



Understanding expert finding systems: domains and techniques

Mohammed Zuhair Al-Taie¹ · Seifedine Kadry² · Adekunle Isiaka Obasa³

Received: 6 April 2018 / Revised: 22 August 2018 / Accepted: 23 August 2018 / Published online: 30 August 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

Expert finding can be required for a variety of purposes: finding referees for a conference paper, recommending consultants for a software project, and identifying qualified answerers for a question in online knowledge-sharing communities, to name a few. This paper presents taxonomy of the task of expert finding that highlights the differences between finding experts, from the type of expertise indicator's point of view. The taxonomy supports deep understanding of different sources of expertise information in the enterprise or online communities; for example, authored documents, emails, online posts, and social networks. In addition, different content and non-content features that characterize the evidence of expertise are discussed. The goal is to guide researchers who seek to conduct studies regarding the different types of expertise indicators and state-of-the-art techniques for expert finding in organizations or online communities. The paper concludes that although researchers have utilized a large number of graph and machine-learning techniques for locating expertise, there are still technical issues associated with the implementation of some of these methods. It also corroborates that combining content-based expertise indicators and social relationships has the benefit of alleviating some of the issues related to identifying and ranking answer experts. The above findings give implications for developing new techniques for expert finding that can overcome the technical issues associated with the performance of current methods.

Keywords Enterprise · Expert finding · Expert finding methods · Community question answering · Online communities · Taxonomy

1 Introduction

The search for people with expertise has been addressed in two major domains. The first one is the enterprise, where knowledge hierarchy is well defined, and information quality is high (Wang et al. 2013). The goal was to manage and optimize human resources (Li et al. 2016). Consider, for example, an enterprise with a large size of employee data. The managers would be happy if the degree of expertise of each employee, on different topics, is determined. This would allow them to find the people who are qualified for a particular job. With the advent of Web 2.0 technologies,

Community Question Answering (CQA) platforms, such as Quora, StackOverflow (SO), and Yahoo! Answers (YA), became another closely related domain to Expert Finding (EF). They are a type of crowdsourcing knowledge services that use various kinds of human knowledge to solve complex and cognitive tasks (Yang et al. 2015).

The task of expert finding is different from the traditional document retrieval systems that search for real and concrete documents in a collection, because EF aims at locating a higher semantic and abstract concept “person” (Fu et al. 2007a, b). Fulfilling the requirements of this task will help us to find the most appropriate collaborators for a project, a research topic, or consultancy (Zhang et al. 2008). In the realm of online communities, identifying prominent contributors can help to evaluate the quality of answers from different users (Yang and Manandhar 2014) and improve the likelihood of an appropriate reply by disseminating new questions to the appropriate experts (Procaci et al. 2016). It can also improve the experience of other community members who have less knowledge, reduce user waiting time, and conserve system resources (Zheng et al. 2012). To achieve

✉ Seifedine Kadry
s.kadry@bau.edu.lb

¹ Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

² Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut, Lebanon

³ Department of Computer Science College of Science and Technology, Kaduna Polytechnic, Kaduna, Nigeria

this aim, searching for experts has been an active research area in many fields such as artificial intelligence, knowledge management, computer-supported cooperative work, and others.

The contribution of this paper can be summarized as follows: First, different kinds of expertise indicators are organized into a taxonomy that is derived from the recent literature about EF methods. Second, different content and non-content features are discussed to represent information sources for EF. Third, methods of EF are surveyed and discussed, which are listed in the taxonomy.

The rest of this paper is organized as follows: Sect. 2 presents the methodology of our study; Sect. 3 describes the taxonomy of expert finding systems based on two criteria: domain and technique. Part one of this taxonomy addresses all major types of expertise indicators identified in the enterprise and the online community. Part two discusses the techniques used by researchers to locate expertise in the enterprise and the online community. These techniques are either graph-based or machine learning-based. Section 4 concludes with the comparison of different approaches and condition of the current work; the conclusion is given in Sect. 5.

2 Methodology

This paper reviews the current literature to identify the major work that has been done to address the task of expert finding in online communities and corporations. It brings types of information sources (expertise indicator) together with methods used to find expert people in the enterprise and online communities.

Our search strategy aims to find a comprehensive and unbiased collection of primary studies from the literature. We leveraged the major databases that are widely used in computer science and information systems research, which are the ACM Digital Library, the IEEE Explore Digital Library, SpringerLink, ScienceDirect, and ISI Web of Science (WoS). These are rich and comprehensive databases containing bibliographic information of a plethora of publications from all major publishers of the computing literature.

However, in the field of software engineering, WoS mainly records publications published in premium journals and only a few conferences, whereas conferences (and sometimes workshops and symposiums) are major outlets for publishing in CS and SE. Thus, we mainly use WoS along with Google Scholar 6 for cross-checking with the search results from the chosen sources and for performing some meta-analyses. Our search terms stem from the research questions and can be categorized into two major dimensions, as shown in Fig. 1: Enterprise and Online Community. On one hand, we aim at exploring the different behavior models or diagrams considered in the primary studies. On the other

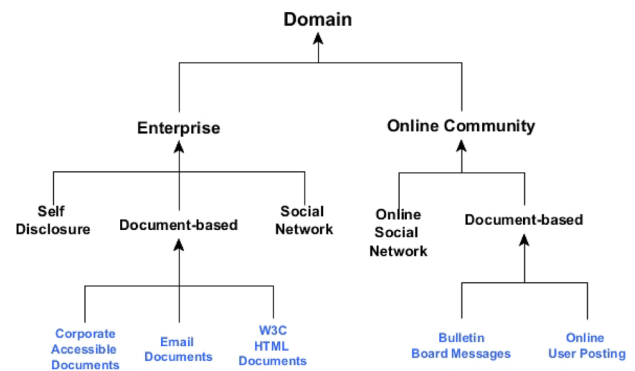


Fig. 1 Classification of EF systems based on domain

hand, we aim to discover relevant types of studies that have been performed for checking different types of inconsistencies. To ensure a sufficient scope of searching, we also consider different alternative words (e.g., behavior/behaviour/behavioural and model/diagram), synonyms (e.g., model, diagram, workflow, and process), and abbreviations for the search terms. Then, we used the Boolean operator “OR” to join the alternate words and synonyms and the Boolean operator “AND” to form a sufficient search string.

3 Expert finding taxonomy

One look at the literature tells us that researchers have used many methods for the task of EF. For example (Balog et al. 2012), is an extensive survey that highlights advances in models and algorithms of expertise retrieval. The goal of expertise retrieval (also known as expertise location or expertise identification) is to link humans to expertise areas, and vice versa, and is often conducted within the scope of a single company. In Yimam-Seid and Kobsa (2003), the different types of EF systems for organizations are discussed. In Maybury (2006), state-of-the-art commercially available tools for EF are described. In Lappas et al. (2011), some algorithms for expertise evaluation, team identification, and expert-location systems in social networks are highlighted. Lin et al. (2017) presented the state-of-the-art methods in expert finding and summarize them into different categories based on their underlying algorithms and models. In another recent survey (Wang et al. 2018), the authors presented of the state-of-the-art techniques for the expert recommendation in CQA, followed by a summarization and comparison of the existing methods, both their advantages and shortcomings.

The current study differs from the previous studies in that it studies these methods in both the organization and online environments. It relates to the different implementation domains of EF with recent advances in EF methods. The

main output is a new taxonomy of EF that comprehensively relates different types and approaches. This leads to classify EF systems based on two criteria: domain and technique.

3.1 Classification based on domain

Based on where the task of EF is applied (domain), the taxonomy classifies EF methods, as depicted in Fig. 1, into two typical domains: enterprise (or organization) and online community. The contents of the figure are based on the work introduced by Wang et al. (2013).

The first category, which is the enterprise, uses three sources of expertise to find one's expertise areas and level of expertise: self-disclosed information, documents, and social networks. The second category, which is online communities, uses two sources of expertise: documents and social networks.

3.1.1 Enterprise

Compared to online knowledge-sharing communities, knowledge in organizational knowledge repositories is well documented, and information quality of all content is sufficient (Wang et al. 2013). Finding expert people in an organization is a frequently encountered problem (Fu et al. 2007a, b). Therefore, most organizations maintain database repositories to store employees' qualifications.

EF in the enterprise has two objectives (Zhu et al. 2010). The first one is "who are the experts on topic X?" which is related to finding experts on a particular topic such as "Scala programming" or "dark ages." The other objective could be "what does person Y know?" or "does person Y know about subject X?" Although knowledge in the enterprise is carried by people, discovering who knows what is often challenging (Campbell et al. 2003).

In the enterprise corporation, there are three main types of information sources that can be utilized to extract evidence of expertise: self-disclosed information (Li et al. 2016; Wang et al. 2013), documents (Wang et al. 2013; Petkova and Croft 2008), and social networks (Zhang et al. 2007a, b).

To summarize, we can say that expertise indicators that are based on self-disclosed information are time-consuming and difficult to keep up-to-date. On the other hand, expertise indicators that use documents and social influence are both important sources for automated EF systems.

3.1.2 Online community

Online communities are Internet-based virtual places that specialize in knowledge seeking and sharing. People with common interests gather and discuss different topics at a variety of depths (Wang et al. 2013; Seo and Croft 2009). EF has also been the subject of many studies in online domains

including email networks, newsgroups, and microblogs. Users in knowledge-based online communities can post questions and then wait for other members to answer, browse history questions that other members have asked, or give answers to particular questions (El-Korany 2013). Not all users in online communities are active in the same way, as it was found that some users are consistently active, some start active but end passive over time, and some start passive but become active over time (Pal et al. 2012).

What makes the identification of experts in online communities of a great significance to researchers is that online communities are large knowledge repositories with millions of participants and millions of text documents. In addition to that, ordinary users can communicate with expert users, which is not as difficult as it is in the real world (Seo and Croft 2009). Due to a significant amount of data that is available in these repositories, different data-mining tasks can be launched, such as looking for trends, user sentiment analysis, or product reviews (Akram et al. 2016).

Evidence of expertise in online communities can be extracted from two sources: social networks (Aslay et al. 2013; Lu et al. 2012; Wang et al. 2013) and documents (Bian et al. 2008; Krulwich et al. 1996; Souza et al. 2013).

Compared to knowledge in organizational knowledge repositories, knowledge in online communities does not have a universal knowledge structure and information tends to have a low quality, which can affect the performance of knowledge management activities that involve information processing (Wang et al. 2013).

3.2 Classification based on technique

Expert finding methods, whether for the enterprise or online communities, can be broadly divided, as depicted in Fig. 2, into two groups: graph techniques and machine-learning techniques (Huna et al. 2016).

3.2.1 Graph techniques

Formally, a social graph is represented as $G = (X, E)$, where X is the set of nodes and E is the set of links that connect them. A link between two nodes would mean a friendship, asker-answer interaction, email communications, or co-authorship. On online communities (e.g., CQAs), users together with their interactions are represented by a graph, called expertise graph (Pal and Konstan 2010), in which nodes are expert and non-expert users, and edges are some relationship. To identify and rank-expert users (Procaci et al. 2016), various graph-based (also called link-based) methods are applied to expertise graphs. They are broadly divided into two main categories: (1) computing measures (e.g., PageRank and HITS) and (2) graph properties (e.g., connectedness and centrality).

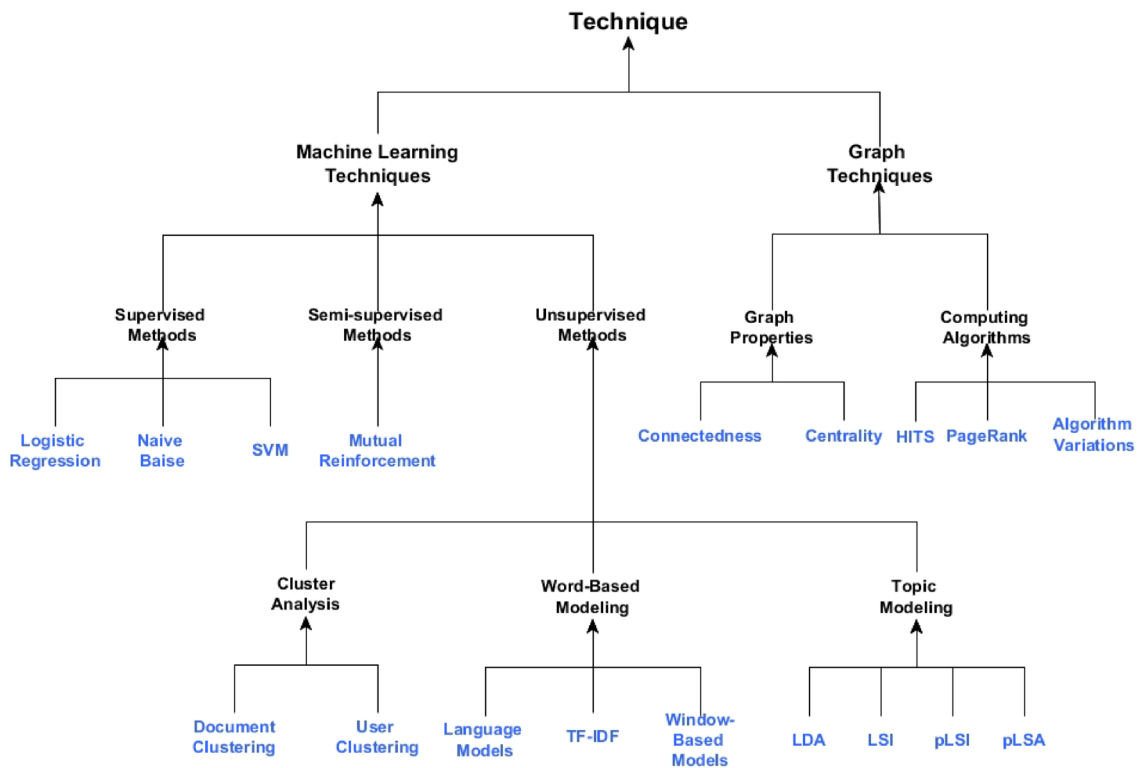


Fig. 2 Classification of EF systems based on technique

We summarize the work discussed about graph-based expert finding methods in Table 1 below.

3.2.2 Machine-learning techniques

Machine learning is the automated detection of meaningful patterns in data (Shalev-Shwartz and Ben-David 2014). Methods of machine learning include Bayes classifier, linear regression, and k -means cluster. They have been applied in a wide range of applications the Web, bioinformatics, medicine, and astronomy. To tackle the task of EF, researchers have been applying machine-learning methods to features extracted from sources of expertise, whether in the enterprise or online social networks, for extracting expertise dimensions (Sahu et al. 2016a, b; Pal et al. 2012). The features can be classified into two groups: content-based and non-content-based. The previous studies have shown that both content-based and non-content-based are valuable sources of information for the task of EF (El-Korany 2013).

We summarize the work discussed about machine-learning-based expert finding methods in Table 2 below.

4 Comparison of approaches

This section provides the advantages and disadvantages of both graph-based and machine-learning algorithms in addressing the problem of this study, whether in the enterprise or the online environment.

4.1 Comparison of approaches based on domain

Based on where the task of EF is applied, EF methods are classified into two domains: enterprise (or organization) and online community.

- *Enterprise* In the enterprise corporation, there are three main types of information sources that can be utilized to extract evidence of expertise: self-disclosed information, documents, and social networks.

Table 1 Summary of the existing graph-based expert finding methods

I	General graph technique	Specific graph technique	References
1	Computing algorithm	PageRank	Java et al. (2006)
		PageRank	Yang et al. (2008)
		HITS	Jurczyk and Agichtein (2007)
		HITS	Farahat et al. (2002)
		PageRank and HITS	Campbell et al. (2003)
		EntityRank, PageRank and HITS	Zhang et al. (2007a, b)
		SALSA	Lempel and Moran (2001)
		ExpertRank	Wang et al. (2013)
		Topic-sensitive PageRank	Zhou et al. (2012a)
		Twitterrank	Weng et al. (2010)
		TextRank	Fang et al. (2012)
		Expert Rank	Akram et al. (2016)
		AuthorRank	Liu et al. (2005)
		Weighted HITS	Yang and Wu (2014)
		2	Graph properties
Centrality	Aslay et al. (2013)		
Centrality	Chen and Nayak (2008)		
Centrality	Fisher et al. (2006)		
Centrality	Zhang et al. (2007a, b)		
Connectedness	Zhang et al. (2007)		
Connectedness	Rode et al. (2007)		
Connectedness	Seo and Croft (2009)		
Connectedness	Serdyukov et al. (2008)		
3	Hybrid approach	Content-based and graph-based	Lu et al. (2012)
		Content-based and graph-based	Horowitz and Kamvar (2012)
		Content-based and graph-based	Horowitz and Kamvar (2010)
		Content-based and graph-based	Fu et al. (2007a, b)
		Content-based and graph-based	Yang et al. (2015)
		Content-based and graph-based	Zheng et al. (2012)
		Content-based and graph-based	Kao et al. (2010)

- Potential experts may be required to disclose their expertise area on their personal Webpage, yellow pages, and directory listings of experts or by filling out questionnaires. However, Since manually building a one's profile in the enterprise is a time-consuming task and the profile is unlikely to remain as is, it is inevitable that we consult other sources of expertise indicators besides self-disclosed information (Wang et al. 2013).
- Document-based EF methods typically use text-mining techniques to capture the author's expertise (Wang et al. 2013). Some issues are associated with using enterprise documents for deriving expertise. For example, in most cases, a candidate's expertise is not fully documented in the organization and only parts of a document are related to the candidate whom it mentions (Fu et al. 2007a, b). Another issue is that documents cannot be used to measure a candidate's importance or influence in the social network that the candidate belongs to (Wang et al. 2013). As such, and to better to describe the knowledge of a particular candidate, it is important to use other sources of expert evidence in representing the expertise. The social influence of people plays a major role in the perception of their expertise.
- To summarize, we can say that expertise indicators that are based on self-disclosed information are time-consuming and difficult to keep up-to-date. On the other hand, expertise indicators that use documents and social influence are both important sources for automated EF systems.

Table 2 Summary of the existing machine learning-based expert finding methods

I	General machine-learning technique	Specific machine-learning technique	References
	Supervised	SVM	Zhou et al. (2012)
		Logistic regression	Li and King (2010)
		Gradient boosted decision trees	Dror et al. (2011)
		A number of classifiers	Pal et al. (2012)
		TrueSkill and SVM	Lappas et al. (2011)
		SVM and Naïve Bayes	Liu et al. (2015)
	Semi-supervised	Coupled mutual reinforcement	Bian et al. (2009)
	Unsupervised	Tensor factorization	Zhao et al. (2016)
		Single-pass clustering algorithm	Macdonald et al. (2008)
		Topic clustering	Xuan et al. (2013)
		WebDCC	Godoy and Amandi (2006)
		GMM, EM, and k-means	Pal and Counts (2011)
		Hierarchical clustering	Liu et al. (2005)
		Clustering of persons	Cao et al. (2005)
		User clustering	Procaci et al. (2016)
		User clustering	Procaci et al. (2016)
		Hierarchical and non-hierarchical clustering	Furtado et al. (2014)
		User clustering	
		User clustering	
		User clustering	Balog and De Rijke (2007)
		Group recommendation	Zhou et al. (2012b)
		Group recommendation	
		Group recommendation	Pal et al. (2012)

- *Online community* Compared to knowledge in organizational knowledge repositories, knowledge in online communities does not have a universal knowledge structure and information tends to have a low quality, which can affect the performance of knowledge management activities that involve information processing (Wang et al. 2013).

4.2 Comparison of approaches based on technique

- *Graph properties* Contrasting the performance of graph-based algorithms and content-based algorithms for measuring expertise, graph-based algorithms were found to outperform the content-based approaches (Campbell et al. 2003). While graph-based algorithms take into account both text and communication patterns, the content-based approaches consider the text only. On the other hand, some graph techniques are stronger than others. For example, simple centrality measures such as degree, closeness, and betweenness were found inferior to the use of Authorrang and PageRank algorithms (Liu et al. 2005), because centrality measures simply represent documents as un-weighted links between users, and neglect the relevance to the query (Serdyukov et al. 2008; Rode et al. 2007). Although graph-based methods are widely used for the task of EF in online communities, they suffer from two major problems: high computational intense (Pal and Counts 2011) and spamming (Chen and Nayak 2008). Because link-based methods ignore the content features, they may give misleading results if the answer has wrong, spam, or incomplete content (Akram et al. 2016). They only give reliable results if users in an online QA system behave properly (Chen and Nayak 2008).
- *Machine-learning techniques* We discuss this task by providing information on the following points:
 1. *User clustering* What makes cluster analysis be of little use for the problem of EF is that the clusters that result from clustering only contain users with the same level of expertise. In other words, candidate experts do not gather into one cluster of the data as they can be split into communities of experts and non-experts. Another issue is that the use of clustering for EF requires cross-validating the resulting clusters. There are many methods proposed for validating the resulting clusters such as conductance and average clustering coefficient. However, the absence of ground truth data sets makes the clustering validation more challenging.

2. *Window-based models* Word-based models have their problems. For example, and because they rely on the Vector Space Model (VSM) to represent words, they suffer from two main issues. (1) High dimensionality (Sahu et al. 2016a, b; Zolaktaf Zadeh 2012), because terms show as if they were independent of each other. (2) Because each word in the VSM model is represented by one vector to capture the semantic and syntactic information, the model cannot grab hold of homonymy and polysemy (Guo 2013). In addition, the word-based model, which is used by word-based models, was considered problematic regarding finding experts in online communities (Chen and Nayak 2008), for the following reasons (Zolaktaf Zadeh 2012). (1) Because questions in CQAs are posted in a short form, achieving enough overlap between a query term and a document, which is necessary to derive similarity, is not possible. (2) Because users have different levels of knowledge and technical competency, the content that the two models produce should not unavoidably use query terms that target one document rather than others.
3. *Topic models* One of the issues with the LDA model is that because it assumes that textual data are unstructured, it fails to capture particular features of the text, because it puts the entire document within one topic distribution (Riahi et al. 2012). Although topic models are widely used for discovering topics in text based on patterns of term co-occurrence, they still suffer from some issues. For instance, LSI cannot deal with polysemy or synonymy, pLSI suffers from variable overfitting (Aggarwal and Reddy 2013), and LDA puts the entire document in one topic distribution (Riahi et al. 2012).

5 Conclusion

This paper reviews the current literature to identify all the major studies that address the task of expert finding in online communities and corporations. The goal is to introduce taxonomy of the task of expert finding that highlights the differences between finding experts, from the type of expertise indicator and technique used point of view.

The paper concludes that although researchers have developed a large number of graph and machine-learning techniques for locating expertise, there are still several technical issues associated with the implementation of these methods. It also corroborates that combining content-based expertise indicators and social relationships has the benefit of alleviating some of the issues associated with identifying and ranking answer experts (Lu et al. 2012). The study limits itself to discussing the main techniques adopted by most remarkable approaches. It does not go deeply into the

architectural design of the various EF systems and how they differ from each other. Another limitation is that the paper does not address other theoretically and technically related research directions such as finding similar experts, expert team formation, or finding potential experts. Finally, among the many types of online social networks, the primary focus of the study is on the use of community question answering platforms for finding experts.

References

- Aggarwal CC, Reddy CK (2013) Data clustering: algorithms and applications. Chapman and Hall/CRC, London
- Akram AU, Iqbal K, Faisal CMS, Ishfaq U (2016) An effective experts mining technique in online discussion forums. In: Computing, electronic and electrical engineering (ICE cube), 2016 international conference on, IEEE
- Aslay Ç, O'Hare N, Aiello LM, Jaimes A (2013) Competition-based networks for expert finding. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval, ACM
- Balog K, De Rijke M (2007) Determining expert profiles (with an application to expert finding). In: Proceedings of the 20th international joint conference on artificial intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 2657–2662
- Balog K, Fang Y, de Rijke M, Serdyukov P, Si L (2012) Expertise retrieval. *Found Trends® Inf Retr* 6(2–3):127–256
- Bian J, Liu Y, Agichtein E, Zha H (2008) Finding the right facts in the crowd: factoid question answering over social media. In: Proceedings of the 17th international conference on world wide web, ACM
- Bian J, Liu Y, Zhou D, Agichtein E, Zha H (2009) Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In: Proceedings of the 18th international conference on world wide web, ACM
- Bouguessa M, Dumoulin B, Wang S (2008) Identifying authoritative actors in question-answering forums: the case of yahoo! answers. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, ACM
- Campbell CS, Maglio PP, Cozzi A, Dom B (2003) Expertise identification using email communications. Proceedings of the twelfth international conference on information and knowledge management, ACM. LA, USA
- Cao Y, Liu J, Bao S, Li H (2005) Research on expert search at enterprise track of TREC 2005. TREC, Ranson
- Chen L, Nayak R (2008) Expertise analysis in a question answer portal for author ranking. In: 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, 9–12December 2008, Sydney, Australia
- Dror IE, Champod C, Langenburg G, Charlton D, Hunt H, Rosenthal R (2011) Cognitive issues in fingerprint analysis: inter-andintra-expert consistency and the effect of a 'target' comparison. *Foren Sci Int* 208(1–3):10–17
- El-Korany A (2013) Integrated expert recommendation model for online communities. arXiv preprint. <http://arxiv.org/abs/1311.3394>
- Fang L, Huang M, Zhu X (2012) Question routing in community based QA: incorporating answer quality and answer content. Proceedings of the ACM SIGKDD workshop on mining data semantics, ACM. Beijing, China

- Farahat A, Nunberg G, Chen F (2002) Augreas: authoritativeness grading, estimation, and sorting. Proceedings of the eleventh international conference on information and knowledge management, ACM, VA, USA
- Fisher D, Smith M, Welser HT (2006) You are who you talk to: detecting roles in usenet newsgroups. In: System sciences, 2006. HICSS'06. Proceedings of the 39th annual Hawaii international conference on, IEEE
- Fu Y, Xiang R, Liu Y, Zhang M, Ma S (2007a) A CDD-based formal model for expert finding. In: Proceedings of the sixteenth ACM conference on conference on information and knowledge management, ACM
- Fu Y, Xiang R, Liu Y, Zhang M, Ma S (2007b) Finding experts using social network analysis. In: Proceedings of the IEEE/WIC/ACM international conference on web intelligence, IEEE computer society
- Furtado A, Oliveira N, Andrade N (2014) A case study of contributor behavior in Q&A site and tags: the importance of prominent profiles in community productivity. *J Braz Comput Soc* 20(1):5
- Godoy D, Amandi A (2006) Modeling user interests by conceptual clustering. *Inf Syst* 31(4):247–265
- Guo L (2013) Semantically enhanced topic modeling and its applications in social media. Drexel University, Philadelphia
- Horowitz D, Kamwar S (2010) The anatomy of a large scale social search engine. Proc WWW, NC, USA, pp 431–440
- Horowitz D, Kamvar SD (2012) Searching the village: models and methods for social search. *Commun ACM* 55:111–118
- Huna A, Srba I, Bielikova M (2016) Exploiting content quality and question difficulty in CQA reputation systems. In: International conference and school on network science. Springer, Berlin
- Java A, Kolari P, Finin T, Oates T (2006) Modeling the spread of influence on the blogosphere. Proceedings of the 15th international world wide web conference. Edinburgh, UK
- Jurczyk P, Agichtein E (2007) HITS on question answer portals: an exploration of link analysis for author ranking. SIGIR, Amsterdam
- Kao W-C, Liu D-R, Wang S-W (2010) Expert finding in question-answering websites: a novel hybrid approach. In: Proceedings of the 2010 ACM symposium on applied computing (SAC '10). ACM, New York, NY, USA, pp 867–871. <https://doi.org/10.1145/1774088.1774266>
- Krulwich B, Burkey C, Consulting A (1996) The ContactFinder agent: answering bulletin board questions with referrals. Proceedings of the thirteenth national conference on Artificial intelligence, vol 1. pp 10–15, August 04–08AAAI/IAAI
- Lappas T, Liu K, Terzi E (2011) A survey of algorithms and systems for expert location in social networks. In: Aggarwal C (eds) Social network data analytics. Springer, Boston, MA, pp 215–241
- Lempel R, Moran S (2001) SALSA: the stochastic approach for link-structure analysis. *ACM Trans Inf Syst.* <https://doi.org/10.1145/382979.383041>
- Li B, King I (2010) Routing questions to appropriate answerers in community question answering services. In: Proceedings of the 19th ACM international conference on information and knowledge management, ACM
- Li W, Eickhoff C, de Vries AP (2016) Probabilistic local expert retrieval. In: Ferro N et al (eds) Advances in information retrieval. ECIR 2016. Springer, Cham
- Lin S, Hong W, Wang D, Li T (2017) A survey on expert finding techniques. *J Intell Inf Syst* 49(2):255–279
- Liu X, Bollen J, Nelson ML, Van de Sompel H (2005) Co-authorship networks in the digital library research community. *Inf Process Manag* 41(6):1462–1480
- Lu Y, Quan X, Lei J, Ni X, Liu W, Xu Y (2012) Semantic link analysis for finding answer experts. *J Inf Sci Eng* 28(1):51–65
- Liu X, Ye S, Li X, Luo Y, Rao Y (2015) ZhihuRank: a topic-sensitive expert finding algorithm in community question answering websites. In: Li F, Klamma R, Laanpere M, Zhang J, Manjón B, Lau R (eds) Advances in web-based learning – ICWL 2015, vol 9412. Springer, Cham
- Macdonald C, Hannah D, Ounis I (2008) High quality expertise evidence for expert search. In: European conference on information retrieval. Springer, Berlin
- Maybury MT (2006) Expert finding systems, technical report MTR06B000040. MITRE Corporation
- Neshati M, Hiemstra D, Asgari E, Beigy H (2014) Integration of scientific social networks. *World Wide Web* 17(5):1051–1079
- Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: bringing order to the web. Stanford InfoLab, Stanford
- Pal A, Counts S (2011) Identifying topical authorities in microblogs. In: Proceedings of the fourth ACM international conference on web search and data mining, ACM
- Pal A, Konstan JA (2010) Expert identification in community question answering: exploring question selection bias. In: Proceedings of the 19th ACM international conference on information and knowledge management, ACM
- Pal A, Chang S, Konstan JA (2012) Evolution of experts in question answering communities. ICWSM, Washington, USA
- Petkova D, Croft WB (2008) Hierarchical language models for expert finding in enterprise corpora. *Int J Artif Intell Tools* 17(01):5–18
- Procaci TB, Nunes BP, Nurmikko-Fuller T, Siqueira SW (2016) Finding topical experts in question & answer communities. In: Advanced learning technologies (ICALT), 2016 IEEE 16th international conference on, IEEE
- Qu M, Qiu G, He X, Zhang C, Wu H, Bu J, Chen C (2009) Probabilistic question recommendation for question answering communities. In: Proceedings of the 18th international conference on world wide web, ACM
- Riahi F, Zolaktaf Z, Shafiei M, Milios E (2012) Finding expert users in community question answering. In: Proceedings of the 21st International conference on world wide web, ACM
- Rode H, Serdyukov P, Hiemstra D, Zaragoza H (2007) Entity ranking on graphs: studies on expert finding. Technical report TR-CTIT-07-81, centre for telematics and information technology, University of Twente, Enschede
- Sahu TP, Nagwani NK, Verma S (2016a) Multivariate beta mixture model for automatic identification of topical authoritative users in community question answering sites. *IEEE Access* 4:5343–5355
- Sahu TP, Nagwani NK, Verma S (2016b) TagLDA based user persona model to identify topical experts for newly posted questions in community question answering sites. *Int J Appl Eng Res* 11(10):7072–7078
- Seo J, Croft W (2009) Thread-based expert finding. In: Agichtein E, Hearst M, Soboroff L (eds) SIGIR 2009 workshop on search in social media. ACM, Citeseer
- Serdyukov P, Rode H, Hiemstra D (2008) Modeling multi-step relevance propagation for expert finding. In: Proceedings of the 17th ACM conference on information and knowledge management, ACM
- Shalev-Shwartz S, Ben-David S (2014) Understanding machine learning: from theory to algorithms. Cambridge University Press, Cambridge
- Smirnova E, Balog K (2011) A user-oriented model for expert finding. In: European conference on information retrieval. Springer, Berlin
- Souza C, Magalhães J, Costa E, Fechine J (2013) Social query: a query routing system for twitter. In: Proceedings of 8th international conference on internet and web applications and services (ICIW)
- Sung J, Lee J-G, Lee U (2013) Booming up the long tails: discovering potentially contributive users in community-based question answering services. Proc 7th int AAAI Conf. Weblogs Social Media, 2013
- Weng J, Lim E-P, Jiang J, He Q (2010) Twittrrank: finding topic-sensitive influential twitterers. Proceedings of the third ACM

- international conference on Web search and data mining, ACM, New York City
- Wang GA, Jiao J, Abrahams AS, Fan W, Zhang Z (2013) ExpertRank: a topic-aware expert finding algorithm for online knowledge communities. *Decis Support Syst* 54(3):1442–1451
- Wang X, Huang C, Yao L, Benatallah B, Dong M (2018) A survey on expert recommendation in community question answering. *J Comput Sci Technol* 33(4):625–653
- Xuan H, Yang Y, Peng C (2013) An expert finding model based on topic clustering and link analysis in CQA website. *J Netw Inf Secur* 4(2):165–176
- Yang B, Manandhar S (2014) Exploring user expertise and descriptive ability in community question answering. In: *Advances in social networks analysis and mining (ASONAM), 2014 IEEE/ACM international conference on, IEEE*
- Yang RR, Wu JH (2014) Study on finding experts in community question-answering system. In: Liu DL, Zhu XB, Xu KL, Fang DM (eds) *Applied mechanics and materials*. Trans Tech Publications, Switzerland
- Yang J, Adamic LA, Ackerman MS (2008) Competing to share expertise: the taskcn knowledge sharing community. *ICWSM*, Washington, USA
- Yang L, Qiu M, Gottipati S, Zhu F, Jiang J, Sun H, Chen Z (2013) Cqarank: jointly model topics and expertise in community question answering. In: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management, ACM*
- Yang J, Bozzon A, Houben G-J (2015) E-WISE: an expertise-driven recommendation platform for web question answering systems. In: *International conference on web engineering*. Rotterdam, The Netherlands
- Yimam-Seid D, Kobsa A (2003) Expert-finding systems for organizations: problem and domain analysis and the DEMOIR approach. *J Organ Comput Electron Commerce* 13(1):1–24
- Zhang J, Ackerman MS, Adamic L, Nam KK (2007a) QuME: a mechanism to support expertise finding in online help-seeking communities. In: *Proceedings of the 20th annual ACM symposium on user interface software and technology, ACM*
- Zhang J, Tang J, Li J (2007b) Expert finding in a social network. In: Kotagiri R, Krishna PR, Mohania M, Nantajeewarawat E (eds) *Advances in databases: concepts, systems and applications*, vol 4443. Springer, Berlin, Heidelberg
- Zhang J, Tang J, Liu L, Li J (2008) A mixture model for expert finding. *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin
- Zheng X, Hu Z, Xu A, Chen D, Liu K, Li B (2012) Algorithm for recommending answer providers in community-based question answering. *J Inf Sci* 38(1):3–14
- Zhou Y, Cong G, Cui B, Jensen CS, Yao J (2009) Routing questions to the right users in online communities. In: *Data engineering, 2009. ICDE'09. IEEE 25th international conference on, IEEE*
- Zhou TC, Lyu MR, King I (2012a) A classification-based approach to question routing in community question answering. In: *Proceedings of the 21st international conference on world wide web, ACM*
- Zhou G, Lai S, Liu K, Zhao J (2012b) Topic-sensitive probabilistic model for expert finding in question answer communities. *Proceedings of the 21st ACM international conference on information and knowledgemanagement, HI, USA*
- Zhao Z, Zhang L, He X, Ng W (2015) Expert finding for question answering via graph regularized matrix completion. *IEEE Trans Know Data Eng* 27(4): 993–1004
- Zhao Z, Yang Q, Cai D, He X, Zhuang Y (2016) Expert finding for community-based question answering viaranking metric network learning. *Proc 7th int joint conf artif intell*, pp 3000–3006
- Zhu X (2011) Semi-supervised learning. In: *Encyclopedia of machine learning*. Claude Sammut and Geoffrey Webb editors *Encyclopedia of Machine Learning*, Springer, pp 892–897
- Zhu J, Huang X, Song D, Rürger S (2010) Integrating multiple document features in language models for expert finding. *Knowl Inf Syst* 23(1):29–54
- Zolaktaf Zadeh Z (2012) Probabilistic modeling in community-based question answering services. Thesis, Dalhousie University Halifax, Nova Scotia. <https://pdfs.semanticscholar.org/f772/f526403f97bd2e81b92f8c926d980e5aa2a5.pdf>