Contextual Factors for Finding Similar Experts (Extended Abstract)*

Katja Hofmann†, Krisztian Balog†, Toine Bogers‡ and Maarten de Rijke†
†ISLA, University of Amsterdam and ‡ILK, Tilburg University
K.Hofmann@uva.nl, K.Balog@uva.nl, A.M.Bogers@uvt.nl, deRijke@uva.nl

ABSTRACT

Expertise seeking research studies how people search for expertise and choose whom to contact in the context of a specific task. An important outcome are models that identify factors that influence expert finding. Expertise retrieval addresses the same problem, expert finding, but from a system-centered perspective. The main focus has been on developing content-based algorithms similar to document search. These algorithms identify matching experts primarily on the basis of the textual content of documents experts are associated with. Other factors, such as the ones identified by expertise seeking models, are rarely taken into account.

In this paper we extend content-based expert finding approaches with contextual factors that have been found to influence human expert finding. We focus on a task of science communicators in a knowledge-intensive environment, the task of *finding similar experts*, given an example expert. Our approach combines expertise seeking and retrieval research. First, we conduct a user study to identify contextual factors that may play a role in the studied task and environment. Then we design expert retrieval models to capture these factors. We combine these with content-based retrieval models and evaluate them in a retrieval experiment.

Our main finding is that, while content-based features are the most important, human subjects also take contextual factors into account, for example media experience and organizational structure. We develop two principled ways of modeling the identified factors and integrate them with content-based retrieval models. Our experiments show that models combining content-based and contextual factors can significantly outperform existing content-based models.

1. INTRODUCTION

The increasing amount of information available is making the need to critically asses information more important. The burden of credibility assessment and quality control is partly shifting onto individual information seekers, but the need for information intermediaries—e.g., experts—has not disappeared and is actually increasing in cases where the credibility of information has to meet high standards [8]. Against this background, *expert finding* is a particularly relevant task: identifying and selecting individuals with specific expertise, for example to help with a task or solve a problem.

The goal of *expertise retrieval* is to support search for experts using information retrieval technology. Following the experimental paradigm and evaluation framework established in the information retrieval community, expertise retrieval has been addressed in world-wide evaluation efforts [9]. Promising results have been achieved, particularly in the form of algorithms and test collections [1, 2]. State-of-the-art retrieval algorithms model experts on the basis of the documents they are associated with, and retrieve experts on a given topic using methods based on document retrieval, such as language modeling [3, 4]. In evaluations of these algorithms user aspects have been abstracted away.

While research into expertise retrieval has primarily focused on identifying good topical matches between needs for expertise and the content of documents associated with candidate experts, behavioral studies of human *expertise seeking* have found that there may be important additional factors that influence how people locate and select experts [11]—such factors include accessibility, reliability, physical proximity, and up-to-dateness. We term these *contextual factors* to distinguish them from content-based factors that have been explored in previous work.

2. RESEARCH QUESTIONS AND METHOD

Our aim in this paper is to explore the integration of contextual factors into content-based retrieval algorithms for finding similar experts. We look at this problem in the setting of the public relations department of a university, where communication advisors employed by the university get requests for topical experts from the media. The specific problem we are addressing is this: the top expert identified by a communication advisor in response to a request is not available because of meetings, vacations, sabbaticals, or other reasons. In this case, communication advisors have to recommend similar experts and this is the setting for our expert finding task. Based on this task we address three main research questions:

1. Which contextual factors influence (human) decisions when finding similar experts in the university setting we study?
2. How can such factors be integrated into content-based algorithms for finding similar experts?
3. Can integrating contextual factors with existing, content-based approaches improve retrieval performance?

To answer our research questions, we proceed as follows. Through a set of questionnaires completed by a university's communication advisors, we identify contextual factors that play a role in how similar experts are identified in this situation, and we construct a test data set to evaluate retrieval performance. From the questionnaire, we identify contextual factors that play a role in the studied setting.

Based on the questionnaire results, we develop models of contextual factors, and integrate these with existing, content-based retrieval methods. We explored modeling factors as input-dependent, similar to content-based similarity methods, and as input-independent,

*The full version of this paper has appeared in [6].

DIR2011, February 4, 2011, Amsterdam.

similar to a prior probability. The intuition between the first model is that candidates with similar characteristics to the given target expert would be likely to be recommended. The intuition behind the second model is that there may be certain characteristics that make a candidate to be more likely to be recommended, independent of the target expert. In our experiments, we evaluate our contextual retrieval models against a baseline consisting of the optimal combination of content-based retrieval models.

3. FINDINGS

Our first goal was to identify contextual factors that play a role in the task of finding similar experts in response to media requests for expertise. We find that, while *topic of knowledge* appears to be the most important factor in the studied setting, contextual factors play a role as well, such as *position*and *contacts*. In addition to these factors that had been identified in previous studies, we were able to identify two new factors that played a role, namely *organizational structure* and *media experience*.

The individual contextual factors that appear to have the most impact are *media experience*, *organizational structure*, and *position*. This finding suggests that there may be a strong task-specific component to the contextual factors that play a role in finding similar experts, and possibly in other retrieval tasks as well. In future work, it would be interesting to perform similar studies of contextual factors in information seeking tasks in other settings. Based on findings from several such studies it may be possible to develop more general models of how tasks relate to other factors, and how these relations influence people's relevance decisions.

Our second research question was how to model contextual factors and integrate them with existing retrieval models. To this end, we explored modeling factors as input-dependent, similar to content-based similarity methods, and as input-independent, similar to a prior probability. We found that both types of models improved upon the baseline using content-based factors only. Overall, input-independent models led to better performance, except for the input-dependent model of *organizational structure*. Thus, the studied setting, it is important that a candidate expert is part of the same department as the topic expert, but in addition to that there are attributes that are common to frequently recommended experts, such as having prior media experience, or being a professor. Best performance was achieved with a run that combined both types of models. These results show that both types of models are useful and that it is not enough to identify a factor, but that it also needs to be modeled appropriately.

The third question was whether integrating contextual factors with content-based retrieval methods would improve retrieval performance. Our results show that our models that include contextual factors indeed achieve significant improvements over the content-based baseline methods.

Overall, our results indicate that identifying contextual factors and integrating them with content-based expertise retrieval models is indeed a promising research direction. The method used for collecting data on contextual factors is an extension of normal relevance assessment and could be applied in other settings where the original topic creators are available for relevance assessment.

4. CONCLUSION

In this paper we started from the observation that contextual factors appear to play a role in expertise seeking. We explored the role of contextual factors in the task of finding similar experts. First, we identified contextual factors that play a role in the task of finding similar experts in the public relations department of a university. The identified factors were modeled in two principled ways and implemented using available data. We integrated the resulting models with existing, content-based models and evaluated them to assess retrieval performance. Our results demonstrate that it is possible to identify and model contextual factors in the studied task of finding similar experts, and we think that this may be the case for other retrieval tasks as well.

In information seeking research, models of how contextual factors play a role have been developed and it has been shown that information seeking behavior changes with, for example, specifics of the task [5, 7] and the user's problem stage [10]. From an information retrieval perspective, these contextual factors are difficult to model and researchers typically design experiments where they abstract from context to make results generalizable. In this paper we have argued that, in order to arrive at generalizable results, we need to model context and develop models of how contextual factors influence expertise seeking. We have shown that the factors can be modeled, that it is possible to integrate them with retrieval models, and that the resulting models can improve retrieval performance.

Acknowledgements We are extremely grateful to the communication advisors at Tilburg University for their participation in our study: Clemens van Diek, Linda Jansen, Hanneke Sas, Pieter Siebers, Michelle te Veldhuis, and Ilja Verouden. We wish to thank Lilian Woudstra for her extensive feedback and Joost Kircz for discussions and feedback.

This research was supported by the Netherlands Organization for Scientific Research (NWO) under project numbers 017.001.190, 640.001.501, 640.002.501, 612.066.512, 612.061.814, 612.061.-815, 640.004.802, by the Dutch and Flemish Governments (http://www.stevin-tst.org) under projects DAESO (STE-05-24) and DuOMAn (STE-09-12) and by the IOP-MMI program of SenterNovem / The Dutch Ministry of Economic Affairs, as part of the À Propos project.

5. REFERENCES

[1] P. Bailey, N. Craswell, I. Soboroff, and A. P. de Vries. The CSIRO enterprise search test collection. *SIGIR Forum*, 41(2): 42–45, 2007.

[2] K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, June 2008.

[3] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR '06*, pages 43–50. ACM Press, 2006.

[4] K. Balog, L. Azzopardi, and M. de Rijke. A language modeling framework for expert finding. *Information Processing and Management*, 45(1):1–19, 2009.

[5] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191–213, 1995.

[6] K. Hofmann, K. Balog, T. Bogers, and M. de Rijke. Contextual factors for finding similar experts. *JASIS&T*, 61(5):994–1014, 2010.

[7] J. Kim. Describing and predicting information-seeking behavior on the web. *JASIS&T*, 60(4):679–693, 2009.

[8] M. Metzger. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *JASIS&T*, 58(13):2078–2091, 2007.

[9] TREC. Enterprise track, 2005. URL: http://www.ins.cwi.nl/projects/trec-ent/wiki/.

[10] P. Vakkari. Changes in search tactics and relevance judgements when preparing a research proposal a summary of the findings of a longitudinal study. *Information Retrieval*, 4(3-4):295–310, 2001.

[11] L. S. E. Woudstra and B. J. Van den Hooff. Inside the source selection process: Selection criteria for human information sources. *Information Processing and Management*, 44:1267–1278, 2008.