TOWARDS A MULTI-FEATURE ENABLED APPROACH FOR OPTIMIZED EXPERT SEEKING

Mariam Abdullah[1], Hassan Noureddine[1], Jawad Makki[1], Hussein Charara[1], Hussein Hazimeh[2], Omar Abou Khaled[2], Elena Mugellini[2]

[1]Lebanese University, Beirut, Lebanon
[2]HES-SO/FR, Fribourg, Switzerland

ABSTRACT

With the enormous growth of data, retrieving information from the Web became more desirable and even more challenging because of the Big Data issues (e.g. noise, corruption, bad quality…etc.). Expert seeking, defined as returning a ranked list of expert researchers given a topic, has been a real concern in the last 15 years. This kind of task comes in handy when building scientific committees, requiring to identify the scholars' experience to assign them the most suitable roles in addition to other factors as well. Due to the fact the Web is drowning with plenty of data, this opens up the opportunity to collect different kinds of expertise evidence. In this paper, we propose an expert seeking approach with specifying the most desirable features (i.e. criteria on which researcher's evaluation is done) along with their estimation techniques. We utilized some machine learning techniques in our system and we aim at verifying the effectiveness of incorporating influential features that go beyond publications.

KEYWORDS

Entity retrieval, Expert seeking, Academia, Information extraction

1. INTRODUCTION

Currently, the Web is in the state of always being active since lots of data are being uploaded constantly. As a result, this has caused the Web to witness so much interactions between different organizations (company, university…etc.) for they weren't satisfied with their own knowledge. For this reason, they used the Web to connect to the outside world for development and improvement (socially, scientifically…etc.). However, despite the interactions' usefulness, they couldn't help but drown the Web with plenty of data, recognized as Big Data which is a very common term nowadays. Retrieving information from the Web is classified as non-trivial for this data is likely to contain noise with no guarantees of good quality. One of the things that's been frequently searched for is experts; seeking for experts is defined as the task of taking the user's query as input and generating a ranked list of expert researchers as output. The query denotes the topic of expertise and the generated list is sorted according to their expertise levels in what concerns the query topic. Fifteen years ago, the scientific community showed its interest in this

task and since then, they became highly dedicated to this domain. In spite of the significance this task possesses, some still wonder "wouldn't it be easier if we counted on human recommendations?" Yet, the fact human judgments might not be based on reasonable criteria answers the aforementioned question.

To estimate a scholar's expertise degree, the key idea is the set of characteristics on which the evaluation is done, i.e. features and the way of estimating them where recent works have focused on scholar's academic publications to extract different features in addition to detecting his relations. As a matter of fact, none took notice of the activities a scholar has done beyond publishing, for instance being in a conference committee, being honored and awarded, his seminars and courses, etc... Incorporating these supplementary features means dealing with more data which sounds a bit more challenging, because of the formerly stated data issues. We aim at verifying that going beyond publications enhances the expertise retrieval performance. Among the non-traditional features, we picked the conference committee evidence, because we believe it is substantial to study how good a scholar is at assessing other scholars' academic works. Concerning a scholar's publications, we will be considering the conferences' ranks in which they were published for experts usually publish their valuable work in top-ranked conferences. The main contributions of this work can be summarized as follows.

- We developed an expert seeking system by combining traditional and new features with the use of supervised and unsupervised learning.

- Incorporation of conference committee memberships of scholars. To the best of our knowledge, we are the first to deal with beyond publication features.

This paper is organized as follows. In section 2, we review the related works including expertise evidences, expertise retrieval models, approaches that utilized these models, then we give a brief discussion. In section 3, we present the proposed approach comprising of architecture and procedure with the desirable features. In section 4, we give the experimental evaluation. In section 5, we give the concluding remarks and future research directions.

2. RELATED WORK

Expertise seeking is an information retrieval task concerned with the search for the most knowledgeable people in a specific research topic. This task involves taking a user's query and returning a list of people sorted by their level of expertise regarding the query topic. In order to rank experts, we must evaluate them according to high-impact criteria termed features. They are extracted from different sources on the Web, for example from publications, one can extract the h-index, number of publications, number of citations...etc. It is also possible to consider other features collected outside publications, for instance a scholar's awards, his courses…etc. All of these features are capable of indicating a scholar's expertise but not with equivalent proportions.

Different models were proposed in this context; Generative, Voting, Graph-based and Learning to Rank models. Generative models rank candidate experts according to the likelihood of a person being an expert on the given query [1]. A researcher is evaluated through his documents, either through language models (LM) which look for the occurrence of query words, i.e. it uses terms to represent the content of publications or through topic models [[2], [3]] that detect the semantics of documents since it learns the topics found in the corpus with its related words through

unsupervised learning [[4], [5]]. In voting models, documents ranked with respect to a query vote for candidate experts in which they are mentioned [1]. Some voting methods are also applied in the aspect of rank aggregation [6]. Graph models are supplied with the ability of modeling various relations among people and documents, both implicit and explicit ones. Such relations are well represented by expertise graphs, where both documents and candidate experts become vertices and directed edges symbolize containment conditions. They can handle modeling document-candidate, document-document, and candidate-candidate associations [7]. Most of the previous models are not aware of handling heterogeneous features which probably needs a well-determined technique and learning to rank was there to fill such a prominent gap. Even though, voting models were somehow useful in the aforementioned task, but they only provided an acceptable result because regardless of data, the features are given the same weights. Learning to rank applies machine learning strategies to learn the ranking model from data in a supervised manner. In other terms, data is free to speak for itself rather than making assumptions regarding the model [8].

Some considered generative models in their work; H. Deng et al. [9] proposed a hybrid model by combining a weighted language model and a topic-based model. E. Smirnova et al. [10] proposed a user-oriented approach that balances two factors that influence the user's choice: time to contact an expert, and the knowledge value gained after. J. Tang et al. [11] proposed three generative topic models by introducing the conference information into the author topic model. Concerning voting models, D. Eddine Difallah et al. [12] developed a system capable of selecting which workers should perform a given task based on worker profiles extracted from social networks by considering the information extracted from the task descriptions and categories liked by the user on social platforms. Others applied graphs models; Zhou et al. [13] proposed a topic-sensitive probabilistic model, an extension of PageRank, which considers both the link structure and the topical similarity among users. J. Zhang et al. [14] introduced a propagation-based approach which estimates researchers' initial scores through their local information, then expertise of researchers is propagated towards the ones he co-authored with based on the intuition if a person knows many experts on a topic or his name co-occurs many times with another expert, then he is likely an expert. A. Kardan et al. [15] proposed a model for expert finding on social networks where people in social networks are represented instead of web pages in PageRank. Recent works have concentrated on learning to rank models; V. Kavitha et al. [16] combined multiple features of research expertise to rank experts; time weighted citation graph by giving significance to recent publications of an author and modified LDA to cope up with newly generated publication terms; Z. Yang et al. [17] used a supervised learning model with features including language models, author-conference-topic model, and other ones. C. Moreira et al. [18] proposed some features; academic indexes, regarding the textual content, Okapi BM25, TF and IDF were suggested in addition to some profile information; Sorg et al. [19] proposed a discriminative model that allows the combination of different sources of evidence in a single retrieval model using Multi-Layer Perceptron (MLP) and Logistic Regression as regression classifiers. The considered features were derived from language models, standard probabilistic retrieval functions and features quantifying the popularity of an expert in the question category.

2.1. Discussion

The previous works' analysis made us detect two key concepts which building an expert finding system relies on; the quality of the features incorporated into the system, i.e. their strengths of indicating the researchers' proficiency, and the way of exploiting them. As a matter of fact, the related work approaches were devoted towards the researcher's academic work, something he is

producing academically, i.e. publications, denoting his contributions and the topical domains he appears to be interested in. Yet, not only this kind of information it offers, but also one can infer the publications content by interpretation and analysis, their statistics and his relations. Apart from publications, a researcher tend to take part in activities revealing to some extent his proficiency level.
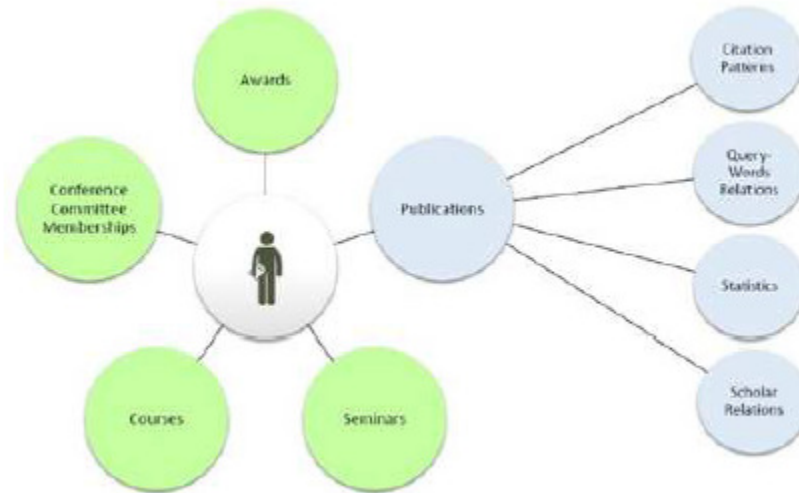

Figure 1. Scholar's Expertise Indicators

Figure 1 shows a scholar's expertise indicators. A scholar might be misjudged when evaluated only through his publications, for there are other unnoticed signs indicative of his expertise for example giving seminars, being part of conference committees, teaching courses and being honored and awarded. We aim at incorporating the conference memberships for it is important to study how good a scholar is at evaluating other scholars' academic work, where the classical case ranks him only based on his own academic work. Our proposed approach uses learning to rank for its ability of combining heterogeneous features optimally in a supervised manner; in addition to using the author topic model (generative model) to estimate the relation between author and topic through unsupervised learning. The reason why we picked this model is because: publications' abstracts are not always available which makes it pretty hard to know the topics of a publication when using LDA, i.e. authors' information are not included, and it links authors to topics in a semantic manner.

3. THE PROPOSED APPROACH

We aim at building an expert seeking system with the purpose of enhancing the retrieval effectiveness through adding non-traditional features by taking advantage of conference memberships following the intuition that such positions are not assigned spontaneously. Additionally, we believe that a scholars' publications conference rank, to some extent, matter; for experts usually publish their valuable work in top-ranked conferences. The more he publish in top-ranked conferences, the more valuable his work is. Our system is called "Multi-Feature Based Expert Seeking System" (FeBES) because it uses multiple features extracted from different sources to estimate scholars' expertise. In this section, the proposed model's concepts are provided, including the model's architecture, procedure and desirable features on which we will count to determine how professional scholars are given a domain.

3.1. Architecture

Figure 2 shows the architecture of the proposed approach, it is a learning to rank architecture. The approach is a mixture of models; it uses learning to rank to combine multiple features, and uses generative model (author topic model) to estimate the relation between author and query through unsupervised learning to link topics with document words and authors.
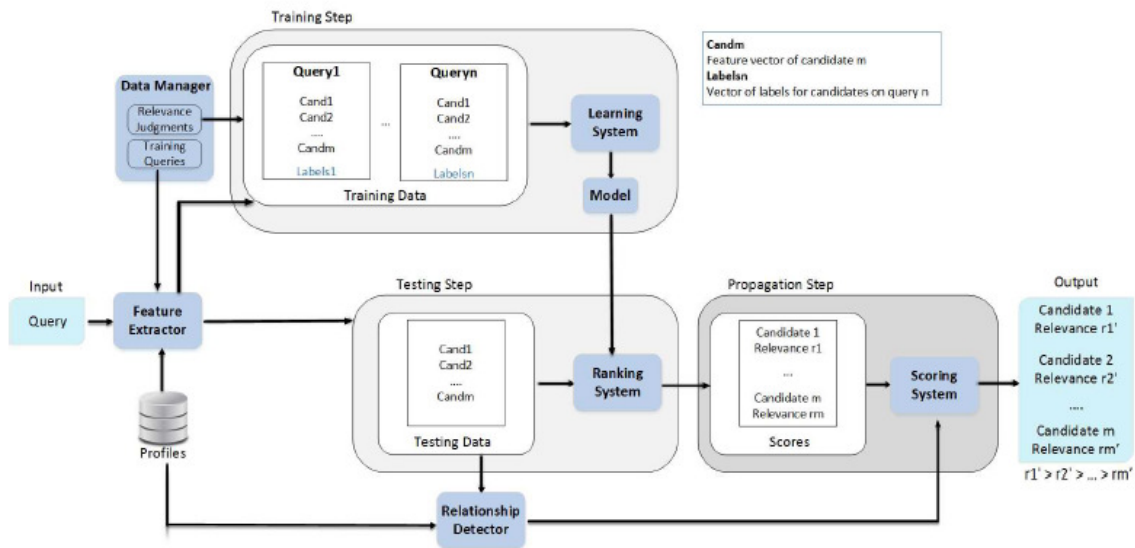

Figure 2. Our Approach Architecture

3.2. Procedure

The input represents the topic of interest submitted by the user. The set of profiles belong to researchers where each profile contains personal information and the academic work done by the researcher, having the following attributes: name, affiliation, e-mail, location, homepage, summary about his career and a list of all his publications. These profiles are generated in [20] through correlating the information extracted from heterogeneous sources, by taking advantage of data repetition in multiple sources and those existing in one source. On one hand, data is being validated and links between sources are created. On the other hand, the missing data issue is solved, and are saved in the repository serving as the system's internal input. The data manager is in charge of maintaining the training queries with relevance judgments of researchers with respect to those queries. Features are extracted based on the researchers' profiles by the feature extractor, they are of two types; query-independent and query-dependent. The input into the learning system is the training instances. Each training instance refers to a query associated with feature vectors resulting from all features for researchers along with their relevance judgments with respect to that query. This component generates the learned ranking model, comprising of features' weights, which serves as an input to the ranking system to handle all user-submitted queries. The ranking system receives the user's input query, then retrieves the candidate experts from the set of all researchers. Afterwards, feature vectors for those candidates are built. Finally, the learned model is applied on the constructed feature vectors to generate a score for each candidate. The output displayed to the user is a ranked list of candidate experts by sorting the scores generated by the ranking system.

3.3. Features

Features are of two types: query-independent and query-dependent. Features belonging to the former category maintain the same value regardless the query topic, however those belonging to the latter one changes as the query changes. We will mention the included features and how to estimate their values.

Query-Independent Category

We have cautiously picked the features which seem to mostly influence a researcher's expertise. The considered features are; h-index, i10-index, total number of citations, total number of publications, number of publications in recent years, total number of memberships in conferences, and conferences' rank. We believe experts are more likely to publish papers in high-ranked conferences. It can be estimated in the following way:

$$\left(\sum\nolimits_{rank} W_{rank} * n_{pubs\,(rank)}\right)/n_t \tag{1}$$

Where rank represents a conference rank and wrank is the weight of this rank. Parameter npubs(rank) is the number of papers published in conferences having this rank. The overall score is divided by the total number of publications of the scholar nt because we are into knowing the distribution of all publications over these ranks.

Query-Dependent Category

We relied on the relation between query and author P(q, a) through Author-Topic model. The more he has published on-topic publications, the stronger P(q, a) is, because as a result, he would be assigned more to on-topic words. The following equation was applied

$$P(q, a) = \sum\nolimits_{wi} \sum\nolimits_{t} P(w_i|t)\, P(a|t)\, P(t) \tag{2}$$

Where w_i is a query word, t is a topic among the set of all topics. $P(w_i|t)$ is the probability of word w_i given topic t, $P(a|t)$ is the probability of author a given topic t and $P(t)$ is the prior probability of topic t.

We also considered the scholar's conference memberships. The intuition is that scholars who very often take part as members in conferences should be thought of as more valuable due to the fact such positions are not assigned spontaneously. Our concerns include how often he is a part of conference committees, how connected these conferences are to the query terms and how dominant these relevant participations are, relative to all other ones. Though the latter point may sound odd, most experts are fully devoted to one domain and partially to related ones. We also counted on the conference rank based on its importance. The following formula is inspired from the document model, but it has been applied only on documents, we projected it to our context, where documents were replaced by conferences and we introduced another value which influences the whole formula.

$$P(q|a) = \alpha \sum\nolimits_{c} P(q|c)\, P(a|c)\, P(c) \tag{3}$$

Where q and a represent the query and author respectively, and c is a conference. P(q|a) is the probability of query q given author a, P(a|c) is the probability of author a given conference c, P(q|c) is the probability of query q given conference c and P(c) is the prior probability of conference c. P(a|c) is either zero or one depending on whether or not author a is in the committee of conference c. P(c) depends on the conference c rank. To estimate P(q|c), conference c needs to be represented as the collection of documents that published in conference c with the use of language and author topic models.

$$P(q|c) = \Sigma_d\, P(q|d) = \Sigma_d \prod_{wi} P(w_i|d) = \Sigma_d \prod_{wi} P_{LM}(w_i|d) \times P_{AT}(w_i|d) \qquad (4)$$

$$P_{AT}(w|d) = \Sigma_t\, \Sigma_a\, P(w|z)\, P(z|a)\, P(a|d) \qquad (5)$$

Where t is a topic and a is an author of document d. regarding the author topic model. P(w|z) is the probability of word w given topic z, P(z|a) is the probability of topic z given author a and P(a|d) defines the relation between author a and document d.

$$P_{LM}(w|d) = (1 - \lambda)\, P(w|d) + \lambda\, P(w) \qquad (6)$$

Regarding the language model, P(w|d) is the probability of word t in document d, P(w) is the probability of t in the whole corpus and λ is a smoothing parameter. As for α, it represents the dominance degree of on-topic memberships. To apply this, we have to estimate P(q|c) for each conference and decide, based on a threshold value, whether or not conference c is highly related to query q.

$$\alpha = n/n' \qquad (7)$$

where n is the number of strongly on-topic participations and n' is the number of all participations including relevant and non-relevant ones.

4. EXPERIMENTATION & EVALUATION

4.1. Implementation & Preprocessing

We have implemented the various system elements, and thus provided web interface for receiving user requests and respond with relevant results. The prototype of our architecture is implemented using JavaEE, where all the tests were performed on 2.60 GHz Processor, 8GB of RAM PC. Additionally, we used Matlab to apply "author-topic model", which is an unsupervised learning technique. Our system is portable and domain-independent, i.e. it can be applied on any dataset regardless of its domain. To test our approach, we have chosen a Computer Science dataset with 297 researchers including around 215 experts from 7 topics.

With regard to author topic model, we considered 54,000 publications. The dataset has undergone some preprocessing (insignificant authors removal, associated publications removal, stop words and insignificant words removal) and the number of topics was set to 80 and number of iterations to 1000. We were left with 8,534 authors, 31,114 publications, 19,607 vocabulary words and 1,640,069 word tokens. The output is a list of learned topics having topic-author and topic-word proportions

We used SVMRank to learn the model and considered three queries. We used the Arnetminer's evaluation dataset comprising of seven queries. For each query, we labeled experts list with 1, and complemented it with equivalent number of non-experts (labeled as 0), containing easy and hard examples. Researchers' feature vectors undergone some preprocessing as well by removing the null and conflicted ones and finally normalizing them.

4.2. Results & Evaluation

Table 1 shows the learned model comprising of the features' weights. The obtained results reveal that conference committee memberships has the biggest impact among all other features, for experts are members of committees just for they have the required expertise. Another high influential information is the h-index, because it represents the scientific impact of his publications on other scholars, it also signifies his productivity. The relation between query and author through author-topic model is significant as well, because this value is based on the author's publications, the more he is assigned to query topic-related words in his publications, the more he is likely one of the experts on the query topic. As for conference ranks, the other introduced criterion, showed that it does influence the scholar's expertise but not in same degrees as the previous ones. Even the number of publications has not a very big impact, because it's not always about the quantity of publications, it's more about the quality. The weight for number of citations has a negative impact, even though the h-index had the opposite case. Well, this is because two scholars might have the same total number of citations, but the way they are distributed over their publications may vary (number of publications may vary too) and when that happens, the h-index varies as well. The number of publications in recent years (2009 and above) showed that it's not important for a person to have published a lot of papers in the recent 6 years, the whole career counts more than these last years. I10-index's weight is the smallest one, and it is that way perhaps because h-index almost contains the i10-index concept, and the former highly influences the expertise.

Table 1. Features Weights

Feature	Weight
# of pubs	0.810033
# of recent pubs	-1.703149
# of citations	-0.440943
H-index	3.464950
I10-index	-1.323053
Conference ranks	1.118827
Author-topic model	2.461189
Conference committee	4.592299

For the evaluation, we used Arnetminer's new dataset (7 queries). Precision at 5, 10, 15, 20 (P@5, P@10, P@15, P@20) and the Mean Average Precision (MAP) are the considered evaluation metrics. Tests were done on two approaches, a baseline approach and our approach where our approach = baseline + {conference rank, conference committee}. The figures 2, 3 and 4 below show the performance of the baseline approach and our approach in Information Extraction, Machine Learning and Intelligent Agents respectively.

In figure 3, we can clearly see the gap between the two approaches. When n was 5, the baseline precision was very low (20%), whereas that of our approach was 60%. Even though when n

became 10, the precision of the former improved and that of the latter declined, but the latter still had a better precision. Later on, both precisions dropped with our approach outperforming the baseline approach. In figure 4, we can notice that baseline approach and our approach have the same initial value at p@5, but when n increased to 10, the baseline precision decreased, however that of our approach maintained the same value. Until n became 20, our approach outperformed the baseline approach. In figure 5, both approaches had the same initial precision value. And when the approaches' precision began to reduce, our approach still outperformed the baseline approach.
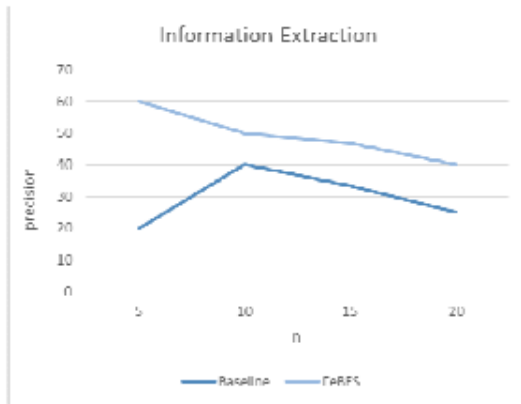

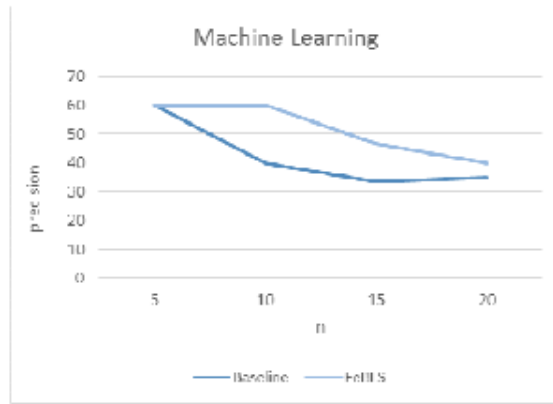Figure 3. Comparison between Baseline and Our Approach in Information Extraction

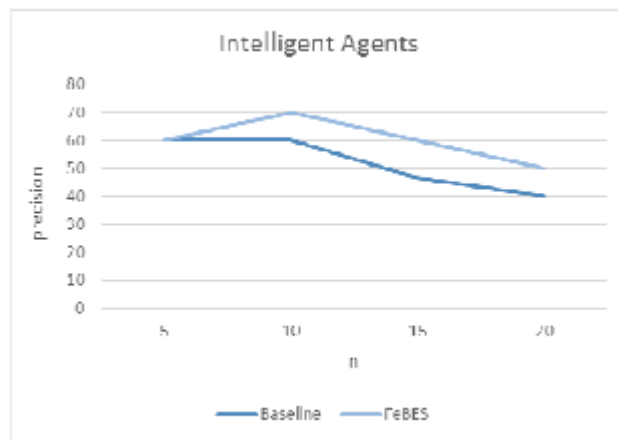Figure 4. Comparison between Baseline and Our Approach in Machine Learning

Figure 5. Comparison between Baseline and Our Approach on Intelligent Agents

Table 2 summarizes the three queries (by taking the average), and estimates MAP for both approaches. We can easily tell the difference between baseline approach and our approach. At the four n values, the latter had a better performance and a more effective result than that of the former. When n was 5, 10 and 15, the difference was around 14%. As for the MAP, the results showed our approach outperformed the baseline by 18%. The explanation behind such results is because publications don't cover the same evidence as that of conference memberships, where the latter helps distinguish the significant researchers according to the topic they are experienced in

based on being a member in conferences. In addition to the papers' conference ranks, where the quality of a scholar's papers is considered, so it is not only about the quantity of these papers but also their quality and value.

Table 2. Comparison between Baseline and Our Approach

	P @ 5	**P @ 10**	**P @ 15**	**P @ 20**	**MAP**
Baseline Approach	46.7%	46.7%	37.9%	33.4%	39.1%
Our Approach	60%	60%	51.2%	43.4%	57.5%

5. CONCLUSION & FUTURE WORK

The previous results show that our approach has 57.5% as mean average precision whereas 39.1% for the baseline. Concerning the three queries, our approach outperforms the baseline. Based on these results, we noticed that including new kind of expertise evidence into expert seeking came in handy because they happen to represent a different aspect of expertise. Therefore, we conclude that a scholar's expertise should be assessed not only through his academic publications, but also through external activities he is involved in because ignoring such an aspect might cause misjudgment.

As future plans, we aim at verifying the effectiveness of other beyond publication features including awards, courses, and seminars. Not to forget to mention that applying the propagation phase, known as propagating one's expertise based on those he has relations with, has shown enhancement as stated in the related work section. For this reason, we intend to do a combination by adding the propagation phase to our approach to improve the retrieval mperformance. We also would like to prove the efficiency of orienting this phase towards time and distinguishing between co-author and supervision relations. Moreover, we believe it is preferable to distinguish between the different roles a scholar is assigned when he is in a conference committee, because he could possibly be in a scientific committee, or in a program committee or even a reviewer.

REFERENCES

[1]   K. Balog, Y. Fang, M. d. Rijke, P. Serdyukov and L. Si, Expertise Retrieval, Foundations and Trends in Information Retrieval, Vol. 6: No. 2–3, pp 127-256., 2012.

[2]   C. Moreira, "Learning to rank academic experts," Master Thesis, 2011.

[3]   C. Zhai, "Statistical Language Models for Information Retrieval A Critical Review," in Foundations and Trends in Information Retrieval, vol. 2, 2008, p. 137–213.

[4]   M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth and M. Steyvers, "Learning Author- Topic Models from Text Corpora," in ACM Transactions on Information Systems (TOIS), 2010.

[5]   K. Balog and M. d. Rijke, "Associating people and documents," in Proceedings of the European Conference on IR Research, (ECIR '08), Berlin, Heidelberg, 2008.

[6]   C. Moreira, P. Calado and B. Martins, "Learning to Rank Experts in Academic Digital Libraries," in 15th Portuguese Conference on Artificial Intelligence, EPIA, Lisbon, Portugal, 2011.

[7]   P. Serdyukov, H. Rode and D. Hiemstra, "Modeling multi-step relevance propagation for expert finding," in Proceeding of the ACM International Conference on Information and Knowledge Management, New York, NY, USA, 2008.

[8]   C. Moreira, "Learning to rank academic experts," Master Thesis, 2011.

[9]   H. Deng, I. King and M. R. Lyu, "Formal models for expert finding on DBLP bibliography data," in Proceedings of the IEEE International Conference on Data Mining, Washington, DC, USA, , 2008.

[10] E. Smirnova and K. Balog, "A User-Oriented Model for Expert Finding," in LNCS 6611, 2011.

[11] J. Tang, J. Zhang, J. L. L. Yao, L. Zhang and Z. Su, "Arnetminer: Extraction and mining of academic social network," in Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD '08, New York, NY, USA, 2008.

[12] D. E. Difallah, G. Demartini and a. P. Cudré-Mauroux, "Pick-A-Crowd: Tell Me What You Like, and I'll Tell You What to Do, A Crowdsourcing Platform for Personalized Human Intelligence Task Assignment Based on Social Networks," in WWW, 2013.

[13] G. Zhou, S. Lai, K. Liu and J. Zhao, "Topic-Sensitive Probabilistic Model for Expert Finding in Question Answer Communities," in CIKM, Maui, HI, USA, 2012.

[14] J. Zhang, J. Tang and a. J. Li, "Expert Finding in a Social Network," in 12th International Conference on Database Systems for Advanced Applications, Bangkok, Thailand, 2007.

[15] A. Kardan, A. Omidvar and Farahmandnia, "Expert finding on social network with link analysis approach," in Electrical Engineering (ICEE), 19th Iranian Conference, 2011.

[16] V. Kavitha, G. Manju and T. Geetha, "Learning to Rank Experts using combination of Multiple Features of Expertise," in Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference, 2014.

[17] M. n. uddin, t. h. duong, K.-j. oh, j.-g. jung and g.-s. jo, "experts search and rank with social network: an ontology-based approach," International Journal of Software Engineering and Knowledge Engineering, vol. 23, no. 1, 2013.

[18] M. Rosen-Zvi, T. Griffiths, M. Steyvere and P. Smyth., "The author-topic model for authors and documents," in UAI '04 Proceedings of the 20th conference on Uncertainty in artificial intelligence.

[19] P. Sorg and P. Cimiano, "Finding the Right Expert: Discriminative Models for Expert Retrieval, in Proceedings of the International," in Conference on Knowledge Discovery and Information Retrieval (KDIR), Paris, France, 2011.

[20] H. Noureddine, I. Jarkass, H. Hazimeh, O. A. Khaled and E. Mugellini, " CARP: Correlationbased Approach for Researcher Profiling," in 27th International Conference on Software Engineering and Knowledge Engineering SEKE, 2015.

AUTHORS

Mariam Abdullah is a Software Developer at Mentis Company in Beirut, Lebanon. She was born in Tyre, Lebanon in 1992. She received her Bachelor's degree in Computer Science in 2013 from the Lebanese University, Beirut, Lebanon. She received her Master's degree in Information and Decision Support Systems in 2015 from the Lebanese University in collaboration with university of Applied Sciences of Western Switzerland (HES-SO).
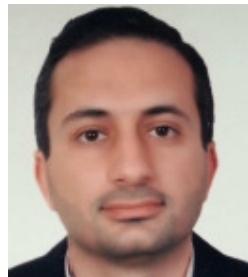
Hassan Noureddine received the Ph.D. degree from the Lebanese University in cooperation with the University Of Applied Sciences Of Western Switzerland (HESSO-FR), Fribourg in 2016. He received Research Master Degree in Signal, Telecom, Image, and Speech (STIP) from the Lebanese University, Doctoral School of Sciences and Technology in 2011. Since 2015, he has been with the American university of education and culture where he is lecturing in Computer Science Studies at the Faculty of Science. His main areas of research interest are Semantic Web & Ontologies, Social Network Analysis, Sentiment analysis, Identity linkage and Information extraction.

Jawad Makki received the Ph.D. degree from the University of Toulouse, France, in 2010. Since 2011, he has been with the Lebanese University where he is lecturing in Computer Science and Information Studies at the Faculty of Science and the Faculty of Information. His main areas of research interest are Semantic Web & Ontologies, Knowledge Representation, NLP & Information Extraction, and Social Network Analysis.

Hussein CHARARA has received an M.S. degree in Computer and Communications Engineering from the Lebanese University, Lebanon in 2002, and an M.S. degree in Networking and Telecommunications from the Institut National Polytehnique (INP-ENSEEIHT), France, in 2003. In 2007, He obtained a Ph.D. degree in Network, Telecom, Systems and Architecture from INP - IRIT, France. From 2006 to 2009 He worked for AIRBUS and THALES AV avionics as a R&D Engineer and PM. He contributed to the implementation & development of the Real time embedded AFDX networks such as for A380, A400M and Soukhoï RRJ programs. From 2009 to 2010, He worked for the ARECS GmbH - München where He was involved in the GPS/Galileo receivers modeling and simulations. Since 2010, He is an assistant professor at the Lebanese University. He is working in several research fields including: Traffic engineering of real time network, Performance Evaluation and QoS, Avionics and wireless sensor networks

Hussein Hazimeh is a PhD student at the University of Fribourg, Fribourg, Switzerland under the supervision of Professor Philippe Cudré-Mauroux, in collaboration with the university of Applied Sciences of Western Switzerland (HES-SO) under the supervision of Professor Omar Abou Khaled and Professor Elena Mugellini. He reviewed his Master 2 from the Lebanese University, Beirut Lebanon, in collaboration with HES-SO. His research interests focus on: Social Media Analytics, Sentiment analysis, Identity linkage and Information extraction.

Omar Abou Khaled is Professor in computer science at HES-SO campus Fribourg (EIA-FR) and member of HumanTech Institute (humantech.eiafr. ch/). He holds a PhD in Computer Science received from the Perception and Automatic Control Group of HEUDIASYC Laboratory of "Université de Technologie de Compiègne", and a Master in Computer Science from the same university. Since 1996 he has been working as research assistant in the MEDIA group of the Theoretical Computer Science Laboratory of EPFL (Ecole Polytechnique Fédérale de Lausanne) in the field of Educational Technologies and Web Based Training research field on MEDIT and CLASSROOM 2000 projects.

He was International Advisor at HES-SO until august 2014. He was Head of Internationale Relations Office at EIA-FR. Head of MRU "Advanced IT Systems Architects" at EIA-FR. Until 2007 He was leader of the MISG (Multimedia Information System Group). He is responsible of several projects in the field of Document Engineering, Multimodal Interfaces, Context Awareness, Ambient Intelligence, Blended Learning, and Content-Based Multimedia Retrieval. He is involved in the teaching of several courses related to Information Systems, Web Technologies & Multimodal Interfaces.

Elena Mugellini is currently Professor at the University of Applied Sciences of Western Switzerland in Fribourg (HES-SO). She holds a Diploma (Bsc and Msc) in Telecommunications Engineering and a Ph.D. in Computer Sciences from University of Florence, Italy. Elena is the leader of HumanTech Institute (Technology for Human well-being, humantech.eia-fr.ch/).Her research expertise lies in Human Computer Interaction (natural interaction, smart spaces, machine learning, serious game and gamification) and Intelligent Data Analysis (multimedia content and knowledge management, semantic technologies, information visualization).

She teaches at graduate and undergraduate level a range of topics, including: project management, human-computer interaction and user experience design, web engineering, software engineering and information systems. Her work is published in journals, conference proceedings and as book chapters and has also been presented in numerous scientific conferences. She is an active reviewer for several international conferences and journals.