Data-driven Techniques for Expert Finding

Veselka Boeva[1], Milena Angelova[1] and Elena Tsiporkova[2]
[1]Department of Computer Systems and Technologies, Technical University of Sofia, Branch Plovdiv, Bulgaria
[2]Data Innovation Group, Sirris, The Collective Center for the Belgian Technological Industry, Brussels, Belgium

Keywords: Data Mining, Expert Finding, Health Science, Knowledge Management.

Abstract: In this work, we propose enhanced data-driven techniques that optimize expert representation and identify subject experts via automated analysis of the available online information. We use a weighting method to assess the levels of expertise of an expert to the domain-specific topics. An expert profile is presented by a description of the topics in which the person is an expert plus the relative levels (weights) of knowledge or experience he/she has in the different topics. In this context, we define a way to estimate the expertise similarity between experts. Then the experts finding task is viewed as a list completion task and techniques that return similar experts to ones provided by the user are considered. The proposed techniques are tested and evaluated on data extracted from PubMed repository.

1 INTRODUCTION

Nowadays, organizations search for new employees not only relying on their internal information sources, but they also use data available on the Internet to locate required experts. As the data available is very dispersed and of distributed nature, a need appears to support this process using IT-based solutions, *e.g.*, information extraction and retrieval systems, especially expert finding systems (expert seeker). Expert finding systems however, need a lot of information support. On one hand, the specification of required "expertise need" is replete with qualitative and quantitative parameters. On the other hand, the expert seekers need to know whether a person who meets the specified criteria exists, how extensive her/his knowledge or experience is, whether there are other persons who have the similar competence, how he/she compares with others in the field, etc. Consequently, techniques that gather and make such information accessible are needed. In this work, we are particularly interested in developing enhanced data-driven techniques that optimize expert representation and identify subject experts via automated analysis of the available online information.

Many scientists who work on the expertise retrieval problem distinguish two information retrieval tasks: *expert finding* and *expert profiling*, where *expert finding* is the task of finding experts given a topic describing the required expertise (Craswell et al., 2006), and *expert profiling* is the task of returning a list of topics that a person is knowledgeable about (Balog et al., 2007) (Balog, 2008). A method that can easily be apply to both expertise retrieval tasks has been proposed in (Boeva et al., 2012). It is concerned with the question of how to quantify how well the area of expertise of an individual expert or a group of experts conforms to a certain subject within the framework where the domain experts are represented by unified profiles. The concept of expertise spheres has been introduced and it has been shown how the subject in question can be compared with the expertise profile of an individual expert and her/his sphere of expertise. The latter ideas can further be exploited by applying enhanced techniques for optimizing expert representation in order to improve the accuracy of matching an expert with the other experts or the subject in query.

In this work a weighting method that assesses the levels of expertise of an individual expert to the domain-specific topics/keywords is further used for constructing richer expert profiles. The aim is to present an expert profile by a concise description of the topics in which he/she is an expert plus an evaluation of the level of knowledge or experience he/she has about the different topics. An important issue in this context is to establish a way to estimate the expertise similarity between experts.

Boeva V., Angelova M. and Tsiporkova E.
Data-driven Techniques for Expert Finding.
DOI: 10.5220/0006195105350542
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 535-542
ISBN: 978-989-758-220-2

Similarity is a fundamental concept in theories of knowledge and behavior. Psychological experiments have shown that similarity acts as an organizing principle by which individuals classify objects, and make generalizations (Goldstone, 1994). In the context of expertise retrieval, on one hand the system can be expected to support the classification of a newly extracted expert based upon the knowledge or expertise that he/she shares with subject categories of known experts.On the other hand, the finding similar experts task can be viewed as a list completion task (Mattox et al., 1999), *i.e.* the user is supposed to provide a small number of example experts and the system has to return similar experts. This scenario would be useful, for example, when given a small number of individuals, the system can help in recruiting additional members with similar expertise. For example, in case of large-scale emergency and crisis situations (floods, earthquakes, etc.) it is often required to find urgently a high number of additional experts to the already involved individuals in order to adequately respond to the current scale of the disaster. Another possible application is the task of recruiting reviewers, *e.g.*, for reviewing conference, journal or project submissions. For instance, in order to select the most suitable researchers who will be eventually involved in the reviewing process of a journal article, the Editor-in-chief may provide a small number of researchers who have been used to review similar articles in the past, and the system returns a list of scientists with similar knowledge or experience.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 introduces the developed data-driven techniques for expert profile construction and expert identification. Section 4 presents the initial evaluation of the developed techniques, which are applied and tested on data extracted from PubMed repository. Section 5 is devoted to conclusions and future work.

2 RELATED WORK

A variety of practical scenarios of organizational situations that lead to expert seeking have been extensively presented in the literature, *e.g.*, see (Cohen et al., 1998), (Kanfer et al., 1997), (Kautz et al., 1996), (Mattox et al., 1999), (McDonald et al., 1998), (Vivacqua, 1999). Several commercial and free tools that automate the discovery of experts have also become available, *e.g.*, see (Foner, 1997), (Vivacqua, 1999), (Kautz et al., 1997).Web-based expert seeking tools that support both type players (applicants and recruiters) at the job market have recently appeared (Majio) (Yagajobs).

Expert finders are usually integrated into organizational information systems, such as knowledge management systems, recommender systems, and computer supported collaborative work systems, to support collaborations on complex tasks. Initial approaches propose tools that rely on people to self-assess their skills against a predefined set of keywords, and often employ heuristics generated manually based on current working practice (Seid et al., 2000). Later approaches try to find expertise in specific types of documents, such as e-mails (Campbell at al., 2003), (D'Amore, 2004) or source code (Mockus et al., 2002). Instead of focusing only on specific document types systems that index and mine published intranet documents as sources of expertise evidence are discussed in (Hawking, 2004). In the recent years, research on identifying experts from online data sources has been gradually gaining interest (Tsiporkova et al., 2011), (Singh et al., 2013), (Hristoskova et al., 2013), (Abramowicz et al., 2011), (Bozzon et al., 2013).

3 METHODS

3.1 Construction of Expert Profiles

An expert profile may be quite complex and can, for example, be associated with information that includes: e-mail address, affiliation, a list of publications, co-authors, but it may also include or be associated with:educational and (or) employment history, the list of LinkedIn contacts etc. All this information could be separated into two parts: expert's personaldata and information that describes the competence area of expert.

The expert's personal data can be used to resolve the problem with ambiguity. This problem refers to the fact that multiple profiles may represent one and the same person and therefore must be merged into a single generalized expert profile, *e.g.*, the clustering algorithm discussed in (Buelens et al., 2011) can be applied for this purpose. A different approach to the ambiguity problem has been proposed in (Boeva et al., 2012). Namely,the similarity between the personal data (profiles) of experts is used to resolve the problem with ambiguity. The split and merge of expert profiles is driven by the calculation of similarity measure between the different entities composing the profile, *e.g.* expert name, email address, affiliations, co-authors names etc.

In this work, we use a Dynamic Time Warping (DTW) based approach to deal with the ambiguity issue. In general, the DTW alignment algorithm finds an optimal match between two given sequences (*e.g.*, time series) by warping the time axis iteratively until an optimal matching (according to a suitable metric) between the two sequences is found (Sankoff and Kruskal, 1983). Due to its flexibility, DTW is widely used in many scientific disciplines and business applications as *e.g.*, speech processing, bioinformatics, matching of one-dimensional signals in the online hand writing communities etc. A detail explanation of DTW algorithm can be found in (Sakoe and Chiba, 1978), (Sankoff and Kruskal, 1983).

Initially, we use the DTW alignment algorithm to match the strings representing the expert names. The advantage of using DTW for string comparison is that it allows to easily detect partial matches when *e.g.*, an abbreviated expert name is still recognized as the same as the full name. If the calculated DTW distance between the two names is zero then it can be deduced that the expert names are identical. However, this is not enough to conclude that these two experts present one and the same person. Therefore, the DTW algorithm is applied to compare the input vectors presenting the affiliation information of the two matched experts. If they have the same affiliation then their official email addresses are matched. If the latter ones are the same then we can conclude that these two expert profiles present the same person. In the opposite case, *i.e.* the experts have different email addresses, then we treat them as two different individuals.

The data needed for constructing the expert profiles could be extracted from various Web sources, *e.g.*, LinkedIn, the DBLP library, Microsoft Academic Search, Google Scholar Citation, PubMed etc. There exist several open tools for extracting data from public online sources. For instance, Python LinkedIn is a tool which can be used in order to execute the data extraction from LinkedIn. In addition, the Stanford part-of-speech tagger (Toutanova, 2000) can be used to annotate the different words in the text collected for each expert with their specific part of speech. Next to recognizing the part of speech, the tagger also defines whether a noun is plural, whether a verb is conjugated, etc. Further the annotated text can be reduced to a set of keywords (tags) by removing all the words tagged as articles, prepositions, verbs, and adverbs. Practically, only the nouns and the adjectives are retained and the final keyword set can be formed according to the following simple chunking algorithm:

- *adjective-noun(s) keywords:* a sequence of an adjective followed by a noun is considered as one compound keyword, *e.g.*, "molecular biology";
- *multiple nouns keywords:* a sequence of adjacent nouns is considered as one compound keyword, *e.g.*, "information science";
- *single noun keywords:* each of the remaining nouns forms a keyword on its own.

In view of the above, an expert profile can be defined as a list of keywords (domain-specific topics), extracted from the available information about the expert in question, describing her/his subjects of expertise. Assume that n different expert profiles are created in total and each expert profile i ($i = 1, 2,..., n$) is represented by a list of p_i keywords.

3.2 Assessing of Expertise

An expert may have more extensive knowledge or experience in some topics than in others and this should be taken into account in the construction of expert profiles. Thus the gathered information about each individual expert can further be analysed and used to assess her/his levels of expertise to the different topics that compose her/his expert profile.

There is no standard and no absolute definition for assessing expertise. This usually depends not only on the application area but also on the subject field. For instance, in the peer-review setting, appropriate experts (reviewers, committee members, editors) are discovered by computing their profiles, usually based on the overall collection of their publications (Cameron, 2007). However, the publication quantity alone is insufficient to get an overall assessment of expertise. To incorporate the publication quality in the expertise profile, Cameron used the impact factor of publications' journals (Cameron, 2007). However, the impact factor in itself is arguable (Hecht et al., 1998), (Seglen, 1997). Therefore, Hirsch proposed another metric, the "H-Index", to rank individuals (Hirsch, 2005). However, this index works fine only for comparing scientists working in the same field, because citation conventions differ widely among different fields (Hirsch, 2005). Afzal et al. proposed an automated technique which incorporates multiple facets in providing a more representative assessment of expertise (Afzal et al., 2011). The developed system mines multiple facets for an electronic journal and then calculates expertise' weights.

In the current work, we use weights to assess the relative levels of knowledge or experience an individual has in the topics he/she has shown to have an expertise. Let us suppose that a weighting method appropriate to the respective area is used and as a result each keyword (domain-specific topic) k_{ij} of expert profile i (i = 1,…, n) is associated with a weight w_{ij}, expressing the relative level (intensity) of expertise the expert in question has in the topic k_{ij}, *i.e.* $\sum_{j=1}^{p_i} w_{ij} = 1$ and $w_{ij} \in$ (0, 1] for i = 1,…, n.

In this way, each expert is described by the topics (keywords) in which he/she is an expert plus the levels (weights) of knowledge or experience he/she has in the different topics.

3.3 Expertise Similarity

The calculation of expertise similarity is a complicated task, since the expert expertise profiles usually consist of domain-specific keywords that describe their area of competence without any information for the best correspondence between the different keywords of two compared profiles. Therefore, it is proposed in (Boeva et al., 2012) to measure the similarity between two expertise profiles as the strength of the relations between the semantic concepts associated with the keywords of the two compared profiles. Another possibility to measure the expertise similarity between two expert profiles is by taking into account the semantic similarities between any pair of keywords that are contained in the two profiles.

Accurate measurement of semantic similarity between words is essential for various tasks such as, document (or expert) clustering, information retrieval, and synonym extraction. Semantically related words of a particular word are listed in manually created general-purpose lexical ontologies such as WordNet. WordNet is a large lexical database of English (Fellbaum, 2001), (Miller, 1995). Initially, the WordNet networks for the four different parts of speech were not linked to one another and the noun network was the first to be richly developed. This imposes some constraints on the use of WordNet ontology. Namely, most of the researchers who use it limit themselves to the noun network. However, not all keywords representing the expert profiles are nouns. In addition, the algorithms that can measure similarity between adjectives do not yield results for nouns hence the need for combined measure. Therefore, a normalized measure combined from a set of different similarity measures is defined and used in (Boeva et al., 2014) to calculate the semantic relatedness between any two keywords.

In the considered context the expertise similarity task is additionally complicated by the fact that the competence of each expert is represented by two components: a list of keywords describing her/his expertise and a vector of weights expressing the relative levels of knowledge/expertise the expert has in the different topics.

Let s be a similarity measure that is suitable to estimate the semantic relatedness between any two keywords used to describe the expert profiles in the considered domain. Then the expertise similarity S_{ij} between two expert profiles i and j, can be defined by using the weighted mean of semantic similarities between the corresponding keywords

$$S_{ij} = \sum_{l=1}^{p_i} \sum_{m=1}^{p_j} W_{lm} \cdot s(k_{il}, k_{jm}), \tag{1}$$

where $W_{lm} = w_{il}.w_{jm}$ is a weight associated with the semantic similarity $s(k_{il}, k_{il})$ between keywords k_{il} and k_{jm}, and $W_{lm} \in$ (0, 1] for l = 1,…, p_i and m = 1,…, p_j. It can easily be shown that $\sum_{l=1}^{p_i} \sum_{m=1}^{p_j} W_{lm} = 1$.

3.4 Expert Identification

As was mentioned in the introduction the experts finding task can be viewed as a list completion task, *i.e.* the user is supposed to provide a small number of example experts who have been used to work on similar problems in the past, and the system has to return similar experts.

The concept of expertise spheres has been introduced in (Boeva et al., 2012). Conceptually, these expertise spheres are interpreted as groups of experts who have strongly overlapping competences. In other words, the expertise sphere can be considered as a combination of pieces of knowledge, skills, proficiency etc. that collectively describe a group of experts with similar area of competence. Consequently, the user may find experts with the required expertise by entering the name(s) of example expert(s) and the system will return a list of experts with close (similar) expertise by constructing the expertise sphere of the given expert(s).

In order to build an expertise sphere of an expert it is necessary to identify experts with similar area of competence, *i.e.* for each example expert i a list of expert profiles which exhibit at least minimum (preliminary defined) expertise similarity with her/his expert profile needs to be generated. An expert profile j will be included in the expertise sphere of i if the following inequality holds $S_{ij} \geq T$, where $T \in$ (0, 1) is a preliminary defined threshold.

The experts identified can be ranked with respect to their expertise similarities to the example expert.

Table 1: Expert MeSH heading profiles.

Experts	MeSH headings
1	Kidney Transplantation; Liver Transplantation
2	Health Behavior
3	Drinking; Health Behavior; Health Knowledge, Attitudes, Practice; Program Evaluation
4	Models, Biological; Temperature; Models, Neurological; Water
5	Computer Simulation; Models, Molecular; Protons; Thermodynamics; Molecular Conformation
6	Vibration; Models, Molecular; Infrared Rays; Hydrogen Bonding
7	Monte Carlo Method; Models, Theoretical; Phase Transition; Thermodynamics
8	Photosynthesis; Quantum Theory
9	Health Behavior; Decision Support Techniques;…(more than 20 MeSH terms)
10	Polymorphism, Genetic

Another possibility is to present the domain of interest by several preliminary specified subject categories and then the available experts can be grouped with respect to these categories into a number of disjoint expert areas (clusters) by using some clustering algorithm, as *e.g.* (Boeva et al., 2014), (Boeva et al., 2016). In the considered context each cluster of experts can itself be interpreted as an expertise sphere. Namely, it can be thought as the expertise area of any expert assigned to the cluster and evidently, the all assigned experts are included in this sphere. In this case, in order to select the right individuals for a specified task the user may restrict her/his considerations only to those experts who are within the expert area (cluster) that is identical with (or at least most similar to) the task's subject. The specified subject and the expert area can themselves be described by lists of keywords (subject profiles), *i.e.* they can be compared by way of similarity measurement. In this scenario, weights can also be introduced by allowing the user to express her/his preferences about the relative levels of expertise the experts in query should have in the specified topics. In addition, the subject profiles that are used to present the different clusters of experts can also be supplied with weights. The experts in the selected cluster can be ranked with respect to the similarity of their expert profiles to the specified subject profile.

In case of a newly extracted (registered, discovered) expert we can classify him/her into one of the existing clusters of experts by determining his/her expertise sphere. Namely we initially calculate the expert's expertise spheres with respect to any of the considered expert areas. Then the expert in question is assigned to that cluster of experts for which the corresponding expertise sphere has the largest cardinality, *i.e.* the overlap between the two sets of experts is the highest.

4 INITIAL EVALUATION AND RESULTS

4.1 PubMed Data

The data needed for constructing the expert profiles are extracted from PubMed, which is one of the largest repositories of peer-reviewed biomedical articles published worldwide. Medical Subject Headings (MeSH) is a controlled vocabulary developed by the US National Library of Medicine for indexing research publications, articles and books. Using the MeSH terms associated with peer-reviewed articles published by Bulgarian authors and indexed in the PubMed, we extract all such authors and construct their expert profiles. An expert profile is defined by a list of MeSH terms used in the PubMed articles of the author in question to describe her/his expertise areas.

4.2 Metrics

Unfortunately, large data collections such as e.g. LinkedIn, the DBLP library, PubMed etc. contain a substantial proportion of noisy data and the achieved degree of accuracy cannot be estimated in a reliable way. Accuracy is most commonly measured by precision and recall. Precision is the ratio of true positives, *i.e.* true experts in the total number of found expert candidates, while recall is the fraction of true experts found among the total number of true experts in a given domain. However, determining the total number of true experts in a given domain is not feasible.

In the current work, we use *resemblance r* and *containment c* to compare the expertise retrieval solutions generated on a given set of experts by using the weighting method introduced in Section 3 with the solutions built on the same set of experts without taking into account the intensity of their expertise.

Let us consider two expertise retrieval solutions

$S = \{S_1, S_2,\ldots, S_k\}$ and $S' = \{S'_1, S'_2,\ldots, S'_k\}$ of the same set of experts, where S_i and S'_i, $i = 1, 2,\ldots, k$, are the corresponding expertise retrieval results. The first solution S is generated on the considered data set without taking into account the expert levels of expertise in different topics while the second one S' is a solution built by applying the proposed weighting method. Then the similarity between two expertise retrieval results S'_i and S_i, which are constructed for the same example expert, can be assessed by resemblance r:

$$r(S_i^{'}, S_i) = \left|S_i^{'} \cap S_i\right| \Big/ \left|S_i^{'} \cup S_i\right| \quad (2)$$

The overall resemblance r for expertise retrieval solutions S' and S can be defined as the mean of r values of the corresponding expertise retrieval results.

We also use *containment* c that assesses how S_i' is a subset of S_i:

$$c(S_i^{'}) = \left|S_i^{'} \cap S_i\right| \Big/ \left|S_i^{'}\right| \quad (3)$$

It is evident that the values of r and c are in the interval [0, 1].

4.3 Implementation and Availability

Publications originating from Bulgaria have been downloaded in XML format from the Entrez Programming Utilities (E-utilities) (Sayers).The E-utilities are the public API to the NCBI Entrez system and allow access to all Entrez databases including PubMed, PMC, Gene, Nuccore and Protein. The E-utilities use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various NCBI software components to search for and retrieve the requested data. The E-utilities are therefore the structured interface to the Entrez system, which currently includes 38 databases covering a variety of biomedical data, including biomedical literature. To access these data, a piece of software first makes an API call to E-Utilities server, then retrieves the results of this posting, after which it processes the data as required. Thus the software can use any computer language that can send a URL to the E-utilities server and interpret the XML response.

For calculation of semantic similarities between MeSH headings, we use MeSHSim which is an R package. It also supports querying the hierarchy information of a MeSH heading and information of a given document including title, abstraction and MeSH headings (Zhou et al., 2016).

In our experiments, we have appliedthe DTW algorithm to resolve the problem with ambiguity (see Section 3.1). For this purpose, we have used a Python library cdtw. It proposes a DTW algorithm for spoken word recognition which is experimentally shown to be superior over other algorithms (Paliwal et al. 1982).

Table 2: Expert MeSH heading weights.

Experts	MeSH heading weights
1	0.5; 0.5
2	1
3	0.25; 0.25; 0.25; 0.25
4	0.166; 0.333; 0.166; 0.333
5	0.285; 0.285; 0.142; 0.142; 0.142
6	0.5; 0.166; 0.166; 0.166
7	0.428; 0.285; 0.142; 0.142
8	0.75; 0.25
9	0.022;...; 0.045;…; 0.068;…; 0.25
10	1

4.4 Results and Discussion

Initially, a set of 4343 Bulgarian authors is extracted from the PubMed repository. After resolving the problem with ambiguity the set is reduced to one containing only 3753 different researchers. Then each author is represented by two components: a list of all different MeSH headings used to describe the major topics of her/his PubMed articles and a vector of weights expressing the relative levels of expertise the author has in the different MeSH terms composing her/his profile. The weight of a MeSH term that is presented in a particular author profile is the ratio of repetitions, *i.e.* the repetitions of the MeSH term in the total number of MeSH terms collected for the author. This weighting technique could additionally be refined by considering the MeSH terms annotating the recent publications of the authors as more important (*i.e.* assigning higher weights) than those met in the old ones.This idea is not implemented in the current experiments.

Examples of 10 expert MeSH heading profiles can be seen in Table 1. The corresponding weight vectors calculated as it was explained above can be found in Table 2.

We build expertise spheres of the ten example experts whose profiles are given in Table 1. Initially,

we construct the expertise spheres of these authors by applying the weighting method introduced in Section 3. Respectively, the expertise spheres of the same authors without taking into account the intensity of their expertise in the different MeSH topics containing in their profiles are also produced. Next the resemblance *r* and the containment *c* are used to compare the two expertise retrieval solutions generated on the set of extracted Bulgarian PubMed authors for the example expert profiles.

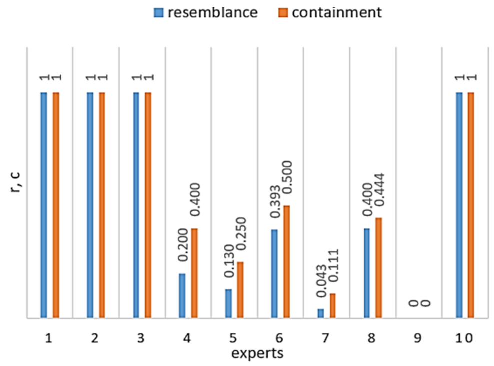

Figure 1: *r* and *c* scores calculated on the expertise retrieval results that are generated for the example experts given in Table 1 by selecting for each expert profile a fixed number (50) of the most similar expert profiles.

Figure 1 depicts *r* and *c* scores which have been calculated on the expertise retrieval results produced for the example experts by identifying for each expert profile a fixed number (50) of expert profiles that are most similar to the given one. As one can notice the obtained results are quite logical. Namely, the returned expertise retrieval results are identical (*r*= 1 and *c*= 1) when the experts have equally distributed expertise in the different MeSH headings presented in their profiles (*e.g.*, see experts: 1, 2, 3 and 10). However, in the other cases (see experts: 4, 5, 6, 7 and 8) the resemblance between the corresponding expertise retrieval results is not very high (maximum 0.4). Evidently, the produced expertise retrieval results can be significantly changed by using a weighting method for assessing expert expertise. The latter is also supported by the results generated for the containment *c*.

Similar results have also been obtained when the expertise retrieval results generated on the example experts are produced by using a preliminary defined similarity threshold (see Figure 2). This is supported by the very close overall resemblance scores generated by the two experiments: 0.57 (a fixed number of authors) and 0.61 (a similarity threshold), respectively. We have tested a list of different thresholds (all values in {0.3, 0.4, 0.5, 0.6, 0.7}). However, many of the expertise retrieval results generated on the example experts for the higher (above 0.3) thresholds were empty. The latter is most probable due to the fact that the extracted Bulgarian PubMed authors have very sparse expertise. Empty expert retrieval results have also been generated for expert 9 in the both experiments (Figure 1 and Figure 2), since as one can notice he/she has a very dispersed and unique expertise.

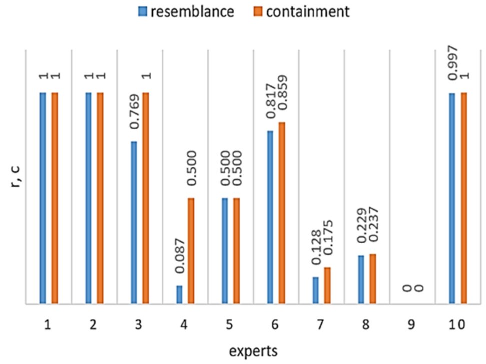

Figure 2: *r* and *c* scores calculated on the expertise retrieval results that are generated for the example experts given in Table 1 by selecting for each expert profile a list of those experts who exhibit at least 0.3 expertise similarity with the given profile.

5 CONCLUSIONS

This paper has discussed enhanced data-driven techniques for expert representation and identification. We have proposed a weighting method to assess the levels of expertise of an expert to the domain-specific topics. An expert profile has been presented by two components: a list of topics in which the person is an expert and a vector of weights presenting the relative levels of knowledge or experience the person has in the different topics. In this context, we have defined a way to estimate the expertise similarity between experts. Further we have considered expert identification techniques that return similar experts to ones provided by the user. The proposed techniques have been tested and evaluated on data extracted from PubMed repository.

Our future plans include the further refinement and validation of the proposed weighting method for assessing expert expertise on data coming from different application areas and subject fields.

REFERENCES

Abramowicz, W. et al. 2011. Semantically Enabled Experts Finding System - Ontologies, Reasoning Approach and Web Interface Design. In *ADBIS*. 2:157-166.

Afzal, M.T., Maurer, H. 2011. Expertise Recommender System for Scientific Community. *Journal of Universal Computer Science*, 17(11): 1529-1549.

Boeva, V. et al. 2012. Measuring Expertise Similarity in Expert Networks. In *6th IEEE Int. Conf. on Intelligent Systems*, IS 2012 IEEE Sofia Bulgaria 53-57.

Boeva, V. et al. 2014.Semantic-aware Expert Partitioning. *Artificial Intelligence: Methodology, Systems, and Applications*, LNAI. Springer Int. Pub. Switzerland.

Boeva, V. et al. 2016. Identifying a Group of Subject Experts Using Formal Concept Analysis. In *The 8th IEEE Int. Conf. on Intelligent Systems*, Sofia 464-469.

Balog, K., de Rijke, M., 2007. Finding similar experts. In *30th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, New York.

Balog, K., 2008. *People search in the enterprise*. PhD thesis, Amsterdam University.

Buelens, S., Putman, M., 2011. *Identifying experts through a framework for knowledge extraction from public online sources*. Master thesis, Gent University, Belgium.

Bozzon, A. et al. 2013. Choosing the Right Crowd: Expert Finding in Social Networks. In *EDBT/ICDT'13*. Genoa, Italy.

Cameron, D. L. 2007. *SEMEF: A Taxonomy-based Discovery of Experts, Expertise and Collaboration Networks*. MS thesis, The University of Georgia.

Campbell, C.S. et al. 2003. Expertise identification using Bibliography 189 email communications. In *12th Int. Conf. on Inform. and Knowl. Manag*. ACM Press.

Cohen, A. et al. 1998. The expertise browser: how to leverage distributed organizational knowledge. In *Workshop onCollaborative Inf. Seeking*. Seatle, WA.

Craswell, N., et al., 2006. Overview of the TREC-2005 Enterprise Track. In *14th Text Retrieval Conference*.

D'Amore, R. 2004. Expertise community detection. In *27th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. ACM Press.

Fellbaum, C., 2001. *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge.

Foner, L. 1997. Yenta: a multi-agent referral system for matchmaking system, In *The First Int. Conference on Autonomous Agents*. Marina Del Ray, CA.

Goldstone, R. L., 1994. The role of similarity in categorization: providing a groundwork. *Cognition*, 52:125–157.

Hawking, D. 2004. Challenges in enterprise search. In *15th Australasian Database Conference*. Australian Computer Society, Inc.

Hecht, F. et al.1998. The Journal Impact Factor: A Misnamed, Misleading, Misused Measure. *Cancer Genet Cytogenet* 04: 77-71, Elsevier Science Inc.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *PNAS* 102(46): 16569-72.

Hristoskova, A. et al. 2013. A Graph-based Disambiguation Approach for Construction of an Expert Repository from Public Online Sources. In *5th IEEE Int. Conf. on Agents and Artificial Intelligence*.

JAGAJOBS: the UK Jobs Data API. http://www.programmableweb.com/api/yagajobs

Kanfer, A. et al. 1997. Humanizing the net: social navigation with a'know-who' email agent. In *The 3th Conf. on Human Factors & The web*. Denver.

Kautz, H. et al. 1996. Agent amplified communication. In *The 13th Nat. Conf. on Artificial Intelligence* Portland.

Kautz, H., Selman, B., Shah, M. 1997. Referral Web: combining social networks and collaborative filtering. *Communications of the ACM* 40(3): 63-65.

Majio. https://maj.io/#/

Mattox, D.et al. 1999. Enterprise expert and knowledge discovery. In*HCI International'99*. Munich.

McDonald, D. W., Ackerman, M. S. 1998. Just talk to me: a field study of expertise location,In *CSCW'98*. ACM Press. Seattle, WA, 315-324.

Miller, G.A., 1995. WordNet: A Lexical Database for English. *Communications of the ACM* 38(11): 39-41.

Mockus, A., Herbsleb, J.D. 2002. Expertise browser: a quantitative approach to identifying expertise. In *24th Int. Conf. on Software Engineering*. ACM Press.

Paliwal, K.K.et al.1983. A modification over Sakoe and Chiba's Dynamic Time Warping Algorithm for Isolated Word Recognition. *Signal Proc.*4:329-333.

Sakoe, H. and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. In *IEEE Trans. on Acoust, Speech, and Signal Proc.*, ASSP-26, 43-49.

Sankoff, D., Kruskal, J. 1983. *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Addison Wesley Reading Mass.

Sayers, E. A general introduction to the E-utilities. http://www.ncbi.nlm.nih.gov/books/NBK25497/

Seglen, P. O. 1997. Why the impact factor of journals should not be used for evaluating research. In *BMJ 1997*. 314(7079):497.

Seid, D., Kobsa, A. 2000.Demoir: A hybrid architecture for expertise modelling and recommender systems.

Singh, H. et al. 2013. Developing a Biomedical Expert Finding System Using Medical Subject Headings. *Healthcare Informatics Research* 19(4): 243-249.

Toutanova, K., 2000. Enriching the knowledge sources used in a maximum entropy part of speech tagger. In *the Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora*. EMNLP/VLC-2000.

Tsiporkova, E., Tourwé, T. 2011. Tool support for technology scouting using online sources. In *ER Workshops 2011*. LNCS 6999:371–376. Springer.

Vivacqua, A. S. 1999. Agents for Expertise Location. In *The AAAI Spring Symposium on Intelligent Agents in Cyberspace*. Stanford, CA.

Zhou, J., Shui, Y., 2016. The MeSHSim package.