

# Considering Information Providers in Social Search

Hans-Jörg Happel and Athanasios Mazarakis  
FZI Research Center for Information Technology  
Karlsruhe, Germany

{happel, mazarakis}@fzi.de

## ABSTRACT

The notions of collaborative information seeking (CIS) and social search have extended the classical model of information seeking and retrieval. In its core, CIS and social search acknowledge the existence of multiple users and study their implicit and explicit interactions across various dimensions. In this paper, we argue to further extend the scope by introducing information providers as a separate role to complement the process of information seeking with information provision. We briefly describe prototype implementations and identify a number of future research challenges.

## 1. INTRODUCTION

The information need of a user is a primary subject of investigation in information retrieval (IR). It can be defined as information which a user requires to complete a specific task [1]. The main purpose of IR systems is to help users satisfying their information needs by providing a set of relevant information – which is typically contained in some kind of document. To use an IR system, the user 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.

Basic elements of this “classic” model of information retrieval are thus the information seeker, her information need, the IR system and the corpus of documents, against which a representation of the information need is matched. While this model is helpful for analyzing many kinds of information retrieval problems, it has turned out that it does not address several aspects of real world information seeking [31].

One missing dimension is *time*. The model described above assumes a synchronous process, in which the information seeker receives results instantly. It returns only those documents which have been analyzed before the time of the query, and can thus be called *retrospective* search [23]. However, in the web as well as in the enterprise, new documents which might be relevant for user queries are created continuously. In order to address this issue, the paradigm of *prospective* search [23][37] has been conceptualized. Prospective search systems allow users to store their queries and notify them, as soon as new results arrive. Popular examples for such systems are Google Alerts [18] or Windows Live Alerts [30].

Another important dimension is *collaboration*. IR systems are used by many different information seekers, whose information needs might be partially identical or overlapping. Implicit relevance ranking [1] is an example for a rather indirect, kind of interaction, which makes users influence the search experience of others. Recently, researchers have coined the terms of “collaborative information seeking” and “social search” to describe additional means for explicit and implicit collaboration throughout the search process [15][16][23].

While prospective search acknowledges the dynamic nature of document creation and social search acknowledges the collaborative nature of the information seeking process, we argue that both do not address how and why new information is fed into the information retrieval process. Retrospective, prospective and social search mainly consider information seekers, their queries and a corpus of documents. In this paper, we thus introduce information providers and their private information space as additional elements in the information retrieval process.

Therefore, we continue by describing our underlying concept of “need-driven knowledge sharing”, which we consider to be complementary to information seeking. Afterwards, we shortly describe some example systems highlighting this concept. We finally line out a number of related research challenges which are also relevant for the core social search community.

## 2. NEED-DRIVEN KNOWLEDGE SHARING

We consider information seeking as a communication process between information seekers and providers [21]. Information seekers typically initiate communication with a verbal representation of their information need (“query”). Information providers return “answers” based on their interpretation of these queries. While in face-to-face settings, a providers’ contribution can be directly tailored towards the request of an information seeker, this is typically not the case in distributed, asynchronous settings, where information providers are decoupled from information seekers and their requests.

Therefore, mediation technologies play an important role in distributed information seeking. In the following, we differentiate between mediation services and mediation spaces. Mediation services can be as simple as a plain communication channel between both actors or involve more complex functionality such as awareness or notification. Mediation spaces consist of persistent information that is not relevant for satisfying an information need as such, but that captures supplementary information for mediation purposes such as clarification discourses, conceptual mappings or descriptive metadata like tags. An example for a mediation space is a newsgroup where people can reference and discuss information needs.

Mediated information seeking approaches have several advantages such as satisfying very specific information needs which are probably difficult to satisfy in collocated settings, allowing for asynchronous information seeking and scaling up to many information seekers by making answers available in a persistent fashion [21].

On the other hand, especially asynchronous information seeking decouples information needs from actual information

provision, which has a number of problematic side-effects. First, it removes most context from the information seeking process and makes it thus difficult to algorithmically bridge different conceptualizations between information needs and content satisfying these needs.

Second, and probably most important, information providers do not receive any more guidance which kind of information should be shared within the information system. While this may turn out less problematic in a web-scale environment, where most information needs will be satisfied by *someone* creating the content, it is a definite issue in enterprise settings, where many information needs are enterprise-specific and resources for information sharing are scarce.

The probably most successful means of distributed information seeking is what we call “document and content sharing” [21] which subsumes web search engines and enterprise search approaches. Typically, the number of mediation services and spaces is rather low in these systems. Information seekers are very limited in expressing their information needs (i.e. using keyword queries) and information providers are not addressed at all, resulting in completely decoupled roles.

We are thus investigating mechanisms to alleviate for the lack of “need-drivenness” in document and content sharing by establishing feedback channels between information seekers and providers. As a major source for information needs, we propose making queries and query logs first order citizens in distributed information seeking.

### 3. PROTOTYPES<sup>1</sup>

In the following, we shortly present two different approaches, which describe concrete realizations for some novel mediation services and spaces in a document and content sharing scenario.

#### 3.1 Woogle – Marrying Wikis with Enterprise Search

Woogle is a prototype system which we developed to combine the advantages of Wikis and Enterprise Search [20][22]. Both are popular IT solutions for information access 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 both systems have a gatekeeper role in accessing organizational information, Wikis have a strong focus on creating information, but neglect efficient retrieval (especially information outside the Wiki) while search engines focus on providing efficient access to existing information, but until recently typically do not address collaboration and information provision at all.

In our combination approach, the Wiki provides the user interface for searching within enterprise information (including the Wiki itself). Each user query is optionally assigned a special Wiki page, which allows persisting the information need (see

Figure 1). Thus, a mediation space is created, where users can further elaborate and discuss about information needs. Also, the built-in awareness and collaboration features such as user notifications can be easily applied to the search process.

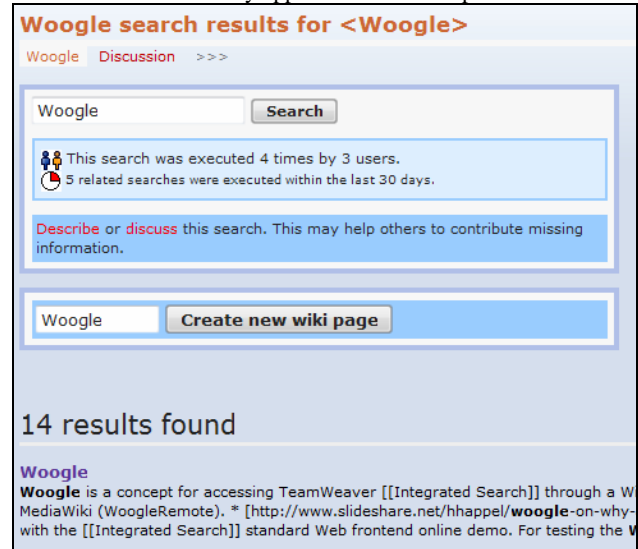


Figure 1: Information need indicators in the search UI

An additional mediation service which addresses information providers is depicted in Figure 2. While Wikis can signal “desired pages” by rendering them in red color, we added an additional pop-up which supports need-driven knowledge sharing by providing meta-information about the actual information needs related to a particular “red link”.

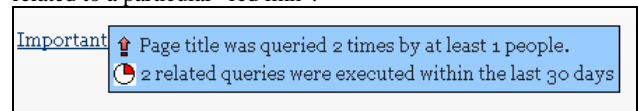


Figure 2: Woogle mouse-over display for “red links”

Woogle thus focuses on providing a mediation space (i.e. Wiki pages for each information need) which are complemented by additional mediation services leveraging these explicit need representations.

#### 3.2 Inverse Search – Diffusing Private Artifacts to Shared Information Spaces

Our second prototype addresses the phenomenon of “hoarding” information in private spaces – especially files on a local desktop resp. in private folders. Typically, such information is not accessible for other users and there is no way for information providers to easily find out which documents would be worth sharing, although these documents could help satisfying other’s information needs.

In order to tackle this issue, we designed Inverse Search [19] to support information providers in sharing documents from their private information space. Basis for our approach is a mediation service which allows information providers to retrieve aggregated unsatisfied information needs from an (enterprise) search engine. Information providers can then match their private documents against the given set of information needs to get a ranked list of documents worth sharing (Figure 3). Inverse Search thus works in

<sup>1</sup> Both tools described in this section are available for download at [www.teamweaver.org](http://www.teamweaver.org) under an Open Source license. For Woogle, there is also an online demonstration available.



meaning – or concerning alternative need expressions such as “red links” in Wikis [22].

#### 4.4 Privacy

While social search approaches leverage social relations among people and maintain extensive data about users’ behavior, most of them do not address privacy issues at all [9]. Experience from our initial interview studies supports that users are very interested in a social search experience – such as e.g. a global aggregation or even individual display of information needs. However, while users are interested in other’s information needs, they hesitate when it comes to revealing their own needs to others.

Since other sub-disciplines of information retrieval such as query log analysis have been addressing privacy concerns in a broad fashion recently [3][27], we think that it is a fundamental concern which should be also addressed in the context of social search more offensively.

#### 4.5 Means of mediation

In order to get users of social search systems involved in collaborative activities, supporting mediation services and spaces need to be in place.

In order to act as a reference point for collaboration, several approaches such as S<sup>3</sup> [32], Google SearchWiki [17] or Woogole [22] try to give information needs a persistent representation. Up to now, this representation is mostly a direct mapping of the keyword-based query string. However, direct keyword mappings are known to suffer from homonymy – i.e. similar strings represent different intentions – and synonymy – i.e. different strings represent similar intentions. We thus believe that research should think about issues such as means for representing persistent information needs or notions of query similarity.

We also believe that social search requires additional mediation services – e.g. for the analysis and aggregation of information needs. Information retrieval has so far developed basic notions to distinguish e.g. navigational and informational searches [7]. However, means for addressing collaborative aspects such as aggregated information needs [19] that could help prioritizing needs are still in an early stage of research.

#### 4.6 User interfaces

Many of the aspects discussed before call for new features in current user interfaces. This rises questions both concerning completely new search paradigms and interfaces and concerning interfaces and interactions combining social search with existing, “classic” search approaches.

While highly automated, algorithmic approaches to search – such as realized by AltaVista and Google – succeeded over human-maintained catalogues – such as the initial Yahoo site – in the early days of the web, small user contributions to Web 2.0 systems such as delicious<sup>4</sup> or Yahoo Answers<sup>5</sup> are beginning the change the game. Ultimately, “socially enhanced” versions of highly automated search engines such as realized by Google SearchWiki [17] might become the best solution.

---

<sup>4</sup> <http://delicious.com/>

<sup>5</sup> <http://answers.yahoo.com/>

## 5. CONCLUSION

Collaborative information seeking and social search address important “blind spots” of the classic information retrieval model by considering collaborative activities among fellow information seekers. In this paper we argued, that social search should also develop a more refined understanding of contributions to the social search process, in particular the contribution of content satisfying the information needs of information seekers.

While we consider many kinds of social search contributions and underlying issues like privacy or motivation highly relevant also on a web scale, we think that enterprise search applications might benefit the most. To this end, we briefly introduced two prototypes – Woogole and Inverse Search – and highlighted a number of open research challenges.

## 6. 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.

## 7. REFERENCES

- [1] Agichtein, E., Brill, E., and Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. In Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Seattle, Washington, USA, August 06 - 11, 2006). SIGIR '06. ACM, New York, NY, 19-26.
- [2] Ames, M. and Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 971-980.
- [3] Amitay, E. and Broder, A. (2008). Introduction to special issue on query log analysis: Technology and ethics. ACM Trans. Web 2, 4 (Oct. 2008), 1-2.
- [4] Angiolillo, J. (2003). Search Log Analysis as a Usability Engineering Tool. In: CHI'03 Workshop on Best practices and future visions for search user interfaces (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003).
- [5] Baeza-Yates, R. and Riberio-Neto, B. (1999). Modern Information Retrieval. ACM Press, 1999.
- [6] Bernstein, M., Marcus, A., Karger, D. and Miller, R. (2009). Understanding and Supporting Directed Content Sharing on the Web. Technical Report, MIT-CSAIL-TR-2009-048, 10/2009.
- [7] Broder, A. (2002). A taxonomy of web search. SIGIR Forum 36, 2 (Sep. 2002), 3-10.
- [8] Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In Proceedings of the 2005 international ACM Conference on Supporting Group Work (Sanibel Island, Florida, USA, November 06 - 09, 2005). GROUP '05. ACM, New York, NY, 1-10.
- [9] Burghardt, T., Buchmann, E., Böhm, K. and Clifton, C. (2008). Collaborative Search and User Privacy: How Can They Be Reconciled? CollaborateCom 2008: 85-99

- [10] Cheshire, C. (2007). Selective Incentives and Generalized Information Exchange. *Social Psychology Quarterly*, 70, 82-100.
- [11] Cheshire, C., & Antin, J. (2008). The Social Psychological Effects of Feedback on the Production of Internet Information Pools. *Journal of Computer-Mediated Communication*, 13, 705-727.
- [12] Costa, P.T., Jr. and McCrae, R.R. (1985). The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources.
- [13] 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.
- [14] Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.
- [15] 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.
- [16] Golovchinsky, G., Qvarfordt, P., and Pickens, J. (2009). Collaborative Information Seeking. *Computer* 42, 3 (Mar. 2009), 47-51.
- [17] Google Inc. Google SearchWiki. (2009). <http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=115764>, 11 2009.
- [18] Google Inc. Google Alerts. (2010) <http://www.google.com/alerts>, 01 2010.
- [19] Happel, H.-J. (2008). Closing Information Gaps with Inverse Search. In *Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM2008)*
- [20] Happel, H.-J. (2009). Woogle - On Why and How to Marry Wikis with Enterprise Search. In *Proceedings of the 2nd Workshop on Integrated Knowledge Management Systems (IKMS2009)*
- [21] Happel, H.-J. (2009). Towards Need-driven Knowledge Sharing in Distributed Teams. In *Proceedings of the 9th International Conference on Knowledge Management (I-KNOW 2009)*
- [22] Happel, H.-J. (2009). Social search and need-driven knowledge sharing in Wikis with Woogle. In *Proceedings of the 5th international Symposium on Wikis and Open Collaboration (Orlando, Florida, October 25 - 27, 2009)*. WikiSym '09. ACM, New York, NY, 1-10.
- [23] Hearst, M.A. (2009). *Search User Interfaces*. Cambridge University Press. 2009.
- [24] Irmak, U., Mihaylov, S., Suel, T., Ganguly, S. and Izmailov, R. (2006). Efficient query subscription processing for prospective search engines. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 1037-1038, New York, NY, USA, 2006. ACM Press.
- [25] Joyce, E., and Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11, 723-747.
- [26] 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.
- [27] Korolova, A., Kenthapadi, K., Mishra, N., and Ntoulas, A. (2009). Releasing search queries and clicks privately. In *Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009)*. WWW '09. ACM, New York, NY, 171-180.
- [28] Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4).
- [29] 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.
- [30] Microsoft Inc. Windows Live Alerts (2010). <http://alerts.live.com/> 01 2010
- [31] 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.
- [32] Morris, M.R. and Horvitz, E. (2007). S3: Storable, Shareable Search. *Interact 2007*. 120-123.
- [33] Rosenfeld, L. and Hurst, M. (2010). Search Analytics. Rosenfeld Media (to appear) [http://www.rosenfeldmedia.com/books/searchanalytics/blog/changing\\_your\\_content\\_creation](http://www.rosenfeldmedia.com/books/searchanalytics/blog/changing_your_content_creation)
- [34] Sponder, M. (2004). Search Engine Effectiveness, Metrics and Scorecarding. Report. [www.searchengineworkshops.com/SearchEngineMetrics.pdf](http://www.searchengineworkshops.com/SearchEngineMetrics.pdf)
- [35] Stamous, S. and Efthimiadis, E.N. (2009). Queries without Clicks: Successful or Failed Searches? In: *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*. Boston, MA, USA. July 23, 2009.
- [36] Teevan, J., Dumais, S. T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Salvador, Brazil, August 15 - 19, 2005)*. SIGIR '05. ACM, New York, NY, 449-456.
- [37] Yang, B. and Jeh, G. (2006) Retroactive answering of search queries. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 457-466, New York, NY, USA, 2006. ACM Press.
- [38] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 425-478.