



Using the web to look for work

Implications for online job seeking and recruiting

Online job
seeking

Bernard J. Jansen and Karen J. Jansen

Penn State University, University Park, Pennsylvania, USA, and

Amanda Spink

University of Pittsburgh, Pittsburgh, Pennsylvania, USA

49

Abstract

Purpose – The web is now a significant component of the recruitment and job search process. However, very little is known about how companies and job seekers use the web, and the ultimate effectiveness of this process. The specific research questions guiding this study are: how do people search for job-related information on the web? How effective are these searches? And how likely are job seekers to find an appropriate job posting or application?

Design/methodology/approach – The data used to examine these questions come from job seekers submitting job-related queries to a major web search engine at three points in time over a five-year period.

Findings – Results indicate that individuals seeking job information generally submit only one query with several terms and over 45 percent of job-seeking queries contain a specific location reference. Of the documents retrieved, findings suggest that only 52 percent are relevant and only 40 percent of job-specific searches retrieve job postings.

Research limitations/implications – This study provides an important contribution to web research and online recruiting literature. The data come from actual web searches, providing a realistic glimpse into how job seekers are actually using the web.

Practical implications – The results of this research can assist organizations in seeking to use the web as part of their recruiting efforts, in designing corporate recruiting web sites, and in developing web systems to support job seeking and recruiting.

Originality/value – This research is one of the first studies to investigate job searching on the web using longitudinal real world data.

Keywords Jobs, Internet, Search engines, Recruitment advertising

Paper type Research paper

Introduction

The web is a worldwide information resource, and it is important to understand emerging searching trends for locating and utilizing this information. It is also important to understand the impact these trends have on the way people conduct the daily business of their lives. For example, the web has dramatically changed the way job seekers find positions. As of July 2002, over 52 million Americans have conducted online searches for information about jobs, with more than 4 million Americans doing so on a typical day, representing a 60 percent increase from 2000 (Boyce and Rainie, 2002). Recent demographics for web job searchers suggest that there is an equal mix of males and females and 61 percent are age 18-29 (Boyce and Rainie, 2002). Furthermore, 57 percent of whites, 43 percent of African-Americans, and 47 percent of Hispanics have web access (Rainie and Packel, 2001), and within minority groups, nearly 60 percent have conducted online job searching (Boyce and Rainie, 2002). These statistics



show that the web is a tool that is capable of reaching various segments of the job searching market. Although these numbers are encouraging, using the web as a recruitment tool can still lead to disparate impact and these demographic trends should be kept in mind when using online recruitment (Stanton, 1999).

Recognizing the increasing and more diverse traffic on the web, companies are beginning to advertise and post position openings online. In fact, Leonard (2000) reported that 75 percent of *Fortune* 500 companies are posting jobs to their corporate sites, and less than a year later, Capelli (2001) reported that 90 percent of large US companies are using the web for recruiting. In addition, more than 75 percent of human resource professionals are now using web job boards to supplement traditional recruiting methods (HR Focus, 2000). It is evident that web-based recruiting and job search is now a major trend, reflecting the growing use of the web for commercial purposes (Lawrence and Giles, 1999; Spink and Jansen, 2004; Spink *et al.*, 2002).

Several articles have addressed web searching in general (Arasu *et al.*, 2001; Montgomery and Faloutsos, 2001; Silverstein *et al.*, 1999), including those focusing on topic-specific searches. Many recent articles have been published focusing on how corporations can enhance their recruitment process by using the web (Boehle, 2000; Dysart, 1999; Gutmacher, 2000; Leonard, 2000). Other articles have examined the role and importance of job boards such as Monster.com, CareerMosaic, and HotJobs (Donovan, 2000; Gordon, 2002; Gutmacher, 2000). With the focus on job boards and corporate recruiters, it is somewhat surprising how little research attention has been paid to the job seekers themselves. With one recent exception (Feldman and Klass, 2002), we could locate no other published research exploring how job seekers locate job-related information and positions on the web and how successful they are in finding relevant information.

In this study, we address this shortcoming in the literature by examining trends in job-related searching across a five-year period. We report the results of this examination as well as the effectiveness of job-related searching. Finally, we examine the extent to which job seekers are meeting their objectives (e.g. job postings, applications) for those searchers specifically searching for positions. Following a review of literature, we present the methodology utilized to obtain and analyze actual web queries. We analyzed these queries to determine trends in searching and term use over time. We also submitted subsets of these queries to major web search engines to evaluate the relevance of retrieved information. The implications of these results are then discussed for both those searching for positions and those desiring to attract job seekers to their organizational web sites.

Related studies

Job-seeking studies

On the whole, there has been very little research on the recent trend to utilize fully the web as part of the recruiting process. The majority of published articles describe how companies are currently using the web and outline the issues and concerns associated with web recruiting (c.f. Boehle, 2000; Capelli, 2001; Dysart, 1999; Epstein and Singh, 2003; Leonard, 2000; Munger, 2002; O'Leary *et al.*, 2000; Stanton, 1999). Most of these articles adopt a corporate or organizational perspective, providing hints and tips for organizations to find candidates (e.g. Capelli, 2001; Leonard, 2000), paying little attention to how potential candidates find information about organizations and

organizations' posted job-related information. On a related note, there appears to be an implicit assumption in many of these prescriptive articles that active job seekers can readily find job postings and applications on corporate web sites (e.g. Dysart, 1999), which is a potentially erroneous conclusion. Much attention is also dedicated to finding passive job candidates (i.e. those who are not actively looking for work but may be interested in a new position). These candidates are least likely to be using the very popular job boards (Boehle, 2000), suggesting that it may be beneficial to find out more information about how individuals broadly search the web for job-related material, beyond the limited scope of job boards. This is especially important since corporate human resource managers prefer their corporate web sites relative to job boards (Boehle, 2000).

Recent empirical articles have examined the attractiveness of a corporate web site in attracting potential job applicants (Dineen *et al.*, 2002; Kuhn and Skuterud, 2000; Scheu *et al.*, 1999; Supjarendeek *et al.*, 2002). These articles focus on attractiveness and the fit between the corporate web site and job seeker characteristics. Similar to the descriptive articles above, this stream of research focuses on the organization's ability to attract candidates once they have arrived at their web site. This is certainly an important research topic, but it tells us very little about how job seekers search for and find these web sites.

One notable exception is a study conducted by Feldman and Klaas (2002), who examined the experiences of managers and professionals searching for jobs via the web. These researchers surveyed recent MBA graduates to ascertain the job search behaviors and strategies they employed, and to examine the perceived effectiveness of web searching and the difficulties encountered. Their measure of perceived effectiveness was somewhat limited (one item, forced ranking), but they found that 29 percent of respondents reported that the web was the most helpful strategy for finding a job, compared to 40 percent who believed that networking was the most useful. Perhaps the most enlightening results were those concerning difficulties encountered when searching the web. Almost 43 percent of respondents reported that there were not enough relevant jobs listed to make its use worthwhile. Thirty-three percent of respondents reported that the companies' web sites lacked relevant data about jobs. Finally, 10 percent of respondents had difficulty even finding company web sites. These findings shed light on the user's experience in both searching for and finding relevant job-related information.

Web studies

Although there is little research of online job searchers, there is a growing body of literature that examines how people in general search on the web (Hölscher and Strube, 2000; Jansen and Pooch, 2001; Spink *et al.*, 2002). This research provides some insight into how people search for information on the web and provides a broad framework for considering the job-related search process. Jansen and Pooch (2001) present an extensive review of the web searching literature, reporting that web searchers exhibit different search techniques than do searchers on other information systems. Hölscher and Strube (2000) report information on sessions, queries, and terms, noting that experts exhibit different searching patterns than novices. Finally, in analyzing trends in web searching, Spink and colleagues (2002) report that web searching has remained relatively stable over time, although they noted a shift from entertainment to

commercial searching. The researchers noted an increase in general e-commerce searching (i.e. commerce, travel, economy, and jobs) from about 11 percent of all queries in 1997 to nearly 25 percent in 2001. This reported trend mirrors results from survey data (Boyce and Rainie, 2002) and corresponds to the increase in the quantity of available commercial information on the web (Lawrence and Giles, 1999). Although not directly applicable to job searching, this stream of research on general web searching techniques provides useful information and a methodology for examining web job searchers.

From a synthesis of the existing literature on the topic, we observe the following. First, one study provides an understanding of how people actually search for jobs online (Feldman and Klass, 2002). Although a corporation may post recruiting material on their corporate web site, we know very little about how potential job applicants locate this material. Second, there appears to be no published information on how successful online job seekers are in obtaining relevant information. We do not know if searchers are actually locating job positions or if organizations that desire to recruit online are effectively utilizing the web as a resource in their recruiting efforts. There seems to be an implicit belief that once an applicant “goes online,” he or she unquestionably locates the desired information. This study challenges these prevailing assumptions by considering the information needs of the job searchers and examining how these information needs are translated into queries that are submitted to web search engines. We then analyze the effectiveness of actual job-related searches in obtaining relevant results. Finally, for those specifically searching for jobs, we examine the extent to which job seekers are finding what they are looking for (e.g. job postings, applications).

Research questions

Our first research question focuses on how people search for job-related information on the web. To investigate this question, we obtained, and qualitatively analyzed, actual job-related queries submitted to a major web search engine. For this research question, we sought to examine the characteristics of job-related queries, investigating areas such as the number of terms in a query, the number of queries in a session, and the use of query operators.

Our second research question examined the effectiveness of these searches in locating job-related information. This involved resubmitting a subset of these real-world queries to a web search engine and evaluating the retrieved results in order to determine the effectiveness of the queries. We extracted these queries from the complete data set using key terms, and the queries encompass a wide range of job-related information needs. We are interested in how effective these queries are at retrieving relevant information. This has implications not only for job seeker using the web, but also search engines and web sites that serve these users.

For the third research question, which questioned the likelihood of job seekers finding a job posting or application, we qualitatively extracted queries from the data set that the researchers deemed were position-specific (i.e. queries relate to directly finding a position). We submitted a subset of these queries to three major web search engines. Independent evaluators judged the retrieved sites to determine whether or not they contained a job posting or application materials.

Research design

Data collection

The queries we obtained for this research had been submitted to Excite (www.excite.com), a major web search engine at the time of data collection. Excite provided us with three transaction logs, each holding a large and varied set of queries (approximately one million). The transaction logs spanned several hours of user searching on the following dates: 16 September 1997 (Tuesday, midnight to 8 a.m.), 1 December 1999 (Wednesday, 9 a.m. to 1 p.m.), and 30 April 2001 (Monday, midnight to midnight)[1]. Excite was the second most popular web site in 1997 (Munarriz, 1997), and was the fifth most popular in 1999 and 2001 as measured by number of unique visitors (Cyber Atlas, 1999, 2001).

Each record within the transaction log contained three fields:

- (1) *Time of Day*: measured in hours, minutes, and seconds from midnight of each day as logged by the web server.
- (2) *User Identification*: an anonymous user code assigned by the Excite server.
- (3) *Query Terms*: terms exactly as entered by the given user.

With these three fields, we located a user's initial query and recreated the chronological series of actions by each user in a session[2].

Data analysis

From the three complete transaction logs, we were interested in only those queries that were job-related. We therefore culled a subset of queries pertaining to job hunting and job related information using a modified snowball sampling technique (e.g. Patton, 1990). More specifically, we started with several seed terms (i.e. job, employment, and hiring) that are central indicators of job-related searching based on a standard human resource textbook (Sherman Jr *et al.*, 2000). Using this set of terms, we extracted all records from the 1997 transaction log that contained these terms. We then reviewed the extracted records identifying other terms that frequently appeared. These new terms were then combined with the set of original terms, and from the original transaction log we extracted all records that contained these terms. The process was repeated until the addition of new terms to the set added less than ten new and unique queries. We repeated the process for the other two transaction logs.

Query analysis. We then qualitatively analyzed the retrieved subset to identify queries that were obviously not job-related. For example, we eliminated queries such as interview with a vampire and work out. We employed the same process on the other two transaction logs. At this point, we were satisfied that we had retrieved a subset of each of the transaction logs that contained solely job-related searches. This set of data was used to address the first of our three research questions.

Effectiveness. For the second research question, a random set of 100 queries was pulled from the 2001 job-related data set for closer examination. Each of the 100 queries was submitted to Microsoft Search (www.msn.com) because it is the most popular web site as measured by number of unique visitors and has one of the largest document collections used by any web search engine (Nielsen/Netrating, 2002). After each query was submitted, the web site addresses for each of the top ten results were saved. We chose this number of results because reported statistics show that approximately 80 percent of web users never view more than the top ten or so documents (Jansen *et al.*,

1998; Silverstein *et al.*, 1999; Spink *et al.*, 2002). If a query retrieved fewer than ten results, then that number of results was utilized. The 100 submitted queries retrieved a total of 969 results.

Three independent reviewers were hired to visit each of the 969 web sites to determine whether the result was a topically relevant document based on the query. The reviewers received training regarding the judgment process and were given written instructions for determining topical relevance. Topical relevance was a binary judgment and coded as 1 for topically relevant and 0 for topically non-relevant. Agreement across the three raters was calculated using r_{wg} , and was found to be fairly high ($r_{wg} = 0.78$), especially considering the subjectivity associated with determining topical relevance.

Job query analysis. The final research question focused on those actually looking for jobs (not just job-related queries) and their ability to find a job posting in the top ten retrieved results. To isolate such queries, the researchers reviewed the job-related queries from the 2001 data set and choose a separate subset of 110 multi-term queries that, in context, clearly indicated interest in locating or obtaining a job. For example, the following queries were included: fashion jobs in Los Angeles, NBA employment, Johns Hopkins University employment, and teenage summer jobs.

These queries were then submitted to three of most popular search engines at the time of the study: America Online, Microsoft Search and Google (Cyber Atlas, 2001). We were again interested in only the top ten results retrieved by each query. This set of 110 queries retrieved a combined total of 3,088 results from the three search engines. Once again, if less than ten results were retrieved, then that number was utilized. An independent reviewer evaluated each of these results, making a relevance judgment, and assigned the appropriate relevance code. In this case, relevance was determined based on whether the retrieved page had a job position posted (1) or not (0).

Results

Table I presents overall descriptive data for both the entire transaction log and the job-related subset for each year. Overall data was reported in Spink *et al.* (2002).

One general observation, before more examining closely the job-related searching trends, is that the number of users held fairly steady at approximately fifteen hundred across the five-year period. In addition, although total queries held constant in the complete transaction log, the number of job-related queries varied over time, with a significant dip in 1999. To help explain this trend, we obtained the unemployment data for the time periods of our data collection (Bureau of Labor Statistics, 2002) because unemployment rates during the period of study may have impacted online job searching. As expected, the unemployment rate decreased from 4.9 percent in September 1997 to 4.1 percent in December 1999, rising again to 4.5 percent in April 2001. This trend correlates with the dip in job-related queries, suggesting that the lower unemployment rate in 1999 may indeed have influenced job-related searching.

Job-related search characteristics

The number of queries in a session and the number of terms in a query are indicators of the complexity of the searcher's information need, with a greater number of queries and terms indicating greater complexity. The number of single queries in a session is substantially higher than for the overall transaction logs. This means that a large

	1997*	1997	1999*	1999	2001*	2001
	Entire	Job only	Entire	Job only	Entire	Job only
Sessions	211,063	1,637	325,711	1,451	262,025	1,568
Queries	1,025,908	2,711	1025,910	1,982	1,025,910	2,265
Terms	1,277,763	9,447	1,500,500	7,021	1,538,120	9,050
Queries in session						
1 query	<i>48.4</i>	<i>63</i>	<i>20.8</i>	<i>77</i>	<i>30.8</i>	<i>75</i>
2 queries	<i>60.4</i>	<i>22</i>	<i>19.8</i>	<i>17</i>	<i>19.8</i>	<i>16</i>
3 + queries	<i>55.4</i>	<i>15</i>	<i>19.3</i>	<i>6</i>	<i>25.3</i>	<i>10</i>
Mean queries per user	<i>2.5</i>	<i>1.7</i>	<i>1.9</i>	<i>1.4</i>	<i>2.3</i>	<i>1.4</i>
Users modifying queries	<i>52.0</i>	<i>37</i>	<i>39.6</i>	<i>23</i>	<i>44.6</i>	<i>25</i>
Terms in query						
1 term	<i>26.3</i>	<i>13</i>	<i>29.8</i>	<i>14</i>	<i>26.9</i>	<i>15</i>
2 terms	<i>31.5</i>	<i>19</i>	<i>33.8</i>	<i>20</i>	<i>30.5</i>	<i>15</i>
3 + terms	<i>43.1</i>	<i>68</i>	<i>36.4</i>	<i>66</i>	<i>42.6</i>	<i>69</i>
Mean terms per query	<i>2.4</i>	<i>3.49</i>	<i>2.4</i>	<i>3.55</i>	<i>2.6</i>	<i>4.00</i>
Boolean queries	<i>5.0</i>	<i>5</i>	<i>5.0</i>	<i>11</i>	<i>10.0</i>	<i>15</i>
Percentage usage of 100 most frequently occurring terms	<i>17.9</i>	<i>61</i>	<i>19.3</i>	<i>59</i>	<i>22.0</i>	<i>59</i>
Terms not repeated in the data set	<i>57.1</i>	<i>10</i>	<i>61.6</i>	<i>14</i>	<i>61.7</i>	<i>13</i>
Result pages viewed per query						
1 page	<i>28.6</i>	<i>31.8</i>	<i>42.7</i>	<i>46.4</i>	<i>50.5</i>	<i>44.7</i>
2 pages	<i>19.5</i>	<i>41.6</i>	<i>21.2</i>	<i>25.3</i>	<i>20.3</i>	<i>21.7</i>
3 + pages	<i>51.9</i>	<i>26.6</i>	<i>36.1</i>	<i>28.2</i>	<i>29.2</i>	<i>33.5</i>

Note: All numbers in italics are percentages

Table I.
Comparative statistics for
entire transaction log and
job-related query data

majority of job searchers are making only one attempt at finding relevant information before ending their search session. Looking at the number of terms within queries, we see that the number of queries with three terms was substantially higher for the job searching queries than for the overall data sets. The use of Boolean operators by job related searchers is also about 5 percent higher than in the general web population. This may indicate that there is something about job-seeking queries that lend themselves to the use of Boolean operators. This is similar to what prior research has noted in the area of multimedia information searching, where image queries have a high occurrence of Boolean operators (Jansen *et al.*, 2003). Job seekers originally viewed fewer results pages (i.e. the number of retrieved uniform resource locators retrieved by the search engine and presented to the user usually in chunks of ten) than the general web searcher, although this trend appears to be reversing based on the 2001 data. Combining these findings, we seem to have contradictory indicators. The majority of the sessions are very short but the queries are relatively lengthy. It could be that the lengthier queries are locating the required information therefore resulting in shorter sessions. Conversely, it could be that even with the longer queries, users are not locating the information they need and are just ending their search on this search engine. We will return to this point when we investigate relevance in the next section.

Terms are the building blocks of queries, representing, to some degree, the information need of the searcher. The distribution of term usage within a large set of web queries generally follows a Zipf distribution (Jansen *et al.*, 2000), with a relatively small set of terms used quite frequently and a large set of terms used relatively infrequently. For the complete

transaction logs, the set of the 100 most frequently utilized terms represented 18 percent to 22 percent of the total term usage. However, for the job-related queries, the 100 most frequently used terms accounted for 59 percent to 61 percent of the total terms. The percentage of terms used only once was quite low relative to the general web population. Importantly, there is a very tight jargon used for online job searching, which implies that there are common information needs across job searching individuals. Armed with this information, companies can design their web sites to include these terms in order to maximize the likelihood of having their site returned in the results list during a job-related search.

Table II presents the term frequencies for all three years from the job-related data sets. To arrive at this list, we first sorted all terms within the data set in descending order by frequency of term occurrence. From the list of highest-ranked terms, we then removed the terms without information content (e.g. and, or, is, the), known as stop words. We then selected the top 25 terms from each of the three data sets. These were merged into one list, resulting in a total of 33 unique terms, as presented in the first

Terms	1997	Frequency 1999	2001
bank	38	40	32
California	50	25	23
Canada	48	30	17
career	169	44	70
careers	30	10	3
education	13	32	56
employers	51	14	10
employment	1,196	570	520
example	1	1	41
experience	1	20	42
federal	46	29	8
Florida	11	24	28
home	24	28	25
human	79	37	26
insurance	6	16	27
job	462	271	299
jobs	477	180	183
listings	114	19	10
monster.com	2	50	108
office	7	15	25
opening(s)	66	19	39
opportunities	277	67	50
positions	38	-	3
recruiters	53	48	36
resources	79	38	29
résumé	211	260	382
résumés	121	89	88
retirement	77	115	104
search	32	32	42
state	41	42	30
Texas	42	22	14
unemployment	50	71	111
work	107	104	153

Table II.
Top occurring terms and
frequencies for 1997,
1999, and 2001

column of Table II. We then provided the frequency of occurrence for all terms from all three data sets (even if it was not in the top 25 in other years) to allow for better interpretation of trends over time.

The most frequently occurring terms for all three time periods were “employment” and “job”, with the frequency of occurrence trending downward. This was generally the case for most of the top terms. In 1997, the top 25 terms accounted for 63 percent of all term occurrences. By 2001, the top 25 terms accounted for 50 percent of all term occurrences, indicating a broadening of the job searching jargon. However, the core set of high-use terms was fairly stable. Of the 25 terms appearing in 1997, 17 also appeared on the 2001 list.

There were also some interesting new additions in the top 25 lists over the five-year period. Of the 25 terms in the 2001 list, nine did not appear in the 1997 list, and four did not appear in 1999 list. The rapid increase in occurrences of “monster.com” (from two in 1997 to 108 in 2001) is indicative of the influence job boards have had on job-related searching. The term “resume” also showed a significant increase, which can be attributed to a combination of factors. First, this trend mirrors the increased use of job boards. Second, many articles have urged recruitment experts to find passive candidates online by searching for “resume” and resume.htm on the web (Gutmacher, 2000; Leonard, 2000). Locations such as California, Canada, and Texas, which appeared in the 1997 and 1999 list, were not in the top 25 of the 2001 list, although their occurrences were still relatively frequent. The term “Florida” with only 11 occurrences in 1997 steadily increased, with 28 occurrences in 2001. The occurrence of these terms indicate that location, and specific locations, are of high interest to job seekers. Overall, these trends provide interesting insights into job-related search terms and trends.

Although a term analysis is worthwhile, it is important to examine how these terms are utilized in conjunction with other terms given that most of the job-related queries are three or more terms. We analyzed the job-related queries using term co-occurrence analysis, which looks for the simultaneous occurrence of terms within queries (Leydesdorff, 1989). Figure 1 presents the term co-occurrences for the 1997 data set for the top 25 most frequently occurring terms in a correlation matrix fashion. The three most frequently occurring pairs were (1) “jobs” and “employment” (204 co-occurrences), (2) “employment” and “opportunities” (181), and (3) “jobs” and “opportunities” (106).

Figure 2 presents the same information for the 1999 data set. Here, the three most frequently occurring pairs were: first, “employment” and “opportunities” (48, a substantial drop from 1997); second, “human” and “resources” (37); and third, “job” and “bank” (35). While there is still a clustering of co-occurrences around the most frequent terms, the clustering is less pronounced and the distribution of co-occurrences is more disperse relative to 1997. Another change is the increase in the co-occurrence of the term “unemployment” from only two co-occurrences in 1997 to 11 co-occurrences in 1999. Overall, the co-occurrences are lower and sparser, indicating a broadening of the job searching terms used.

Figure 3 presents the term co-occurrences for the 2001 data set for the top 25 most frequently occurring terms. Generally, the co-occurrence distribution trend has continued, with a sparser allocation. The most frequently occurring pairs in 2001 were “resume” and “example” (157); “resume” and “experience” (100); and “resume” and “education” (100). However, the most prevalent pair in 1997 (“employment” and “job”),

	bank	California	Canada	career	careers	employers	employment	federal	human	job	jobs	listings	openings	opportunities	positions	recruiters	resources	resume	resumes	retirement	search	state	texas	unemployment	work	
bank	--																									
California	--	--																								
Canada	1		--																							
career	3	3		--																						
careers			2	9	--																					
employers		2		11	7	--																				
employment	3	26	22	66	10	16	--																			
federal	1	1	5		1		11	--																		
human		4	12	1	1				4	--																
job	34	14	2	49	8	13	92	13	2	--																
jobs		18	6	62	11	27	204	19	2	74	--															
listings		4	2	15	4	9	68	2	55	44		--														
openings		4		7	2	9	28			49	30		--													
opportunities	4	8	3	51	9	13	181	9		84	106	25	19	--												
positions		1		8	2	10	19			16	20	9	13	21	--											
recruiters				10	4	3	9			5	4	2	2	5	2	--										
resources		4	12		1			4	74	3	2						--									
resume		4	3	5		2	3	1	1	10	3			1	2	1		--								
resumes				8	3	3	11			5	16	1		8	5		15	--								
retirement		1	1						3			2		1						--						
search			3	4	6			10			18	6	6	2			3	1			--					
state			4	1			18		1	12	7	6		3			1			2		--				
texas	2			1			18			12	6		1	2						9			--			
unemployment																						2	2	--		
work			2	1		3	3	2		2	8	2	4	7	5			1					1	2	--	--

Figure 1.
Frequency of term
co-occurrence for top 25
terms for 1997

occurs only 20 times in 2001. Combined with the appearance of other terms on the 2001 list (e.g. “education”, “home”, “insurance”), it appears that the job-searching language may be moving from broad terms (e.g. “employment”, “opportunities”) to narrower and more specific terms. Thus, although employment was still the most prevalent term (Table II), the terms paired with it are getting broader and more dispersed, suggesting that searchers are getting more sophisticated and specific in their searching.

Effectiveness of job-related searches

Our second research question addressed the effectiveness of job-related searches in obtaining relevant results. Using the random subset of queries from the 2001 data set, we submitted them to a web search engine and had three independent raters evaluate the results to determine topical relevance. This analysis helps address the question of whether job-related search sessions are short because the searchers are finding the information that they need or that they are not finding the information they need and just giving up or going elsewhere.

Relevance is the key measure for determining relative precision, which is a standard metric to evaluate information system performance (Korfhage, 1997). Relative precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved at a certain position in a results list. So, if, for example, one relevant document was retrieved out of ten, with the other nine being not relevant, precision would be 0.01 (i.e. $0.01 = 1 \text{ relevant} / 10 \text{ retrieved}$). This specific metric is

	bank	California	Canada	career	education	employment	federal	Florida	home	human	job	jobs	monster.com	openings	opportunities	recruiters	resources	resume	resumes	retirement	search	state	texas	unemployment	work
bank	--																								
California	1	--																							
Canada	6		--																						
career	1	1		--																					
education		2		2	--																				
employment	1	7	13	1	5	--																			
federal						13	--																		
Florida					1	9	1	--																	
home				2	13				--																
human		1			1					--															
job	35	3	6	4	2	5	2	1	2		--														
jobs		3	6	6	2	30	3	5	5	2	1	--													
monster.com													--												
openings		1		1	1	4				15	3		--												
opportunities	1		1	4	4	48			1	14	14		3	--											
recruiters			1							1	1	1			--										
resources		1			1				37	3			1	--											
resume	1	1			1		1			2	2					--									
resumes			1								1						19	--							
retirement	2	1				6	4												--						
search				1		3				22	2				3		1			--					
state		2				13	2	1	2	7	7								7		--				
texas	1				1	9				3	4		1			4		1		1	--				
unemployment	1	5	3		1		1	5		1	1					1				8	1	--			
work		2				1	1		9		1	4									1	1	2		--

Figure 2.
Frequency of term co-occurrence for top 25 terms for 1999

referred to as P@10 (i.e. precision at 10). The topical relevance evaluations from the three raters were averaged to create one measure of relevance, which was used to calculate relative precision for the queries. Of the 969 documents retrieved within the top ten results, 506 (52 percent) results were judged to be topically relevant and 434 (48 percent) were judged to be topically non-relevant. Thus, the relative precision for the entire set of results is 0.52, meaning that on average just over half of the obtained results from job-related queries were topically relevant.

Locating a job posting

While making their relevance judgments, the raters also noted whether or not there was a job posting or a link to a job posting on the web site that was retrieved. Results revealed that only 23 percent (224) of the 969 documents contained job postings. This seemingly low number surprised us, although it should be noted that these were job-related, not necessarily job-specific searches. Therefore, our third research question focused on job-specific searches to determine the frequency by which searchers found a job posting in the retrieved results.

In this portion of the analysis, we used a separate sample of 110 job-specific queries from the 2001 data set which the researchers deemed were queries seeking job positions. These queries were submitted to three popular search engines (MSN, Google, and AOL) and retrieved web sites containing job postings were coded as relevant. From the total number of relevant documents, we again calculated P@10 for these queries. The results of this evaluation are displayed in Table III. The relative precision of these results was just over 0.39, meaning that approximately 40 percent of the

	bank	career	education	employment	example	experience	Florida	home	human	insurance	job	jobs	monster.com	office	opening	opportunities	recruiters	resources	resume	resumes	retirement	search	state	unemployment	work	
bank	--																									
career		--																								
education		1	--																							
employment	3		2	--																						
example			35		--																					
experience			42		35	--																				
Florida				9			--																			
home				7				--																		
human	1								--																	
insurance	1		2	6		2				--																
job	28	2	4	4			4	1	2	3	--															
jobs	1	1	4	16			6				17	--														
monster.com											1		--													
office				8				2			3			--												
opening											30	30			--											
opportunities		3		20						1	24	15			30	--										
recruiters				1						1							--									
resources	1								26		3							--								
resume		1	100		157	100		2		2	16	16			30	16			2	--						
resumes					1					2	2	1									--					
retirement							3	1		2												--				
search		1	1	7							30	4							1	1		--				
state	2			15			3				6	8				1					2		--			
unemployment							5			10			11										4	--		
work	1	2					1	15			2	5	3											2	--	

Figure 3.
Frequency of term
co-occurrence for Top 25
terms for 2001

Table III.
Job postings in results of
job-specific queries
across three popular
search engines

	America Online	Google	Microsoft Search	Total
Posting	422	427	373	1,222
No posting	624	592	645	1,866
Precision	40.3%	41.9%	36.6%	39.6%

retrieved results contained job postings. This percentage of relevant documents was fairly consistent across the three search engines.

Table IV provides a more in-depth examination of these queries and their topical relevant results. Results show that about 50 percent of the queries retrieved less than five relevant results within the top ten results, and research has shown that the vast majority of web searchers look at no more than the top ten results (Jansen and Spink, 2003). The most common occurrences were between five and eight relevant results, occurring 44 percent of the time.

We also wanted to examine this set of job-specific queries to see if there were recurring concepts that appeared as refiners of the search. Although various ontologies have been developed to classify web queries (Spink *et al.*, 2002), we could not locate any that specifically dealt with job seeking queries. Therefore, we performed a linguistic classification of the queries, modifying a technique from Enser (1995). The set of queries contained 419 terms, not including Boolean operators such as AND and OR.

Number of topical relevant results	Queries retrieving that number of results	Topical relevant results retrieved (%)
10	6	11.3
9	5	8.5
8	18	27.1
7	8	10.5
6	13	14.7
5	15	14.1
4	8	6.05
3	6	3.4
2	7	2.6
1	9	1.7
0	5	0
	100	100

Table IV.
Number of topical relevance results by occurrences

From these terms, we removed stop words (e.g. the, and, in, etc.) and general topic terms such as “jobs” and “employment”. We then collapsed terms that were obvious phrases, such as “United” and “States” into one phrase (e.g. “United States”). This resulted in 179 remaining query terms. We then classified each of these terms, resulting in seven job-related refiner groupings.

Over 45 percent of all refiners were “location” (e.g. city, state or country), far more prevalent than the next two most common refiners, “industry (e.g. airlines, hospital; 17 percent) and “skill set” (e.g. teaching, janitorial; 11 percent). “Company” refiners referred to specific organizations (e.g. IBM, Ford), accounting for 8.9 percent. The remaining refiners were “job sites” (e.g. classified ads), “government”, and “temporal” (e.g. summer employment), accounting for 8.4 percent, 7.3 percent, and 2.2 percent respectively.

Discussion

In reviewing the analysis of job-related searching over the last five years, some trends in the ways job seekers search for information on the web are apparent. First, their session lengths (i.e. the number of queries they submit) are very short, with the over 60 percent of the sessions containing only one query. Second, the number of terms in this query is relatively high, with the majority of job seekers submitting queries with three or more terms. Given that session and query length are indicators of the complexity of information need and searching expertise, these two findings are contradictory. Perhaps the shorter sessions are due to longer queries locating relevant information on the first search. However, our analysis suggests this is not likely given the 52 percent relevance rate we found. Another possibility is that the percentage of relevant documents obtained was enough to satisfy the searchers. Once again, this does not seem likely, especially considering other researchers’ findings that 43 percent of surveyed individuals could not locate enough relevant information on the web, 33 percent were dissatisfied with the information on companies’ web sites, and 10 percent had difficulty locating company web sites (Feldman and Klass, 2002). It may be that job seekers are using a variety of information sources.

Certainly, these results indicate that organizations that want to maximize the number of searchers finding their web sites must do a better job of designing these sites so that job applicants can find them. It would be useful for companies to take into account the terms that job seekers are utilizing to locate job information and positions on the web as they design their web sites. The data we provide in Table II and Figures 1–3 should prove beneficial in this regard.

In our analysis of the terms that job seekers utilize, there currently appears to be a very tight language that job seekers employ when searching, but there are also indications that this language is expanding. The terms “employment” and “job” were the most frequently occurring terms across the three time periods. Over that time, nine new terms appeared in the top 25 terms, with four appearing in the most recent time period. The core set of high-use terms was fairly stable, with 17 appearing throughout the five-year period; however, the percent usage of the top terms is generally trending downward. The top 25 terms accounted for 63 percent of all term occurrences in 1997 but only accounted for 50 percent of all term occurrences by 2001. This 13 percent drop indicates a broadening of the job searching language over time.

Similarly, in analyzing the co-occurrence of terms, the most prevalent term, “employment”, is being paired with a widening number of other terms over time. Combined with the appearance of other narrower terms on the 2001 list (e.g. “education”, “home”, “insurance”), it appears that the job searching language may be moving from broad terms (e.g. “employment”, “opportunities”, “jobs”) to more specific terms. This trend also suggests that searchers are getting more sophisticated and specific in their searching. Thus, although there are some search words that appear consistently, the speed of change in search patterns across the five years necessitates continual examination of search trends over time to ensure that job seekers can locate relevant job-related information.

Two terms showed increased usage over time: “monster.com” and “resume”. The emergence of “monster.com” as a frequent search word is certainly an indication of the influence that job boards have had on web job searching. Similarly, the increase in the occurrences of the term “resume” mirrors the job board trend. It may also reflect the increased recruitment of passive candidates, as companies are learning to “source” potential candidates by searching for “resume” and “resume.htm” across the web (Gutmacher, 2000; Leonard, 2000).

Although job boards are playing an increasingly important role in online recruitment, it is apparent from our term co-occurrence analysis that significant numbers of online job seekers continue to utilize general web search engines to locate job-related information on “resume examples” and “resume education”, which may be useful information for some web-based companies and service providers. Companies desiring to recruit online should also consider submitting their web documents containing job information to the major web search engines for indexing (i.e. the web page is added to the search engine’s information base), ensuring their pages will appear in search results.

Location terms, such as “California”, “Canada”, and “Texas”, had high occurrences in all sampling periods, and “Florida” steadily increased over time. Although three US states and Canada were the only locations appearing in the top terms, it is important to note that there were numerous occurrences of other US states and other countries. This corresponds with the finding that location was the most common refiner used in

searches. The obvious implication is for companies to include location descriptors for job positions and announcements posted on their web sites.

Our findings also suggest that the relative precision (i.e. the ability to find topically relevant results) of job-related searches was relatively low (52 percent). It seems apparent from this low percentage that the terms job seekers use do not necessarily correspond to the terms that are used on corporate web sites. While this percentage is low, the results are even less encouraging with respect to finding actual job postings. Only 23 percent of job-related searches and 40 percent of job-specific searches contained job postings. Of course, this may be the result of there being fewer available jobs in general rather than a shortcoming of corporate web design, but it would seem like a good recruiting strategy to link actual job postings or even a statement of current hiring needs directly to job-related search results as often as possible.

Strengths and limitations

This study contributes to web research and online recruiting literature in important ways. First, the data come from actual job seekers submitting genuine queries and looking for job information. Accordingly, it provides a realistic glimpse into how web users actually search, without the self-selection issues or altered behavior that can occur with lab studies or survey data. Second, our sample is quite large, with approximately 1,500 users within each time period, and more importantly, spans a five-year period, permitting us to examine and report on trends in searching over time. Finally, we obtained data from a very popular search engine at the time of data collection and conducted our relevance analysis on three of the largest search engines on the web to ensure that our results were generalizable.

As with any research, there are limitations that should be recognized. The sample data comes from one major web search engine, introducing the possibility that the queries do not represent the queries submitted by the broader online job-seeking population. However, Jansen and Pooch (2001) have shown that characteristics of web sessions, queries, and terms are very consistent across search engines. Another potential limitation is that we do not have information about the demographic characteristics of the users who submitted queries, so we must infer their characteristics from the demographics of web searchers as a whole.

The set of query terms is also an imprecise measure of actual information need. We have attempted to mitigate this shortcoming by using a modified snowball sampling technique for choosing job-related terms and by utilizing multiple independent reviewers in determining relevance. Finally, these queries may not reflect the type and kind of job searching that is occurring on the popular job banks, such as Monster.com or CareerBuilder.com. From our term analysis, it is apparent that these job banks will continue to play a major role in job searching activities. However, it is important to note that research shows that over 70 percent of online users utilize a general web search engine to locate other web sites and web documents (CommerceNet/NielsenMedia, 1997), suggesting that the data we report provides insight into the terms and specific sites for which individuals are searching, regardless of the particular system.

Conclusion

Overall, the data reported in this study provide a useful characterization of job-related information searching and give companies insight into the terms and pairs that are

most frequently used. Equipped with this information, companies can design their web sites to include these terms, provide more direct access to job postings and hiring needs, and reach a larger pool of job candidates. Further research should continue to examine the changing trends in searching and begin to explore more directly the manner by which individuals use job boards in an attempt to find job-related information.

Notes

1. Times are Pacific Time as recorded by the Excite web server.
2. Jansen and Pooch (2001) provide useful definitions for key terminology. A term is any series of characters separated by white space. A query is the entire string of terms submitted by a searcher in a given instance. A session is the entire series of queries submitted by a user during one interaction with the web search engine

References

- Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A. and Raghavan, S. (2001), "Searching the web", *ACM Transactions on Internet Technology*, Vol. 1 No. 1, pp. 2-43.
- Boehle, S. (2000), "Online recruiting gets sneaky", *Training*, Vol. 37 No. 5, pp. 66-74.
- Boyce, A. and Rainie, L. (2002), "Online job searching", Pew Internet and American Life Project, available at: <http://pewinternet.org> (accessed October).
- Bureau of Labor Statistics (2002), *Labor Force Statistics from the Current Population Survey*, Bureau of Labor Statistics, Washington, DC, available at: http://data.bls.gov/servlet/SurveyOutputServlet?data_tool=latest_numbers&series_id=LFS21000000 (accessed October).
- Capelli, P. (2001), "Making the most of online recruiting", *Harvard Business Review*, Vol. 79 No. 3, pp. 139-46.
- CommerceNet/NielsenMedia (1997), "Search engines most popular method of surfing the web", Commerce Net/Nielsen Media, available at: www.commerce.net/news/press/0416.html (accessed 30 August 2000).
- Cyber Atlas (2000), "US top 50 internet properties", *At Home/Work Combined Cyber Atlas*, available at: <http://cyberatlas.internet.com> (accessed December).
- Cyber Atlas (2001), "US top 50 internet properties", *At Home/Work Combined, Cyber Atlas*, May, available at: <http://cyberatlas.internet.com> (accessed 1 July).
- Dineen, B.R., Ash, S.R. and Noe, R. (2002), "A web of applicant attraction: person-organization fit in the context of Web-based recruitment", *Journal of Applied Psychology*, Vol. 87 No. 4, pp. 723-34.
- Donovan, D. (2000), "Opportunities online", *Forbes*, Spring, p. 72.
- Dysart, J. (1999), "HR recruiters build interactivity into web sites", *HR Magazine*, Vol. 44 No. 3, pp. 106-10.
- Enser, P.G.B. (1995), "Progress in documentation: pictorial information retrieval", *Journal of Documentation*, Vol. 51 No. 2, pp. 126-70.
- Epstein, R. and Singh, G. (2003), "Internet recruiting effectiveness: evidence from a biomedical device firm", *International Journal of Human Resources Development and Management*, Vol. 3 No. 3, pp. 216-25.
- Feldman, D.C. and Klass, B.S. (2002), "Internet job hunting: a field study of applicant experiences with online recruiting", *Human Resource Management*, Vol. 41 No. 2, pp. 175-92.

-
- Gordon, J. (2002), "Battle of the boards", *Forbes*, August, pp. 50, 170.
- Gutmacher, G. (2000), "Secrets of online recruiters exposed!", *Workforce*, Vol. 79 No. 10, pp. 44-50.
- Hölscher, C. and Strube, G. (2000), "Web search behavior of internet experts and newbies", *International Journal of Computer and Telecommunications Networking*, Vol. 33 No. 1-6, pp. 337-46.
- HR Focus (2000), "Online recruiting: what works, what doesn't", *HR Focus*, No. 3, pp. 11-13
- Jansen, B.J. and Pooch, U. (2001), "Web user studies: a review and framework for future work", *Journal of the American Society of Information Science and Technology*, Vol. 52 No. 3, pp. 235-46.
- Jansen, B.J. and Spink, A. (2003), "An analysis of web information seeking and use: documents retrieved versus documents viewed", *Proceedings of the 4th International Conference on Internet Computing*, Las Vegas, NV, 23-26 June, pp. 65-9.
- Jansen, B.J., Spink, A. and Pederson, J. (2003), "An analysis of multimedia searching on AltaVista", *Proceedings of the 5th ACM SIG Multimedia International Workshop on Multimedia Information Retrieval*, Berkeley, CA, pp. 186-92.
- Jansen, B.J., Spink, A. and Saracevic, T. (2000), "Real life, real users, and real needs: a study and analysis of user queries on the web", *Information Processing and Management*, Vol. 36 No. 2, pp. 207-27.
- Jansen, B.J., Spink, A., Bateman, J. and Saracevic, T. (1998), "Real life information retrieval: a study of user queries on the web", *SIGIR Forum*, Vol. 32 No. 1, pp. 5-17.
- Korfhage, R. (1997), *Information Storage and Retrieval*, Wiley, New York, NY.
- Kuhn, P. and Skuterud, M. (2000), "Job search methods: internet versus traditional", *Monthly Labor Review*, No. 123, pp. 3-11.
- Lawrence, S. and Giles, C.L. (1999), "Accessibility of information on the web", *Nature*, No. 400, pp. 107-9.
- Leonard, B. (2000), "Online and overwhelmed", *HR Magazine*, Vol. 45 No. 8, pp. 36-42.
- Leydesdorff, L. (1989), "Words and co-words as indicators of intellectual organization", *Research Policy*, Vol. 18, pp. 209-23.
- Montgomery, A. and Faloutsos, C. (2001), "Identifying web browsing trends and patterns", *IEEE Computer*, Vol. 34 No. 7, pp. 94-5.
- Munarriz, R.A. (2002), "How did it double?", Daily Double, available at: www.fool.com/DDouble/1997/DDouble970812.htm (accessed 10 November).
- Munger, R. (2002), "Technical communicators beware: the next generation of high-tech recruiting methods", *IEEE Transactions on Professional Communication*, Vol. 45 No. 2, pp. 276-90.
- Nielsen/Netrating (2002), "Top web properties", Nielsen/Netrating, available at: www.nielsen-netratings.com (accessed 27 September)
- O'Leary, B.S., Lindholm, M.L., Whitford, R.A. and Freeman, S.E. (2000), "Selecting the best and brightest: leveraging human capital", *Human Resource Management*, Vol. 41 No. 3, pp. 325-40.
- Patton, M.Q. (1990), *Qualitative Evaluation and Research Methods*, 2nd ed., Sage Publications, Inc., Newbury Park, CA.
- Rainie, L. and Packel, D. (2001), "More online, doing more: 16 million newcomers gain internet access in the last half of 2000 as women, minorities, and families with modest incomes continue to surge online", Pew Internet and American Life Project, available at: www.pewinternet.org (accessed 2 January 2002).

Scheu, C., Ryan, A.M. and Nona, F. (1999), "Company web sites as a recruiting mechanism: what influences applicant impressions?", *Proceedings of 14th Annual Meeting of the Society for Industrial and Organizational Psychology*, Atlanta, GA,.

Sherman, A. Jr, Bohlander, G. and Snell, S. (2000), *Managing Human Resources*, South-Western College Publishing, New York, NY.

Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. (1999), "Analysis of a very large web search engine query log", *SIGIR Forum*, Vol. 33 No. 1, pp. 6-12.

Spink, A. and Jansen, B.J. (2004), *Web Search: Public Searching of the Web*, Kluwer, New York, NY.

Spink, A., Jansen, B.J., Wolfram, D. and Saracevic, T. (2002), "From E-sex to E-commerce: web search changes", *IEEE Computer*, Vol. 35 No. 3, pp. 107-11.

Stanton, J.M. (1999), "Validity and related issues in web-based hiring", *The Industrial-Organizational Psychologist*, Vol. 36 No. 3, pp. 69-77.

Supjarerndee, S., Temtanapat, Y. and Phalavonk, U. (2002), "Recruitment filtering with personality-job fit model", *Proceedings of the International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, 8-10 April, pp. 46-51.