Information Technology & Tourism, Vol. 10 pp. 000–000
Printed in the USA.

1098-3058/08 $60.00 + .00

www.cognizantcommunication.com

AN ANALYSIS OF TRAVEL INFORMATION SEARCHING ON THE WEB

BERNARD J. JANSEN,* CHRISTOPHER C. CIAMACCA,* and AMANDA SPINK†

*College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA
†Faculty of Information Technology, Queensland University of Technology, Brisbane, QLD, Australia

In this article, the phenomenon of searching for travel information on the Web is reported. The issues of how predominant travel searching is on the Web, how people are searching for travel information on the Web, and what terms people are using to express their travel-related information needs are investigated. In this research, 2,465,145 interactions from 534,507 users of the commercial Web search engine, Dogpile.com, on May 6, 2005 are analyzed employing both quantitative and qualitative approaches. Findings show that, at most, approximately 6.5% of Web queries are for travel searching. Geographical information accounts for nearly 50% of this travel searching, with general travel information accounting for just less than 10%. An analysis of individual terms in travel queries shows there is substantial searching for travel-specific websites such as *mapquest*, *travelocity*, and *orbitz*. Travel searchers appear to ~~be~~ target-specific events, again showing a strong geographical bias along with a temporal component of the underlying information intent. The distribution of travel topics is skewed, with several topics being "very focused" and others being "very general." A classification scheme for travel-related Web queries was developed, which should be helpful for other researchers in the online travel searching area. The implications for both content providers of travel information and for searchers of travel information on the Web are discussed.

Key words: Web search engines; Web searching; Travel searching; Travel queries; Travel searching terms; Travel information search; Travel query terms

Introduction

Web searching has become a daily behavior for many people, with the Web now the first choice for many of those seeking online information (Cole, Suman, Schramm, Lunn, & Aquino, 2003; Madden & Rainie, 2003). Their tool of choice is a search engine, with more than 73% of people using a search engine to locate information on the Web (~~Nielsen Media, 2006~~). In order to understand, predict, and influence this usage, it is important to understand both how people use Web search engines and the Web-searching trends that are emerging within specific domains. Examining Web searching within specific domains is an important area of research that has the potential to

Address correspondence to Bernard J. Jansen, College of Information Sciences and Technology, The Pennsylvania State University, 329F IST Building, University Park, PA 16802, USA. E-mail: jjansen@acm.org

increase the understanding of Web searching, to advance the knowledge of Web searchers' information needs, and to influence positively the design of Web information systems.

This is especially true in the travel and tourism domain. The Pew Internet and American Life Project reported that 73% of American Internet users have obtained travel-related information on the Web (Madden & Rainie, 2003). A simple search on any major Web search engine for travel information returns billions of results. The high use of the Web to locate travel information points to the need for information systems that can assist people in finding relevant information. In order to design these systems and to make them effective, there must be an understanding of how people search for travel information on the Web.

However, there is limited knowledge of how people actually express their information needs when using Web search engines to locate relevant travel information. In this study, this shortcoming in the literature is addressed by examining how prevalent travel searching is on the Web and how it is conducted. Fodness and Murray (1999) presented a model of travel information searching. The research in this article addresses the operational aspects of this model's information searching strategy by investigating specifically how searchers express their traveling information needs to Web search engines. Pan and Turner (2006) also developed a framework for tourist information search, consisting of five phases (i.e., ongoing search, prepurchase search, planning search, enroute search, and after-trip search). The research in this article addresses all phases of this model. As such, the research presented in this article is a valuable contribution to the refinement of existing models and is applicable to other aspects of travel information searching research.

Following a review of literature, the research questions and the research design utilized to analyze actual queries from a Web search engine are presented. These queries are analyzed to determine the occurrence of and manner of searching, and term usage over time. The results of this examination are reported with a specific focus on how people conducted travel-related information searching on a Web search engine. The implications of these results are then discussed for both travel firms and those desiring to locate travel Websites.

Related Studies

Many researchers have examined demographic usage and behavioral characteristics of online travel behavior (Bonn, Furr, & Susskind, 1999; Morrison, Jing, O'Leary, & Cai, 2001; Weber & Roehl, 1999). Prior work on investigating travel information searching has primarily focused on single travel Websites, lab studies, and surveys.

Several research studies focused on individual travel websites, many times via surveys of customers of these sites. Vogt and Fesenmaier (1998) reported that leisure and recreation are key motivators for travel-related information searching, along with decision making. Lexhagen (2004) investigated why the online tourism market has not experienced the growth rate expected. The researcher hypothesized that one reason could be that travel websites do not provide enough support for customer searching during the purchase process. The results reported by Lexhagen also indicated that no value-added services were perceived as important in the postconsumption phase. It could also be that existing systems do not adequate support user needs in the tourism domain.

Laboratories studies have also investigated aspects of travel searching. Morrison et al. (2001) surveyed participants in order to measure their likelihood of booking travel online. The researchers developed a conceptual model depicting the process by which people booked travel on the Web. Toms, Freund, Kopak, and Bartlett (2003) assessed how people search within the travel domains to identify unique searching needs. Using 48 participants interacting with a modified version of Google, results indicated those searchers seeking travel information exhibited significant searching differences. Those doing travel tasks spent significantly more time looking at the content of a website relative to those in the health domain (Toms et al., 2003). This would indicate that searching for travel information is inherently difficult or that existing systems do not adequately support this mode of searching.

A few studies have examined a broader spec-

trum of travel searchers, typically using survey methodology. Weber and Roehl (1999) examined aspects of people using the Web for travel information. Using data from a November 2001 survey conducted by the Canadian Tourism Commission, Beldona, Morrison, and O'Leary (2005) investigated the purchase motives for complex and noncomplex travel in a Web environment. The researchers stated that purchases of activities, accommodation, events, and attractions demanded informational contexts. The researchers also reported that the purchases of car rentals and airline tickets were driven by transactional contexts. Grouping travel searchers in cohorts by age, Beldona (2005) reported that older baby boomer cohorts reported greater increases in the likelihood of travel information searching than younger Generation Xers, supporting earlier findings by Weber and Roehl (1999).

Johnson, Moe, Fader, Bellman, and Lohse (2004) examined searching across competing e-commerce sites using panel data from over 10,000 Internet households. The researchers presented results indicating that online travel search is actually quite limited, with households visiting only 1.8 travel sites during a typical active month. Johnson et al. (2004) reported similar findings that shoppers searched across very few sites in a given shopping month. Cai, Feng, and Breiter (2004) examined the relationships between tourists' purchase decision involvement and their information search behaviors, including perceived value of different types of information contents and their use of the Web as an information channel. The researchers reported that tourist information preferences and Web usage significantly differed based on level of purchase decision involvement.

There have been some reports of travel searching statistics on major Web search engines. Jansen and Spink (2005b) qualitatively analyzed approximately 2,500 queries from multiple Web search engine transaction log. The researchers assigned each query into one or more classifications, including a *commerce*, *travel*, *employment*, or *economy* grouping. The researchers report that this grouping, relative to the total set, represented 13.3% in 1997, 24.5% in 1999, and 24.7% in 2001 on the Excite search engine. Luo, Feng, and Cai (2005) examined the relationships between tourists use of a variety of information sources, including the Web. The researcher reported that both demographic characteristics (e.g., gender and household income) and situational factors (e.g., trip purpose and travel party type) are significant to the information source chosen.

Pan and Litvin (2006) analyzed keywords used in destination searches and reported that the use of keywords is significantly higher than for average searches. The researchers concluded that this might indicate that travel searching is a cognitively intensive activity. Pan and Litvin also reported that city is the most searched destination. Nearly half of all travel-related searches in this study involved a search based upon a city name. The research also reported that city often appeared in the search along with hotel, attraction, and activities. This study partially motivated the selection of key terms in this research.

Using a survey, Changfeng (2006) developed a model that incorporated consumer trust in the use of an online travel site. One would expect similar elements of trust in the use of Web search engines. Pan and Fesenmaier (2006) categorized travel information spaces according to users' personal view of term meanings based on a travel planning exercise. The researchers concluded that users have a variety of semantic mental models in regards to travel destination searching. Pan and Fesenmaier also reported that users tend to search for information hubs on the Internet. Because this study involved a small sample laboratory study, it would be interesting to see the behaviors exhibited by Web searchers.

Using a qualitative meta-analysis methodology, Y. A. Park and Gretzel (2007) proposed website measures to assist travel organizations in assessing website effectiveness, evaluating the return on investment, and making website improvements. The researchers proposed a framework of nine factors (i.e., information quality; ease of use; responsiveness; security/privacy; visual appearance; trust; interactivity; personalization; and fulfillment). Y. A. Park, Gretzel, and Sirakaya-Turk (2007) identified six core dimensions of website quality, reporting that ease of use was the most important of these dimensions.

Beritelli, Bieger, and Laesser (2007) examined the effect of the Web on the make-up of information sources. The researchers report that the traveler's Web use was related to age, education, and income level. However, the researchers also reported that the Web was a secondary resource in some travel situations. Jun, Vogt, and MacKay (2007) investigated the relationship between searching for travel information and product purchasing in both online and offline contexts. The researchers report that both travel information searching and product purchasing differ in the pretrip stage and vary by the category of travel product. The researchers also conclude that previous travel experiences affected how users executing approach searching for travel information.

From a synthesis of the available research on travel and tourism information searching, it is apparent that the travel domain has certain inherent characteristics that may influence how people search for travel-related information. There also appears to be a growing body of work on how people interact with travel websites and services. Lab studies and survey research have documented some aspects of online traveling searching. However, there does not appear to be any large-scale studies of how prevalent searching for travel information is on the Web. This lack of prior work leaves many open questions. How much of Web search is devoted to locating travel information? What terms do people use when searching for travel information on the Web? How do these people structure their queries? These unanswered questions motivate this research.

Our underlying premise for this research is that travel information searching is a specific instance of information searching (Wilson, 2000). Although there are various methods of slicing the human information process, information searching for this research is viewed in line with Wilson's perspective, which is:

> the "micro-level" of behavior employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information. (p. 49)

Within the field of information science, there are various models describing the information searching process (c.f., Belkin, Oddy, & Brooks, 1982; Kuhlthau, 1993; Wilson, 1999). Within the travel domain, distinct strategies of information seeking have been reported (Fodness & Murray, 1999). Gursoy (2001) and Gursoy and McCleary (2004) state that travel information search strategies are dependent on familiarity with websites and products, along with the level of expertise in search. Mitsche (2005) provides an overview of search patterns for travel information on a domain-specific search engine. Fodness and Murray's (1999) model of travel searching addresses information strategies involved in searching for travel information. The results reported in this study extend this model by focusing on the operational aspects. The assumption is that travel information searchers have two areas of uncertainty, (1) *knowledge uncertainty* and (2) *choice uncertainty*, which they are trying to reduce by searching for information. How searchers express this uncertainty when seeking travel information from Web search engines is specifically investigated in this research.

These two aspects of uncertainty have been highlighted in prior ecommerce work (Urbany, Dickson, & Wilkie, 1989). Few studies have explored uncertainty in searching for travel information. The results from this research speak to this shortcoming.

Research Questions

The research questions driving this study are:

1. **How prevalent is travel searching on Web search engines?** To address this research question, the percentage of travel-related queries based on pairs of keywords contained within queries were identified using the mutual information statistic, which is described in detail in the research design section. The mutual information statistic was used because it allowed aggregate key term pairs used across the entire

dataset, and it facilitated qualitative analysis that would not be possible at the individual query level due to the size of the dataset. Because the mutual information statistics does not account for queries of single terms, all single term queries are also aggregated and classified based on the single term within these queries. Using percentages of travel-related term pairs and single term queries, this percentage was compared to the complete data set to get an overall percentage for travel-related searching.

2. **How are people searching for travel-related information on Web search engines?** To address this research question, a major portion of the identified travel-related queries were classified to gain insight into the nature of their search topics using a mixed methods analysis of both quantitative (e.g., exploratory data analysis) and qualitative (i.e., open coding) approaches. The focus of this analysis was primarily at the term level. A comprehensive classification scheme for travel-related queries was developed.
3. **What is the usage of travel-related terms in Web searching?** For this research question, individual terms and term pair usage and occurrences for multiterm and single term queries was examined. Terms that have the highest degree of association (i.e., those terms are most likely to appear together) were also examined to provide greater insight into the underlying information need of travel searchers.

Research Design

Dogpile

For this study, a sample of queries submitted to the Web search engine Dogpile (http://www.dogpile.com/) was collected. A division of Infospace, Dogpile is a meta-search engine that incorporates the results from other search engines into aggregate search engine result pages (SERPs). The Dogpile SERP includes results from the four leading Web search indices (i.e., Ask Jeeves, Google, MSN, and Yahoo!), along with many others. Functionally, when a searcher submits a query, Dogpile simultaneously submits this query to multiple other Web search engines, collects the results from each Web search engine, removes duplicates results, and aggregates the remaining results into a combined ranked listing using a proprietary algorithm. Dogpile has tabbed indexes or verticals for federated searching of Web, Images, Audio, and Video content. Dogpile also incorporates query reformulation assistance. Query suggestions appear in the "Are You Looking for?" section of the interface (http://www.dogpile.com/).

Data Collection

For data collection, the searcher–system interactions were recorded in a transaction log that represented a portion of the searches executed on Dogpile, on May 6, 2005. Given the stability of Web searching (Jansen & Spink, 2005b), analysis of these data could provide significant insights into how people are actually conducting travel searching on the Web. The original general transaction log contained 4,056,374 records. Each record contained several fields including:

- **User Identification:** a user code automatically assigned by the Web server to identify a particular computer.
- **Cookie:** an anonymous cookie automatically assigned by the Dogpile server to identify unique users on a particular computer.
- **Time of Day:** measured in hours, minutes, and seconds as recorded by the Dogpile server.
- **Query Terms:** terms exactly as entered by the given user.

Most of the major search engines provide information on popular searches. Such services include AOL Hot Searches (http://hot.aol.com/), Dogpile SearchSpy (http://www.dogpile.com/info.dogpl/searchspy/), Google Trends (http://www.google.com/trends/hottrends), and Yahoo! Buzz Index (http://buzz.yahoo.com/). However, access to the data in these services is limited.

Data Preparation

The original flat ASCII transaction log file of 4,056,374 records was imported into a relational database. A unique identifier for each record was generated. Four fields (*Time of Day, User Identification, Cookie,* and *Query*) were used to locate the initial query and then recreate the chronological

series of actions in a session. A sample of the transaction log is shown in Table 1.

The terminology similar to that used in other Web transaction log studies (Jansen & Pooch, 2001; S. Park, Bae, & Lee, 2005).

- **Term:** a series of characters separated by white space or other separator.
  1. **Unique term: a term submitted one or more times in the data set.**
  2. **Term pair:** two terms that occur within the same query.
- **Query:** string of terms submitted by a searcher in a given instance.
  1. **Initial query:** first query submitted in a session by a given user.
  2. **Identical query:** a query within a session that is a copy of a previous query within that session.
  3. **Repeat query:** a query submitted more than once during the data collection period, irrespective of the user.
  4. **Query length:** the number of terms in the query, including traditional stop words.

Removing Agent Queries

For this research, the interest was only in queries submitted by humans and the transaction log contained queries from human users, agents, and common access terminals. There is no methodology for specifically identifying human from nonhuman searchers in a search log. Therefore, researchers interested in human sessions usually use a temporal or interaction cut-off (Montgomery & Faloutsos, 2001; Silverstein, Henzinger, Marais, & Moricz, 1999).

An interaction cut-off was used by separating all sessions with 100 or fewer queries into an individual transaction log to be consistent with the approach taken in previous Web searching studies (Jansen & Spink, 2005b; Jansen, Spink, & Pedersen, 2005; Spink & Jansen, 2004). This cut-off is substantially greater than the mean search session (Jansen, Spink, & Saracevic, 2000) for human Web searchers. This approach increases the probability that we were not excluding any human searches. However, this cut-off probably introduced some agent or common user terminal sessions. Overall, it seemed reasonable that most of the queries submitted primarily by human searchers were included.

Removing Duplicate Queries

Search log applications typically record result pages viewed as separate records with an identical user identification and query, but with a new time stamp (i.e., the time of the second visit). This permits the calculation of results page viewings. Unfortunately, it also introduces duplicate records that skew other calculations. To correct for these duplicate queries, the transaction log was collapsed upon user identification, cookie, and query. The number of identical queries was calculated using the fields of user identification and cookie, storing this number in a separate field within the transaction log. This collapsed transaction log provided the records by user for analyzing sessions, queries, and terms.

Term and Term Co-occurrence Analysis

A field for the length of the query was also incorporated, measured in number of terms. From the collapsed data set, a table for term data and a table for co-occurrence data were generated. The term table contains fields for a term, the number

Table 1
Sample of the Transaction Log Data

User Identification	Cookie	Time of Day	Query Terms	Location	Source	Feedback
xx.0.101.132	26JMB84A498WVZM	05:03:56	cng powered cars	usa	Web	0
xx.0.101.132	26JMB84A498WVZM	05:08:45	fuelmaker	usa	Web	1
xx.0.101.19	NotDef	12:47:01	messieurs	usa	Image	0
ww.0.170.2	82MU6A4BW1BPY	13:07:00	“mickey rourke”	usa	Web	0
ww.0.170.2	82MU6A4BW1BPY	13:05:48	“presa canario dogs”	usa	Web	0
ww.0.170.2	82MU6A4BW1BPY	11:49:28	osteopenia	usa	Video	0

of times that term occurs in the complete data set, and the probability of occurrence. The co-occurrence table contains fields for term pairs and the number of times that pairs occur within the data set irrespective of order.

Data Analysis: Query Topic Analysis

A broad approach in what was travel related was taken, including other tourist information needs such as concerts, dining out, and movies. A set of seed terms obtained from a literature review of travel, tourism research (e.g., Pan & Litvin, 2006), and a review of the transaction log was first developed (i.e., travel, trip, travel, hotel, motel, cruise, mapquest, etc.). Using this set of seed terms and a modified snowball technique (Patton, 1990), the search log was continually accessed to isolate travel-related queries. After each iteration, selected queries were manually reviewed to obtain new terms. These new terms were added to the selection query and then the search log was requeried. This process continued until the new iteration added less than five new queries. We believe at this point that we had an adequate representation of the data log.

With hundreds of thousands of queries still in the search log, it was impossible to classify manually all travel-related queries in a timely manner. Therefore, the mutual information statistic reduced the number of queries needing individual classification. The mutual information statistic measures term association and does not assume mutual independence of the terms within the pair. The mutual information statistic was calculated for all term pairs within the data set. Many times, a relatively low frequency term pair may be strongly associated (i.e., if the two terms always occur together). The mutual information statistic identifies the strength of this association. The mutual information formula used in this research is:

$$I(w_1, w_2) = \ln \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

where $P(w_1)$, $P(w_2)$ are probabilities estimated by relative frequencies of the two words and $P(w_1, w_2)$ is the relative frequency of the word pair and order is not considered. Relative frequencies are observed frequencies (F) normalized by the number of the queries:

$$P(w_1) = \frac{F_1}{Q'}; P(w_1) = \frac{F_2}{Q'}; P(w_1, w_2) = \frac{F_{12}}{Q'}$$

Both the frequency of term occurrence and the frequency of term pairs are the occurrence of the term or term pair within the set of queries. Because a one-term query cannot have a term pair, the set of queries for the frequency base differs. The number of queries for the terms is the number of nonduplicate queries in the data set. The number of queries for term pairs is defined as:

$$Q' = \sum_{n}^{m} (2n - 3) Q_n$$

where Q_n is the number of queries with n words ($n > 1$), and m is the maximum query length.

Therefore, queries of length one have no pairs. Queries of length two have one pair. Queries of length three have three possible pairs. Queries of length four have six possible pairs. This continues up to the queries of maximum length in the data set. The formula for queries of term pairs (Q') account for this term pairing. For each term pair, there was now the mutual information statistic (i.e., a measure of the association between terms within the pair) and the frequency of occurrence for each term pair within the dataset. Because a large-scale qualitative analysis was needed, the dataset had to be reduced to a manageable number. Therefore, all term pairs with a negative mutual information statistic (i.e., the terms were negatively associated) and with a frequency of less than five were removed. Term pairs that were positively associated and had a reasonable occurrence remained. Specifically, this was 65,094 term pairs with a total frequency of occurrence in the dataset of 1,306,609.

However, the mutual information statistic does not include single term queries, which prior research shows is a common characteristic of Web queries (Jansen et al., 2000). Therefore, all one-term queries within the dataset were isolated, resulting in 123,026 unique one-term queries with 281,283 occurrences within the dataset. Each of

these term pairs were manually classified using a categorization scheme that was developed for this research. Each term pair or single term query could have up to three categories. Table 2 shows the categories developed post priori.

Each term pair or query was classified at as fine a granularity as possible. For example, if the query referred to a hotel (e.g., *holiday inn, comfort suite*), it was classified as such; otherwise, it was classified as accommodations (e.g., *rental house, beach house*).

Results

Overall Results

Table 3 presents the results of an analysis for the overall dataset in order to compare findings with prior work on Web searching. There were 2,465,145 interactions of both queries and SERP views during the data collection period. Of these interactions, there were 1,523,793 queries submitted by 534,507 users (identified by unique IP address and cookie) containing 4,250,656 total terms. There were 298,796 unique terms in the 1,523,793 queries. Most of the users (84%) came from the US. The mean query length was 2.79 terms and nearly 50% of queries contained three or more terms. The mean session length was 2.85 queries per user. More than 46% of users modified their queries, and 29.4% of the sessions contained three or more queries. Nearly 10% of the queries in the data set were repeat queries submitted by 10.8% of the searchers. The 898,393 unique queries represent 58.96% of the 1,523,793 total queries. The

Table 2
Travel Searching Codes Used in Classifying Term Pairs and Single Term Queries

	Codes
Accommodations	
Sporting events	Baseball event—relating to a baseball game at any level
	Basketball event—relating to a basketball game at any level
	Football event—relating to a football game at any level
	Hockey event—relating to a hockey game at any level
	Soccer event—relating to a soccer game at any level
Transportation	Airplane—relating to transportation by airplane
	Car—relating to transportation by a car or truck
	Motorcycle— relating to transportation by motorcycle
	Ship— relating to transportation by ship
	Train—relating to transportation by train
Geographical	City—relating to a geographical interested focused on a city
	Country—relating to a geographical interested focused on a country
	Location information—relating to a geographical interested focused on a general area
	Religious Location—relating to a geographical interested focused on a religious site
	Specific event—relating to a geographical interested focused on a specific event held at that location
	Specific location—relating to a geographical interested focused on a particular location that is not a city, country, general area, religious location, centered on a specific event, a specific business, or a state
	Specific store—relating to a geographical interested focused on a particular business
	State—relating to a geographical interested focused on a state
Packages	Cruise—relating to a cruise
	Land Tour—relating to a tour on land
Cultural or general entertainment	Concert—relating to a music concert
	Movie—relating to a movie
	Museum—relating to museum
	Musical—relating to a musical
	Play—relating to a play
General travel	Business—relating to travel for business
	Date—relating to travel on a particular date
	Escort Service—relating to scheduling an escort service
	Price—relating to the price of travel
	Restaurant—relating to a dining establishment
	Shopping—relating to shopping
	Travel equipment—relating to equipment for travel
	Travel information—relating to general travel information, such as travel sites, travel agencies, and governmental notices

Table 3
Aggregate Statistics From the Dogpile Transaction Log

Sessions	534,507	
Queries	1,523,793	
Terms		
Unique	298,796	7.03%
Total	4,250,656	
Location (USA)	1,282,691	84.1%
Mean terms per query	2.79 SD = 1.54	
Terms per query		
1 term	281,639	18.5%
2 terms	491,002	32.2%
3+ terms	751,152	49.2%
Mean queries per user	2.85 SD = 4.43	
Users modifying queries	246,276	46.08%
Repeat queries (queries submitted more than once by two or more searchers: 57,651 searchers)	151,413	9.9%
Unique queries (queries submitted only once in the entire data set)	898,393	58.9%
Queries generated via feedback	128,126	8.4%
Session size		
1 query	288,231	53.9%
2 queries	88,875	16.6%
3+ queries	157,401	29.4%
Results pages viewed per query		
1 page	1,052,554	69.07%
2 pages	253,718	16.6%
3+ pages	217,521	14.2%
Mean results pages viewed per query	1.67 SD = 1.84	
Boolean queries	33,403	2.1%
Other query syntax	116,905	7.6%
Terms not repeated in data set (57.7% of the unique terms)	172,488	4.06%
Use of 100 most frequently occurring terms (0.03% of the unique terms)	752,994	17.7%
Use of other 126,208 terms (42.24% of the unique terms)	3,325,174	78.2%
Unique term pairs (occurrences of terms pairs within queries from the entire data set)	2,209,777	

remaining 473,987 were queries to multiple data sources. In 1,052,554 queries (69.07%), the searchers viewed only the first results page. There were a very small percentage of Boolean queries (2.19%) or queries containing advanced query syntax (7.6%). Of these advanced query syntax queries, most were for phrase searching. Of the total terms, there were 4.06% used only once in the data set, representing 57.7% of the unique terms. The top 100 most frequently used terms accounted for 17.71% of the total terms. There were 2,209,777 term pairs.

In general, these results are comparable to that reported from prior work on Web searching, including research from the Excite search engine (Jansen et al., 2000; Spink & Jansen, 2004; Spink, Jansen, Wolfram, & Saracevic, 2002), the Alta Vista search engine (Jansen et al., 2005; Silverstein et al., 1999; Spink & Jansen, 2004), and AlltheWeb (Jansen & Spink, 2005a). Generally, queries are short as measured by the number of terms, and sessions are short as measured by the number of queries. The usage of terms follows a power law distribution with a small number of terms used quite frequently and a large number of terms used very infrequently. Given the similarity of this dataset to others reported in prior work, there is confidence that this dataset is similar in characteristics to the larger Web population. Therefore, the results are probably similar to those that one would obtain from data of other Web search engines.

Research Question 01: How Prevalent Is Travel Searching on Web Search Engines? Based on the procedure outlined, about 6.5% of all Web searching is travel related, as shown in Table 4. Of the 1,306,609 term pair occurrences within the dataset, just more than 6% of these were travel related. Of the single term queries, 6,655 (just under 0.5%) of the 1,523,793 single term queries within the dataset were travel related. Term pairs

Table 4
Percentage of Travel-Related Searching

Classification	Occurrence of Classification	Total of Occurrences in Dataset	Percentage
Total travel-related term pairs identified	3,310	79,490	6.1%[a]
Total travel-related single term queries identified	929	6,655	0.4%[b]
			6.5%

[a]Percentage of 1,306,609 term pairs within the dataset.
[b]Percentage of 1,523,793 queries within the dataset.

do not directly translate into queries; however, one can show that it is a good approximation. There were 491,002 two-term queries and 751,152 queries of three or more terms, for a total of 1,242,154 queries. The remaining 281,639 queries were all single term queries, given the 1,523,793 total queries within the dataset. Therefore, it is reasonable to assume that had we analyzed the complete dataset at the query level, the percentage would be approximately 6–6.5% of the total dataset. This is an upper bound, as some of the terms and term pairs selected might not be travel related.

Research Question 02: How Are People Searching for Travel-Related Information on Web Search Engines? All term pair co-occurrence with an appearance rate of greater than five and a positive mutual information statistic score were qualitatively analyzed. Each occurrence was categorized according to the code legend shown in Table 2. Frequencies of occurrences of each code and the occurrence of the coded term pair in the database were then aggregated. Table 5 displays the topical term pair analysis results. In Table 5, column one is the category. Columns two and three are the number of term pairs that were classified in this category and the corresponding percentage, respectively. Columns four and five are the number of occurrences of that category in the dataset and corresponding percentage, respectively. Table 5 shows that the most frequently occurring travel-related term pairs are queries about *cities* (27% of all travel-related pairs). In fact, all three of the top categories are *locations*, with *state* (18%) and *specific location* (9.5%) being the second and third most popular categories, respectively. The queries *concerts* and *travel information* (nearly 8% each) and *hotels* (just under 7%) follow the top three categories. *Country* and *movie* searching are next, with just more than 4% each. These eight categories accounted for more than 85% of all the travel searching term pairs. The remaining 29 categories account for less than 15% of travel-related query term pairs. These findings point to the need for increased research concerning the use of online mapping applications such as Google Maps and MSN Virtual Earth in supporting online travel searching. There has been limited work in investigating the effects of these geo-spatial applications (Santanche, Nath, Liu, Priyantha, & Zhao, 2006).

Term pairs do not take into account queries that are a single term. It has been well noted in prior literature that a large percentage of Web queries are single term (Jansen et al., 2000; Silverstein et al., 1999; Wang, Berry, & Yang, 2003). To address this missing segment of the data set, all single terms queries were aggregated and then classified using the codes in Table 2. Table 6 presents the results of this analysis of single term queries.

Table 6 shows that there are some changes in ordering of the frequency of classification occurrences. This would be expected, as these are only single terms. General *travel information* is by far the most common query (nearly 32%), with *cars* (10%) and *country* (8.5%) at second and third most frequently occurring queries. *Shopping* and *states* queries are also popular, at about 7% and 6%, respectively. Therefore, generally, there are no major surprises at the single term queries, although there is a tightening of occurrences and clustering around certain categories.

Research Question 03: What Is the Usage of Travel-Related Terms in Web Searching? To ad-

Table 5
Term Pair Category

Category	Occurrence of Category	Percentage	Total of Occurrences in Dataset	Percentage
1. City	1,029	31.1%	21,450	27.0%
2. State	391	11.8%	14,571	18.3%
3. Specific location	319	9.6%	7,569	9.5%
4. Concert	68	2.1%	6,089	7.7%
5. Travel information	308	9.3%	6,076	7.6%
6. Hotel	290	8.8%	5,416	6.8%
7. Country	93	2.8%	3,440	4.3%
8. Movie	110	3.3%	3,369	4.2%
9. Location information	89	2.7%	1,340	1.7%
10. Religious shrine	54	1.6%	1,293	1.6%
11. Airplane	64	1.9%	1,217	1.5%
12. Travel equipment	76	2.3%	1,013	1.3%
13. Restaurant	39	1.2%	855	1.1%
14. Price	38	1.1%	851	1.1%
15. Accommodations	42	1.3%	589	0.7%
16. Football event	39	1.2%	528	0.7%
17. Baseball event	23	0.7%	467	0.6%
18. Cruise	31	0.9%	437	0.5%
19. Shopping	25	0.8%	417	0.5%
20. Train	24	0.7%	311	0.4%
21. Basketball event	19	0.6%	297	0.4%
22. Museum	15	0.5%	287	0.4%
23. Specific store	16	0.5%	258	0.3%
24. Business travel	16	0.5%	232	0.3%
25. Escort service	22	0.7%	188	0.2%
26. Ship	17	0.5%	151	0.2%
27. Play	12	0.4%	161	0.2%
28. Soccer event	10	0.3%	121	0.2%
29. Car	7	0.2%	135	0.2%
30. Specific event	6	0.2%	124	0.2%
31. Hockey event	10	0.3%	97	0.1%
32. Date	1	0.0%	86	0.1%
33. Land tour	4	0.1%	26	0.0%
34. Musical	2	0.1%	23	0.0%
35. County	1	0.0%	6	0.0%
36. Total	3,310	100%	79,490	100%
Overall			79,490	6.1%
Complete dataset			1,306,609	

dress this research question, the use of term pairs and the terms with the strongest association were examined.

Terms

Table 7 shows the results of the term analysis. Table 7 shows the most frequently occurring terms and corresponding percentages for both term pairs and single term queries. The percentages are based on the total number of terms (i.e., 4,250,656). The term pair and single term queries are reported separately because the usage was different. We note several findings. First, even the most frequently occurring terms represent a small percentage of overall term usage. The most frequently used content term (*music*) accounted only for approximately 0.213% of all term usage. Second, natural groupings appear. For the term pairs, these include location (*county, city, state, texas, florida, beach, california*), adjectives (*new, black, young, 2005, best*), and general entertainment (*music, art, park*). For the single term queries, most of the terms relate to general travel information, most notably

Table 6
Single Term Query Category

Category	Occurrence of Category	Percentage	Total of Occurrences in Dataset	Percentage
1. Travel information	164	17.65%	2,118	31.8%
2. Car	80	8.61%	674	10.1%
3. Country	77	8.29%	567	8.5%
4. Shopping	45	4.84%	461	6.9%
5. State	66	7.10%	412	6.2%
6. Specific location	71	7.64%	333	5.0%
7. Movie	15	1.61%	294	4.4%
8. City	131	14.10%	284	4.3%
9. Price	52	5.60%	288	4.3%
10. Concert	12	1.29%	286	4.3%
11. Hotel	61	6.57%	210	3.2%
12. Airplane	35	3.77%	195	2.9%
13. Religious shrine	0	0.00%	140	2.1%
14. Business travel	11	1.18%	98	1.5%
15. Football event	6	0.65%	78	1.2%
16. Accommodations	30	3.23%	76	1.1%
17. Cruise	21	2.26%	62	0.9%
18. Travel equipment	10	1.08%	53	0.8%
19. Restaurant	17	1.83%	43	0.6%
20. Train	8	0.86%	35	0.5%
21. Museum	5	0.54%	9	0.1%
22. Baseball event	4	0.43%	4	0.1%
23. Basketball event	1	0.11%	4	0.1%
24. Escort service	2	0.22%	2	0.0%
25. Location information	2	0.22%	2	0.0%
26. Specific event	2	0.22%	2	0.0%
27. Ship	1	0.11%	1	0.0%
Overall	929	100%	6,655	100%
Complete dataset			1,523,793	0.04%

website travel hubs (*mapquest, maps, weather, travelocity, mapquest.com, travel*, and *orbitz*).

Term Co-Occurrence

Although individual terms provide insight in user intent, a term co-occurrence (Leydesdorff, 1989) is more helpful in determining the specific usage of a term intended by a searcher. Table 8 shows the most frequently occurring term pairs. Table 8 shows that the top 24 term pairs are all pairs that one would see in natural language queries. There are several term pairs of *entertainers* (8) and *location* (11). In examining the mutual information for these term pairs, we see that most had highly positive scores. However, there were a few exceptions (e.g., *hotels in, county florida, washington state*, and *york state*) with term pairs containing a common term. These few exceptions aside, the combination of the *entertainer* and *location* categories and the high mutual information statistic scores show that travel seekers on the Web do a lot of searching using proper names. In calculating the mutual information statistics for the data set, there was range of more than eight between the maximum degree of association (max = 9.46) and the minimum degree of association (min = 1.00). The mean association of term pairs in the data set was 3.67 with a standard deviation of 1.81. Therefore, term pairs with associations of approximately 3.5 or more would indicate a high degree of association for this data set. In Table 8, we see that 19 of the 24 most frequently occurring term pairs had high positive degrees of association.

Table 9 shows the term pairs with highest degrees of positive association. Table 9 shows again a high occurrence of proper names and phrases.

Table 7
Most Frequently Utilized Terms

Term Pairs			Single Term Queries		
Term	Frequency	%	Term	Frequency	%
1. music	9,067	0.213%	mapquest	619	0.0146%
2. new	9,056	0.213%	maps	535	0.0126%
3. county	6,455	0.152%	movies	265	0.0062%
4. black	6,267	0.147%	weather	233	0.0055%
5. city	5,126	0.121%	beyonce	222	0.0052%
6. state	4,891	0.115%	beach	115	0.0027%
7. map	4,806	0.113%	travelocity	101	0.0024%
8. young	4,673	0.110%	mapquest.com	99	0.0023%
9. world	4,399	0.103%	travel	92	0.0022%
10. texas	4,238	0.100%	orbitz	72	0.0017%
11. florida	4,084	0.096%			
12. movie	4,052	0.095%			
13. art	4,048	0.095%			
14. boys	3,830	0.090%			
15. 2005	3,789	0.089%			
16. park	3,758	0.088%			
17. beach	3,375	0.079%			
18. people	3,186	0.075%			
19. california	3,161	0.074%			
20. best	3,133	0.074%			

Table 8
Most Frequently Occurring Term Pairs

	Mutual Information Statistic	Frequency of Occurrence	%	Term	Term	Category
1	4.05	2,770	3.48%	new	york	State
2	5.68	1,131	1.42%	britney	spears	Concert
3	5.56	919	1.16%	kentucky	derby	Specific event
4	3.93	898	1.13%	new	jersey	State
5	5.25	875	1.10%	san	diego	City
6	5.53	864	1.09%	united	states	Country
7	6.46	561	0.71%	hilary	duff	Concert
8	4.03	520	0.65%	south	carolina	State
9	6.50	465	0.58%	gwen	stefani	Concert
10	1.01	362	0.46%	hotels	in	Hotel
11	6.57	323	0.41%	christina	aguilera	Concert
12	4.95	311	0.39%	holiday	inn	Hotel
13	2.83	311	0.39%	washington	state	State
14	2.30	258	0.32%	york	state	State
15	5.91	251	0.32%	pink	floyd	Concert
16	4.66	244	0.31%	washington	dc	City
17	4.30	235	0.30%	public	library	Specific place
18	6.84	230	0.29%	led	zeppelin	Concert
19	7.17	227	0.29%	snoop	dogg	Concert
20	3.44	226	0.28%	west	virginia	State
21	7.06	216	0.27%	puerto	rico	Country
22	1.50	214	0.27%	county	florida	State
23	4.88	212	0.27%	backstreet	boys	Concert
24	5.85	204	0.26%	driving	directions	Travel information

Table 9
Term Pairs With Highest Positive Mutual Information Statistic

Mutual Information Statistic	Number of Occurrence	Term Pair		Category
9.46	7	sao	paolo	City
9.44	7	champs	elysee	City
9.44	12	shinto	shrines	Religious shrine
9.42	13	tgi	fridays	Restaurant
9.35	12	originated	czechoslovakia	Country
9.31	20	dunkin	donuts	Restaurant
8.92	38	saudi	arabia	Country
8.89	29	airbus	a380	Airplane
8.74	14	trains	diesels	Train
8.80	18	abu	dhabi	Country
8.67	6	blazing	saddles	Movie
8.64	7	boeing	767	Airplane
8.53	41	willy	wonka	Movie
8.50	15	landlord	tenant	Accommodations
8.45	49	burma	myanmar	Country
8.45	13	papua	guinea	Country
8.42	43	burma	pepsico	Business travel
8.41	73	mardi	gras	Specific event
8.41	16	neiman	marcus	Shopping
8.26	8	mardi	gra	Specific event
8.15	7	dj's	philly	Concert
8.13	10	godfather	trilogy	Movie
8.13	16	forrest	gump	Movie
7.99	6	metropolitan	arts	Museum
7.97	20	kodak	easyshare	Travel equipment
7.92	9	smyrna	airshow	Specific event
7.91	7	des	plaines	City
7.87	7	pacers	pistons	Basketball event
7.86	6	kodak	Dx7590	Travel equipment
7.82	21	chartres	cathedral	Religious location
7.78	155	hong	kong	Country
7.77	18	ho	minh	City

Silverstein et al. (1999) reported the co-occurrence of the top 10,000 terms from approximately 313,000,000 million queries. Silverstein et al. also reported highly correlated phrases. In a study of Excite users, Wolfram (1999) notes high clustering of several term pairs around entertainment that we do not necessarily see in this analysis. In Table 9, we see a general spread among all categories, although with some clustering around location (i.e., *country* and *city*).

Discussion

Based on the analysis from this research, approximately 6% of Web search is composed of travel searching. As a comparison, a random sample of 2,500 queries from the Dogpile transaction log was qualitatively analyzed (Jansen, Spink, Blakely, & Koshman, 2006), assigning these queries into one of 11 general topic categories developed by Spink et al. (2002). Within the 11 categories, five had a higher percentage. Therefore, travel searching at 6% of Web queries represents a significant amount of the total Web searching. Frequency data using term pairs showed an interesting distribution of travel information interests. The most frequently used terms are geographical locations, such as city, state, specific location, and country. These geographic categories represent nearly 60% of travel searching. Therefore, travel searching appears to be very location focused. Advertising campaigns and website development aimed at Web travel searchers should highlight key location aspects over other travel attributes. In

this analysis, the 31 nongeographic travel categories accounted for only approximately 40% of our term pairs.

Interestingly, the analyzed geographic clustering apparent in the term pairs analysis did not hold for single term queries. For these queries, general travel information accounted for more than 31% of traveling searching. Geographic travel queries accounted for only 24% of single-term travel queries. Although still a notable percentage, it is much lower than the more than 60% of travel term pairs. Much of the single-term queries were for specific travel websites, such as *mapquest* and *travelocity*. For travel content providers, these travel-specific sites should certainly be part of an overall marketing plan. At the term level of analysis, even the most popularly used terms occurred very infrequently, with the most popular term at a mere 0.2% of all term usage. This shows that a diverse term usage among travel information searchers with query topics clustered around geographic and travel information. There was a significant usage of location and proper names, resulting in both a diverse term collection and high positive association of a sizeable number of terms. This would point to the need for a wide set of meta-tags on travel websites and selection of keywords for sponsored search (Jansen, 2006).

There are three overall implications resulting from this research. First, Web travel information searching represents a significant percentage of Web searching. This is somewhat surprising in that the dataset came from a general purpose search engine and not a travel-specific search site. In other areas (e.g., pornography, medical), niche search engines resulted in lower percentages of searching in the particular domain (Spink et al., 2002; Wolfe, Jansen, & Spink, 2006). This would imply that a significant portion of people seeking travel-related information use non-travel-specific sites at some point in their search. This makes the general purpose search engines rich targets for those advertising travel-related information.

Second, the distribution of travel-related topics is extremely diverse and the manner in which travel searchers express themselves in attempting to reduce their uncertainty. In fact, the diversity of topics and modes of expression may indicate a great deal of uncertainty in the travel information searching process. Uncertainty in traveling information searching is supported by prior work in lab experiments (cf., Toms et al., 2003) and other ecommerce research (Urbany et al., 1989). This finding points to the need for more improvements in the interfaces and systems providing travel-related information.

Finally, the distribution of topics is skewed toward the edges, with some topics being "very focused" and others being "very general." On the "very focused" edge, there are queries for specific *hotels, website, car rentals*, and *locations*. On the "very general" side, there are queries for travel information and countries. This dichotomy might imply that there is a lack of technological linkage between the two extremes, and searchers are having difficulty expressing themselves or are uncertain how to express these information needs. Possible solutions might be cross-platform linkages among travel systems. For example, if a customer books a ticket on an airline travel site, targeted links might appear for travel books on that city from online book sellers and for restaurant review sites in that area.

As with any research, there are limitations to the findings. This study is restricted to data from one commercial Web meta-search engine and possibly does not represent the queries submitted by the broader Web searching population. Additionally, we collected the data on only 1 day; therefore, the data may not be representative of overall usage through out the year. Jansen and Spink (2005b) have shown that Web searching is fairly consistent across days and search engines. However, research also sows that temporal and contextual situations can affect term usage (Jansen et al., 2005; Wang et al., 2003). Therefore, the findings concerning tactics are probably applicable to general Web searching. However, travel searching is certainly seasonal, so future research is needed to determine if the topically classification holds at various periods of the year. Finally, a shortcoming of the transaction log analysis is that one cannot discern the underlying intent of the searcher. However, this research compliments the numerous laboratory studies that have been conducted in the travel research domain (e.g., Pan & Turner, 2006; Y. A. Park & Gretzel, 2007; Toms et al., 2003).

The strengths of the study are that a large set

of search engine queries to identify travel searching on the Web was quantitatively and qualitatively analyzed. This is one of the first studies of this type to focus on travel searching using real data from a Web search engine that services millions of queries per day. Therefore, these queries represent real searches by real users trying to address real travel information needs. A classification scheme for travel information searching was also developed. Thus, these research findings may make a significant contribution to the field of travel information searching.

Conclusion and Further Research

The research results provide important insights into the current state of Web travel searching and Web usage for developers of search engines, website designers, and ecommerce sites aimed at travel customers and tourists. This study represents the first major research of human interaction with a major commercial meta-search Web search engine within the travel domain. The results from this research have the potential to influence positively the future design of search engines and websites related to travel information and services. There are several avenues for future research. Certainly, there is the need for more analysis of search data over a longer period and among a wider range of search engines. Researchers should continue to examine and track Web search trends and characteristics within the travel domain. Transaction log analysis or lab studies will help assess future behavior and identify future user needs.

Acknowledgment

We thank Infospace, Inc for providing the Web search engine data set without which we could not have conducted this research. We encourage other commercial Web search engine companies to find way to collaborate with researchers in the academic community.

Biographical Notes

Dr. Bernard J. (Jim) Jansen is an assistant professor in the College of Information Sciences and Technology at The Pennsylvania State University, USA. His specific areas of expertise are Web searching, sponsored search, and personalization for information searching. He is coauthor of the book, *Web Search: Public Searching of the Web* and coeditor of the book *Handbook of Weblog Analysis*. Jim is a member of the editorial boards of six international journals.

Christopher C. Ciamacca is a student in the College of Information Sciences and Technology at The Pennsylvania State University, USA.

Dr. Amanda Spink is Professor in the Faculty of Information Technology at the Queensland University of Technology and Co-Leader of the Information Science Cluster. Her primary research and education interests include information science studies, including evolutionary and developmental theories, models and experiments related to information behavior, cognitive information retrieval; Web retrieval, including relevance, feedback, and multitasking models.

References

Beldona, S. (2005). Cohort analysis of online travel information search behavior: 1995–2000. *Journal of Travel Research, 44*(2), 135–142.

Beldona, S., Morrison, A. M., & O'Leary, J. (2005). Online shopping motivations and pleasure travel products: a correspondence analysis. *Tourism Management, 26*, 561–570.

Belkin, N., Oddy, R., & Brooks, H. (1982). ASK for information retrieval, parts 1 & 2. *Journal of Documentation, 38*(2), 61–71, 145–164.

Beritelli, P., Bieger, T., & Laesser, C. (2007). The impact of the Internet on information sources portfolios: Insight from a mature market. *Journal of Travel & Tourism Marketing, 22*(1), 63–80.

Bonn, M. A., Furr, H. L., & Susskind, A. M. (1999). Predicting a behavioral profile for pleasure travelers on the basis of Internet use segmentation. *Journal of Travel Research, 37*(4), 333–340.

Cai, L. A., Feng, R., & Breiter, D. (2004). Tourist purchase decision involvement and information preferences. *Journal of Vacation Marketing, 10*(2), 138–148.

Changfeng, C. (2006). Identifying significant factors influencing consumer trust in an online travel site. *Information Technology & Tourism, 8*(3–4), 197–214.

Cole, J. I., Suman, M., Schramm, P., Lunn, R., & Aquino, J. S. (2003). *The UCLA Internet report surveying the digital future year three*. Retrieved February 1, 2003, from http://www.digitalcenter.org/pdf/InternetReportYearThree.pdf

Fodness, D., & Murray, B. (1999). A model of tourist information search behaviour. *Journal of Travel Research, 37*(2), 220–230.

Gursoy, D. (2001). *Development of a traveler's information search behavior model*. Unpublished doctoral dissertation, Virginia Polytechnic Institute.

Gursoy, D., & McCleary, K. (2004). An integrative model

of tourists' information search behaviour. *Annals of Tourism Research, 31*(2), 353–373.

Jansen, B. J. (2006). Paid search. *IEEE Computer, 39*(7), 88–90.

Jansen, B. J., & Pooch, U. (2001). Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology, 52*(3), 235–246.

Jansen, B. J., & Spink, A. (2005a). An analysis of Web searching by European Alltheweb.com users. *Information Processing & Management, 41*(2), 361–381.

Jansen, B. J., & Spink, A. (2005b). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management, 42*(1), 248–263.

Jansen, B. J., Spink, A., Blakely, C., & Koshman, S. (2006). Web searcher interactions with the Dogpile.com meta-search engine. *Journal of the American Society for Information Science and Technology, 58*(4), 1875–1887.

Jansen, B. J., Spink, A., & Pedersen, J. (2005). Trend analysis of AltaVista Web searching. *Journal of the American Society for Information Science and Technology, 56*(6), 559–570.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing & Management, 36*(2), 207–227.

Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science, 50*(3), 299–308.

Jun, S. H., Vogt, C. A., & MacKay, K. J. (2007). Relationships between travel information search and travel product purchase in pretrip contexts. *Journal of Travel Research, 45*(3), 266–274.

Kuhlthau, C. (1993). *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex Publishing.

Lexhagen, M. (2004). The importance of value-added services to support the customer search and purchase process on travel websites. *Information Technology & Tourism, 7*(2), 119–136.

Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research Policy, 18*, 209–223.

Luo, M., Feng, R., & Cai, L. A. (2005). Information search behavior and tourist characteristics: The Internet vis-a-vis other information sources. *Journal of Travel & Tourism Marketing, 17*(2/3), 15–25.

Madden, M., & Rainie, L. (2003, December 23). *America's online pursuits: The changing picture of who's online and what they do*. Retrieved May 25, 2004, from http://www.pewinternet.org/reports/toc.asp?Report=106

Mitsche, N. (2005). *Understanding the information search process within a tourism domain-specific search engine*. Paper presented at the International Conference Information Technologies in Tourism, Innsbruck.

Montgomery, A., & Faloutsos, C. (2001). Identifying web browsing trends and patterns. *IEEE Computer, 34*(7), 94–95.

Morrisonn, A. M., Jing, S., O'Leary, J. T., & Cai, L. A. (2001). Predicting usage of the Internet for travel bookings: An exploratory study. *Information Technology & Tourism, 4*(1), 15–30.

Nielsen Media. (2006). *Search engines most popular method of surfing the Web*. Retrieved 30 January, 2006, from http://www.commerce.net/news/press/0416.html

Pan, B., & Fesenmaier, D. R. (2006). Online information search and trip planning process. *Annals of Tourism Research, 33*(3), 809–832.

Pan, B., & Litvin, S. (2006). *Real users, real trips, and real queries: An analysis of destination search on a search engine*. Paper presented at the Annual Conference of Travel and Tourism Research Association (TTRA 2006), June 16–18, Dublin, Ireland.

Pan, B., & Turner, G. B. (2006). *Tourist information search and acquisition: An extended framework*. Paper presented at the Annual Conference of Atlantic Marketing Association.

Park, S., Bae, H., & Lee, J. (2005). End user searching: A Web log analysis of NAVER, a Korean Web search engine. *Library & Information Science Research, 27*(2), 203–221.

Park, Y. A., & Gretzel, U. (2007). Success factors for destination marketing Web sites: A qualitative meta-analysis. *Journal of Travel Research, 46*(1), 46–63.

Park, Y. A., Gretzel, U., & Sirakaya-Turk, E. (2007). Measuring Web site quality for online travel agencies. *Journal of Travel & Tourism Marketing, 23*(1), 15–30.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications, Inc.

Santanche, A., Nath, S., Liu, J., Priyantha, B., & Zhao, F. (2006). *SenseWeb: Browsing the physical world in real time*. Paper presented at the 5th International Symposium on Information Processing in Sensor Networks (ACM/IEEE IPSN 2006), April 19–21, Nashville, TN, USA.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a very large Web search engine query log. *SIGIR Forum, 3*(1), 6–12.

Spink, A., & Jansen, B. J. (2004). *Web search: Public searching of the Web*. New York: Kluwer.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From E-sex to E-commerce: Web search changes. *IEEE Computer, 35*(3), 107–111.

Toms, E. G., Freund, L., Kopak, R., & Bartlett, J. C. (2003). *The effect of task domain on search*. Paper presented at the 2003 conference of the Centre for Advanced Studies on Collaborative Research, October 6–9, Toronto, Ontario, Canada.

Urbany, J. E., Dickson, P. R., & Wilkie, W. L. (1989). Buyer uncertainty and information search. *Journal of Consumer Research, 16*(2), 208–214.

Vogt, C. A., & Fesenmaier, D. R. (1998). Expanding the functional information search model. *Annals of Tourism Research, 25*(3), 551–578.

Wang, P., Berry, M., & Yang, Y. (2003). Mining longitudinal Web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology, 54*(8), 743–758.

Weber, K., & Roehl, W. S. (1999). Profiling people searching for and purchasing travel products on the World Wide Web. *Journal of Travel Research, 37*(3), 291–298.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249–270.

Wilson, T. D. (2000). Human information behavior. *Informing Science, 3*(2), 49–55.

Wolfe, R., Jansen, B. J., & Spink, A. (2006). Semantics and the medical Web: A review of barriers and breakthroughs in effective healthcare query. In E. Li & T. C. T. Du (Eds.), *Advances in electronic business* (Vol. II, pp. 267–279). Hershey, PA: Idea Group Publishing.

Wolfram, D. (1999). Term co-occurrence in Internet search engine queries: An analysis of the excite data set. *Canadian Journal of Information and Library Science, 24*(2/3), 12–33.