OPENGEOCODING.ORG – A FREE, PARTICIPATORY, COMMUNITY ORIENTED GEOCODING SERVICE

Franz-Josef Behr[a, *], Astrit Rimayanti[b], Hui Li[c]

[a]Department of Geomatics, Computer Science and Mathematics, University of Applied Sciences Stuttgart
Schellingstraße 24, D-70174 Stuttgart - franz-josef.behr@hft-stuttgart.de
[b]National Coordinating Agency for Surveys and Mapping (BAKOSURTANAL), Indonesia, Jl. Raya Jakarta-Bogor
KM 46, Cibinong 16911, Indonesia - rimayanti@yahoo.com
[c]M-Way Solutions GmbH, Leitzstr. 45, D-70469 Stuttgart - leeglanz@gmail.com

Commission IV, WG IV/5

KEY WORDS: GIS, Internet/Web, Services, Georeferencing, Interoperability, Developing Countries Spatial Infrastructures

ABSTRACT:

One of the most important features in many applications, especially in geographic information systems (GIS), is the capability to locate addresses, i.e. to geocode the address. During the last year there has been also a significant progress to geocode addresses by commercial Internet mapping APIs and Internet services. Geocoding became more popular and online mapping tools have popularized the concept of using an address as an initial map navigation tool. In contrast to these commercial solutions with focus on industrialized countries, the project described in this paper is an implementation of a free, participatory, community oriented geocoding service. The application offers a free geocoding tool as a substitute for commercial solutions and can be found at *http://www.opengeocoding.org/.* While commercial solutions concentrate on industrial countries – and offer only their service for postal addresses in these countries – this service offers geocoding for developing countries. Hence, the application can be an alternative to the existing commercial geocoding services, independent from the commercial solutions provided that a critical mass of participants can be achieved.

1. INTRODUCTION

1.1 Motivation

"Bridging the widening gap between the 'information-rich' North and 'information-poor' South required developing countries to invest in skills training and upgrade information technology infrastructure" – with these words, exploring creative partnerships for research and development, Benjamin Aggrey Ntim, Ghana's Communications Minister, has been cited by the United Nations (2008). From a geospatial point of view one particular data set for such an information infrastructure is address data, the base of many administrative and economic processes in all countries. While in few countries such data are already provided by state agencies or are offered from commercial vendors, in most of the countries such data is not available at all.

Based on such data, online mapping tools offer since a few years geocoding services, but limited to developed countries. Today, when the Internet has revolutionized the world of Information Technology, users from developing countries should not be left behind: Even in less developed countries addresses should be available as well as a geocoding application to provide locational information (i.e. coordinates) for postal addresses. Such a geocoding application will not only integrate open data with open source code, but also will have its own database to support the demand of users in both developed and developing countries (The International Monetary Fund 2008).

1.2 Objectives

The overall aim of this project is to design and to implement an open source geocoding service for free, participatory, community oriented geocoding of postal addresses. The specific objectives of this research were focused on

- Evaluation of different addressing standards and schemas (see Davis 2003, Xuehu 2006).
- Design and development of a database structure that contains information about the addresses and of a concept to validate data against a reference dataset.
- Implementation of a geocoding application with emphasis on the query execution on heterogeneous addressing schemas, based on manual input or on service requests.
- Visualization of the results by available Geobrowser APIs.

2. RELATED WORK

Davis et al. (2003) discuss how geocoding can be implemented over incomplete and possibly inaccurate addressing data (which is the prevailing situation in Brazilian cities) and how users can benefit from the results of such a process.

Nicoara (2005) investigates the lack of standardization in addressing, differing ways of defining location, ownership and property definitions, the complexities of multi-source location and data address maintenance, differing address table schemas and models, and the variety of algorithms used.

---

* Corresponding author.

Gruber (2006) mentions that several geocoding services like Google Maps, Yahoo Maps and MapQuest have been popularised the concept of using address as an initial map navigation tool.

Hill (2006) discusses fundamentals of geospatial referencing.

2.1 Geocoding applications

An overview about an extensive list of geocoding and reverse-geocoding applications can be found at http://groups.google.com/group/Google-Maps-API/web/-resources-non-google-geocoders. By analyzing this list it can be noted that most of theses services are offered mainly for industrialized countries like USA, Canada, Japan, and the EU states.[1]

Majewski (2006) explains the geocoding features of Google's API. Addresses can be submitted as a single string and are parsed on server side, expanding abbreviations. The geocoder response uses JSON, KML or xAL format (see below). Generally, developers are encouraged to build their own client-side caches[2]. Because geocoding via HTTP is supported, it is possible to use AJAX technology on client-side or to call the service from server side applications.

The Geocoding API of Yahoo! Maps supports REST-(Fielding (2000) based queries.. The result is delivered in XML format[3]. Similar to Google this service also provides a precision response about the result of the geocoding process.

Both services have severe limitations concerning the frequent usage. Yahoo's geocoding service is limited to 5,000 queries per IP address per day. Google's geocode requests are limited to 50,000 requests per day per API key.

Geonames.org offers a GUI based geocoding and mapping service for country and place names as well as postal codes. Geocoding is also supported by REST web services. Since a few weeks, the originally free service has a 50'000 credits limit per day and IP address. For geocoding services, 1 – 4 credits are needed.

Geocoder.us is basically a Perl frontend to the TIGER/Line file from the US Census Bureau. Here it is evident that the data are based on governmental data providing for the whole country.

Universal Address System developed by NAC Geographic Products Inc. (http://www.nacgeo.com/) is designed to provide a highly efficient unified representation of an address, postal code, area code, geographic coordinates and map grid coordinates in the world.

It has to be noted that geocoding is often based on approximation and interpolation. http://www.ontok.com/geocode/compare presents some results of accuracy assessment based on a sample of 41 addresses. According to this website the RMS is at least approx. 30 meters.

There are also several free projects of geocoding applications.

Geokit (http://geokit.rubyforge.org/) is a Ruby on Rails plugin for mapping applications geocoding, location finders, and distance calculation. For geocoding Geokit provides a uniform interface for geocoders (Google, Yahoo, Geocoder.us, Geocoder.ca, with failover!), and IP-based geocoding.

Another geocoding library for Ruby as a wrapper for Yahoo!Maps and Geocoder.us is available from http://geocoder.rubyforge.org/.

Theurer (2005) describes a mashup using Yahoo!Maps which offers geocoding address, overlay the traffic information and overlay the local search information. He uses REST to call Yahoo! API.

Stark (2008) reports about an application in Switzerland (http://www.openaddresses.ch/en/) to collect address data on a collaborative approach.

2.2 Address standards

There is a great variety of international addressing system which might lead to interoperability problems. These problems have encouraged many organizations and developers to create specifications for name and address data. The Organization for the Advancement of Structured Information Standards (OASIS, http://www.oasis.org) has quite successfully developed the extensible Customer Information Quality Specifications Version 3.0 (CIG) which includes formal definitions for Name (xNL), Address (xAL), Name and Address (xNAL) and Party (xPIL). xAL is used as part of Google Gecoder's response. Based on xNL and xAL specifications, xAL uses hierarchical nested XML elements to encode address information and is also available as one of OpenGeocoding's output options.

xNAL is listed together with other XML based address http://xml.coverpages.org/namesAndAddresses.html.

The Government of Victoria (2006) published an own address standard.

Issues related to an international address standard, its relevance for postal and Spatial Data Infrastructures and other topics are currently part of ISO/TC 211's agenda during a workshop in Copenhagen at end of May 2008 (http://www.isotc211.org/-Calendar_Copenhagen.htm).

3. THE GEOCODING APPLICATION

3.1 Design phase

Generally, as mentioned before, there has been a huge advancement in geocoding technology, but limited to developed countries. It would be beneficial if an application could incorporate addressing systems and data of both developed *and* developing countries. The geocoding application described here makes an attempt to address this problem

During the project the following research objectives have been persecuted:

- Analyze the users' information about their addressing system through questionnaire.

---

[1] A list of supported countries for Google's geocoder can be found at http://mapki.com/wiki/Read_This_First#-Geocoding_UK.2C_China_or_Japan

[2] http://code.google.com/apis/maps/documentation/services.-html#Geocoding_Caching [2008-05-09]

[3] http://developer.yahoo.com/maps/rest/V1/geocode.html [2008-05-09]

- Identify data sources, methods, and technologies for designing the address database.

- Design of architecture for addressing database with validation from reference data and of a query page for geocoding.

- Implementation of the geocoding application using open-source software, emphasizing on query execution on heterogeneous addressing systems and identifying potential problem in the implementation phase.

The development was driven by a survey performed with international Master students and alumni from the Photogrammetry and Geoinformatics course of Stuttgart University of Applied Sciences to gather information about addressing systems used in different countries. The database structure has been setup accordingly.

3.2 Data used as reference information

There exist a huge amount of publicly available datasets on the Internet which might be useful as reference data sets. Theses data are evaluated according their fitness for use and their availability. Data suited for this purpose are downloaded and stored in a MySQL database. As main data sources the following should be mentioned:

- List of place names with coordinates and bounding boxes of countries worldwide: http://www.geonames.org.
- Global list of cities: http://en.wikipedia.org/wiki/-List_of_cities
- List of ISO codes of countries worldwide: https://www.-cia.gov/cia/publications/factbook/appendix/appendix-d.html.

3.3 Functional aspects

There are two usage aspects for geocoding service in this project.

- Data collection: Registered users submit address information and corresponding coordinates. Special attention is given to the validation to prevent spamming, by comparison of user data to existing data on country, province and city level. Supported by AJAX (Garret 2005) based techniques for map display and user friendly auto-completion, users can enter addresses and the corresponding coordinates which could be derived from maps or GPS measurements. Alternatively, users can use a Web map to digitize the location,. After successful data validation, address and coordinates are saved in the database where they can be used for geocoding later on.

- Geocoding: The geocoding service for postal addresses is offered based on this data free of charge through a REST based web service. The geocoding process itself includes the transformation of an unstructured address into a structured address, the establishment of a correspondence between the structured address and the addressing database, and the assignment of coordinates to the address. The geocoding process itself needs access to the database in which the reference data and user generated information is stored. The geocoding application does not rely only to its internal database filled by the community. If no relevant results can be retrieved from its database, the request will be forwarded to other publicly available geocoding services. The result will be checked, parsed, and forwarded in the format specified to the user.

Table 1 lists the service parameters that can be used by clients for submission of addresses.

Parameter	Value	Description
Country	string (country code, iso-3166-optional)	country name (optional)
Province	string	province name (optional)
City	string	city name
District	string	district name (optional)
Streetnames	string	street name
Postcode	number	poastal code
Output	string	output format for the result

Table 1: Parameters for geocoding service request

The example below show sthe request for the service to geocode Acra, Ghana, with JSON output format.

http://opengeocoding.org/geocoding/geoservice.php?country=GH&province=&city=Accra&district=&streetnames=&postcode=&output=json&submit.x=32&submit.y=24&submit=geocode

Besides offering REST services, the application additionally offers a Web based user-friendly front-end for interactive address submission (Figure 1).

Figure 1: User interface for submission of an address.

Results are displayed in a XHMTL page offering the opportunity to display the location based on the Mapstraction library (see section 3.7).

3.4 Development

For quality assurance, the application has been tested rigorously before uploading it for the general use. Errors and bugs that were discovered during the development process were eliminated carefully. Because different patterns of addresses exist and all of them should work properly in the application, various testing schemes were prepared to check the application in various scenarios and for various patterns of address.

Different types of data were entered and checked for their accuracies.

To give access to the public the application was deployed at the domain www.opengeocoding.org. As the application was implemented for the community under participative aspects, the application contains a platform to exchange views within different users. To achieve it the application uses the free forum package *MyBB* developed in PHP and MySQL. It is believed that using such a platform the community can share freely their ideas about usage and usability of this application. Moreover, entries from the community provide better understanding of the application.

3.5 Technologies used

The portal site is based on Google's API, JavaScript and AJAX. PHP is the corresponding server-side scripting language for for database access, geocoding and creation of web pages. MySQL, since February 2008 owned by SUN Microsystems, Inc. (http://www.mysql.com/, Williams & Lane 2002) is used as a database management system.

3.6 Output formats

The service provides results in XML, KML, GML, JSON, and CSV format. The XML format corresponds to the aforementioned xAL standard, embedded in a KML envelop generated according to the OpenGIS® KML Encoding Standard (http://www.opengeospatial.org/standards/kml/). GML output contains coordinates according to the OpenGIS Geography Markup Language (GML) Encoding Standard (http://portal.opengeospatial.org/files/?artifact_id=20509).

The following code fragment show s the result of geocoding of Accra, Ghana in XML format:

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://earth.google.com/kml/2.0">
 <Response>
  <name>Ghana</name>
  <Status>
   <code>200</code>
   <request>geocode</request>
  </Status>
  <Placemark id="p1">
   <address>Ghana</address>
   <AddressDetails Accuracy="1"
    xmlns="urn:oasis:names:tc:ciq:xsdschema:xAL:2.0">
    <Country>
     <CountryNameCode>GH</CountryNameCode>
    </Country>
   </AddressDetails>
   <Point>
    <coordinates>-1.023194,7.946527,0</coordinates>
   </Point>
```

```
  </Placemark>
 </Response>
</kml>
```

Another format in which results can be transferred to the client is called JavaScript Object Notation (JSON, http://json.org/) which uses a quite conventional JavaScript notation to define objects, in the context discussed here to define geographical features. JSON is meanwhile used quite intensively throughout the web, for example several of Google's web-based applications and services can provide feed data in JSON format.

The following code fragment show s the result of geocoding of Accra, Ghana in JSON format:

```
{ "name":"Accra, Ghana",
  "Status":
  {"code":200,
   "request":"geocode"},
  "Placemark":[
   {"id":"p1",
    "address":"Accra, Ghana",
    "AddressDetails":
     {"Country":
      {"CountryNameCode":"GH",
       "Locality":
       {"LocalityName":"Accra"}
      },"Accuracy": 4
     },
    "Point":
    {"coordinates":[-0.20738,5.54009,0]
    }
   }
  ]
}
```

The JSON-encoded information is executed as JavaScript statement by JavaScript's eval() function. So all the objects defined in the JSON string are instantiated.

CSV format (comma separated values, an implementation of a delimited text file) is also supported. Because this format is quite simple and supported by almost all spreadsheets and database management systems, it has become a pseudo standard even for GI systems.

3.7 Presentation of results

With the interactive format option, results will be displayed in HTML (Figure 2) which shows all match results found – placenames, even in combinations with state or country names do not always uniquely specify a location (Hill 2006:111). In this case, the user can decide which location should be displayed on the map or should be forwarded to the Mapstraction part of the application.

Besides using Google Maps, geocoded addresses are also visualized using the Mapstraction library (http://-www.mapstraction.com), a common API for different online mapping tools allowing users to switch from one map presentation to another smoothly. Principally it is possible to display maps in independent views, including MapQuest, Map24 and OpenLayers. With Mapstraction library however, there is *one* effective way to display certain address in three different views at a time (Figure 3).
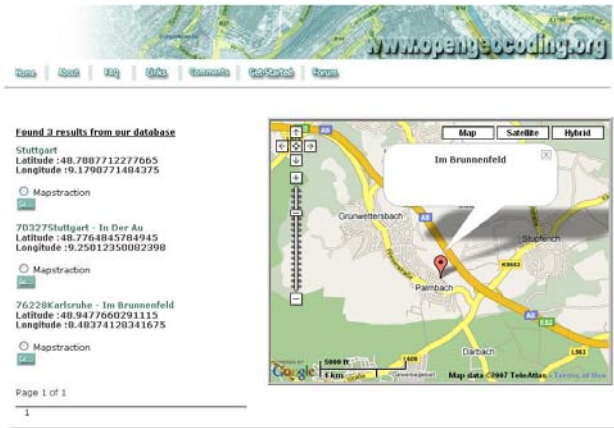
Figure 2: Result from Database in interactive mode.

4. CONCLUSION AND RECOMMENDATION

4.1 Conclusion

The major result and principal benefit of this project is the design and implementation of a portal site with offers a user interface for the collection of georeferenced address information as well as for the geocoding of addresses. Additionally the geocoding operation is offered as a REST based service. The application gathers data from the users and share it freely to the community.

Figure 3: Result of geocoding process. Mapstraction. displays the result in Yahoo!Maps, Google Maps and Live Local.

The application has been tested carefully with different addressing alternatives to identify potential problems that might occur during its launch for a wider audience.

Results will be displayed in HTML format, with map visualization, but are alternatively provided in XML, KML, GML, JSON and CSV format. The rationale behind providing different output formats is to give users the specified output result that might be useful for their further work.

The geocoding application in this project does not rely only to its database which is filled by the users but also it make use of publicly geocoding services such as Google Maps and Yahoo!Maps.

The application offers free geocoding as a substitute for commercial solutions and can be found at http://www.opengeocoding.org/. While commercial solutions concentrate on industrial countries – and offer only their service for postal addresses in these countries –, this service aims offering geocoding especially also for developing countries. Hence, the application can be an alternative to the existing geocoding services provided that a critical mass of participants can be achieved.

4.2 Limitations

However the application has got a few limitations.

- Because the digitization operation relies on Google Maps, this operation fails in areas or countries where no detailed map information is available like China, Albania, Azerbaijan, or Canada(!).

- The spatial reference given by the Google Maps can be changed by Google without any preceding notice – digitized location will be shifted compared to Google's satellite imagery.

- Due to availability of reference data, the validation is partially only applied for country, city and coordinates submitted by the community.

- The more geocoded addresses are submitted by the community, the more this application will be applicable.

4.3 Recommendation

However there are still some possible improvements:

- There are cities and even countries having several names or names changing during time. The application should be able to recognize a placename independent from the form the user used. Currently the similar_text function in PHP was used in the validation part to calculate the similarity between two strings,. This could be enhanced with another functions such as soundex., a phonetic algorithm for identification of names by sound, as pronounced in a certain language.

- Due to the availability of reference data, the validation is only applied to country and city and coordinates. It could be better to have references data for other data such as for street name and postcode to have better methods and algorithms to be developed and to have more accurate geocoded address.

REFERENCES

Davis C. A., Fonseca, F.T., De Vasconcelos Borges, K.A., 2003. A flexible addressing system for approximate geocoding. In: *GeoInfo 2003: Proceedings of the Fifth Brazilian Symposiumon GeoInformatics*, Campos do Jordão, São Paulo, Brazil. http://www.geoinfo.info/geoinfo2003/papers/geoinfo2003-25.pdf [2008-05-09]

Fielding, R. T., 2000. Architectural Styles and the Design of Network-based Software Architectures. http://www.ics.uci.-edu/~fielding/pubs/dissertation/top.htm [2008-05-10]

Garrett, J. J. (2005). AJAX: A New Approach to Web Applications. Technical report, Adaptive Path. http://www.-adaptivepath.com/publications/essays/archives/000385.php. [Accessed on 09.09.2005]

Geoorigin 2008. *Geocoding*. http://www.georigin.com/-services.asp?id=2. [2008-05-09]

Government of Victoria, 2006. Whole of Victorian Government Standard - Street Address Data Standard. A standard for the storage, interchange and validation of street address data by Victorian Government agencies. http://www.dtf.vic.gov.au/CA25713E0002EF43/WebObj/DIStreetAddressData/$File/DI%20Street%20Address%20Data.pdf [26. Febr. 2006]

Gruber, F. 2006. *Comparing the Mapping Services.* http://www.techcrunch.com/2006/04/17/comparing-the-mapping-services. [2008-05-09]

Hill, L. L., 2006. Georeferencing: The Geographic Associations of Information. MIT Press, Cambridge, Massachusetts

Majewski B., 2006. *Geocoding at last!* http://googlemapsapi.-blogspot.com/2006/06/geocoding-at-last.html.[2008-05-09]

Nicoara, G., 2005. *Exploring the Geocoding Process: A Municipal Case Study using Crime Data.* http://charlotte.utdallas.edu/mgis/prj_mstrs/2005/Summer/greta/Nicoara_Masters/Website/main.html. .[2008-05-09]

Stark H.-J., 2008. OpenAdresses.ch. Proc. FOSSGIS 2008. http://www.fossgis.de/wiki/images/5/55/OpenAddresses.ch.pdf [2008-05-10]

Taylor (2007). *FREE mapping service from Yahoo and Google.* http://www.enginesforwebsites.com/yahoo_google_maps_-geocode.aspx. [13. Febr. 2007]

The International Monetary Fund, 2008: List of advanced and emerging and developing economies . http://www.imf.org/external/pubs/ft/weo/2008/01/weodata/groups.htm. [2008-05-10]

Theurer, D., 2005. *How to build a Maps Mash-up.* http://www.theurer.cc/blog/2005/11/03/how-to-build-a-maps-mash-up/ [2008-05-10]

United Nations, 2008: Press Release TAD/2043, 24 April 2008, http://www.un.org/News/Press/docs/2008/tad2043.doc.htm [2008-05-10]

Xuehu Zhang, Haoming Ma, Qi Li, 2006. An address geocoding solution for Chinese cities. in: Geoinformatics 2006: Geospatial Information Science, edited by Jianya Gong, Jingxiong Zhang, Proc. of SPIE Vol. 6420, 64200J, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.712698

ACKNOWLEDGEMENT