Library Hi Tech

From query analysis to user information needs: a study of campus map searches

Zoe Chao

Received 1 December 2014
Revised 23 October 2015
Accepted 29 October 2015

From query analysis to user information needs: a study of campus map searches

Zoe Chao
Penn State University, University Park, Pennsylvania, USA

Abstract

Purpose – Search engines and web applications have evolved to be more tailored toward individual user's needs, including the individual's personal preferences and geographic location. By integrating the free Google Maps Application Program Interface with locally stored metadata, the author created an interactive map search for users to locate, and navigate to, destinations on the University of New Mexico (UNM) campus. The purpose of this paper is to identify the characteristics of UNM map search queries, the options and prioritization of the metadata augmentation, and the usefulness and possible improvement of the interface.
Design/methodology/approach – Queries, search date/time, and the number of results found were logged and examined. Queries' search frequency and characteristics were analyzed and categorized.
Findings – From November 1, 2012 to September 15, 2013, the author had a total 14,097 visits to the SearchUNM Maps page (http://search.unm.edu/maps/). There were total 5,868 searches (41 percent of all the page visits), and out of all the search instances, 2,297 of them (39 percent) did not retrieve any results. By analyzing the failed queries, the author was able to develop a strategy to increase successful searches.
Originality/value – Many academic institutions have implemented interactive map searches for users to find locations and navigate on campus. However, to date there is no related research on how users conduct their searches in such a scope. Based on the query analysis, this paper identifies user's search behavior and discusses the strategies of improving searches results of campus interactive maps.

Keywords Information retrieval, User interfaces, Data analysis, Computer applications, Geographic information systems, Information searches
Paper type Research paper

Introduction

Students often come to a library's reference desk to ask for directions to a specific building or classroom (Bishop, 2011). Typically, they will be shown a print map and offered a verbal explanation, helping them to their destination. Yet, wayfinding interfaces on the web have evolved dramatically; search engines like Google and Microsoft Bing have developed interactive maps for users to locate themselves in real-time and made them freely available as smart phone apps. The prevalent use of smart phones among college students underscores a need for institutions to create truly digital campus maps in order to provide non-static, interactive location searches.

To understand a user's geographic information (henceforth denoted geo-info) need, it is important first to define what geo-info is. Raper (2007) described geo-info "as information whose representation and communication is dependent on the constitutive role of place, space, and time." In principle, geo-info can be represented and communicated in either geo-centric coordinate referencing or places and landmarks (Raper, 2007). For example, Google Maps uses maps, places, imagery, directions, and distance/time to represent the information and to communicate results with users. In the past decade, researchers have

Special thanks go to Randy Johnson and Christin Sraha at the University of New Mexico for their hard work on this project.

Library Hi Tech
Vol. 34 No. 1, 2016
pp. 104-129

0737-8831
DOI 10.1108/LHT-12-2014-0110

proposed the concept of geographic relevance (GR) to address the spatial-temporal context in geographic information retrieval. Raper (2007) defines GR as "a spatial-temporally extended relation between information need and geographic information object." In order to be geographic relevant, i.e. to meet user's geo-info need, three aspects need to be considered. First, a geographic entity in the real world must be well represented in a computer system. Second, a user's spatial-temporal context, e.g. location, mobility, etc., must be taken into account, in addition to the normal relevance ranking. Third, the interface must be able to communicate the relevant geographic information effectively (Raper, 2007; De Sabbata and Reichenbacher, 2012).

In 2012, the campus Information Technologies (IT) department and the University Libraries (UL) at the University of New Mexico (UNM) created an interactive map search for users to locate, and navigate to, destinations on the UNM campus by integrating the free Google Maps Application Program Interface (API) with locally stored metadata. The Google Maps API has been applied to provide geo-info in many different contexts: to optimize resource distribution (Wang *et al.*, 2009); to refer hospitals (Kobayashi *et al.*, 2010); and to provide real-time traffic flow (Wu *et al.*, 2007). It is also more light-weight and less burdensome, compared to OpenLayers API and ArcGIS API (Fernandes *et al.*, 2013). More importantly, it is an economically sound choice (Pan *et al.*, 2007; Kobayashi *et al.*, 2010). We made the assumption that users reveal their current geo-info interest and information need by submitting a query to a search engine. By identifying what people search for and the words used in such searches, we hoped to gain a better understanding of the interaction between users and our map search system, as well as to evaluate the representation of geographic entities on campus, the impact of user's context, and the effectiveness of our interface.

The research questions we hope to answer are the following:

RQ1. What are the characteristics of a UNM map search query?

RQ2. Do the data and metadata of the geographic entities on campus align with the user's information needs?

RQ3. Does our map search interface communicate effectively enough for users to retrieve the information?

Literature review

To retrieve information from general web resources, the classic information retrieval system (CIRS) generally evaluates relevance based on the relation between queries and the text in web documents. If often does not consider a user's spatial-temporal context, i.e. GR. To facilitate the geo-info retrieval in a CIRS, researchers have studied the characteristics of geographic queries (henceforth denoted geo-query), the geographical content of web resources, and the indexing and processing power of the search tools. Additionally, there is research focussing on systems primarily designed for geo-info known as geographic information retrieval systems (GIRS), for example, a Google-Maps-based traffic information system.

Though two different systems, CIRS and GIRS are not exclusive of each other, as a matter of practical use. Often, CIRS can serve as a gateway to GIRS: for example, a user does a Google search (interacts with a CIRS) and clicks a link to Google Maps (interacts with a GIRS). In many cases, however, the connections between GIRS and CIRS are absent. As a result, the rich geo-info developed in GIRS is not exposed to the more widely used CIRS. Although our main research scope is limited to the GIRS within the UNM campus, we review the progress of geo-info retrieval in both CIRS and GIRS to provide a more comprehensive overview.

Characteristics of geo-query

Analyzing query logs is an effective approach to gain insight on users' search behaviors and information needs. Several researchers have analyzed queries from web search engines, such as Excite (Sanderson and Kohler, 2004) and AOL (Gan *et al.*, 2008), in an attempt to differentiate geo-queries from general web queries. The task of differentiating between a geo- and non-geo-query can be tricky because the inclusion or exclusion of a location name cannot be the sole indicator to categorize the geo- or non-geo-query. For example, "New York pizza" can be interpreted either as a request for pizza locations in the New York area, or the topic of New York style pizza (Gravano *et al.*, 2003). On the other hand, a question like "what is the name of that mountain?" certainly has a geographical component, because the answer depends on where and when the user is situated (Mountain and MacFarlane, 2007). Although studies show that about one-fifth of the queries submitted to a general web search engine can be categorized as geographically related (Sanderson and Kohler, 2004; Zhou *et al.*, 2005), the number could rise to 60 percent when simulating a situation in which queries were conducted via mobile devices in outdoor settings (Mountain and MacFarlane, 2007). As Mountain and MacFarlane (2007) point out, "'mobile individuals' information needs are more likely to be a product of their surroundings, and the environment in which they are interacting."

Noticeably, in the general web search environment – i.e. a CIRS – geo-queries tend to be longer than general queries. The average length of a geo-query is between three to four words, compared to two to three words for a non-geo-query (Spink *et al.*, 2002; Sanderson and Kohler, 2004; Gan *et al.*, 2008; Jones *et al.*, 2008). Because queries are often categorized as geo-related based on the additional geographical term(s), they are forced to be longer. Also, it is difficult to perceive the underlying geographical intent of a very short query when examining the web log from general web search engines (Gan *et al.*, 2008). However, in the study of Microsoft Live Maps, which is a GIRS, the average query length is 4.4 words, significantly longer than either geo- or non-geo-queries in a general web search (Xiao *et al.*, 2010). According Xiao *et al.* (2010), this is because "people understand that map search systems are designed for geographic information retrieval. Therefore, they intentionally specify places, sometimes very detailed addresses, in the query, to search for driving directions, for example" (pp. 8, 14). Interestingly, Taghavi *et al.* (2012) conclude that the average length of general queries has grown steadily over time, indicating that in order to have better precision in the search results, users tended to type in more words in response to the ever-increasing online resources. Both studies suggest that users may have continued to modify their search behavior over time based their experiences with search engines.

Geographical data and metadata

Because a GIRS is largely based on a framework of existing geo-data with related information depending on the nature of its purposes and geographic scope, its data and metadata are generally normalized and consistent. For example, the WebPark pilot project, in which a mobile information system for Swiss National Park was created, Mountain and MacFarlane (2007) collected and organized data and documents based on the park's geography and supplemented the content with information such as places of interests, hiking routes, fauna and flora information, and even a plant identification tool. There are two ways to handle data and metadata for the Google Maps API based GIRS: either the data and metadata were collected and stored locally (e.g. categorized information about medical providers (Kobayashi *et al.*, 2010) and tourist information (Pan *et al.*, 2007)), or they were retrieved constantly from an outside data source (e.g. real-time traffic flow information (Wu *et al.*, 2007)). In both approaches, the information then was overlaid on top of Google Maps.

Unlike GIRS, the data and metadata in CIRS are not standardized. To improve geo-info retrieval in CIRS, researchers have identified and added geo-related metadata to web resources. According to Amitay *et al.* (2004), more than one-third of the geographic-names mentioned in web documents have more than one meaning. The gazetteer approach – matching text with a list of geographic terms – is often used to disambiguate terms (e.g. "reading" and "Reading" (PA)) to further pinpoint the locational focus of the web resources (Amitay *et al.*, 2004; Zhou *et al.*, 2005). Besides the textual information of web documents, the locations of their web environments and hyperlinks, as well as the significance of the web content and their geographic scope, are also taken into account in enriching the metadata (Ding *et al.*, 2000; Wang *et al.*, 2005a; Silva *et al.*, 2006; Jones *et al.*, 2008). For the broad and diverse sources of CIRS, GIRS can be complementary in facilitating geo-info retrieval.

Indexing and query processing

Different ways of indexing and query processing in CIRS have been studied to improve geo-info searches. After classifying queries into local and global, based on the prevalence and diversity of locations in search results, Gravano *et al.* (2003) proposed to further "localize" or "globalize" the search results by appending the user's location to the query and re-ranking the results. Considering that the gazetteer approach has limits on incorporating context information, Wang *et al.* (2005b) looked into improving search relevance by detecting the query's dominant location, ascertained from the query, transaction log, and search result. Simple keyword matching does not take into account geographical relationships, such as a township within a county within a state. Geographic ontologies were developed to detect spatial relationships and to retrieve and rank web pages based on these textual and spatial indices (Abdelmoty *et al.*, 2005; Fu *et al.*, 2005). Researchers have designed location-aware search engines by integrating spatial indexing and text indexing (Lee *et al.*, 2005; Vaid *et al.*, 2005; Zhou *et al.*, 2005; Chen *et al.*, 2006).

Although the sources of data in GIRS may not be as expansive as the web, database design, and indexing performance are critical issues nonetheless (Wu *et al.*, 2007; Wang *et al.*, 2009). Wu *et al.* (2007) described how, by improving the database design and the performance of indexing, they were able to cut down the processing time from over an hour to 0.05 seconds to query the traffic information for the whole city of Bellevue. Similar to detecting a query's intent and its dominant location in CIRS, a user's spatial-temporal context in GIRS is crucial in ranking GR. In short, users find information more geographically relevant if it is: closer (spatial proximity), more visible (visibility), reachable in shorter period of time (temporal proximity), and dovetailing with the person's recent spatial behavior (speed-heading prediction surfaces) (Mountain and MacFarlane, 2007). These criteria serve as effective filters for refining the alignment between a system's GR ranking and an individual's information needs.

Interface

No matter how robust the data and algorithm may be, a GR will not succeed if the interface cannot communicate effectively with users. Given that the communication starts even before a query is formed, a good interface design can guide users in articulating their intent.

For location-aware web search engines, various researchers have designed an interface that allows users to draw a region for their search scope, in addition to entering a text query (Lee *et al.*, 2005; Zhou *et al.*, 2005). Mountain and MacFarlane (2007) distill the four filters, described in the previous section, to user-friendly labels – "search around me," "visible places," "accessible places," and "search ahead" – in the WebPark mobile information system. The feedback from their users verifies that such filtering options on a GIRS mobile interface are extremely helpful. There is a direct relationship between the user's motivation and an optimal degree of interface complexity. For example, in a nature preserve area, a new-comer, a regular visitor, and a researcher are each motivated by different levels of information need, which directly affects their preferred quantity of presented information (Roth and Harrower, 2008). The information-to-interface ratio is critical in designing an interface, especially for mobile devices.

Studies have also shown that a map-based interface is not always the user's preferred wayfinding interface (Cheung, 2006; Church *et al.*, 2010). Church *et al.* (2010) found that situational context and a user's information needs strongly influence the user's satisfaction with an interface, as users tend to favor, on mobile devices, a map-based interface in unfamiliar surroundings and a text-based interface in a familiar area. Many GIRSs have developed functions that allow users to customize personal geo-info environments and to share geographic references and observations within the community (Mountain and MacFarlane, 2007; Pan *et al.*, 2007; Wu *et al.*, 2007; Church *et al.*, 2010).

Background

The UNM used the Google Search Appliance (GSA) to provide search for the University's Web content from 2006 to 2011. When the GSA license expired, the campus IT and the UL took the opportunity to collaborate on a re-launch of the service as a joint project, called "SearchUNM." When the project "kicked off" in 2011, we envisioned three main categories of searches for users to find information on the SearchUNM page: websites (for campus websites), people (for the employee directory), and maps (for building locations). In the first year after the launch, the planning and campus development's webpage, with printable PDF map links, was the only option when users clicked the "Maps" tab. Perceiving the increasing need for students to search and pinpoint locations interactively, a small team (one programmer from IT and two librarians from UL) within the larger SearchUNM team decided to explore other options.

In early November 2012, the team launched the interactive map search for locations on UNM's main campus and branch campuses. The SearchUNM Map consists of two main components: the base map information and functions from Google Maps, and the data and metadata of campus locations created and stored locally. The free Google Maps API provides the basic Google Maps information, which includes streets and building outlines. We also took advantage of Google Maps' basic navigation functions, such as panning and zooming, which users are familiar with, as well as the options of regular map view and satellite imagery. The data on the Google Maps cannot always be guaranteed to be correct, since the map information is maintained by Google, and we do not have control over errors or omissions on the actual Google Maps. The UNM team put together step-by-step instructions on how to report problems directly to Google Maps, in response to people's reporting errors on the map.

System implementation

The data and metadata of campus locations are stored in two types of files: Keyhole Markup Language (KML) and JavaScript Object Notation (JSON). KML is a file format based on XML for expressing geographic data and visualization in Google Maps and Google Earth (https://developers.google.com/kml/). JSON is a light-weight, human-readable data-interchange format based on JavaScript Programming Language. We used a web server (HTTP) and PHP Hypertext Preprocessor (PHP) for our web applications. The client side uses CSS, HTML, JavaScript, the Google API, and KML to render the map display. PHP is used in the server-side processing to search JSON files and pass location info to the client in order to display information and icons on the map.

Data and metadata

In our SearchUNM Map interface for the UNM main campus (see Figure 1), the team decided to offer nine basic categories for users to "browse" in addition to the search function. Each category (e.g. "Libraries" or "Dining") was encoded into a KML file, which is a data layer containing the coordinates and the designated icon for display on the Google map. For example, when a user selects the category, "Dining," the browser interface will display the layer of "Dining" with different locations (the coordinates) and the icon (in this case, a fork and a knife) based on the metadata in its KML file. More than one category/layer can be selected and displayed simultaneously. For the locations that are not listed in the categories, the coordinates are grouped in one KML file with a red pushpin as the default icon.

In the JSON file, each location consists of metadata for the title, building number, abbreviation, campus, keywords, latitude/longitude, an image URL, a link for a building page, an internal unique ID, and a description. Only the title, ID, latitude, and longitude are the required elements. The location title, home page URL, description, and image would display in the pop-up window when users select a result or click the icon (Figure 2). The title, keywords, and description are the indexed fields. Currently our search function is carried out by matching queries with the indexed metadata in the JSON files. It is a light-weight application that requires an exact match to retrieve results; it does not have the flexibility, such as Boolean logic, found in a regular rational database and Structured Query Language. Therefore, for example, it would retrieve no result if students were searching for the "Science and Mathematics Learning Center" with the query "science math learning center."

Initially, the programmer on the project painstakingly collected all the metadata by looking up the information online. Early on, the team contacted different campus units for contact points to review and to update the metadata, but only a few of those contacted gave us feedback.

Interface

The layout of SearchUNM Map is very simple: it shows a search box along with nine categories listed on the left-hand side with a Google Maps background. Originally, the categories of Libraries and Visitor Parking were checked by default (Figure 3). When an icon is clicked, the pop-up window opens on the right-hand side and the red pushpin icon is centered (see Figure 2). Within the pop-up window, in addition to the location's information, options for direction links ("to here" or "from here") and the location URL for bookmarking the page are provided. The design aligns the boxes on the right and left edges and leaves the map area as open as possible to the users.

Methodology

Search log analysis

Online searching has become a daily norm for many people. The web is a constant destination for information seekers. Web servers can be set up to record the details of

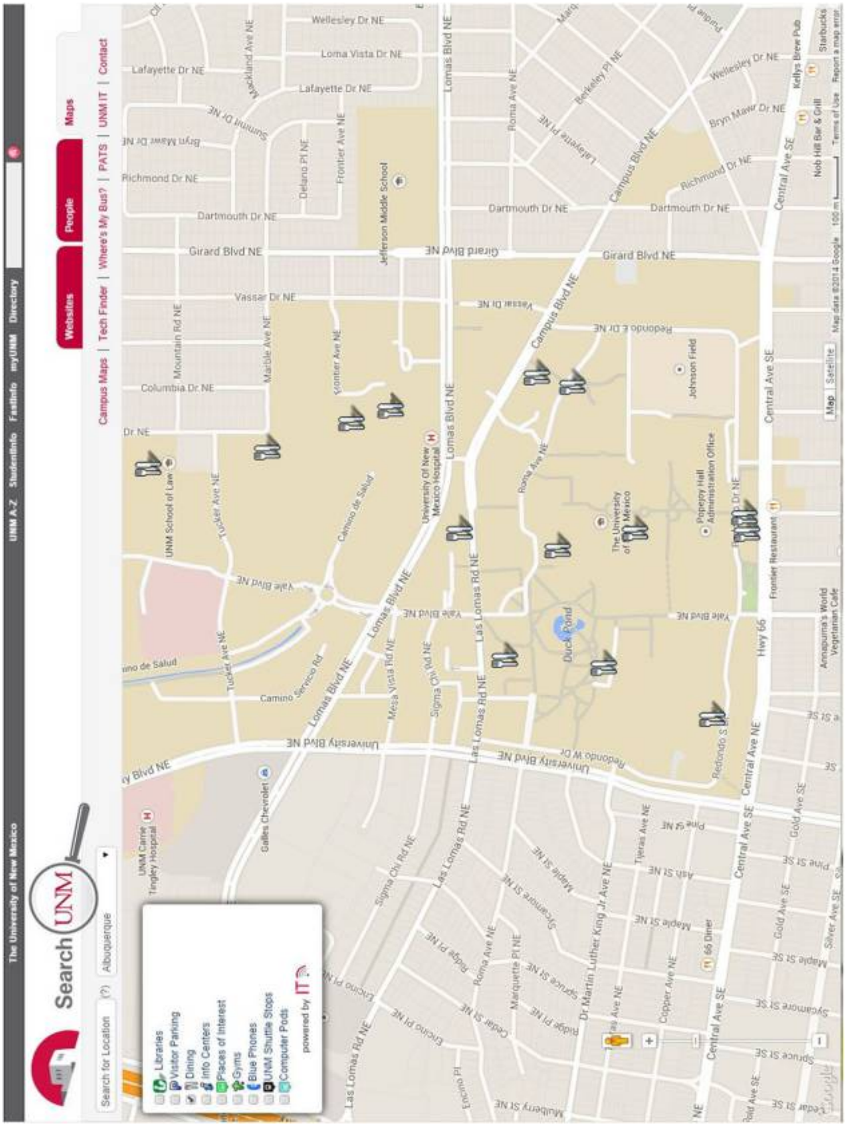


Figure 1.
The interface of SearchUNM Map (http://map.unm.edu/), with Dining category checked

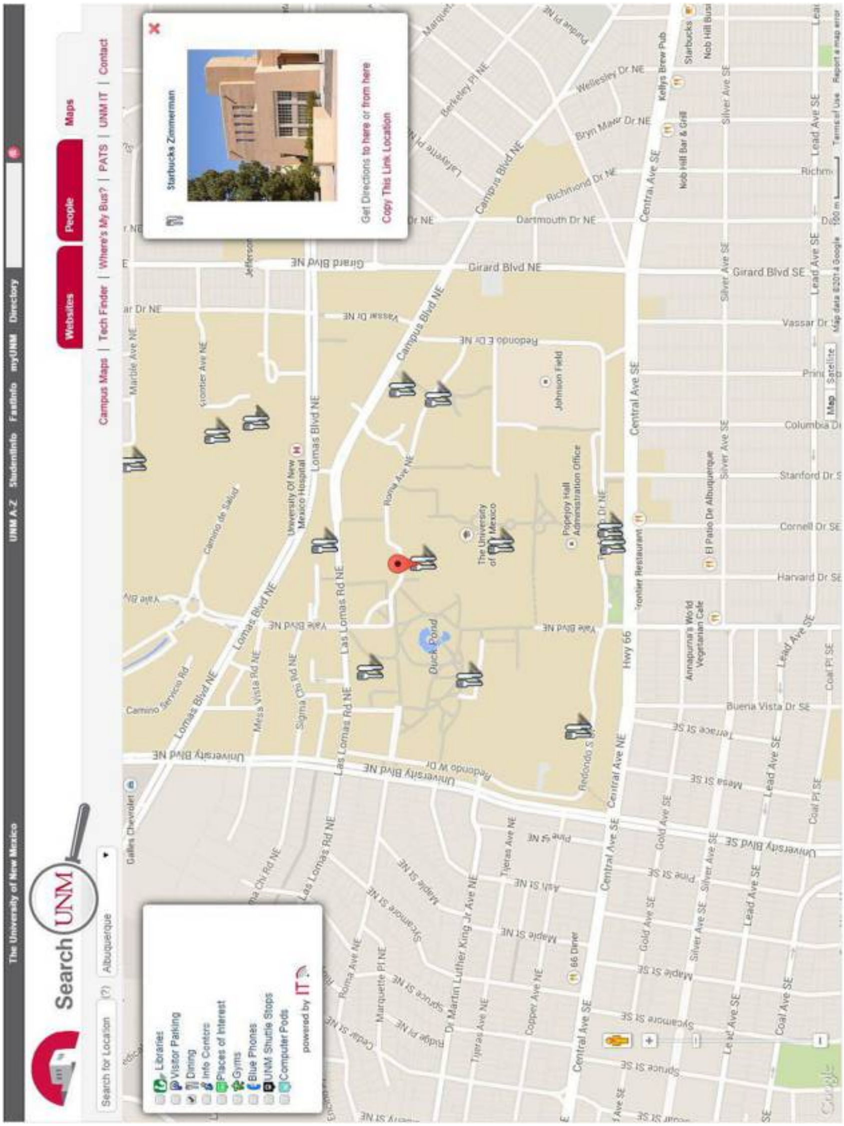


Figure 2.
The interface of SearchUNM Map, with Dining category checked, and the detailed information in the pop-up window when an icon is clicked

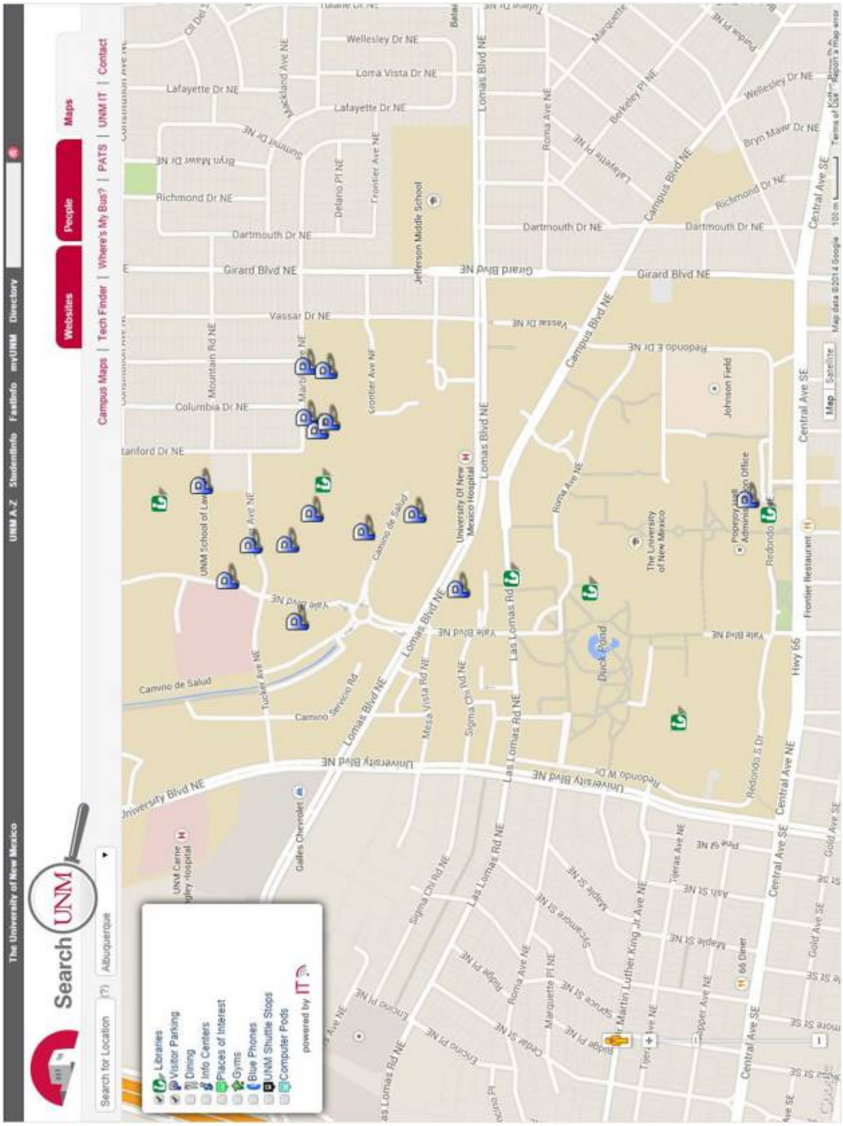


Figure 3.
The original default interface of SearchUNM Map, with "Libraries" and "Visitor Parking" categories checked

interactions automatically between users and online search tools. The logs of such electronically recorded interaction are generally referred as web server logs, transaction logs or search logs (Peters, 1993; Jansen, 2006; Asunka *et al.*, 2009). The data collected in search logs vary depending on the needs of those accessing the logs. Typically, the client computer's internet protocol (IP) address, user query, and timestamp are recorded. (Jansen, 2006; Asunka *et al.*, 2009). In the context of research on web searching, log analysis is defined as "the use of data collected in a transaction log to investigate particular research questions concerning interactions among Web users, the Web search engine, or the Web content during searching episodes" (Jansen, 2006). Here, for the purposes of this paper, a web search engine can be a general-purpose search engine, a niche search engine, or a search application on a website.

Strengths and limitations

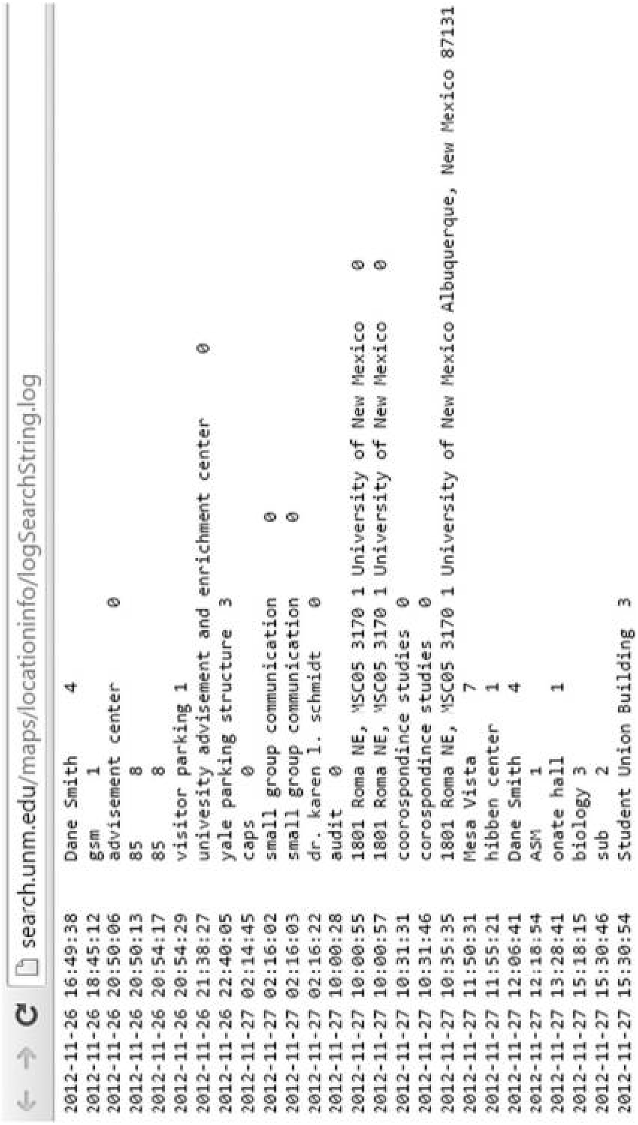
Search logs consist of interactions between users and the system. These interactions may include a user submitting a query, the system responding with a result page, and the user clicking a URL listed in the result. Therefore, the interactions can be seen as mechanical expressions of underlying information needs or motivations (Jansen, 2006). By examining the search log, one may identify common characteristics of interactions between users and the system and, consequently, address issues such as system performance, information structure, or measurements of user interactions.

One main advantage of analyzing search logs is that significant amounts of search data for a sizable number of system users can be collected inexpensively and unobtrusively. Researchers have critiqued such log analysis, because it deals with only the interactions, and fails to record the personal context of individual users, such as, for example, their background, perceptions of the search, information needs, and satisfaction with the system (Kurth, 1993). However, as Jansen (2006) points out, "these are issues with many, if not all, empirical methodologies." Although log analysis on its own is incapable of giving the complete picture of user experience with a system, it is an invaluable approach to better understanding the online information search process. Moreover, it can be extremely helpful when used in conjunction with surveys, usability studies, and so forth (Jansen, 2006). Several studies reviewed for this paper have developed insight into user-system interactions from log analysis (Sanderson and Kohler, 2004; Gan *et al.*, 2008; Xiao *et al.*, 2010).

Research design

This study analyzed search logs collected from November 1, 2012 to September 15, 2013, for the SearchUNM Map page of the SearchUNM website (http://search.unm.edu/). Typically, search logs record data such as the client computer's IP address (or anonymous user ID), user query, and search engine access time (Jansen, 2006). For this study, we added functions in the PHP files to record data in our server. We collected and displayed the date/time of each search performed, its query, and the number of results retrieved with the query (Figure 4). We also created a simple visualization using colors to demonstrate whether or not any result was found by each unique query, and the total number of times the query occurred (Figure 5). We omitted IP address and client ID data to mitigate privacy issues. Also, we logged our activities to record what files had been updated and our correspondence with the contact persons for the campus buildings, if identified.

Though there are a few fairly comprehensive research projects on web queries using search log analyses (Jansen *et al.*, 2000; Spink *et al.*, 2002; Wang *et al.*, 2003), a standard


search.unm.edu/maps/locationinfo/logSearchString.log

```
2012-11-26 16:49:38	Dane Smith	4
2012-11-26 18:45:12	gsm	1
2012-11-26 20:50:06	advisement center	0
2012-11-26 20:50:13	85	8
2012-11-26 20:54:17	85	8
2012-11-26 20:54:29	visitor parking 1
2012-11-26 21:38:27	univesity advisement and enrichment center	0
2012-11-26 22:40:05	yale parking structure	3
2012-11-27 02:14:45	caps	0
2012-11-27 02:16:02	small group communication	0
2012-11-27 02:16:03	small group communication	0
2012-11-27 02:16:22	dr. karen l. schmidt	0
2012-11-27 10:00:28	audit	0
2012-11-27 10:00:55	1801 Roma NE, MSC05 3170 1 University of New Mexico	0
2012-11-27 10:00:57	1801 Roma NE, MSC05 3170 1 University of New Mexico	0
2012-11-27 10:31:31	coorospondince studies	0
2012-11-27 10:31:46	corospondince studies	0
2012-11-27 10:35:35	1801 Roma NE, MSC05 3170 1 University of New Mexico Albuquerque, New Mexico 87131
2012-11-27 11:50:31	Mesa Vista	7
2012-11-27 11:55:21	hibben center	1
2012-11-27 12:06:41	Dane Smith	4
2012-11-27 12:18:54	ASM	1
2012-11-27 13:28:41	onate hall	1
2012-11-27 15:18:15	biology 3
2012-11-27 15:30:46	sub	2
2012-11-27 15:30:54	Student Union Building	3
```

Figure 4.
The date/time of each search performed, its query, and the number of results retrieved with the query

Notes: Text highlighted in green means that results were found. Text highlighted in red means no results were produced. The numbers before the colons signify the total number of query occurrences to date. For example, at the time this screencast was generated, both "continuing education" and "math" had been searched 31 times, independently

Figure 5.
The number of occurrences of each unique query

methodology has not yet been developed. To address this shortcoming, Jansen (2006) proposed a three-stage process as follows:

- collection: the process of collecting the interaction data for a given period in a transaction log;
- preparation: the process of cleaning and preparing the transaction log data for analysis; and
- analysis: the process of analyzing the prepared data.

As the team conducted the research, this process was adopted in a broader sense.

Since log files can often get very large, there are software products developed to automate the analyzing process. Nevertheless, some researchers have found manually studying the search log can uncover more information about website use (Asunka *et al.*, 2009). Our team decided to manually analyze the data, since the amount of data was manageable. The search logs were converted to comma-separated values, and ingested into Microsoft Excel for further analysis.

Results and discussions

Page traffic and query distribution

From November 1, 2012 to September 15, 2013, we had a total 14,097 visits to the SearchUNM Map page. There were a total of 5,868 searches (41 percent of all the page visits), and out of all the search instances, 2,297 of them (39 percent) retrieved no results (see Figure 6).

For clarity, the term "search" is used to describe a search instance, and the term "query" is the text submitted by a user in a search instance. One query can be used in several search instances. For example, during this period of time, "library" was searched 240 times. Therefore, there were a total of 240 search instances with the query "library." In the alternative, we can say the search frequency of the query "library" was 240.

There were total of 2,357 unique queries. Of all the unique queries, 644 (27 percent of total queries) accounted for 3,571 successful search instances (61 percent of total search instances). The 2,297 failed search instances consisted of 1,713 unique queries (73 percent of total queries; see Figure 7).

Figure 8 shows the distribution of unique queries based on the search frequency. From the graph, we can infer that, first, not all the unique queries are equally significant. About 70 percent (1,692 out of 2,357) appear only once. Second, some queries were used much more repetitively, indicating there are pockets

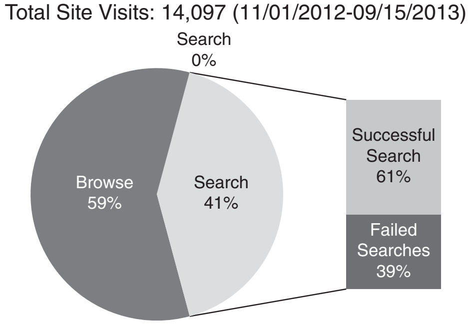


Figure 6.
The browses and searches of total site visits during the study time

of shared geo-info needs among users on UNM campus. Third, although queries with high-search frequency are comparatively fewer, their recurrences are expected to grow exponentially.

When looking at the query success rate for different search frequencies, we find that queries searched more than ten times, for the most part, have successful results. Afterward, the successful rate dropped as the search frequency declined (see Figure 9). When the number of search instances were added to the search frequency and search outcomes, the search results were largely successful, except for the queries with a very low search frequency (see Figure 10, and Table AI). We suspect many of these infrequent failed searches resulted idiosyncratically, from personal mistakes, such as typos, or misconceptions of the system. Users may have learned from the failed searches and modified their search behavior as they became familiar with the system (Xiao *et al.*, 2010; Taghavi *et al.*, 2012). On the other hand, for the clusters of common geo-info needs on UNM campus, the system more or less aligns with users' perceptions of the map search, and the associated information is usually retrieved successfully.

Query analysis

Although search logs analysis is inexpensive, unobtrusive and scalable, the log does not offer any information about a user's intent or context. For example, when a

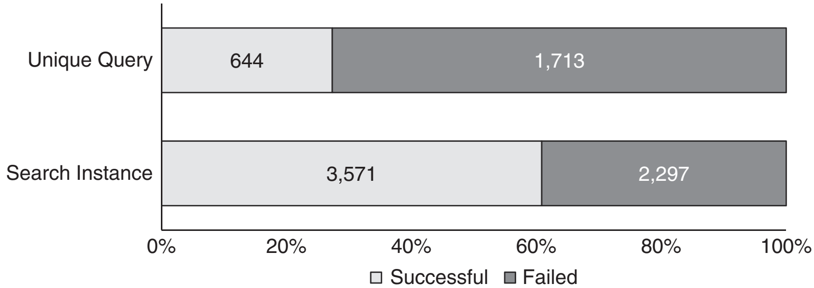
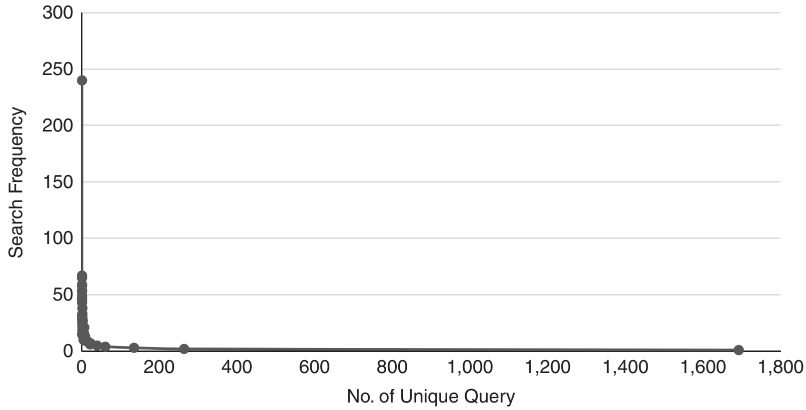


Figure 7. The success rates of the unique queries, and the search instances



Notes: On the left end of the curve, the most searched unique query was searched 240 times. On the right end, a total of 1,692 unique queries were searched only once

Figure 8. The distribution of unique queries and their search frequency

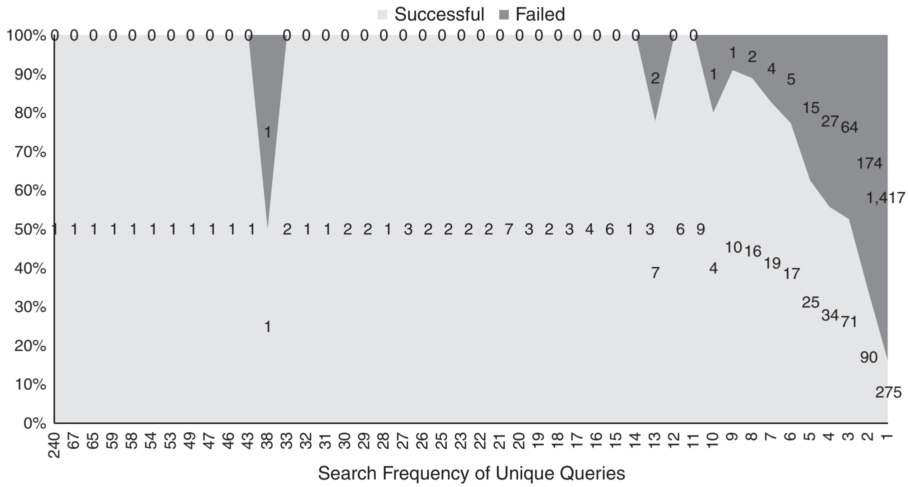


Figure 9.
The relationship between the search frequency of unique queries and the outcome of searches (successful or failed)

Notes: For example, out of the total searches, there are two unique queries that have been searched 38 times (i.e. search frequency), respectively. One of them returned search result, the other failed, meaning that for the queries searched 38 times, the success rate is 50 percent

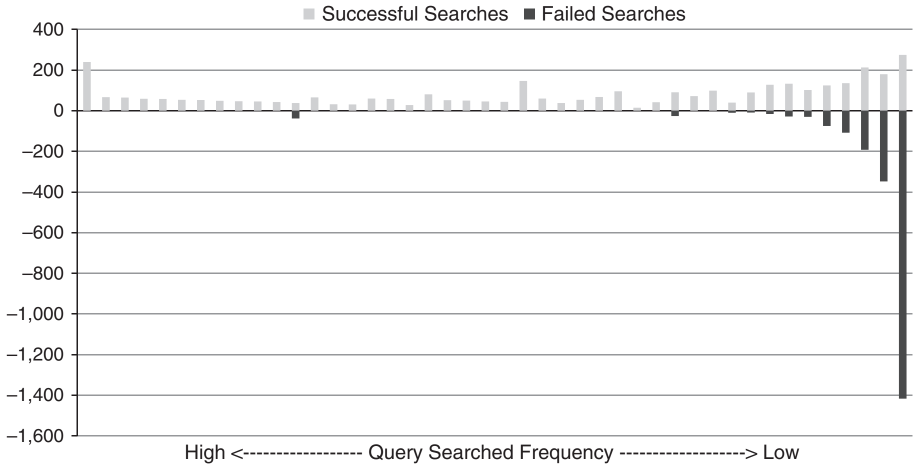


Figure 10.
Distribution of successful and failed searches based on the search frequency

user searches the word "registration," does the term refer to an office in a building, a campus unit, or the function of a campus unit? Likewise, we are unable to know the user's opinion of, and satisfaction with, the system (Kurth, 1993; Jansen, 2006). Nevertheless, we know that users generally are seeking geo-info; when they search "food," they are most likely looking for places providing food. The most frequently searched query in SearchUNM Map is "library," with 240 search instances during the period of analysis. Table I lists the top ten searched queries, all of which resulted in successful searches. Highly similar queries are not combined in the table – for example, "dane smith hall," "dane smith," and "dave smith" – in an attempt to understand the exact queries users have in mind.

Rank	Query	Search instances	Main usage
1	library	240	Library
2	mitchell hall	67	Classrooms
3	woodward	65	Classrooms
4	education	59	Classrooms
5	dane smith hall	58	Classrooms
6	dane smith	54	Classrooms
7	mesa vista hall	53	Administration/classrooms
8	center for the arts	49	Departments/classrooms
9	humanities	47	Departments/classrooms
10	smlc	46	Departments/classrooms

Table I. The top 10 queries in UNM map search

Not all queries with high-search frequency brought back results. Since these queries represented a shared geo-info need among users, examining and fixing them will be the strategically sound first step to increase successful searches. From the list of top ten searched queries, we can see that, first, most of these queries are the names of buildings used primarily for classrooms, suggesting that is the principal need of users. Second, although "hall" is part of a building name, it is not a keyword that helps retrieve correct results. However, students tend to include it in the query in most cases (except "woodward" for "Woodward Lecture Hall"), possibly because terms with the word "hall" appear in the course catalog, or on a course syllabus, as part of the names for these buildings. Such querying also means that not all the users take advantage of the autocomplete list, which was implemented about a month after launch. The correct building names and possible keywords are listed in a JSON file. The algorithm generates a list of suggested terms after the user types in the first two letters. The suggested terms in our list do not include the word "hall," which implies that many users tend to type faster than the list populated, or just copy and paste and miss the list. Third, acronyms are commonly used for long queries, for example, "smlc" for "Science and Mathematics Learning Center."

According to researchers, geo-searches tend to have more words in the query, compared with regular web search queries (Sanderson and Kohler, 2004; Gan *et al.*, 2008). Table II shows the numbers of queries and searches based on the word count of queries.

In our study, the average query length is two words. As shown in Table II, the one-word and two-word queries dominate both the total unique queries and the total search instances. The numbers here, however, cannot be interpreted as less than the average geo- or non-geo-queries from previous researchers, for the following reasons. First, the scope of our data is defined within the UNM campus. Users don't need to add an extra word(s) – for example, a city name – to indicate their location intent. Second, queries can be influenced by the suggestions generated from our autocomplete, and thus do not truly reflect users' original query forms. Third, because our search function requires an

Word(s) in query	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No. of unique queries	869	848	362	161	70	13	19	4	2	6	1	1	0	1
% out of total unique queries	37	36	15	7	3				2					
No. of search instances	2,832	1,840	715	315	116	14	20	4	2	7	1	1	0	1
% out of total searches	48	31	12	5	2				1					

Table II. The length of queries and their query counts and search instances counts exact match with our JSON files in order to retrieve results, we listed some "hints" to help users avoid failed searches (see Figure 11). One of the hints is "Try a single word, rather than a phrase," which could influence how users construct their queries.

Data and metadata

A failed search results from a user's query finding no matching text from our indexed data and metadata, including the title, keywords, and description of the locations. About 70 percent of the total unique queries failed to retrieve any result. More than 80 percent of failed searches result from queries that have only one search instance. Only 3.3 percent (58 out of 1,713) of the failed searches are caused by queries that were searched more than three times. Ensuring that queries with high-search frequency result in no failures is a key priority for improving the system.

Most of the queries with failed searches fit in the following principal categories. The first category is, the wrong scope. Some users searched for locations outside of the UNM campus, which is outside our scope. An even more typical "problem" query was a building name with a room number. Somehow, many users expect results featuring floor plans, with room numbers mapped out within buildings. To help improve success rates for such queries, we continue to add functional units and offices into the metadata. For example, searching for "registrar" now brings up the building where the office of the registrar is located. We did not, however, include the specific room number of the "Office of the Registrar" in the metadata. We plan to facilitate this type of query by trimming off the numbers that follow the building names in the queries.

The second type of search failure was address queries. Some users copy and paste whole or partial addresses and expect the same search function found in Google Maps (https://maps.google.com/). Although the Google Maps API is in use, the metadata of UNM locations are created and stored locally. Not every building has an address in the metadata. In addition to the standard address format, the department name and mail stop code (MSC) are key components for mail delivery on the UNM campus. For example, "University Libraries, MSC05 3020" is the MSC for the UNM Libraries, and includes four different buildings. This is the type of address that a user might find on any UNM website. In our data, queries that start with "MSC" count for 155 search instances.

Misspelling or misremembering a name is the third category of search failure. Although misspellings account for fewer than 10 percent of our total queries, it is nonetheless interesting, given that – as mentioned previously – the autocomplete function is in place. Again, users' typing speeds may be too fast to allow query suggestions to pop-up in time. Also, the first two letters need to be correct to trigger the appropriate suggestion. For example, if user spells "Re" instead of "Ra," the spelling suggestion "Raynolds" will not appear. The difference in correct and incorrect building names sometimes can be subtle, such as misremembering "Center for the Arts" as "Center of the Arts."

The fourth category involves variations on building names. In addition to acronyms, people tend to use the more familiar unofficial name to search for a building or office instead of the official one. For instance, many people are unaware that the official name of the UNM "lock shop" is "Physical Plant Department." Also, people use many non-keywords like "building," "office," "suite," and "hall" in queries, which increase the probability of a failed search, such as searching "dsh building" for "Dane Smith Hall." These non-keywords should be truncated before processing. Finally, new buildings caused search failures. During the summer of 2013, two major construction sites were completed on the UNM main campus. The metadata for these two locations was not created before users began searching for the buildings.

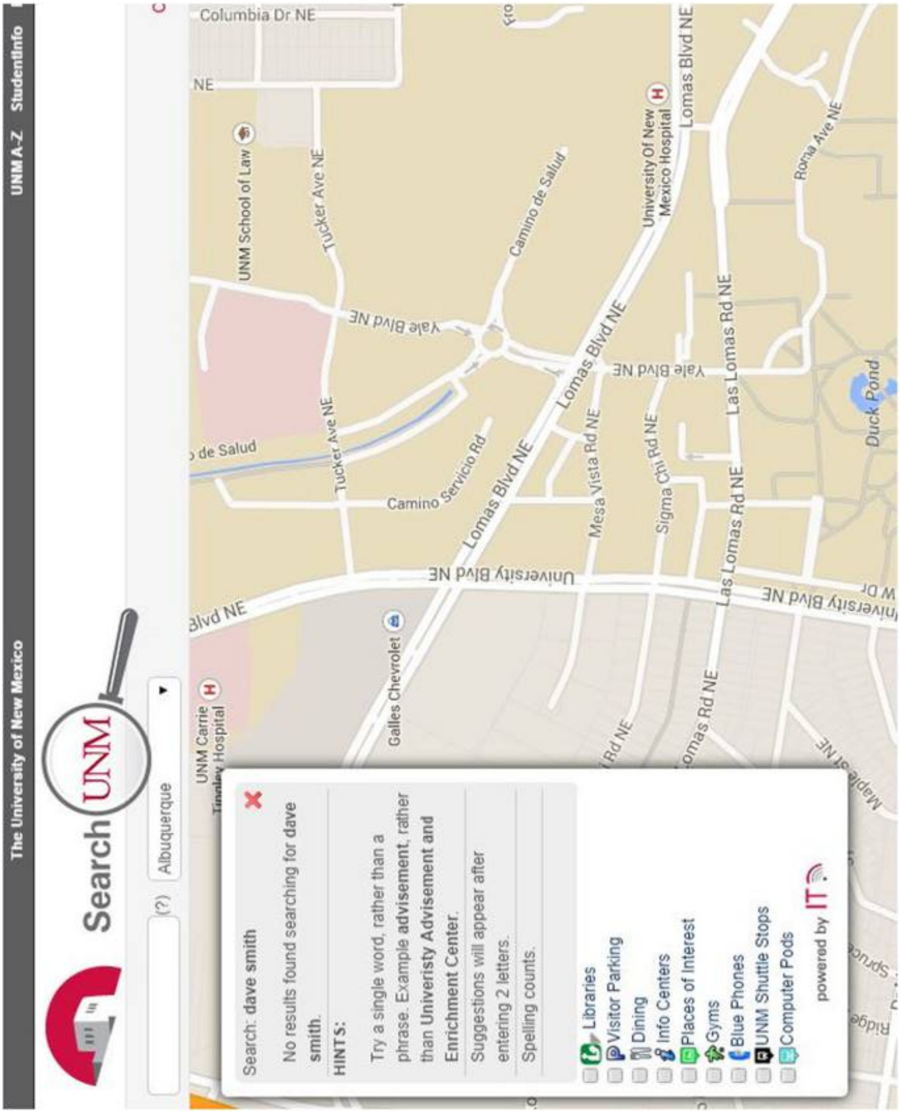


Figure 11.
Hints appear after
a failed search

Indexing and query processing

One main drawback of using JSON and KML instead of a database for our data and metadata is that JSON and KML limits the search function to only exact matches between the query and our data and metadata. At this current stage, we do not have plans to modify the query processing except to improve the system by enhancing the indexed data and metadata of the failed queries. But there are approaches we may consider should we decide to go forward in the future.

First is the issue with the search scope. The levels of geo-info on campus can generally be seen as room, building, campus, and city; each level may well develop its own GIRS. From the result, we know that many users are not aware of the scope of the SearchUNM Map. As users may go beyond our search capacity in either direction, we can "localize" or "globalize" the results by detecting the queries (building name with room number or postal address). This approach would inform users and take them to the appropriate GIRS (possible an interactive floor map application or the Google Maps).

Second is the ranking of the search results. Our light-weight, exact match search function does not have a ranking algorithm. With more participation from campus units and more details of departments, offices, and buildings added to our metadata, the result list has grown rapidly. For example, searching "education" produces a list of 24 alphabetically ordered building/department names with the most likely relevant location, "Education Classrooms," placed in the middle of the result list. Re-arranging the ranking after detecting a user's current location is one remedy to consider: the closer the location is, the higher GR it is to the user (Mountain and MacFarlane, 2007). We can also come up with options/filters to take account of a user's context in the results. As the research by Mountain and MacFarlane (2007) has shown, users find "search all" is the least relevant option when much information is available.

As mentioned previously, CIRS can serve as a gateway to GIRS. For example, if a reference to GIRS is provided in a web document, it is likely the information will be indexed in CIRS. To facilitate the discovery of SearchUNM Map (i.e. GIRS) in the overarching SearchUNM (i.e. CIRS), the two librarian members used iframe to embed the link, http://map.unm.edu/m/?searchterm=library, in the UL Map webpage of the library website. This link basically takes users to the same result as occurs by searching "library" in the SearchUNM Map search. Consequently, when users search "library map" in SearchUNM, they can see the UL Maps in the results (see Figure 12) and access the interactive map (see Figure 13). Users can also link to either the SearchUNM Map page or the campus PDF map on the page.

Interface

The selection of the nine browsing categories (which now have become fifteen) was based on the team's decision on what users might need. Originally, "Libraries" and "Visitor Parking" were checked by default because we expected these two categories were what students and visitors need the most, perhaps showing our own bias as there were two librarians on the team. The data collected do suggest that student were looking for the library most frequently. Although 60 percent of the site traffic did not involve searches, the high frequency of "library" search instances shows that the default "Libraries" layer was not noticed. It could be the library icon we used fits too well within the Google Map background and did not grab the user's attention (see Figure 3).

A variety of factors can influence users to choose between searching and browsing. It is clear that this decision-making process is often based on the amount of effort the user predicts will be required for results, which is a cost-benefit consideration

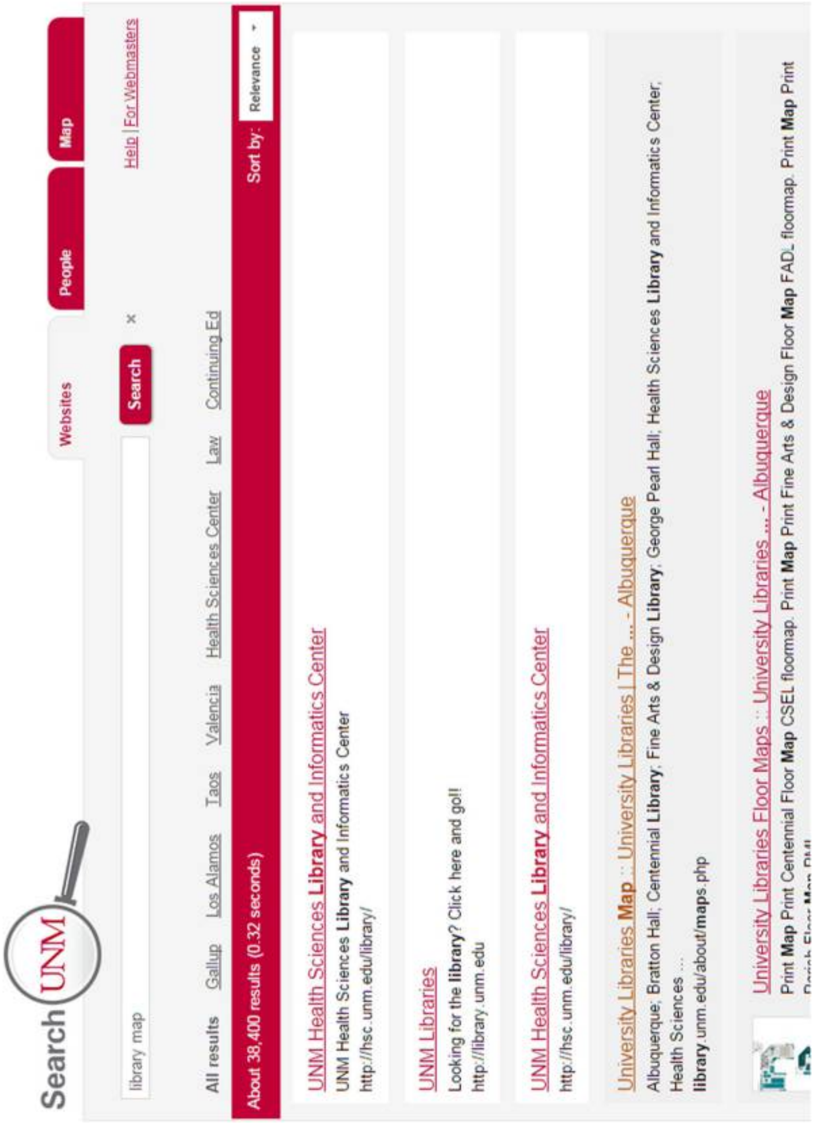


Figure 12.
The search result for query "library map" in SearchUNM

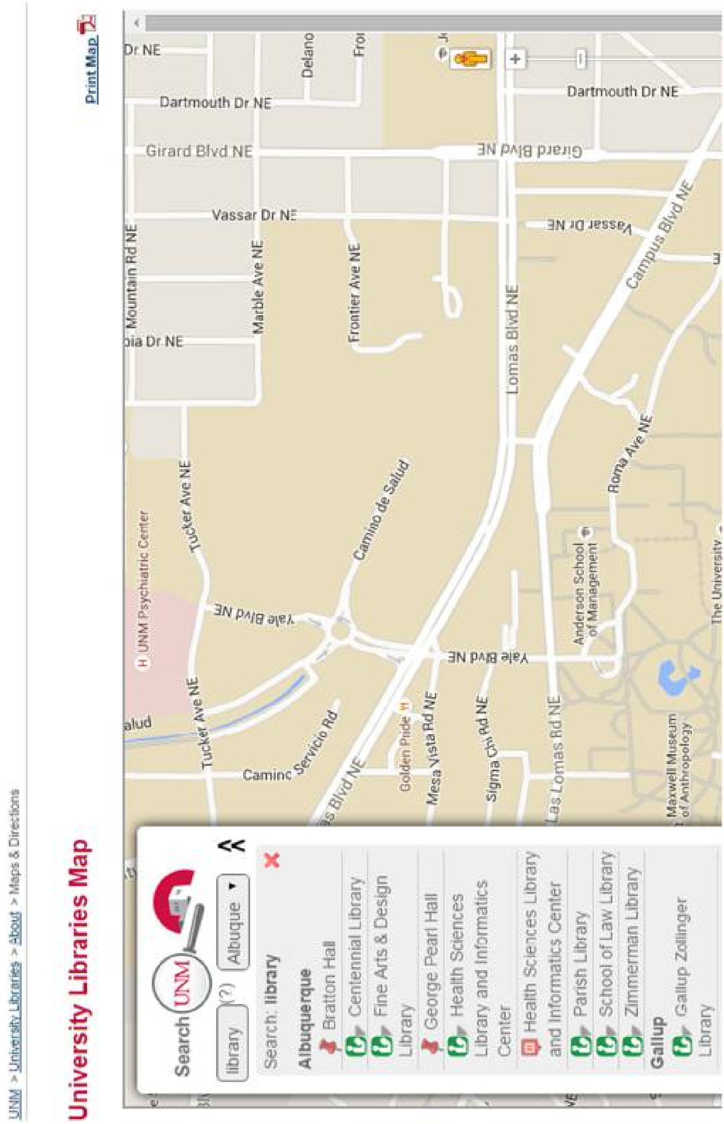


Figure 13.
The embedded iframe in the University Libraries Map webpage

(Katz and Byrne, 2003). If the menu or site structure does not fit into a user's information need, the cost of browsing will likely be higher than the cost of searching. On the other hand, a prominent interface component that captures a person's attention can change the user's mind as well.

For this project, it have been difficult to decide which category should be listed as its own layer without more data collection. Based on the query log, it appears that most users are hoping to locate classrooms or lecture halls, which currently are not an option in our browsing. However, it may not be necessary to offer such an option, since these users have specific queries in mind when visiting the page. Thus, it is more likely that searching is the preferred path for them to get information. The sensible approach is to use these categories for visitors or for people who do not have a specific location in mind – for example, emergency blue phones or bus stops. One concern is how many categories we should provide. Both humans and mobile devices have limited capacity for information processing. Efficiency in the decision-making process hinges on the information presented in a limited display space. A well-designed interface aims to achieve a balance of the information-to-interface ratio (Roth and Harrower, 2008; De Sabbata and Reichenbacher, 2012). As our interactive map attracts increasing attention from users and units on campus, we have received more requests/suggestions to create new browsing categories. It will be helpful to establish a policy in this regard.

In addition, a more eye-catching "hints" area is needed (see Figure 11) in order to communicate effectively with users. Roth and Harrower (2008) found that minimalist design of widgets on an interactive map often does not provide enough cues for users to infer their functions. As we were reluctant to add the variety of misspelled forms in our metadata out of concern that it might reinforce incorrect spellings, the "Do you mean?" function is a possible solution in the future. At the present, an effective "hints" panel is a straightforward approach to make known to users that autocomplete is in place and a one-word search is better than a phrase search.

Conclusions and future study

In this study, we tried to identify the characteristics of queries submitted to the SearchUNM Map page, the options and prioritization of the metadata augmentation, and the usefulness and possible improvement of our interface. Our main findings are as follows. First, the term "library" is searched most often in the SearchUNM Map search (In early 2015, we documented that "library" appeared in more than 2000 search instances, compared to the other popular queries that numbered in the low hundreds). However, classrooms and lecture halls combined make up the major proportion of the search instances. Many queries contain building names and room numbers, which may be a result of users' copying-and-pasting from the course catalog. Floor plans of buildings with room numbers are not in our map search scope, nor in our anticipated plans, considering that most buildings on campus have more than one floor, making floor-plan representations of these buildings difficult on the Google map. In addition to trimming off room numbers from queries, we have worked with other units to merge the building IDs in our system with the UNM XML Schedule (http://xmlschedule.unm.edu/docs/xml-schedule-usage-guideline.html). The developers and web masters can now generate URLs for locations on our map with the building IDs when transforming the XML file for web documents.

Second, from the query log, we know that more than just the official building name should be in the metadata. Acronyms, shorthands (e.g. "advisement center" for "University advisement and enrichment center"), common names (e.g. "gym" for "recreation center"), functional units (e.g. financial aid office), departments (e.g. foreign languages), and addresses need to be in the metadata, as these have all shown up in queries. However, not all queries are equally important. The search frequency of a query will be used as a guideline to decide what needs to be added to the metadata for a location.

Third, we need to explore a more effective design and mobile friendly layout for the interface. As we include more data and metadata to increase successful searches, we need to balance the ascending interface complexity with the efficiency of human information processing. What is the ideal number for the browsing categories? What will these categories be? And how do we take account of a user's context and provide the most relevant results? Search log analysis has given us a broad overview of common user behavior and the geo-info need for our map search. Next, we will need more in-depth research, such as a study of the user experience, to learn an individual user's context and their perception of, and satisfaction with the search.

From our study, it is evident that users have the information need to navigate inside buildings. With library and classrooms as the most searched queries, we can envision the need for getting real-time directions via mobile devices inside libraries as well as in buildings used mainly for classrooms. At a different scale, the same concept of GR can be applied: the geographical entity is well presented (data and metadata); a user's context is considered (indexing and processing); and the result is communicated effectively (interface design).

Compared to queries entered at a desktop, queries submitted by mobile phones are more likely to be geographically related (Mountain and MacFarlane, 2007). As mobile web access increases, it is foreseeable that an interactive map like SearchUNM Map will serve as a starting point for students to fulfill information needs, in addition to location needs. Based on the most common user scenario, the related information for a location should accompany its geo-info. In the case of determining libraries' information inside the pop-up windows, the two librarian members decided that along with a link to the website and directions, the "Ask a Librarian" phone numbers and the building hours were the most needed information for users. Thus, by looking up the location of a library, a student is able to find out more about the library, beyond its mere location. Likewise, each location found in the map can be the gateway for other, related, information.

Though this project was a collaboration between campus IT and the library, ideally the ongoing updating and augmenting of metadata needs to reside in the unit with the most accurate and up-to-date information about campus buildings. We hope that in the next stage, in addition to implementing what we have learned from this study, we can bring in discussion with additional partners, and engage in a genuine university-wide collaboration on this valuable project.

References

Abdelmoty, A.I., Smart, P.D., Jones, C.B., Fu, G. and Finch, D. (2005), "A critical evaluation of ontology languages for geographic information retrieval on the internet", *Journal of Visual Languages & Computing*, Vol. 6 No. 4, pp. 331-358.

Amitay, E., Har'El, N., Sivan, R. and Soffer, A. (2004), "Web-a-where: geotagging web content", *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM*, pp. 273-280.

Asunka, S., Chae, H.S., Hughes, B. and Natriello, G. (2009), "Understanding academic information seeking habits through analysis of web server log files: the case of the teachers college library website", *The Journal of Academic Librarianship*, Vol. 35 No. 1, pp. 33-45.

Bishop, B.W. (2011), “Location-based questions and local knowledge”, *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 8, pp. 1594-1603.

Chen, Y.-Y., Suel, T. and Markowetz, A. (2006), “Efficient query processing in geographic web search engines”, *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, ACM*, pp. 277-288.

Cheung, A.K.L. (2006), “Representational issues in interactive wayfinding systems: navigating the Auckland University Campus”, in Carswell, J.D. and Tezuka, T. (Eds), *Proceedings of the 6th International Conference on Web and Wireless Geographical Information Systems, Springer-Verlag, Hong Kong, December 4-5*, pp. 90-101.

Church, K., Neumann, J., Cherubini, M. and Oliver, N. (2010), “The Map Trap?: an evaluation of map versus text-based interfaces for location-based mobile search services”, *Proceedings of the 19th International Conference on World Wide Web, ACM*, pp. 261-270.

De Sabbata, S. and Reichenbacher, T. (2012), “Criteria of geographic relevance: an experimental study”, *International Journal of Geographical Information Science*, Vol. 26 No. 8, pp. 1495-1520.

Ding, J., Gravano, L. and Shivakumar, N. (2000), “Computing geographical scopes of web resources”, *Proceedings of the 26th VLDB Conference*, *Cairo*, *September 10-14*.

Fernandes, A.I., Goulão, M. and Rodrigues, A. (2013), “A comparison of maps Application Programming Interfaces”, available at: arXiv preprint arXiv:1305.3485 (accessed December 1, 2014).

Fu, G., Jones, C.B. and Abdelmoty, A. I. (2005), “Ontology-based spatial query expansion in information retrieval”, *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*, Springer Berlin Heidelberg, pp. 1466-1482.

Gan, Q., Attenberg, J., Markowetz, A. and Suel, T. (2008), “Analysis of geographic queries in a search engine log”, *Proceedings of the First International Workshop on Location and the Web, ACM*, pp. 49-56.

Gravano, L., Hatzivassiloglou, V. and Lichtenstein, R. (2003), “Categorizing web queries according to geographical locality”, *Proceedings of the Twelfth International Conference on Information and Knowledge Management, ACM*, pp. 325-333.

Jansen, B.J. (2006), “Search log analysis: what it is, what’s been done, how to do it”, *Library & Information Science Research*, Vol. 28 No. 3, pp. 407-432.

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 & Management*, Vol. 36 No. 2, pp. 207-227.

Jones, R., Zhang, W.V., Rey, B., Jhala, P. and Stipp, E. (2008), “Geographic intention and modification in web search”, *International Journal of Geographical Information Science*, Vol. 22 No. 3, pp. 229-246.

Katz, M.A. and Byrne, M.D. (2003), “Effects of scent and breadth on use of site-specific search on e-commerce web sites”, *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 10 No. 3, pp. 198-220.

Kobayashi, S., Fujioka, T., Tanaka, Y., Inoue, M., Niho, Y. and Miyoshi, A. (2010), “A geographical information system using the Google Map API for guidance to referral hospitals”, *Journal of Medical Systems*, Vol. 34 No. 6, pp. 1157-1160.

Kurth, M. (1993), “The limits and limitations of transaction log analysis”, *Library Hi Tech*, Vol. 11 No. 2, pp. 98-104.

Lee, R., Shiina, H., Tezuka, T., Yokota, Y., Takakura, H., Kwon, Y.J. and Kambayashi, Y. (2005), “Map-based range query processing for geographic web search systems”, *Digital Cities III. Information Technologies for Social Capital: Cross-Cultural Perspectives*, Springer Berlin Heidelberg, pp. 274-283.

Mountain, D. and Macfarlane, A. (2007), "Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals", *Journal of Information Science*, Vol. 33 No. 5, pp. 515-530.

Pan, B., Crotts, J.C. and Muller, B. (2007), "Developing web-based tourist information tools using google map", in Sigala, M., Mich, L. and Murphy, J. (Eds), *Information and Communication Technologies in Tourism 2007*, Springer, Vienna, pp. 503-512.

Peters, T.A. (1993), "The history and development of transaction log analysis", *Library Hi Tech*, Vol. 11 No. 2, pp. 41-66.

Raper, J. (2007), "Geographic relevance", *Journal of Documentation*, Vol. 63 No. 6, pp. 836-852.

Roth, R.E. and Harrower, M. (2008), "Addressing map interface usability: learning from the Lakeshore Nature Preserve interactive map", *Cartographic Perspectives*, Vol. 60, pp. 46-66.

Sanderson, M. and Kohler, J. (2004), "Analyzing geographic queries", paper presented at the SIGIR Workshop on Geographic Information Retrieval, available at: www.geo.uzh.ch/~rsp/gir/abstracts/sanderson.pdf (accessed December 1, 2014).

Silva, M.J., Martins, B., Chaves, M., Afonso, A.P. and Cardoso, N. (2006), "Adding geographic scopes to web resources", *Computers, Environment and Urban Systems*, Vol. 30 No. 4, pp. 378-399.

Spink, A., Jansen, B.J., Wolfram, D. and Saracevic, T. (2002), "From e-sex to e-commerce: web search changes", *Computer*, Vol. 35 No. 3, pp. 107-109.

Taghavi, M., Patel, A., Schmidt, N., Wills, C. and Tew, Y. (2012), "An analysis of web proxy logs with query distribution pattern approach for search engines", *Computer Standards & Interfaces*, Vol. 34 No. 1, pp. 162-170.

Vaid, S., Jones, C.B., Joho, H. and Sanderson, M. (2005), "Spatio-textual indexing for geographical search on the web", *Advances in Spatial and Temporal Databases*, Springer Berlin Heidelberg, pp. 218-235.

Wang, A., Jun, Z. and Jiang, W. (2009), "Useful resources integration based on google maps", *Proceedings of 2009 4th International Conference on Computer Science & Education, ICCSE*, pp. 1044-1047.

Wang, C., Xie, X., Wang, L., Lu, Y. and Ma, W.-Y. (2005a), "Web resource geographic location classification and detection", *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, ACM*, pp. 1138-1139.

Wang, L., Wang, C., Xie, X., Forman, J., Lu, Y., Ma, W.-Y. and Li, Y. (2005b), "Detecting dominant locations from search queries", *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM*, pp. 424-431.

Wang, P., Berry, M.W. and Yang, Y. (2003), "Mining longitudinal web queries: trends and patterns", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 8, pp. 743-758.

Wu, Y.-J., Wang, Y. and Qian, D. (2007), "A Google-map-based arterial traffic information system", *Intelligent Transportation Systems Conference, ITSC 2007, IEEE*, pp. 968-973.

Xiao, X., Luo, Q., Li, Z., Xie, X. and Ma, W.Y. (2010), "A large-scale study on map search logs", *ACM Transactions on the Web*, Vol. 4 No. 3, pp. 1-33.

Zhou, Y., Xie, X., Wang, C., Gong, Y. and Ma, W.-Y. (2005), "Hybrid index structures for location-based web search", *Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM*, pp. 155-162.

Further reading

Reichenbacher, T., Crease, P. and De Sabbata, S. (2009), "The concept of geographic relevance", *Proceedings of the 6th International Symposium on LBS & TeleCartography*, *Nottingham*, *September*.

Appendix

Search frequency	Number of unique queries Succeed	Fail	Total search instances Succeed	Fail
240	1	0	240	0
67	1	0	67	0
65	1	0	65	0
59	1	0	59	0
58	1	0	58	0
54	1	0	54	0
53	1	0	53	0
49	1	0	49	0
47	1	0	47	0
46	1	0	46	0
43	1	0	43	0
38	1	1	38	38
33	2	0	66	0
32	1	0	32	0
31	1	0	31	0
30	2	0	60	0
29	2	0	58	0
28	1	0	28	0
27	3	0	81	0
26	2	0	52	0
25	2	0	50	0
23	2	0	46	0
22	2	0	44	0
21	7	0	147	0
20	3	0	60	0
19	2	0	38	0
18	3	0	54	0
17	4	0	68	0
16	6	0	96	0
15	1	0	15	0
14	3	0	42	0
13	7	2	91	26
12	6	0	72	0
11	9	0	99	0
10	4	1	40	10
9	10	1	90	9
8	16	2	128	16
7	19	4	133	28
6	17	5	102	30
5	25	15	125	75
4	34	27	136	108
3	71	64	213	192
2	90	174	180	348
1	275	1,417	275	1,417

Table AI. The descending order of the search frequencies and their associated unique query counts and the outcomes of searches (successful or failed)

About the author

Zoe Chao is the User Experience Librarian at the Penn State University. Her research interests include user experience, design, data analytics, and information architecture. She has an MLIS from the University of Illinois at Urbana-Champaign, and an MS in fisheries and wildlife from the Michigan State University. Zoe Chao can be contacted at: szc19@psu.edu