Towards Geospatial Search for Honduras

Dirk Ahlers
UNITEC – Universidad Tecnológica Centroamericana, Sistemas Computationales
Tegucigalpa, Honduras
and
Independent Search Consultant, Oldenburg, Germany
ahlers@dhere.de

***Abstract*—Local Web search has become a major and popular means of gathering information with a geographic relation. Local search not only relies on powerful algorithms from Geographic Information Retrieval and other fields, but also on the broad availability of location-referenced data on the Web and in other sources. While available commercial systems work satisfactory in well-developed countries, other regions of the world pose challenges in all areas of geospatial search. This article examines the potential for local search for the country of Honduras. The challenges, among others, are the very sparse Web coverage, few reliable commercial or public databases, challenging informal address schemes, uncertain and only coarse location data as well as privacy and security issues. This article aims to discuss these challenges and provide solutions within ongoing work towards the development of a geospatial search engine for Honduras.**

***Index Terms*—Geospatial Web Search, Search Engine, Information Retrieval, Information Extraction, Entity Retrieval, Feasibility Study**

I. Introduction

Along with the immense growth of the World Wide Web, search engines have developed alongside to index its vast amount of documents and make it accessible to users. In the same way, more and more location-based information is put online. To make sense of it and make it accessible to search, many research efforts have been undertaken [1], [2], [3], [4] and commercial services such as Google Maps, Google Earth, Yahoo! Maps are building location-based search applications and creating indexes of geospatial information [5]. The geospatial properties of this information allow searches different from common Web search. Instead of searching with keywords and getting back a list of matching documents, location-based search allows searching by placing results on a map, searching for distances or regions, and retrieving more rich and structured information about the georeferenced entities. In this sense, local search takes a sole document search over to an entity search in the sense that it not only models georeferenced documents, but the actual georeferenced entities described in the documents [1].

To implement such a system, instead of general purpose Web search engines, the concept of vertical or specialized search engines is employed. Vertical search engines aim to unlock information for specific tasks or for domain- or topic-specific search [6]. In vertical search, specific restrictions are combined with topic-specific processing, indexing and search. This allows to take specific characteristics and potential metadata into account and build tailored search interfaces and applications. Local search is thus a vertical search as it concerns a special non-keyword characteristic of data, requires special processing and indexing structures and allows interactions beyond the pure keyword paradigm.

Geospatial search has a huge potential because location is a very powerful organizing principle in people's lives due to its strong correlation to the real world. However, most research and development efforts go towards developed countries, primarily the US and Europe in correspondence with the home of search companies and perceived high-impact economies. The Honduras search engine market it is currently dominated by US commercial search engines without any discernible independent players. Some smaller directories are either outdated and unmaintained or currently being set up so that no sufficient coverage is reached. Yet, Web search in itself is rather universal and thus, general search engines can easily also gather information about smaller countries, which they in fact do and thus are often sufficiently useful. The example of google.hn provides a localized service that prefers pages dealing with Honduras, but still accesses all other pages in the index. Obviously this already uses some method for geo-referencing of documents at a coarse granularity, apparently mostly based on domain names, language, and keywords.

The situation becomes fiercer when looking at local search for Honduras. 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 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.

The remainder of this paper will explore the individual challenges posed by the situation in Honduras and offer potential solutions within a research and development project. The research concerns the investigation into the use of location information to build up a geospatial search engine with a possible mobile usage with a large mobile network provider. This work details the initial stages of the project, potential

data sources, the approach and proposed architecture and the future steps and developments.

II. Challenges

A. Requirements for location-based search and mobility

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. A user study and survey is being prepared to gain insight into people's use of location data and how they search and find out about local information, what the preferred modes and sources of information search are and how this can be incorporated into the final search engine. The survey will deal with both discovery search and recovery search aspects as well as technological ability, equipment, and knowledge. As a preliminary result, it can be said that the preferred information seeking mode remains word-of-mouth or existing knowledge about locations, together with a knowledgeable social circle. Search engines are used, but less so for local information.

B. Developing a geographical search engine for Honduras

As discussed before, the need for a focused Honduras local search exists, but cannot be well met by any existing services. Therefore, a geospatial search engine with a country-level scope for Honduras will have to be developed. Due to the data source situation, it has to follow a hybrid approach of both Web search with georeferencing of documents and additionally database access and merging for specific data sources. This is explored in more detail in Section IV. This system is by far the largest challenge to face, but the following ones are also dependent on this and need to be solved as they discuss prerequisites.

C. Analysis of Honduras local data sources and the Honduran Web graph

The first steps of market analysis have been finished without significant findings as discussed above. Since no search engine market was found, the examination moved over to a survey of potential data sources for a geospatial search. As a search engine should primarily be an automatic retrieval system and only secondly, if ever, be supported by user-contributed content, the identification and analysis of useable data sources is of the utmost importance. A huge number of sources have been identified that may provide data and be a target for entity retrieval apart from a general Web search engine. This is detailed in Section III. For the development of local search based on the Web itself, Web graph and link structure analysis can provide interesting insights into the available data of the Honduran Web, inform a ranking method and help to gather information more efficiently.

D. Exploiting coarse location data

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 [7], [8], [9].

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 [10]. Yet, geoparsing is a widely examined issue in geographic information retrieval and has been extensively researched. Therefore, performing region-based geoparsing would be possible. To improve on this result, Section IV will explore possibilities to still reach a sufficiently high granularity even of location references of only coarse addresses.

E. Granularity issues in mobile applications

The aforementioned issue of lacking exact location data not only proves difficult for the extraction of information, but also for its use. User's location data is used in many applications as an initial and most important hint for the retrieval system. If information is then displayed on a map, the location is sufficiently marked. But without a map interface, such as giving a description of a place, the lack of an exact address proves difficult. If an original navigational description was available, it can be given as part of the results. But in other cases, a description might have to be generated based on nearby buildings, major streets etc., once these are available [11]; similar to a landmark navigation [12], [13].

F. Privacy and security in geolocalization

Common privacy issues are that revealing one's location or also aggregate locations could allow attackers to infer one's home, route, or favourite locations. On the other hand, using location services for applications such as having an overview of the locations of close friends or colleagues has many benefits. However, besides the privacy implications, Hondurans need to consider actual security concerns. Due to a high level of targeted criminality, many people prefer to keep their personal information, especially their location, very private. Yet, many entries in location-sharing services in Honduras explicitly concern people's own houses ("Mi casa", "My house"). In these cases, the functionality seems to override security concerns. Still, the services being developed will

have to take care to implement techniques and processes that consider the requirements for security and location-privacy.

III. Potential Data Sources

A survey of available data sources for Honduras was done to achieve an initial birds-eye-view of existing data as discussed in Section II-C. The survey was started by known data sources for other countries or global sources or search engines. Expanding from these by general Web search and specialized search engines, more sources were uncovered. This lead to a large amount of potential data sources. The following considers specialized data sources and the Deep Web [14], [15] and subsequently discusses the general Web.

A. Directories and geospatial databases

For structured entity information a number of sources were identified. A classification of the sources was performed and following, an initial probing was performed into information depth, information density and other properties. A short overview of the information amount is given in Fig. 1 in a logarithmic scale. The amount of entities was estimated by querying for locations or the keyword for Honduras, where applicable. In some cases, the straightforward approach does not work and there have not yet been built any implementations of connectors for the data source, which explains some empty bars. The sources are extremely heterogeneous. While some show very detailed information, others only show a name and partial addresses. This influences the uses for the data that can range from actual data points to just supporting information or to a discarding of the data in some cases. Various mapping services exist, that not only can provide good satellite imagery, but also road networks.

Apart from actual entity information from, e.g., OpenStreetMap, Maps, travel sites etc., there is some gazetteer information that will be used to ground locations in the proposed search engine. The most comprehensive was found to be geonames, with others having only a fraction of its coverage. However, one interesting finding is that geonames has the highest amount of locations. This means that all other services have much less coverage than simply all place names in Honduras. When removing rivers, mountains etc. and only retaining populated places' names, geonames ranks still higher than most sources which indicates that these do not even have one entity per city or town, leaving much work to be done.

This is ongoing work into learning more about the data sources and extracting structured entity data.

B. Expected estimates for Web coverage

A geospatial search should not only take in data from other directories, but actually retrieve Web pages dealing with the location by finding these geo-relevant unstructured documents and analyzing them. For this, a first step is to assess where and how pages dealing with Honduras locations can be found. Therefore, the number of potential domains with Web information pertaining to Honduras is being estimated.

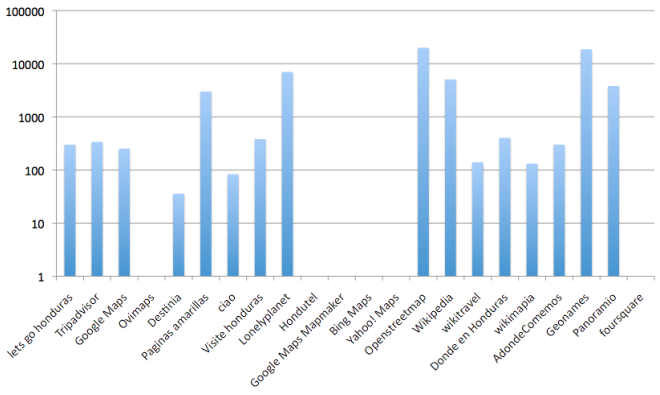

Fig. 1. Overview of information amount for examined sources

domain name	amount
.hn	4260
.com.hn	1253
.gob.hn	85
.org.hn	84
.net.hn	56
.edu.hn	41
.mil.hn	1
sum	5780

Fig. 2. Assigned Honduran domain names (as of June 2011)

The first step was to examine the Honduras IP space and Honduran domain names. Data from an IP block database[1] shows a number of about 129,000 IPs in the blocks assigned to Honduras. This number only counts address blocks, not assigned IP addresses or actual hosts, so the actual host number is much smaller. The regional Network Coordination Center LACNIC has assignment percentages for the overall address space handled[2]. For those blocks in which Honduras (through nic.hn) has a share, the average assignment rate lies about 90% and would thus come at about 116,000 hosts. Note that not all of these IP addresses are actually assigned, and not all those assigned are active, and not all those active run a Web server. Thus the expected number of actual Web servers is much lower.

More detailed data on domain names comes from the national domain name registrar[3]. For data protection reasons, the DNS system that provides domain names does not allow to search for them by IP; reverse DND is not always reliable. It also does not give out all domain names that it would have available. This is not specific to Honduras, but to all countries. However, some statistics are available about the assigned domain names. Honduras uses a system of structured second-level namespaces, but it is not mandatory and direct use of a top-level domain name is possible, too. Table 2 shows that there are about 5780 domain names registered in

[1] http://www.countryipblocks.net/country-blocks/select-formats/
[2] http://www.lacnic.net/en/registro/espacio-disponible-ipv4.html
[3] http://nic.hn/english/index.html

Honduras. Apart from reverse DNS, it is possible to query each of these IP addresses to find out if there is a Web server running on the default port. However, in many cases, the servers use a virtual host technology to run multiple hosts on one server. In this case, the domain name has to be provided, otherwise no information is available. For example, the IP address of Unitec (200.107.212.133) provides multiple domains, not only for unitec.edu, but also for www.unitec.edu, www.ceutec.unitec.edu etc. Also, a domain with a .hn ccTLD may be registered in Honduras, but physically hosted somewhere else and thus not have a Honduran IP address. This is a general problem in trying to assess content location by infrastructure hints, which furthermore can be rather imprecise [16], [17].

A related problem, and potentially the largest one, is that much information about honduras is hosted on pages outside a .hn ccTLD. For example, unitec.edu has a US education gTLD. However, it is physically hosted in Honduras. In this case, only a content analysis of the Web pages could provide the hint that this institution is actually in Honduras and should be included in a search. A further problem with using only Honduras Web pages would be that many international pages would be left out even if they often have rather good coverage, especially of tourist destinations and hotels, or general pages dealing with the area. Additionally, much information, especially of a tourism nature, is more comprehensively available in English than in Spanish. While the official language is Spanish, English as a recognized regional language is spoken on the Bay Islands in the Caribbean, a major tourist destination. This is normally not a problem, but in the case of Honduras, these often contain the only mentions of certain places.

A further source of initial seeds for a search engine is the DMOZ Open Directory. However, it has only little coverage for Honduras. Yet, it confirms the previous statements. 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. As a first rough estimate on the sites that are available, the Spanish DMOZ links are correlated with the 5780 domains available for .hn. This gives a number from a conservative 6200 to a probably heavily overestimated 225,000 domains, with its geometric mean below 40,000 domains. Finally, general search engines such as Google, Yahoo, or Bing can be used to retrieve individual pages, but these limit the amount of results they give, so that also only gives limited insight. A search for only unique domain names is not possible. Result estimates of 60 to 500 million pages are therefore an overestimation, but hint at the potential pages about the country. Overall, this means that there are quite a few issues that prevent a system from getting all domain names for Honduras. An educated guess would initially put the number of potential Web servers at about 40,000 to 100,000, a number well manageable by modern Web crawlers [18], [19]. The challenge lies rather in the discovery and selection of relevant domains and documents by an adapted crawling strategy [20].

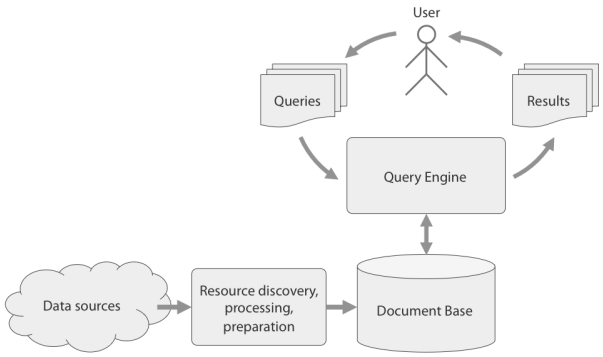

Fig. 3. Information Retrieval model (extended from [25])

The steps described up to this point are basically the same as for building a general Web search engine on a country level. A similar project has been described for Chile [21]. Therefore, the work done in the project might even be the basis for a general search engine for the Honduran Web if those pages dealing with Honduras are identified, without entity extraction and validation.

C. Coarse granularity location references

Additionally, examples of addresses and addressing schemes are being collected as they appear on the Web to gather training data for subsequent entity-oriented location extractors. A strong issue here is that the addressing system is very informal as already discussed in Section II-D. The format with a postal code described by the UPO is virtually unheard of [22]. Addresses may be rather incomplete and only mention the district or nearby landmarks, i.e., “Zona Jacaleapa, frente a Colonia Honduras, Tegucigalpa”, “final de Bulevar Morazán”, “3a Calle,Tegucigalpa Honduras”. This poses large issues to address recognition, toponym resolution, geoparsing, and grounding [23]. Overall, the granularity of location references is rather low and other methods have to be employed to extract and geocode location with high precision since previous methods [24] are of only limited applicability.

IV. Approach

The previous sections have shown that there are available data sources for Honduras on the Web. However, structured data sources are very sparse and Honduras Web pages are rather scattered, with many outside the .hn domains. This section describes the general approach and methodology towards a Honduran local search engine. The two main data sources will be the freely available Web pages as described before, the other the structured data sources. Additionally, future applications might also invite user-supplied data in a crowdsourcing approach. This might allow to create a more dynamic system that might be able to provide such data that is not yet available in other sources and may be used for corrections.

The general process of an information retrieval system is shown in Fig. 3. It describes how data sources are gathered

and processed into a document index so that users can make queries to the search engine. The proposed local search solution is based on this, and employs a specific architecture described later in Fig. 4. The first part concerns a Web search engine to gather Web pages concerning the country. It will comprise components of crawling, indexing, and searching of Honduran Web pages and specific information extraction suited to the desired type of search, interesting search queries, and Honduran characteristics. As discussed above, Honduran Web pages are rather scattered. The requirement is that they all should be found, but without having to crawl large part of the rest of the Web. For this, the technique of focused crawling [20] will be employed. Since geographic aspects are not well catered for in purely textual search, adaptations are needed throughout the search engine, as discussed in the following.

The extraction of location references by a geoparser [23], [7] not only assesses keywords occurring in a document, but aims to assign semantics to location references. It is aided by a gazetteer, a geographic thesaurus of known place names and their relation, to help in identification and disambiguation of place names. The most comprehensive gazetteer was found to be geonames. Due to the issues of low granularity (Section III-C), the geoparser will not only aim at extracting textual location references, but also at exploiting non-address locations. To overcome the low granularity, pages often make use of coordinates, directions or navigation hints, croquis (self-drawn map sketches), or embedded maps. To some extent, these can be extracted automatically.

A pure web search engine would miss some of information that is hidden away in the mentioned specialized databases, the so-called Deep Web. While the content of these databases is available on the Web for a human user, it is difficult to retrieve by a common Web crawler. However, some of the sources are accessible to automatic approaches connectors to their APIs, others might be crawled by the Web crawler and treated as unstructured data or be analyzed by specific wrappers. Initial extraction implementations are already being developed for selected sources to be part of an overall location search architecture to be adapted to this purpose. Some specific search engines such as aggregators use a similar approach but additionally exploit specific structures on domains to retrieve associated metadata or pages as structured content [26]. By this, disambiguities in extraction are kept to a minimum and the data needs only to be normalized for use within the system.

Naturally, the Web crawler and the connectors will gather a lot of duplicate entity information. This mandates systems for reconciliation, merging, and fusion of entity data. As a potential correction method and an initial proposal towards a solution, a new intelligent data merging system is proposed. It will take in both types of data and thus, Web results might get validated and grounded by the structured data sources. The features on which to match include location, names, phone codes etc. The data fusion would allow a geospatial search of the Honduran Web by fusing these different data sources by mostly automated processes and gather much of the data that is available. Since a lot of entities will not be described the same

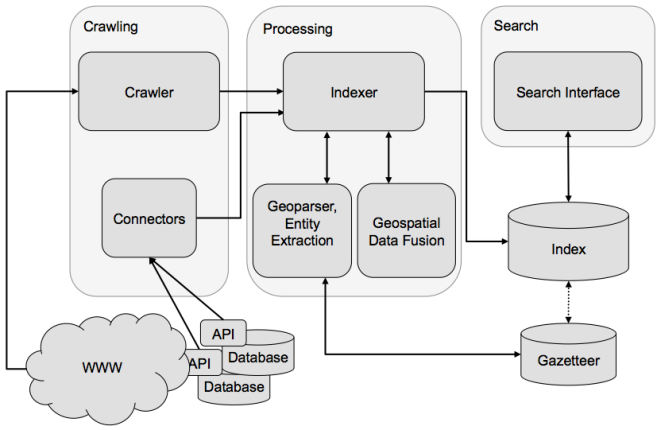

Fig. 4. Architecture of the geospatial search engine (adapted from [1])

way, there is a need for non-exact matching. Furthermore, the addressing system allows for varying location descriptions for one and the same place, which is another challenge in the development of the data merging. However, the location can also aid in merging named entities in various ways. By better understanding location references, even differently described places could be identified. In most cases, what the geocoder or the wrappers extract are actually only textual references, not yet exact coordinates that could be drawn onto a map. To make this step, a geocoder is used. There exist some geocoding services and the geonames gazetteer, but they again suffer from the low granularity issue, which will sometimes only place things at the level of a city district. The issue of inexact coordinates can further be improved by using coordinate-providing sources. In some cases, addresses are already described by embedded maps, which allow to extract exact coordinates. For other cases, other sources can be put to use not just as a data source, but also as a gazetteer and geocoder. An investigation into OpenStreetMap shows that there exists a lot of data for Tegucigalpa and Honduras. Especially the data that lies behind the map contains much more information, often only as a point with a name, but with a well-defined geographic coordinate. The idea then is to use OpenStreetMap and similar sources for grounding and geocoding purposes. For found names in other sources, a similarity measure can identify matches in the database and then the coordinates can be extracted to be used for the grounding of entities.

With the main components defined, the architecture of the proposed search engine is laid out in Fig. 4. Based on the IR process of Fig. 3, it takes the general structure of a Web search engine, specifically adapted to geospatial search [1]. The main conceptual difference is that the former graphics shows the process as seen from the user interaction, while the latter focuses on the technical aspects and the data preparation. For this project, it furthermore includes aggregator components for connecting to structured data sources of the Deep Web which are used to gather initial knowledge about the country. It features the main components of focused Web crawler, connectors, geoparser and named entity extraction, geocoder

and geospatial data fusion; an index and a search interface round of the architecture.

V. Conclusion

The first analysis has shown that there is sufficient data available to start building a Honduras geospatial search engine. Initial probings and implementations have been successful. The work that has been done will be continued towards better in-detail analysis of the data sources and their overlap. Extraction modules will be implemented that access the data sources directly. A Web search engine will be put up with an initial seed set of identified Honduran domains and the identification of domains outside the .hn namespace will be addressed. Further research will go into the geoparsing, especially of descriptions and directions. This can then also be used to give out location references to users for navigation. A data fusion method to combine data from the different sources will be developed. New methodology will be developed to combine text- and entity-based merging methods with geographic conflation techniques to reliable merge vaguely positioned geospatial entities. Furthermore, different interfaces towards this data will have to be developed. This will especially be influenced by the product ideas and research impulses given by the industry partner. The article does not deal yet with questions of limited or intermittent bandwidth [27] or limited availability of computers [28]. However, mobile low-bandwidth or SMS-only solutions [29] are considered as interfaces towards the search engine.

Overall, Honduras provides an ideal ground for generating a research prototype 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 the research prototype will be able to cover a huge fraction of the Honduran Web, thus building up a comprehensive index. This will also facilitate an easy movement into a public service, thus improving the local search engine coverage.

Acknowledgment

We thank out students Isaac Martinez and Jose Matute who work within the research project and provided initial implementations and data analysis.

References

[1] D. Ahlers and S. Boll, "Location-based Web search," in *The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society*, A. Scharl and K. Tochterman, Eds., 2007.

[2] R. S. Purves, P. Clough, C. B. Jones, A. Arampatzis, B. Bucher, D. Finch, G. Fu, H. Joho, A. K. Syed, S. Vaid, and B. Yang, "The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the internet," *International Journal of Geographical Information Science*, vol. 21, no. 7, pp. 717–745, 2007.

[3] A. Markowetz, Y.-Y. Chen, T. Suel, X. Long, and B. Seeger, "Design and Implementation of a Geographic Search Engine," in *WebDB 2005*, 2005, pp. 19–24.

[4] K. A. V. Borges, A. H. F. Laender, C. B. Medeiros, A. S. D. Silva, and J. Clodoveu A. Davis, "The Web as a Data Source for Spatial Databases," in *Anais do V Brazilian Symposium on Geoinformatics*, 2003.

[5] D. Ahlers, "Local Web Search Examined," in *Web Search Engine Research*, D. Lewandowski, Ed. Emerald, 2012, to appear.

[6] R. Steele, "Techniques for Specialized Search Engines," in *Proceedings of Internet Computing 2001*. Las Vegas: CSREA Press, 2001.

[7] D. Ahlers and S. Boll, "Retrieving Address-based Locations from the Web," in *GIR '08: Proceedings of the 5th International Workshop on Geographic Information Retrieval*. New York, NY, USA: ACM, 2008.

[8] K. A. V. Borges, A. H. F. Laender, C. B. Medeiros, and C. A. Davis, Jr., "Discovering Geographic Locations in Web Pages Using Urban Addresses," in *GIR'07: Proceedings of the 4th ACM Workshop on Geographical Information Retrieval*, ACM, 2007, pp. 31–36.

[9] W. Cai, S. Wang, and Q. Jiang, "Address Extraction: Extraction of Location-Based Information from the Web," in *APWeb 2005*, ser. Lecture Notes in Computer Science, vol. 3399. Springer, 2005, pp. 925–937.

[10] S. E. Overell and S. M. Rüger, "Identifying and grounding descriptions of places," in *Proceedings of the 3rd ACM Workshop on Geographic Information Retrieval, GIR 2006*, 2006.

[11] L. A. Souza, C. A. Davis, Jr., K. A. V. Borges, T. M. Delboni, and A. H. F. Laender, "The Role of Gazetteers in Geographic Knowledge Discovery on the Web," in *LA-WEB '05: Proceedings of the Third Latin American Web Congress*. IEEE Computer Society, 2005, pp. 157–165.

[12] H. Hile, R. Grzeszczuk, A. Liu, R. Vedantham, J. Košecka, and G. Borriello, "Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning," in *Proceedings of the 7th International Conference on Pervasive Computing*, ser. Pervasive '09. Springer, 2009, pp. 59–76.

[13] D. Caduff and S. Timpf, "The Landmark Spider – Representing Landmark Knowledge for Wayfinding Tasks," in *AAAI 2005 Spring Symposium*. AAAI Press, 2005.

[14] D. Mundluru and X. Xia, "Experiences in Crawling Deep Web in the Context of Local Search," in *GIR '08: Proceedings of the 5th International Workshop on Geographic Information Retrieval*. New York, NY, USA: ACM, 2008, pp. 35–42.

[15] B. He, M. Patel, Z. Zhang, and K. C.-C. Chang, "Accessing the Deep Web," *Communications of the ACM*, vol. 50, no. 5, pp. 94–101, 2007.

[16] M. Freedman, M. Vutukuru, N. Feamster, and H. Balakrishnan, "Geographic Locality of IP Prefixes," in *Internet Measurement Conference (IMC) 2005*, Berkeley, CA, October 2005.

[17] K. S. McCurley, "Geospatial Mapping and Navigation of the Web," in *WWW '01: Proceedings of the 10th international conference on World Wide Web*. New York, NY, USA: ACM Press, 2001, pp. 221–229.

[18] B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, 1st ed. Addison Wesley, 2009.

[19] M. Najork and A. Heydon, "High-performance web crawling," in *Handbook of massive data sets*. Kluwer, 2002, pp. 25–45.

[20] D. Ahlers and S. Boll, "Adaptive Geospatially Focused Crawling," in *CIKM '09*. ACM, 2009, pp. 445–454.

[21] M. Mendoza, H. Guerrero, and J. Farias, "Inquiro.CL: a New Search Engine in Chile," in *WWW '09: 18th International World Wide Web Conference (WWW in Ibero-America track)*. ACM, 2009.

[22] Universal Postal Union, "Postal addressing systems in member countries – Honduras," Universal Postal Union, Tech. Rep., 2004.

[23] J. Leidner, *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. Universal Publishers, 2008, PhD thesis.

[24] D. Ahlers and S. Boll, "On the Accuracy of Online Geocoders," in *Geoinformatik 2009, Osnabrück*, ser. ifgiprints, W. Reinhardt, A. Krüger, and M. Ehlers, Eds., vol. 35, Münster, 2009, pp. 85–91.

[25] C. J. van Rijsbergen, *Information Retrieval*, 2nd ed. London: Butterworths, 1979.

[26] A. Arasu and H. Garcia-Molina, "Extracting Structured Data From Web Pages," in *SIGMOD '03*. New York, NY, USA: ACM, 2003, pp. 337–348.

[27] J. Chen, L. Subramanian, and J. Li, "RuralCafe: Web Search in the Rural Developing World," in *18th International World Wide Web Conference*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 411–420.

[28] S. Amershi and M. R. Morris, "Cosearch: a system for co-located collaborative web search," in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2008, pp. 1647–1656.

[29] J. Chen, L. Subramanian, and E. Brewer, "SMS-Based Web Search for Low-end Mobile Devices," in *MobiCom '10*. New York, NY, USA: ACM, 2010, pp. 125–136.