

# People Search in the Enterprise

Krisztian Balog  
ISLA, University of Amsterdam  
Kruislaan 403, 1098 SJ Amsterdam  
kbalog@science.uva.nl

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.4 [Information Systems Applications]: H.4.2 Types of Systems; H.4.m Miscellaneous

## General Terms

Algorithms, Measurement, Performance, Experimentation

## Keywords

Enterprise search, People search, Expertise finding

The large increase in recent years in the amount of information available online has led to a renewed interest in a broad range of IR-related areas that go beyond standard document retrieval. Some of this new attention has fallen on *entity retrieval*. This emerging area of entity retrieval differs from traditional document retrieval in a number of ways. Entities are not represented directly (as retrievable units such as documents), and we need to identify them “indirectly” through occurrences in documents. This brings new, exciting challenges to the information retrieval and extraction fields.

In the proposed research, I focus on one particular type of entity: *people*. The need for people search has been recognized by many commercial systems, who offer facilities for finding individuals or properties of individuals. These include locating classmates and old friends, finding partners for date and romance, white and yellow pages, etc. My interest is different, and focuses on “professional” or “work-related” people search applications. I propose two information access tasks, both within an enterprise (or organizational) setting: (i) people finding, which is concerned with the retrieval of individuals that meet some criteria, and (ii) people profiling, which is about characterizing a specific person. Both tasks are explored along two main axes: topical and social.

In an enterprise setting, a key criterion by which people are selected and characterized is their level of expertise with respect to some topic. The concept of “being an expert” is not defined explicitly, instead, it is assumed that people closely associated with a topic are experts on that area. As a result, the main challenge in topical search is to infer the association between a person and an expertise area from the supporting document collection. *Topical expert finding* involves the task of identifying people with the appropriate skills and knowledge: Who are the experts on topic X?

The task of *topical expert profiling* seeks to answer a related question: What topics does person Y know about? A topical profile of a person is a record of the types and areas of skills of that individual together with an identification of levels of proficiency in each. People finding and profiling together form a complete picture, and in an operational people finder system users alternate between these two types of search [4].

In 2005 and 2006 the TREC Enterprise track introduced the W3C collection and provided a common platform for researchers to empirically assess methods and techniques devised for expert finding. In my research I developed models for topical expert finding and profiling, and made extensive experiments using the W3C collection [1, 3, 4]. While nearly all of the expert search work performed has been validated experimentally using the TREC platform, it only represents one type of intranet. In [2], we focus on expertise retrieval in an intranet that differs from the W3C setting.

So far, most of my research has been on topical search. What I would like to do next is to complement this work by bringing in more social aspects. In the *social search* task we look beyond individuals and are interested in exploring the connections and social relations between them. Who is related to X? What is the nature of their connection? The task can be decomposed into two steps. The first step is to discover the connections between people, while the second step is about characterizing their relation.

There is a number of research questions to be addressed: How to model these people search tasks? How to represent topics, documents, and candidates? What is the appropriate level of granularity? How to represent and make use of the structure that may be available: internal and external document structure, topic categorization, organizational hierarchy? How to build document-people associations? How does the quality of associations affect the overall performance? Which is more important: quality or quantity?

To answer these research questions, I have proposed, and use, a probabilistic retrieval framework based on language modeling techniques [1, 2]. We collect evidence from multiple sources, and integrate it with a restricted information extraction task — the language modeling setting allows us to do this in a transparent manner, and provides a particularly convenient and natural way of modeling the tasks we consider.

## References

- [1] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings SIGIR '06*, 2006.
- [2] K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch. Broad expertise retrieval in sparse data environments. In *Proceedings SIGIR '07*, 2007.
- [3] K. Balog and M. de Rijke. Finding experts and their details in e-mail corpora. In *Proceedings WWW-2006*, 2006.
- [4] K. Balog and M. de Rijke. Determining expert profiles (with an application to expert finding). In *Proceedings IJCAI-2007*, 2007.