MOPSI Location-based Search Engine: Concept, Architecture and Prototype

Pasi Fränti
Speech and Image Processing Unit
University of Eastern Finland
PO Box 111, FI-80101 Joensuu
+358132517931
franti@cs.joensuu.fi

Juha Kuittinen
Speech and Image Processing Unit
University of Eastern Finland
PO Box 111, FI-80101 Joensuu
+358505817639
jkuitti@cs.joensuu.fi

Andrei Tabarcea
Faculty of Automatic Control and Computer Engineering
Technical University of Iaşi
Dimitrie Mangeron Blvd., nr. 53A, 700050, Iaşi
+40744531425
tabarcea@tuiasi.ro

Lomanzi Sakala
Speech and Image Processing Unit
University of Eastern Finland
PO Box 111, FI-80101 Joensuu
+358449517980
lsakala@cs.joensuu.fi

ABSTRACT

Traditional *location-based services* use databases where all entries have been explicitly georeferenced beforehand. We propose an alternative approach based on web search and using ad-hoc georeferencing of web-pages. We denote it as *location-based search engine* and emphasize its seemingly small but significant distinction from traditional location-based services. We outline how to construct such search engine and prove its effectiveness using a prototype called MOPSI search.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval.

General Terms

Design, Experimentation, Human Factors

Keywords

Search engine, LBS, mobile device, location information, WWW.

1. INTRODUCTION

Current search engines fail to utilize one important aspect of the relevance: the location of the user because of two main reasons. Firstly, the location of the user was not as widely available as nowadays due to GPS phones being less frequent. Secondly, the information in web is rarely attached with the location for which it would be relevant [5].

In existing systems, known as *location-based services* (LBS), the location of data must be explicitly coded in the database. These are usually databases such as hotel directors, restaurants lists or similar common services, e.g. *Google Maps* [2] and *Yellow Pages* [1]. Their main drawback is that the data must be collected beforehand, which makes the systems depend on the services providers and their efforts to keep the information up to date.

SAC'10, March 22-26, 2010, Sierre, Switzerland.

We introduce an alternative solution referred to as *location-based search engine* based on the ideas in [3] with the following practical solutions. For a given location (e.g. from GPS), we perform location-restricted web query, analyze the web-pages found (relevant by keyword), extract potential address information and compare them to the entries in a gazetteer. Positive results are presented according to their distance relative to the user location, plotting the target location on map or giving navigational information to the location.

Figure 1: The web interface of the MOPSI search engine. Available on-line: http://cs.joensuu.fi/mopsi/

We have constructed a working prototype called MOPSI search, and demonstrate its efficiency against existing LBS systems. The prototype is available as web page (Figure 1) and also as Java application suitable for S60 mobile phones (Figure 2).

2. COMPONENTS OF THE MOPSI SEARCH ENGINE

MOPSI search engine consists of the following:

1. User interface for web and mobile devices.
2. Core server software.
3. Geocoded street ↔ name database with spatial indexing.

2.1 Overall Structure

The mobile interface is based on Java ME and is compatible with most smart-phones, using HTTP protocol to communicate with the MOPSI search engine. The coordinates and keyword(s) are sent to the server and the search results are displayed as an ordered list according to their distance to the user or on a map. Functionalities for user tracking and data collection also exist.

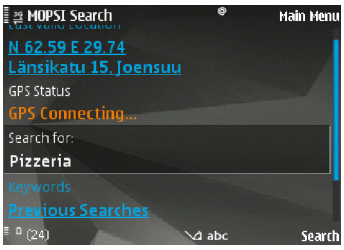

Figure 2: Mobile interface (Java for Symbian S60).

The web interface is designed using PHP with AJAX capabilities and it is used for accessing the search engine from a computer. The location can be input by typing the street address or by pointing location on map. The search results are shown both as a list and on Google Maps, optionally with a route to the location.

2.2 Core Server Software

Our approach is to detect address portions from web-pages. We use an external search engine for query-based search and then post-process the search results by extracting street or postal addresses, which are translated into coordinates using *geocoded* street-name database. The core server-side software performs these operations and is the main part of our location-based search engine. Its architecture is outlined in Figure 3.

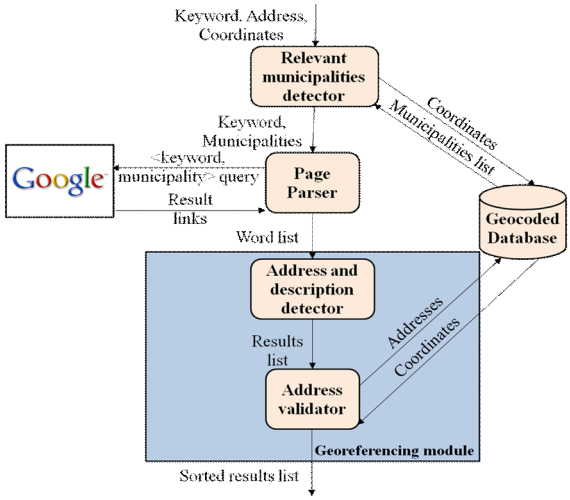

Figure 3: Core server software architecture.

The *Relevant municipalities detector* finds municipalities within a predefined range (e.g. 10 km) with the use of the *geocoded* database. The *Page parser* uses an external search engine to perform a <keyword, municipality> query for every municipality in that range, downloads the resulted web-pages and extracts a word list from the text by stripping the html tags. The *Address and description detector* searches for address blocks in the word list. The *Address validator* verifies the addresses from the result list using *geocoded* database, retrieving geographical coordinates. The server-side database is constructed beforehand from all addresses and coordinates of the target region. We use MySql with spatial extensions, whereas other options were be PostgreSQL with PostGIS and Oracle Spatial.

3. EXPERIMENTS IN FINLAND

We tested the search using 10 randomly selected urban and 10 rural municipalities in Finland. The search was performed using 5 keywords for commercial services (hotel, restaurant, pizzeria, cinema, car repair) and 5 non-commercial services (hospital, museum, police station, swimming hall, church). The results were compared against two traditional location-based services: *Google Maps* [2] and *Finnish Yellow Pages* [1].

The results (Table 1) show that MOPSI provides more results than Google Maps or Yellow Pages. The results were then evaluated by users rating them as (1) *relevant*, (2) *somewhat relevant* or (3) *not relevant*. We measured the mean relevance using the methodology in [4]. In the case of rural municipalities, MOPSI results are more relevant than Google Maps results in case of commercial services (1.71-vs-1.33) but less relevant in case of non-commercial services (2.35-vs-2.48). In case of urban municipalities, the differences were insignificant (2.20-vs-2.17 and 1.46-vs-1.41).

Table 1. Number of results for our test scenario

Query type	MOPSI prototype	Google Maps	Yellow Pages
Rural non-commercial	69	29	0
Rural commercial	245	92	189
Urban non-commercial	148	413	37
Urban commercial	1412	813	1337
Total number of results	2352	1405	1597
Overall mean relevancy	1,59	1,66	1,28
Overall std. deviation	0,84	0,89	0,54
Overall std. error	0,02	0,02	0,01

4. CONCLUSIONS

Concept of location-based search engine was outlined and its design issues were discussed. A prototype implementation in Finland was demonstrated with qualitative comparison using typical search examples that show its potential for practical applications.

5. REFERENCES

[1] Finnish Yellow Pages. http://en.02.fi/yellow+pages/

[2] Google Maps. http:/maps.google.com/

[3] Hariharan G., Fränti P., Mehta S. Data Mining for Personal Navigation. SPIE Conf. on Data Mining and Knowledge Discovery, Orlando, Florida, April 2002.

[4] Jansen, B. J., Molina, P. R. The effectiveness of Web search engines for retrieving relevant ecommerce links. Information Processing and Management, Vol.42: 1075–1098. 2006.

[5] McCurley, K. S. Geographic mapping and navigation of the web. 10th WWW Conference, Hong Kong, May 2001.