The Information Commons Gazetteer

A Public Resource of Populated Places and Worldwide Administrative Divisions

Peter Lucas, Magesh Balasubramanya, Dominic Widdows and Michael Higgins

MAYA Design, Inc.
Pittsburgh, Pennsylvania, 15203 USA
{lucas,magesh,widows,higgins}@maya.com

Abstract

Advances in location aware computing and the convergence of geographic and textual information systems will require a comprehensive, extensible, information rich framework called the Information Commons Gazetteer, that can be freely disseminated to small devices in a modular fashion. This paper describes the infrastructure and datasets used to create such a resource. The Gazetteer makes use of MAYA Design's Universal Database Architecture; a peer-to-peer system based upon bundles of attribute-value pairs with universally unique identity, and sophisticated indexing and data fusion tools. The Gazetteer primarily constitutes publicly available geographic information from various agencies that is organized into a well-defined hierarchy of worldwide administrative divisions and populated places. The data from various sources are imported into the Commons incrementally and are fused with existing data in an iterative process allowing for rich information to evolve over time. Such a flexible and distributed public resource of the geographic places and place names allows for both researchers and practitioners to realize location aware computing in an efficient and useful way in the near future by eliminating redundant time consuming fusion of disparate sources.

1. Introduction

Recent advances in location-aware computing and search technology are enabling a synthesis of traditional geospatial technology (such as Global Positioning Systems) and human language technology (such as Named Entity Recognition). A rich GPS system needs to give not only the latitude and longitude of a device, but also should be able to state the geopolitical region in which the device is located and to list nearby populated places and features. A rich NER system needs to recognize not only that a given noun phrase refers to a populated place, but also to state which populated place and where it is located. These developments have led to a wider appreciation of the need for rich geospatial resources within both the GIS and the language technology communities.

This paper describes an initiative by MAYA Design, a design and technology research laboratory in Pittsburgh, Pennsylvania, to collect and publish a universally available resource of geographical information to fill this need. This resource is part of the Information Commons, a collaborative effort to gather and fuse publicly available information into a global semantic network. The traditional name for a list of place names is a Gazetteer, and this term is gradually becoming applied to much richer resources that include spatial coordinates, demographic and political information. For this reason, the resource described in this paper is called the Information Commons Gazetteer.

The presentation will proceed as follows. First, we describe the engineering and social requirements on such a system. Second, we describe the framework of tools and resources that were made use of to create such a system. Thirdly, we describe the source datasets used, the process by which the Information Commons Gazetteer was assembled from these resources, and how it is published, indexed, and disseminated.

2. Semantic, Social, and Engineering Requirements

Before embarking upon this project, it was determined that the resource and the platform created should be:

1. Comprehensive. The information provided should be as complete as possible, to enable researchers and developers to accept such a resource as a de facto standard.
2. Free to Use. This is vital not only for idealistic reasons, but also for practical ones — however comprehensive a resource, if it is not free, it will not become standard.
3. Information Rich. Some researchers are interested in demographic information, some in political information, others in business and economic information.
4. Extensible. To be information rich without being overloaded, it must be possible for researchers and developers to add information to the resource, if not directly then by reference.
5. Distributed. A comprehensive and information rich resource will gradually become larger and larger, yet few researchers (and no small devices at all) will want to use the entire resource at any one time. The architecture must therefore be modular, to enable different users to select and use the parts that are relevant to their application.

Our initial research determined that no single resource satisfied all of these goals. The Getty Thesaurus [2], though comprehensive, is not free and so has not enjoyed wide use. Parts of the Alexandria Gazetteer [3] are available for free download, but the information freely disseminated is strictly limited (for example, latitudes and longitudes are not given in the basic version) which is not suitable for many purposes. However, the Alexandria Gazetteer is compiled from many of the same sources as the Information Commons Gazetteer, and it is possible that with some reengineering it could be united with the Information Commons Gazetteer.

3. Design Framework

This section provides an overview of the process involved in the making of the framework for the Information Commons Gazetteer and the resources used for that end.

3.1. Universal Database Architecture

The platform chosen for representing the Information Commons Gazetteer was the Universal Database architecture, an extensible peer-to-peer architecture developed over several years by MAYA Design and its collaborators (see [1]). In this architecture, all information is represented in *u-forms*, a u-form being an extensible bundle of attributes and values indexed by a unique identifier. This architecture enables information to be replicated in different venues, without losing its identity. This modularity enables small mobile devices to use parts of the information space, even when disconnected from the Internet. This functionality was deemed vital for our purposes, which is one of the reasons why a Web Services architecture such as the Semantic Web, though similar in some respects, had to be rejected. (A Universal Database peer can easily be used to provide Web Services where appropriate — we merely believe that this should be a function, not a limit, of the system).

3.2. Data Import and Fusion

The data import process initially creates "shadow" u-forms, which are representative of the source data as obtained from the source. In other words, these shadows act as proxies to the source data in the Universal Database architecture. The attribute value pairs of the shadow change as and when the source data changes on subsequent re-import of the source.

The uniquely identified definitive place u-forms are either looked up using names and geographic information if they already exist or created with new unique identifiers by copying the relevant information in the appropriate format from the shadow sources. This process also involves fusing existing places or sub-divisions using index lookup, updating those indices and / or creating new indices.

Various attribution information is added to the definitive place u-form. The 'source' attribute provides a reference back to all shadows and other u-forms that contributed to the content in the definitive u-form. This also provides a mechanism for information update as and when the source data changes. The 'creator' attribute attributes the content to the respective organizations or agencies that provided the source data. The 'language' attribute defines the language in which the information in the u-form is published. Other attribution such as the publisher, latest date of publication and publishing rights etc., enhances the quality and usefulness of the meta-data of the definitive place u-forms.

3.3. Indexing Tools and Indices

With millions of geographic places and place names in the Gazetteer, it is imperative that any given place be found in an efficient and timely fashion. Indices enable searching for places in the Gazetteer both for general purposes and also for fusion attempts. Since the indices co-exist with the data on the same Universal database platform, the index structures too are designed to be distributed and extensible. Just as the place data is available freely in a distributed peer-to-peer architecture, the indices are accessible similarly without the need for conventional centralized index database servers. This fluid nature of the indices is essential, as it allows for instant lookups and index updates by the data import tools without any centralized monitoring and maintenance of large databases. Furthermore, the distributed nature of the indices and the places allows for fusion to be achieved in parallel and in multiple venues. For the Information Commons Gazetteer, two kinds of general-purpose distributed indices are employed to achieve fusion and to consistently search for places and place names.

1) Name index: A distributed B+ tree index [4] with all the nodes as persistent u-forms serves as the place name index, offering efficient lookup characteristics. The keys of the name index are the place names and alternate names in Unicode, which supports multilingual search. The keys are tokenized and canonicalized on white space characters and using lower case. The index, upon searching for a given key, returns a list of unique identifiers of all the places whose name or alternate name matches the query name either in part or completely. The index supports prefix name matching, thereby providing the option to look for a wide range of values and subsequently use other means to narrow down the target while attempting fusion.

2) Spatial index: An R tree [5] index structure is an efficient data structure for indexing n-dimensional data and is especially well suited for geographic data of two dimensions (latitude and longitude). The spatial index used in the Information Commons Gazetteer is a distributed equivalent of the R tree data structure indexed on latitude and longitude. Apart from the spatial distribution of the places, for each place the index stores a normalized importance score, which is a function of the population of the place. The index can be used in two modes. In the *simple* mode, it behaves as a regular R tree index i.e., for a query such as " get all places within this co-ordinate rectangle [[lat1, long1], [lat2, long2]]", the index returns all the unique identifiers of the places whose geographic co-ordinates fall within the query rectangle. Whereas in the *priority* mode, which makes use of the importance score of the places, the index ranks the results based on the population and queries can be limited to a certain number of results. E.g., the query would look like "get me the 100 most important places (based on population) that fall within this co-ordinate rectangle [[lat1, long1], [lat2,long2]]". Such a ranking is of great importance in the fusion process, so we can eliminate unlikely targets when the source data is based on population. Other variations of the above basic queries are used to generate richer queries such as "get all places within a 10 km radius from a given place [lat,long]"

Other specific indices based on numerous codes and ids from various data sources were also created to facilitate the fusion attempts. Such indices are again implementations of the earlier mentioned B+ tree data structure. E.g., an ISO 3166-1A2 code index, a FIPS 5-2 code index, a FIPS 55 index etc.

An hierarchical tree based data structure that makes use of the semantic information of the places such as the sub-division hierarchy is the primary data structure that is used to maintain most shadow sources and also to maintain the main Information Commons Gazetteer Populated Places Index. Such a hierarchical structure allows for easy iteration and quick access of specific selected data.

3.4. Inference and Disambiguation Tools

The results of a name lookup or a spatial lookup may span a large number of places depending on the query. During the fusion process this subset of matching results needs to be further narrowed down to a smaller set and ultimately to one single place to fuse with the target. There is no single approach that works for disambiguation of multiple matches, as it largely depends on the nature of the data. The Information Commons Gazetteer employs multiple schemes and tools to decide on the single match whenever possible.

The inference techniques rely on extra information in the source that is related to the target place names in direct or indirect ways. E.g., when trying to fuse a source datum having the name 'pittsburgh' with the Gazetteer places, a simple name match results in 23 results. If the source just had the name alone, it would be hard to find the right match from the multiple targets. But if, say, the source also had extra information such as the level 3 division it belongs to, then we can filter the name match results to see which of them have the same level 3 division. Other inference parameters that were used include the sub-division type and the municipality type. The level of disambiguation and its success largely depends on the quality and quantity of data in the query source that could be used for such inference.

When no such inference could be made or if, even after using them, multiple matches exist (of a manageable number) then we resort to manual inspection methods. The tools for manual disambiguation were developed and employed to select the best option from the multiple matching results of the index lookups and other automatic disambiguation techniques. These visual tools are configurable to show the relevant and rich details of the matching results as compared with the query source to allow a human to attempt an educated guess with very high confidence. If very high confidence cannot be reached even after human inspection, no fusion is attempted and instead the probable results are marked for future attempts if new data added from other imports enriches the Gazetteer in this context. One of the tools allows for soundex matching of the names of the source and the targets, ranking the results with normalized confidence values.

These soundex and phonetics based disambiguation techniques are also employed when a regular name-matching attempt produced zero results. It is very common for data from various sources to be of bad quality (such as mis-spelt names and diacritical errors during transliteration from other languages). In such cases, prefix search is used to provide a wide range of results that are then narrowed down to a smaller subset using inference and spatial matching. Then the soundex ratings are used to find a good match. The Civium Workbench Framework [6] provides a rich set of visual tools and interfaces to complement manual fusion techniques. For example, the results of a spatial matching can be viewed in the map view and the place name that matches the source and which is geographically closest can be chosen with little or no effort at all. Also, all the information of both the source and probable targets are just one click away to further confirm the choice.

4. Information Commons Gazetteer Implementation

The Information Commons Gazetteer was built upon the platform described in the previous section by fusing data from the following resources:

1. The National Geo-spatial Agency's GEOnet Names Server (GNS) database [7] provided the place names (with lat/long coordinates) for all non-U.S. places. This resource contains many name variants in over 20 languages, which were fused as alternative names in a single u-form representing the place itself.

2. The Unites States Geological Survey's Geographic Names Information Systems (GNIS) [8] provided the place names (with lat/long) of geographic features in the United States and its territories.

The result of fusing these datasets is a combined dataset of over 5.5 million populated places worldwide, with names, latitudes, and longitudes. This forms one of the main building blocks of the Gazetteer.

Rich information population data was fused into the Gazetteer where possible by identifying matching names and checking latitude and longitude coordinates to see if a match was plausible. The auxiliary resources that were used to enrich the Gazetteer places are explained later in the paper. Other regional resources will be sought out and used where possible to create richer population information where possible, which is one of our reasons for wanting to make contact with more European researchers and institutes.

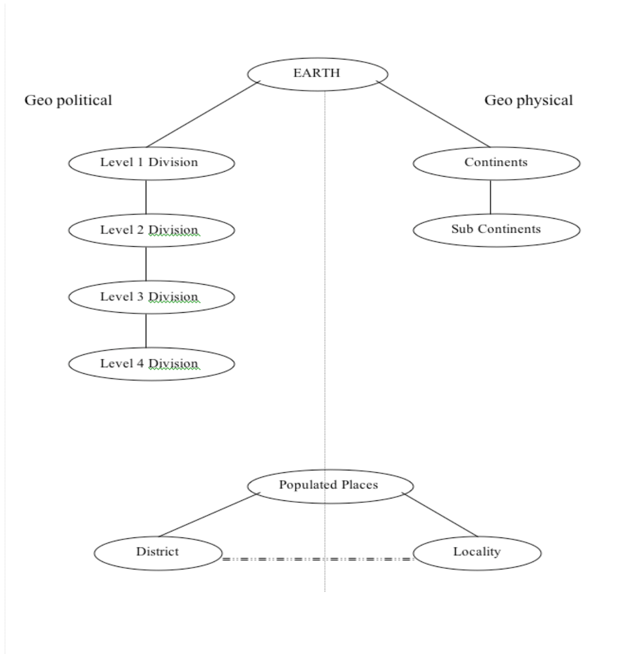

Fig. 1

In the Information Commons Gazetteer, any given place can exist as a geophysical entity and / or a geo political entity. Fig. 1, in the first part describes this dichotomy starting from Earth. The Geo political divisions correspond to the political hierarchy where level 1 corresponds to countries, level 2 to states and their equivalents and so on. This hierarchy defines a strict parent member relationship where the sum of all members in a given level would add up to the previous level without overlap. The second part of the figure illustrates that the Populated places too exist as either political 'Districts' or

Physical 'Localities' and they are cross-referenced. It is to be noted that these cross-references need not be one-to-one though in most cases they tend to be.

4.1. Political Sub-division Hierarchy

The level 1 divisions (countries) of the world and level 2 divisions of the countries were created from the ISO 3166 -1A2 and ISO 3166-2 code lists [9]. This provided basic information such as names and alternate names of the subdivisions in multiple languages and the ISO codes which was used to create an ISO code lookup index. The CIA world fact book [10] provided further details such as latitude, longitude, population data, flag images, iso_alpha_3 code and IANA code which were fused with the already created level 1 divisions. The NGA's GNS data of feature type 'A' designated as 'PCLI' provided additional alternate names for the level 1 divisions. Also, the GNS data (except for U.S. and its territories which used USGS GNIS data) designated as 'ADM1' and 'ADM2' provided the place name, latitude, longitude, alternate names for the level 2 subdivision and level 3 subdivisions respectively. The level subdivisions were referenced back to the corresponding parent and child divisions of the level hierarchy as applicable by making use of the administrative codes and ISO codes in the GNS source data, thereby building the political hierarchy structure. For the U.S., the FIPS PUB 5-2 [11] codes uniquely identify the level 2 and level 3 divisions which are part of the GNIS resource from which the corresponding U.S. subdivisions were created. A FIPS code index was also created using these codes.

4.2. Populated Places

4.2.1. World Populated Places (Except U.S.)

The place names, alternate names and geo locations of world places, except the U.S. and its territories, are derived primarily from NGA's GEOnet Names Server [6] data of feature class 'P', out of which those designated as PPLA or PPLC are treated as geo political entities and the rest are considered as geo physical entities. An attempt to fuse the geo physical and geo political entities representing the same place is undertaken as part of the import process by using name lookups and spatial localization. For a geo political entity derived from GNS (designated as PPLA or PPLC), if an equivalent geo physical entity is not found, then a derived geo physical entity for the place is created using the data from the geopolitical entity. This is not true for the converse, since it can be inferred that a political entity presupposes the existence of a physical entity whereas the converse isn't always true. The level 1 division (country) for these places was fused using the ISO 3166-1A2-code lookup index. The Populated Places hierarchical index based on the sub division levels and first two characters of the place name served as a semantic index for the Gazetteer places with around 2 million places in its multi level hierarchy. The previously described place name index with approximately 5.5 million place names and alternate names were added to the Information Commons Gazetteer during this import along with an R tree based spatial index that allowed for spatial queries on the 2 million places. The R tree also enables importance based ranking of the results. Fig. 2 is the pictorial representation of the Gazetteer spatial index of all Gazetteer places where each rectangle in the figure contains 20 populated places. Rectangles form the leaf nodes of the spatial index tree where the populated places are grouped together based on the minimal bounding box of those place entries. The darker regions in the figure indicate a greater density of rectangles of smaller area that is proportional to the density of the number of populated places in that region. The sparseness of human incorporation of places is clearly visible through the spatial index in the cases where a leaf node rectangle extends over a large area as seen in the middle of Australia, northern Canada and central Asia.

Fig. 2

4.2.2. U.S. Populated Places

The USGS GNIS data of feature type 'populated place' is the primary resource for approximately 200,000 geophysical entities in the US and its territories, providing the place name and name variants, latitude, longitude, elevation and population data. Looking up the FIPS code from the GNIS source in the FIPS index that was created earlier fused level 1, 2 and 3 divisions of these places. The population size was used to characterize the place in terms of its population, thereby aiding in disambiguation with similarly named places of lesser importance (in terms of population). These places are indexed into the Populated places sub division level hierarchical index created during the GNS data import. The GNIS data of feature type 'civil' is the main source of geopolitical entities that are either geopolitical subdivisions, part of the political hierarchy, or are populated places of type 'district' grouped under different municipality types. Based on manual research from various sources, the places were grouped into different subdivision levels and municipality types by hierarchical inference and name matching of the level/ type qualifier available as part of the place name. Table 1 lists the type qualifiers associated with the sub division level for the geo political sub divisions and Table 2 lists the type qualifiers and municipality types for the Populated places of type 'district'.

Sub division Level	Type qualifier as part of the place name
2	state, commonwealth
3	county, borough, parish, municipio, census area
4	township

Table 1

Municipality Type	Type qualifier as part of the place name
Same as type qualifier	town, grant, city, borough, tract, election precinct, village, colony magisterial district, election district
None or No type	any other type

Table 2

As an example of hierarchical inference, the type 'borough' appears in both subdivision level (3) and also in municipality type. The disambiguation in this case, makes use of the fact that only in the state of Alaska are level 3 divisions known as 'boroughs' (and also none of Alaska's municipality types are called 'boroughs'). Using this inference, the data is grouped appropriately. Since the level 2 and level 3 divisions were already created as described in section 4.1, for data corresponding to these two levels, the FIPS code from the source is used to lookup the existing unique subdivision from the FIPS index and the source data (name variants, latitude, longitude, elevation, population etc) are fused. In case the lookup failed to return an existing subdivision, a new one is created using the GNIS data and fused with its parent subdivision(s), and the FIPS index is updated. In this way, the new imports add and enrich existing entities without creating duplicate entries thus providing for a truly unique identity space. The level 4 divisions are created from the GNIS data and fused with the existing subdivisions (level 1,2 and 3) from the FIPS index lookup. Also, the 'members' of the level 3 divisions are updated with the corresponding level 4 subdivision children. This builds the political hierarchy structure and associates the newly created subdivisions with their parent. This is also true with any new discovery of the level 2 and level 3 subdivisions. A similar procedure is carried out for the municipalities but they are added to the 'adm_regions_misc' of the level 3 division only if they have a municipality type. All newly created Populated places are added to the Populated places hierarchical index and the name and spatial indices are updated accordingly.

4.2.3. Political and Physical Place Fusion

Whenever a new geopolitical entity of type district is created, an attempt is made to fuse it with a corresponding geophysical locality. This is primarily done using name matching and by scoping the search over a specific geopolitical/ geographical region by hierarchical inference from the subdivision level information. For example, when trying to fuse the political entity 'Pittsburgh, City of' with the geophysical 'Pittsburgh', the search is scoped to look for names and alternate names that match 'Pittsburgh' at the level 3 division corresponding to 'Pittsburgh, City of' (as they both have the same level 3 subdivision) in the Populated Places hierarchical index. On a unique match, references are added to either of the entities about the other. In the case of no results on name matching, a new geophysical location is created borrowing the values from the geopolitical entity for latitude, longitude, name and population. For the case of multiple name matches, further disambiguation techniques such as spatial matching, feature_type differentiation (e.g., a subdivision and a district) and manual inspection are employed to resolve the tie and a match is chosen appropriately or a new geophysical place created if none of the partial matches seems to be the right fit. In all the above cases, the Populated places hierarchical index, the name index and the spatial index are updated with all the changes and additions of both the newly created geophysical and geopolitical entities. Fig. 3 shows the u-form of 'Pittsburgh, City of' displayed as a list of attribute - value pairs as seen in the Civium Workbench application [6]

Fig. 3

4.3. Fusion of Other Data Sources for World Places

4.3.1. World Gazetteer Population Data

Around 80,000 entries from the World Gazetteer [12] website provided current population data, past historical data, and alternate names in multiple languages for populated places around the world. Some of them had location codes (latitude and longitude) and some others also had the subdivision level names along with ISO country codes. The fusion of these data with the Information Commons Gazetteer involved name lookup and spatial matching (when location codes were available). When a unique match was found, the World Gazetteer data were fused with the Commons Gazetteer place. In case of multiple matches, the best match was chosen upon manual inspection of the options after applying spatial and other filtering criteria. When no match was found, such entries were marked for future lookups and matching.

4.3.2. VMAP Boundary Shape Data

Around 36000 built-up area boundary shapes of world places obtained from NGA VMap0 [13] vector data were treated as the boundaries of the geophysical entities as they are seen from air or space. Here again, both name matching and spatial inference were employed to find the correct match. Also, in this case, where there were multiple matches for a given name within a spatial radius, then the place with a higher population was chosen through indirect inference that the built-up area is related to the population by virtue of its spread across the area in most cases and the implied fact that this dataset contained built-up areas of the most populous places. Therefore, we could make use of the population based importance ranking of the priority mode spatial queries to find a high probability match of the target.

4.4. Fusion of Other Data Sources for U.S. Places

4.4.1. FIPS 55 Code Fusion

FIPS55 [14] is the Federal Information Processing Standard (FIPS), which contains codes for named populated places, primary county divisions, and other location entities of the United States and areas under the jurisdiction of the United States. Since the maintenance of the FIPS 55 standard has been changed to the USGS, they have added their unique GNIS_feature ID to all the FIPS place codes. Therefore, the fusion with this data was straightforward and required only a lookup tool between the place code and the GNIS_feature ID, which already exists in the USGS GNIS source shadows. Furthermore, as the FIPS 55 5-digit place codes are only unique within a given state (level 2 division in the U.S.), this has been fused with the 2-digit FIPS PUB 5-2 [11] numeric state code to form a 7-digit countrywide unique identifier as part of the fusion effort. This FIPS 55 fusion renders future fusion efforts more structured, especially with respect to data referring the FIPS 55 codes as in the case of the Census 2000 Places Gazetteer import described later in this section.

4.4.2. NACO fusion

The National Association of Counties [15] in their website provides the details for the level 3 divisions of the U.S. (counties and their equivalents), such as area, capital, year established, and subdivision type. These were extracted from the web and fused with the corresponding sub division u-form by matching the FIPS code. The Year Established provides date markers to scope the start and end times of the entity for time based tracking and the capital place names were fused with the u-form for that place name using name matching and other spatial disambiguation methods discussed above. Also, the entity that is the capital of the subdivision is updated appropriately to characterize its association as capital of the subdivision. A similar procedure added the temporal and capital information to the level 2 subdivision from data obtained from other sources that maintained the date of U.S. level 2 divisions

4.4.3. U.S. Census Bureau Cartographic Boundary Shape Fusion

The boundary shapes of Incorporated Places / Census Designated Places provided by the U.S. Census Bureau [16] is fused with the geopolitical entities by methods involving name matching and scoping the search to within a subdivision level (using FIPS lookup from the source data) and by use of spatial disambiguation such as spatial lookup of all places within a radius of the target location.

4.4.4. U.S. Census Bureau Gazetteer of Places Fusion

The Census 2000 gazetteer of places published by the U.S. Census Bureau has population data, land area and hydro area data for around 25,000 places that was fused with the Information Commons gazetteer data. This fusion attempt was straightforward by making use of the fused FIPS 55 codes from the previous fusion process, thereby demonstrating the ability of the Information Commons Gazetteer to evolve incrementally by making use of a truly unique identity space.

5. Conclusion

This information rich Gazetteer is an essential component of the larger Information Commons agenda of creating a Universal Database of feature rich inter-linked information. Future imports of data could make use of the Gazetteer to provide more content about the places. For example, we have already fused all the public and private schools in the U.S. and its territories (more than 100,000) to their respective level 2 and level 3 subdivisions This will further be extended to Postal ZIP codes, again making use of the fused data from the FIPS 55 fusion process.

6. References

[1] Peter Lucas, Jeff Senn, and Dominic Widdows. Distributed knowledge representation using universal identity and replication. Technical Report MAYA-05007, MAYA Design, 2005.

[2] The Getty Institute. The Getty Thesaurus. http://www.getty.edu/research/conducting_research/vocabularies/tgn/.

[3] Alexandria DigitalLibrary Project. Alexandria gazetteer. http://www.alexandria.ucsb.edu/gazetteer/.

[4] Donald Knuth. The Art of Computer Programming, volume III. Addison-Wesley, 1973.

[5] Antonin Guttman. 1984. R-trees: a dynamic index structure for spatial searching. In *Proceedings of SIGMOD*, pages 45–47.

[6] MAYA Design, Inc. Civium Workbench. http://civium.maya.com/.

[7] National Geo-spatial Agency. GEOnet Names Server (GNS). http://earth-info.nga.mil/gns/html/.

[8] Unites States Geological Survey, Geographic Names Information Systems (GNIS). http: //geonames.usgs.gov/index.html.

[9] ISO 3166 codes. http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html

[10] CIA world fact book. http://www.cia.gov/cia/publications/factbook/

[11] FIPS PUB 5-2. http://www.itl.nist.gov/fipspubs/fip5-2.htm

[12] World Gazetteer. https://www.world-gazetteer.com/.

[13] NGA VMAP0. http://geoengine.nima.mil/geospatial/SW_TOOLS/NIMAUSE/webinter/vmap0_legend.html

[14] FIPS 55 standard. http://geonames.usgs.gov/fips55.html.

[15] National Association of Counties. www.naco.org.

[16] U.S. Census Bureau. http://www.census.gov/geo/www/cob/pl2000.html.