COMPARING PEOPLE IN THE ENTERPRISE

Gianluca Demartini
L3S Research Center, Leibniz Universität Hannover, Appelstrasse 9a, D-30167 Hannover, Germany

Keywords: Enterprise Search, Similarity measure, Semantic technologies.

Abstract: Enterprise Search Systems are requested to provide more and more functionalities for supporting decision at the management level. An important aspect to consider is the human power and the knowledge which is available. For this reason, in this paper, after extracting a list of requirements out of a specific scenario, and presenting the previous work, we describe an improved approach to compare experts in order to retrieve and present to the user the most appropriate candidates for a given project.

1 INTRODUCTION

The knowledge workers are starting to be the most important competitive advantage that an enterprise can have in the age of information. For this reason, it is every day more important that Enterprise Search (ES) enables an easy managing of human resources. In order to do an efficient planning and delegation, a manager must be able to know all the abilities and skills of her employees and also be able to make comparisons among them for an easier decision making process. The main contribution of this paper is an improved approach to compare experts in order to retrieve and present to the user the most appropriate people for a given project. After presenting the scenario, we describe a previously proposed set of similarity measures and show how they do not meet simple requirements. Then, we describe our semantic based approach on expert search.

In the previous work, some first steps the direction of defining formal models for comparing people and retrieving the most expert ones have been made in the context of Expert Search: probabilistic models (Fang and Zhai, 2007) and language models (Azzopardi et al., 2006; Balog et al., 2006; Balog and de Rijke, 2006a) have been proposed. Our work continues this line of work, and shows how to model expert search leveraging the hierarchical structure of expertise topics. Another model for expert search proposed in (Macdonald and Ounis, 2006) views expert search as a voting problem. The documents associated to a candidate are viewed as votes for this candidate's expertise. Again, the relationships between candidates and documents are only binary and not continuous. In (Macdonald and Ounis, 2007b) the same authors extended the model including relevance feedback techniques, which is an orthogonal issue. A interesting distinction has been made between *expert finding* and *expert profiling* in (Balog and de Rijke, 2006b). The former approach aims at first retrieving the documents relevant to the query and then extract the experts from them. The latter first builds a profile for each candidate and then matches the query with the profiles without considering the documents anymore (Balog and de Rijke, 2007). When an expert profile for each enterprise employee is built is possible to make comparisons among them.

The rest of the paper is structured as follows. In the next section we present the specific scenario of finding experts in the enterprise for which our techniques have been designed. In section 3 we extract out of the described scenario a list of requirements that we aim to address with the proposed solution. In section 4 we present a previous attempt to define a similarity measure among people and how we improve on that. In section 5 we describe how we can improve expert search effectiveness using semantic technologies. Finally, in section 6 we conclude the paper outlining some possible future work.

2 A MOTIVATIONAL SCENARIO FOR EXPERT SEARCH IN ENTERPRISE

In the rest of the paper we focus on one specific scenario. We consider the situation of a human resource manager that has to deal with employees in a big enterprise. Out of the many tasks, she has to hire new

Demartini G. (2008).
COMPARING PEOPLE IN THE ENTERPRISE.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - AIDSS*, pages 455-458
DOI: 10.5220/0001688404550458

employees for filling certain positions described by a profile, she has to decide who to promote to an higher position, and so on. The knowledge available to the manager, for making decisions, is the skill profile of each person she has to deal with.

The manager would highly benefit from a system supporting her in these tasks. Here we list some possible tasks that can be solved using a systems that provides the user with a list of people ranked according their expertise on the query topic.

- Find the most suited candidate for a position
- Build a new project team
- Find someone for solving a problem
- Identify qualification gaps in the enterprise

The most important tool for the manager and for the system to solve these tasks is a *similarity measure* being able to compare people, thus allowing to create a ranking out of the set of possible candidates and their profiles. In the next section we state which are the most important requirements of such comparison measure, for solving the tasks presented in this specific scenario.

3 REQUIREMENTS FOR A SIMILARITY MEASURE

When we want to compare people according to their expertise, we need a special type of similarity measure. The goal here is not to present a comprehensive list of requirements, but to highlight the most important aspects for the scenario considered in this paper. Therefore, we focus on the following important requirements:

Assume Continuous Scores of Expertise. When we want to effectively compare people according to their skills, a binary measure of expertise (e.g., Mr.X is/isn't expert on "Ontology engineering") is not enough. We need a score which has value in $[0,1]$ for each topic.

Deal with Topic Ambiguity. The comparison of people should consider that there are ambiguous topics of expertise (e.g., "Bank") and, therefore, it should not make the mistake of considering one employee more skilled than another, if the topic is not the same.

Leverage on the Hierarchical Nature of Expertise. The topics of expertise are, obviously, more or less specific: some of them include others when they are very general (e.g., "Computer Science" is more general than "Programming Languages"). A similarity measure should not make the error of considering the topics as a flat list of fields where people can be skilled but it should exploit this taxonomy for performing better comparison.

In the following we describe a set of similarity measures, proposed in the past, underlining their weak points. After, we propose our solution, taking into account the requirements extracted so far, also using semantic technologies and natural language processing.

4 EXISTING SKILL-PROFILE SIMILARITY MEASURES

After explaining the motivations and related work in the context of expert search, in this section we critically analyse a previous work which proposed several similarity measures for skill-profile matching (Biesalski and Abecker, 2006). The authors describes the module of an ESS and the four types of similarity measures used:

Direct Skill Comparison. An exact match between skills of people and those needed for a certain position;

Proportional Similarity. It identifies partially fulfilled requirements;

Compensatory Similarity. It considers also overqualifications to compensate partially fulfilled requirements;

Taxonomic Similarity. It uses an expertise taxonomy to find close matches between skills;

The first criticism to these measures is that they use a four level scale of expertise. While this is surly better than a binary distinction between expert / non-expert, it is not enough to model the continuous aspect of expertise. That is, a skill or expertise could be better identified, for example, by a real number $e \in [0,1]$.

The first three measures do not take into account a string similarity score between skill names assuming that they comes from the same dictionary. More, the *Compensatory similarity* assumes value 1 if the expertise level required is the same of the one of the candidate, and values greater than (less than) 1 if it is lower (higher). This results in having a similarity measure which, differently from the others, does not assume values in $[0,1]$ making them not interchangeable.

The most advanced technique is the *Taxonomic similarity* which leverages relations among topics. This is also related with what we propose in the section 5.1 where we suggest that the expertise topics are not orthogonal as assumed for now in the Information Retrieval community. One straightforward extension to this could be to use a similarity based on the lowest common ancestor between two nodes (Harel and Tarjan, 1984) given a suffix tree (Ukkonen, 1995) based on the specificity of topics.

In conclusion, we want to stress the point of having non-binary and bounded similarity measures which would also enable an easier comparison of ESSs for a faster decision making process.

5 USING ONTOLOGIES INSTEAD OF VECTOR SPACE FOR MODELING EXPERTISE

In the scenario of expert profiling that we are tackling, there are several ways to improve the retrieval effectiveness using different evidences. One of such ways is the use of semantics. As done for the web context (Demartini, 2007), annotations can help to identify the correct articles to consider for expertise extraction, knowledge taxonomies can help in finding the correct experts, and ontologies can help in disambiguating multi senses topics.

5.1 Using Ontologies as Expertise Taxonomies

The expert finding task is usually performed in enterprises where the significant knowledge areas are limited. For this reason the expert finding system usually adopt customized and manually built taxonomies to model the organization's most important knowledge areas (Becerra-Fernandez, 2006).

In days where the big enterprises cover several markets, the expertise areas are much more wide than in the past. For this reason finding expert in the enterprise will require much more effort to manually develop a universal expertise taxonomy. We propose to use the Yago ontology (Suchanek et al., 2007), that is, a combination of notions from WordNet[1] and Wikipedia[2], to model the expertise and to identify the knowledge areas used to describe people's knowledge. In this way we can better define the expert profiles according to Yago. For example, knowing that "Macintosh computer" *is a subclass of* "Computers" can help the system when there are no results for the query "Find an expert on Computer". The system can proceed looking for experts in the relative subcategories. More, if we know that "Eclipse" *is a* "Java tool" we can assume that an expert on Eclipse will be an expert (with score proportional to the number of children of the class "Java tool") on Java tools.

[1] `http://wordnet.princeton.edu/`

[2] `http://wikipedia.org/`

5.2 Using Wordnet to Disambiguate Expertise Topics

In the enterprise context there is one more problem to take into account: the topic ambiguity. Multi sense terms might represent topics of expertise. For example, an expert on "Bank" might be expert on only one of the several senses of this noun: slope/incline | financial institution/organization | ridge | array | reserve | ...[3]

Using, for example, the algorithm JIGSAW (Semeraro et al., 2007) for word sense disambiguation we can disambiguate between different topics of expertise. JIGSAW calculates the similarity between each candidate meaning for an ambiguous word and all the meanings in its context defined as words with the same POS tag in the same sentence. The similarity is calculated as inversely proportional to path length between concepts in the WordNet IS-A hierarchy. The assumption in this case is that the appropriate meaning belongs to a similar/same concept as words in the context belong to. For example, if the sentence "John Doe manages the Citizen Bank that has good availability of cash." is an evidence of the expertise on the topic "Bank", we can disambiguate its sense using the context and, in this case, the meaning of "cash". The distance between all the meanings of "Bank" and all the meanings of the nouns in the context (defined as a window of text surrounding the term) can be used in order to find the intended sense. We can then add the sense "financial institution" to the expertise profile of the candidate "John Doe".

It is also possible to use co-occurrence statistics to improve the quality of the profiles. If we take a user profile we can disambiguate the topics looking at the context in the related articles. For example, according to the profile, the user is an expert on "Jaguar" and we find that in the articles considered in his profile the word "Car" often co-occur with the word "Jaguar". In this way we add the topic "Car" to the expertises of the user always with the final goal of disambiguation.

When performing profile extension or relevance feedback, we should anyway pay attention to cases of expertise drift where a candidate "can have several or many unrelated areas of expertise" as shown in (Macdonald and Ounis, 2007a).

[3] from WordNet 3.0

6 CONCLUSIONS

In this paper we presented possible improvements on similarity measures between employees, and how we can improve the effectiveness of expert search tasks adopting semantic technologies. As future steps, we will deploy the designed techniques in a real-world enterprise scenario in order to assess the effectiveness of our methodology.

ACKNOWLEDGEMENTS

This work was supported by the Nepomuk and the Okkam projects funded by the European Commission under the 6th Framework Programme (IST Contract No. 027705) and the 7th Framework Programme (IST Grant Agreement No. 215032) respectively.

REFERENCES

Azzopardi, L., Balog, K., and de Rijke, M. (2006). Language modeling approaches for enterprise tasks. *The Fourteenth Text REtrieval Conference (TREC 2005)*.

Balog, K., Azzopardi, L., and de Rijke, M. (2006). Formal models for expert finding in enterprise corpora. *Proceedings of the 29th SIGIR conference*, pages 43–50.

Balog, K. and de Rijke, M. (2006a). Finding experts and their Details in e-mail corpora. *Proceedings of the 15th international conference on World Wide Web*, pages 1035–1036.

Balog, K. and de Rijke, M. (2006b). Searching for people in the personal work space. *International Workshop on Intelligent Information Access (IIIA-2006)*.

Balog, K. and de Rijke, M. (2007). Determining Expert Profiles (With an Application to Expert Finding). *Proceedings of IJCAI-2007*, pages 2657–2662.

Becerra-Fernandez, I. (2006). Searching for experts on the Web: A review of contemporary expertise locator systems. *ACM Transactions on Internet Technology (TOIT)*, 6(4):333–355.

Biesalski, E. and Abecker, A. (2006). Similarity measures for skill-profile matching in enterprise knowledge management. In *ICEIS (2)*, pages 11–16.

Demartini, G. (2007). Finding experts using wikipedia. In *Finding Experts on the Web with Semantics (FEWS) Workshop at ISWC*, pages 33–41.

Fang, H. and Zhai, C. (2007). Probabilistic Models for Expert Finding. *Proceedings of 29th European Conference on Information Retrieval (ECIR'07)*, pages 418–430.

Harel, D. and Tarjan, R. (1984). Fast algorithms for finding nearest common ancestors. *SIAM Journal on Computing*, 13(2):338–355.

Macdonald, C. and Ounis, I. (2006). Voting for Candidates: Adapting Data Fusion Techniques for an Expert Search Task. *Proceedings of the 15th ACM Conference on Information and Knowledge Management (CIKM'06)*, pages 387–396.

Macdonald, C. and Ounis, I. (2007a). Expertise drift and query expansion in expert search. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 341–350, New York, NY, USA. ACM.

Macdonald, C. and Ounis, I. (2007b). Using Relevance Feedback in Expert Search. *Proceedings of 29th European Conference on Information Retrieval (ECIR'07)*, pages 431–443.

Semeraro, G., Degemmis, M., Lops, P., and Basile, P. (2007). Combining learning and word sense disambiguation for intelligent user profiling. *Twentieth International Joint Conference on Artificial Intelligence*.

Suchanek, F., Kasneci, G., and Weikum, G. (2007). Yago: a core of semantic knowledge. *Proceedings of the 16th international conference on World Wide Web*, pages 697–706.

Ukkonen, E. (1995). On-line construction of suffix trees. *Algorithmica*, 14(3):249–260.