Online Job Search: Study of Users' Search Behavior using Search Engine Query Logs

Behrooz Mansouri [1], Mohammad Sadegh Zahedi [1], Ricardo Campos [2], Mojgan Farhoodi [1]

[1]Web Azma Lab, Information Technology Faculty, Iran Telecommunication Research Center, Tehran, Iran
[2] Polytechnic Institute of Tomar, Smart Cities Research Center, LIAAD - INESC TEC, Portugal
{b.mansouri,s.zahedi,farhoodi}@itrc.ac.ir, ricardo.campos@ipt.pt

ABSTRACT

Over the last few years, an increasing number of user's and enterprises on the internet has generated a global marketplace for both employers and job seekers. Despite the fact that online job search is now more preferable than traditional methods - leading to better matches between the job seekers and the employer's intents - there is still little insight into how online job searches are different from general web searches. In this paper, we explore the different characteristics of online job search and their differences with general searches, by leveraging search engine query logs. Our experimental results show that job searches have specific attributes which can be used by search engines to increase the quality of the search results.

KEYWORDS

Query understanding; Query log analysis; Job Search

ACM Reference format:

B. Mansouri, M.S. Zahedi, R. Campos and M. Farhoodi. 2018. SIG Proceedings Paper in word Format. In *Proceedings of ACM SIGIR conference, Ann Arbor, MI, USA, July 2018 (SIGIR'18)*, 4 pages. https://doi.org/10.1145/3209978.3210125

1 INTRODUCTION

The Internet plays a vital role in our daily life with people of all ages and experiences constantly browsing the Internet and using search engines to meet their information needs. For example, an elderly person may issue a query to get information about a disease, while a student may do the same to collect information about a practical research. In general, the Internet has facilitated people's lives in different aspects and everyday tons of millions of queries are issued to find the right answer for personal and social problems. One of the basic concerns of people, from the old time till now, is having a job that is appropriate to their individual skills. In the past, finding jobs was done through recruiting advertisements, mostly on newspapers. But today, with the help of the Internet, companies have their own ads on their websites. In addition to this, there are several websites to find a job known as employment websites, such as *Indeed.com* or even social networks such as *linkedIn.com* that make it easy for job seekers and enterprises to connect to.

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07…$15.00
https://doi.org/10.1145/3209978.3210125

A study by Kuhn and Mansour [12] showed that, the unemployed persons who used the internet to find a job, were re-employed about 25% faster than comparable workers who do not search for a job online. Another research by Prakash [15], shown that for those employees who found their jobs online, the exit rates are lowered by at least 28%. The availability of online job searches, has also lead to a bunch of records concerning the job seeking behaviors ready to be studied. For instance, Jansen et al. [10] studied how people search for job-related information on the web and how effective are these searches.

The results of their study show that, individuals, seeking job information, submit only a single query consisting of several terms. The study of Stevenson [17] focused on the effect that the Internet has on the job search behavior and stated that, those who use the Internet for job seeking purposes were mostly employed and were likely to leave their current jobs to find a new one. Baker and Fradkin [2] in turn, used Google search data to study the impact of unemployment insurance (UI) on job search. They showed that individuals receiving UI, search less for jobs than individuals who are unemployed and who are not receiving UI. While the majority of previous researches have focused on the impact of online job searches on labor market, only a few works have studied the properties of job searches from the information retrieval viewpoint. In particular, Spina et al. [16] compared users' search behavior for job and talent search, i.e. when a company is searching across resumes to identify candidate that might be interested in applying for vacant positions. They used the SEEK (Australian employment website) log for job search and Yandex (Russian search engine) log for web search analysis. The results of their study revealed that user models, ranking factors and success metrics are different for job and talent searches compared to general web ones. Faberman and Kudlyak [8] in turn, examined the relationship between job seekers' search effort and the duration of this search. Their findings were two-fold: first, that an individual job seeker's search effort increases with search duration; second, the fact that the longer job seekers look for a job, the more applications they send per week throughout the duration of search. Different from previous web search, we focus on studying job searching behavior by using a two-year query log records of a Persian general-purpose search engine. Using this data, we plan to answer the following research questions:

RQ1: How often are search engines used for searching for a job?
RQ2: How is users' effort different in searches related to jobs?
RQ3: Which websites are mostly clicked regarding job searches?
RQ4: How users formulate their queries in job searches? How the length of these queries are different from general ones? How often temporal expressions and named entities such as organization or location are used in job searches?
RQ5: May search engines help users find their desired job?

The overall goal of such analyses is to understand how online job seeking is done and what are the differences between these type of searches and general ones. By understanding users' behavior, search engines will be able to provide better results and search services. For instance, our analysis on query formulation shows that, in 35% of job searches, users add temporal expressions to their queries, of which 93.5% were referring to recent time. This may indicate that users are targeting fresh information on job searches which may lead search engines to consider different ranking algorithms for these searches. The remainder of this paper is organized as follows. Section 2 presents our experimental setting. Section 3 describes the different results of our experiments. Finally, Section 4 provides some conclusions and outlines future work.

2 EXPERIMENTAL SETTING

To conduct our analysis, we resort to a Persian general-purpose search engine (parsijoo.ir) query log consisting of 27M queries and corresponding users' interaction with the search engine collected from March 2015 to March 2017. To select the job search-related queries from the logs we rely on keyword-based techniques. Following the work of Beitzel et al. [4], we asked 5 students, for each one to suggest a list of 20 terms commonly used for job searches. We then selected the top-20 frequent seed terms (such as hiring, employment and job) and from the query log records, extracted the queries containing these terms. In the next step, we explored the extracted records to identify other frequent terms that are used in job searches queries. For this purpose, we start by considering the top-100 frequent terms and selected only those related to the job search task. In particular, 30 more terms were selected, which added to the previous 20, totalized 50 seed terms. Any query containing these terms is recognized as a job search related query. With this method, 512,483 job search records were extracted from query logs. To compare the job searches with general searches (those not related to job search) we randomly selected the same amount of records not containing any of these terms (512K) from the query log.

3 EXPERIMENTAL RESULTS

In this section we present the results of our analysis on job searches and attempt to provide answers to each one the research questions.

3.1 Frequency of Job Search

As stated in the previous section, 512,483 of queries issued by users are related to the job seeking task, which account for 1.89% of the total queries. This answers our first research question, and clearly shows that job search queries are just a tiny part of the search task. Although 1.89% seems to be a small percentage, still it represents a considerable number of queries, which we should look at carefully. To better understand the frequency of job search queries, we plotted their distribution by week days in Figure 1.

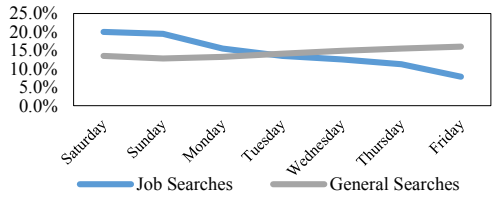

Figure 1.The distribution of job search related and other type of queries during week days.

As this figure suggests, job searches are mostly done within the first days of the week (in Persian calendar, Thursday and Friday are weekend days). By looking at the figure one can observe that 20.7% of the queries related to this task are issued on Saturday to a minimum of 6.5% on Friday. This contrast with the frequency of general searches which show a reverse trend as more queries are issued during the weekends as opposed to the beginning of the week. This can be leveraged by content providers and companies who may plan to publish their hiring advertisements during the first days of the week, when job seekers tend to more active in searching for a job. Another interesting observation in this research was the distribution of Job related queries by cities. To do this, we considered the users' IP address. Table 1 indicates the distribution of job search related queries in the five cities with most job searches. Each city has a unique IP range.

Table 1. The distribution of job search related in the five cities having the most queries issued.

#	city	Percentage of Job search	Population (in million)
1	Tehran	32.3%	8.6
2	Esfahan	9.7%	5.12
3	Mashhad	6.3%	3
4	Tabriz	2.8%	1.73
5	Ahwaz	2.1%	1.13

As can be seen, job searches were dominantly done in the industrial cities. More than 30% of the job queries were issued by users in Tehran (the capital of Iran). Although one can argue that this is the biggest city in Iran with more than 8.6M of people, still we can find a few other cities which are not within the top-5 biggest cities in Iran, and where a search for a job is still a realm. One such information can be used by search engines to better target their ads for a few cities, where job search is predominantly carried out. The impact of local aspects of web search on advertising – associating geographic queries with ads – is not a new topic and has been studied over the last few years [1,3,7].

3.2 Users' Effort

In this section, we aim to understand the difference, if any, in the user's effort when posing queries concerning the search of a job or a general query. Our assumption, is that unlike a normal search, users looking for a job aim to browse all possible pages related to their query as a way to review all job positions that suits them, which makes job search a recall-oriented searching process. This leads users to issue more queries within the same search session regarding a job position. To verify this assumption, we began by comparing the number of queries per session in job search related sessions against normal search sessions. The obtained results show that, on average, 1.62 general Persian web search queries are issued by a user, while 2.45 were submitted in search sessions related to job search (any search session that contains at least one query related to the job seeking task), thus suggesting that users try to reformulate their query to better find the desired information. To consubstantiate our assumption, we also studied the number of clicks on search engine result pages (SERPs) in both job and normal searches. The obtained results show that, for job searches the average number of clicks per search session sums up to 3.69 clicks, while for general searches this figure is reduced to only 1.30 clicks, which may also be related to the fact that some of this web pages are of navigational nature. Though, this tends to confirm that users prefer to visit more webpages related to their queries when looking for information regarding the job search

task. Finally, we studied the search session duration. For this purpose, we consider the time between the first and the last query submitted to the search engine within the same search session. Therefore search sessions with only one query are not considered. Our analysis indicates that, the average search duration for job searches is 15.98 minutes, while for general search is only 3.04 minutes. This shows that job search sessions are, on average, approximately 12 minutes longer than general searches and that users spend more time to review the results related to job searches when compared to general ones. Overall, it can be seen that the users' effort is considerable higher in job searches when compared to general searches, as more queries are issued, more result pages are viewed and more time is spent by users to seek their desired job. This answers our second research question and confirms that job search is a recall-oriented process, which may, per se, justify a different result retrieval strategy by search engines.

3.3 Clicked Pages

In this section, we will study the popular clicked pages in job searches in order to answer research question number three. Our goal is to study if users prefer to view the organizations website (who are looking for employees) or employment websites, such as indeed.com. To do so, we considered the top 300 clicked URLs in job searches, which account for 71% of the total clicked pages in job searches. Next, we manually categorize them into one the four types below:

- Employment: vertical web search engines web pages belonging to job agents such as "www.bazarekar.ir".
- News: news websites such as "www.yjc.ir".
- Organization: webpages belonging to specific organization such as "www.police.ir".
- Other: any webpages that cannot be put into one of the previous types.

The results obtained show that employment websites were the most clicked pages with 51% of the total. This may indicate that, users prefer to continue their job searches through vertical search engines that may provide better job search services for them, such as constraint for cities, sex, age or salary. The second type of clicked pages are Organizations which accumulate 29% of the total clicked pages. By further exploring these webpages, it was observed that they belong to well-known organizations which hire on a regular basis and usually have a dedicated web page to that purpose. The submitted queries that led into clicking such pages show that users were aware of the recruitment process and were searching for more precise information about the employment conditions. In turn, news pages account for only 15% of the clicked pages, mostly referring to governmental companies displaying information about the hiring process. The other type of pages only account for 5% of total clicked web pages. Overall, from this obtained data, it can be concluded that the majority of job seekers, prefer to be forwarded to specialized websites as opposed to general web search engines which lack in providing job-dedicated search options. This should be taken into account by search engines, which have a great opportunity here to increase the users' satisfaction.

3.4 Query Formulation

In this section, we aim to study the different aspects related to query formulation, in particular the differences between job search and general searches regarding query length and the use of temporal expressions in queries. In addition, we also aim to study, how often named entities such as location and organization are used when issuing queries related to the job search task. We begin by analyzing the query length of both job and general searches, as several experiments [5,6,11] have proven that, in general, existing retrieval methods, perform worse for long queries than for shorter ones. Figure 2 shows the boxplot of query length distribution for both types of queries. As can be seen from this figure, the most frequent queries related to the job search task have a length between 4 and 6 terms, with 5.3 terms per query on average, which is 2.1 terms longer when compared to general searches.

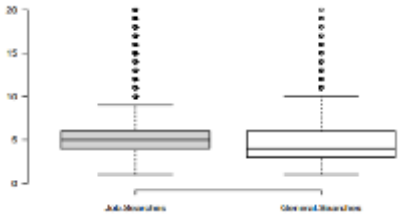

Figure 2. Boxplot of query length distribution for job and general searches.

Next, we studied the use of temporal expressions in job searches. By running our Persian Time tagger [14] on job search queries, we observed that 35% of these queries contain temporal expressions. This contrasts with the results of our recent study (over the same query log dataset used here) on the use of temporal expressions in web searches [13] where we were able to show that a minority of only 1.89% of the queries contain temporal expressions. To better understand the use of temporal information within this context, we studied to which time these temporal expressions refer to. Noticeably, the majority of temporal expressions, 93.5% to be more precise, refer to current time, 5.2% to future and only 1.3% to the past. This emphasis on the freshness of results for job searches justifies and supports the use of special retrieval techniques by search engines. Finally, we aim to understand whether Location and Organization named entities are used when posing queries related to the job search task. For this purpose, we used the Persian named entity recognizer toolkit (http://31.184.132.168/ner/index.html) and labelled the queries regarding the job searches. Our analysis shows that in 22% of all the job searches, users have mentioned the place where they are seeking for a job (e.g., "English Teacher in Tehran"). Going into details, not all the locations, as expected, are used equally in job searches queries. Indeed, locations in the industrial cities are more referred to. This information can be used by search engines in tasks such as query suggestion, by suggesting the location where users may find a job. Figure 3 shows the distribution of locations mentioned in job searches in each city in the top-10 most frequent referred cities.

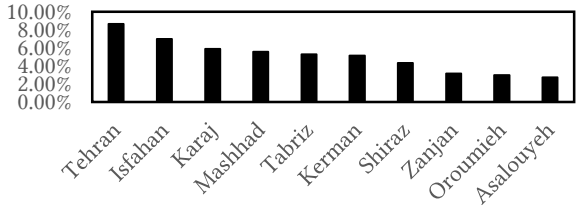

Figure 3. The distribution of locations mentioned in job searches by cities. (Top-10 popular cities)

All these cities are well-known industrial cities in Iran providing many job opportunities. Interestingly, a further analysis showed that 85% of the job searches containing a reference to a

city, were referring to a local city, which is to a city where the query was issued, meaning that, only a minority of the users (15%) look for a job outside their hometown. From this information, one may conclude that search engines should restrict the job search results to the related local job opportunities. In contrast to these results, a minority of only 5% of the queries includes organizations mentioned by the users. This shows that users prefer to be offered job search results from all possible organizations and confirms that search engines should pay more attention to recall measure. Overall, our analysis in this section proved that, job seekers formulate their queries differently when compared to issuing general web queries, either by using more terms, temporal expressions and constraints such as location. This answers research question number four.

3.5 Users' Satisfaction

Our final analysis was to understand how current search engine may help users seeking for a job. To do so, we asked three editors to submit 1000 random job queries to 3 different search engines (namely Google, Bing and Parsijoo) and check the top-5 result pages to verify whether or not their information need is satisfied. They were asked to score search engines by one of the three scores: 2, 1 or 0, where 2 means that the search engine was able to find at least one document that satisfies all of the constraints in the query, 1 means constraints in the query were partially satisfied and 0 indicates that no constraints in the user query were satisfied. For instance, editors will give a score of 2 for the results of the query "Part time software engineer in Tehran", if both "Part time" and "Tehran" - job type and location constraint – are mentioned in one of the top-5 results. If only one of the two constraints are satisfied, they would score 1 and if none is satisfied a score of 0 would be given. To be able to compare job searches with general searches, we also randomly selected 1000 queries from the general ones. An inter-rater reliability analysis using the Fleiss Kappa statistics [9] was performed to determine consistency among the editors. Overall, the annotators obtained about 0.72 of agreement level, which represents a high agreement between editors. Figure 4 indicates the average scores for each of the three search engines for job and general searches.

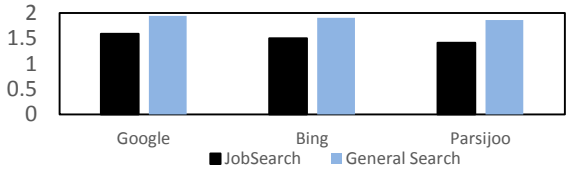

Figure 4. The average scores for Google, Bing and Parsijoo for job and general searches.

As can be seen from this plot, and regardless the search engine, results were noticeably more satisfiable for general search then for job search ones. This may raise the question of whether general purpose search engines such as Google, Bing or Parsijoo can provide good job-search results as our observation in section 3.3 also showed that users prefer to click on employment websites. To better understand the reason for lower satisfaction in job searches, we reviewed the queries with the lowest scores. Our exploration shows that search engines fails to satisfy hard constraints, that is constraints which may be rare in general searches but can be casually used in job searches. For instance, for the query "miner employment at Khvaf", search engines fail to satisfy "Khvaf" location constraint, as this is a small not well-known city in Iran. On the other hand, the result for job searches with constraints related to famous cities or companies such as "Iran air" or "National Oil Company" were satisfiable for users. Moreover, adding too many constraints to the job search query will also lead into dissatisfaction as these three search engines were mostly capable of considering a part of them. For example, the query "Female accountant Mashhad part time" has 3 different constraints including sex (female), location (Mashhad) and job type (part time). In summary, our analysis revealed that search engines are not yet capable of providing good search results for job searches compared to general ones.

4 CONCLUSIONS

Despite the fact that online job search has become common mean of job seeking process, users' job search behavior was not fairly studied. In this paper, we explored online job searchers' behavior and compared job and general searches. Our analysis showed that, job seekers formulated their query differently when compared to general searches and target fresh data in their searches. By investigating three different general-purpose search engines, our finding shows that job search result are not yet as satisfactory as general searches. As for future work, we plan to use our experiment findings and study how search engines can provide better job searches facilities and increase users' satisfaction.

ACKNOWLEDGMENTS

This research was supported by Iran Tele-communication Research Center (ITRC) and partially funded by the ERDF through the COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT as part of project UID/EEA/50014/2013.

REFERENCES

[1] Anggono, I. N., Haddadi, H., Chaabane, A., & Kaafar, M. A. (February-2016) Preserving Privacy in Geo-Targeted Advertising. In International Workshop on Ad Targeting at Scale,WSDM (pp).ACM
[2] Baker, S. R., & Fradkin, A. (2014). The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. Review of Economics and Statistics, (0).
[3] Barford, P., Canadi, I., Krushevskaja, D., Ma, Q., & Muthukrishnan, S. (2014, April). Adscape: Harvesting and analyzing online display ads. In Proceedings of the 23rd international conference on World wide web (pp. 597-608). ACM.
[4] Beitzel, S. M., Jensen, E. C., Chowdhury, A., Frieder, O., & Grossman, D. (2007). Temporal analysis of a very large topically categorized web query log. Journal of the American Society for Information Science and Technology, 58(2), 166-178.
[5] Bendersky, M., & Croft, W. B. (2008). Discovering key concepts in verbose queries. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ACM'08, 491-498.
[6] Broder, A. Z., Carmel, D., Herscovici, M., Soffer, A., & Zien, J. (2003). Efficient query evaluation using a two-level retrieval process. In Proceedings of the twelfth international conference on Information and knowledge management, ACM'03, 426-434.
[7] Dan, O., Parikh, V., & Davison, B. D. (2016, February). Improving IP geolocation using query logs. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (pp. 347-356). ACM.
[8] Faberman, R. J., & Kudlyak, M. (2014). The intensity of job search and search duration.
[9] Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among many Raters. In Psychological Bulletin, 76(5), 378-382.
[10] Jansen, B. J., Jansen, K. J., & Spink, A. (2005). Using the web to look for work: Implications for online job seeking and recruiting. Internet research, 15(1), 49-66.
[11] Kuhn, P., & Mansour, H. (2014). Is Internet job search still ineffective?. The Economic Journal, 124(581), 1213-1233.
[12] Kumaran, G., & Allan, J. (2007). A Case For Shorter Queries, and Helping Users Create Them. In HLT-NAACL, 220-227.
[13] Mansouri, B., Zahedi, M. S., Campos, R., Farhoodi, M., & Yari, A. (2018, April). Understanding the use of Temporal Expressions on Persian Web Search. In Companion of the The Web Conference 2018 on The Web Conference 2018 (pp. 1743-1748). International World Wide Web Conferences Steering Committee.
[14] Mansouri, B., Zahedi, M. S., Campos, R., Farhoodi, M., & Rahgozar, M. (2018, March). ParsTime: Rule-Based Extraction and Normalization of Persian Temporal Expressions. In European Conference on Information Retrieval (pp. 715-721). Springer, Cham.
[15] Prakash, A. (2015). Three essays on labor market outcomes. The University of Arizona. 10150/560807.
[16] Spina, D., Maistro, M., Ren, Y., Sadeghi, S., Wong, W., Baldwin, T., Cavedon, L., Moffat, A., Sanderson, M., Scholer, F. and Zobel, J. (2017) Understanding user behavior in job and talent search: an initial investigation. In Proceedings of the 2017 SIGIR Workshop on eCommerce (eCom 2017). CEUR-WS. org.
[17] Stevenson, B. (2008). The Internet and job search (No. w13886). National Bureau of Economic Research.