

# Geographic Information Retrieval: An Overview

Øyvind Vestavik

Dept. Computer and Information Science  
Norwegian University of Technology and Science  
Trondheim  
Oyvind.Vestavik@idi.ntnu.no

## Abstract

*Geographic Information Retrieval can be seen as a specialized branch of traditional Information Retrieval. It includes all of the research areas that have traditionally made up the core of research into Information Retrieval, but in addition has an emphasis on spatial and geographic indexing and retrieval. This article present some of the challenges in Geographic Information Retrieval like the ambiguities involved in the use of place names and phrases linking a document to a location and the influence of human perception of and reasoning about space. It also explains the role of gazetteers, thesauri and ontologies of place names in Geographic Information Retrieval and reviews some relevant projects.*

## 1. Introduction

Geographic Information retrieval can be seen as a specialized branch of traditional Information Retrieval. It includes all of the research areas that have traditionally made up the core of research into Information Retrieval, but in addition has an emphasis on spatial and geographic indexing and retrieval [18]. Geographic Information Retrieval is not excluded to objects with a physical manifestation in geographic space and a geographic extent like for instance rivers, cities, lakes or countries but deals with any kind of information that has some sort of relation to one or more locations on the earth's surface. Information that has such relationships to geographic space is often called georeferenced information and therefore an frequently used term for the field is Georeferenced Information retrieval. Georeferenced Information is in all kinds of media, not only structured data like maps, land surveys, airborne and satellite images and tabulated observations. The focus of this paper will be on indexing and retrieval of text taking location into account.

Approaching information based on their georeferences can be important in several contexts. Researchers looking for information about the environmental changes in a certain area over time, information about the wildlife in a given

area or population growth in a certain city might want to limit their searches to geographic areas. Students searching for information on youth hostels in cities they plan to visit or on which attractions there are in a given area might also want to limit their searches based on geographic properties. The availability of Global Positioning Systems (GPS) in handheld devices like mobile phones, Personal Digital Assistants and even cameras has triggered a need for finding information about objects in the vicinity of the device. Targeted information can be delivered to such devices based on the location of the device. Although some research has already been done in geographic information retrieval, a lot stands to be achieved. Some of the challenges and possible solutions will be described in this article.

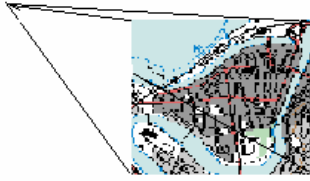
The rest of this article is structured as follows: First we examine some of the properties of georeferenced textual information. Next, we examine some of the properties of the cognitive handling of geographic space by humans and how this influences the creation of texts with georeferences. Then we examine the role of gazetteers, ontologies and Thesauri in georeferencing textual documents and how users can search for georeferenced information before reviewing some project relevant to geographic information retrieval.

## 2. Properties of georeferenced information

Much of the information available in digital libraries and on the Internet is georeferenced, although mostly it is not denoted in terms of geographic coordinates. Often, the link(s) between the information in a textual document and the location(s) it refers to manifests itself as place names and/or phrases that suggest a geographic location and. The geographic location and extension of a place name is often called a *geographic footprint* [14] and is given by coordinates (latitude, longitude). Geographic Information Retrieval requires that place names and phrases that include direct or indirect references to place names be resolved and translated into footprints that can be indexed. There are, however, some problems using place names and phrases that contain place names for assigning footprints to

documents that uniquely identifies a location for the document.

### Trondheim Sentrum



**Figure 1:** Geographic footprint of a place name

Place names occurring in texts can be misleading, ambiguous and/or imprecise. Firstly, several places can share the same name, making the place names unique only within a limited geographic area. There are for instance 26 places in Norway sharing the place name *Lade*. Secondly, some place names occurring in texts are temporal or cultural conventions rather than official names, requiring the reader to have an understanding of the time, context or cultural environment the place names are used in to be able to link it to some geographic location. For instance, during a military campaign commanders often map a battlefield into areas, naming each area according to some schema other than their official place names. Thirdly, some place names change over time. For instance, the Norwegian capitol Oslo was once named Christiania. Fourthly, the geographic extension that the place name denotes can be extended, reduced or changed over time. For instance, the geographic extension of countries may change during/after wars and the extent of a city may change according to population growth or decrease. Fifthly, the borders of a location can be fuzzy. For instance, Norway and Russia do not agree on the border between the two countries in the Barent Sea. And finally, some place names and spatial designators denote a state of mind or associations linked to an area rather than an actual defined location or area as for instance “southern California” or the Norwegian “syden”. In addition, the same place name can be written differently in different text, either because the author has misspelled the name or because there are different legal spellings of the same place name. Some of these spelling variations can be official variants of the same name within a language or variations between languages.

Phrases that reference locations relatively to a place name or pseudo place name can also be ambiguous or imprecise. For instance, the phrase “200 kilometers south of the capitol of Russia” certainly denotes a geographic location. However, appearing in a text, the distance 200 kilometers might be an approximation suitable only in the contexts it was used. In a small scale environment where the topic of the text is giving an overview over Russian cities the approximation might be feasible, whereas using this reference as a basis for a footprint in a larger scale environment might give an error of certainly up to 20-30 kilometers. Also, the direction given might be erroneous.

Often, people are imprecise in giving geographic direction, using one of the four general directions north, south, east or west, when the actual direction might be somewhere in between. Again, in small scale environments this imprecision might be not be important, where as in a larger scale environment the error in the footprint created on the basis of this phrase might lead to a loss of important information in a search context. The given phrase is also ambiguous because of the reference to the capitol of Russia. The capitol of Russia has been placed in todays Moskow *and* in todays St.Petersburg, so the footprint associated with “the capitol of Russia” is dependent of the time-frame of the text the phrase appears in.

Often, phrases that reference places do so indirectly without actually mentioning the place at all. Some of these phrases reference place names or geographical designators mentioned elsewhere in a text. An example could be “Lake Mjøsa is the biggest lake in Norway. About 200 km south of the lake...” The phrase “About 200 km south of the lake..” denotes a location, but would give no meaning without considering the context it occurs in because it references a place name found in the first sentence. In some texts the context is not given in the text itself but rather in the topic of the text. For instance in a short text about climbing Mount Everest, Mount Everest could very well not be mentioned at all, because it was implicit in the setting of the text (for instance if it was a web page in a site about climbing Mount Everest). Resolving a phrase like “on the south peak” would then be problematic.

There are also phrases that are georeferenced by mentioning objects or phenomena that are not locations, but rather *have* a location in geographic space. For instance, a reference to “Rosenborgs homefield” would suggest that there is a link to Trondheim in the document and “The 49ers” would provide a link to San Francisco. A text mentioning the Norwegian mountain fox would also provide a link to the last mountain area where this species still breed in the wild. Such information can only be utilized for georeferencing if the algorithms used have access to and can use large bodies of knowledge.

### 3. The Impact of Cognitive Models of Space

Formal handling of spatial information is mostly based on Euclidean geometry. But the way people think about and reason about geographic and spatial information differs from such formal mathematical models based on coordinates and exact distance and angle measurements.

According to [10] humans develop spatial understandings of their surroundings by two different methods: procedural and survey. Procedural spatial knowledge is based on exploration of geographic space by navigating it physically. Through this navigation we conceptualize it from different views and construct our view of geographic

space mentally from small pieces of information and observations (What is here, what is next to where I am now, where do I turn right to get to the mall and so on) Such geographic knowledge obtained by traveling around in an area could be called learning by “feel”. Spatial knowledge based on survey on the other hand includes looking at maps and obtaining survey and overview knowledge. Both approaches offer geographic knowledge, but information obtained using a procedural approach cannot easily be transferred to the kind of information obtained by survey and vice versa.

The field that is interested in how people handle spatial information is in [8] called “Naïve geography” and argue that this should be accepted as a separate object of study like naïve physics has been. “Naïve geography capture and reflect the way people think and reason about geographic space and time, both consciously and subconsciously. Naïve stands for instinctive or spontaneous” [8]. Some of the aspects of naïve geography might be important to take into account when georeferencing texts as they may have significant impact on how humans express spatial knowledge in texts. Some of these are described below.

*People seem to link time to geographical distance.* People seem to have a subjective perception of distance in relation to effort over time. The perception of distance between Trondheim and Vienna going by car might be different from the perception of distance between the two same cities going by plane. Also, if the effort involved in traveling from point A to point B is different than the effort involved in going from B to A, the distance between the two points can be perceived as asymmetrical. This means that a reference to distance in the text of a document might be distorted if the distance is an estimate based on a persons perception of a given space.

*Topology Matters, Metric Refines.* People seem to be able to organize space fairly well when it comes to topological relationships such as inclusion (what contains what), coincidence (what is located at the same place as what), neighborhood relations (what is next to what), and left of/right of relations. Topological relationships can be seen as first class information, where as metric relation such as distance between to neighboring phenomena is second class information in the organizing of geographic space in the human mind. Human errors are rarely topological, most often metrical. This means that topological relations between to geographic entities expressed in texts is often more reliable than geometrical distances between geographical entities occurring in texts.

*People have Biases toward North-South and East-West Directions.* When people give directions or explain where some geographical feature is placed relative to some other geographical feature they mostly use one of the four directions north, south, east and west. People tend to express a direction as for instance due south, when the correct direction should have been south-southwest. This reflects that people tend to relate things to each other by

the general directions east, west, north and south. South America is for instance often perceived as being due south of North America, even though a quick look at a map establishes that due south of North America is nothing but the pacific ocean. South America is located south-east of North America. Such biases might lead to misleading footprints when using for instance the phrase “200 km south of Moskow” for georeferencing a text if the intended location is in fact 200 km south-southwest of Moskow.

People seem to have the same kind of biases towards right angles when navigating in space.

The fact that humans treat space differently from our formal models of the world based on latitude/longitude, exact distance measurements and exact directions, implies that texts written by humans can be imprecise or ambiguous when considering the georeferences within them. In addition, the terminology people use for spatial concepts and relations may also vary according to language and culture.

#### **4. Georeferencing texts with Gazetteers, Thesauri and Ontologies**

Having identified the ambiguities of place names and human expressions of geographic references rooted in the theory of naïve geography it is clear that georeferencing documents based on place names and phrases has a certain uncertainty and that the geographic footprint of a document can be seen as correct only with a certain probability.

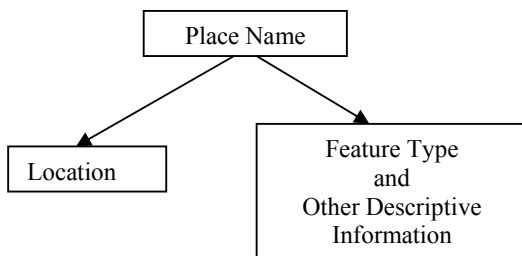
There are many alternative approaches to georeferencing, including identification of server IP, analysis of linking topology to find out if a certain page has links to information about certain areas, analysis of server logs to find out the location of users that access the information (people tend to access local pages more often then other pages) and more, but such methods really georeference the files/documents carrying the information rather than the information itself.

One can also georeference documents by considering similarities in keyword occurrences between a document to be georeferenced and documents that have already been georeferenced and by manual indexing by human experts. The Mammal Networked Information Systems, a network of institutions involved in collaborative georeferencing maintains one of the most extensive normative descriptions of georeferencing methods available [20]. Here we are going to focus on automatic georeferencing based on the contents of the documents text alone.

In an automated approach most projects have based their approaches to georeferencing on a combination of place name identification and natural language processing to identify phrases that modifies the location pointed to by occurrences of place names (“200 km south of the

Moskow”) or that provides georeferences that indicates a georeference without actually mentioning a specific place name (“Rosenborgs homefield”).

To resolve place names into geographic coordinates traditionally projects have utilize the services of one or more gazetteers. The Oxford Compact Dictionary defines a gazetteer as “a geographical index or dictionary” and [13] defines a gazetteer as “.. a list of geographic names, together with their geographic locations and other descriptive information”. [14] identifies three core components of a gazetteer: names, location and feature types.



**Figure 2:** Core components of gazetteers

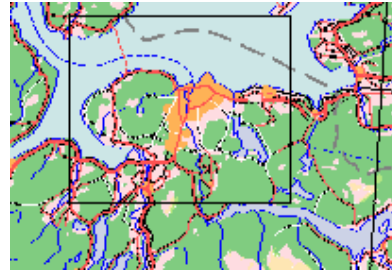
The name is the textual designator of a geographic location, the location is the coordinates of a point, line or area on the earth’s surface pointed to by a name, and the feature type is the type of location that a name points to (Forrest, agricultural area, river, inhabited location etc). The location that a place name refers to (the place names footprint) can be given as a point, a bounding box or a polygon, all represented by coordinates. Each of these alternatives has its drawbacks in terms of accuracy, datastorage requirements and processing requirements.

Using a point representing the centroid of an area requires little datastorage, but such centroids carries no representation of a locations geometry and size.



**Figure 3:** Trondheim represented by a centroid point

Using bounding boxes requires the bounding box to be larger than the footprint itself in order to cover the entire footprint, most likely overlapping with or even covering other areas as well.



**Figure 4:** Trondheim represented by a bounding box

The most accurate representation of an area is as a polygon representing an area as a set of vertexes with arcs between them. The arcs can be straight lines or be given by a function for drawing a line between the vertexes. However, this is also the most costly approach in terms of storage and processing requirements. Also, with polygon representation the question of what is accurate enough emerges. With enough vertexes and arcs a geographic area can be represented fairly accurate, but depending on the context this improved accuracy might not be needed and a more generalized polygon requiring less storage and processing capabilities would satisfy the need of the user.



**Figure 4:** Trondheim represented by an approximated polygon

In later years projects have recognized the need for more advanced gazetteer services. One of the major drawbacks of a traditional gazetteer is that it does not encode spatial or semantic relationships between places. Spatial relationships can support a user’s need for finding information in the vicinity of the original location, possibly ranked according to spatial distance. Semantic relationships can support a user’s need for finding information based on semantic relationships between place names like synonymy and hyponymy,

Traditional gazetteers do not explicitly encode spatial relationships between places/locations supporting spatial reasoning. Performing spatial relevance assessment on the footprints in the gazetteer is possible, but probably too inefficient for information retrieval purposes. [22] proposed an interesting approach to qualitative spatial reasoning in gazetteers. By combining polygon based representation of footprints with a neighborhood graph imposed on the set of geographic footprints, spatial

relevance queries can be handled by traversing the neighborhood graph rather than accessing the polygons themselves, thereby saving computation on the server. This resembles more an ontology approach to spatial reasoning than a traditional map based approach.

Also, traditional gazetteers seldom encode semantic relationships between places. A user searching for information about a location/place would often like to have information about places/locations denoted by names that are super ordinate terms, subordinate terms, synonyms or related terms. The Getty Thesaurus of Geographic Names [11] [12] encodes such relationships by recording hierarchical relationship between place names based on administrative areas and some spatial features. The Getty Thesaurus of geographic names also records versions of place names and geographic coordinates. Being a thesaurus, The Getty Thesaurus of Geographic names places more emphasis on the relation between places than their actual location on the earth in terms of map coordinates. [16] proposes a model of and conceptualization of place (ontology) that supports the locational similarity between named places. And [23] describes a system called Geographic Knowledge Representation System (GKRS) modeling the semantic relationship between place names in a semantic network.

A problem with gazetteers is that they usually cover only parts of the world. Gazetteers usually cover nations or regions very well, but there are few gazetteers that cover the whole world, and those which do cover large parts of the world have a less intensive coverage than those which only covers a defined, less extensive are. An approach to overcoming such limitations is the exchange of information between gazetteers. The ADL community has therefore proposed the ADL Gazetteer Content Standard and the ADL Gazetteer Service Protocol for supporting the exchange of gazetteer data [3].

Some projects only use place names to georeference documents. Other projects take the place name approach one step further. For instance the geoXwalk project has developed a rule base that lets the parser recognize phrases that modify the location referenced by a place name (for instance “west of Paris”, “200 km north of Oslo”).

Further analysis of text to identify phrases that does not contain place names at all, but still have a georeferencing dimension (for instance “The 49ers had free agent offensive lineman Scott Rehberg in for a visit”) is possible given a large knowledge base, but has not to my knowledge been attempted.

## 5. Searching for Georeferenced Information

There seems to be essentially two ways of letting users specify geographic constraints in a search context. The first

is to let the users specify one or more place names as keywords in a traditional keyword based query along with other key words indicating the concept part of a query. When parsing the query, the GIR/IR system then treats the found place names as special keywords by the GIR/IR system, indicating the geographical scope of the information need of the user (In typical information retrieval algorithms today place names are treated as any other keyword). The second is to let users specify the geographic constraint to a query by drawing on one or more maps.

The obvious queries to be asked in a georeferenced information retrieval system are “What is here” asking for place names, feature types and information linked to this location and “Where is it” resulting in a reference in a map. But geographic queries can be classified further as defined in [6] and [19] and summarized by [18]. The concrete query types that can be used in a Geographic Information Retrieval environment are identified as *Point in Polygon*, which asks for any georeferenced information that contains, surrounds or refers to a particular geographic point location, *Region Queries*, which asks for information regarding anything that is contained in, adjacent to, or overlaps the region, *Distance and Buffer Zone Queries*, which asks for information within some fixed distance of a geographic object (point, line, polygon), *Path Queries*, which require the presence of a network structure that can be queried for network traversal information and *Multimedia Queries* which combine multiple georeferenced information sources in resolving a query.

Finding ways to effectively combine these query types with concept based queries as in traditional IR could lead to powerful search engines capable of searching for information based on keywords but with the additional advantage of being able to search within a defined geographical area. This area can be drawn on a map using points, bounding boxes, or polygons. The hits of a search can also be represented in a map if necessary.

## 6. Some Relevant Projects

This section presents some relevant projects for geographic information retrieval. The list is by not meant to be exhaustive, but gives some indications to the activity in the field in terms of projects working with georeferenced IR.

The first system to automatically index documents based on places name and phrases was *GIPSY: The Georeferenced Information Procession System*. [9] Gipsy used a three-step algorithm which relied on a gazetteer or thesaurus containing place names and the names of other geographically significant object. First all content-bearing geographic words and phrases were identified. Then these words and phrases were analysed and geographic coordinate data were assigned to them. Finally the resulting polygons from step one were stacked on top of

each other, forming a topology. The peaks or areas above a threshold height of this topology were then chosen as the footprint for the document.

Other projects that have been oriented towards parsing text to find geographic references are the ADL Geospatial Integration Project [2] and the Berkeley Going Places in the Catalog Project [4].

Several projects are involved in construction of gazetteers, and geographic thesauri and ontologies for use in geographic information retrieval.

SPIRIT (spatially-aware information retrieval on the internet) [17] is a project funded under the EC Fifth Framework Programme aimed at improving search capabilities on the Internet by using geographical and conceptual ontologies to model both the vocabulary and the spatial structure of places for purposes of information retrieval. This ontology, which is envisioned as an extension to traditional gazetteers is expected to help finding footprints of place names and related locations as well as help ranking hits based on geographic properties.

The geoXwalk project [21] is a project aimed at providing a british and Irish gazetteer service. geoXwalk is now in its third phase focusing on three main features of geographic information retrieval; a gazetteer server / database supporting spatial searches, a set of middleware components comprising APIs supporting open protocols to issue spatial and/or aspatial search queries, a semi-automatic document 'scanner' that can parse non-geographically indexed documents for placenames, relate them to the gazetteer and return appropriate geo-references (coordinates) for confirmed matches – a 'geoparser'.

GeoVSM [5] is an abbreviation for Geographic Vector Space Model. The project integrates coordinate based geographic indexing with the key-word based vector space model in representing information space. Relevance measures are based on both geographic measures and on thematic measures which can be combined into one single measure system.

Alexandria digital library (University of California Santa Barbara) [1] aims at developing a globally distributed georeferenced digital library. Research is focused at the development of distributed, peer-to-peer digital library architectures and production-ready software, design, implementation, and evaluation of real-time 3D interfaces for visualizing geospatial digital library content, development of an integrated learning environment based on ADL geospatial digital library technology, investigation of technology-supported learning, focusing primarily on the classroom use of ADEPT by examining its impact on student learning and teaching behavior, formalization of the structure and services of Knowledge Organization Systems and integration of KOS into digital libraries and investigation of methods of deriving georeferencing by placenames and references to places from text

DLESE, The Digital Library for Earth System Education is an effort involving educators, students, and scientists working together to improve the quality, quantity, and efficiency of teaching and learning about the Earth system at all levels. The project provides access to high-quality collections of educational resources, access to Earth data sets and imagery, including the tools and interfaces that enable their effective use in educational settings, Support services to help educators and learners effectively create, use, and share educational resources and Communication networks to facilitate interactions and collaborations across all dimensions of Earth system education

## 7. Summary and Conclusions

This article was written to get an initial understanding of Geographic Information Retrieval. Indexing and retrieving documents based on their georeferences in addition to other keywords seems to provide a powerful base for searching within given areas or locations and attempts to do this in a variety of ways have existed for about a decade. However, there seems to be some unresolved issues as well. In indexing place names has to be identified, but there is no one gazetteer service that provides all possible place names that may appear in a text and there is little support for content exchange or cross-server searching. Even if a place name can be identified, there are a lot of issues to consider when using place names like for instance non-uniqueness, spellings and representation form. Also, the identification of phrases that contains direct or indirect references to place names are problematic. The typology and structure of such phrases are different between different languages and a rule base or learning algorithm has to be constructed for each language to capture the important phrases. How people reflect and reason about geographic space also have great impact on how they express geographical knowledge in texts and thereby what we can extract from them and use as the basis for georeferencing. Gazetteers have long stood their ground as lookup services for place names. But it seems that the relative simple structure of a Gazetteer is now being updated for a much more advanced usage. Geographic Thesauri and ontologies are terms that often is connected to this trend.

## References

- [1] Alexandria Digital Library Project  
University of California, Santa Barbara  
<http://www.alexandria.ucsb.edu/>
- [2] Frew J. "ADL Textual-Geospatial Integration Project"  
<http://nkos.slis.kent.edu/2002workshop/frew.ppt>

- [3] ADL Gazetteer Content Standard  
Alexandria Digital Library Project  
University of California, Santa Barbara  
<http://www.alexandria.ucsb.edu/gazetteer/>
- [4] Buckland M., Gey F.C. and Larson R., "Going Places in the Catalog: Improved Geographical Access"  
<http://www.sims.berkeley.edu/~buckland/catplace.pdf>
- [5] Cai, G., "GeoVSM: An integrated Retrieval model for geographic information" *M.J. Egenhofer and D.M. Marks (Eds): GIScience 2002, LNCS 2478 pp 65-79, 2002 Springer*
- [6] De Floriani, L., Marzano, P. & Puppo, E., "Spatial Queries and Data Models"  
*A. Frank & I. Campari (Eds). Spatial Information Theory: A Theoretical Basis for GIS. (Lecture Notes in Computer Science # 716). Berlin: Springer-Verlag 1996.*
- [7] Vice President Al Gore "The Digital Earth: Understanding our planet in the 21<sup>st</sup> Century"  
*Talk given at the California Science Center, Los Angeles, California, on January 31, 1998*  
<http://www.digitalearth.gov/VP19980131.html>
- [8] Egenhofer, M. J., Mark, D. M., "Naïve Geography"  
*Frank A. U., Kuhn, W.(Eds.): Spatial Information Theory: A theoretical foundation for GIS. Berlin 1995, Springer Verlag (LNCS 998) pp. 1-15*
- [9] Woodruff, A.G., Plaunt, C., "GIPSY: Geo-referenced Information Processing System"  
*Journal of the American Society for Information Science vol 45(9) pp 645-655*
- [10] Gluck, M., "Geospatial Information Needs of the General Public: Texts, Maps, and User's Tasks"  
*Linda C. Smith, Mike Gluck (Eds) Geographic Information Systems Patrons Maps and Spatial Information, 1995 pp 151-172*
- [11] Harping P., "Proper words in proper places: The Tesauros of Geographic Names"  
*MDA Information vol 2(3) pp 3-5*
- [12] Harping P., "The Limits of the World: Theoretical and Practical Issues in the construction of the Getty Thesaurus of Geographical Names"  
*ICHIM 97: The Fourth International Conference on Hypermedia and Interactivity in Museums, September 1997. Paris France.*
- [13] Hill I., "Introduction to Georeferencing in Digital Libraries Tutorial Document"  
*University of California, Santa Barbara Aug. 2003*  
*Contact the author at [lhill@alexandria.ucsb.edu](mailto:lhill@alexandria.ucsb.edu)*
- [14] Hill, L., "Core elements of digital gazetteers: Placenames, categories and footprints"  
*Borbinha, J. and Baker, T. (Eds.) Research and Advanced Technology for Digital Libraries, proceedings 2000*
- [15] Hill, L., Frew, J., Zheng, Q., "Geographic Names: The implementation of a gazetteer in a georeferenced digital library" *D-Lib Magazine vol 5(1)*
- [16] Jones, C.B., Harith, A., Tudhope, D., "Geographic Information Retrieval with ontologies of place"  
*D.R. Montello (Ed.): Spatial Information Theory. Foundations of Geographic Information Science. International Conference, COSIT 2001, Springer 2001*
- [17] Jones et al, . "Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project" *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval August 11-15, 2002, Tampere, Finland', ACM Press, pp.387 – 388*
- [18] Larson, Ray R., "Geographic Information Retrieval and Spatial Browsing" *Linda C. Smith, Mike Gluck (Eds) Geographic Information Systems Patrons Maps and Spatial Information, 1995 pp 81-123*
- [19] Laurini, R., Thompson d., "Fundamentals of Spatial Information Systems"  
*Academic Press, 1992*
- [20] Manis Georeferencing Guide  
<http://elib.cs.berkeley.edu/manis/GeorefGuide.html>
- [21] Reid J., A., "geoXwalk – A Gazetteer Server and Service for UK Academia" *Koch/Sølyberg (Eds) Research and Advanced Technology for Digital Libraries: 7th European Conference, ECDL 2003 Trondheim, Norway, August 17-22, 2003 Proceedings*
- [22] Schlieder C., Vögele T., Visser, U., "Qualitative spatial reasoning for information retrieval by gazetteers"  
*In Mortello, D., editor, Conference on Spatial Information Theory (COSIT) 2001, Morro Bay, California.*
- [23] Zhu, B., Ng, T. D., Schatz, B., Ramsey, M., Chen, H. "Creating a Large-Scale Digital Library for Georeferenced Information"  
*D-Lib Magazine vol 5(7/8)*