Finding spatial equivalences across multiple RDF datasets

Juan Martín Salas[1] and Andreas Harth[2]

[1] FRLP, Universidad Tecnológica Nacional, Argentina juan.salas@ieee.org
[2] Institute AIFB, Karlsruhe Institute of Technology (KIT), Germany harth@kit.edu

Abstract. The importance of geospatial information is being reflected on the growing amount of spatial datasets on the Semantic Web. However, the high variability of the data presents challenges for integration. In this paper, we address the problem of finding spatial equivalences between geospatial RDF datasets. First, we present mappings between our NeoGeo vocabulary and the vocabularies used by some well-known spatial RDF datasets. Second, we describe a method to find spatially co-located features across spatial RDF datasets. To find equivalences, we rely on analyzing the Hausdorff distance distribution in the compared datasets, with the objective of finding a sensible criterion that aids the recognition of equivalent regions.

1 Introduction

Geospatial data is ubiquitous in information management, supporting scientific, industrial or just everyday activities. The relevance of geospatial data is reflected by the growing amount of geospatial datasets on the Web.

A feature is an abstraction of a real world phenomenon (e.g. a building, a mountain or an administrative region). A geographic feature is a feature associated with a location relative to the Earth, which is usually represented by a certain geometric shape (e.g. a point, a curve or a polygon). Features that are spatially co-located (i.e. share the same location) are not necessarily always the same. However, finding spatially co-located regions is a powerful measure of similarity between features.

Factors like rounding effects, different scales and different formats, present a challenge when attempting to elicit equivalences between geospatial resources. We define a method for obtaining a criterion that best fits the differences between the datasets merged.

This work was successfully applied to integrate our RDF representations of the NUTS nomenclature of the European Union [3] and of the GADM project [4] to other datasets describing spatial information on the Semantic Web, and also between each other.

Our contributions are as follows:

– Representation and modeling of datasets: we survey the representation of existing geospatial datasets, and distill an integration vocabulary which covers the

[3] http://nuts.geovocab.org/
[4] http://gadm.geovocab.org/

core set of classes and properties in existing data. We also integrate existing vocabularies and publish two geospatial datasets (Section 2).
– Integration and mapping of multiple datasets: we develop an algorithm for finding equivalences for geometric shapes across multiple datasets (Section 3).
– Evaluation of the presented approach: we conduct experiments in which we evaluate the accuracy of the results (Section 4).

We discuss the related work in Section 5. Finally, we identify areas for future work and conclude in Section 6.

2 Representing geospatial data on the web

2.1 Analyzed datasets

We start with providing a brief summary of the analyzed datasets.

– UN FAO Geopolitical Ontology[5]: The Food and Agriculture Organization of the United Nations (FAO) is a specialized agency of the UN. The UN FAO Geopolitical Ontology provides the FAO and its associated partners with a master reference for geopolitical information.
– OS OpenData[6] [8]: The Ordnance Survey (OS) is the national mapping agency for Great Britain. OS has released a number of its products as Linked Data.
– GeoLinkedData.es [2, 3]: The initiative provides geospatial information about the national territory of Spain. The information provided as RDF at their website is gathered from different national sources. However, the integration process is based on string matching.
– LinkedGeoData.org [22]: The project provides data from OpenStreetMap[7] as Linked Data.
– GeoNames.org: GeoNames is a geographical database that covers all countries and provides Linked Data under a Creative Commons attribution license.
– Uberblic.org: Uberblic provides an integration service that includes data from GeoNames, Wikipedia, MusicBrainz, Freebase, Last.fm and Foursquare.
– RAMON NUTS[8]: The Nomenclature of Units for Territorial Statistics (NUTS) is a geocoding standard for referencing the subdivisions of countries for statistical purposes developed by the European Union and published as Linked Data.
– DBpedia.org: The community effort extracts structured information from Wikipedia for publication as Linked Data.
– NeoGeo: We provide an integration vocabulary, described in more detail in Sections 2.4 and 2.5.

[5] http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/
[6] http://data.ordnancesurvey.co.uk/
[7] http://openstreetmap.org/
[8] http://rdfdata.eionet.europa.eu/ramon/nuts2008/

2.2 Representing location

The analyzed spatial datasets represent the location of features in different ways. We identified four main kinds of representation: point, bounding box, points in lists, points using a single property and literals. Geometric shapes are not only described using different vocabularies, but also these vocabularies are based on different structures, which increases the difficulty of working with GeoData across datasets.

- **Point** Location of objects is merely represented by a geographic point. The most common vocabulary to do so is W3C Geo[24], sometimes complemented with a GeoRSS representation [21], such is the case of the UK Ordnance Survey, even if GeoRSS is not a proper RDF vocabulary but an XML-Schema. In some cases, neither W3C Geo nor GeoRSS is used, but an own vocabulary, as is the case of the Uberblic Ontology, which uses its own "latitude", "longitude" and "altitude" predicates.
- **Bounding box** The location is represented by two points or four line segments forming a georeferenced rectangle (on cylindrical projections). This is the case of the FAO Geopolitical Ontology, which uses four predicates (hasMinLongitude, hasMinLatitude, hasMaxLongitude, hasMaxLatitude) to represent a rectangle. The rectangle is represented by line segments, which should be tangential to the region at some point.
- **Points in lists** The geometric shape of a region is represented by a collection of points, each being described as a single RDF node. The whole collection of points is then linked together using either an RDF Collection or an RDF Container. LinkedGeoData.org represents geometric shapes by using a "hasNodes" object property, which links to a rdf:Seq container. The rdf:Seq container describes the nodes of a shape, which are represented using the W3C Geo Vocabulary.
- **Points using a single property** In the GeoLinkedData.es ontology, rivers are represented by a group of "Curva" (curve) RDF resources (similar to a GML LineString). "Curva" resources use a single "formadoPor" object property to link each of their nodes, which at the same time contain the WGS-84 coordinates (represented with the W3C Geo Ontology) and an "orden" (order) value property, defining the position of each node within the geometric shape.
- **Literals** Both Ordnance Survey and GeoLinkedData.es (for rivers only) ontologies include a predicate allowing to include a GML representation of the geometric data, which is coded in RDF as a literal. A "geometry:extent" property links a feature to its geometric representation.

2.3 Representing spatial relations

A spatial relation states the location of an object in relation to another. We created a set of vocabulary mappings to the NeoGeo vocabulary using the `rdfs:subPropertyOf` predicate. Table 1 shows which predicates are used in each dataset to describe spatial relations.

Dataset	Disjoint	Touches	Overlaps	Within	Contains	Equals	Nearby
UN FAO		hasBorder-With		isInGroup			
Ordnance Survey	disjoint	touches	partially-Overlaps	within	contains	equals	
GeoLinkedData.es				forma-ParteDe	formado-Por		
LinkedGeoData.org							
GeoNames.org		neighbour / neigh-bour-ingFea-tures		parent-Feature	children-Features		nearby / nearby-Features
Uberblic.org		adjoining-_location		containing-_location			
RAMON NUTS				partOf			
DBpedia.org				locatedIn-Area			
NeoGeo	DC	EC	PO	PP	PC	EQ	

Table 1: Equivalent properties for spatial relations across multiple vocabularies.

2.4 NeoGeo ontologies

Given the lack of a standardized spatial vocabulary, we developed our own set of spatial ontologies, which we call NeoGeo[9]. We manually created a set of mappings between our vocabularies and the vocabularies used by some acknowledged spatial datasets like the Ordnance Survey and LinkedGeoData.org. Both the GADM and NUTS datasets use the NeoGeo vocabularies.

The Geometry Vocabulary[10] is an RDF vocabulary for the description of geo-referenced geometric shapes. It is based on the Core Profile of the Spatial Schema [12] and the General Feature Model [11]. Hopefully, the lack of a standardized RDF vocabulary in this domain will probably be addressed by GeoSPARQL[16] shortly. For experimentation reasons, the Geometry vocabulary allows to encode geometric shapes in a representation fully based on RDF or as a WKT representation [10] embedded in an XMLLiteral.

The Spatial Ontology[11] provides a vocabulary for the representation of the spatial relations used in the Region Connection Calculus (RCC) [19]. It also provides monotonic reasoning by mapping most of the semantics of RCC into OWL.

2.5 NeoGeo datasets

We provide two datasets, NUTS and GADM, containing geospatial information as Linked Data.

[9] `http://geovocab.org/`
[10] `http://geovocab.org/geometry`
[11] `http://geovocab.org/spatial`

The Nomenclature of Territorial Units for Statistics (NUTS) is a classification defined by the Eurostat office of the European Union. It is intended to divide the administrative regions of the European Union, in a way that the resulting regions are demographically equivalent.

The RDF representation of the NUTS nomenclature contains a 1:60,000,000 geospatial representation of the NUTS statistical units mapped to RDF. The resources representing NUTS regions in our dataset include (among others) links to resources in DBpedia, FAO Geopolitical Ontology, GeoNames, Ordnance Survey and GeoLinkedData.es.

The Global Administrative Areas (GADM)[12] is a project seeking to become a collaborative effort on building a spatial database containing information about all of the administrative regions in the world. GADM aims to provide high resolution mappings for all administrative areas in the world, along with additional information about them. The latest version of GADM (0.9) maps 226.439 administrative areas. The information can be downloaded at their website in the following formats: Shapefile, ESRI geodatabase, RData and KMZ.

Given the value of the GADM project, we have created an RDF representation of the information contained in the original GADM project, which we seek to enrich with additional capabilities like the materialization of spatial relations, mappings to other RDF datasets and SPARQL querying support.

3 Instance mappings

The alignment of the vocabularies is only the first step for the integration of the datasets. The second step in the process is to find matchings between the features.

We can classify the features into three general categories, in relation on how their location is represented in the datasets. First are the resources that present no quantitative spatial information at all. Second, the features that approximate their location by using only a single point. And finally, features which present rich information about their location (i.e. include a description of their geometric shape).

3.1 Resources with no quantitative geospatial information

Resources which include no quantitative information about their location can be integrated by relying either on text matching [14], or object property matching [23] [4]. These techniques are also suitable for the disambiguation of spatially obtained mappings [9]. This kind of resources is not the topic of this work.

3.2 Resources with poor quantitative geospatial information

Sometimes the location of a feature is approximated by using a single point (e.g. using the W3C Geo vocabulary) instead of representing its actual extent (i.e. the geometric shape). Examples of this kind of representation are DBpedia, LinkedGeoData.org, GeoNames and LinkedGeoData.es.

[12] `http://gadm.org/`

This kind of representation can lead to false assertions while performing comparisons against a spatial index if these features are not especially considered. For example, DBpedia uses the W3C Geo Vocabulary to describe the latitude and longitude coordinates of features as points. The resource for Germany in DBpedia `http://dbpedia.org/resource/Germany` is spatially represented by a point with latitude 52.516666 and longitude 13.383333. If we intended to obtain the containment relations for such resource by comparing it with a spatial index, the result would be that Germany is part of Berlin, which is false. Therefore, even though these relations can be obtained using the coordinates represented in DBpedia, first it is necessary to ensure that the process will not return such false statements. The false matches can be avoided, for example, by filtering the features that will be compared by its class, in a way that ensures that the feature will be properly contained in the features that it will be compared to (e.g. cities in provinces or restaurants in cities).

3.3 Resources with rich quantitative geospatial information

Resources that include an accurate description of their extent as a geometric shape can be compared using this information. We will focus on obtaining links between spatially co-located regions (we use the `spatial:EQ` predicate). Whether `owl:sameAs` links can be deduced from the obtained links depends on the modeling of the datasets (e.g. the class the resource belongs to).

To perform the comparison, we will adopt a Plate Carrée projection for both of the compared datasets. Being this projection equirectangular, we can treat latitude and longitude coordinates as if they were cartesian. Therefore, the units will be presented in centesimal degrees.

The benefit of using an equirectangular projection is that it simplifies the calculations by avoiding local reprojections (e.g. to UTM), and also allows to use a global spatial index, improving the performance of the process. In our approach it is not important if the projection distorts the real size or the actual geometric shapes on the surface of the Earth, as long as the geometric data is equally distorted for all datasets.

Due to a series of factors like rounding effects and different scales, there is no guarantee that both geometric shapes will be vertex by vertex identical. Figure 1 exemplifies these differences by showing the boundaries for Luxembourg as they are represented in the GADM and NUTS datasets.

An effective method of determining how similar two geometric shapes are is to compute the Hausdorff distance between them. The Hausdorff distance is the "maximum distance of a set to the nearest point in the other set" [20]. More formally, given two sets of points $A = \{a1, a2, ..., a_n\}$ and $B = \{b1, b2, ..., b_n\}$, the Hausdorff distance is defined by:

$$d_H(A, B) = max\left(\{\arg\max_{a \epsilon A} \arg\min_{b \epsilon B} d(a, b), \arg\max_{b \epsilon B} \arg\min_{a \epsilon A} d(a, b)\}\right)$$

It can be deduced from the formula that in the particular case of calculating the Hausdorff distance between points, the Hausdorff distance matches the Euclidean distance $d(a, b)$.

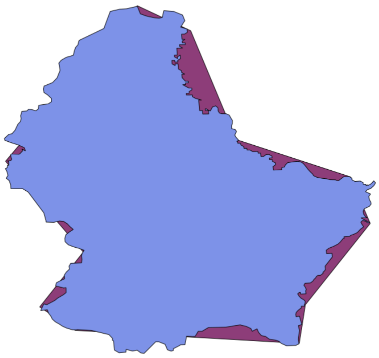

Fig. 1: Incongruency of the geometric data (GADM: blue, NUTS: violet) due to differences in resolution.

Figure 2 shows the values of correct and wrong guesses for similar regions in both datasets. In order to better appreciate the variability of the values, only small areas are plotted in the chart. From the figure it can be deduced that smaller regions (e.g. boroughs) require greater precision than larger regions (e.g. countries), in order to differentiate them from each other. Therefore, the Hausdorff distance margins allowed for regions which are suspected to be spatially co-located must be different, depending on the area size of the regions being compared.

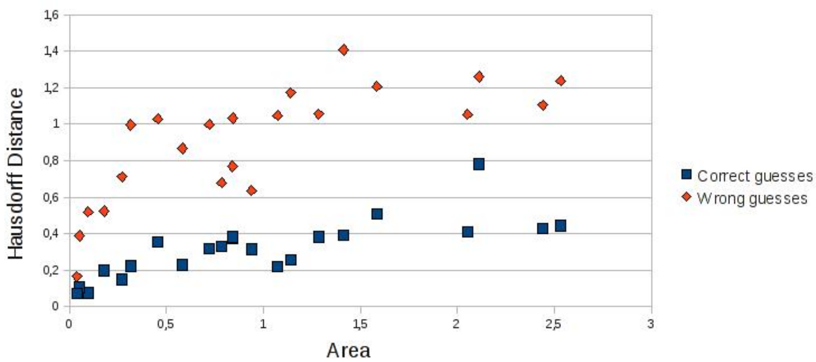

Fig. 2: Values for correct and wrong guesses for similar regions in NUTS and GADM.

To address this issue, it is desirable to obtain a function for a Hausdorff distance threshold for a given area size. In order to do this, first we calculate the midpoint between the lowest and second lowest Hausdorff distance values, for a representative set of features in both datasets. Afterwards we perform a quadratic regression from the midpoint Hausdorff distance values. This produces a formula that allows to determine the maximum Hausdorff distance allowed between two regions, in order to consider them similar. The resulting function has the following form:

$$MaxHDist(x) = A \cdot x^2 + B \cdot x + C$$

Where MaxHDist is the maximum Hausdorff distance allowed between two regions in order for them to be considered similar. The x variable is the area of the region. The quadratic function gives more precision for small regions while allowing a greater margin for large regions. The A,B and C constants are tunable parameters for the integration procedure.

A yet unresolved issue of using a quadratic function is that the samples must include values for the approximated maximum area for which the integration will be performed. This is because the values of the function will tend to decrease after reaching a maximum value. We are performing experiments with logarithmic functions to solve this issue.

Table 2a shows sample execution times and Hausdorff distance values between features in the NUTS and GADM datasets.

Region Name	NUTS Id	Area	Hausdorff Distance	Time (ms)
Finland	FI	62.2835	1.3996	30353
Iceland	IS	19.3357	0.4163	567
Croatia	HR	6.2139	1.1374	7830
Schleswig-Holstein	DEF	2.1126	0.7281	1870
Karlsruhe	DE12	0.8433	0.1062	35
Seine-Saint-Denis	FR106	0.0358	0.0812	1

(a) for the original geometric shapes

NUTS Id	Hausdorff Distance	Time (ms)
FI	1.3483	2504
IS	0.4613	66
HR	1.1366	1108
DEF	0.7257	296
DE12	0.1906	13
FR106	0.0716	2

(b) simplified with a separation of 0.2 degrees

NUTS Id	Hausdorff Distance	Time (ms)
FI	1.3483	2257
IS	0.4863	49
HR	1.1366	1053
DEF	0.7801	278
DE12	0.3762	14
FR106	0.0716	2

(c) simplified with a separation of 0.5 degrees

Table 2: Sample Hausdorff distance values and execution times

Calculating the Hausdorff distance between the original geometric data is quite expensive, especially for large regions. In order to increase the performance of the process, as an optional step, we chose to simplify the geometric shapes using the

Ramer-Douglas-Peucker algorithm [18] [5], prior to the calculation of the Hausdorff distances.

The Ramer-Douglas-Peucker algorithm starts by considering a line segment between the first and last points of the line. Then, it finds the furthest point from the line segment between the first and last points of the line. If the point found is closer than a predefined distance ε to the line segment, all other points that were not chosen to be used in the solution can be discarded. If the point furthest from the line segment is greater than ε, then the point is used in the solution. The algorithm then calls itself recursively with the found point and the last point as parameters.

Tables 2b and 2c show the Hausdorff distance between the NUTS regions and their matching GADM region, as well as execution times for different levels of simplification. As it can be seen, execution times are dramatically reduced, especially for large regions.

A further refinement of the process is to calculate the simplification distance for the Ramer-Douglas-Peucker algorithm depending on the Hausdorff distance threshold and therefore of the area of the regions. This is based on the same principle applied for the Hausdorff distance, where small areas require greater precision than large areas.

Given two spatial datasets A and B, the algorithm can be summarized as Algorithm 1.

4 Experiments

4.1 Implementation

We implemented the algorithm presented in Section 3.3 using the PostGIS 1.5.2 extension running on PostgreSQL 8.4.8. The computer is running on Ubuntu 10.04 on an Intel SU7300 processor with 4GB DDR3 RAM.

PostGIS includes the ST_HausdorffDistance function, which implements an approximation to the original algorithm. This approximation can be thought of as the "Discrete Hausdorff distance", which is the Hausdorff distance restricted to discrete points for one of the geometric shapes. If more precision is needed, the function receives also an optional "denistyFrac" parameter which performs a segment densification before computing the discrete Hausdorff distance.

Since we are not concerned about the actual Hausdorff distance values, but just use it as a measure to determine if two regions are similar enough to be considered spatially co-located, this approximation is sufficient.

For the simplification of geometric data we will use the ST_SimplifyPreserveTopology function included in PostGIS. This function is a refined version of ST_Simplify, which is based on the Ramer-Douglas-Peucker algorithm [18] [5].

The query used with PostGIS to find regions which are supposed to be spatially co-located is very similar to the one presented below. To avoid having to perform the same calculations repeatedly, the values of the maximum Hausdorff distance function are cached into the "max_hausdorff_dist" column. The "geometry" column in both tables belongs to the "Geometry" datatype provided by PostGIS.

input : Datasets A, B
output: Equivalent regions from A and B
Convert the compared resources to a shared coordinate reference system.
Project the data into an equirectangular projection.
Obtain a representative set of regions in dataset A which intersect regions in dataset B and have a maximum arbitrary Hausdorff distance between each other.
foreach *region a of a representative set of regions in dataset A* **do**
  Get the minimum Hausdorff distance to a region in dataset B.
  Get the second minimum Hausdorff distance to a region in dataset B.
  Calculate the midpoint between the minimum and second minimum Hausdorff distances.
end
Perform a regression on the midpoints between the Hausdorff distances to calculate the Hausdorff threshold function.
foreach *region a in A* **do**
  foreach *region b in B* **do**
    if *a intersects b* **then**
      Calculate the Hausdorff distance between a and b.
      if *Hausdorff distance between a and b is lower than the threshold for the area of a* **then**
        a and b can be considered as spatially co-located.
      end
      else
        a and b cannot be considered as spatially co-located.
      end
    end
  end
end

Algorithm 1: Matching algorithm

```
SELECT g.gadm_level, g.gadm_id, n.nuts_id
  FROM nuts n INNER JOIN gadm g ON (n.geometry && g.geometry)
  WHERE
    n.shape_area BETWEEN (g.shape_area*0.9) AND (g.shape_area*1.1)
    AND ST_HausdorffDistance(ST_SimplifyPreserveTopology(n.geometry, 0.5),
ST_SimplifyPreserveTopology(g.geometry,0.5)) < g.max_hausdorff_dist;
```

Basically this query selects the identifiers for the GADM region (level and id), and for the NUTS region (id). The && operator matches an intersection between the bounding boxes of the of the regions. Since similar regions will also have a similar area size, the first condition in the "where" clause filters regions that have a similar area size with an error of 10%. The second condition checks if the discrete Hausdorff distance between the simplified geometric shapes is within the limit calculated by the function presented in Section 3.3.

4.2 Evaluation

We can analyze the effectiveness of the method by looking at the results of the process of finding spatial equivalences between the NUTS and GADM datasets.

The NUTS dataset codes the geometric shapes fully in RDF using the NeoGeo vocabulary, and the coordinate system used is WGS-84. The data is retrieved by using a Construct SPARQL query and then converted into WKT using XSLT.

After retrieving the geometric data, it is merged with the GADM dataset using the method presented in Section 3.3.

Not all NUTS regions are expected to match a GADM region, since many NUTS regions represent parts or aggregations of administrative boundaries. Also a GADM administrative region in a certain level should be able to match different NUTS regions in different levels, and vice-versa.

From the existing 1,671 NUTS regions of the 2008 nomenclature that were included in the comparison, the algorithm detected 965 matches, from which 13 were false positives, as Table 3 shows.

NUTS Region	Incorrect guess	GADM Id	GADM Level	Area	Hausdorff Distance
UKM34	East Renfrewshire	14084	2	0.0214	0.1862
FR106	Val-De-Marne	13799	2	0.0334	0.1644
BE321	Soignies	2691	2	0.0654	0.3521
BE353	Thuin	2692	2	0.1188	0.2834
CH061	Aargau	531	1	0.1672	0.3653
LT	Latvija	136	0	9.5204	2.5098
LI	Appenzell Innerrhoden	533	1	0.0205	0.2783
UKM28	North Lanarkshire	14095	2	0.0689	0.3478
BE331	Lige	2696	2	0.1013	0.335
BE353	Thuin	2692	2	0.1188	0.2834
CH061	Aargau	531	1	0.1672	0.3653
SE3	Norge	168	0	60.585	7.8658
BE321	Soignies	2691	2	0.0654	0.3521

Table 3: False positives resulting on the application of the method

These false positives are due to the fact that the threshold is set too high for very small and very large areas. It is desirable to produce a larger gradient for small areas and a smaller Hausdorff distance threshold for large areas. This is still a matter for further research.

5 Related work

The problem of aligning spatial datasets is not new in the Semantic Web community and much work has been put on finding sensible solutions both at T-Box and A-Box level.

Ontology alignment is a heavily researched topic. Proposed solutions have been based on the terminological [15], structural [7], semantic [6] and extensional [17] aspects of the aligned ontologies. The last two works consider the alignment of spatial T-Boxes in particular.

Algorithms have also been proposed for feature matching across spatial datasets. [1] presents a series of algorithms for the integration of features, for which the location is approximated by a single point. These approaches have the problems exposed in Section 3.2 and therefore, are not suitable for all cases. [13] considers feature matching as an assignment problem based on a minimization of the Hausdorff distance between the geometric shapes. However, being this a case of Linear Programming, the method can only be applied for all the geometric shapes of both datasets at the same time, making it more difficult to integrate to live crawling.

6 Conclusion

We have presented a generic method that can be used to map multiple spatial datasets. We also showed its functioning by describing the integration between our two spatial datasets and analyzed its results.

Although the method has been used successfully to align the GADM and NUTS datasets, the false positive rate can still be improved when analyzing regions covering a wide spectrum of area sizes. However, the presented method has proven to be usable in Semantic Web applications.

Since the first experiments showed promising results, we are developing a tool that automates the whole mapping process. We are also further refining the algorithm to improve its precision and performance.

Acknowledgements

The authors acknowledge the support of the European Commission's Seventh Framework Programme FP7/2007-2013 (PlanetData, Grant 257641).

References

1. C. Beeri, Y. Kanza, E. Safra, and Y. Sagiv. Object fusion in geographic information systems. In *Proceedings of the thirtieth international conference on very large data bases. Volume 30*, pages 816–827. VLDB Endowment, 2004.
2. L.M.V. Blázquez, B. Villazón-Terrazas, V. Saquicela, A. de León, O. Corcho, and A. Gómez-Pérez. Geolinked data and inspire through an application case. *ACM SIGSPATIAL GIS*, pages 446–449, 2010.
3. A. de León, V. Saquicela, L.M. Vilches, B. Villazón-Terrazas, F. Priyatna, and O. Corcho. Geographical Linked Data: a Spanish use case. In *Proceedings of the 6th International Conference on Semantic Systems*, pages 1–3. ACM, 2010.
4. X. Dong, A. Halevy, and J. Madhavan. Reference reconciliation in complex information spaces. In *Proceedings of the 2005 ACM SIGMOD international conference on management of data*, pages 85–96. ACM, 2005.
5. D.H. Douglas and T.K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, 10(2):112–122, 1973.
6. M. Dube and M. Egenhofer. Establishing similarity across multi-granular topological–relation ontologies. *Quality of Context*, pages 98–108, 2009.

7. J. Euzenat. An API for ontology alignment. *The Semantic Web–ISWC 2004*, pages 698–712, 2004.
8. J. Goodwin, C. Dolbear, and G. Hart. Geographical Linked Data: The administrative geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12:19–30, 2008.
9. F. Hakimpour, B. Aleman-Meza, M. Perry, and A. Sheth. Spatiotemporal-thematic data processing for the Semantic Web. *The Geospatial Web*, pages 79–89, 2007.
10. J.R. Herring. OpenGIS® Implementation Specification for Geographic Information - Simple Feature Aaccess - Part 1: Common Architecture. *Open Geospatial Consortium*, 2006.
11. International Organization for Standardization. ISO 19101. Geographic information Reference model, 2002.
12. International Organization for Standardization. ISO 19137.Geographic information Core profile of the spatial schema, 2007.
13. L. Li and M.F. Goodchild. Optimized feature matching in conflation. In *Geographic Information Science: 6th International Conference, GIScience 2010, Zurich, Switzerland, September 14-17, 2010. Proceedings*, 2010.
14. X. Li, P. Morie, and D. Roth. Semantic integration in text: From ambiguous names to identifiable entities. *AI magazine*, 26(1):45, 2005.
15. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the IEEE CS International Conference on Data Engineering*, pages 117–140. IEEE, 2002.
16. Open Geospatial Consortium Inc. GeoSPARQL - A geographic query language for RDF data, 2011.
17. Rahul Parundekar, Craig A. Knoblock, and José Luis Ambite. Aligning Ontologies of Geospatial Linked Data. In *Proceedings of the Workshop on Linked Spatiotemporal Data*, 2010.
18. U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3):244–256, 1972.
19. D.A. Randell, Z. Cui, and A.G. Cohn. A spatial logic based on regions and connection. *KR*, 92:165–176, 1992.
20. Günter Rote. Computing the minimum Hausdorff distance between two point sets on a line under translation. *Information Processing Letters*, 38:123–127, 1991.
21. Raj Singh, Ron Lake, Josh Liberman, Mikel Maron, and Carl Reed. An Introduction to GeoRSS: A Standards Based Approach for Geo-enabling RSS feeds. 2006.
22. Claus Stadler, Jens Lehmann, Konrad Höffner, and Sören Auer. LinkedGeoData: A Core for a Web of Spatial Open Data, 2011.
23. S. Tejada, C. Knoblock, and S. Minton. Learning object identification rules for information integration. *Information Systems*, 26(8):607–633, 2001.
24. World Wide Web Consortium (W3C) - Semantic Web Interest Group. W3C Geo Vocabulary. `http://www.w3.org/2003/01/geo/`, 2003.