Heat Diffusion Based Search for Experts on World Wide Web

Shoban Babu Sriramoju

Department of Computer Engineering, SR Engineering College, 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 identity 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 search application which is web based that uses heat difusion technique to identify web resources that contain details about human experts. Ranking is used for presenting results meaningfully. Our prototype application demonstrates the proof of concept. The empirical results are encouraging.*

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

1. 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 search engines is that they provide billions 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 the exact results as expected by the 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. Expert search is allowed in the applications where the user can search for persons with expertise in different areas. 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 are implementing an application which is web based 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.

2. Related Work

Expert search has been around for several years on the web. Human experts with certain different skills can be searched over web [2]. Due to this many approaches came into existence. Building profiles of people automatically is one of the approaches as explored in [4]. TREC enterprise track was built in 2005 for expert search [5]. Model 1 and 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 resuls 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 the web search results is quite less when compared with community search that targets specific information. In [15] and [16] researches tried to provide mechanisms to identity 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 for identifying human experts over web and provide search results accurately.

3. Heat Diffusion Method for Expert Search

Heat diffusion model will construct a matrix to analyze co-occurrence information and aggregate it to deal with noises on the web. This method is used for identifying high expertise people whose information is available in web. The co-occurrence based heat diffusion model studies the query string and names with strong connection to relevant things.

Based on this criteria, appropriate ranking is provided which improves the quality of expert search result. More information can be found on the approach in [1].

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.

Figure 2: Web based UI for authentication

As can be seen in Figure 2, authentication can help users to gain access to the search facilities of the application.

Figure 3: UI for viewing data

As shown in Figure 3, presents a list of topics with other attributes. This will help in building large synthetic dataset which is used in future work.

Figure 4: Provision for manipulating registered users

As seen in Figure 4, it is evident that the application supports manipulation of users' data.

Figure 5: Search interface of the proposed application

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.
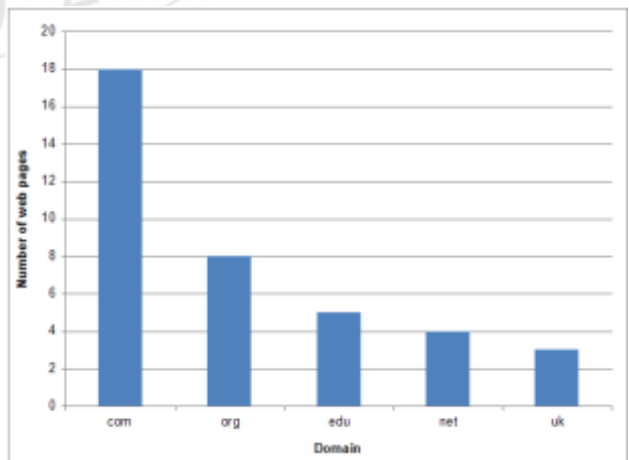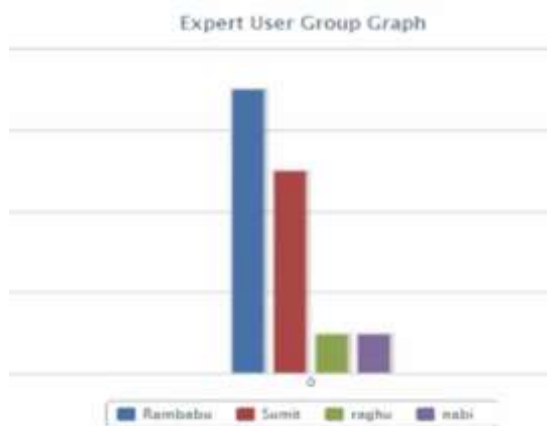

Figure 6: Domains and number of web pages

When search is carried out using he proposed system, relevant results are presented as output. The summary of search results is presented din Figure 7.

Expert User Group Graph

As can be viewed in Figure 6, the summary of search results is presented in. The graph shows expertise of various users.

6. 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 result contains irrelevant noise information that are confusing. The web pages may also produce the results that provide vast evidence of expertise on different 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 by using heat difusion technique to rank the results and identify very useful research that reflects users search intention. We built a prototype application to will demonstrate the expert search concepts. The results specify that it is possible to rank the result based on expert search to make the search results more meaningful. Such results can be directly used by users which contains the expert information.

References

[1] Ziyu Guan, Gengxin Miao, Russell McLoughlin, Xifeng Yan, Member, and Deng Cai. (2013). Co-Occurence Based Difusion 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] Yimam Seid and Kobsa, "Expert Finding Systems for Organizations: Problem and Domain Analysis and the Demoir Approach," Organizational Computing and Electronic Commerce, vol. 13, no. 1, pp. 1-24, 2003.

[4] N. Craswell, Hawking, A.M. Vercoustre, and P. Wilkins, "Pnoptic 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.

[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] Balog and Rijke, "Non Local Evidence for Expertise Findings," 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] Fang, L. Si, and 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, 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] Yang, "Mining Social Networks Using Heat Difusion Processes for Marketing Candidates Selection," Proc. ACM Conf. Information and Knowledge Mgt, 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] Babu, Sriramoju Ajay and Babu, S Shoban. "International Journal of Research and Applications Jan-Mar© 2016 Transactions 3 (9): 422-426 eISSN: 2349–0020." International Journal of Research In Science & Engineering e-ISSN: 2394-8299 Volume: 3 Issue: 4 July-August 2017 p-ISSN: 2394-8280 IJRISE JOURNAL| www.ijrise.org|editor@ijrise.org [56-65]

[23] Ajay Babu Sriramoju, Dr. S. Shoban Babu. "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays For Image Processing" (2013).

[24] Dr. Shoban Babu Sriramoju, Prof. Mangesh Ingle, Prof. Ashish Mahalle "Trust and Iterative Filtering Approaches for Secure Data Collection in Wireless Sensor Networks" in "International Journal of Research in Science and Engineering" Vol-3, Issue-4,July-August 2017 [ISSN : 2394-8299].

[25] Dr. Shoban Babu, Prof. Mangesh Ingle, Prof. Ashish Mahalle "HLA Based solution for Packet Loss Detection in Mobile Ad Hoc Networks" in "International Journal of Research in Science and Engineering" Vol-3, Issue-4,July-August 2017 [ISSN : 2394-8299].

[26] Shoban Babu Sriramoju. "A Framework for Keyword Based Query and Response System for Web Based Expert Search" in "International Journal of Science and Research" Index Copernicus Value(2015):78.96 [ISSN : 2319-7064].

[27] Sriramoju Ajay Babu, Dr. S. Shoban Babu. "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications" Vol-1, Issue-1,Jan-Mar 2014 [ISSN : 2349-0020].

[28] Dr. Shoban Babu Sriramoju, Ramesh Gadde. "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol-1, Issue-1,Jan-Mar 2014 [ISSN : 2349-0020].

[29] 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]

[30] 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]

[31] Shoban Babu Sriramoju, Madan Kumar Chandran. "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]

[32] Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao. "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol-V,Issue-II,Febrauary'2014 [ISSN : 2321-3469]

[33] Ajay Babu Sriramoju, Dr. S. Shoban Babu. "Analysis in Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and Management" Vol-VII, Issue-X, Nov' 2014 [ISSN : 2249-4510]

[34] 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]

[35] Ajay Babu Sriramoju, Dr. S. Shoban Babu. "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays for Image Procesing" in "International Journal of Information Technology and Management" Vol-V, Issue-I, Aug' 2013 [ISSN : 2249-4510]

[36] Dr. 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]

[37] Shoban Babu Sriramoju, Dr. Atul Kumar. "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]

[38] Shoban Babu Sriramoju, Dr. Atul Kumar. "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]

[39] Shoban Babu Sriramoju, Dr. Atul Kumar. "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]

[40] Shoban Babu Sriramoju, Dr. Atul Kumar. "Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts" in "International Journal of Information Technology and management"