[DM-04-078] Geospatial Semantic Queries

Abstract

The increasing accessibility of geospatial data in the form of knowledge graphs, developed in alignment with the Semantic Web vision and employing Linked Data principles, is becoming a prominent feature of the Web. The multitude of available geospatial knowledge graphs demonstrates their indispensable role within the Web of Data Cloud. Such graphs serve as central nexuses, interconnecting events, people, and objects, offering an ever-growing semantic representation of the geospatial information wealth. The resulting resources and capabilities are being leveraged to utilize, consume, and capitalize on geospatial information through the strategic deployment of knowledge graphs. The geospatial graphs are stored and managed by triple stores, which are also known as RDF stores or knowledge bases. As examples, SPARQL and GeoSPARQL are semantic query languages that are used to retrieve and process knowledge graphs. Some developments and experiences in the GIScience community have demonstrated the feasibility of expressing queries across diverse knowledge graphs to retrieve and process geospatial data from disparate and distributed sources. These efforts have facilitated the consumption of geospatial knowledge graphs through lightweight web applications or GIS applications.

Keywords: GeoSPARQL, Knowledge Graph, ontology, queries, RDF, semantics, SPARQL

Author & citation

Vilches-Blázquez, L. (2025). Geospatial Semantic Queries. The Geographic Information Science & Technology Body of Knowledge (Issue 1, 2025 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2025.1.7.

Explanation

Many application domains related to GIScience have historically used spatially extended relational databases to perform geospatial reasoning (Battle & Kolas, 2012). The combination of efficient, stable data storage and retrieval with geospatial computation and indexing has been made possible by spatially extended databases. These advantages enable fast and accurate responses to queries such as "Which locations within a 2 km radius of a city's neighbors are likely to be flooded?". Notwithstanding their traditional adoption, relational databases continue to present challenges, including queries with numerous joins across entities, queries with variable properties used to connect data within a table or combine data from several tables, inference on datasets to support reasoning capabilities to identify errors and infer new conclusions based on the available data, qualitative spatial reasoning (Weiss, Karras, & Bernstein, 2008; Battle & Kolas, 2012), as well as the ability to share data schemas across multiple repositories for a common domain understanding.

These limitations pose significant challenges in the utilization and integration of data from disparate relational databases. In most cases, data are meticulously compartmentalized from each other, employing distinct data models, utilizing ad hoc vocabularies, and occasionally disseminated in non-standard formats by disparate owners (Vilches-Blázquez et al., 2014). However, as an increasing number of datasets are becoming available online,

the ability to extract information from combinations of these datasets is becoming increasingly important in order to present a more complete understanding of the data.

Over the past two decades, the Semantic Web, an extension of the existing Web in which information is given well-defined meanings to better enable computers and people to work together (Berners-Lee, Hendler, & Lassila 2001), has been increasingly used to address some long-standing problems in the geospatial domain (Huang & Harrie 2019). In more recent times, Knowledge Graphs (KGs) have emerged as an extension of Semantic Web practices and have been adopted by various companies such as Google, IBM, Facebook, or Microsoft (Noy et al. 2019). KGs capture and convey knowledge of the real world using graph-based representations to facilitate the creation, reuse, and retrieval of human- and machine-readable structured data (Cimiano & Paulheim 2017; Hogan et al., 2021). In this way, KGs have become one of the main ways to integrate diverse data (Hitzler et al. 2020), enabling the handling and linking of multiple heterogeneous datasets within a single system (Krötzsch & Thost 2016; Bellomarini, Sallinger, & Vahdati 2020; Vilches-Blázquez & Saavedra, 2022; Rowland et al., 2022).

The vision of the Semantic Web is that any representation should be represented as a graph associated with the Resource Description Framework (RDF), which is the standard knowledge representation language for this Web. Using RDF as the normal form of knowledge representation has certain advantages, such as avoiding the use of proprietary formats, harmonizing formats (e.g., databases, shapefiles, spreadsheets, and CSV files), facilitating data integration by breaking down data silos, among others (Vilches-Blázquez et al., 2014). This provides a reliable infrastructure for sharing, publishing and querying structured data on the Semantic Web (McDonald and Levine-Clark 2017). In this context, a basic triple in a graph can be interpreted as two nodes representing tangible entities (e.g., "River" and "Sea") linked by a relation (e.g., "flow into"). The aforementioned triples (nodes and relations) (e.g., :River :flowInto :Sea) facilitate the interconnection of disparate datasets, thereby generating graph-based representations, so-called KGs (Bellomarini, Sallinger, and Vahdati 2020).

An increasing amount of geospatial data is now available in the form of KGs that adhere to the Linked Data principles (Berners-Lee, 2006): (1) use of Uniform Resource Identifiers (URIs) as names for things; (2) use of the HyperText Transfer Protocol (HTTP) so that people can look up those names (URIs); (3) when someone looks up a URI, provide useful information using the RDF and SPARQL Protocol and RDF Query Language (SPARQL) standards; and (4) include links to other URIs so that they can discover more things.

This novel approach to publishing geospatial data adheres to the best practices for publishing, discovering, and utilizing spatial data on the Web (Tandy et al., 2023; van den Brink et al., 2019). These practices include the adoption of the Linked Data principles by geospatial resources, as well as their modeling through ontologies, their representation through RDF, their indexing by search engines, and their integration with other resources (Heath & Bizer, 2011). These actions are fundamental to the transformation of geospatial data into geospatial KGs. Furthermore, the aforementioned best practices (Tandy et al., 2023) support the FAIR principles (Wilkinson et al., 2016), a set of guidelines that emphasize the ability of computational systems to Find, Access, Interoperate, and Reuse data with no or minimal human intervention. For a detailed description of the FAIR principles, see FAIR Principles - GO FAIR.

Given these best practices and (Linked Data & FAIR) principles for the GIScience community, a plethora of resources, capabilities, and methodologies are being leveraged to create, exploit, consume, and capitalize on geospatial information through the strategic deployment of KGs. The multitude of available geospatial KGs exemplifies their indispensable role within the Web of Data Cloud acting as central nodes that connect events, people, and objects (Mai et al., 2019), and providing an ever-growing semantic representation of the richness of geospatial information (Jovanovik, Homburg & Spasić, 2021).

The geospatial KGs are stored and managed by triple stores, which are also known as RDF stores or knowledge bases. These triple stores are better equipped to handle several types of issues that relational databases are not designed to handle, including the aforementioned issues related to queries with multiple joins across entities, inference on datasets (Weiss, Karras & Bernstein, 2008), qualitative spatial reasoning, or the ability to share data schemas across multiple repositories.

In this context, SPARQL and GeoSPARQL are semantic query languages used to retrieve and process KGs. Regarding SPARQL, it is a semantic query language defined by the World Wide Web Consortium (W3C) that is capable of retrieving and manipulating data stored in RDF. The language is capable of expressing queries across a range of data sources, including those stored in native RDF format or viewed as RDF through the use of middleware. The following example SPARQL query is presented to illustrate the retrieval of cities and their associated countries from DBpedia, a cross-domain knowledge base that remains a core component of the Web of Data Cloud (Bizer et al., 2009).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX schema:<http://schema.org/>
PREFIX dbo: <http://dbpedia.org/ontology/>

select ?City ?Country
where {?City rdf:type schema:City .
?City dbo:country ?Country .
}

LIMIT 100
```

The SPARQL 1.1 Federated Query version enables users to construct queries and share data schemas through the utilization of ontologies (see related topics) on the Web. This process employs the same language layer (RDF) and enables the retrieval of data from disparate KGs distributed across multiple SPARQL endpoints. These endpoints are services that facilitate queries against data stored in RDF using the SPARQL language. The subsequent example presents a federated query that retrieves information from two knowledge bases, such as Wikidata and LinkedGeoData. This query employs geometrical information to ascertain the location of Automated Teller Machines (ATMs) belonging to the Bankcard-Servicenetz interbank network situated in Munich, Germany. The Wikidata initiative has compiled a collection of different examples of (federated) queries.

```
PREFIX lgdo: <http://linkedgeodata.org/ontology/>
PREFIX geom: <http://geovocab.org/geometry#>
PREFIX bif: <bif:>

SELECT ?atm ?geometry ?bank ?bankLabel WHERE {
  hint:Query hint:optimizer "None".

  SERVICE <http://linkedgeodata.org/sparql> {
    { ?atm a lgdo:Bank; lgdo:atm true. }
    UNION { ?atm a lgdo:Atm. }

    ?atm geom:geometry [geo:asWKT ?geometry];
        lgdo:operator ?operator.
    FILTER(bif:st_intersects(?geometry, bif:st_point(11.5746898, 48.1479876), 5)) # 5 km
around Munich
  }

  BIND(STRLANG(?operator, "de") as ?bankLabel)
  ?bank rdfs:label ?bankLabel.
```

The Open Geospatial Consortium (OGC) has defined GeoSPARQL (Car et al., 2024) as a means of describing and querying geospatial data on the Semantic Web. Next, a dummy example from the GeoSPARQL standard is shown to represent a feature and its associated geometry.

```
  # bank is part of the Bankcard service network, either via an explicit statement or
implicitly due to its legal form (unless explicitly excluded)

  { ?bank wdt:P527 wd:Q806724. }
  UNION { ?bank wdt:P1454 wd:Q5349747. }
  MINUS { wd:Q806724 wdt:P3113 ?bank. }
}
```

This example is retrieved from the GeoSPARQL standard.

The GeoSPARQL 1.1 specification comprises three principal elements: (i) a compact spatial domain OWL (Web Ontology Language) ontology, which enables the direct representation of geometrical forms in conjunction with spatial entities and the establishment of relationships between features through the use of spatial relations; (ii) SPARQL extension function definitions, which facilitate the calculation of relations between spatial objects; and (iii) a range of supplementary resources, including vocabularies of Simple Feature types and data validators.

Some developments and experiences in the GIScience community have demonstrated the feasibility of expressing federated queries over diverse KGs to retrieve and process geospatial data from disparate and distributed sources. In this line, Páez and Vilches-

Blázquez (2022) provided an approach to request, retrieve, and consume (geospatial) KGs available on different and distributed platforms supporting SPARQL 1.1 and GeoSPARQL standards. Various examples of federated queries using GeoSPARQL are provided herein. In addition, this approach enables the consumption of geospatial KGs through lightweight web applications or a GIS application, such as QGIS.

References

Battle, R., & Kolas, D. (2012). Enabling the geospatial Semantic Web with Parliament and GeoSPARQL. Semantic Web, 3(4), 355-370.

Bellomarini, L., Sallinger, E., Vahdati, S. (2020). Chapter 2 Knowledge Graphs: The Layered Perspective. In: Janev, V., Graux, D., Jabeen, H., Sallinger, E. (eds) Knowledge Graphs and Big Data Processing. Lecture Notes in Computer Science(), vol 12072. Springer, Cham.

Berners-Lee, T. (2006). Linked Data Design Issues.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 34-43.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia - A crystallization point for the Web of Data. Journal of Web Semantics, 7(3), 154-165.

Car, N.J., Homburg, T., Perry, M., Knibbe, F., Cox, S.J.D., Abhayaratna, J., Bonduel, M., Cripps, P.J., & Janowicz, K. (Eds.) (2024). GeoSPARQL – A Geographic Query Language for RDF Data. Open Geospatial Consortium.

Cimiano, P., and Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web. 8(3):489-508

Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. SpringerCham.

Hitzler, P., Janowicz, K., and Cudré-Mauroux, P. (2020). Leveraging Knowledge Graphs for Big Data Integration: the XI Pipeline. Semantic Web. 2020;11(1):13-17.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge Graphs. ACM Computing Surveys (CSUR), 54(4), Article No.: 71, Pages 1 - 37.

Huang, W., & Harrie, L. (2019). Towards knowledge-based geovisualisation using Semantic Web technologies: a knowledge representation approach coupling ontologies and rules. International Journal of Digital Earth, 13(9), 976–997.

Jovanovik, M., Homburg, T., & Spasić, M. (2021). A GeoSPARQL Compliance Benchmark. ISPRS International Journal of Geo-Information, 10(7), 487.

Krötzsch, M., Thost, V. (2016). Ontologies for Knowledge Graphs: Breaking the Rules. In: Groth, P., et al. The Semantic Web – ISWC 2016. ISWC 2016. Lecture Notes in Computer Science(), vol 9981. Springer, Cham.

Mai, G., Janowicz,K, Yan, B., and Scheider, S. (2019). Deeply integrating Linked Data with Geographic Information Systems. Transactions in GIS. 23: 579–600.

McDonald, J., & Levine-Clark, M. (Eds.) (2017). Encyclopedia of Library and Information Sciences. 4th ed. Boca Raton, FL: CRC Press.

Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., & Taylor, J. (2019). Industry-scale Knowledge Graphs: Lessons and Challenges. Communications of the ACM, 62(8):36 - 43.

Páez, O., & Vilches-Blázquez, L. M. (2022). Bringing Federated Semantic Queries to the GIS-Based Scenario. ISPRS International Journal of Geo-Information, 11(2), 86

Rowland, A., Folmer, E., Beek, W., & Wenneker, R. (2022). Interoperability and Integration: An Updated Approach to Linked Data Publication at the Dutch Land Registry. ISPRS International Journal of Geo-Information, 11(1), 51.

Tandy, J., van den Brink, L., Barnaghi, P., & Homburg, T. (Eds.) (2023). Spatial Data on the Web Best Practices. Open Geospatial Consortium & World Wide Web Consortium.

van den Brink, L., Barnaghi, P., Tandy, J., Atemezing, G., Atkinson, R., Cochrane, B., ... & Troncy, R. (2019). Best practices for publishing, retrieving, and using spatial data on the web. Semantic Web, 10(1), 95-114

Vilches-Blázquez, L. M., & Saavedra, J. (2022). A graph-based representation of knowledge for managing land administration data from distributed agencies – A case study of Colombia. Geo-Spatial Information Science, 25(2), 259–277.

Vilches-Blázquez, L. M., Villazón-Terrazas, B., Corcho, O., & Gómez-Pérez, A. (2013). Integrating geographical information in the Linked Digital Earth. International Journal of Digital Earth, 7(7), 554–575.

Weiss, C., Karras, P., & Bernstein, A. (2008). Hexastore: sextuple indexing for semantic web data management. Proceedings of the VLDB Endowment, 1(1), 1008-1019.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1):1-9.