



## Web-based geographic search engine for location-aware search in Singapore

Flora S. Tsai\*

School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

### ARTICLE INFO

#### Keywords:

Web search  
Geographic search  
Location-aware  
Geographic Web  
Singapore  
Local search

### ABSTRACT

The proliferation of advanced, online, three-dimensional mapping programs provides Web users with detailed aerial and satellite maps, high-powered graphics and animation, and customization capabilities. This mapping revolution can transform the landscape of the Web into a geographic Web, where information can be organized geographically, so that users can understand information in the proper context. In this paper, we describe a geographic information retrieval system that is able to search for location-specific information in Singapore-based Web sites. The user is able to view their search locations on a satellite map instead of the two-dimensional maps currently used in street directories. The Web-based search engine is able to search for locations based on area names, building names, and groups of landmark types, business names, and business categories. Furthermore, the user is also able to use their current coordinates as a parameter so that the search engine is able to return results in order of the distance from the user's current location.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

Search engines are information retrieval systems designed to help find information stored on a computer system such as on the Web, inside corporate or proprietary networks, social networks (Tsai, Han, Xu, & Chua, 2009), or blogs (Chen, Tsai, & Chan, 2008). The search engine allows the user to locate content meeting specific criteria (usually based on a given word or phrase) and retrieves a list of items matching the given criteria. This list is often sorted with respect to some measure of relevance of the results. Search engines use regularly updated indexes to operate quickly and efficiently.

Despite the popularity of search engines, there have not been many comprehensive location-based search engines that enable the user to search for locations around their area regardless of which part of the world they are from. Location-based search engines (sometimes referred to as local search) are the most prevalent in the United States where abundant sources of information are available. Examples of commercial local search engines include Google (Google Earth, 2009, <http://earth.google.com>; Google Local, 2009, <http://local.google.com>), Yahoo (Yahoo! Local Maps, <http://maps.yahoo.com>), and Microsoft (Microsoft, <http://maps.live.com>). These systems are based on a map-and-hyperlink architecture, where the Web content and geographic information are weakly bonded, such that the Web pages only refer to geographic locations in their content (Tezuka, Kurashima, & Tanaka, 2006). Other local search engines that are based on the map-and-hyperlink structure

include geospatial mapping and navigation of the Web (McCurley, 2001), SPIRIT spatial search engine (Jones, Abdelmoty, Finch, Fu, & Vaid, 2004), and GeoSphereSearch (Graupmann & Schenkel, 2006).

Many of these search engines, besides being based on a map-and-hyperlink structure, are primarily targeted at US locations. Within other parts of the world, there are currently not many viable location-based search engines available.

In this paper, we describe a Web-based local search system which returns locations by parameters that the user enters and allows the user to enter coordinates to list out the results in order of the distance from the user. Although the system is currently limited to Singapore-based Web sites, the user is able to view their search locations on a Google Earth (Google Earth, 2009, <http://earth.google.com>) satellite map instead of the two-dimensional maps currently used in street directories. Haversine's algorithm is used as a method of determining the distance between an entered coordinate and each of the location's coordinates. The Web-based search engine is able to search for locations based on area names, building names and drop-down menus of landmark types, business names and business categories. Furthermore, the user is also able to use their current coordinates as a parameter so that the search engine is able to return results in order of the distance from the user's current location.

### 2. Review of related work

#### 2.1. Google earth

Google Earth (<http://earth.google.com>) is a virtual globe program initially called Earth Viewer and was created by Keyhole

\* Tel.: +65 6790 6369; fax: +65 6793 3318.  
E-mail address: [fst1@columbia.edu](mailto:fst1@columbia.edu)

Inc. Keyhole Inc. was later acquired by Google in 2004 and the product was renamed Google Earth. The satellite images found in Google Earth was obtained by the superimposition of images obtained through the use of satellite imagery, aerial photography and Geographical Information System (GIS, 2008, <http://www.gis.com/whatisgis/index.cfm>) or GIS over a three-dimensional globe. Images on Google Earth are maintained at more than two years, to maintain privacy, and less than three years, to maintain relevance. We choose Google Earth over other 3D satellite mapping programs because it currently has the best resolution of locations and buildings within Singapore.

Google Earth is basically a three-dimensional mapping tool that provides the user with the ability to freely rotate or scale the model by changing the viewing angle and position of the user. Google Earth follows the visually accurate representation of the Earth rather than simplified graphical depiction of a globe.

Google Earth introduced a whole new idea in location-based search by enabling the user to search for locations and present the user with a satellite view of their current location. Although Google Earth comes with its own search engine that allows the user to search for countries over the world, locations in the countries are not comprehensive. In particular, Singapore and much of Asia currently have few locations marked out under the basic version of Google Earth. Google Earth also differs in the resolution of the pictures received. For example, Las Vegas as well as many of the major US cities have much higher resolution than most of Asian cities such as Singapore.

Google Earth Plus adds Global Positioning System (GPS) integration by reading the track and waypoints from a GPS device and adding them into temporary places on the Google Earth program. These temporary places can be saved and more information can be added to the locations that are relevant to the coordinates of the waypoints downloaded.

Google Earth allows the user to enable street map overlays and three-dimensional building models over the basic satellite image of the globe through the use of Keyhole Markup Language (KML). Google Earth's allowance for layers to be added onto it such that locations are better marked out also makes an allowance for the user to enable or disable the layer. Google Earth also has an easy to use interface for the user to create their own place marks, or mark a location with information on the particular location, in kml or kmz format. These place marks are also exportable and allows for a user to spread information about a particular place in the Google Earth platform.

## 2.2. Keyhole Markup Language

Keyhole Markup Language (KML) is a file format used to display geographic data in an earth browser such as Google Earth, Google Maps and Google Maps for mobile. KML uses a tag-based structure with nested elements and attributes and is based on the XML standard (Google Earth, 2009, <http://earth.google.com>).

A KML file is processed in much the same way that HTML (and XML) files are processed by Web browsers. KML's tag-based structure with names and attributes used for specific display purposes follows that of HTML. Google Earth therefore acts as a browser for KML files. KML files contain specifications for a set of features, such as place marks or 3D models, for display in Google Earth. Each place has a longitude and latitude and optional data such as tilt, heading and altitude.

KML can be used to specify icons and labels to identify locations, create different camera positions to define unique views, use image overlays attached to the ground or screen, write HTML descriptions of features, including hyperlinks and embedded images, dynamically fetch and update KML files from remote or lo-

cal network locations, and fetch KML data based on changes in the 3D viewer.

Inside Google Earth, KML layers can be added in using the Graphical User Interface (GUI) itself in the form of placemarks, paths, polygons and image overlays using the options on the toolbar itself, taking away the need to write KML code in the form of long coding. A form based method of entering information also ensures that users need not know the exact coding in order to enter KML layers and increases the ease of entering layers.

## 2.3. Street directory

Streetdirectory.com (<http://www.streetdirectory.com>) is an online Web map Service founded by Virtual Map in 2000, a company with its headquarters located in Singapore. Initially it used licensed data from Singapore Land Authority under a non-exclusive agreement until July 2004. On top of providing maps for personal use, it also offers a variety of ad-supported services. The Web site has since expanded to include maps in Malaysia and Indonesia.

Streetdirectory.com recently tied up with Google Earth so that in searching for a location, the user is able to bring up a satellite image of the current location as an alternative to a two-dimensional map that it currently provides. Currently, users to streetdirectory.com are able to search for a location, attain its longitude and latitude, view the location on a two-dimensional map or a satellite image, attain directions to travel from a specific address to another address, and get a list of amenities or locations around the specified area. Being Singapore-based, streetdirectory.com does bring out very sufficient and updated information about any particular location in the whole of Singapore.

However, streetdirectory.com only shows the basic satellite image of the location without the amenities or locations around the location being searched being shown on the satellite image, unlike in the two-dimensional maps that they show. The locations are in this instance listed externally rather than on the map itself. In this case, having a Google Earth KML file would be a better reference if the user has Google Earth installed.

The satellite images on streetdirectory.com are useful for clients not using the Web site on a computer not equipped with Google Earth; however they are limited. Certain areas of Singapore are not visible in satellite image form on streetdirectory.com, most like being that they are of poor resolution.

## 2.4. Global Positioning System

Developed by the United States Department of Defence, this system was officially named NAVSTAR GPS (NAVigation Satellite Timing And Ranging Global Positioning System). Since then, with its name shortened to GPS, it has been an indispensable aid to navigation around the world and is an important part of map-making and land surveying. Although the cost of maintaining the system, containing more than 24 satellites, is approximately US\$400 million per annum, including the replacement of ageing satellites, GPS is free for civilian use as a public good.

The basis of GPS is the mathematical method of "triangulation" from its position to the satellites. A GPS receiver measures the distance using the travel time of radio signals and has in-built knowledge of the confidential exact location of the various satellites. By using all the parameters it acquires and has built-in, the receiver calculates the latitude and longitude of the current location. However, for the receiver to be able to have contact with the satellite, the receiver has to be used under a clear sky. Rainy conditions or after a rain has fallen would mean that the satellite would be unable to connect to the GPS device.

In order to determine the location of the user, we use the GPS device called Garmin 60, produced by a manufacturer that holds

a large share of the market together with Magellan. Both these manufacturers are also the only ones supported by Google Earth. Google Earth allows download of tracks and waypoints from GPS devices manufactured by these two companies and creates KML files for the waypoints and tracks downloaded (Google Earth, 2009, <http://earth.google.com>).

### 3. Design and algorithms

#### 3.1. Database design

The table used is named *Business\_Address* and has the columns shown in Table 1.

For the database, one table was used to combine all the results into one area for better linkage and viewing. Of all the columns, *BusinessID* is the unique identification number that is auto-incremental for each new entry. Each ID has unique *BusinessName* and *BusinessAreaAddress* combination. The *Business Address* is broken down into:

- *BusinessAreaAddress*, where the address is stored without the postal code;
- *BusinessAreaPostal*, where the postal code is stored;
- *Area*, where the keywords of the area are stored, e.g. Causeway Point;
- *General Area*, where the General Area of the location is stored, e.g. Yishun.

The *BusinessPhone* stores the Business Phone of the company in that outlet. The *Area* is further broken down by a listing of its surrounding landmarks under *LandmarkType*. The *BusinessName* is categorised under *BusinessCat* which stores the Category of the Business that it provides.

From the *Address*, the location is found in Google Earth and a KML file is produced into a folder in the server. The folder address on the Web server is saved under *Folder* and the filename is saved under *Filename*. A complete file destination address is stored under *IP* as backup. From the *Address*, the coordinates are acquired through the Web in decimal degree format and the values are inserted into latitude and longitude. When a user enters in their own coordinates in degree decimal. The values are entered into the columns *Current\_latitude* and *Current\_longitude*. At the same time the user coordinates are entered into the database, the distance between the user coordinates and each of the coordinates of the locations in the database are used to calculate the distance between the locations and the user's current coordinates, the distance calculated is saved under the *Distance* column.

**Table 1**  
Database schema.

Field	Type
<i>BusinessNameID</i>	int
<i>BusinessName</i>	text
<i>Area</i>	text
<i>BusinessAreaAddress</i>	text
<i>BusinessAreaPostal</i>	text
<i>BusinessPhone</i>	text
<i>BusinessCat</i>	text
<i>IP</i>	text
<i>Filename</i>	text
<i>Folder</i>	text
<i>Current_latitude</i>	float
<i>Longitude</i>	float
<i>Latitude</i>	float
<i>Current_longitude</i>	float
<i>Distance</i>	float
<i>Deeplink</i>	text
<i>GeneralArea</i>	text

Finally, from the locations searched on [streetdirectory.com](http://streetdirectory.com)'s Web site. The deeplinks to each of the locations were obtained and inserted into the table in HTML format under the column *Deeplink*.

#### 3.2. Calculation algorithms

There are two methods for calculating the distance from two coordinates, one being the more accurate but with a longer processing time of Vincenty's Theorem and the other being the faster Haversine's Formula. In this project, we use the Haversine's Formula (Sinnott, 1984) for faster processing.

The Haversine's Formula is an equation important in navigation, giving great-circle distances between two points on a sphere by using the decimal degree form of their longitude and latitude coordinates. It is a special case of a more general formula in spherical trigonometry, the law of haversines, relating the sides and angles of spherical "triangles".

For two points on a sphere (of radius  $R$ ) with latitudes  $\phi_1$  and  $\phi_2$ , latitude separation  $\Delta\phi = \phi_1 - \phi_2$ , and longitude separation  $\Delta\lambda$ , where angles are in radians, and the distance  $d$  between the two points (along a great circle of the sphere) is related to their locations by the formula:

$$h = \text{hav}(\Delta\phi) + \cos(\phi_1) \cos(\phi_2) \text{hav}(\Delta\lambda) \quad (1)$$

Let  $h$  denote  $\text{hav}(\sin(d/R))$ , given from above.  $d$  can then be solved either by simply applying the inverse haversine (if available) or by using the arcsine (inverse sine) function:

$$d = (R) \text{hav}^{-1}(h) = (2R) \arcsin(\sqrt{h}) \quad (2)$$

This formula is only an approximation when applied to Earth as Earth is not a perfect sphere, its radius  $R$  varies from 6356.78 km at the poles to 6367.45 km at the equator. The error is therefore 0.1% depending on the location due to this slight ellipticity. (Assuming that the geometric mean of  $R = 6367.45$  km is used.)

### 4. Results and discussion

The database from which these results are taken contain 1652 entries with the following categories: Apparel, Bank, Cinema, Department Store, Duty Free Shop, Electronics, F&B (food and beverage), Fast Food, Food Court, Furniture, Health and Beauty, Minimart, Musical Instruments, Restaurant, Snack Bar, Sports, Stationery, Seafood, and Supermarket.

The landmark type searched for are Building, Road, MRT stations, Schools and Shopping Centres. General Area searched under Advanced have various roads grouped into one big area, e.g. Tanjong Katong and Haig Road are both grouped under the Katong area.

#### 4.1. Simple search

The initial page is the main page which uses only HTML as it is a form page for entering the area keywords to search for locations. There are links created to other search options, Business Category links, Business Name links, and Landmark links. These four pages are the pages from which the simple search starts from and have similar layouts and coding. The only exception being the section in bold, signifying which page the user is in and which method the user is searching.

The results page of the simple search function comes with a header field that enables the user to either search a new query under the default "Area" section or provides links to the same pages under the main page. The search results are obtained through PHP coding. The method of searching is first obtained by setting four IF

sections to determine which method the user is searching for, by determining which of the POST fields out of the four differently named fields have a value. If the user chose to search through Area keywords but has not entered any queries, an error message will appear on the next page requesting the user to go back and enter a query.

When a valid query is entered, the POST field or the query is passed into a MySQL query statement that searches for results that contain the query passed through in the column searched. The results are then displayed by Name, KML file link, Address, Phone, Category and the deeplink on the right of the Result. A screenshot of the result screen is shown in Fig. 1.

When the user clicks on the link “Google Earth KML File”, the KML file is downloaded and the user can open it on their desktop to view the location in Google Earth. The Street Directory deeplinks on the right of each location will bring the user directly to the location map of that location found in streetdirectory.com.

The top of the page shows the search query that the user entered to get to this page. Moreover, the default value of the area search at the top of the page also shows the search query the user entered for this set of results.

#### 4.2. Advanced search

When the user submits the page using the search button, the advanced search page (Fig. 2) is loaded.

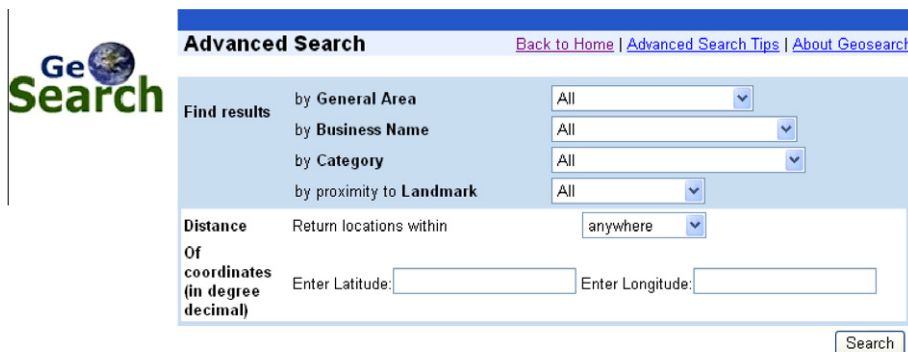
The first four parameters are the original parameters used by the simple search. In the advanced search, the user can choose to limit the search by one or all four of the parameters. All four parameters are displayed in the following drop-down boxes:

- The first is by the general area in which the location searched is situated in. e.g. Jurong, City Hall.
- The second is by Business Name, the distinct business names in the database queried using a MySQL statement and displayed in the drop-down menu.
- The third is by Category, the distinct business categories queried by an MySQL statement and the results displayed in a drop-down box.
- The fourth is by a location’s proximity to a Landmark; the type of Landmark is chosen via the drop-down box.

If one or more of the four parameters in the first section is filled in, the results would be limited by one or more of the limits using



Fig. 1. Search results for Cathay.



Click [here](#) for a coordinate calculator)

Fig. 2. Advanced search interface.

### Latitude / Longitude Coordinate Converter

To precisely locate points on the earth's surface, degrees longitude and latitude have been divided into minutes (') and seconds ("). There are 60 minutes in each degree. Each minute is divided into 60 seconds. Seconds can be further divided into tenths, hundredths, or even thousandths. For example, the U.S. Capitol is located at 38°53'23"N , 77°00'27"W (38 degrees, 53 minutes, and 23 seconds north of the equator and 77 degrees, no minutes and 27 seconds west of the meridian passing through Greenwich, England).

The following calculator will convert Lat / Long coordinates from the format of Degree Minutes Seconds to the format of Decimal Degrees. Assume that we start with a coordinate in (Deg Min Sec). This would be expressed as: (DD MM SS) or (39 54 32)

Starting with  
**Degrees Minutes Seconds** - DD MM SS (Example = 39 54 32 W).

We convert to:  
**Decimal Degrees** (Example = 39.909deg) - The minutes and seconds (54 32) are converted to decimal format (.909).

For the calculator below, negative values of Longitude are not necessary, Colorado, for instance can be expressed as [39deg, -108deg] or [39 N, 108 W]. The calculators assume positive numbers.

Enter Degree:

Enter Minute:

Enter Second:

Fig. 3. Coordinate calculator page.

an AND term. Therefore, only results fulfilling all the parameters entered would be acquired and displayed.

The section below these four parameters hold a drop-down menu for selecting the distance from a current coordinate which can be entered in the fields for latitude and longitude. At the bottom is a link to a coordinate calculator where the user is given an explanation of coordinates and the different versions. The user can then convert their coordinates from degrees minutes seconds to decimal degrees. The coordinate calculator page is seen in Fig. 3.

When the user further enters their current location coordinates in decimal degrees format in the latitude and longitude field, the data is entered into the database and the distance between the user's current location and the locations inside the database is calculated and entered into the database. The results are then displayed in a sorted format with the smallest distance being at the top of the list. The Radius drop-down box further allows the user to limit the radius with which the results are returned.

## 5. Conclusions and recommendations

In this paper, we describe a Web-based geographic search engine which returns locations by parameters that the user enters and allows the user to enter coordinates to list out the results in order of the distance from the user. Although the system is currently limited to Singapore-based Web sites, the user is able to view their search locations on a satellite map instead of the two-dimensional maps currently used in street directories. The Web-based search engine is able to search for locations based on area names, building names and drop-down menus of landmark types, business names and business categories. Furthermore, the user is also able to use their current coordinates as a parameter so that the search engine is able to return results in order of the distance from the user's current location.

A database of 1652 Singapore-based locations has been produced with deeplinks and associated KML file to every location

available for download in the server. The coordinates of every location in latitude and longitude format have also been entered. All of this information have been acquired from the Internet and Google Earth itself. Users are able to use the simple search to search by area keywords, landmark types, business categories and business names. The advanced search enables the user to restrict the query by up to four restrictions and further allows the order of the results to be returned with regards to proximity to the user's current location.

Using the Haversine's Formula, the user can enter in any coordinates and the results will subsequently be returned in order of the closest proximity to the coordinates entered.

### 5.1. Recommendations

A possible upgrade to the Web page is for a Geocoding API to be added to grab the coordinates of the location when the user enters in their addresses. However if a method of finding out the coordinates through an external program and entering it into the database is found then it could help the site out in reducing the manual entry of the coordinates into the database. Currently each of the entries are searched using Street Directory and entered into the database manually.

Because the database currently stands at 1652 entries, further work could be done on this database by either adding to the entries already in the complete database or finding a method with which the database could come from another source and updating the database from an internet source. Further additions to the database could be the inclusion of informational links or information on the places themselves inside the KML files, like photos, videos, blog articles or historical articles. More 3D models could also be created for buildings in areas where there are few models.

With future work, further research could be performed to pave the way for a mapping revolution in which Web pages are tied to geographical coordinates, allowing for more relevant

location-aware searches, thus transforming the physical as well as the virtual world.

### Acknowledgment

The author thanks Lianna Tan for her contributions.

### References

- Chen, Y., Tsai, F. S., & Chan, K. L. (2008). Machine learning techniques for business blog search and mining. *Expert Systems with Applications*, 35(3), 581–590.
- Graupmann, J., & Schenkel, R. (2006). GeoSphereSearch: ContextAware Geographic Web Search. In *Proceedings of the 2006 workshop on geographic information retrieval (GIR 2006)*, Seattle, WA, USA, August 2006.
- Jones, C. B., Abdelmoty, A. I., Finch, D., Fu, G., & Vaid, S. (2004). The spirit spatial search engine: Architecture, ontologies and spatial indexing. In *Proceedings of the third international conference geographic information science (GIScience 2004)*, Adelphi, MD, USA, October 2004.
- McCurley, K. S. (2001). Geospatial mapping and navigation of the web. In *Proceedings of the 10th international conference on world wide web (WWW 2001)*, HongKong, China, May 2001.
- Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, 68(2), 159.
- Tezuka, T., Kurashima, T., & Tanaka, K. (2006). Toward tighter integration of web search with a geographic information system. In *Proceedings of the 15th international conference on world wide web (WWW 2006)*, Edinburgh, Scotland, May 2006.
- Tsai, F. S., Han, W., Xu, J., & Chua, H. C. (2009). Design and development of a mobile peer-to-peer social networking application. *Expert Systems with Applications*, 36(8), 11077–11087.