Knowledge Explorer: Exploring the 12-Billion-Statement KnowWhereGraph Using Faceted Search (Demo Paper)

Zilong Liu
UC Santa Barbara &
University of Vienna
USA & Austria
zilongliu@ucsb.edu

Zhining Gu
Arizona State University
USA
zhiningg@asu.edu

Thomas Thelen
UC Santa Barbara
USA
thelen@nceas.ucsb.edu

Seila Gonzalez Estrecha
Michigan State University
USA
seilage@msu.edu

Rui Zhu
UC Santa Barbara &
University of Bristol
USA & UK
ruizhu@ucsb.edu

Colby K. Fisher
Hydronos Labs
USA
cfish25@gmail.com

Anthony D'Onofrio
Michigan State University
USA
anthon70@msu.edu

Cogan Shimizu
Wright State University
USA
cogan.shimizu@wright.edu

Krzysztof Janowicz
UC Santa Barbara &
University of Vienna
USA & Austria
janowicz@ucsb.edu

Mark Schildhauer
UC Santa Barbara
USA
schild@nceas.ucsb.edu

Shirly Stephen
UC Santa Barbara
USA
shirlystephen@ucsb.edu

Dean Rehberger
Michigan State University
USA
rehberge@msu.edu

Wenwen Li
Arizona State University
USA
wenwen@asu.edu

Pascal Hitzler
Kansas State University
USA
hitzler@k-state.edu

ABSTRACT

Knowledge graphs are a rapidly growing paradigm and technology stack for integrating large-scale, heterogeneous data in an AI-ready form, i.e., combining data with the formal semantics required to understand it. However, toolchains that support data synthesis and knowledge discovery through information organization, search, filtering, and visualization have been developed at a pace lagging knowledge graph technology. In this paper, we present Knowledge Explorer, an open-source faceted search interface that provides environmentally intelligent services for interactively browsing and navigating KnowWhereGraph. Currently one of the largest open knowledge graphs, KnowWhereGraph contains over 12 billion statements with rich spatial and temporal information from more than 30 data layers. With an extensive collection of facets, Knowledge Explorer enables spatial, temporal, full-text, and expert search with dereferencing functionality to support "follow-your-nose" exploration, and it allows users to narrow their search by selecting facets. Given the size of the underlying graph and dependency on GeoSPARQL, we have improved query performance by implementing Elasticsearch indexing, spatial query generation, and caching. Knowledge Explorer is capable of retrieving information within seconds, answering a wide variety of competency questions posed by researchers, humanitarian relief organizations, and the broader public, thus helping better perform tasks such as cross-gazetteer place retrieval and disaster assessment from global to local geographic scales.

CCS CONCEPTS

• **Information systems** → Search engine indexing; *Geographic information systems*; **Information retrieval query processing**; **Ontologies**; • **Human-centered computing** → **User interface programming**.

KEYWORDS

KnowWhereGraph, faceted search, knowledge graph, spatial query generation, environmental intelligence

SIGSPATIAL '22, November 1–4, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9529-8/22/11.
https://doi.org/10.1145/3557915.3561009

ACM Reference Format:
Zilong Liu, Zhining Gu, Thomas Thelen, Seila Gonzalez Estrecha, Rui Zhu, Colby K. Fisher, Anthony D'Onofrio, Cogan Shimizu, Krzysztof Janowicz, Mark Schildhauer, Shirly Stephen, Dean Rehberger, Wenwen Li, and Pascal Hitzler. 2022. Knowledge Explorer: Exploring the 12-Billion-Statement KnowWhereGraph Using Faceted Search (Demo Paper). In *The 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22), November 1–4, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3557915.3561009

1 INTRODUCTION

Knowledge graphs are a rapidly growing paradigm and technology stack for integrating Web-scale, heterogeneous data with rich formal semantics to ease data reuse and retrieval, and enable machine interpretability. However, the development of tools that support data synthesis and knowledge discovery through information organization, search, filtering, and visualization lags the pace of knowledge graph development, in general, and with respect to spatial and temporal data, in particular. This is especially apparent when dealing with extremely large knowledge graphs, and in this work, we focus on the exploration of KnowWhereGraph [5], currently one of the largest open knowledge graphs. In order to effectively utilize, navigate, and understand the statements contained within the graph, we aim to develop new tools and provide environmentally intelligent services on top of them.

In this work, we present Knowledge Explorer, a faceted search interface for interactively browsing and navigating KnowWhereGraph. Efforts to create similar, graph-backed semantic portals and frameworks have led to a number of open-source products whose goals are to be extensible and adaptable. Examples include SemFacet [1], Sampo-UI [3], Sextant [6], and LodLive [2]. Compared with many of these that also support faceted search over knowledge graphs, Knowledge Explorer enables more search functionalities under complex search conditions created by the facet selection of users. It sits on top of KnowWhereGraph, which consists of over 12 billion statements that provide rich spatial and temporal information from more than 30 data layers, including common spatial references such as ZIP codes, and thematic layers on topics such as soil condition. Empowered by custom information retrieval and visualization techniques, Knowledge Explorer can help address a wide variety of geographic tasks ranging from a search for local experts in humanitarian relief to natural disaster assessment and management.

Concretely, our contributions are as follows:

- Knowledge Explorer, an open-source faceted search interface for interactive navigation and browsing of the 12-billion-statement KnowWhereGraph through spatial, temporal, full-text, expert, and follow-your-nose search.
- A discrete global grid-based spatial query generation method for finding geospatial relations that are not explicitly stored in the graph.
- An extension to the selected dereferencing framework that supports customized information display about densely-connected entities, combined with faceted search.

The rest of our paper is organized as follows. Section 2 describes the architecture of Knowledge Explorer as well as our novel methods for improving its functionality: discrete global grid-based spatial query generation and Elasticsearch enhancement. In Section 3, we discuss how to use Knowledge Explorer. Finally, we conclude and propose future directions in Section 4.

2 KNOWLEDGE EXPLORER

2.1 Architecture

The faceted search interface was built using the AngularJS[1] framework and deployed with Node.js[2], backed by GraphDB[3] as the triplestore enhanced by Elasticsearch[4] (see Figure 1). NGINX[5] was used as the server and reverse proxy to direct traffic to the appropriate application. Leaflet[6] was used for map display and control. We extended our original Phuzzy.link [7] framework and deployed it for URI dereferencing. Both faceted search and dereferencing service retrieve information from the latest version of KnowWhereGraph through a publicly-accessible SPARQL endpoint[7]. The source codes, setup instructions, and further documentation of Knowledge Explorer can be accessed on GitHub[8].

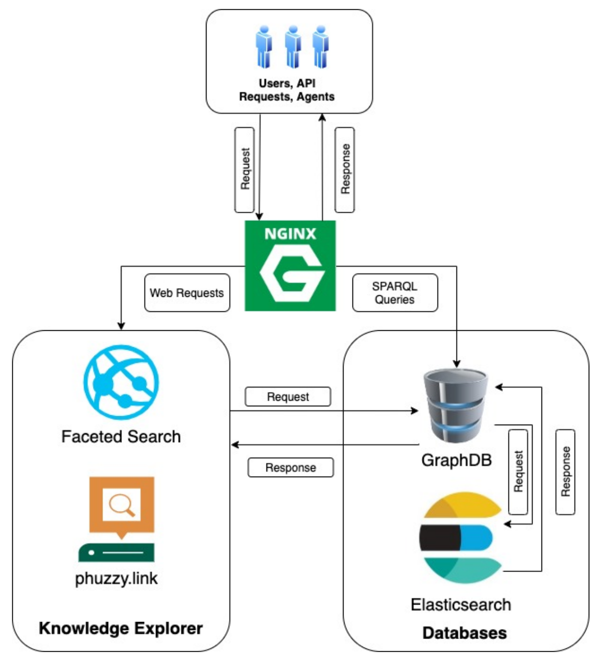

Figure 1: Architecture of Knowledge Explorer

2.2 Query Generation

Given a user selection of facets, SPARQL queries are automatically generated to retrieve information from the graph by following the current graph schema[9]. Faceted search queries are generated modularly by a function that calls sub-functions to construct their subsets,

[1]https://angularjs.org
[2]https://nodejs.org
[3]https://www.ontotext.com
[4]https://www.elastic.co
[5]https://www.nginx.com
[6]https://leafletjs.com
[7]https://stko-kwg.geog.ucsb.edu/graphdb
[8]https://github.com/KnowWhereGraph/kwg-faceted-search
[9]https://stko-kwg.geog.ucsb.edu/lod/ontology

and these queries can also be expanded to more complex ones. Because certain relations may not be explicitly materialised—and in our case, not all geospatial relations that directly link places from different gazetteers have been computed and stored in the triplestore—we implemented an information retrieval method based on multiple S2 cell[10] levels (up to level 13) that connect place instances through a subset of pre-computed RCC8 relations in our graph [9]. The SPARQL query shown in Listing 1 serves as an example for how we retrieve airports, buildings, and schools in Washington, USA (`kwgr:Earth.North _America.United_States.USA.48_1`) when there are no `kwg-ont:sfWithin` relations that link places specified in the Geographic Names Information System[11] (GNIS) with `kwg-ont:AdministrativeRegion`. The S2 cells where GNIS features of interest are located were identified first and then examined to determine whether they are spatially, and in this example, topologically linked to the administrative region of interest. To sum up, our querying strategy is to consider hierarchical grids as intermediate nodes to obtain possible answers, and we leverage discrete global grids as a unified spatial reference system to link disconnected geographic entities in the graph.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX usgs: <http://gnis-ld.org/lod/usgs/ontology/>
PREFIX kwg-ont: <http://stko-kwg.geog.ucsb.edu/lod/ontology/>
PREFIX kwgr: <http://stko-kwg.geog.ucsb.edu/lod/resource/>
select distinct ?entity ?label ?quantifiedName ?type ?typeLabel
{
    ?entity rdf:type ?type;
            rdfs:label ?label.
    ?type rdfs:label ?typeLabel.
    values ?type {usgs:Airport usgs:Building usgs:School}
    ?entity kwg-ont:sfWithin ?s2cell.
    ?s2cell rdf:type kwg-ont:KWGCellLevel13;
            kwg-ont:spatialRelation ?placesConnectedToS2.
    ?placesConnectedToS2 kwg-ont:sfWithin ?superPlacesConnectedToS2.
    filter (?placesConnectedToS2 in (kwgr:Earth.North_America.United_States.USA
        ↪ .48_1) || ?superPlacesConnectedToS2 in (kwgr:Earth.North_America.
        ↪ United_States.USA.48_1))
}
```

Listing 1: An example SPARQL query retrieving airports, buildings, and schools in Washington, USA

2.3 Elasticsearch Index Management

Although GraphDB does not support native text retrieval, we integrated it with Elasticsearch to enable full-text search and result ranking. An Elasticsearch GraphDB connector instance[12] was created to store and index fields and types mapped from the graph's properties (or property chains) and classes, respectively. In addition to enabling all literal fields to be analyzed, we synchronized all values of those properties with multiple values. The index is set to be automatically updated to be synchronized with the graph as it evolves.

2.4 Dereferencing

One of the visions for the World Wide Web is to link data together by providing landing pages and machine-readable representations of Uniform Resource Identifiers (URIs). URIs that link to more data are referred to as *dereferencable* URIs. We adopted the dereferencing framework Phuzzy.link, and extended its support for information display about densely linked entities and customized its display for specific triple patterns. This system also supports a "follow-your-nose" type of browsing for both machines and humans, with content negotiation in place for easy exploration of related entities. We linked faceted search results to the dereferencing interface so that users can find additional information about a particular entity which, in turn, opens more possibilities for subsequent searches.

3 USER INTERFACE INTERACTION

We provide public access to Knowledge Explorer[13], where users can explore data of three types: *People*, *Place*, and *Hazard*. Each tab has its own set of facets relevant to its theme. The *People* tab has facets that allow users to filter by a hierarchy of expertise topics ranging from *Business & Management Topic* to *Social Science- & Social Service-Related Topic*. Within the *Place* tab, users can search by *Administrative Region*, *Zip code*, *FIPS code*, *US Climate Division*, *National Weather Zone*, and feature types consistent with the data and ontology of an authoritative Linked Data version of the Geographic Names Information System, i.e., GNIS-LD [8]. These facets are specific to the United States and will be extended in the future to support place-based search involving other countries. The place-related facets are also available for selection within the *Hazard* tab, where users can narrow their search for hazards based on a start date, end date, and hazard types, which include *Earthquake*, *Fire*, *NOAA Hazard*, and *Smoke Plume Snapshot*. Both *Fire* and *NOAA Hazard* contain a hierarchy of hazard classes, respectively. When users select *Fire* or *NOAA Hazard*, facets with observable properties [4] (e.g., acres burned, mean difference Normalized Burn Ratio (dNBR), cases of direct death, cases of injury) will be displayed to allow users to continue narrowing their search for hazards.

Due to the stability and size of the graph, the facets are currently cached. Figure 2 shows an example of the main search interface under the *Hazard* tab. The interface is divided into three panels, including (1) facets, (2) a table display of search results with estimated count and clickable table cells linked to the dereferencing interface, and (3) an interactive map showing clustered markers illustrating the locations associated with the search results.

Knowledge Explorer supports five kinds of faceted search: spatial search, temporal search, full-text search, expert search, and follow-your-nose search. These searching functions can be combined to help answer one or a composite of the following questions and beyond: (1) *Where* is Place X? (2) *Where* did Hazard Y occur? (3) *When* did Hazard Y start and end? (4) *Who* has expertise in Expertise Topic Z?

Spatial Search. The S2 cell-based spatial query generation and the wide variety of place facets enables two kinds of geographic information retrieval tasks: cross-gazetteer place retrieval and hazard spatial search. In cross-gazetteer place retrieval, a user can either select or provide a place name as input for a place type under the *Place* tab, and multiple place type selection is enabled. Additionally, a user can select GNIS feature types to narrow the search scope for places, and place entities that match both place type and feature type selections will be returned. Hazard spatial search is built

[10] https://s2geometry.io/devguide/s2cell_hierarchy
[11] https://www.usgs.gov/us-board-on-geographic-names
[12] https://graphdb.ontotext.com/documentation/9.8/enterprise/elasticsearch-graphdb-connector.html
[13] https://stko-kwg.geog.ucsb.edu

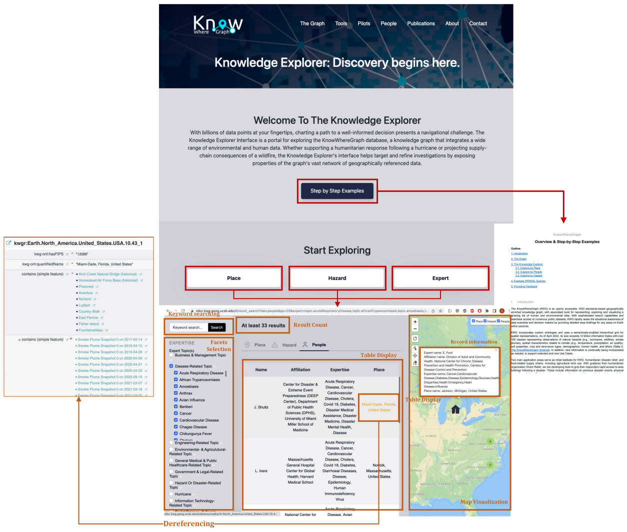

Figure 2: User interface interaction of Knowledge Explorer

upon cross-gazetteer place retrieval, allowing users to find hazards within places of interest under the *Hazard* tab.

Temporal Search. Besides spatial search for hazards, we also enable users to conduct a temporal search, i.e., to find specific hazard events within a given period of time, indicated by start and end dates. By coupling spatial, temporal, and observable property facet selection, along with the map visualization, Knowledge Explorer can help answer many geographic questions such as *"What fires occurred in California before 2022 and had fully-described burn scars over 100 acres?"*.

Full-Text Search. Given a keyword input by users, search results are ranked based on entity match scores that measure the similarity among entity labels so that the most relevant results are returned.

Follow-Your-Nose Search. The dereferencing service supports follow-your-nose search, providing a summary page with a map showing the location of an entity (if applicable) and a list of relations and interlinked entities. Users are allowed to navigate the links connected to an entity from the faceted search results.

Expert Search. While Knowledge Explorer has a dereferencing service similar to that of our Expert Finder System [10], the faceted search exposes more information about experts to end users and lets them search by keywords or expertise topics.

4 CONCLUSIONS AND FUTURE WORK

Knowledge Explorer is deployed on top of KnowWhereGraph, a knowledge graph with over 12 billion statements involving a diverse set of places, disaster occurrences, environmental variables, and experts, providing users with a wide range of pre-integrated, cross-domain data in an AI-ready form. The open-source explorer is built based on the AngularJS framework, Node.js, Phuzzy.link, Elasticsearch, and Leaflet to support efficient information retrieval with intuitive user interface design and interaction. Its components, including tree-hierarchy facets, an overview table, and a map visualization, help users grasp the graph's content quickly. The variety of search approaches, including spatial search, temporal search, full-text search, follow-your-nose exploration, and expert search, support knowledge discovery and navigation from multiple perspectives. In addition, we are using the interface together with other tools (e.g., knowledge graph-based GeoEnrichment service[14]) in the context of disaster mitigation and relief together with our project partner Direct Relief[15].

Future development of Knowledge Explorer will add more facets and filters, not only to broaden the scope of information that Knowledge Explorer can access, but also to refine its search results. As KnowWhereGraph evolves, we will also improve scalability of the system to handle more complex search scenarios and functionalities. More considerations will be given to the improvement of query efficiency, e.g., comparing geospatial query efficiency of Elasticsearch to that of GraphDB and benchmarking different GeoSPARQL functions. Finally, while we focused on KnowWhereGraph here, the explorer can be customized to other graphs.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. 2033521: "KnowWhereGraph: Enriching and Linking Cross-Domain Knowledge Graphs using Spatially-Explicit AI Technologies".

REFERENCES

[1] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Šarūnas Marciuška, and Dmitriy Zheleznyakov. 2016. Faceted search over RDF-based knowledge graphs. *Journal of Web Semantics* 37 (2016), 55–74.
[2] Diego Valerio Camarda, Silvia Mazzini, and Alessandro Antonuccio. 2012. LodLive, exploring the web of data. In *Proceedings of the 8th International Conference on Semantic Systems.* 197–200.
[3] Esko Ikkala, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2022. Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. *Semantic Web* 13, 1 (2022), 69–84.
[4] Krzysztof Janowicz, Armin Haller, Simon JD Cox, Danh Le Phuoc, and Maxime Lefrançois. 2019. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics* 56 (2019), 1–10.
[5] Krzysztof Janowicz, Pascal Hitzler, Wenwen Li, Dean Rehberger, Mark Schildhauer, Rui Zhu, Cogan Shimizu, Colby K Fisher, Ling Cai, Gengchen Mai, et al. 2022. Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. *AI Magazine* 43, 1 (2022), 30–39.
[6] Charalampos Nikolaou, Kallirroi Dogani, Konstantina Bereta, George Garbis, Manos Karpathiotakis, Kostis Kyzirakos, and Manolis Koubarakis. 2015. Sextant: Visualizing time-evolving linked geospatial data. *Journal of Web Semantics* 35 (2015), 35–52. https://doi.org/10.1016/j.websem.2015.09.004 Geospatial Semantics.
[7] Blake Regalia, Krzysztof Janowicz, and Gengchen Mai. 2017. Phuzzy. link: A SPARQL-powered Client-Sided Extensible Semantic Web Browser.. In *VOILA@ ISWC.* 34–44.
[8] Blake Regalia, Krzysztof Janowicz, Gengchen Mai, Dalia Varanka, and E Lynn Usery. 2018. Gnis-ld: Serving and visualizing the geographic names information system gazetteer as linked data. In *European Semantic Web Conference.* Springer, 528–540.
[9] Cogan Shimizu, Rui Zhu, Gengchen Mai, Colby Fisher, Ling Cai, Mark Schildhauer, Krzysztof Janowicz, Pascal Hitzler, Lu Zhou, and Shirly Stephen. 2021. A Pattern for Features on a Hierarchical Spatial Grid. In *The 10th International Joint Conference on Knowledge Graphs.* 108–114.
[10] Rui Zhu, Ling Cai, Gengchen Mai, Cogan Shimizu, Colby K Fisher, Krzysztof Janowicz, Anna Lopez-Carr, Andrew Schroeder, Mark Schildhauer, Yuanyuan Tian, et al. 2021. Providing Humanitarian Relief Support through Knowledge Graphs. In *Proceedings of the 11th on Knowledge Capture Conference.* 285–288.

[14] https://knowwheregraph.org/tools/geoEnrichment
[15] https://www.directrelief.org