

Collecting Expertise of Researchers for Finding Relevant Experts in a Peer-Review Setting

Delroy Cameron, Boanerges Aleman-Meza, I. Budak Arpinar

LSDIS Lab

Computer Science Department

University of Georgia

Athens, GA 30602-7404

{cameron, boanerg, budak}@cs.uga.edu

Abstract

We present ideas for determining the expertise of researchers across various areas of computer science and for finding relevant experts/reviewers in a peer-review setting. We explain how Semantic Web techniques for data collection and data representation using ontologies can be used in addressing this specific “ExpertFinder” problem.

1. Introduction

The task of assigning reviewers for scientific papers in a peer-review setting is quite demanding on the person that performs such a task; usually the conference or workshop chair(s). Existing conference management systems, such as Confious¹ and OpenConf² facilitate this task by using a variety of methods. However, in spite of their successes, a more challenging problem (indirectly related to paper assignment) is that of putting together the Program Committee (PC) of reviewers. PC members must possess the necessary and relevant expertise to review papers in the Conference (or Workshop). In many cases, the selection of PC members is based on the conference chair’s (and/or conference organizers’) knowledge of experts in the field. Quite often, previous interaction and/or collaboration with such experts, suffices for composing a qualified review committee. However, due to an increasing number of emerging communities and diversification of research areas, it is likely that many experts are unknown to the conference chair and hence may be overlooked.

The problem is then to find experts in a seamless fashion, pre-empting having previous knowledge or interaction with them. Our approach to this problem consists of two aspects. First, it is necessary to know the topics of expertise of a given researcher. Second, given a

list of topics, determine the relevant experts. Addressing these two aspects involves in many cases non-exact matches of expertise. For example, a researcher with expertise on “Semantic Web Processes” might be a good match for a conference on “Web Services.” Hence, the use of semantics is a promising way of finding expertise, by relying on ontologies to match topics of expertise. The collection of data for and representation of expertise are aspects directly related to the ExpertFinder Initiative.³

2. Collecting Expertise

The approach that we envision to this end builds upon our recent work on a large populated ontology of researchers in computer science called SwetoDblp [1], created mostly from data of DBLP.⁴ The aim is to relate researchers listed in such ontology to various topics they might have expertise on. In our preliminary work, we collected the expertise of a subset of researchers who have published papers in World Wide Web and Semantic Web Conferences. This dataset includes 1,200+ researchers and 1,504 relationships to topics (about 100 unique topics). We anticipate that an extensive taxonomy of expertise, similar to that created in [10], will aid in extrapolation of expertise, particularly for cases involving non-exact matches. Similarly, we have done preliminary work on creating a taxonomy of the 100 topics in our dataset. We have found this to be a laborious and time consuming task that led us to conclude that it is quite difficult to achieve the creation of a taxonomy for all research topics appearing in DBLP. In fact, our dataset was quite small, consisting of 2.5% of all researchers appearing in DBLP. We believe that the construction of taxonomies of topics is a key research challenge towards making the ExpertFinder vision become real.

Even at small scale, our dataset of researchers and their topics of expertise has demonstrated applicability in

¹ <http://www.confious.com/>

² <http://www.zakongroup.com/technology/openconf.shtml>

³ <http://www.rdfweb.org/topic/ExpertFinder>

⁴ <http://dblp.uni-trier.de/>

a recent application. A live demo of semantic annotation on the ISWC-2006 website shows how we used this dataset to indicate the expertise of various researchers. For example, the snippet in Figure 1 shows that Dr. Kunal Verma's expertise includes "Semantic Web Services." In the same way, Dr. Manfred Hauswirth's includes "P2P Systems" (not shown). Further details on such semantic annotation demo and the datasets used are available⁵ such as the dataset of topics of expertise.⁶



Figure 1. Researcher Expertise Profile from a Semantic Annotation Demo

Outside of analytics on topic taxonomies, we plan to consider other approaches to estimate expertise. For example, there exists information in conference series from DBLP that could indicate that authors in such conferences have expertise in given topics. For example, authors of papers in Semantic Web Conferences⁷ have expertise on the topic "Semantic Web." A similar approach to [7] could be adopted to compute expertise atoms for researchers across different topics. Additional metrics such as number of publications, publication impact and publication history could be taken into account to provide more complete expertise profiles. Of course, the integrity of expertise profiles largely depends of the nature and quality of the data. Some data integration issues might need to be addressed.

The extensive efforts frequently required for building semantic web applications should not be in vain. Thus, one of our objectives is to make publicly available the datasets created in our efforts. We believe that making the dataset that relates researchers listed in the SwetoDblp ontology to topics of expertise publicly available is a step towards support and participation in the ExpertFinder Initiative.

3. Finding Experts for Peer-Review Assignments

There exists previous work for determining peer reviewers (e.g., [9]) but the issues that we aim to address are in respect to large scale applicability and automation or semi-automation of user-centric duties. This is a common problem with expert finder systems in general [6, 7, 10]. Most ExpertFinder systems are based on highly localized, privatized and specialized datasets, beneficial only in small settings. By facilitating the task of finding suitable reviewers, we anticipate that the quality of an overall conference could improve, since both the number of reviewers available for consideration would be larger and the extent of their expertise would be determined and used in the selection process. Additionally, as in [5], the use of Semantic Web techniques creates computer-interpretable data, limiting the extent of manual user input. This provides a new dimension for existing peer-review systems (e.g., [8]) that rely on extensive user input. Further, modeling a researchers' expertise can prove important in recognizing and analyzing collaboration networks within clusters of research communities. We anticipate that recommendations for inclusion on PC lists could be affected by the growth or lack thereof within such clusters.

Our previous work on detecting conflict of interest [2] (between reviewers and authors of papers) considered data both from DBLP and FOAF. Such work focused on relationships among reviewers and authors but it did not consider the various issues involved with reviewer selection and paper assignments. We feel that these components are critical for a holistic assessment of the peer-review process. FOAF data, for example was considered for finding relationships among persons but not for persons and their particular interests. We have seen that the 'interests' relationships in FOAF has been used in a number of applications, for example to match music preferences of people [3] to enrich user profiles. Thus, we suspect that expertise information can be drawn from a number of disparate data sources, including FOAF to augment existing expertise profiles. Work in [4] for example, develops an architecture for crawling and indexing data from diverse data sources across the web, enabling querying of semantic content. Such techniques can then be used for augmenting expertise profiles.

4. Expert Finder Evaluation

The evaluation of techniques for finding experts is not straightforward. However, data of Program Committee members from previous years could be used to observe the extent of concurrence and/or disparity with computer-

⁵ <http://lsdis.cs.uga.edu/projects/semdis/iswcdemo2006/>
⁶ <http://cs.uga.edu/~cameron/expertise.html>
⁷ <http://www.informatik.uni-trier.de/~ley/db/conf/semweb/index.html>

based techniques. Of course, this raises issues once again of the integrity/quality of a dataset. For example, data recently collected from DBLP would indicate skewed expertise information because potential PC candidates would have more published material since last serving on a previous Program Committee. Similarly, new PC members would have emerged through published research. For example, ICDE conferences have a large number of researchers on its program committee, which includes new members every year. To address these issues we make two observations. First, we note that one of the benefits of adding expertise data to existing ontologies such as SwetoDblp is that further details can be provided when results of potential reviewers are listed. For example, the relevant publication titles and/or publication venues could be provided to a PC Chair who is trying to determine whether or not to invite a researcher for the PC of a conference. Second, we are afforded an opportunity to perform expertise analytics on PC members over several conferences by observing expertise growth of seasoned researchers in particular domains.

5. Conclusions

Finding both expertise and experts is a topic of importance in practical applications. In industrial settings it is particularly important because there are significant economic implications involved with locating and employing the most qualified experts in a project. In academia, it is also important to facilitate the tasks involved in peer-review. In this paper, we described our preliminary efforts and ideas for collection of expertise. We also discussed some of the benefits and challenges involved. We described the importance of finding PC members for a conference and listed possible ways for evaluating computer-based methods by using on data of PC members in past conferences. We believe that techniques based on semantic technologies will prove useful in ExpertFinder applications.

References

- [1] Aleman-Meza, B., Hakimpour, F., Arpinar, I.B., Sheth, A.P.: SwetoDblp Ontology of Computer Science Publications (submitted for publication, <http://lsdis.cs.uga.edu/projects/semdis/swetodblp/SwetoDblp-AHAS06.pdf>)
- [2] Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.P., Arpinar, I.B., Joshi, A., Finin, T.: Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection, 15th International World Wide Web Conference (WWW2006), Edinburgh, Scotland, UK, May 2006
- [3] Celma, O.: Foafing the Music: Bridging the Semantic Gap in Music Recommendation, 5th International Semantic Web Conference, (ISWC2006) Athens, Georgia, USA, November 5-9, 2006
- [4] Harth, A., Umbrich, J. and Decker, S.: MultiCrawler: A Pipelined Architecture for Crawling and Indexing Semantic Web Data, 5th International Semantic Web Conference Athens, GA, USA, November 5-9, 2006
- [5] Kraines, S., Guo, W., Kemper, B. and Nakamura Y.: EKOSS: A Knowledge-user Centered Approach to Knowledge Sharing, Discovery and Integration on the Semantic Web, 5th International Semantic Web Conference Athens, GA, USA, November 2006
- [6] Liu, P. and Dew, P.: Using Semantic Web Technologies to Improve Expertise Matching within Academia, Proceedings of I-KNOW, Graz, Austria, June 2004
- [7] Mockus, A., Herbsleb, J.D.: Expertise Browser: A Quantitative Approach to Identifying Expertise, ICSE (2002), Orlando Florida, USA, May 2002
- [8] Papagelis, M., Plexousakis, D. and Nikolaou, P.N.: CONFIOUS*: Managing the Electronic Submission and Reviewing Process of Scientific Conferences, 6th International Conference on Web Information Systems Engineering, New York, NY, USA, 2005.
- [9] Rodriguez, M.A. and Bollen, J.: An Algorithm to Determine Peer-Reviewers, (submitted), LA-UR-06-2261, December 2005 <http://www.cse.ucsc.edu/~okram/papers/referee-identification.pdf>
- [10] Song, X., Tseng, B.L., Lin, C.-Y. and Sun, M.-T.: ExpertiseNet: Relational and Evolutionary Expert Modeling, 10th International Conference on User Modeling, Edinburgh, Scotland, UK, July 2005