

A Framework for Keyword Based Query and Response System for Web Based Expert Search

Shoban Babu Sriramoju

Professor, Department of Computer Science and Engineering, Varadha Reddy College of Engineering, India

Abstract: *Querying for human experts over WWW became an important activity. The query made through search engines produces numerous results that are huge in number and the users often lose track in finding suitable results. This is the problem when expert search is made in general. There are many issues related to this problem. Many researchers came up with ranking algorithms in order to overcome this problem. Page Rank and other algorithms played an important role in finding quality results. Recently there is research on heat diffusion techniques. In this paper we presented a framework that makes use of heat diffusion based ranking algorithm for more accurate ranking. This algorithm provides best ranked results to end users. This can help users to have highly relevant expert information available over web. We built a prototype application to demonstrate the proof of the concept. The results are taken and observed that the proposed method is useful in finding highly relevant experts over WWW.*

Keywords: Web search, web mining, expert search and ranking.

1. Introduction

Mining World Wide Web (WWW) became an important research activity due to the source of information available over Internet. In fact, there is wealth of information over WWW which can be analyzed to gain business intelligence for different fields and organizations. In other words, web mining is the process of exploring web to obtain useful and previously not known facts or patterns or trends. These trends when utilized by an expert and interpreted become business intelligence. Such business intelligence can be used in order to make well informed decisions. Since WWW is a collection all web servers across the globe, WWW became a goldmine for researchers. Different domains can have a common search for information or domain specific search. Recently there is need for making expert search. It is the search for human experts. When an expert is needed in any field, WWW allows us to have specific search that can help in getting results.

The problem with aforementioned search is that search engines retune plethora of records or pages that are to be verified again for most relevant expert details. Sometimes the pages returned might have irrelevant information. These problems are faced by end users. To overcome these issues, it is important to have some sort of correlation, filtering and then ranking. When ranking is made, it is possible to improve the quality of results. Many researchers contributed towards ranking and other parts of the expert search. In this paper, we proposed a framework that helps in finding experts and then rank them based on heat diffusion based ranking algorithm. The framework proposed in this paper has the underlying algorithm that takes different matrices and performs web mining in order to have ranked expert results. The remainder of the paper is structured as follows. Section II provides review of literature. Section III presents the proposed system in detail. Section IV presents experimental results while section V concludes the paper.

2. Related Works

This section provides review of literature pertaining to expert search problem and the present state of the art in this area. Expert search is the research area which is growing faster. As explored in [1] organizations are building knowledge databases that contain important skill information of employees. However, making knowledge base with global information and doing it with automation is very important. Such automated approaches came into existence as presented in [2] and [3]. Since 2005, expert search became very hot topic in the academics and research domain [4]. There was a language model framework proposed in [5] for expert search. Similar kind of work was carried out in [6] with multi-lingual support in search for experts. In order to improve the performance of search results researchers explored many techniques such as data fusion techniques [7], proximity between occurrences [8], [9], non local evidence approach [10], [11], query expansion and relevance feedback [12], [13], similarity between persons [6] and URL length of retrieved documents [14].

Probabilistic random walks is another approach used in [15]. Relevance-feedback based discriminative learning and a framework related to it are done in [16]. Towards doing this kind of experiments two benchmark datasets are available. They are known as CSIRO [17] and W3C [4]. These datasets are widely used in order to have expert search research. There are many issues with expert search. For instance, finding similar experts is a problem as explored in [18]. Finding high expertise individuals is another problem as discussed in [19]. The problem of community based query and answer kind of framework was explored in [20]. Problem of matching research papers and reviews and its related algorithms were explored in [21]. In the same fusion, in [22] and [23] the matching problem with single and multiple aspects of expertise was discussed. Topic based model and language modelling were explored in [24] while coupled random walks approach is followed in [25].

The concept of heat diffusion was used by many researchers in applications such as online advertisement matching's, social network marketing, ant spamming, classification and dimensionality reduction as explored in [26], [27], [28], [29] and [30]. In this paper we proposed a framework that takes care of ranking expert results using an algorithm in order to improve user experience and improve quality of results pertaining to expert search.

3. Proposed Framework for Expert Search

In this paper we proposed a framework that takes care of heat diffusion based ranking of expert search results. There is much importance given to search of experts in the contemporary world. The rationale behind this is that experts available in different fields are located geographically and reaching them with the virtual world is mandatory. In this context, the search results for experts should provide high quality results in order to help users to gain suitable candidates.

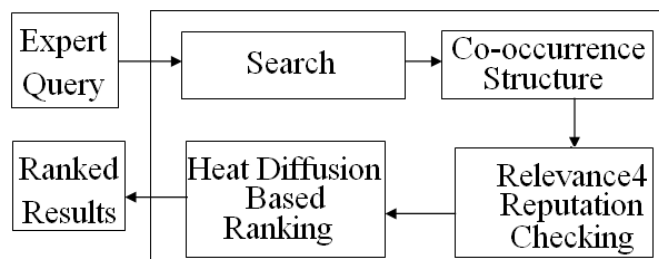


Figure 1: The proposed framework for ranked results

As shown in Figure 1, it is evident that the expert query is given by people in the application which provide web search user interface. Once the search results are returned by search engine, the results are subjected to finding co-occurrence structure. Once co-occurrence structure is found, it is further used to check relevance and reputation checking. The results are then subjected to heat diffusion based ranking which is very essential in order to give ranking results. Thus the ranked results are very useful to end users. It can save their time besides presenting high quality information. Heat diffusion is a mechanism that is used with the data related web. Relationship matrix between people and pages, Relationship matrix between words and pages and Matrix containing page rank score of pages are involved in the algorithm that can finally rank results accordingly and provide results with high quality. Such quality results are the take away to users of expert search.

Notations Used

Table 1: Notations used in Heat Diffusion Algorithm

Notation	Description
R_p	Relationship matrix between people and pages
R_w	Relationship matrix between words and pages
R_e	Matrix containing page rank score of pages
R	Ranked results
HM	Heat matrices
H	Heat distribution vector

As shown in Table 1, the notations are used in the proposed algorithm. The algorithm is based on heat diffusion concept

that is used to have improved ranking and present expert search results with more quality.

Heat Diffusion Algorithm

Algorithm: Heat Diffusion Algorithm (HDA)

Inputs: R_p , R_w , R_e

Outputs: Ranked results R

```

01 Find number of distinct co-occurring people using  $R_p$ 
02 Create degree matrices using  $R_p$ ,  $R_w$ ,  $R_e$ 
03 Create heat matrices  $HM$ 
04 For  $i = 1$  to  $n$  do
05   Update heat distribution vector  $H$  using  $HM$ 
06 End
07 Compute  $R$  based on  $H$ 
08 Return  $R$ 
    
```

Algorithm 1: Heat Diffusion Algorithm

Algorithm 1 is used to take different relationships among web pages and web users and compute ranked results. The ranked results are computed based on heat diffusion method that caters to the needs of quality ranking. When ranking process is improved, it is understood that the ranked results can satisfy the users. The users when use improved expert search results, the ranking of results can help users to pick ideal experts that meet their requirements precisely.

4. Experimental Results

This section provides results after making experiments. The observations are mainly related to the quality of ranked results due to the heat diffusion based framework. We built a prototype application to demonstrate proof of the concept. The results of experiments are presented in this section.

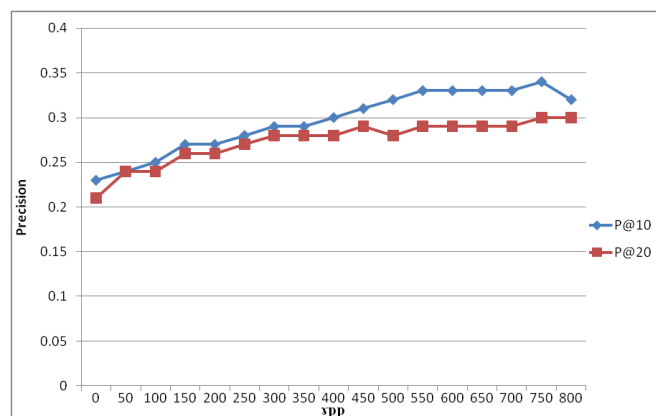


Figure 2: Heat conductivity among people vs. precision

As shown in Figure 2, the horizontal axis shows the heat conductivity among people and the precision in vertical axis. The results show that the performance of P@10 is better than that of P@20.

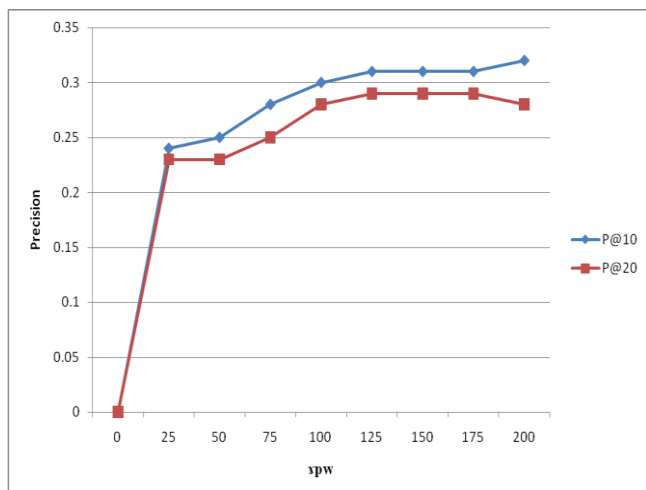


Figure 3: Heat conductivity among people and words vs. precision

As shown in Figure 3, the horizontal axis shows the heat conductivity among people and words and the precision in vertical axis. The results show that the performance of P@10 is better than that of P@20.

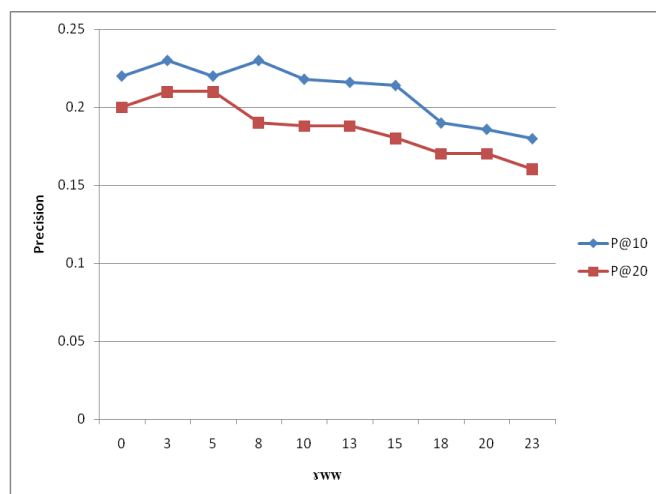


Figure 4: Heat conductivity among words vs. precision

As shown in Figure 3, the horizontal axis shows the heat conductivity among words and the precision in vertical axis. The results show that The performance of P@10 is better than that of P@20.

5. Conclusions and Future Work

The usage of search engines has been around for many years. However, using search engines for finding human experts is relatively new. Nevertheless, this trend is increasing as the skilled people are playing vital role in the growth of organizations. Keeping this in mind web mining has been used for finding human expert information over web. In spite of the utility of expert search information over Internet, it is understood that the search engines are providing very vast information and sometimes most of the results show irrelevant details. Overcoming this problem is challenging. Many researchers came up with ranking algorithms in order to overcome this problem. Page Rank and other algorithms played an important role in finding quality results. Recently

there is research on heat diffusion techniques. In this paper we presented a framework that makes use of heat diffusion based ranking algorithm for more accurate ranking. This algorithm provides best ranked results to end users. Our empirical results revealed that the proposed heat diffusion approach is useful in finding best results related to expert search. In future we intend to improve this research by using different algorithms and comparing them with heat diffusion algorithm.

References

- [1] P.R. Carlile, "Working Knowledge: How Organizations Manage What They Know," Human Resource Planning, vol. 21, no. 4, pp. 58- 60, 1998.
- [2] N. Craswell, D. Hawking, A.M. Vercoustre, and P. Wilkins, "P@noptic Expert: Searching for Experts not Just for Documents," Proc. Ausweb Poster, 2001.
- [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, A.P. de Vries, and I. Soboroff, "Overview of the Trec 2005 Enterprise Track," Proc. Text Retrieval Conf. (TREC), 2005.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] K. Balog, L. Azzopardi, and M. de Rijke, "A Language Modeling Framework for Expert Finding," Information Processing & Management, vol. 45, no. 1, pp. 1-19, 2009. Fig. 9. Running time when varying the number of relevant webpages (Local Ranking).
- [9] Y. Fu, W. Yu, Y. Li, Y. Liu, M. Zhang, and S. Ma, "THUIR at Trec 2005: Enterprise Track," Proc. Text Retrieval Conf. (TREC), 2005.
- [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] 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.
- [12] K. Balog and M. De Rijke, "Associating People and Documents," Proc. IR Research, 30th European Conf. Advances in Information Retrieval (ECIR), pp. 296-308, 2008.
- [13] 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.
- [14] 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.
 - [15] P. Serdyukov, H. Rode, and D. Hiemstra, "Modeling Multi-Step Relevance Propagation for Expert Finding," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 1133-1142, 2008.
 - [16] 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.
 - [17] 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.
 - [18] 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.
 - [19] X. Liu, W.B. Croft, and M. Koll, "Finding Experts in Community- Based Question-Answering Services," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 315-316, 2005.
 - [20] D. Mimno and A. McCallum, "Expertise Modeling for Matching Papers with Reviewers," Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 500-509, 2007.
 - [21] 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.
 - [22] M. Karimzadehgan, C. Zhai, and G. Belford, "Multi-Aspect Expertise Matching for Review Assignment," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 1113-1122, 2008.
 - [23] H. Deng, I. King, and M.R. Lyu, "Formal Models for Expert Finding on DBLP Bibliography Data," Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 163-172, 2009.
 - [24] D. Zhou, S. Orshanskiy, H. Zha, and C. Giles, "Co-Ranking Authors and Documents in a Heterogeneous Network," Proc. Int'l Conf. Data Mining (ICDM), pp. 739-744, 2007.
 - [25] H. Bao and E.Y. Chang, "Adheat: An Influence-Based Diffusion Model for Propagating Hints to Match Ads," Proc. Int'l Conf. World Wide Web (WWW), pp. 71-80, 2010.
 - [26] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003
 - [27] 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.
 - [28] 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.
 - [29] 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.
 - [30] Shoban Babu Sriramoju "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Reasearch and Applications" Vol-1, Issue-1, Jan-Mar 2014.
 - [31] Shoban Babu Sriramoju "Risk Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management" Vol-VI, Issue-I, Feb'2014 [ISSN : 2249-4510]
 - [32] Shoban Babu Sriramoju "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management" Vol-VI, Issue-I, Feb'2012 [ISSN : 2249-4510]
 - [33] Shoban Babu Sriramoju "An Application for Annotating Web Search Results" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-3, March'2014 [ISSN (online): 2320-9801, ISSN(print) : 2320-9798]
 - [34] Shoban Babu Sriramoju "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-3, March'2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
 - [35] Shoban Babu Sriramoju "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering" Vol-4, Issue-2, February'2014 [ISSN : 2277 128X]
 - [36] Shoban Babu Sriramoju "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol-V, Issue-II, February'2014 [ISSN : 2321-3469]
 - [37] Shoban Babu Sriramoju "Mining Big Sources Using Efficient Data Mining Algorithms" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-1, January'2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
 - [38] Shoban Babu Sriramoju "A Review on Processing Big Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-1, January'2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
 - [39] Shoban Babu Sriramoju "An Analysis around the study of Distributed Data Mining Method in the Grid Environment : Technique, Algorithms and Services" in "Journal of Advances in Science and Technology" Vol-IV, Issue No-VII, November'2012 [ISSN : 2230-9659]
 - [40] Shoban Babu Sriramoju "An Analysis on Effective, Precise and Privacy Preserving Data Mining Association Rules with Partitioning on Distributed Databases" in

“International Journal of Information Technology and management” Vol-III, Issue-I, August’2012 [ISSN : 2249-4510]

[41] Shoban Babu Sriramoju “A Competent Strategy Regarding Relationship of Rule Mining on Distributed Database Algorithm” in “Journal of Advances in Science and Technology” Vol-II, Issue No-II, November’2011 [ISSN : 2230-9659]

[42] Shoban Babu Sriramoju “Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts” in “International Journal of Information Technology and management” Vol-I, Issue-I, August’2011 [ISSN : 2249-4510]