LT Expertfinder: An Evaluation Framework for Expert Finding Methods

Tim Fischer **Steffen Remus** **Chris Biemann**

Language Technology Group
Department of Informatics
Universität Hamburg, Germany
{5fischer,lastname}@informatik.uni-hamburg.de

Abstract

Expert finding is the task of ranking persons for a predefined topic or search query. Most approaches to this task are evaluated in a supervised fashion, which depends on predefined topics of interest as well as gold standard expert rankings. However, manually ranking experts can be considered highly subjective and small variants in rankings are hardly distinguishable. Particularly for dynamic systems, where topics are not predefined but formulated as a search query, we believe a more informative approach is to perform user studies for directly comparing different methods in the same view. In order to accomplish this in a user-friendly way, we present the LT Expertfinder web application, which is equipped with various query-based expert finding methods that can be easily extended, a detailed expert profile view, detailed evidence in form of relevant documents and statistics, and an evaluation component that allows a qualitative comparison between different rankings.

1 Introduction

Human expertise is one of the noteworthy resources in the world. However, true experts may be rare and their expertise difficult to quantify due to multiple continuously changing factors. The goal of *expert finding* is to rank people regarding their knowledge about a certain topic, which is a challenging, yet rewarding task with many real-world applications. Just to name a few, some applications might be: Companies may require highly trained specialists whose consultancies can be requested for specific problems (Balog et al., 2006), conference organizers may need to assign submissions to reviewers which best match their expertise (Fang and Zhai, 2007), recruiters may search for talented employees and emerging experts in a particular field (Serdyukov et al., 2008).

1.1 Problem statement

While it is common to cast approaches to expert finding in a supervised learning framework, this requires respective datasets, necessarily limited to a narrow set of topics. Some of such datasets are: the enriched versions of DBLP[1] provided by the ArnetMiner project (Tang et al., 2008), used by e.g. Deng et al. (2008); Yang et al. (2009); Moreira et al. (2011), or the W3C Corpus[2] of TREC[3], used by e.g. Balog et al. (2006). Obviously, however, the subjective nature of attributed expertise makes expert ranking quality hard to quantify. A certain value of an evaluation measure based on gold standard dataset comparison with respect to supervised or unsupervised system outputs does not necessarily guarantee a better or worse performance of one system compared to another. Also, depending on the targeted domain, a supervised setting might not be a viable option for evaluation. In real-world settings, where underlying data changes dynamically and expert finding is rather an interactive approach than a one-shot query evaluation, we find it more adequate to facilitate an evaluation procedure based on user studies, where alternative approaches are comparatively judged.

1.2 Motivation & Contribution

In this paper, we address this issue and present the LT Expertfinder web application, which is equipped with several query-based expert finding methods and can be easily extended. A detailed expert profile helps users to eventually select one expert in favor of another. Methodological evidence, in form of relevant documents and various

[1] https://aminer.org/lab-datasets/expertfinding/
[2] https://tides.umiacs.umd.edu/webtrec/trecent/parsed_w3c_corpus.html
[3] TextREtrieval Conference: https://trec.nist.gov/

Proceedings of NAACL-HLT 2019: Demonstrations, pages 98–104
Minneapolis, Minnesota, June 2 - June 7, 2019.

statistics, as well as a view of the query-dependent citation graph, is provided. Finally, we added an evaluation component that allows the qualitative comparison between different rankings. To the best of our knowledge, this is the first tool that provides evidences and a direct comparison to multiple expert rankings. We apply our system to the *ACL Anthology Network*[4] in order to find experts on various computational linguistics topics.

2 Related Work

Early expert finding systems relied on manually crafted, and manually queried, databases. Maintaining these databases is obviously a time consuming and complex task on the administrative and user side. Early automatic expert finding systems usually focused on specific domains like email (Campbell et al., 2003) or software documentation (Mockus and Herbsleb, 2002). One of the first approaches that automatically extracts expertise from any kind of document was the P@NOPTIC system by Craswell et al. (2001).

Shared tasks, such as the *Enterprise Track* of TREC (Craswell et al., 2005; Soboroff et al., 2006; Bailey et al., 2007; Balog et al., 2008) resulted in many automatic methods for predefined topics. Those systems can be grouped into four major categories: *a*) generative models (Balog et al., 2006; Fang and Zhai, 2007; Deng et al., 2008), *b*) voting models (Macdonald and Ounis, 2006), *c*) graph-based models (Serdyukov et al., 2008; Campbell et al., 2003; Jurczyk and Agichtein, 2007; Zhang et al., 2007), and *d*) discriminative models (Hashemi et al., 2013). For an extensive survey on expert finding methods, we refer to Lin et al. (2017).

Hawking (2004) highlights the importance of expert profiling mainly because the results should provide more information than simply a ranked list of person names. Balog and De Rijke (2007) emphasizes the importance of the social network, i.e. colleagues and collaborators contribute greatly to the value of an expert.

Thus, systems, such as the ArnetMiner[5] tool (Tang et al., 2008), aim at modeling entire academic social networks by automatically extracting researcher profiles from the Web. Moreover, ArnetMiner models topical aspects of papers, authors, and publication venues. The CareerMap[6] (Wu et al., 2018), which is now a component of ArnetMiner, visualizes a scholar's career trajectory, which is extracted from ArnetMiner's publication database. The CL Scholar[7] system (Singh et al., 2018) mines textual and network information for knowledge graph construction and question answering using natural language or keywords. CSSeer[8] (Chen et al., 2013) is a keyphrase-based recommendation system for expert finding based on the CiteSeerX[9] digital library and Wikipedia. It extracts keyphrases from the title and abstract of documents in CiteSeerX and utilizes this information to infer the author's expertise. The Expert2Bólè system (Yang et al., 2009) features generic expert finding as well as bólè search, which aims at identifying the top supervisors in a given field. The authors argue that generic expert finding methods are insufficient for finding specific experts for different purposes. In their application, bólè search, it is for example more important to find persons who are able to judge and nuture other experts than to assess their own expertise. Hence, generic expert finding methods cannot be applied to this problem.

3 LT Expertfinder

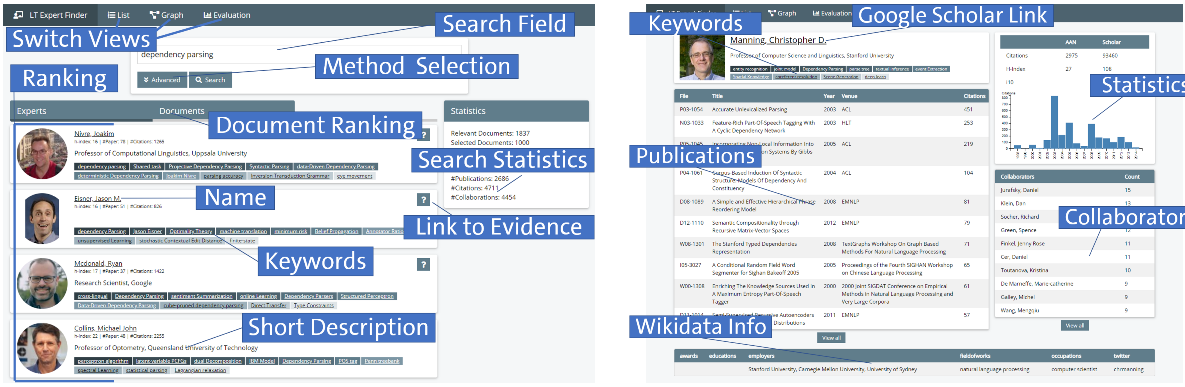
The LT Expertfinder features detailed expert profiles and various expert finding methods in an extendible framework. Moreover, it provides user-friendly evaluation methods: a view of the query-dependent citation graph, an evaluation view combining different results and provenances in form of relevant documents and related statistics.

3.1 Dataset

For the purpose of this paper, we use the *ACL Anthology Network* (AAN, Radev et al., 2013) as our underlying corpus, but note that it can be easily exchanged. The application is not limited to this particular data source. The AAN is based on papers in *ACL*[10] *Anthology* – a digital archive of conference and journal papers about natural language processing and computational linguistics – and provides citation and collaboration information. In its current version from December 2016, the AAN in-

[4] `http://tangra.cs.yale.edu/newaan/`
[5] `https://aminer.org/`

[6] `https://aminer.org/mostinfluentialscholar/ml`
[7] `https://github.com/CLScholar`
[8] `http://csseer.ist.psu.edu/`
[9] `http://citeseerx.ist.psu.edu`
[10] Association of Computational Linguistics



(a) List view: select a ranking method in the 'Advanced' tab, is-sue a search query, and get the results in a list view with concise author information as well as detailed result set statistics.

(b) Profile view: list all publications ranked by number of cita-tions, view collaboration information and external information from Wikidata as well as Google Scholar (if available) and link to the respective source pages.

Figure 1: LT Expertfinder Application

cludes more than 23K papers including their full text, 18K authors, 124K citations and 142K distinct co-authorship relations.

We further enriched the data with more detailed author information by heuristic Wikidata and Google Scholar entity page look-ups that match an author's name. Note that not every author has a Wikidata or Google Scholar entry, and some authors have multiple entries. In total we count approximately 9K authors with a matching Google Scholar entity, and 14K authors with matching Wikidata entities, of which 1.5K authors can be linked to exactly one Wikidata entity. Our heuristic does not distinguish between Wikidata entities and shows them all, whereas only the first Google Scholar entity is selected.

3.2 Application

The main contributions of this tool are to provide different expert search methods, detailed expert profiles and evidence features, which support a user's decision making process. The application's main page is shown in Figure 1a, which contains a simple search field for query input and a list of the retrieved ranked list of experts. The ranked expert list consists of condensed expert profiles showing the name, an image, a description, statistics and keywords representing expertise areas. The result is obtained by the particular method that is selected beforehand from a range of different expert finding methods.

The profile view (cf. Figure 1b) can be accessed by clicking on an expert's name (anywhere in the application). It shows publications as well as collaborators, various statistics like citations, citations over time, h-index and i10-index, and more. Keywords are extracted for each document in the corpus using a keyword extractor tool (Wiedemann et al., 2018)[11], which provides results as a ranked list of keywords[12]. In order to provide keywords for each author, the keywords of each document that an author has written are aggregated and ranked by document frequency. Lastly, the profile view shows information such as awards, educational degrees, employers (current and previous) as extracted from Wikidata.

3.3 Expert Finding Methods

We implemented three initial expert finding methods for the LT Expertfinder.

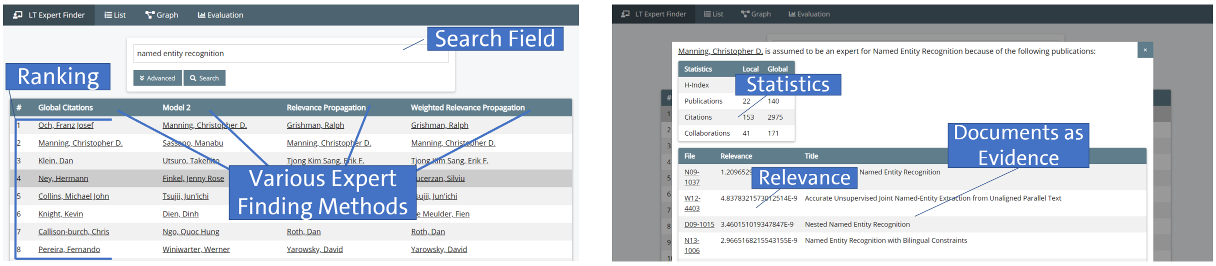
3.3.1 Model2

The document generation model by Balog et al. (2012) is widely used as a baseline to compare expert finding methods. In their original paper, Balog et al. (2012) present two models: Model1, the candidate generation model and Model2, the document generation model. Their experiments reveal that Model2 performs better.

The main challenge of the expert finding task is the accurate estimation of $p(q|c_i)$ as a ranking function of a candidate expert $c_i \in C$ and query terms q (Balog et al., 2012). The probability $p(q|c_i)$ is estimated by using a simple generative process: Given a candidate c_i, select a document d associated with c_i and generate the query q with the probability $p(q|d, c_i)$, which is obtained

[11] https://github.com/uhh-lt/lt-keyterms

[12] We only keep the top ten multi-term keywords per document.



(a) Evaluation view: in this example, four different methods rank experts for the query ”Named Entity Recognition”. Click-ing authors opens the evidence view (cf. Figure 2b).

(b) Evidence view: this view is opened from the evaluation view (cf. Figure 2a). It shows the author, which is linked to the au-thor profile, several query dependent statistics and, and relevant documents (which open via a click).

Figure 2: Evaluation with LT Expertfinder

using a language modeling approach $p(q|d, c_i) \approx p(q|d)$.

3.3.2 Relevance Propagation (RP)

Serdyukov et al. (2008) proposed graph-based approaches to expert finding. They introduced so-called expertise graphs, which consist of candidate experts and documents connected by authorship relations. Expertise graphs are query-dependent, as they are constructed from the relevant documents that are retrieved by a standard document retrieval for a given query. Serdyukov et al. model expert finding as a random walk through the expertise graph where authors are ranked by the ‘number of visits‘ of the random walker. In their paper, they present different random walk techniques, with incremental improvements. We re-implemented the k-step random walk as well as the infinite random walk. The infinite random walk is based on the assumption that the walk to find experts is a non-stop process. This technique is run iteratively until the expert rankings converge as opposed to the k-step random walk, which applies the calculations a fixed number of times.

3.3.3 Weighted Relevance Propagation

We further improve RP by introducing additional edges and edge weights. We include document citations, co-authorship relations and various weighting schemes for every edge type. Document citations are weighted by recency since we argue that a random user will most likely decide to read the most recent paper. Co-authorship relations are weighted by the number of total co-authorships, i.e. all outgoing edges to other author nodes. Authorships are weighted by a combination of the local and and the global h-index, where the local h-index refers to the h-index that is computed on the current result set of documents, and the global h-index refers to the to corpus wide h-index. As the h-index represents both the number of publications as well as the number of citations per publication, it is a suitable choice for determining the query-independent relevance of an author. Finally, the expert ranking is obtained by an infinite random walk through the weighted expertise graph. The main difference to RP is, that this method’s infinite random walk is applied with respect to the calculated weightings of the expertise graph.

3.3.4 Other methods

The tool also supports several other methods. It includes basic ranking methods based on simple statistics like h-index and citation count. These methods basically find all authors that dealt with the query topic and then rank the authors by their global or local h-index or citation count. In addition to that, the tool includes PageRank (Page et al., 1999), which ranks authors based on their incoming citations and co-authorships. Lastly, the tool contains a ranking method based on relevance scores obtained from a document retrieval on the query topic. Simply put, this method utilizes the sum of the relevance scores of an author’s documents to create an expert ranking.

3.4 Comparison & Evidence

One of the major contributions of our tool is to provide a user-friendly comparison method. The evaluation view (cf. Figure 2a) executes the major expert finding methods and presents columnar results. With this view, it is easy to identify differences as well as to qualitatively compare the results. Clicking on an expert’s name in this component will open the evidence view (cf. Figure 2b) for further investigation. It shows the documents

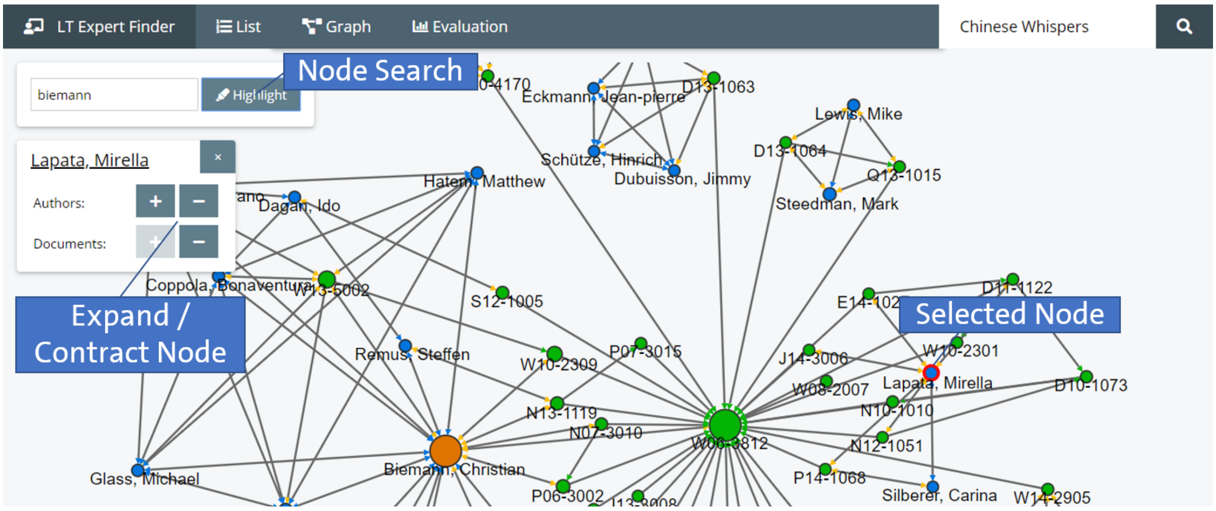


Figure 3: Graph view: authors are rendered as blue nodes, documents are rendered as green nodes, the highlighted node is rendered in red. The size of a node reflects its relevance. The graph is initially filtered by relevance to reduce cognitive overload and can be expanded or reduced for particular nodes.

that are relevant to the query and written by the candidate expert including their document relevance score calculated by the respective method, and several query dependent statistics such as h-index, number of citations, etc.

The graph view, which is shown in Figure 3 visualizes the query-based citation network for a particular method. Documents and authors are rendered as nodes whereas citations, authorship relations and co-authorship relations are represented as edges. Thus, it allows a quick peek into the data and an even better understanding of the results.

4 Conclusion

We presented the LT Expertfinder, a user-friendly tool for expert search, expert profiling, and most of all it enables the qualitative comparison of different ranking approaches and provides evidence for the ranking process. We implemented several ranking methods that can be easily extended with more methods. Also, it provides detailed expert profiles, which are linked to Wikidata and Google Scholar. Additionally, an explorable graph view is provided, which helps for further analysis of the results. This combination of features in a single tool is, to the best of our knowledge, still unexplored and helpful for the community for further development and evaluation of expert finding methods. For the future, we plan to expand our corpus using automatic crawling methods of scientific papers, which are analyzed and indexed on a daily basis. Crawling the ACL Anthology has already been successfully performed by Singh et al. (2018) with the help of their PDF Extraction tool OCR++ (Singh et al., 2016), which we also intend to use. Further, we plan to utilize the LT Expertfinder to develop methods for finding emerging experts in a field. We release the LT Expertfinder as freely available, open source application, under a permissive license.[13,14,15,16] A short demonstration video is also available[17].

References

Peter Bailey, Arjen P. De Vries, Nick Craswell, and Ian Soboroff. 2007. Overview of the TREC 2007 Enterprise Track. In *Proceedings of the Sixteenth Text REtrieval Conference*, Gaithersburg, MD, USA.

Krisztian Balog, Leif Azzopardi, and Maarten de Rijke. 2006. Formal Models for Expert Finding in Enterprise Corpora. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–50, Seattle, WA, USA.

Krisztian Balog and Maarten De Rijke. 2007. Determining Expert Profiles (with an Application to Expert Finding). In *Proceedings of the 20th Inter-*

[13]Demo: http://ltdemos.informatik.uni-hamburg.de/lt-expertfinder/ui/

[14]Docker: https://cloud.docker.com/u/uhhlt/repository/docker/uhhlt/xpertfinder

[15]Source Code: https://github.com/uhh-lt/lt-expertfinder

[16]License: Apache License, Version 2.0

[17]https://youtu.be/A4yRZezWUvE

national Joint Conference on Artifical Intelligence, pages 2657–2662, Hyderabad, India.

Krisztian Balog, Yi Fang, Maarten de Rijke, Pavel Serdyukov, and Luo Si. 2012. Expertise Retrieval. *Foundations and Trends in Information Retrieval*, 6(2):127–256.

Krisztian Balog, Paul Thomas, Nick Craswell, Ian Soboroff, Peter Bailey, and Arjen P. De Vries. 2008. Overview of the TREC 2008 Enterprise Track. In *Proceedings of the Eighteenth Text REtrieval Conference*, Gaithersburg, MD, USA.

Christopher S. Campbell, Paul P. Maglio, Alex Cozzi, and Byron Dom. 2003. Expertise Identification Using Email Communications. In *Proceedings of the 12th ACM International Conference on Information and Knowledge Management*, pages 528–531, New Orleans, LA, USA.

Hung-Hsuan Chen, Pucktada Treeratpituk, Prasenjit Mitra, and C. Lee Giles. 2013. CSSeer: An Expert Recommendation System Based on CiteseerX. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 381–382, Indianapolis, IN, USA.

Nick Craswell, Arjen P. De Vries, and Ian Soboroff. 2005. Overview of the TREC 2005 Enterprise Track. In *Proceedings of the Fourteenth Text REtrieval Conference*, pages 199–205, Gaithersburg, MD, USA.

Nick Craswell, David Hawking, Anne-Marie Vercoustre, and Peter Wilkins. 2001. P@NOPTIC expert: Searching for experts not just for documents. In *Ausweb Poster Proceedings*, pages 21–25, Coffs Harbour, Queensland, Australia.

H. Deng, I. King, and M. R. Lyu. 2008. Formal Models for Expert Finding on DBLP Bibliography Data. In *Proceedings of the Eighth International Conference on Data Mining*, pages 163–172, Pisa, Italy.

Hui Fang and ChengXiang Zhai. 2007. Probabilistic models for expert finding. In *Proceedings of the 29th European Conference on Information Retrieval Research*, pages 418–430, Rome, Italy.

Seyyed Hadi Hashemi, Mahmood Neshati, and Hamid Beigy. 2013. Expertise retrieval in bibliographic network: A topic dominance learning approach. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pages 1117–1126, San Francisco, CA, USA.

David Hawking. 2004. Challenges in Enterprise Search. In *Proceedings of the 15th Australasian Database Conference*, pages 15–24, Dunedin, New Zealand.

Pawel Jurczyk and Eugene Agichtein. 2007. Discovering Authorities in Question Answer Communities by Using Link Analysis. In *Proceedings of the 16th ACM International Conference on Information and Knowledge Management*, pages 919–922, Lisbon, Portugal.

Shuyi Lin, Wenxing Hong, Dingding Wang, and Tao Li. 2017. A survey on expert finding techniques. *Journal of Intelligent Information Systems*, 49(2):255–279.

Craig Macdonald and Iadh Ounis. 2006. Voting for Candidates: Adapting Data Fusion Techniques for an Expert Search Task. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 387–396, Arlington, VA, USA.

Audris Mockus and James D. Herbsleb. 2002. Expertise Browser: A Quantitative Approach to Identifying Expertise. In *Proceedings of the 24th International Conference on Software Engineering*, pages 503–512, Orlando, FL, USA.

Catarina Moreira, Bruno Martins, and Pável Calado. 2011. Using Rank Aggregation for Expert Search in Academic Digital Libraries. In *Simpósio de Informática, INForum*, Portugal.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Dragomir R. Radev, Pradeep Muthukrishnan, Vahed Qazvinian, and Amjad Abu-Jbara. 2013. The ACL Anthology Network corpus. *Language Resources and Evaluation*, 47(4):919–944.

Pavel Serdyukov, Henning Rode, and Djoerd Hiemstra. 2008. Modeling Multi-step Relevance Propagation for Expert Finding. In *Proceedings of the 17th ACM International Conference on Information and Knowledge Management*, pages 1133–1142, Napa Valley, CA, USA.

Mayank Singh, Barnopriyo Barua, Priyank Palod, Manvi Garg, Sidhartha Satapathy, Samuel Bushi, Kumar Ayush, Krishna Sai Rohith, Tulasi Gamidi, Pawan Goyal, and Animesh Mukherjee. 2016. OCR++: A Robust Framework For Information Extraction from Scholarly Articles. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3390–3400, Osaka, Japan.

Mayank Singh, Pradeep Dogga, Sohan Patro, Dhiraj Barnwal, Ritam Dutt, Rajarshi Haldar, Pawan Goyal, and Animesh Mukherjee. 2018. CL Scholar: The ACL Anthology Knowledge Graph Miner. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 16–20, New Orleans, LA, USA.

Ian Soboroff, Arjen P. De Vries, and Nick Craswell. 2006. Overview of the TREC 2006 Enterprise Track. In *Proceedings of the Fifteenth Text REtrieval Conference*, Gaithersburg, MD, USA.

Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, pages 990–998, Las Vegas, NV, USA.

Gregor Wiedemann, Seid Muhie Yimam, and Chris Biemann. 2018. A Multilingual Information Extraction Pipeline for Investigative Journalism. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 78–83, Brussels, Belgium.

Kan Wu, Jie Tang, Zhou Shao, Xinyi Xu, Bing Gao, and Shu Zhao. 2018. CareerMap: Visualizing career trajectory. *Science China Information Sciences*, 61:1–3.

Zi Yang, Jie Tang, Bo Wang, Jingyi Guo, Juanzi Li, and Songcan Chen. 2009. Expert2Bólè : From Expert Finding to Bólè Search. In *Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining*, Paris, France.

Jun Zhang, Mark S. Ackerman, and Lada Adamic. 2007. Expertise Networks in Online Communities: Structure and Algorithms. In *Proceedings of the 16th International Conference on World Wide Web*, pages 221–230, Banff, AB, Canada.