Towards a Representation of Uncertain Geospatial Information in Knowledge Graphs

Lucie Cadorel
lucie.cadorel@inria.fr
Université Côte d'Azur, Inria, CNRS, I3S, KCityLabs
Sophia-Antipolis, France

Andrea G. B. Tettamanzi
andrea.tettamanzi@univ-cotedazur.fr
Université Côte d'Azur, Inria, CNRS, I3S
Sophia-Antipolis, France

Fabien Gandon
fabien.gandon@inria.fr
Inria, Université Côte d'Azur, CNRS, I3S
Sophia-Antipolis, France

ABSTRACT

This paper highlights the challenges of representing uncertain geospatial information in knowledge graphs. We propose to use Real Estate advertisements since professionals use a lot of vernacular and vague places in order to promote a house to their target audience. Then, we suggest to model local place names using fuzzy set theory. Finally, we discuss how to build a knowledge graph that represents extracted geospatial objects and their uncertainty.

CCS CONCEPTS

• **Information systems → Geographic information systems**.

KEYWORDS

Uncertainty, Fuzzy Sets, Knowledge Graph

ACM Reference Format:
Lucie Cadorel, Andrea G. B. Tettamanzi, and Fabien Gandon. 2022. Towards a Representation of Uncertain Geospatial Information in Knowledge Graphs . In *The 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs (GeoKG '22) (GeoKG '22), November 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3557990.3567588

1 INTRODUCTION

Digital gazetteers are dictionaries, often organized into a graph, that provide place names, their attributes and relations of millions of entries, and are often developed by government agencies from structured data. However, gazetteers do not contain all places because of their relative insignificance to a gazetteer covering a large area (e.g., world gazetteer) or their vernacular nature (e.g., abbreviations, non-official names). Vernacular places refer to place names with vague boundaries defined by regional culture (e.g., a "city center", "the old downtown", "West of Nice"). On the other hand, a myriad of unstructured (textual) data using natural language, such as travel blogs, social media or Real Estate advertisements, qualitatively refer to locations, and could be used to enrich gazetteers. Nevertheless, the geospatial knowledge extracted from textual data is often uncertain and imprecise since vague and unofficial place names are used. For example, Real Estate advertisements contain a lot of vernacular places because professionals do not often give the exact position of a house, and use unofficial place names and vague spatial expression to locate it. Also, location is one of the most valuable factors of purchasing and, Real Estate agents exaggerate boundary of place names that are popular and well reputed to promote a house to their target audience [6].

Although geospatial knowledge graphs (KGs) offer a good way of modeling and representing the knowledge extracted from texts and reason over it [1, 3–5], most existing geospatial KGs represent geospatial objects with sharp and established boundaries (e.g., polygons, points, lines, etc.). On the other hand, fuzzy set theory, that uses membership functions to define the degree to which each point in the space belongs to a fuzzy location, could deal with geospatial objects with imprecise boundaries. In this paper, we suggest to represent uncertain geospatial objects extracted from Real Estate advertisements, using fuzzy set theory, in a geospatial KG. We first present fuzzy sets to model uncertain boundaries of vernacular place names. Then, we outline the main concepts to represent the uncertain information in a geospatial KG.

2 FUZZY GEOSPATIAL OBJECTS

In fuzzy sets theory, elements of a set have degrees of membership, generally in the interval [0,1]. A major advantage of fuzzy set theory is that it returns an approximation of the location instead of a sharp area that should be the location. A fuzzy set A is characterized by its membership function μ_A , which describes the degree of membership of a point in the space to a fuzzy geographic set. Another advantage is that we can retrieve crisp sets from the membership function, such as core, support and α-cuts, defined as:

$$\begin{aligned} Core(A) &= \{x \in A; \mu_A(x) = 1\}, \\ Supp(A) &= \{x \in A; \mu_A(x) > 0\}, \\ \tilde{A}_\alpha &= \{x \in A; \mu_{\tilde{A}}(x) \geq \alpha\}, \\ \tilde{A}^+_\alpha &= \{x \in A; \mu_{\tilde{A}}(x) > \alpha\}. \end{aligned}$$

An α-cut of a set A, $\tilde{A}_\alpha$, is a crisp set where all the points belonging to the set have a degree of membership greater than or equal to α. While, a *strong* α-cut of A, $\tilde{A}^+_\alpha$, is a crisp set where all the points belonging to the set have a degree of membership greater than α The support is a particular strong α-cut where $\alpha = 0$. The core is a crisp set where all the points have a degree of membership equal to 1. A fuzzy set may have an empty core.

In our application on Real Estate advertisements, we extract a large number of vernacular place names. Thus, we estimate local place names applying the Kernel Density Estimation (KDE) algorithm to geotagged housing advertisements. KDE is a non-parametric estimation method that infers the shape of a variable from a sample, and gives a probability (density) for each point of the support. In our study, we chose Gaussian kernels to approximate the boundary of geospatial object. We then transform Gaussian kernels into a fuzzy set by taking the normalize density function as a membership function. Although this method is very suitable for vernacular places, it requires a great amount of data to get a reliable estimated footprint.

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
GeoKG '22, November 1, 2022, Seattle, WA, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9538-0/22/11…$15.00
https://doi.org/10.1145/3557990.3567588

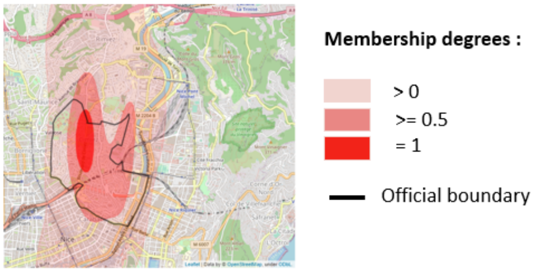

Figure 1: Estimated fuzzy representation from RE listings VS official boundary of *Cimiez* neighborhood in Nice, France

3 KNOWLEDGE REPRESENTATION OF UNCERTAIN AND VAGUE PLACES

In this section, we discuss the representation of geospatial information extracted from Real Estate advertisements, in the form of an RDF knowledge graph. The main challenge is to represent uncertain information and store spatial objects as fuzzy sets. We propose to add properties to *GeoSPARQL* vocabulary and to use *RDF-star* to express uncertainty. *GeoSPARQL* is a standard to represent and query geospatial linked data. Comparing to other geographic vocabularies[1] *GeoSPARQL* offers a flexible way to represent extracted knowledge thanks to its small ontology. Also, many spatial relations have been defined and implemented that facilitate geospatial applications. Regarding *RDF-Star*, it has been introduced to overcome the limitations of previous mechanisms such as RDF reification or singleton properties. It also provides a concise syntax extension to easily annotate RDF statements themselves. The key concepts and relations are described below.

Spatial Object. A spatial object is anything being or having a shape or position according to *GeoSPARQL*. In our study, we have to represent houses, place names, place types and footprints as *geo:SpatialObject*. A place type is a subclass of *geo:Feature* while a place name (or a house) is an instance of *geo:Feature*. Sometimes, Real Estate professionals refer to a place only with its place type (e.g., "near the train station"). So, we represent the place name as a blank node. Finally, we link a place to its location through the *geo:hasGeometry* property.

Uncertainty. Uncertainty arises when imperfect or unknown information is observed. In our situation, we have uncertainty at different levels of information. First, we automatically extract place names and place types that are not always official (e.g., "*La Banane*" in Cannes, France). Also, we estimate a footprint using Kernel Density Estimation. However, this method is not very precise and depends on the amount and reliability of data. We propose to model and add a confidence value to the relations using *RDF-Star*.

Vague Spatial Footprint. A spatial footprint is a geometrical representation of a spatial object. A spatial footprint is vague if its boundary is not crisp. In our work, we represent a spatial footprint as a fuzzy set, while the class *geo:Geometry* provides only geometries with sharp boundaries (e.g., Polygon, Point, Line). Furthermore, *GeoSPARQL* defines spatial relations (e.g., *geo:sfContains*, *geo:sfOverlaps*) between spatial objects that facilitate geospatial applications. In [2], the authors extend spatial relations to vague objects, especially using the core, support and α-cut of a fuzzy set. Also, GeoSPARQL makes it possible to add several geometries to the same place. So, we propose to define *ex:AlphaCut* and *ex:StrongAlphaCut* as subclasses of *geo:Geometry* and three properties *ex:hasCore*, and *ex:hasSupport*, *ex:hasAlpha*. Then, we could add the core, support and several α-cuts to a place to represent as reliable as possible its fuzzy boundary.

[1] e.g., https://schema.org/Place, https://dbpedia.org/ontology/Place or http://data.ign.fr/def/topo/20190212.htm#EntiteTopographique

```
@prefix  geo: <http://www.opengis.net/ont/geosparql#> .
@prefix  rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix  rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix  owl: <http://www.w3.org/2002/07/owl#> .
@prefix  xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix  ex: <http://example.org/> .

############# Classes
ex:Neighbourhood rdfs:subClassOf geo:Feature.

ex:AlphaCut rdfs:subClassOf  geo:Geometry.

ex:StrongAlphaCut rdfs:subClassOf  ex:AlphaCut.

ex:hasAlpha a rdfs:Property ;
    rdfs:domain ex:AlphaCut  ;
    rdfs:range xsd:double .

ex:Core rdfs:subClassOf ex:AlphaCut,
    [ a owl:Restriction ;
        owl:onProperty ex:hasAlpha ;
        owl:hasValue "1"^^xsd:double ] .

ex:Support rdfs:subClassOf ex:StrongAlphaCut ,
    [ a owl:Restriction ;
        owl:onProperty ex:hasAlpha ;
        owl:hasValue "0"^^xsd:double ] .

ex:hasCore rdfs:subPropertyOf geo:hasGeometry ;
    rdfs:range ex:Core .

ex:hasSupport rdfs:subPropertyOf geo:hasGeometry ;
    rdfs:range ex:Support .

############# Instances
<<ex:PlaceName1 a ex:Neighbourhood>>
    ex:confidence "0.8"^^xsd:double .

<<ex:PlaceName1 ex:hasCore ex:Core1; ex:hasSupport ex:Support1;
    geo:hasGeometry ex:AlphaCut1, ex:AlphaCut2 >>
    ex:confidence "0.4"^^xsd:double.

ex:AlphaCut1 a ex:AlphaCut ; ex:hasAlpha "0.7"^^xsd:double;
     geo:asWKT "MULTIPOLYGON␣(((43.6957␣7.280889,...,␣43.69578␣
          7.280882)))"^^geo:wktLiteral.

ex:AlphaCut2 a ex:AlphaCut ; ex:hasAlpha "0.3"^^xsd:double;
     geo:asWKT "MULTIPOLYGON␣(((43.695␣7.2808,...,␣43.6955␣
          7.280892)))"^^geo:wktLiteral.
```

Listing 1: Example RDF syntax of an uncertain place name and its fuzzy representation

4 CONCLUSION AND FUTURE WORK

In this paper, we devised a formal representation of uncertain and imprecise geospatial objects extracted from Real Estate advertisements in a knowledge graph. This paper constitutes an initial step in the knowledge representation of vague geospatial information, and we plan to continue this work by implementing it on a large dataset.

REFERENCES

[1] G. A. Atemezing and R. Troncy. 2012. Comparing Vocabularies for Representing Geographical Features and Their Geometry. In *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web, Boston, USA, November 12, 2012 (CEUR Workshop Proceedings, Vol. 901)*, Dave Kolas, Matthew Perry, Rolf Grütter, and Manolis Koubarakis (Eds.). CEUR-WS.org, 3–14.

[2] Arta Dilo, Rolf A. de By, and Alfred Stein. 2007. A System of Types and Operators for Handling Vague Spatial Objects. *Int. J. Geogr. Inf. Sci.* 21, 4 (jan 2007), 397–426. https://doi.org/10.1080/13658810601037096

[3] Alishiba Dsouza, Nicolas Tempelmeier, Ran Yu, Simon Gottschalk, and Elena Demidova. 2021. WorldKG: A World-Scale Geographic Knowledge Graph. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management.* ACM. https://doi.org/10.1145/3459637.3482023

[4] S. Wang et al. 2019. Geographic Knowledge Graph (GeoKG): A Formalized Geographic Knowledge Representation. *ISPRS Int. J. Geo Inf.* 8, 4 (2019), 184. https://doi.org/10.3390/ijgi8040184

[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] Grant McKenzie and Yingjie Hu. 2017. The "Nearby" Exaggeration in Real Estate (Position Paper). In *Proceedings of the Cognitive Scales of Spatial Information Workshop (CoSSI 2017)* (L'Aquila, Italy), Werner Kuhn, Dan Montello, Scott Freundschuh, Crystal Bae, Thomas Harvey, Sara Lafia, and Daniel Phillips (Eds.). 4–8. https://www.grantmckenzie.com/academics/COSSI_2017.pdf Workshop Article.