

Enhanced Algorithms for Enterprise Expert Search System

Valentin Molokanov, Dmitry Romanov and Valentin Tsibulsky
National Research University "Higher School of Economics",
Moscow, Russia

ABSTRACT

We present the results of our enterprise expert search system application to the task introduced at the Text Retrieval Conference (TREC) in 2007. The expert search system is based on analysis of content and communications topology in an enterprise information space. An optimal set of weighting coefficients for three query-candidate associating algorithms is selected for achieving the best search efficiency on the search collection. The obtained performance proved to be better than at most TREC participants. The hypothesis of additional efficiency improvement by means of query classification is proposed.

Keywords: TREC, expert search, enterprise information management

1. INTRODUCTION

In large organizations information repositories are characterized by an extreme heterogeneity of structured as well as unstructured information in a vast amount of variously formatted documents. Indeed, different reports, meeting protocols, descriptions of working groups, projects, etc. are valuable sources which can be found via some usual search functionality. Despite of that, a user has to spend too much time in viewing such documents in order to find employees who are competent at a given theme at the present time. The task of finding people with concrete professional experience arises unavoidably in the need of asking anything in some professional area as well as in performing a series of other more difficult tasks; among them are, for example, finding all members of a specified project or finding all employees that are working with a specified customer. In similar scenarios using an enterprise expert search system is more advantageous in comparison with a simple search engine, as the user can find the appropriate people much faster. An expert search system delivers a list of people who might have knowledge and be useful as experts at a given topic. So an expert search system can be an effective means of organization management in the purposes of improving performance and collaboration quality by presenting information about the employees who possess knowledge in requested areas.

We continue our research of expert ranking algorithms using the TREC Enterprise track 2007 corpus. The corpus represents the crawl of the open-access information from the official site of the Commonwealth Scientific and Industrial Research Organization (CSIRO) [1]. Although in 2009 TREC Enterprise track was replaced by Entity track to change a research direction towards finding arbitrary entities in web data, the Enterprise track collections remained accessible, so we decided to use the 2007 corpus to optimize our expert search system performance.

The main problem for any automated expert search system is to associate a query with people. To identify such associations, various techniques are proposed. Expert search methods in modern enterprise systems are rather different, so there is no conventional expert search model for enterprise systems. However, most of them could be classified into two principally different approach types: document-based and candidate-based.

The document-based approach became the first acceptable approach to expert search. It imitates expert search process with the use of an ordinary search system. Here, the primary retrieval of relevant documents and the following people search in such documents are implied. The approach is just referred to as a two-stage model and is described in details in [2]. It became widely used in its several variants at TREC 2007 [3]. Another quite natural approach to expert search is a candidate-based one. It supposes building a special description (so-called profile) for each candidate, after that candidate ranking is produced with the help of simple search technologies. With the help of various methods candidates' profiles are filled with their expertise information. The examples of such methods are presented in [4], [5], [6], [7]. Despite the variety of interpretations of query-candidate connections in these two approaches, TREC 2008 results showed that the mentioned approaches have no advantages over each other, and the best expert search efficiency at TREC was achieved by regarding structured information in documents [8] and intranet structure [9], as well as by handling additional information from beyond the collection [10].