Chapter 20
Finding Information Faster by Tracing My Colleagues’ Trails: A Reference Algorithm for Enterprise Search

Patrick Winter
University of Marburg, Germany

Michael Schulz
University of Marburg, Germany

Tobias H. Engler
University of Marburg, Germany

ABSTRACT
Knowledge workers are confronted with the challenge of efficient information retrieval in enterprises, which is one of the most important barriers to knowledge reuse. This problem has been intensified in recent years by several organizational developments such as increasing data volume and number of data sources. In this chapter, a reference algorithm for enterprise search is developed that integrates aspects from personalized, social, collaborative, and dynamic search to consider the different natures and requirements of enterprise and web search. Because of the modular structure of the algorithm, it can easily be adapted by enterprises to their specificities by concretization. The components that can be configured during the adaptation process are discussed. Furthermore, the performance of a typical instance of the algorithm is investigated through a laboratory experiment. This instance is found to outperform rather traditional approaches to enterprise search.

INTRODUCTION
The ability of an enterprise to integrate and reuse the sometimes highly specialized knowledge of its employees has been identified as a major chance for gaining competitive advantages (Grant, 1996). Therefore, many enterprises have established internal repositories to support the explication of this

DOI: 10.4018/978-1-5225-0559-4.ch020
knowledge, its storage, its transfer, and, eventually, its reuse (Alavi & Leidner, 2001). However, only few employees utilize the knowledge stored in these repositories (Davenport et al., 2003; Desouza, 2003). In search for an explanation, difficulties in finding suitable documents efficiently have been identified as the major barrier to knowledge reuse (Davenport & Prusak, 2000), leading to high search costs and opportunity costs for enterprises (Feldman & Sherman, 2003). Functionalities for information retrieval within an enterprise (enterprise search) have long been undeveloped and highly inefficient (Hawking, 2004). In addition, recent changes in the organizational environment such as the availability of more data and data sources (McAfee & Brynjolfsson, 2012), the greater number of employees working with data (Hänel & Schulz, 2014), and the democratization of information in the enterprise (Li et al., 2014) further emphasize the need for new enterprise search functionalities that can help to overcome this barrier in addition to alternative approaches (e.g., improved information storage, (Schulz et al., 2015)). This can also be seen by the fact that three out of six constituting technology characteristics of Enterprise 2.0 (search, links, and tags) directly relate to enterprise search (McAfee, 2006).

Early search algorithms have mostly relied on a simple pattern matching between the search query and a document’s content. Later, search engines have improved this approach by incorporating the link structure to rank the relevance of web content (Page et al., 1999). In the last years, four streams have emerged that each address one of the disadvantages of this approach: Personalized and social search taking into account the querying user’s personal characteristics and social relationships (resp.) to adjust the ranking of the results, collaborative search aiming to exploit the information provided by historic search sessions (by potentially other users), and dynamic search considering search sessions which consist of multiple search queries. However, most of the algorithms originating from these streams were designed especially for web search. Attempts have been made to transfer such algorithms to enterprise search, but it soon has been recognized that this is hard to accomplish (O’Leary, 1997) given the different nature of these domains (e.g., the strongly differing numbers of potential users and results, no organic link structure on the intranet, etc. (McAfee, 2006)). The few algorithms that have especially been designed for enterprise search (e.g., Ronen et al., 2009) have two important limitations:

First, while some of them combine more than one of the four search streams described above, the unique chance to do so in the domain of enterprise search is often overlooked. In the environment of an enterprise search engine (ESE), users can easily be identified by their account, so that their activities can be tracked and logged across various systems. This enables a special form of collaborative search: For a search session, it can be predicted using historical information on previous search sessions:

1. Whether it will be successful,
2. Which document the querying user is likely to search for, and
3. How she will refine her search queries (integrating dynamic search).

The available information can be weighted by the strength of her similarity and relationship with the previous querying users (integrating personalized and social search).

Second, many existing enterprise search algorithms were designed specifically for the enterprise in which they are deployed. They often are proprietary and, therefore, kept a secret. Even if their code is disclosed they still cannot simply be deployed in other enterprises if the latter exhibit different characteristics (e.g., use a different file structure). Thus, elaborate transfer processes are necessary to adapt the algorithm.