

# Ranking Based Mining for Expert Human Resources on the World Wide Web

K. Ashesh<sup>1</sup>, G. Vamsi Krishna<sup>2</sup>

<sup>1,2</sup> Gitam University, Department of Computer Science, Visakhapatnam, India

**Abstract:** *In all walks of life, human resources play an important role in the success of any organization in the real world. Finding expert human resources is an open problem for enterprises besides retaining good resources to excel in chosen business. Web Search engines like Google can help in finding expert human resources. However, in such case, the results are in millions of records with varying quality and noise. Browsing search results is again time consuming as the identification of suitable resources is deemed to be tedious provided the voluminous search results. Recently Guan et al. presented an approach based on a ranking model which proved to be effective in finding expert human resources with ease. Influenced by this work, in this paper, we implement a prototype application with web interface that uses a ranking technique in order to provide best results that are interested to end user. Our application demonstrates the proof of concept and the empirical results show encouraging results.*

**Keywords:** Heat diffusion, data mining, expert search, ranking and co-occurrence

## 1. Introduction

There is no alternative for human experts as they are knowledgeable and instrumental for the success of enterprises in the real world. Naturally there is tendency in retaining expert resources so as to achieve goals. There is talent hunt that can bestow plethora of advantages to businesses. In this connection, there is increased activity in web search for human expert identification. Search engines like Google are widely used to find experts. There are some community web search engines that need prior registration in order to perform search for experts. There is common problem conceived with any web search engine. They produce results as billions of records that are mixed with noise and quality of different levels. Generally it is not possible to get the right experts directly without further effort in surfing search results. This makes the search for human experts a tedious task. Full of noise exposed in results is the main issues that needs to be addressed.

As information is growing in exponential fashion over WWW, the bulk of results have become common in web search results. Obtaining exact results from the search engine results is a time consuming task. Search engines are not perfect in the area to show exact results. For instance when a search operation is made on Ivanovic who is a famous tennis player in the world, the Google search results are astonishing as it shows that the candidate is expert in swimming.

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

As can be seen in Figure 1, the vague results or the results with noise are evident. This problem needs to tackle for identifying human experts with ease. To overcome this problem, there are many community search engines that help in solving the problem. However, the reality is that such search mechanisms do not provide exact results. Another problem with such tools is that they are bound to particular field and not aware of other fields. The quality of the human expert search results is the challenging problem needs to be addressed. Recently Guan et al. [1] proposed a solution for this problem. The solution was based on co-occurrence based diffusion model and the ranking. This solution could find relevance and reputation of human experts and the ranking was able to provide best results at the top facilitating end users to pick results directly for further action. Instead of providing millions of records to web user, the solution provides most accurate results. Influenced by this solution, we built a prototype application that facilitates the human expert search in intuitive fashion. The empirical results are encouraging.

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.

## 2. Related Work

Searching World Wide Web for required information has become a common phenomenon of people of all walks of life. However, many organizations and enterprises use search engine over WWW for finding human experts in order to exploit their services for faster growth and expansion of their businesses. Many prior works in this kind of research were found in the literature. In [2] it was found that there are mechanisms that can help in making human expert search on WWW. This will help in finding human resources with given skill set. Later many other researches were attracted by the problem of human expert search over web. Profile management with underlying search approach was presented in [3] and [4].

In 2005 an expert search approach was built specifically to unearth talented human beings across the globe. Thus TREC enterprise track came into existence [5]. Later on other approaches such as Model 1 and Model 2 were conceived for human expert search [6]. Tilburg University provided similar kind of research results in its web site with a new model for searching human experts [7]. In many researches retrieval performance was another problem addressed. Page Rank is the solution for retrieval performance as found in the literature [8]. There were other approaches for the same which include non-local evidence as explored in [10] and [9], voting techniques and data fusion [13], query expansion [11], learning based framework [14] and so on. These solutions focused on the retrieval performance.

A proven fact and that is observed by Internet users often is that the search results are poor with lot of noise that will waste time of end users. Moreover spam is another problem that mimic like good results taking the user to unintended web pages. Even in community web applications where search is facilitated the search results are not up to mark. High expertise humans were explored in [15] and [16] with good search mechanisms. In [18] and [17] multiple aspects matching was implemented. Many researchers focused on graph based solutions and heat diffusion approaches in order to solve the problem of expert search over web [19], [20], [21] and [22]. Recent approach which is very close to our solution is in [1] which is based on co-occurrence based diffusion method for talent search over WWW with reasonable accuracy.

### 3. Proposed Model For Expert Search

Research revealed that co-occurrence information plays an important role in obtaining search results without noise. The concept can be coupled with the heat diffusion model that can aggregate information to deal with noise problem in search results. With this method human experts can be identified intuitively. The study of query string and the relevancy of results were studied using co-occurrence based approach. Moreover ranking the final results can make them more useful and users can directly get required results. Thus the search results for human experts are expected to have highest quality. More technical details used for implementation of the prototype in this paper can be found in [1].

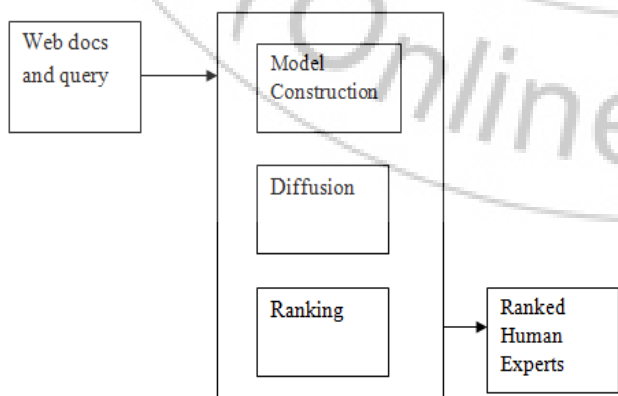


Figure 2: Overview of the proposed approach

As can be seen in Figure 2, it is evident that that proposed framework takes web documents and query as inputs and performs operations like model construction, diffusion and ranking. The ranking function is applied as the user needs to have quality list of human experts.

### 4. Prototype Application

A prototype web application is built to demonstrate the concept. The application is built using Java programming language and technologies like Servlets and JSP. The web application provides user-friendly interface that can help in interacting with WWW and get the results as per the query string. The results are ranked based on the co-occurrence method. The environment used for the experiments is a PC with 4 GB RAM, core 2 dual processor running Windows 7 operating system.

### 5. 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.

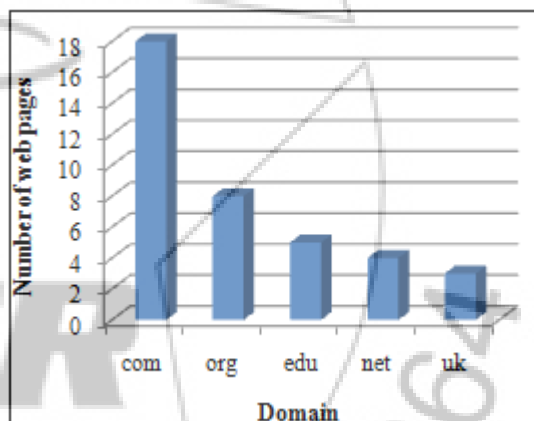


Figure 3: Domains and number of web pages

When search is carried out using the proposed system, relevant results are presented as output. The summary of search results is presented in Figure 4.

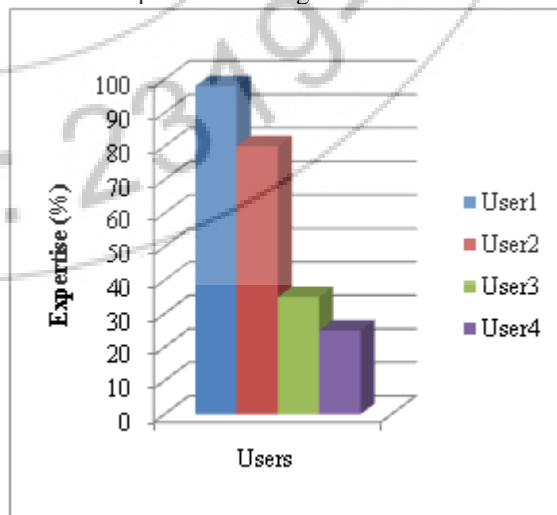


Figure 4: Summary of search results

As can be viewed in Figure 4, the summary of search results is presented in. The graph shows expertise of various users.

## 6. Conclusion and Future Work

Expert search over web is the main study of this paper. The traditional search engines and specialized community search engines do have their drawbacks in providing accurate results pertaining to human experts. The main problem with them is that the quality of the results is poor and noise is more. As the search applications bestow millions of records, finding desired results is time consuming and needs further surfing of results. This does not provide rich user experience in human expert search. Vague evidence in human expert results is another problem which makes the results not reliable. To overcome this problem, Guan et al. [1] proposed a novel approach based on diffusion and ranking model. This model makes use of co-occurrence dynamics so as to estimate accurate human experts thus reducing noise in the results. Moreover, the underlying ranking mechanism helps in increasing relevance in search results. The same approach is used by us. We built a prototype application that helps in finding human experts with ease. The experimental results reveal that the proposed application is useful in finding human experts without noise. This tool can be improved further in future to ensure that vagueness disappears in search results.

## 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@nopic 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.