Automatic Integration of Spatial Data into the Semantic Web

Claire Prudhomme[1,2], Timo Homburg[1], Jean-Jacques Ponciano[1], Frank Boochs[1], Ana Roxin[2] and Christophe Cruz[2]

[1]*Mainz University of Applied Sciences, Lucy-Hillebrand-Str. 2, 55128 Mainz, Germany*

[2]*Le2i FRE2005, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, Bâtiment i3M rue Sully, 21500 Dijon, France*

Keywords: Geospatial Data, Linked Data, Natural Language Processing, Ontology, R2RML, SDI, Semantic Web, Semantification.

Abstract: For several years, many researchers tried to semantically integrate geospatial datasets into the semantic web. Although, there are many general means of integrating interconnected relational datasets (e.g. R2RML), importing schema-less relational geospatial data remains a major challenge in the semantic web community. In our project SemGIS we face significant importation challenges of schema-less geodatasets, in various data formats without relations to the semantic web. We therefore developed an automatic process of semantification for aforementioned data using among others the geometry of spatial objects. We combine Natural Language processing with geographic and semantic tools in order to extract semantic information of spatial data into a local ontology linked to existing semantic web resources. For our experiments, we used LinkedGeoData and Geonames ontologies to link semantic spatial information and compared links with DBpedia and Wikidata for other types of information. The aim of our experiments presented in this paper, is to examine the feasibility and limits of an automated integration of spatial data into a semantic knowledge base and to assess its correctness according to different open datasets. Other ways to link these open datasets have been applied and we used the different results for evaluating our automatic approach.

1 INTRODUCTION

Integration of heterogeneous datasets is a persisting problem in geographical computer science. Many classical GIS approaches exist making use of relational databases to achieve a tailormade integration of geospatial data according to the needs of the current task. In the SemGIS project we are aiming at integrating heterogeneous geodatasets into a semantic web environment to take advantage of the flexibility of semantic data structures and to access a variety of related datasets that are already available in the semantic web. We intend to use the so-formed geospatial knowledge base in the application field of disaster management in order to predict, mitigate or simplify decision making in an event of a disaster. As in our project we are facing a large number of heterogeneous geodatasets of which we often do not know the origin nor intention nor the author and therefore lack an appropriate domain expert to help us understand data fields, we as non-domain experts are be left with a manual integration approach of said data. Dataset descriptions, if available, are often in natural language only which may give us hints but are hard to process in general and contain often hard to resolve ambiguities. However, despite mentioned obstacles we believe that a at least rudimentry classification and interlinking of our given datasets by means of the data values and data descriptions, is feasible. To achieve this goal we will in this paper present our fully-automated approach to analysing and integrating geospatial datasets into the current semantic web ecosystem and we will highlight its success as well as shortcomings of its various steps.

2 STATE OF THE ART

Our work is based on Natural Language Processing which is a major field of computational linguistics. In particular our approach is using but not limited to Language Recognition approaches to recognize keywords and terms in our datasets to relate them to already existing concepts in our semantic web knowledge base. We mostly rely on the following es-

Prudhomme, C., Homburg, T., Ponciano, J-J., Boochs, F., Roxin, A. and Cruz, C.
Automatic Integration of Spatial Data into the Semantic Web.
DOI: 10.5220/0006306601070115
In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017)*, pages 107-115
ISBN: 978-989-758-246-2

tablished techniques of the NLP community: Part-Of-Speech Tagging on limited language resources in common languages, the usage of several versions of Wordnet(Miller, 1995) along with BabelNet(Navigli and Ponzetto, 2010) and the multilingual labels of ontologies like DBPedia (Auer et al., 2007) and Wikidata (Vrandečić and Krötzsch, 2014). In addition we try to enrich our results with traditional reverse geocoding methods (Guo et al., 2009), which have been established for many years in the geospatial community.

2.1 Related Work

Related work has been done on the integration of interconnected database table and published by the W3C as the R2RML[1] standard. This standard automatically creates a local ontology of a given database schema once a given mapping is provided. Further research has been conducted by (Bizid et al., 2014) in which they use GML schemas to convert GML datasets to local ontologies and provide automated interlinking strategies for similarly structured database resources. In contrast to our work, this work focuses on similar datasets of a similar format only and does therefore not take into account a wider range of possible input formats.

In a more general context, the task we are approaching can be seen as an interlinking or a link discovery task in a specific geospatial domain, whereas we try to link a generated local ontology to existing resources in the semantic web and try to identify concepts that represent the contents of our respective datasets. (Nentwig et al., 2015) gives an overview on tools in the link discovery domain.

2.2 Related Projects

Among the different tools already available, some projects, as the Karma project (Knoblock et al., 2012) or the Silk Framework (Volz et al., 2009) have a semi-automatic linking approach for a variety of domains with a geospatial reference. In contrast to our work most of these tools described in the references above require human assistance in the form of experts or administrators of the corresponding databases. Another project very close to ours, is the Datalift project (Scharffe et al., 2012). This project allows for taking many heterogeneous formats (databases, CSV, GML, Shapefile,...) as Input in order to convert them into RDF and link them to the semantic web. Their goal is exactly the same, as ours but the approach not. Their approach is in two steps. The first is to convert the input format to raw RDF, which means the creation of triples with subject which corresponds to an element of a row, a predicate which has the same name as the column, and the object which is the content of the cell corresponding to the intersection between the row of the subject and the column of the predicate. The second step is to convert these raw RDF triples in a ”well-formed” RDF according to a choosen vocabularies. This second step is done thanks to the use of a SPARQL Construct queries. We have tested this approach on one of our shapefile to see what means ”well-formed” RDF and see if we can compare it with our approach. The content of the raw RDF triples was changed in an annotation of the element. That is why, we cannot compare our approach with this approach, we have no result concept to compare. We however want to examine if a predefined process can deliver reasonable results for the geospatial domain in a fully automated fashion. Some other projects, as Logmap (Cuenca Grau and Jimenez-Ruiz, 2011) are specialized in the ontological matching in an automated fashion and will therefore be used as a benchmark in a later section of the paper.

[1]https://www.w3.org/TR/r2rml/

Table 1: Example File ”example” represented as a database table.

ID	the_geom	Fe1	Fe2	... FeN
Ex.1	POINT(..)	123	”String”	3.4

3 SEMANTIC EXTRACTION FROM GEODATA SETS

A geodataset describes in our context, a database table including one column for the to be described geometry and *n* Featurecolumns in which values related to the geometry are stored. An example is presented in Table 1. Every feature (Fe) comes with a feature descriptor, which is usually present in a one-worded string description.

In the real world, a geodataset can take on many vector and raster data formats (e.g. GML[2] (+dialects), KML[3], SHP, POSTGIS Database Table, GEOTIFF (Ritter and Ruth, 1997)). For a more detailed description on data integration please refer to (Homburg et al., 2016). Even though in reality schema descriptions for example in GML 3.2 might reveal a columns schema type by delaring a non-native type description in the XML schema, we assume for our experiments that no such information is given. We proceed in this manner because in our experience a significant

[2]http://www.opengeospatial.org/standards/gml

[3]http://www.opengeospatial.org/standards/kml/

amount of spatial data is schemaless and sometimes even their origin or authors are not known. In reality we could in some cases take advantage of (XML) schema type information if we create a local ontology out of (XML) schema descriptions as described in (Homburg et al., 2016). However even if we find a data type in a schema we are still left in relating this type to other types in the semantic web, as URIs from types in XML schemas do not usually refer to URIs of the semantic web. This task is in essence an equivalent task to relating schemaless data with descriptors.

3.1 Categorizing Features

When receiving a geodata set from an unknown source, little is known about how the features semantically relate to the given geometry. Features can describe a geometry better e.g. a column ”height” of a dataset ”Tree”. It could also, be database related information such as ID columns or a relation to a different concept, e.g. column ”partner city” of dataset ”City”.

In conclusion: If we can determine the semantic functions of feature columns of the given dataset, we are able to relate them to the given SpatialObject and extend our knowledge base. For this purpose we intend to define classes of columns of our relational datasets that describe us information about the geometry itself (e.g. address information), metadata information such as IDs of databases or describing elements, hints about foreign key relations, hints about semantic descriptions of the geometry and in the case of numerical values the context of the numerical value in accordance with its unit if available. Neither of the aforementioned classifications is guaranteed to exist, however if we can determine classifications like these in our dataset we are able to take this information into account for our integration process.

3.2 Relation to Existing Datasets

We might encounter that the dataset we are about to import into our knowledge base already exists in another knowledge base. However, its description in the already existing knowledge base might vary in detail, contain additional information or even contradicting information to the dataset we are about to import. The challenge therefore is to recognize an already existing dataset, to extend it accordingly and to determine false information entered into the knowledge base in order to correct it. Management of contradicting statements of data will be discussed in a later section.

4 EXTRACTION PROCESS

In this section we describe our semantic extraction process by the combination of three components: Geometry matching, Feature Value Analysis and Feature Description Analysis.

4.1 Preliminary Tasks

Before an evaluation of the different criteria is done we try to detect the language of the dataset first. In our use cases we can assume that datasets are only present in one language, so we adapt this assumption in our further analysis. In order to do this we are detecting the languages of column names using the Google Translate API[4]. Because of possible disambiguations across languages we are assuming the most occuring language result as the language used in the dataset.

4.2 Geometry and Dataset Specification

Chances are, that the geometries we intend to import from our dataset are already present in the semantic web or through the import of a previous datasource (Fig. 1). In this case we can make use of already existing geospatial ontologies to verify our assumption. By using the Geonames[5] and LinkedGeoData Ontology[6] we can ask for existing concepts in a small enough buffer around the centroid of the geometry we are about to analyze as shown in Listing 1.

Listing 1: Example query to detect geometries in a buffer using LinkedGeoData.

```
SELECT DISTINCT ?class ?label ?s WHERE {
?s rdf:type ?class .
?s rdfs:label ?label .
?s geom:geometry ?geom .
?geom ogc:asWKT ?g .
Filter(bif:st_intersects
(?g, bif:st_point
    (6.862689989289053,50.97576136158093), 1.0
    E−6))
FILTER NOT EXISTS { ?x rdfs:subClassOf ?class.
    FILTER (?x != ?class) }}
```

If the buffer is set small enough and the coordinates are precise, we will get a unique result of a representing class for the geometry, if it exists. By analyzing each geometry of our given dataset in this way we can increase the chance to get the correct class if the same class occurs with many matches. In addition we are verifying our assumption of found geometries

[4]https://translate.google.com/
[5]http://www.geonames.org/ontology/documentation.html
[6]http://linkedgeodata.org/About

by comparing associated properties found in LinkedGeodata/Geonames with our dataset. In many cases we may discover the label of the found entity to be a value in one of the columns of our dataset, therefore assuring our assumption to have found the right entity in the ontology. On top of that we may as well classify certain columns in our dataset just by finding an equivalent property in one of the corresponding ontologies.

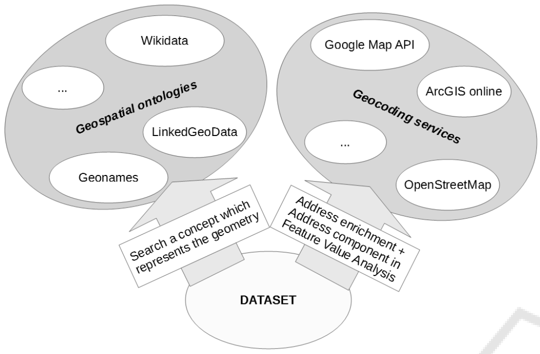

Figure 1: Geometry and dataset specification.

4.2.1 Address Enrichment

For every row within our dataset we try to geocode the geometry that is present and add this information to our knowledge space under a common name. Using this way we are able to use a unified vocabulary to query address data for the geometries we are about to add to our knowledge base. In addition we achieve an information base that we can compare to possible features of the dataset in order to identify address parts represented.

4.3 Feature Value Analysis

Several kinds of information appear frequently in our dataset, that is why, a first step is to identify common information thanks to a Feature Value Analysis (Figure 2). During this step, a sequence of process specific to each kind of information searched are applied. The explanation of these processes is presented below.

- **Address Components:** The specificity of geodatasets is that they contain a geometry for each spatial object. The usage of the spatial object geometry with a geocoding service (in our case, Google Map API), allows for address enrichment which has been explained previously (cf. 4.2.1). The information which has been retrieved is compared with the different value of the cell, in order to know which column contains information concerning the geographic address of the object.
- **ID:** The process of an eventual ID discovery corresponds to an analysis of values in order to identify a column which fulfils the following constraints: the value has to be an Integer and unique.
- **Unit:** We can safely assume that a double represent a quantification, that is why an analysis of all columns determines that a column represents something with a unit if all value are Double or Double and Integer. Something that is usually measured in any unit (e.g. 2.5°C) or is a description of an amount (2.5 apples). If we can identify the column type from its descriptor, then we may be able to draw conclusions about the unit associated with this type. Work on integrating e.g. DBPedia with unit ontologies has been done by (Rijgersberg et al., 2013) and may also be extended manually by our projects work for most commonn units.
- **Regular Expression:** A set of regular expression has been defined for: a date, a phone number, an email address, a website and a uuid. This set of regular expressions is then applied on all strings in order to check whether the string matches with one of those regular expressions. The elements identified as a date are stored thanks to a data property with the name of the column and the type *xsd:Date*. Information corresponding to a phone number, an email address, and a website is stored using FOAF ontology properties foaf:phone, foaf:mbox, foaf:homepage respectively. The uuid is stored as a data property.
- **Remaining String:** Natural language processing is applied on all strings which have not yet been identified. For the moment, this natural language processing is specific to German and English and may be extended to further languages in the future. It is aiming to determine if the string is an adjective or a noun. The values of the column, which contain a majority of adjective will become an instance of concept link to the general concept with an object property. When a column contains a set of nouns which occur frequently, we assume the column describes a type of the general object. The value of this column is processed in order to identify a set of nouns without redundancy and then, the nouns which composed this set are added as a subclass of the general concept which represents the file. When all values have been analyzed the process of Feature Descriptor Analysis (cf. 4.4) begins and is applied on all column names which have not yet been processed by the value analysis, on the adjective column, and on the nouns which become a subclass.

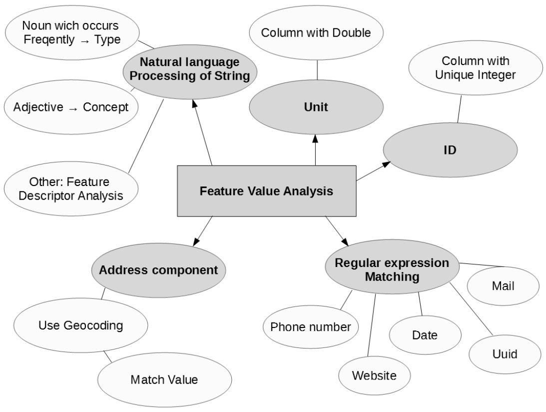

Figure 2: Identification by value analysis.

4.4 Feature Descriptor Analysis

The Feature Descriptor Analysis (Fig. 3), in the case of our datasets the column name, can give us valuable information about Properties and Classes in Ontologies that represent the columns content. However, column names are represented in natural language and with a limited context to parse from, which can limit disambiguation methodologies if needed. In addition before an analysis of the feature descriptor can be conducted, the following preprocessing steps need to be done:

- Recognition of common abbreviations and replacement of those with their long form
- Detection of the language being used in the columns name

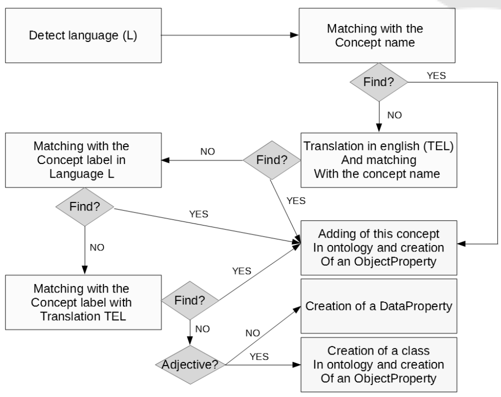

Figure 3: Process of linking with semantic web resources.

4.4.1 Analysis

We conduct the analysis of column names as follows: For a list of to be examined triple stores we try to match a concept first using its basic URI and in case this fails using a Label matching approach. In case there is no concept after these two steps we translate the given column name to English and try the aforementioned steps again. We discovered that using an English translation is not always possible as the translation of the full term is not necessarily respresenting a word that can be found in a dictionary or ontology. More often than that compound words needed to be split and investigated separately. In that regard we have been analysing the parts of compound nouns from their ending to their beginning and try to resolve possible concepts from those noun parts.

Listing 2: Splitting of compound nouns.

```
Bauarbeiter −> Arbeiter
primary school −> school
```

In case we cannot find any concept for the columns name using all of the mentioned methods we declare the column as unresolvable. If we have many results for the concerned column we will rank the results using the Levensthein Distance to find out the concepts name which is most near to the columns name. This concept will be taken to describe the column in the local ontology.

4.5 Combination of the Different Steps and Creation of the Local Ontology

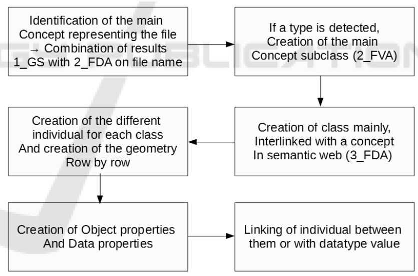

Figure 4: Combination of the different step in order to build the interlinked ontology.

After executing the aforementioned four steps we receive four sets of concepts which are used to build the resulting local ontology (Fig. 4) as follows:

- If a class has been detected using the geometry detection, this class (or the highest ranked class) will be taken to describe the dataset
- If a class has been detected via analysis of the files name, this class will be taken if no appriate geometry class has been detected
- Properties and their respective ranges as detected by the Feature Descriptor Analysis and for address columns as determined by its respective analysis are created

- Individuals are created according to the recognized classes and values that can be resolved to URIs will be created as the corresponding individuals

5 EXPERIMENTAL SETUP

The experimental setup explains what are the different datasets and the different approaches which have been used and what are their goals.

5.1 Datasets

For testing our approach, we need to apply it on a set of data which provide a diversity of information with different quality. Our project SemGIS being in the context of disaster management, we have chosen five files: two files about schools, two files about hospitals and a file about rescue organisation which will serve as buildings to be evacuated in case or emergency unit buildings respectively. Files about schools and hospitals come from two different regional sources, one of each concerns the city of Cologne and the other concerns the region of Saarland in Germany. The two different files have some similarities, but don't provide the same type of information. They allow to evaluate the integration of a similar dataset (with a same subject) from different data sources and with different content. Thanks to three fields (school, hospital and rescue organisation), and different data sources, we obtain a diversity of information with different quality as the column names and its contents which describe the information are built differently. Sometimes, these column names are an abbreviation, a complete name, a composed name separated by an underscore or an abbreviation of two words, still representing the same semantic meaning.

5.2 Experiments

For assessing the relevance and the efficiency of our automatic approach, we have applied two others approaches on dataset.

5.2.1 Experiment 1: Manual Approach by Non-expert

This manual approach consists in creating an ontology for each file of the dataset. There exists several ways to create an ontology and to understand the meaning of a file when you are non-expert [7]. That is why, we have chosen to realize this approach two times with two different persons. For doing so, we have provided the dataset to these two persons and asked them to create an ontology which contains all information of the file.

During the experiment, we have evaluated the similarity between these two manual approaches and then, each of them with our automatic approach. The comparison between a manual approach and our approach allows to compare a human method with a computational method.

5.2.2 Experiment 2: Other Automatic Approach with LogMap

LogMap allows to match two ontologies according two automatic way: One using string matching and one using a matching repair algorithm. A part of our automatic process is to match our local ontology with the semantic Web (DBpedia and Wikidata). In order to compare our automatic way with LogMap we have created a very simple local ontology with the name of the column as Concept and apply the different matchings of LogMap on this simple local ontology with DBpedia and as a comparison with Wikidata.

6 EVALUATION

To evaluate our results we created the ontologies we expect as results manually for each data set. We therefore consider a manual annotation by a non-expert human being as our gold standard. We compare the manually created ontologies to the automatically generated ontologies using a point based scoring system as follows:

Agreement Score:
Award one point for:

- Each correct assigned range for a property
- The correctly assigned class for the dataset

In addition we award a correct assignment if a class we received in the generated result is a subclass of the class we expect from the gold standard. We also award a fraction of a point for non-recognized classes which have a similarity score according to the measure of Resnik.(Resnik, 1995) If classes we have found are semantically similar but not depicted in the ontology we are using for classification, then we will add the resnik measure to our agreement score.

[7] A Non-Expert is someone who knowsSsemantics and GIS but doesn't know the context and the goal of the dataset.

Table 2: Evaluation Results - Generated Ontology vs. LogMap.

Dataset Name	Agreement Score
H.C. DBPedia	18.7%
H.S. DBPedia	18.7%
S.C. DBPedia	23%
S.S. DBPedia	23%
F. DBPedia	5.8%

7 RESULTS

7.1 Comparison with LogMap

We present our evaluation results in table 2. The data sets used for our experiments are about Hospital (H), School (S) and Firebrigade (F). The data set on Firebrigade corresponds to the north of germany, whereas the data sets for hospitals and schools corresponds on for the city of Cologne (C) and one for the german *Bundesland* Saarland (S).

7.1.1 Interpretation

LogMap is based on a repair matching in order to link two ontologies. Table 2 presents a comparison of the matching result with DBpedia between LogMap and our automatic approach. The result of LogMap with DBpedia has obtained a low matching, only few concepts have been detected with the dataset. This low matching illustrates the difficulty to identify concepts from the value of datasets and explains the low matching between both results. Our automatic approach has identified more concepts than Logmap. Although LogMap was not specified in the interpretation of the meaning, but rather in the matching, we use it to compare the part of our approach which allows to match a column name with a concept. So we can say that the step of feature value analysis obtains a good result which is due to its combination of several steps of matching based on the natural language processing. We have also, tried the matching with Wikidata, but LogMap has found no matching. In comparison with DBpedia, Wikidata has the particularity to have a URI with an identifier which is not a string similar to a label. We assume this particularity can imply some difficulties for LogMap. Moreover, LogMap uses also the hierarchy of the ontology, but with this type of data, there is no hierarchical information which is another problem with a tool for ontology matching. Our advantage is that our approach considers and combines several types of information to identify a concept as for example a label or a comment. In the case of Wikidata, taking into account the label of the concept during the analysis, allows for obtaining a good result.

Table 3: Evaluation Results - Agreement Score.

Data sets	AGO vs. MO1	AGO vs. MO2	MO1 vs. MO2
H.C.DB	61.5%	33%	15.3%
H.C.WD	86%	0%	53%
H.S.DB	62.5%	0%	18.7%
H.S.WD	91.3%	0%	40%
S.C.DB	38%	0%	30.7%
S.C.WD	92%	0%	52%
S.S.DB	64.7%	0.05%	18%
S.S.WD	89%	0%	45%
F.DB	41.0%	0%	30%
F.WD	78.9%	0%	31.5%

7.2 Comparison with Manual Ontologies

We present our evaluation results in table 3. This table presents in the first column the comparison between our automatic generated ontology (AGO) and the first manual ontology (MO1). The second column presents the same comparison but with the second manual ontology (MO2). The last column shows the result of comparison between the two manual ontologies. The process has been applied two times for each data set: one time to link concepts to DBPedia (DB) and one time with Wikidata (WD).

7.2.1 Interpretation

Thanks to the comparison between two manual ontologies, we can see that two different persons can create two very different ontologies from a same dataset, since we have obtained an average of similarity of 35,4%. When we compare our automatic approach with the two manual ontologies, we can observe that the result is close to the first ontology but is very distant to the second manual ontology. As the first manual ontology has been created with the same way of reasoning than the automatic process, we obtain an average of 71,48% of similarity. However, we can say our approach is very influenced by our way to create an ontology. The creation of an ontology is generally influenced by its creators because it depends directly of their approach of building and the meaning which they want to represent. That is why, two ontologies creates for the same goal but by two different people can be very different. Moreover, we can say that we obtain a better result with Wikidata than with DBpedia which taxonomy is a more developed. DBpedia is very rich in individuals, but a very flat class taxonomy compared to Wikidata concerning our Dataset. It also needs to be considered, that because of ambigious ways column names and values are defined in the ontology (e.g. abbreviations, shortened names etc.) misunderstandings of the human annotators about their meanings have been occuring and could often only partially resolved by extensive web research on their part.

8 CONCLUSION

In this paper we presented a new approach to integrate geospatial data into the semantic web using a fully automated way. By focusing on the geometry as a new central point for concept detection, we can build a local ontology structured around this main concept which allows to gather all information about it. We encountered that mapping our datasets using the Wikidata Ontology we can achieve on average better results than using DBpedia, due to a lack of a developed class hierarchy in DBPedia. Overall the results for our tested datasets seem very close to at least one of our annotators and we believe that our results depict one kind of interpretation of the datasets quite well. However, we as well experienced a lack of detection of several attributes of the datasets such as double values or meaningful integer values. In most of the cases such information is referring to the context of other given columns in which case with a more sophisticated approach we can hope to resolve the meaning of the corresponding properties automatically. In more rare cases not even we as humans can draw a distinctive connection to the meaning of such columns and therefore we see little chance for the computer to figure out a solution. We see this research as a basis of discussion for further automatic integration approaches of heterogeneous geospatial data. The approach we have presented is in our opinion likely to work on many common geospatial concepts and can therefore lead to an automated enrichment process for OpenStreetMap and/or Semantic Web geospatial data in the future.

8.1 Future Work

In our future work we want to explore how we can better resolve conflicts in our datasets. We therefore intend to create statistical profiles of typical geometries for the classes we are about to encounter and therefore create plausability criteria for the relations between classes representing geometrical objects both type and possibly area-based. Having discovered typical relations between typical together occuring geometry concepts we may as well achieve a more powerful basis for later reasoning experiments. In addition we may be able to deduce the class of geometries by their environment once we have been conducting an in-depth statistical analysis of geometry relations. This prospect could even be extended to be applied on moving objects that might be in a dangerzone of some kind.

ACKNOWLEDGEMENTS

This project was funded by the German Federal Ministry of Education and Research[8].

REFERENCES

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). *Dbpedia: A nucleus for a web of open data*. Springer.

Bizid, I., Faiz, S., and Boursier, Patriceand Yusuf, J. C. M. (2014). *Advances in Conceptual Modeling: ER 2013 Workshops, LSAWM, MoBiD, RIGiM, SeCoGIS, WISM, DaSeM, SCME, and PhD Symposium, Hong Kong, China, November 11-13, 2013, Revised Selected Papers*, chapter Integration of Heterogeneous Spatial Databases for Disaster Management, pages 77–86. Springer International Publishing, Cham.

Cuenca Grau, B. and Jimenez-Ruiz, E. (2011). Logmap: Logic- based and scalable ontology matching.

Guo, H., Song, G.-f., Ma, L., and WANG, S.-h. (2009). Design and implementation of address geocoding system. *Computer Engineering*, 35(1):250–251.

Homburg, T., Prudhomme, C., Würriehausen, F., Karmacharya, A., Boochs, F., Roxin, A., and Cruz, C. (2016). Interpreting heterogeneous geospatial data using semantic web technologies. In *International Conference on Computational Science and Its Applications*, pages 240–255. Springer.

Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., Muslea, M., Taheriyan, M., and Mallick, P. (2012). Semi-automatically mapping structured sources into the semantic web. In *Extended Semantic Web Conference*, pages 375–390. Springer.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Navigli, R. and Ponzetto, S. P. (2010). Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 216–225. Association for Computational Linguistics.

[8]https://www.bmbf.de/en/index.html Project Reference: 03FH032IX4

Nentwig, M., Hartung, M., Ngonga Ngomo, A.-C., and Rahm, E. (2015). A survey of current link discovery frameworks. *Semantic Web*, (Preprint):1–18.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*.

Rijgersberg, H., van Assem, M., and Top, J. (2013). Ontology of units of measure and related concepts. *Semantic Web*, 4(1):3–13.

Ritter, N. and Ruth, M. (1997). The geotiff data interchange standard for raster geographic images. *International Journal of Remote Sensing*, 18(7):1637–1647.

Scharffe, F., Atemezing, G., Troncy, R., Gandon, F., Villata, S., Bucher, B., Hamdi, F., Bihanic, L., Képéklian, G., Cotton, F., et al. (2012). Enabling linked-data publication with the datalift platform. In *Proc. AAAI workshop on semantic cities*, pages No–pagination.

Volz, J., Bizer, C., Gaedke, M., and Kobilarov, G. (2009). Silk-a link discovery framework for the web of data. *LDOW*, 538.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.