A GEOSPATIAL FOOTPRINT LIBRARY FOR VALIDATING VOLUNTEERED GEOGRAPHIC INFORMATION

by

Ahmad Omar Aburizaiza

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Earth Systems and Geoinformation Sciences

Committee:

	Dr. Matthew T. Rice, Dissertation Director
	Dr. Nigel Waters, Committee Member
	Dr. Chaowei Yang, Committee Member
	Dr. Andrew Crooks, Committee Member
	Dr. Anthony Stefanidis, Department Chairperson
	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
	Dr. Peggy Agouris, Dean, College of Science

Date: 11/22/17

Fall Semester 2017
George Mason University
Fairfax, VA

A Geospatial Footprint Library for Validating Volunteered Geographic Information

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Ahmad Omar Aburizaiza
Master of Science
Department of Geosciences at Idaho State University, Pocatello ID, 2009
Bachelor of Science
Department of Computer Science at University of the Pacific, Stockton CA, 2002

Director: Matthew T. Rice, Associate Professor
Department of Geography and Geoinformation Science

Fall Semester 2017
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my great parents Salwa and Omar, my loving wife Boshra, my three wonderful children Mohammad, Maria and Omar, and my academic advisor Dr. Rice.

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. My great parents and my loving wife for their prayers and support. Dr. Rice for his amazing help and support throughout my PhD journey in both academic and personal matters. Drs. Waters, Yang, and Crooks were of invaluable help. Finally, thanks go out to the Fenwick Library for providing a clean, quiet, and well-equipped repository in which to work.

TABLE OF CONTENTS

LIST OF TABLES

CLICK CUTLIST OF FIGURES

LIST OF ABBREVIATIONS

ABSTRACT

A GEOSPATIAL FOOTPRINT LIBRARY FOR VALIDATING VOLUNTEERED GEOGRAPHIC INFORMATION

Ahmad Omar Aburizaiza, Ph.D.

George Mason University, 2017

Dissertation Director: Dr. Matthew T. Rice

This research presents an innovative process of text-based Volunteered Geographic Information (VGI) validation by programmatically constructing geoparsed geospatial footprints. Place names are collected in a detailed, locally-centered gazetteer of landmarks, street names, and building names at George Mason University Fairfax Campus and Fairfax City area. The implementation uses distinct algorithms of various geometric computational operations based on the number and the type(s) of place names found in the text-based VGI entry, and the geospatial orientation of the place names. There are more than fifteen types of geospatial footprints defined in this dissertation. These geospatial footprints can be utilized as a reference library for other geoscientists to append newly developed types and/or to enhance the existing ones. This research addresses the evolution of the algorithms implemented to create the geospatial footprints and the technologies used to build the geospatial footprint library, and their role in

validating the position of VGI entries, which is a primary issue of concern for scientists and professionals that rely on geospatial data contributed by the public.

1. INTRODUCTION

Over the past decade, a significant new data source has emerged into the discipline of geosciences known as Volunteered Geographic Information (VGI). VGI has become an important research topic in different areas including emergency response, public safety, neighborhood maintenance, and driving navigation. It also became a part of many research projects at George Mason University (GMU). Since Dr. Michael Goodchild introduced the concept of VGI (Goodchild 2007a), geoscientists started debating and investigating its validity, reliability, and future use.

There are many different methodologies found in peer-reviewed literature to perform quality assessment and validate VGI entries. Such methodologies depend on the specific approach used to collect the VGI entries. One of these approaches to collect VGI is via text-based media, which frequently includes event and location descriptions. Text-based media can be any of the following sources: SMS, email, social media, or web-based form. Toponyms, or place names, can be found within text when people are describing a place, an emergency, or an event. Other text entries can have more than one place name to describe location better. In addition, specific wording in text can be used to describe the geospatial orientation between the place names mentioned in the text.

Geoparsing can be applied to extract all possible names in text. But the process of geoparsing needs a reference with which to check for place names. A gazetteer can be utilized as a reference of place names on a local, national, or global scale. Extracting only place names is not enough to encapsulate the geospatial footprint of the location described. One example would be a person describing an incident between two buildings. It might be that the two buildings are close by to each other or further apart. The person might also mention that the incident is closer to one of the buildings. Another example is a person mentioning an incident using a preposition to describe a geospatial proximity to a specific place. For instance, a volunteer explaining an incident's location as: "A curb crack near King Hall", vs. another entry: "A curb crack next to King Hal".

There are various descriptive scenarios of place names and their geospatial orientation found in text-based media. The George Mason University Geocrowdsourcing Testbed (GMU-GcT) is a team effort VGI system built at the Geography and Geoinformation Science (GGS) department at GMU, and has been the primary vehicle for a variety of published research, including the material in this dissertation. The GMU-GcT permits users to enter text-based reports as one of the options to collect VGI. Over a period of three years, the system collected more than 300 text-based reports with a variety of scenarios describing location. Other than the geospatial footprints library, the original gazetteer for this system was created by the author of this dissertation using PostgreSQL and PostGIS. In addition, the author created a web application using Mapbox.js (https://www.mapbox.com/mapbox.js) to navigate throughout reported obstacles by volunteers.

In the literature, there are systems that geoparse place names as point footprints and some geoparse them as very simple polygon footprints. An example of a geoparsing system is described from Tulane University referred to as the GEOLocate project (Bart & Rios 2015, Ellwood et al. 2016). The project description and details can be found in their website: http://www.museum.tulane.edu/geolocate. In addition to other examples, the GEOLocate system is explained in detail in the literature review, chapter 2.

The research presented in this dissertation has brought the focus of this work to the generation of a reference library of possible polygonal footprints, generated programmatically based on the different scenarios of describing location in text based VGI entries. In this dissertation, the evolution of the algorithms and the geospatial technologies utilized to build the polygon geometries are described thoroughly, as well as research showing the use of different positional validation approaches for geocrowdsourced data. The following sections of this introduction short summaries of the key research and technological topics, followed by the dissertation's objectives and a literature review.

1.1. VGI Definition

Dr. Michael Goodchild introduced the concept of VGI into literature in 2007 (Goodchild 2007a). VGI can be defined as the geospatial data created by non-professional geographers through Web 2.0. In Web 2.0, public users disseminate information using wikis, blogs, social networks, and open web maps. VGI has an impact on science in general, not only on the geosciences. It is shifting the data acquisition standard from top-down by mapping agencies and companies, to bottom-up through crowdsourcing geographic knowledge. There are many VGI web applications out there such as OpenStreetMap (OSM) (https://www.openstreetmap.org) and Wikimapia (http://wikimapia.org). There are also many VGI mobile applications, for instance Waze (https://www.waze.com), SeeClickFix (https://seeclickfix.com), and FourSquare (https://foursquare.com).

1.2. Public participation in Creating VGI

Public users have many reasons to generate VGI. Some volunteers contribute VGI purely for the benefit of other users. Other volunteers disseminate it to fulfill a course requirement at school (Coleman et al. 2009). VGI has the benefit of lower overall costs for mapping projects, because data is collected and disseminated through end-user computers, tablets, and smartphones. A significant factor in this lower cost was the end of Selective Availability of GPS through President Clinton's May 2000 Executive Order, and the subsequent price drop of GPS units and the integration of GPS in almost all smartphones and current location detection in computer browsers. With advancements in technology specifically offering user's location on smartphones, VGI dissemination is easier for the public users (Goodchild 2009). The public are attracted to input VGI to improve services around them. Also, they want to report emergency incidents that can harm their neighborhoods. In other scenarios, VGI is used to motivate political movements. Another reason the public are motivated to input VGI is pride of place. Self-promotion is also a motivation for end users to create VGI entries on the internet (Goodchild 2007a).

In addition to disseminating VGI entries in many applications, the public can validate VGI entries collected by other volunteers. This is referred to as aggregation of VGI. OSM is an evident example where contributors can edit entries previously created by other contributors. New edits can also be reviewed and edited by others (Dobson 2013). Google Map Maker is another known example that permits contributors to edit Google Maps' geographic and attribute data (Coleman 2013). The Google Map Maker was closed by Google but integrated into Google Maps. Qin et al. (2015) explained SeeClickFix (https://seeclickfix.com/) as an application that allows users to vote on problems reported in their neighborhoods to promote the importance of the reports. Waze (https://www.waze.com/) is an example where volunteers can edit traffic reports (Rice et al. 2015). Waze has the same functionality of voting on reported problems while driving on the road.

1.3. Common VGI Collection Approaches

There are three common approaches implemented in VGI applications to collect data. One of the most common ways is map digitization. In this approach, the volunteer draws a point, a line, or a polygon on a map, and then adds attributes to the digitized feature. OSM is a very good application example that uses the map digitization method to collect VGI entries.

The second approach is geotagging. Geotagging can be defined as assigning location data to multimedia sources such as images, videos, or audio. Such data mostly consist of latitude and longitude points. Flickr is famous of their geotagging technique in their media-based hosting (Van Laere et al. 2014).

The third approach is geoparsing and it is the primary methodological focus of this dissertation. Geoparsing can be defined as the process of extracting geospatial location from text. Figures 1, 2, and 3 illustrate three examples of the three mentioned approaches. Some VGI applications permit volunteers to use one approach while others allow more than one option such as the GMU-GcT System.

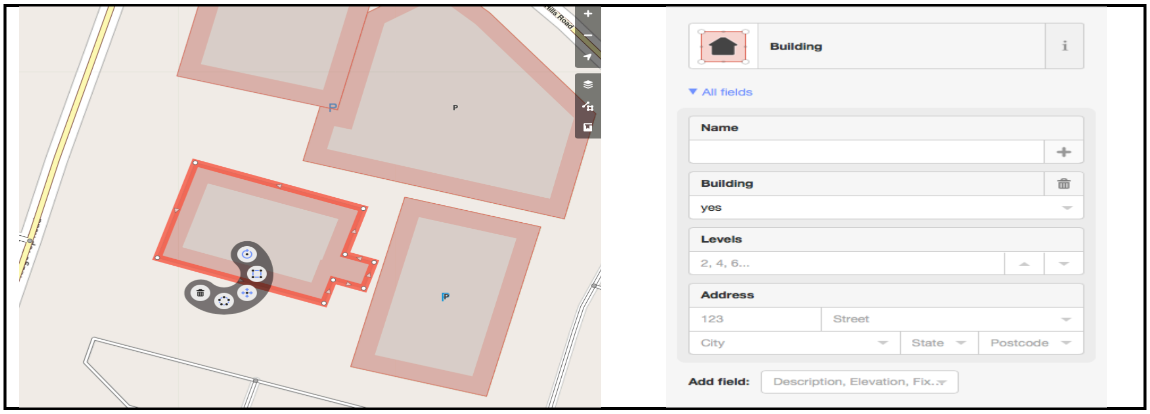

Figure 1 Digitizing a building in OSM by the author of this dissertation. The right-side window are attributes and the left-side is the map digitization.

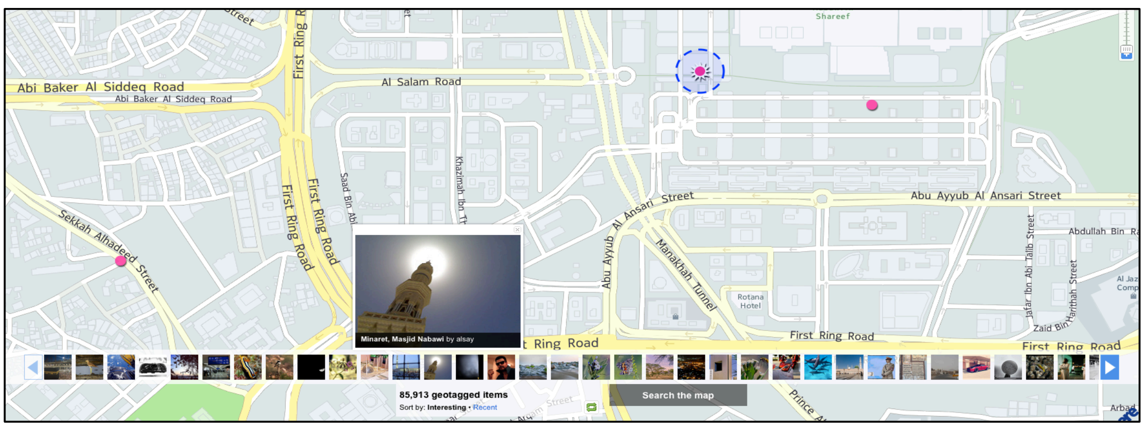

Figure 2 A geotagged image from Flickr by the author of this dissertation. The image location is the highlighted pink point.

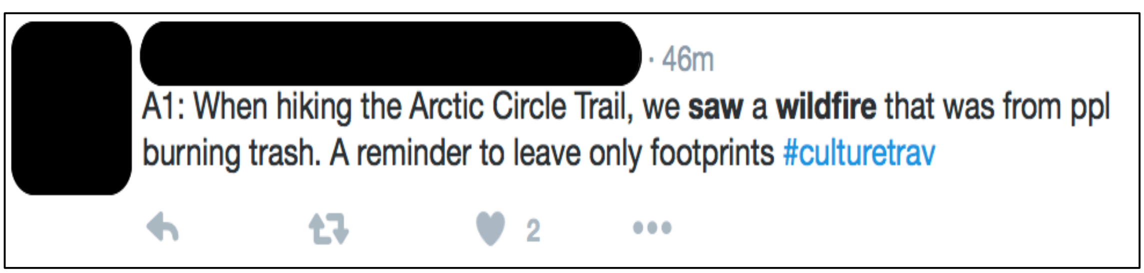

Figure 3 A VGI entry that can be geoparsed from Twitter.

1.4. Dynamic vs. Static VGI

Some VGI entries can be considered static or permanent e.g. town names, street names, and building names. Static VGI entries are common in collaborative mapping projects such as OSM and Wikimapia. Other VGI entries can be dynamic or in other words mutable. Such entries are considered to be temporary. For instance, fire incidents, crime scenes, temporary maintenance sites, broken manhole covers, street light outs, and curb cracks. Dynamic VGI entries are more relevant to emergency response, natural disaster, and driving navigational applications. Some examples of applications in these categories are Waze, SeeClickFix, and GMU-GcT. It is important to mention that in both static and dynamic VGI entries, the entries can be corrected or edited by other users to improve integrity and accuracy. This concept is supported by Linus's Law as discussed by Haklay et al. (2010) and Goodchild and Li (2012).

1.5. Gazetteers

Gazetteers are the dictionaries of geographical names associated with location and type (Hill 2009). They are considered as both a formal and an informal method of georeferencing. Formal gazetteers frequently originate from a formal, authoritative naming body, such as the US Board of Geographic Names, established in 1890. Other gazetteers, such as those used for georeferencing in social media, are created and curated by private companies, and include informal naming, abbreviations, and occasionally slang variants on placenames.

There are many examples of developed gazetteers in different countries. Some gazetteers are considered local while others are national or even global gazetteers. The North Carolina Gazetteer (https://www.ncpedia.org/gazetteer) is an example of a local gazetteer specific to the state. Gazetteer Scotland (http://www.geo.ed.ac.uk) is a national level gazetteer. It is the first thorough gazetteer for Scotland including details of tourist, industrial, and historical sites; in addition to family names and clans. The Columbia Gazetteer of the world (http://www.columbiagazetteer.org/) is a gazetteer of global place names with comprehensive details including place type, coordinates, and population.

Gazetteer entries contain three core elements recognized as the tuple N, F, T, where N is the name(s), F is footprint, and T is type. As noted, one gazetteer entry can have multiple names including formal, slang, abbreviations, and jargon names. It is a good practice to include all possible names in gazetteer entries. The temporal dimension can be very beneficial in gazetteer entries because some place names change overtime. For instance, Czechoslovakia was divided into Czech Republic and Slovakia. The proceedings report for the National Center for Geographic Information and Analysis (NCGIA) workshop on Digital Gazetteer Research and Practice, notes temporal gazetteer development as a fundamentally important area of research (Goodchild and Hill 2006). Extensions of digital gazetteer research, addressed in this NCGIA workshop i.e. temporal, have become important in this area of research.

1.6. The GMU-GcT System

GMU-GcT is a team-effort system developed to collect VGI entries at GMU Fairfax campus and Fairfax City area. In the many iterations of the GMU-GcT, volunteers were permitted to insert at least three components for each VGI entry: an image, an explanatory text, and a digitized point. Some volunteers do not include all three components in their VGI entries. The system is still functional and to date has collected more than 300 VGI reports about dynamic obstacles in the local area.

This research started after finding limitations of electronic tactile maps. Tactile maps are used by blind and vision-impaired people to help them navigate in their neighborhoods. These maps explain the surrounding geospatial features via the sense of touch. The electronic tactile maps interpret places orientation, using special mouse devices that signal terrain texture while panning the map.

Electronic tactile maps are very beneficial for blind and vision-impaired people, but the problem is that temporary obstacles are not plotted in tactile maps (Rice et al. 2012a). Such temporary obstacles can be very dangerous for blind and vision-impaired people e.g. maintenance sites, potholes, and steep stairways. An ideal solution is to utilize Volunteered Geographic Information (VGI) as a geospatial source for the electronic tactile maps. The GMU-GcT builds a layer of crowdsourced information, georeferenced with user-entered points and location text, that can be added to any electronic map resource.

1.7 Dissertation Objective

In order to provide the highest quality of information, VGI entries must be validated to ensure their integrity. Validating the VGI entries collected using the GMU-GcT system diverged into different concentrated research problems. Some team members researched general quality assessment for VGI (Qin 2017), while others explored the validation process through the image components (Rice et al. 2018). Other team members explore the validation through investigating the socially-moderated location component

(Rice 2015). My focal research for this dissertation was to validate the text-based VGI location component.

The dissertation objective is to explore the possibility of creating more accurate geospatial footprints explaining the location of a reported incident in a text-based VGI entry. There are various scenarios of spatial orientation of place names found in text-based VGI such as an intersection of two lines, proximity to a place, bearing description in relation to a place, and so on. This leads to creating a library of different cases of spatial orientation. Similar trials are found and discussed in the literature review chapter but with lower accuracy and higher ambiguity, i.e. a city bounding box boundary or a town centroid point. Another contribution of this dissertation is providing the implementation code of the geospatial footprint library in Github for geoscientists to explore, configure, or add new scenarios.

The methodology used in this work is to geoparse the explanatory location text and generate polygonal footprints to cover the potential location of the temporary obstacle using the geospatial orientation of the geoparsed place names. The polygonal footprints reference library developed in this dissertation is based on various scenarios of the number of place names found in text and their geospatial orientation. One scenario example of this library is illustrated in figure 4. In this scenario, the text is referring to a street segment between two intersecting streets.

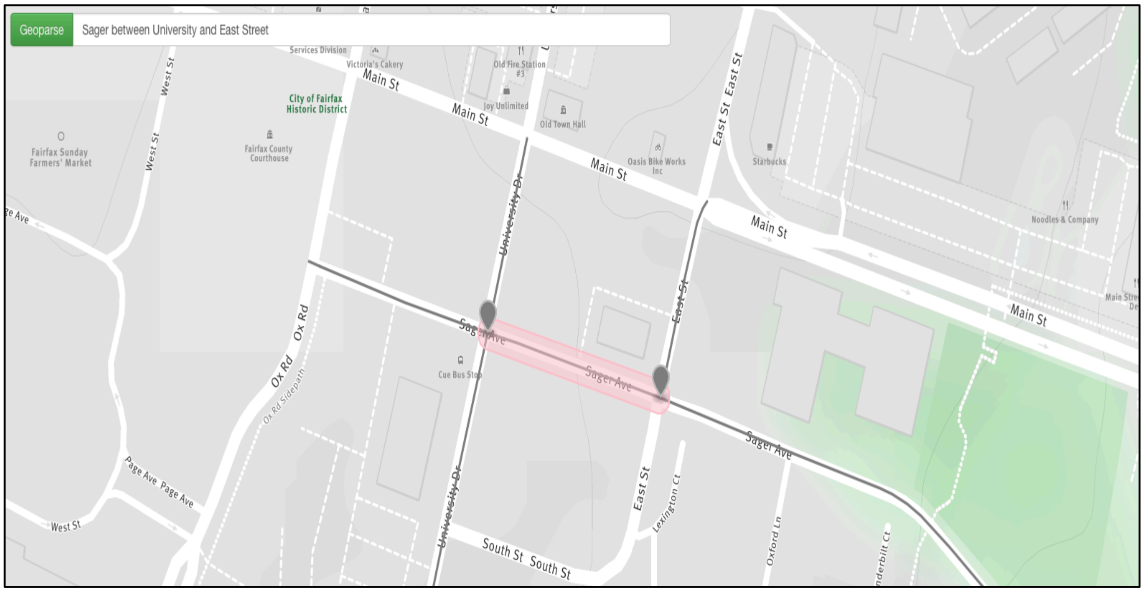

Figure 4 One scenario in the dissertation's footprint library: an obstacle location described as on Sager between University Drive and East Street.

As mentioned, there are similar systems to generate footprints based on geoparsing but the footprints are ambiguous. These systems are discussed in chapter 2. The geospatial footprints library in this dissertation generates different polygonal structures for the footprints based on the location description found in text-based media, and is demonstrated to be more useful as a form of position validation compared to other simple geospatial footprints found in literature.

2. LITERATURE REVIEW

2.1. VGI and Its Motives

The term Volunteered Geographic Information, or VGI, was termed by Michael Goodchild in 2007 (Goodchild 2007a), stemming from previous work on the implications of geographically-distributed, web-enabled information sharing communities (Goodchild et al. 2005). It means the creation of geographic information by neogeographers or non-professional geographers, enabled with the revolution of Web 2.0. In Web 2.0, public users can contribute information through wikis, blogs, social networks, and open web maps. There are many examples of VGI applications including Wikimapia and OSM. Kuhn (2007) published a very good example of VGI, emphasizing that GI is like a shower's water and VGI is the hot water, in other words VGI is an appealing research topic for geoscientists. Sui and Goodchild (2011) stated that VGI is the language of citizens to express things and events surrounding them. VGI can be thought of as bottom-up data collecting (Sui et al. 2013). VGI is shifting the data acquisition standard from top-down via mapping agencies and companies, to bottom-up through crowdsourcing geographic knowledge (Jiang 2013). Elwood (2008b) emphasized that VGI is receiving more attention from geographers and other scholars as potential sources for research. Rice et al. (2012b) provide a summary of the general data production changes initiated by VGI such as OSM, NGA's PLACES program, and USGS National Map Corps.

Public users have many reasons to generate VGI. VGI has the benefit of lower cost of production, and can be disseminated by computers, tablets, and mobile phones owned and maintained by end-users and volunteers (Goodchild and Glennon 2010). Another reason for public participation in VGI is the price drop of GPS units and their integration with smartphones (Zook et al. 2010). Users are attracted to input VGI to improve services around them. VGI has positive motivations from the public such as pride of place and allowing friends and family to search for their entries (Chow 2013). Users are attracted to share their knowledge and opinions whether they are amateurs, professionals, or hobbyists (Howe 2006). People are encouraged to disseminate data in Web 2.0 to create self-recognition and self-pride. In addition, people are motivated to contribute VGI due to altruism and societal benefit (Goodchild 2007; Coleman 2010). Another reason on why public users contribute VGI is to motivate political movements (Flanagin and Metzger 2008) and initiate change and awareness.

2.2. VGI Benefits

A comparison of authoritative data to VGI shows that authoritative data are slow in updates while VGI is much faster (Sui and Goodchild 2011). Rice et al. (2012b) review the production of geospatial data, contrasting the traditional top-down approaches with bottom-up and hybrid approaches, describing the likely future as being a hybrid approach where the public contributes attribute information and error checking. This is being explored by the US Geological Survey. Goodchild and Glennon (2010) state that authoritative data mapping in remote areas takes much longer time than via VGI collection. They discuss the utilization of VGI in the four major wildfire events in Santa

Barbara, CA between 2007-2009. A lesson learned from their study is that VGI data production is much faster than authoritative data, and therefore more valuable in time-critical applications, even though the information quality and precision may be lower.

Linus's Law states that with enough eyeballs, all bugs are shallow (Heipke 2010). More eyes on data means less chance of errors (Sui and Goodchild 2011). VGI can reach quality levels exceeding authoritative data. Based on Tobler's First Law of Geography, everything is related to everything else, but near things are more related than distant things (Waters 2017). In the same context, VGI is mostly similar in same and close by locations (Goodchild and Glennon 2010), and local geographic expertise (innate knowledge of nearby locations) is a benefit favoring VGI over authoritative data.

Feick and Roche (2013) explain that government acceptance of VGI data input represents democracy in a civil society. The authors emphasized that citizens are partners in the co-production of decision making. Moreover, VGI fills the gaps within authoritative data. (Goodchild and Glennon 2010) mention that authoritative emergency responders cannot be everywhere all the time but the crowd is always available. In 2009, there were 27 mapping services collecting VGI about Santa Barbara, CA fires, and the most popular one of them collected 600,000 hits within a period of 13 days. Zook et al.'s (2010) discussion of the interplay between the government, business, open source organizations, and the public, during the emergency response to the 2010 Haitian Earthquake is a useful point of reference, and together with discussion about the evolution of data production in Rice (2012b), presents a likely view of the future use of VGI by government mapping agencies.

VGI is created and updated more rapidly than traditional, authoritative data. Li and Goodchild (2012) state that authoritative data are not adequate in emergency crisis since it has long production cycles. VGI might be the only model of GI that is applicable for emergency response and crisis management. Its value cannot be estimated since it can save human lives (Feick and Roche 2013). Map revision cycles in authoritative agencies are usually slow which gives the advantage of VGI to fill in the gaps (Coleman et al. 2010). Moreover, Feick and Roche (2013) mentioned that the “report a problem” features in Google Maps, Garmin, and TomTom enhanced their GI data through VGI.

Some would argue that VGI is not accurate and is low-quality, or at a minimum, suggest that accuracy and reliability are the biggest challenges in the use of VGI. This is sentiment tends to be domain specific, and is not always the primary concern for individuals working in emergency response. Zook et al. (2010) explain that for VGI used in emergency response, responders do not need high quality and highly accurate data to help people. Some incidents need very urgent responses and cannot wait for highly accurate data, or a detailed quality assessment workflow. The authors articulate how utilizing VGI during the earthquake in Haiti saved many people’s lives.

Local knowledge is very important in emergency response and management activities. No one would know their areas or neighborhoods better than local people. In the Haiti’s earthquake, Haitians are more familiar with their country and they can better explain their locations than official emergency responders and humanitarians. Goodchild (2007) claimed that drivers would trust directions from local amateur residents more than professional geographers who do not live locally; this means that the local amateurs are considered professionals in such scenarios. This can be referred to as local knowledge (Sui and Goodchild 2011). Johnson and Sieber (2013) state that citizens are usually closer to phenomena than authoritative employees and commonly more familiar with their areas. Local knowledge, such as that embedded in VGI, helps in better decision-making. Hecht & Moxley (2009) examined the First Law of Geography through the crowdsourced contributions of Wikipedia. In their study, they found that nearby Wikipedia entities have a higher probability of similarities than farther entities, even some far apart entities have some relationships.

2.3. VGI Validation

It is very difficult to claim that any geospatial data is 100% accurate either created by professional or non-professional geographers. But still VGI introduces many concerns about accuracy and quality. Waters (2009) presents several valid critiques of VGI and related systems including the reliability of the volunteer in regard to his/her reputation of past records, locality of the volunteer, and the volunteer's morality.

Several methods have been developed to increase its quality and accuracy. Starting with VGI aggregation, (Sui and Goodchild 2011) state that synthesizing VGI entries can be used to manage credibility and uncertainty. VGI can be synthesized to resolve incompleteness by using multiple entries of a specific location to complement each other. VGI can also be utilized to increase location accuracy. Passive and repetitive VGI in similar locations increase accuracy of position (Coleman 2013). As more users contribute and correct each other's data, data is considered more credible. Agreement to a specific entry should increase the level of credibility (Flanagin and Metzger 2008).

The ability to update VGI entries among the volunteer community ensures strengthening the quality and accuracy. VGI applications enable vetting, editing, and correcting entries also by end-users, such as happens in Wikimapia (Goodchild 2007). Local expertise, frequently present in VGI, adds accuracy comparable to local authoritative data, and adds information about change that might not be known to producers of authoritative data (Goodchild 2008). When users edit VGI entries, errors are commonly reduced over time, though the dynamic (discussed in Haklay et al. 2010) is more complex, with long-term edits of some features resulting in edit-reversion loops. This improvement of VGI quality over time is referred to as aggregation of VGI by Dobson 2013). In addition, VGI applications such as OpenStreetMap and Google Maps Maker additional error checking workflows utilizing specialized moderators to review new or updated entries. Entries will be pending until a moderator reviews and approves it. This approach, based on moderator expertise, works well in many scenarios, including the GMU-GcT, and has been reviewed comprehensively by Goodchild and Li (2012), and by the GMU-GcT Research Team: Rice (2015), Rice et al. (2014, 2015, 2016) Qin et al. (2015, 2016), and Qin (2017).

Identifying the time and space components of a VGI entry helps in validating the integrity of the data. For instance, a user from a certain country correcting a street network in another country is questionable (Coleman et al. 2010). By looking at the user's profile, if the user is living within the areal extent of the feature being modified, there is a higher probability that the entry will be accurate.

Creating a coloring scheme based on editing history is helpful to determine the frequency of data editing and correction and hence data credibility (Flanagin and Metzger 2008). Other applications use an attribute of values between 0 and 1 to determine if the location is certain or not. This is known as fuzzy-set approach. Other approaches use grayscale to represent the certainty level. Another method is using a point with a radius where the point is in the center and the radius length represents the level of certainty and the circular area defines the footprint of all possible locations (Hill 2009). Google Maps plot uncertain locations with a transparent blue circle symbol while features with certain locations are plotted with a pin symbol.

Comparing VGI data to authoritative sources is vital (Coleman et al. 2010), and is the basis for most of the OpenStreetMap (OSM) quality assessment studies, reviewed by Ruitton-Allinieu (2011), Rice (2015). Haklay (2010) compared OSM coverage in London to authoritative data, and found that road features in OSM are generally within 6 meters of their true location. Only roads and streets were used in this comparison study. In addition to positional accuracy, completeness was compared as well. First regarding positional accuracy, the testing method used buffers around motorways to determine the shift in OSM roads. The test resulted in 80% overlap with OSM data. Another test was between two different road types and it resulted of 88% overlap. Secondly for completeness, 93% of England coverage was used in the study. The first method was calculating the difference of the sums of road lengths and the second was using specific SQL queries. The result was 69% OSM completeness.

Girres and Touya (2010) applied a similar study of the French OSM data. They also found out that positional accuracy is also generally within 6 meters. In their study, they also looked at broader data quality parameters such as lineage, temporal consistency, attribute accuracy, and semantics. Because of the broad coverage of different key aspects of data quality, the Girres and Touya research paper is an exemplar and model for this work.

Arsanjani et al. (2015) evaluated the accuracy of OSM data created in 4 German cities: Berlin, Frankfurt, Hamburg, and Munich, in comparison to the Global Monitoring for Environment and Security Urban Atlas (GMESUA) data. GMESUA datasets are high-resolution land use maps in various European cities. The evaluation considered the following five criteria: thematic accuracy, positional accuracy, temporal accuracy, logical consistency, and data completeness. The overall accuracy calculated of the OSM data is 75.9% for Berlin, 76.5% for Frankfurt, 63.9% for Hamburg, and 67.1% for Munich. And the GMESUA over all accuracy is barely over 90%.

2.4. Gazetteers

Hill (2009) defines gazetteers as dictionaries of geographic place names, their locations, and their types. Gazetteers entries contain three core elements recognized as the tuple N, F, T, where N is name(s), F is footprint(s), and T is type(s). Time is an essential element that should be implemented in gazetteers entries since place names can change over time (Goodchild and Hill 2008).

Guptill (2006) provides a useful explanation on the historical events leading to the development of the US government naming authorities and the subsequent development of gazetteers. In 1890, president Benjamin Harrison established the US Board of Geographic Names because of confusion and controversy associated with geographic names. In 1906, president Theodore Roosevelt ordered the implementation and standardization of geographic naming for federal use. In 1947, president Truman signed the establishment of today's organization of the US Board of Geographic Names. To implement a standard tool for searching formal geographic names, the United States Geological Survey (USGS) has built a GNIS "Geographic Names Information System" for the US Board of Geographic Names. It contains physical, cultural, and historical names of 1.9 million features in the US. Other attributes of the entries are different spellings, feature classification, and geometric boundaries. Features in GNIS are associated with location and not with spatial extent. Other geodatabases are associated with GNIS such as NHD "National Hydrography Dataset". When a user enters a new entry in the NHD, a name entry in GNIS should match to fulfill the policy requirements of the US Board of Geographic Names.

Interoperability is very important to sync different types of gazetteers for better searching mechanisms. There are various types of gazetteers such as official, local, and historical. In additions, gazetteers with authoritative names cannot be used to understand or reference slang and vernacular names. Also, gazetteers with generalized boundaries can be less useful (Goodchild and Hill 2008). There are different types of gazetteers, not a single one is comprehensive for every condition and every use. Some are local and others are global. Interoperability between them is very beneficial for GIS applications (Johnson and Sieber 2013). Some queries might need interoperability with multilingual gazetteers as well (Goodchild and Hill 2008). Bekisz (2015) and Cave (2015) address these multi-lingual problems in masters theses, while McDermott (2017) addresses a variety of issues related to temporal inconsistency and relevance in gazetteer-based geoparsing of travel narratives in a recent doctoral dissertation.

2.5. Geoparsing footprints

Beaman & Conn (2003) discussed a web service prototype to geoparse biological collections data. Their system can access a gazetteer of 330,000 Malesian place names. In the paper, they represented the automated geoparsed locations as points. At the University of Edinburgh, a geoparsing system was developed as well. The system has two components, the first component collects place names from a given text as an XML file, and the second component search the place names in three different gazetteers. The place names are digitized as points on the map (Tobin et al. 2010; Alex et al. 2015). (Horák et al. 2011) demonstrated a geoparsing system to extract place names from media news. Again, the geoparsed locations are only points on the map.

Another example from Tulane University is the GEOLocate project (Bart & Rios 2015, Ellwood et al. 2016), found in: http://www.museum.tulane.edu/geolocate. There are two versions of the project: desktop and web. The GEOLocate system geoparses place names as points in most cases and in polygons footprints in less cases. The polygon footprints highlight some geographic feature such as lakes and city boundaries. The system is also capable of georeferencing a point based on a spatial orientation description explained with directions. But the boundary of such footprint is big in area and not trimmed to the real possible location.

McDermott (2017), a research collaborator, developed geoparsing algorithms using frequency analysis and geographic clustering to create and plot the travel trajectories for a travel journal, and in the process, developed methods for aggregating temporally-connected point locations into larger clusters, which were represented with a polygon. While the scope of this research was narrow and specific to travel journals, the use of complex geospatial representations for collections of geoparsed toponyms and points-of-interest names is notable.

2.6. Dissertation Objective Synthesis

To summarize the composed literature review, there are different approaches to collect VGI entries from the World Wide Web. VGI is disseminated by the public via various methodologies based on the system or application they are using. There are systems that permit the process through map digitization. OSM and WikiMapia are good example of systems using this method. The volunteers can use such systems to digitize points, lines, or polygons. Geotagging is another method to collect VGI entries. Some famous applications that utilize geotagging are Flickr, and Instagram. In these applications, geolocation is attached to photos and videos, usually as a simple x, y point location. Another method to gather VGI entries is through text media sourced from SMS, email, or social media.

Geoscientists have concerns regarding VGI's accuracy. In literature, there are many scholars studying different techniques to validate the integrity and accuracy of VGI. Text-based VGI need more than one component to implement its validity. Creating geospatial footprints based on geoparsing toponyms found in gazetteers is a common approach.

The literature review in this dissertation lists examples of systems using this common approach of creating geospatial footprints through geoparsing and gazetteers. The geospatial footprints in these systems are mostly created as simple geometries. Some of these applications will create a georeferenced point feature based on a toponym found in text e.g. the centroid of a city. Other systems highlight simple polygons as bounding boxes, covering more area than the location explicitly explained in text.

Identifying the boundaries of the geospatial footprints is very challenging due to many reasons. One of the reasons for this challenge is having more than one toponym in a single text entry, a problem discussed by Leidner (2017) and McDermott (2017). Leidner suggests that toponymic resolution and errors related to disambiguation of toponyms, following a named entity recognition process, is a major challenge in finding the most likely referenced item in a gazetteer. Another difficulty is finding details about the spatial relation between multiple toponyms. Directional words also can add further refinement in creating the geometry of the geospatial footprint, as noted in Moncla et al. (2014) paper addressing hiking descriptions, where directional descriptions are key. As noted in the final summary and future directions for this dissertation research (section 7.2), directional words and spatial prepositions will be a major focus as the research workflows are implemented in a municipal messaging and alert system, where directional words are common. One more reason is having proximity terms related to toponyms explained by the volunteer. Finally, ambiguity is a major dilemma in creating a geospatial footprint. One example of an ambiguous location is having only one toponym of a major highway (or any other long linear object) in a text entry without further details about intersection points or segments of interest.

Based on the mentioned challenges and difficulties of constructing geospatial footprints through geoparsing and gazetteers, this dissertation's focal goal is to build a library to automatically generate geospatial footprints with simple, complicated, or ambiguous geometries. In addition, the library will be available on Github for other geoscientists to contribute and use. This library can be utilized by geoscientists as a reference to extend and/or to improve the algorithms applied for each scenario. Fifteen different types of geospatial footprints have been implemented in the library. Each geospatial footprint type has its own distinct algorithm based on the number of toponyms, types of toponyms, spatial orientation, and spatial relationship between the toponyms.

2.7. Contributions to Research Papers

The following chapters 3, 4, 5, and 6 of this dissertation are from peer-reviewed published papers in different journals, and were presented in well-known geoscience conferences. The author was the principal or a major research contributor for each of these papers, along with Dr. M. Rice at the GMU-GcT research team. Chapter 7 is the conclusion of the scientific findings of this dissertation, and the proposed future work from this research.

Chapter 3 discusses the first draft of the gazetteer structure and the first algorithmic phase of the geospatial footprints. The system was initiated as a desktop GIS application built with MapWindow, an open source desktop GIS framework. The footprints at phase one were simply bounding boxes covering all possible names found in a text message. This chapter was written jointly by the author, Dr. M. Rice, and Curt Hammill, and was published in conjunction with the 25th International Cartographic Conference and the edited volume of selected papers, Advances in Cartography and GIScience. All of the computational work was implemented by the author, but the writing was done jointly. The conceptual design of the gazetteer was originally Hammill's, but later was modified and expanded by the author. Hammill was graduating and this was one of his major contributions as an MS student, so he was listed above the author in the authorship credits, though their contributions were approximately equal in importance. Since Dr. Rice did the majority of the editing and communicating with the publisher, he was listed first. This chapter is an important first incarnation of the computational framework of the dissertation, and forms an important point of reference for later developments.

Chapter 4 presents the second phase of this research. In this phase, the footprints were trimmed geometrically to minimize location error found in the location text message. The system in this phase migrated from the desktop platform onto the Web. The development framework consisted of HTML, CSS, JavaScript, PHP, KML PostGIS/PostgreSQL, and Google Maps API. In addition, the gazetteer had expanded with more toponyms including some foreign names of places and landmarks at GMU campus. This chapter was written jointly by the author and Dr. M. Rice, with the author doing all of the computational work, and a number of minor contributors who helped compile data and add material to the literature review. This paper was produced from a conference proceedings paper from FOSS4G 2011 (Free and Open Source Software for Geospatial), that the author attended few months prior. The paper was selected from the academic committee at the conference for review and publication by Transactions in GIS, based on the authors presentation at the conference.

Chapter 5 is the third phase of this research. In this phase, new polygonal geometries were implemented such as street intersections and a buffered street segment between two intersections. The development framework in this phase was advanced with the integration of HTML5, CSS3, JavaScript, jQuery, CartoDB.js, and PostGIS/PostgreSQL. This chapter was jointly written by the author, R. Rice, and Dr. M. Rice, but the computational aspects of the research presented was the author's. This paper is the most difficult to untangle, in terms of the allocation of credit. The author was a major contributor to the work with gazetteers and the web application that generates footprints for VGI position validation. Recent GMU Doctoral graduate student Han Qin developed the geocrowdsourcing testbed and quality assessment procedures to which this research was applied. The reason for inclusion of the paper in this dissertation was the nature of the development of the gazetteer-based geoparsing and its use in moderator-based position validation.

Chapter 6 is about the final phase of the geospatial footprint library. The system is now capable of constructing fifteen different geospatial footprints. The development framework was advanced again to the mashup of HTML5, CSS3, JavaScript, jQuery, AJAX, Leaflet, MapBox, Turf.js, and GeoJSON. This chapter was nearly completely the author's work, though co-author Rice helped write the introduction and literature review.

The paper was presented initially at FOSS4G 2015 in Seoul, South Korea, and selected for post-conference publication by the Spatial Information Research journal. The author made the FOSS4G presentation, wrote the conference abstract and paper, and acted as the communicating author.

3. INTEGRATING USER-CONTRIBUTED GEOSPATIAL DATA WITH ASSISTIVE GEOTECHNOLOGY USING A LOCALIZED GAZETTEER

Abstract

We present a methodology for using cartographic-based processes to alert the vision-impaired as they navigate through areas with transitory hazards. The focus of this methodology is the use of gazetteer-based georeferencing to integrate existing local cartographic resources with user-contributed geospatial data. User-contributed geospatial data is of high interest because it leverages local geographic expertise and offers significant advantages in dealing with hazard information in real-time. For blind and vision-impaired people, information about transitory hazards encountered while navigating through a public environment can be contributed by end-users in the same public environment, and quickly integrated into existing cartographic resources. For this project, we build collections of user-contributed geospatial updates from email, voice communication, text messages, and social networks. Other necessary technologies for this project include text-to-voice software, global positioning devices, and the wireless Internet. The methodology described in this paper can deliver usable, cautionary reports of hazards, obstacles, or other time-variable concerns along a pedestrian network. Using the George Mason University campus as a study area, this paper describes how transitory events can be presented in usable form to a vision-impaired pedestrian within a usably short period of time after the event is reported. Buildings and other destinations of

interest can be registered in a robust, eXtensible Markup Language (XML)-based, localized gazetteer. Walking networks, parking lots, roads, and landmarks are mapped as vector-based digital information. Any events or changes to the base map, whether planned and disseminated through official channels or reported by end-users, can be linked to a location in the network as established by the attributes cataloged in the localized gazetteer, and presented on an existing base map or in an assistive technology environment. For mobile applications, a vision-impaired pedestrian with a Geographic Information System (GIS) and a Global Positioning System (GPS)-enabled assistive device can receive an alert or warning about proximity to reported obstacles. This warning might include other information, such as alternative paths and relative directions to proceed, also referenced through the localized gazetteer. This research provides insight into challenges associated with integrating user-contributed geospatial information into a comprehensive system for us e by the blind or vision-impaired.

Published in 2011 at the Advances in Cartography and GIScience, volume 1 (279-291). Authors Rice, M.; Hammill, W.; Aburizaiza, A.; Schwarz, S.; and Jacobson, D.

3.1. Background and Objectives

The International Cartographic Association (ICA) Commission on Maps and Graphics for the Blind and Vision-Impaired People has been an important outlet for publication associated with tactile map production and use. Many papers focus on design, production, and evaluation (i.e., Tatham 1991, Eriksson 2001, and Perkins 2001) while a few others focus more broadly on issues such as standardization (Tatham 2001) and assistive geotechnology (Coulson et al. 1991). In Coulson et al., the authors emphasize that existing geotechnology can be used to automate the production of tactile maps. This is an important starting point, because assistive environments that can quickly and automatically incorporate additions and changes to an environment are particularly useful for individuals navigating through space. Transitory obstacles that present a hazard or barrier to navigation are a distinct challenge, because they generally appear as unplanned events and cannot generally be depicted with a standard tactile map environment, where updates may take several days or perhaps weeks. Because hazards and obstacles appear and change often, it is important to use sources of data that are frequently contributed and have temporal relevance.

Dr. Michael Goodchild, in his 2009 keynote address to the Association of American Geographers (AAG), described the value of user-contributed or volunteered geographic information (VGI), citing two of the most important aspects 1) its leveraging local geographic expertise for wider purposes and 2) its temporal relevance (2009a). Goodchild developed the concept of user-contributed geographic information and identified it as an important trend in environments with large, active end-user communities (2009b). Although typical end-users are not trained cartographers or tactile map experts, the geographical expertise and temporal relevance of contributed geospatial information makes it extremely valuable. Based on the ideas of Coulson et al. (1991) and Goodchild (2009), we are creating a methodology to incorporate real-time user-contributed geospatial information into existing accessibility-oriented mapping systems. The functional centerpiece of this methodology is the use of a localized gazetteer, which allows mapping of placename-based descriptions into geographically referenced map locations. These georeferenced end-user contributions can be incorporated into existing mapping systems oriented towards blind and visually-impaired persons, and triggered by proximity as a blind or vision-impaired end-user navigates through a mapped location. Our methodology focuses on near real-time incorporation of environmental obstacles or hazards, but recognizes that the approach can be useful in joining a variety of user-contributed information to existing mapping systems.

3.2. Approach and Methods

Our methodology for delivering user-contributed geographic information to blind and vision impaired individuals follows a process shown in Figure 5. This flow of information starts with observations in a geographic environment. These observations are time-stamped and contributed through voice communication, email, text message, or through social media updates. The observations are analyzed for geographic content, generally through geoparsing for local placenames, distances, spatial prepositions, and temporal information. The observations are then matched to entries in a localized gazetteer developed for the university campus with features associated to a geographic location. The location, consisting of a georeferenced footprint connected to a placename, is plotted on a network or integrated into a map display. The location is available as a text-to-voice prompt on an accessibility map of the local campus. The location can be explored using standard mouse interaction from a fixed location or can be triggered using a location sensitive application that issues a proximity alert and a text-to-voice prompt. This allows for the display of transient obstacles or hazards on a map, and for those obstacles or hazards to be communicated to the blind and visually-impaired using assistive mapping interfaces described by Golledge et al. (2006).

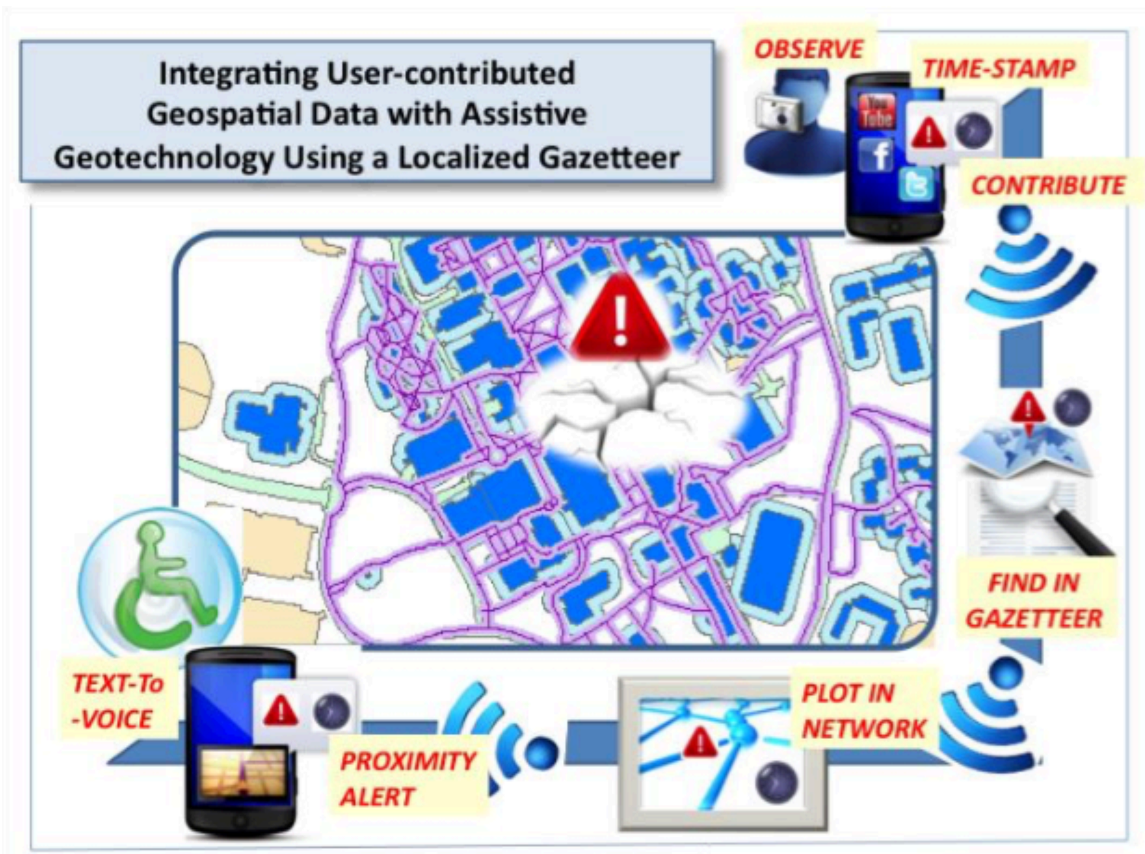

Figure 5 Process for integrating user-contributed geospatial data with assistive geotechnology using a localized gazetteer (Rice et al. 2011, 282)

Figure 5 captures a specific scenario in which this functionality might find utility. A pedestrian in the environs of campus might discover a temporary obstacle and desire to share this easily with other pedestrians. Possible obstacles might range from sidewalk construction to a large student rally. With a basic text message using commonplace and natural descriptions (e.g., “near the Engineering building”) and perhaps a camera image, he could register the obstacle. That information is time-stamped and compared to a localized gazetteer for locating the obstacle in GIS-understandable geographic space. Alternative paths are identified using standard network path algorithms, as reviewed in Waters (1999), who discusses the significance and role of transportation GIS within a variety application areas, particularly those that include automated routing, navigation, and dispatch services. This synthesized information is made available on a subscribed syndication for any GIS/GPS-equipped users. Other technologies for text-to-voice can specifically alert and inform the blind and visually-impaired.

Our localized gazetteer forms the linkage between user-contributed observations that generally use placenames and existing cartographic resources contained primarily in ArcGIS. The result is a map-based display system that contains both existing geospatial data and updates contributed by end-users.

Observations contributed by volunteers or end-users take many forms, but generally end up translated into text, which is geoparsed for relevant placenames, directions, distances, and geographically-relevant prepositions and prepositional phrases, such as ‘nearby’, ‘next to’, and ‘on top of’. Observations can also be obtained using a technology such as GeoRSS (Geographically Encoded Objects for Really Simple Syndication)[1]. GeoRSS is a process of extracting geographical information, such as latitude and longitude, from any geographically tagged feed. These feeds are useful when one wants to keep track of regularly changing information such as news and traffic conditions. There are two encodings of GeoRSS. GeoRSS-Simple is the simplified version of encoding whereas GeoRSS Geographic Markup Language (GML) has more features, including the availability of more than one specific coordinate system. In order to extract geographical data from social network feeds such as Twitter and Facebook, the locational information has to be in the feed itself. This requires third party software since both aforementioned applications are still developing geo-tagging on their formats. Much controversy exists over privacy concerns as well as location accuracy, but GeoRSS and related geotechnologies form an important aspect of the observation collection process, and we have developed some procedures for masking sensitive or private information associated with end-users that contribute information. Kwan et al. (2004) describe many of the more important geoprivacy concerns and evaluate the effectiveness of geomasking processes.

After observations are collected, a list of relevant placenames is obtained from our localized gazetteer, which contains a comprehensive list of official campus feature names, colloquial variants of placenames, abbreviations, coded placenames, and common foreign-language variants of placenames. The observations are then associated with

---

[1] See http://www.georss.org/ for a full description of the GeoRSS format and specification, which is a lightweight method for encoding geographic information within other information feeds.

spatial footprints of the features from the localized gazetteer and plotted on a map.

The Geoparsing software tool being developed for this project uses VB.NET together with MapWindow GIS OCX (Object Linking and Embedding Control eXtension) control. MapWindow GIS is an open source GIS application under the Mozilla Public License, started at Utah State University, Logan, Utah (Ames et al. 2008). Since MapWindow GIS is an open source software package, GIS programmers are permitted to configure, use, and improve the software code for their specific needs. There are two GIS programming paradigms in MapWindow GIS: standalone applications and plug-ins (Aburizaiza and Ames 2009). For this project, the standalone development approach was utilized. A screenshot of the Geoparsing tool developed with MapWindow GIS is shown in Figure 6.

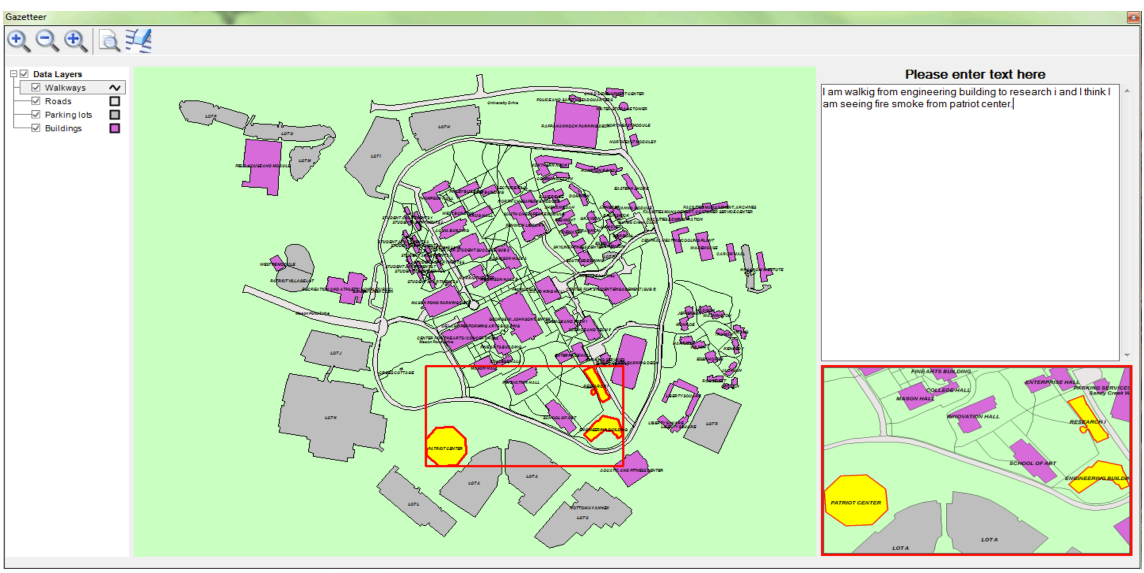

Figure 6 Screen capture from Geoparsing Tool showing a geographic selection from a text entry as keyed to the localized gazetteer (Rice et al. 2011, 284)

The geoparsing tool accepts text entries and generates a list of all possible strings associated with placenames in a localized gazetteer. Concerns regarding insufficient or inaccurate text-entry, such as character case, special characters, and duplicates are accounted for in the processing used in the tool. Intensive testing for text entry errors and special cases has significantly improved the association rate of text entries to locations in the localized gazetteer. Geoparsing of the data pinpoints the geographical feature placenames with location coordinates for further spatial analysis (Hill 2008). The tool then identifies those features on a geographically referenced map, displayed using the MapWindow OCX control. Added functions for the tool include capturing the path between the spatial features identified in the localized gazetteer, and parsing useful distances and directions from texted propositions, modifiers and cardinal directions, such as near, towards, or north.

As mentioned, we use a localized gazetteer to match text-based or voice-based descriptions containing placenames associated with observations to a spatial footprint. Our localized gazetteer is built using a data model, structured on concepts from Hill (2006). The gazetteer data model contains items of primary interest to blind individuals navigating through the local environment; namely, buildings, roads, walkways, parking lots, and landmarks. Table 1 and Table 2 provide summary descriptions of our gazetteer data model for buildings. Table 2contains entries for feature naming characteristics, which describe how any feature is referred to in a variety of different settings, and Table 2 contains entries for the feature association characteristics, which describe how the feature is related to larger groups of features and sub-elements within a single feature.

Table 1 Gazetteer data model – building naming characteristics (Rice et al. 2011, 285)

Item	Description	Note
Official_Name	Designated official name	To georeference with official map data sources
Original_Drawing_Name	Text label from official map or drawing	To georeference with official map data sources
Abbreviation	Standard feature abbreviation	To use for linkage in the gazetteer if used as a common verbal descriptor
Vernacular1	Slang or informal name for feature	To use for linkage in the gazetteer if used as a common verbal descriptor
Vernacular2	Jargon or technical name for feature	To use for linkage in the gazetteer if used as a common verbal descriptor
Vernacular3	Coded or numbered name for feature	To use for linkage in the gazetteer if used as a common verbal descriptor
Formerly_Named	Previous feature name	To account for re-naming of buildings (e.g., for memorials or re-purposing)
Foreign_Language1-5	Name in foreign language	5 common translations based on student demographics

Table 2 Gazetteer data model – building association characteristics (Rice et al. 2011, 285)

Item	Description	Note
Contains_subelement1	Office or Department Name	Include as many sub-elements to describe those contents of the building which have common verbal descriptors
Contains_subelement2	Retail or dining facility	-
Contains_subelement3	Entertainment locale	-
Contains_subelement4	Special work center	-
Part_of1	Enclosed within or attached to another feature	Denotes physical enclosure or attachment
Building_Cluster_Name	Formal building clusters	From generalized areal descriptors
Event_Grouping	Functional or event-based groupings	For transitory but pre-planned obstacles

The naming and association characteristics shown in Tables 1 and 2 are modified for landmark, walkway, roadway and parking lot features that contain specific linear referencing and linear network information, entrance names, and unique association characteristics.

Since the localized gazetteer is intended to be the information engine for the assistive process, its construction required unique tailoring. Well-known gazetteers such as the U.S. Geographic Names Information System (GNIS) from the U.S. Geological Survey (USGS) assign to a geographic feature its official name, including variants or former names, with geographic reference data (Hill 2006). Geographic features can be classified in notional groups or given additional designations or textual descriptions, such as historical site (which might refer to a feature that is no longer in existence), as is performed in the Geographic Names Project from USGS. For this localized gazetteer to optimize its benefits to assistive geotechnologies, its data model, shown in Tables 3 and 4, focuses on additional ways that a feature might be orally or verbally expressed. This structure which captures naming alternatives as described in Tables 1 and 2 more robustly connects a name to its geo-referenced location.

Sources of naming alternatives come from the members of the local population themselves. George Mason University (GMU) is a linguistically diverse campus, attracting faculty and students from over 130 nations. In 2009, of the 32,500 students, 1700 enrolled as non-resident aliens with only 80 students from native English speaking countries. The largest non-English single language student population is Chinese, numbering 283 non-US citizens from countries that natively speak Chinese (GMU 2010). The faculty is linguistically diverse; over 9% of the 5300 staff and faculty are non-resident aliens. Large numbers of American students and faculty are non-native English speakers. Local demographics indicate that as many as 35% of the American students would speak Spanish, Chinese, Korean, Arabic, Hindi or another as their first language (US Census Bureau 2010).

Students, both English and non-native English-speakers were queried about common vernacular usage. Foreign students and faculty provided cultural and linguistic perspectives on the placenames used to refer to locations around campus. One cross-cultural observation was the widespread use of a building's English name even in discussions held in languages other than English. This was less common in discussions about areas that were functionally described, like parking lots. These student and faculty inputs augmented naming data from other more official sources, such as university campus facility mapping products, university offices, web-listings, and the campus telephone book. Some limited site surveys confirmed these official sources.

University locations and the means to identify them follow both a geographic and a functional hierarchy. As an example shown in Figure 7, twelve individually named dormitories are clustered into Presidents Park. Students of all linguistic backgrounds commonly refer to this area by this name, because of its natural association for a set of buildings which are geographically co-located, functionally equivalent, and carry thematically similar names, such as Jefferson, Roosevelt and Truman. Because of its large population of students, references to Presidents Park would be a common cluster term identified in the operational use of the localized gazetteer.

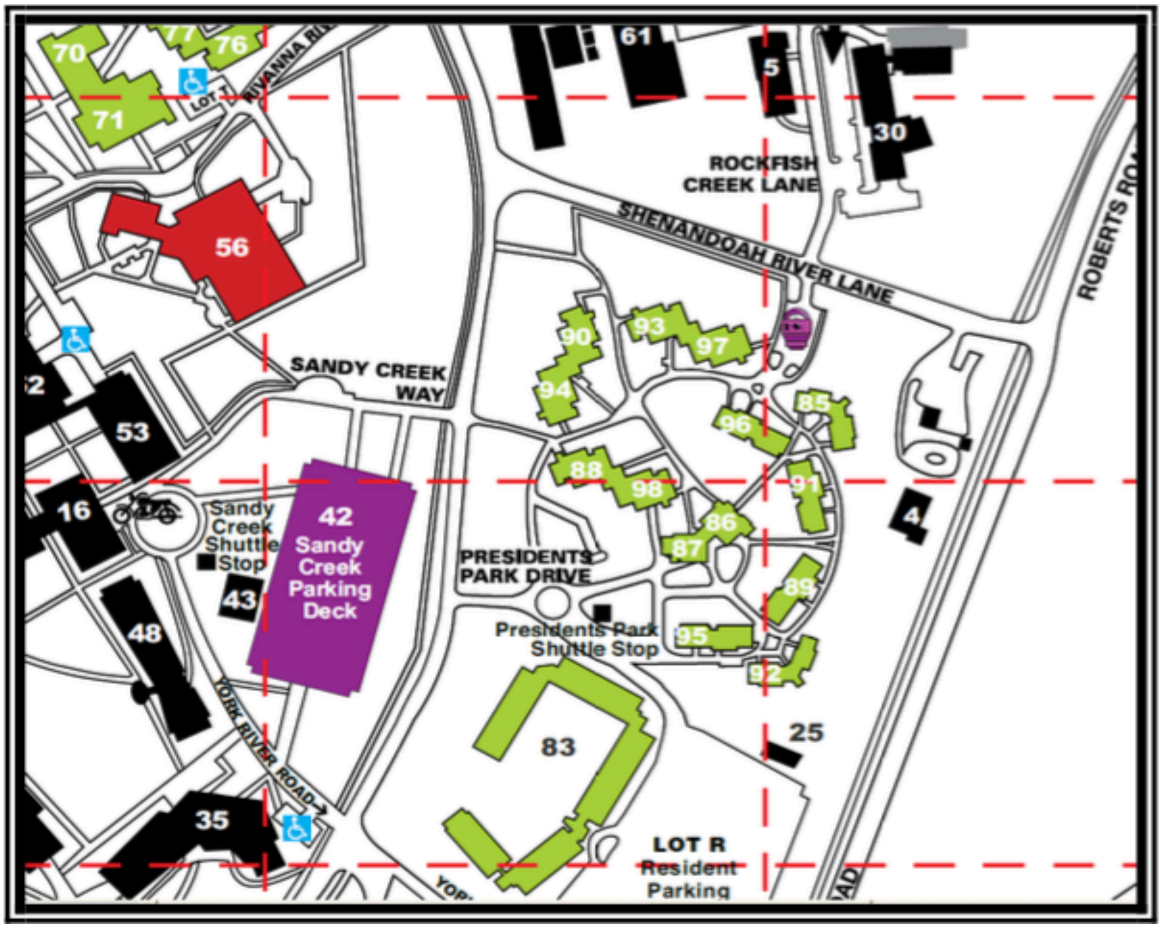

Figure 7 Map detail from George Mason University, Fairfax campus. Note the Residence Hall complex along Presidents Park Drive, comprised of buildings numbered 85 through 98. Building 48 us Research I. Building 56 is Student Union Building II (GMU 2011) (Rice et al. 2011, 287)

For an effective localized gazetteer, the association of University buildings to their contents should be captured. Reference to a known location inside a named building is a common means to locate oneself. The data model shows this attribute as "Contains_ ..." (Table 3). As an example, a University building named Research I (shown in Figure 7) contains, besides offices and classrooms, the following entities:

1. College of Science Dean's Office
2. Department of Geography & Geoinformation Science

3. Geographic Information Science Center of Excellence
4. Center for Earth Observing & Space Research
5. Center for Geospatial Intelligence
6. Joint Center for Intelligent Spatial Computing
7. Center for Spatial Information Science and Systems
8. Laboratory for Natural Hazards
9. Super Computing Facility

Another example of content is the Student Union Building II shown in Figure 4. Besides lounges and student study areas, it contains the Meal Plan Office, the Photo Identification Office and the student mail service center. Each of these sub-elements of their buildings is included in the localized gazetteer because of their utility in providing a georeference.

Other attributes of naming include former names which endure in vernacular usage for years after a building has been renamed. Some structures adjoin buildings but have no distinct name. The observatory connected to Research I (shown in Figure 7 as a noticeable appendage to building 48) is an example of this. It would be included in the localized gazetteer. More transient events might be associated with a building and could be added as an event grouping in the localized gazetteer.

A further consideration for the application of this localized gazetteer is what structures to include in it beyond the obvious buildings and parking lots. For the localized gazetteer to provide its greatest value, it should cover the widest areas of highest use as well as those with the highest safety or security risk for the population. Although a university campus may have a well-defined boundary, the access to the campus and adjacent areas should be considered for inclusion. Additionally, unnamed areas should be added to the localized gazetteer based on an analysis of safety. For example, the western edge of the Fairfax campus at GMU holds numerous athletic fields which are isolated. These places should be described in terms that the students use, even if no official name exists. Prominent landmarks are common reference points and should be included, such as a clock or water tower.

Although our gazetteer data model forms a starting point for a functional system, we are constantly discovering new ways to modify it to make our system better. User feedback is critical, particularly because the system relies on end-users to become involved with communicating about changes to navigation corridors and the presence of obstacles or hazards to navigation. We are analyzing ways to provide positional privacy in the system, and masking of user identities, preferring an opt-in approach to self-identification, being aware of the many issues associated with motivations of end-users and concerns about negative social dynamics. A major ongoing effort is the integration of time stamps and temporal relevance into the system.

3.3. Future Plans[2]

At present, our two primary sources of user-contributed geospatial information are the campus alert system and pedestrians who transit campus on daily basis. We plan on growing the community of end-users and geospatial data contributors by advertising the presence of our system through our local campus disability services office and through the local campus planning office, which routinely provides information about sidewalk closures associated with construction. We also plan on integrating other auditory and haptic cues for obstacles and hazards, based on earlier work reported by Rice, Jacobson, and Golledge (2005). We also intend to refine the temporal aspects of the methodology described here. Endusers typically lack an understanding of the temporal dynamics of obstacles or hazards other than a present-tense existence, i.e., "there is a hazard here right now". In many VGI-based systems, end-users don't have a way of specifying a temporal endpoint for their contributions and there appears to be few resources directed at follow-up. A few authors, including Goodchild and Glennon (2010) have discussed temporal issues, noting the primacy of temporal relevance and the benefit of VGI, while maintaining a balance between errors associated with false positives and false negatives. A general approach to filtering and managing end-user contributions is to treat the most recent update as the most relevant, and to phase out contributions after a set

[2] This section, written in 2011, reflects the progress in the research workflow at the earliest stages of the project, and as published in the cited article. The author recognizes the value in noting that many of these future plans have been achieved and have been documented in subsequent chapters. For the sake of overall consistency with published work, and to preserve the significant progression in approaches to the research problem, the author of this dissertation has chosen to leave this section (and similar "Future Work" sections in Chapters 4 and 5) intact as originally written and published. An overall summary of the research contributions and a common "Future Work" section for all chapters appears at the end of this dissertation in Chapter 7.

period of time. This general approach, however, presents a number of problems when updates come from sources where material has been rebroadcast or repackaged by news aggregation websites, causing it to appear more recent and therefore more relevant. We are still working on temporal issues and hope to have a better way of defining relevance and end-points for contributions.

Other future goals for this research include the refinement of the methodology to deliver in-situ obstacle or hazard information to blind and visually-impaired individuals, and to improve the locational aspects of the information to suit the cognitive needs of the blind or visually-impaired individual transiting across our local university campus. Refinements to the geoparsing tool will include improved text recognition and functional interpretation of distance and direction. Other technologies to incorporate are text to voice web services similar to Google Voice functionality. Additional goals with extensibility into broader areas of user contributed geographic information include an assessment of accuracy.

Volunteered geographic information and participation from end-user communities may have a significantly transformative effect on GIS and applications oriented toward accessibility. As end-users become more inclined to transform their observations into VGI contributions, visually-impaired and blind individuals will benefit. We hope to provide significant contributions toward this evolving process and look forward to future developments.

4. SUPPORTING ACCESSIBILITY FOR BLIND AND VISION-IMPAIRED PEOPLE WITH A LOCALIZED GAZETEER AND OPEN SOURCE TECHNOLOGY

Abstract

Disabled people, especially the blind and vision-impaired, are challenged by many transitory hazards in urban environments such as construction barricades, temporary fencing across walkways, and obstacles along curbs. These hazards present a problem for navigation, because they typically appear in an unplanned manner and are seldom included in databases used for accessibility mapping. Tactile maps are a traditional tool used by blind and vision-impaired people for navigation through urban environments, but such maps are not automatically updated with transitory hazards. As an alternative approach to static content on tactile maps, we use volunteered geographic information (VGI) and an Open Source system to provide updates of local infrastructure. These VGI updates, contributed via voice, text message, and e-mail, use geographic descriptions containing place names to describe changes to the local environment. After they have been contributed and stored in a database, we georeference VGI updates with a detailed gazetteer of local place names including buildings, administrative offices, landmarks, roadways, and dormitories. We publish maps and alerts showing transitory hazards, including location-based alerts delivered to mobile devices. Our system is built with several technologies including PHP, JavaScript, AJAX, Google Maps API, PostgreSQL, an Open Source database, and PostGIS, the PostgreSQL's spatial extension. This article

provides insight into the integration of user-contributed geospatial information into a comprehensive system for use by the blind and vision-impaired, focusing on currently developed methods for geoparsing and georeferencing using a gazetteer.

Published in 2012 at Transactions in GIS, volume 16(2) (177-190). Authors Rice, M.; Aburizaiza, A.; Jacobson, D.; Shore, B.; and Paez, F.

4.1. Introduction

Few geospatial activities are as directly essential and fundamental as the regular, daily navigation and wayfinding tasks that help a person travel between workplace, appointments, errands, and home. For blind and visually impaired people these daily navigation tasks are nearly always challenging and occasionally impossible, particularly in unfamiliar environments. Our modern built environment is, with very few exceptions, designed for the sighted.

During a commencement address delivered at Simon Fraser University in 2001, Reginald Golledge, a pioneer in assistive geotechnology (figure 7), outlined the various obstacles to navigation and wayfinding encountered by the blind and visually impaired. Described by Golledge as movement barriers, these obstacles consist of real impediments in the built environment as well as metaphorical barriers that present disincentives to full participation in employment and society. With this context in mind, our goal is to use Open Source geospatial software, tools, and techniques to help blind and visually impaired members of society overcome these movement barriers and therefore, more fully participate in society. This article presents our work in designing a system with Open Source software that uses volunteered geographic information (VGI) to augment officially sourced infrastructure datasets. This is accomplished through the use of a gazetteer and geoparsing system that obtains and displays spatial footprints for VGI stored as text in a database.

4.2. Approaches for accessibility mapping and assistive geotechnology

The International Cartographic Association's (ICA) Commission on Maps and

Graphics for Blind and Partially Sighted People has been an important outlet for research on maps, graphics, and geotechnology for the blind, with historic focus on tactile maps and graphics. Perkin's (2001) review of tactile map production techniques, and Tatham's (1991) review of tactile map design principles are primary sources of information on the subject of tactile map production and non-visual map symbolization for the blind and visually impaired. Taylor (2001) and more recently Przyszewska and Szyszkowska (2011) described the significant challenges and difficult production tasks in large tactile mapping projects. As they and many ICA presenters have attested over the last 20 years, production of tactile maps and graphics can be difficult, time-consuming, and expensive.

Figure 8 Dr. Reginald Golledge using the UCSB Personal Guidance System, circa 1997 (Rice et al. 2012a, 179)

In an earlier ICA conference proceeding, Coulsen et al. (1991) described a different approach to mapping and accessibility for the blind and visually impaired. Coulsen described a future where geographic information systems (GIS) could be used to more quickly and efficiently generalize, simplify, and prepare geospatial data for inclusion on a tactile map. The vision espoused by Coulsen et al. (1991) is one in which geotechnology is used to assist with the difficult production of tactile maps and graphics for the blind and visually impaired. Noteworthy for its time period – before GIS was being used as a tool for mapping for the blind – Coulsen et al. described a future assistive geotechnology that has begun to emerge. In an approach similar to Coulsen et al., Miele and Marston (2005) described an approach for automated tactile map design using GIS street centerline files and a Braille embosser, a central component of Miele's Tactile Map Automated Production project at the Smith-Kettlewell Eye Research Institute (2011). The approaches suggested by Coulsen et al. (1991) and those presented by Miele and Marston (2005) are significant because of their integration of standard tactile mapping approaches with GIS and geotechnology, and in the case of Miele and Marston, the extension and use of this production approach for navigating through unfamiliar environments.

Extending the idea in Coulsen et al. (1991) from the domain of tactile navigation aides to general aides for navigation and wayfinding, a few researchers, notably Reginald Golledge, have used GIS in innovative ways to provide technological navigation aids for the blind (Golledge et al. 1998). Golledge's Personal Guidance System (PGS) is built using a combination of simple GIS functionalities, a global positioning system (GPS) for positioning, an electronic compass for orientation, and earphones for receiving auditory cues (Figure 8). This system allows blind and visually impaired people to navigate through unfamiliar environments, receiving auditory and text-to-speech cues for positioning relative to the sidewalk centerlines, warning cues for objects, and text-to-speech announcements for buildings and infrastructure directly in front of the end-user. While under development, the Personal Guidance System and related system extensions incorporated some novel sensory communication systems such as a vibro-tactile pointer interface (Marston et al. 2008, Golledge et al. 2007), auditory infrared signage (Marston et al. 2006), and advanced auditory displays.

The Personal Guidance System was a significant advance in navigation for blind and visually impaired individuals, particularly with respect to innovations in interface design and methods of robust field testing. However, it was, in essence, a closed system with respect to the underlying geospatial data. Similarly, much of the work presented and published by the ICA, including Perkins (2001) and Tatham (1991) adopted an approach to tactile map design based on static data and a static product. An ideal system for mapping and navigation for the blind and visually impaired would incorporate some element of rapid update, reflecting the constant changes in the built environment. Two research efforts that address dynamic content in assistive geotechnology systems are Nuernberger (2008) and Barbeau et al. (2010), who devised practical approaches for improving accessibility through the use of cell phones. Nuernberger (2008) devised methods for rapidly disseminating information about changes in an environment to mobility impaired individuals through the use of supplemental verbal information delivered over a cell phone. Similarly, Barbeau et al. (2010) demonstrated the effectiveness of a travel assistance device based on GPS-enabled smart phones by public transportation users with disabilities. Both research projects demonstrate the feasibility of delivering real-time geospatial information and location cues to disabled end-users, and as such, represent an improvement from past paradigms associated with static tactile maps and assistive devices using fixed base data.

4.3. Volunteered geographic information and assistive geotechnology

A contemporary approach for building a more dynamic assistive geotechnology system involves the use volunteered geographic information (VGI) contributed by community members, to augment official sources of geographic information. The general usefulness and effectiveness of VGI has been highlighted by a number of authors. Goodchild (2007b) introduced the term volunteered geographic information to define the emerging inexpensive and functional methods for the collection of geographic data through the active participation of end-users and community members. Furthermore, Zook et al. (2010) and Goodchild and Glennon (2010) have examined the results of using community members as an immediate and real-time source of information during emergencies. Both of these articles support the notion that user-generated content can effectively supplement traditional and official sources of geospatial information during emergencies and disaster management. Transitory hazards, such as construction barricades and frequent changes in the landscape, put blind and vision-impaired people in danger and represent the same type of significant hazard envisioned by these authors.

Like many public institutions, George Mason University has a population of blind, visually impaired and mobility impaired students and visitors. It is also the setting for several very large constructions projects, resulting in frequent obstructions within the pedestrian corridors and a serious disruption for students, staff, and visitors that navigate across campus. Our interest and experience with assistive geotechnology led to the early conceptual design of a system that would overcome the weakness of having a closed and static design. Specifically, we planned a system for geospatial data collection, analysis, routing and delivery, built with free and Open Source software, that would allow for rapid ingestion of information about temporary obstacles in the public rights-of-way.

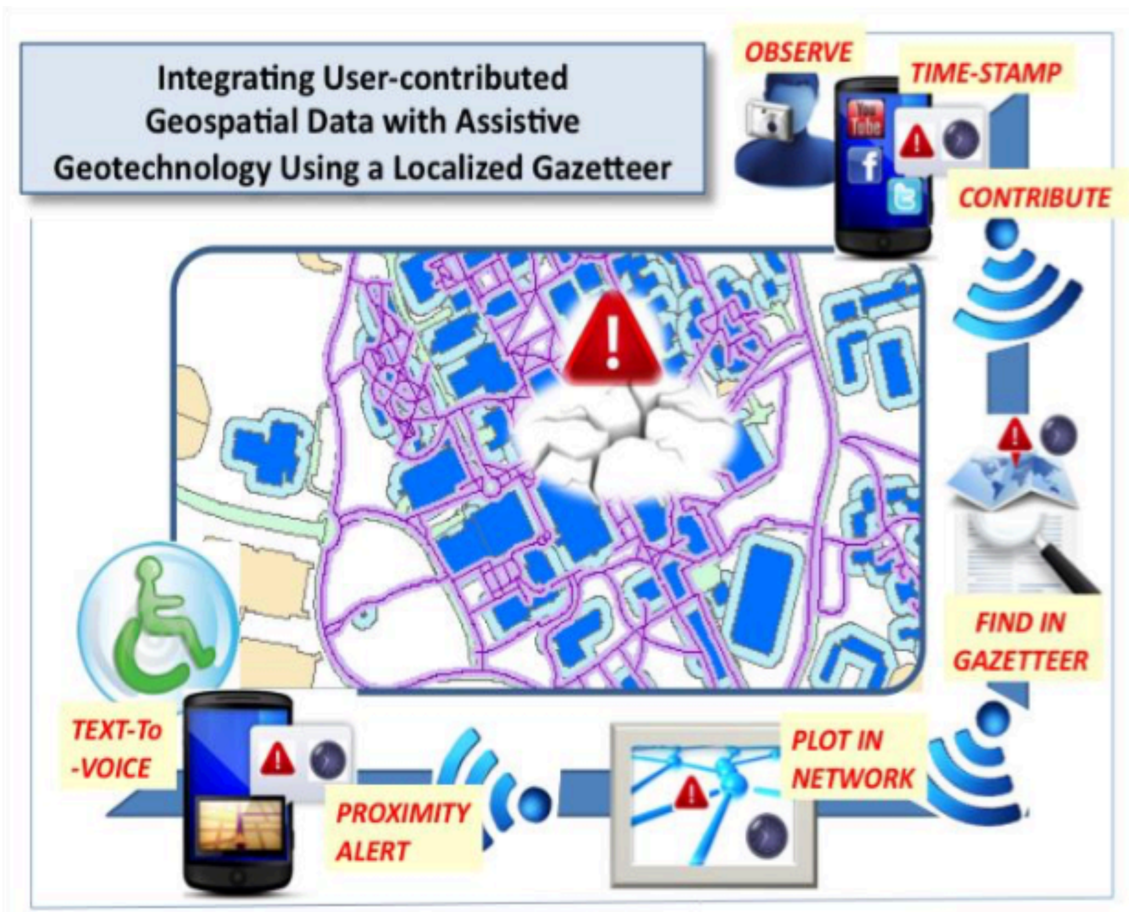

Figure 9 Conceptual process flow for integrating VGI with an assistive geotechnology system (Rice et al. 2012a, 181)

After considering the many options, we decided to design our system using VGI, noting the many benefits of such an approach, such as local geographical expertise in data collection and description, and increased temporal coverage (Goodchild 2007a). We developed a conceptual design to show what the flow of information would look like (Figure 9). Our system would focus on the integration of volunteered or user-contributed geospatial data with assistive geotechnology, using a detailed, locally-centered gazetteer. The sequence of events in such a system would include observation, contribution, geoparsing and georeferencing using a gazetteer, plotting and routing using infrastructure and network data, and finally, the delivery of maps and proximity alerts to mobile assistive devices, including alerts using auditory cues and text-to-voice translation. We have used this conceptual design to guide our development.

Because most of the VGI incorporated into our system is based on unstructured, text-based descriptions of obstacles received via email and text messaging, we recognized a critical need for geoparsing and georeferencing capabilities. During the early development of our system, we developed a comprehensive locally-focused gazetteer for use in georeferencing VGI that incorporates detailed place-naming for the local area. Because humans rely heavily on place-naming for geographic descriptions, and few potential contributors would have a detailed understanding of metric georeferencing systems, the gazetteer was perceived to be, and continues to be, a central element in our system.

The use of VGI, and our development of geoparsing and georegistration algorithms for use with unstructured VGI, is a central focus of this research, and an important contribution within the framework of assistive geotechnology. Our current system extends approaches suggested by Coulsen et al. (1991) and the state-of-the-art work by Golledge et al. (1998) and recent advances by Nuernberger (2008) and Barbeau et al. (2010), by incorporating VGI contributed by community members. This allows information about temporary obstacles to be incorporated into maps and mobile alerts delivered through our system, and provides a demonstration of the utility of VGI within the assistive geotechnology community. The important elements of our system, including our gazetteer development, geoparsing capability, and Open Source design will be presented, followed by a discussion of our results, experiences, and conclusions.

4.4. Gazetteer data model and geoparsing

The two most important functions of our present system are: (1) the ability to identify the use of place names within VGI; and (2) the assignment of metric georeferencing to volunteered content through the use of a detailed local gazetteer. Because people frequently communicate about their surroundings using place names and because the place name usage is sometimes non-standard, we have spent several months building and populating a comprehensive gazetteer for the college campus and surrounding neighborhoods, based on the georeferencing and gazetteer development principles outlined by Hill (2006).

Our evolving gazetteer data model includes naming characteristics for buildings, sidewalks, streets, intersections, parking lots, neighborhoods, landmarks, parks, gardens, plazas, and a variety of miscellaneous features used in local geographic descriptions. As implemented, we are currently providing gazetteer data for nearly 1,200 separate geographical entities in a 2-square mile area centered on the George Mason University Campus in Fairfax, Virginia. Table 3 shows some of the naming characteristics for buildings, the primary geographic feature used in local geographic descriptions related to navigation. Being both the origin and destination of most pedestrian events on our campus, establishing a comprehensive set of naming characteristics and naming variations for buildings is an important precursor to integrating VGI with officially-sourced campus datasets. In our gazetteer, we record each building's official name (Table 3, Official Name), as well as the name used on campus drawings and publications (Table 3, Official Drawing Name), which is, in most cases, a shortened form. We also record numerical designations, abbreviated variants of the building name, slang names, jargon or technical names, and any previous or former names, reflecting the frequent name changes as building occupants change and features are renamed. We also store several commonly-used non-English language variants of feature names in our gazetteer, focusing on the most commonly spoken languages by students on our diverse campus whose national origins and primary languages are from China, India, Korea, Pakistan, Saudi Arabia, Egypt, Vietnam, and Latin America.

An important, but often overlooked aspect of naming conventions and gazetteers in small projects, is the need to store association characteristics for features. Buildings on our local campus are often aggregated into geographical clusters (buildings in the same area with a common group name) and functional clusters (buildings sharing a common functional purpose, such as dormitories). Buildings and other geographic features are commonly referenced by their individual name and, where they exist, by the name of any functional and geographical groupings. Buildings also commonly contain separately named constituent parts, which we reference in our gazetteer using a contains element.

Table 3 Gazetteer data model: building naming characteristics (Rice et al. 2012a, 183)

Item	Description
Official Name	Designated official name
Original Drawing Name	Text label from official map or drawing
Building Code	Numerical designation for building
Geographic Cluster	Name of geographical grouping of buildings
Contains	Names of sub-elements of the building
Functional Cluster	Name of functional grouping of buildings
Abbreviation	Standard feature abbreviation
Vernacular1	Slang or informal name for feature
Vernacular2	Jargon or technical name for feature
Vernacular3	Coded or numbered name for feature
Formerly named	Previous feature name
Foreign language 1–8	Name in foreign language

In order to provide a method for detecting spelling variation, which is common in unstructured VGI, we also store a name variant based on a consonant-only version of the official feature name. The naming characteristics and footprints for campus geographic features are stored in a database, and used in conjunction with geoparsing algorithms, which look for uses of the names or any name variants within text. The geoparsing algorithms allow us to identify relevant geographical entities whose footprints are recorded in our gazetteer and, therefore, facilitates the routing, mapping, and alert functionalities of our system. Some additional details about our gazetteer entity relationship and association characteristics can be found in Rice et al. (2011).

4.5. Open source system implementation and design

4.5.1. Free and open source software

We have designed our system using free and Open Source software due to the flexibility inherent in the Open Source geospatial tools and the ease with which they can be obtained and utilized by non-profit groups that are commonly interested in accessibility issues. The low-cost of implementing a system in free and Open Source software is a significant factor in our own development and at least anecdotally, in the ability of local non-profit groups interested in mirroring our approach. The balance between free and Open Source geospatial technology and off-the-shelf commercial and proprietary software is a complex, delicate, and difficult one to achieve. For the wider user community, the initially prohibitive cost of purchasing commercial software is potentially offset by the benefits of support from the software vendor. The financial cost of commercial software and the requisite utility it may or may not provide for a project such as ours needs to be offset against the "invisible" cost of free and Open Source geospatial technology. This invisible cost is the payment for an in-house expert, consultant or similar person to develop and implement the system. Along this continuum of software with overt financial costs and free software with hidden costs, the development decisions can be difficult, as noted by Steiniger and Bocher (2009). In our case, we have the requisite expertise to deploy Open Source tools and the expertise to support end-users and have adopted them for use based on the low cost. We routinely reconsider the possible benefit to be derived from a mature commercial product and associated support structure and explore those options when they arise, and have no aversion to using commercial products in our development when they present a clear advantage in utility or total cost.

4.5.2. System implementation and design

There are five general areas of functionality in our geoparsing and georeferencing system: (1) reading the VGI message (often a warning message about an obstacle); (2) manipulating the warning message; (3) scanning the gazetteer database; (4) editing the database; and (5) outputting the results. The technologies utilized to implement the system are: HTML, JavaScript, AJAX, Google Maps API, PHP and PostgreSQL//PostGIS.

The heart of the system is a PostgreSQL database, an Open Source database that has the advantage of being free to install and use. The PostgreSQL extension, PostGIS, enables spatial capabilities and compared to other Open Source spatial databases, has advanced functionality that we have found useful. PostGIS spatial functions are used inside the SQL statements, as spatial queries, and range from geometry construction and editing functions to spatial operations. PostGIS is also relatively easy to implement and capable of being mashed-up with other GIS desktop and web technologies. In our system, we use a variety of spatial functions such as point extraction from polygons, convex hull creation, polygon buffering, binary geometry format to KML string conversion, binary geometry format to text conversion, and multipolygon object creation from smaller polygons.

To date, our system database contains four tables: the gazetteer table, the contains table, the convexhulls table, and the users' table. The gazetteer table has 26 possible fields for feature description, including the primary building name characteristics noted in Table 1. Moreover, the gazetteer contains feature names in eight languages, allowing geoparsing of feature names from the common languages used on campus. The contains table provides detailed information about feature entity relationships and allows for georeferencing objects that are contained as a sub-element within another feature, which helps associate building footprints with commonly referenced constituent parts. The convexhulls table is used to store convex hull geometry of geographic entities geoparsed from user-contributed VGI, in both binary and KML string formats. It also stores the time stamp of the VGI message, to enable the system to distinguish between old and new VGI entries and to assist in establishing temporal relevance. The users' table is for registering end-users who are interested in receiving any relevant generated warning messages. Cell phone numbers and their carriers are also stored to send the warning messages to the users.

Figure 10 End-user contribution text and associated spatial footprints (Rice et al. 2012a, 185)

PHP, along with JavaScript and AJAX, plays the major role of reading and manipulating warning messages, editing the database, and creating the output. When a VGI message (or warning message) is received, the gazetteer table and contains tables in the database are scanned to match any feature names in the warning message. The IDs of the named features are collected in the code. Points are then collected from relevant feature geometries and used to create and buffer a convex hull to cover all possible areas relevant to any obstacles or warnings reported by the originator of the VGI. The buffered convex hull polygon is then inserted in the convexhulls table in the database in both KML string and binary formats.

Figure 10 shows an example of four different hypothetical warning messages to our system, to highlight some of the designed features. Because the local college campus is very diverse, and because we intend to capture verbal and textual information from the largest possible community of VGI contributors, we felt a need to incorporate foreign place name variants in our gazetteer and to develop robust geoparsing capabilities for place names, including misspelled place names. For instance, the Arabic words in the first message (Figure 10) reference the Fenwick Library and the Chinese logograms in the second message represent Lecture Hall. In the third message, an end-user references the computer store, which is contained inside the Johnson Center building, and in the fourth message, an end-user references the Fine Arts Building using the abbreviation FAB. In each case, including those with misspellings, the system identifies a geographic feature named in the message and forms a buffered convex hull from the relevant feature geometries. We have trained the system to catch a number of common spelling errors in English by storing the consonant-only variants of the place names and using PHP to match them to possible place name references.

As an additional step, our system generates several KML files from the convexhulls table based on the time stamp of each convex hull. The KML files are posted on a map using Google Maps API V3. Each KML file is associated with a specific temporal domain, helping establish relevance and allowing for visual sorting of the most recent messages. At present, one of the KML files is used to represent warnings received within the past 24 hours. A second KML file is for warnings received within the last week, and so on. Figure 11 shows a scenario where two warning messages were received through VGI sources and processed according to the procedures outlined in this article. If the first warning message was received five days ago, and the second warning message was received only today, the messages can be sorted and displayed according to temporal relevance. Both messages are currently found in the KML file for the past week, but only the second message is found in the KML file of the past 24 hours. Warning messages are removed after a period of 5 days, unless system moderators flag the message to be removed after a field check confirms resolution of the situation, or if a system moderator determines, through the same field check, that the message needs to remain for a longer time in the system.

4.6. Discussion and conclusion

The system we have outlined and discussed was designed to fill a gap in some useful assistive geotechnology systems, such as the Personal Guidance System developed by Golledge and colleagues (Golledge et al. 1998), as well as to develop an approach suggested Coulsen et al. (1991) where a geographic information system could be used to enhance the utility and speed of tactile map production. In our case, the focus of the system is not in generating, designing, and using tactile maps or other accessible maps (a topic addressed by Tatham 1991, Rice et al. 2005 and Golledge et al. 2005), but rather to develop a way of incorporating information about temporary obstacles and barriers into an existing system for the purposes of identifying hazards to blind, visually impaired, and mobility impaired individuals. Our approach uses volunteered geographic information to identify temporary obstacles in our local environment. Because most accessibility maps and related systems are designed with a closed architecture using static data or using an analog cartographic process, they cannot provide information about the temporary hazards that present a significant impediment to navigation and wayfinding. Our system is built with official infrastructure data but allows for the contribution and display of VGI and benefits from local geographic expertise, a characteristic of VGI noted by Goodchild (2007a). Two central capabilities that are required to effectively use VGI in our system are: (1) the ability to geoparse contributed information for the presence of geographical place names and other geographical identifiers; and (2) the ability to georeference the contributed information using a comprehensive local gazetteer, facilitating mapping and other GIS functionality.

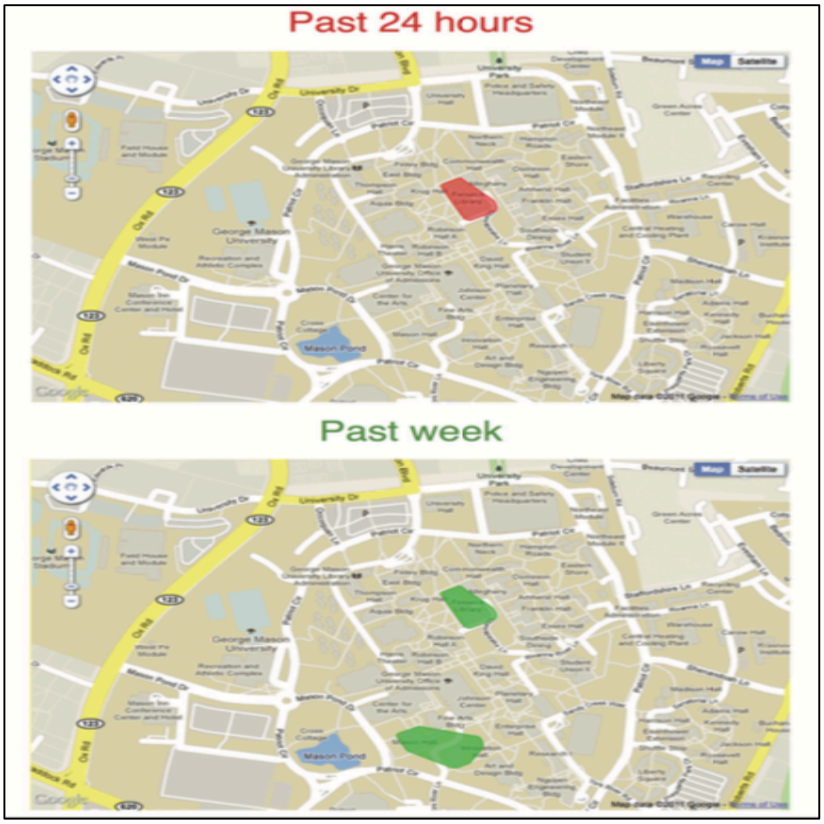

Figure 11 Red and green tagging of the KML footprints for temporal relevance (Rice et al. 2012a, 187)

A required element of our system (or any system based on voice or text-based geographical descriptions) is a gazetteer containing, among many other things, the names, abbreviated name variants, and footprints or any object in the built environment used in referencing geographic position. In our local application, the primary features in the gazetteer include buildings, parking lots, walkways, and prominent landmarks (such as bell towers, clocks, fountains, etc.). The naming characteristics for these items are extensive, but as a basic minimum include the official name, an abbreviated form of the name, and a footprint or geometry for the feature. Assembling a gazetteer containing the level of detail described here is not a trivial task, even for a small geographical area with a strong central authority with up-to-date infrastructure data and associated metadata.

We conclude that the most difficult task in utilizing our approach is building a detailed gazetteer for a local area. This task can be accomplished in any area with existing GIS data for buildings, streets, and landmarks, but many of the useful features of our system, such as the ability to detect misspelled placenames, slang variations of placenames, and foreign language variants of placenames requires a significant investment in time, effort, and expertise.

Our research demonstrates, within our limited geospatial domain, the mechanisms for using volunteer geographic information, georeferenced using a local gazetteer, to providing information about temporary barriers and obstacles in the local environment. Our approach, using free and Open Source geospatial software, could be implemented in any similar local geographic setting, given the existence of enough interested support personnel and developers to create a detailed gazetteer. Our approach could be further extended to build upon information in local datasets rapidly during times of disasters or natural hazards where volunteer geographic information is both necessary and needs to be validated in a critical manner to ensure its reliability.

4.7. Future development directions[3]

Our system is being refined and extended in a number of ways, recognizing

[3] As noted in Chapter 3, the discussion of future development directions in this chapter and in others has been preserved to show the state of the workflow at that time, and to underscore the progression of the research, which is summarized in full at the end of this dissertation (Chapter 7). The future work and development directions are consistent with the published version of the paper, and are updated significantly in the remaining chapters.

important elements which have not been developed. First, we have developed proficiency in identifying regions of interest from VGI based on the convex hull geometry of the named features, and the creation of a KML and binary representation of the convex hull geometry. We recognize that local geographic descriptions contain references to objects that the VGI contributor is nearby but not necessarily inside, so as a positional accommodation we use a simple buffer to increase the diameter of the convex hull around the region of interest. We plan on developing methods for creating more geographically specific regions of interest based on other infrastructure items such as sidewalks, which typically constrain the movement of pedestrians between buildings in the local area. We also plan on creating methods for validating the position of a VGI contributor based on ancillary data present in the geotags of contributed photographs, as well as location data communicated with consent through GPS-enabled mobile applications.

A second area of future development is in our need to accommodate complex spatiotemporal dynamics and spatiotemporal relationships contained in geographical descriptions from VGI contributors making observations as they are moving. We are also interested in geographical descriptions where the points of reference are changing, and geographical descriptions of obstacles and barriers that use positioning descriptions relative to the observers position and orientation, which may be unknown. As noted, we can capture some information about a VGI contributor's position with permission from mobile device parameters, and that information can be used to validate geographical descriptions. Capturing user orientation, however, is more challenging due to uncertainties in bearing and the lack of embedded orientation information in contributed media and orientation data from mobile devices. Text-based descriptions of geographic phenomena referencing positions relative to user orientation will be more difficult to capture and use in our system.

A third important area for future work is in user recruitment, user training, data validation, and the development of reliability measures. We plan on implementing a user registration system and user rating system to allow development of methods for tracking the reliability of VGI contributions to our system. At present, the system is open and most contributors are students, staff, and faculty who are aware of the project and are moderately trained. We have had little problem with malicious content or misleading entries, but with an open system that will change as the numbers of contributors and potential liabilities for malicious content increase. At present, we rely on the general feelings of altruism expressed by our contributors, who have an interest in contributing useful information about obstacles and hazards. Elwood (2010) notes how the social dynamics and boundaries between experts and non-experts change and adjust in projects like this, and we plan on documenting any noteworthy developments associated with the behavior of the community of end-users and the community of VGI contributors, and interactions between the two groups, noting that this would be a positive contribution of our project.

Acknowledgements

We would like to thank the George Mason University College of Science for their support in funding the early stages of this research, the Army Engineer Research and Development Center (ERDC), for generously funding the continued work under BAA

#AA10-4733, and William C. Hammill (George Mason University), for his assistance with the conceptual system design. Finally, we acknowledge the advice and encouragement of Dr. Reginald Golledge (1937–2009) during the inception of the project.

5. POSITION VALIDATION IN CROWDSOURCED ACCESSIBILTY MAPPING

Abstract

We live in a society in which instant gratification is expected: we demand constantly up-to-date information, which is reflected in our reliance on maps for navigation. Volunteered geographical information (VGI) and geocrowdsourcing make this demand attainable, with popular examples being Waze and OpenStreetMap, where maps are updated quickly by citizen contributors with current base data and features. At George Mason University (in Fairfax, Virginia), the Office of Disability Services releases a traditional paper accessibility map once annually. Owing to its production methods and format, this accessibility map does not capture the transient obstacles that occur frequently throughout campus, rendering it less useful to disabled pedestrians. To fix this dilemma and establish a more useful accessibility system, we have created an application in which contributors report transient obstacles that may impede pedestrian navigation, including sidewalk obstructions, construction detours, and other obstacles that may affect pathway walkability. One of the concerns associated with VGI and geocrowdsourced information is quality assurance, which is imperative when the usage scenarios (including blind, visually impaired, and mobility-impaired navigation) depend on positional accuracy. This study attempts to address the

concerns related to the quality assurance of VGI, specifically quality assessment of the positional accuracy of the geocrowdsourced spatial data. We present our quality assessment techniques and novel methods for assessing the consistency of positional characteristics of geocrowdsourced spatial data related to accessibility. These methods rely on moderated positional assessments, geotags extracted from contributed images, and gazetteer-based geoparsing of location descriptions. Finally, we base our methods and approaches on research contributions and best practices from past and current efforts in accessibility mapping.

Published in 2016 in Cartographica: The International Journal of Geographic Information and Geovisualization, volume 51(2) (55-66). Authors Rice, R.; Aburizaiza, A.; Rice, M.; and Qin, H.

5.1. Introduction

About 10 miles outside of Washington, DC, George Mason University (GMU) has been recognized repeatedly as one of the top ‘‘up-and-coming’’ universities in the United States (GMU 2012). GMU prides itself on being the largest and most diverse university in the Commonwealth of Virginia, which has resulted in increased congestion and constant construction projects to accommodate the steadily growing campus. The diverse student body includes many individuals with accessibility issues, namely 300 mobility impaired and roughly 50 visually impaired students, faculty, and staff. To assist these disabled individuals and comply with regulations, GMU produces and prints an accessibility map, released once annually in PDF format. This map, printed in high-contrast colours, captures permanent obstacles such as stairways, steps, and steep paths that are subject to the accessible design scrutiny contained in the Americans with Disabilities Act. Many of the significant obstacles that actually hinder navigation in the local area are transient – construction detours, poor surface conditions, sidewalk obstructions, entrance or exit problems, crowds, and events that appear and disappear frequently (see Figure 12).

Any traditional printed map provided by GMU to the student body and local community is not capable of capturing these dynamic obstacles, owing to the practical constraints of the cartographic production processes. GMU’s printed accessibility map captures fixed obstacles and pathways that are non-compliant with the Americans with Disabilities Act, including steps and steep and narrow pathways, but does not possess the capability of relaying information regarding dynamic hazards (such as the obstacle shown in Figure 12) owing to its long production time. Recognizing the need, we have created a system that utilizes map-based, geocrowdsourced data collection and traditional GIS quality assessment techniques. This system, the GMU Geocrowdsourcing Testbed, serves as an important local information resource and a dynamic aid with which to study the changing accessibility conditions of the local area. The current development of this system is presented here, along with two research studies that look at the quality of map-based positioning of obstacle reports and methods for validating the position of obstacles within our GMU Geocrowdsourcing Testbed.

Figure 12 Transient navigation obstacle near the George Mason University campus (Rice et al. 2016, 56)

5.2. Literature review

For nearly 30 years, the International Cartographic Association's Commission on Maps and Graphics for Blind and Partially-Sighted Persons has been an outlet for outstanding research in tactile mapping and accessibility. Tatham's 1991 paper on theoretical and practical design considerations for tactile maps is a standard reference, as are Perkins's (2001; 2002) review of tactile map production technologies and Coulsen et al.'s insightful 1991 paper suggesting that GIS would be useful for future tactile mapping efforts. More recent work has centred on colour blindness (Pugliesi and Decanini 2011) as well as large-scale tactile map production efforts (Przyszewska and Szyszkowska 2011; Taylor 2001).

Figure 13 UCSB Personal Guidance System, 9 May 2003. Courtesy of the University of California, Santa Barbara (Rice et al. 2016, 57)

With Coulsen et al. (1991) in mind, another recent thread of research looks at the way that GIS and contemporary geotechnology are updating and replacing traditional tactile maps. Miele (2007) looks at tactile mapping through scalable vector graphics and automated Web-based production techniques. Laakso and others (2013) develop data models of pedestrian infrastructure to enhance accessibility, and Rice et al. (2011) and Rice, Jacobson and others (2013) look at geocrowdsourcing and gazetteer-based geoparsing for accessibility.

Research in geoassistive technology often requires dynamic positioning and mobile devices. Loomis et al. (2001) and Marston et al. (2006, 2007) introduce the University of California, Santa Barbara (UCSB) Personal Guidance System and innovative audio and tactile displays. Their approach uses the Global Positioning System (GPS), a fluxgate compass for orientation, GIS base map data on a mobile computer, headphones for auditory cues, and a tactile pointer interface worn on the finger. Figure 13 shows the UCSB Personal Guidance System, circa 2003 Rice et al. (2005) and Golledge et al. (2005) introduce tactile and auditory mapping techniques to augment accessibility systems and more recently, Rice et al. (2012) and Rice et al. (2013) have introduced geocrowdsourcing and gazetteer-based geoparsing to augment traditional accessibility maps and accessibility system. The advantage in their approach is that transient obstacle information can be administered to the public through crowdsourcing, and that information, including its position and attributes, can be quality-assessed based on the social approach introduced by Goodchild and Li (2012).

This work expands on these existing research contributions by introducing the GMU Geocrowdsourcing Testbed, a pilot study on map-based obstacle crowdsourcing, and our quality assessment methods for validating position. Karimi et al. (2014) also employ geocrowdsourcing to create a personalized accessibility map for a local area but comment on the difficulties of employing a quality assessment method owing to the lack of an official data source. The quality assessment methods employed by the GMU Geocrowdsourcing Testbed are based on the traditional standards for data quality in volunteered geographic information (VGI) as outlined by Girres and Touya (2010), who based their work on the traditional GIS data quality metrics discussed by Guptill and Morrison (1995). We consistently refer to the OpenStreetMap (OSM) accuracy studies conducted by Girres and Touya (2010) and Haklay (2010), acknowledging that OSM data typically lies within 6 m of an authoritative data source, while Arsanjani and others (2013) and Zheng and Zheng (2014) choose to perform OSM accuracy checks by buffering OSM features, finding in some cases that there is only about a 60% overlap. While there is still much research to be done regarding the positional accuracy of VGI in services other than OSM, we have used the current research as a guideline for our quality assessment accuracy standards in the GMU Geocrowdsourcing Testbed.

5.3. Methodology

5.3.1. GMU Geocrowdsourcing testbed

The GMU Geocrowdsourcing Testbed was developed from ideas introduced by Rice et al. (2011), who sought to improve systems such as the UCSB Personal Guidance System by incorporating transient obstacle information. This system, developed in the following year and outlined in subsequent publications (Rice et al. 2013; Rice et al. 2014), uses contributors from the public to identify and report transient obstacles, and moderators to check, correct, modify, and quality-assess the information contributed to the system.

The GMU Geocrowdsourcing Testbed consists of a Web application (see Figure 14) developed with HTML5, JavaScript, JQuery, CSS, and PHP and a mobile data-reporting tool built with JQuery. The transient obstacle data contributed to the system are passed to the server side using Ajax and stored in PostgreSQL tables. The GMU Geocrowdsourcing Testbed has visualization capabilities built with JQuery and routing functionality from ArcGIS Server. An important part of the system is a detailed PostgreSQL gazetteer database of building names, street names, and landmarks, used with address points for geoparsing and georeferencing user-supplied text descriptions.

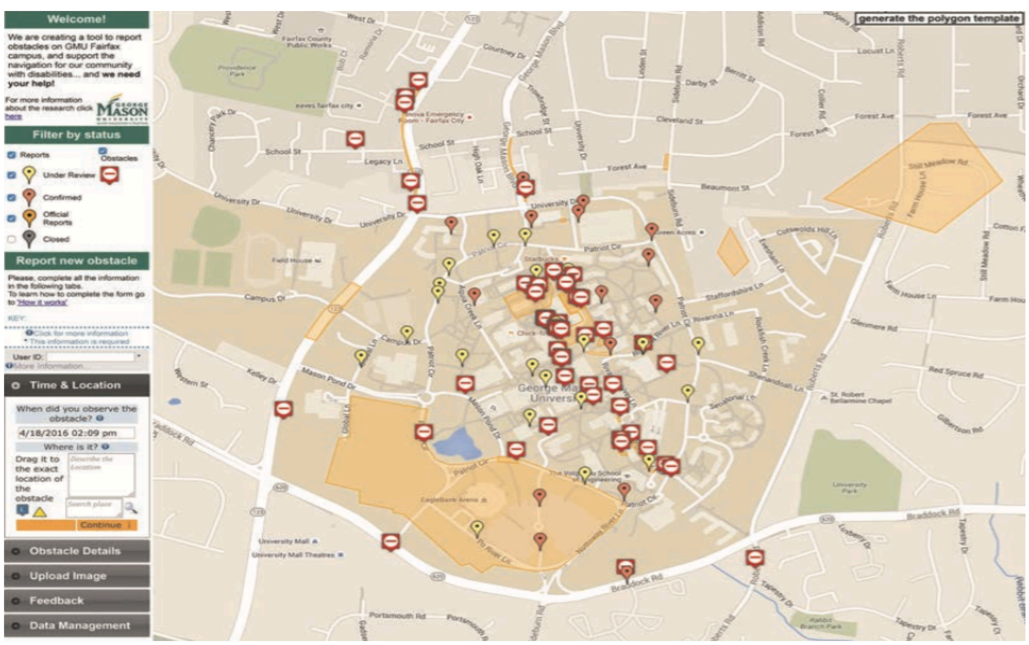

Figure 14 The GMU geocrowdsourcing testbed (http://geo.gmu.edu/vgi, accessed 15 November 2014) (Rice et al. 2016, 58)

5.3.2. The reporting and moderation process

When a contributor identifies an obstacle they would like to report, either they will locate the obstacle on a map by click-dragging a locator icon to the correct location, or their mobile device will provide GPS coordinates if they are using the mobile app. Contributors are also expected to provide a text description of the location, as well as obstacle details such as the type of obstacle and expected duration, and then asked to provide an image. Since the launch of the Testbed, over 200 potential users have been trained to generate a total 356 reports (Paez 2014). Following submission, these reports are field-checked and moderated by project staff, quality-assured using a quality assessment metric, and then clustered and used to generate obstacles.

5.3.3. Quality assurance of geocrowdsourced data

As of November 2014, the GMU Geocrowdsourcing Testbed has generated a total of 88 quality-assessed obstacles throughout GMU's campus and nearby Fairfax City. We first implemented the quality assessment metrics in May 2014 for the purpose of a pilot study conducted as part of a master's thesis (Pease 2014). In this pilot study, 23 students were asked to contribute obstacle reports using the Web-platform version of the tool on an iPad. This allowed us to gain insight into the georeferencing capabilities of end users using a base map containing buildings and walkways, as well as the consistency of our quality assessment statistics and their ability to assure the quality of geocrowdsourced data.

In efforts to gain more understanding about the quality assurance of position, we will further examine the positional accuracy of the reports compiled in the pilot study.

We will also use the geocrowdsourced data from our current testbed and examine the positional error of location data obtained in four ways: (1) through visual selection of the location by the end user using the Web platform, through click-dragging the locator icon; (2) through the GPS coordinates generated by the mobile device, when using the mobile platform; (3) through the geotagged coordinates from images when provided by the end user; and (4) through geoparsing of text descriptions of the location, using a gazetteer consisting of street names, building names, and landmarks. Exact obstacle positioning will be obtained either through method 1 or 2, depending on the method of contribution. Inclusion of an image is optional (method 3), though, when provided, it provides insight into an obstacle's location, given some offset since the geotag is based on the photographer's location.

Figure 15 Pilot study route (left) and reported obstacles (right) (Rice et al. 2016, 59)

5.3.4. Crowdsourcing transient obstacles: pilot study

In May 2014 we conducted a pilot study in which we first implemented our quality assessment statistics. Twenty three students were asked to contribute and were individually guided through a predetermined loop on campus containing construction projects, poor surface conditions, and sidewalk obstructions (see Figure 15). The students were instructed on how to use the map-based data collection tools and given an iPad to create four or five reports, though a handful of students submitted up to eight reports.

At the end of the study, 128 reports had been generated, though the moderators deemed 87 of those reports to be useful. The remaining 41 reports were not used owing to their lack of adequate positioning and missing text descriptions, which made them impossible to match to any obstacle when field-checked. The reports that contained useful information were matched to an actual obstacle location, and the positional error was calculated and stored in a PostgreSQL database. The Haversine formula for determining the distance between two points on a sphere was used to calculate the positional error of reported obstacles with respect to the actual obstacle location, as determined by project moderators[4]. Project moderators conducted an extensive field checking and moderation process to create ground truth against which the 87 reports could be compared. This ‘‘social moderation’’ process, discussed by Goodchild and Li (2012), has produced good results for our project and is discussed in detail by Rice et al.

[4] At the scale of this project (and at the associated distances) the curvature of the Earth makes no significant difference to more standard planar distance calculation methods. The Haversine formula was used in the codebase so that future extensions of this project for much larger geographic areas would not be negatively impacted. With efficient coding, this implementation is not significantly slower than traditional distance calculation methods.

(2014).

Positional accuracy characteristics for the 87 obstacle reports from the pilot study show a mean positional accuracy of 22.5 m and a median positional accuracy of 15.6 m, which appears to be much higher than similar findings by Haklay (2010), Girres and Touya (2010), and others (see Figure 16). In his study of positional accuracy for OSM data in the UK, Haklay (2010) reported positional accuracy to be close to 6 m.

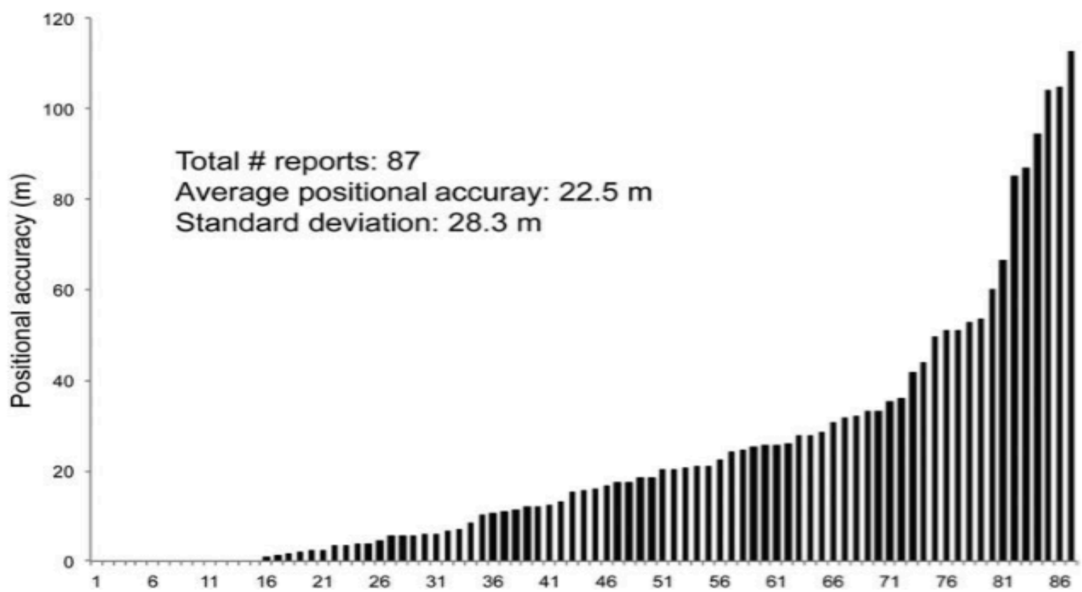

Figure 16 Positional accuracy of pilot study obstacle reports (Rice et al. 2016, 60)

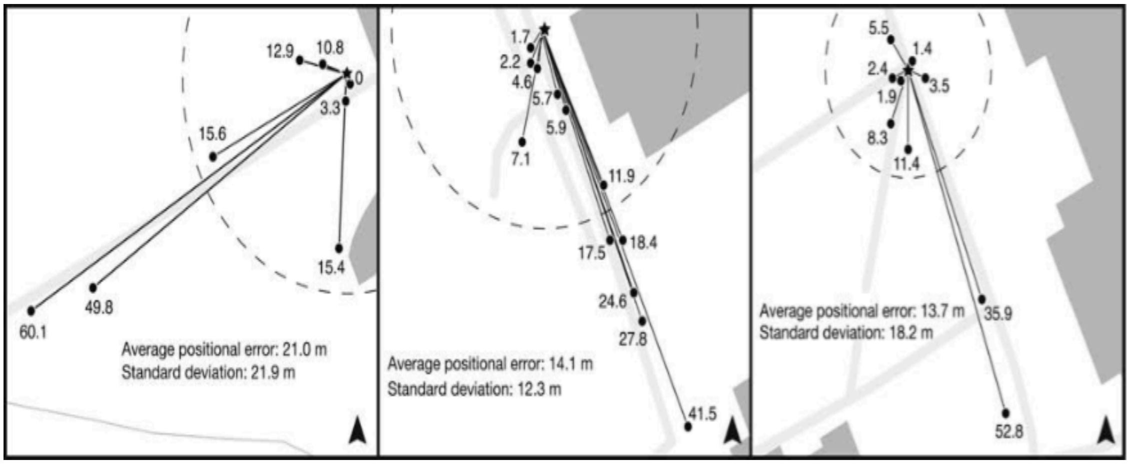

Figure 17 Average positional errors and report distributions for three pilot study obstacles (Rice et al. 2016, 60)

The high positional error encountered in this pilot study can be attributed to two issues: (1) contributors were using an iPad and had to click-drag the locator icon using their finger as opposed to a computer mouse, which can be tricky if the end user is not acclimated to using a touchscreen, and (2) the student contributors were not personally invested in contributing to this project but were participating for course credit. This lack of personal investment or interest in geocrowdsourcing or the goal of the project defeats a premise of citizen science, though an external incentive for participants is not unheard of in VGI projects. Brown et al. (2014) discuss a VGI project in which contributors were offered a $10 gift card as an incentive to crowdsource potential areas of interest in the Sierra National Forest. However, Goodchild (2007) notes that the personal agenda of the end user who is contributing is tied to the quality of information provided, and this pilot study accentuates that detail – lack of an invested interest yields low-quality data. Coleman et al. (2009) also address the motivations of contributors to projects using VGI, including altruism and social reward, and present a useful taxonomy of VGI contributors.

Three obstacles that were field-checked by moderators before the pilot study were also heavily reported by student contributors, with 8, 9, and 12 reports associated with each (see Figure 17). For these three obstacles, we calculated the average positional error associated with the aggregated obstacle reports (21.0 m, 14.1 m, and 13.7 m) and found these error levels to be consistent with the positional error of the overall data set. Figure 17 shows the position and distribution of the individual reports for each obstacle, with the average error shown as a bounding ring. The presence of two or more outliers in each case is clearly seen (confirmed by position errors between one and three standard deviations away from the mean), as is the clustering of five to seven reports tightly around the true obstacle position, at a distance of less than 10 m. The average positional error for reports shown in Figure 17 might be entirely consistent with positional error reported by other authors (Haklay 2010; Girres and Touya 2010) with removal of the two to three outliers. This removal of outliers may already happen in projects such as OSM and Waze, where data are being contributed and edited collectively, whereas in studies such as ours the outliers are preserved.

5.4. Position validation in geocrowdsourcing of transient obstacles

Positional accuracy characteristics for all current obstacles in our GMU Geocrowdsourcing Testbed are show in Figure 18, both for all 88 current obstacles (top) and for a reduced set of 84 obstacles where anomalous obstacles with unusually high positional errors have been removed. These reports with unusually high positional errors are, upon inspection, due to user mislocation of a report's location icon. The entire set of geocrowdsourced obstacles has an average positional accuracy of 22.7 m (consistent with the results of the pilot study). Removal of outliers (see Figure 18, bottom) results in an average positional accuracy of 3.0 m, which is similar to, and perhaps better than, those reported in other studies (Haklay 2010; Girres and Touya 2010). While the Haklay (2010) and Girres and Touya (2010) studies focused on OSM data, which are generated by a much larger contributor base, their studies provide a useful benchmark for comparison.

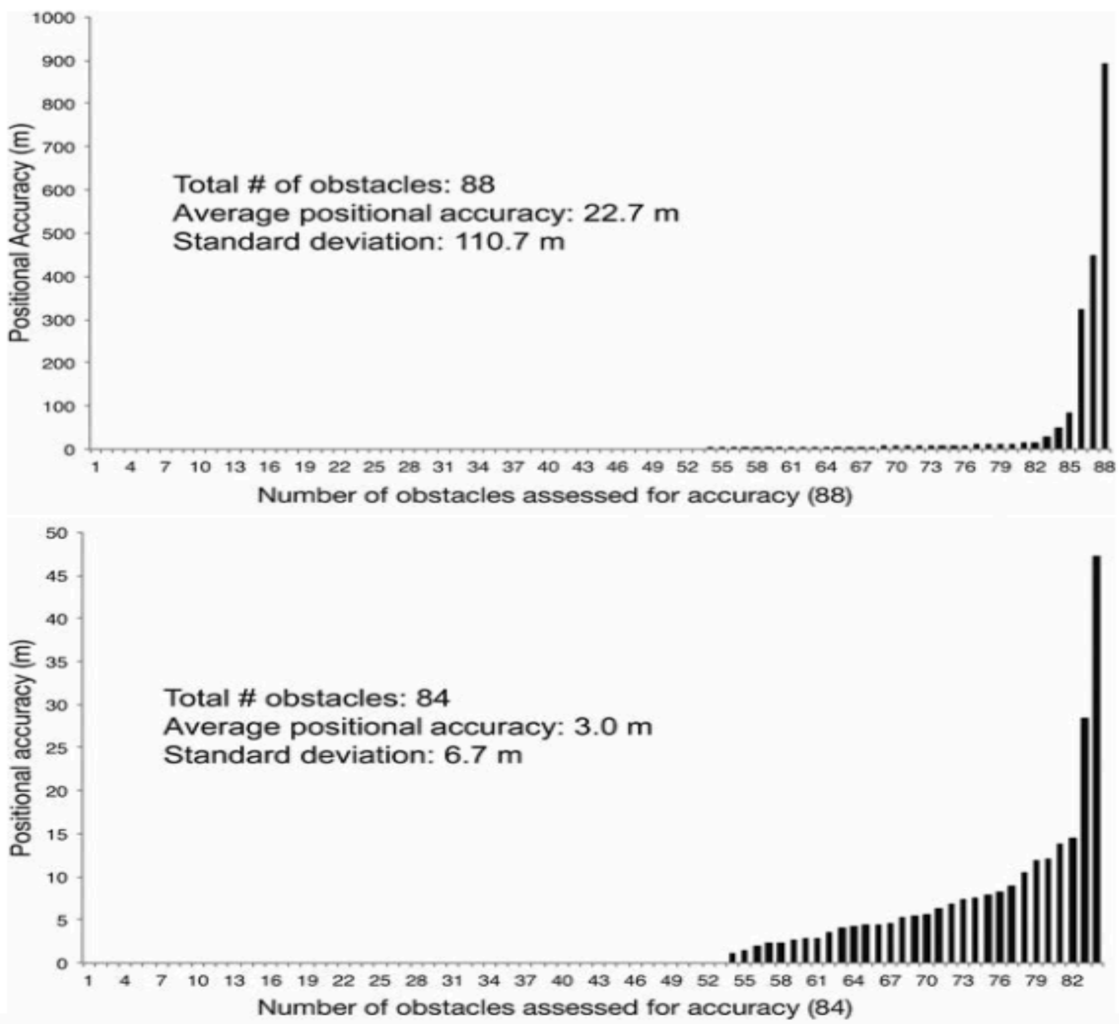

Figure 18 Positional accuracy for 88 GMU Testbed obstacles (top) and with outliers removed (bottom), collected in November 2014 (Rice et al. 2016, 61)

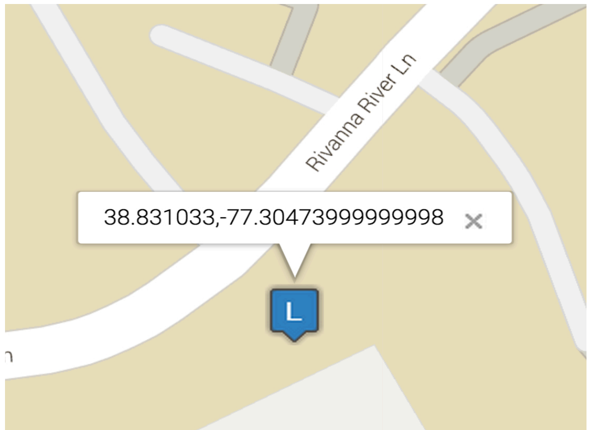

Figure 19 Locator-icon use to assign position to reports (Rice et al. 2016, 62)

Each transient obstacle report contributed to our GMU Geocrowdsourcing Testbed has a possibility of four separate methods of positioning that can be used for cross-checking and validation. First, reports contributed through the desktop data contribution portal are positioned by the user with a locator icon (see Figure 19) or, alternatively, with the user's mobile device GPS if the mobile data collection tools are being used. When relying on a mobile device's GPS, it is critical to be aware of associated positional inaccuracy, which can range from an average of 10 m on WiFi to 74 m on a cellular network (Zandbergen 2009). Second, the user provides an image from a mobile device, which contains embedded geotags that can be extracted and saved. Third, the user provides a text-based description of an obstacle's position, which can be geoparsed for place names and the creation of a spatial footprint (Hill 2006, 100–103). Fourth, every obstacle report that is contributed to the GMU Geocrowdsourcing Testbed is subjected to a moderation review that establishes a ''groundtruthed'' version of the position. This fourth position is established by a team of staff moderators with extensive training, and it represents the best of the four methods for establishing an obstacle report's position. Collectively, the four methods for establishing position allow for a robust, redundant methods for validation that include replacement (where the other positioning methods do not exist) and the ability to cross check for positional consistency.

Early efforts to validate obstacle positions through geoparsing are shown in Figure 20, which uses a detailed gazetteer of local feature names to automatically geoparse toponyms from obstacle location descriptions. A detailed spatial footprint, in the form of a convex hull, is created using PostGIS and displayed on a map. This work, published by Rice et al. (2011, 2012a), is notable for including slang, abbreviated, colloquial, and foreign-language-based variant toponyms and for being executable in real time.

More recent work, shown in Figure 21, extends previous geoparsing and spatial footprints to include options for both convex and concave hull geometry for geoparsed place names, produced with PostGIS. However, the footprints from this work are limited to polygon-type features such as buildings and major landmarks.

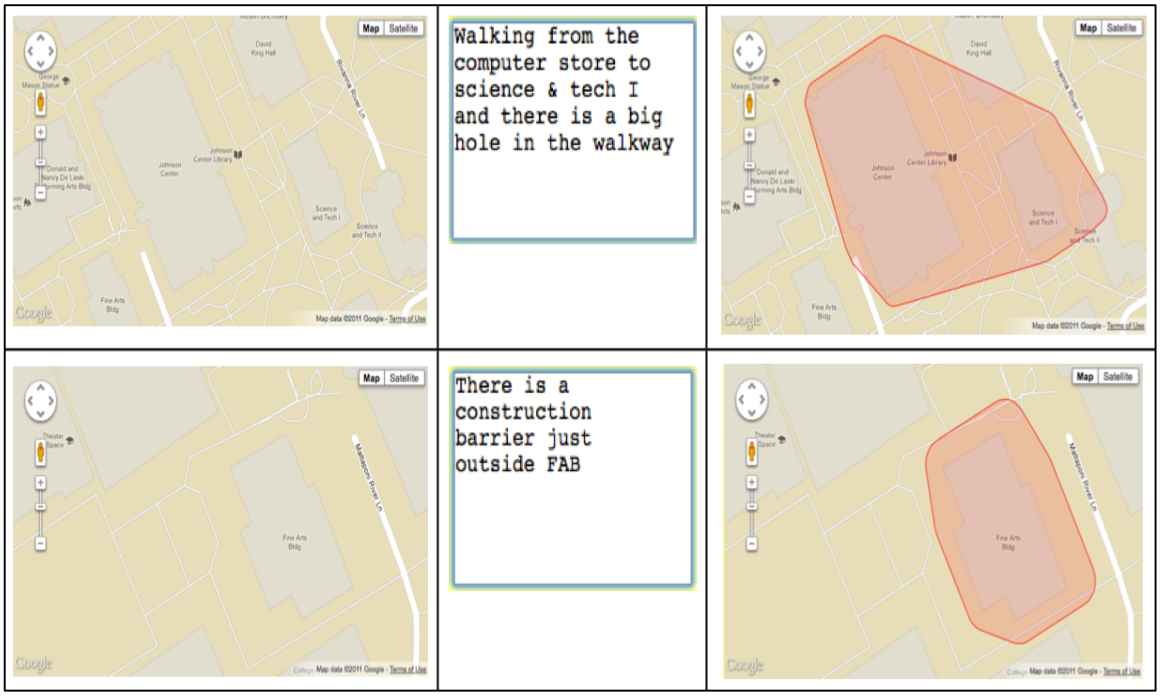

Figure 20 Obstacle footprints from geoparsed obstacle text (Rice et al. 2012). The top example shows a footprint around the Johnson Center, where the computer store is located, and Science & Tech I, to include the walkway in between. The bottom example shows a footprint around the Fine Arts Building (FAB) (Rice et al. 2016, 62)

Figure 21 Convex (left) and concave (right) hull geometries for "Fenwick Library" and "Krug Hall" (Rice et al. 2016, 63)

Figure 22 shows our current work on validating the four versions of geocrowdsourced obstacle positioning, building on the positional accuracy assessment techniques, image geotag extraction, obstacle location text, and associated spatial footprints from geoparsed toponyms. Currently, the methodology includes linear features in addition to polygon features, enabling the creation of spatial footprints around road intersections, as well as a segment between two roads. The three figures contain a red (grey in print) convex hull built automatically from geoparsed toponyms in the contributed obstacle location description, a yellow dot (light grey dot with white border in print) showing the image geotag location, a blue dot showing the original report location as established by the user, and a green dot showing the obstacle report location as determined by the moderator. In each case, these three location characteristics fall inside the red convex hull, indicating consistency in positioning. Distances between the points are displayed in a table along with the obstacle description text and can be used to assess validate and assess the quality of positioning for each report.

All the computational aspects of this work are implemented with HTML5, PostGIS/PostgreSQL, CartoDB, and MapBox. The spatial functionality is achieved with the PostGIS spatial functions inside the SQL statements. The communication with the database is through the CartoDB.js library, which utilizes node.js as its base for server-side communication.

5.5. Summary and conclusion

The GMU Geocrowdsourcing Testbed has been created to facilitate a more accessibly university campus. It was built with past assistive geotechnology exemplars and VGI applications as examples. The system is designed to explore the contributions that crowdsourcing and VGI can make in capturing transient obstacle information, which is a significant problem in the local area owing to construction and growth. The information contributed to this system through crowdsourcing has a level of positional accuracy (excluding outliers) similar to that reported by notable studies of VGI quality (Haklay 2010; Girres and Touya 2010). The four primary location parameters for each contributed obstacle can be analyzed for consistency using geoparsing, spatial footprints, and distance measurements, which help the authors determine the usefulness and reliability of the information and, ultimately, explore the dynamics and limitations of geocrowdsourcing, an important emerging trend in cartography and GIScience.

While this research employs a method that works for a local study area and small data set, it is important to consider the bigger picture. Other universities and communities can gain insight from this research and utilize geocrowdsourcing to improve their own accessibility systems. However, as the size and complexity of geocrowdsourcing projects increase, different quality assessment techniques must be considered. Deployment of a geocrowdsourcing testbed similar to the one presented in this article, as well as future studies of positioning, could adopt a geocrowdsourced moderation paradigm, which replicates the structure of larger projects such as Wikipedia, where a hierarchal structure of moderators is used for checking content. Future development of the testbed, including expansion to larger geographical areas, will include the implementation tools that allow users to validate and refine the position of pre-existing reports and view the resulting improvement in positional accuracy. Future research implementing geocrowdsourcing will benefit from lessons learned and quality assessment practices from earlier efforts such as this one and ultimately will provide greater benefits to end users and communities of interest.

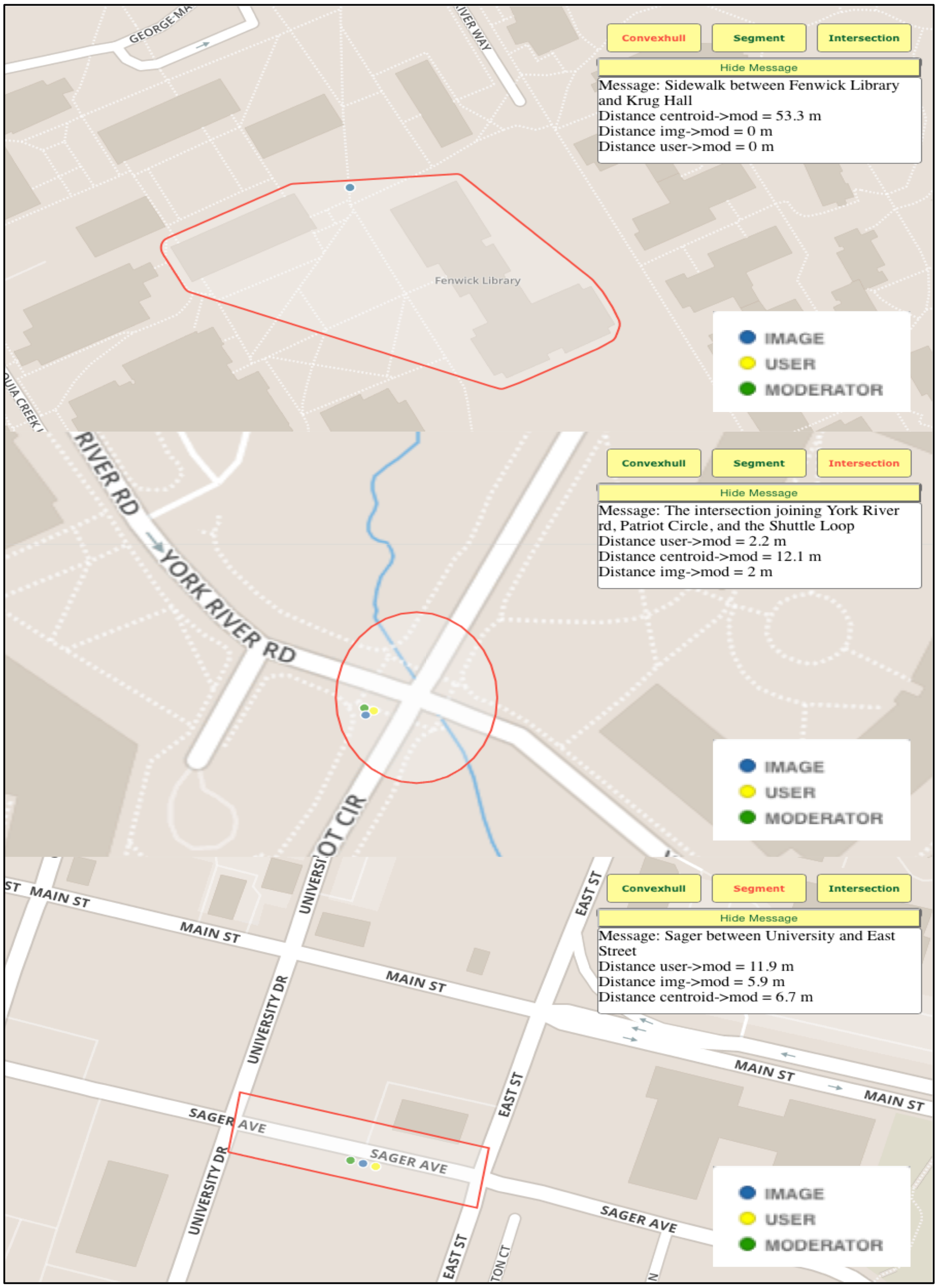

Figure 22 Geoparsing location text: footprints and obstacle locations for three examples, including a convex hull for polygon features (such as buildings and major landmarks) and spatial footprints that include geoparsed linear features, such as street intersections and road segments between two streets (Rice et al. 2016, 64)

Acknowledgements

We thank Ms. Patricia Pease for the collaborative sharing of data with our project. This material is based on work supported by the United States Army Corps of Engineers, Engineer Research and Development Center, under Broad Area Announcement no. AA10-4733, contract no. W9132V-11-P-0011. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Army Corps of Engineers. Mr. A. Aburizaiza is funded by the Department of Geomatics, King Abdulaziz University, Jeddah, Saudi Arabia.

6. GEOSPATIAL FOOTPRING LIBRARY OF GEOPARSED TEXT FROM GEOCROWDSOURCING

Abstract

The research paper reports on the generation of geospatial footprints from geoparsed text associated with geocrowdsourced platial data collected and stored in the George Mason University Geocrowdsourcing Testbed (GMU-GcT). The GMU-GcT facilitates study of social dynamics, quality assessment, data contribution patterns, and position validation for geocrowdsourced spatial data, with a primary purpose of mapping transient obstacles and navigation hazards in a dynamic urban environment. This paper reports on the automated generation of geospatial footprints using open-source software, and discusses the role of automated geospatial footprints in quality assessment for automated position validation. A detailed, local gazetteer is used to store placenames and placename variants including abbreviated, slang, former, and jargon-based instances. Obstacle reports containing location descriptions are geoparsed and processed with the help of the GMU-GcT gazetteer to generate geospatial footprints, which are used in a quality assessment process to validate the position of obstacle reports. Continuing research with the GMU-GcT has produced fifteen characteristic footprints types, which are generated and grouped into simple, complex, and ambiguous categories. The open source tools used for generating these general footprints include MapBox, TURF.js, jQuery, and Bootstrap.

Published in 2016 at the Spatial Information Research Journal, volume 24(4) (409-420). Authors Aburizaiza, A.; and Rice, M.

Keywords
Geospatial footprints, Gazetteers, Geoparsing, Volunteered geographic information validation, Open source GIS

6.1. Introduction

George Mason University, is the largest public university in the Commonwealth of Virginia, hosts 33,000 students and several thousand faculty and staff, in a dynamic urban environment outside Washington, D.C. The campus, adjacent to the City of Fairfax, is the site of near-constant construction and expansion. This changing urban environment presents difficulty for students, faculty, and staff who are visually or mobility-impaired and depend on familiar navigation pathways to get to and from work or home. Construction barricades, sidewalk obstructions, and detours are commonplace. The temporary, unplanned nature of these disruptions makes them nearly impossible to capture and map using traditional GIS workflows, and a crowdsourced approach is one of the only practical ways to provide the information in a timely manner. Geocrowdsourcing and volunteered geographic information (VGI), introduced by Goodchild (2007), and reviewed by Elwood (2008a), Haklay (2010) and others, represents an opportunity to extend traditional mapping through open-source software and novel geocrowdsourcing workflows. The GMU Geocrowdsourcing Testbed (GMU-GcT), developed by Rice et al. (2012a, 2012b, 2013a, 2013b, 2014), introduces a comprehensive geocrowdsourcing approach for collecting and quality-assessing transient obstacle data. Related work by Karimi et al. (2014) provides a useful look at how the accessibility domain benefits from novel open-source mapping applications and data modeling. The GMU-GcT is based on the early work in geoparsing, gazetteers, and geocrowdsourcing (Rice et al. 2012a, Aburizaiza 2011, and Rice et al. 2011) and has been extended to include quality assessment, routing, and visualization (Rice et al. 2013a, Rice et al. 2013b, and Rice et al. 2014).

Contributors to the GMU-GcT include students, faculty, staff, and members of the public, who submit obstacle reports containing the location, the basic characteristics, and images of the obstacles (Figure 23). This information is processed in a preliminary quality assessment procedure and displayed on a map as a provisional obstacle report. Student moderators field check the reports and provide a comprehensive quality assessment through ground truth. The valid reports are maintained in the system and displayed to the public as confirmed reports. Each report is maintained in the system as long as it remains relevant. The relevancy is based on the estimated duration of the report as well as the reviewer's quality assessment procedure. During the preliminary quality assessment and moderator field check, geospatial footprints are created. Geospatial footprints are polygonal representations of the combined locations of platial terms from the contributor's text-based contributions (Figure 24). These geospatial footprints are discussed in more details later in this paper.

6.2. Position validation and location description text in the GMU-GcT

Qin et al. (2015) present the quality assessment procedures in the GMU-GcT, including assessments of location, time, and attribute, the three primary facets of the atomic view of geographic information noted by Longley et al (2011) Modeled after the comprehensive quality assessment of geocrowdsourced data by Haklay (2010), Qin et al. (2015) develop assessments of position and categorical attribute agreement, and use this quality assessment procedure to create a composite score for geocrowdsourced data.

Rice et al. (2016) discuss the details of the positional validation procedures in the GMU-GcT, by introducing a concept of multi-position validation in the GMU-GcT. Reports contributed by the public are checked for position through three comparisons. First, they are compared to moderated ground truth, where the mapped position of a report contributed by the end-user is compared with the location established through a moderated field check. Second, the images contributed by the end-user are processed to extract embedded geotags and orientation data, which can be combined with geotags and orientation data from multiple reports to create a footprint based on image geotags. Finally, the location description provided by the contributed data is processed to extract placenames, prepositions, distances, directions, landmarks, addresses, and other feature names.

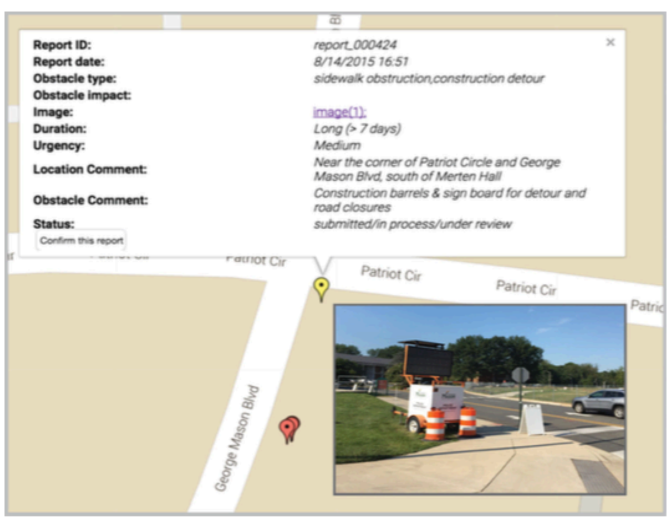

Figure 23 A GMU-GcT obstacle report with image (Aburizaiza and Rice 2016, 410)

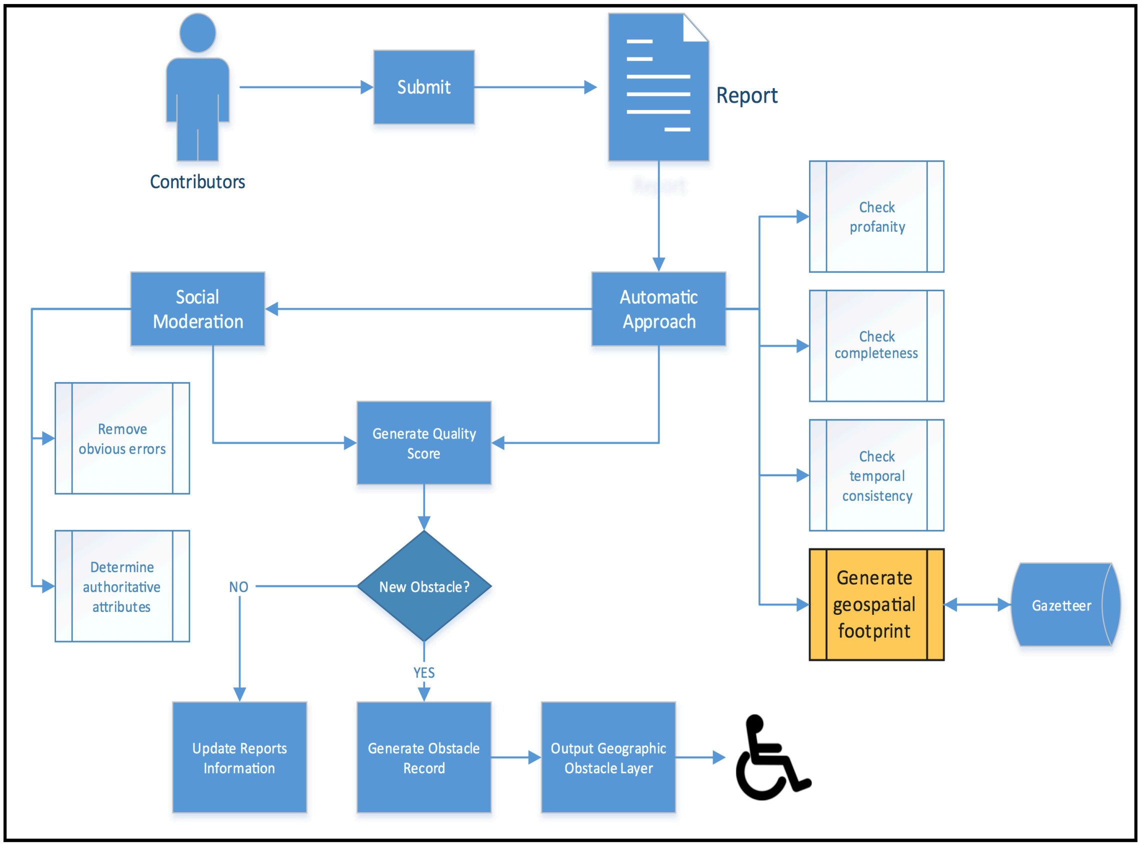

Figure 24 Generation of geospatial footprints within the GMU-GcT quality assessment workflow (Aburizaiza and Rice 2016, 411)

This geospatial footprint developed from geoparsed placenames and other information is a valuable tool for position validation. When the user-contributed map location and the moderator location established through field check do not coincide with the features named in the location description, the inconsistency can be used to flag the report for closer inspection. In such cases, the geospatial footprints reinforce the validity of the location. Earlier work by the authors Aburizaiza et al. (2011), Rice et al. (2011, 2012a) used this general technique, but used the simplest cases based on proximity to a named feature. The work was extended by Rice et al. (2016) to include intersection points from two named linear features, a convex hull formed by two named polygonal features, and a linear segment cut by two named linear features. These general cases were developed and implemented so that any crowdsourced contribution to the GMU-GcT with a location description containing named features with this configuration could have a geospatial footprint developed automatically. The ongoing recent work in this paper extends this general technique to include several more cases of general named or platial feature layout, as contained in the obstacle description text in the GMU-GcT. A platial, or place-based perspective on geographic information, is articulated by Gao et al. (2013), Miller and Goodchild (2014) and others, who recognize the importance of location expressed through placenames, in social media and non-traditional data sources that can be used in addressing geographical problems.

Location text descriptions in the GMU-GcT contain references to one, two, three, or more places, and these place references can be in various forms including abbreviations, slang, former (old) names, and colloquial variants of standard names. Many of these name variants are contained in the GMU-GcT gazetteer. In additional to placenames or variations of placenames, the GMU-GcT frequently contains location descriptions with directions, distances, and spatial prepositions. Moreover, a description can contain features that have a distinct group identity and associated name, or a named feature contained within another named feature, or an unnamed place, e.g., walkways, with a geospatial relationship to another named feature.

The combinations of features in the GMU-GcT's location description text is understandably complex. This paper focuses on extending the geoparsing capabilities based on simple and complex placename or platial orientations to generate more cases of geospatial footprints, and instantiating a reference library of the footprints for future research. Currently, fifteen different cases are identified, and more are being discovered. In the methodology section below, the cases are explained in three categories: simple, complex, and ambiguous. The GMU-GcT gazetteer used in this work is actively updated to reflect naming changes and instances of abbreviation, slang, and colloquial place references. The individual entries for the gazetteer are stored as a JSON array value stored in the GeoJSON properties.

6.3. Methodology

A Web application was developed to run and test the geoparsing algorithms for the fifteen cases. The application was built as a mobile Web application formatted to permit mobile and tablet users to utilize it through their mobile browsers. HTML5, jQuery, and Twitter's Bootstrap were used to build the interface and adjust the site components according to the screen size.[5]

MapBox is a well-known and powerful Web mapping technology with a fast map tiling service. MapBox has a JavaScript Library called MapBox.js which permits programmers to build Web mapping applications for geospatial data visualization. It also allows programmers to customize its map object with the essential tools.

MapBox uses the Turf.js JavaScript Library, which is a geospatial analysis library

[5] The codebase for this application is archived at: A link to dissertation code and applications can be found at: http://geo.gmu.edu/archive/aburizaiza

that can run on client side, server side, or both. It uses the GeoJSON format as the input and the output. It is a very rich library and runs efficiently since geospatial analysis can run on the client side without connecting to the server.

MapBox.js and Turf.js were used to build and visualize the geospatial footprints after extracting place names, prepositions, and bearing words (north, south, southwest, etc.) using jQuery and AJAX. Currently, campus data is stored in a GeoJSON file residing in the server. The plan in the future is to store the campus data in a MongoDB database since MongoDB stores data in JSON format rather than relational database table format.

The geospatial footprints developed are categorized into three categories: simple, complex, and ambiguous. Each category is explained in detail. Some cases are similar in concept but are different in the algorithm structure. Such cases are explained together while their differences in the algorithm are covered later.

6.3.1. Simple geospatial footprints

The first simple case (Figure 26) is only one point or one polygonal place mentioned in a text message with no preposition nor bearing. Some landmarks are digitized as either point or polygon features. Other polygonal places are buildings or groups of buildings. The point or the polygon is buffered and then both the place itself along with the buffer are plotted on the map. The only difference between point and polygon cases is the zoom level set after the geoparsing process is finished.

The system also creates a Bootstrap dialog box (Modal Object)[6], similar to JavaScript's alert, informing the user that one landmark, or building, etc., was found and its name used in the message. Also, the dialog box informs the user that there was no preposition or bearing words in the message. In other words, the algorithm's detailed steps are explained to the user. This explanatory Bootstrap dialog box is illustrated in each footprint case. It also notifies the user if no text was entered, or if place names are not found in the message. Figure 26 displays an example of the first simple case with its explanatory dialog box.

The next simple case (Figure 27) is finding only two names of intersected linear places. The algorithm also searches the message for terms such as ‘‘intersection’’ or ‘‘corner of’’. Even if two linear places are specified without intersectional terms, the algorithm would find the intersection of the two linear places. The explanation modal will clarify if intersection terms are found or not in addition to the two linear place names. The intersection point is buffered and then the two linear places, the intersection point, and the intersection buffer are all plotted on the map. Rarely, two linear places have two intersections instead of one. The algorithm also takes care of this and the Web application will zoom to the center point between the two intersections with smaller zoom level. Figure 27 illustrates two examples of this algorithm of one intersection and two intersections.

---

[6] The Bootstrap alert object is referred to as a Modal Object, which is a responsive, JavaScript alert object, and not associated with the statistical terminology. For more information, see https://v4-alpha.getbootstrap.com/components/modal/

6.3.2. Complex geospatial footprints

The first complex case (Figure 28) is extracting either one point-type place and one polygonal place or two polygons. The vertices of the polygonal places are extracted to an array, using a Turf function named ‘‘explode’’. In the case of a point and polygon, the point is added to the array afterward. A convex hull is created based on the array points and then buffered with a negative distance. The reason for the negative distance is to avoid areas not in between the two places, hence wrong obstacle location. In the case of one point and one polygon, the buffer gets clipped with the polygon since the location would be between the polygon and the point and not inside the polygon itself. As for the case of two polygons, the buffer is clipped twice with the two polygons. After clipping, the result is a multipolygon GeoJSON object. Only one polygon is extracted from the multipolygon object; it intersects with a linestring connecting the polygon’s centroid and the point feature or the two polygons’ centroids in case of two polygons. This polygon is the in-between polygon and it covers the possible location indicated by the user. Figure 28 explains the details of selecting the in-between polygon. The explanatory dialog box, as for all cases, describes the details of the algorithm steps and the place names found.

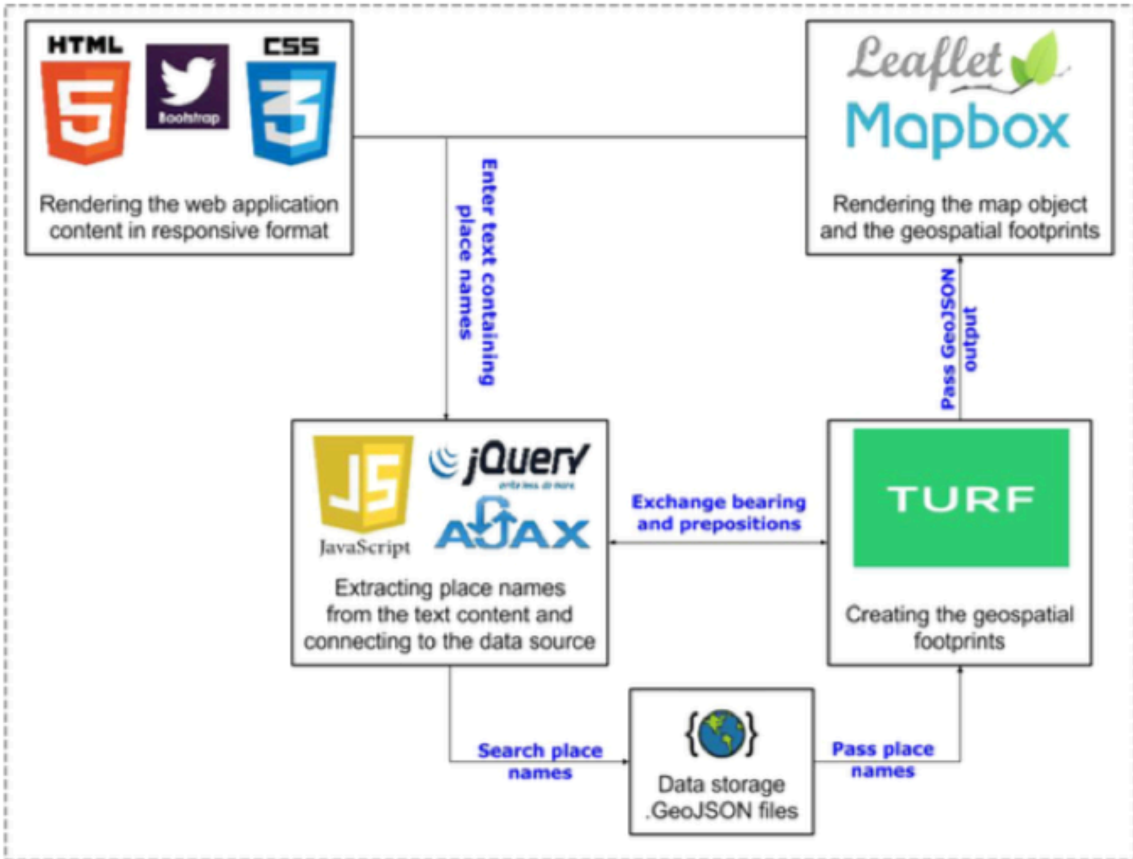

Figure 25 The web application architecture (Aburizaiza and Rice 2016, 413)

There are unnamed places such as walkways used to describe a location in relation to named places (Figure 29). The algorithm to select the unnamed places begins similarly to the previous case. Volunteers made several comments about obstacles on walkways between two polygonal places. The points of the two places are exported to an array. Then a convex hull is created and negatively buffered. The buffer is also clipped with the two polygonal places. The result is a multipolygon and the in-between polygon is selected using the linestring connecting the two places' centroids. The unnamed walkways are stored in a separate GeoJSON file residing on the server. Using jQuery and AJAX, the code iterates the walkways file and uses turf geospatial functions to collect the walkways within the in between polygon. Figure 29 describes an example of selecting walkways between two places with its explanatory modal. One quick note is that the official building names are David King Hall and Robinson Hall A. The gazetteer has other possible names such as slang placenames, jargon-based placenames, and colloquial or information names. Rob A is a jargon-based name used by students at GMU. King Hall is a common name not the official name.

The fourth complex case (Figure 30) is explaining a location in a linear place with proximity to a polygonal place. The algorithm starts with finding the nearest point on the linear places to the centroid of the nearby polygon. Currently the algorithm will select the previous two segments and the next two segments starting from the nearest point selected before. After that, the points of both the polygonal place and the extracted linear segments are extracted to generate the convex hull. The polygon of the polygonal place is then cut from the convex hull which results in a multipolygon object. From the multipolygon, only the polygon part between the polygonal place and the extracted segment from the linear place is plotted to the map. Figure 30 demonstrates the result of the following volunteer's comment: ''Directly in front of GMU commerce building, university Drive, chipped pavement along the sidewalk''.

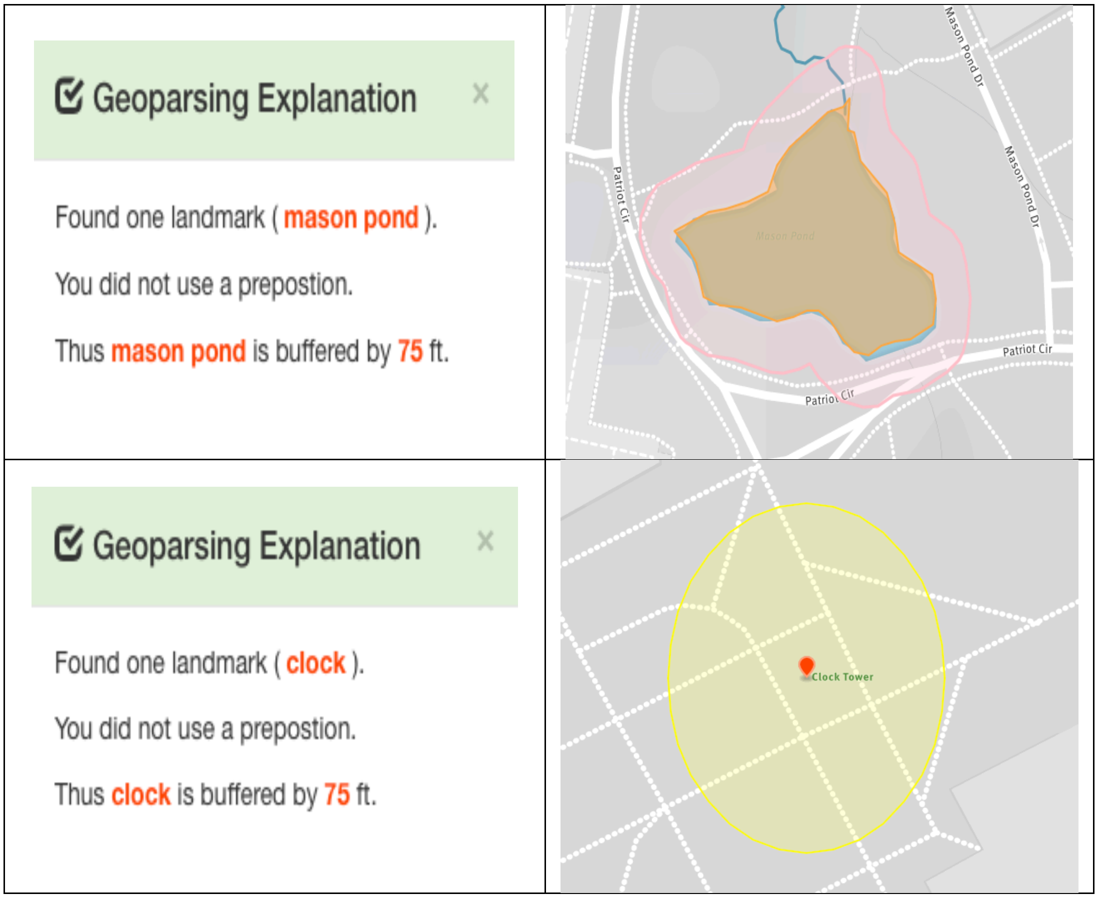

Figure 26 The first simple case of one polygonal or point-type place with no preposition or bearing words (Aburizaiza and Rice 2016, 414)

Finding a bearing word such as north, northwest, west in relation to a polygonal place is challenging (Figure 31). The algorithm begins by creating an envelope around the polygon. Then the width and length of the envelope are computed to calculate the diagonal of the rectangular envelope. This is again done through Turf.js. A point is created in the direction of the bearing word with half the diagonal as the distance. The point is then buffered, but the buffer could intersect with the polygonal place. If so, the algorithm will clip the polygonal place part out to give the final geospatial footprint. Figure 31 demonstrates the algorithm's results for three directions: north, northwest, and west, for comparison. The polygonal place part is clipped out in all three cases. A fourth example is also illustrated showing the use of a bearing word with a point type place.

The final complex case (Figure 32) describes a location on a linear place between two intersections with another two linear places. A common example message would be ‘‘An X obstacle on Main Street between 1st Street and 2nd Street’’. The algorithm first determines which linear place should be highlighted. With reference to the previous example, a volunteer could say instead ‘‘Between 1st Street and 2nd Street, I was driving on Main Street and saw X obstacle’’. The order of the linear places is different between the two messages. Different examples were tested and the algorithm was capable of determining the correct order. After the two intersections are created, the segments between them are selected and buffered as the footprint. The algorithm is still missing two scenarios, if Main Street is intersecting with 1st Street more than once, and if Main Street has a major curve between 1st Street and 2nd Street. Both cases are being implemented but not working yet. Figure 32 illustrates an example of this case. The original volunteer’s comment is ‘‘Sager between University and East Street, fractured concrete covered by plywood and orange cones...’’.

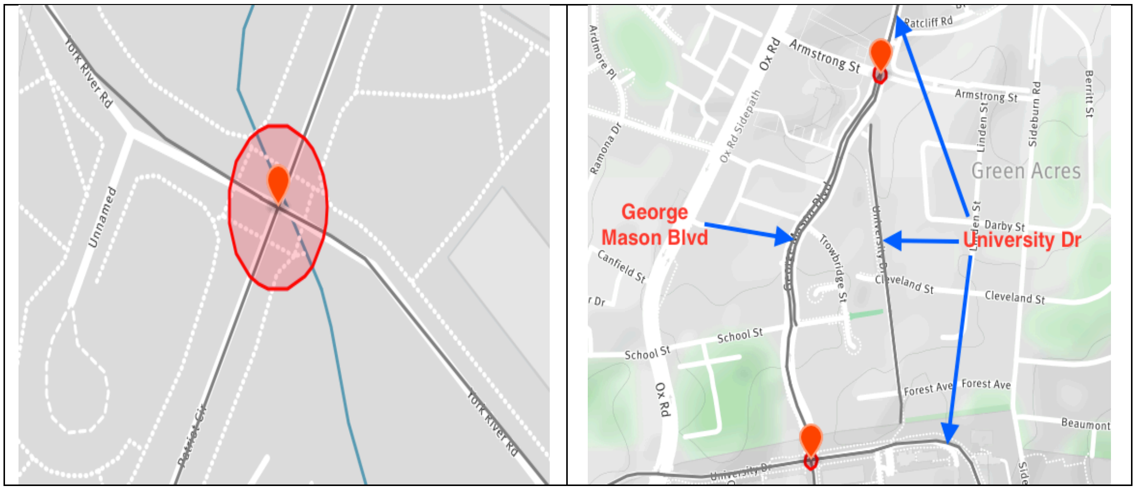

Figure 27 The diagram on the left shows one intersection between two roads. The one on the right demonstrates two intersections because one of the roads has three separate segments (Aburizaiza and Rice 2016, 415)

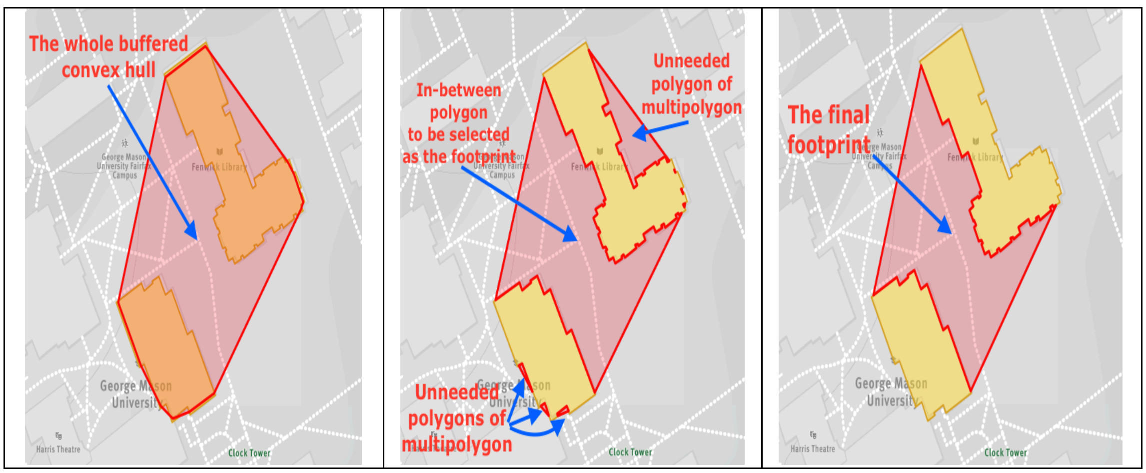

Figure 28 The final transitions in the algorithm of highlighting a location between two polygonal places (Aburizaiza and Rice 2016, 415)

6.3.3. Ambiguous geospatial footprints

There are many examples on the GMU-GcT where the location text and geoparsing code yield ambiguous results. Those cases will be discussed here. Finding a preposition with one point or one polygonal place is an ambiguous case. Currently the algorithm searches for prepositions in text and then accesses a JSON object that stores buffer distances based on the preposition found. The prepositions are categorized in proximity ranges. For instance, the preposition ‘‘next to’’ has a closer proximity than the preposition ‘‘near’’. Further research on spatial prepositions and natural language processing is needed in order to determine what processing steps should be undertaken in this case, and specifically, what buffer distances should be used to buffer the JSON object. A surveying of GMU-GcT contributors may be able to help determine what proximity is intended with certain spatial prepositions and some patterns may emerge. Figure 31 emphasizes different examples based on different prepositions. Rice et al. (2011) note that buffer distances associated with spatial prepositions may be related to factors such as visibility and lighting that are not captured in the GMU-GcT.

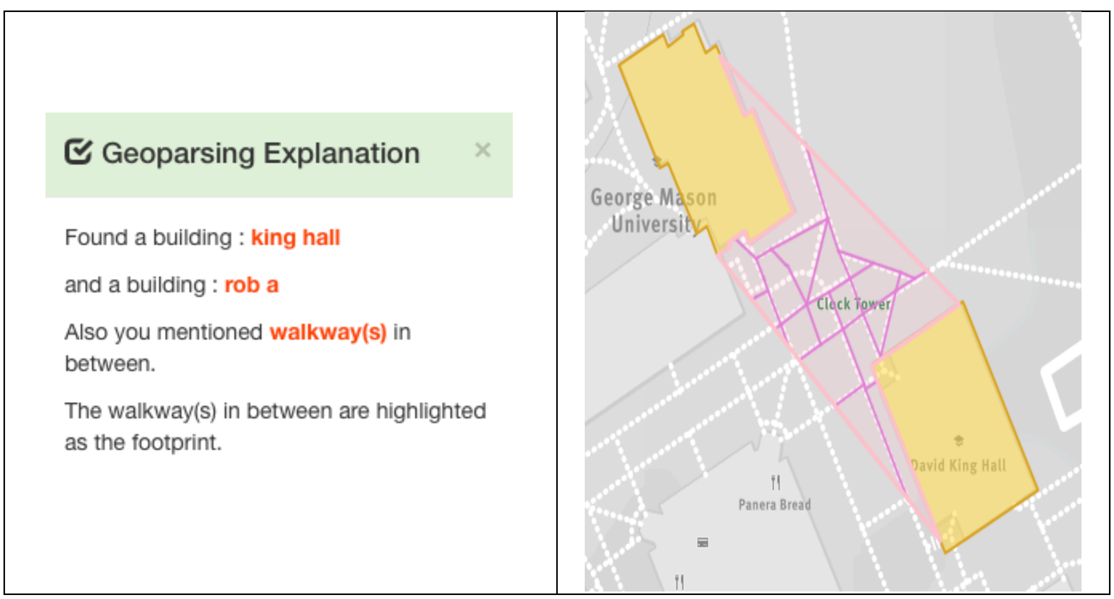

Figure 29 The footprint of walkways between Robinson Hall A and David King Hall (Aburizaiza and Rice 2016, 416)

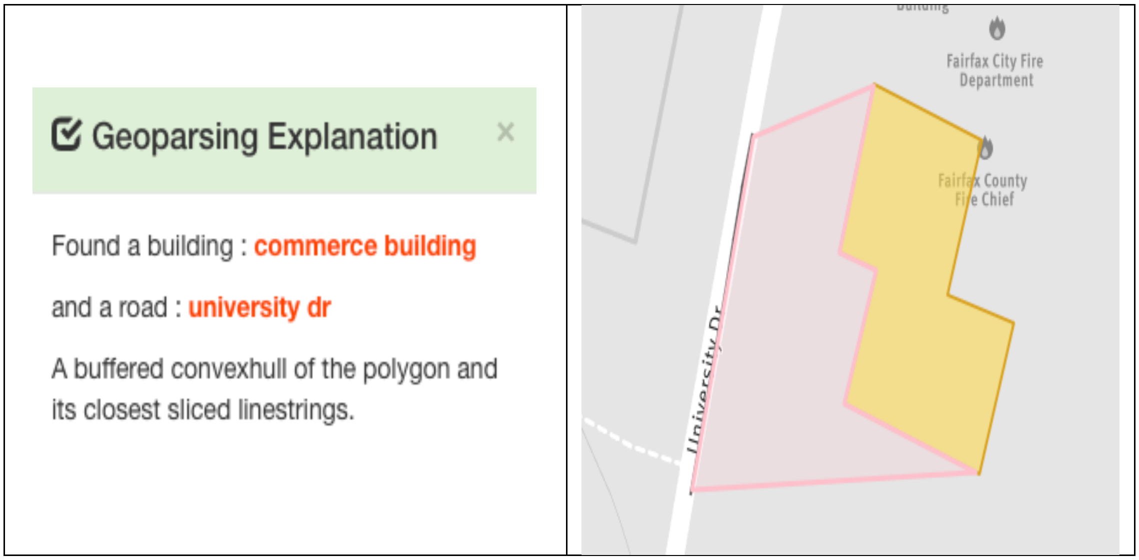

Figure 30 Creating a footprint between a linear place and a polygonal place (Aburizaiza and Rice 2016, 416)

The other ambiguous case occurs when a user defines a location by mentioning only one linear place: for instance, a volunteer reporting an obstacle on Main St. Main St in Fairfax VA is roughly 5 km long. The location of the obstacle cannot be determined unless the user gives another place name to specify the location. The code will inform the user about this information in a warning dialog box without generating the geospatial footprint. Figure 34 shows an example of such warning modal. This warning modal may be unnecessary or may be suppressed if additional positional information (such as an image geotag or user-asserted map position) is present. In these instances, the ambiguous spatial footprint provides only general confirmation that the obstacle is close to or associated in some way with the linear place (Figures 33, 34).

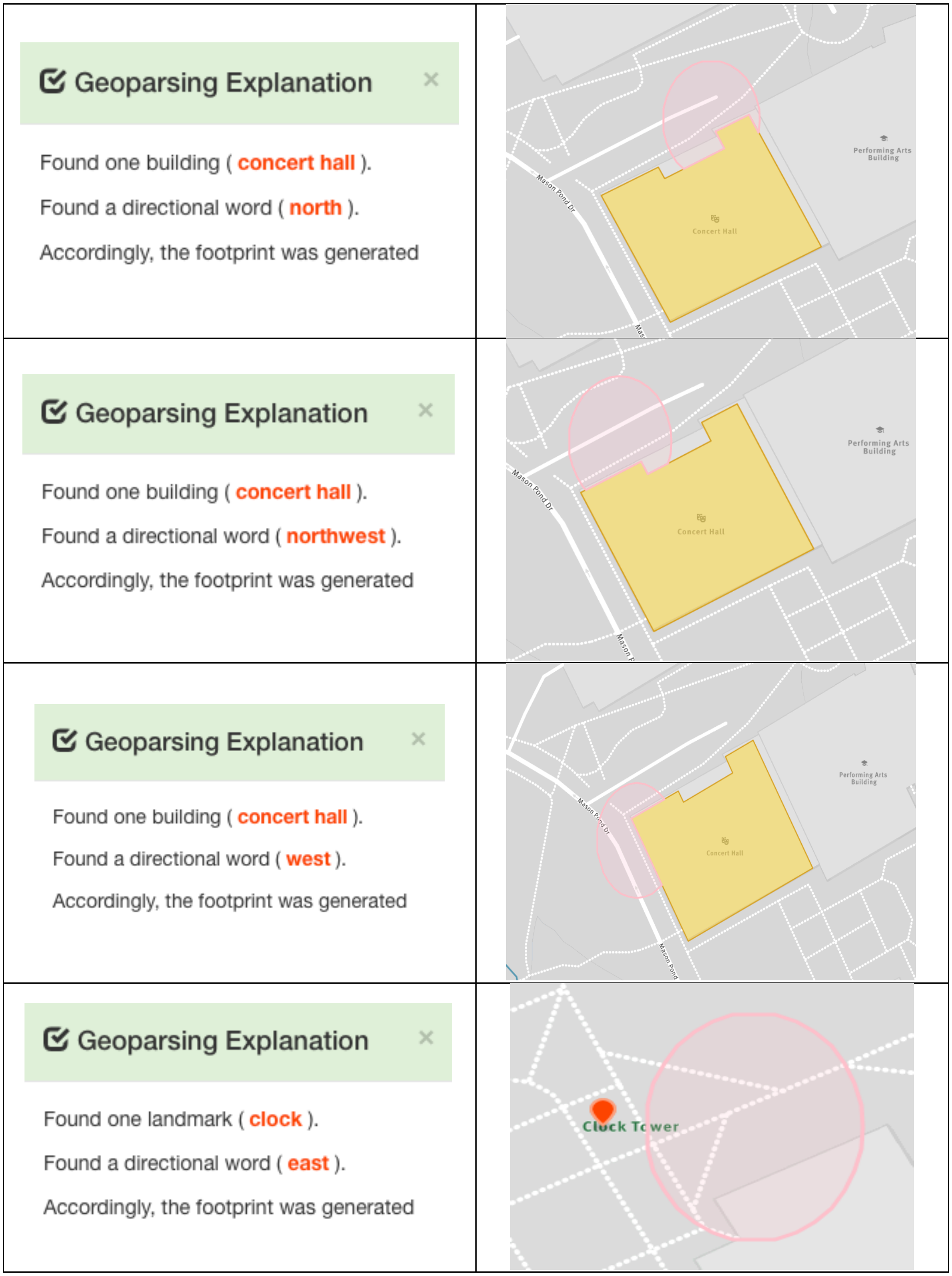

Figure 31 The bearing geospatial footprint in reference to polygonal and point-type places (Aburizaiza and Rice 2016, 417)

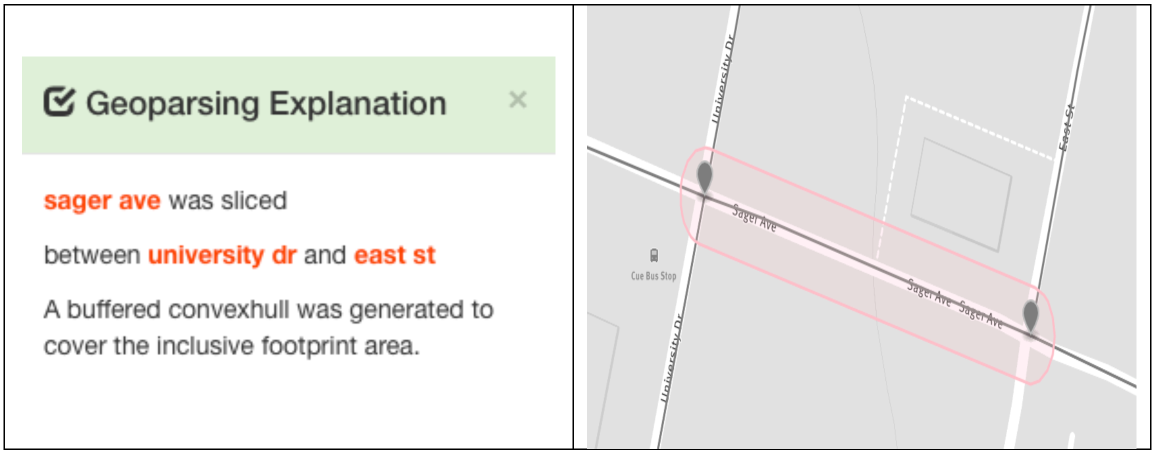

Figure 32 A footprint of a sliced linear place between two linear intersections (Aburizaiza and Rice 2016, 418)

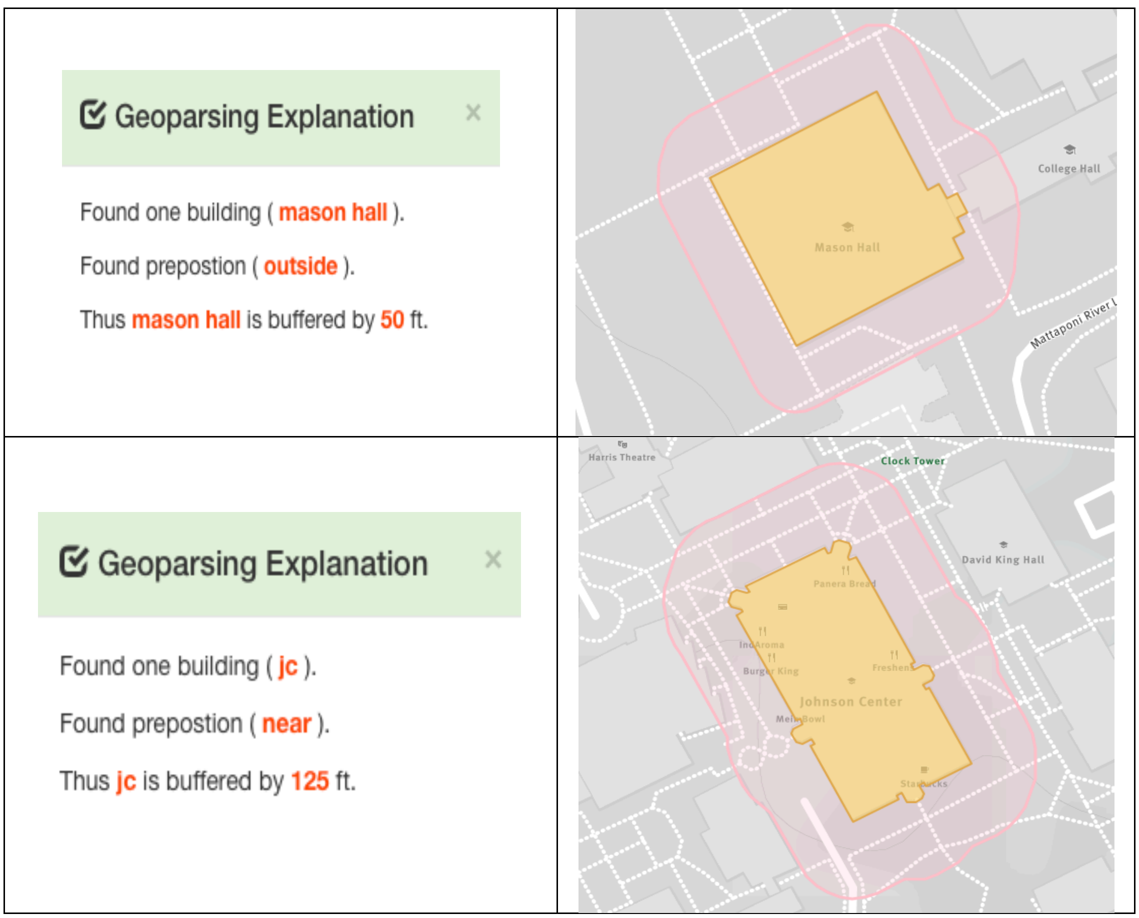

Figure 33 Examples of prepositions found in messages explaining proximity of place names (Aburizaiza and Rice 2016, 418)

Figure 34 The warning modal if only one linear place is indicated in a message. Patriot Circle is highlighted on the right image (Aburizaiza and Rice 2016, 419)

6.4. Discussion and conclusion

The GMU Geocrowdsourcing Testbed (GMU-GcT) was developed to provide a mechanism to map transient navigation obstacles in real time. A comprehensive quality assessment system has been developed which uses validation of position through multiple sources. A useful way to validate position, and one based on natural human expression, is to look at the descriptive location text entered by a contributor. Humans use placenames and concepts of place to organize and describe the location of objects. The text-based descriptions of obstacle locations in the GMU-GcT provide a way of automatically checking the asserted, map-based positioning of crowdsourced obstacle contributions and the position of obstacles as determined through image geotags, which by nature have a relatively high level of imprecision (Rice 2015). In order to use the text-based location descriptions contributed to the GMU-GcT, gazetteer-based geoparsing and processing have been developed to create spatial footprints for geocrowdsourced obstacle reports, which are used to validate position. Hundreds of obstacle reports to the GMU-GcT have been processed and analyzed to develop fifteen general cases, which have been presented. These general cases extend the general point-based geolocation mechanisms to relevant polygon regions that can be used for data integration, fieldbased data validation, quality assessment, and obstacle interaction. The algorithms implemented were tested with a variety of geoparsed location descriptions collected with the GMU-GcT. For reports containing obstacle location descriptions in the GMU-GcT, these fifteen cases address approximately 65 % of the total cases. We are currently analyzing the GMU-GcT database for new footprint types to add to our library, and ways of increasing the rate of identification for the current set of fifteen footprint types.

Future work will extend the algorithms to other location descriptions from other geocrowdsourcing and geosocial media applications, and will address new types of footprints not addressed by the fifteen examples presented here. An additional area for future work are the cases of geoparsed text which remain ambiguous. Currently the code, is available at GitHub under the link: https://github.com/TipsForGIS and on geo.gmu.edu/archive/aaburiza/diss.zip

Acknowledgments

The authors wish to thank George Mason University Researchers H. Qin, F. Paez, and R. Rice, for their contributions to the GMU-GcT, and Dr. M. F. Goodchild, for his feedback and review of an early draft of this paper.

7. CONCLUSIONS AND FUTURE WORK

The following sections present a summary of the important conclusions and future work of this dissertation, which have been addressed in previous chapters, but which will be presented here in complete form.

7.1. Conclusion

Throughout this dissertation, Volunteered Geographic Information (VGI) has been discussed as a relatively new source of geospatial data for geoscientists. VGI cannot be neglected just because it is disseminated by non-professional geographers. In research literature, there are several different methodologies for effectively collecting VGI data e.g. through geoparsing. For each approach, geoscientists have been researching various techniques to validate the integrity and accuracy of VGI. One of the techniques, specifically for the geoparsing approach, is creating geospatial footprints based on place names found in text-based VGI.

In this dissertation, the research focus was building a geospatial footprints library, based on different scenarios of place names and their spatial orientation. The algorithms of the different geospatial footprints were implemented based on the text-based VGI collected through the GMU Geocrowdsourcing Testbed (GMU-GcT). The GMU-GcT is an application built over several years, with several iterations with multiple purposes. On the surface, it is for collecting volunteered reports about obstacles on GMU campus and Fairfax City area. Additional purposes of the GMU-GcT are the study of data validation, quality assessment, and the social dynamics of geocrowdsourcing. The system utilizes the three mentioned approaches for collecting VGI. A volunteer has an option of using one of the methods or all of them. In some cases, only text-based VGI is entered by volunteers. This underscores the importance of using the methodology of creating geospatial footprints from text-based VGI, when the other two methods are not found, or more importantly, as a method for validating the position of data entered by some other means.

Another scenario where the methodology of this dissertation is important to implement was during the Haiti earthquake in 2010. Almost all the emergency reports during the earthquake were via text. This was because at the time of the disaster, many Haitians own ordinary cell phones and not smartphones. A system of automated geoparsing and generation of footprints is a very useful way to validate data and information generated through GPS or tapping locations on a mobile device or through digitizing from a map or from imagery. As noted by Longley et al. (2015) and Hill (2008), humans are language-oriented and nearly universally unfamiliar with exact positioning using metric georeferencing. Verbal descriptions of place and location are common, and can be processed though geoparsing and detailed gazetteers to generate footprints associated with specific events.

This dissertation presents several steps in the evolution of georeferencing location descriptions in geocrowdsourced data, contained in a geocrowdsourced accessibility system. The current, and most sophisticated methodology for georeferencing location descriptions is presented in chapter 6, with the evolution of a geospatial footprint library from simple polygonal footprints into more complicated polygonal geometries. The initial research initiative of this dissertation, contained in chapter 3, implemented the geospatial footprints as bounding boxes encapsulating all place names found in text, without any geometric manipulation to limit or to minimize the possible location of an incident.

The second stage of this research, presented in chapter 4, transformed the algorithmic environment from desktop-based to Internet based. In this stage, the geometries of the geospatial footprints were implemented as convex hulls. The convex hulls include specific coverages that were inadequate for the location of an incident. A gazetteer was established in this stage as a reference for extracting place names of different features for the George Mason University campus and City of Fairfax. The gazetteer included different toponyms including formal names, slangs, abbreviated names, jargons, and some foreign names as well.

The third phase in this dissertation, presented in chapter 5, formed a new set of geometries for the geospatial footprints library and demonstrated their usefulness in data validation processes contained in the GMU-GcT. An example of a newly complicated geometry in this phase was extracting line segments from a specific street between two intersections with two other streets. The technologies of implementing the geometries and visualizing the new structures of the geospatial footprints, were migrated into a more advanced mechanism and easier for other developers to manipulate and improve. Another important addition to this phase was plotting the user's point, image point, and moderator's point for use in moderated quality assessment. Moreover, the distances between the three mentioned points were calculated as a reference for another validation technique presented in Rice (2015).

The final stage in this research again migrated the implementation and the visualization methodology into an easier coding structure for other developers in the geosciences community to build new geospatial footprints, based on other scenarios of spatial orientation found in text-based VGI. A total of fifteen different footprints are the base of this dissertation's geospatial footprints library. The geospatial footprints in this library were categorized into simple, complex, and ambiguous categories. The complexity of the algorithms depends on the number of place names found in a text-based VGI and their orientation and relationship to each other. One of the new examples in this final stage is a geospatial footprint that selects a specific feature, e.g. walkways, between two place names. Such geospatial footprint is applicable when a volunteer is specifying an incident on a walkway between two. Another example in the final set of the library is extracting a line segment from a street or a road alongside a building. The rest of the street segments are excluded since they are outside the scope of the location explanation in text-based VGI.

As mentioned on the final stage of this research, there are ambiguous geospatial footprints. One example of an ambiguous case found in GMU-GcT was a volunteer defining a location by mentioning a street name. In the message, the volunteer reported an obstacle on Main St. Main St in Fairfax VA is roughly 5 km long. The location of the obstacle cannot be determined unless the user gives another place name to highlight the exact location along a segment, or by mentioning a nearby or an intersecting feature. The code informs the user about the ambiguity in the message, and the GMU-GcT data validation process would rely only on the other methods for georeferencing, as outlined in Chapter 5.

In summary, the major contributions of this dissertation are the innovative creation of geospatial footprints through geoparsing and gazetteers, the code scalability of the library on Github for geoscientists, and the elasticity to integrate the library into other domains of applications.

7.2. Future work

The nature of the geospatial footprints library in this dissertation makes them readily expandable. Currently, there are fifteen types of geospatial footprints in the library implemented for fifteen different scenarios of place names orientation and relationship found in text-based VGI. Developers in the geoscience community should be able to use this library as a reference to add new geospatial footprints. One of the domains where developers can experiment in the future is to define proximity related to different prepositions.

Prepositions are used frequently in text-based VGI entries as a proximity measure related to place names. Volunteers have different visions of inferring proximities to places. For instance, the preposition ‘‘next to’’ usually is referring to a more direct proximity than the preposition ‘‘near’’. Further research on spatial prepositions and natural language processing is needed in order to determine what processing steps should be undertaken in this case, and specifically, what buffer distances should be used to create the geospatial footprint. In addition, resending the created footprints and requesting feedback from volunteers to confirm the validity of the buffer coverage can be implemented. Such technique can be used for better understanding of the relationship between spatial proximity and prepositions.

Another scope of future work is to improve the gazetteer utilized in this dissertation by adding new entries of place names, and defining the level of geographic scale at which these names need to be collected. Typical gazetteer data includes names for populated places, administrative regions, and some public facilities, and is simply not detailed enough to support the processes in this dissertation. Chapter 4 addresses some of these important gazetteer support issues, which should be a focus of future work.

The frequent change in general points-of-interest (POI) names are difficult to collect and use in a geoparsing and georeferencing process, as noted by McDermott (2017). New POI names and place names as well as new abbreviated forms of names, slang names, and common foreign language variants of placenames should be collected to support the processes outlined in this thesis. A framework should be developed to guide the collection and update of such names, including by volunteers, who could assist researchers by recording common local place names in their own language. The gazetteer, as presented in Chapter 4, is expandable to accept foreign names, new names, and even old names (former names) of different places at GMU campus and Fairfax City area. Lessons learned from the local gazetteer development activities should be generalized to support municipalities and organizations interested in connecting text-based or language-based messaging and reporting systems with maps and geographic data. This process would be facilitated by a detailed, localized gazetteer, built and maintained with lessons learned from this research.

The gazetteer utilized in this dissertation is a local gazetteer for an area of almost 1 square km. In order to implement the geospatial footprints library for larger geospatial coverages such as mega cities, a very large and rich gazetteer is needed to construct the geospatial footprints more efficiently. The implementation in such case will also need to integrate the location of the volunteer reporting an incidence to eliminate duplicate footprints. For instance, when a volunteer reports an accident close to a Starbucks on Main Street, and there are several locations of Starbucks on Main Street itself, the closest Starbucks to the recorded current location of the volunteer will have a higher probability for the geospatial footprint correct coverage. Moreover, in such large areas the implementation of the geospatial footprints library should consider spatial indices to improve the querying performance, and additional methods for resolving ambiguities.

An additional area of future work is the experimental implementation of the geoparsing code and footprint library with authoritative alert systems such as the City of Fairfax, Virginia's alert system, which generates widely-distributed text messages about impact of maintenance, construction, and special events on the accessibility of public roadways, sidewalks, parks, and buildings. The implementation should generate short URL links for footprint maps associated with the named features in the public alerts. Initial implementation should be private and tested by City Staff before it is used in a public setting. One additional improvement for this proposed future work is to generate the geospatial footprints as URL links associated in alert tweets on Twitter. This can be done through the development of a Twitter bot. This future research area has been proposed for funding and may be a near-term development that could showcase the usefulness of the techniques demonstrated in this dissertation.

The scope for future work in geoparsing and footprint-based data validation is much broader and wider than the domain of geocrowdsourcing and volunteered geographic information. One example is utilizing the geospatial footprint library with the National Geographic’s mapping platform called MapMachine. MapMachine is designed to visualize photos, videos, sound file, and articles of place names found through geoparsing (Carroll 2006). While this dissertation has focused on a narrow, controlled domain, future work will likely be in other domains and application areas, including security and infrastructure protection in megacities, public alert systems, smart vehicle navigation systems, future smart transportation networks, and logistics systems. These systems are complex, as noted by Curtin et al. (2014), who documented the difficulty in obtaining optimal solutions to even small logistics problems. Future work will inevitably require significantly deeper analysis of computational methods, and the implementation of this work in high performance computing systems, where solutions to complex problems are more likely to be found.

REFERENCES

Aburizaiza, A. O., & Ames, D. P. (2009). GIS-Enabled Desktop Software Development Paradigms. In: Advanced Geographic Information Systems & Web Services, International Conference (pp. 75–79). IEEE. http://doi.org/10.1109/GEOWS.2009.28

Aburizaiza, A. O., & M. T. Rice, (2011). VGI and geotechnology for supporting blind and vision-impaired people using a localized gazetteer. FOSS4G: Free and Open Source Software for Geospatial, 2011. Denver, Colorado: OSGeo. http://www.osgeo. org/.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh Geoparser for Historical Georeferencing. *International Journal of Humanities and Arts Computing*, *9*(1), 15–35. https://doi.org/10.3366/ijhac.2015.0136

Ames, D. P., Michaelis, C., Anselmo, A., Chen, L., & Dunsford, H. (2008). MapWindow GIS. In S. Shekhar & H. Xiong (Eds.), *Encyclopedia of GIS* (pp. 633–634). Boston, MA: Springer US. Retrieved from http://link.springer.com/10.1007/978-0-387-35973-1_753

Arsanjani, J. J., Barron, C., Bakillah, M., & Helbich, M. (2013). Assessing the quality of OpenStreetMap contributors together with their contributions. In *Proceedings of the AGILE*., Leuven, Belgium, May 14–17

Barbeau, S. J., Winters, P. L., Georggi, N. L., Labrador, M. A., & Perez, R. (2010). Travel assistance device: utilising global positioning system-enabled mobile phones to aid transit riders with special needs. *IET Intelligent Transport Systems*, *4*(1), 12. http://doi.org/10.1049/iet-its.2009.0028

Bart, H. L., & Rios, N. E. (2015). GEOLocate Web Application [Computer Software]. Tulane University, New Orleans, LA, USA.

Beaman, R., & Conn, B. (2003). Automated geoparsing and georeferencing of Malesian collection locality data. *Telopea*, *10*(1), 52–43. https://doi.org/10.7751/telopea20035604

Bekisz, M. J. (2015). *Discovering The Place Where All They Had Ended: A Study In Holocaust Toponym Georeferencing And Spatial Association.* George Mason University Masters Thesis. http://mars.gmu.edu/handle/1920/9204

Brown, G., Kelly, M., & Whitall, D. (2014). Which "public"? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *Journal of Environmental Planning and Management*, *57*(2), 190–214. http://doi.org/10.1080/09640568.2012.741045

Bucher, Joep Crompvoets, 1–4. Dresden, Germany: The Association of Geographic Information Laboratories in Europe (AGILE).

Carroll, A (2006) "A Case-Study-In-Progress: How a Media Organization Tackles the Georeferencing Challenge/Opportunity", in Position Papers, NCGIA Workshop on Digital Gazetteers Research and Practice, December 7-9, 2006. Upham Hotel and Conference Center, Santa Barbara, California. http://ncgia.ucsb.edu/projects/nga/docs/carroll-paper.pdf

Cave, A. (2016). Reconstructing Paths of Travel Based on Location References in Unstructured Text. GMU Masters Thesis, Fairfax, Virginia. http://mars.gmu.edu/handle/1920/10374

Chow, T. E. (2013). "We Know Who You Are and We Know Where You Live": A Research Agenda for Web Demographics. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 265–285). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-4587-2_15

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*, *4*(1), 332-358.

Coleman, D. J., Sabone, B., & Nkhwanana, N. (2010). Volunteering Geographic Information to Authoritative Databases: Linking Contributor Motivations to Program Effectiveness. *Geomatica*, *64*(1), 383–396.

Coleman, D. J. (2013). Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 245–263). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-4587-2_14

Coulson, M. R C., Rieger, M. and Wheate, R. 1991 Progress in creating tactile maps from Geographic Information Systems (GIS) output' Proceedings, ICA 15th

Conference on 'Mapping the Nations', Vol. 1, ed K. Rybaczuk and M. Blakemore, 167–174.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. IEEE Transactions on Dependable and Secure Computing, 9(6), 811-824. https://doi.org/10.1109/TDSC.2012.75

Curtin, K. M., Voicu, G., Rice, M. T., & Stefanidis, A. (2014). A comparative analysis of traveling salesman solutions from geographic information systems. *Transactions in GIS*, *18*(2), 286-301.

Dobson, M.W. (2013). VGI as ta Compilation Tool for Navigation Map Database. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge (pp. 207-327).

Ellwood, E., Bart Jr, H., Doosey, M., Jue, D., Mann, J., Nelson, G., ... & Mast, A. (2016). Mapping Life–Quality Assessment of Novice vs. Expert Georeferencers. Citizen Science: Theory and Practice, 1(1). https://theoryandpractice.citizenscienceassociation.org/article/10.5334/cstp.30/

Elwood, S. (2008a). Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal*, *72*(3-4), 133–135. http://doi.org/10.1007/s10708-008-9187-z

Elwood, S. (2008b). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, *72*(3-4), 173–183. https://doi.org/10.1007/s10708-008-9186-0

Elwood, S. (2010). Geographic information science: emerging research on the societal implications of the geospatial web. *Progress in Human Geography*, *34*(3), 349–357. http://doi.org/10.1177/0309132509340711

Eriksson, Y. (2001). How to make tactile maps discriminable to touch. In *Proceedings of the 20th International Cartography Conference, Beijing, China* (pp. 2914-2922).

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, *72*(3–4), 137–148. https://doi.org/10.1007/s10708-008-9188-y

Gao, S., Janowicz, K., McKenzie, G., & Li, L. (2013). Towards platial joins and buffers in place-based GIS. In Proceedings of the 1st ACM SIGSPATIAL international workshop on computational models of place (COMP'2013) (pp. 1–8).

George Mason University Office of Institutional Research and Reporting (2010) GMU

Factbook 2009-2010 http://irr.gmu.edu/factbooks/0910/Factbook0910 _Enrollment.pdf. Accessed 2 February 2011

George Mason University Official Website (2011) http://www.gmu.edu /resources/welcome/Directions-to-GMU.html. Accessed 2 February 2011

George Mason University (2012) ''Mason #1 on U.S. News 2013 Up-and-Comers List.'' 12 September. Available at http://www.connect2mason.com/content/us-news-world-report-ranked-mason-1-and-coming-university/

Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset: Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, *14*(4), 435–459. http://doi.org/10.1111/j.1467-9671.2010.01203.x

Golledge, R. G., Klatzky, R. L., Loomis, J. M., Speigle, J., & Tietz, J. (1998). A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, *12*(7), 727–749. http://doi.org/10.1080/136588198241635

Golledge, R. G. (2001). Convocation address. *Simon Fraser University*, *6*.

Golledge, R. G., Rice, M., & Jacobson, R. D. (2005). A Commentary on the Use of Touch for Accessing On-Screen Spatial Representations: The Process of Experiencing Haptic Maps and Graphics. *The Professional Geographer*, *57*(3), 339–349. http://doi.org/10.1111/j.0033-0124.2005.00482.x

Golledge, R. G., Rice, M. T., & Jacobson, R. D. (2006). Multimodal interfaces for representing and accessing geospatial information. In *Frontiers of Geographic Information Technology* (pp. 181-208). Springer Berlin Heidelberg.

Golledge, R. G., Marston, J. R., Loomis, J. M., & Klatzky, R. L. (2007). Wayfinding without sight: The haptic pointer interface. *Invited paper presented at Annual Meeting of the Association of American Geographers, San Francisco, California*.

Goodchild, M. F., Kyriakidis, P., Rice, M., & Schneider, P. (2005). Spatial Webs, Final Report and Position Papers. https://escholarship.org/uc/item/46z721n2

Goodchild, M., & Hill, L. (2006). *Digital Gazetteer & Practice Workshop, Summary Report* (Specialist Research Meetings - Papers and Reports) (pp. 1–37). Santa Barbara, CA: UC Santa Barbara. Retrieved from http://escholarship.org/uc/item/6fr9g5nv

Goodchild, M. F. (2007a). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221. http://doi.org/10.1007/s10708-007-9111-y

Goodchild, M. F. (2007b). Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* 2: 24–32

Goodchild, M. F. (2008) Assertion and Authority: The Science of UserGenerated Geographic Content. *In Proceedings of the Colloquium for Andrew U. Frank´s Birthday, 3:82–96. Department of Geoinformation and Cartography, Vienna University of Technology, 2008*. http://www.tandfonline.com/doi/abs/10.1080/17489720902950374

Goodchild, M. (2009). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, *3*(2), 82–96. http://doi.org/10.1080/17489720902950374

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, *3*(3), 231–241. http://doi.org/10.1080/17538941003759255

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. http://doi.org/10.1016/j.spasta.2012.03.002

Gregory, I., Donaldson, C., Murrieta-Flores, P., & Rayson, P. (2015). Geoparsing, GIS, and textual analysis: Current developments in spatial humanities research. *International Journal of Humanities and Arts Computing*, *9*(1), 1-14. https://doi.org/10.3366/ijhac.2015.0135

Guptill, S. C., & Morrison, J. L. (Eds.). (1995). *Elements of spatial data quality*. Elsevier.

Guptill, S. C. (2006). *Digital Gazetteers Research & Practice Workshop, Position Paper* (Specialist Research Meetings Papers and Reports) (pp. 36–37). Santa Barbara, CA. Retrieved from http://escholarship.org/uc/item/2w19j1j6

Haklay, M. (2010). How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682–703. http://doi.org/10.1068/b35097

Haklay, M. (Muki), Basiouka, S., Antoniou, V., & Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, *47*(4), 315–322. https://doi.org/10.1179/000870410X12911304958827

Hecht, B., & Moxley, E. (2009). Terabytes of Tobler: evaluating the first law in a massive, domain-neutral representation of world knowledge. Spatial information theory, 88-105. http://brenthecht.com/papers/cosit2009.pdf

Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(6), 550–557. https://doi.org/10.1016/j.isprsjprs.2010.06.005

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

Hill, L. (2008) Geoparsing. In: Kemp K (ed) Encyclopedia of Geographic Information Science. Sage Publications, Thousand Oaks, California

Horák, J., Belaj, P., Ivan, I., Nemec, P., Ardielli, J., & Růžička, J. (2011). Geoparsing of Czech RSS News and Evaluation of Its Spatial Distribution. In R. Katarzyniak, T.-F. Chiu, C.-F. Hong, & N. T. Nguyen (Eds.), *Semantic Methods for Knowledge Management and Communication* (Vol. 381, pp. 353–367). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23418-7_31

Jiang, B. (2013). Volunteered Geographic Information and Computational Geography: New Perspectives. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 125–138). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-4587-2_8

Johnson, P. A., & Sieber, R. E. (2013). Situating the Adoption of VGI by Government. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 65–81). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-4587-2_5

J., Arsanjani, J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. In J. J. Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience (pp. 37–58). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-14280-7_3

Karimi, H. A., Zhang, L., & Benner, J. G. (2014). Personalized accessibility map (PAM): a novel assisted wayfinding approach for people with disabilities. *Annals of GIS*, *20*(2), 99–108. http://doi.org/10.1080/19475683.2014.904438

Kuhn, W. (2007). *Volunteered geographic information and GIScience* (pp. 86–97). Santa Barbara, CA: NCGIA and Vespucci Workshop on Volunteered Geographic Information.

Kwan, M. P., Casas, I., & Schmitz, B. (2004). Protection of geoprivacy and accuracy of spatial information: how effective are geographical masks?. Cartographica: The International Journal for Geographic Information and Geovisualization, 39(2), 15-28.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information systems and science*. John Wiley & Sons Ltd.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems*. John Wiley & Sons Ltd.

Laakso, M., Sarjakoski, T., Lehto, L., & Sarjakoski, L. T. (2013). An Information Model for Pedestrian Routing and Navigation Databases Supporting Universal Accessibility. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *48*(2), 89–99. http://doi.org/10.3138/carto.48.2.1837

Leidner, J. L. (2017). Georeferencing: From Texts to Maps. *The International Encyclopedia of Geography*. http://doi.org/10.1002/9781118786352.wbieg0160

Li, L., & Goodchild, M. F. (2010). The Role of Social Networks in Emergency Management: A Research Agenda. *International Journal of Information Systems for Crisis Response and Management*, *2*(4), 48–58. https://doi.org/10.4018/jiscrm.2010100104

Loomis, J., R.G. Golledge, and R. Klatzky. 2001. ‘‘GPS-Based Navigation Systems for the Visually Impaired.’’ In Fundamentals of Wearable Computers and Augmented Reality, ed. W. Barfield and T. Caudell, 429–46. Mahwah, NJ: Lawrence Erlbaum.

Marston, J. R., Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Smith, E. L. (2006). Evaluation of spatial displays for navigation without sight. *ACM Transactions on Applied Perception*, *3*(2), 110–124. http://doi.org/10.1145/1141897.1141900

Marston, J. R., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2007). Nonvisual route following with guidance from a simple haptic or auditory display. *Journal of Visual Impairment & Blindness*, *101*(4), 203-211.

Marston J R, Golledge R G, Loomis J M, Klatzky R L, and Giudice N A 2008 Substituting for vision: Field experiments with several orientation and navigation devices. Invited presentation at the Vision Research Seminar Series, Department of Ophthalmology, Emory University School of Medicine, Atlanta, Georgia

McDermott, Scott D. (2017) *Frequency and Proximity Clustering Analyses for*

Georeferencing Toponyms and Points-of-Interest Names from a Travel Journal. George Mason University Doctoral Dissertation, Fairfax, Virginia.

Miele, J. A., & Marston, J. R. (2005, March). Tactile map automated production (TMAP): Project update and research summary. In *Proceedings CSUN*.

Miele, J. A. (2007, August). Automated creation and detailed annotation of audio/tactile maps using scalable vector graphics (SVG). In *International Cartographic Conference, Moscow, Russia*.

Miele, J. A. (2011). Tactile map automated production (TMAP): Using GIS data to automatically generate raised-line street maps. *WWW document, http://www. ski. org/Rehab/TMAP/*

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461. http://doi.org/10.1007/s10708-014-9602-6

Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J., & Gaio, M. (2014, November). Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 183-192). ACM. http://doi.org/10.1145/2666310.2666386

Nuernberger, A. (2008). *Presenting accessibility to mobility-impaired travelers*. (Ph.D. Dissertation, University of California Santa Barbara, Santa Barbara, California).

Paez Wulff, F. I. (2014). *Recruitment, Training, and Social Dynamics in Geo-Crowdsourcing for Accessibility* (Master's thesis, George Mason University, Fairfax, Virginia). http://mars.gmu.edu/handle/1920/9042

Pease, P. A. (2014). *The Influence of Training on Position and Attribute Accuracy in Volunteered Geographic Information* (Master's thesis, George Mason University, Fairfax, Virginia). http://mars.gmu.edu/handle/1920/10731

Perkins, C. (2001). Tactile campus mapping: Evaluating designs and production technologies. In *20th International Cartographic Conference* (Vol. 5, pp. 2906-2913).

Perkins, C. (2002). Cartography: progress in tactile mapping. *Progress in Human Geography*, *26*(4), 521–530. http://doi.org/10.1191/0309132502ph383pr

Przyszewska, K., & Szyszkowska, K. 2011 Atlas of the world for the blind and visually impaired. In *Proceedings of the Twenty-fifth International Cartographic Conference*, Paris, France

Pugliesi, E., & Decanini, M. (2011). Cartographic design of in-car route guidance for color-blind users. In *International Cartographic Conference. Enlightened view on Cartography and GIS, Paris*.

Qin, H., Aburizaiza, A. O., Rice, R. M., Paez, F., & Rice, M. T. (2015). OBSTACLE CHARACTERIZATION IN A GEOCROWDSOURCED ACCESSIBILITY SYSTEM. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-3/W5*, 179–185. http://doi.org/10.5194/isprsannals-II-3-W5-179-2015

Qin, H., Rice, R. M., Fuhrmann, S., Rice, M. T., Curtin, K. M., & Ong, E. (2016). Geocrowdsourcing and accessibility for dynamic environments. *GeoJournal*, *81*(5), 699–716. http://doi.org/10.1007/s10708-015-9659-x

Qin, H. (2017). *Modeling Accessibility through Geocrowdsourcing. Fairfax, VA: George Mason University*. George Mason University, Fairfax, Virginia, USA.

Research Unit on Spatial Cognition and Choice 2011 Reg Walking with Original System (1997; Archived Image). WWW document, http://www.geog.ucsb.edu/pgs/images/Reg_1997.jpg

Rice, M., Jacobson, R. D., Golledge, R. G., & Jones, D. (2005). Design Considerations for Haptic and Auditory Map Interfaces. *Cartography and Geographic Information Science*, *32*(4), 381–391. http://doi.org/10.1559/152304005775194656

Rice, M. T., Hammill, W. C., Aburizaiza, A. O., Schwarz, S., & Jacobson, R. D. (2011). Integrating User-contributed Geospatial Data with assistive Geotechnology Using a localized Gazetteer. In A. Ruas (Ed.), *Advances in Cartography and GIScience. Volume 1* (pp. 279–291). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-19143-5_16

Rice, M. T., Aburizaiza, A. O., Jacobson, R. D., Shore, B. M., & Paez, F. I. (2012a). Supporting Accessibility for Blind and Vision-impaired People With a Localized Gazetteer and Open Source Geotechnology. *Transactions in GIS*, *16*(2), 177–190. http://doi.org/10.1111/j.1467-9671.2012.01318.x

Rice, M. T., Paez, F. I., Mulhollen, A. P., Shore, B. M., & Caldwell, D. R. (2012b). Crowdsourced geospatial data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data (Annual No. AA10-4733). Fairfax, VA: George Mason University. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf

Rice, M. T., Jacobson, R. D., Caldwell, D. R., McDermott, S. D., Paez, F. I., Aburizaiza, A. O., … Qin, H. (2013). Crowdsourcing techniques for augmenting traditional accessibility maps with transitory obstacle information. *Cartography and Geographic Information Science*, *40*(3), 210–219. http://doi.org/10.1080/15230406.2013.799737

Rice, M.T., K.M. Curtin, F.I. Paez, C.R. Seitz, and H. Qin. 2013. Crowdsourcing to Support Navigation for the Disabled: A Report on the Motivations, Design, Creation and Assessment of a Testbed Environment for Accessibility. US Army Corps of Engineers, Engineer Research and Development Center, US Army Topographic Engineering Center Technical Report, Data Level Enterprise Tools Workgroup. Broad Area Announcement no. AA10-4733, Contract no. W9132V-11-P-0011. Fairfax, VA: George Mason University.

Rice, M. T., Curtin, K. M., Paez, F. I., Seitz, C. R., & Qin, H. (2013a). Crowdsourcing to support navigation for the disabled: A report on the motivations, design, creation and assessment of a testbed environment for accessibility (US Army Corps of Engineers, Engineer Research and Development Center, US Army Topographic Engineering Center Technical Report, Data Level Enterprise Tools Workgroup No. BAA: #AA10-4733, Contract: # W9132V-11-P-0011) (pp. 1–62). Fairfax, VA: George Mason University. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA588474.

Rice, M.T., F.I. Paez, R.M. Rice, E. Ong, H. Qin, C.R. Seitz, J.V. Fayne, K.M. Curtin, S. Fuhrmann, D. Pfoser, and R.M. Medina. 2014. ''Quality Assessment and Accessibility Applications of Crowdsourced Geospatial Data: A Report on the Development and Extension of the George Mason University (GMU) Geocrowdsourcing Testbed.'' Broad Area Announcement no. AA10-4733, contract no. W9132V-11-P-0011. Assessing the Value of Neo-Geographic Information. Pp.1-91. Fairfax, VA: George Mason University. http://www.dtic.mil/get-tr-doc/pdf?AD=ADA615952

Rice, R. M. (2015). *Validating VGI data quality in local crowdsourced accessibility mapping applications: A George Mason University Case Study*. (Master's thesis, George Mason University, Fairfax, Virginia). http://mars.gmu.edu/handle/1920/10522

Rice, R. M., Aburizaiza, A. O., Rice, M. T., & Qin, H. (2016). Position Validation in Crowdsourced Accessibility Mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *51*(2), 55–66. http://doi.org/10.3138/cart.51.2.3143

Rice, M., Jackobson, D., Curtin, K. Pfoser, D., Qin, H., Coll, K., Rice, R., Paez, F., Aburizaiza, A. (2018) "Quality Assessment and Accessibility Mapping in an

Image-based Geocrowdsourcing Testbed" Cartographica, ed. Emmanuel Stefanakis. [Accepted, In press]

Sieber, R. E. (2000). GIS implementation in the grassroots. *URISA journal*, *12*(1), 15-29. http://www.iapad.org/wp-content/uploads/2015/07/Siebervol12no1-2.pdf

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, *78*(2), 319-338. http://dx.doi.org/10.1007/s10708-011-9438-2

Steiniger, S., & Bocher, E. (2009). An overview on current free and open source desktop GIS developments. *International Journal of Geographical Information Science*, *23*(10), 1345–1370. http://doi.org/10.1080/13658810802634956

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748. https://doi.org/10.1080/13658816.2011.604636

Sui, D., Goodchild, M., & Elwood, S. (2013). Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 1–12). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-4587-2_1

Tatham, A. F. (1991). The design of tactile maps: theoretical and practical considerations. *Proceedings of international cartographic association: mapping the nations*, 157-166.

Tatham AF (2001) Achievable goal or chimaera? Standardisation in tactile mapping. In: Mapping the 21st Century: Proceedings, The 20th International Cartographic Conference, ICC 2001, Beijing, China, August 6-10th, 2001. Chinese Society of Geodesy, Photogrammetry, and Cartography, ICC 2001, Beijing, vol. 5, p. 2923

Taylor, D. R. F. (2001, August). Tactile Atlas of Latin America. In *Proceedings of the Twentieth International Cartographic Conference: Mapping the 21st Century, Beijing, China* (Vol. 5, pp. 2894-2898)

Tobin, R., Grover, C., Byrne, K., Reid, J., & Walsh, J. (2010). Evaluation of georeferencing (p. 1). ACM Press. https://doi.org/10.1145/1722080.1722089

U.S. Census Bureau (2010) 2009 American Community Survey 1-Year Estimates http://www.census.gov/acs/www/. Accessed 2 February 2011

Van Laere, O., Quinn, J., Schockaert, S., & Dhoedt, B. (2014). Spatially Aware Term Selection for Geotagging. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 221–234. https://doi.org/10.1109/TKDE.2013.42

Waters, N. 1999. Transportation GIS: GIS-T, Chapter 59, pp. 827-844, in *Geographical Information Systems: Principles, Techniques, Applications and Management*, 2nd Edition, Edited by Paul Longley, Michael Goodchild, David Maguire, and David Rhind, John Wiley and Sons.

Waters, N. (2009). EDGE NODES-Reputation Systems: The Dark Side of VG. *GeoWorld*, *22*(8), 18.

Waters, N. (2017). Tobler's First Law of Geography. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds.), *International Encyclopedia of Geography: People, the Earth, Environment and Technology* (pp. 1–13). Oxford, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118786352.wbieg1011

Zandbergen, P. A. (2009). Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. *Transactions in GIS*, *13*, 5–25. http://doi.org/10.1111/j.1467-9671.2009.01152.x

Zheng, S., & Zheng, J. (2014). Assessing the Completeness and Positional Accuracy of OpenStreetMap in China. In T. Bandrova, M. Konecny, & S. Zlatanova (Eds.), *Thematic Cartography for the Society* (pp. 171–189). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-08180-9_14

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, *2*(2), 6–32. http://doi.org/10.2202/1948-4682.1069

BIOGRAPHY

Ahmad Omar Aburizaiza received his Bachelor of Science in Computer Science from the University of the Pacific, Stockton CA in 2002. He was employed as an ArcSDE Administrator at Saudi Aramco, Dhahran Saudi Arabia between 2003 and 2006. He received his Master of Science in Geographic Information Science from Idaho State University in 2009. During his time at Idaho State University, he worked as GIS programmer at the Geospatial Software Lab in the same university. During his PhD degree at George Mason University, he worked in three different roles in the department of Geography and Geoformation Science and the Fenwick Library: as a graduate research assistant, a GIS instructor, and a geospatial consultant. He was also employed as a GIS Analyst III at the City of Alexandria VA in 2014. He was also selected at NASA Langley Research Center, Hampton VA, as a co-team lead for summer 2016. Currently, he is a Full Stack GIS Developer (Software Engineer III) at NCI Inc., a contractor with the Federal Communication Commission (FCC).