In search of Honduras – Case report of developing local search for a developing country

Dirk Ahlers
NTNU – Norwegian University of Science and Technology
Trondheim, Norway
dirk.ahlers@idi.ntnu.no

Abstract

This paper reports on the efforts to establish a research project on a geospatial search engine for the Latinamerican country of Honduras as well as establishing an encompassing research group on information retrieval during the author's stay there 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 might be encountered. These include low Web coverage in a low-resource country, cross-language information retrieval, and generally, work in challenging circumstances. The specific focus on geospatial information uncovers further issues that need to be addressed, such as informal addressing systems, broad or incorrect location references, or insufficient ground truth in databases. The research stay yielded valuable experiences, even if the tangible results of the project stay behind the original goals.[1]

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. 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. The basic usefulness of mobile phones and their applications in developing countries have already been shown[Duncombe and Boateng, 2009; Donner, 2008; Hagan *et al.*, 2012; Frias-Martinez *et al.*, 2012; Frohlich *et al.*, 2012], but search in general has received less attention [Chen *et al.*, 2010; 2009; Kothari *et al.*, 2009]. In this paper, we want to examine the special case of geospatial search in developing regions and we explore it by means of the Latinamerican country of Honduras. The basis question then is, how can one make local search (or even Web search in general) work in a low-resource country with only very little Web coverage where few people even have Internet access? The available potential has to be assessed and a roadmap drawn for the realization.

An overview of the project's challenges of a country-specific search engine has previously been published, discussing the challenges at the beginning of the project [Ahlers, 2011]. We elaborate on these challenges with some added hindsight of research performed in the meantime in the form of this consolidated report.

1.1 Social background

Some socio-economic statistics abridged from [Ahlers and Henze, 2012] should help to better understand the expected background and population. Honduras is a developing country which is classified as a lower-middle-income economy[2], ranks 121st in the Human Development Index worldwide, and is the sixth-poorest country in Latin America[3] with 23% of the population below the poverty line[4] and 60% of the population below the national poverty line[5].

Despite these numbers, Internet use is rising fast, with 11% of Internet users in the population in 2010 up from 1% in 2000[6]. Very little reliably data is available on mobile phone or mobile Internet use in Honduras. Even a project partner, a local telecommunications company, could not readily provide such information. While computer ownership is at only 2.5%, peoples' mobile subscriptions are much more promising. From 3% in 2000, they have surpassed 100% in 2008 and were at 125% in 2010[7]. This oversaturation can be explained by the practice of having mobile phones for different providers to take advantage of lower calling cost. More useful is the estimate of people actually owning a mobile phone at about 75% in 2010[8]. Informal estimates for the share of smartphones with Internet

[1] The research described here was carried out at UNITEC – Universidad Tecnológica Centroamericana, Tegucigalpa, Honduras.

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.

[2] http://data.worldbank.org/country/honduras

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

[4] http://www.unicef.org/infobycountry/honduras_statistics.html

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

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

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

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

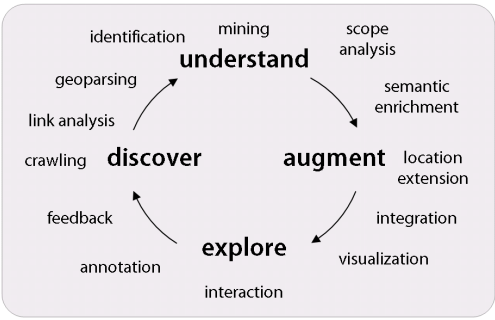

Figure 1: Process life cycle of geospatial Web information (from [Boll and Ahlers, 2008])

access are around 10%, which hints that a lot of Internet use happens on mobile devices.

While the very unevenly distributed use of communication technology denotes a strong digital divide in Honduras, mobile phone and Internet use is rising. This 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. Overall, this means that Honduras is a latecomer but is quickly catching up, which makes local search a viable option for future applications. With increasing use of online services, this might also reach larger parts of the population [Chong and Micco, 2002].

This might 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 small shops in their neighborhood. These people mostly will not need these services. 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, especially in view of a large poor population, the adoption of services for the population capable of affording the necessary technology to participate, can still be a worthwhile undertaking and may, by uncovering and presenting available information, improve the general data situation and might lead to other undertakings that might benefit a larger amount of the population.

2 Development process for geospatial search

Many research projects have been undertaken to extract the location information for a special vertical geospatial Web search, e.g., [Ahlers and Boll, 2007; Purves *et al.*, 2007; Markowetz *et al.*, 2005; Borges *et al.*, 2003]. 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 [Ahlers, 2012a]. However, no tailored geospatial search for Honduras exists and the existing services showed some shortcomings. The initial project idea therefore was to work towards a geospatial database of Honduran locations, places, points of interest and Web pages in close cooperation with a local mobile telecom provider [Ahlers, 2011].

The main challenges concern the requirements and data situation; and the analyzing, extracting and indexing of location data. We base this on the general process life cycle of finding and using geospatial Web information presented in [Boll and Ahlers, 2008]. This includes processes of discovery, understanding, augmentation, and exploration as shown in Fig. 1. Extending this data-centric viewpoint towards the process for a whole search engine, we arrive at a development process for a geospatial search engine [Ahlers, 2013c]:

General situation assessment and understanding: The first step in creating a new geospatial search engine is to get an overview of what applications and services may be interesting and relevant. Specific undertakings range from market analysis, viability analysis, and data source investigations over requirements engineering to data analysis and user studies, regarding search and/or mobile applications. These can provide initial insight into the needs and gaps in the currently available systems and would also provide requirements for subsequent steps.

- **Market:** Quite an obvious step, a market analysis should find out if the intended solution is actually new and needed or if other participants can provide some help.
- **Data:** The data situation is the most important, as it plays an important role in the feasibility analysis. This includes how much data is available, what its characteristics and quality is, which additional sources are available.
- **Users:** As a search engine is an offer towards users, their requirements and situation have to be taken into account. Also, the target group needs to be identified or selected first. Informal interviews, user studies, or usage observation can provide valuable insight.

Building a knowledge base: To aid in the extraction of geospatial data, a bootstrapping of known geographical placenames is normally used. Such gazetteer data can either be directly available or needs t be collected and combined from multiple sources.

Discovery and analysis of data sources: To get a good overview, many data sources have to be explored to understand the type of information they offer and estimate the amount of data available.

Extraction and analysis of data: Specific extraction methods have to be developed. These especially concern geoparsing, i.e., the identification and extraction of location references, and geocoding, i.e., the grounding of location references to geographic coordinates.

Web crawling: For crawling a country-specific Web, the characteristics and boundaries of the country in the Web have to be defined to setup parameters for the crawl.

Source integration: The different identified sources need to be integrated. Local search is not just a document search but is also an entity search in the sense that it models georeferenced documents as well as the actual georeferenced entities described in the documents. For this cross-correlation and entity resolution across all results is performed.

Building interfaces: Finally, based upon the available data and potential augmentations from additional analysis steps, interfaces for users to actually access the index are to be developed.

Then we further have cross-sectional issues such as problems encountered along the way in all the different steps,

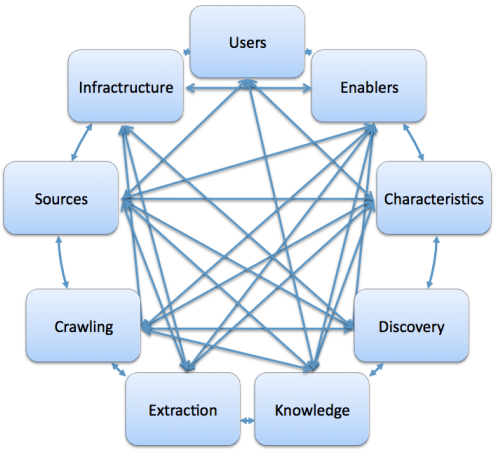

Figure 2: Components in the development of geospatial search (from [Ahlers, 2013c])

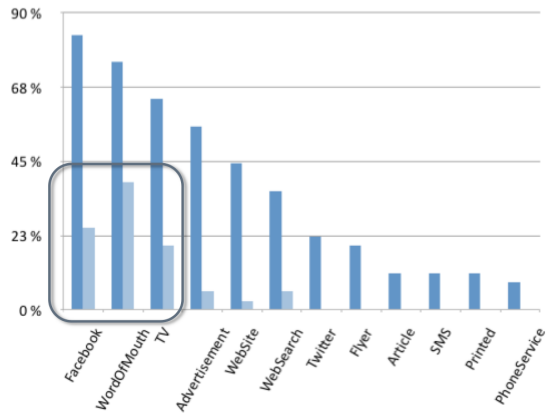

Figure 3: Information seeking behaviour: used sources. Answers to the question "How do you find out about local things?" Dark blue: multiple choice answers, light blue: only primary choice. (from [Ahlers and Henze, 2012])

which have to be solved and integrated. A major point here, especially for a researcher who is a foreigner in the country, is to be aware of own biases and preconceived notions or assumptions. To this end, one should try to talk to as many people as possible and try to collect as much information as possible, especially if it is contradictory. Actually, this is a good hint that something is not as easy as it might seem. As in good journalism, it is better to ask at least two sources so that assumptions cannot go wrong that easily. This classification and more formal description is a work in progress [Ahlers, 2013c] which is visualized with its strong interdependencies in Fig. 2.

3 Challenges and realization

We will shed some light onto this process by examining it from the viewpoint of geospatial search in Honduras. We discuss challenges identified in [Ahlers, 2011] and organize them along the development process. We also expand the presentation with research results gathered in the meantime.

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 very low, often offering nothing more than a name and a rough location. Instead of waiting for other players to take up the market, the current 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 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[9]. Local search [Ahlers, 2012a] has initial data, but is far from a comprehensive coverage – which is in part the topic of this research. Similar research projects have been described, for example, for Chile [Mendoza *et al.*, 2009], Portugal [Gomes *et al.*, 2008], Brazil [Borges *et al.*, 2003], or Germany [Markowetz *et al.*, 2005; Ahlers and Boll, 2007].

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

3.2 Users

The most comprehensive overview on mobile phone use in developing countries was a literature study [Donner, 2008] reviewing about 200 studies. An ethnographic study of 26 participants by means of interview and shadowing was done to examine the use of mobile phones to maintain a social network for migrant workers in cities in China [Lang *et al.*, 2010]. It was noted that social interactions happen throughout the day, with little distinction between work and spare time, for non-factory workers. The use of mobile phones can also be understood as a method of empowerment in developing nations. [Blumenstock and Eagle, 2010] analyzed patterns of mobile phone use in Rwanda with a joint approach of using demographic surveys and call detail records analysis of a mobile operator, additionally discussing other surveys done in the developing world.

For insights focused on Honduras, we did a user study on the use of local search and local information seeking behavior in general [Ahlers and Henze, 2012]. 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 as seen in Fig. 3. This social aspect may also explain why the most used online source is Facebook, followed with a distance by search engines. Overall, local search is not that prevalent and social aspects are very strong. Due to rather little search happening on the move, there is less of the usually associated here-and-now mentality [Ahlers and Henze, 2012]. One implication of these findings is that the search engine most probably would have to 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 and security in geolocalization

In Honduras, privacy concerns are connected to bigger security concerns. Due to a high level of targeted criminality, many people prefer to keep their personal information, especially their location, very private [Ahlers and

Figure 4: Entity example with street-level address from different sources, counter-clockwise: Facebook, yellow pages, foursquare, OpenStreetMap (from [Ahlers, 2012b])

Henze, 2012]. Yet, many entries in location-sharing services explicitly concern people's own 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 not walk and rather take cars, taxis, or buses and would only get their phone out in safe places, but not just on the street. This influences usage, which is not as spontaneous as in other countries and happens less on the move. Therefore, developed services need to ensure safe handling of people's location data and also consider a less spontaneous mobile use.

3.3 Country-specific characteristics

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 sometimes 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. Sometimes a description is accompanied by a sketched map, a so-called croquis to help with orientation. The usually encountered 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. 4 shows varying descriptions and given locations for a place.

3.4 Data

The official language of Honduras is Spanish, and normally, the articles from this language would be expected to sufficiently cover the country. However, at the Caribbean coast and Bay islands of the country, English is a recognized regional language and is more frequently spoken, mainly by the Garifuna population. Additionally, this is the main area for foreign tourism and much information of a tourism nature is more comprehensively available in English than in Spanish. Therefore, a search engine that should cover the whole of Honduras needs to use cross-language retrieval techniques.

Furthermore, there is an interesting anomaly for Honduras in that much information about the country exists in English instead of Spanish. For example, while examining Wikipedia, we noted that there exist more English

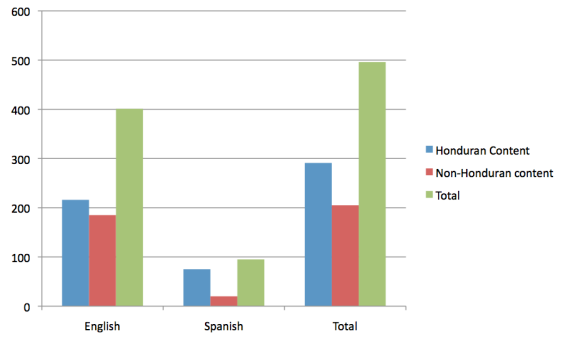

Figure 5: Provenance of pages with topic Honduras in DMOZ [Ahlers *et al.*, 2012]

geotagged articles than Spanish ones. This is shown as an example in Section 4.

In trying to define the Honduran Web, we took two separate approaches as discussed in [Ahlers *et al.*, 2012]. First, we had a look at DMOZ, and second, we built a list of all Honduran domains under a .hn ccTLD. The DMOZ Open Directory has relatively little coverage for Honduras. In the English hierarchy, it contains 421 entries, with 10 from .hn (2.5%) and 411 others; for Spanish, there are only 96 entries, but 46 are from .hn (48%), only 50 from others. The English part contains mainly travel sites and general descriptions while the Spanish contains actual local pages. Fig. 5 shows the provenance of the DMOZ data for the English and Spanish categories, classified by being from inside or outside the country.

We were able to obtain a list of registered domains through inofficial channels. Of the 5780 registered domains, we were able to only find 893, meaning that probably a lot of them are not used. Trying to confirm a suspicion of a majority of Web hosting happening outside the country, we used a commercial database to estimate the hosting locations of the reachable domains. The country assignment is shown in Fig. 6 on a logarithmic scale. A large amount is actually assigned in Honduras, but the majority in the US, with other American countries following behind. A deeper inspection of the US hosts revealed these to be mostly in southern countries associated with a large Latinamerican population, while the other countries are often related to the owners or investors of businesses. However, we also found 25% of governmental domains hosted in the US, confirming the suspicion of a 'digital divide' [Nakahira *et al.*, 2006].

Due to unknown deterrents, possibly high server costs or similar, a very high number of Honduran businesses not only host outside the country, but also do not use the .hn domain and instead opt for a generic .com domain. Further complications arise from the fact that many businesses forego an own Web presence and instead create a Facebook page. This makes it more difficult to gather all relevant domains for Honduras and actually include all relevant information. As a first rough estimate on the sites that are available, all Honduran DMOZ links are correlated with the known domains registered for .hn. This gives a number from a conservative 6200 to a probably heavily overestimated 225,000 domains overall, with its geometric mean below 40,000 domains. This means that .hn domains only represent an estimate of between 5% to 61% of all relevant domains for the country.

We are looking into ways to reliably identify .com do-

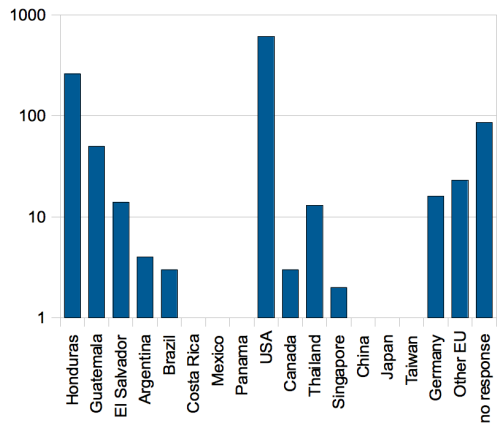

Figure 6: Distribution of hosting countries for .hn domains [Ahlers *et al.*, 2012]

mains carrying Honduran content, using a mixture of location analysis, language, and interlinking.

3.5 Building a gazetteer

For a small country with very limited Web coverage, the geotagged Wikipedia articles can provide some reliable initial knowledge (cf. Section 4). We injected the Wikipedia data into a larger gazetteer that we are generating for Honduras. For the gazetteer, we use initial data from geonames.org to serve as a bootstrapping of the search engine's knowledge about places and placenames.

3.6 Source integration

Since the Web shows only very low coverage for Honduras, we aim to additionally include 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 mandates 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. However, even for only Web-based location search, cross-correlation and entity resolution is needed to identify identical entities (cf. [Ahlers, 2013a; Sehgal *et al.*, 2006]). The multiple sources also may provide a remedy for low-granularity locations, as they may be combined in some cases to improve the accuracy of geocoding [Ahlers, 2012b] (cf. Fig. 4).

4 Wikipedia cross-language linkage

Wikipedia includes a vast amount of articles about places, many of which include a geographic coordinate that locates the content in the real world. This makes it a good starting point to bootstrap out knowledge about a country. Cross-language links are possible between articles describing the same place in multiple languages. Ideally, any article about a place of interest would include its correct coordinate and links to other Wikipedia language versions. This poses the question of how we can identify geotagged Wikipedia articles that describe the same place across different languages and what are useful similarity measures? This section is based on [Ahlers, 2013b].

The official language of Honduras is Spanish, and normally, the articles from this language would be expected to sufficiently cover the country. However, much information about the country comes from outside. We therefore also looked into English articles to see if these would increase the coverage. This prompted the discovery of an interesting anomaly: Honduras has more English geotagged articles than Spanish ones.[10] Our aim is to merge both language versions and identify identical places on the article level.

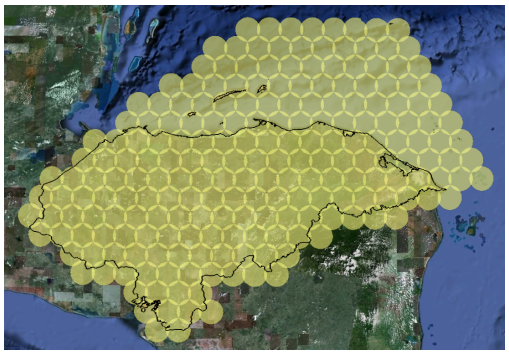

Figure 7: Hexagonal grid arrangement of packed circles for API queries

Table 1: Examples of sibling articles

Spanish title	English title
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
Virginia (Honduras)	Virginia, Lempira
Parque nacional Pico Bonito	Pico Bonito National Park

The approach we are following is called, varyingly, record linkage, entity fusion, entity resolution, or duplicate detection. [Sehgal *et al.*, 2006] gives an overview on geospatial entity resolution. [Overell and Rüger, 2006] use Wikipedia to ground and disambiguate place names. Merging geonames data to Wikipedia, [Hoffart *et al.*, 2012] use a simple approach that if more than one entity exists in geonames with the same name, the closest entity within a distance of 5km is chosen. [Liu and Yoshioka, 2011] add a translation approach to improve the title matching. This work is closest to ours, however, is still lacking in a graded consideration of both textual and positional similarity.

We use the APIs of geonames and wikilocation to retrieve articles, because we do not want to retrieve the whole Wikipedia dump for the relevant languages. Since the services put a limit of 20km in the radius for simple circle search, we have to use a covering of queries for the region, which is shown in Fig. 7.

4.1 Wikipedia Language Fusion

We define the data fusion method in terms of finding language *siblings*. We combine text- and entity-based merging methods with geographic conflation techniques. For each article, we select and rank candidate siblings in the respective other language. The merging is based on the title and the location as shown in the examples in Table 1. The geographic feature type is rarely present, so it can only used as second-level evidence. For a comparison of two potential siblings, there are four cases to consider, 1. Names

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

and coordinates match, 2. Names match, coordinates do not match, 3. Names do not match, coordinates match, 4. Names do not match, coordinates do not match. The first case is obviously trivial. All other cases are modeled by similarity measures based on non-exact matching.

Coordinates can vary due to different interpretations of the center of an area or variations in user-generated coordinates, especially for larger entities [Ahlers and Boll, 2009]. 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*, inside of which all candidates are examined.

For all candidates' *titles* within the radius, three cases would constitute a match, 1) titles match exactly, 2) titles match with small variations, 3) title can be translated and transposed to match. We define a title translation distance TTD as an editing distance similarity measure based on partial translations and permutations. The first case is easy, the second case only needs to account for spelling variations, which we do 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 be well accounted for with the adapted Levenshtein distance. 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[11] (e.g., airports, islands, mountains, stadiums, cities, parks, etc.). Heuristics were generated about some conventions that we observed for both languages. 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*$_{EN}$ uses the administrative type without a hint towards the country. This is helpful as often, departments and capital cities have the exact same coordinates.

To cover permutations, we employ a list of transposition heuristics as part of the translation. The inverted-first-pair translation swaps the first two terms: *Congreso Nacional de Honduras*$_{ES}$ $\rightarrow$ *National Congress of Honduras*$_{EN}$. The inverse order translation swaps first and last terms: *Río coco*$_{ES}$ $\rightarrow$ *Coco river*$_{EN}$; and the inverted-first-pair-moved translation inverts the order of the first two words and moves them to the end: *Parque nacional Pico Bonito*$_{ES}$ $\rightarrow$ *Pico Bonito National Park*$_{EN}$. We generate all potential variations of the title, including translations, and chose the variation with the minimum TTD and the smallest location distance as a sibling.

4.2 Evaluation

Honduras had 342 Spanish and 405 English articles, an 18% English overrepresentation. We use the Wikipedia language interlinks as a ground truth for the evaluation. For all articles, the Wikipedia page and its interlinks were manually examined to determine siblings.

The algorithm resulted in 317 article pairs, 25 only Spanish articles, and 88 only in English (Fig. 8). Of these, 99.4% are correct pairs. The articles without siblings are 84% correct, with 16% false negatives. Only two pairs were false positives. The first wrongly identifies *Comayaguëla*$_{ES}$ and *Comayagua*$_{EN}$ because they have both

[11] http://www.geonames.org/export/codes.html

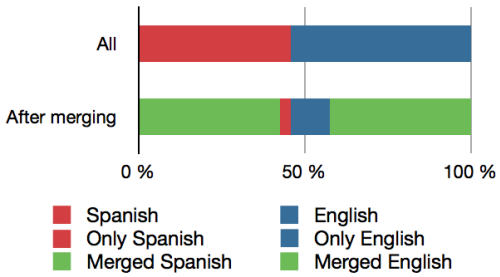

Figure 8: Results of merging Wikipedia articles

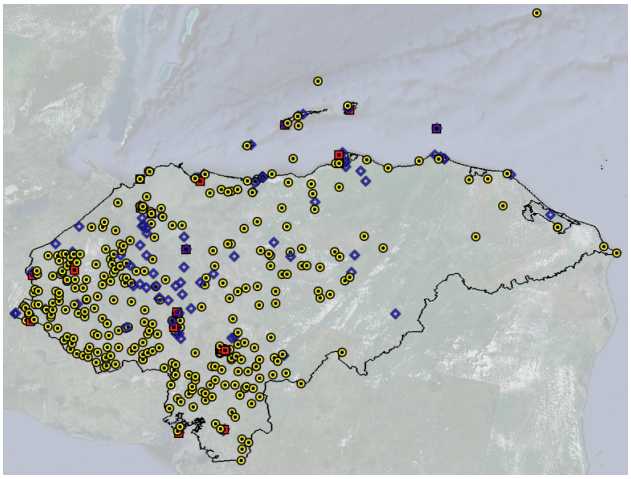

Figure 9: Mapping of geotagged Wikipedia articles, yellow ⊙: merged articles, red ⊡: Spanish, blue ◇: English

the exact same coordinates, even if the cities are about 80km apart. In this case the error lies with the incorrect coordinate in the article. The second assigns the department *Comayagua*$_{ES}$ to the city *Comayagua*$_{EN}$, which surprisingly is also wrong in the interlinks. This induces a subsequent error in the false negatives: *Comayagua (ciudad)*$_{ES}$ and *Comayagua Department*$_{EN}$ each had no siblings, but should have been matched to the previous pair. The other false negatives concern mostly slight mismatches paired with distanced coordinates, but also some more debatable ones, such as *Roatán (municipio)*$_{ES}$ and *Coxen Hole*$_{EN}$. When mapping articles as shown in Fig. 9, we see no language dominating certain regions but both languages distributed rather similarly.

The approach is to be extended by using the learned characteristics in an entity fusion approach for gazetteer data as well, which will make stronger use of the feature type. This is expected to help in cleaning up and linking geonames data to other sources.

5 Conclusion and sustainability

Overall, Honduras provides an ideal ground for research due to its numerous challenges that will require the combination of many 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, thus building up a comprehensive and usable index.

Even with the complications described above such as low coverage, ambiguous or insufficient location references, non-local Web hosting, etc., there exists sufficient data to develop a prototypical search engine, starting with

some "easier" aspects of the data. 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. In some cases, resources had to be procured in a very backhanded way. For example, there did not seem to be an official way to get certain numbers, but a student knew someone who might have access to certain data. There were little official or formal ways of interaction and in terms of doing research, Honduras very clearly shows the signs of a developing country. A rather sad aspect of the project is that, even while there was enthusiastic support of the general idea, 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 until this point for the full project, but the partial solutions described here are still useful for future projects. However, the whole project was a great experience and due to some students moving into related industry jobs working on similar ideas, some knowledge will remain in the country and be developed further.

Overall, the project offered very good potential for research. It also drives one to challenge certain assumptions, as many factors have to be established here that can be taken for granted in other regions. We hope that the results will be used further in the country and we also hope that may be applicable to other regions as well. Furthermore, the work carries a large potential for follow-up research, as many interesting questions are still open.

Acknowledgment

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.

References

[Ahlers and Boll, 2007] Dirk Ahlers and Susanne Boll. Location-based Web search. In Arno Scharl and Klaus Tochterman, editors, *The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society*. Springer, London, 2007.

[Ahlers and Boll, 2009] Dirk Ahlers and Susanne Boll. On the Accuracy of Online Geocoders. In *Geoinformatik 2009, Osnabrück*, volume 35 of *ifgiprints*, pages 85–91, Münster, 2009.

[Ahlers and Henze, 2012] Dirk Ahlers and Niels Henze. ¿Donde está? – Surveying Local Search in Honduras. In *MWB2012 – Workshop on Mobility and Web Behavior at MobileHCI2012*, 2012.

[Ahlers *et al.*, 2012] Dirk Ahlers, Jose Matute, Isaac Martinez, and Chandan Kumar. Mapping the Web resources of a developing country. In *GI Zeitgeist 2012, Proceedings of the Young Researchers Forum on Geographic Information Science*, volume 44 of *IfGi prints*, pages 117–122. AKA, 2012.

[Ahlers, 2011] Dirk Ahlers. Towards Geospatial Search for Honduras. In *LACNEM 2011*, San José, Costa Rica, 2011.

[Ahlers, 2012a] Dirk Ahlers. Local Web Search Examined. In Dirk Lewandowski, editor, *Web Search Engine Research*, volume 4 of *Library and Information Science*, pages 47–78. Emerald, 2012.

[Ahlers, 2012b] Dirk Ahlers. Multi-source conflating index construction for local search in a low-coverage country. In *LA-WEB 2012 (8th Latin American Web Congress)*, pages 28–31, 2012.

[Ahlers, 2013a] Dirk Ahlers. Business Entity Retrieval and Data Provision for Yellow Pages by Local Search. In *Integrating IR technologies for Professional Search Workshop @ ECIR2013*, 2013.

[Ahlers, 2013b] Dirk Ahlers. Lo mejor de dos idiomas – Cross-lingual linkage of geotagged Wikipedia articles. In *ECIR2013*, 2013.

[Ahlers, 2013c] Dirk Ahlers. Towards a development process for geospatial information retrieval and search. WWW '13, pages 143–144, Geneva, Switzerland, 2013. IW3C.

[Blumenstock and Eagle, 2010] J. E. Blumenstock and N. Eagle. Mobile divides: Gender, socioeconomic status, and mobile phone use in rwanda. In *4th International Conference on Information and Comunications Technologies and Development (ICTD)*, 2010.

[Boll and Ahlers, 2008] Susanne Boll and Dirk Ahlers. A Web more Geospatial: Insights into the Location Inside. In David De Roure and Wendy Hall, editors, *Workshop on Understanding Web Evolution: A Prerequisite for Web Science (WebEvolve2008) held at WWW08*, volume 300, pages 40–43, Beijing, China, 2008. Web Science Research Initiative.

[Borges *et al.*, 2003] Karla A. V. Borges, Alberto H. F. Laender, Claudia B. Medeiros, Altigran S. Da Silva, and Jr. Clodoveu A. Davis. The Web as a Data Source for Spatial Databases. In *Anais do V Brazilian Symposium on Geoinformatics*, 2003.

[Chen *et al.*, 2009] Jay Chen, Lakshminarayanan Subramanian, and Jinyang Li. RuralCafe: Web Search in the Rural Developing World. In *18th International World Wide Web Conference*, WWW '09, pages 411–420, New York, NY, USA, 2009. ACM.

[Chen *et al.*, 2010] Jay Chen, Lakshmi Subramanian, and Eric Brewer. SMS-Based Web Search for Low-end Mobile Devices. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 125–136, New York, NY, USA, 2010. ACM.

[Chong and Micco, 2002] Alberto Chong and Alejandro Micco. The internet and the ability to innovate in latin america. RES Working Papers 4291, Inter-American Development Bank, Research Department, January 2002.

[Donner, 2008] Jonathan Donner. Research approaches to mobile use in the developing world: A review of the literature. *The Information Society*, 24:140–159, May 2008.

[Duncombe and Boateng, 2009] Richard Duncombe and Richard Boateng. Mobile phones and financial services in developing countries: a review of concepts, methods, issues, evidence and future research directions. *Third World Quarterly*, 30(7):1237–1258, 2009.

[Frias-Martinez *et al.*, 2012] Vanessa Frias-Martinez, Jesus Virseda, and Aldo Gomero. Educamovil: a mobile learning tool for low-income schools. MobileHCI '12, pages 169–172, New York, NY, USA, 2012. ACM.

[Frohlich *et al.*, 2012] David Frohlich, Simon Robinson, Kristen Eglinton, Matt Jones, and Elina Vartiainen. Creative cameraphone use in rural developing regions. MobileHCI '12, pages 181–190, New York, NY, USA, 2012. ACM.

[Gomes *et al.*, 2008] Daniel Gomes, André Nogueira, João Miranda, and Miguel Costa. Introducing the Portuguese web archive initiative. In *Proceedings of the 8th International Web Archiving Workshop*, Aaarhus, Denmark, September 2008.

[Hagan *et al.*, 2012] Margaret Hagan, Nan Zhang, and Joseph 'Jofish' Kaye. Safe mathare: a mobile system for women's safe commutes in the slums. MobileHCI '12, pages 47–52, New York, NY, USA, 2012. ACM.

[Hoffart *et al.*, 2012] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence Journal, Special Issue on Artificial Intelligence, Wikipedia and Semi-Structured Resources*, 2012.

[Kothari *et al.*, 2009] Govind Kothari, Sumit Negi, Tanveer A. Faruquie, Venkatesan T. Chakaravarthy, and L. Venkata Subramaniam. SMS based Interface for FAQ Retrieval. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 852–860, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[Lang *et al.*, 2010] Xueming Lang, Elisa Oreglia, and Suzanne Thomas. Social practices and mobile phone use of young migrant workers. In *Proceedings of the 12th international conference on Human Computer Interaction with mobile devices and services*, MobileHCI '10, pages 59–62, New York, NY, USA, 2010. ACM.

[Liu and Yoshioka, 2011] Yiqi Liu and Masaharu Yoshioka. Construction of large geographical database by merging Wikipedia's Geo-entities and GeoNames. Technical Report SIG-SWO-A1102-03, 2011.

[Markowetz *et al.*, 2005] Alexander Markowetz, Yen-Yu Chen, Torsten Suel, Xiaohui Long, and Bernhard Seeger. Design and Implementation of a Geographic Search Engine. In AnHai Doan, Frank Neven, Robert McCann, and Geert Jan Bex, editors, *WebDB 2005*, pages 19–24, Baltimore, Maryland, USA, 2005.

[Mendoza *et al.*, 2009] Marcelo Mendoza, Hipolito Guerrero, and Julio Farias. Inquiro.CL: a New Search Engine in Chile. In *WWW '09: 18th International World Wide Web Conference (WWW in Ibero-America track)*, WWW '09. ACM, 2009.

[Nakahira *et al.*, 2006] Katsuko T. Nakahira, Tetsuya Hoshino, and Yoshiki Mikami. Geographic locations of web servers under african domains. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 989–990, New York, NY, USA, 2006. ACM.

[Overell and Rüger, 2006] Simon E. Overell and Stefan M. Rüger. Identifying and grounding descriptions of places. In Ross Purves and Chris Jones, editors, *Proceedings of the 3rd ACM Workshop on Geographic Information Retrieval, GIR 2006*, Seattle, WA, USA, 2006. Department of Geography, University of Zurich.

[Purves *et al.*, 2007] Ross S. Purves, Paul Clough, Christopher B. Jones, Avi Arampatzis, Benedicte Bucher, D. Finch, Gaihua Fu, Hideo Joho, Awase Khirni Syed, Subodh Vaid, and Bisheng Yang. 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):717–745, 2007.

[Sehgal *et al.*, 2006] Vivek Sehgal, Lise Getoor, and Peter D Viechnicki. Entity Resolution in Geospatial Data Integration. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, GIS '06, pages 83–90, New York, NY, USA, 2006. ACM.