A Flexible Addressing System for Approximate Geocoding

CLODOVEU AUGUSTO DAVIS JR.[1]
FREDERICO TORRES FONSECA[2]
KARLA ALBUQUERQUE DE VASCONCELOS BORGES[1,3]
[1]PRODABEL- Empresa de Informática e Informação do Município de Belo Horizonte,
Av. Presidente Carlos Luz, 1275, 31230-000, Belo Horizonte, MG, Brasil
{clodoveu, karla}@pbh.gov.br
[2]Pennsylvania State University
State College, PA, United States
ffonseca@ist.psu.edu
[3]Departamento de Ciência da Computação da UFMG,
Av. Presidente Antônio Carlos, 6627, 31270-000, Belo Horizonte, MG, Brasil
kavb@dcc.ufmg.br

Abstract. One of the most important features in an urban geographic information system is the ability to locate addresses, in any form employed by the population, in a quick and efficient way. From an organized set of georeferenced address data, local governments, infrastructure companies, and the various agents that need to work in urban spaces can reliably and precisely locate points of their interest, in fields of application as diverse as public health services, crime fighting, product distribution, postal delivery and many others. This activity is known as *geocoding*, which can be defined as the process of locating points on the surface of the Earth from alphanumeric addressing data associated to events. This paper shows 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. A geocoding quality indicator is proposed, with the intention of providing an indication of the degree of certainty that is associated with each individual result of the geocoding process. This indicator can then be used in spatial analysis, by giving analysts the opportunity to include the uncertainty associated with the data in their work. This project is under development, and a set of (as yet unimplemented) geocoding tools is proposed and described in the paper.

1. Introduction

One of the most important features in an urban geographic information system (GIS) is the ability to locate addresses – in every form known and used by the population – in a quick and precise manner. This is useful in a large number of situations, in special those that require georeferencing legacy alphanumeric databases and information systems. In most of these conventional information systems, addressing information exists, but probably in a semi-structured fashion. Since urban (postal) addresses are the first and foremost way used by citizens to express the location of points of their interest, including their own places of work and residence, it is natural that they get to be employed as a sort of *indexing key* for urban spaces [5][3]. From an organized set of georeferenced urban addresses, local government managers or private sector companies can quickly and reliably locate events or phenomena of their interest, in fields of application such as health services, crime fighting, product distribution, postal delivery and many others, in an activity known as *geocoding*.

In developed countries, especially in the United States, there are usually governmental nation-wide efforts to generate and maintain an addressing network that can be used for a number of different purposes, such as census activities and mail distribution. The U.S. Census Bureau, for instance, has been maintaining and distributing at a very low cost its well-known TIGER (Topologically Integrated Geographic Encoding and Referencing) files [19], in which addressing information is coded as a set of attributes for segments of street centerlines. Private sector companies have access to this material, and invest in its improvement, thus demonstrating that address databases can be a valuable economic asset. In the UK, the Ordnance Survey produces and sells licenses on an address point database that contains over 25 million locations, along with a coordinate list for the 1.6 million distinct postal codes within the country [14].

However, emergent countries such as Brazil have no such organized addressing infrastructure. The consequences of this for urban geographic applications are manifold, since georeferencing event data can take much longer, with poorer quality and several consistency and precision problems. Furthermore, large cities in emergent countries often contain slums, shanty-towns and other types of low-income areas that are characterized by chaotic occupation, for which, in some cases, there is not even a physical address marked in each dwelling, let alone an organized digital database containing those addresses.

Nevertheless, the usefulness of georeferenced address bases is such that, in many places, local government departments and infrastructure service providers constantly invest in the generation of address-based georeferencing information based on alphanumeric cadastres and conventional cartographic sources. Since there is often no established standard for the creation of such information resources, regional or national efforts that need to work with point-georeferenced event data, especially in the epidemiologic and crime fighting segments, have been severely hindered.

Notice that different geographic applications require different levels of accuracy regarding the location of events. Also, as mentioned, existing addressing information infrastructure can be quite variable from one place to another. This calls for generic geocoding methods in which not only the degree of success in the location of events is maximized, but also the quality of the result can be assessed and taken into consideration during the spatial analysis activities that usually follow the geocoding effort.

This paper describes a work in progress. The concepts and basic parameters for the development of an uncertainty-driven geocoding tool are described here, even though this implementation has not yet been undertaken. Such a tool will be useful to many projects that require point-georeferenced urban data from social or environmental events or phenomena, such as public health and crime analysis. In special, the SAUDAVEL (an acronym in Portuguese for an epidemiologic surveillance support system) project [10] will directly benefit from it.

The remainder of this paper is structured as follows. Section 2 presents the variety of socio-cultural contexts in which addresses are employed, highlighting the differences and similarities among them. Section 3 proceeds to describe the geocoding process from the conceptual and operational points of view. Section 4 presents our proposal for an indicator of the quality of the geocoding results. Finally, section 5 presents some conclusions and discusses possibilities for future work.

2. Socio-cultural Aspects of Urban Addresses

The idea of numbering buildings within cities arose from the need to guide visitors (or even residents) as to the direction of a given dwelling or commercial activity. The numbering system that most Westerners are accustomed to, *i.e.*, sequentially increasing numbering along the street, with odd numbers to a side and even numbers on the other, is widely accepted but is far from being dominant. According to [17], many cities implement different numbering systems for historical reasons, since standardization and modernization did not (or could not) occur everywhere.

The first addressing initiatives took place in Western Europe and China in the 18th Century [17]. Numbering every building was not a general rule until government realized how a more efficient addressing would help cadastral and fiscal initiatives. For that and other reasons, numbering in Paris did not start until 1779, and was met with resistance from the population, in special from the dominant classes. There are some places, however, in which a name assigned to a building can also have addressing value. This is the case of Istanbul, in Turkey, and of Maceió and Salvador, in Brazil. In the latter, for instance, ancient Portuguese occupation produced sequential odd and even numbers at each side of the streets, but failed to foresee the need for intermediate numbers, in case of new buildings or reparceling. Thus, large or important buildings had to be known by name within the street, and currently even the phone book lists building names in association to addresses. The sequential odd-even system is also widely used in France, and has the disadvantage of not indicating directly the relative distances between buildings.

But there are many other addressing systems in the world. Japanese and Korean systems, for instance, can be quite different from their Western counterparts. According to a Web-based tourism information source [13], “easily the most maddening thing about Seoul is the system of street addresses. If you've spent time in Japan, you might have some luck understanding the system, but probably have no better chance of actually finding anywhere without a detailed map; even the locals seem baffled by the system. Addresses are not given according to consecutive numbers along a road, but as random numbers within an arbitrary block of land. You will need to wander around looking for your destination, as there are virtually no street signs (let alone street names). Most business cards have maps printed on the back.” In fact, most streets actually have no name at all. Numbers are assigned inside neighborhoods (called *dong*) within urban sectors (called *gu*), a hierarchy of areas that are named, not numbered [18]. The Korean numbering system is similar to Western land parcels numbering, something that has been successfully used for a long time in cadastral applications – but that is far from being the most intuitive system for urban navigation [12]. It is a very old system, which has been kept because of tradition and because it would probably be very difficult to change it. In Kyoto, Japan, the Digital City project has been conceived with this kind of limitation in mind: a map-based user interface facilitates the location of points of interest for tourists and locals alike, since the addressing system seems to be somewhat complicated to navigate without detailed mapping information [8].

One of the most important addressing systems in the world is the metric numbering system, combined with the odd-even rule. Buildings are assigned numbers according to their metric distance from the beginning of the street, rounded up to the nearest odd or even number, or approximated in a way that every building gets a different number. There are variations in which the numbering is sequential and block-oriented (for instance, 100 numbers to a block), but the numbering is not distance-oriented. This system has the advantage of allowing an easy assessment of the distance between two addresses in the same street, while allowing for simple adaptation to new developments along each street.

The metric numbering system requires two additional factors to be implemented and managed: a clear definition of the starting point of each street, and active coordination by local government as to the "official" assignment of numbers. Starting points have to be well chosen to avoid the need for renumbering whenever a street is extended; thus, these points are usually defined as being the closest to the city's center. In some cities, such as Venice [17], this point is defined in reference to well-known regional landmarks, a criterion which seems more logical for large cities: it is not always easy to determine which end of a 200-meter street is closest to the center if the urbanized area extends for tens of kilometers from the city's original landmarks. As to governmental coordination, the idea is to prevent any deterioration of the system's working logic, so house numbers would have to be assigned by some municipal authority. Nevertheless, clandestine and illegal developments constantly appear in Brazilian large cities, and even though the population tries to follow the addressing rules as a way to include themselves more easily in the public services network, irregular numbering frequently occurs.

While there is certainly a great variety of addressing systems throughout the world, there seem to be too few standardizing initiatives. In the United States, a standard for address data has been opened for public review in the spring of 2003 by the FGDC [6], as a proposal for the creation of a national spatial data infrastructure standard. It attempts to establish a basic terminology in order to create a semantic agreement regarding address data. For instance, the proposal defines addresses as "locators to places where a person or organization may reside or receive communications, but excluding electronic communications". This initiative is still at an early stage of development. It is also the case of the European Geographical Information Infrastructure Initiative, which decided on the need to include basic data sets that will allow for the automatic geocoding of any address in Europe [1].

Having demonstrated that there is no universal standard, we can now point out some common traits among all these addressing systems, considering the intention to assign coordinates to every address that can be recognized by postal authorities. The concepts of street (or, more generally, thoroughfare), building numbering, neighborhood, and city seem to be approximately the same all around, even though in some situations concepts of a more cadastral nature, such as parcel or block, are used as address references. Incomplete, inaccurate, or hard-to-use addresses also require a number of indirect references, which can be thought of as distinct landmarks within the city or points widely known and recognized. In the future, we intend to create an ontology of addresses, in which all those concepts can be made explicit and interrelated, considering the cultural context.

3. Geocoding

In this section, we will limit the discussion to the metric addressing system with the odd-even rule, as implemented in many Brazilian cities. Of course, since governmental action in the assignment of addresses is slim to none, many irregularities and exceptions occur. These will be treated here.

Geocoding is an expression that has been usually associated with the idea of locating points in the surface of the Earth from alphanumeric address information (Figure 1). Geocoding is often performed as a preliminary step to spatial analysis of point data, and the requirements as to the geographic accuracy of the process must be defined by the application. Sources for addresses include conventional information systems, cadastres, directories, web sites, and specialized search engines [4, 7, 9, 20, 21].

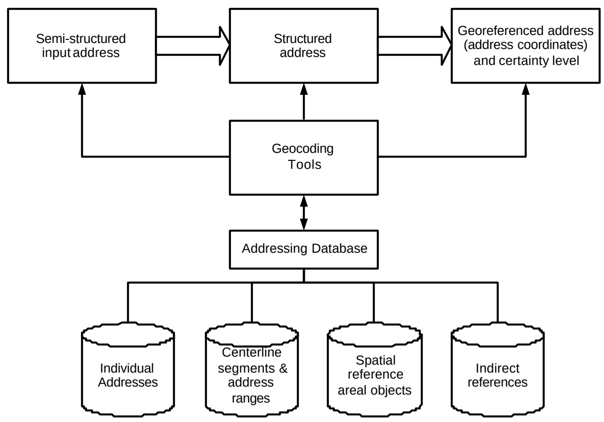

Figure 1 - General geocoding schema

We assume that geocoding is being performed in order to establish a location for a real world event or phenomenon – even though this location can be approximate. Indications of the actual location are to be gathered from the address that is associated with the event or phenomenon. We also assume that input addresses may be provided in a semi-structured way, *i.e.*, addresses may or may not have separate fields for each addressing component, such as thoroughfare type (street, avenue, plaza, and so on), thoroughfare name, building number, zip code, city, state, country and others. This structure has to be obtained from the input data, in order to allow us to proceed.

Starting from structured addresses, actual geocoding can be performed in several ways, depending on the available addressing information. These different geocoding strategies will be described later. In each of the strategies, some discussion on the accuracy of the geocoding process will be included. We argue that, considering that the accuracy requirements from each application can be quite variable, and that there can exist places for which the existing addressing information infrastructure is not suitable to all of the strategies, the geocoding process will probably generate results of variable quality.

The result of each geocoding operation is a pair of coordinates that are to be associated with the input address. In this paper, we propose the calculation of an indicator of geocoding quality, a number that is a function of the uncertainty associated with the geocoding method that has been employed, which in turn depends on the completeness of the available addressing information.

3.1 Geocoding Process

The geocoding process includes the treatment of semi-structured alphanumeric addresses of events (which we call *parsing*), the establishment of a correspondence between the structured address and the addressing database (*matching*), and the actual assignment of coordinates to the event (*locating*). In order to be able to perform the parsing, matching, and locating tasks, the geocoding process needs to have access to a database in which information about the addressing system are stored.

The parsing phase consists on the analysis of character fields containing address information and their transformation into a structured and standardized database tuple, ready for the matching phase. This can be done using various text- and string-oriented algorithms [15], something we will not discuss in this paper. In this process, there is often the need to deal with incorrections and mistakes that are quite common in unstructured address data, seeking to establish a correspondence between them and "official" place names. Naturally, information on place names must be available for querying in the database. It is also useful to have information on popular (unofficial) and outdated place names.

The matching phase tries to find, in the database, the most precise reference that can be positively associated with the given address. The results of the parsing phase may be incomplete, and depending on the missing information a decision must be made as to which available reference to a location will generate the most precise results. For this to happen, there should be two additional groups of data in the database. The first one is comprised of the actual addressing infrastructure, with object classes such as point-georeferenced individual addresses and street centerlines with address ranges. The second includes any additional information items that can be used to resolve ambiguities or as a rough geographic reference in case the address, for any reason, cannot be located in the first category. This includes elements such as all sorts of spatial reference units (area objects that correspond to artificial borders, such as neighborhood limits, districts, ZIP areas, municipal divisions, and so on), along with a catalog of reference points known by the citizens. This catalog can contain what we call "reference places", *i.e.*, popularly known spots in a city that extend our common notion on addresses, since they can be identified by name. These points are so easily recognized by the population that finding their location does not require a formal address.

Finally, the locating phase receives the results from the matching and determines the actual coordinates that are to be assigned to the address. Depending on the object or objects that have been found to be in association with the event, this assignment of coordinates may be trivial (i.e., by simply copying the coordinates of a point object) or complex (for instance, by generating a random point inside an area object).

3.2 Conceptual Schema for Geocoding

From the description of the process in the previous section, a database schema for the addressing system database is proposed (Figure 2). The diagram has been developed using OMT-G [2].

The schema has been prepared considering the full range of addressing components; however, the actual addressing database can be rather incomplete, depending on the available data about a given city or location. It includes, therefore, objects such as municipalities, neighborhoods, thoroughfares (and alternative names), street centerlines and address ranges, street crossings, individual addresses and reference places. A gazetteer is also included, as a tool to resolve generic or uncharacterized place names [16].

The source for alphanumeric semi-structured addresses of events has been intentionally left out of the schema, since it can take on many forms, depending on the problem that requires geocoding. As a general measure, though, the results of the parsing and of the matching phases should be stored, along with the final location of the event. This should be done as a measure to allow future refinements on the location, in case the addressing database is updated or expanded.

Considering, as mentioned, that the database may not be fully populated, the matching phase of the geocoding process must try to geocode at the most precise level first. If that is not possible, it must successively resort to less precise geocoding methods until some location can be established. These methods are discussed next.

3.3 Strategies and Results

Observing the conceptual schema in the previous section, we can design a number of geocoding strategies. The accuracy of the results is a consequence of the method employed and of the quality of the original data. Several strategies that can be implemented from such an addressing database will be described next.

Individual point address. Naturally, the most precise method is the matching of the input address to an object from the `Building Address` class (Figure 3). If a matching individual address is not found, it is possible to resort to the numerically closest address in the same street. There is also the possibility of matching the given address to a reference point (i.e., to a member of the `Reference Point` class). Building names, like in the case of Salvador, may also be matched, if buildings are included as reference points.

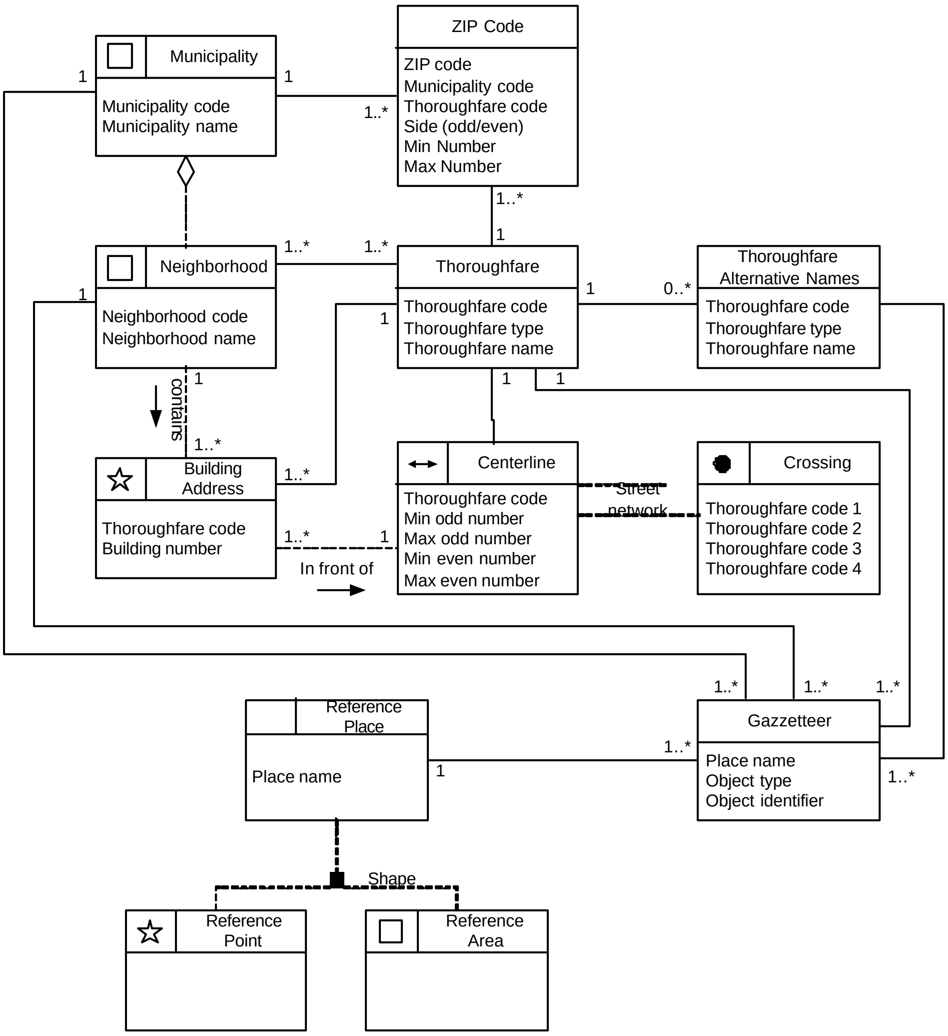

Figure 2 - OMT-G Addressing Database Schema - Class Diagram

Centerline range interpolation. An approximate location can be established by interpolating along a street segment, observing the addressing ranges that are associated with each one of them (Figure 4). This may not be possible in the case of irregular numbering, but is viable in case the numbering is non-metric.

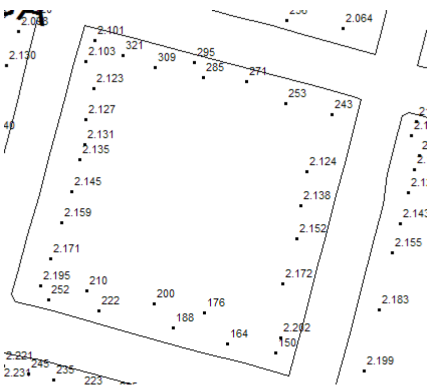

Figure 3- Individual addresses

Thoroughfare interpolation. If the street network exists but is not properly segmented, it is possible to establish an approximate location by assuming the metric system. The problem is the identification of the starting point. A digitizing rule in which the starting point (i.e., the end of the street in which numbering starts) is plotted first may help.

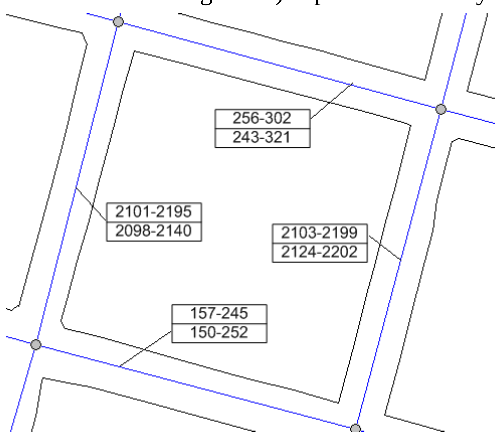

Figure 4 - Street centerlines and address ranges

Reference area. These are objects that have been included in a catalog of urban references, but cannot be associated to a precise point. Examples include large urban equipments, such as sports arenas, parks, or transportation terminals. In each case, an area object corresponding to the limits of the referenced place will be included in the `Reference Area` class.

Street Crossing. Many times, a location within a city is specified as "at the crossing of Street A and Road B". This situation is included in the conceptual schema we presented, and such situations can be adequately treated. If the street network has been completely fed into the database, it is also possible to locate thoroughfare segments contained within two crossings, as in "Street A between Streets B and C".

Neighborhood centroid. If there is an indication as to the name of the neighborhood the address belongs to, the location may be associated to the neighborhood polygon's centroid. In this case, as in all cases of area-based geocoding, an alternative for the location phase is the generation of a random point inside the neighborhood boundaries.

Postal area. In many parts of the world, postal areas receive codes (such as the ZIP code in the U.S. and the CEP in Brazil). These can be mapped and their boundaries used just like the neighborhoods in the preceding strategy. Special postal delivery points, which receive special codes, may be included in the database as points.

Municipality. If no other addressing reference is usable, the generation of a random point inside the municipal boundaries or the association of the address to the centroid of the municipal boundaries may suffice for many applications.

Naturally, the idea behind this proposal is to try to geocode at the most accurate level first, and then to successively resort to less precise methods as the results turn out to be negative. We assume that the detail level and the reliability of addressing data may vary within a given city – a very common case in Brazil and other developing countries, in which low-income areas present not only deficient infrastructure, but are also poorly mapped. We must also observe that many important applications do not require high accuracy. For instance, health applications often only need to narrow down the location of a disease case to the identification of a census sector within which it was recorded. Crime fighting applications, on the other hand, may benefit from accurate recording of cases, which can help in the identification of patterns of criminal activity.

Therefore, we propose enhancements on the geocoding process, by generating an indication of the uncertainty that can be associated with the location of each event, when this location has been obtained from geocoding.

4. Geocoding Quality Indicator

We define the Geocoding Quality Indicator (GQI) as a number that is to be associated with each geocoded event, and that indicates how close we are to the actual location of the event on the real world. The GQI is calculated to be a number between 0 and 1, 0 being the most accurate location and 1 the less accurate. The GQI is not intended to be analyzed as a distance or as an area, since its upper limit is associated with the less accurate of all methods, which is the simple identification of a municipality. Since municipal boundaries vary in area and shape, the GQI acts as a relative indicator within a given municipality.

The lowest GQI value (zero) is to be associated with the most accurate method, i.e., the matching of the event's address to an individual point address as recorded in the database (`Building Address` class). Intermediate values for the GQI are calculated as the ratio between the limits of an area within which we are sure the address is located, and the total municipality area. If this area is larger, the GQI will be greater, but will of course be limited to 1.

With this principle in mind, we proceed to determine formulae for the calculation of the GQI in three different "generic" geocoding strategies: area-based, line-based, and point-based.

4.1 GQI in Area-based Geocoding

This is the case in which the exact address could not be found, but the location has been narrowed down to an area in which the address is known to be. Since the object corresponding to that reference area is represented as a polygon, GQI should be calculated as:

$$GQI = \frac{ObjectArea}{MunicipalityArea}$$

In the case of the proposed strategies, the object area can be either the area of the neighborhood, of the postal area, or of any other object from the `Reference Area` class that has been found to correspond to the given address.

4.2 GQI in Line-based Geocoding

If the reference object is represented as a line, we propose the concept of *standard width*. The standard width is to be established as a parameter for the entire geocoding procedure, and corresponds to an average thoroughfare width. Of course, if the line object is a thoroughfare, the actual width information may be used if it exists; if not, each thoroughfare class (street, highway, avenue, plaza, square, alley, and so on) can be associated to a different standard width. Therefore, the GQI can be calculated as:

$$GQI = \frac{LineLength \times StdWidth}{MunicipalityArea}$$

4.3 GQI in Point-Based Geocoding

As already mentioned, the GQI will be zero if the given address matches an individual point address. However, there are other possibilities for point-based geocoding, using information such as street crossings or using the numerically closest individual address. In any case, the general idea for the calculation of GQI holds, with a few adaptations.

Numerical approximation. Consider the numerical difference between the given and the found numbers to be the diameter of a circle:

$$GQI = \frac{p \times |GivenNumber - FoundNumber|^2 / 4}{MunicipalityArea}$$

Street crossing. If only one crossing has been identified between the two given thoroughfares, use

$$GQI = \frac{p \times WorstCaseStdWidth^2 / 4}{MunicipalityArea}$$

If two crossings have been found, use the line rule, with extremes at each crossing point and the worst case standard width. If more than two have been found, use the area rule on the area of the convex hull of the located crossings.

Numbering range within a street segment. If the thoroughfare presents regular numbering, use the line rule with a width of half the standard width. This is to account for the distance between the street centerline and the actual location of the building.

On the other hand, if the numbering is irregular, and the given number falls within the numbering range of more than one segment, use the line rule on the union of all segments that contain the given number, along with all segments between them in the thoroughfare.

It is clear that adaptations of the point, line, and area general rules must be conceived in order to accommodate different sets of concepts and rules, such as the ones that are peculiar to a non-metric numbering system. For instance, in the sequential odd-even rule, the interpolation technique may work, but the GQI calculation based on the numbering difference makes no sense. In the Korean system [13, 12], the area rule may be applied to the *gu* or the *dong*, in the absence of individual point-based address data.

4.4 Interpretation and Use in Spatial Analysis

We intend the GQI to be used not only as an assessment of the reliability and the accuracy (relative to the accuracy of the addressing database) of the geocoding process, but as an indication of the range of situations in which the geocoded data may or may not be used.

Specifically in the case of spatial analysis efforts, the GQI may be used to filter out that portion of the data for which the certainty as of the location may be insufficient for an adequate interpretation of the results. For instance, in the case of crime data, a spatial clustering analysis requires accurate positioning of the events, otherwise significant clusters might be undetected, or false clusters could appear to be valid. In this case, only low-GQI events would be taken into consideration for analysis. For epidemiology, on the other hand, depending on the disease under study, the precise positioning of each event might not be as important as in the case of crime data; there are many analyses that are performed by aggregating events at the Census sector level, and therefore a higher GQI could be tolerated.

Alternatively, the GQI can also be taken into consideration in the analysis, by assigning greater importance to the events that have been more accurately positioned. This would allow for the development of spatial statistical methods in which the event is weighted according to the expected accuracy of its location. Clustering methods, for instance, could be adapted to take this factor in consideration.

5. Conclusions and Future Work

Even though we cannot, at this point, quote specific numeric results, this line of research shows much promise, along with large applicability in real problems of underdeveloped countries. Initial results show that the implementation of the geocoding tools described here can be achieved quite easily, and that the main source of concern is, of course, the data. Since there is no established standard in Brazil as to urban addressing data, several data collecting efforts have to be undertaken in order to fill out some of the object classes included in our conceptual model.

Another interesting line of work is, of course, in the definition and promotion of a Nation-wide addressing standard, something that can be backed by strong and affordable technology but must be guided by legitimate public sector agents. This may be greatly helped by the ontological study we mentioned earlier in the paper.

As to the insertion of the results from this work in the SAUDAVEL project, our intention is to implement the tools defined here as an extension for the TerraLib initiative by INPE [11]. This would ensure affordability and openness, two characteristics that are very important for public sector work. With the proposed geocoding tools, SAUDAVEL applications will be able to more easily locate points from their usual data sources, which include semi-structured address data from Federal information systems such as SINAN (an acronym in Portuguese for the Brazilian system that collects data on epidemic and some kinds of infecto-contagious diseases). SINAN data cover the entire country, but there are usually many difficulties in the geocoding of the events that take place in poorer cities – which are precisely where aid is more urgently needed.

6. Acknowledgements

Clodoveu Davis' work is partially supported by CNPq, the Brazilian governmental agency in charge of fostering scientific and technological development. CNPq also sponsors the SAUDAVEL project [10]. Frederico Fonseca's work is partially supported by the National Science Foundation under NSF ITR grant number 0219025 and by the generous support of Penn State's School of

Information Sciences and Technology. Karla Borges acknowledges the support provided by Prodabel, as a doctoral student in an institutional program.

7. References

[1] Arnaud, A. M. Quality of Non-Direct Georeferencing for More Democratic Access and Use of Spatial Information. *Proceedings of the International Conference on Public Participation and Information Technologies (ICPPIT99)*, Lisbon, Portugal, 1999.

[2] Borges, K. A., Davis Jr., C. A., Laender, A. H. F. OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica* 5(3):221-260, 2001.

[3] Davis Jr., C. A. Address Base Creation Using Raster-Vector Integration. *Proceedings of the URISA 1993 Annual Conference*, 45-54, Atlanta (GA), 1993.

[4] DotGeo Web Site. Available at http://www.dotgeo.org. Last access Sept. 15 2003.

[5] Eichelberger, P. The Importance of Addresses – The Locus of GIS. *Proceedings of the URISA 1993 Annual Conference*, 200-211, Atlanta (GA), 1993.

[6] Federal Geographic Data Committee. *Draft Proposal for a National Spatial Data Infrastructure Standards Project – Address Standard.* Available online at http://www.fgdc.gov/ standards/status/sub2_4.html. Last access May 02 2003.

[7] GeoSearch Web Site. Available at http://www.geosearch.com. Last access Sept. 15 2003.

[8] Hiramatsu, K., Ishida, T. An Augmented Web Space for Digital Cities. Proceedings of the IEEE/IPSJ Symposium on Applications and the Internet (SAINT-01), 105-112, 2001.

[9] InfoSpace Web Site. Available at http://www.infospace.com. Last access Sept. 15 2003.

[10] Instituto Nacional de Pesquisas Espaciais. *SAUDAVEL Project Web Site.* Available at http://www.saudavel.dpi.inpe.br. Last access Sept. 12 2003.

[11] Instituto Nacional de Pesquisas Espaciais. *TerraLib Project Web Site.* Available at http://www.terralib.org. Last access Sept. 12 2003.

[12] Kim, U.-N. A Historical Study on the Parcel Number and Numbering System in Korea. Proceedings of the International Conference of the International Federation of Surveyors. Available online at http://www.fig.net/figtree/pub/ proceedings/korea/full-papers/pdf/session10/ kim.pdf. Seoul, Korea, 2001. Last access May 08 2003.

[13] Lonely Planet Web Site. *Destination Seoul.* Available online at http:// www.lonelyplanet. com/destinations/north_east_asia/seoul. Last access May 05 2003.

[14] McCurley, K. S. Geospatial Mapping and Navigation on the Web. *Proceedings of the Tenth International World Wide Web Conference (WWW10)*, 221-229, Hong Kong, China, 2001.

[15] Navarro, G. A Guided Tour to Approximate String Matching. *ACM Computing Surveys*, march 2001.

[16] Pazinatto, E., Baptista, C. S., Miranda, R. A. V. GeoLocalizador: Um Sistema de Referência Espaço-Temporal Indireta Utilizando um SGBD Objeto-Relacional. *Proceedings of the IV Brazilian Symposium of GeoInformatics (GeoInfo 2002)*, Caxambu (MG), Brazil, 49-56, 2002. (in Portuguese).

[17] RuaVista Magazine. *The Numbering System of Buildings.* Available online at http://www.ruavista.com/numbering.htm. Last access May 05 2003.

[18] Teachkoreanz & Teachkorea-USA Program Web Site. *Postal Addresses in Korea.* Available online at http://www.teachkoreanz.com/living/ address.htm. Last access May 08 2003.

[19] United States Census Bureau. 108th CD Census 2000 TIGER/Line Files Technical Documentation. Available online at http://www.census.gov/geo /www/tiger/tgrcd108/tgr108cd.pdf. Washington, DC, 2000. Last access Mar 01 2003.

[20] WhereOnEarth Web Site. Available at http://www.whereonearth.com. Last access Sept. 15 2003.

[21] Yahoo Yellow Pages Web Site. Available at http://www.yp.yahoo.com. Last access Sept. 15 2003.