

Social Search and Need-driven Knowledge Sharing in Wikis with Woogle

Hans-Jörg Happel

FZI Research Center for Information Technology
Karlsruhe, Germany

happel@fzi.de

ABSTRACT

Wikis have emerged as easy and flexible tools for accessing and sharing knowledge on the internet and within organizations. Due to their highly collaborative nature, users can typically modify and create content without any access restrictions since these are replaced by social practices. On the other hand, due to the lack of central guidance, knowledge sharing is driven by the individual choice of contributors. We claim that – especially in enterprise settings – the lack of guidance for content creation is a major problem. We therefore analyze search and knowledge sharing in current Wikis. To improve knowledge sharing, we present the Woogle concept, which introduces “Social Search” and “Need-driven Knowledge Sharing” into Wikis. We describe our prototype implementation Woogle4MediaWiki and results from initial evaluation studies.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*search process*; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

General Terms

Documentation, Design, Experimentation, Human Factors

Keywords

Wikis, Knowledge Sharing, Social Search

1. INTRODUCTION

Exchanging information is vital for the success of modern organizations. Especially in large organizations or distributed collaboration settings, when direct communication is often not feasible, information technologies play a key role for information exchange.

Wikis can be considered as a category of next generation, web-based groupware systems which allow for easily capturing and disseminating information in a community or organization. While initially rooted in the web, such as the design patterns community or the Wikipedia, there has been an increasing adoption rate concerning corporate Wikis in recent years [7][16].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '09, October 25–27, 2009, Orlando, Florida, U.S.A.
Copyright © 2009 ACM 978-1-60558-730-1/09/10...\$10.00.

Wikis are distinct from conventional groupware systems by stressing the web aspect (e.g. by requiring a web-browser only and heavily relying on hyperlinks), favoring the free editing of content instead of access permissions and by fostering the accumulation/maturing of information by enforcing conceptual integrity between URL, page title and content [34].

In enterprises, Wikis are typically used to collect and refine small pieces of unstandardized, immature information [9] and thus as a common entry point when searching information for which no specific other information system exists. However, this makes it difficult for many users to find out why and what information they should put into the Wiki, which often leads to user adoption problems [28].

Due to its collaborative editing features, Wikis offer several discussion and awareness mechanisms such as a recent changes list, change notifications and discussion pages. Wikis provide space for discussing, commenting and linking to (web) resources outside the Wiki such as intranet pages or files, and can thus help to “glue” together distributed information.

However, while Wikis are a place for information seeking and knowledge sharing, both aspects are only loosely related. Wiki collaboration features are typically not available for search activities and knowledge sharing is not directly guided by the information needs of the users.

Recent research however reports, that knowledge sharing is a fundamental part of the information seeking process [32]. In a study with 150 participants, Evans and Chi found out that 58.7% of them were sharing knowledge related to their search results with others [17]. Therefore it seems as if the decoupling of information seeking and information provision, as we can observe in Wikis, blocks an important communication and feedback channel.

We argue that information access and knowledge sharing in Wikis are different, but closely related aspects of information processing which can significantly benefit from each other. To this end, we designed Woogle, a concept to improve search with collaboration features (“social search”) and to guide knowledge sharing with actual information needs of the user community (“need-driven knowledge sharing”).

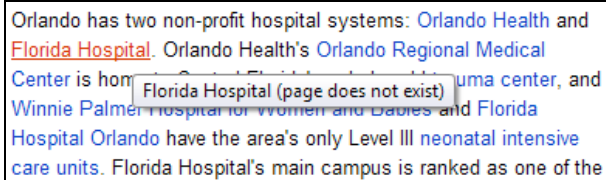
In the following, we first discuss the current state of search and knowledge sharing in Wikis. Then we describe Woogle as a concept to improve both issues. Therefore we line out the general design principles and present our Woogle reference implementation, Woogle4MediaWiki. Afterwards, we describe our evaluation approach and discuss initial results. We also shortly summarize related work and give an outlook to ongoing and future work.

2. SEARCH AND KNOWLEDGE SHARING IN WIKIS

We first review different modes of a) expressing information needs and b) knowledge sharing within Wikis. Since our reference implementation is based on MediaWiki and many of our example screenshots stem from the Wikipedia (which is also using MediaWiki), our discussion will mainly draw from this system. However, most other Wiki engines are based on the same or very similar concepts.

2.1 Information Need

The information need of a user is a primary subject of investigation in information retrieval (IR). The main purpose of IR systems is to help users satisfying their information needs by providing a set of relevant documents. A personal information need can be defined as information which a user requires to complete a specific task [6]. To use an IR system, the user typically has to express this information need in terms of the query language which can be interpreted by the search system. In most systems, this is a textual, “keyword-based” representation of the information need.

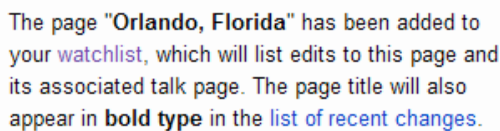


Orlando has two non-profit hospital systems: [Orlando Health](#) and [Florida Hospital](#). Orlando Health's [Orlando Regional Medical Center](#) is home to the [Orlando Cancer Center](#), [Orlando Health's Regional Cancer Center](#), and [Winnie Palmer Hospital for Women and Babies](#) and [Florida Hospital Orlando](#) have the area's only Level III neonatal intensive care units. Florida Hospital's main campus is ranked as one of the

Figure 1: “Red link” in the English Wikipedia

Wikis offer various means of expressing information needs, which differ with respect to explicitness and expressivity. The most obvious one is of course the *search* function. However, a vanilla MediaWiki instance does not maintain a query log to record such information needs.

The second, more explicit means is related to linking to other Wiki articles. Wikis are unique in allowing to *reference non-existing pages*. In contrast to normal web links to non-existing URLs, these references are not rendered with a “404 – not found” error message, but the Wiki provides an empty edit form to write a new article. Within the referencing text, such links are typically rendered in red color to denote that this page is still to be written and thus implies an information need (c.f. Figure 1).



The page "[Orlando, Florida](#)" has been added to your watchlist, which will list edits to this page and its associated talk page. The page title will also appear in bold type in the [list of recent changes](#).

Figure 2: Adding an article to the watchlist

A more personal form of expressing an information need is to “watch” pages in the Wiki (c.f. Figure 2). In this case, the user will be notified about changes – either by highlighting or by E-Mail. While this feature is primarily used for awareness and vandalism protection, some people might also use it to learn about

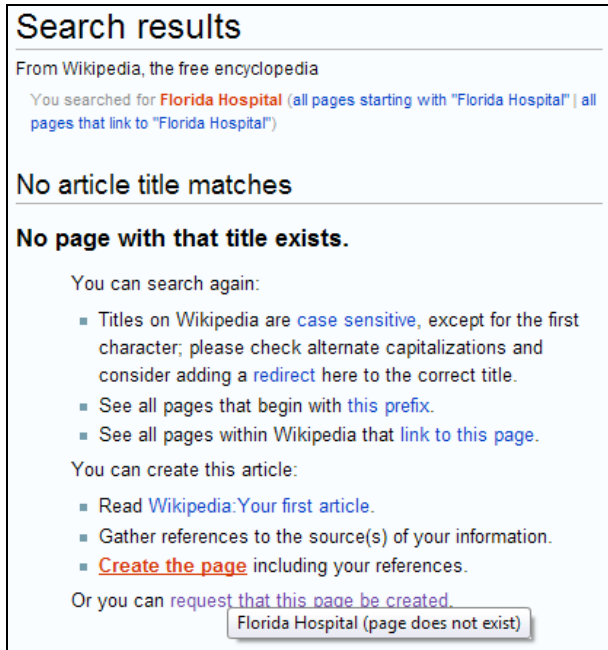
new information about a topic. This makes sense, since information on a topic might be evolving [37], especially in a Wiki system. Also, MediaWiki allows to watch empty, “red link” articles.

Finally, some communities provide means for explicitly *requesting articles*. For instance, the English Wikipedia community maintains lists with “requested pages”¹. Uttering such requests is also suggested from within the Wikipedia search interface, if no results can be found (c.f. Figure 3). However, this is not a built-in feature of the Wiki engine, but rather a community-maintained effort.

2.2 Knowledge Sharing

The topic of knowledge sharing is widely investigated in the fields of organizational studies and information systems. It can be defined as a “dual problem of searching for (looking for and identifying) and transferring (moving and incorporating) knowledge across organization subunits” [19].

Personalization and codification are typically described to be the two core strategies for knowledge sharing [14]. The personalization strategy primarily relies on personal communication to share implicit knowledge, which only exists in the “heads” of individuals (see e.g. [14]). In turn, the codification strategy targets explicit knowledge which is captured in documents. For this scenario, information technology is an important enabler, especially in cases when regular personal communication is not feasible, such as in large organizations or distributed settings [4][27].



Search results
From Wikipedia, the free encyclopedia
You searched for **Florida Hospital** (all pages starting with "Florida Hospital" | all pages that link to "Florida Hospital")

No article title matches

No page with that title exists.

You can search again:

- Titles on Wikipedia are **case sensitive**, except for the first character; please check alternate capitalizations and consider adding a [redirect](#) here to the correct title.
- See all pages that begin with [this prefix](#).
- See all pages within Wikipedia that [link to this page](#).

You can create this article:

- Read [Wikipedia:Your first article](#).
- Gather references to the source(s) of your information.
- [Create the page](#) including your references.

Or you can [request that this page be created](#).
Florida Hospital (page does not exist)

Figure 3: Unsatisfied search in the English Wikipedia

However, the success of knowledge management systems (KMS) such as Wikis largely depends on individuals contributing their knowledge. One of the major challenges of KMS is to overcome

¹ http://en.wikipedia.org/wiki/Wikipedia:Requested_articles

barriers that prohibit people from doing so. The most important barriers are:

- **Motivation:** People usually have a low motivation to contribute knowledge to public repositories. Reasons are a lack of personal benefit [12][11][36] and privacy, since people do not like to expose their information and expertise to others [5][24].
- **Effort:** Computer-supported knowledge sharing initiatives require effort for creating and maintaining central knowledge repositories. This includes the cost of knowledge capturing, categorization and setting access rights for documents [24][25].

Knowledge sharing in Wikis can occur in various forms. A plain Wikis come only with a single entry page, containing some introductory information. The main way to share knowledge is thus typically to *contribute to an existing article*. However, this way only additional information on an existing topic can be added, but no totally new information.

Therefore, *red links* – i.e. pages without content (c.f. previous section) – are a second entry point to share knowledge. There are several ways to encounter red links. The obvious one is within an existing article (c.f. Figure 1). Furthermore, MediaWiki provides a page “Special:WantedPages” which lists empty pages and also displays its number of inlinks. However, this listing does not take into account particular topic areas. Larger communities such as the Wikipedia address this aspect by maintaining own lists of “requested articles” (c.f. section 2.1). Unfortunately, these lists, while providing some topical order, do not denote information about the size of demand for a given article.

Finally, also *search* can be an entry point for knowledge sharing. Since Wikis are about easy content sharing, MediaWiki offers the possibility to directly create a page from within the search dialog (see Figure 3).

2.3 Summary

We consider search as a communication process between information seekers and information providers. While in face-to-face settings, a providers’ contribution is typically tailored towards the request of the information seeker, this is typically not the case in distributed, asynchronous settings, where information providers are usually decoupled from search interfaces.

Although one would expect that most information is put into search systems in order to get retrieved, publishing information is seldom driven by concrete demand. On the other hand, search interfaces are usually very limited in their modes of interaction such as issuing a query, refining a query and clicking results. There is no way in which search interfaces allow information seekers to signal potential information providers the need for information. Queries, as the most important communication artifact in these settings, do typically not have a persistent, first order representation.

Thus, in computer-assisted settings, there is a fundamental gap between the creation of information in information systems and its retrieval and access. This gap results in a problem of representing desired information from the perspective of the information seeker, and in a problem of knowing how much certain information is desired from the perspective of the information provider.

Our main research question is thus how to establish a feedback loop between the processes of information retrieval and information provision.

3. WOOGLE

In this section we describe Woogle, a prototype system which realizes what we call “need-driven knowledge sharing” within a Wiki environment. We first describe the conceptual foundations and design rationales for Woogle. Afterwards, we discuss the architecture and implementation of the system.

3.1 Concept

3.1.1 Social Search

Our core idea is to bridge the artificial separation of information seeking and information provision in current information systems by seamlessly integrating and mutually improving both processes. In particular, we strive to improve information access in terms of better representing desired information and supporting the disambiguation of information needs. Information provision in turn should be improved by providing an explicit notion of sought information needs and by providing means for an easy sharing of information within the information seeking process. We thus derive the following design principles:

Improving collaboration

- **Give queries a first order representation** to serve as a common point of reference during information seeking and information provision.
- **Provide means for communication and awareness** such as discussion and notification mechanisms.

Improving provision

- **Seamless transition** from information seeking to information provision in order to achieve tight integration and holistic support for user’s information behavior.
- **Provide different modes of information provision**, such as creating explicit information need descriptions, annotations and adding new information easily and without entry barriers.

The intended mechanisms should generally be able to cover the whole collaboration process in information seeking which can be separated into the three phases “before search”, “during search” and “after search” [17].

3.1.2 Need-driven Knowledge Sharing

In section 2.2, we have described that time and effort are among the major barriers for knowledge sharing. It is clear, that in an enterprise setting – but also in large communities such as Wikipedia – resources for knowledge creation and sharing are limited. Thus, knowledge sharing activities should be prioritized towards knowledge which satisfies the most information needs.

Furthermore, studies have shown that people generally like to share knowledge with others, but only if they know, that it will be useful for them [15]. In the past, many knowledge sharing

initiatives failed since users could not see any value in sharing with others [31].

We there introduce the concept of “need-driven knowledge sharing” (NKS) to address both of these issues [22]. NKS is based on actual information needs of users of a system and thus makes sure that shared knowledge meets a concrete demand. Furthermore, analyzing information needs of different users helps to prioritize which knowledge would be most relevant for the whole community. We thus expect that a system based on NKS will improve and foster knowledge sharing.

We introduce the following major design principles for NKS:

- **Minimize effort:** In order to achieve the most impact with the least effort – i.e. efficient knowledge sharing – it needs to be guided in terms of demands. Since covering all information needs would exceed the knowledge sharing capacities in most environments, we want to provide clues which are the most important information needs that cover a large aggregate information need.
- **Motivate to share:** Providers should be motivated to share relevant knowledge with information seekers. Traditional KMS and knowledge sharing practices often require sharing information without signaling any benefit to the provider. Thus, those practices are often perceived as self-purpose with unclear value. In opposite to this, we want to give the potential information provider more concrete information that can help to estimate the benefit of sharing certain knowledge.
- **Retain privacy.** KMS implementations often lack acceptance, since contributing knowledge to the public space means losing control about it. However, many information providers want to retain such control, since information might be premature or sensitive [31]. Also, the storage and analysis of information needs required careful attention within our approach [8].

As already stated, these principles are not implemented by state-of-the-art Wiki engines. However, as we will show in the following paragraphs, only slight modifications to existing architectures are required in order to realize the potential benefits for knowledge sharing.

3.2 Implementation

For our reference implementation we chose to develop a plug-in for the MediaWiki engine, since it is the most popular Open Source Wiki and since it comes with powerful extension mechanisms. We are currently also developing a Woogle implementation for the Atlassian Confluence Wiki engine. However, discussion and screenshots in this section will be based on Woogle4MediaWiki.

3.2.1 General architecture

Woogle4MediaWiki is realized as an extension for the MediaWiki engine. It employs various so called “hooks” to extend the functionality and user interface of MediaWiki. The major modification regarding the user interface is to replace the build-in MediaWiki search and result presentation forms. Other modifications will be discussed in the course of this section.

Regarding backend functionality, Woogle4MediaWiki introduces two major features. One is a keyword-document index, which is used to improve the search experience. Furthermore, additional database tables are introduced to store query and result click information (c.f. section 3.2.3.1).

As depicted in Figure 4, the backend can be instantiated in two different ways. One is called “WoogleNative” and realizes the described modifications within the MediaWiki system itself. The index is realized using the Zend Lucene library.²

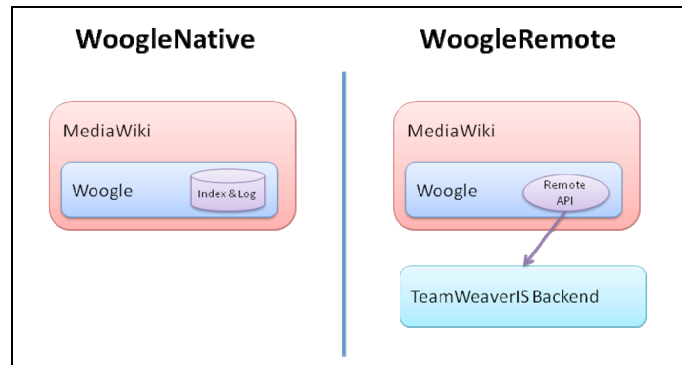


Figure 4: General architecture

The second option called “WoogleRemote” connects the MediaWiki instance with a remote installation of a software called “TeamWeaver Integrated Search” (TeamWeaverIS³). This is an advanced information retrieval framework, which allows to index and search documents from a large set of different repositories and/or data formats. Woogle can access any TeamWeaverIS backend via a web service interface. Wiki pages are also indexed within the remote backend and changes are propagated in near-realtime.

While requiring a more complex technical setup, WoogleRemote allows realizing Enterprise Search functionality using the Wiki as a user interface [23]. WoogleNative in turn does not support other content than Wiki pages and will probably not be suitable for very large Wikis, but provides easy to install and easy to use benefits for smaller MediaWikis.

3.2.2 Social Search

To realize our approach we created a special “namespace” in MediaWiki, meaning that all pages with an URL-prefix “Woogle:” are processed by our extension. Accordingly, queries are represented as “Woogle:query”, yielding bookmarkable query pages which do not interfere with the regular Wiki content.

Once calling such a “Woogle-Page” – which could be launched by the URL or via the MediaWiki search box – a Wiki page which contains a list of search results at its bottom is shown (c.f. Figure 5). For WoogleRemote, the results can also include external documents which are not part of the Wiki itself.

Besides the results, a “Woogle-Page” presents a number of immediate search-related “actions” to the user. Current possible actions are:

² <http://framework.zend.com/manual/en/zend.search.lucene.html>

³ <http://www.teamweaver.org>

- **Freely editable text box** at the top of the page to describe or disambiguate the information need. The text box is restricted in size in order to prevent the creation of too much “original content” on Woogle pages resp. to force users moving such content to regular Wiki pages.
- Immediate possibility to **create a new Wiki page**, if no suitable results exist.
- **Watchlist** to receive notifications when new results arrive or changes to the “Woogle-Page” occur.

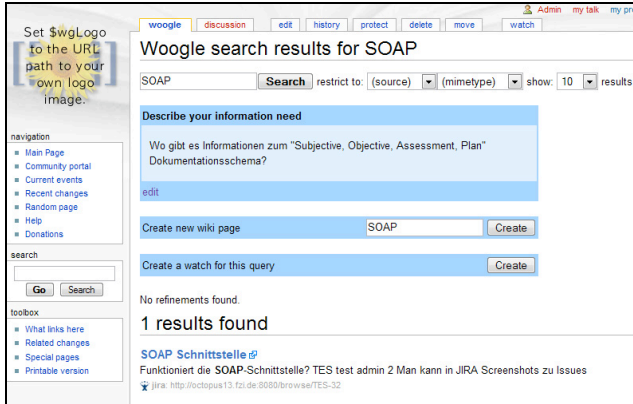


Figure 5: Woogle4MediaWiki screenshot

Besides these actions and the result list, “Woogle-Pages” are normal MediaWiki content pages with the following advantages:

- **Referenceability of query pages** from within the Wiki and from external applications.
- MediaWiki **discussion page** for each “Woogle-Page”.

As mentioned before, Woogle4MediaWiki can be configured to retrieve and preview results from different sources and formats.

3.2.3 Need-driven Knowledge Sharing

We will describe the Woogle implementation of NKS in two steps. First, we will describe how we derive and process information needs in the context of the Wiki system. Afterwards, we show how we support knowledge sharing with this information.

3.2.3.1 Collecting Information Needs

As stated in section 3.1.2, our concept of NKS is based on user’s *information needs*, for which we identified “red links”, watches, requests and queries as potential sources in section 2.1. While the former three are of rather qualitative nature, they might not occur in a sufficient scale in real world systems. “Red links” and watches can be easily retrieved from the MediaWiki API. Requests are currently not considered by our approach, since they are not collected in a standard, formalized way.

Queries are typically available in large amounts, but far more difficult to interpret. Therefore, based on this definition of personal information need in section 2.1, we conceptualize an aggregate information need (AIN) as an aggregate of the personal

information needs of members in a community or system. Thus, the aggregate information need denotes the overall amount of information, which the members of a community require to complete their particular tasks. Concerning keyword searches, an AIN should thus represent the most frequent queries that have been executed throughout the system.

Table 1. Aggregate query log

QueryId	Hashed value of the query
Needstring	Normalized query string
Needexec	Number of query executions
Needexecclick	Number of query executions without any result click
Recency	Average timestamp across all query executions
Firstdate	First time the query was executed
Lastdate	Last time the query was executed
Resulthits	Number of results for the query
Needexecpages	Number of query executions for which at least one further result page was browsed
Needexecavg-page	How many result pages did the users see in average for that query
Queryusers	Id’s of users executing the query
Sizequeryusers	Number of users listed in queryusers
Publicusers	Users who agreed to be publicly associated with the query

Aggregating queries provides two major benefits for us. First, an online processing of individual query instances would be computationally too expensive and logging individual queries would flood our database. Second, by aggregating information needs, those shared by a large number of users can be prioritized more easily.

Table 2. Aggregate query/click log

QueryId	QueryId
Subject	URI of the clicked result
Clicks	Number of clicks for a result after a certain query
Position	Position within the result list
Recency	Average timestamp across all clicks
Repold	Id of the repository where the result is stored (for WoogleNative, this is currently always the Wiki itself)
Firstdate	First time the result was clicked for this query
Lastdate	Last time the result was clicked for this query

Our logging scheme is presented in Table 1 and 2. The scheme and its respective instance data is persisted within the MediaWiki database. Besides the user queries we also observe and log result clicks, since they allow us to apply heuristics concerning the satisfaction of information needs. For deriving certain pieces of information (e.g. the number of searches without a following result click), we also implemented a simple session mechanism which enables us to observe series of user interactions.

3.2.3.2 Supporting Knowledge Sharing

Now, we finally describe how we apply the collected and aggregated information needs to drive the knowledge sharing process within the Wiki. In section 2.2, we identified

- a) contributions to existing pages
- b) red links
- c) search and
- d) overview pages

as main starting points for knowledge sharing. Our current implementation primarily addresses b) and c).

Regarding red links, we implemented functionality to provide Wikis users with direct feedback about how much a certain red linked page is sought. Therefore, we aggregate information about “inlinks” of a red link page and queries to a priority value. This is reflected in an icon which symbolizes the relative priority in its three different forms (low, medium and high priority). The number of people watching the red link page and the number of users searching for its title are aggregated into another icon denoting the “organizational breadth” of the information need (few people, average, many people). The icons including textual information are currently displayed in a JavaScript mouse-over popup (see Figure 6).

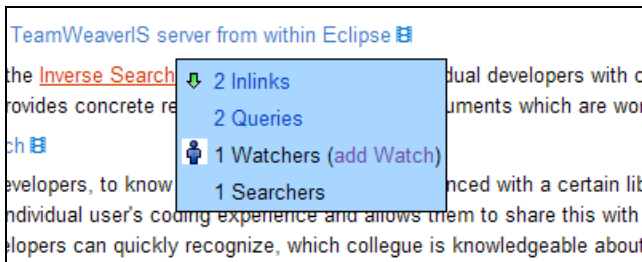


Figure 6: Woogle mouse-over display for “red links”

Our second focus is related to the search process. As described in section 3.1.1, search is a highly collaborative activity. Therefore, our “social search” features, offering collaboration and knowledge sharing facilities right within the search environment (c.f. section 3.2.2) are complemented by meta-information which signals the need for certain information.

Parts of this meta-information are similar to those described in the previous paragraphs related to red links. Besides icons for overall priority and breadth of information need, we add two further signals which characterize the information need. One is a clock icon, which denotes how the information need is distributed within time (recent, average or outdated). This can help potential contributors to estimate if there is recent demand for some information.

Finally, we add a traffic light icon, which symbolizes the result quality for the given query. To estimate this, we devised a number of heuristics. First, we assume a good result quality, if the number of clicks per query execution is near to 1. A lower result quality is indicated, if it is significantly below 1 (which means, that there are no result clicks for some queries), above 1 (which indicates that information may be scattered across many results) or if users tend to browse across result pages (which indicates a low satisfaction with information on the first result pages resp. a high information need).

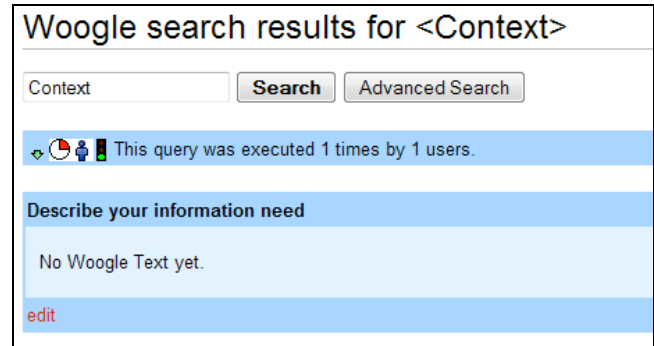


Figure 7: Information need indicators in the search UI

All these icons are currently displayed for each query in the search interface as shown in Figure 7. However, feedback from our initial user studies (c.f. section 4.3.1.2) might let us try alternative visualizations of information needs.

Among the aspects currently not addressed by our implementation are a) contributions to existing pages and d) overview pages. However, solutions for both approaches are planned and will be included in forthcoming releases of our system.

Regarding a) we want to provide an explicit, formal “request” action, which allows users to demand information concerning a given page (or query) complemented by a short rationale for their request. This information can be used to either route these request to appropriate contributors (e.g. similar to [13]) or by displaying it as meta-information for a page/query.

3.2.4 Privacy concerns

When discussing social search and knowledge sharing, privacy is a major issue for many users [8]. For the realization of our system, we had to address it in two different ways – concerning the operational part and concerning the evaluation.

The operational part is about the nature of our systems design, which is heavily based on user queries and result clicks (c.f. section 3.2.3). Therefore we addressed privacy by introducing a client-specific randomized userId. This Id, which can also be changed by the user at any time, is only used for query logging and thus keeps data sent to the remote query service anonymized. Users may also disable the submission and logging of query and click information at all. However, if too many users decide to do so, this would spoil our core idea of need-driven knowledge sharing.

Scientific evaluation of our system (c.f. section 4) is even more critical, since it needs to gather additional data beyond the operational core for scientific analysis. In order to take user’s privacy concerns serious, we implemented a participation dialog

which in detail explains the data logged by the system and which asks for explicit agreement of the user. The user may revisit this decision at any time in the preferences, where also the Woogle extension as such can be disabled.

4. EVALUATION

This section describes the evaluation goals, the process for evaluating the Woogle system and its current state of execution.

4.1 Evaluation goals

The main goal of our system is to improve search and knowledge sharing within Wikis by creating a feedback loop between information needs and knowledge sharing.

Therefore, there are two major claims that need to be validated:

- a) Is the system effective in improving/motivating knowledge sharing within a Wiki?
- b) Do users understand the system and are they willing to use it?

Clearly, question b) is *both* a prerequisite and a successor for question a). Therefore, we decided to design a three-step evaluation process, which will be lined out in the following.

4.2 Evaluation process

The evaluation of a system like Woogle is a rather difficult challenge, since it requires a large number of users and queries (to allow for meaningful aggregation) and a long time frame (to allow for collecting a broad number of unsatisfied needs). Therefore, laboratory experiments are not well suited, since it would be difficult to simulate real world information needs and to obtain a sufficient number of participants. We finally chose a mix-method evaluation approach which we describe in the following.

4.2.1 Qualitative evaluation

We decided to start with conducting qualitative interviews with a limited numbers of potential users. Goal of these interviews is to:

- a) Collect users' opinion on critical issues regarding contribution behavior and privacy awareness,
- b) Get feedback on the system as such,
- c) Devise design changes from users' feedback.

Therefore, we designed the qualitative interview with two major building blocks. The first one is a semi-structured interview, lasting around 30 minutes which covers a number of questions regarding information needs, knowledge sharing and privacy as drawn from existing literature.

The second phase, also lasting around 30 minutes, provides a walkthrough of the current Woogle4MediaWiki system. The system features are explained to the users and they are asked to describe if they understood the system, criticize it and to optionally provide additional ideas or feature suggestions.

4.2.2 Online field experiment

For the second stage of our evaluation, we chose the instrument of an online field experiment. In an online field experiment, users within a community are given different experimental treatments in order to investigate if a statistical difference regarding certain behavior can be observed among these groups. The advantage of this method is that it tries to combine the experimental control of

laboratory studies with the field study advantage of operating in real world environments [26].

Our major evaluation goal for the online field experiment is to investigate if Woogle can help to motivate users to share knowledge. Therefore we create three different experimental groups. Group "alpha" is our control group which receives the normal, plain MediaWiki search. Group "beta" receives an improved Woogle search interface, with the social search features described in section 3.2.1. However, this group does not receive any automatically derived meta-information about actual information needs. This information, depicted in Figure 7 is only shown to group "gamma". The distinction between "beta" and "gamma" is necessary to identify which effects can be accounted to the raw social search interface and which to the information need indicators.

As described in section 3.2.4, our implementation includes an instrumentation framework which allows to a) collect detailed log data about user's behavior and b) receive informed consent from participants. The instrumentation component can be deployed to any Woogle4MediaWiki instance at any time. Once activated, registered users will be presented a participation dialog when accessing the Wiki search for the next time. If they agree to participate, they are randomly assigned to one of the experimental groups. Users are free to revoke their participation at any time in the MediaWiki preferences menu.

4.2.3 Sharing survey

In a third step, we aim to conduct a broader empirical study, investigating if a system like Woogle would meet a more general acceptance of users. Furthermore, we want to check feature suggestions which were made during development and were not yet implemented in our current release.

While we received initial feedback on these issues in our qualitative interviews, we want to broaden our quantitative and cultural selection by addressing a larger field of participants. Since similar limitations apply to existing studies on sharing behavior and privacy concerns (e.g. [15][30][8]) we expect more generalizable insights into these aspects from our survey.

4.3 Evaluation results

In this section we describe initial results and the current state of our evaluation process. The sharing survey is omitted, since it will be conducted at a later time.

4.3.1 Qualitative evaluation

During the last weeks, we conducted an initial set of qualitative interviews with five users. All of these users were experienced in the field of computer science. Three were graduates and two were students of computer science or related disciplines. All users reported to at least occasionally use Wikis.

We will first summarize the initial semi-structured interview and then describe the feedback the users gave for the Woogle prototype.

4.3.1.1 Semi-structured interview

The interview started with questions related to the information search infrastructure and information need behaviour. Users primarily relied on searching in E-Mail or Desktop search (for their private space) or Google (for the public space). "Organizational spaces" such as Wikis were far less frequently

searched – mostly due to inconvenient search features. Concerning frequency, users search the private sphere approximately 5 times a day and the web from 10 to 50 times a day.

Users reported that they typically had short term information needs – i.e. were typically only interested in immediate results. As for the internet search, users indicated that they typically continue to search until they find at least a rough solution for their problem. A major problem described for private and internet search was formulating and finding the most appropriate query terms. All users described their search as an explorative trial- and error process in such situations.

Regarding their sharing behavior, the majority of our users followed an “on-demand” approach – i.e. if they were directly approached by other people. Only two users shared knowledge in organizational systems occasionally without explicit triggers.

The most interesting results appeared in the privacy part of our interview. All five subjects could be rated to be slightly concerned about privacy issues – i.e. they were aware of potential privacy problems but they typically made a trade-off with benefits they receive in turn and effort for retaining privacy.

On the other hand, users differed regarding their treatment of content. Two people shared virtually all of their work-related information openly (e.g. in shared folders), while three people rather kept most of their work-related information private. This was mainly justified with the fear of being attributed for some immature or sloppy drafts of content.

Finally, regarding privacy in the context of search, only one user was really concerned about revealing search queries within the work environment. While our other interview partners were more relaxed, they however preferred a selective revelation of such information (e.g. reciprocally, only if another user searches for the same term) over a general listing of all searches carried out within the organization. They were even less concerned, when the query information would not be directly associated with their names.

4.3.1.2 Evaluation of the Woogle prototype

The Woogle4MediaWiki prototype was generally appreciated by all of our interview partners. Two users explicitly mentioned their impression that the features might motivate people to use the Wiki – although we did not reveal this as an explicit goal of our system.

Critical feedback was raised concerning the current style of representations. Three users were confused by the “information overload” created by the different icons and demanded a more simple representation. One user demanded the same for the social search features attached to the search result page.

Although not explicitly asked by us, four users were interested in getting more information about *who* was seeking certain information. While such features are foreseen in our systems design, we did not yet incorporate them into the live system. However, if feedback remains constantly positive concerning this issue, we might consider it for a second iteration release.

While these initial results are promising and give us some important feedback, qualitative studies are not finished yet. We aim to conduct such interviews with at least 3-5 additional users for each online field experiment site (c.f. next section).

4.3.2 Online field experiment

Our first online field experiment which is currently deployed, takes place within our institute’s internal Wiki system. The Wiki is used in a daily or weekly fashion for organization and knowledge management within our research group. It currently contains 1,875 content pages and is used by 160 users, 46 of them denoted active by the MediaWiki metrics (Special:Statistics).

Due to the status of deployment, we cannot yet make any reasonable statements concerning experimental results. In order to broaden the basis for forthcoming empirical analysis, we are discussing with several commercial organizations which are using MediaWiki to launch additional online field experiments.

5. RELATED WORK

Collaboration within the information seeking process has been discussed under the label of “collaborative information retrieval” earlier this decade [11][18] and very recently [33][29]. “Social search” is a related term which is widely used for various different approaches. Evans and Chi define it as:

“...an umbrella term used to describe search acts that make use of social interactions with others. These interactions may be explicit or implicit, co-located or remote, synchronous or asynchronous.” [17]

However – in contrast to our work – most of these approaches focus on synchronous collaboration aspects in mostly collocated settings. Also, while the model of Evans and Chi highlights various collaboration points within information seeking [17], the role of information providers is not explicitly addressed.

Nevertheless, some “social search” approaches investigate the suitability of Wikis to support search tasks. In a demonstration prototype⁴, the creators of Wikiseek make the whole result page for a specific query available as an editable Wiki page while the initial set of results is bootstrapped with Google results. This idea resembles the idea of human maintained “catalogues” such as the initial Yahoo Web-search using Wiki principles. Wikia search allows several actions such as adding, deleting, and commenting search results⁵ – a feature which has been recently adopted by Google as the “SearchWiki”⁶.

Q&A systems, such as Yahoo Answers⁷ or Ask.com⁸ acknowledge the social nature of information exchange processes by giving its users the opportunity to discuss questions and different answers. They also have an explicit representation of information needs. However, similar needs might be expressed independently of each other and allow for different competing answers. This leads to a diverse quality of answers depending on the type of request [3] and also lacks the “consensual spirit” of Wikis, which makes it costly for users to extract desired information from the results. Furthermore, these Q&A systems do typically not consider external result content.

Earlier works on organizational memory systems such as Answer Garden [1][2] connect information seekers and information

⁴ <http://community.wikiseek.com>

⁵ <http://search.wikia.com/>

⁶ <http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=115764>

⁷ <http://answers.yahoo.com>

⁸ <http://www.ask.com>

providers by triggering potential contributions based on user information needs. However, Answer Garden lacks advanced discussion features (in the initial version), restricts content creation and does not incorporate existing document repositories.

Finally, this work is complementary to other work of ours. In [21] we described the application of aggregated information needs for sharing documents from private information spaces. The general idea of adding social search features to a Wikis has been presented in [23]. In [20], we have discussed the application of need-driven knowledge sharing to semantic Wikis.

6. CONCLUSION

In this paper, we argued that search and knowledge sharing within a Wikis can be mutually beneficial for each other. We presented the Woogle concept which introduces social search and need-driven knowledge sharing to this end.

We described our reference implementation Woogle4MediaWiki and lined out our evaluation strategy. Initial feedback from these studies makes us confident, that Woogle will achieve its main goal to improve knowledge sharing in Wikis. However, we are excited concerning the results of our online field studies which are currently ongoing.

Although we believe that our approach is generally viable for larger Wikis on the internet (although e.g. targeting the Wikipedia would require further considerations regarding scalability), our main concern are Wikis in an enterprise setting. Limited resources make it necessary to focus and prioritize knowledge sharing which can be achieved by our approach. Also, Woogle can ease the problem of bootstrapping Enterprise Wikis by presenting search results from external sources.

Further steps regarding the implementation of Woogle are possible in several directions. As described in this paper, the identification of querying users and a mechanism for “requesting” content are already foreseen. Other issues could be the introduction of social ranking mechanisms (i.e. allowing users to annotate/re-rank results) and the extension towards semantic Wiki search [35][20].

7. ACKNOWLEDGEMENT

This work is partially supported by the THESEUS project, which is funded by the German Federal Ministry of Economics (BMWi) under grant 01MQ07019, and the GlobaliSE project, which is funded by the Landesstiftung Baden-Württemberg foundation.

8. REFERENCES

- [1] Ackerman, M.S. and Malone, T.W.: Answer garden: a tool for growing organizational memory. In: Proceedings of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems, New York, NY, USA, ACM (1990) 31–39.
- [2] Ackerman, M.S. and McDonald, D.W. 1996. Answer Garden 2: merging organizational memory with collaborative help. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (Boston, Massachusetts, United States, November 16 - 20, 1996).
- [3] Adamic, L.A., Zhang, J., Bakshy, E., and Ackerman, M.S. 2008. Knowledge sharing and yahoo answers: everyone knows something. In Proceeding of the 17th international Conference on World Wide Web (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, 665-674.
- [4] Alavi, M. and Leidner, D.E.: Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 25(1), 107–136 (2001)
- [5] Ardichvili, A., Page, V., and Wentling, T.: Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management* 7(1), 64–77 (2003)
- [6] Baeza-Yates, R. and Riberio-Neto, B.: *Modern Information Retrieval*. ACM Press, New York (1999)
- [7] Bughin, J. and Manyika, J.: *How businesses are using web 2.0*. McKinsey Quarterly, 03 2007.
- [8] Burghardt, T. and Buchmann, E., Böhm, K., and Clifton, C.: Collaborative Search and User Privacy: How Can They Be Reconciled? In Proceedings of the 4th International Conference on Collaborative Computing (CollaborateCom), Orlando, 2008.
- [9] Braun, S. and Schmidt, A.: Wikis as a technology fostering knowledge maturing: What we can learn from Wikipedia. In Proceedings of the 7th International Conference on Knowledge Management (I-KNOW 2007), 2007.
- [10] Cabrera, A. and Cabrera, E.F.: Knowledge-sharing dilemmas. *Organization Studies* 23, 687–710 (2002)
- [11] Churchill, E. F., Sullivan, J. W., and Snowdon, D. 1999. Collaborative and co-operative information seeking: CSCW'98 workshop report. *SIGGROUP Bull.* 20, 1 (Apr. 1999), 56-59.
- [12] Cress, U. and Hesse, F.W.: Knowledge sharing in groups: experimental findings of how to overcome a social dilemma. In: *ICLS 2004: Proceedings of the 6th international conference on Learning sciences*, pp. 150–157 (2004)
- [13] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In Proceedings of the 12th international Conference on intelligent User interfaces (Honolulu, Hawaii, USA, January 28 - 31, 2007). *IUI '07*. ACM, New York, NY, 32-41.
- [14] Davenport, T. and Prusak, L.: *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1997.
- [15] Dearman, D., Kellar, M., and Truong, K. N. 2008. An examination of daily information needs and sharing opportunities. In Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work (San Diego, CA, USA, November 08 - 12, 2008). *CSCW '08*. ACM, New York, NY, 679-688.
- [16] Economist Intelligence Unit: *Serious business: Web 2.0 goes corporate*, 06 2007.
- [17] Evans, B. M. and Chi, E. H. 2008. Towards a model of understanding social search. In Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work (San Diego, CA, USA, November 08 - 12, 2008). *CSCW '08*. ACM, New York, NY, 485-494.

- [18] Fidel, R., Bruce, H., Pejtersen, A.M., Dumais, S., Grudin, J., and Poltrock, S.: Collaborative Information Retrieval (CIR) (2000) Information Behaviour Research, Vol. 1, 235-247.
- [19] Hansen, M.T.: The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Adm. Sci. Quarterly* 44, 82–111 (1999)
- [20] Happel, H.-J.: Growing the Semantic Web with Inverse Semantic Search. In Proceedings of the 1st Workshop on Incentives for the Semantic Web (INSEMTIVE '08)
- [21] Happel, H.-J.: Closing Information Gaps with Inverse Search. In Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM2008)
- [22] Happel, H.-J.: Towards Need-driven Knowledge Sharing in Distributed Teams. In: Proceedings of the 9th International Conference on Knowledge Management (I-KNOW 2009), 2009 (to appear).
- [23] Happel, H.-J.: Woogle - On Why and How to Marry Wikis with Enterprise Search. In Proceedings of the 2nd Workshop on Integrated Knowledge Management Systems (IKMS2009)
- [24] Desouza, K.C.: Barriers to effective use of knowledge management systems in software engineering. *Commun. ACM* 46(1), 99–101 (2003)
- [25] Desouza, K.C. and Evaristo, J.R.: Managing knowledge in distributed projects. *Commun. ACM* 47(4), 87–91 (2004)
- [26] Konstan, J.A. and Chen, Y.: Online Field Experiments: Lessons from CommunityLab. In Proceedings of e-Social Science 2007, Ann Arbor, Michigan, 2007.
- [27] Maier, R.: Knowledge Management Systems. Springer, Heidelberg (2003)
- [28] Majchrzak, A., Wagner, C., and Yates, D. 2006. Corporate wiki users: results of a survey. In Proceedings of the 2006 international Symposium on Wikis (Odense, Denmark, August 21 - 23, 2006). WikiSym '06. ACM, New York, NY, 99-104.
- [29] Morris, M. R. 2008. A survey of collaborative web search practices. In Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 1657-1660.
- [30] Olson, J. S., Grudin, J., and Horvitz, E. 2005. A study of preferences for sharing and privacy. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 1985-1988.
- [31] Orlikowski, W. J. 1992. Learning from Notes: organizational issues in groupware implementation. In Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work (Toronto, Ontario, Canada, November 01 - 04, 1992). CSCW '92. ACM, New York, NY, 362-369.
- [32] Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., and Pejtersen, A. M. 2003. Information seeking and sharing in design teams. In Proceedings of the 2003 international ACM SIGGROUP Conference on Supporting Group Work (Sanibel Island, Florida, USA, November 09 - 12, 2003). GROUP '03. ACM, New York, NY, 239-247.
- [33] Pickens, J., Golovchinsky, G., and Morris, M.R.: Report on the 1st Collaborative Information Retrieval Workshop Held in Conjunction with the Joint Conference on Digital Libraries (JCDL) 2008. *D-Lib Magazine*. July/August 2008. Volume 14 Number 7/8
- [34] Romberg, T. and Happel, H.-J.: Wikis - Die Wissensmanagement-Lösung für Agile Unternehmen? In: Haasis, Klaus / Zaboura, Nadia (Hrsg): A digital lifestyle: Leben und Arbeiten mit social software.
- [35] Haase, P., Herzig, D., Musen, M.A., and Tran, T.: Semantic Wiki Search. In Proceedings of the 6th European Semantic Web Conference (ESWC 2009): pp. 445-460.
- [36] Wasko, M.M. and Faraj, S.: Why should I share? examining social capital and knowledge contribution in electronic networks. *MIS Quarterly* 29(1), 35–57 (2005)
- [37] Yang, B. and Jeh, G.: Retroactive answering of search queries. In: WWW 2006: Proceedings of the 15th international conference on World Wide Web, pp. 457–466. ACM Press, New York (2006)