Exploiting Linked Spatial Data and Granularity Transformations

Heidelinde Hobel[1,2] and Andrew U. Frank[1]

[1] Institute for Geoinformation and Cartography
Vienna University of Technology
{hobel,frank}@geoinfo.tuwien.ac.at
[2] Doctoral College Environmental Informatics
Vienna University of Technology
heidelinde.hobel@tuwien.ac.at

Abstract. Geographic information is one of the fundamental core data sources for various applications. Freely available geographic information knowledge bases are emerging and the spatial dimension has become part of the Linked Open Data initiative. However, geographic information is stored as abstract geographic objects and exploring, extracting, and understanding the information must be facilitated for different user perspectives and use cases. We propose to use a semantic model and an extraction methodology which is aimed at allowing the consumption of geographic information in an intuitive way. We illustrate our approach based on previous work of a highway navigation conceptualization and present a functional approach to exploit granularity extractions targeted at enabling the user to change the point of view in navigation tasks.

1   Introduction

Geospatial information is becoming more and more important for a variety of applications in our everyday life. From supply chain management to e-Commerce support and from navigation tools through to complex recommendation systems, Geographic Information Systems are important interfaces to build suitable applications with a local or global perspective. Efforts in the fields of the Semantic Web and Linked Data [1] led to the emergence of the *Web of Data* comprising various geospatial data sources, e.g. LinkedGeoData[3] or GeoLinked Data[4]. The whole approach relies on Semantic Web technologies, especially the Semantic Web's language, which is referred to as the Resource Description Framework (RDF).

However, one challenge in Geographic Information Science is to develop suitable conceptual models that facilitate the understanding of geographic data, relating this data with information sources of other domains, and using this combined data to solve a specific task. Therefore, we have to consider the user's

---

[3] http://linkedgeodata.org
[4] http://geo.linkeddata.es

perspective and the goal he wants to achieve with the information provided in distributed and heterogeneous data sources. OpenStreetMap[5], and the semantic counterpart LinkedGeoData, are based on the following abstract elements: nodes, ways, relations, and tags. LinkedGeoData is aimed at linking the spatial dimension with knowledge bases from other domains. Although semantics have become an integral part of the geographic perspective, the required semantics to understand and explore geographic information in an intuitive way and from different user perspectives are not yet integrated in the geographic linked data efforts.

The goal of our research is to efficiently handle geographic data on different granularity levels by enabling the user to change the point of view. The contribution of our paper is summarized as follows:

– We implemented a descriptive model for topological navigation tasks based on different levels of detail by modeling different graph representations in the semantic model.
– We formalized the granularity transformations by using canonical projection functions.
– We highlight future challenges and compare our approach with the approach from Timpf and Kuhn [11].

The remainder of this paper is structured as follows: In Section 2, we describe shortly the fundamental conceptual model for wayfinding as well as related work. In Section 3, we present our ontology and the extractions to retrieve the data for the navigation tasks at the previously proposed granularity levels. We continue with the comparison of our approach with the approach from Timpf and Kuhn [11] (Section 4) and discuss some remaining challenges (Section 5). We conclude our work in Section 6.

2   Related Work

The geographic conceptual model we have chosen to investigate was firstly introduced by Timpf et al. [12], where navigation tasks are modeled at three levels of detail: (1) planning, (2) giving and receiving instructions, and (3) driving. Timpf and Kuhn [11] extended the previous hierarchical model by taking graph granulation theory into account. Their main goal was to build a theory of granularity transformations for wayfinding processes. To achieve their goal, they formalized conceptual models for each of the previously introduced task levels (referred to as conceptual levels or levels of detail).

Since the emergence of the Web of Data, there have been major efforts to populate the Linked Data cloud with useful information. The advantages of the Linked Data cloud are evident in many areas such as research, collaboration, and creation of value. Also in the field of geospatial data, initiatives are pushing ahead and populate the Web with interlinked geospatial data [3–5]. Together with the

---

[5] http://www.openstreetmap.org/

"Semantic Geospatial Web" [4], a plethora of tools, extensions, and optimizations were developed, e.g. GeoSPARQL[6] and stRDF/stSPARQL [8], facilitating the creation, publication, and processing of enriched geospatial data. According to Egenhofer [4], geospatial ontologies and query languages should be optimized to deal with synonyms, algebraic treatment of properties, and mapping of spatial terms onto corresponding geometries. Visser et al. [14] illustrated the need for formal ontologies of geospatial data and demonstrated how these ontologies enhance the retrieval of information. However, these geospatial characteristics are not directly suitable for the wayfinding graph model described by Timpf and Kuhn [11]. In [13], Tomko introduced a case study, where web-accessible data was used to enrich navigation routes.

The idea of multiple levels in the presentation of geospatial data, referring to the concept of different levels of detail, is already explored in various application fields of GI science. Weibel and Dutton [16] address the need of generalizing geospatial data as well as the models and algorithms dealing with these issues. In [2], Buttenfield proposed an algorithm for transmitting vector data on less detailed representations that are refined at finer levels. Stell and Worboys [10] proposed a framework to deal with generalizations on graph based data structures. An approach to model the levels of detail of spatial processes based on partial function application was proposed by Weiser et al. [17].

3   Wayfinding Ontology and Granularity Transformations

In this section, we introduce our idea of how to develop a descriptional semantic model, which allows different users to explore and understand geographic information of road networks based on a topological network data model and extract the corresponding information for the use cases introduced by Timpf and Kuhn [11].

3.1   The Topological Network Model

Road networks are typically represented as graphs. Cities, interstates, and exits or entrances are modeled as nodes and roads between the nodes are described as links. For instance, the RDF Graph Modeling Language (RGML) illustrates the opportunities of RDF to describe graph structures [9] corresponding with road networks. Based on the principles of Linked Data, we describe the objects of interest as entities and add semantic annotations to add information. In [7] it is described how to use Linked Data's Graph structure to naturally represent network topologies. Based on Timpf and Kuhn's ontology, we designed an ontology that is aimed at facilitating the consumption of highway network graphs, where the graph amalgamations are interwoven in the data set (see Figure 1). One problem of automatically generated amalgamations is that the world is versatile and concepts such as intersections, exits, and entrances are not easily generalizable due to their different properties.
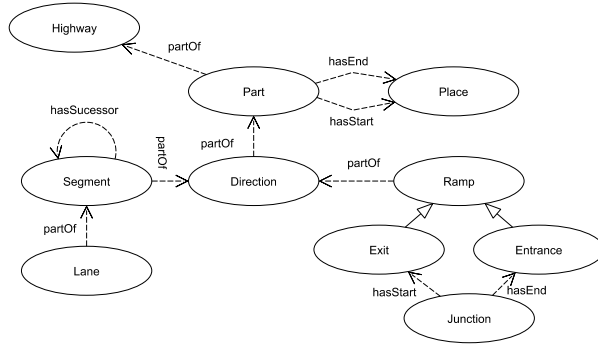
---

[6] http://geosparql.org

Fig. 1. Mapping of Timpf and Kuhn's [11] informal ontology to RDF structure

The following section illustrates how the ontology can be used to extract the information required to solve previously introduced [12] navigation tasks.

3.2   Granularity Transformations

In the Semantic Web Science typically SPARQL[7] is used to select, filter, and query information. We used an independent approach to illustrate the granularity transformations for a wider public.

In our model, instead of granularity mappings, we use coarsening functions, mapping from one conceptual level to another by removing information, given in form of triples. This means that we explicitly suppress some information which results in a more abstract level of detail. This process can be iterated (i.e., changing from one level to another) and based on different criteria. Starting with the following notation for RDF graphs derived from the W3C RDF Semantics specification [15], we describe an RDF graph G as a set of triples (which are often denoted as sentences) $G = (\mathcal{E}, \mathcal{P}, \mathcal{V})$, where

$\mathcal{E}$ is the set of entities, which describe anything in *the universe of discourse* and can be seen as a vocabulary [6]. Entities denote the *things* we want to describe with our RDF graph and are referred to as subjects.

$\mathcal{P}$ is the set of properties (i.e., the predicate of *sentence*), which describe the relations between entities and values.

$\mathcal{V}$ is the set of values, which are either entities or atomic values (i.e., literals). The value of a triple is denoted as object of a sentence.

We start with a set $\mathcal{E}$ of entities, a set $\mathcal{P}$ of properties and a set $\mathcal{V}$ of values, that allow us to express sentences. We represent a world $\mathcal{W}$, i.e. *universe of discourse*, in form of an RDF graph, which is formalized as a set of triples, where each triple $\boldsymbol{t}$ consists of an entity $e \in \mathcal{E}$, a property $p \in \mathcal{P}$, and a value $v \in \mathcal{V}$, i.e. $\boldsymbol{t} = (e, p, v)$, and therefore we obtain: $\mathcal{W} := \{\boldsymbol{t}_i : i = 1, 2, \ldots\} \subseteq \mathcal{E} \times \mathcal{P} \times \mathcal{V}$.

---

[7] http://www.w3.org/TR/rdf-sparql-query/

By employing the concept of canonical projection functions on $\mathcal{E} \times \mathcal{P} \times \mathcal{V}$ (e.g., π_1 retrieves the entity of triples), we can extract information and thus precisely describe certain entities which appear in triples (which can be subject to further conditions) in a world: e.g., to get all entities with property p_0 and value v_0 we can write:

$$A := \pi_1 \left(\pi_2^{-1}(p_0) \cap \pi_3^{-1}(v_0) \cap \mathcal{W} \right) = \{e \in \mathcal{E} : (e, p_0, v_0) \in \mathcal{W}\}$$

Therefore, properties and values allow us to search using different criteria. The canonical projection functions are defined on the whole space $\mathcal{E} \times \mathcal{P} \times \mathcal{V}$, which requires to intersect all appearing preimages with our *universe of discourse* $\mathcal{W}$ (as can be seen in the example above). In the following step, all triples including the identified entities as subject or object have to be removed. For instance, let us consider a country with several cities. We can use functions to find these cities, but if we consider a more granular world, e.g. only cities with a high population, we have to remove all information about these small cities, i.e. all triples concerning these smaller cities. Hence we are interested in all triples with the entity e and values v not in A (with $p =$ "is", $v =$ "small"), (A^c denotes A complement) so we consider the set:

$$\{\boldsymbol{w} \in \mathcal{W} : \boldsymbol{w} = (e, p, v) \wedge e \notin A \wedge v \notin A\} = \mathcal{W} \cap \pi_1^{-1}(A^c) \cap \pi_3^{-1}(A^c)$$

Based on the requirements of Timpf and Kuhn [11], wayfinding requires three different conceptual models: $\mathcal{W}_{driver}, \mathcal{W}_{instruction}$ and $\mathcal{W}_{planning}$. Whereas $\mathcal{W}_{driver}$ corresponds with the full descriptional model $\mathcal{W}$. According to our coarsening process the following relation holds:

$$\mathcal{W}_{planning} \subseteq \mathcal{W}_{instructional} \subseteq \mathcal{W}_{driver} \subseteq \mathcal{E} \times \mathcal{P} \times \mathcal{V}$$

According to our construction, an operation that is executed on a given level of detail can also be executed on a more detailed level, e.g. the function *route planning* has to be executable at the planning level as well as the driver level. However, it is not possible to execute the function *driving* at the planning level.

In order to map from the driving level to the instructional level, we remove triples from $\mathcal{W}_{driver}$ describing the lanes and road segments, resulting in a predicate containing a disjunction (denoted as $\vee$):

$$C_1 := \{e \in E : (e, \text{"isA"}, \text{"Lane"}) \vee (e, \text{"isA"}, \text{"Segment"}) \in \mathcal{W}_{driver}\}$$
$$\implies \mathcal{W}_{instructional} = \{\boldsymbol{z} \in \mathcal{W}_{driver} : \boldsymbol{z} = (e, p, v) \wedge e \notin C_1 \wedge v \notin C_1\}$$

Similarly, the coarsening function which maps from the instructional to the planning level includes removing the directions of the highways. We assume that in this case this comprises all "Ramps", "Junctions", and "Directions":

$$C_2 := \{e \in E : (e, \text{``}isA\text{''}, \text{``}Ramp\text{''}) \vee (e, \text{``}isA\text{''}, \text{``}Junction\text{''}) \vee$$
$$\vee (e, \text{``}isA\text{''}, \text{``}Direction\text{''}) \in \mathcal{W}_{instructional}\}$$
$$\implies \mathcal{W}_{planning} = \{\boldsymbol{w} \in \mathcal{W}_{instructional} : \boldsymbol{w} = (e, p, v) \wedge e \notin C_2 \wedge v \notin C_2\}$$

4   Comparison

In this section, we compare our implementation for granularity transformations based on Linked Data with the Graph Transformations of Timpf and Kuhn [11].

The first advantage of Linked Data is that every relation and information can be mapped in a flexible and iterative way. Hence, instead of using different graphs, where we have to use graph amalgamations to map from model to model, we can describe the connectivities of the entities in one descriptional model. Hence, Timpf and Kuhn's approach does not support the semantics between the described entities. For instance, when exploring a lane, we want to know the name of the highway the lane belongs to. Extractions facilitate in addition the retrieval of smaller subsets of information, which can be compared to humans' natural abstraction abilities. The second advantage we could identify is that the Linked Data model can be easily enriched with further information that can be used for various applications. For instance, nodes could be enriched with sights or landmarks to improve navigation tools or the links could be enriched with risk data that can be used to decide about the best route in supply chain management.

5   Discussion and Limitations

The proposed approach leaves some open challenges. In this section, we discuss challenges arising when using RDF graphs for wayfinding.

Incompleteness Modeling the real world in simple triples is a time-consuming and challenging task, since the annotation of all conceivable conditions requires significant efforts in creating descriptive content. The knowledge base we create is, therefore, mostly incomplete and reasoning based on this knowledge base will not always reveal the best solution. Arguable, the Semantic Web, Linked Open Data, and an open data community represent the the first step to mitigate the problem of incomplete knowledge, since everyone can contribute and extend the fundamental knowledge for reasoning. For instance, with the mentioned concepts, linking a concrete route by using an array of location points (a snapshot of the route) is a standard task. Linking nodes of OpenStreetMap enables OpenStreetMap contributors to edit the information when exploring the network graphs.

Imperfect Models While modeling the ontology introduced in Section 3.1, it became apparent that modeling the concepts and relations is highly dependable on the application area and the use cases to be considered. Creating the perfect model for all kinds of application tasks is therefore not possible. In Section 3.2, we have introduced our idea of granularity transformations. Extracting, transforming, and mapping information out of a comprehensive knowledge base into a suitable model is, due to RDF's flexible nature, a justifiably fitting approach. In this paper, we have redesigned the formalization firstly introduced in [12] to match with our expectations of conceptual levels of detail in wayfinding. By altering the connections of our entities and adding additional information by annotations as well as our granularity transformations, we showed, based on a use case example, the benefits of a simple RDF notation and transformation operations.

Oversized Search Space In [11], the authors evaluated their implementation based on a shortest path search in the formalized wayfinding graph. One problem we could not solve was identifying an appropriate extraction method for the highway graphs in order to find the best heuristic for the search of shortest paths. Under realistic geospatial workloads, i.e. all real world entities are mapped in the RDF triple store, and thus without extraction of a suitable subset reasoning is a time-consuming and costly task. Geospatial operations, e.g., finding all location points in a given area, also formed part in extensions of RDF and SPARQL (cf. Section 2). How these functions could be used to extract heuristics for our model, is a further research topic, especially when considering that Linked Data could be used to find perfect sightseeing trips.

6   Conclusion and Future Work

The Web of Data is gaining importance, offering users the possibility to utilize it as an open knowledge base serving information for various applications. Publishing and consuming data for geographical reasoning is still in an early stage, since the application fields are versatile and the development of fundamental ontologies and query languages has not been finished. The data of proprietary navigation tools is kept on inaccessible company servers. This major drawback inspired us to analyze a previously introduced formal model for wayfinding on different levels of detail. We used the derived results to develop an ontology and functional transformation operations, which should facilitate the consumption of topological network data.

We aim to extend our first approach to a framework that should enable to explore geospatial data and useful links in an intuitive way, which should facilitate data retrieval, linking, visualization, and hence the consumption of open data with a geospatial perspective.

References

1. C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.
2. B. P. Buttenfield. Transmitting vector geospatial data across the internet. In M. J. Egenhofer and D. M. Mark, editors, *GIScience*, volume 2478 of *Lecture Notes in Computer Science*, pages 51–64. Springer, 2002.
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 A. Paschke, N. Henze, and T. Pellegrini, editors, *I-SEMANTICS*, ACM International Conference Proceeding Series. ACM, 2010.
4. M. J. Egenhofer. Toward the semantic geospatial web. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, GIS '02, pages 1–4, New York, NY, USA, 2002. ACM.
5. E. Giaccardi and D. Fogli. Affective geographies: toward a richer cartographic semantics for the geospatial web. In S. Levialdi, editor, *AVI*, pages 173–180. ACM Press, 2008.
6. T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.
7. G. Hart and C. Dolbear. *Linked Data: A Geographic Perspective*. CRC Press, 1 edition, Jan. 2013.
8. M. Koubarakis and K. Kyzirakos. Modeling and Querying Metadata in the Semantic Sensor Web: the model stRDF and the query language stSPARQL. In *7th Extended Semantic Web Conference (ESWC2010)*, June 2010.
9. J. R. Punin and M. S. Krishnamoorthy. Describing structure and semantics of graphs using an rdf vocabulary. In *Extreme Markup Languages*, 2001.
10. J. G. Stell and M. F. Worboys. Generalizing graphs using amalgamation and selection. In *SSD*, pages 19–32, 1999.
11. S. Timpf and W. Kuhn. Granularity transformations in wayfinding. In C. Freksa, W. Brauer, C. Habel, and K. F. Wender, editors, *Spatial Cognition*, volume 2685 of *Lecture Notes in Computer Science*, pages 77–88. Springer, 2003.
12. S. Timpf, G. S. Volta, D. W. Pollock, and M. J. Egenhofer. A Conceptual Model of Wayfinding Using Multiple Levels of Abstraction. In A. U. Frank, I. Campari, and U. Formentini, editors, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, International Conference GIS - From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning*, volume 639 of *Lecture Notes in Computer Science*, pages 348–367. Springer, 1992.
13. M. Tomko. Case study-assessing spatial distribution of web resources for navigation services. In *Proceedings of the 4th International Workshop on Web and Wireless Geographical Information Systems*, pages 90–104, 2004.
14. U. Visser, H. Stuckenschmidt, G. Schuster, and T. Vögele. Ontologies for geographic information processing. *Computers & Geosciences*, 28(1):103–117, 2002.
15. W3C. RDF Semantics. Website, 2004. Technical Report, `http://www.w3.org/TR/rdf-mt/`; last visited on July 20th 2013.
16. R. Weibel and G. Dutton. Generalising spatial data and dealing with multiple representations. *Geographical information systems*, 1:125–155, 1999.
17. P. Weiser, A. U. Frank, and A. Abdalla. Process composition and process reasoning over multiple levels of detail. In *GI Science 2012 Extended Abstract*, 2012.