad clusters by the system (R 1.7). We got more into detail in Chapter 3.1 but will not implement it in our prototype. Finally, we expect higher structured individual clusters as a result of people using our framework and therefore getting better feedback for tagging.

3.4.1.1 Overview of Requirements

Ref.	Requirement	In Prototype?	Comment
1.1	The system shall cluster individual resources by similarity of their tag sets.	Yes	
1.2	The system shall rank each tag in an individual cluster to create a tag cloud.	Yes	
1.3	The system should order resources of a cluster by congruence of their tag sets to the most important tags of the cluster.	No	
1.4	If affiliation to an individual cluster is not unambiguous the system shall assign resources to any adequate cluster.	No	Fuzzy clustering takes to much time to implement in prototype.
1.5	The system shall adjust the number of clusters according to structure contained in the individual archive.	Yes	

1.6	The system should provide the owner the ability to manipulate her clusters.	No	
1.7	The system shall identify clusters with insufficient density and notify the owner.	Partly	very loose structured clusters are not stored

3.4.2 Mapping Individual Interests to Foci

This step aims at bringing together people with similar interests and similar ways of expression to support the creation of shared contexts about related topics. Therefore, we introduced the core function of joining (clustering) similar individual clusters - i.e. those with similar tag sets - of different people to foci (R 2.1). Different degrees of interrelation in clusters are considered in this process while those individual clusters that do not have a sufficient degree of density are neglected completely (R 2.2).
Properties of a focus, like maturity (R 2.3) and constraint (R 2.4), are used in the later process of building social structures.

3.4.2.1 Overview of Requirements

Ref.	Requirement	In Prototype?	Comment
2.1	The system shall join personal clusters of all persons with similar main tags to foci.	Yes	Consider different levels of structure (cf. 3.4.1)
2.2	The system shall exclude personal clusters with insufficient density.	Partly	Depending on 1.7
2.3	The system should state the maturity of a focus (including breadth, height, activity and age).	Partly	very loose structured foci are removed
2.4	The system should state how constraining a focus is for each member.	No	

3.4.3 Building a Social Network

In the process of **building a social network within and between foci** we can summarize the following main functions: ties have been proposed in the prior step by joining people to foci. Tie strength between all members of each focus is determined based on the degree of friendship, the level of interaction and the number of foci people share (R 3.1). Friendship, the first of these three properties, may be adjusted manually (R 3.2) to emphasize a relation - for example as a result of acquaintance from outside the system. Just like friendship, interaction is a directed property which is updated whenever feedback is given. However, for tie strength both values are combined. Thus, an update of the tie strength with regard to changes in friendship values and interaction values has to be performed constantly (R 3.3). We emphasized in Chapters 2 and 3 that interaction is the crucial element that allows for shared contexts. As just mentioned it is shaped by feedback to recommendations. What is not considered here is the temporal quality of interaction. If people do not interact frequently or have not communicated for a long time, one has to assume that this is harmful for their shared context. Therefore, we evaluate whether a recipient ignores recommendations by a certain sender permanently and lower the interaction value in her direction accordingly (R 3.4).

Once relations among people within a focus are evaluated those who are tied most strongly are joined to groups in another clustering process (R 3.5). As a by-product the most central person of a group can be determined which is useful for bridging groups and foci by connecting their most active and influential persons to each other (R 3.6). As groups contain members of a focus who are tied very strong new ties could be proposed based on transitivity (R 3.7, cf. Chapter 2.3). Yet, we decided to consider all members of a focus as weakly tied (R 3.8). Though we will not rely on transitivity here it could be interesting to foster weak ties in foci that are quite strongly connected in terms of transitivity.

In Chapter 3.3 generation of an **overall social network** was proposed that is based on the average of all tie strength values of all foci two people share (R 3.12). For any strong relation from this network it could be examined if there are foci the related people do not share yet. A tie could be proposed as local bridge (R 3.9). We do not follow this exactly but forward recommendations in this manner (cf. 3.3.4.1). However weak ties are only established if positive feedback is given. Local bridges between foci may be proposed based on analysis of similarity of foci (cf. Chapter 3.3.4.1). Identification of foci with relevant similarity would be a precondition (R 3.10). Similar to groups central persons as well as those being located at the border of a focus (due to their archive's similarity to the focus and consequently their few relations to other members) could be identified to connect them (R 3.11). This way, topics that are not too popular in one focus (and therefore probably not suitable) could be proposed to other foci as innovation (cf. Chapter 3.3.3.2).

3.4.3.1 Overview of Requirements

Ref.	Requirement	In Prototype?	Comment
3.1	The system shall weigh the tie strength between every pair of members of a focus based on the degree of friendship, the level of interaction and the number of foci both share.	Yes	Depending on 3.2, 3.3
3.2	The system should provide the ability to set the degree of friendship to all related people.	No	
3.3	The system shall constantly update every tie strength due to changes in directed interaction values or changes in directed friendship values.	Partly	Without friendship calculation, depending on 3.2 and 4.10
3.4	If a person does not consider another person's recommendations and both already have a high level of interaction, the system should decrease the level of interaction towards this person over time.	No	
3.5	The system shall join strongly tied people within a focus to a group.	Yes	
3.6	The system should provide information about the most central person within a group.	No	

3.7	The system should propose ties between unconnected group members based on transitivity.	No	
3.8	The system shall propose weak ties between all members of a focus	Yes	
3.9	If two people have a high degree of friendship and share several foci, the system should propose a tie (local bridge) between two foci they do not share.	No	depending on 3.12
3.10	If two foci are similar but not connected, the system should propose a local bridge between two persons of each foci.	No	
3.11	If a bridge between two foci is proposed, the system should direct the new tie from the most central person of one focus to a person at the border of the other focus	No	
3.12	The system shall calculate the average of all tie strength values of all foci two people share to build an overall social network	No	

3.4.4 Recommendations and Feedback

In the fourth step we identified the following core functions. The **recommendation engine** will select items that are not yet recommended from the personal archive (one or more individual clusters). Those items are forwarded within the focus to all members (R 4.1). On any recipient's side a recommendation will be evaluated regarding the tie value to the sender and a comparison with the recipient's personal archive of a focus (R 4.2). The results will be weighted and averaged and inserted into a recommendation for the focus (R 4.3). We will realize all these functions at least roughly in our prototype whereas the following additional tasks of the recommendation engine will not be implemented: sorting of recommendations into an overall recommendation list (including weighting by constraint of foci, R 4.4), dealing with multiple recommendations of a single item as shortly discussed in 3.3.4.1 (R 4.5) and forwarding of recommendations on the overall network to detect and use potential local bridges (R 4.6). For any item recommended the predecessor of a sender has to be stored to allow creation of weak ties based on feedback (R 4.7).

Our **feedback mechanism** combines several functions as well. Feedback is obtained by evaluation of tags chosen for archival and optional rating (R 4.8). We will not follow our proposal of implicit measurement of reading frequency in our prototype (R 4.10). Once feedback values are at hand they are weighted and averaged and the interaction value to the sender is adjusted accordingly (R 4.11). As we will not implement bridges from the overall network we will not need indirect feedback (R 4.12) and storage of tie strength of bridges between foci (R 4.13) as well.

Summing up, we will concentrate on recommendations over weak and strong ties within foci and basic feedback.

3.4.4.1 Overview of Requirements

Ref.	Requirement	In Prototype?	Comment
4.1	The system shall forward any items that have not yet been recommended from each person's archive to all members within her focus.	Yes	
4.2	The system shall evaluate received recommendations regarding the tie value in direction of the sender and a comparison with the recipient's personal archive of a focus resulting in a feedback value for each.	Yes	
4.3	The system shall weigh and average both feedback values and arrange it to a recommendation list accordingly.	Yes	Depending on 4.2
4.4	The system will sort recommendations into an overall recommendation list (including weighting by constraint of foci).	No	
4.5	The system should deal with multiple recommendations of a single item by giving weighted feedback to the recommender's predecessors.	No	
4.6	The system shall forward every person's items for recommendation also to her direct ties on the overall network to detect and use potential local bridges.	No	Depending on 4.1
4.7	The system shall remember the predecessor of its sender for each recommendation sent on the overall network.	No	Precondition for 4.6
4.8	The system shall provide people the ability to archive (including tagging) and rate recommendations.	Yes	
4.9	The system shall calculate feedback for any archived recommendation by evaluation of tags chosen and an optional rating resulting in one or two feedback values.	Yes	Only one feedback level
4.10	The system shall provide a feedback value from reading frequency of a recommendation.	No	
4.11	The system shall weigh and average feedback values and adjust the directed interaction value to the sender accordingly.	Yes	Depending on 4.9 and 4.10
4.12	The system shall provide for indirect feedback (feedback also to a sender's predecessor).	No	Includes 4.9 through 4.11
4.13	The system shall store the tie strength of bridges between foci.	No	

3.4.5 Social Awareness: Basic Requirements for Interface Design

We developed requirements from an analysis of how to create social awareness in the area of foci as an example. They will guide the interface design in the next chapter.
We also identified interface modules which we consider as building blocks that are reassembled according to different goals.

3.4.5.1 Summary of Requirements (for foci)

Ref.	Requirement	In Prototype?	Comment
5.1	Whenever a focus or a group is visited the system shall make transparent that items are recommended by certain other people and response to recommendations is possible in terms of feedback.	Yes	
5.2	The system shall enable the user to conceive that recommendations and feedback are exchanged between other members of a group or focus.	Yes	Only implicitly, cf. Chapter 3.3.4.2
5.3	The system shall present the (constantly updated) social network of a person's group members according to communication.	Yes	Only for groups due to simplified display
5.4	The system should reveal whether recent communication brought people closer together or degraded their relationship.	Yes	Not explicitly, but traceable from network changes
5.5	The system shall show the social network from a third person perspective.	Partly	View not entirely the same from all viewers' perspectives

In the next chapter we will describe design, implementation and evaluation of a prototype realizing the requirements mentioned.

4. *GROOP.US* - Development of a Prototype

In this chapter we will document the development of a prototype called *GROOP.US* implementing the basic requirements mentioned in the previous chapter. We will start with introducing design considerations regarding architecture and interface of *GROOP.US*. Next, technical fundamentals and constraints of the practical implementation are described, followed by a detailed overview of the functionality and interfaces in form of a step-by-step guide. Finally, we will verify if the prototype is able to fulfill the stated goals and requirements and specifically mention important observations and drawbacks.

4.1 Design

4.1.1 Architecture

In the following we will describe an approach to realize the social navigation framework introduced in Chapter 3. We will concentrate on central aspects of each of the framework's steps. For each step we will give an overview of expected modules and their interaction realizing the requirements stated in the previous chapter. Constraints regarding the implementation are not defined, this will happen in Section 4.2.

Relying on data from weblogs and social web services accessible through HTTP requests and offering a web interface for user interaction the core system will be a server side web application. We will not go into detail explaining modules and techniques for server side interface generation and client side interface display.
Our envisioned system architecture is divided into two major parts: the Framework Controller and the Core Framework (Fig. 4.1). The Framework Controller stores all instances of all classes of the Core Framework and provides means of communication between those instances. Furthermore the Framework Controller controls user interaction and interface generation based on data provided by the Core Framework. A CRON-Job module starts repetitive processing tasks. Finally, the User-Management module controls user registration and relates authenticated users to their profiles.

The User Profile module is located in the Core Framework package. This package provides the main functionality of the framework. A user profile holds tie information (i.e. friendship and multiplexity, cf. 3.3.3.1) to all other related people in the system and provides a ranked inbox for incoming recommendations. Furthermore pointers to all individual clusters a person has, the foci and groups she is in and the feeds she is subscribed to are provided.
A person's feeds are frequently updated by the Feed Grabber module. Sources for such feeds are RSS ([Brickley2000]), Atom ([Nottingham2005]) or RDF ([Beckett2004]) streams. The feed grabber extracts creation date, title, description, tags, subjects or categories and the URL of the annotated resource and stores these data for each feed in the Feed module. Alternatively, plugins for APIs of web services can be provided that extract comparable information using REST-compatible queries (c.f [Prescod2002]). The third option is to load a locally stored XML file, providing a RSS-like notation.

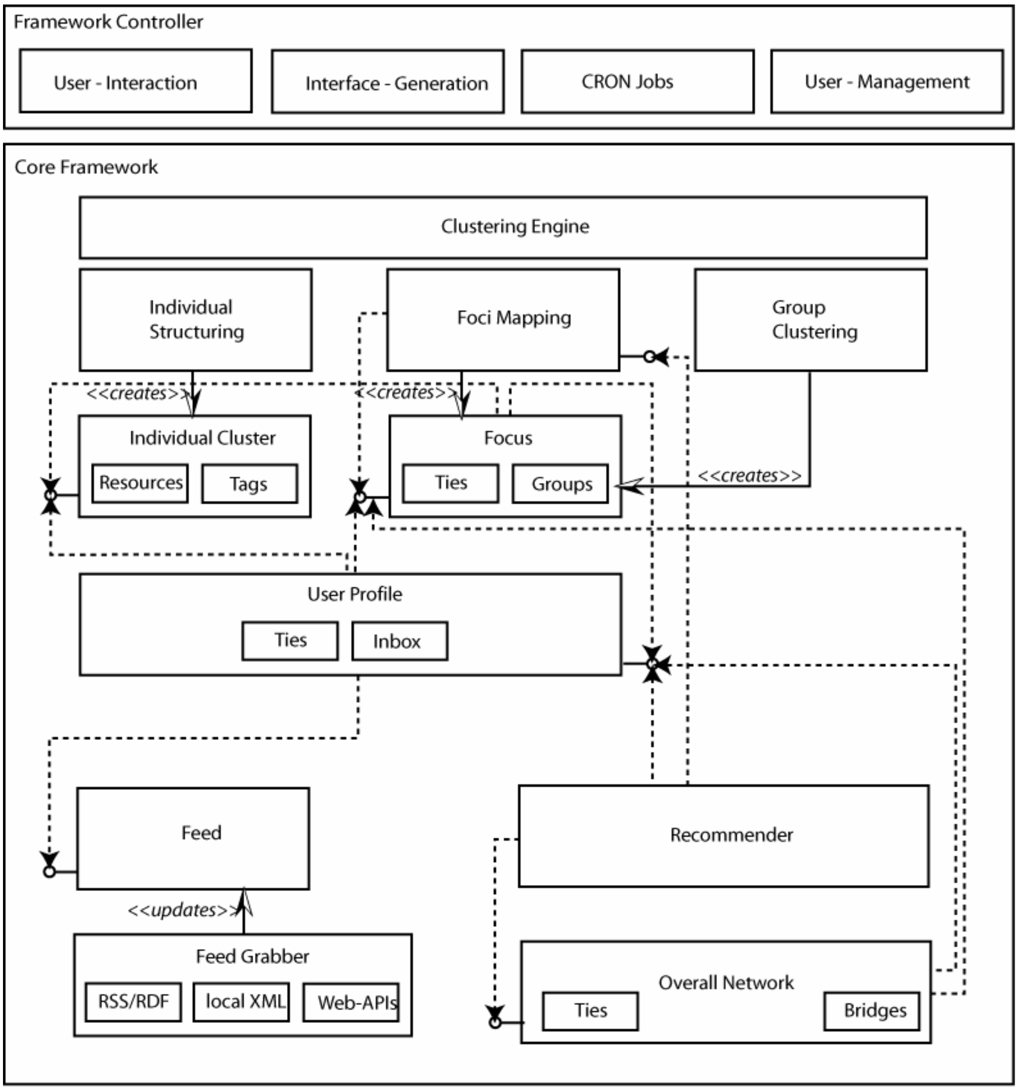

Figure 4.1: Overview of the prototype's system architecture

Based on information extracted from a person's feeds the first step of our framework is performed (cf. 3.3.1): the structuring of individual resources. Like the Foci Mapping and Group Clustering modules the Individual Structuring module is driven by a clustering engine, implementing a hierarchical clustering approach. The result of the individual structuring process is a number of individual clusters holding resources with similar tags. The Framework Controller starts clustering for each person in the system and relates individual clusters to her user profile.

The second step of the framework - mapping individual clusters to foci - is also initiated by the Framework Controller, which provides all individual clusters of all people to the Foci Mapping module. This module creates foci of individual clusters with a similar tag distribution. Pointers to foci a person's individual clusters are part of are stored in her user profile, additionally pointers to all generated foci are stored in the Foci Mapping module for user independent access.

After creation of foci each focus initiates the third step of the framework: proposing a social network. The Focus module provides methods to calculate the initial interaction value as one part

of the tie strength between all people in a focus by comparing the similarity of their tags in their individual clusters, accessible through their user profiles. The second part of a person's tie strength - the multiplexity value - is stored in the person's user profile and is increased by each new focus that points to the profile. The friendship value - the third part of the tie strength - is set manually by the user, which is controlled by the Framework Controller. After the foci mapping process is finished, the Group Clustering module calculates groups of the strongest tied persons within a focus based on the initial tie strength values. Groups are stored within foci and are additionally accessible through a person's user profile.
The Overall Network module stores the average tie strength of two people. It utilizes the User Profile module to access people's multiplexity and friendship values, as well as the Focus module to calculate the average interaction value between two persons in all shared foci. Additionally, the Overall Network module proposes bridges between related foci.

The fourth step - deriving recommendations and feedback - is triggered by the Recommender module. Through the Foci Mapping module each user profile related to a focus is accessed and a list of recommendation candidates from all individual clusters within that focus is compiled. Then, the recommender sends that list to all inboxes of people within the focus in question. After receiving a list of new recommendations the user profile ranks all items in its inbox according to the tie strength value to its sender and in accordance to the item's tag set with the tags within the user profile's individual clusters related to the focus the item originates from.
Recommendations may not only be sent within foci, but also between foci. For this reason the recommender utilizes the overall network to send recommendations from one focus to all foci it has bridges to.

Finally feedback is provided by tagging and rating recommended resources. The Framework Controller provides the interlink between a user's action and the adequate response within the focus the recommendation originates from. For this reason the Focus module provides methods to estimate the congruence of tags the receiver has chosen compared to the tags the sender had previously used. Based on tag congruence and a rating the interaction value from receiver to sender within this focus is adjusted.

The CRON-Job module of the Framework Controller does an hourly recalculation of recommendations and group affiliation. Re-clustering of individual clusters and foci is done on a daily basis.

4.1.2 Interface Design

From the requirements definition (and the restrictions defined regarding the implementation) four major use cases for our prototype can be derived which are: **(A) watching your areas of interest**, **(B) observing foci (in terms of tag clouds and the social network within)** you belong to, **(C) receiving recommendations** for foci and **(D) giving feedback to recommendations.**
The first step in interface design is to decide which interfaces are required to enable users to fulfill the use cases identified. Afterwards their functionality will be refined.

Designing for Social Awareness

In Chapter 3.3.5 we elaborated two guiding questions which shall ensure that the social affordance and a changing social environment of the application is supported by each interface. For interfaces that deal with a focus we already developed basic requirements (based on those questions) which have been defined in 3.4.5.1.

Use Case A does not contain any social activities. It shall help people to grasp their individual memory and show how their tagging behavior so far lead to piles of tags with a certain structure and related resources (cf. Chapter 3.3.1.1). As there will be no means of manipulation of personal clusters in our prototype (cf. 3.4.1.1) this interface is quite simple. It consists of a tag clouds for each cluster. Tag clouds are often represented as boxes with tags separated by space. The size of each tag is related to its importance, that is the number of occurrence (cf. Figure 3.2, Chapter 3.3.1.2). We chose a circular representation where tags that are most important are in the center and less important tags are distributed on "orbits". Important tags are bigger in addition.
Their position in space does *not* express the distance of a pair of tags. Figure 4.2 shows a personal cluster as designed for our prototype.

Figure 4.2: Draft of a personal cluster as circular tag cloud.

This kind or representation allows to get important tags of a cluster at a glance. Moreover people can judge the structure of a cluster from its appearance. Highly hierarchical clusters will have few main tags in the center and many tags on the outer orbits. The space within will be empty. Clusters with less hierarchy will show an evenly distributed pattern. The icon in the upper left points up that the cluster is related to an individual (as opposed to a focus cluster, see below).

Having said that watching your individual memory does not contain social activity does not mean the use case is not important for the overall social scenario. Changes in the structure of the individual memory over time as a result of communication shall be pointed out to the user. The user might experience this by just looking at her personal clusters occasionally. Illustrating changes in the tag clouds might foster comprehension but we will not elaborate this here. Moreover individual areas of interest manifest people's relation to a focus. For reasons of transparency and trust into recommendations later they shall be able to understand why they are in a focus and how their relationship to this focus is. This is why we will combine the interfaces for Use Case A and **Use Case B**, observation of your foci. But before introducing the combined interface we will have a closer look at the representation of foci.

Basically we can distinguish two views on a focus. Its topics and their structure on the one hand and its social network on the other - corresponding to the building blocks foci and foci network mentioned before. For the former we will use the circular tag cloud again. Figure 4.3 shows a tag cloud of a focus. It contains the number of people in the focus which is printed next to the icon in the upper left which symbolizes that we are now dealing with a tag cloud that represents several people. In respect to social awareness, this is the first indicator of the social capabilities of the application.

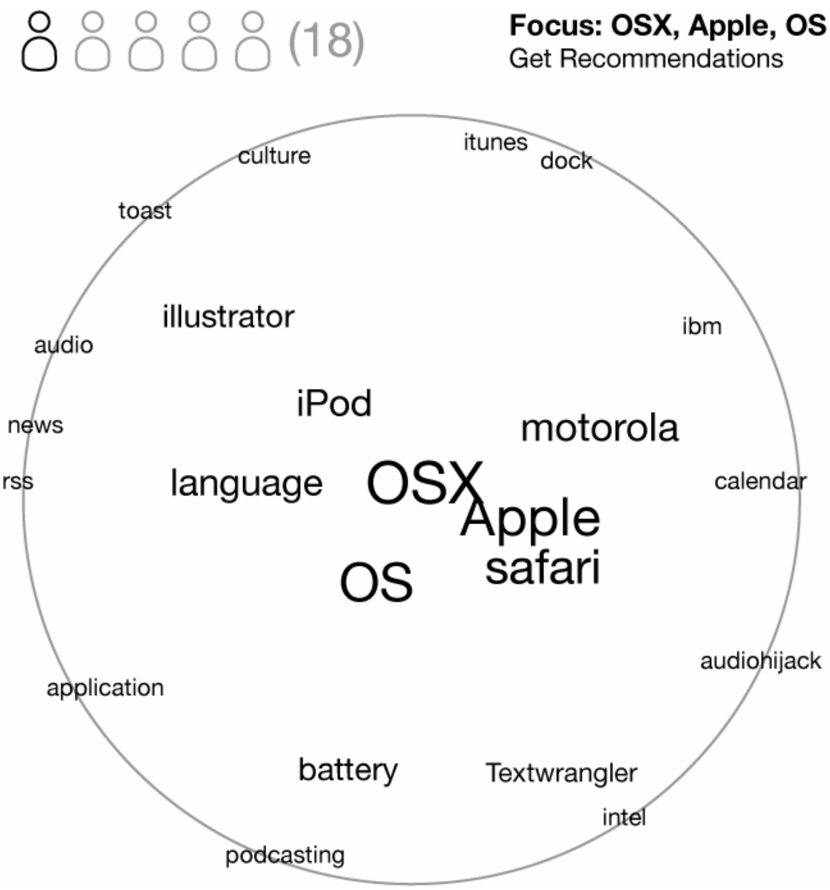

Figure 4.3: Draft of a focus as a circular tag cloud.

Main Interface 1: Personal Clusters and Foci

In our opinion observing the social network of a focus is not too helpful when evaluating the relation of personal clusters and one's foci. Therefore we will elaborate on the foci network later when coming to use cases C and D. For now, the first of two basic interfaces of our prototype shall be presented called "Personal Clusters and Foci" (see Figure 4.4).

The screen is divided into three parts. On the left personal clusters are shown. In the middle tag clouds of foci a person belongs to are displayed. They are drawn bigger to emphasize that they contain personal clusters. Finally to the right resources of selected personal clusters are shown. The confrontation of personal clusters and foci allows for comfortable evaluation of relations. Personal tag clouds can be selected and related foci will be highlighted. Tags from the personal cluster will be highlighted in the focus so that conclusions about relation to a focus can be drawn. In the right

column resources archived in the selected cluster are displayed with their tags.
Sometimes more than one individual area of interest was assigned to a focus. In this case related personal clusters will be marked as well in a weaker color including their tags in the focus as illustrated in Figure 4.4. However this feature will not go into the first prototype.

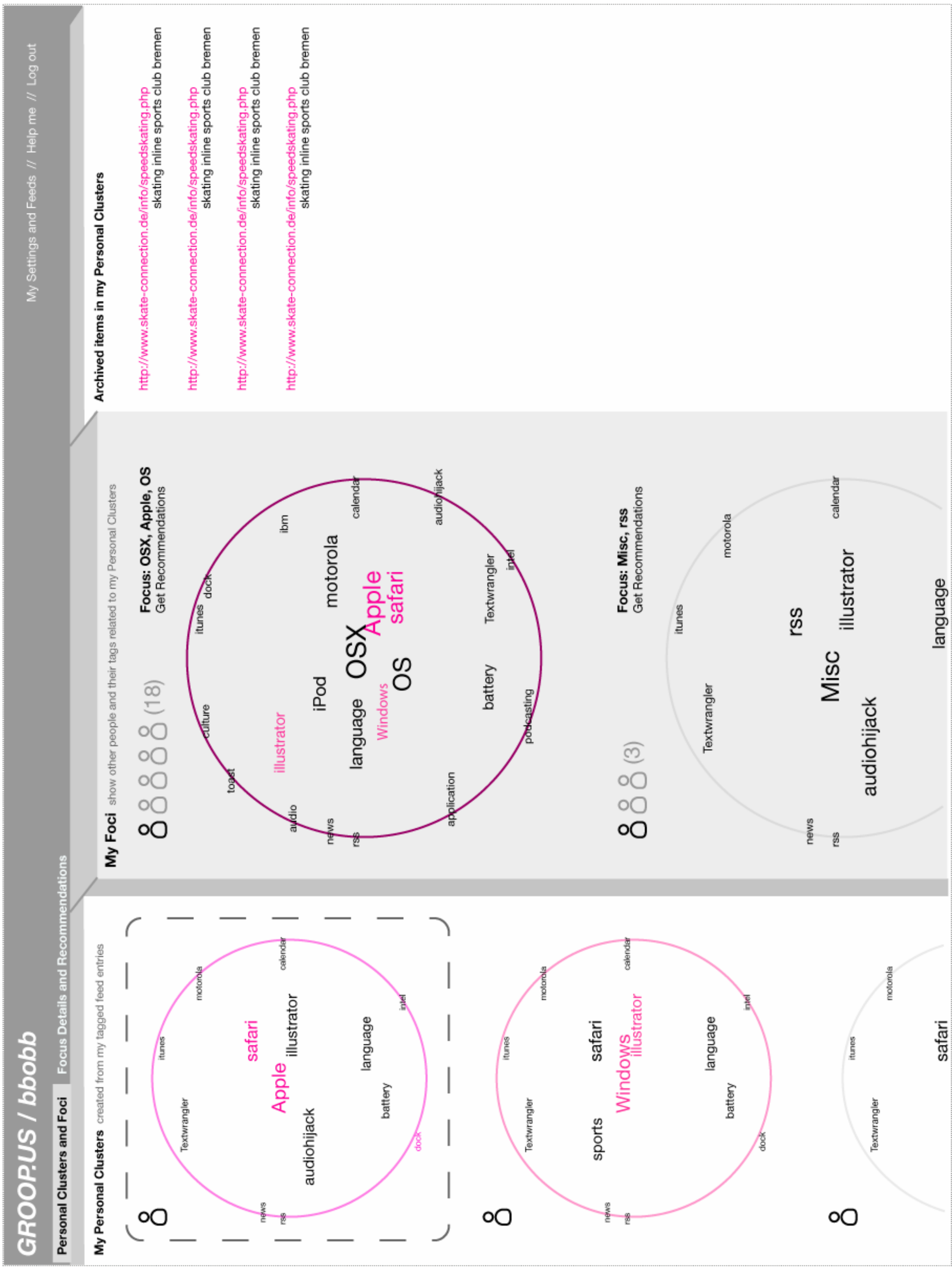

Figure 4.4: Draft of a "Personal Clusters and Foci" interface.

The left and right frame have the same background color to indicate that they belong together. The middle frame is accentuated by a different background color and the slight 3D effect. This way we want to indicate the importance of the foci level and its location above the personal level.

People can leave this interface by using the main navigation in the upper left or by clicking "Get Recommendations" at each focus. This will lead them to our second interface that contains the social network part of Use Case B, as well as the use cases C and D.

Main Interface 2: Focus Details and Recommendations

The focus and its social network are the central element in this interface. They ensure that the user is aware that recommendations she receives are based on her presence and activities around certain topics. Therefore we will now discuss how the social network of a focus may be represented (Use Case B).

As discussed in earlier chapters people within a focus are divided into groups of people who are most similar and interact most. The interface has to show these groups, relation within groups and between groups. In Chapter 3.4.5.1 we defined a third person perspective on the social network as a requirement (cf. also Chapter 3.3.5). Therefore ties between people originate from bidirectional interaction values. Thin lines show a low overall interaction value while thick lines represent relations with high interaction in both directions. Figure 4.5 shows our draft for the social network of a focus.

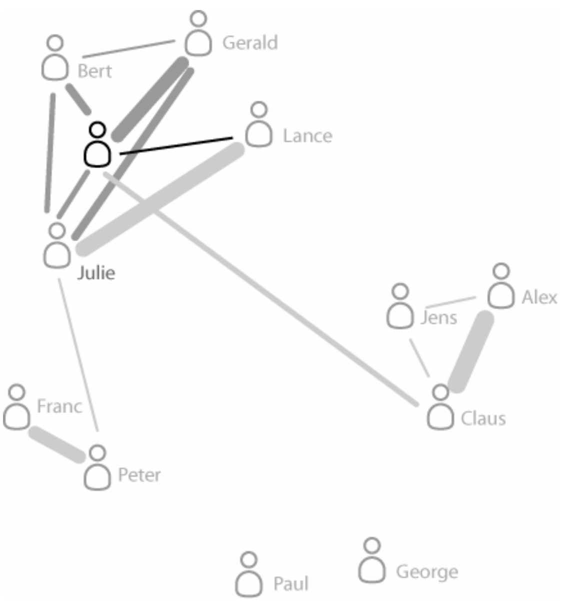

Figure 4.5: Draft of a social network diagram of a focus.

The dark person is the user (viewer). Optionally the line color might be used to give information about the tie strength directed from the user to her neighbors in the network (as opposed to the bidirectional value). An example: a thick light line shows high averaged bidirectional interaction while a thick dark line indicates that you already gave a lot of positive feedback.
Displaying networks dynamically in this manner is quite complex to realize. Thus, for a first prototype we chose a simpler solution as illustrated in Figure 4.6.

Similar to circular tag clouds people are located on orbits around the viewer. People who are close to the viewer are on a close orbit, those who are far on far orbits. Relations between other people in the group are not contained. Other groups are shown in lists of people. Relations between groups are not displayed either. Even though the network is viewed from a third person perspective it does not display the same situation to all users any more. The person in the middle is always the viewer and only relations between the viewer and others are displayed. Yet the viewer still gets the impression to be a part of the group which might lead to social behavior as discussed in Chapter 3.3.5.
Even though this solution is inferior to the former one it should be sufficient to create social awareness and support the following use cases.

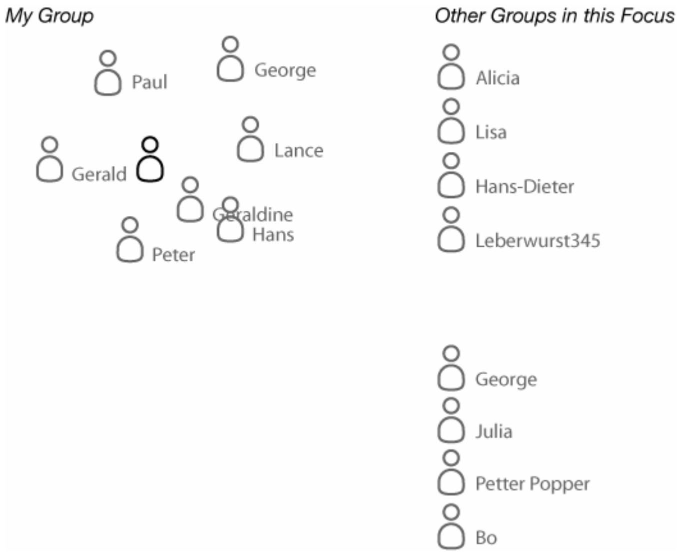

Figure 4.6: Refined Draft for implementation of a social network diagram of a focus.

Use Case C - receiving recommendations - requires three different interfaces: a reader which includes a listing of incoming recommendations, the focus tag cloud and the focus network. Incoming recommendations are listed by title and name of the sender. They are ranked by their relative importance which will be displayed as percentage value in *GROOP.US* later. Displaying the content of a recommendation is quite simple. Any data coming with the URL such as title, description and date is shown. Moreover the tags used by the sender are displayed. The URL is a link to an external window. It might be an option to show the target in an inline window.

Figure 4.7 shows the recommendation interface we plan for our prototype. In the left column the focus is displayed as tag cloud (1) and as social network (2). In the right column we have the reader interface (3) with the list of recommendations on top and the detail view for each item below.
In requirement 5.1 (cf. Chapter 3.4.5.1) transparency of the recommendation functionality is claimed. Therefore, whenever a recommendation is selected the sender is marked in the focus network interface (2). This way the recipient is able to evaluate why he got a recommendation (because the sender is in the same focus) and how to judge its relevance (higher for people in her own group and for people who are closest). To further evaluate relevance the tags used by the sender of a recommendation are highlighted in the focus (1). These tags (golden/green) can be compared to the central topics of the focus or the tags that the recipient uses (pink) which are still highlighted from the ”Personal Clusters and Foci” interface.

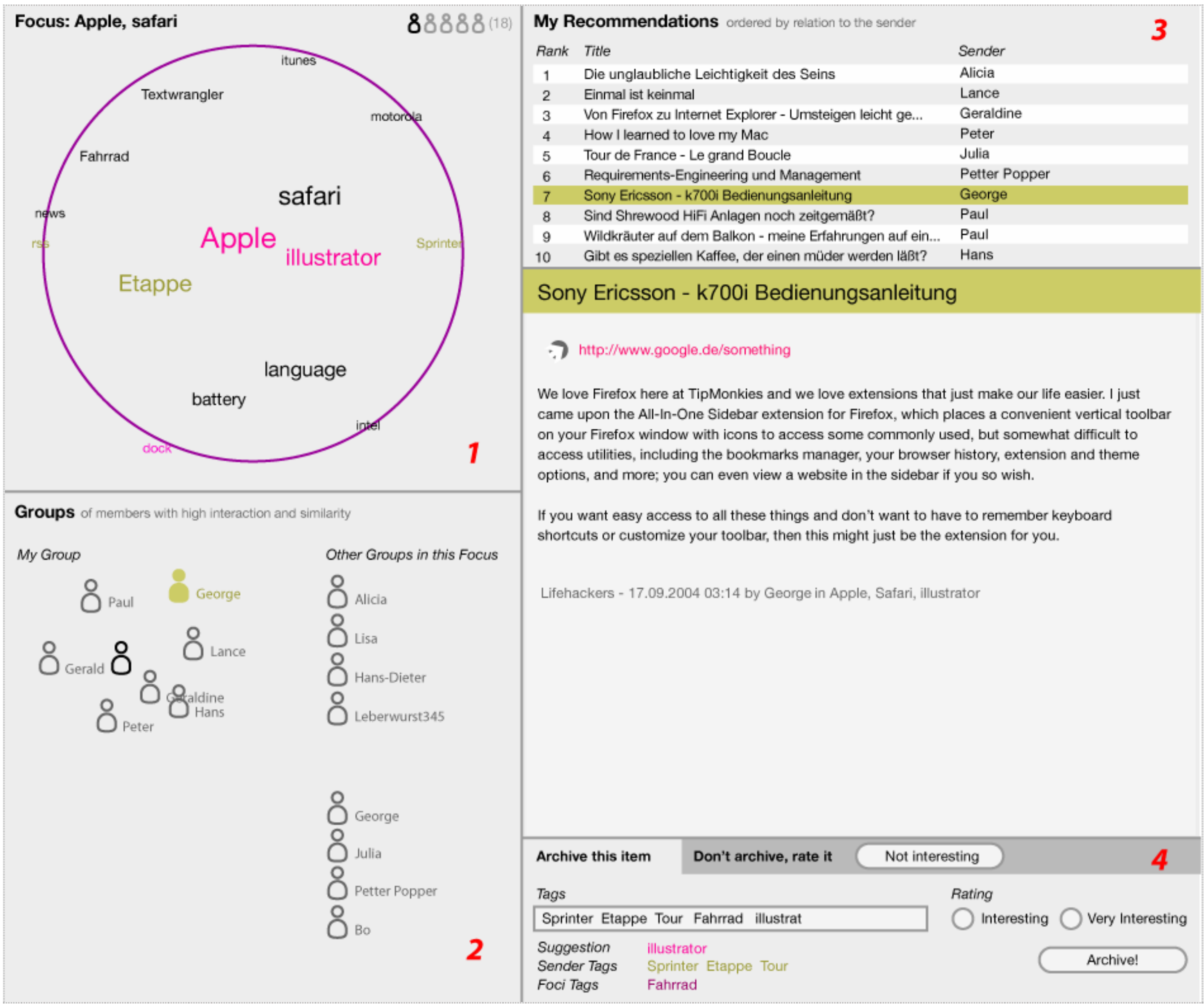

Figure 4.7: Draft of a recommendation (1-3) and feedback (4) interface.

As postulated in requirement 5.3 (cf. Chapter 3.4.5.1) the social network (2) is updated after each feedback (see below). People may get closer or move away within the recipient's group, leave the group or join the group. In *GROOP.US* we will at first only realize changes within the group because moving across groups requires a new clustering which is not done automatically at the moment. By updating the network after each action requirement 5.4 is fulfilled as well: whenever the feedback triggered a decisive change in the relation to a sender, this is comprehensible in the social network immediately. It can also be experienced in the list of recommendations because items of an interaction partner will be ranked higher or lower in the future.

Figure 4.8: Draft of a feedback interface.

Figure 4.7 has a feedback interface (4) included, too (see also Figure 4.8). It basically differentiates between appreciating a recommendation (and following archiving and rating) and rejecting it voting it "not interesting" (cf. Chapter 3.3.4.2). These two sections are differentiated by two different background colors.
If the item is appreciated it can be archived and optionally rated. The input field for tags is initialized with the sender's tags for adoption. Tags may be erased and added arbitrarily. Below the input field tags that belong to the focus are suggested when typing. This way different usage of a word stem shall be prevented and focus tags shall be boosted. The two lines below called "Sender Tags" and "Foci Tags" are constantly updated. Whenever a tag is inserted in the input field that belongs to the sender's tags or the set of tags belonging to the focus respectively it is listed. This function will help the user to comprehend the effect of a feedback. Many tags you have in common with the sender will foster the relationship and the number of foci tags allows to predict if the resource fits to the current focus - in the recipient's opinion - or will be assigned to a different personal cluster later.
In our initial prototype we will only have a plain input field.

The rating interface is kept simple. According to our framework an explicit statement about interest in the recommendation is possible. We go for a textual representation in *GROOP.US*. Icons ranging from "thumbs down" to "thumbs up" might be an alternative.

Finally figure 4.10 shows the complete second main interface. It is divided in three columns as well. The middle and right column contain the focus environment with reader and feedback interface. The left column shows all foci of a user and the personal clusters that are related (indented) in a short overview. Here only the most important tags are displayed. The whole column is set back by a 3D effect and has a different background color to distinguish it from the functions related to the focus that is selected at the moment. The foci may be selected and the right part will show the details of the selected focus. This way a navigation through all foci of a user is realized. In the personal clusters tags that occur in the focus as well are highlighted as in the "Personal Clusters and Foci" interface. Moreover tags of an incoming recommendation are also highlighted in the personal clusters. This complements the highlighting of tags in the focus mentioned above because the recipient gets direct feedback about common used tags.

Conclusion

Finally let us recapitulate how the social affordance of *GROOP.US* and the property of foci - being an ever changing social place - is conveyed. Social affordance is mainly realized by creating a relation between individual areas of interest and foci and then making clear that items are recommended by other people in a focus. The consequence of social interaction for the place it happens in is conveyed by constantly updating the social network as well as the tag cloud for a focus.

These social navigation capabilities will be described in more detail in the next chapter (4.2.2) when will talk about the implementation of *GROOP.US*.

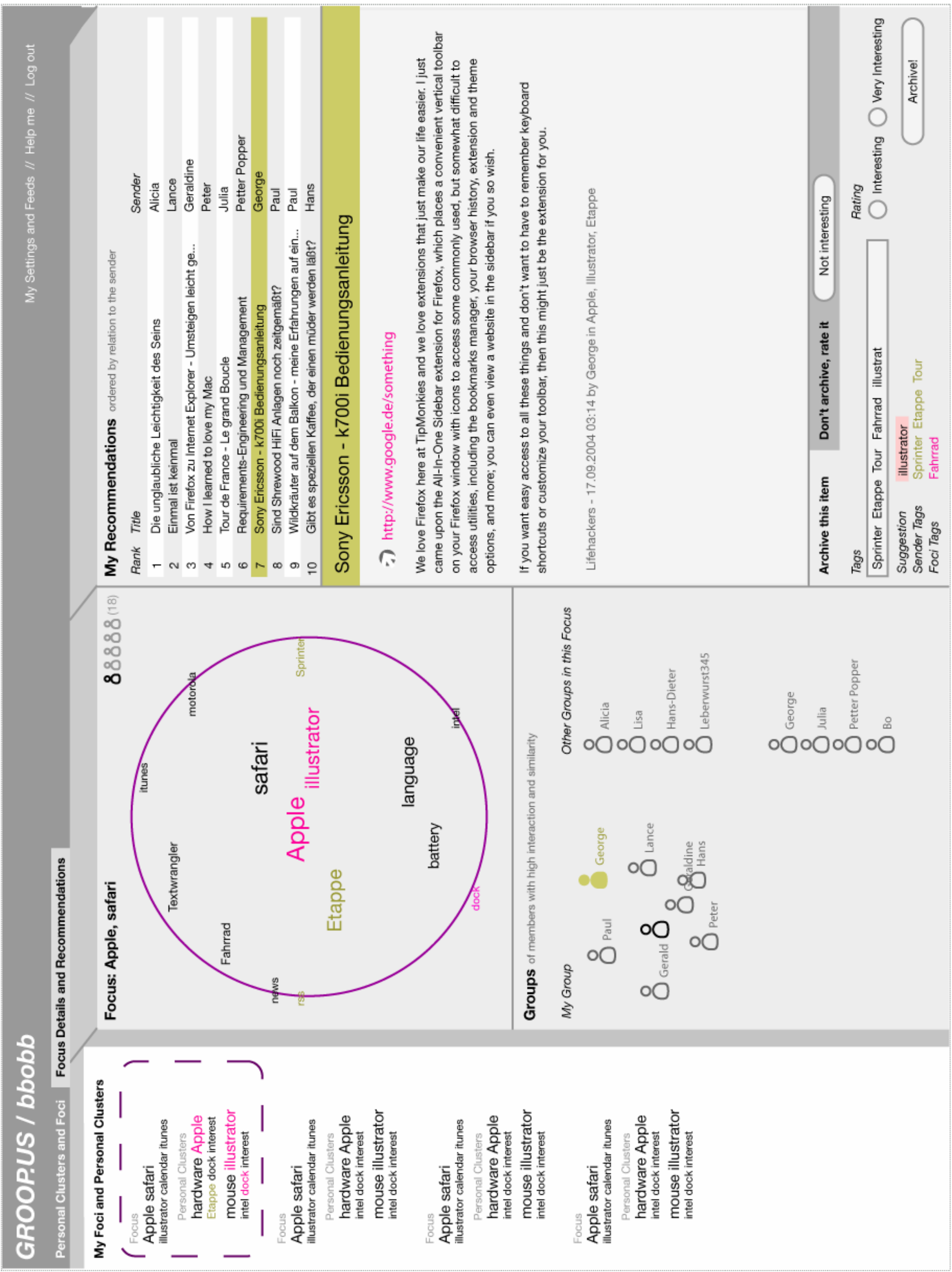

Figure 4.10: Draft of a "Focus Details and Recommendations" interface.

4.2 Implementation

4.2.1 Technical aspects

For realizing *GROOP.US* as a web application we rely on the following systems and programs. On the **server side** only open source software is used. The programming language of choice is Python (http://www.python.org). Advantages are on the one hand the possibilities for script-like rapid software development and on the other Python's strong support for object oriented programming paradigms.
To reduce efforts we base our development on an existing application server called Zope (zope.org) and a content management system named Plone (plone.org). Both are based on Python. We mainly reuse the user management and the templating system. Furthermore Zope is driven by an object oriented database called ZODB, which allows persistent object storage without writing any specific database storage or retrieval methods (like SQL adaptors). The hierarchical clustering algorithm used for clustering individual clusters, foci and groups is based on the Orange data mining library (http://www.ailab.si/orange) with specific clustering extensions (http://ai.fri.uni-lj.si/%7Ealeks/orng/).

The **front-end** runs in modern web browsers and is based on W3C standards like XHTML 1.0 and CSS. Javascript will be used for dynamic interface manipulation. The system was successfully tested in Firefox 1.0.X, Internet Explorer 6, Safari 2.0 and Opera 8.0.

4.2.1.1 Implementation details and constraints

The technical implementation follows our architecture design very closely (cf. 4.1.1). Figure 4.11 gives an overview of the used classes. A few variations and constraints are worth mentioning:

Besides controlling functions already mentioned in the design of the architecture the FrameworkController connects the Core Framework and the application server. For this reason the FrameworkController inherits directly from Zope classes. It holds instances of most classes including the GroupClustering, IndividualStructuring, FociMapping and the FeedGrabber. Furthermore it stores a list of all UserProfiles. The listed methods are called by Zope page templates mediating information exchange between these instances and user interaction.

The most obvious difference is - as already stated in the requirements definition in 3.4 - that due to time restrictions the module for calculating the overall network is missing. Accordingly the Recommender module only evaluates information for one single focus at a time. For this reason we integrated recommender functionality directly in the Focus class and the FrameworkController. Regular repetition of the clustering processes and recommendation of resources is not automatically done by a CRON-Job module, but has to be triggered manually through an administration interface.
Regarding the different methods for retrieving streams of annotated resources as input data for creating individual clusters only the import of locally stored XML files is supported. Developing a parser for various different file formats, adding validation routines and implementing a web

crawler for collecting and updating externally stored feeds is not directly related to the core problems we are dealing with.

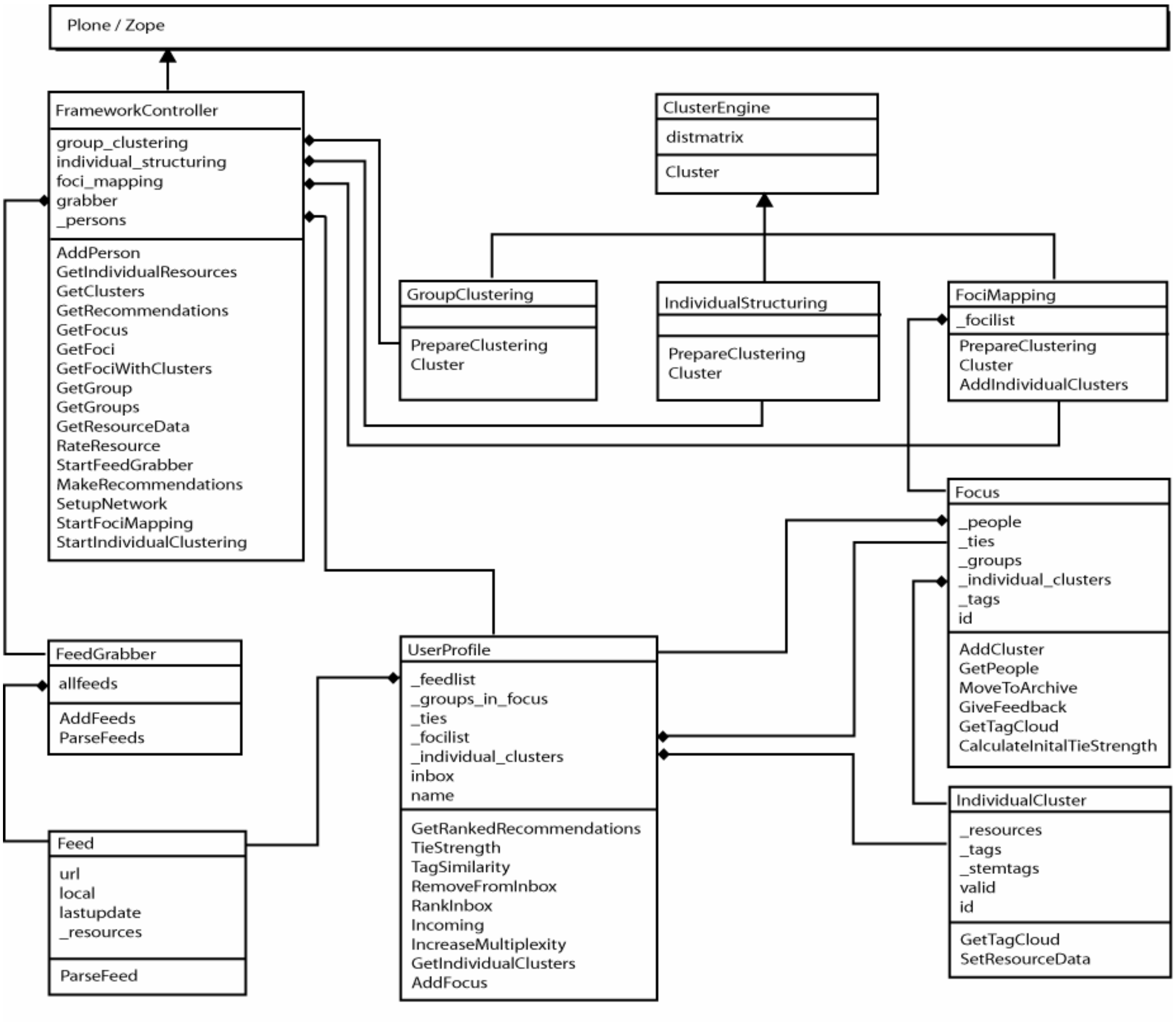

Figure 4.11: Class diagram of the prototype's system architecture

Although not integrated into the first version of *GROOP.US* the import of tagged resources form the social web service del.icio.us using its web API was possible in one of our first case studies (still available at http://laurie.informatik.uni-bremen.de/clusty) to evaluate the potential of tagged based clustering. The locally mirrored data is the basis for building individual clusters in the final prototype.

The current data in the online version of *GROOP.US* includes data sets of 1000 people, imported from the social web services del.icio.us. These data sets include over 130.000 annotated bookmarks (note that the import was restricted to 500 bookmarks per person). Calculating individual clusters, foci, groups and an initial number of recommendations takes between 1.5 and 4 hours depending on server performance (calculations were done using a Mac G4 1.2 GHz and a Xeon 2.4 GHz).

4.2.2 Documentation of Functionality and Interfaces

The purpose of this chapter is twofold. It is a high level documentation of the functionality of our prototype on the one hand and shall serve as a guide to our application as well. For this reason we will go through our interfaces step by step explaining all functions and how to handle them. The rest of this chapter addresses you as a *GROOP.US* user directly even though the system may only be used with one of the sample user at the moment. As *GROOP.US* will be evaluated in detail in later sections this guide also serves as an introduction for people not knowing the online version.

4.2.2.1 Getting Started

To visit our web application "GROOP.US", go to http://groop.us.
Our application is best viewed at a screen resolution of 1024 x 768 pixels or higher and is successfully tested in the latest versions of Firefox (Version 1.0.X), Internet Explorer 6 (Windows) and Safari 2.0 (Mac OSX).
On the start page you will find a login box. In this test version you can only log in as one of the sample users. Please, select one from the list and state the password you were given before. You are forwarded to an interface that contains your personal clusters and related foci (see Figure 4.12).

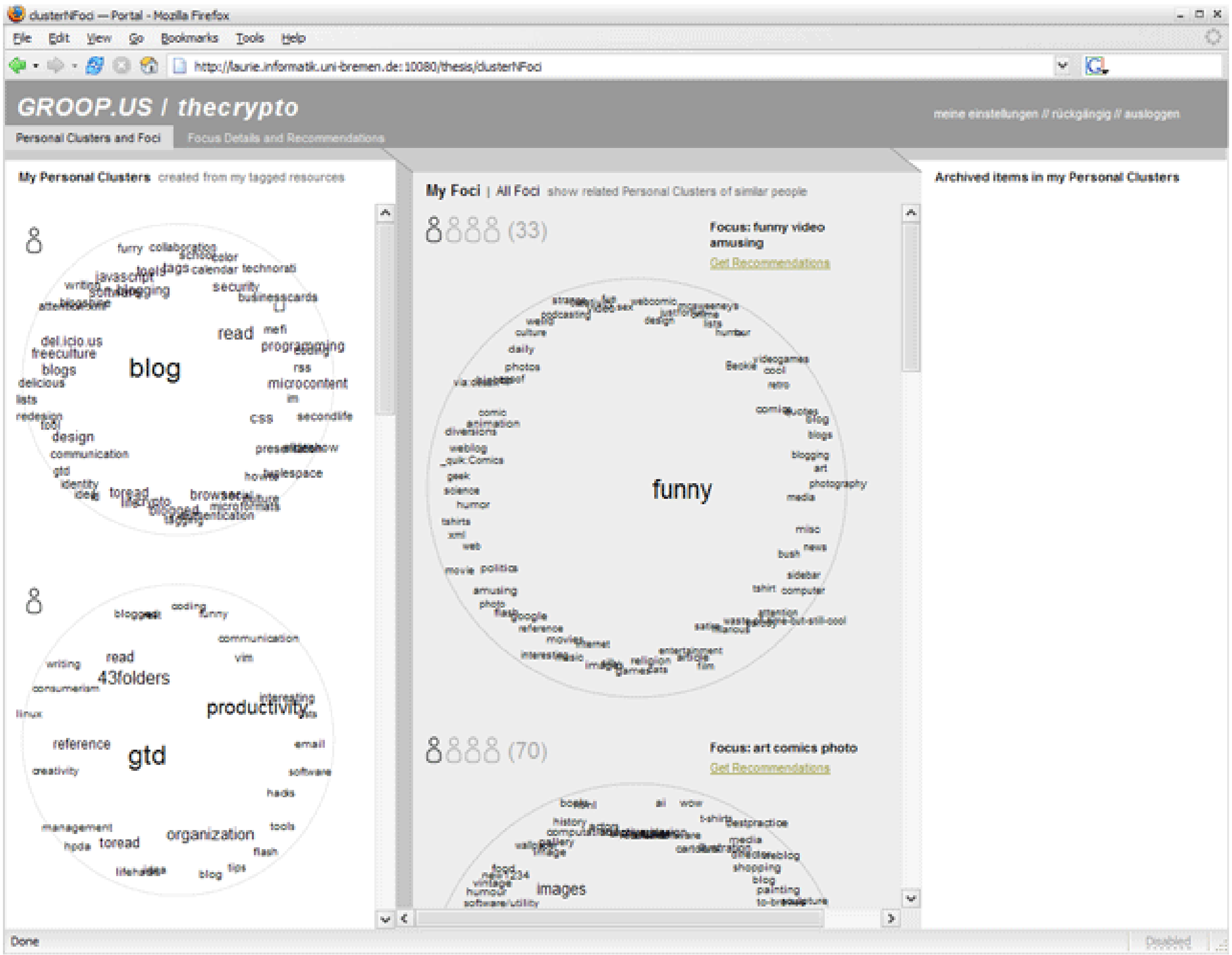

Figure 4.12: The interface "Personal Clusters and Foci".

4.2.2.2 Exploring Personal Clusters and Related Foci

In this section you can watch your areas of interest as calculated by the system based on your tags in the left column. Each circle represents an area of interest which is best described by the main tag(s) in the center. You can see if your clusters combine many topics, if there is a strong hierarchy (few tags in the center that are used often and many tags in the periphery that are only used occasionally) or if tags are more evenly distributed indicating a flat hierarchy. If you cannot read a tag, try to hover it with your mouse. The text will be scaled to a bigger colored font size. The tag can also be read from the small info box that pops up (see Figure 4.13).

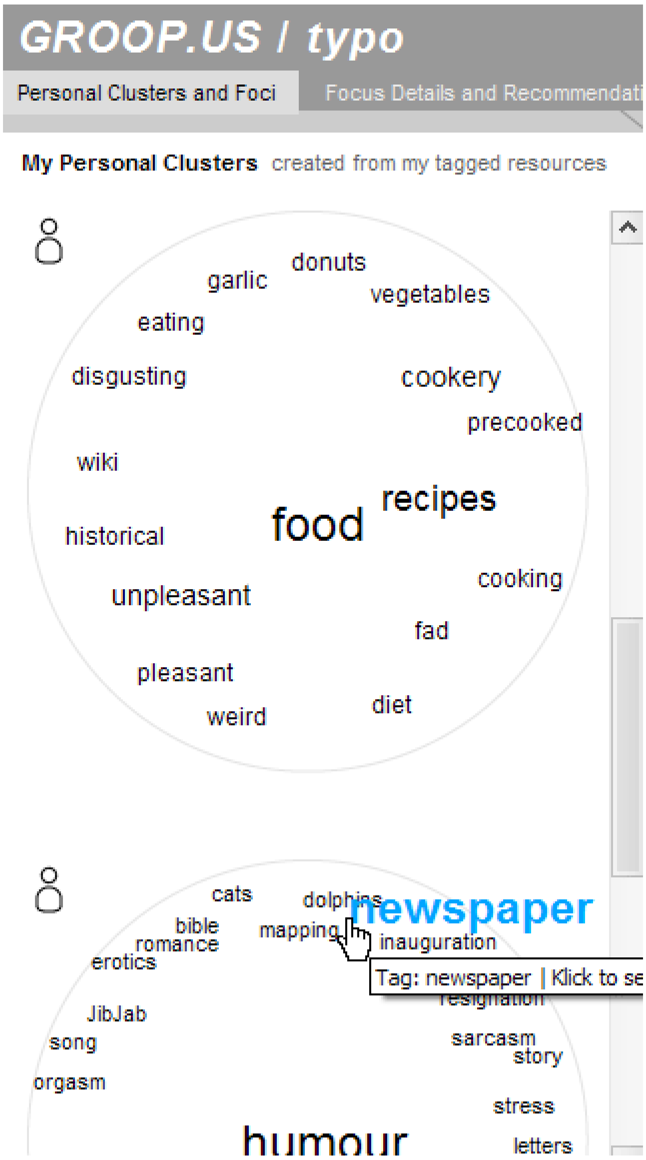

Figure 4.13: Individual Clusters of food and humour related resources.

The size and position of each tag in a cluster is not related to an absolute measure but shows the importance of each tag related to the overall number of tags in the cluster. In the food cluster there are only 27 tags assigned to 8 resources. Thus no strong hierarchy is identifiable yet. In the humour cluster 48 tags have been used on 13 resources and the tag "humour" is definitely predominant (example from Fig. 4.13).

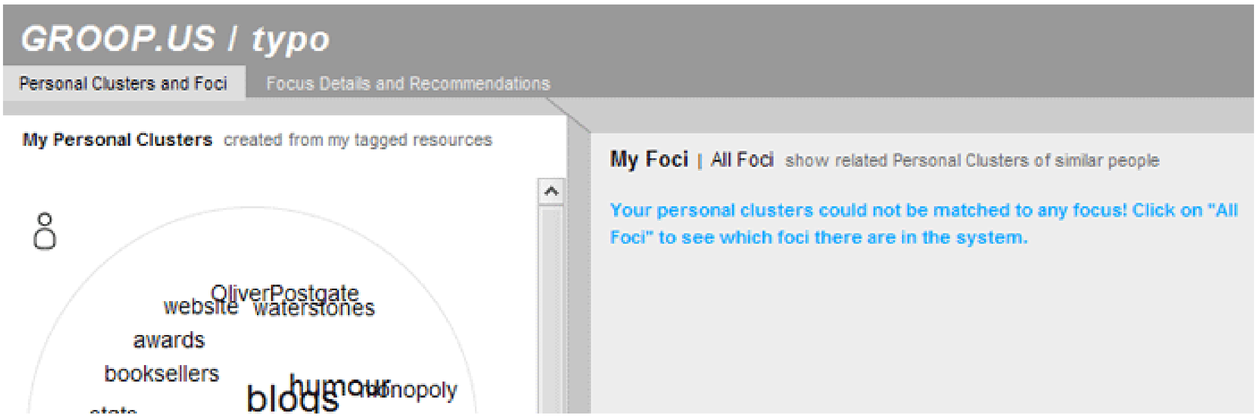

Figure 4.14: Message: no personal cluster could be matched to a focus.

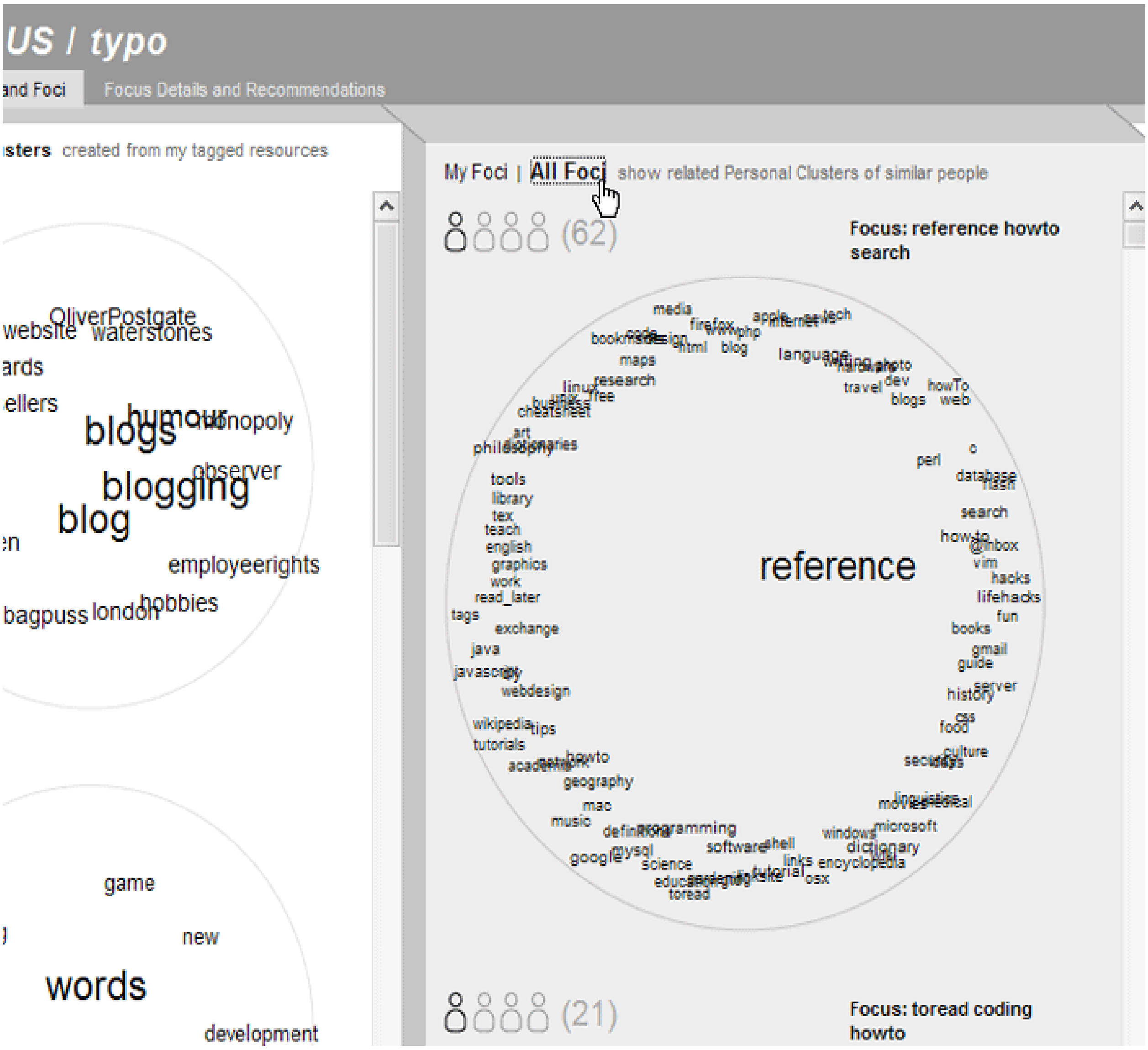

Figure 4.15: Toggle between ”My Foci” and ”All Foci”.

Your foci

Now, have a look at the column in the middle. If the system was able to match any of your personal clusters to foci all your related foci will be shown here. Otherwise you will be informed that this was not possible (see Fig. 4.14, cf. Chapter 3.3.2.1). In this case you can look at all clusters that are in the system to see which tags might bring you closer.

It is always possible to toggle between an overview of foci a relation is at hand (My Foci) or all Foci in the system (All Foci). Click on the Header as illustrated in Figure 4.15.

Selecting Personal Clusters

Now, click on one of your personal clusters. It will get selected which is indicated by a different color of the circle (pink) and a dashed frame around the circle. In the very right column all resources belonging to the cluster are listed with their tags.
If a related focus is available it is scrolled to the top of the middle frame and gets a colored (lilac) outline (see Fig. 4.17). You can see the number of members for each focus in the upper left next to the icon. If you move your mouse over the icon you will see some names of people in the focus (see Fig. 4.16). In the upper right of each focus the three most important tags are listed.
Personal clusters no focus is available for are labelled with a corresponding icon in the upper left of the selected cluster. You can select them and watch resources in it but no focus will be selected.

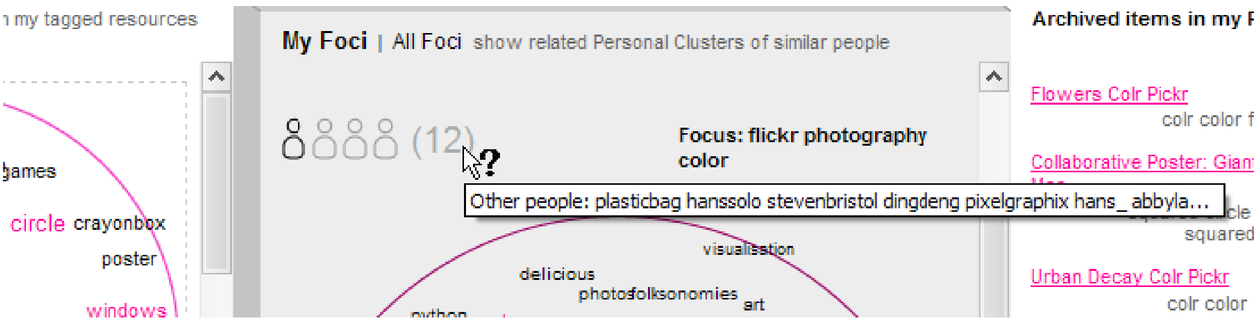

Figure 4.16: Above each focus the number of members is listed and the main tags describing the focus.

Select one of your clusters that is related to a focus. All occurrences of tags from your personal cluster are highlighted in the focus. You can see how many of your tags are popular enough to be shown in the focus tag cloud. Moreover you can compare the importance of your tags to the importance in the focus. Figure 4.17 shows a personal cluster with the main tags ”tagging” and ”folksonomy”. From these tags only ”folksonomy” is predominant in the focus, too. Tagging is still quite important, but recommendations you will get from the 34 other people in the focus are more likely to deal with folksonomy than with tagging.

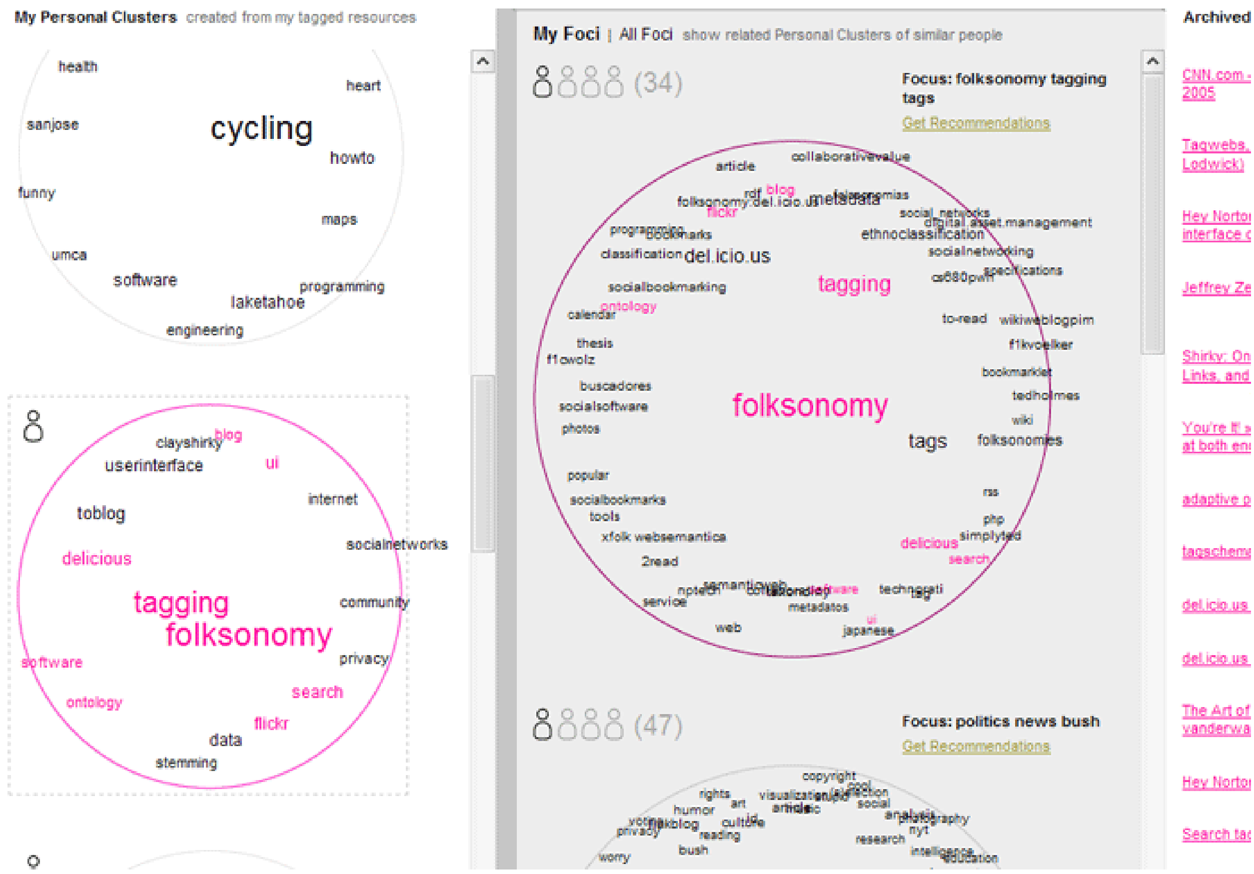

Figure 4.17: A selected personal cluster and its related focus with common tags highlighted.

Click on the link "Get Recommendations" next to a focus (see Fig. 4.18) to change to the detailed focus environment and recommendations.

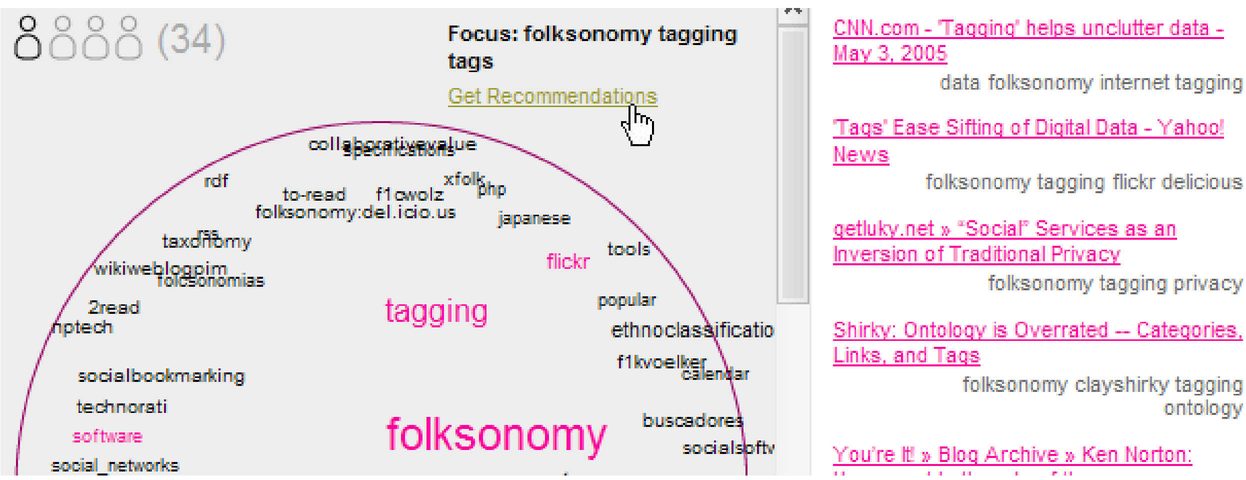

Figure 4.18: Change to details and recommendations for the current focus.

4.2.2.3 Recommendations and Feedback in a Detailed Focus Environment

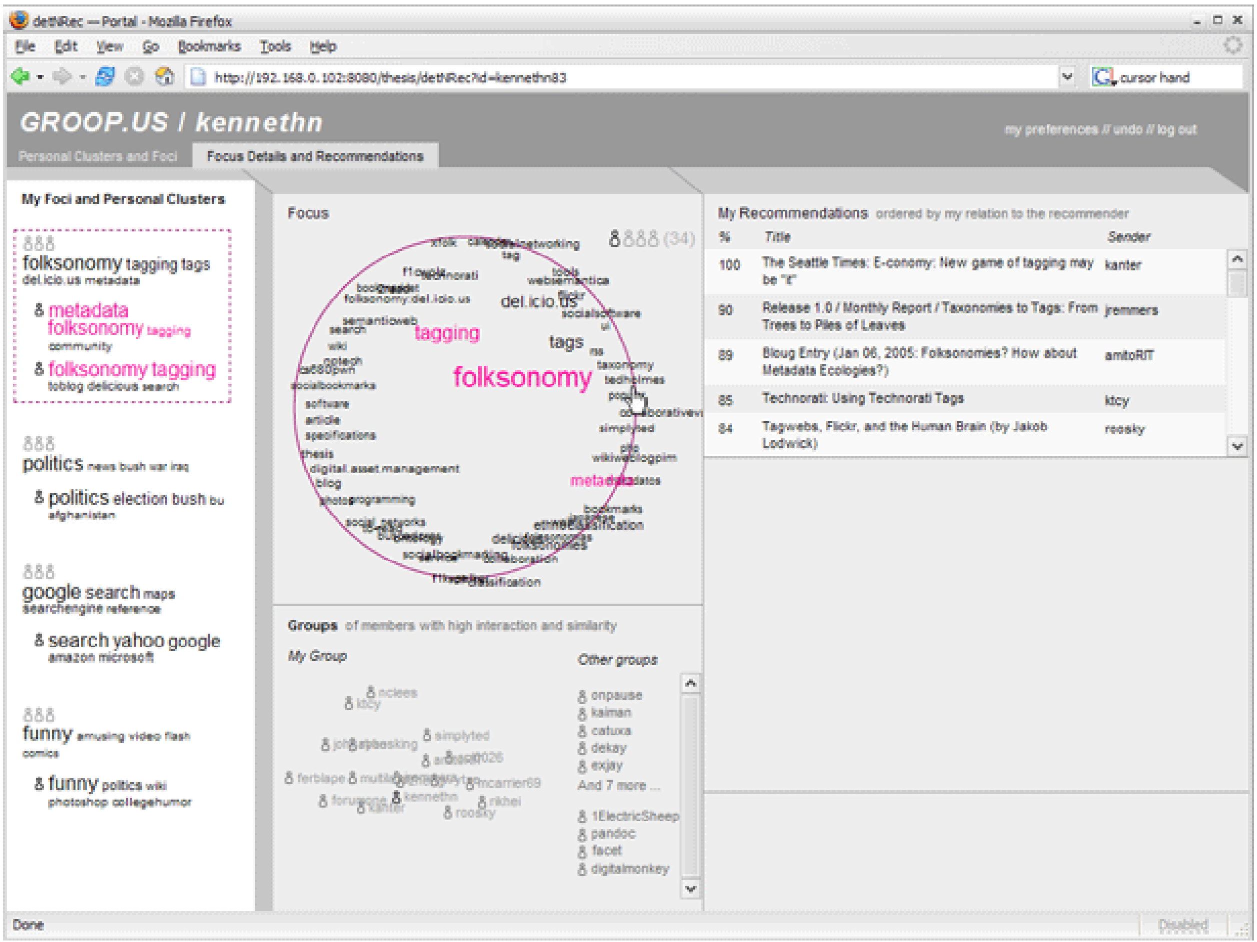

Figure 4.19: The interface "Focus Details and Recommendations".

You have entered the Section "Focus Details and Recommendations" which is indicated by the navigation tab in the header. The interface now looks like illustrated in Figure 4.19.
In the small left column you get an overview of the foci you are in and (indented) your related personal cluster(s). Both foci and clusters are represented by their main tags in size of their relative importance. You can switch between your foci to get recommendations in other areas. A closer look at the selected focus (marked lilac - which is the focus color) of the current example reveals that the sample user has two personal clusters that match the focus you are looking at (see Figure 4.20).

Obviously the two personal clusters are similar enough to be matched to a single focus about folksonomy and tagging. The main tags of the personal clusters that are also in the focus are highlighted pink in the overview and in the tag cloud view in the middle column (see Figure 4.20). In the example this indicates that the cluster that features folksonomy and tagging as most important tags might be more fitting than the other one. Possibly the user might decide to use folksonomy as main tag for all future resources to combine both clusters.

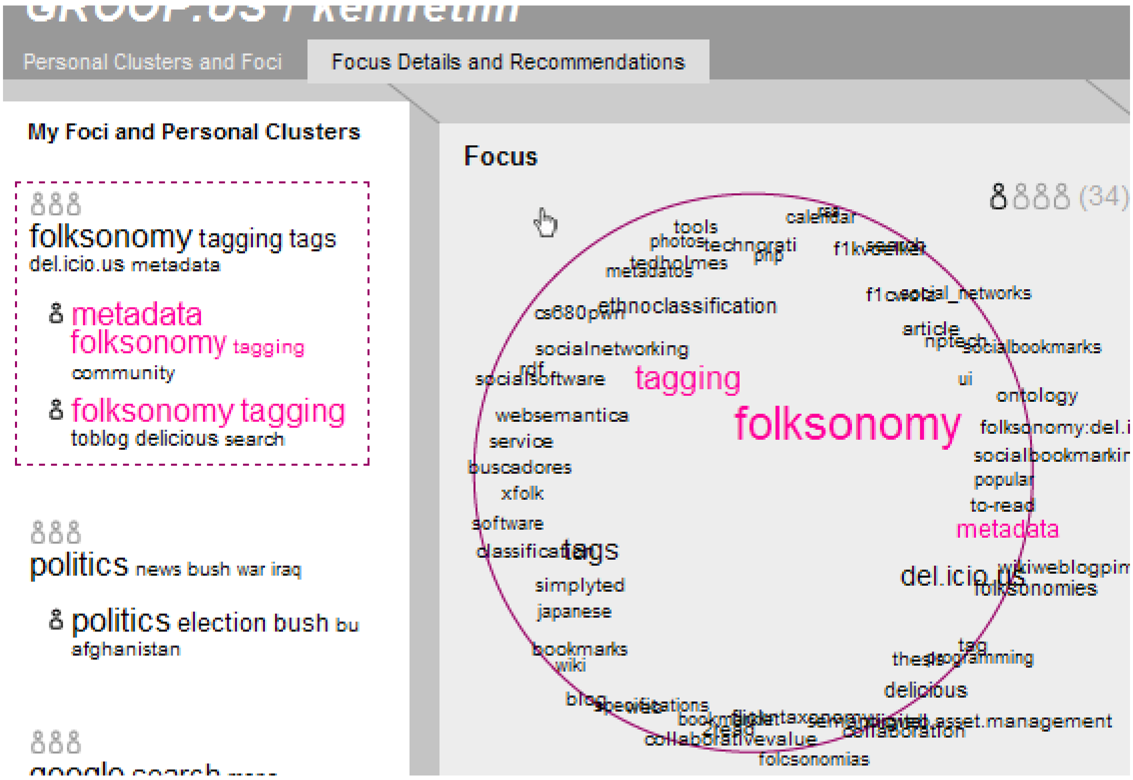

Figure 4.20: Overview of foci and related personal clusters.

Now let us draw your attention to the lower half of the central ”Focus” column (cf. Figure 4.21). It shows all members of the focus. On the left those members of the focus’ social network who are close to you but distant to other members are shown under ”My Group”. The closer a neighbor is to you, the higher is your shared tie strength and your multiplexity value (shared foci). The display might become quite crowded depending on the number of people in the focus and in your group. You can hover the mouse over each name to zoom it and to get information about your tie strength (cf. also Figure 4.21):

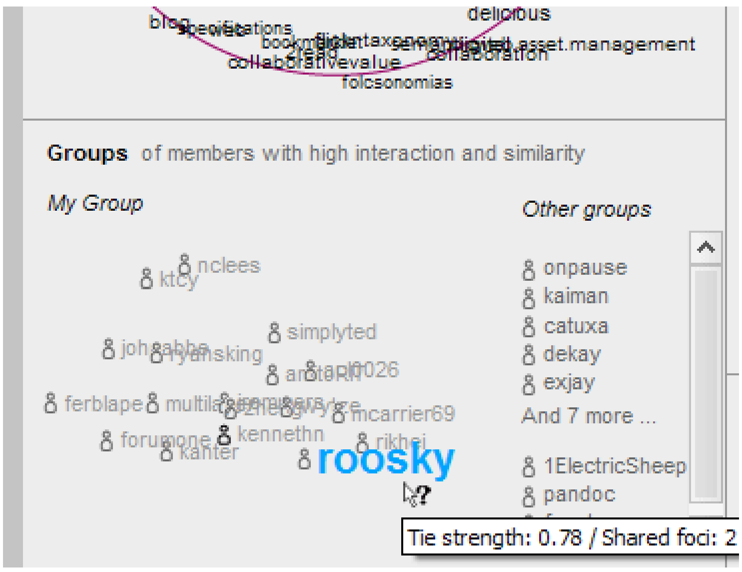

Figure 4.21: The groups of a focus.

Optionally other groups are listed to the right. They group people who are close to each other but not close to you. You get a list of the first five people in each group. Hover over the information about more members of a group to get their names (see Figure 4.22).

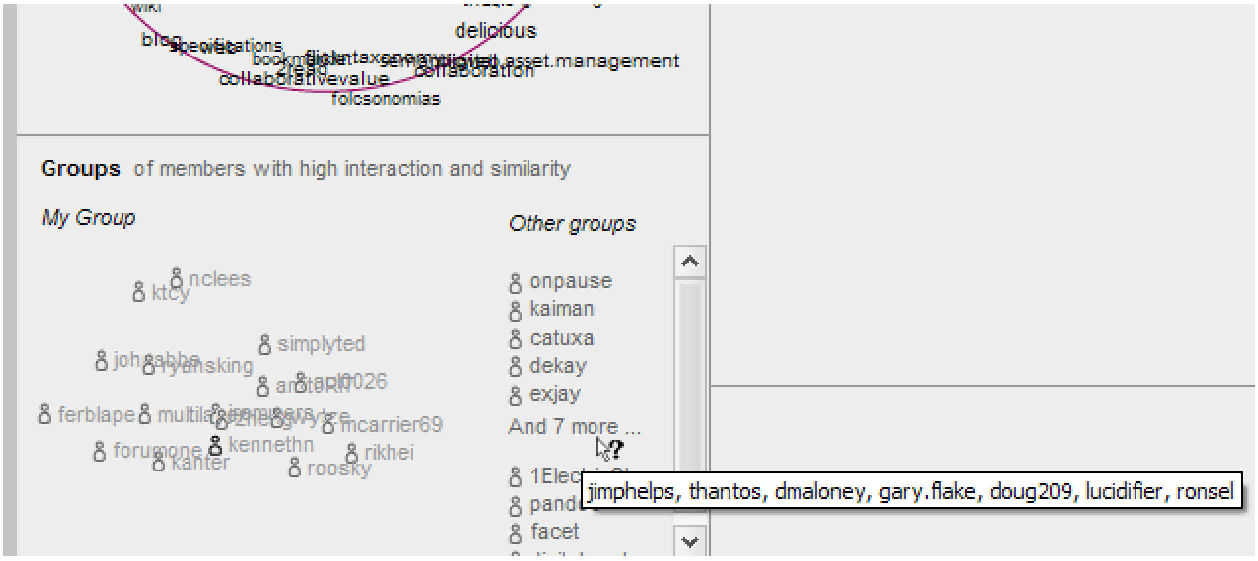

Figure 4.22: Get more information about other groups in a focus.

Explore recommendations

After getting an impression of the focus and its social network have a look at the right column. It is divided in three parts. Topmost you get a list of recommendations. Select the first recommendations by clicking on its title. The focus and recommendations part of the interface will now look like illustrated in Figure 4.23.

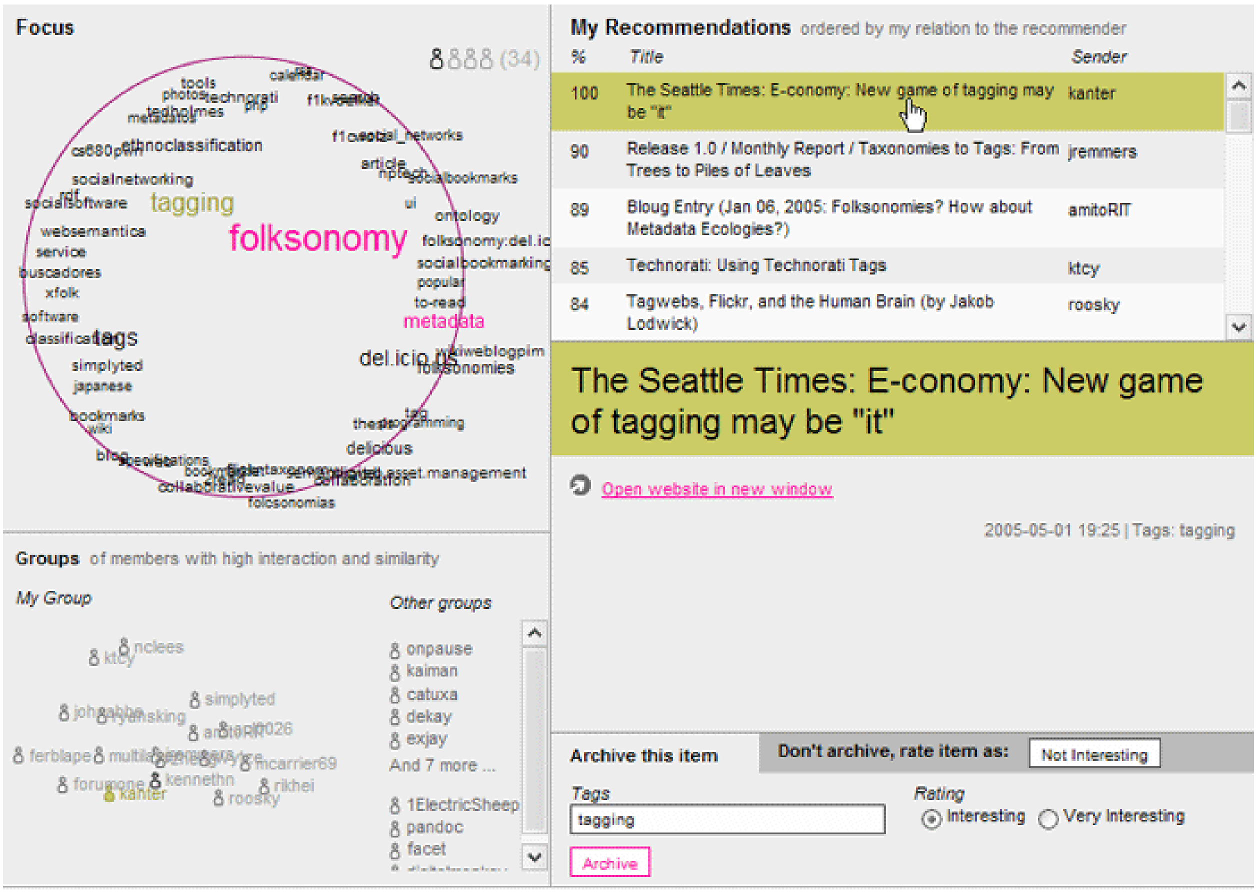

Figure 4.23: A selected recommendation.

The middle part of the right column shows any data that is available for the recommendation (title, description, date, tags). You can open the actual website in a new window by clicking the link "Open website in a new window".
You can see the sender of the current recommendation in the list of recommendations and highlighted in the Groups section. In our example the user "kanter" who sent the recommendation is located in "My Group" very close to the sample user "kennethn" (see highlight in Figure 4.23). This is no surprise because items of group members are favored in the process of prioritization of incoming recommendations. Moreover tie strength is taken into account and similarity to the recipient's archive. Have a look at the feedback area at the bottom of the recommendation column (cf. Figure 4.23). In the input box for tags "tagging" is already inserted for adoption. It is the tag that the recommender chose for the item. And not surprisingly it is one of the main tags of our sample user. In order to emphasize how tags of a recommender match the tags of a focus and those of your personal clusters, they are colored golden/green (the color of recommendations) in the focus tag cloud and in the overview of foci and personal clusters in the left column (see Fig. 4.24):

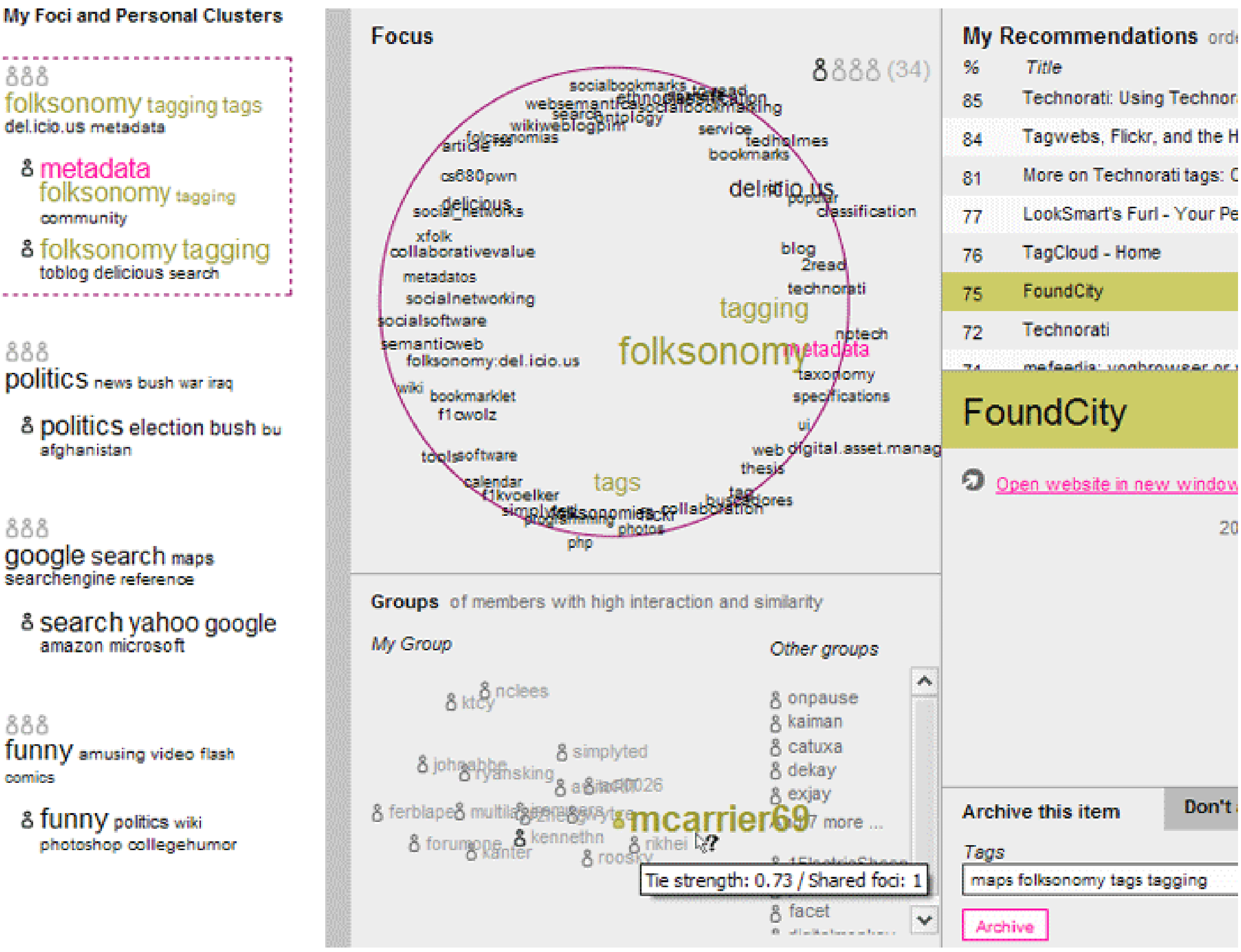

Figure 4.24: All tags that are assigned to an item by a recommender are highlighted.

In the case shown in Figure 4.24 the recommendation "Found City" matches both main tags of the recipient. Yet the item is only rated 75% in relation to the top recommendation because the tie strength to recommender "mcarrier69" is lower than that to "kanter" from the example illustrated in Figure 4.23.

Recommenders are also marked when they are in "Other Groups". If a person is not among the five people of a group that are listed, the first person is exchanged against the current recommender (Figure 4.25).

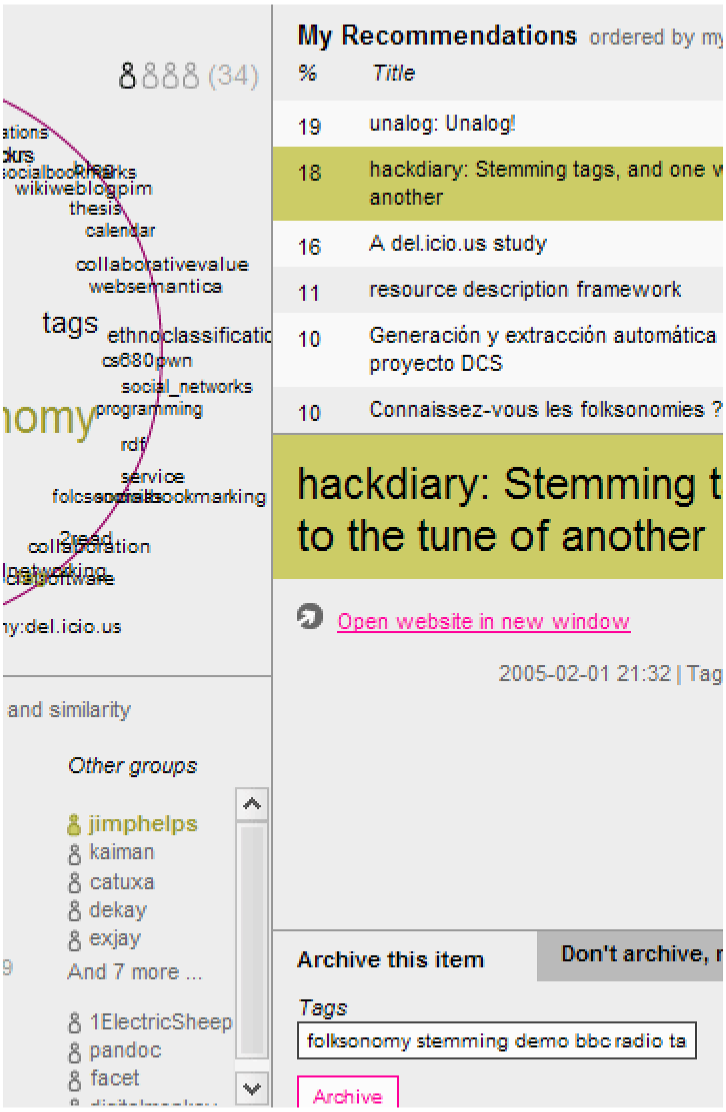

Figure 4.25: The sender of a recommendation is also highlighted when he is in another group.

Feedback

You can give feedback to any article using the feedback form below the content window in the right column (see Figure 4.26). If you do not like the item you can give negative feedback in the area with darker background by rating the recommendation ”not interesting”.
For positive feedback use the form on the lighter grey background entitled ”Archive this item”. The tags of the recommender ”roosky” are already shown in the input field (Figure 4.26):

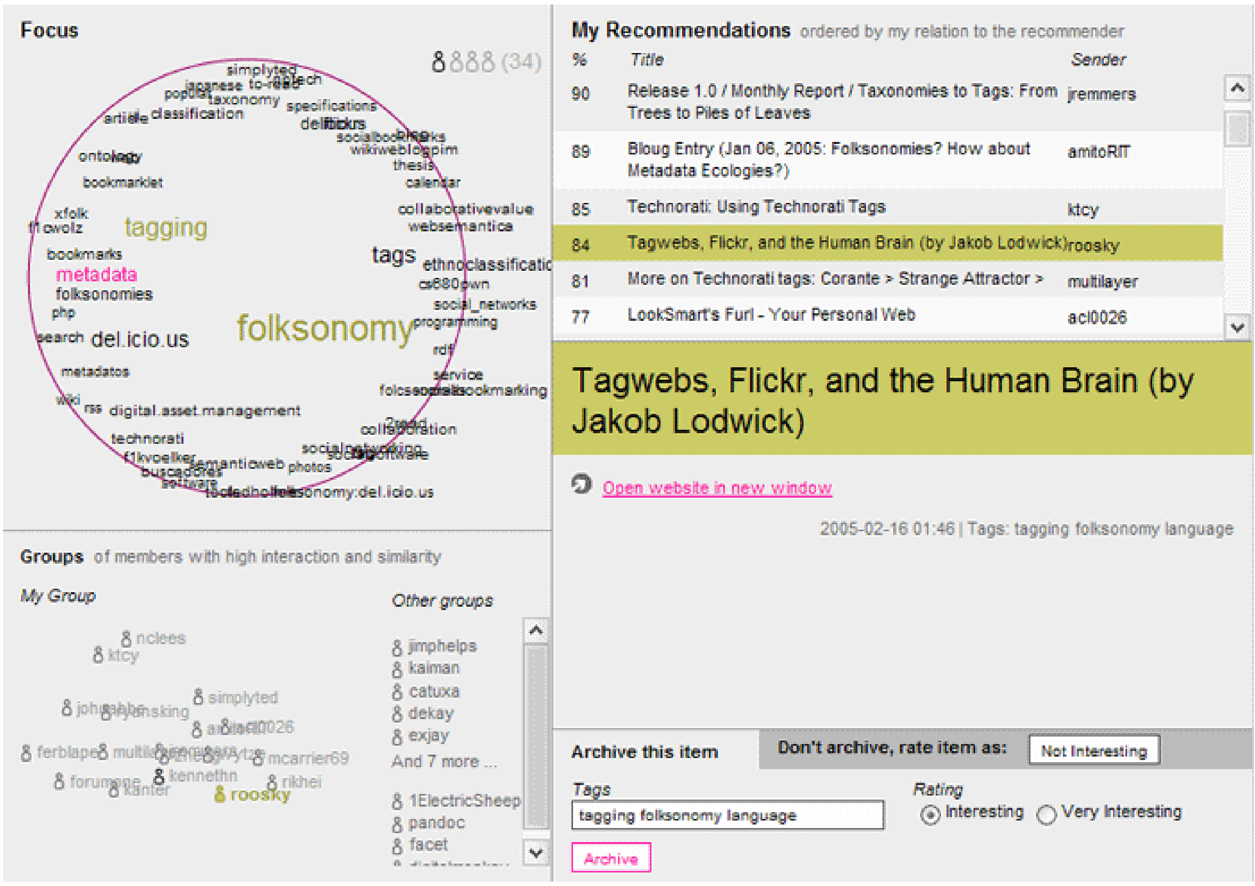

Figure 4.26: A recommendation was opened. Feedback may be given. The input field is initialized with the sender’s tags.

Suppose you are the sample user and you want to adopt all tags except of ”language” which you would not assign to this resource. You erase this tag and add two new: ”flickr” and ”community”. You can additionally rate the recommendation, in our example we rate it ”very interesting” (Figure 4.27):

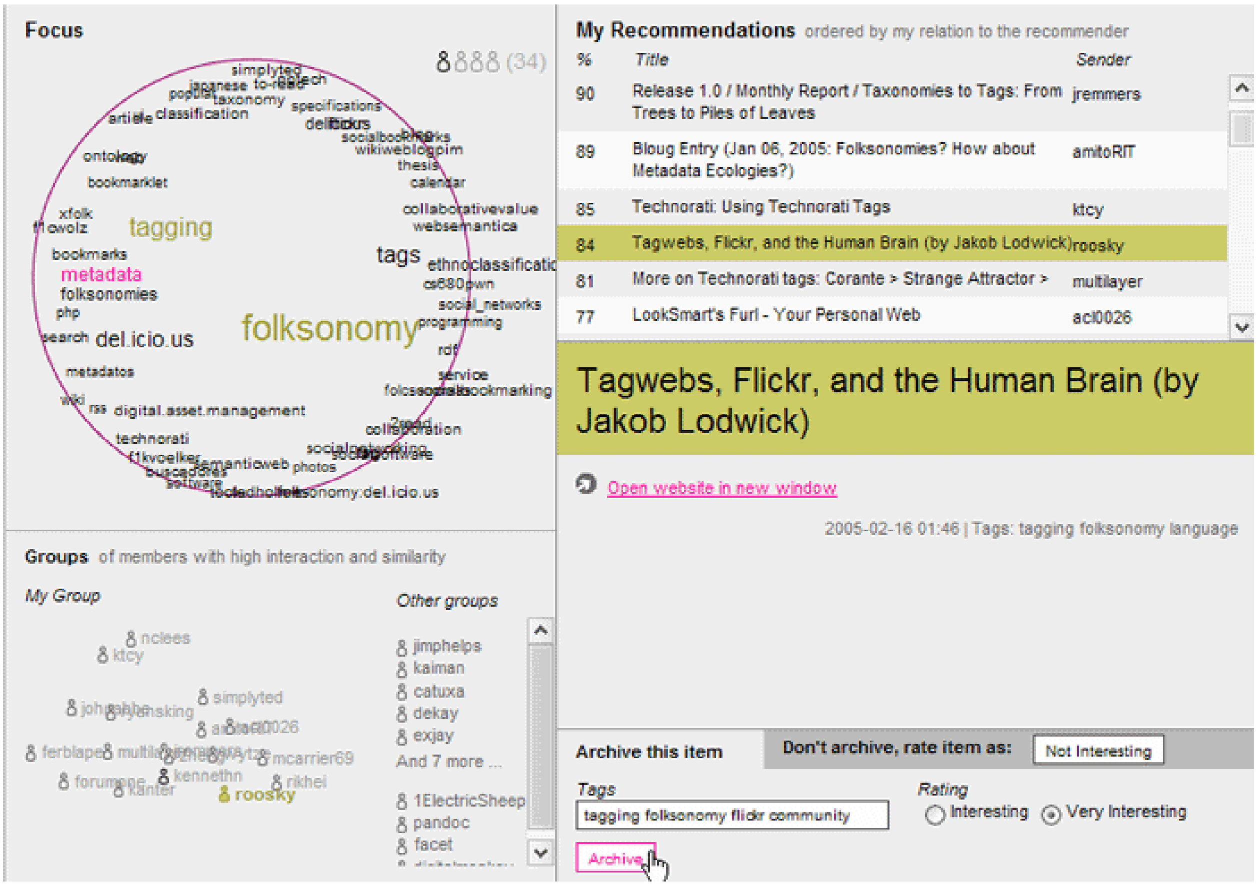

Figure 4.27: Tags are changed as desired and rating "very interesting" was selected.

Before you press the "Archive" Button have a look at the social network to the left. The recommender "roosky" is quite near to you but not the closest. Now press the "Archive" Button and watch the result in Figure 4.28.

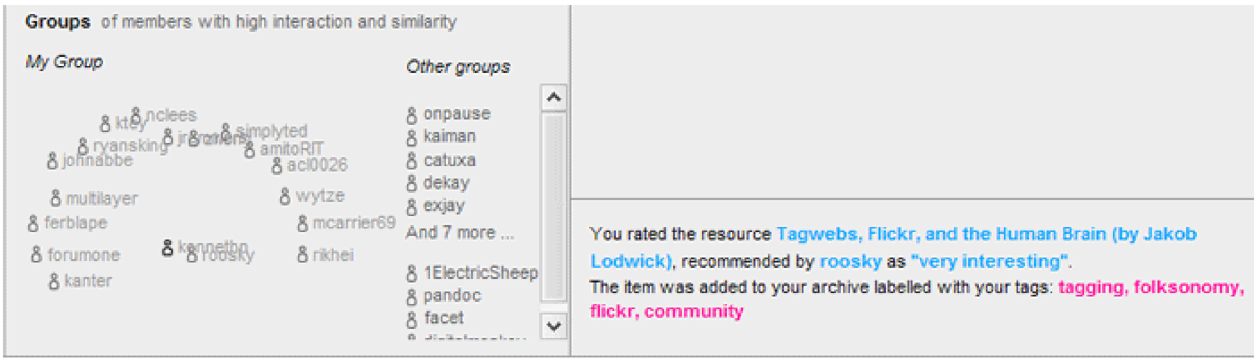

Figure 4.28: Feedback sent. The social network has changed.

The archival of the recommendation to your personal archive is confirmed. The social network in our sample user's group has changed because it shows the relative distance to other group members: the communication that just happened has been positively valued by the system by increasing the tie strength to the recommender "roosky". Now he is the only one who is really close, while the other group members are relatively far. You can check the new tie strength by hovering over the recommender (Fig. 4.29):

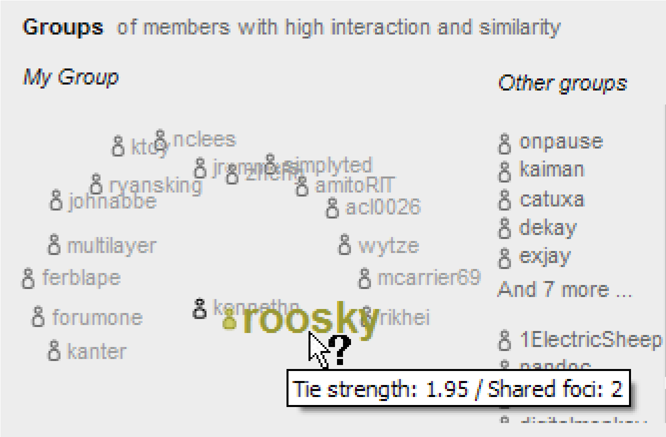

Figure 4.29: Check tie strength in the changed social network.

The new value is 1.95 in contrast to "rikhei" (to the lower right) to whom the bidirectional tie strength is 0.7. Of course this new network situation influences the list of recommendations. The list is immediately updated. After adding new recommendations (done manually by the administrator) a recommendation by "roosky" is on top of the list (see Fig. 4.30). Whenever a recommendation was archived it is removed from the list of recommendations.

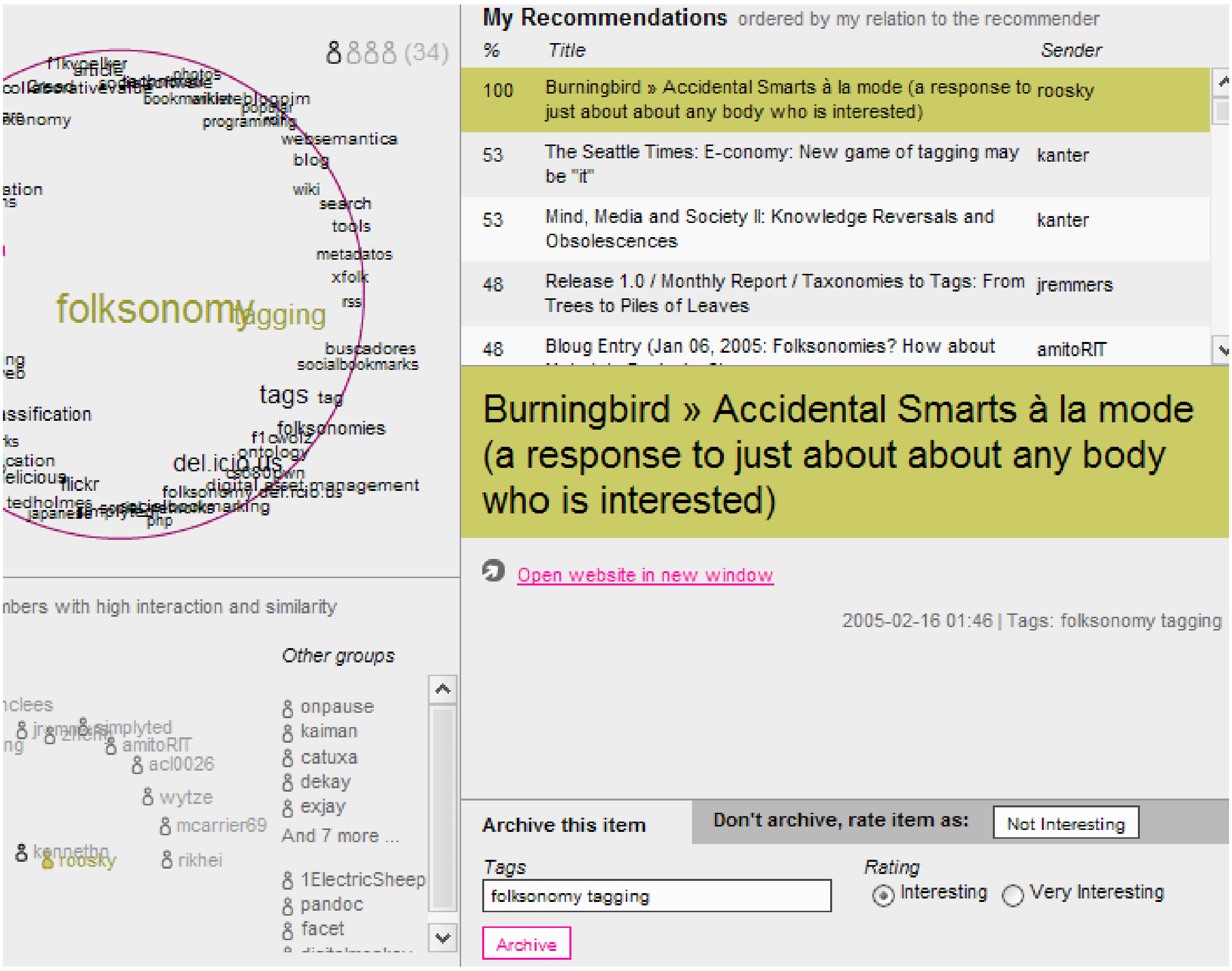

Figure 4.30: The interaction took effect: another item by the recommender is ranked highest.

4.3 Conclusion

Concluding this chapter we will examine our practical work in two steps. First, we will once again go through the four steps of the framework looking at noteworthy observations and details to conclude if the previously stated goals and requirements are met.
Secondly, we conduct a small case study showing how to utilize *GROOP.US* for knowledge exchange based on research information for this thesis.

4.3.1 Evaluation of *GROOP.US*

Before going into detail, an overview of *GROOP.US*' data set shall be given (in order to view the current data set call http://group.us/stats). Currently 1000 people share about 130.000 bookmarks. About 150 people have too few or too unstructured tagged resources to form significant (i.e. dense) individual clusters and are therefore not included in further calculations (cf. 3.3.1.4). Effectively about 850 people form the data basis with over 5.500 individual clusters. These clusters are combined to nearly 100 foci. All following observations are based on this data. The data represents a momentarily snapshot and may change over time.

4.3.1.1 Structuring Individual Resources

The main goal of the framework's first step is to structure a person's individual resources according to similarities between their tag sets to a number of individual clusters. These clusters should be displayed in a well-conceivable way accentuating the most frequently used tags within a cluster.

The clustering based on tag similarity yields good results. On the average a person in the system has 6.5 individual clusters with about 24 resources per cluster. We observed three different kinds of tag clouds: broad clusters (Fig. 31, left) show a number of various main tags repeatedly used in conjunction, explainable by assigning several tags in the form of categories or attributes to a resource. Hierarchical or folder-like clusters show only one dominating main tag (Fig. 31, center). This main tag is consistently used for a large number of resources with many varying minor tags adding further details. The third, often seen type of clusters contains a large number of resources with only one, sometimes two tags (Fig. 31, right). Although only very few tags are shown in such a tag cloud there may be quite a lot of resources within the cluster.
All types of personal clusters mentioned can be found within the system, no type is particularly dominating. Concluding, the similarity calculation foregoing the clustering does not predefine the outcome, instead the type of a cluster depends on frequently used tag combinations. Thus, the visual appearance of tag clouds allows direct conclusions regarding a person's tagging behavior.

Figure 4.31: a) broad personal cluster b) hierarchical cluster c) cluster dominated by resources with one or two tags

A minor number of personal clusters shows only very few tag correlations, mainly caused by incoherent and unsystematic tagging, resulting in tag clouds without comprehensible structure. The density of some of these clusters was too low to identify any tag correlations, consequently they were completely ignored (that is why some of the user profiles contain no data).
The counter-example are tag clouds with very strong interrelations of a few tags. Common tag combinations for example are ”windows” and ”software” or ”webdesign” and ”css”.

So far most of the clusters analyzed center around a specific topic or an area of interest (e.g. art, webdesign & css, max & osx, programming, etc.). That is because the majority of people try to find tags with a thematic reference to the resource described. But some of them mix tags describing the topic of a resources with planned actions regarding a resource (e.g. ”ToRead”, ”ToDo”) or evaluating tags (like ”***” or ”cool”). Our clustering approach will lead in most cases to meaningless results if tags with differing intentions are combined (the left cluster in Figure 32 shows a cluster that includes tags from one to five stars, which is obviously not helpful). Nevertheless it may show positive results, if applied consistently (e.g. the right cluster in Fig. 32 can be seen as a to-do list).

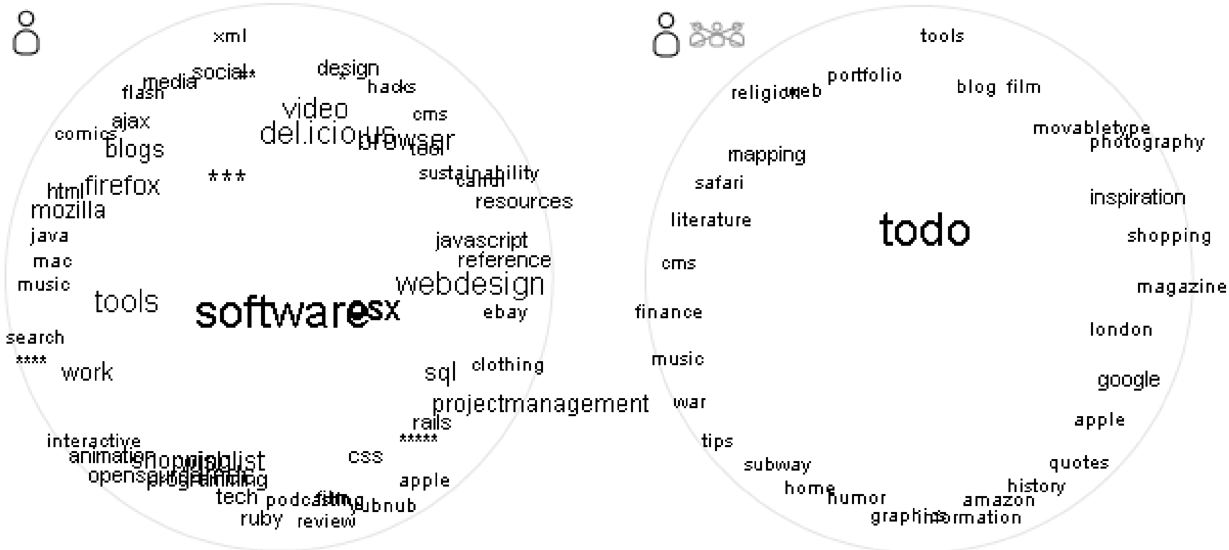

Figure 4.32: a personal cluster with tags * to ***** b) to-do list like personal cluster

The representation of circular tag clouds (cf. 4.1.2) supports the identification of main tags of a personal cluster on the first glance. Only very large personal cluster tend to get complex, especially in the periphery of a cluster (Fig. 31, b). Sometimes tags overlap, reducing readability especially if a large number of similar important and long tags have to be displayed. To reduce these presentational problems the tag cloud display is limited to a specific number of the most important tags (at the moment 100). Of course this may limit a person to grasp the complete structure of her clusters. One may ask if in such cases a cluster consists of too many resources and should basically be limited to a smaller, more traceable size instead of doing this at the level of representation only.

4.3.1.2 Mapping Individual Cluster to Foci

The second step's main goal is to join similar individual cluster to foci. The relation between personal clusters and the corresponding focus should be made obvious, as well as the structure of the focus.

On the average a person is part of about four foci. The average focus has 35 members. The resulting foci are surprisingly well-structured. From 100 created foci only one is discarded caused by low density and missing structure (note that this is the reason why some individual clusters are not related to a focus). About 85% of the foci are strong hierarchically clustered and show only one main tag (e.g. "software", "reference, "blog", "design", "music" etc.). The other foci are only little broader with three to ten main tags (e.g "folksonomy, del.icio.us, tags, tagging, socialsoftware" or "google, searchengine, visualization, maps"). Very broad tag clouds, like seen among individual clusters, are not present. The reason can be seen in the similarity algorithm used for calculating the distance of individual clusters. Only main tags of two individual clusters are considered and - even more important - frequent main tags have higher influence than less frequent. Thus, personal cluster with similar high ranked tags are joint, strengthening even more frequent tags while unimportant tags are ignored. On the other hand including infrequent used tags would have no noticeable influence on cluster distances if their tag rank is considered. In our opinion the dominance of very few main tags within a focus has the advantages of a clear thematic structure and provides a strong basis of a few tags for a shard group vocabulary.

We stated in Chapter 3.3.1.4 that resources should probably be included in different individual clusters if their affiliation is not completely clear. The same is true for merging individual clusters to foci. The following example clarifying the problem was taken from *GROOP.US*. Figure 33 shows and individual cluster (big) and two foci (grey background). The individual cluster contains over 50 resources, about 70% of these resources are tagged with "china" and "internet". The cluster is related to a focus dominated by the tag "internet" although a focus exists that is specifically related to china. But besides the tag "china" the focus also contains the main tag "japan" not found in the personal cluster, leading to a weakening dissimilarity between personal cluster and focus. Nevertheless strong similarity to both foci is given. Ideally the cluster would have been part of both foci, as possible with fuzzy clustering.

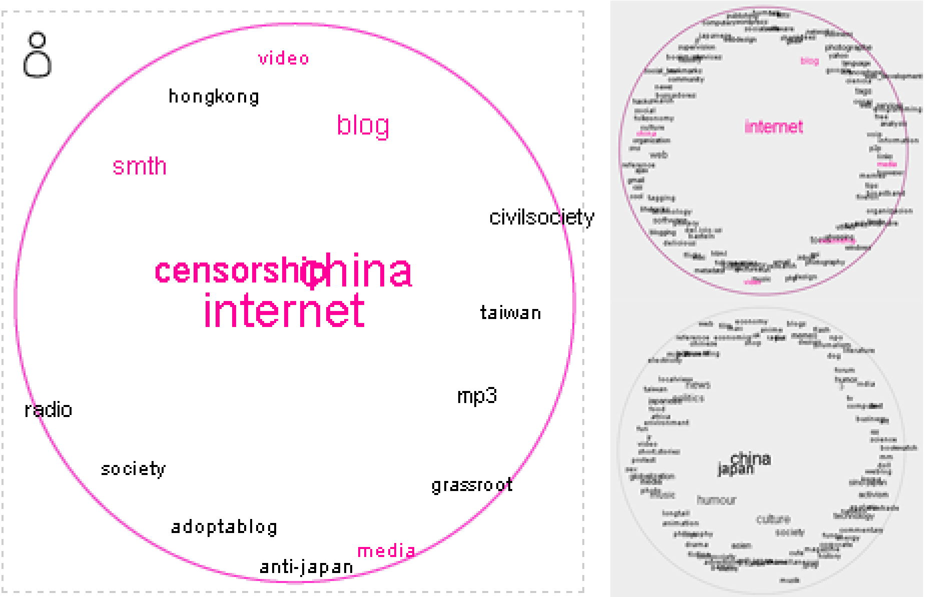

Figure 4.33: relation between individual cluster and its focus. The cluster on the right is related to the left above focus.

The different foci show a big variety of topics, ranging from very broad topics like ”shopping”, ”books”, ”art” or ”programming” to more specific topics like ”php”, ”java”, ”python”, ”ruby” or ”javascript” which could all be seen as subtopics of ”programming” (in fact all of these tags are present in the ”programming” focus).
The combination of the few main tags of a focus gives further clues about the thematic direction the focus has. For example one focus has only one main tag called ”javascript”, while another has the two tags named ”javascript” and ”ajax”. A further examination of minor tags shows that the first focus deals with the programming language javascript in general, while the other is specifically concerned with a special field of javascript programming: asynchronous javascript and XML (short: ajax). This example shows that a specific context is needed to understand the meaning of a tag. ”Ajax” is a relatively new term for client side dynamic manipulation of websites, probably only known by experts of this knowledge domain. And even for those experts the unambiguous meaning of the tag is only obvious with the additional tag ”javascript” (otherwise it could be confused with a football team or dish liquid).
Another example for context dependency of foci is the language used for tagging resources. At the moment the majority of people tag their resources in English. Similarity of individual clusters for building foci solely depends on syntactical similarity of tags. As a result main tags in individual clusters in other languages than English - e.g. German or Japanese - are in most cases completely insignificant for relating personal clusters to foci. For example an individual cluster with the most important tag ”anleitung” is not matched to a the focus ”howto” or ”tutorials” but to the focus ”mac”, although the tag ”mac” is only minor important.
Mechanisms to define translations of tags manually - or even automatic tag translation - would allow language independent foci. But it is questionable if this is a desirable effect. Resources are probably tagged in German because they are written in German and therefore not interesting for non German-speaking people.

We have already mentioned the problem of synonyms and plural forms or different spellings of tags in Chapter 2.1. At the moment tag stemming is not activated for mapping personal clusters to foci based on their tags. Consequently the tags ”tool” and ”tools” are main tags of two different foci instead of being merged to one (another example are the main tags ”blog” and ”blogs”).

Furthermore different spellings of American and British English are not seen as synonymic (e.g. ”humor” and ”humour”) as well as different word combinations, like ”social_software” and ”socialsoftware”. Tag clouds visualize these problems, showing different spellings in a cluster, but without any easy option (despite changing all tags in question for each resource) to change the tags. Manipulation interfaces have been mentioned in Chapter 3.3.1.4 but are not considered in this early version of *GROOP.US*.

The circular tag cloud representation visualizes very well differences between the very few main tags of a focus and the large number of insignificant tags. But compared to individual clusters foci tend to be very large with a huge number of tags. Thus, although foci are drawn bigger than individual clusters they are getting very complex. Alternatively foci could be limited to display main tags only, as shown in the navigation bar of the recommendation interface.
But the main reason for drawing foci with all contained tags is to show the relations between a focus and its individual clusters. Although a tag may be important within an individual cluster its significance in the corresponding focus may be very low. The tag highlighting mechanism reveals these contrasts very well and gives clues which of a person’s individual tags are part of the frequently used group vocabulary.
What is missing is direct feedback how an individual cluster and consequently the corresponding focus changes if new recommendations are archived. Clustering of resources cannot be done in real-time. Therefore influences of one’s tagging behavior and the influences on foci are not immediately visible, but happen delayed after the next clustering. Consequently the action that led to a rearrangement of clusters may not be obvious. In future optimized clustering algorithms may allow nearly real time clustering. Alternatively the transitions between different states of individual clusters and foci could be visualized.

4.3.1.3 Building a Social Network

The third step of the proposed framework calculates social relations between people within a focus based on the degree of interaction and the number of other foci both share. The resulting tie strength is used to identify groups of strongly connected people. The relations a person has to her other group members should be visualized.

There are about 280 groups in the system, on average two groups per focus. The average group has 12 members.

At this time distinct conclusions about the effectivity of the proposed social network cannot be made. To evaluate the accuracy of initial ties communication - or in our case interaction - between related people has to take place. Although we have about 1000 people in the system only two of them - the authors of this work - interact actively. All other relations are proposed on calculated similarities and are not verified by people - given that *GROOP.US* is not public yet. For this reason we will exemplary analyze interaction between both in a case study at the end of this chapter. The relation of all others can be observed and partly simulated, but allows no reliable evaluation.

What can be observed is that the initial tie strength values between members of a group are very similar. The initial tie strength relies on similar tags two persons use and the number of foci they share. People are in the same focus because they use similar main tags and the few differences that are present are split into different groups. More influence has the number of shared foci. If two people share more than one focus it is very likely that both also share groups within these foci.

Overall initial tie strength values are very low and a few positive interactions between two people lead to big changes.

The visualization of groups shows only the distance of the current logged in person to her group members and not the interrelation between other group members. Accordingly, a person has never a complete understanding about the group structure. Furthermore groups may get quite big, making it hard to distinguish the tie strength of particular users. We introduced a mouse over action on a person's symbol allowing it to compare numerical tie strength values, but better solutions could include zooming or scrolling to grasp more details. An enhanced social network was introduced in Chapter 4.1.2 that would allow to get more information about relations between other group members, its realization is high on the agenda for future work.

4.3.1.4 Recommendations and Feedback

In the last step of the framework resources of each person of a focus are recommended to all other members of a focus. For each member incoming recommendations are ranked according to the tie strength towards the sender and in accordance with her individual clusters. Archiving and rating of a recommendation influences the tie strength between receiver and sender. How the current relation to a sender is, which of her tags assigned to the resource already are in my individual clusters and how the relation changes after rating or declining the recommendation should be made visible.

Currently recommendations are frequently generated for all members of a focus. In small foci three resources from each individual cluster of each member of a focus are selected for recommendation, in big foci one per cluster is selected. The newest and not yet recommended resources are chosen. In future a high rating should be favored over date when selecting resources.
The functionality of the ranking algorithm is easy to verify. A high rank mainly depends on a strong tie to the sender. Accordingly people drawn very close in the group visualization are in the upper part of the recommendation list. Especially in the beginning a few positive ratings catapult further recommendations of the same sender to the top ranks of the list and lead to a very close position in the group. If the sender is not in the same group, the probability is high that the next group clustering will merge both in one group.
The second value influencing the rank of a recommendation is the accordance of its tags with the tags in the receiver's archive. The effect can be observed by looking at two recommendations of the same sender at significantly different ranks. Fig. 34 shows several recommendations of the sender "mabahamo". The different ranks originate from the recommendations' tags. Resource one is ranked 79% of the most suitable recommendation, because 66% of its tags could be found in the receiver's archive (xml and xsl vs. itunes). The next resource of this sender is only ranked 74% because only 50% matched the archive (xml vs. php).

%	Title	Sender
80	たのしいXML: XML/XHTML入門ページです	e_luck
79	Playing with playlists: Using XSLT and XPath with the iTunes library	mabahamo
77	Bleb TV Devel list - RE: [TV] ABC1 listings available	leinster
75	OPML to iTunes playlist xsl from 0xdecafbad	leinster
74	Parsear XML con PHP :: Archivos :: 32 de Diciembre	mabahamo
74	XML.com: Hacking iTunes	mabahamo
72	XSL Concepts and Practical Use	benhiller

Figure 4.34: ranked recommendations, showing differing ranks of the same sender.

At the moment some recommendations appear twice. This is caused by duplicates in the imported data sets which are not filter out. Furthermore it is not checked if the same resource is recommended by several people at the same time (and therefore appears twice in the list, cf. 3.3.4.1). In both cases control mechanisms have to be added.

Visualization of the recommendation process happens through highlighting of the sender in the group network and pointing out tags of her recommendation matching the focus or any individual cluster of the receiver. Furthermore the network within a group and the order of the ranked recommendations is immediately updated if positive or negative feedback influences the tie strength value between sender and receiver.
The question is whether this is enough to make people understand that recommendations are exchanged between others and that the application - or the focus respectively - as a place is shaped by these interactions (cf. Chapter 2.4.2). This is important because observation of other is an important part of judging relevance of an item and in order to develop a Transactive Memory System (cf. Chapter 2.2). We think that a more detailed network view as described above is important here (cf. Chapter 4.1.2). Such an enhanced network could contain even dynamic elements that give an impression about negative or positive feedback between others.

Finally the question has to be asked if the recommendations are helpful for the receiver. Again, without making *GROOP.US* accessible to a larger audience and performing adequate questionnaires we can only answer from our personal perspectives, as shown in the following section.

4.3.2 Case Study

To get a first impression of the effectiveness of *GROOP.US* we will consider a short case study as an example for utilizing *GROOP.US* for knowledge management. In preparation for our master thesis we have collected a variety of online resources as the basis for our theoretical elaborations. We bookmarked these resource using the social web service del.icio.us, each under his own account (see http://del.icio.us/kaiman and http://del.icio.us/bbobb). We agreed previously to tag all bookmarks specifically dealing with topics concerning our thesis with the tag "thesis", followed by further tags describing the topic in more detail. From 1000 data sets imported for testing *GROOP.US* we have chosen our own, because only for our own data the context and tagging behavior is known precisely.
Before looking at the initial clustering results the expected outcome shall be predicted. 50% of all bookmarks of the user "kaiman" are tagged with the tag "thesis" compared to 30% of the user "bbobb". Consequently both tags should play an important role in individual clusters of both user profiles. The tag "thesis" should be a main tag in at least a few personal clusters. We do not expect

that all resources tagged with "thesis" form only one personal cluster because diversity between additionally added tags is very big (ranging from "rdf" to "socialnetwork" and "blog"). But we expect one or more foci that emerge around the tag "thesis" as a result of our pre-decided tagging approach.
After clustering the user "kaiman" has nine personal cluster. Seven of them contain the tag "thesis", in four clusters the tag is among the three most important tags, which are:

- cluster 1: rss, thesis
- cluster 2: rdf, thesis
- cluster 3: thesis, tags, del.icio.us
- cluster 4: folksononmy, thesis

Cluster 1 is matched to the focus with the main tag "rss", cluster 2 to "rdf", cluster 3 to "del.icio.us" and cluster 4 to the focus with them main tags "tagging" and "folksonomy".

The user "bbobb" has seven personal clusters, four of them containing the tag "thesis". In three of them "thesis" is one of the three main tags:

- cluster 1: thesis, foaf, networking
- cluster 2: tagging, del.icio.us, thesis
- cluster 3: socialsoftware, thesis, blog

Cluster 1 is not related to any focus. Cluster 2 to the focus containing the tags "tagging" and "folksonomy" we already mentioned above. Cluster 3 is part of the focus "blog".

As predicted the tag thesis plays an important role in many personal clusters. But in contrast to the predicted results only one shared focus is formed. Having a closer look at the individual clusters this is not surprisingly.
First of all the individual clusters give hints about the different areas of research both authors were concentrating on. Although the tag thesis is dominant, there is only few overlapping between the additional tags of the different clusters. Consequently both authors share only one focus showing the topic both authors worked on together, all other research was not done conjointly but split-up between both.
Overall the tag "thesis" did not have any influence on the foci building process. Looking at the different foci the personal clusters are related to shows why: the tag "thesis" is one of the unimportant tags, probably only used by the authors. Accordingly its overall influence is low, only the focus both authors share shows a slightly - but still negligible - increase of its importance. Both authors share eight tags, especially two main tags.

The shared focus consists of 28 people, divided into three groups. The strongest tied people form a group. Both authors are in one group with seven members overall. Although the tie strength between all group members is very similar (ranging from 0.56 to 0.69) both others are the most strongly tied people.
Recommendations of both are on the first place of each of their recommendation lists (Fig. 35). The reason is not only the high tie strength but also a high similarity of the tags the recommendation is tagged with in the receiver's archive (100% and 75% respectively). In both lists recommendations of the person "lucidifier" are on ranks two to four. To both "lucidifier" is strongly related (although

not as strong as both authors are tied). Additionally the correspondence of all recommendations to the receivers' archives is exactly 100%, even though up to five tags per recommendation are used. That is why further recommendations of "kaiman" and "bbobb" are at the bottom of the list - here correspondence is only between 50% and 66%.

Concluding the tag "thesis" had very little influence on the process of mapping individual cluster to foci. But one can assume that within the shared focus the tag "thesis" was an important factor controlling the group clustering. Calculating initial tie strength values depends mainly on similarities of personal clusters - which in this case were dominated by the tag "thesis" - and only secondly on the number of foci people share. Thus, although "lucidifier" shares two foci with both authors tie strength between the both is still higher. Interestingly on the average "lucidifier's" recommendations show higher conformance with the focus and also with the personal clusters of both authors within the focus. But in the initial tie strength calculation all tags of two people in a focus are compared - even those which are not frequent in a focus (like "thesis") - and not only the few tags assigned to a recommendations. The similarity of both authors is therefore related to accordances outside of the focus - confirming at least partly our predictions.

%	Title	Sender
100	Metadata, Mark II	kaiman
97	i d e a n t: Tag Literacy	lucidifier
97	P.S.:	lucidifier
97	Folksonomy - Wikipedia, the free encyclopedia	lucidifier
92	de.lirio.us :: : entries	mcarrier69
88	Metadata? Thesauri? Taxonomies? Topic Maps!	kaiman
86	qootas.org : sekimura's blog: folksonomy 民主的分類の面白さ	facet

%	Title	Sender
100	Burningbird - Cheap Eats at the Semantic Web Cafe?	bbobb
96	i d e a n t: Tag Literacy	lucidifier
96	P.S.:	lucidifier
96	Folksonomy - Wikipedia, the free encyclopedia	lucidifier
94	de.lirio.us :: : entries	mcarrier69
92	qootas.org : sekimura's blog: folksonomy 民主的分類の面白さ	facet
87	CC Mixter - The remix family tree	bbobb

Figure 4.35: recommendations for the focus "folksonomy" for "bbobb" (left) and "kaiman" (right)

The recommendations themselves are highly related to the topic in question - folksonomies and tagging, some of them actually quoted in this thesis, e.g. [Powers2005] recommended by "bbobb", [Merholz2004] recommended by "kennethn", [Doctorow2001] recommended by "kennethn" and [Mathes2004] recommended by "rjjjsp". The person named "kennethn" is mentioned twice although he or she is not in the same group as the receiver. Positive ratings as respond to these high quality recommendations would bring both together in one group after the next re-clustering.
It has to be pointed out that one recommendation in the top ten was in Japanese, hence - also probably related to the topic in question - not useful for the receivers. This example shows that a share group context did not emerge - which is not surprisingly taking the fact that interaction did not take place so far. Rating unsuitable recommendations negatively will lower the importance of further recommendations of the sender and - if ratings are continuously negative - lead to a separation of sender and receiver into different groups.

This short and by no means complete case study shows the potential of *GROOP.US*. Within this chapter we concentrated on the details of our solution. Finally, the next chapter should again take a look at the whole picture going back to our initially stated goals.

5. Conclusion

5.1 Recapitulating our work

Concluding, we want to look again at our initial goals for this thesis, summarized in its title: ”Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation”. Basically, three main questions can be identified:

1. Can similar individual contexts be successfully identified and combined on the basis of tagging?
2. Can joint individual contexts - as starting point for creating a social network - lead to shared group contexts?
3. Can shared contexts be utilized for deriving recommendations to realize social navigation?

5.1.1 Summary

To answer these questions we proceeded in three parts. First, we looked at the **theoretical foundations**. We started with analyzing the potential of tagging by introducing collaborative metadata creation in the form of folksonomies. The term ”context” was defined, concentrating on the transition from individual to shared contexts by the examples of sign processing, transactive memories and the judgement of relevance of information. Then, basics of social networks were explained, introducing the concept of foci and groups. Finally, principles of social navigation were presented, concentrating on concepts of recommender systems and the importance of social awareness for plausibility of recommendations.
In the second part, we introduced a conceptual **framework for social navigation** in four major steps, proposing a solution for the questions stated above. Identification and joining of individual contexts was dealt with in the steps *Structuring of Individual Resources* and *Mapping Individuals' Interests and Resources to Foci*. *Proposing a Social Network* - including the identification of groups - was described in the framework’s third step to answer the second question. Utilizing the emerging shared group context for Social Navigation was described in the framework’s fourth step: *Recommendations and Feedback*.
Finally, the third part describes the **practical implementation** of *GROOP.US*, a prototype realizing the main aspects of our theoretical framework.

5.1.2 Conclusions

At the end of our work, can we give positive answers to the question raised above?

Regarding our framework, question 1 has to be split into two parts: **are the central topics of an individual context - we refer to as individual cluster - identified correctly? And can similar individual clusters be identified and combined to a focus?**
After evaluating the prototype both questions - at least regarding the authors’ experience (cf. 4.3.2) - can be answered positively.

The dominant tags in individual contexts and foci show strong correspondence, although unusual individual tags are insignificant - simply because no shared consensus exists for these tags.

An interesting question is in which way relations and structure immanent in tagged resources are shaped by the context they were created in. Currently, *GROOP.US'* data set is solely taken from one source - the del.icio.us bookmark manager (cf. 1.2). People using this service often bookmark their resources specifically having in mind the shared value that may arise. Accordingly, tags may be chosen to meet a specific target group (for example people following the discussion about "folksonomies"). This behavior could be an explanation why our approach works well for data from del.icio.us. But what about tagged resources from other sources, probably less target-oriented, for example from authors of weblogs categorizing their own articles? It is not sure if a combination of different data sources (with different intentions and target groups in mind) yields to reasonable results.
A related question is how *GROOP.US'* way of showing relations between individual contexts and foci shapes tagging behavior. Instead of tagging resources for oneself only or for an unspecific target audience (e.g. all users of del.icio.us) tagging happens directly in the context of a focus, showing how a person's individual context is situated in the focus and how her personal tagging influences relations to others. We predict that a known group context in a focus will lead at least to a few group specific interpretations of tags, but this hypothesis has to be proven in a daily usage of *GROOP.US*.
Although visualization of foci and their group contexts is a central aspect of *GROOP.US* we have to face the criticism that, at the moment, individual contexts constituting a focus are not visible to all its members (in fact, it is partly visible when recommendations are shown with their tags). Thus, it is not known why other people are part of a focus which interferes with a main idea of social awareness.

Again, the second question has to be answered differentiated by evaluating **if the proposed social networks and accordingly groups within foci are reasonable. And secondly, does a shared understanding within such groups emerge?**

Starting with the first question, initial ties within the proposed social network of a focus represent similarities in the tags people used. We have shown that similar tagging behaviors may hint for similar individual contexts and a shared understanding about a topic (cf. 2.1.2). Furthermore, the initial tie strength is higher the more varying topics - referred to as individual cluster - two people share. The few results observed (cf. 4.3.2) support the impression of an operative and reasonable network, but one has to keep in mind that no proposed tie relies on any direct interaction the system knows about. But that does not mean that there has not been any interaction outside of *GROOP.US*. Integrating information about relations in existing social networks could improve the network building process. Sources for such relations could be other social web services or weblogs providing machine readable data in formats like FOAF [Brickley2005] and XFN [Celik2005].
A different option would be to allow people to define friendship information to other people manually, as originally proposed in our framework (cf. 3.3.3.1) but not realized in the prototype. This would allow closer relations between people in *GROOP.US* without having to interact frequently. Nevertheless, the creation of foci and therefore the decision which people are joint in a social network depends on tagging behavior only, without considering tie strength values. Accordingly, high friendship values would increase the probability of two people in a focus two join a group, but only if they are already in a shared focus. This constraint is intended, prohibiting that two friends share several foci based on friendship values only with one of them actually not being interested in the topic in question or without using the vocabulary (the tags) common in that

focus.
Of special interest are groups in social networks. A cluster of strongly tied people within a focus forms a group. Interaction between group members is particularly fostered. Looking at the initial proposed groups strong similarity is given, but the difference to members of other groups is not significant. As a result, interaction between members of different groups quickly leads to a rearrangement of groups. The idea is to base group membership strongly on interaction to validate imprecise initial values (the example given in Section 4.3.2 shows that this done for good reason). On the other hand groups should not be the subject of frequent changes after a solid number of related and frequently interacting people is found. Otherwise a shared understanding cannot emerge.

We base our interpretation of a shared context on the concept of folksonomies (cf. 2.1.2) and transactive memories (cf. 2.2.2). We assume that people in a shared context have similar understandings about the used tags to indicate the same information. Only through frequent interaction - in our case tagging and rating of recommendations - and observation, perceiving the usage of tags of others, a shared context can be maintained.
GROOP.US supports perception of shared contexts in various ways: by highlighting tags of recommendations in foci and individual clusters and by emphasizing the relation to people someone is interacting with. It is the question if this is sufficient to create an enduring shared context. Taking the example of transactive memories the precondition for evolving a shared understanding is direct communication. We interpreted rating and tagging of recommendations as a form of communication between sender and receiver. Archiving a recommendation in *GROOP.US* influences the relation (positively or negatively) to the sender, resulting in an adjusted tie strength value. Accordingly, the group display for representing tie strength values is updated, but without indicating the sender of the recommendation why the relation changed. Even with displaying feedback one would not talk about direct communication.
A more far-reaching approach would be the introduction of means of direct communication, for example through allowing personal recommendations (and not automated recommendations only, like shown in *GROOP.US*). This would allow direct social navigation besides the present form of indirect social navigation.
A deep shared understanding within a group is shown by high symmetric tie strength values. High values are reached, if a strong correlation between the tags sender and receiver used for tagging a resource is given. The counter-example is if the receiver completely ignores the sender’s tags or rates the recommendation as ”not interesting” without tagging it.
It may happen that initially two people share a group because they use the same tags and have a similar understanding of them, but overall both have a differing standard of knowledge. Examples are beginners and experts or adolescents and adults. As a result, both would probably rate the other’s recommendations as ”not interesting” and consequently both would not share the same group in the long run. As interaction was rated negatively - although interests are similar, however unbalanced - the system would probably stop mutual recommendations. The worst-case scenario would be a number of isolated groups with similar members, all sharing the same knowledge without any kind of innovation.

To allow innovation flow across borders of closed social circles we introduced the concept of bridges (cf. 3.3.3.2). Although not realized in the prototype, bridges could connect unrelated groups or foci with similar topics. Concepts for proposing bridges on the basis of the overall social network are shown in Chapter 3.3.4.1. Bridges are also an option to relate groups with similar interests, but a differing vocabulary hindering the creation of a shared context. By forwarding tagged recommendations between such groups the creation of a shared vocabulary can be mediated probably resulting in the emergence of groups or even foci.

A final conclusion regarding the effectiveness of the proposed social network and whether the emergence of a shared group context can be ensured cannot be drawn. To receive more than a loose impression, *GROOP.US* has to be used by many in parallel. Only a large and active user base can verify the questions stated on a broad level.

Question three asked, **if the emerging shared contexts can be successfully utilized for deriving recommendations for social navigation.**

First of all, deriving high quality recommendations is the overall goal of our framework. The recommendation engine relies on all previously mentioned steps. Inaccuracies in foregoing steps will sum up and prevent reasonable results. The case study (cf. 4.3.2) showed suitable recommendations, but this conclusion cannot be generalized without further verification on a broader level.
To ensure trust in the recommendations the reason why a recommendation is given has to be transparent. That is, what is my relation to the sender and how does his choice of tags to label a recommendation match my own perception? Tag highlighting shows, if correspondence of tags is given at all and if matching tags are of major importance in the receiver's group and individual context.
Even more important is to indicate the sender's position within the given group context. Not only the relation between sender and receiver - as shown in *GROOP.US* - has to be visualized, but also the relation of other group members to the sender. As pointed out in Chapter 2.2.3, the overall relevance of information has to be evaluated not only in the receiver's individual context, but also by taking into account the opinions of other relevant parties - in our example the members of a shared group context. At the moment, *GROOP.US* lacks possibilities to point out all interrelations within a group (a proposal was given in Chapter 4.1.2). Previous ratings of other group members are not displayed for a recommendation.

Right now, 1000 people are registered in *GROOP.US*. The number of people in foci and groups shows a perceivable size. Use case scenarios for such a small and closed number of people could be information management in universities or companies.
But what about opening *GROOP.US* to "the whole world", facing 100.000 or even millions of people using the system? Apart from performance problems, is a shared context still possible with such a big number of people? Are differences in tagging behaviors so subtle to find significant variations to form foci and groups with a manageable number of people? Or would one big macrocosm be separated into many small microcosms with only few bridges between them? What about problems many big communities face like spamming (cf. 2.1.2) or free-rider problems (cf. 2.4.4.4)?

The number of open questions and problems shows, that research surely is not finished. We plan to open *GROOP.US* to the public in the next few months to get more reliable information, especially about network formation and shared contexts.
In the end, we want to position our concept for social navigation in an overall vision: the Social Web.

5.2 Outlook: From Information Overflow to the Social Web

In this final chapter we are going to relate our hypothesis and the proposed solution to the ongoing discussion of the future of information processing addressing information overflow on the internet.

Like many other research projects in knowledge management and related fields, we took the possibility for everybody to publish - leading to the inevitable unstructuredness and consequently an overflow of information - as a starting point for our thesis.

Our hypothesis and answer to the question, "how do I find relevant information effectively?", was to identify a person's social network and therein groups who share a context regarding a particular topic. Members of these groups are integrated into social navigation processes (like chat, collaborative filtering) which finally lead to the information desired. The semantics of topics and related resources can be predicted because varied experiences with one's group members likely lead to similar interpretations.

The Semantic Web Versus a "Web 2.0"

Other efforts have been undertaken to apply semantics to the information on the internet, the Semantic Web being the most important initiative since 1999. It basically proposes standardized machine-readable formats for metadata and ontologies complemented by programs (agents) that contain logic in order to make computers able to infer semantics for humans (cf. [Passin2005], [Berners-Lee2001]).

There are two important issues regarding the Semantic Web that let many researchers and developers doubt its success ([Norvig2005], [Shirky2003a]). First, the techniques proposed are very complex addressing information architecture professionals. Secondly, even though its techniques are ready to process any semantics, the Semantic Web does not propose means to gather them from a people's interaction online. Examples and scenarios are given but they only illustrate that the techniques work *if* semantics are already implied or are related to a topic where ambiguity is not possible, such as science.
Concluding, the lack of wide adoption of a concept that has been around for six years now shows that something is wrong. We assume that it is the lack of concepts to integrate the interpretation of data contributed by people on a broad basis. The existing techniques plus adequate concepts would possibly lead to wide-spread semantic data creation, resulting in a real *semantic* Web.

As we have argued in the introductory part of our thesis there is already lots of information about people and their relation to things and other people on the web today in communities, online-shops, weblogs and social web services. Being inspired by the rising success of weblogs and social web services developers and researchers in this field start to promote their crucial features: structured means of publishing and simple exchange formats as a better starting point towards a semantic web. Instead of storing information in demanding external data formats such as RDF, semantics shall be encoded in simple techniques that are successfully applied already. Examples are Microformats, that is a set of extensions to (X)HTML that allows for structured encoding of social

network information (XFN, hCard) or tags (relTags) [Celik2004], simple exchange formats such as RSS and ATOM (cf. [Brickley2000], [Nottingham2005]) or REST which allows to call data in various formats (XML) from documented APIs of web services (cf. [Prescod2002]).
A web of services that provide their data in such a structured manner can be regarded as a platform where everybody can "remix" data according to her needs. Be it as a non-technical user creating a weblog with streams of bookmarks, pictures and music tracks from various web services as an "API to herself" or be it as a developer integrating information about relation of people and things or other people in order to derive semantic meaning by an algorithm to create a new service like a social navigation system. Some call this new understanding of the World Wide Web complemented with the techniques described above the "Web 2.0" (cf. [MacManus2005], [Ruby2005]).
Concluding, the difference between the Web 2.0 and the Semantic Web is that techniques are quite simply to apply and are actually used today. Seen from the perspective of a professional in information architecture and computer science having various and less powerful techniques than those proposed by the Semantic Web is inferior. Yet the data is at hand and we can concentrate on the actual problem: predicting semantics to help people.

Inferring Semantics from a Social Web

To infer its semantic every information or statement has to be reinterpreted in a person's context, first of all her individual context. The individual context includes basically every experience and memorizing capabilities. Any interaction with a computer system such as buying a book can be tracked to evaluate a person's individual context. Yet, predicting semantics of a newly encountered information based on the individual context is difficult as it requires to comprehend how people think, an effort that is hopelessly pursued by the scientific field of AI.
This is why we propose the "Social Web". The context of a person is also made up by her social network. If we manage to identify people from the social network who are most similar regarding a matter or who even have a shared context (cf. Chapter 2.2), we can use their assured experience to predict semantics in a newly encountered environment.
Social networks can be inferred from a variety of information available on the internet including links, blogrolls and FOAF files. Starting from a weblog, pointers to related people's weblogs, quoted articles and FOAF files can be used to build the social network of a person in an iterative process because each related resource may have pointers to other related people as well. *GROOP.US* implemented a different approach, we created a new network based on similar tagging behavior and evaluated the results afterwards. Other approaches are conceivable and already in use, for example based on viewing, rating or buying items in recommender systems. Yet social network information is lost when data or user profiles are exported from those systems.
Therefore context information must be added to the data streams obtainable from social web services.

To summarize this, services of the Social Web relate people in different contexts based on data from the Web 2.0 in order to help them finding what is relevant in a certain situation. Ideally, the social network is conveyed and shaped during the process of information processing to be made available to other applications, again.

This scenario raises two important issues. Global identity management on the one hand and storage of newly created social network data on the other.
A person must be globally identifiable when network information shall be reused in another service where she has no account or a different username. There are description formats like FOAF

and Global identity solutions such as Light-Weight Identity [Ernst2005]. Both allow for username mapping. But everybody can fake and upload a FOAF file which raises the issue of trust. Therefore, a secure identity management solution is a better choice. Who is able to retrieve which identity information must is controlled by the owner. FOAF and many global identity management solutions can hold differentiated profile information including relations to other people (cf. [Koch2002]).
This brings us to the second issue, the storage of social network information of a person. Similar to usernames this can be seen as a unique property of a person that has to be stored globally. An ontology, for example an extended version of the RDF format FOAF, seems to be a good solution here. If every web service updates global relation information of its user, a Semantic Web of contextualized relations could emerge.

But we have to consider the lesson learned from the attempt to spread the Semantic Web. If the organizational and technical overhead is too large and the return of invest is poor, developers will not contribute. Let us consider the exchange of data from *GROOP.US*. A variety of information can be obtained: one can access every person's resources including the context in terms of number of tags, members, related people, their tie strength and a good deal more. Moreover, all contexts (foci) could be retrieved in terms of their main tags, people and resources involved. Resources could also be a starting point to retrieve contexts they are contained in with their main tags and so on. Modelling these relations in an ontology as well as understanding and retrieving the data from the ontology is a challenging task.
Therefore, we think that a "Web 2.0 approach" is a promising alternative to formal ontologies: The availability and meaning of data is transferred to other developers from knowing the application of its origin. Lightweight API calls or exchange formats are used to retrieve data. An example in *GROOP.US* would be an RSS export of all foci and related members which is referenced to from the corresponding interface. Other developers reuse the data by means of self-designed data mining and collaborative filtering techniques.

Again, the formal solution proposed first leads to more correct results and has greater potential, but the second approach is easier and already applied as shown by GROOP.US and other social web services.

We are not pessimistic about the Semantic Web but we believe that the Web 2.0 as described is a good intermediate step to show the social potential of the web - or the Social Web - for inferring semantics. Once the Social Web is filled with many services and a huge amount of data available, it might be a logical step to formalize it by means of Semantic Web technologies.
We hope to have contributed to this effort with the considerations and proposed solutions in this thesis.

Bibliography

[Baier2004]
Baier, Toby, Harald Weinreich, Frank Wollenweber. "Verbesserung von Social Navigation durch Identitätsmanagement". Mensch und Computer 2004: Allgegenwärtige Interaktion. Ed. R. Keil-Slawik, H. Selke, G. Szwillus. München: Oldenbourg Verlag, 2004: 189-198.
07.Mar 2005 http://vsis-www.informatik.uni-hamburg.de/publications/view.php/205

[Beckett2004]
Beckett, Dave. RDF/XML Syntax Specification (Revised). W3C. 10. Feb. 2004
01. Aug. 2005 http://www.w3c.org/TR/rdf-syntax-grammar/

[Berners-Lee2001]
Berners-Lee, Tim, James Hendler, Ora Lassila. "The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". ScientificAmerican.com. 17 May 2001
10. Aug. 2005 http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

[Boyd2005]
Boyd, danah. "questions of classification (a response to Clay)".Many2Many. 30. Jan 05.
15. Apr 2005
http://www.corante.com/many/archives/2005/01/30/questions_of_classification_a_response_to_clay.php

[Borgatti1997]
Borgatti, Stephen P., Martin G. Everett. "Network analysis of 2-mode data". Social Networks. Vol. 19. 1997: 243-269

[Brickley2000]
Brickley, Dan, et al. RDF Site Summary (RSS) 1.0. Dec. 2000.
01. Aug. 2005 http://web.resource.org/rss/1.0/spec

[Brickley2005]
Brickley, Dan, Libby Miller. "FOAF Vocabulary Specification ". 19. May 2005
01. Jun 2005. http://xmlns.com/foaf/0.1

[Celik2004]
Celik, Tantek, Kevin Marks. "Real World Semantics". eTech 2004. 2004
11. Aug 2005 http://www.tantek.com/presentations/2004etech/realworldsemanticspres.html

[Celik2005]
Celik, Tantek, Eric Meyer, Matthew Mullenweg. "XHTML Friends Network". Global Multimedia Protocols Group. 2003-2005.
15. Mar 2005 http://gmpg.org/xfn/

[Chalmers1999]
Chalmers, Matthew. "Informatics, Architecture and Language". Social Navigation of information space. Ed. A.Munro, K.Höök, D. Benyon. London: Springer-Verlag. 1999

[Chandler2002]
Chandler, Daniel. "Semiotics - the basics". Routledge. London. 2002.

[Coates2005]
Coates, Tom "An addendum to a definition of Social Software". PLASTICBAG.ORG, 05. Jan 2005
13. May 2005
http://www.plasticbag.org/archives/2005/01/an_addendum_to_a_definition_of_social_software.shtml

[Dieberger1999]
Dieberger A. "Social Connotations of Space in the Design for Virtual Communities and Social Navigation" Social Navigation of Information Space. Ed. A. Munro et al. Computer Supported Cooperative Work, Springer, London, 1999.
[Dieberger2000]
Dieberger A., P. Dourish, K. Höök, P. Resnick, A. Wexelblat. "Social Navigation - Techniques for Building More Usable Systems". Interactions 7(6). Nov+Dec 2000: 36-45
[Doctorow2001]
Doctorow, Cory. "Metacrap: Putting the torch to seven straw-men of the meta-utopia". 26. Aug 2001. 20. Apr 2005 http://www.well.com/~doctorow/metacrap.htm
[Donath1995]
Donath, Judith. "Sociable Information Spaces". Presented at the Second IEEE International Workshop on Community Networking. Princeton, NJ. 20. Jun 1995.
22. Mar 2005 ftp://ftp.media.mit.edu/pub/donath/SociableInfoSpace.ps
[Donath2004]
Donath, Judith, danah boyd. "Public displays of connection". BT Technology Journal Vol 22 No 4. Oct 2004: 71-82.
22. Mar 2005 http://smg.media.mit.edu/papers/Donath/PublicDisplays.pdf
[Dourish1993]
Dourish, P., V. Bellotti, W. Mackay and C.-Y. Ma. "Information and Context: Lessons from a Study of Two Shared Information Systems". Proceedings of the ACM Conference on Organisational Computing Systems COOCS '93 (Milpetas, CA), 42-51. New York: ACM. 1993 25. Apr 2005
http://www.ics.uci.edu/~jpd/publications/1993/coocs93-context.pdf
[Dourish1994]
Dourish, P. and M. Chalmers. "Running Out of Space: Models of Information Navigation". Short paper presented at HCI'94, Glasgow, UK. 2004
25. Apr 2005 http://www.ics.uci.edu/~jpd/publications/1994/hci94-navigation.pdf
[Dourish2004]
Dourish, Paul. What We Talk About When We Talk About Context. Personal and Ubiquitous Computing. Vol. 8(1). 2004: 19-30.
[Eco1976]
Eco, Umberto. "Semiotics and the Philosophy of Language". Basingstoke [u.a.] : Macmillan. 1984
[Erickson2003]
Erickson, T., Kellogg, W. A. "Social Translucence: Using Minimalist Visualizations of Social Activity to Support Collective Interaction." Designing Information Spaces: The Social Navigation Approach Ed. K. Höök, D. Benyon, A. Munro. Springer, 2003. 17-42.
[Ernst2005]
Ernst, Johannes. "White Paper: Light-Weight Identity 2.0". NetMesh Inc., 01. Aug 2005
11. Aug 2005 http://lid.netmesh.org/docs/NetMesh-LID.pdf
[Feld1981]
Feld, Scott L. "The Focussed Organization of Social Ties". The American Journal of Sociology, Vol. 86, No. 5. The University of Chicago Press. Mar 1981: 1015-1035.
[Feld1997]
Feld, Scott L. "Network analysis using the method of path lists: testing the weak ties hypothesis". Social Networks 19. Department of Sociology, Louisiana State University, USA. 1997: p. 97-111
[Freeman1977]
Feld, Linton C. "A Set of Measures of Centrality Based on Betweeness". Sociometry. Vol. 40, No.1. Mar 1977: p. 53-41

[Friedkin1980]
Friedkin, Noah, "A Test of the Structural Features of Granovetter's 'Strength of Weak Ties' Theory". Social Networks. Vol. 2. 1980: p. 411-422
[Girvan2002]
Girvan, M., M. E. J. Newmann. "Community structure in social and biological networks". PNAS. Vol. 99, No. 12. 11. Jun 2002: 7821-7826.
25. Apr. 2005 http://www.pnas.org/cgi/content/full/99/12/7821
[Golbeck2004]
Jennifer Golbeck. "The Trust Ontology". Trust and Reputation in Web Based Social Networks. 2004.
15. Mar 2005 http://trust.mindswap.org/trustOnt.shtml
[Golbeck2004a]
Jennifer Golbeck, Hendler, J. "Accuracy of metrics for inferring trust and reputation in semantic web-based social networks". Proceedings of EKAW. 2004.
08. Aug 2005 http://www.mindswap.org/papers/GolbeckEKAW04.pdf
[Granovetter1982]
Granovetter, M. "The Strength of Weak Ties: A Network Theory Revisited". Social Structure and Network Analysis. Ed. P. Marsden, L. Nan. Sage Publications, California. 1982
[Hammond2005]
Hammond, Tony. Timo Hannay, Ben Lund, and Joanna Scott "Social Bookmarking Tools (I) - A General Review". D-Lib Magazine. Vol. 11, No. 4. Apr. 2005
[Harper1999]
Harper, R. "Information that counts: A sociological view of information navigation". Social Navigation of information space. Ed. A.Munro, K.Höök, D. Benyon. London: Springer-Verlag. 1999
[Hillmann2003]
Hillmann, D. "Using Dublin Core ". Dublin Core Metadata Initiative, 23. Aug 2003
http://www.dublincore.org/documents/usageguide/
[Höök2003]
Höök, Kristina, David Benyon, Alan Munro. "Editors' Introduction: Footprints in the Snow". Social Navigation of information space. Ed. A.Munro, K.Höök, D. Benyon. London: Springer-Verlag. 1999
[Jain1999]
Jain, A.K., M.N. Murty, P.J. Flynn. "Data Clustering: A Review". ACM Computing Surveys. Vol 31, No.3. Sep. 1999
[Kautz1997]
Kautz, H., B. Selman, M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering". Communications of the ACM, Vol. 40, no. 3, Mar 1997. 63-65
08. Aug 2005 http://citeseer.ist.psu.edu/kautz97referralweb.html
[Kanfer1997]
Kanfer, Alaina, Jim Sweet, Ann Schlosser. "Humanizing the Net: Social Navigation With A 'Know-Who' Email Agent". Education & Outreach Technology Research Group. 12. Jun. 1997. National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.
04. Mar 2005 http://archive.ncsa.uiuc.edu/edu/trg/human/
[Kanfer2000]
Kanfer, A., B. C. Bruce, C. Haythornthwaite, N. Burbules, J. Wade, G. C. Bowker, J. Porac. "Modelling distributed knowledge processes in next generation multidisciplinary alliances" Information Systems Frontiers: A Journal of Research and Innovation, 2(3/4). 2000

[Koch2002]
Koch, Michael. "Global Identity Management to Boost Personalization". Proceedings of the Ninth Research Symposium on Emerging Electronic Markets 2002. Ed. Petra Schubert, Uwe Leimstoll. Basel, Switzerland: University of Applied Sciences Basel (FHBB), Institute for Business Economics, 2002: 137-148.
07. Mar 2005 http://www.e-business.fhbb.ch/eb/publications.nsf/id/137

[Konstan1997]
Konstan, J., A. Miller, B. N. Maltz, D. Herlocker, J. L. & Gordon, L. R. Riedl, J. "GroupLens: Applying Collaborative Filtering to Usenet News". "special section: recommendation systems". ACM March 1997, Vol. 40, No. 3. 77-87.
08. Aug 2005 http://citeseer.ist.psu.edu/konstan97grouplens.html

[Konstan2002]
Konstan, J., J. Riedl. "Collaborative Filtering: Supporting Social Navigation in Large, Crowded Infospaces". Designing Information Spaces: The Social Navigation Approach. Ed. K. Höök, D. Benyon, A.J. Munro. London: Springer Verlag. 2002. 43-83

[Lawley2005]
Lawley, Liz. "social consequences of social tagging".Many2Many. 20. Jan 05.
15. Apr 2005
http://www.corante.com/many/archives/2005/01/20/social_consequences_of_social_tagging.php

[Linden2003]
G. Linden, B. Smith, J. York "Amazon.com recommendations: item-to-item collaborative filtering" Internet Computing, IEEE Volume 7, Issue 1, Jan. - Feb. 2003. 76 - 80

[MacManus2005]
MacManus, Richard. "Web 2.0 Definition and Tagging" Read/Write Web - Web 2.0 News, Views and Ideas. 01.02.2005 18 Mar. 2005 http://www.readwriteweb.com/archives/002645.php

[Mathes2004]
Mathes, Adam. "Folksonomies - Cooperative Classification and Communication Through Shared Metadata". Computer Mediated Communication - LIS590CMC
Graduate School of Library and Information Science, University of Illinois Urbana-Champaign. Dec. 2004
18 Mar. 2005 http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

[Merholz2004]
Merholz, Peter. "Metadata for the Masses ". Adaptive Path. 19 Oct 2004,
18 Mar. 2005 http://www.adaptivepath.com/publications/essays/archives/000361.php

[Morris1938]
Morris, Charles William. "Grundlagen der Zeichentheorie". In: Ullstein Materialien. Carl Hanser Verlag. München. 1972: 17-68

[Michal2005]
Michal, Amir. "CollaborativeRank - Motivating People to Give Helpful and Timely Ranking Suggestions". 23. Apr 2005
08 Aug. 2005 http://collabrank.web.cse.unsw.edu.au/collabrank.pdf

[Minar1999]
Minar, N., J. Donath. "Visualizing the Crowds at a Web Site" CHI'99 Extended Abstracts, 1999, 186-187.

[Nadin1988]
Nadin, Mihai. "Interface design: A semiotic paradigm". Semiotica 69-3/4. Amsterdam: Mouton de Gruyter, 1988: 269-302

[Norvig2005]
Norvig, Peter. "Semantic Web Ontologies: What Works and What Doesn't". AlwaysOn - the insider's network. 12. Jan 2005.
11. Aug. 2005 http://www.alwayson-network.com/comments.php?id=7480_0_4_0_C
[Nottingham2005]
Nottingham, M., R. Sayre. The Atom Syndication Format, 11. Jul. 2005.
1. Aug. 2005 http://ietfreport.isoc.org/idref/draft-ietf-atompub-format/
[NISO2004]
National Information Standards Organisation. "Understanding Metadata ". Bethesda (USA): NISO Press. 2004
[Nöth2004]
Nöth, Winfried. "Handbuch der Semiotik". Stuttgart - Weimar. Verlag J.B. Metzler. 2000
[Palau2004]
Palau, J., M. Montaner, B. Lopez. "Collaboration Analysis in Recommender Systems using Social Networks". To be presented to the Eighth International Workshop on Cooperative Information Agents (CIA'04). Erfurt (Germany). 27-29 Sept. 2004.
[Passin2004]
Passin, Thomas B. "Explorer's Guide to the Semantic Web". Manning Publications, Greenwich. 2004.
[Pierce1931-74]
Pierce, Charles Sanders. "Collected Papers of Charles Sanders Pierce". Volume II, Elements of Logic. Harvard University Press. 1931-74
[Powers2005]
Powers, Shelly. "Cheap Eats at the Semantic Web Café". Burningbird. 27 Jan. 2005
18 Mar.2005 http://weblog.burningbird.net/archives/2005/01/27/cheap-eats-at-the-semantic-web-cafe/
[Porter2005]
Porter, Joshua. "Controlled Vocabularies Cut Off the Long Tail". Bokardo. 9 Mar. 2005
16 Apr.2005 http://bokardo.com/archives/controlled_vocabularies_long_tail/
[Prescod2002]
Prescod, Paul. Second Generation Web Services. O'Reilly Media. 06. Feb. 2002
01 Aug. 2005http://webservices.xml.com/pub/a/ws/2002/02/06/rest.html
[Resnick1994]
Resnick, Paul, N. Iacovou, M. Suchak, P. Bergstorm, J. Riedl "GroupLens: An Open Architecture for Collaborative Filtering of Netnews". Internal Research Report. MIT Center for Coordination Science, March 1994.
06. May 2005 http://citeseer.ist.psu.edu/resnick94grouplens.html
[Richardson2005]
Leonard Richardson. "The Ultra Gleeper: A Recommendation Engine for Web Pages". 02. Jun 2005
08. Aug 2005 http://www.crummy.com/software/UltraGleeper/IntroPaper.html
[Riedl2001]
Riedl, Mark O. "A computational model and classification framework for social navigation". Intelligent User Interfaces. 2001: 137-144.
08 Mar.2005 http://citeseer.ist.psu.edu/mark01computational.html
[Rosenfeld2005]
Rosenfeld, Louis. "Folksonomies? How about Metadata Ecologies?". LouisRosenfeld.com. 06. Jan 2005.
16. Apr 2005 http://www.louisrosenfeld.com/home/bloug_archive/000330.html
[Ruby2005]
Ruby, Sam. "Agile Web 2.0 Development". Intertwingly - It's just data. 08. Aug 2005.
11. Aug 2005 http://www.intertwingly.net/blog/2005/08/08/Agile-Web-2-0-Development

[Rupp2002]
Rupp, Chris, SOPHIST GROUP. "Requirement-Engineering und -Management - Professionelle, iterative Anforderungsanalyse für die Praxis". Hanser, München 2002.
[Sarwar2001]
Sarwar, B. M., G. Karypis, J. A. Konstan, J. Riedl. "Item-based collaborative filtering recommendation algorithms". In Proceedings of the 10th International World Wide Web Conference (WWW10), Hong Kong. May 2001.
04. Mai 2005 http://www.grouplens.org/papers/pdf/www10_sarwar.pdf
[Shirky2003]
Shirky, Clay. "Social Software and the Politics of Groups".
Clay Shirky's Writings About the Internet - Economics & Culture, Media & Community, Open Source. 09 Mar 2003.
18 Mar 2005 http://shirky.com/writings/group_politics.html
[Shirky2003a]
Shirky, Clay. "The Semantic Web, Syllogism, and Worldview". Clay Shirky's Writings About the Internet - Economics & Culture, Media & Community, Open Source. 07 Nov 2003.
11 Aug 2005 http://www.shirky.com/writings/semantic_syllogism.html
[Shirky2005a]
Shirky, Clay. "Tags run amok!".Many2Many. 1. Feb 2005.
15. Apr 2005 http://www.corante.com/many/archives/2005/02/01/tags_run_amok.php
[Shirky2005b]
Shirky, Clay. "Folksonomies are a forced move: A response to Liz".Many2Many. 22. Jan 2005.
15. Apr 2005
http://www.corante.com/many/archives/2005/01/22/folksonomies_are_a_forced_move_a_response_to_liz.php
[Smith2005]
Smith, Gene. "Sorting Out Social Classification ".
Information Arichtecture Summit. Montreal. 4 Mar. 2005.
15 Apr. 2005 http://atomiq.org/etc/folksonomies_smith_ia_summit_2005.pdf
[Speroni2005]
Speroni, Pietro. On Tag Clouds, Metric, Tag Sets and Power Laws. 25. May 2005
09. Jun. 2005 http://blog.pietrosperoni.it/2005/05/25/tag-clouds-metric/
[Sterling2005]
Sterling, Bruce. "Order out of Chaos". Weird. Issue 13. 04. Apr. 2005.
15 Apr. 2005 http://wired-vig.wired.com/wired/archive/13.04/view.html?pg=4
[Svensson2003]
Svensson, Martin, Kristina Höök. "Social Navigation of Food Recipes: Designing Kalas" Designing Information Spaces: The Social Navigation Approach Ed. K. Höök, D. Benyon, A. Munro. Springer, 2003. 201-222.
[Udell2004]
Udell, John. "Collaborative knowledge gardening". Infoworld. 20 Aug. 2004.
15 Apr. 2005 http://www.infoworld.com/article/04/08/20/34OPstrategic_1.html
[VanderWal2004]
Vander Wal , Thomas. "Explaining and Showing Broad and Narrow Folksonomies".
vanderwal.net - Off the Top. 21 Feb. 2005.
15 Apr. 2005 http://www.vanderwal.net/random/entrysel.php?blog=1635

[Viégas2004]
Fernanda Viégas, danah boyd, David H. Nguyen, Jeffrey Potter, Judith Donath. "Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments". HICSS-37. Hawaii, HI. 5-8 January, 2004.
[Wegner1987]
Wegner, d.M. "Transactive memory: A contemporary analysis of the group mind." Theories of Group Behavior. Ed. B. Mullen and G.R.Goethals. New York: Springer-Verlag, 1986: 185-208
17 Mar. 2005 http://www.wjh.harvard.edu/~wegner/pdfs/Wegner%20Transactive%20Memory.pdf
[Wegner1995]
Wegner, D. M. "A computer network model of human transactive memory". Social Cognition, 13. 1995: 1-21
08 Mar. 2005
http://www.wjh.harvard.edu/~wegner/pdfs/Wegner%20Computer%20Network%20Model%201995.pdf
[Weinberger2005]
Weinberger, David. "By their tags shall ye know them". Operating Manual for Social Tools. 19 Jan. 2005
15 Apr. 2005 http://www.corante.com/om/archives/032470.html
[Wellmann1990]
Wellmann, Barry, Scott Wortley. "Different Strokes from Different Folks: Community Ties and Social Support". The American Journal of Sociology, Vol.96, No. 3, Nov. 1990: 558-588
[Wellmann1996]
Wellmann, Barry, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gulia, Caroline Haythornthwaite. "Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community". The American Journal of Sociology, Vol.22., 1996: 213-238
[Wolf1999]
Wolf, J., C. Aggarwal, K-L. Wu, P. Yu. "Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering". Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, San Diego, CA. 1999.
09. May 2005 http://citeseer.ist.psu.edu/aggarwal99horting.html
[WordNet2003]
WordNet - a lexical database for the English language. Cognitive Science Laboratory, Princeton University, 2003
14 Apr. 2005 http://wordnet.princeton.edu