

Hybrid Geo-spatial Query Methods on the Semantic Web with a Spatially-Enhanced Index of DBpedia

Eman M.G. Younis, Christopher B. Jones, Vlad Tanasescu, and Alia I. Abdelmoty

School of Computer Science & Informatics, Cardiff University,
CF24 3AA, United Kingdom

{E.Younis,C.B.Jones,V.Tanasescu,A.I.Abdelmoty}@cs.cardiff.ac.uk

Abstract. Semantic Web resources such as DBpedia provide a rich source of structured knowledge about geographical features such as towns, rivers and historical buildings. Retrieval from these resources of all content that is relevant to a particular spatial query of, for example, containment or proximity is not always straightforward because there is considerable inconsistency in the way in which geographical features are referenced to location. In DBpedia some geographical feature instances have point coordinates, some have qualitative properties that provide explicit or implicit locational information via place names, and some have neither of these. Here we show how structured geo-spatial query, a form of question answering, on DBpedia can be performed with a hybrid strategy that exploits both quantitative and qualitative spatial properties in combination with a high quality reference geo-dataset that can help to support a full range of geo-spatial query operators.

Keywords: GIS, maps, question answering, query, linked data, sparql.

1 Introduction

The Web can be regarded as a rich source of geographical information but much of that information can be difficult to retrieve because it is embedded in natural language text. Conventional search engines can access documents that contain place names in a user's query but unless the required content is a yellow pages listing, which may be picked up with so-called local search methods, it remains for the user to sift through the retrieved documents to find relevant information. Research into spatially-aware search engines has produced systems that can improve the quality of retrieval results but those results are still typically unstructured text documents (Purves et al, 2007). Semantic Web technologies have been motivated by the objective of machine-readable access to structured data on the Web, which holds the promise of much more focused responses to user queries. The SPARQL query language for example can be used to formulate queries on RDF (Resource Description Framework) data records on the Web in the form of subject-predicate-object triples. If semantic equivalence between the data items in different triple stores can be asserted then they can be linked together to support more complex queries across multiple RDF data sources. As increasing amounts of geo-referenced information become encoded in this way the Semantic Web is becoming a valuable source of structured geo-information.

Probably the currently richest source of geo-referenced information on the Web, with regard to the semantics, is to be found in Wikipedia and its Semantic Web version DBpedia (an RDF resource: <http://dbpedia.org/>) which contains hundreds of thousands of entities that are geo-referenced with point geometry, but not with lines or polygons. Increasing quantities of digital map data that can complement the geo-semantic content of RDF resources such as DBpedia are also appearing, with OpenStreetMap (OSM) being perhaps the most substantial source of freely available topographic features encoded as points, lines and areas. OSM has been converted to RDF format and links have been determined between some of its geometric features and corresponding topics in DBpedia in the LinkedGeoData project (Stadler et al. 2011). That project has exploited the links with a mapping system that supports pan and zoom of maps annotated with DBpedia features, but not structured spatial query.

The work presented in this paper is motivated by the objective of supporting high quality spatial query to rich semantic RDF content, such as DBpedia, which has many records describing the semantics of geographical features but has limited geo-spatial data. Thus we wish the user to be able to perform typical geo-information queries such as to find specified types of content within named regions, such as a city, or within a specified distance of the centre or the boundary of a region and to find features that hold overlap and crossing relations with a reference place or feature. Such queries cannot be performed using only the single point geometry associated with geo-referenced features in DBpedia. To achieve our objective we maintain a high quality, detailed reference digital map dataset for the entire region of interest, so that queries that name any feature in the region can be instantiated with the relevant feature geometry. The reference geo-data are stored in a spatially indexed database in combination with a spatial index of geo-referenced entities in RDF content, i.e. DBpedia.

The approach may be seen as a spatially intelligent index of Semantic Web content comparable with the role of an inverted index in a conventional web search engine, but in the work presented here we only currently access the single RDF resource of DBpedia. The analogy is with regard to the need for fast access to Web resources that satisfy the user's query constraints, which in our case can include spatial qualifiers. Unlike a conventional search engine for web documents which returns the URLs (uniform resource locators) of matching documents, we use the URIs (uniform resource identifiers) retrieved from the local index to formulate SPARQL queries on the DBpedia endpoint (a URL) if that is required to provide an explicit answer to the query.

In the present version of our experimental system the spatial index of DBpedia content is based on the point coordinates associated with DBpedia instances. Because not all geographical DBpedia instances actually have coordinates, we exploit other qualitative properties of the instances to associate them with contained places. Thus there are multiple properties that name the containing town, city or other administrative area, such as *dbpo:administrativeDistrict*¹, *dbpo:location* and

¹ The prefix *dbpo* stands for <http://dbpedia.org/ontology/>, while *dbpedia* stands for <http://dbpedia.org/resource/> and *dbpp* for <http://dbpedia.org/property/>

dbpo:locatedInArea in which the property name is either implicitly or explicitly spatial. These types of properties are particularly valuable in processing containment queries where the user-specified containing place may match the object of one of these DBpedia properties. There are also properties that indicate other spatial relations such of cardinal direction and proximity, but these latter properties have no consistent spatial interpretation and are not used in any consistent way. It is also the case that, just as many DBpedia place instances do not have coordinates, there are many place instances that do not have properties that imply containment at a useful granularity. For purposes of containment search our approach is therefore a hybrid one that combines exploitation of the qualitative spatial properties with results obtained from geo-spatial processing methods that test for containment of DBpedia point coordinates within polygons provided by the reference geo-data. Exploitation of coordinates and spatial properties cannot be guaranteed to find all place instances that may be relevant to a containment query, as some instances have neither of these types of property. Successful access to such instances may depend upon automated geo-referencing using methods such as those described in De Rouck et al. (2011), which was applied to Wikipedia articles. Our system is designed on the assumption that this will be achieved in the future, enabling DBpedia instances to be maintained in the local spatial index.

It should be noted that because the approach we present depends for its effectiveness upon the presence of rich and detailed digital map data corresponding to the RDF semantic content, we employ a national mapping agency dataset, with consistent high quality spatial coverage, to demonstrate the approach for a single country. As OSM evolves it may well be able to serve that role and is of course international in coverage.

In the remainder of the paper we summarise related work in section 2 before providing an overview in section 3 of the architecture of our experimental system. Section 4 presents some experimental results that demonstrate the effectiveness of the approach including an analysis of the availability of coordinates and qualitative spatial containment properties. The paper concludes in Section 5 with a summary of the progress to date and directions for further development of public access spatial query of geographical information on the Semantic Web.

2 Related Work

Our work on structured query of geographic information can be regarded as a form of geographic question answering system (GQAS), but it differs from much existing work in that area in focusing on structured data rather than free text and in the use of geo-spatial processing in addition to exploiting some qualitative data. We review first briefly some work in this area that is largely based on language processing. The START system (Katz and Lin, 2002; Lin and Katz, 2003) accepts natural language questions and can provide some properties of geographic places such as their population and distances between places, but it is not able to satisfy typical geo-spatial questions regarding proximity and topological relationships between places.

The Geo-Logica system (Waldinger et al, 2004) incorporates an automated deduction system with spatial and temporal reasoning capabilities, whereby having formulated the natural language query in a logical form, geographic information is extracted from text documents from various sources. It cannot process spatial relations explicitly or compute spatial properties from geo-data. A voice activated GQAS is proposed in Luque et al (2008) that allows speech input of questions about Spanish geography. It is based on language analysis of free text Web resources in combination with place name gazetteers and a training corpus of geographical questions. The QUASAR system (Buscaldi, 2007) uses language processing to access free text sources, including use of Wikipedia, to extract geographic information, with a focus on word sense disambiguation. The GeoCLEF and GikiCLEF events have resulted in publication of geographic question answering systems but these are mostly based on information extraction from free text documents. The work described by Hartrumpf and Leveling (2010) combines text information extraction with geographical information retrieval (GIR) methods that work with a spatial index of documents, and does exploit DBpedia, alongside Wikipedia, as a source, but it converts the RDF to natural language expressions for processing by the non-GIR methods. Mishra et al (2010) employ the user's query to retrieve documents from a search engine that are then subject to information extraction, results of which can be viewed on a map.

The increasing quantity of Semantic Web resources, including the Geonames gazetteer, OSM and DBpedia, has led to several initiatives to provide spatially enabled access to their content. The LinkedGeoData project (Stadler et al, 2011) contributed to the transformation of OSM to RDF and has presented methods to determine links between map features in OSM and equivalent instances documented in DBpedia, as well as between OSM and Geonames. Their matching is based on a combination of the Jaro-Winkler string distance between the text of the respective place names and the geographic distance between the entities. They have illustrated exploitation of the links with an interactive map that supports pan and zoom but that application does not support geo-spatial query with conventional spatial relationships. Linking between equivalent entities is a critical issue in exploiting Semantic Web data. Examples of other work on linking geo-data on the Semantic Web are Hahmann and Burghardt (2010), which uses Levenshtein string distance, and Sala and Harth (2011) which employs the Hausdorff distance to establish similarity between spatially extensive linear or polygonal features. We employ similar methods in our work to establish links between DBpedia place instances and geo-features in our reference geo-data store.

The availability of geospatial data on the web has motivated various developments to enhance SPARQL, the main language for access to RDF, with spatial functionality which may be supported by the various triple stores of RDF content. GeoSPARQL reflects a W3C supported initiative to create such a language (Battle and Kolas, 2011). It provides a full range of spatial operators accessed via a spatial index of the corresponding RDF data store. Rather than modifying the SPARQL language for geo-spatial query, Brodt et al (2010) present an approach that confines spatial functionality to SPARQL filter functions. A hybrid approach is presented in Della Valle et al (2010) in which spatially enhanced SPARQL queries are mapped to a

PostGIS spatial database that implements spatial indexing and spatial query operators. They demonstrated the approach with queries that employ a mix of polygonal data from OSM data with point referenced data such as from DBpedia. Our work differs from such approaches in creating a centralized spatially-enabled index of Semantic Web content, in the manner of a web search engine, and employing a locally-stored reference geo-dataset to mediate queries that require line and polygonal representations of named features. We implement hybrid query methods that exploit qualitative spatial properties in addition to quantitative geo-data. The system also performs queries on an RDF triple store if required

Because many Semantic Web resources such as Geonames, OSM and DBpedia are volunteered resources, contributed to by individuals with only informal procedures for validating the content, there can be considerable variation in the quality and coverage of the data (see for example Hackley 2010 and Mooney et al 2010, for some analyses of OSM). Although OSM is improving greatly in coverage, for the purposes of our experimental system we have used a national mapping agency dataset which, while restricted to a single country for purposes of our experimental system, has high quality linear and polygonal geometric representation of many map features including city boundaries. For international coverage, OSM clearly has tremendous potential to serve a similar function.

3 System Architecture

The key components of the architecture we present (see Figure 1) consist of a query interface, a query processor, a local spatially indexed geo-data repository (referred to as the reference geo-data or reference geometry), a local spatially indexed store of place URIs extracted from DBpedia, and access via SPARQL to the external RDF store which is DBpedia.

The current version of the experimental user interface supports structured query for question answering enabling the user to enquire about properties of places, and to find places subject to spatial constraints of containment, crossing and proximity. When enquiring about properties of places the user is given a drop down menu that lists the DBpedia properties associated with the place the user names. A SPARQL query to DBpedia is used to retrieve the answer. For spatial queries, that may include a feature type constraint, the following methods have been implemented for comparative purposes:

- Within distance of a point, line or area geometry object, where the buffer is created from reference geo-data;
- Crossing a line or area feature, where the DBpedia instance may be represented by reference geo-data, such as a line, and the feature is also represented by reference geo-data;
- Containment using a hybrid approach that combines the results from geo-data search in which the containing area is represented by reference geometry, with results from a SPARQL query that uses qualitative spatial (or implied spatial) properties.

In all the above, feature types may be selected from available types used with DBpedia.

The query processor performs tasks of query planning according to the nature of the user query, generation of a query footprint, retrieval of relevant URIs from the local spatial index of DBpedia content and formulation and execution of SPARQL queries to the external RDF store, followed by return of the result to the user interface.

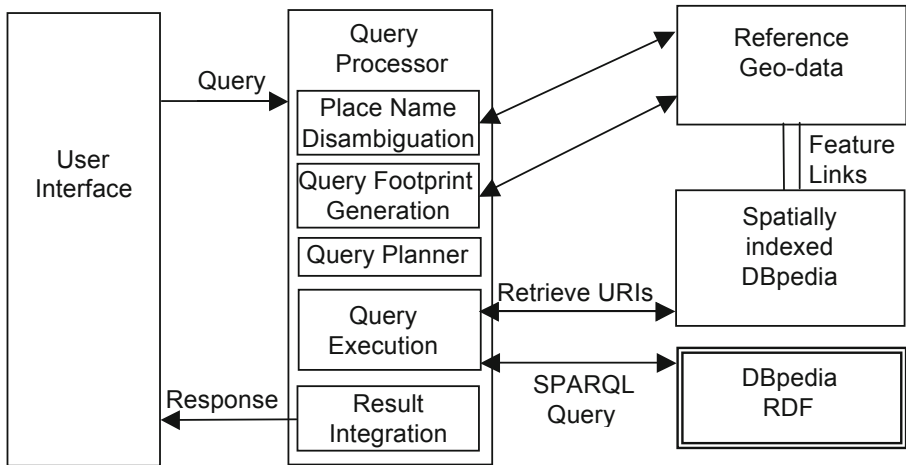


Fig. 1. Architecture of system for geo-spatial structured query of DBpedia

Query planning classifies a user query into one of three types:

1. A non-spatial query selecting by properties of places: requires SPARQL access to DBpedia only;
2. A proximity (within distance) and crossing/overlap query that selects relative to named geo-features: requires reference geo-data, and spatial access to local DBpedia index and may require a SPARQL query for DBpedia properties;
3. A containment query that selects within a named region: combines results from a spatial query that uses reference geo-data, with results of SPARQL queries using spatial containment properties. May require SPARQL query to access specialised properties of retrieved place instances.

The first situation is relatively trivial in that it requires only a SPARQL query to retrieve required properties of the named DBpedia instances. The second situation for proximity and overlap/crossing queries requires the creation of a query footprint based on the local reference geo-data store. Thus, for example, a query for DBpedia instances within a specified distance of a named river requires access to the river geometry from the local geo-data. This is then used to issue a query to the local

spatial index of DBpedia instances corresponding to the river objects, in combination with constraints on feature type (which is recorded in the index). If some specialised property of the retrieved places is required then a SPARQL query to DBpedia is performed using the URIs of relevant instances that were previously retrieved.

The third situation of a containment query relative to a named regional place (such as a city) will find results using both local spatial indexing of geo-referenced content and deduction of containment from appropriate DBpedia properties via a SPARQL query. The results are then merged, as there may be some duplication for instances that maintain both coordinates and the implied spatial properties. The use of the local geo-spatial data is similar to the previous strategy in that, having disambiguated the query place name, a boundary for the place is retrieved from the local geo-data if there is one. Our local geo-data is notable for maintaining boundaries for a large number of named settlements. The boundary is then used to perform a PostGIS query on the spatially indexed DBpedia content in combination with feature type constraints.

Identification of properties that specify or imply containment is a semi-automated process. For given feature classes that correspond to regions of space, such as a city, we select representative, well known place instances in DBpedia and retrieve all properties for which the respective place is the object of the property. The resulting properties are then filtered manually to remove those which do not in fact imply spatial containment, such as *dbpedia:birthPlace*. To perform a containment query the resulting list can be used to filter the results from a SPARQL query in which the object is the named place and the subject is constrained to a user specified type. Our strategy is to perform the filtering on the results of the SPARQL query, in the query processor, though it would also be possible to formulate a more complex SPARQL that included this filtering process.

The results from both types of containment query are then merged and if the query requires some other named property of the retrieved places then a further SPARQL query is executed to retrieve those properties of the previously found place instances.

The local store of reference geo-data consists in our experimental system of UK Ordnance Survey named features which are stored in a PostGIS spatial database that provides OGC compliant spatial queries for topological relations as well as distance (buffer) searches. When the query processor obtains a reference toponym from the user's query this must be matched to a name in the local reference geometry database (which may require disambiguation via the user interface). Using PostGIS we have a full set of OGC spatial operators and by maintaining the index locally we have faster response than if the reference geo-data needed to be retrieved from a spatially-enabled SPARQL endpoint.

The local store of georeferenced DBpedia content was obtained by performing SPARQL queries on DBpedia to access all places of particular types, which were filtered via their coordinates, where present, to confine much of the content for our experimental system to the British Isles. To do this requires knowledge of all feature classes of geographical instances. Our approach to this was to identify representative instances of different broad classes of place and to ascend the hierarchies of their

respective class properties in order to identify relevant parent classes. For each general category we then performed SPARQL queries to retrieve all instances of these parent classes and their children as illustrated by the following query for the category Museum.

```
define input:inference
"http://dbpedia.org/resource/inference/rules/yago#"
PREFIX yago:      <http://dbpedia.org/class/yago/>
PREFIX dbpo:     <http://dbpedia.org/ontology/>
SELECT DISTINCT ?s ?lat ?lon ?geom ?point
FROM <http://dbpedia.org>
WHERE {
  {?s a dbpo:Museum }
  UNION
  {?s a ?t . ?t rdfs:subClassOf dbpo:Museum}
  UNION
  {?s a yago:Museum103800563}
  UNION
  {?s a ?t . ?t rdfs:subClassOf yago:Museum103800563}
  OPTIONAL { ?s geo:lat ?lat}
  OPTIONAL { ?s geo:long ?lon}
  OPTIONAL { ?s geo:geometry ?geom}
  OPTIONAL { ?s grs:point ?point}
}
```

Note that the query uses both YAGO (Suchanek et al. 2007) and DBpedia ontology parent classes and retrieves coordinates, if present, in whatever form they may be stored.

The resulting data for inclusion in the local index, which includes the URI that contains the place name, are stored similarly to the local geo-data, within a PostGIS database. For the purposes of our experimental system the DBpedia coordinates were converted to the UK National Grid map projection (i.e. in metres) to match the native coordinates of the geo-spatial data. In order to scale the system to work globally all coordinates could be geodetic (latitude and longitude), which systems such as PostGIS can support for purposes of spatial query.

Our system is designed to maintain links between DBpedia objects and corresponding features in reference geo-data. These links are required for automated processing of queries where SPARQL queries are required to determine properties of DBpedia instances that are represented in the spatial aspect of the query by reference geometry. This occurs for example when querying the properties of linear or area objects that hold a specified spatial relationship to some other named feature. Currently implemented methods for matching are similar to those described by Stadler et al (2011) in their work on linking OSM to DBpedia, being a combination of name matching and distance between to the two geometries, of which the DBpedia

geometry will always be a point. In our method, place names are normalised before performing an exact match, as the use of methods such as Levenshtein edit distance was found to result in too many false positives. These methods will continue to be refined through exploitation of additional evidence such as feature type.

4 Example Queries and Experimental Results

In this section we present examples of the different sorts of query that can be processed using the various methods that we employ, specifically for proximity from lines and areas, crossing (overlap) and containment. The method of combining qualitative DBpedia properties that infer containment with quantitative methods based on reference geo-data are examined in more detail than the others in order to reveal the balance between the numbers of results produced by the two containment methods.

4.1 Non-spatial Queries to DBpedia

Here the user can specify a property, selected from a drop down menu, and a target location. To find such properties for the query “Find the Capital of the United Kingdom”, a simple SPARQL query is constructed as follows,

```
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX dbpo:   <http://dbpedia.org/ontology/>
SELECT ?o WHERE {dbpedia:$s dbpo:$p ?o}
```

where to \$s and \$p are substituted by 'United_Kingdom' and 'capital', respectively.

4.2 Proximity Queries

The use of detailed geo-data enables proximity queries on DBpedia geographic instances to be performed relative to point, line and area features representing geographical features named by the user in the query. In this procedure the reference feature named in the query is represented by the reference geo-data, while the subject of the query is represented geometrically either by the point coordinate geometry of georeferenced DBpedia features or reference geo-data geometry that has been matched to the DBpedia instance. Here will illustrate examples of both situations.

For the query “Find churches within 1km of the River Thames” (i.e. return the references to the relevant DBpedia instances), the user’s query term “River Thames” is represented by geometry from the reference geo-data, while the locations of the churches are those from DBpedia, found here via the local spatial index of DBpedia instances. Figure 2 illustrates a map of the results with some of the retrieved instances listed.

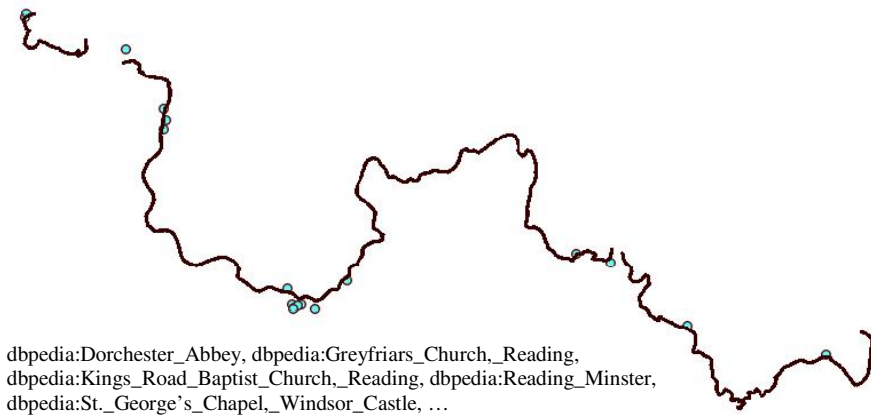


Fig. 2. Some results for a query to find DBpedia churches within 1km of the River Thames. Geometry data courtesy UK Ordnance Survey.

Another example of using point geometry from DBpedia in combination with reference geometry is for the query “Find hospitals outside and within 10km of the city of Cardiff”. Here the polygonal boundary of Cardiff is obtained from the reference geo-data and the point locations of the hospitals are obtained from the spatial index of DBpedia. The spatial query with PostGIS uses a combination of a distance constraint and a negation of inside in order to obtain locations that are outside the city. Figure 3 illustrates the results.



Fig. 3. Results for query for DBpedia hospitals outside and within 10km of the city of Cardiff. Geometry data courtesy UK Ordnance Survey.

4.3 Crossing Queries

We illustrate the use of reference geo-data to represent the retrieved DBpedia instances with the query “Find the mouths of the rivers that cross Oxford”. Here the PostGIS spatial database is used to find rivers that satisfy the spatial constraint, where the geometry of the rivers and the city boundary come from the reference geometry, while the corresponding DBpedia instances, that match the river names, are queried with SPARQL to find the mouths of the rivers. Automation of this query requires links between reference geo-data features and corresponding DBpedia instances as explained in Section 3.

River	Mouth
"Oxford Canal",	dbpedia:River_Thames,
"River Cherwell",	dbpedia:North_Sea,
"River Thames or Isis"	dbpedia:Thames_Estuary



Fig. 4. Results for query to find the mouths of the rivers that cross Oxford. Geometry data courtesy UK Ordnance Survey.

4.4 Containment Queries

Containment queries relative to named regions with known boundaries can be performed by a combination of quantitative and qualitative methods and, as indicated previously, in the absence of full quantitative geo-referencing of geographical DBpedia instances both methods are required to maximise the completeness of the response. In order to gain some insight into the balance between the use of coordinates and of properties that imply spatial containment, we selected ten UK cities (Bath, Birmingham, Bristol, Cardiff, Durham, Leeds, Liverpool, Manchester, Newport (South Wales), Nottingham) as the target of a set of queries to retrieve instances of the following eight feature types: Churches, Historic Buildings, Hospitals, Hotels, Libraries, Museums, Shopping Malls, and Stadiums. Spatial queries used the DBpedia coordinates to determine containment in city boundaries obtained from the reference geo-data, while qualitative containment was determined as explained previously, using all properties that have the respective city as object, subject to filtering via the predetermined list of non-containment properties. The containment properties (after filtering) that were used in this study are listed below:

Containment = {dbpo:district, dbpo:homeport, dbpo:location, dbpo:locationCity, dbpo:municipality, dbpo:owner, dbpo:principalArea, dbpo:region, dbpp:district, dbpp:city, dbpp:location, dbpp:locationTown, dbpp:mapCaption, dbpp:municipality, dbpp:owner, dbpp:parish, dbpp:postTown, dbpp:region}

Note that some of the results from the second method could duplicate instances in the local spatial index, where a DBpedia instance has both coordinates and spatial containment properties. Table 1 presents a summary of the results in which we report, for each place type, the numbers of instances found within the respective containing city using spatial containment relations (S) and geographic coordinates (G). For a set of instances retrieved using spatial containment relations (S), the sets SG, SG_i and SG_n are constructed by filtering the proportion of S with geographic coordinates (SG), then using the coordinates to test those which fall within the boundaries of the reference area (SG_i) and those which do not (SG_n).

Table 1. Per feature and per settlement area, number of features related to the area with spatial containment relations (S), those of S with geographic coordinates (SG), those of SG that fall within the boundaries of the reference area (SGI), those and that do not (SGN). G represents all features with geographic coordinates which fall within the boundary.

	Bath					Birmingham					Bristol					Cardiff					Liverpool				
	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G
churches	9	7	7	-	8	3	3	3	-	18	35	2	2	-	2	2	2	-	4	22	18	18	-	38	
historicbuildings	3	3	3	-	3	2	1	1	-	28	-	-	-	-	-	1	1	1	-	3	16	16	16	-	22
hospitals	2	1	1	-	1	4	3	3	-	13	10	7	6	1	9	3	3	3	-	4	3	3	3	-	6
hotels	1	1	-	1	1	3	3	3	-	7	1	1	-	1	-	2	2	2	-	3	3	3	3	-	9
libraries	-	-	-	-	-	2	2	2	-	3	1	1	1	-	1	-	-	-	-	1	-	-	-	-	1
museums	13	11	9	2	9	7	6	6	-	24	17	11	10	1	8	8	7	4	3	4	10	9	4	5	9
shoppingmalls	1	1	1	-	1	3	2	2	-	11	3	3	3	-	3	5	5	5	-	5	5	4	4	-	5
stadiums	2	2	2	-	2	10	9	8	1	13	2	2	2	-	2	12	12	12	-	12	4	4	3	1	6
theatres	5	5	5	-	5	6	6	6	-	13	5	3	3	-	5	3	2	2	-	5	3	3	3	-	7
Total:	36	31	28	3	30	40	35	34	1	130	74	30	27	3	30	36	34	31	3	41	66	60	54	6	103

	Durham					Newport					Leeds					Manchester					Nottingham				
	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G	S	SG	SGi	SGn	G
churches	-	-	-	-	2	1	1	1	-	1	6	5	4	1	7	2	2	2	-	33	14	13	13	-	46
historicbuildings	-	-	-	-	-	1	1	1	-	1	9	9	7	2	8	2	2	2	-	25	16	16	16	-	53
hospitals	3	2	2	-	2	2	2	1	1	1	5	5	5	-	5	5	5	5	-	18	3	1	1	-	2
hotels	-	-	-	-	-	1	1	1	-	1	2	2	2	-	4	5	2	2	-	6	-	-	-	-	-
libraries	1	1	1	-	1	-	-	-	-	-	1	1	1	-	2	2	1	1	-	6	-	-	-	-	-
museums	1	1	-	1	3	-	-	-	-	2	1	1	1	-	7	8	7	6	1	19	3	2	2	-	2
shoppingmalls	2	2	1	1	1	2	2	2	-	2	4	2	2	-	5	4	2	2	-	10	-	-	-	-	2
stadiums	-	-	-	-	-	4	4	3	1	3	4	4	4	-	4	9	8	8	-	27	4	3	3	-	3
theatres	-	-	-	-	-	-	-	-	-	1	2	2	2	-	2	13	7	7	-	16	4	3	3	-	3
Total:	7	6	4	2	9	11	11	8	3	12	34	31	28	3	44	50	36	35	1	160	44	38	38	-	111

Table 2. Totals for each category for all cities and associated percentages. |G/(G+S-SG)| are the instances retrieved using spatial coordinates and geographic boundary only, |SG/G| the percentage of instances with coordinates that also have spatial relation properties. |SG/S| the percentage of instances with spatial relations that also have coordinates. |SGn/SGi| the number of features that are linked by containment to the area but are not in the area according to our geo-data reference city boundary, and |SGi/G| the percentage of instances with coordinates that also have spatial relations and are within the city boundary.

	TOTALS										
	S	SG	SGi	SGn	G	G+S	G/(G+S-SG)	SG/G	SG/S	SGn/S	SGi/G
churches	94	53	52	1	159	200	79.5%	33.3%	56.4%	1.9%	32.7%
historicbuildings	50	49	47	2	143	144	99.3%	34.3%	98.0%	4.1%	32.9%
hospitals	40	32	30	2	61	69	88.4%	52.5%	80.0%	6.3%	49.2%
hotels	18	15	12	3	31	34	91.2%	48.4%	83.3%	20.0%	38.7%
libraries	7	6	6	-	15	16	93.8%	40.0%	85.7%	0.0%	40.0%
museums	68	55	42	13	87	100	87.0%	63.2%	80.9%	23.6%	48.3%
shoppingmalls	29	23	22	1	45	51	88.2%	51.1%	79.3%	4.3%	48.9%
stadiums	51	48	45	3	72	75	96.0%	66.7%	94.1%	6.3%	62.5%
theatres	41	31	31	-	57	67	85.1%	54.4%	75.6%	0.0%	54.4%
Total:	398	312	287	25	670	756	88.6%	46.6%	78.4%	8.0%	42.8%

It may be noted that there is considerable disparity in the proportion of instances of particular types that are found only due to coordinates and those found only due to qualitative properties, where the number of results using geographic coordinates are nearly always higher in this analysis (but for a notable exception see Bristol in Table 1). Using only geographic coordinates and a geographic footprint retrieves 88.6% of all results, the other 11.4% being provided by instances without geographic coordinates but with qualitative relations to the containment instance. When compared with pure SPARQL queries that use only qualitative relations, our method almost doubles, on average, the number of instances retrieved, i.e. 398 against 756. These results can therefore be regarded as providing strong validation for the benefits of combining both methods.

5 Conclusions and Future Work

We have presented the design of an experimental system designed to demonstrate hybrid methods for performing spatial query on Semantic Web resources such as DBpedia, with the intention to maximise the completeness of the answers with respect to finding relevant content and obeying user specified spatial operators. The approach is novel with regard to the combination of exploiting high quality geo-data and mixing quantitative and qualitative methods to obtain results. In addition to providing examples of how the high quality geo-data can be employed to find results based on a variety of spatial relations, we have demonstrated that, for the case of containment queries, the combination of quantitative geo-spatial query with qualitative query produces greatly superior results to the use of these methods in isolation. The work presented here is only concerned with the quality of the results. Performance issues, particularly timing, will be addressed in future work.

It may be noted that, even with the methods presented, some spatial instances will be omitted from the results, where these instances do not have either coordinates or explicit or implicit spatial property relations. Future work will focus on geo-referencing of these instances to enable them to be accessible via geo-spatial query. The present system employs a simple text-based user interface. This will be enhanced with map-based feedback. Reference geo-data will be extended to include resources such as OSM and Geonames and hence gain international coverage and improved disambiguation facilities. Our approach is intended to scale up to support access to multiple Semantic Web resources that provide geographical information. Thus the local index will be extended to include reference to other geographically informative RDF resources, which will be linked to reference geo-data, using similarity matching methods.

References

1. Buscaldi, D.: Resource Integration for Question Answering and Geographical Information Retrieval. Research project report, Polytechnic University of Valencia Valencia, Spain: The Department of Information Systems and Computation (2007), <http://users.dsic.upv.es/~proso/resources/BuscaldiIDEA.pdf>

2. Battle, R., Kolas, D.: GeoSPARQL: Enabling a Geospatial Semantic Web. Submitted Semantic Web Journal (2011), http://www.semantic-web-journal.net/sites/default/files/swj176_1.pdf
3. Brodt, A., Nicklas, D., Mitschang, B.: Deep integration of spatial query processing into native RDF triple stores. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (2010)
4. Della Valle, E., Qasim, H.M., Celino, I.: Towards Treating GIS as Virtual RDF Graphs. In: Proceedings of 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services (WebMGS (2010)
5. De Rouck, C., Van Laere, O., Schockaert, S., Dhoedt, B.: Georeferencing Wikipedia pages using language models from Flickr. In: Proceedings of the Terra Cognita Workshop (ISWC), pp. 3–10 (2011)
6. Hackley, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B*, 682–703 (2010)
7. Hahmann, S., Burghardt, D.: Connecting LinkedGeoData and Geonames in the Spatial Semantic Web. In: 6th International GIScience Conference (2010)
8. Hartrumpf, S., Leveling, J.: Recursive Question Decomposition for Answering Complex Geographic Questions. In: Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G. (eds.) CLEF 2009, Part I. LNCS, vol. 6241, pp. 310–317. Springer, Heidelberg (2010)
9. Katz, B., Lin, J.J.: Start and Beyond. In: Proceedings of 6th World Multiconference on Systemics, Cybernetics, and Informatics (2002)
10. Lin, J., Katz, B.: Question Answering Techniques for the World Wide Web. In: 11th Conference of European Association of Computational Linguistics, EACL 2003 (2003)
11. Luque, J., et al.: GeoVAQA: a voice activated geographical question answering system. IV Jornadas en Tecnología del Halba (2008), http://jth2006.unizar.es/finals/4jth_158.pdf
12. Mishra, A., Mishra, N., Agrawal, A.: Context-aware restricted geographical domain question answering system. In: International Conference on Computational Intelligence and Communication Networks, CICN 2010, IEEE, Washington (2010)
13. Mollá, D., Vicedo, J.L.: Question Answering in Restricted Domains: An Overview. *Computational Linguistics* 33(1) (2007)
14. Mooney, P., Corcoran, P., Winstanley, A.C.: Towards quality metrics for OpenStreetMap. In: Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (2010)
15. Purves, R.S., Clough, P., Jones, C.B., et al.: The design and implementation of SPIRIT : a spatially-aware search engine for information retrieval on the internet. *International Journal of Geographical Information Systems* 21(7), 717–745 (2007)
16. Salas, J.M., Harth, A.: Finding spatial equivalences across multiple RDF datasets. In: Terra Cognita 2011, co-located with ISWC, Bonn, Germany (2011)
17. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: LinkedGeoData: A core for a web of spatial open data. To appear in *Semantic Web Journal* (in press), http://svn.aksw.org/papers/2011/SWJ_LinkedGeoData/public.pdf
18. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web, pp. 697–706 (2007)
19. Waldinger, R., Appelt, D.E., Fry, J., Israel, D.J., Jarvis, P., Martin, D., Riehemann, S., Stickel, M.E., Tyson, M., Hobbs, J., Dungan, J.L.: Deductive Question Answering from Multiple Resources. In: *New Directions in Question Answering*. AAAI (2004), <http://www.ai.sri.com/pubs/files/986.pdf>