ExperTime: Tracking Expertise over Time

Jan Rybak
Norwegian University of Science and Technology
jan.rybak@idi.ntnu.no

Krisztian Balog
University of Stavanger
krisztian.balog@uis.no

Kjetil Nørvåg
Norwegian University of Science and Technology
kjetil.norvag@idi.ntnu.no

ABSTRACT

This paper presents ExperTime, a web-based system for tracking expertise over time. We visualize a person's expertise profile on a timeline, where we detect and characterize changes in the focus or topics of expertise. It is possible to zoom in on a given time period in order to examine the underlying data that is used as supporting evidence. It is also possible to perform visual and quantitative comparison of two arbitrarily selected time periods in a highly interactive environment. We invite profile owners to evaluate and fine-tune their profiles, and to leave feedback.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval

Keywords

Temporal expertise profiling, expert profiling, expertise retrieval

1. INTRODUCTION

Over the past decade, the increased availability of online evidence that can be traced and mined has led to a renewed interest in two types of expertise retrieval system: *expert finding* systems (to locate individuals who are experts on a given topic) and *expert profiling* systems (to describe the expertise of a given person) [1]. In practice, the two tasks—finding and profiling—are strongly connected; once potential expert candidates are identified, a closer look into their profile is needed before they are actually contacted.

Our focus lies in two important issues related to expertise profiling that have received little attention so far. First, the expertise of a person on a given topic should be considered in context; that context is comprised of the larger landscape of knowledge areas within the domain and of the distribution of the person's expertise across. Second, expertise is not static, it changes over time. In scientific fields that are rapidly developing, such as computer science, "once an expert always an expert" does not hold. Therefore, it is vital to be able to track changes in a person's expertise over time.

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.
ACM 978-1-4503-2257-7/14/07.
http://dx.doi.org/10.1145/2600428.2611190.

We build on prior work [3] and propose to use domain-specific topic taxonomies that enable us to capture and conveniently visualize a person's expertise at any given point in time. By displaying a series of profile snapshots on a timeline, we can have a complete overview of the development of expertise over time. In addition, we identify and characterize important changes that occur in these profiles. Profile owners are invited to refine the profiles we automatically generated for them and are encouraged to leave feedback.

The system is available at `http://bit.ly/expertime`. A short video tutorial introducing the main functionality can be found at `http://bit.ly/expertime-video`.

2. RELATED WORK

Publications are key indicators of expertise in an academic context. Bibliographic databases, like DBLP, have been used to mine information networks, including clustering, ranking and profiling of conferences and authors based on research sub-fields [5, 6]. Bibliographic collections have also been used for expert finding [2, 4]. ArnetMiner [6], a tool for automatic extraction of information about researchers, addresses, among other things, the problem of analyzing expertise evolution over time. Their solution, however, is based on keyword statistics projected on a time axis and lacks higher organizational structure.

3. DATA AND METHODS

In this section we briefly introduce the temporal expertise profiling method underlying our system. For a full account we refer the reader to [3]. We model the expertise of an individual as a series of timestamped hierarchical profiles (referred to as *profile snapshots* from now on). We assume that a (domain-specific) taxonomy of expertise areas is provided and weigh nodes in this tree-shaped hierarchy based on documents supporting evidence of expertise. This representation not only allows us to measure how broad, narrow, or diverse one's expertise is, but also facilitates the formal comparison and characterization of changes between two profile snapshots. Changes are divided into two main groups, depending on which level of the topic hierarchy is affected: changes in the field of research (i.e., top-level nodes) and changes in the topics of research (i.e., lower-level nodes). Our method has two main parameters: (1) the *decay factor* controls the rate at which expertise "fades away," and (2) the *focus threshold* sets the sensitivity of the method to changes. We allow the user to tune these values (see §4).

In this work, we focus on the computer science domain, for two reasons. One, we are reasonably familiar with the domain, and, two, there are publicly available sources of publications and topic classification systems in this area. Specifically, we use DBLP as our bibliographic database and the ACM 1998 Computing Classification System as our taxonomy. We use the provided ACM classi-

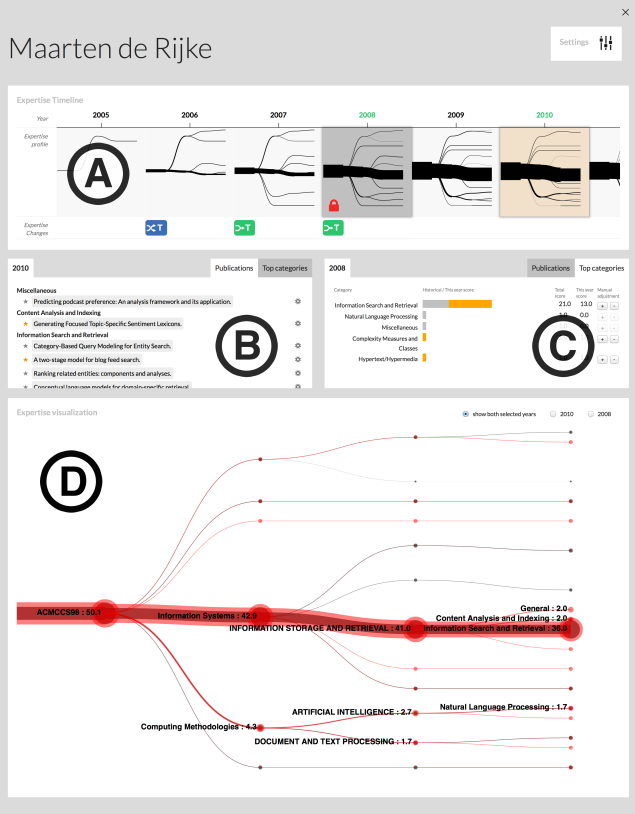

Figure 1: An example expertise profile in ExperTime.

fication, where available, and use an automatic classifier for the rest of the publications. (We offer our users the possibility to refine this categorization.) Note that our the system is not tailored to any specific domain; topic taxonomies are available for other disciplines as well (e.g., PubMed in medicine), which makes our system easily extensible to other fields.

4. FUNCTIONALITY

A number of researcher profiles have already been processed and made available. Those who are not yet in our system can sign up and claim their profiles. Figure 1 displays an example profile. We distinguish between two user modes: (1) the user is only a *viewer* of a profile, and is not allowed to make changes, and (2) the user is the *owner* of the profile. In the latter case, the user is allowed (and encouraged) to change the settings and to adjust her expertise profile to what she believes is best reflecting reality. A key design objective was that such changes shall take effect immediately, i.e., we update the underlying model when the parameter settings (such as the decay factor and focus threshold) or the data (assignments of publications to categories and weighting of publications and/or categories) are modified. When displaying a profile, we indicate whether it is an automatically generated one or one that has already been curated by its owner. Next, we briefly discuss the main components of the interface, indicated with circled letters on Figure 1.

(A) Timeline We create yearly profile snaphots and display them on a timeline. Each year is expressed as a tree (see §3). By clicking on the snapshots, the user can select one or two years for more details. The timeline can be understood as the main controlling instrument—the page is updated upon the selection of years. Detected changes in one's expertise are displayed below the timeline as small icons.

(B) Supporting documents We present a list of publications, categorized according to the topic taxonomy. Users have the option to change the category assignments for each paper if they feel that it is not classified correctly. It is also possible to promote publications that are deemed more important than others by giving them a "star."

(C) Top categories We show the list of top-ranked categories from the topic taxonomy, along with the estimated importance of each category (subdivided into past period and current year). Similarly to publications, categories can also be promoted if the user believes that her contribution is not projected correctly in the category's score.

(D) Detailed profile snapshot(s) The largest area is occupied by the visualization of a profile snaphot—a tree of categories with branches weighted according to the person's expertise on the corresponding topics. It is also possible to have two years selected for comparison on the timeline (A), in which case the corresponding profile snaphots are visualized by two semitransparent overlaying tree graphs (so that the user can more easily identify differences).

5. IMPLEMENTATION

The system consists of two main building blocks. The back-end part, using mostly Java and MySQL technologies, is responsible for the extraction, processing, and storage of information. The front-end is an HTML5 website using servlets for communication with the server and JavaScript with additional libraries, such as jQuery and D3.js, for visualization and interactivity.

6. CONCLUSIONS

In this paper, we have described a demonstration of a temporal expertise profiling system, called ExperTime. We presented our initial motivation, used methods and underlying data, as well as the main functionality of the system. The next step on our research agenda is the evaluation of temporal expertise profiles. This is a challenging problem for two reasons: (1) the notions of focus and topic changes are subjective and are likely to vary from person to person, and (2) the complexity of the task is beyond the point where TREC-like benchmark evaluations are feasible. The feedback we plan to obtain with the ExperTime system, both implicit and explicit, will provide invaluable information to guide the development of appropriate evaluation methodology.

7. REFERENCES

[1] K. Balog, Y. Fang, M. de Rijke, P. Serdyukov, and L. Si. Expertise retrieval. *Found. Trends Inf. Retr.*, 6(2-3):127–256, 2012.
[2] H. Deng, I. King, and M. R. Lyu. Formal models for expert finding on dblp bibliography data. In *Proceedings of ICDM'08*, 2008.
[3] J. Rybak, K. Balog, and K. Nørvåg. Temporal expertise profiling. In *Proceedings of ECIR'14*, 2014.
[4] E. Smirnova and K. Balog. A user-oriented model for expert finding. In *Proceedings of ECIR'11*, 2011.
[5] Y. Sun, T. Wu, Z. Yin, H. Cheng, J. Han, X. Yin, and P. Zhao. BibNetMiner: Mining bibliographic information networks. In *Proceedings of SIGMOD'08*, 2008.
[6] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of SIGKDD'08*, 2008.