Analysis of Query Entries
of a Job Search Engine

Yeolib Kim

2815 San Gabriel St. Austin, TX 78705
`yeolib.kim@utexas.edu`

Abstract. The amount of interest in analyzing log files is growing with the expanding volume of information on the Web. The primary mode for information retrieval without a doubt has been search engines which can help users filter their information needs. While the population is used to general-purpose search engines, domain-specific search engines exist to segment a certain market and cater to underserved minority interests. Job search engines are widely used by job seekers ranging from new college graduates to those that are attempting to change their jobs. As part of this research, we investigate job search engine, indeed.com. We accumulated the query entries of 10 participants and analyzed the trends. We specifically examine their input queries and how they deal with output search results. Results indicate that the participants, on average, enter 1.66 queries into the main search box. The main query type for a job search engine is entering a job title. Participants rarely look beyond the first page for output search results. Instead of checking the content past the first page, they would refine their search terms and focus again on the first page. Overall, interesting patterns emerge from this research on a job search engine, but more studies are required.

Keywords: Job search engine, query length, query type, location, result page.

1   Introduction

Given the prevalence that search engines have been part of our online experience, there has been significant increase in the research and commercial section to analyze how users overall function inside search engines. Web search engines can be broadly classified into general-purpose search engine, meta-search engine, search engine inside a single website, and domain-specific search engine. Nowadays, there is a need to understand the behavior of users who use domain-specific search engines and the performance of these search engines. Domain-specific search engines are specialized search engines that focus on returning Web pages that are relevant to certain domains. These systems offer sophisticated search function because they establish their own local databases and can apply various machine learning or knowledge representation techniques to the data.

Job search is a vital part of pursuing employment. Job seekers look for jobs caused by involuntary job loss, reentry into the workforce, completion of job training, or the

desire to pursue new career opportunities. Recently, the trend is to find employment online due to the proliferation of dot-coms. A New York Times articles mentioned that the "Internet is rapidly changing the way workers search for jobs and employers recruit workers. The resulting speed and ease of filling jobs have significant implications for unemployment, pay and productivity" [1]. Increasingly, that instrument to find online jobs have shifted from job boards toward job search engine. Some of the reasons include the efficiency in browsing available jobs and the sheer amount of jobs posted that dwarf job boards. The significance of this study is straightforward. In this paper, we examine user behavior related to indeed.com, a domain-specific job search engine. Our study analyzes input method, input query length, input query type, output result page, and output result positions.

2   Background

Studies on domain-specific search engines have seen little publication up to this point. One example is a research on a medical search engine called MedicoPort [2]. It aims to retrieve useful medical information for people with no medical training. One of the main advantages of MedicoPort is that it responds to the user with only medical related Web sites and gives priority to medical portals among the Web sites with similar relevance value. Second, it can include sites that are closely related to the query concept even though the user query does not contain the exact terms. To compare the results with search engine Google, they recruit ten people to test out MedicoPort and Google. The group consisted of medical specialists and also individuals with no medical expert. They all agree that the relevance of the information retrieved by Medico-Port outperforms Google by a significant extent in finding health contexts. Next, the authors evaluate the effectiveness of the retrieval performance. They enter 10 query terms which retrieves 1,011 documents with 41,705 unique terms. The metric they use is called precision, the ratio of relevant documents retrieved to all of the documents retrieved. The average precision of the MedicoPort system for the 10 query terms is 96.3 percent. The ability to generate related concepts before starting the search is one of the merits of using MedicoPort. Excluding the usage of concept generation drops the precision rate from 96.3 percent to 93 percent.

   The second study is Deadliner, a search engine that catalogs conference and workshop announcements, and also monitors and extracts a wide range of academic convocation from the Web [3]. The system extracts speakers, locations, dates, paper submission deadlines, topics, program committees, abstracts, and affiliations. The extractor performance looked at how many correct extractions are made. They use 500 documents from DB-World, a site encompassing 208 lists of interesting topics, 338 conference titles, 906 deadlines, and 197 program committees. Testing on deadline extractions, a total of 300 deadline data fields are used. Of these, 214 deadlines are correctly detected and extracted. A total of 31 non-deadline dates are detected and extracted but they are considered extraneous. The overall deadline extraction is correct 70 percent of the time. The performance of extracting committee and affiliation corresponds to 87 percent accuracy. With regards to extracting parts of the title, they

correctly identify 90 percent of conference names and the type of meeting. However, extracting titles that include date fields and country and city names resulted in disappointing performances.

3   Methodology

Our methodology utilized a screen recording device to capture the Web activity inside indeed.com and the subsequent hyperlink connections. Indeed.com is a job search engine that aggregates from job boards, headhunter sites, online newspapers, and university job postings. The search engine box is used for job title, company name, location, and job keywords. Therefore, the terminology utilized by users has different characteristics than what a general-purpose search engine would perform by restricting the crawlers to harvest only job-related posts. While traditional log analysis on search engines is collected automatically by the machine, the researcher opted to use screen recording and code user behavior to understand the experience of using a job search engine.

Participants were recruited via email primarily through the researcher's snowball technique at the University of Texas at Austin. All were compensated $15 for their time. During the experiment, participants were asked to search for jobs as if they are actually looking for jobs. This provides the most natural environment similar to actually applying for a job.

3.1   Sample Composition

Ten participants held a Master's degree or were currently in graduate school pursuing a degree in one or more of the following fields: Information studies, risk management, and law. Six of the participants were female and six were between the ages of 26 and 30. All participants were considered ideal candidates for this study because they were about to enter the job market after graduation.

Prior work experience is an average of 6.2 years ranging from one year to sixteen years. They mostly coveted jobs in the technology sector followed by jobs in design, digital records, and research-related. Most do not know what establishments they want to work at, but three mentioned Fortune 500 companies. Some would like to get employed at mid-size or venture technology companies. The detailed positions participants were looking for follow the same trajectory as their respective focus of area. In terms of salary, they would like to meet in the $50,000 to $80,000 range. All of them would like to find their jobs in Austin, Texas with a couple considering other Texas cities, Los Angeles, and New York City. Most of them had experience with indeed.com, but a few primarily used Craigslist, Monster, or the University job board.

3.2   Measures

The input by participants relate to the following terms:

- Input Method: Method in which participant found job (ex. typed, refinement, etc.)
- Query Length: Number of terms inputted inside search box
- Query Type: what kind of information inputted inside search box (ex. keyword, education, etc.)
- Location: What kind of location inputted inside location box (ex. state, city, zip) and which locations participants inputted inside search box (ex. Austin, New York, etc.)

The output search results relate to the following terms:

- Result Page: Page in which result job target is found
- List Rank: Position of target job
- Job Position in First Page: Job position that participants clicked in the first page

4   Results

A total of 264 searches are performed. 211 of them are found by the result of using indeed.com. 53 of the searches are reached via hyperlink from indeed.com and they subsequently used the internal search or job board at the destination site.

4.1   Search Inputs

Input Method
Input method refers to the mode in which participants found a job. Table 1 displays the results.

Table 1. Input method to find jobs

Input Method	Number of Cases	Percent
Typed in main search box	82	69%
Refinement by menu option	19	16%
Linked by "Salary Search" to jobs	7	7%
Other options	10	8%

69 percent of all searches are conducted by typing words inside the main search box. Refinement by menu options contributed to 16 percent of all searches. For instance, after their initial search, on the left hand side, if they want to narrow down their search to jobs with salaries that exceed $50,000, this would be coded as refinement by menu options. In one of the browser tools, indeed.com offers "Salary Search", an option to look at how much salary a job pays. 7 percent of the jobs are reached by an internal hyperlink from "salary search" to the destination job. Other options include using other browser tools and the "Advanced Job Search".

Query Length

Query length represents the number of terms inputted inside the search box. Table 2 shows the results.

Table 2. Query length

Terms per Query	Number of Cases	Percent
1	108	50%
2	80	37%
3	24	11%
4	3	2%

The mean terms per query are 1.66. From the total, approximately 50 percent of searches are reached by typing in one term per query. Even though it is not shown, there are 9 instances of using Boolean operators. Those include using commas, addition sign, and the quotation mark.

Query Type

Query type denotes what kind of information is inserted into the search box. Examples are job title, keywords, and education. Job title describes a title such as junior analyst or senior manager. Keywords can be marketing, usability, health care, and education.

Table 3. Query type

Query type	Number of Cases	Percent
Job title	52	42%
Keyword	51	41%
Company name	10	8%
Salary level	9	7%
Education level	2	2%

According to Table 3, there are five different methods to search for job information: job title, keyword, company name, education level, and salary level. There were 94 cases of entering a unique input, but also 14 cases of a combination of different input types (ex. keyword + company name). As an example, "doctor information science" would be classified as education level and keyword. Some of the searches are very specific to the participant, but most of them used general terms.

Location

Compared to other search engines, job search engines have a location box next to the main search box. Because most job seekers are exact about where they want to work, job search engines provide the option to customers. Based on the participant's results, the majority of searches are city names (98 percent) followed by state and zip code. They had a clear desire to find a job in Austin (85percent). Very occasionally, participants would find jobs in Texas cities encompassing San Antonio, Houston, and Dallas (3 percent). Additional cities involve major metropolitan cities such as New York, Los Angeles, and San Francisco.

4.2   Search Outputs

Result Page

Result page refers to the page in which the job target is found and clicked. Table 4 contains the results.

Table 4. Result page

Result page	Number of Cases	Percent
Page 1	142	67%
Page 2	31	14%
Page 3	13	6%
Page 4	13	6%
Page 5	5	2%
Page 6+	7	3%

During the browsing session, participants clicked on the job hyperlink in the following pages. The mean value of the job target result page is 1.77 pages. 67 percent of the target jobs are found in the first page after a search is conducted.

List Rank

Within a result page, the job posting is positioned among the ten available hyperlinks from top to bottom. Table 5 has the mean value of job position that is clicked.

Participants tend to click on job in the upper echelon. In the first page, the mean value of job position that is clicked is 4.26. This tendency is also displayed in the second and third pages. From the fourth page and afterwards, participants clicked on jobs in the middle or the bottom echelon. Sponsored jobs displayed on the top, bottom and right hand side are clicked 12 times. There was strong trend to observe the jobs closely in the first few pages, but after a while, participants would automatically scroll the mouse toward the middle in the latter pages. This may possibly be a reason why the latter pages have a higher mean value of job positions that are clicked.

Table 5. List rank

Result page	Mean Value of Job Position that is Clicked
Page 1	4.26
Page 2	4.96
Page 3	4.38
Page 4	5.61
Page 5	6.20
Page 6+	6.57

Job Position in First Page

Since most participants clicked on job results in the first page, we decide to look at what position it is located in the first page. Table 6 shows the results.

Table 6. Job position in first page

Job Position in First Page	Number of Cases	Percent
1	28	19%
2	19	13%
3	17	11%
4	14	9%
5	10	7%
6	10	7%
7	8	5%
8	9	6%
9	6	4%
10	9	6%
Sponsored Job	12	8%

From Table 6, within the first page, we can see that 19 percent of jobs clicked come from the first result displayed. The number of clicks drops subsequently after the first job position in the first page.

5   Limitations

A primary limitation of our study is the small sample size that completed the experiment. Traditional log file analysis on search engines use machines to detect input and output search results from a large population. Since our intentions also involved observing behavioral aspects, we end up with a small sample. Larger groups still produce useful results, but the cost of recruiting and the extra effort needed to run, test and analyze the results leads to rapidly diminishing returns. In conclusion, we understand that a sample size of 10 may not be representative of the entire population. Future research should address these issues with a larger sample size for data mining studies and behavioral studies.

Also, our sampling method has inherent bias because of the snowball technique. From this experiment, because most of the participants are about to enter the job market after graduation, it misses out on those who desire to change their jobs from a previous job. Because the researcher has no sampling frame, participants that would have met our inclusion criteria are excluded. Another problem is that via the snowball technique, only including participants with education levels with masters and above does not represent the general population.

Another potential problem is that job search does not occur at a single time, but occurs over a period of time. Participants mentioned that whenever they feel like they have to find a job, they would come back to look for jobs on the Web. In that case, new jobs would appear in the job search engine. Then, there is no need to find jobs thoroughly on the Web. They would wait for the ideal job to appear. Every case is different depending on the person, but job search normally is a longitudinal process.

6   Conclusion

Domain-specific Web search engines are effective tools for reducing the difficulty in acquiring information from the Web. In contrast to general-purpose search engines, domain-specific search engines indexes Web pages that contain only a small subset of documents relevant to the field. We assume that the domain is constrained and documents within this field have common elements. By extracting meaningful information out of this domain can yield valuable evidence pertaining to the field. An understanding of how searchers utilize these systems is critical for further refinement and evaluation of domain-specific search engines.

Job search engines play a big role for job seekers and we need to understand the trends in how they conduct their search. In this study, we investigated queries from a job search engine, indeed.com. We specifically analyzed input and output search queries. There is still territory undiscovered on studies involving job search engines. Future research can address these issues.

References

1. Krueger, A.: The Internet is lowering the cost of advertising and searching for jobs. New York Times C2 (July 20, 2000)
2. Can, A.B., Baykal, N.: MedicoPort: a medical search engine for all. Computer Methods and Programs in Biomedicine 86(1), 73–86 (2007)
3. Kruger, A., Giles, C.L., Coetzee, F., Glover, E., Flake, W.G., Lawrence, S.: Deadliner: Building a new niche search engine. In: CIKM 2000, Washington, DC, USA (November 2000)