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 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 search application which is web based that uses heat diffusion 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 browse 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.

![Figure 1: Vague expertise evidence](image)

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 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 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](image2)

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](image3)

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.

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 4: Provision for manipulating registered users](image4)

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

![Figure 5: Search interface of the proposed application](image5)

![Figure 6: Domains and number of web pages](image6)

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.
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 diffusion 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


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