

Critical Reading

WorldKG: A World-Scale Geographic Knowledge Graph

(Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021.)
(In Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21))

Presented By:
Andrew LaFortune and Zishuo Li

Spring 2022 CSCI 8715 Spatial Data Science Research
Prof. Shashi Shekhar



Background

Limitations of state of the art:

- “the representation of geographic entities ... within OpenStreetMap is highly heterogeneous, diverse, and incomplete”
- “cross-domain knowledge graphs are not well-populated with up-to-date geographic information”

Background:

- Gazetteer (master list of place names)
- GeoDataOnt (Sun et al.)
- OSMOnto - OSM tags description in an ontology
- OSM Semantic Network (Ballatore et al.)
- Neural Class Alignment (NCA)(Dsouza et al.)

Goal:

- Knowledge graph - the network of world entities in specifying relationships
- Automatic class inference of OSM nodes using NCA
- SPARQL can query knowledge graph to derive new information
 - distance between two points or two polygons



Problem Definition

Inputs:

- OpenStreetMap data (7+ billion **nodes**, 800+ million ways, 9 million relations)
 - Ways: linear collection of nodes (e.g. [road](#), [building](#))
 - Relations: collection of nodes and/or ways (e.g. bus route, [university](#))
- WikiData Knowledge Graph (1000+ **classes** organized hierarchically)
- DBPedia Knowledge Graph (1000+ **classes** organized hierarchically)

Output:

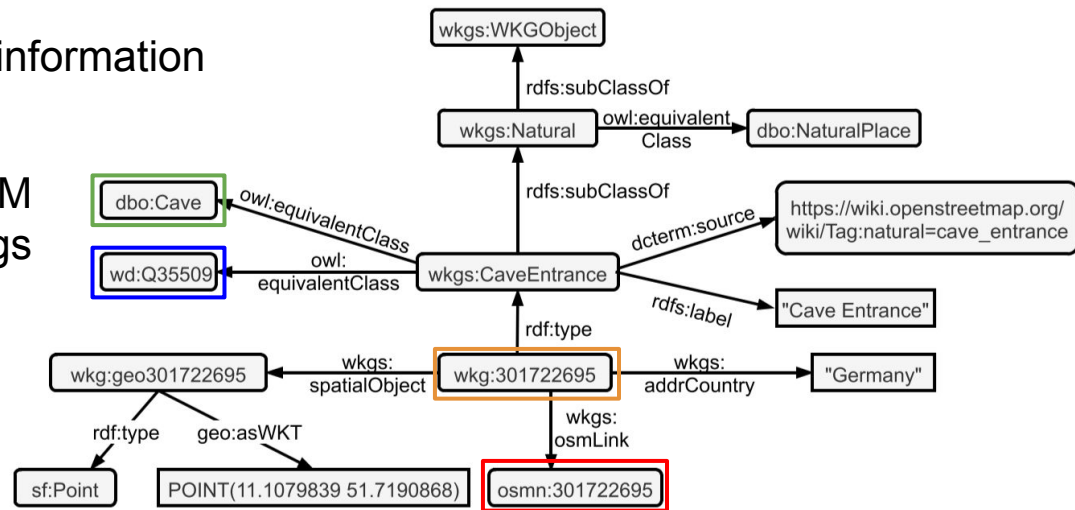
- WorldKG Knowledge Graph
 - OSM nodes linked to WikiData or DBPedia classes = **WorldKG entities**

Objectives:

- Large coverage on geographic information

Constraints:

- Only use Node entries from OSM
- Discard OSM Nodes with no tags



OSM stats: https://www.openstreetmap.org/stats/data_stats.html



Contributions

- *Proposed Contribution:*
 - The knowledge graph - WorldKG
 - WorldKG ontology in describing relationships between geographic entities
- *Unique Features:*
 - Hierarchical representation of OpenStreetMap ontology
 - Linkage of WikiData and DBpedia entities to OSM classes
- *Metrics:*
 - **Number of geographic entities:** 100 million plus
 - OSM has 6.8 BILLION
 - WikiData has 8.6 million
 - DBpedia has 1.2 million
 - **Wikidata and DBpedia class prediction accuracy:**
 - 99%
 - Is this a significant contribution?





DBpedia



yago
select knowledge



Linked
GeoData



	OpenStreetMap	Wikidata, DBpedia, YAGO	LinkedGeoData, YAGO2geo	WorldKG
knowledge graph	No	Yes	Yes	Yes
coverage on geographic entities	6.8 billion	8.6 million/1.2 million	few	0.1 billion
coverage on geographic classes	N/A	1000+	N/A	1000+

OSM Logo: <https://en.wikipedia.org/wiki/OpenStreetMap>, WikiData logo: <https://commons.wikimedia.org/wiki/File:Wikidata-logo-en.svg>, DBpedia logo: <https://en.wikipedia.org/wiki/DBpedia>, YAGO logo: [https://en.wikipedia.org/wiki/YAGO_\(database\)](https://en.wikipedia.org/wiki/YAGO_(database)), Linked GeoData Logo: <http://linkedgeodata.org/>, WorldKG logo: <https://www.worldkg.org/>



Key Concepts

Key Concepts	Simple Example
OpenStreetMap	Section 3.1 (pp. 2-3)
Knowledge Graphs	Section 3.2 (pp. 3)
WorldKG Ontology	Figure 1 (pp. 4)
Ontology Creation	Section 5.1, Figure 3 (pp. 5-6)
Knowledge Graph Creation	Section 5.2, Figure 3 (pp. 5-6)




Key Concept: OpenStreetMap

- OSM captures over 6.8 billion geographic entities
 - Nodes - geographic points (e.g. mountain peaks)
 - Ways - geographic lines (e.g. roads and rivers)
 - Relations - groups of elements (e.g. boundaries and bus routes)
- Each entity can be identified by:
 - Unique numeric identifier (id)
 - Element Type (Node, Way, Relation)
 - Location (latitude and longitude)
 - Tags (key-value pairs)
 - Collection of nodes (for Way and Relation)

Key	Value
<i>id</i>	<i>27384190</i>
<i>name</i>	<i>Zugspitze</i>
<i>natural</i>	<i>peak</i>
<i>summit:cross</i>	<i>yes</i>
<i>ele</i>	<i>2962</i>



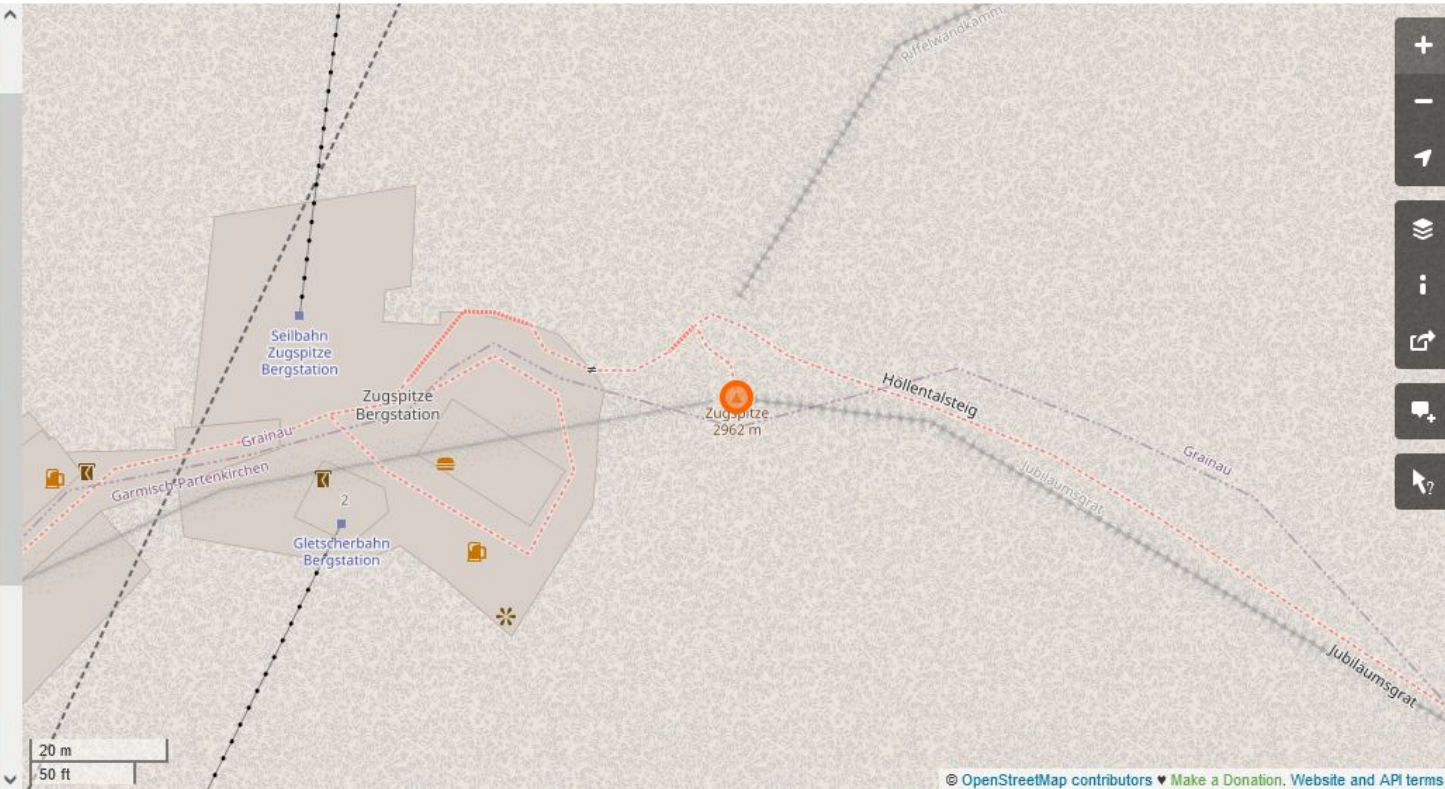
Element Type, **id**, **Location**, **Tags**,

 **OpenStreetMap** Edit History Export GPS Traces User Diaries Copyright Help About Log In Sign Up

Node: Zugspitze
(27384190)
Version #19
(no comment)
Edited over 3 years ago by [ManelG](#) · Changeset #64152490
Location: 47.4212150, 10.9862970

Tags

alt_name	Zugspitze (Ostgipfel)
ele	2962
importance	national
name	Zugspitze
natural	peak
prominence	1746
summit:cross	yes
wikidata	Q3375
wikipedia	de:Zugspitze



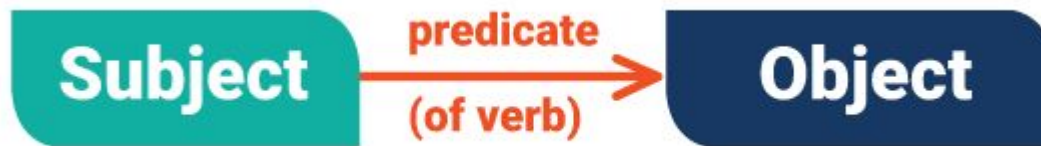
The map shows the Zugspitze peak at 2962m, with the Zugspitze Bergstation and Gletscherbahn Bergstation. It also shows the Grainau area, the Jubiläumssgrat, and the Höllemtalsteig. A scale bar indicates 20m and 50ft.

© OpenStreetMap contributors ♥ [Make a Donation](#), [Website](#) and [API terms](#)

Source: <https://www.openstreetmap.org/node/27384190>

Key Concept: Knowledge Graphs

- Represent real-world entities via nodes and edges in a graph
- Nodes describe objects and edges represent the relationships between those objects
- RDF (Resource Description Framework) is a common format for Knowledge Graph data



Subject	Predicate	Object
<i>Q3375</i>	<i>label</i>	<i>Zugspitze</i>
<i>Q3375</i>	<i>instance of</i>	<i>mountain</i>
<i>Q3375</i>	<i>coordinate</i>	<i>47°25' N, 10°59' E</i>
<i>Q3375</i>	<i>parent peak</i>	<i>Q15127</i>



Statements

instance of



mountain

 edit

► 1 reference

+ add value

image



 edit

Zugspitze Westansicht.JPG

3,264 × 2,448; 3.79 MB

media legend

Zugspitze, la muntanya més alta d'Alemanya. (Catalan)

▼ 0 references

+ add reference

+ add value

coordinate location



 edit

47°25'16.381"N, 10°59'10.730"E

<https://www.wikidata.org/wiki/Q3375>

Main page

Community portal

Project chat

Create a new item

Recent changes

Random item

Query Service

Nearby

Help

Donate

Lexicographical data

Create a new Lexeme

Recent changes

Random Lexeme

Tools

What links here

Related changes

Special pages

Permanent link

Page information

Concept URI

Cite this page

Key Concept: WorldKG Ontology

WorldKG prefixes

- **dct**terms: dublin core metadata initiative terms
 - e.g. source, accessRights, publisher
- **rdf**: Resource Description Framework
- **rdfs**: RDF Schema
- **owl**: Web Ontology Language (extends rdf)
 - e.g. equivalentClass, ObjectProperty
- **geo**: GeoSPARQL, maps geospatial data to rdf
- **osmn**: OpenStreetMap node
- **dbo**: DBpedia object
- **wd**: Wikidata
- **wkg**: WorldKG
- **wkgs**: WorldKG Schema
- **sf**: Simple Feature
 - e.g. Point, line, polygon
- **uom**: Units of Measure

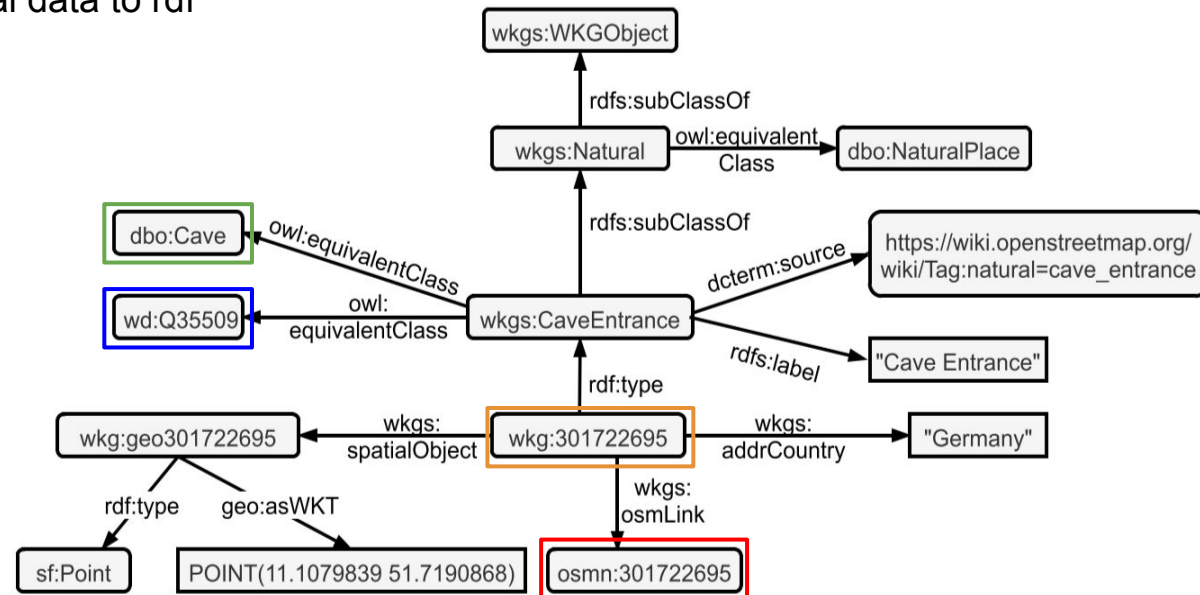
Node: Lessinghöhle (301722695)
Version #2
Bad Suderode
Edited over 12 years ago by hpod · Changeset #2054046
Location: 51.7190868, 11.1079839

Tags

name	Lessinghöhle
natural	cave_entrance



Right image source: <https://mapio.net/pic/p-45859522/>
Left: <https://www.openstreetmap.org/node/301722695>



Example: Restaurant Recommendation

- Three closest restaurants to Brandenburger Gate in Berlin, Germany
 - Use GeoSPARQL functions such as `bif:st_distance`

```
PREFIX uom:
<http://www.opengis.net/def/uom/OGC/1.0/>

SELECT ?closeObject ?restaurant
(bif:st_distance(?cWKT, ?fWKT, uom:metre)
 AS ?distance)
WHERE {
  ?poi rdfs:label "Brandenburger Tor".
  ?poi wkg:spatialObject [
    geo:asWKT ?cWKT
  ] .
  ?closeObject rdf:type wkg:Restaurant.
  ?closeObject rdfs:label ?restaurant.
  ?closeObject wkg:spatialObject ?fGeom.
  ?fGeom geo:asWKT ?fWKT .
}
ORDER BY ASC(
  bif:st_distance(?cWKT, ?fWKT, uom:metre))
LIMIT 3
```

Restaurant	Distance
"Hopfingerbräu im Palais"	0.128322
"Restaurant Quarré"	0.243953
"Lorenz Adlon Esszimmer"	0.247478



Exercise on Key Concepts

Exercise: What do knowledge graphs provide that are not captured in OSM?

1. Location data like Latitude and Longitude
2. Tags representing attributes of geographic entities
3. Structured contextual information including properties and relations between entities
4. An interactive map with geographic entities drawn on it

Exercise: Which of the following queries could be answered by WorldKG?

1. How long is the Washington Avenue Bridge?
2. What are the closest bus stops to Keller Hall?
3. What is the total area of the Sahara Desert?
4. When is it going to snow next?



Validation Method

- Sampling: 100 WorldKG entities from 10 classes, manual judgement of correct or wrong

(a) Wikidata

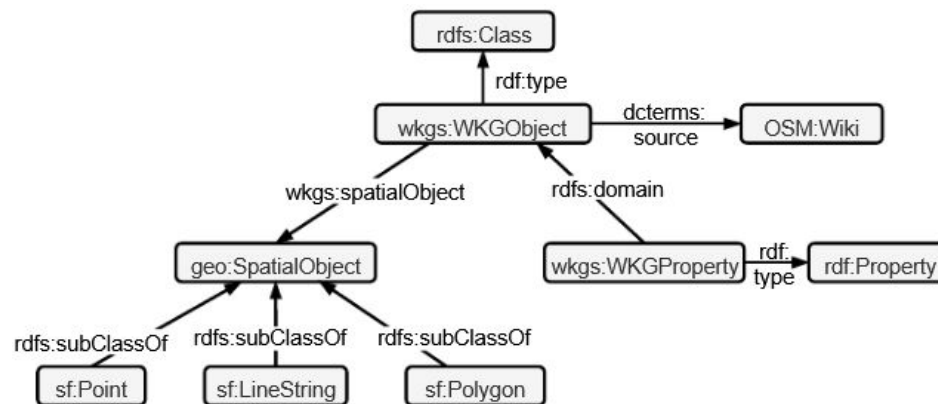
WorldKG class	WorldKG entities	Wikidata class	Wikidata entities	Correct	Wrong	Non-verifiable	Accuracy
Tomb	12849	Q381885	3076	97	1	2	98.98%
Monument	44503	Q4989906	23320	91	0	9	100.00%
Mineshaft	8453	Q556186	677	95	2	3	97.94%
BicycleRental	40914	Q61663696	1757	96	0	4	100.00%
TourismHotel	204291	Q27686	11152	97	0	3	100.00%

(b) DBpedia

WorldKG class	WorldKG entities	DBpedia class	DBpedia entities	Correct	Wrong	Non-verifiable	Accuracy
ManMadeTower/ PowerTower	2769981	Tower	2533	97	0	3	100.00%
City	10465	City	22600	100	0	0	100.00%
Museum	46955	Museum	7422	94	2	4	97.92%
AmenitySchool	424236	School	31867	100	0	0	100.00%
CaveEntrance	39525	Cave	615	91	0	9	100.00%

Assumptions and Refinement

- **Assumptions:** List assumptions made by the authors. Critique an assumption that you believe is unreasonable. What is the impact of removing this assumption on the solution proposed by the authors?
 - Dependency on OpenStreetMap
 - Continuing support on the project and other utility work
- **Refinements:** If you were to rewrite this paper today, what would you preserve and revise?
 - Keep one example throughout section 4 to show the progression of an OSM entry to a Turtle listing and ontology graph in WorldKG
 - Quantify high accuracy and large scale in the Abstract and Conclusion (over 97% accuracy and over 100 million entities)
 - remaining OGC simple feature types (collections of points, lines, and polygons)



Conclusions

- This work is useful because it creates a structure to map the largely unstructured data within OSM to a knowledge graph
- The emphasis on connection to WikiData and DBpedia is not abundantly useful since only the class name is connected
- Only a small portion of data in OSM is linked as Ways (lines) and Relations (collections, regions) are discarded
- The usability claim in the paper is not clear at the moment as the SPARQL endpoint is down
 - OSM and WikiData both have more intuitive user interfaces including a SPARQL query builder on WikiData and click-and-drag map on OSM

