Datenbank-Spektrum manuscript No.
(will be inserted by the editor)

Applying Geographic Information Retrieval

An Experience Report on Developing Local Search for a Developing Country

Dirk Ahlers

Received: date / Accepted: date

Abstract This article reports on the efforts to establish a research project on a geospatial search engine for the Latin American country of Honduras during the author's research stay at a local university. Honduras is an interesting example of the challenges for information and knowledge management in developing countries as it combines many of the issues that are otherwise encountered more isolated or with less impact. These include low Web coverage in a low-resource country with limited Web infrastructure, and generally, work in challenging circumstances. The specific focus on geospatial information uncovers further issues that need to be addressed, such as informal addressing schemes or landmark-oriented location references, broad or incorrect locations for places, or insufficient ground truth in databases. While the tangible results of the project stay behind the original goals, several interesting results were achieved, which are condensed here as an experience report.

Keywords Geospatial Web Search · Geographic Information Retrieval · Geocoding · Developing Countries · Case Study

This paper was written during the tenure of an ERCIM Alain Bensoussan Fellowship Programme which has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 246016. The research described here was in large parts carried out at UNITEC – Universidad Tecnológica Centroamericana, Tegucigalpa, Honduras.

D. Ahlers
NTNU – Norwegian University of Science and Technology
Sem Sælands vei 7–9, 7491 Trondheim, Norway
Tel.: + 47 735 90519
E-mail: dirk.ahlers@idi.ntnu.no

The final version is published by Springer as:
Applying Geographic Information Retrieval, Dirk Ahlers
Datenbank-Spektrum, March 2014, Volume 14, Issue 1, pp 39-46
http://dx.doi.org/10.1007/s13222-014-0148-z

1 Introduction

Location is as an important organizing principle for many Web search tasks. In most industrialized nations the search for locations features prominently within search engines and users are used to seamlessly working local search with a multitude of correct results [2]. It works this well because there is both good data available and tailored technology to make use of it.

But in many developing countries, the situation is gravely different. Local search may not be as accessible, important places are missing, or the information density is rather low, only mentioning a name without more in-depth information. Errors or inaccuracies may further complicate the situation, if information is even available in the first place. This view of insufficient available information is quite different from the usual situation where we are used to having to manage a huge and ever growing amount of available information on the Web. However, some of the challenges can occur in any country independent of development status.

In this paper, we examine the special case of geospatial search and country-specific search engines in developing regions and more precisely, explore it for the Latinamerican country of Honduras. The activities were part of an extended research stay of the author that concerned geospatial search and location-based services. The research questions then are, how can local search, or even Web search in general, work in a low-resource country with only very little Web coverage and an informal and imprecise addressing scheme where few people even have Internet access?

An overview of the challenges and potential was given at the beginning of the project [1]. We discuss these challenges and research performed in the meantime in the form of this retrospect case report. The

present article is a revised version of a contribution to LWA2013 [6], focusing less on the development process and instead aiming to illuminate selected activities. Some aspects specifically concerning geoparsing and geocodingwere discussed separately [9].

1.1 Social Background

Some socio-economic statistics abridged from [13] should help to better understand the expected background. Honduras is a developing country which ranks 121st in the Human Development Index worldwide, and is the sixth-poorest country in Latin America[1] with 60% of the population below the national poverty line[2].

Little reliably data is available on mobile phone or mobile Internet use. Indicators also diverge across sources; we have selected those deemed most credible. While computer ownership is at only 2.5%, mobile subscriptions are much more promising. From 3% in 2000, they have surpassed 100% in 2008 and were at 125% in 2010[3]. This oversaturation can be explained by the practice of having mobile phones for different providers. More useful is the estimate of people actually owning a mobile phone at about 75% in 2010[4]. Internet use is rising fast, with 11% of Internet users in the population in 2010 up from 1% in 2000[5]. Informal estimates for the share of smartphones with Internet access are around 10%, which hints that a lot of Internet use happens on mobile devices.

The divide in the use of communication technology is reflected in the strategies of mobile phone providers, who aim to cater to the bimodal use by offering initial smartphone solutions, but chiefly providing texting and USSD (menu-based dialogs) services. Honduras can be considered a latecomer that is now quickly catching up, which makes local search a viable option for future applications. With increasing use of online services, it could also reach larger parts of the population [20].

These points can also help answer the question of whether such a service is really needed and whether there are not more pressing needs. Currently a large part of the population cannot afford expensive services, lives mostly by subsistence farming and might only buy at local neighborhood shops. But with a slowly growing middle class and more people expecting local search to work, many users would benefit. Thus, while local search might be regarded as a luxury, the adoption of services for the population capable of affording the necessary technology to participate, is a worthwhile undertaking. It may, by uncovering and presenting available information, improve the general data situation and thus facilitate further undertakings that might benefit growing parts of the population.

[1] http://hdrstats.undp.org/en/countries/profiles/HND.html

[2] http://data.un.org/Data.aspx?d=MDG&f=seriesRowID\%3A581

[3] http://data.worldbank.org/indicator/IT.CEL.SETS.P2/countries/HN?display=graph

[4] http://www.latinobarometro.org/

[5] http://devdata.worldbank.org/ict/hnd_ict.pdf

2 Introduction to Geographic Information Retrieval

Many research projects have undertaken to extract location information for geospatial Web search, e.g., [11, 36,31]. Also commercial services such as Google Maps, Google Earth, Bing Maps, or Yahoo! Maps are building location-based search applications and creating indexes of geospatial information [2].

All these have their background in Geographic Information Retrieval (GIR), the identification, extraction, interpretation, indexing, ranking, and processing of various forms of geographical references in unstructured documents [27,30]. A good introduction and overview of different approaches and methods of GIR, including disambiguation of toponyms, is given in [24]. Some approaches focus on the entities that are located at a place [5], which requires geospatial entity resolution [37]. Others use Wikipedia for location disambiguation [35] or merging [26] (cf. Section 5.2).

In a wider sense, country-specific search engines have to identify and maintain a certain geographic scope. They have been described, for example, for Chile [32], Portugal [23], Brazil [17], or Germany [11,31]. The general value of mobile phones and their applications in developing countries has been shown extensively [22,21], but search as such has also received some attention [18, 19,28].

A very relevant point was raised regarding the limits of generalization of many approaches and data sources [33], which usually reflect the views of the industrialized countries, an issue highlighted in this work.

3 Country-Specific Characteristics and Challenges

No tailored geospatial search for Honduras exists and the existing services showed some shortcomings. The project idea therefore was to work towards a geospatial index of Honduran locations, places, points of interest and Web pages in close cooperation with a local mobile telecom provider. We will discuss selected country-

specific challenges identified in [1] along with solutions and results generated in the meantime.

The main challenges concern the requirements and data situation and the extraction, analyzing, and indexing of location data. We start with the data-centric process life cycle of geospatial Web information [16] comprising processes of discovery, understanding, augmentation, and exploration. It was extended to describe a development process for a geospatial search engine with multiple interconnecting steps as a work in progress [8].

The first step should be to get an overview of what applications and services may be relevant and which are already available. Specific undertakings range from market analysis (cf. Section 3.1), viability analysis, and data source investigations (Section 3.4) over requirements to data analysis and user studies (cf. Section 3.2), regarding search and mobile applications. The building of a knowledge base from gazetteers (Section 5) or Wikipedia data (Section 5.2) is part of the work as well as the extraction of geospatial data (Section 3.3). For Web crawling [10], the characteristics of the national Web are shown in Section 4.

3.1 Search engine market

The big search engines provide map data, sometimes at very good quality and also provide some map-based local search. However, compared to other regions of the world, there is very little local information available and its depth is limited, often offering nothing more than a name and a rough location. The situation offers a unique potential to build a Honduran geospatial search engine. While the Web coverage is still low and the address scheme makes exact location extremely difficult, Web usage is rising, potential scattered data sources exist, and people begin using location-based services which creates sufficient demand and support. Honduras has no own search engine, and people regularly use the big international search engines, mainly Google. There exists a local domain, google.hn, which seems to give a slight preference to pages about Honduras as part of the location customization[6]. Local search has initial data, but is far from a comprehensive coverage – which is in part the topic of this research. Therefore building country-level geospatial search and services is still a worthwhile undertaking.

[6] http://support.google.com/websearch/bin/answer.py?hl=en&answer=179386

3.2 Users

An important aspect is to understand the target user group, their requirements, and situation. The most comprehensive overview on mobile phone use in developing countries is a literature study [21]. Mobile phone use can also be understood as a method of empowerment in developing nations [15].

For insights focused on Honduras, we carried out a survey on the use of local search and local information seeking behavior in general [13]. Among other things, we found that the preferred modes and sources of information search are word-of-mouth or existing knowledge about locations, combined with a knowledgeable social circle. This social aspect may also explain why the most used online source is Facebook, followed with a distance by less prevalent search engines. One implication of these findings is that the search engine preferably would follow a hybrid approach to access a variety of data sources, also including the social networks, or even employing crowdsourcing to establish both relevant information and trust.

Privacy concerns were found to be connected to concrete security concerns. Due to a high level of targeted criminality such as robberies or kidnappings [38], many people prefer to keep personal information, especially their location, very private. Yet, many entries in location-sharing services explicitly mention people's houses ("Mi casa", "My house"). In these cases, the functionality seems to override security concerns. Additionally, in part due to the security situation, people will avoid walking in most areas and only get their phone out in safe places. This influences usage, which was found to be not as spontaneous as in other countries and to happen less on the move, which leads to less of the usually associated here-and-now mentality [13].

A very motivating result was that all participants warmly welcomed the project idea or even any improvement of the data situation.

3.3 Geoparsing and Geocoding

A very challenging characteristic of Honduran location references is that exact locations in the form of addresses with house numbers in a formal, high-granularity addressing scheme are usually not given. This seriously impedes a high-granularity approach that would try to map information to individual buildings. There are some areas or smaller cities where a rectangular street grid exists, which usually also allows for a better addressing scheme. However, in most regions, location references are given by city name, city district and some-

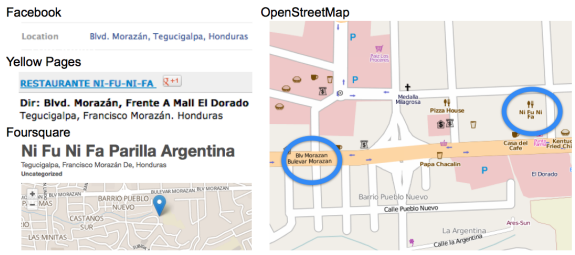

Fig. 1 Entity example with street-level address from different sources, counter-clockwise: Facebook, yellow pages, foursquare, OpenStreetMap (from [3])

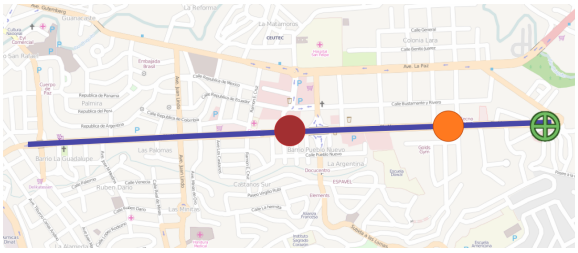

Fig. 2 Example of possible geocoding results (from [3])

times the street name. Various other forms of descriptions have evolved that allow finding a certain building. Often these are given additional directional information such as nearby landmarks or well-known buildings [9]. Examples are "Colonia Kennedy, Tegucigalpa" (a neighborhood in the capital), "En el centro" (the city center), or "Al final de Bulevar Morazán" (at the end of a major street). The low-granularity location references – in common Web pages as well as in databases – pose a particular problem to geoparsing, the extraction of location references from general text. The example in Fig. 1 shows varying descriptions and given locations for a place. Geoparsing can in part be solved with adapted methods discussed in Section 2 and the gazetteer built in Section 5. If street names are not available in the gazetteer, approaches based on patterns with less external knowledge can be used [29].

Geocoding, i.e., the grounding of location references to geographic coordinates, is much more demanding. It can target different scales of spatial granularity. Finding the location as within Tegucigalpa may not be sufficiently accurate. We have to find a balance between this and the ideal of building-exact results. Fig. 2 shows possible results for "Tigo, Al final de Bulevar Morazán". We may be able to identify the street as the marker in the middle. We might further understand "Al final" and actually know which end of the street is "the end" to move the marker east. The final crosshair is at the actual Tigo building. We might find this by using entity extraction to get the name and then try to improve the geocoding by external sources as proposed in [3].

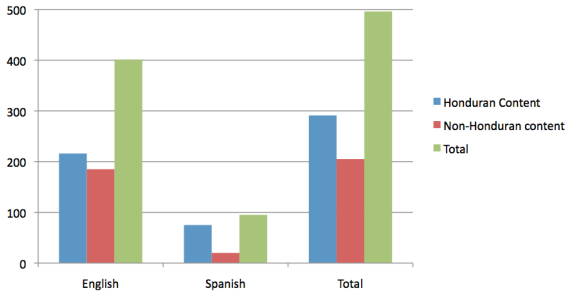

Fig. 3 Provenance of pages with topic Honduras in DMOZ (from [14])

3.4 Source availability and integration

Since the Web shows only very low coverage for Honduras, we aim to additionally integrate specific structured datasources into the search engine index. In view of the user behavior, we also aim to include social networks, which carry a lot of location-relevant information. This suggests that the search engine follows a hybrid approach of both Web search with georeferencing of documents and additionally database access and merging for specific data sources to perform entity cross-correlation is needed (cf. [5,37]). Using multiple sources may also help to improve the accuracy of geocoding (cf. Section 3.3).

The official language of Honduras is Spanish. However, at the Caribbean coast, English is a recognized regional language and is more frequently spoken. Additionally, this is the main area for foreign tourism and much information is more comprehensively available in English. Therefore, future work also includes cross-language retrieval. Initial work is discussed in Section 5.2.

4 Characteristics of the Honduran Web

In trying to identify relevant Web sources and define the "Honduran Web", we took two separate approaches as discussed in [14]. First, we looked at the DMOZ Open Directory[7], and second, we built a list of all Honduran domains under a .hn ccTLD.

DMOZ has relatively little coverage for Honduras. In the English hierarchy, it contains 421 entries, with 10 from .hn (2.5%); for Spanish, there are only 96 entries, but 46 are from .hn (48%). The English part contains mainly travel sites and general descriptions while the Spanish contains actual local pages. Fig. 3 shows the provenance of the DMOZ data for the English and Spanish categories, manually classified as from inside or outside the country.

[7] `http://www.dmoz.org/`

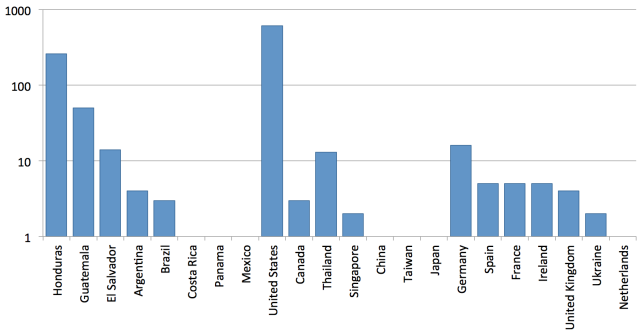

Fig. 4 Hosting countries for .hn domains (adapted from [14])

For .hn, we could obtain a count of domains through inofficial channels. Of the reported 5780 registered domains, we were able to only find around 900, which indicated a lot of them are not used for Web hosting or not at all. We found them by using a divide-and conquer strategy with queries to common search engines.

We additionally brute-forced our way through all IP blocks assigned to Honduras (about 140,000 IPs) to see if a HTTP server was available. Some response pages also give hints about the respective domain, but the majority of responses came from network equipment, cameras, printers, etc. We cross-checked with the discovered domain names, but only gathered very few domains from this. Conversely, we also found some .com domains that were hosted inside Honduras.

Trying to confirm the reverse suspicion of domains being hosted outside the country, we queried the domains' IP hosting locations. The result by country is shown in Fig. 4 on a logarithmic scale. A large amount is assigned in Honduras, but the majority in the US, with other American countries following behind. US hosts were mostly in the south in areas with a larger Latin American population, while other countries are often related to the owners or investors of businesses. We also found 25% of governmental domains hosted in the US, confirming the suspicion of a 'digital divide' [34]. Additionally, a very high number of Honduran businesses opt for a generic .com domain. Some even forego an own Web presence and instead create a Facebook page. This makes it more difficult to gather all relevant domains for Honduras and include all relevant information. As a first rough estimate on the number of available sites, all Honduran DMOZ links were correlated with the known domains registered for .hn. Adjusted with some known Honduran .com domains, we arrive at an estimate of below 40,000 relevant domains overall.

Future work involves a better identification of generic .com domains carrying Honduran content, using a mixture of location analysis, language, and link structure, similar to, e.g.,[23, 40].

5 Building a gazetteer

A gazetteer contains knowledge about places, their names, coordinates, and geographical or administrative hierarchies [25]. As external knowledge it supports the extraction of geospatial references by grounding them to known places [39] (cf. Section 3.3).

5.1 Geonames

Gazetteer data can either be directly available or needs to be collected and combined from multiple sources. In this case, we use data from geonames.org to serve as a bootstrapping of the search engine's knowledge about places. An initial assessment of the geonames dataset for Honduras [4] showed, for example, that it was lacking the neighborhoods in the cities, which are an important part of address references. We therefore manually generated a list of known neighborhoods for the capital Tegucigalpa as a pilot region. To gather additional structured data about places, we also added geotagged Wikipedia articles to (cf. Section 5.2). As discussed in Section 3.3, additional information about entities or street names was also planned to be included.

5.2 Wikipedia cross-language linkage

To improve the gazetteer data, we are interested in those Wikipedia articles about places that include a geographic coordinate which grounds the content in the real world. Looking into additional languages prompted the discovery of an interesting anomaly: Honduras has more English geotagged articles than Spanish ones.[8] With 342 Spanish and 405 English georeferenced articles, English was overrepresented by 18%. Our aim is to merge both language versions and identify identical places on the article level. As cross-language links between articles were not always reliable, we developed a translation approach to pair geotagged Wikipedia articles about the same place. We report the main parts here, the full description is found in [7].

We use the GeoNames API to retrieve articles instead of downloading the huge Wikipedia dumps for the relevant languages. Since the service limits the radius search to 20km, the country has to be covered with queries. We chose a tiling of the country borders with a small tolerance and the addition of islands and international waters, as seen in Fig. 5.

[8] Such anomalies exist in many countries:
`http://www.zerogeography.net/2012/10/dominant-wikipedia-language-by-country.html`

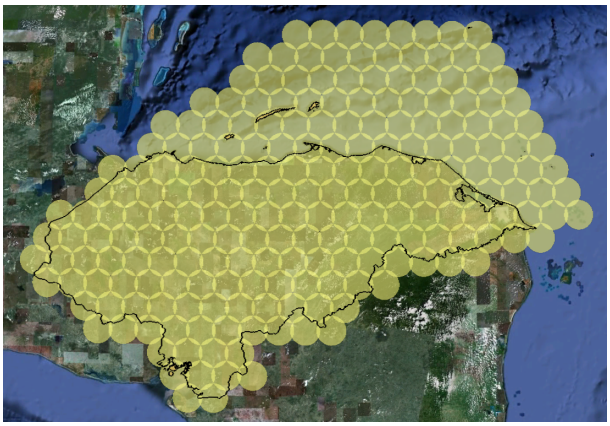

Fig. 5 Hexagonal grid arrangement of packed circles for API queries

Table 1 Examples of sibling articles

Spanish title (ES)	English title (EN)
Tegugigalpa	Tegucigalpa
Aeropuerto Internacional Toncontín	Toncontín International Airport
Santa Bárbara (Santa Bárbara)	Santa Bárbara, Honduras
El Paraíso	El Paraíso Department
Departamento de Copán	Copán Department
Juticalpa	Juticalpa, Olancho
Virginia (Honduras)	Virginia, Lempira
Nuestra Señora de Suyapa	Virgin of Suyapa
Parque nacional Pico Bonito	Pico Bonito National Park

We define the data fusion method in terms of finding language *siblings*, as seen in Table 1, based on entity-merging and geographic conflation methods. For each article, we select and rank candidate siblings in the respective other language based on title and location.

Coordinates can vary due to different interpretations of the center of an area or variations in user-generated coordinates, especially for larger entities [12]. We limit the amount of candidate siblings we have to examine by cutting off the location similarity with a perimeter of 10km around an article's location.

We define a title translation distance TTD as an editing distance similarity measure based on partial translations and permutations. We account for spelling variations with a Levenshtein editing distance adapted with a weight relative to the term length and with a reduced penalty for accents and tildes. Interestingly, most proper nouns are identical or very similar in both languages and can already be well accounted for with this simple approach. However, common nouns have to be translated and the order of terms within a placename also be changed. The translation table was filled mostly with relevant geographical feature types, taken from GeoNames[9] (e.g., airports, islands, mountains, stadiums, cities, parks, etc.). Heuristics were generated from some observed conventions. For example, for municipality and department names, *Santa Bárbara (Honduras)*$_{ES}$ puts the higher-level administrative body, in this case the country name, in brackets, while *Santa Bárbara Department*$_{EN}$ only uses the administrative type. This helps to distinguish departments and capitals that often have the exact same coordinates.

To cover permutations, we employ a list of transposition heuristics. They can swap the first two terms: *Congreso Nacional de Honduras*$_{ES}$ → *National Congress of Honduras*$_{EN}$; swap first and last terms: *Río coco*$_{ES}$ → *Coco river*$_{EN}$; or swap and move the first two words to the end: *Parque nacional Pico Bonito*$_{ES}$ → *Pico Bonito National Park*$_{EN}$. We generate all potential variations and chose the one with the minimum TTD and the smallest location distance as a sibling.

For all articles, an extendive manual assessment was performed to determine the ground truth for siblings. The algorithm resulted in 317 article pairs, 25 only Spanish articles, and 88 only in English. 99.4% were correct pairs, articles without siblings were 84% correct. Two pairs were false positives. *Comayaguëla*$_{ES}$ and *Comayagua*$_{EN}$ were wrongly merged as both had the exact same coordinates, even if the cities are about 80km apart; following, the department *Comayagua*$_{ES}$ was matched to the city *Comayagua*$_{EN}$, which surprisingly was also wrong in the interlinks. This induced subsequent false negatives where the respective articles could not match to their true siblings. Other false negatives concern mostly title mismatches with distanced coordinates, but also very differing names, which brings the approach to its limits. When mapping articles as shown in Fig. 6, both languages are distributed rather similarly. The cluster of English articles in the west of the center contains mostly stubs for departments.

As future work, the approach should be transferred to other Latin American countries and to other language pairs. Additionally, the learned characteristics of feature types can be used to improve entity fusion approaches for gazetteer data.

6 Search Engine Prototype

Towards the realization of an actual search engine "Geobuscador HN", a first search engine was set up based on a default Nutch install[10] with the seeds defined in Section 4. In a first step, the crawl was restricted to .hn. The performance was mostly capped by the available

[9] `http://www.geonames.org/export/codes.html`
[10] `http://nutch.apache.org/`

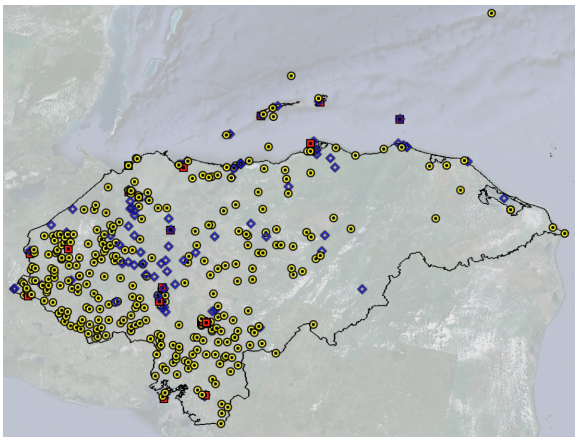

Fig. 6 Mapping of geotagged Wikipedia articles, yellow ⊙: merged articles, red ⊡: Spanish, blue ◇: English (from [7])

network speed, but a number below 100,000 pages was retrieved. This of course is easily handled by modern search engines. The next steps would have included an adapted crawling and indexing stage, as well as a large crawl to identify Honduran .com domains. The findings of Section 3.2 show that the best approach would be to use a mobile-first strategy. This would drive the building of the tailored interfaces. Additionally, a second prototype was implemented within the infrastructure of the phone operator to connect the USSD services, localization infrastructure being deployed into the network, and our initial location index.

7 Conclusion and sustainability

Overall, Honduras provides an interesting ground for research due to its numerous challenges that require the combination of different fields of search engine technology and geographic information retrieval. Furthermore, due to the small size of the country, even a research prototype can be expected to cover a huge fraction of the Honduran Web. Even with the described challenges such as low coverage, ambiguous or insufficient location references, non-local Web hosting, etc., there exists sufficient data to develop a prototypical search engine.

The whole project was a great personal experience. A major point, especially for a researcher who is a foreigner in the country, was to become aware of own biases. It also drives one to challenge preconceived assumptions, as many factors had to be established that are usually taken for granted. In that respect, work in challenging circumstances has a great learning effect.

Unfortunately, the results stayed behind the original goals. A rather sad aspect of the project is that, even while there was enthusiastic stakeholder support in the beginning, there was not enough motivation to continue the project during a funding issue or even to properly recover it afterwards so that no substantial sustainability could be reached for the full project. A more organizational issue was that doing research in the country was very different as there was not a strong background of research or even development present. There were little official or formal ways of interaction. In some cases, resources had to be procured in a very backhanded way or through a link of acquaintances.

Yet the partial solutions and prototypes described here can be useful for future projects. Due to some students moving into related industry jobs working on similar ideas, some knowledge will remain in the country and be developed further. We thus hope that the results will be used further in the country. On another note, many examined data sources have greatly improved since the start of the project, so future information access in Honduras will become easier.

Furthermore, the work carries a large potential for follow-up research, as many interesting questions are still open. One aspect is the applicability to other regions. Other aspects include the definition of a national Web outside the country-specific ccTLD as hinted at in Section 4, improved entity-based geocoding [3] as well as a hybrid search including the rich information in social networks and, especially for Honduras, the use of crowdsourcing for data gathering. Current ongoing work concerns the closer inspection of gazetteer data [4] for different countries.

Acknowledgements We thank our students Isaac Martinez and Jose Matute of UNITEC – Universidad Tecnológica Centroamericana, Tegucigalpa, Honduras, who worked within the research project and provided initial implementations, as well as the academic heads of computer science, the dean of engineering and the research department at tigo.

References

1. Ahlers, D.: Towards Geospatial Search for Honduras. In: Proceedings of the 3rd Latinamerican Conference on Networked and Electronic Media LACNEM 2011 (2011)
2. Ahlers, D.: Local Web Search Examined. In: D. Lewandowski (ed.) Web Search Engine Research, *Library and Information Science*, vol. 4. Emerald (2012).
3. Ahlers, D.: Multi-source conflating index construction for local search in a low-coverage country. In: LA-WEB 2012 (8th Latin American Web Congress), (2012).
4. Ahlers, D.: Assessment of the Accuracy of GeoNames Gazetteer Data. In: GIR '13: Proceedings of the 7th International Workshop on Geographic Information Retrieval (2013)
5. Ahlers, D.: Business Entity Retrieval and Data Provision for Yellow Pages by Local Search. In: Integrating IR technologies for Professional Search @ ECIR2013 (2013)
6. Ahlers, D.: In search of Honduras – Case report of developing local search for a developing country. In: LWA 2013 – Lernen, Wissen, Adaption (2013)

7. Ahlers, D.: Lo mejor de dos idiomas – Cross-lingual linkage of geotagged Wikipedia articles. In: ECIR2013 (2013)
8. Ahlers, D.: Towards a development process for geospatial information retrieval and search. In: WWW '13. (2013)
9. Ahlers, D.: Where the streets have no name – Experiences in GIR for a developing country. In: GIR '13: 7th International Workshop on Geographic Information Retrieval (2013)
10. Ahlers, D., Boll, S.: Geospatially Focused Web Crawling. Datenbank-Spektrum **7**(23), 3–12 (2007)
11. Ahlers, D., Boll, S.: Location-based Web search. In: A. Scharl, K. Tochterman (eds.) The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society. Springer, London (2007)
12. Ahlers, D., Boll, S.: On the Accuracy of Online Geocoders. In: Geoinformatik 2009 (2009)
13. Ahlers, D., Henze, N.: ¿Donde está? – Surveying Local Search in Honduras. In: MWB2012 – Workshop on Mobility and Web Behavior at MobileHCI2012 (2012)
14. Ahlers, D., Matute, J., Martinez, I., Kumar, C.: Mapping the Web resources of a developing country. In: GI Zeitgeist 2012, *IfGi prints*, vol. 44. AKA (2012)
15. Blumenstock, J.E., Eagle, N.: Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda. In: ICTD. (2010)
16. Boll, S., Ahlers, D.: A Web more Geospatial: Insights into the Location Inside. In: D. De Roure, W. Hall (eds.) Workshop on Understanding Web Evolution: A Prerequisite for Web Science (WebEvolve2008) held at WWW08, vol. 300. Web Science Research Initiative (2008)
17. Borges, K.A.V., Laender, A.H.F., Medeiros, C.B., Silva, A.S.D., Clodoveu A. Davis, J.: The Web as a Data Source for Spatial Databases. In: Anais do V Brazilian Symposium on Geoinformatics (2003)
18. Chen, J., Subramanian, L., Brewer, E.: SMS-Based Web Search for Low-end Mobile Devices. In: MobiCom '10. (2010).
19. Chen, J., Subramanian, L., Li, J.: RuralCafe: Web Search in the Rural Developing World. In: WWW '09. (2009).
20. Chong, A., Micco, A.: The internet and the ability to innovate in Latin America. RES Working Papers 4291, Inter-American Development Bank (2002)
21. Donner, J.: Research approaches to mobile use in the developing world: A review of the literature. The Information Society **24**, (2008).
22. Duncombe, R., Boateng, R.: Mobile phones and financial services in developing countries: a review of concepts, methods, issues, evidence and future research directions. Third World Quarterly **30**(7), (2009).
23. Gomes, D., Nogueira, A., Miranda, J., Costa, M.: Introducing the Portuguese web archive initiative. In: 8th International Web Archiving Workshop (2008)
24. Gräf, J., Henrich, A., Lüdecke, V., Schlieder, C.: Geografisches Information Retrieval. Datenbank-Spektrum **6**(18), 48–56 (2006)
25. Hill, L.L.: Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In: ECDL '00 (2000)
26. Hoffart, J., Suchanek, F.M., Berberich, K., Weikum, G.: YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence Journal (2012)
27. Jones, C.B., Purves, R.S.: Geographical Information Retrieval. International Journal of Geographical Information Science **22**(3) (2008)
28. Kothari, G., Negi, S., Faruquie, T.A., Chakaravarthy, V.T., Subramaniam, L.V.: SMS based Interface for FAQ Retrieval. In: ACL '09, (2009)
29. Kumar, C., Ahlers, D., Boll, S.: LocateThisPage: Driveby Location-Aware Browsing. In: GI Zeitgeist 2012, Proceedings of the Young Researchers Forum on Geographic Information Science, *IfGi prints*, vol. 44. AKA (2012).
30. Larson, R.R.: Geographic Information Retrieval and Spatial Browsing. GIS and Libraries: Patrons, Maps and Spatial Information (1996)
31. Markowetz, A., Chen, Y.Y., Suel, T., Long, X., Seeger, B.: Design and Implementation of a Geographic Search Engine. In: WebDB 2005. (2005)
32. Mendoza, M., Guerrero, H., Farias, J.: Inquiro.CL: a New Search Engine in Chile. In: WWW '09. (2009)
33. Murdock, V.: Your mileage may vary: On the limits of social media. SIGSPATIAL Special **3**(2) (2011)
34. Nakahira, K.T., Hoshino, T., Mikami, Y.: Geographic Locations of Web Servers under African Domains. In: WWW '06. (2006).
35. Overell, S.E., Rüger, S.M.: Identifying and grounding descriptions of places. In: GIR 2006 (2006)
36. Purves, R.S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A.K., Vaid, S., Yang, B.: The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the internet. International Journal of Geographical Information Science **21**(7) (2007)
37. Sehgal, V., Getoor, L., Viechnicki, P.D.: Entity Resolution in Geospatial Data Integration. In: GIS '06. (2006).
38. Serrano-Berthet, R., Lopez, H.: Crime and Violence in Central America: A Development Challenge. Tech. rep., The World Bank (2011)
39. Souza, L.A., Davis Jr., C.A., Borges, K.A.V., Delboni, T.M., Laender, A.H.F.: The Role of Gazetteers in Geographic Knowledge Discovery on the Web. In: LA-WEB '05: Proceedings of the Third Latin American Web Congress. (2005).
40. Tolosa, G., Bordignon, F., Baeza-Yates, R., Castillo, C.: Characterization of the Argentinian Web. Cybermetrics **11**(1) (2007)