Representing spatial uncertainty and allowing for probabilistic topological functions with SUFF, an extension to GeoSPARQL

Nicholas J. Car[1,*]

[1]KurrawongAI, Brisbane Australia

Abstract

The Spatial Uncertainty for Features & Functions (SUFF) Model is a small extension ontology to GeoSPARQL that allows for spatial uncertainty, sometimes called fuzziness, in the representation of spatial geometries and for probabilistic topological calculations using them. It introduces ontological elements that allow standard GeoSPARQL geometries to be associated in new ways, allowing for different forms of handling uncertainty, including for visualisation. It describes how standard topological functions, such as those made available in Semantic Web form by GeoSPARQL, may be applied to collections of geometries and yield probabilistic results.

Keywords

GeoSPARQL, geospatial, Semantic Web, OWL, OGC, uncertainty, probabilistic, Levels of Measurement

1. Introduction

The GeoSPARQL standard [1] provides classes and properties for the representation of spatial information in Semantic Web [2] form. The main classes are `Feature` and `Geometry` with the former holding conceptual information about a spatial objects and the latter representations of its spatial projection, such as a polygon defined with coordinates. Predicates for indicating different forms of geometry projection serialization, topological relations between features or geometries and scalar spatial values, such as area, are also provided.

In addition to its ontology, GeoSPARQL defines a set of functions - topological, spatial aggregate and others - that systems can implement to calculate the relations between features and other spatial information.

GeoSPARQL does not specifically cater for spatial uncertainty or probabilistic topological functions: the only forms of geometry handled are the the *simple features* - [3] types of `POINT`, `LINESTRING`, `POLYGON`, `MULTIPOLYGON` *etc.*. - and topological functions return binary results - for example either a polygon is disjoint with another or it is not.

GeoSPARQL is not, by itself, completely sufficient for any particular task and does not attempt to provide a user with all the ontological elements they would need for specialised work, instead, GeoSPARQL is expected to be used with other Semantic Web models which together provide the

---

GeoLD2024: 6th International Workshop on Geospatial Linked Data at ESWC 2024

*Corresponding author.

nick@kurrawong.ai (N. J. Car)
https://kurrawong.ai (N. J. Car)
0000-0002-8742-7730 (N. J. Car)

necessary elements. For example, if a user wanted to represent the area of a lake, GeoSPARQL provides the class `Feature` to represent the lake - a spatial object - the `hasArea` predicate to indicate a scalar area value but no predicates to indicate units of measure or particualr measures: these may be taken from a dedicated metrology ontology, perhaps *Quantities, Units, Dimensions & Types* [1].

This allows GeoSPARQL to be extended by uses to meet particular needs and this paper does this for spatial uncertainty.

This paper describes how, by use with a small extension to GeoSPARQL of only few predicates known as the *Spatial Uncertainty for Features & Functions* (SUFF) ontology that we have defined, [2], GeoSPARQL can represent fuzzy spatiality and allow for probabilistic topological functions as well as providing several other capabilities, such as position obfuscation with specific tolerances.

2. Motivation

The motivation for this work is encountered regularly in spatial data work: many projects need to represent spatial uncertainty - fuzziness - and would benefit from topological functions that work well with uncertain spatial data.

Some examples of spatial objects for which the author has recently needed to represent with uncertain positions are: Australian Indigenous (Aboriginal) peoples' traditional land areas; mineral occurrence areas; informal, named geographical objects; and species distribution maps.

None of these are well served with crisp - certain - representations of position. For example, using a crisp polygon to represent the extent of an Aboriginal traditional land area will not capture the fact that no such areas ever had/have precise boundaries.

A second example is mineral occurrences in the earth's crust that change in concentration over distance. Their extents are usually estimated by sampling - surface collection or drillhole sample extraction. No simple points or individual polygons can represent gradients and extent estimations.

Data analysers may be called on to work with information that would best be represented with uncertainty but none is given. This forces them to add fuzziness to visualisations by blurring polygon boundaries or by representing point locations with a polygon - perhaps circle - perhaps itself with a blurred boundary. This addition of fuzziness on top of the original data may be done differently by different analysts or by using different spatial data software and no two representations of the data including these fuzzy additions are guaranteed to be the same.

This paper, and the SUFF ontology, aim to both provide the ontological elements necessary for representations of several forms of uncertain spatial information within data and instructions on how to extend common topological functions to work with those representations. This will allow data analysts to be provided with all the information necessary to produce consistent fuzzy representations of spatial information, regardless of who they are or what tools they are using.

---

[1] https://qudt.org

[2] https://w3id.org/profile/suff

3. SUFF Model

The SUFF Model is a small ontology that describes hot to represent spatial uncertainty. To do this it defines only a few new predicates and no new classes and instead indicates how to use existing classes and predicates from GeoSPARQL and other ontologies to do this.

An overview of the SUFF Model, showing reused and new model elements, is given in Figure 1.

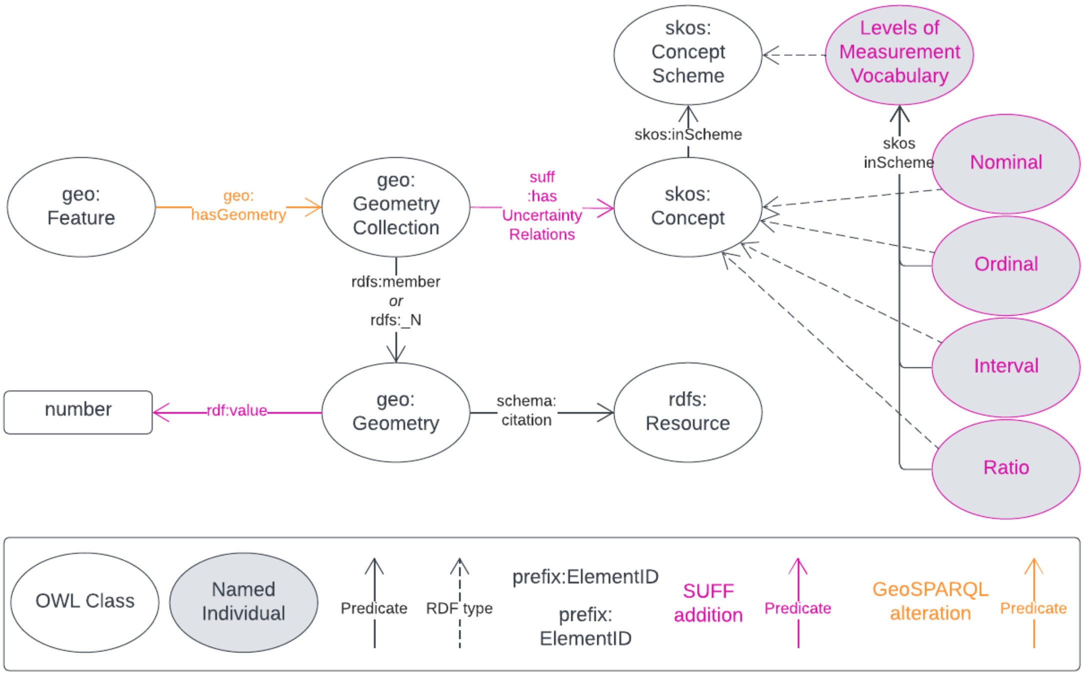

Figure 1: An overview of the SUFF Model.

The SUFF Model is available online at https://w3id.org/suff and detailed information about the model, it's element definitions, examples of use and validation of data created conforming to it are all given there.

3.1. Model Logic Overview

The basic premise of the SUFF Model is that multiple GeoSPARQL `Geometry` objects can be linked to a `Feature` to contain different representations of its position which may correspond to different confidences in the position. Different relationships between the multiple geometries can also be used to allow for different interpretations of their relative confidence of position. These different interpretations are also mapped to specific visualisations so that the relationship from data to visualisation is deterministic.

The model describes how topological functions to determine spatial relationships, such as the such such as the *simple features* [3] family of relations, can be performed against the features linked to multiple geometries.

3.2. Use of multiple geometries

Multiple GeoSPARQL `Geometry` objects are grouped in a GeoSPARQL `Geometry Collection` which is then linked to a GeoSPARQL `Feature` through GeoSPARQL's `hasGeometry` predicate but the range value of the predicate must be loosened to allow for this use, i.e. not just `Geometry`.

Fuzziness in the feature's position is then given by blurring (interpolating) between the individual geometry's boundaries. The manner of blurring can be indicated by linking a concept from the Levels of Measurement vocabulary to the geometry collecting with the SUFF `hasUncertaintyRelations` predicate and certainty at a particular geometry's boundary may be indicated directly by the data creator using the standard `rdf:value` predicate, or it may be calculated by the geometry's relative position in the collection.

Figure 2 shows multiple polygons, A, B, C & D giving different estimates for the position of a `Feature`. The reasons for particular geometries might be given as citations of scholarly works, or descriptive text, or as complex data objects, such as sampling result .

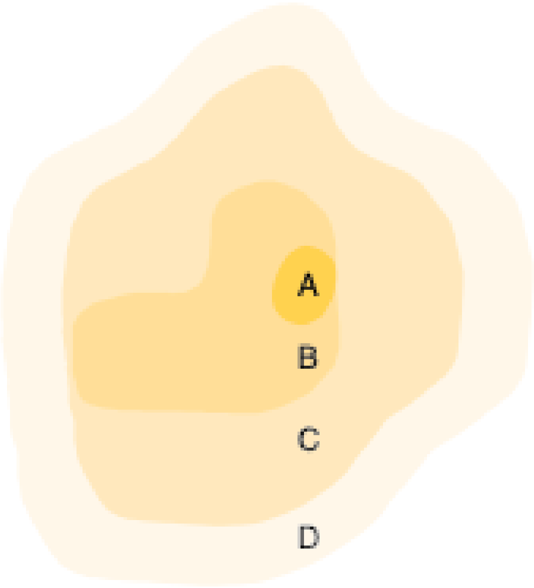

Figure 2: Multiple polygons - one per Geometry linked to a single Feature can indicate variation in fuzziness: the further apart the boundaries are in any place, the greater the spatial uncertainty there. The right side A/B boundary is comparatively certain.

3.3. Levels of Measurement

Levels of Measurement is a classification system used to indicate the nature of information within variables [4]. Four levels are defined: *Nominal*, *Ordinal*, *Interval* and *Ration* (see the Levels of Measurement vocabulary for the definitions) and different calculations about the relative values of geometries in a collection can be made, based on the level indicated for the collection (see the SUFF Model's Rules Section for details).

If the level for a given geometry collection is not given, the *Nominal* is assumed and each geometry is assigned equal certainty with the 0 to 1 range divided by the number of geometries. Visually these would be stacked so coincident geometries' opacities would be added.

If ordering is specifically assigned to the geometries, such as https://www.w3.org/TR/shacl/'s `order` predicate, then the level of the collection is at least *Ordinal* and certainty per geometry is the 0 - 1 range divided by no. geometries + 1 times the geometry order, so for 3 geometries: 0.25, 0.5 & 0.75. *Interval* ordering sees specific uncertainty values given by the user and *Ratio* values are given but a 0 and a 1 are required too.

3.4. Evidence for position

Evidence for a particular geometry may be given using schema.org's `citation` predicate or other conventional form of referencing. For evidence to be used for ordering for the *Ordinal* form, it must be numeric and the value indicator predicate as well as order direction must be given. This will replace use of `rdf:value`.

3.5. Functions

Topological relations defined in GeoSPARQL may be computed between spatial objects with fuzzy geometries and the method is described in the SUFF Model's Rules Section. A summary is that for the calculation of a relation between a feature `A` with fuzzy position and a standard feature `B`, the relation must be calculated for each geometry of `A`'s in its linked collection and `B` and then an answer may be given with a probability based on the certainty of the highest geometry / `B` result, or the lowest, depending on the particular relation.

Consider the topological relation *Simple Features* `Within` [3]: If polygon `A` is *within* the outermost 1 of 4 polygons with the *Ordinal* level for feature `B`, then we may say `A` is *within* `B` with a probability of 0.25 since 1 / 4 (polygons) is 0.25. If `A` is *within* the two outermost polygons of feature `B` and second outermost polygon is assigned the certainty of 0.4 using SUFF's `certainty` predicate, then we can say `A` is *within* `B` with a probability of 0.4.

A graphical representation of some *within* calculations is given in Figure 3.

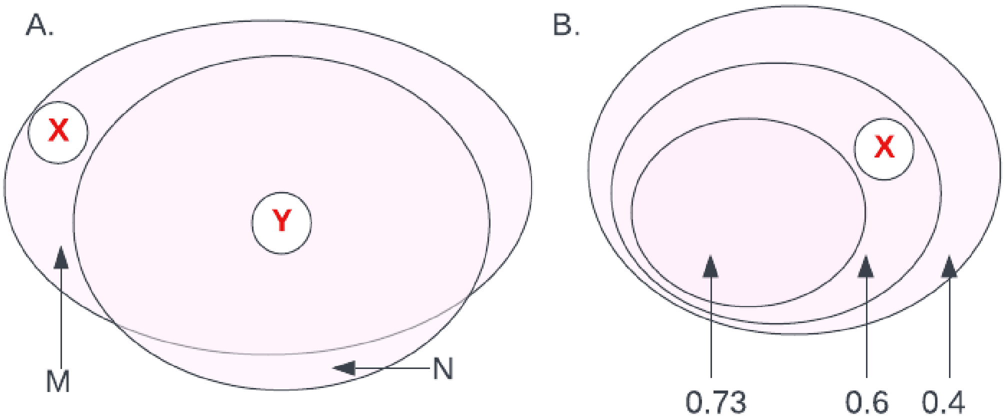

Figure 3: A: Polygon 'X' is *within* the feature represented with multiple pink geometries with a probability of 0.5 since it is within 1 of 2 geometries and no certainty information is give. Polygon 'Y' is *within* the feature with a probability of 1 since it is within 2 of 2 geometries. B: 'X' is *within* the feature with probability 0.6 since the most certain geometry it is within is that of 0.6.

4. Conclusions

The SUFF Model is a very small extension to GeoSPARQL. It introduces very few novel modelling elements and addresses problems long solved by many spatial analysis and many software packages. However, it moves the point of uncertainty definition from the user of spatial data to the data itself and thus has the potential to allow for systematically-presented fuzzy data. It also provides a method for standard topological functions calculations with fuzzy data.

References

[1] Nicholas J. Car, Timo Homburg, Matthew Perry, John Herring, Frans Knibbe, Simon J.D. Cox, Joseph Abhayaratna, Mathias Bonduel, OGC GeoSPARQL - A Geographic Query Language for RDF Data, OGC Implementation Standard OGC 22-047, Open Geospatial Consortium, 2023. URL: http://www.opengis.net/doc/IS/geosparql/1.1.

[2] T. Berners-Lee, J. A. Hendler, O. Lassila, The semantic web: A new form of web content that is meaningful to computers will unleash a revolution, Scientific American (2001). URL: https://doi.org/10.1038/scientificamerican0501-34.

[3] J. R. Herring, Simple feature access - Part 1: Common architecture, OpenGIS Implementation Standard, Open Geospatial Consortium, 2010. URL: http://www.opengis.net/doc/is/sfa/1.2.1.

[4] W. Kirch (Ed.), Level of MeasurementLevel of measurement, Springer Netherlands, Dordrecht, 2008, pp. 851–852. doi:10.1007/978-1-4020-5614-7_1971.