

Webology, Volume 4, Number 4, December, 2007

Home	Table of Contents	Titles & Subject Index	Authors Index
----------------------	-----------------------------------	--	-------------------------------

Location-Based Search Engines Tasks and Capabilities: A Comparative Study

[Saeid Asadi](#)

School of Information Technology & Electrical Engineering, the University of Queensland, Brisbane, Australia, E-mail: asadi (at) itee.uq.edu.au

[Xiaofang Zhou](#)

School of Information Technology & Electrical Engineering, the University of Queensland, Brisbane, Australia, E-mail: zxf (at) itee.uq.edu.au

[Hamid R. Jamali](#)

School of Library, Archive and Information Studies, University College London, UK, E-mail: h.jamali (at) gmail.com

[Hossein Vakili Mofrad](#)

Dept. of Medical Library and Information Sciences, Medical University of Hamedan, Hamedan, Iran, E-mail: vakili (at) umsha.ac.ir

Received November 10, 2007; Accepted December 27, 2007

Abstract

Location-based web searching is one of the popular tasks expected from the search engines. A location-based query consists of a topic and a reference location. Unlike general web search, in location-based search it is expected to find and rank documents which are not only related to the query topic but also geographically related to the location which the query is associated with. There are several issues for developing effective geographic search engines and so far, no global location-based search engine has been reported. Location ambiguity, lack of geographic information on web pages, language-based and country-dependent addressing styles, and multiple locations related to a single web resource are notable difficulties. Search engine companies have started to develop and offer location-based services. However, they are still geographically limited and have not become as successful and popular as general search engines. This paper reviews the architecture and tasks of location-based search engines and compares the capabilities, functionalities and coverage of the current geographic search engines with a user-oriented approach.

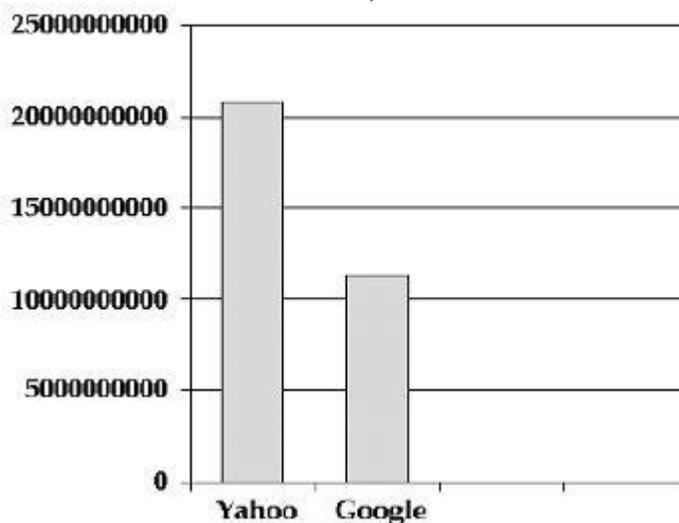
Keywords

Location-based search; Web search; Geographic search engines

Introduction

Since the development of the World Wide Web in early 1990s, search engines have played the main role in locating and searching the resources on the Web. On October 2007, 31 billion searches were conducted on Google alone, more than one billion queries each day ([Lipsman, 2007](#)). Figure 1 shows the number of web pages indexed by Google and Yahoo search engines. In 2005, Yahoo claimed that its index covered more than 20 billion web resources, the largest search engine ([Terdiman, 2005](#)). It is believed that the actual size of the Web is at least several times bigger than what search engines currently cover.

Figure 1. The number of web pages indexed by Yahoo and Google in 2005 ([Terdiman, 2005](#)).



The earlier generation of search engines inherited the techniques and rules from traditional information retrieval systems. They gathered web pages, made textual indexes and tried to retrieve relevant pages according to a query. Soon it was found out that traditional models are not capable of handling web searches properly. The World Wide Web turned to become a different information system because of the dynamic gigantic scale of the Web, the heterogeneous nature of the resources on the Web, and hypertext links among web pages.

Specialized Search Engines

Many techniques and tools were later developed to improve the web search. Major search companies used these techniques which resulted in development of globally popular search engines such as Google and Yahoo. Although general search engines have had many successes in performing effective and efficient search, they have not been able to cover all needs expected by their users. The variety in language, medium type, dynamic contents, hidden information, and various users' needs are the main difficulties for search engines. To overcome these problems, specialized search engines have been established to capture and search different types of resources e.g. videos and sound files. Furthermore, some search engines have been dedicated to specific languages, topics and interests. For example, [VIVISIMO](#) can cluster the search results in different generated categories according to the semantic relationship among the web pages.

Location-Based Web Search

One of the main specializations on web search is considering the location of web resources when it is intended by users. In geographic or location-based web search, the results must not only be related to the topic of a query but also they must be geographically related to a location which is associated with a query.

Location-based search engines must associate locations to web resources in order to answer location-based queries. A location-based query is a query which asks for a product, business or service in a particular area. Unlike general queries which often reflect a topic, location-based queries consist of two parts: a *topic* and a *reference location*. For example, the query $q=\{\text{"restaurants in Brisbane"}\}$ associates a location $L=\{\text{"Brisbane"}\}$ with a subject or service name $s=\{\text{"restaurants"}\}$.

Previous work such as [Sanderson and Kohler \(2004\)](#), [Spink and Jansen \(2004\)](#), and [Asadi et al. \(2005\)](#) have revealed that a significant portion of the queries on general search engines can be considered as location-based queries. General search engines are not able to handle location-based search properly. Although location-based queries follow several patterns which can make them distinguishable for the search engines ([Asadi et al., 2005](#)), the search engines treat them in the same way that they do the general queries. Consequently, the search results for location-based queries are often poor with low precision. Geographic search engines instead are able to consider both dimensions i.e. the topic and the location of the queries and web resources.

Location-based web search is still in its infancy. In addition to general problems that all search engines are faced with, geographic search engines in particular face specific difficulties and problems among them the followings are notable:

- To distinguish and disambiguate different locations, search engines must employ huge geographic gazetteers. Providing and updating such databases is often expensive; and such data is not available for all countries.
- Location names are often ambiguous. This ambiguity can be geo/geo i.e. a geographic name might refer to two or more location names. For example, Singapore can be a country as well as a city name. Geo/non-geo ambiguities are also common in which a geographic name refers to other things like human names. Washington, Paris and London are instances of the names which refer to people and geographic locations.
- Only the web pages indexed in commercial lists such as Yellow Pages are included in the current location-based search engines that are a small proportion of the web resources. Many non-profit or personal pages contain valuable geographic information which makes them eligible for location-based search. When an ambiguous geographic name is submitted, it is a challenging and hard task for geographic search engines to guess the user's intention.
- Web pages can have multiple locations. Extracting, disambiguating and assigning all relevant locations to web resources is a difficult task and current techniques are not completely reliable.
- Current indexing models on general search engines are not able to consider a right weight for location names in a page and treat them similar to other terms. This often leads to loss of the weight of geographic names in competition with non-geographical words. Geographic search engines must develop geospatial indexes to consider all of the corresponding locations for the stored web pages ([Martins, Silva & Andrade, 2005](#)).

In this paper, we review tasks and architecture of location-based search engines and then compare several prototypes of the geographic search engines from a user's point of view.

Related Work

General Web Search

Wandex was the first Web robot and search engine established in 1993 ([Wall, 2004](#)). Early search engines such as AltaVista and Excite followed the traditional information retrieval

techniques. They retrieved results from their indexed repositories and ranked them based on keyword matching and proximity. However, the gigantic size of the Web, different types of media and the link structure of web pages soon revealed that the traditional techniques were insufficient for web search.

The hypertext link structure of the Web was a reach source to improve the quality of the search results. [Brin and Page](#) (1998) introduced PageRank algorithm for finding high quality web pages. It was assumed that in a particular topic, the important web pages are cited more than other pages. Therefore, the quality of a web page can be measured through the number of back links to that page as well as the average quality of the citing pages. Other works on Web link structure lead to interesting models e.g. HITS ([Kleinberg](#), 1999) in which both inlinks and outlinks are used to find virtual communities of web resources or people in a topic. Link analysis algorithms have improved Web information retrieval through connectivity-based results ranking, crawling priority, web page reputation computing, and searching-by-example ([Henzinger](#), 2000).

Since the employment of link analysis algorithm, more comprehensive and effective search engines have appeared. Using PageRank and some other novel techniques, Google could capture a gigantic amount of web pages, more than 24 million pages in the beginning ([Brin & Page](#), 1998). Today, major search engines are able to collect, index and search billions of web pages. In 2005, Yahoo claimed more than 20 billion indexed web resources ([Terdiman](#), 2005).

While the number of indexed web pages is a quantitative measurement for comparison of search engines, the quality of search depends on many factors. Search engines are known for having high recall and low precision. Though they have a reliable coverage over the web resources which lead to a high recall, they face serious problems in presenting the results effectively. Web queries are often short, 2.4 words in average ([Spink & Jansen](#), 2004). As a result, search engines return many results for a particular query. Users are often frustrated by navigation of search engine results ([Sullivan](#), 2000).

Location-Based Web Search

The World Wide Web made it possible to have access to information resources globally without geographical limits and borders. However, beside this global accessibility, many web sites and web-based services have become available on the internet, which often are designed for local users rather than the global users. Several academic as well as commercial projects have been reported on location-based web search. [Google Maps](#) and [Yahoo Local](#) are two examples of commercial location-aware search engines. These engines use a keyword-based as well as a map-based interface to find commercial information related to a specific address or location. Literally, they can cover the entire world: however, because the geographic gazetteers and digital maps are not available for all countries they are often limited to few countries e.g. USA, Canada, UK and Australia. Another problem is that they often rely on well-structured commercial databases such as Yellow Pages to search the queries. This is different from real web search in which web pages are searched.

Different studies such as [Asadi et al.](#) (2005) and [Sanderson & Kohler](#) (2004) show that at least one out of five queries on general search engines have geospatial dimensions and can be regarded as location-based queries. [Gravano, Hatzivassiloglou and Lichtenstein](#) (2003) divided the queries into global and local groups based on the best results from the search engines. If most of the related results for a query refer to a specific location the query is local; otherwise, it is a global query. [Asadi et al.](#) (2005) described nine common patterns for location-based queries. The essential part of each pattern is a topic which is normally

location-dependent. For example, a query such as "cheap hotels" is more likely to refer to a specific city, country etc. even if no geographic name is explicitly mentioned in the query.

Geo-tagging or assigning geographic information to web resources is a basic task for geographic search engines as it is necessary to know the locations each page refers to when a geographic query is searched. Basically, gazetteers have been used in most of the studies such as [Markowetz et al.](#) (2005), [Uryupina](#) (2003) and [Wang et al.](#) (2005) in order to extract geographic names and addresses from web page contents, control them with the gazetteer entries, and add this controlled geographic information to web pages. Web-a-Where ([Amitay et al.](#), 2004) is a geo-tagging system that uses a geographic database of location names and abbreviations and tags the web pages in two steps: *Spotting*, in which a web page is compared with the gazetteer to match as many names as possible; and *Disambiguation* of locations by using the gazetteer information as well as lexical rules and patterns. GeoSearcher geo-tagger ([Watters & Amoudi](#), 2003) assigns geographic coordinate (longitude and latitude) to web resources in two steps: *Geo-parsing*, or analyzing a web page to match geographic feature names; and *Geo-coding*, or disambiguation and assigning the coordinates to a page.

Besides content locations which can be extracted from the web page content, other resources have been used to assign locations to web resources. Geographic scope ([Ding, Gravano & Shivakumar](#), 2000) is the geographic area that most of the back links on a page come from or refer to. Target location ([Asadi et al.](#), 2006) instead, considers the location of the visitors or the people who use a web resource. Both Geographic Scope and Geographic Target are content-independent. As a result, they can be applied to textual and non-textual web resources. However, accessing such data and calculation of the scope or the target is often hard for search engines.

More details and experiments on geographic web search can be found in [Chen et al.](#) (2006), [Vaid et al.](#) (2005) and [Olga](#) (2002). In summary, the past works on location-based web search have mostly focused on extraction of geographic features and geo-tagging web resources with proper locations. Our focus is much more on the interface and output of the geographic search engines.

Geographic Search Engines

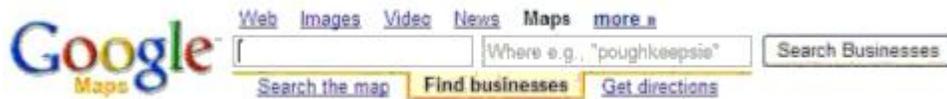
Interface

General search engines often have a text box where users can put their queries and a button to perform the *search*. The query interface of location-based search engines is different and it often consists of two text boxes: A text box for entering the topic and a second text box for entering an address, zip code or location name. Figure 2 compares the interface of Google and Google Maps. Most of the location-based search engines have a query interface similar to Google Map.

Figure 2. Query interface difference in Google (a) and in Google Maps (b)



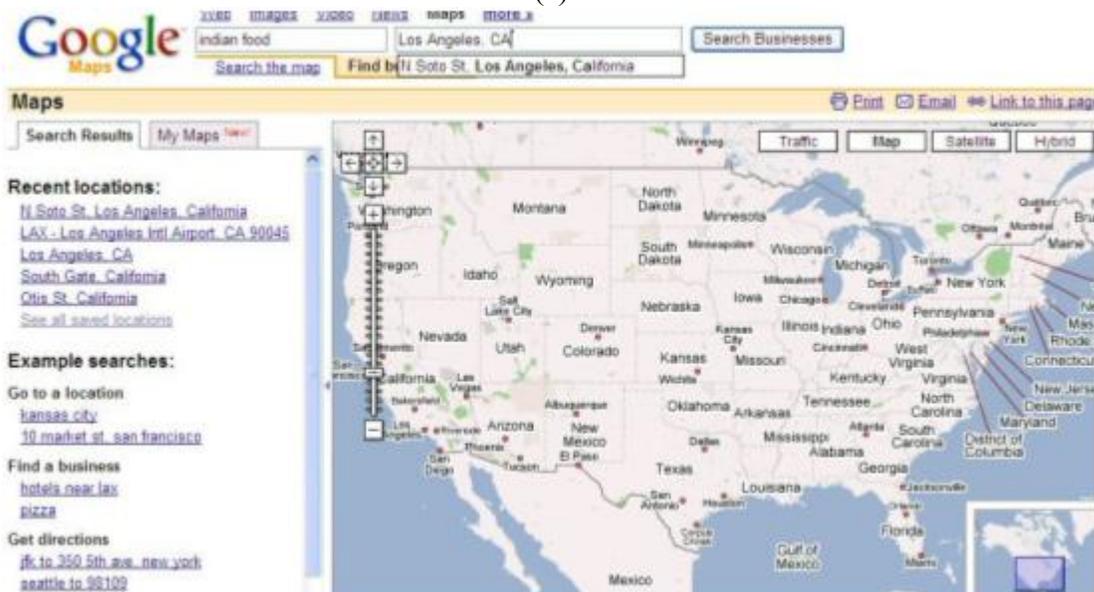
(b)



In addition to the traditional text-based querying, location-based search engines often provide a more user-friendly query interface which uses graphical maps. Interactive maps have been successfully added to web search engines and users can simply navigate their locations. Existing digital maps can be joined with databases - including WWW - to provide a user-friendly approach for searching location-based information. Figure 3 shows the interface of Google Maps which combines a text-based and a map-based query input. The first subfigure shows the interface before querying. A user can first locate a specific location on the map and then input the name of a service or business which he is looking for in that region. The next subfigure shows the interface after returning the results for the submitted query. For more details of the application of digital maps on web search see ([McCurlley, 2001](#)).

Figure 3. The interface of Google Maps: (a) Before sending a query; and (b) After performing the query.

(a)



(b)

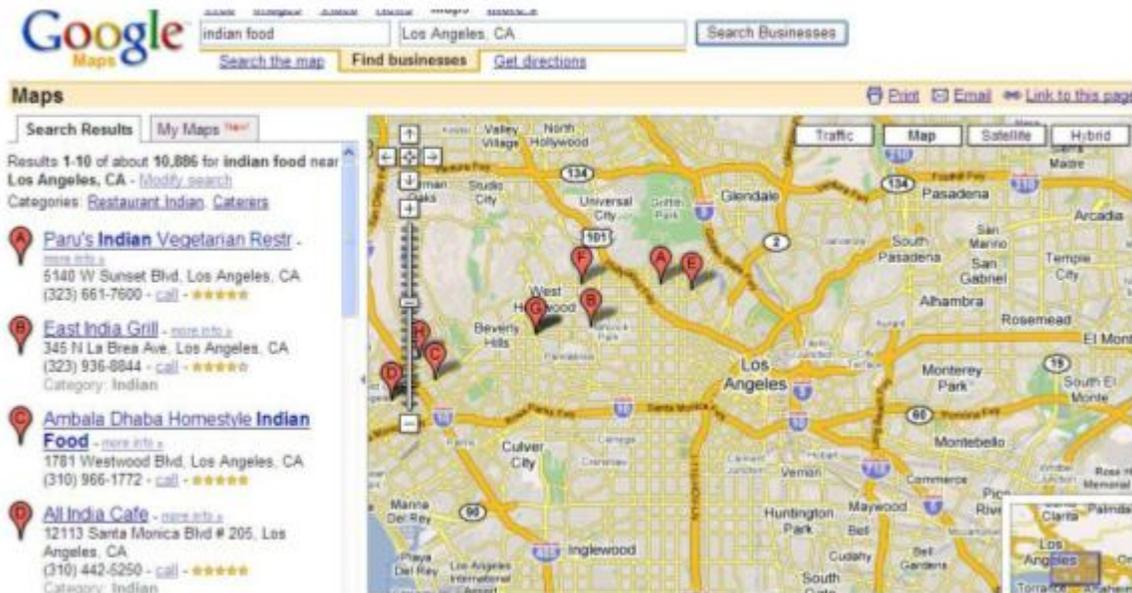
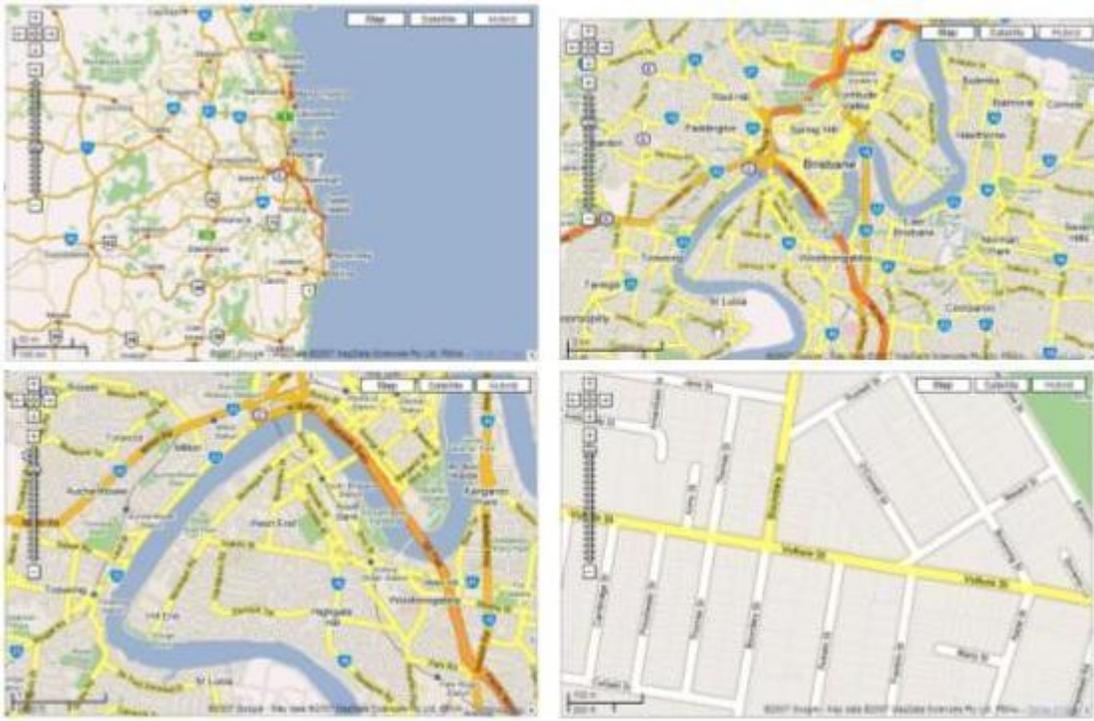


Figure 4 shows how a digital map can be used to navigate through different locations. The granularity of digital maps varies for different countries. This is mainly because of availability of data in some countries. In the best conditions, digital maps can be used from a global level to a street level view based wherever the maps and geographic data are available. Most of the digital maps have been created from or accompanied by satellite photos of the earth. As a result, a user can select a map, satellite or hybrid view for better navigation.

Figure 4. An example of navigation and granularity on digital maps from global view to street level view





Digital Maps and Geographic Search

Joining textual information to digital maps has already become possible and popular by using tools and languages such as GML. Geography Mark-up Language or GML is an XML-based language for expressing geographical features. GML is a modeling language for geographic information systems which acts as an interchange grammar for transfer of information on the Internet. GML lets Internet connected computers to access geographic information, for example, user's locations and traffic conditions. GML sentences are XML-based and they can be embedded in HTML as well. Once documents in a database are marked up with geographical information using GML or similar languages, they can be used to answer geographical queries. For example, the related documents which refer to a particular address can be shown on a digital map. This technique is used by search engines.

A digital map can be used in geographic search engines in three ways:

1. Querying. By locating a specific town, street, region and so on a map, the *reference location* of a location-based query can be determined by the user. The *topic* of the query can be written in a text box or chosen from a category.
2. Result Presentation. The results for a location-based query can be presented on a digital map to ease the decision for a user.
3. Query Modification. If there are not enough results on a specific location, the user can modify a location-based query's reference point by zooming out or changing the current location and navigating for a new location. If many results are shown on the current window, again the query could be modified by zooming in or changing the current location.

Figure 5 shows the result of a location-based query on a map. The first subfigure presents the results on a large scale map, i.e. in a city level view. This view is useful if there are not enough results on a smaller area. However, users often zoom in and look for a more detailed map in a suburb or street level (the second subfigure). If the results on the current window are not satisfying or if the user is interested in another location, he can navigate to another location in the same level (the third subfigure).

Figure 5. Presentation and navigation of results on a map from a large scale view (a) to a smaller scale view (b) and navigating to a new window in the same level (c).

(a)



(b)



(c)



Geographical Retrieval and Ranking

Search engines started their work since the emergence of the Web with imitating the techniques and tools used in traditional information retrieval. Keyword searching is the basic method used in many well-known search engines in which a user sends one or more terms as a query. The search engine task is finding and showing relevant and useful web pages according to the query terms. The numerous documents on the Web as well as the

short length of queries makes the search engines retrieve many results for a typical query. Search engines are known for their high recall and low precision. The success of any web search tool depends on its ability to show the best results and avoid unwanted results. This goal has been achievable through weighing web pages. For a particular query, a retrieval score S is calculated and the n best results are shown based on their S . Retrieval score is dynamic and might change whenever the search engine's collection is updated. Although the formula and algorithm for calculating the retrieval score is different on every search engine and it is often hidden to the public, the retrieval score S can be equal to the sum of different weights considered by the search engine:

$$S = \sum W$$

It was discussed in the introduction and the related work sections that general search engines retrieve relevant web pages for a query and rank the more popular or important pages higher. As a result, the rank score S often consists of two major sub-scores i.e. $W_{(relevance)}$ and $W_{(importance)}$.

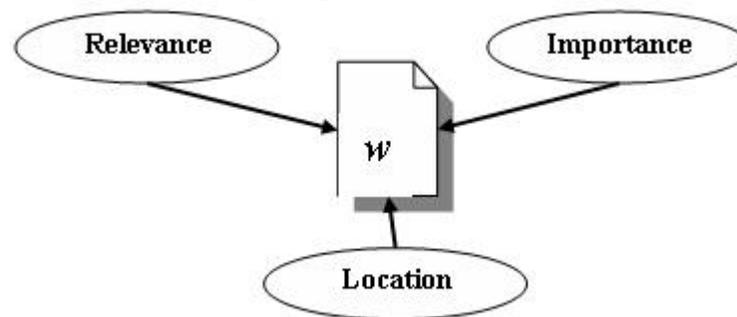
$$S = W_{(relevance)} + W_{(importance)}$$

In many cases the location(s) associated with web pages are significant for users. With sending a location-based query, a user intends to find good quality results not only relevant to a topic but also related to a particular area. For example, in the case of the query '*cheap restaurant in Brisbane*' it is expected that a search engine find the pages which are not only relevant and popular but also geographically related to *Brisbane*. It could be claimed that in location-based search three dimensions are important and must be considered by search tool:

1. Topic relevancy
2. Web page importance or link popularity
3. The location(s) associated with web pages.

Any location-aware search engine must be able to cover these dimensions of web pages appropriately (Figure 6).

Figure 6. Different weighting dimensions in location-based search



The first two dimensions namely *relevance* and *importance* have been addressed by general search engines. As a result, the third dimension or calculation of web page's location is the specific task of any geographic search engine. The location of a web page could be considered as a new score to modify the retrieval score:

$$S = W_{(relevance)} + W_{(importance)} + W_{(location)}$$

Comparative Study of the Geographic Search Engines

We have run a survey on several geographic search engines to understand their advantages and shortcomings through measurable characteristics. The actual algorithms and criteria that they use are confidential and not available to the public. Eleven location-based search engines have been studied from the user's side. Geographically, these engines are mostly related to USA, UK, Canada and Australia.

The Criteria for Comparison

The selected location-based search engines have been studied and compared regarding three different criteria: Interface, search performance and limitations.

Search Interface. Search Interface is a critical and important part of location-based search engines as it should be able to facilitate querying, navigation among the results and query refinement. We have used seven criteria to evaluate the search interface of the studied search engines:

Tex: The interface has a text box for entering textual query.

Add: The interface has a separate text box for entering the location or address i.e. the input boxes of the topic and address are separate.

St: Search engine supports street level address search. Some search engines only can search city names.

Zip: Search engine supports Zip code search.

Map: An interactive digital map is provided to support querying.

Pers: Personalized search and customized querying is supported.

Exac: Exact/Expand search is supported. A user can ask for searching only in the mentioned location or letting the search engine to include surrounding areas.

Search Performance. The actual algorithm, tools, techniques and data collections which are used by search engines are not often secret. However, from the user's side, the criteria used in searching and ranking can be found out more and less. The compared criteria for search tasks that we have considered are as following:

Relevance Ranking: This is a basic task for general search engines and also it is expected that geographic search engines also be able to distinguish relevant documents to a particular query (topic) and avoid irrelevant results.

Distance Ranking: A geographic search engine must be able to rank the results based on their distance from a reference point mentioned in a query.

Categorizing: If there are many results, it is expected that the similar results be clustered or grouped in different categories.

Rating: Users can rate each business and search engine can re-rank the results based on the user's feedback.

Limitations and Shortcomings. Current location-based search engines are often limited to a particular location e.g. a country. They also usually use the geographic information acquired from the commercial databases rather than from the World Wide Web. In addition to the above-mentioned measures for comparing the search interface and the search performance, we have considered other specifications and abilities to compare the limits and shortcomings of the search engines. The five measures used for comparison are as follows:

Geo Cov. : Refers to geographic coverage of a search engine.

Database: Indicates that the search engine uses structured databases to answer a geographic query.

WWW: Indicates that the search engine uses web resources to answer a query.

Map: The results are presented on a map.

Map Q. Refine: User can refine a query by navigating on the map.

Results and Discussion

Search Interface. Google, Yahoo, Microsoft and other major search engines have developed local services and map-based search tools. Most of the existing search engines have almost a similar interface for querying and presentation of the search results. Most of the location-aware search engines such as Yahoo Local, [Local](#) and [InfoSpace](#) have a text-based query interface and user must enter an address or a location name. As it was already shown in Figure 3, Google Maps has an extra interactive map in the start page and a user can first navigate on a desired location and then only enter the topic. This is a more user-friendly way of search as many users may not remember a specific address.

Table 1 compares the studied search engines' interfaces for different tasks and capabilities. All of the studied geographic search engines have a textual input as a basic way to get the query similar to general search engines. Almost all of them have a separate box to enter the address. In this way, the topic and the location are entered separately and this makes it easier for the search engines to find more accurate results. Most of the studied search engines support zip code search instead of entering an address. Zip codes are more accurate and easier to match with digital maps. The search engines often support personalized search i.e. the users can initially enter their desired locations in their profiles and then the search engine retrieves the results based on the specified address. Street level search is not supported in all of these engines. As a result, the granularity of the search scope is often city level or at least suburb level. Map-based querying is still not supported by many search engines. The users need to write down the address which often must be very accurate otherwise the search engine can not detect the correct location. Finally, some of these geographic search engines let the users to determine whether they are looking for the results exactly in the mentioned location or the surrounding areas can also be included. This geographic query expansion is often useful if the reference location is street or suburb level.

Table 1. A comparison between search interface capabilities of different geographic search engines

Search Engine	Tex.	Add.	St.	Zip	Map	Pers.	Exac.
Google Maps	√	√	√	√	√	√	-
Yahoo Local	√	√	√	√	√	√	-
MSN Live	√	√	√	√	√	√	-
Local.com	√	√	√	√	-	√	-
InfoSpace	√	√	-	-	-	√	-
Ask City	√	√	√	√	√	√	-
Yell	√	√	√	√	-	√	-
Nine MSN	√	√	-	√	√	√	√
Yahoo!7	√	√	-	√	√	√	√
True Local	√	√	-	√	√	-	√
Search.net.au	√	-	-	-	-	-	√

Search criteria and capabilities. In this section we go in detail how probably each search engine works and what is their coverage and scope. The performing search algorithm varies on different search engines and it is always hidden from the public. There are few papers about the details of search algorithms on search engines. However, here we review some specifications which can help compare different search engines' work.

A big difference between general search engines and location-based search engines is that the first group search the web pages collected from the entire World Wide Web; while location-based search engines search on a smaller collection which could be referred as business data collection. This means that geographic search engines in fact do not search the Web and they only search a separate collection of business information which is not obtained from the Web. This business information often is obtained from business information collections such as Yellow Pages.

Differences between the WWW and business information databases can be discussed as following:

1. A commercial service such as Yellow Pages often contains structured data which are manually controlled and inserted in pre-defined fields; while the World Wide Web is known as a massive unstructured collection of information. Retrieving information from structured databases is easier than the Web and database search and web search follow different rules.
2. The WWW is a gigantic-sized collection of heterogeneous information resources. Business collections are often limited to very smaller collections of similar documents.
3. The World Wide Web has a global scope and covers all web resources regardless of their location. Commercial collections like Yellow Pages are often limited to a city or even smaller area.
4. According to the structured nature of business information collections which often have explicit data, using their geographic information for web search is much easier compared to web pages. Web resources often have vague geographic information or do not come with any address at all.

The above comparison shows that existing geographic search engines in fact do not search the Web and they only search a database of paid entries.

Table 2 shows different tasks of search and result presentation on the examined engines. Relevance-based ranking is the essential criterion in all search engines including the location-based search engines. This means that the retrieved pages should be relevant to the query topic. For the query *Sydney Hotels* the results must be relevant to the topic i.e. hotel; otherwise, they are irrelevant even though they are related to Sydney. Distance ranking is a basic task for location-based search engines. A location-aware retrieval system must be able to rank the results based on their distance from the reference location. In the previous example, the hotels in Sydney must be ranked before the hotels in the other cities around Sydney. Most of the studied search engines can support distance-based ranking of the search results. Several geographic search engines are able to categorize the results in different groups according to their locations. For example, if there are many results for the query *Sydney Hotels* then the search engines might cluster or group them based on the suburbs they are located in. Rating is a non-geographical scale used by some location-based search engines. Rating is often a measurement for comparing different local services e.g. coffee shops or hotels in an area. Because of the fact that location-based web search is closely related to local services, rating can be an advantage to attract more users.

Table 2. A comparison between search engines performing tasks and result presentation

Search Engine	Relevance Ranking	Distance Ranking	Categorizing	Rating
Google Maps	√	√	√	√
Yahoo Local	√	√	√	√
MSN Live	√	√	√	-
Local.com	√	√	√	√

InfoSpace	√	√	-	√
Ask City	√	√	√	√
Yell	√	√	√	-
Nine MSN	√	√	-	-
Yahoo!7	√	√	-	-
True Local	√	√	-	√
Search.net.au	√	-	-	-

Limitations and shortcomings. Location-based web search relies on digital maps and geographic gazetteers. These geographic information resources are not equally available for all countries. As it is shown in Table 3, the geographic coverage of the location-based search engines is often limited to more developed countries. Developing and updating comprehensive geographic gazetteers and digital maps is not economic for many less developed countries as they still do not have enough infrastructure and applications for these digital facilities.

As it was mentioned before, the ambiguous nature of the location names as well as the difficulties in extracting and assigning proper locations to web resources, are the main obstacles for search engines to cover the resource on the World Wide Web for handling geographic queries. Table 3 also shows that all of the studied search engines rely on human-developed databases such as Yellow Pages rather than web pages on the WWW to answer the geographic queries. Currently, this is one of the basic issues for location-based search engines. Table 3 also shows that not all of the studied search engines are able to present the results on a map even though they have a map-based interface for entering a location. The last column shows that the reference location of a query can be revised in some geographic search engines with a map-based query input.

Table 3. A comparison between search engines coverage and result presentation

Search Engine	Geo Cov	Database	WWW	Map	Map Q. Refine
Google Maps	US, Uk, Au, Cn	√	-	√	√
Yahoo Local	US, Cn	√	-	√	√
MSN Live	US, Uk, Cn, Jp, Au	√	-	-	√
Local.com	US	√	-	√	-
InfoSpace	US	√	-	√	-
Ask City	US	√	-	√	-
Yell	UK	√	-	-	-
Nine MSN	Au	√	-	-	√
Yahoo!7	Au	√	-	-	√
True Local	Au	√	-	√	-
Search.net.au	Au	√	-	-	-

Conclusion

Location-based search is becoming popular for search engines and their users since many local services have developed their web sites and web-based services. So far, the geographic search engines have not been successful compared to the general search engines because of the difficulties and ambiguities they face with. In this paper, we reviewed some of the previous studies on location-based web search. The tasks and basics of the location-based search engine were discussed and then we run a comparative study on the existing location-based search engines to analyze their abilities and shortcomings

with a user's point of view. In summary, current location-based search engines are limited in three ways:

1. Geographic Limitation. Most of the current geographic search engines are only limited to USA and Canada. Some of them have a wider coverage including UK, Australia, Japan, Taiwan and few more countries. Other search engines which have not been mentioned here are also often limited to a city or country or even to a specific language.
2. Data Limitation. Existing location-based engines only cover commercially collected information of local businesses e.g. those mentioned in Yellow Pages. They do not cover World Wide Web although they might have link to those pages.
3. Performance Limitation. An ideal geographic search and presentation has not been supported by the mentioned search engines and they can not match interactive maps with textual geographic data properly. For example, map-based query refinement is not guaranteed.

More research is needed to facilitate the extraction and assignment of addresses and locations to web resources. This geographic information is not necessarily mentioned in the web page content. As a result, more sophisticated techniques and algorithms are needed to analyze and process the web resources to make them usable in location-based search engines.

References

- Amitay, E. et al. (2004). Web-a-where: Geotagging web content. *SIGIR'04*, Sheffield, UK: 273-280.
- Asadi, S. et al. (2005). Searching the World Wide Web for local services and facilities: A review on the patterns of location-based queries. *WAIM'05*, Hong Zhou, China.
- Asadi, S. et al. (2006). Calculation of target locations for web resources. *WISE'06*, pages 277-288.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107-117.
- Chen, Y. et al. (2006). Efficient query processing in geographic web search engines. In *SIGMOD'06*, pp. 277-288.
- Ding, J., Gravano, L., & Shivakumar, N. (2000). Computing geographical scopes of web resources. *VLDB'00*, Cairo, Egypt.
- Gravano, L., Hatzivassiloglou, V., Lichtenstein, R. (2003). Categorizing web queries according to geographical locality. *CIKM'03*, New Orleans, Louisiana, USA.
- Henzinger, M. (2000). Link analysis in web information retrieval. *IEEE Data Engineering Bulletin*, 23(3) Sep. 2000:3-8.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *JASM*, vol. 46.
- Lipsman, A. (2007). [61 billion searches conducted worldwide in August](#). *ComScore: Measuring the Digital World*, October 10, 2007. Retrieved December 2, 2007, from <http://www.comscore.com/press/release.asp?press=1802>
- Markowetz, A. et al. (2005). Design and implementation of a geographic search engine. *WebDB'05*, Baltimore, Maryland, USA.
- Martins, B., Silva, M. J. & Andrade, L. (2005). Indexing and ranking in Geo-IR systems. *GIR'05*, Bremen, Germany.
- McCurley, K. S. (2001). Geospatial mapping and navigation of the web. *WWW'10*, Hong Kong: 221-228.
- Olga, O. (2002). Extracting geographical knowledge from the internet. Proceedings of *ICDM-AM International Workshop on Active Mining ACDM-AM*, Maebashi,

- Japan.
- Sanderson, M. and Kohler, J. (2004). Analyzing geographic queries. In *Proceedings of Workshop on Geographic Information Retrieval SIGIR'04*, Sheffield, UK.
 - Spink, A., & Jansen, B.J. (2004). [A study of web search trends](#). *Webology*, 1(2), Article 4. Retrieved December 2, 2007, from <http://www.webology.org/2004/v1n2/a4.html>
 - Sullivan, D. (2000). [Survey reveals search habits](#). The Search Engine Report. *Search Engine Watch*. Retrieved December 2, 2007, from <http://www.searchenginewatch.com/sereport/00/06-realnames.html>
 - Terdiman, D. (2005). Yahoo passes Google in search index capacity. *News.com*, August 8, 2005.
 - Uryupina, O. (2003). Semi-supervised learning of geographical gazetteers from the internet. In *Proceedings of the HLT-NAACL Workshop on the Analysis of Geographic References*, Edmonton, Canada.
 - Vaid, S. et al. (2005). Spatio-textual indexing for geographical search on the Web. *9th International Symposium on Spatial and Temporal Databases*. Angra dos Reis, Brazil.
 - Wall, A. (2004). [History of search engines & web history](#). Retrieved December 2, 2007, from <http://www.searchmarketing.info/search-engine-history/>
 - Wang, C. et al. (2005). Detecting geographic locations from web resources. In *GIR '05: Proceedings of the 2005 workshop on Geographic information retrieval*, pages 17-24, New York, NY, USA.
 - Watters, C. and Amoudi, G. (2003). GeoSearcher: Location-based ranking of search engine results. *JASIST*, 54(2), 140-151.
-

Bibliographic information of this paper for citing:

Asadi, Saeid, Zhou, Xiaofang, Jamali, Hamid R., & Vakili Mofrad, Hossein (2007). "Location-Based Search Engines Tasks and Capabilities: A Comparative Study" *Webology*, 4(4), Article 48. Available at: <http://www.webology.org/2007/v4n4/a48.html>

Alert us when: [New articles cite this article](#)

Copyright © 2007, Saeid Asadi, Xiaofang Zhou, Hamid R. Jamali, and Hossein Vakili Mofrad.