WorldKG: A World-Scale Geographic Knowledge Graph

Presented By:
Andrew LaFortune
Zishuo Li

University of Minnesota
Driven to Discover℠

Problem Statement

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*
- Abstract: OpenStreetMap is a rich source of openly available geographic information. However, the representation of geographic entities, e.g., buildings, mountains, and cities, within OpenStreetMap is highly heterogeneous, diverse, and incomplete. As a result, this rich data source is hardly usable for real-world applications. This paper presents WorldKG - a new geographic knowledge graph aiming to provide a comprehensive semantic representation of geographic entities in OpenStreetMap. We describe the WorldKG knowledge graph, including its ontology that builds the semantic dataset backbone, the extraction procedure of the ontology and geographic entities from OpenStreetMap, and the methods to enhance entity annotation. We perform statistical and qualitative dataset assessment, demonstrating the large scale and high precision of the semantic geographic information in WorldKG.

	Closest sentence in Abstract	**Closest paragraph in Introduction**
Problem Definition		
Problem Significant		

Problem Statement

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*
- Abstract: OpenStreetMap is a rich source of openly available geographic information. However, the representation of geographic entities, e.g., buildings, mountains, and cities, within OpenStreetMap is highly heterogeneous, diverse, and incomplete. As a result, this rich data source is hardly usable for real-world applications. This paper presents WorldKG - a new geographic knowledge graph aiming to provide a comprehensive semantic representation of geographic entities in OpenStreetMap. We describe the WorldKG knowledge graph, including its ontology that builds the semantic dataset backbone, the extraction procedure of the ontology and geographic entities from OpenStreetMap, and the methods to enhance entity annotation. We perform statistical and qualitative dataset assessment, demonstrating the large scale and high precision of the semantic geographic information in WorldKG.

	Closest sentence in Abstract	**Closest paragraph in Introduction**
Problem Definition	2 ... representation of geographic entities… within OpenStreetMap is highly heterogeneous, diverse, and incomplete.	
Problem Significant	3. … the rich data source (OSM) is hardly usable for real-world applications	

Problem Statement

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*
- Abstract: OpenStreetMap is a rich source of openly available geographic information. However, the representation of geographic entities, e.g., buildings, mountains, and cities, within OpenStreetMap is highly heterogeneous, diverse, and incomplete. As a result, this rich data source is hardly usable for real-world applications. This paper presents WorldKG - a new geographic knowledge graph aiming to provide a comprehensive semantic representation of geographic entities in OpenStreetMap. We describe the WorldKG knowledge graph, including its ontology that builds the semantic dataset backbone, the extraction procedure of the ontology and geographic entities from OpenStreetMap, and the methods to enhance entity annotation. We perform statistical and qualitative dataset assessment, demonstrating the large scale and high precision of the semantic geographic information in WorldKG.

	Closest sentence in Abstract	**Closest paragraph in Introduction**
Problem Definition	2 ... representation of geographic entities… within OpenStreetMap is highly heterogeneous, diverse, and incomplete.	para 1: … representations of geographic entities in OSM are highly diverse, including few mandatory properties and numerous heterogeneous tags…
Problem Significant	3. … the rich data source (OSM) is hardly usable for real-world applications	Para 1: The tag-based structure of OSM data does not follow a well-defined ontology, significantly limiting automatic interpretation and use of OSM data in real-world applications.

Contributions

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*

	Closest sentences in Abstract	**Closest paragraph in Introduction**
Contribution Claim	4: This paper presents WorldKG - a new geographic knowledge graph aiming to provide a comprehensive semantic representation of geographic entities in OpenStreetMap	Para 4: … WorldKG - a novel comprehensive geographic knowledge graph build from the OSM dataset …
Evidence of Novelty and/or superiority over state of the art	5: … its ontology that builds the semantic dataset backbone, the extraction procedure of the ontology and geographic entities from OpenStreetMap, and the methods to enhance entity annotation.	Para 6: … semantically describes geographic entities and links them to the specific classes in Wikidata and DBpedia ontologies … access to WorldKG through a SPARQL endpoint … source code of the whole pipeline for WorldKG creation publicly available on GitHub.

Key Concepts

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*

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**

- Definition 3.1. An OSM corpus C = (N, T) consists of a set of nodes N representing geographic entities, and a set of tags T. Each tag $t \in T$ is represented as a key-value pair, with the key $k \in K$ and a value $v \in V$: $t = \langle k, v \rangle$. A node $n \in N$, $n = \langle i, l, T_n \rangle$ is represented as a tuple containing an identifier i, a geographic location l, and a set of tags $Tn \subset T$.

Key	Value
id	27384190
name	*Zugspitze*
natural	*peak*
summit:cross	*yes*
ele	2962

Key Concept: **Knowledge Graphs**

- *Definition 3.2.* A knowledge graph KG = (E, C, P, L, F) consists of a set of entities E, a set of classes $C \subset E$, a set of properties P, a set of literals L, and a set of relations $F \subseteq E \times P \times (E \cup L)$.
- Definition 3.3. The class of the entity $e \in E$ in the knowledge graph KG = (E, C, P, L, F) is denoted as: class(e) = $\{c \in C \mid (e, \text{rdf:type}, c) \in F\}$
- "Zugspitze" in Wikidata:

Subject	Predicate	Object
Q3375	*label*	*Zugspitze*
Q3375	*instance of*	*mountain*
Q3375	*coordinate*	$47°25'N, 10°59'E$
Q3375	*parent peak*	*Q15127*

Key Concept: **WorldKG Ontology**

Table 1: List of prefixes and namespaces used by WorldKG.

Prefix	Namespace
dcterms	http://purl.org/dc/terms/
geo	http://www.opengis.net/ont/geosparql#
osmn	https://www.openstreetmap.org/node/
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sf	http://www.opengis.net/ont/sf#
uom	http://www.opengis.net/def/uom/OGC/1.0/
wd	http://www.wikidata.org/wiki/
wkg	http://www.worldkg.org/resource/
wkgs	http://www.worldkg.org/schema/

- Classes: constructed hierarchically from the OSM feature list's key-value pairs.
- All keys are considered top-level classes and their values are sub-classes
  - e.g. natural=peak has class:natural and subclass:peak

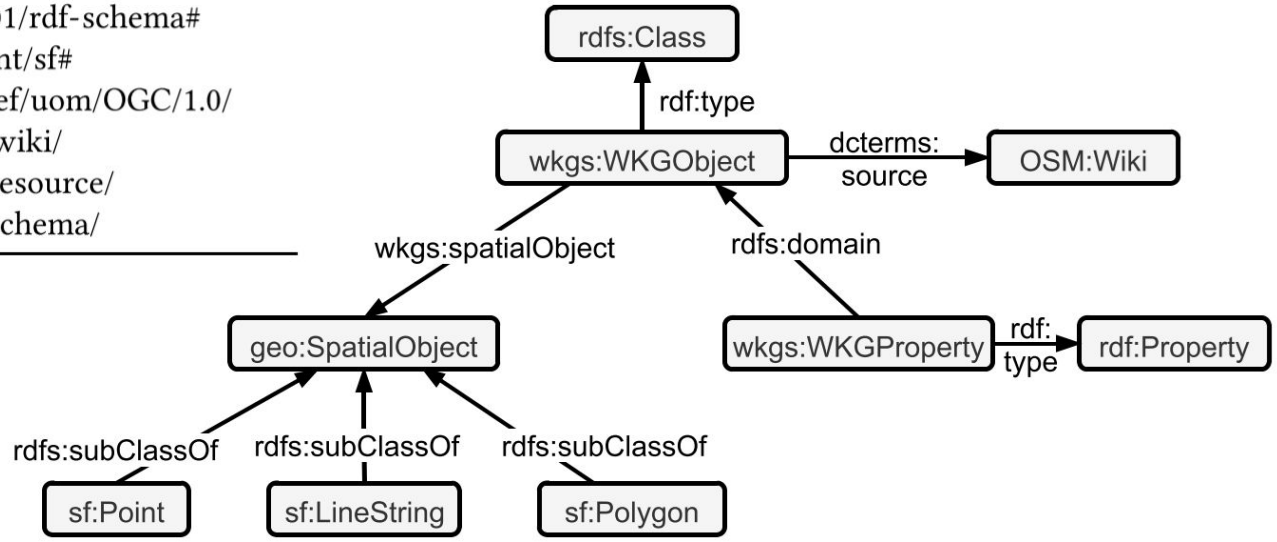

Figure 1: The WorldKG Ontology.

Key Concept: **WorldKG Ontology**

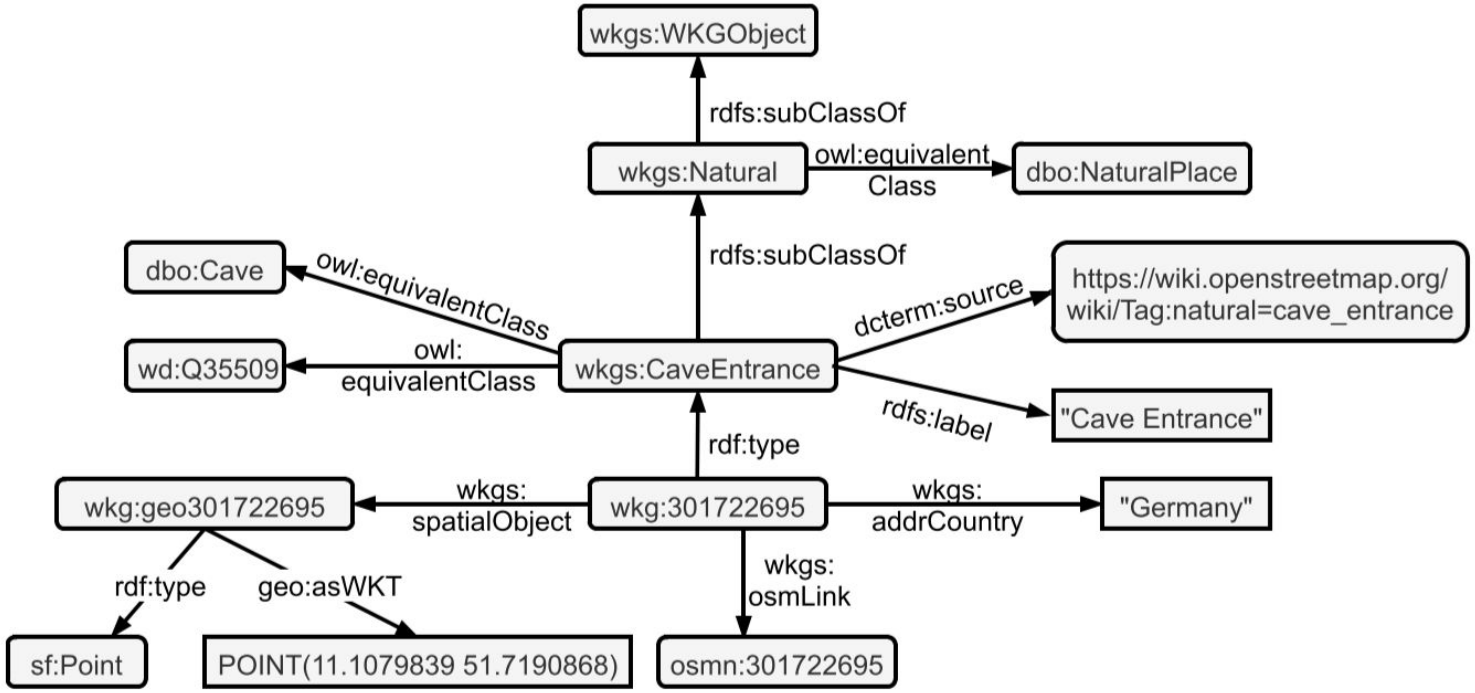

Figure 2: Example instantiation of the WorldKG ontology for a specific instance of `wkgs:CaveEntrance`.

Key Concept: **WorldKG Ontology Creation**

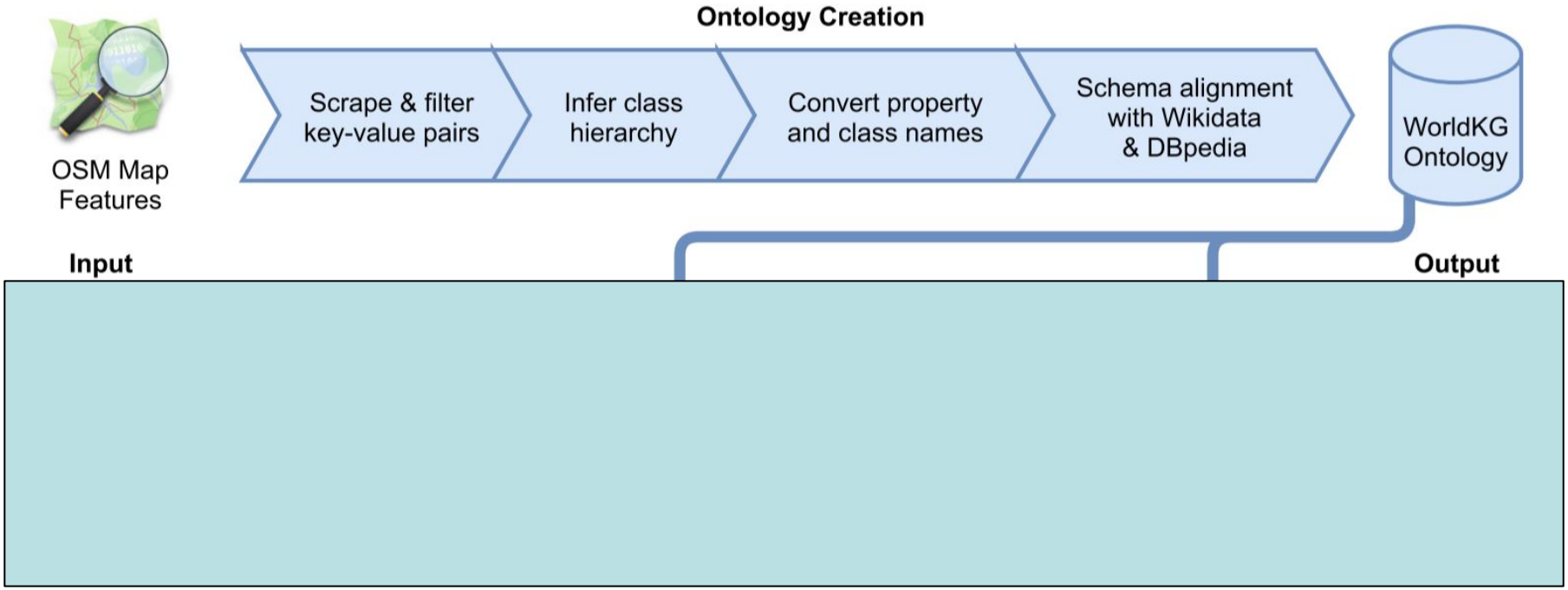

Figure 3: WorldKG ontology and knowledge graph creation process.

Key Concept: **WorldKG Graph Creation**

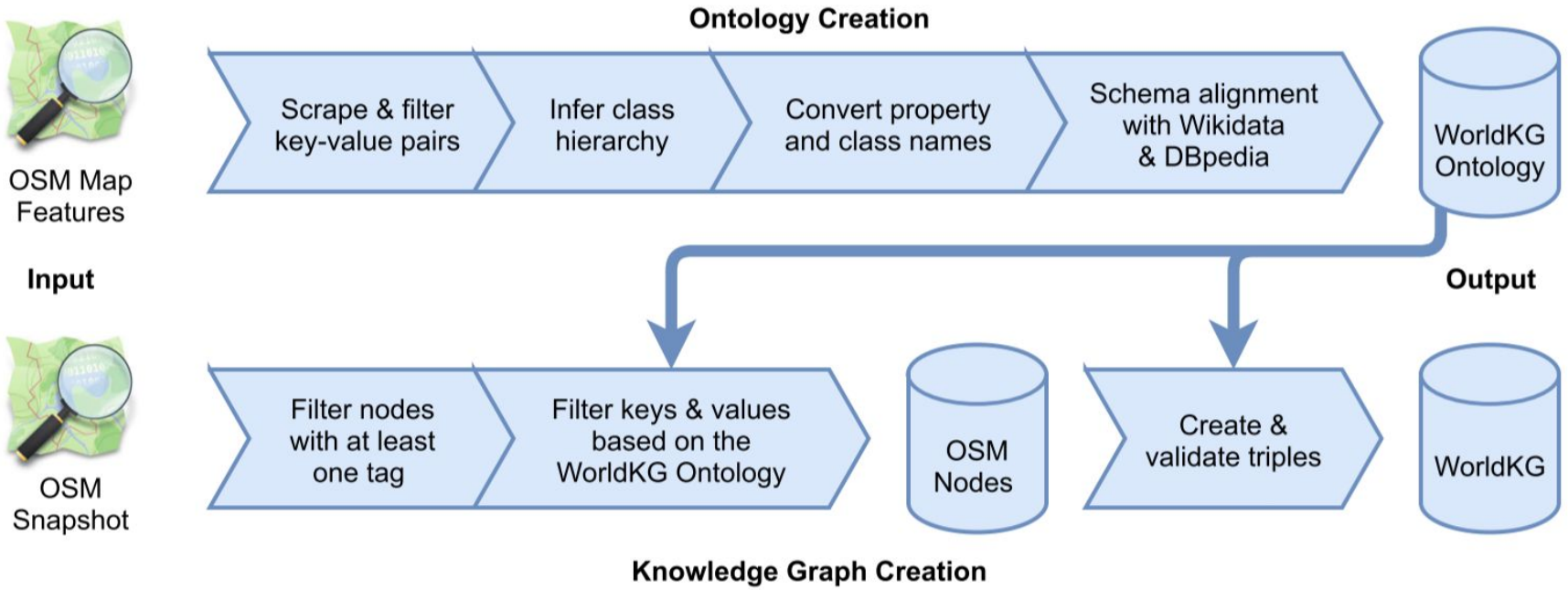

Figure 3: WorldKG ontology and knowledge graph creation process.

Exercise on Key Concepts

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*

Exercise: What is the primary source for the location data on WorldKG?

1. Wikidata
2. DBpedia
3. OpenStreetMap
4. Google Earth

Exercise: What would the Class be for an OSM object with the tag “railway=station” be?

1. railway
2. station
3. infrastructure
4. building

4. Validation methodology

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*

- Scientific Methodology:
  1. Statistics and Evaluation
  2. Use Cases

- What are the strengths and weaknesses of the methodologies?
  - Strengths: Statistics and evaluation results support the conclusion with facts; Use cases demonstrate how the contribution of the paper can be put in use in the real world.
  - Weaknesses: Statistics are limited to only prove what the paper currently did; Use cases are too specific to cover or generalize different scenarios.

- Is the choice of methodology appropriate?
  - Data model papers usually aim at more convenient representations
    This is demonstrated via examples and case studies

Statistics and Evaluation

- WorldKG knowledge graph statistics:

Quantity	Count
Total triples	828,550,751
Total entities	113,444,975
Top-level classes	33
Subclasses	1,143
Unique properties	1,820
Links to Wikidata classes	40
Links to DBpedia classes	21

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Statistics and Evaluation

- Evaluation results:
  - Quality of class alignment: Wikidata - 70% vs DBpedia 90% vs worldKG 100%
  - Quality of type assertion:
    - 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%

- 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%

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Use Cases

- Point-of-Interest Recommendation

```
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 wkgs:spatialObject [
    geo:asWKT ?cWKT
  ] .
  ?closeObject rdf:type wkgs:Restaurant.
  ?closeObject rdfs:label ?restaurant.
  ?closeObject wkgs:spatialObject ?fGeom.
  ?fGeom geo:asWKT ?fWKT .
}
ORDER BY ASC(
  bif:st_distance(?cWKT, ?fWKT, uom:metre))
LIMIT 3
```


UNIVERSITY OF MINNESOTA
Driven to Discover℠

4. Assumptions and 5. Refinement

- *WorldKG: A World-Scale Geographic Knowledge Graph. Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. Proceedings of the 30th ACM International Conference on Information & Knowledge Management.*
- **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?
  - Long-term-support on OpenStreetMap
  - Sustainability over 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)