Woogle – On Why and How to
Marry Wikis with Enterprise Search

Hans-Jörg Happel

FZI Forschungszentrum Informatik
Karlsruhe, Germany
happel@fzi.de

Abstract: Enterprise Search engines are strong in locating existing documents and information in an organization while Wikis are designed to capture new information in a lightweight and collaborative fashion. Conversely, Wikis are rather bad in locating information (especially from external documents) while Enterprise Search does not address the provision of new information and socializing around information needs. Therefore we argue that both systems focus on specific parts of the organizational information process, which should indeed be combined in order to improve enterprise information exchange. We discuss "Woogle" as a concept to integrate Enterprise Search into Wikis and describe its reference implementation "Woogle4MediaWiki".

1 Introduction

Exchanging information is vital for the success of a modern organization. Several researchers have shown, that direct, "face to face" communication is the most efficient way for information exchange and knowledge transfer in a group or organization [All77, OO00, DP97]. In turn, mediated and/or asynchronous information exchange has proven to be less efficient for a number of reasons such as differences in mental models and contexts of users or a lower communication bandwidth [OO00].

However, distributed work settings are getting increasingly common due to specialization and global collaboration. In distributed collaboration, direct communication is often not feasible due to the limited capacity of communication channels and people [AM96]. Therefore, information technologies play a key role in distributed, asynchronous information exchange.

This paper deals with the differences and limitations of enterprise search engines and Wikis as two popular IT solutions for information access and exchange within organizations. Enterprise Search targets to make available existing information for reference and reuse. On the other hand, Wikis as web-based, collaborative authoring tools are getting increasingly popular to acquire and accumulate new information.

While these two tool categories seem to be disjoint at a first glance, we argue in this paper that Enterprise Search and Wikis are indeed addressing tightly related issues in the organizational information exchange process. Furthermore, we argue that information access and information provision can be significantly improved by considering certain design elements from Wikis in enterprise search solutions and vice versa – i.e. by "marrying" Wikis with Enterprise Search. To this end, we designed Woogle, an extension for the MediaWiki engine which allows embedding enterprise search services within a Wiki environment.

In the following, we shortly discuss Enterprise Search and Wikis and line out some weaknesses. Afterwards we describe the Woogle design principles targeting these weaknesses and present our Woogle reference implementation Woogle4MediaWiki. We also shortly summarize related work and give an outlook to ongoing and future work.

2 Foundations

2.1 Information seeking

We consider search as a communication process between information seekers and information providers. The main steps in this process are
   a) finding suitable information providers
   b) bridging different conceptualizations between information seekers and providers
   c) pulling the sought information

In a face to face setting, step a) is hard, when people do not know whom to ask. However, once they found possible information providers, steps b) and c) can be resolved by interpersonal communication. Notably, it is also possible to exchange tacit information, which has not been explicitly captured before.

In distributed, IT-assisted settings, step a) is well supported, since search systems allow to query information from various providers instantly. However, search systems are weak in bridging different conceptualizations [FL+87] and especially in pulling sought information which is not (yet) formalized.

This is due to the fact, that information providers are usually decoupled from search systems. 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 systems 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 systems 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. For this purpose we will consider enterprise search systems and enterprise Wikis in the remainder of this paper, since they are popular solutions for enterprise information retrieval resp. provision.

2.2 Enterprise Search

Searching for information has become one of the main activities of knowledge workers to such an extent, that the verb "to google" even made it into contemporary dictionaries[1]. Similar to the vast amount of information on the Web, many enterprises harbour a large set of documents and other information which can be useful for its employees.

However, search in an enterprise setting suffers from a number of problems [FK+03]. First, the number of cross-links, which is an important building block of popularity-based ranking algorithms, is typically rather low. Secondly, enterprises often have a heterogeneous information systems infrastructure with sophisticated permission schemes which makes it difficult to retrieve certain information.

Enterprise Search is particularly cumbersome from a collaboration perspective. Recent research has shown that information seeking is a highly interactive process which can benefit significantly from collaboration [EC08, Mor08]. However, if e.g. two employees have a similar information need within the same timeframe, typical search systems do not allow to interact on this information need. Unsatisfied queries are captured in query logs and only considered from time to time, although they could form an interesting basis for user interaction.

---

[1] e.g. http://www.merriam-webster.com/dictionary/google

Furthermore, enterprise search systems neglect the dynamic character of information provision and assume a stable corpus which only occasionally grows when crawling new results that appeared arbitrarily. Users can typically not directly influence search results, even if there are only few or bad results. While the early days of the Web favoured human-maintained "catalogues" which did not scale up [KT00], automated indexing and retrieval approaches largely keep human users out of the loop and can thus become increasingly cluttered and irrelevant – i.e. do not scale down well.

2.3 Enterprise Wikis

Wikis can be considered as a category of next generation, web-based groupware systems which allow for easily capturing and disseminating information in an organization. There has been an increasing adoption rate in recent years [BM07, Eco07]. 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), favouring direct editing of content instead of difficult access permissions and fostering the accumulation/maturing of information by enforcing conceptual integrity between URL, page title and content [HR08].

In enterprises, Wikis are typically used to collect and refine small pieces of unstandardized, immature information [BS07], for which no specific information system exists. Wikis are therefore also a common entry point when searching for information for which no specific 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 [MWY06].

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 enterprise information.

2.4 Wikis and search

In summary, it seems as if (enterprise) search engines and Wikis focus on different, but closely related aspects of information processing. While both systems have a gatekeeper role in accessing organizational information, Wikis have a strong focus on creating information, but neglect efficient retrieval (especially concerning information outside the Wiki) while search engines focus on providing efficient access to existing information but do not address collaboration and information provision at all.

Recent research however reports, that information sharing is a fundamental part of the information seeking process [PG+03]. In a study with 150 participants, Evans and Chi found out that 58.7% of them were sharing information related to their search results with others [EC08]. Therefore it seems as if the decoupling of information seeking and information provision, as realized by the application genres of Wikis and search engines, blocks an important communication channel.

3 Woogle

As lined out in the previous section, combining Wikis with Enterprise Search could form a synergetic combination. In this chapter we introduce some basic design principles for such a system and describe our reference implementation "Woogle4MediaWiki".

3.1 Design principles

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.

In order to realize synergies between Enterprise Search and Wikis, we thus derive the following design principles:

- Improving access

    o **Introduce advanced information retrieval functions** to the Wiki to allow for efficient information access. Most existing Wiki engines provide only limited out of the box search support.

    o **Provide integrated Enterprise Search** across enterprise information artifacts to make all existing explicit information easily accessible. This should help to raise Wiki acceptance and usage by bootstrapping and complementing information stored within the Wiki.

- Improving provision

    o **Seamless transition** from information seeking to information provision in order to provide tight integration and holistic support for user's information behaviour.

    o **Provide different modes of information provision**, such as creating explicit information need descriptions, annotations and adding new information easily and without entry barriers.

- Improving collaboration

    o **Give queries a first order representation** to serve as a common point of reference during information seeking and information provision.

    o **Provide means for communication and awareness** such as discussion and notification mechanisms.

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" [EC08].

3.2 Reference implementation

For implementing our approach, we can either integrate Wiki features into a search user interface or Enterprise Search into a Wiki engine. We decided to go for the latter option, since our approach focuses on user-facing interactions, for which existing Wiki engines offer good support, while enterprise search engines are typically strong concerning backend functionality.

For our reference implementation we thus chose to develop a plug-in for the MediaWiki engine, since it is the most popular Open Source Wiki and since it comes with poor built-in search functionality.

Figure 1: Woogle4MediaWiki screenshot

To realize our approach we created a special "namespace" in MediaWiki, meaning that all pages with an URL-prefix "Woogle:" are processed by our plug-in. 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 pages is displayed, which contains a list of search results at its bottom (c.f. Figure 1). 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:

- **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.

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".

Actual search results in Woogle4MediaWiki can stem from two different sources. As a default, the plugin can be configured to retrieve results from a remote instance of TeamWeaver[2], an Open Source enterprise search framework specifically targeting the software engineering domain. We also implemented native PHP indexing and retrieval functionality using a PHP port of the Lucene library, which can be used as an alternative backend.

4 Related work

Collaboration within the information seeking process has been discussed under the label of "collaborative information retrieval" earlier this decade [CSS99, FB+00] and very recently [PGM08, Mor08]. "Social search" is a related term which is widely used to label various different approaches. Evans and Chi define it as:

---

[2] http://www.teamweaver.org

> *"…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."* [EC08]

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 [EC08], 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 prototype[3], 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 results[4] – a feature which has been recently adopted by Google[5].

Q&A systems, such as Yahoo Answers[6] or Ask.com[7] 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 [AZ+08] 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.

Finally, earlier works on organizational memory systems such as Answer Garden [AM90, AM96] connect information seekers and information 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.

5 Conclusion

In this paper, we argued that Enterprise Search and Wikis should be combined to provide a more holistic coverage of the organizational information process. We therefore presented Woogle as a concept for integrating Enterprise Search into a Wiki and its reference implementation Woogle4MediaWiki.

---

[3] http://community.wikiseek.com
[4] http://search.wikia.com/
[5] http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=115764
[6] http://answers.yahoo.com
[7] http://www.ask.com

Woogle4MediaWiki enables users to search across different information sources – even outside the Wiki itself. Furthermore, search functionality is tightly integrated with the Wiki concepts which allow using collaboration and awareness features such as discussion pages and notifications.

While we think that the Woogle concept can also be beneficial on an Internet scale, we think that it is particularly useful for enterprises, where it can serve as a kind of editable metadata layer on top of a full-text index. However, analyzing the concrete real-world benefits of our approach is subject to ongoing evaluation.

We claim two major benefits of our approach. First, Woogle helps to improve enterprise information provision, by allowing to immediately capturing information in a Wiki-style within the enterprise search environment where it is typically accessed. Users can also directly influence and comment on top the automatically create search index. Second, Woogle can ease the problem of bootstrapping Enterprise Wikis. Those usually suffer from sparse content, which impedes their adoption by members of the organization. By locating the Enterprise Search within the Wiki, people are "lured" into the Wiki which raises the chance of contributing information.

While user requirements discussion and first end user feedback indicates the usefulness of the described approach, we are currently considering several further points of improvement. Natural issues are the introduction of social ranking mechanisms (i.e. allowing users to annotate/re-rank results), leveraging information need descriptions for better retrieval [KF07] and the inclusion of semantic relations [VK+06] among queries as well as between queries and content.

References

[AM90]   Ackerman, M.S., 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.

[AM96]   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).

[All77]   Allen, Thomas J.: Managing the Flow of Technology. MIT Press, 1977.

[AZ+08] 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.

[BM07]   J. Bughin and J. Manyika. How businesses are using web 2.0. McKinsey Quarterly, 03 2007.

[BS07]   S. Braun and A. Schmidt. 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.

[CSS99] 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.

[DP97]   T. Davenport and L. Prusak. Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, 1997.

[Eco07]   Economist Intelligence Unit. Serious business: Web 2.0 goes corporate, 06 2007.

[EC08]   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.

[FB+00]   R. Fidel, H. Bruce, A.M. Pejtersen, S. Dumais, J. Grudin and S. Poltrock, Collaborative Information Retrieval (CIR) (2000) Information Behaviour Research, Vol. 1, 235-247.

[FK+03]   Fagin, R., Kumar, R., McCurley, K. S., Novak, J., Sivakumar, D., Tomlin, J. A., and Williamson, D. P. 2003. Searching the workplace web. In Proceedings of the 12th international Conference on World Wide Web (Budapest, Hungary, May 20 - 24, 2003). WWW '03. ACM, New York, NY, 366-375.

[FL+87]   Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. 1987. The vocabulary problem in human-system communication. Commun. ACM 30, 11 (Nov. 1987), 964-971.

[HR08]   Tim Romberg, Hans-Jörg Happel: Wikis - Die Wissensmanagement-Lösung für Agile Unternehmen? In: Haasis, Klaus / Zaboura, Nadia (Hrsg): A digital lifestyle: Leben und Arbeiten mit social software.

[KF07]   Kelly, D. and Fu, X. 2007. Eliciting better information need descriptions from users of information search systems. Inf. Process. Manage. 43, 1 (Jan. 2007), 30-46.

[KT00]   Kobayashi, M. and Takeda, K. 2000. Information retrieval on the web. ACM Comput. Surv. 32, 2 (Jun. 2000), 144-173.

[Mor08]   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.

[MWY06] 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.

[OO00]   Olson, G. M. and Olson, J. S. 2000. Distance matters. Hum.-Comput. Interact. 15, 2 (Sep. 2000), 139-178.

[PG+03]   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.

[PGM08] Jeremy Pickens, Gene Golovchinsky, Meredith Ringel Morris. 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

[VK+06]   M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer. Semantic wikipedia. In WWW '06: Proceedings of the 15th international conference on World Wide Web, pages 585–594, New York, NY, USA, 2006. ACM.

SHORT PAPERS