Building a Geographical Ontology for Intelligent Spatial Search on the Web

Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty
School of Computer Science, Cardiff University, Cardiff, UK
{scmgf, C.B.Jones, A.I.Abdelmoty}@cs.cf.ac.uk

Abstract

All aspects of human activity are rooted in geographic space in some respect. As a consequence, many web resources include references to geographic context. In order to assist in recognising spatial terms employed in a query, it is proposed to use a geographical ontology. A geo-ontology play a key role in the development of spatially-aware search engine, with regards to providing support for query disambiguation, query expansion, relevance ranking and web resource annotation. This paper describes the geo-ontology designed for the SPIRIT system, before focussing on the problem of integrating multiple datasets for constructing such an ontology. Similarity checking of datasets is an essential step in the process of integration. The validity and effect of the different measures are studied by building a prototype geo-ontology utilising different datasets. The experimental results obtained confirmed the effect of quality of the datasets and the importance of the flexibility of the technique proposed for adjusting the similarity measures to handle such an effect.

Keywords: Ontology, Spatial Integration, Semantic Web

1 Introduction

The `World Wide Web` holds vast amounts of information. However, users do not always get information they expect when searching the web. One reason for this is that existing web resources are rarely augmented with semantic annotation that describe their content, which would make them more easily accessible to automated search facilities. The *Semantic Web* is one of solutions that are proposed to enrich Web resources with some well-defined meaning (meta-data), and it is recognised that ontologies play a key role by acting as a shared vocabulary for such meta-data.

As an effort to this direction, in this paper we study the design and construction of a geographical ontology (or geo-ontology) to assist spatial search on the Internet. This research is part of the `SPIRIT` project [5]. The main motivation of SPIRIT is that a large portion of web resources may be regarded as geographically located. For example, many Web documents report on activities that take place in some locations on the Earth's surface. However, existing web facilities are poorly adapted to find information that relate to a particular location. When supplied with a spatial query, a typical search engine only return web pages that include that place name involved. Other web documents that are associated to that place, e.g. that refer to places in the proximity of the subject place or are contained within it, may also be of interest but usually not returned.

The potential of using geographical information to assist spatial search has been recognised in literature [4, 18, 9, 19]. Most research is based on the use of a gazetteer in which a place is normally represented by point locations. The limited spatial semantics associated with these approaches narrows the scope of their ability to effectively retrieve useful resources for spatial queries. This research studies how to build a geo-ontology which plays a central role in intelligent spatial search on the Web. It serves as a shared vocabulary for spatial mark-up of Web resources. It facilitates disambiguation and interpretation of users queries. It also enables the generation of spatial index to support efficient retrieval of web resources as well as relevance ranking of retrieved results.

The process of building the geo-ontology involves first developing an appropriate underlying conceptual model to support the required functionalities, and secondly populating the ontology with enough detail to realise its full potential. The later task is a significant challenge especially when integration of several different data sources is involved. Data for populating the geo-ontology may come from a variety of sources, and they may differ in their underlying structures, their accuracy and the levels of details on representations of the places. This paper focuses on the problem of building a geo-ontology and in particular the challenge of utilising multiple data sources for its populations. The remainder of the paper is organized as follows. The conceptual design of the geo-ontology is described in Section 2. Section 3 reviews related work. An approach for integrating multiple resources for constructing the geo-ontology is reported in Section 4 and experimental results are given in Section 5. Conclusions are given in Section 6.

2 The Conceptual Design of the Geographical Ontology

A spatial query in SPIRIT can be formalised as a triple $< what, rel, where >, where$ *What* specifies a general non-geographic object; *where* specifies a query term which is geographically referenced; *rel* is a spatial relationship which relates the *what* and *where* terms. An example of such a query is "universities near Cardiff". In this section, we investigate various uses of the geo-ontology in SPIRIT, and present a conceptual design of the geo-ontology. Figure 1 illustrates how the geo-ontology communicates with other components of SPIRIT.

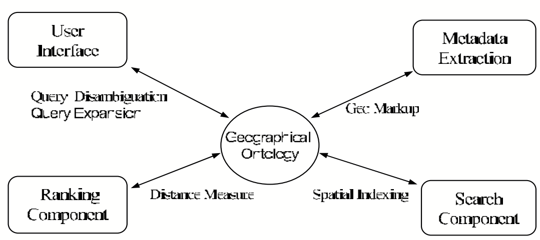

Figure 1. Roles of Geo-Ontology in SPIRIT

The *metadata extraction* component interacts with the geo-ontology to recognise the presence of place names in a document for geo-markup of web documents. Once the annotation of Web documents is achieved, the search component interacts with the geo-ontology to retrieve the geometric footprint associated with the places in documents. These footprints will then be used for building up the spatial index of the web collection. The *user interface* component interacts with the geo-ontology to interpret users' queries. The geo-ontology is first used to detect if there are multiple occurrences of the place name involved in the query. Once the place name is disambiguated, the query is expanded using the ontology, and a new interpreted geographic search extent satisfying the spatial relationship is derived. For example, if the query is "castles near Cardiff", then the spatial search extent may cover "Caerphilly", a region near Cardiff. Finally, the *ranking component* interacts with the geo-ontology to measure spatial relevance of the retrieved documents basing on the comparison of the query footprint and the document footprint.

To fulfil the above objectives, we proposed a conceptual design of the geo-ontology for SPIRIT. To meet the needs of the recognition of the geographical terminologies in the user interface and for metadata extraction of Web documents, the geo-ontology encodes the vocabulary of place names for its region of applicability. Multiple names are supported for each place. This is necessary as a geographic place may have variant names, e.g. "Abertawe" and "Swansea" refer to the same place. By maintaining multiple place names, the ontology facilitates the retrieval of resources that employ alternative place names to that in the query.

To support spatial indexing, query expansion and relevance ranking, the geo-ontology maintains geometric footprints associated with a place. More than one geometric footprints for each place are supported, each one can take the form of *Points*, *Polylines* and *Polygons*. Footprints of different types can be used for different purposes, e.g. a detailed *Polygon* may be used to build up spatial indexing of documents, and a central point may be used by the interface for plotting the search results in an interactive map.

To support spatial query expansion, the geo-ontology maintains classification categories of places, such as *counties, districts*. This is because some relationships in spatial queries, e.g. *near* are fuzzy and may have different interpretation according to the types of places involved. For example, the *near* relationship in a query which searches for "rivers near Westerminster" need to be treated differently from one which searches for "rivers near Lancashire". As "Lancashire" is a County, its neighbourhood region is larger than that of "Westminister" which is a Ward.

Finally, to disambiguate the place names involved in queries, the geo-ontology encodes *containment* relationships between places. This is useful as a place name may be shared by multiple places. By using the *containment* relationship, the geo-ontology is able to derive the broader spatial contexts of a place. For example, using following information, a user should be able to disambiguate which "Newport" s/he is interested in.

$UK, Wales, Newport$
$UK, England, Devon, NorthDevon, Newport$
$UK, England, Leicestershire, Melton, Newport$

3 Related Research and Resources

Using geographical information to assist spatial information retrieval has been pursued in several studies citeRie02a,GMV99a,STV01a,JAT01a. Most of these studies use gazetteers and geographical thesauri, such as citeADL04a,Getty04a,NIMA04a, to help handle spatial search. The essential components for the majority of gazetteers or thesauri are place names, geographic locations and type designation for each place. Existing geographical datasets are useful to SPIRIT, however, they may not be sufficient for the several reasons. Some datasets may only cover a specific area of space, such as [14], whereas the initial aim of SPIRIT is to support spatial search in the scope of Europe. Some datasets are specialized for encoding places of specific types, such as [15], whereas the

SPIRIT geo-ontology needs to encode places that occur at multiple levels of detail ranging from continents and oceans down for example to small villages and streets. Moreover, datasets vary in the degree of semantic richness, e.g. in some datasets a place is only associated with a single name [7], while in other datasets alternative names are also encoded [13, 3]. Finally, datasets may also vary in the type of spatial information stored for a place. In some cases, detailed representation of object boundaries are available [20], while in others abstracted representations in the form of centre points, for example, are stored [8].

As a consequence, a geo-ontology which satisfies the design and system requirement as described in Section 2 needs to be constructed in SPIRIT. The construction of such a geo-ontology requires integrating data from various datasets. An available dataset for SPIRIT is the SABE dataset [7]. The SABE is a digital map dataset representing the geometry of administrative boundaries for Europe. The main advantage of using the SABE is that it contains the detailed geometric footprints for each place it stores. However, the SABE only encodes places of administrative type and very limited thematic information is modelled in the SABE. Therefore using it as the single resource for constructing the geo-ontology will make it impossible to fully interpret the types of spatial query that SPIRIT intends to support. A possible dataset, which complements the SABE is the TGN [8]. The TGN is a structured vocabulary containing information about places with a Global scope. Though each place in the TGN is represented only by a point footprint, the TGN is rich in the thematic information, i.e. it includes names of physical and administrative places not only from the modern world as well as from historical and linguistic origins.

The availability of diverse data resources in one respect enables us to construct a geo-ontology required by SPIRIT, but it also imposes considerable difficulties when integrating them into the geo-ontology. For example, a same place may be encoded in two datasets using different names, and different place types may be assigned to it depending on the classification schema adopted. Furthermore different datasets may encode footprints in different levels of detail. All of this makes the data integration process a difficult one.

Some research has been carried out to investigate the integration of geographical data coming from different resources, of which a critical step is to measure the similarity of two places. We could classify these studies into two categories. The first category of approaches investigates the use of thematic properties, such as descriptors of concepts, or their inter-relation, e.g. RT, BT, NT etc, to measure the similarity of places [2, 1]. The second category of approaches uses the spatial properties as the main criteria to measure whether two places match or not [11, 17, 6, 10]. The main information used is the geometric footprint and various spatial relations between places, such as topological and directional relationships. The use of a combination of thematic and spatial information in similarity measure is investigated in work such as [16].

The approach proposed in this paper builds on the techniques proposed in [16], i.e., both thematic and spatial information are used for data integration. However, it is different from previous research in the following. Firstly, we study the geographical integration with the complexity of the existence of multiple place names, classification types, hierarchies and footprints, which have not been approached previously. Secondly, previous methods focus on measuring the geographical *closeness* of places. If two places are geographically close to each other, they usually have a high similarity matching score even if they are entirely separate places. In this research, we do not care about the geographical *closeness* between different places, rather we are concerned with finding places which are potentially the same by using spatial information, which allows our spatial measurement criteria to be slightly different from existing ones.

4 Integrating Multiple Datasets for Geo-Ontology Construction

An important step in the integration process is to apply similarity checking procedures to identify whether places in the new dataset already exist in the ontology. Four similarity measures between geographic places are proposed in this paper. Two are related to the thematic properties of a place, namely, place name and place type. The other two are related to the spatial properties of a place: footprint and geographical hierarchy which is derived from containment relationships between places.

Given a place P_1 in a dataset, we look for in the geo-ontology the places which match P_1. This is achieved by performing the similarity measure between P_1 with a place P_2 in the geo-ontology. Each similarity measure of $< P_1, P_2 >$ generates a set of matching scores (each for name matching, type matching, hierarchy matching and footprint matching). A set of thresholds (corresponding to name, type, hierarchy and footprint matching respectively) is used to determine whether P_1 matches P_2. If a match is found, then P_2 is updated with the information from P_1 as appropriate, otherwise, P_1 is considered to be a new place to be added to the ontology. By tightening and relaxing the thresholds, the similarity matching process can be adapted according to the characteristics of the datasets considered.

4.1 Place Name Matching

Place names can be considered as a primary factor for indicating the similarity of two places. Names for a same place may either match exactly or have some simple syntactic differences. A normalization step needs to be applied if two names do not match exactly. Typical normalization operations include discarding letter case differences; dropping stop words, e.g. “a”, “of”, “the”, “and”, “-”, “(”,“)” and white spaces etc; dropping prefix or a suffix words in a place name, usually used to indicate the type of the place, e.g. the “District” in the name “Christchurch District”. Usually more than one normalisation operation may need to be applied to place names. Accordingly, the following criteria is used to determine the matching score of two places according to their names N_1 and N_2.

$$\sigma(N_1, N_2) = \begin{cases} 1.0 & if\ N_1 = N_2 \\ 1.0 - m/(n+1) & if\ N_1 = N_2\ after \\ & normalization \\ 0.0 & otherwise \end{cases}$$

That is, $\sigma(N_1, N_2)$ is 1.0 if N_1 and N_2 match exactly. If normalisation is needed, then $\sigma(N_1, N_2)$ is calculated as a function of both the number of normalization steps m required in order for N_1 and N_2 to match and the total number of available normalization steps n that can be performed. For example, if we have 3 different types of normalization operations and 2 steps are required for N_1 and N_2 to match, then $\sigma(N_1, N_2) = 1 - \frac{2}{3+1} = 0.50$.

In the cases of a place with multiple names, the above measure is applied for every possible pair of names and the highest score is retained. Two place are considered to be matching if the matching score is greater than or equal to the name threshold (the similar process applies to other similarity measures as well).

4.2 Place Type Matching

Place type matching used here is based on the traditional semantic measuring techniques where match score is calculated by using the minimum number of semantic relationships that must be traversed to connect two terms [12]. Possible semantic relationships include BT (broader-than), NT (narrower-than), RT (related-to), etc. As different datasets may encode places using different classification schemes, some equivalence links should firstly be established between classification hierarchies [2]. Several approaches in the literature have been proposed for measuring the semantic distance of two terms [2, 1]. In this work, we limit the place similarity measure to utilise the BT/NT relationships only. The formula below is used to calculate the matching score of two places according to their types T_1 and T_2.

$$\sigma(T_1, T_2) = \begin{cases} 1.0 & if\ T_1 \equiv T_2 \\ 1/n & otherwise \end{cases}$$

That is, $\sigma(T_1, T_2)$ is 1.0 if T_1 and T_2 are equivalent, otherwise $\sigma(T_1, T_2)$ is calculated using the number of steps n required to traverse from T_1 to T_2. For example, given hierarchies shown in Figure 2 (where a directed edge indicates a BT/NT relationship between two types and a dashed edge links equivalence types among hierarchies), the matching score for (T_3, T_{12}) is 0.5 as only 2 steps are required to travel from T_3 to T_{12}, and 0.33 for (T_4, T_{11}) as 3 steps are required to travel from T_4 to T_{11}.

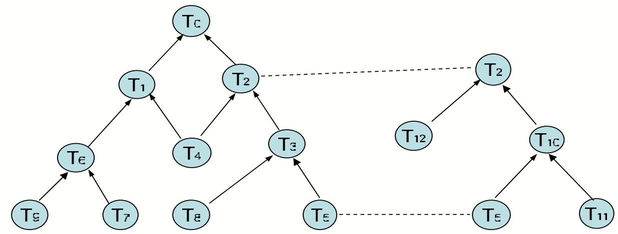

Figure 2. Matching Types

4.3 Hierarchy Matching

The hierarchical information can be derived by using geometric footprints or from the *containment* relationship between places if exist. If two places have the same hierarchy, then there is a strong possibility that they refer to the same place. If two places have different hierarchies, they can be considered to be different even if they have the same name or place type. Hierarchy matching used here is based on the work in [4, 16]. While these studies measure the *closeness* of two places even if they are separate places, we only care about places that are potentially same. Accordingly we have a slightly different formula to compute the matching score of two places by using their hierarchies H_1 and H_2.

$$\sigma(H_1, H_2) = \begin{cases} 1.0 & if\ H_1 \equiv H_2 \\ n_1/n_2 & if\ (H_1 \subset H_2) \vee (H_2 \subset H_1) \\ 0.0 & otherwise \end{cases}$$

That is, $\sigma(H_1, H_2)$ is 1.0 if H_1 and H_2 are equivalent. If H_1 and H_2 are subsets of each other, e.g. “ United Kingdom, Wales, Cardiff” contains “United Kingdom, Wales”, $\sigma(H_1, H_2)$ is computed using n_1 and n_2, where n_1 is the number of places that are common to both H_1 and H_2 and n_2 is number of places in the bigger hierarchy. In the above example, n_1 is 2 and n_2 is 3, therefore the matching score of the two hierarchies is 0.667.

4.4 Footprint Matching

Intuitively footprints are the strongest indication for similarity of two places. That is, if a two places are exactly in the same location on the ground, then they can be considered to be similar. However, due to the quality of the data and the different spatial representations that may be used, locational matches may not be as reliable as expected. In this work, the footprint similarity is based on the distance between representative points for the places involved. If the place is represented by a polygon, then a centroid is used as the representative point and the *origin* is used in the case of polylines. Accordingly, we have the following criteria to measure the similarity of two footprint F_1 and F_2.

$$\sigma(F_1, F_2) = \begin{cases} 1.0 & if\ F_1 \equiv F_2 \\ \mathbf{min}(1.0, tol/dis) & otherwise \end{cases}$$

If F_1 and F_2 are the same, they have the matching score 1.0. Otherwise, a tolerance tol is used in conjunction with the distance dis between F_1 and F_2 to determine the matching score. Since tol is a constant, the bigger dis, the lower is $\sigma(F_1, F_2)$. If $dis < tol$, $\sigma(F_1, F_2) = 1.0$.

5 Experiments

A set of experiments have been carried out to test the feasibility of the method proposed. Firstly, a set of algorithms has been implemented to allow the automatic data integration based on the method proposed. Secondly, a geo-ontology which contains geographical places of UK has been constructed. The data resources used for building the UK geo-ontology is the SABE and TGN. The initial dataset we exploit is SABE, and 11569 places have been extracted to build the initial image of the geo-ontology. We then extracted from TGN the full list of places of UK, and total 7715 places are extracted.

Table 1. Matching Result

experiment setting					matching results		
	t_1	t_2	t_3	t_4	N_1	N_2	N_3
1	1.0	0.0	0.0	0.0	3395	2539	2719
2	0.25	0.0	0.0	0.0	3429	2567	2751
3	0.25	0.5	0.0	0.0	2043	1965	1929
4	0.25	0.65	0.0	0.0	1979	1940	1874
5	0.25	0.5	0.5	0.0	103	100	98
6	0.25	0.65	0.25	0.0	1979	1940	1874
7	0.25	0.65	0.25	1.0 (a)	1667	1661	1611
8	0.25	0.65	0.25	1.0 (b)	1714	1705	1650
9	0.25	0.65	0.25	1.0 (c)	1737	1726	1670
10	0.25	0.0	0.25	1.0 (d)	1675	1667	1618

$(a)\ tol = 10\ km \quad (b)\ tol = 15\ km$

$(c)\ tol = 20\ km \quad (d)\ tol = 10\ km$

Table 1 shows some matching results, where t_1, t_2, t_3 and t_4 are thresholds for the name, hierarchy, type and footprint matching. N_1 is the number of matching pairs, N_2 and N_3 are the number of distinctive places in all the matching pairs coming from TGN and SABE respectively.

The $experiment\ 1$ checks for precise name matching. Out of 3395 pairs of matching places found, 2539 are from the TGN and 2719 are from the geo-ontology. The figure indicates many places share same names in the datasets and this results in multiple mappings. For example, "Woodford" is used by two places both in TGN (tgn_{159} and tgn_{172}) and SABE ($sabe_{278}$ and $sabe_{597}$). Hence, four matching pairs are for the name "Woodford" are found: $< tgn_{159}, sabe_{278} >$, $< tgn_{172}, sabe_{278} >$, $< tgn_{159}, sabe_{597} >$ and $< tgn_{172}, sabe_{597} >$. Imprecise name matching in the $experiment$ 2 finds a further 34 pairs of place matches. The findings suggest that in this case, both datasets utilises mostly similar place names.

Using the matching result from the $experiment$ 2, the $experiment$ 3 studies how places match to each other with regards to their geographic hierarchies. 2043 matching pairs were found. The reduction is primarily due to filtering out all wrong multiple matches. For example, for the "Woodford" example described above, two wrong matches, $< tgn_{159}, sabe_{278} >$ and $< tgn_{172}, sabe_{597} >$ are eliminated according to the following hierarchical information:

$$\begin{array}{l} tgn_{159} : UK, England, Stockport, Woodford \\ tgn_{172} : UK, England, GreaterLondon, Woodford \\ sabe_{278} : UK, England, GreaterLondon, Woodford \\ sabe_{597} : UK, England, Stockport, Woodford \end{array}$$

Another exercise with the hierarchy matching, but with tighter thresholds (0.65 in the $experiment\ 4$) did not have a significant impact on the results. Manual checks of the difference in matching pairs between $experiment$ 3 and 4 revealed that an error of approximately 10% is resulted by using higher hierarchy threshold. Hence, in this case, a hierarchy threshold between 0.5 and 0.65 is sufficient. The $experiments$ 5 and 6 studies type matching. A strong type match in the $experiment\ 5$ failed to identify most of the matching pairs. When the type threshold was relaxed in the $experiment$ 6, all matching pairs from the $experiment$ 4 were found. Hence, we can conclude that type matching is not a significant similarity measure in this case.

The $Experiment\ 7$ studies the effects of footprints matches by using the result obtained from $Experiment$ 6, and a tol of 10 $kilometres$ was used. 1667 matching pairs were found. A manual check found that 80% out of the missing matching pairs were in fact correct and the remaining 20% were uncertain. Hence, this indicates that the tol of 10 kilometres is too rigid and needs to be relaxed even further. Further experiments ($8\ and\ 9$) using different tol confirm our findings. In $experiment\ 10$, the effects of a combined name and footprint matching are studied. The tolerance of 10 kilometres is maintained. A similar result from experiment 7 is obtained with a only a few of the missing matching pairs found. This confirms that the name and hierarchy thresholds used are sufficient.

Our experiments reveal the importance of a combined name and hierarchy matching for checking place similarity. Although footprint matching allowed the discovery of further missed similar places, finding the appropriate distance tolerance needs further investigation. The limitation of the footprint measure, in this case, is the result of the approximation in spatial representations used in TGN.

6 Conclusions

This paper presents a geographical ontology which is constructed to assist spatial search on the Internet. In particular, the paper focuses on the problem of integrating different datasets in the process of populating the ontology. Similarity measures have been identified to measure the similarity of places coming from different datasets. To study the effects of utilising those measures in isolation and in conjunction, a prototype ontology was developed using two different datasets and experiments were carried out to identify the effect of the measures on a sample of common geographic places in the datasets. The experiments confirmed the suitability of the place name matching in conjunction with the hierarchy matching in identifying similar places in the SABE and TGN datasets. The results also confirmed the effects of the quality of the datasets on the process of integration. Further investigations with other types of datasets need to be carried out to confirm our findings.

Acknowledgement

SPIRIT is funded by EC Fifth Framework Programme.

References

[1] A. Rodrguez and M. Egenhofer. Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure. International Journal of Geographical Information Science Vol. 17 (in press).

[2] A. Rodrguez and M. Egenhofer. Determining Semantic Similarity Among Entity Classes from Different Ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2):442–456, 2003.

[3] Alexandria Digital Library Project. ADL Gazetteer. http://alexandria.sdc.ucsb.edu/ lhill/adlgaz/, 2004.

[4] C. B. Jones and H. Alani and D. Tudhope . Geographical Information Retrieval with Ontologies of Place. In *Proceedings of the International Conference on Spatial Information Theory*, pages 322–335. Springer-Verlag, 2001.

[5] C. B. Jones and R. Purves and A. Ruas and M. Sanderson and M. Sester and M. J. van Kreveld and R. Weibel. Spatial Information Retrieval and Geographical ontologies: an Overview of the SPIRIT Project. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 387–388, 2002.

[6] E.A. El-Kwae and M.R. Kabuka. A robust framework for content-based retrieval by spatial similarity in image databases. *ACM Transaction Information System*, 17(2):174–198, 1999.

[7] EuroGeographics. Seamless Administrative Boundaries of Europe (SABE) . http://www.eurogeographics.org/eng/01_about.asp, 2004.

[8] Getty. Thesaurus of Geographic Name. http://www.getty.edu/research/tools/vocabulary/tgn/, 2004.

[9] N. Guarino, C. Masolo, and G. Vetere. OntoSeek: Content-Based Access to the Web. *IEEE Intelligent Systems*, 14(3):70–80, 1999.

[10] J.M. Ware and C.B. Jones. Matching and Aligning Features in Overlayed Coverages. In *Proc. of 6th International Symposium on Advances in Geographical Information Systems*, pages 28–33, 1998.

[11] M. Egenhofer and R. Franzosa. On the Equivalence of Topological Relations. *International Journal of Geographical Information Science*, 9(2):133–152, 1995.

[12] M. W. Bright and A. R. Hurson and S. Pakzad. Automated Resolution of Sematic Heterogeneity in Multidatabases. *ACM Transaction on Database Systems*, 19(2):212–253, June 1994.

[13] NIMA. GEOnet Names Server. http://164.214.2.59/gns/html/index.html, 2004.

[14] Ordnance Survey. 1:50000 Scale Gazetteer. http://www.ordnancesurvey.co.uk/oswebsite/products/50kgazetteer/, 2004.

[15] Ordnance Survey. Code-point. http://www.ordnancesurvey.co.uk/oswebsite/products/codepoint/, 2004.

[16] D. Papadias, N. Karacapilidis, and N. Arkoumanis. Processing Fuzzy Spatial Queries: a Configuration Similarity Approach. *International Journal of Geographical Information Science*, 13(2):93–128, 1999.

[17] R. Goyal and M. Egenhofer. Similarity of Direction Relations. In *Proceedings of the 7th International Symposium on Spatial and Temporal Databases*, pages 36–55, 2001.

[18] W. Riekert. Automated Retrieval of Information in the Internet by Using Thesauri and Gazetteers as Knowledge Sources. *Journal of Universal Computer Science*, 8(6):581–590, June 2002.

[19] C. Schlieder, T. Vgele, and U. Visser. Qualitative Spatial Representation for Information Retrieval by Gazetteers. In *Proceedings of the International Conference on Spatial Information Theory*, pages 336–351. Springer-Verlag, 2001.

[20] Tele Atlas. Teleatlas multinet. http://www.ordnancesurvey.co.uk/oswebsite/products/50kgazetteer/, 2004.