

A Methodology for Expert Search on World Wide Web

Mr. Bhaludra R Nadh Singh,
Research Scholar, Dept. of. CSE,
ANU, Nagarjuna Nagar, A.P, India.

Dr. B. Raja Srinivasa Reddy,
Research Guide,
ANU, A.P, India.

Abstract

Academic institutions, software development companies and enterprises in the real world need expert human resources. These organizations generally take help of search engines like Google. The results of the search are millions of records with noise varying quality. Further browsing is required in order to identify expert people. This causes ambiguity and time consuming. Recently Guan et al. studied a general expert search problem and provided a solution using co-occurrence based diffusion. In this paper we implement a web based search application that makes use of heat diffusion technique to identify web resources that have details about human experts. Ranking is used for presenting results meaningfully. Our prototype application demonstrates the proof of concept. The empirical results are encouraging.

Index Terms – Heat diffusion, data mining, expert search, and co-occurrence

INTRODUCTION

Searching for human experts has gained attention in research circles and industries. This kind of search is increasing day by day for various fields. Many search engines are able to provide such search mechanism. Many community search engines need registration in order to perform search operations. A general problem with the search engines is that they provide millions of results when a search query is given. The results are confusing, ambiguous and time taking to identify truly experts in the chosen area. Full of noise in the search results make it confusing and the users need to browser further to get the desired results. As World Wide Web has become a place for searching any kind of information, the general search engines throw such problem. Obtaining exact results as expected by end users is a challenging problem. Ana Ivanovic is a famous tennis player. However, the search result in Google has shown her as an expert in swimming as shown in Figure 1.

Ivanović picked up a racket at the age of five after watching [Monica Seles](#), a fellow [Yugoslav](#), on television. She started her career after memorizing the number of a local tennis clinic from an advertisement. At the time, she was forced to train during the morning to avoid bombardments. Later, she admitted that she [trained in an abandoned swimming pool](#) in the winter, as

Figure 1 – Vague expertise evidence

There are many existing solutions that target select community. Organizational expert search is allowed in these applications where user can search for persons with expertise in some area. The search results are not actually reflecting the desired result. Moreover they are specific to a particular field. Obtaining human expert results as per the search query with reasonable quality is the challenging problem to be addressed. Recently in [1] co-occurrence based was implemented to have expert search on the web. Co-occurrences help in finding relevance and reputation of human experts in chosen area. Then ranking based on heat diffusion helped to provide most appropriate results instead of presenting millions of records.

In this paper we implement a web based application that allows expert search based on the idea conceived from [1]. The remainder of the paper is structured as follows. Section II provides review of literature. Section III describes the proposed system. Section IV presents experimental results while section V concludes the paper.

RELATED WORK

Expert search has been around for some years on the web. Human experts with certain skills can be searched over web [2]. Towards this many approaches came into existence. Building profiles of people automatically is one of the approaches as explored in [3] and [4]. TREC enterprise track was built in 2005 for expert search [5]. Model 1 and Model 2 came into existence for expert search [6]. In [7] a new model was built for searching Tilburg University's web site. PageRank was explored in [8] for retrieval performance. Other approaches used for extracting best results from web include non-local evidence [9], [10], query expansion [11] and relevance feedback [12]. Data fusion and voting techniques were explored in [13] for best search performance. A learning framework for discrimination was proposed in [14] for human expert search.

One important observation is that web search results are poor and they are confusing and wasting the time of users. The quality of web search results is quite less than that of community search that targets specific information. In [15] and [16] researches tried to provide mechanisms to identify humans with high expertise.

Matching of multiple aspects concept was implemented in [17] and [18]. Graphs and heat diffusion concepts were tried out in many applications as explored in [19], [20], [21] and [22]. Recently in [1] co-occurrence method is used to identify human experts over web and provide search results accurately.

HEAT DIFFUSION METHOD FOR EXPERT SEARCH

Heat diffusion model constructs a matrix to analyze co-occurrence information and aggregate it for dealing with noises on the web. This method is used to identify high expertise people whose information is available over web. The co-occurrence based heat diffusion model intuitively studies the query string and the names with strong connection to other relevant things. Based on this appropriate ranking is provided thus improving the quality of expert search result. More information can be found on the approach in [1].

EXPERIMENTAL RESULTS

Experiments are made in terms of number of web pages in top domains. The results are also observed to find expert users who appear in the search results. On a locally available input dataset which has been synthesized the experiments are made.

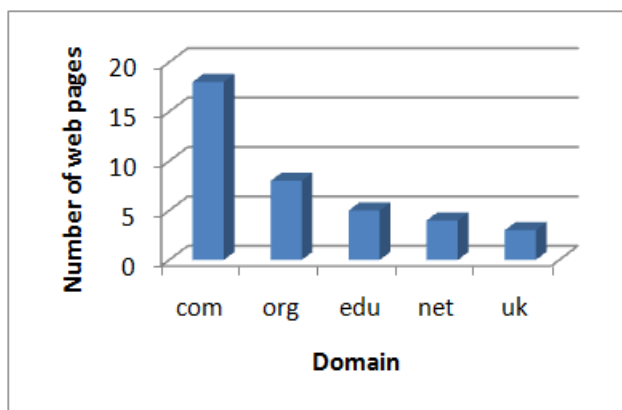


Figure 2 – Domains and number of web pages

When search is carried out using the proposed system, relevant results are presented as output. The summary of search results is presented in Figure 2.



As can be viewed in Figure 3, the summary of search results is presented in. The graph shows expertise of various users.

CONCLUSION AND FUTURE WORK

In this paper we studied the problem of expert search over web. The traditional search engines and even specialized search applications belonging to various domains provide millions of web pages as search results. The results contain noise and irrelevant web pages that are confusing. The web pages may also provide results that provide vague evidence of expertise on various individuals. This is the problem that is difficult to address. Recently Guan et al. [1] presented a model known as co-occurrence based diffusion. According to this model the co-occurrence of the human expert names and various other aspects of the person are considered. A graph is prepared and heat diffusion mechanism is used to rank the results and identify very useful research that reflects users' search intention. We built a prototype application to demonstrate the expert search concepts. The results reveal that it is possible to rank results of expert search to make the search results more meaningful without vagueness. Such results can directly be used by users which contain expert details.

REFERENCES

- [1] Ziyu Guan, Gengxin Miao, Russell McLoughlin, Xifeng Yan, Member, IEEE, and Deng Cai. (2013). Co-Occurrence-Based Diffusion for Expert Search on the Web. IEEE. 25 (5), p1-14.
- [2] P.R. Carlile, "Working Knowledge: How Organizations Manage What They Know," Human Resource Planning, vol. 21, no. 4, pp. 58-60, 1998.
- [3] D. Yimam-Seid and A. Kobsa, "Expert-Finding Systems for Organizations: Problem and Domain Analysis and the Demoir Approach," J. Organizational Computing and Electronic Commerce, vol. 13, no. 1, pp. 1-24, 2003.
- [4] N. Craswell, D. Hawking, A.M. Vercoustre, and P. Wilkins, "P@noptic Expert: Searching for Experts not Just for Documents," Proc. Ausweb Poster, 2001.
- [5] N. Craswell, A.P. de Vries, and I. Soboroff, "Overview of the Trec 2005 Enterprise Track," Proc. Text Retrieval Conf. (TREC), 2005.
- [6] K. Balog, L. Azzopardi, and M. de Rijke, "Formal Models for Expert Finding in Enterprise Corpora," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 43-50, 2006.
- [7] K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch, "Broad Expertise Retrieval in Sparse Data Environments," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information

Retrieval, pp. 551-558, 2007.

[8] J. Zhu, X. Huang, D. Song, and S. Ru" ger, "Integrating Multiple Document Features in Language Models for Expert Finding," Knowledge and Information Systems, vol. 23, no. 1, pp. 29-54, 2010.

[9] P. Serdyukov and D. Hiemstra, "Being Omnipresent to be Almighty: The Importance of the Global Web Evidence for Organizational Expert Finding," Proc. SIGIR Workshop Future Challenges in Expertise Retrieval (fCHER), pp. 17-24, 2008.

[10] K. Balog and M. de Rijke, "Non-Local Evidence for Expert Finding," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 489-498, 2008.

[11] C. Macdonald and I. Ounis, "Expertise Drift and Query Expansion in Expert Search," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 341-350, 2007.

[12] K. Balog and M. de Rijke, "Combining Candidate and Document Models for Expert Search," Proc. 17th Text Retrieval Conf. (TREC), 2008.

[13] C. Macdonald and I. Ounis, "Voting for Candidates: Adapting Data Fusion Techniques for an Expert Search Task," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 387-396, 2006.

[14] Y. Fang, L. Si, and A.P. Mathur, "Discriminative Models of Integrating Document Evidence and Document-Candidate Associations for Expert Search," Proc. 33rd Int'l ACM SIGIR Conf.

Research and Development in Information Retrieval, pp. 683-690, 2010.

[15] K. Balog and M. de Rijke, "Finding Similar Experts," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 821-822, 2007.

[16] J. Zhang, M.S. Ackerman, and L. Adamic, "Expertise Networks in Online Communities: Structure and Algorithms," Proc. Int'l Conf. World Wide Web (WWW), pp. 221-230, 2007.

[17] M. Karimzadehgan and C. Zhai, "Constrained Multi-Aspect Expertise Matching for Committee Review Assignment," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 1697-1700, 2009.

[18] D. Horowitz and S.D. Kamvar, "The Anatomy of a Large-Scale Social Search Engine," Proc. Int'l Conf. World Wide Web (WWW), pp. 431-440, 2010.

[19] H. Ma, H. Yang, M.R. Lyu, and I. King, "Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 233-242, 2008.

[20] H. Yang, I. King, and M.R. Lyu, "Diffusionrank: A Possible Penicillin for Web Spamming," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 431-438, 2007.

[21] R.I. Kondor and J. Lafferty, "Diffusion Kernels on Graphs and Other Discrete Input Spaces," Proc. 19th Int'l Conf. Machine Learning (ICML), pp. 315-322, 2002.

[22] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.