

# Information Retrieval and Reasoning about Spatial Relevance with Place Name Structures

Thomas Vögele<sup>1</sup>  
Christoph Schlieder<sup>2</sup>

<sup>1</sup> Bremen University, TZI

<sup>2</sup> Bamberg University

## Conceptual and Spatial Relevance

References to geographic location play an important role for the classification of information and data. On the Internet, as in other digital and non-digital data sources, an estimated 80% of the available data are geo-referenced. Some of these data, like digital maps and other cartographic products, are directly geo-referenced with geographic coordinates. A large section of the available data, however, do not use coordinates but are indirectly geo-referenced with place names and other plain text descriptors of geographic objects and features (Hill et al., 2000).

In information retrieval, a central task is the evaluation of relevance: among the many members of a result-set of a query, which are the ones that are most relevant with respect to the query? Meaningful ranking of query results according to their relevance is an essential feature of each good search tool.

Typical search tools are designed to resolve conceptual queries: A query like “Show me an *italian restaurant*” will trigger a search for those data sets that are relevant to the concept of “*italian restaurant*”. Because it became obvious that simplistic approaches like full-text searches and keyword matches do not suffice, some state-of-the-art systems work with keyword thesauri and/or ontologies. They typically formalize the semantics of concepts through attributes and/or the definition of relations between concepts. Equipped with tools for terminologic reasoning, advanced search tools are capable to find out that concept “*osteria*”, for example, is a sub-concept of “*italian restaurant*”, or that a place where they sell pizza may be of interest as well.

It is obvious that there are many situations where the geographic location of a query result is at least as important as its conceptual relevance. For example, a person using a mobile device to find a place to eat is not interested in *all* data sets pertaining to italian restaurants, osterias etc., but wants to know only about those instances of these concepts that are relevant with respect to a specific location (i.e. his or her current location). Consequently, in many cases queries should be of the type **concept @ location**, i.e. spatial relevance as well as conceptual relevance have to be evaluated.

## Spatial Reference Models and Place Name Structures

To support such reasoning, expressive models of the geographic domain have to be built analogous to the ontologies in the domain of concepts. We call these models *place name structures* (PNS). The PNS we have in mind do not work on a schema level like the spatial ontologies described for example by Smith (Smith et al., 1998). *Place name structures* describe the spatial relations between concrete geographic locations (i.e. geographic locations denoted with place names). In essence, place name structures work on an instance level.

A place name structure may contain standardized place names like the ones published in official place name lists (or gazetteers), or so-called “ad-hoc place names”, i.e. place names with a limited validity in terms of user community and/or time. Examples for ad-hoc place names are distribution areas used by a firm to partition their area of operation, or colloquial place names understood only within a specific geographic region.

## Integration of Place Name Structures

One of the key problems of semantic data integration is the problem of how to integrate heterogeneous ontologies: How can concept *a*, described in terms of ontology *A*, be compared to concept *b*, described in terms of ontology *B* ? Some systems adopted the approach of a *common vocabulary*: If ontology *A* uses the same basic vocabulary as ontology *B*, an automatic translation of concept *a* into concept *b* is possible (Visser et al., 2002).

In the place name structures, we use *Polygonal Standard Reference Tessellations* (PSRTs) as a common vocabulary. Examples for PSRTs are administrative units, postal code areas, or census districts. In many cases, PSRTs are built on standardized place names (administrative units) or other well-known descriptors (postal codes). This allows them to be used in an intuitive and user-friendly way.

The place name structure geo-references each place name either with respect to a reference unit of the underlying PSRT, or by its relation to other place names in the PNS. If two PNS use the same frame of reference (i.e. the same standard reference tessellation) they can be integrated easily. If they use different frames of reference, a mapping between the underlying reference tessellations has to be performed prior to integration. However, there is evidence that the number of suitable polygonal standard reference tessellations is small, with administrative units being probably the most important one.

To make place name structures and spatial reference model independent from geographic coordinates and all the associated complexities, we use qualitative abstractions of polygonal reference tessellations (Schlieder et al., 2001). These graph-based abstractions encode enough information to support complex spatial queries while drastically reducing the size of the reference data and minimizing the amount of data-processing needed. As a result, place name structures and spatial reference models are lighter and, because they can be entirely encoded in XML or RDF, better interoperable than GIS-based reference models. On the other hand, they offer more modeling flexibility and support more powerful spatial reasoning than place name lists (gazetteers).

Hill L, and Goodchild M F (2000) Digital Gazetteer Information Exchange (DGIE)

Schlieder C, Vögele T and Visser U (2001) Qualitative Spatial Reasoning for Information Retrieval by Gazetteers. Conference on Spatial Information Theory (COSIT) 2001, Morro Bay, California <http://www.tzi.de/buster/papers/cosit.pdf>.

Smith B and Mark D M (1998) Ontology and Geographic Kinds. 8th International Symposium on Spatial Data Handling (SDH'98), Vancouver, International Geographical Union.

Visser U, Stuckenschmidt H, Schuster G and Vögele T (2002) Ontologies for Geographic Information Processing. *Computers & Geosciences* 28: 103-117