

## INTEGRATION OF MOBILE GIS AND LINKED DATA TECHNOLOGY FOR SPATIO-TEMPORAL TRANSPORT DATA MODEL

B. Margan<sup>1,\*</sup>, F. Hakimpour<sup>1</sup>

<sup>1</sup> School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran –  
(behnammargan,fhakimpour)@ut.ac.ir

**KEY WORDS:** Semantic Web, Linked Data, GeoSPARQL, RDF, Transportation Data, Mobile GIS

### ABSTRACT:

Linked Data is available data on the web in a standard format that is useful for content inspection and insights deriving from data through semantic queries. Querying and Exploring spatial and temporal features of various data sources will be facilitated by using Linked Data. In this paper, an application is presented for linking transport data on the web. Data from Google Maps API and OpenStreetMap linked and published on the web. Spatio-Temporal queries were executed over linked transport data and resulted in network and traffic information in accordance with the user's position. The client-side of this application contains a web and a mobile application which presents a user interface to access network and traffic information according to the user's position. The results of the experiment show that by using the intrinsic potential of Linked Data we have tackled the challenges of using heterogeneous data sources and have provided desirable information that could be used for discovering new patterns. The mobile GIS application enables assessing the profits of mentioned technologies through an easy and user-friendly way.

### 1. INTRODUCTION

Transportation data is attractive and vital data in urban management challenges. It is collected via various data providers and it contains different features such as volume of traffic, hour of the day, vehicle class, etc.

Semantic web (Berners-Lee et al. 2001) presents a prospect for Content inspection and deriving insights from data. By interlinking transportation datasets, we can recognize new factors in road traffic accidents and congestion that were unseen before (Heath and Bizer 2011). Linked Data is a machine-readable data on the Web, defined accurately and interlinked with other data sources efficiently. Linking properties and interlinking transportation information from various resources make it efficient to visualize and analyze linked transport data. This research benefits from Linked Data technology to enable multisource data fusion and visualization. Accordingly, we can introduce variations into the dataset that result in developing new transport applications which will be provided from other interlinked datasets.

In a real-life case study of urban environments, a new approach is proposed by (Fotopoulou et al. 2015) to use Linked Data Analytics. Based on the ability to exploit, manage and process Linked Data, their approach is used to examine the health impact of outdoor air pollution in urban areas. Their principles of design and implementation offer a good solution for raising the standard of living in smart city environments by taking into account the diversity of heterogenous data sources.

For Linked Geospatial Data (LGD), a multidimensional and quantitative interlinking approach was proposed. (Zhu et al.

2017) The data interlinking types are defined in the approach based on geospatial data features and their roles in data exploration, including two aspects (inherent and morphological characteristic links) and three levels (elementary, compound, and overall characteristic links).

Information such as travel time and real traffic time is made available by navigation services through provided application program interfaces (APIs) (Kolyvakis et al. 2018). Data provided by navigation services have characteristics like lower price and easy to acquire overcome traditional data gathered through cameras and sensors.

### 2. SEMANTIC WEB

Semantic Web is a web extension standardized by the World Wide Web Consortium (W3C). These standards provide data formats for data sharing and define information sharing protocols. In addition to supporting web-based documentation, W3C has put forward this plan in order to build technology that supports "web data", meaning exactly the kind of data that is stored in databases. The ultimate goal of Web Data is to enable computing machines to perform Web transactions more reliably and efficiently. The Semantic Web is a perspective on web of data, which is interconnected and stored in related structures.

The Semantic Web enables users and ordinary people to create web caches of data, define vocabulary to understand them, write rules of use to manage them. This structured relational data is used in formats such as Resource Description Framework (RDF), Web Ontology Language (OWL). In general, it can be said that the Semantic Web consists of two parts. First,

\* Corresponding author

collecting common templates to combine and integrate data from different sources, while focusing more on the main web focusing on document exchanges.

Also keeping track of how much data is connected to each other and how the data relates to real-world objects is a semantic web topic. This allows a person or processor to start from a database and then move to a set of databases. If the goals pursued by the Semantic Web are to be achieved, it is very desirable for users. But in a general judgment, it can be argued that the pace of semantic web development has been slower than expected because the complexities of its underlying model require a great deal of effort so that an application can hide these complexities from users' eyes.

One of the benefits of using Semantic Web in accordance with classic Web is to offer mutual format for data processing. The geospatial information management accompanying with Semantic Web is known as the Spatio-Semantic Web. To serialize geospatial data, the most popular standards are Well-known text (WKT), Geography Markup Language (GML), Keyhole Markup Language (KML), and so on, and the standards used in GeoSPARQL are WKT and GML. GeoSPARQL's default coordinate reference system is CRS:84, necessarily it could be altered in the query.

On the Semantic Web, vocabulary defines the ideas and interactions used to explain and describe an area of interest (also referred to as "terms"). Vocabularies are used to classify terms that can be used in a specific implementation, to characterize possible interactions, and to identify possible restrictions on the use of those terms. Practically, vocabularies can be very complicated (with several thousand terms) or very straightforward (with just one or two ideas being described). There is also no clear distinction from what is called "vocabulary" and "ontology." It is common to use the term "ontology" to gather words more complicated and potentially somewhat formal, while "vocabulary" can be used when such rigid formalism is not generally used.

## 2.1 Linked Data

Linked Data relates to connections between information sources and the practice of linking web-based data. Unlike classic Web isolated information silos, the Web seeks to interconnect this information so that all datasets contribute to worldwide information integration, linking information from various domains and sources. (Sakr et al. 2018).

A semantic query language is required to enable querying and manipulating Linked Data on the Web. Such a language has been implemented by the World Wide Web Consortium, called SPARQL Protocol and RDF Query Language. RDF is generally a directly marked information format of the graph. Therefore, SPARQL is primarily a query language that matches the graphs. SPARQL could be used to formulate queries that range from simple graph patterns to very complicated analytical queries. (Sakr et al. 2018).

## 2.2 Geospatial Information

The volume of geospatial data on the Web has multiplied over the past few years. This is due to the fact that geospatial data come from different domains (e.g. earth researchers, surveyors, civil engineers) engaged in the production of geospatial data also publish them as RDFs to increase their

value by integrating them with other datasets. (Bereta et al. 2016).

Although much effort has been made to represent and query geospatial RDF information, there are not many publications in the corresponding linked data exploration. Within the scope of Spatial Linked Exploration, cutting edge methods focus on discovering only spatial equivalences among entities, abandoning other types of relationships, topological relationships, unknown and rich geospatial information stored in many datasets unfound. (Smeros and Koubarakis 2016). Consequently, Semantic Web community became involved in offering data models, query languages and applications to display, model and visualize linked geospatial data (Koubarakis et al. 2012). The development of the Open Geospatial Consortium (OGC) standard GeoSPARQL (Perry and Herring 2012), a geospatial extension of RDF and SPARQL, has reinforced these attempts.

The GeoSPARQL is a standard Geographic Query Language from Open Geospatial Consortium (OGC) for retrieving and manipulating the geospatial data stored in the ontology. It provides the ability to query and reason about the spatial properties (such as topological binary relationships) and non-spatial properties of the entities in the ontology. Also, it enables to perform the spatial processing functions such as distance, buffer, etc.

In Figure 1 spatial relations of two features illustrated. They could be used in the GeoSPARQL filter functions for geospatial analysis. One of the basic concepts for spatial data analysis is the formal understanding of geometric relationships between spatial features. Topological relationships are specific subsets of geometric relationships that do not change under topological transformations. Queries in spatial databases are often based on relationships between spatial sequences, so spatial relevance of the sequences needs to be considered in order for the user to obtain the desired response. In such situations, where the geometrical relationships are usually complex between complexities, spatial topology relationships are used instead of multiple simple spatial queries and their results compared. These relationships combine the results of simple spatial queries and return the optimal answer by filtering on them.

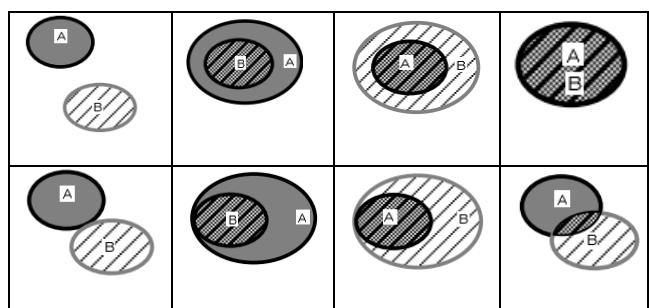


Figure 1. Spatial relations between two features

## 3. PROPOSED METHOD

In this paper, the sample data is collected from Google maps API in a CSV data structure and time length of three weeks as input for Linked Data visualization and exploration. We also

made use of Sophox service, which meets our needs in formulating and executing simultaneous queries over both OpenStreetMap and Wikidata to retrieve transportation network information from OpenStreetMap. To define an appropriate ontology for sample data based on the vocabulary specification for each column, the open source software OpenRefine and its RDF extension makes ontology modeling and conceptualization available. To record the date and travel time between two specific location Time: hasduration and time: inXSDDateTime vocabulary is used. Moreover Geo:asWKT Vocabulary used for POINT and LINESTRING Geometry. Parliament server as a high-performance triplestore is used to enable executing geospatial queries regarding GeoSparql Standards.

In this paper, we present a web and mobile application that enables the visualization and exploration of linked transport data. It enables retrieval of public transport data such as travel time between two nodes, traffic index and vertex information in the user's current position or selected position on the map by executing GeoSPARQL queries on SPARQL endpoints. Apache Cordova as a cross-platform technology was used for building the Android application.

Firstly, user's vertex would be distinguished by determining a sufficient buffer radius from user's current or selected location through use of GeoSPARQL capabilities depicted in Figure 2.

```

PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX ex: <http://www.example.org/POI#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX ogc: <http://www.opengis.net/ont/geosparql#>
PREFIX ogcf: <http://www.opengis.net/def/geosparql/functions/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX time: <http://www.w3.org/2006/time#>

#(bif:st_distance (?fwkt,?vv) AS ?Distance) ?f

SELECT DISTINCT ?address (geof:distance("POINT (35.728951,51.3848)"^^geo:wktLiteral, ?vertex, uom:metre) AS ?dist)
WHERE {
?p rdfs:label ?st;
time:inXSDDateTime ?datepol;
ex:co ?co.

?p2 ex:node ?fk.
?p2 time:hasDuration ?travet.
?p2 time:inXSDDateTime ?timetraf.
?fk rdfs:label ?address.
?fk geo:asWKT ?vertex.
FILTER(geof:sfCrosses(geof:buffer("POINT (35.728951,51.3848)"^^geo:wktLiteral, 500, uom:metre),?vertex))

FILTER(?datepol = ?timetraf)
}
    
```

Figure 2. Example of a GeoSPARQL query to extract label and distance of ways to user's position.

After executing this query, way with the lowest distance from the input position has retrieved to present transportation data such as travel time collected from Google Map API, name, type of route, max speed and whether it is one way or not provided by OpenStreetMap and illustrated in Figure 3.

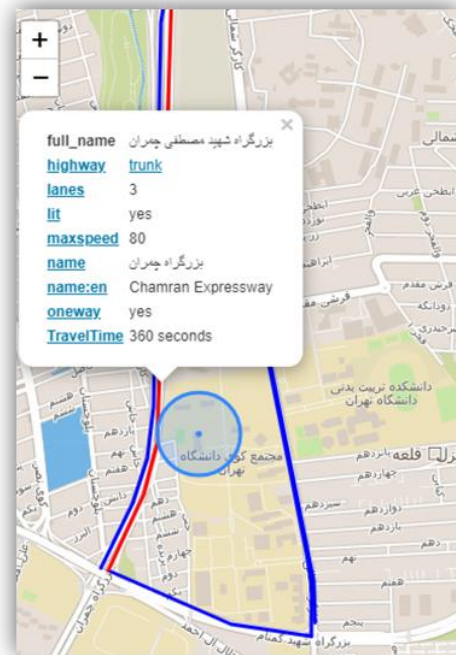


Figure 3. Visualization of nearest way to the user's current position and relevant transportation information

From a technological perspective, this scenario presented how the different enabling technologies mentioned in former sections used to retrieve, visualize and explore linked transportation data and relevant information of each element from different resources. Figure 4 illustrates an overview of the proposed solution.

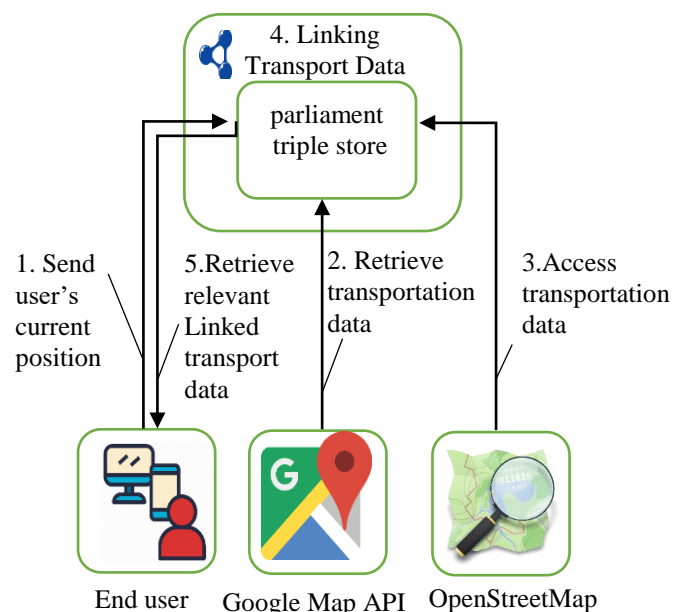


Figure 4. Overview of proposed application

#### 4. CONCLUSIONS

Semantic Web and Linked Data technologies are among the key enabling technologies to tackle the challenges of making use of heterogeneous data sources. In this paper, we presented an approach and documented the implementation of a Semantic Web application, using a web and mobile programming and leveraging heterogeneous data sources that follow semantic Web standards. Spatio-temporal queries executed over linked transport data and resulted in traffic information in accordance with user's position. The application enables assessing the profits of mentioned technologies through an easy and user friendly technique. In our future work we will apply potential of these technologies over other individual transportation data.

#### REFERENCES

- Bereta K, Xiao G, Koubarakis M, Hodrius M, Bielski C, Zeug G. 2016. Ontop-spatial: Geospatial data integration using GeoSPARQL-to-SQL translation. In: Proceedings of the 15th International Semantic Web Conference, Posters & Demonstrations Track (ISWC). Available at: <http://ceur-ws.org>. Vol. 1690.
- Berners-Lee T, Hendler J, Lassila O. 2001. The semantic web. *Scientific american*. 284(5):28–37.
- Fotopoulou E, Zafeiropoulos A, Papaspyros D, Hasapis P, Tsiolis G, Bouras T, Mouzakitis S, Zanetti N. 2015. Linked data analytics in interdisciplinary studies: The health impact of air pollution in urban areas. *IEEE Access*. 4:149–164.
- Heath T, Bizer C. 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*. 1(1):1–136.
- Kolyvakis P, Mader C, Kiritsis D. 2018. Semantic Interoperability and Open IoT APIs for Smart Cities Applications. In: *IFIP International Conference on Advances in Production Management Systems*. Springer. p. 146–154.
- Koubarakis M, Karpathiotakis M, Kyzirakos K, Nikolaou C, Sioutis M. 2012. Data models and query languages for linked geospatial data. In: *Reasoning Web International Summer School*. Springer. p. 290–328.
- Perry M, Herring J. 2012. OGC GeoSPARQL-A geographic query language for RDF data. *OGC implementation standard*. 40.
- Sakr S, Wylot M, Mutharaju R, Le Phuoc D, Fundulaki I. 2018. *Linked Data: Storing, Querying, and Reasoning*. Springer.
- Smeros P, Koubarakis M. 2016. Discovering Spatial and Temporal Links among RDF Data. In: *LDOW@ WWW*.
- Zhu Y, Zhu A-X, Song J, Yang J, Feng M, Sun K, Zhang J, Hou Z, Zhao H. 2017. Multidimensional and quantitative interlinking approach for Linked Geospatial Data. *International Journal of Digital Earth*. 10(9):923–943.