

Intersection-based Spatial Annotation of Trajectories with Linked Data

Tales P. Nogueira¹, Hervé Martin², Rossana M. C. Andrade¹

¹ Group of Computer Networks, Software Engineering and Systems
Federal University of Ceará
Fortaleza, Brazil

² Univ. Grenoble Alpes
CNRS, Grenoble INP, LIG
F-38000 Grenoble, France

tales@great.ufc.br, herve.martin@imag.fr, rossana@great.ufc.br

Abstract. *Smart cities are characterized by providing new services through Information and Communications Technologies. However, it is important to gather data from citizens to discover new knowledge about certain aspects of a city. One example of a rich domain for collecting data in a smart city is exploring the use of mobile fitness applications. Users usually record outdoor activities in the form of trajectories, which can later be acquired for further analysis. In this work, we leverage Semantic Web technologies to propose an annotation algorithm that segments trajectories according to their spatial context. We demonstrate how the method works and the impact of OpenStreetMap related ontologies in the annotation process.*

1. Introduction

Smart cities can monitor and integrate critical infrastructures, e.g., streets, bridges, tunnels, rails, subway, airports, ports, communications, water, power, buildings) to optimize its resources, plan maintenance, and monitor security risks while it maximizes the quality of services to their citizens. A crucial component for the fulfillment of smart cities is the Digital City concept, where Information and Communication Technologies (ICT) play an essential role (Cocchia 2014).

The multiple services that may be available in a smart city can generate a significant volume of data. Other factors contribute to the complexity increase of both generation and use of data in smart cities. One of them is the heterogeneity of formats that can be adopted by the various information sources. Besides, many smart city applications use cases require the association between the main information to the moment when it was sensed (e.g., a temperature sensor not only stores a place's temperature value in degrees but also the moment in time when that data was captured). Another data type that is usually valuable to many kinds of analysis is the location where something is sensed. From the characteristics explained above, one can conclude that spatiotemporal analysis of data generated from the various sensors that form the IoT infrastructure of a smart city is an essential task in the process of delivering useful services to citizens.

Some smart city applications are focused on the improvement of specific aspects of the user's daily routine. For instance, a mobile phone user can choose among a myriad

of applications for tracking her fitness activities, for selecting a touristic itinerary based on her preferences, and for helping to cope with traffic during her daily commute. In this context, new techniques for modeling, processing, analysis, and visualization of spatiotemporal data generated by these applications are necessary.

Trajectory data mining comprises many activities (Zheng 2015) that prepare collected data from location devices for further analysis. A critical task consists of annotating raw data with relevant information. This semantic enrichment process can focus on different aspects of trajectories according to the application domain. A usually useful aspect for trajectories is the spatial context, that aims to describe the environment where the trajectory happened.

FrameSTEP (Nogueira et al. 2018) is a framework for trajectory annotation that leverages Semantic Web technologies for enriching trajectories based on episodes. It is built around the STEP ontology¹ and allows the construction of annotation algorithms that segment trajectories into instances of their ontology.

In this work, we focus on annotating trajectories according to their spatial surroundings. We present the Intersection-based Spatial Annotation algorithm (ISA), which takes advantage of Semantic Web technologies for selecting terms related to keywords informed by the user to retrieve important contextual elements to enrich the trajectory. The method also considers spatial and temporal constraints to select the most important geographic features.

The remainder of this paper is organized as follows: Section 2 presents the theoretical background on trajectory annotation. Section 3 defines important Semantic Web tools used in this work: the LinkedGeoData project and the Open Semantic Network. Section 4 contains related work on trajectory annotation focused on spatial context annotation. Section 5 introduces the main contribution of our work, the Intersection-based Spatial Annotation (ISA) algorithm. Section 6 concludes this work with a final discussion on the results and perspectives for future work.

2. Trajectory Annotation

The term “annotation” is rather general and can be the objective of many steps in trajectory analysis. Some basic movement characteristics movement have been elicited by Andrienko et al. (2008), that divided attributes according to whether they could be expressed in time instants or intervals. At a given moment (i.e., an instant in time) we can measure the timestamp itself, the entity’s position in space, the movement’s direction, the speed at which the movement happens, the moving object’s change of the direction (turning angle), the acceleration, among other attributes. As *overall characteristics*, Andrienko et al. (2008) cite attributes that can be computed over the whole trajectory or its parts, such as its shape, traveled distance, duration, and various statistics over these characteristics. It is also of interest the relationships among trajectories. Being from the same moving object or distinct entities, the possible analysis includes similarity measures, spatial and temporal relations like co-incidence in space and time in many forms (ordered and unordered colocation, full or partial co-existence in time, full or lagged co-incidence in space-time). A similar classification has also been explored by Dodge et al. (2008)

¹www.purl.net/net/step

as primitive movement parameters that can be used to calculate primary and secondary derivatives.

Several trajectory definitions can be used to express the many steps that are needed to fulfill the semantic enrichment pipeline. Spaccapietra et al. (2013) argue that usually there is a *movement track* representing the whole lifespan of an object that may originate *trajectories*, which are the segments of an object's movement that are of interest for a given application. Still according to Spaccapietra et al. (2013), among the possible representations of a trajectory, one can have a continuous, a discrete, or a stepwise representation, being the latter the result of the implementation of a step function that maps a time interval to a finite set of annotations.

Another possible view of this process allows the definition of different trajectory types, e.g. *Spatio-temporal Trajectory*, *Structured Trajectory*, and *Semantic Trajectory*. Yan et al. (2012) have used these terms to describe the various transformations that lead to the enrichment of trajectory data with semantic information.

3. LinkedGeoData and Open Semantic Network

OpenStreetMap (OSM) aims to build a free geographic database of the world (Bennett 2010) based on three data structures – nodes, ways, and relations – and a tagging system. Tags are key-value pairs that can be associated with any node, way or relation. There is no formal enforcement on which tags can be used and how they are used, meaning that anyone can create a new tag when needed. However, the community of contributors generally consults the OSM Wiki² where guidelines are described for the most popular tags. Examples of tags are: *highway = primary*, *amenity = parking*, *waterway = river*.

The LinkedGeoData (LGD) project translates OSM (OpenStreetMap) data originally stored in a relational database into RDF (Resource Description Framework) triples following a lightweight ontology (Stadler et al. 2012). The project proposes two different namespaces: `lgd`³ and `lgdo`⁴ in order to separate instances from the ontology. The mapping from OSM data to RDF is performed based on the unique IDs of OSM features. For instance, the Eiffel Tower has ID 5013364 in OSM and it is represented as the resource `lgd:way5013364` in LGD.

The LinkedGeoData ontology was built through a semi-automated process that consisted of the creation of *tag mappers* that match specified tag patterns to transformation functions written in Java for creating RDF descriptions. For instance, a tag mapper could specify that the OSM features containing the tag `{amenity: school}` had to be mapped to the `lgdo:School` concept or, in a more general case, any feature containing the tag key `tourism` had to be mapped to the `lgdo:TourismThing` concept in the LGD ontology.

Taking into consideration the large variety of tags in OSM, this mapping process can also be viewed as a filtering mechanism if we consider that only tags that hold important semantic value are selected. In this sense, features containing no tags or only informative tags like *name*, *source*, *comment*, *note*, etc. can be discarded as they have no

²https://wiki.openstreetmap.org/wiki/Map_Features

³<http://linkedgeo.org/triplify/>

⁴<http://linkedgeo.org/ontology/>

important semantic information about the feature (e.g. whether it is a building, a park, a river, etc.).

There are two ways of directly querying the LGD database apart from direct downloads of data dumps. The first possibility is a Virtuoso endpoint⁵ and the other one is a web-based interface⁶. Both LGD endpoints allows the submission of SPARQL queries, however, for spatial queries, Virtuoso built-in functions⁷ should be used as this triple store's version do not support OGC's GeoSPARQL standard for spatial queries yet. Another option for querying OSM is LOSM, proposed by Ragone et al. (2016), which is in a beta version.

In our automatic annotation algorithm, we enrich trajectories based on OSM geographic features with relevant tags. In order to achieve this, we have used the OSN RDF dataset proposed by Ballatore et al. (2012). By exploring the hyperlinks among the pages that describe OSM tags in the OSM Wiki⁸, the authors have conceptualized a directed graph containing OSM keys and tags, lexical descriptions, relationships between tags, general internal links, and links to Wikipedia pages. More importantly for our case, the OSN graph also contains links to equivalent classes in the LGD ontology.

By linking the spatial annotation episodes to OSN vocabulary, it is possible to benefit from the semantic similarity rankings between the tags created with the help of co-citation algorithms. For instance, it becomes possible to infer that different geographic features tagged with *waterway = river* and *natural = lake* have some level of similarity as both concepts are related to the concept *natural = water*.

4. Related work

Annotating trajectories with its spatial features is a basic and important task. For many applications, this allows building higher level abstractions that describe the spatial context of the path besides the pure latitude and longitude information.

In the work of Yan et al. (2011), the authors propose an algorithm for trajectory annotation with RoIs (Regions of Interest). However, no standard format of external data is mentioned, a rather limited dataset is used (Swisstopo from Switzerland) and, from what can be inferred of their work, all intersections are computed features disregarding the type of polygons. This last aspect can be especially time-consuming for long trajectories.

Yan et al. (2012) have proposed a classification of spatial annotation types according to the geometric shape of the element to be annotated. Annotation of Lines of Interest consisted of identifying correct road segments for constrained network movement of cars, and classifying modes of transportation for mixed movement (e.g., walking, public transport, cycling). The first challenge was handled with map matching techniques and part of its output was used along with average velocity and acceleration to infer transportation modes.

The *Trajectory annotation with RoIs* algorithm works by receiving a trajectory

⁵www.linkedgeodata.org/sparql

⁶www.linkedgeodata.org/snorql

⁷<http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtGeoSPARQLEnhancementDocs>

⁸http://wiki.openstreetmap.org/wiki/Map_Features

and a set of semantic regions as input. These regions can be either in free form (e.g., OpenStreetMap polygons), or in a grid-like format, which is the case of the Swisstopo⁹ dataset. The algorithm proceeds to compute the intersection among the inputted trajectory and all episodes and then creates tuples in the form $\langle region, entering\ time, leaving\ time, type\ of\ region \rangle$ that is attached to the trajectory afterward.

The authors highlight that the spatial join can be computed only for selected episodes depending on requirements. However, this aspect is not further explored. Moreover, it is implicit that all intersections between the trajectory and polygons or grid cells are computed regardless of the geographic feature type or metadata associated with the external data, which may increase processing time. While this can be useful to annotate episodes in a parameter-free setting and to discover information previously unknown, it is desirable to allow some kind of filtering for the cases of applications that are only concerned with a well-defined set of feature types.

In other research, Furletti et al. (2013) have proposed an algorithm for automatic annotating trajectories with user activities. PoIs (Points of Interest) at stops are inferred based on categories, opening hours, characteristics of the moving object (e.g. maximum walking distance), among other parameters. Then, a probability of the PoI being visited is computed.

Nogueira et al. (2018) have annotated trajectories using OSM data with the help of the LGD ontology. The work presented in this paper can be considered an improvement over the annotation process shown by the authors as it served as the basis for our proposal. The main difference being the integration of the OSN into the annotation process. This way, the algorithm allows the discovery of OSM tags that might be relevant for the application's domain and that might have been unknown before the algorithm's application.

Regarding other related work, one can notice that most of them aim at the identification of RoIs of a moving object when it is not moving. On the other hand, our method does not focus on any particular aspect of the trajectory such as its stops or PoIs.

5. ISA: Intersection-based Spatial Annotation

In this section, we present the ISA (**I**ntersection-based **S**patial **A**nnotation) algorithm for segmenting trajectories based on the intersection of a trajectory with its geographic context. Additionally, the user has the option of filtering which types of features are selected based on OSM tag keys. The data flow of the algorithm is depicted in Figure 1 where the shaded parallelograms represent the parameters of the ISA algorithm.

The ISA method starts with the verification of the presence of keys informed by the user. If one or more keys are inputted, it checks whether they should be expanded. In the positive case, a query in the OSN RDF graph is performed to retrieve all other terms that are related to each of the informed keys.

For instance, if the “keys” parameter contains only the *leisure* key and the “expand” option is enabled, then the expansion mechanism that uses the OSN would incorporate 53 new keys to be used in the remainder of the process. However, some of

⁹<https://www.swisstopo.admin.ch/>

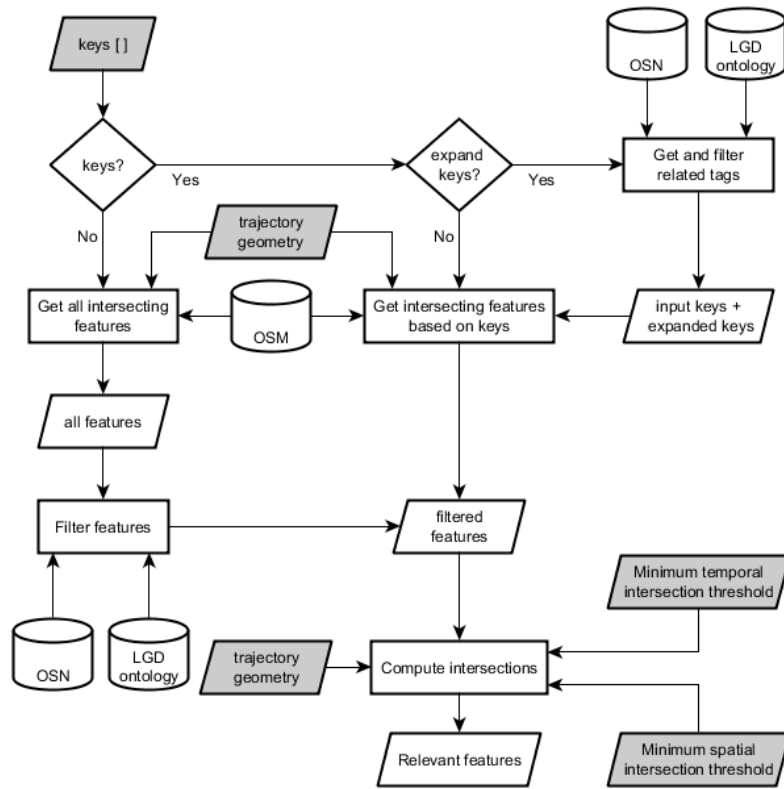


Figure 1. Data flow of the intersection-based spatial annotation algorithm with OSM features.

these new keys might not have a useful semantic value for a spatial context description – e.g. *opening_hours*, *addr:housenumber*, *contact:website* – therefore, we have incremented this query with a filter to include only OSN concepts that have an exact match with some concept from the LGD ontology. In the case of the example, the *leisure* key would be expanded with the following other keys that have a match in LGD: [*'highway'*, *'playground'*, *'route'*, *'tourism'*, *'man_made'*, *'sport'*, *'waterway'*, *'harbour'*, *'shelter'*, *'building'*, *'natural'*, *'fishing'*, *'landuse'*, *'boundary'*, *'amenity'*, *'barrier'*, *'shop'*], totaling 17 keys instead of the 53 without the LGD filter. The query that was built by the ISA algorithm for this example is shown in Listing 1.

The next step of ISA consists in retrieving intersecting OSM features which have tags matching the keys informed by the user – and the expanded keys if this option was chosen. If no keys are previously informed by the user, it means that there is no interest in narrowing down the types of features returned from OSM and the algorithm should make no distinction of retrieve features based on their tags.

Then, OSM is queried to retrieve all geographic features that have polygons intersecting the trajectory. This implementation can be done using the Overpass API¹⁰ or through the LGD SPARQL endpoint.

Once the intersecting features are retrieved, each one is verified to check if at least one of its tags have a corresponding concept mapped in OSN and if that concept

¹⁰<http://overpass-api.de>

```

1 PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
2 PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
5
6 SELECT ?subject ?relatedTerm ?lgdConcept
7 {
8   ?subject skos:related ?relatedTerm .
9   ?relatedTerm skos:exactMatch ?lgdConcept .
10
11  FILTER (REGEX(STR(?subject),
12              "http://spatial.ucd.ie/lod/osn/term/k:(leisure).*"))
13  FILTER (STRSTARTS(STR(?lgdConcept),
14              "http://linkedgeodata.org/ontology"))
15 }

```

Listing 1. SPARQL query that finds all OSN concepts that are related to the *leisure* key. Additionally, only related terms that have an exact match in LGD ontology are returned.

has an exact match in the LGD ontology. This process is similar to the one performed in the keys expansion procedure and helps to maintain only features that have a defined semantic meaning in the context of the LGD ontology.

A polygon representation for each OSM feature is created considering a buffer region of 20 meters to account for GPS accuracy errors. Then, the intersections among the trajectory and the polygons are computed. The result of this computation yields, for each polygon, which trajectory points are contained in them.

The final result of the ISA method retrieves the OSM features that reach at least one of the two criteria represented by the Minimum Temporal Intersection Threshold and the Minimum Spatial Intersection Threshold. These are parameters that range from 0 to 1 and represent the percentage that each feature must meet to be considered as relevant. This avoids the selection of OSM features where the moving object has passed quickly or has barely intersected.

With the output of the ISA algorithm, it is possible to create instances of the STEP ontology. The OSM features can be seen as contextual elements and OSN concepts can be easily linked to these elements as they follow a well-defined link structure, e.g., for the tag *waterway = riverbank*, the corresponding OSN link is <http://spatial.ucd.ie/lod/osn/term/k:waterway/v:riverbank>. An example of RDF code generated by the FrameSTEP framework can be seen in Listing 2.

ISA has been implemented in the FrameSTEP framework (Nogueira et al. 2018) and is available in a public repository¹¹ containing its demonstration. In Figure 2, it is possible to see some examples of OSM features that are selected in different situations. For these cases, no keys have been informed to the ISA algorithm and the spatial and temporal thresholds have both been set to 0.2, meaning that at least 20% of the trajectory's duration and distance should be intersected by an OSM feature in order to that feature to be considered as relevant.

¹¹<https://github.com/talespaiva/framestep>

```

1 <rdf:Description
2   rdf:about="http://example.com/resource/2njWC77BrXkr26ZXKxPZZA">
3 <owl:sameAs
4   rdf:resource="http://linkedgeo.org/triplify/way4014810"/>
5 <rdf:type
6   rdf:resource="http://purl.org/net/step#ContextualElement"/>
7 <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
8   Parc Paul Mistral
9 </rdfs:label>
10 <skos:related
11   rdf:resource="http://spatial.ucd.ie/lod/osn/term/k:leisure/v:park"/>
12 </rdf:Description>

```

Listing 2. Excerpt of RDF/XML file automatically generated by the ISA algorithm.

It is also shown the LGD ontology concepts that have been identified based on the OSM tags. It is possible to notice that the LGD concepts also helps to describe the spatial context, besides serving its role of filtering important categories of geographical features, such as parks, forests, beaches, farms, vineyards, etc.

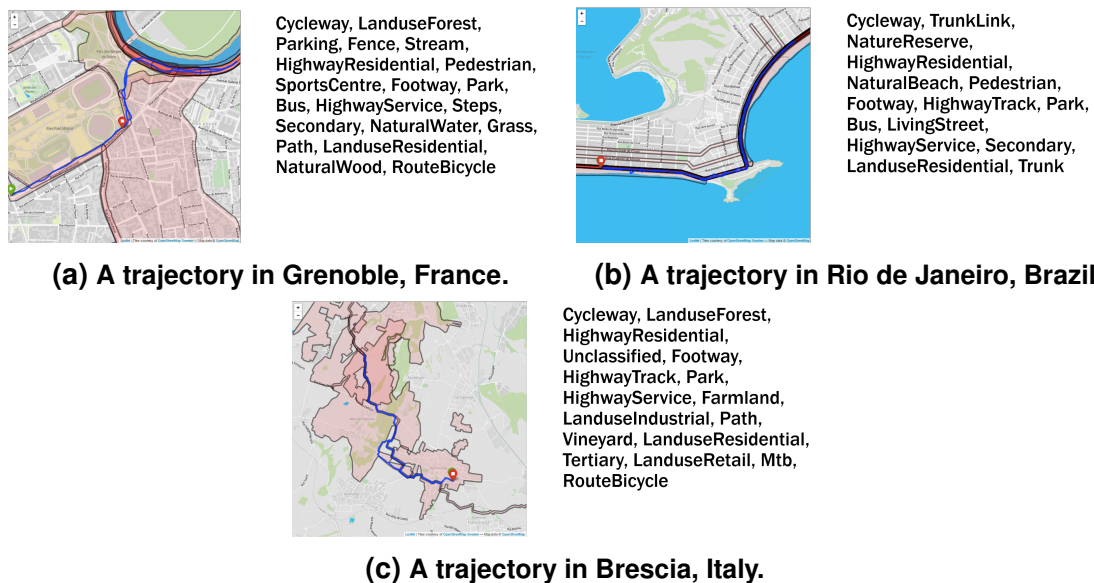


Figure 2. Some example trajectories along with the features selected by the ISA algorithm and the corresponding LinkedGeoData concepts.

6. Conclusions

We have demonstrated that our method allows not only the discovery of new tags taking into consideration an initial set of keys informed by the user, but it is also capable of creating a high-level description of the environment where the trajectory took place based on its spatial join with OSM polygons. This algorithm explores the relationships among OpenStreetMap tags issued from the Open Semantic Network. Such kind of annotation can be useful to user profiling as well as to highlight how citizens use the city to exercise.

Some aspects of ISA can be further improved as future work. First, it relies on the input of the tags keys of interest, excluding the tag values. Depending on the application,

the semantic information can be more valuable in the tag value instead of the tag key. This has not been done due to limitations in the Overpass API, which does not accept complex regular expressions.

Regarding the intersection among polygons and trajectory, the method has room for improvements if one considers the types of features by their semantic similarity. For instance, consider a group of three geographic features with the following respective tags: *landuse = forest*, *landuse = grass*, and *landuse = wood*, as ISA is concerned only with individual polygons, this group of semantically related features may not achieve the temporal and spatial thresholds set by the user. However, the group might still be relevant for describing the trajectory's spatial context.

Another aspect is the previous knowledge about the tag keys that should be informed by the user. Considering that this algorithm can be scaled to a large number of trajectories, it can be extended to infer which kind of tag keys and values are more frequent and recommend tags that seem to be more relevant. This is partially solved by the keys expansion mechanism aided by the filtering process that relies on the LGD ontology, but it can be further studied how upper ontologies such as DOLCE¹² or OpenCyc¹³ can help to identify more adequate concepts for a particular domain.

Acknowledgements

Rossana M. C. Andrade has a researcher scholarship “DT Level 2” sponsored by CNPq. This work is partially supported by INES (www.ines.org.br), CNPq grant 465614/2014-0 and FACEPE grants APQ-0399-1.03/17 and PRONEX APQ/0388-1.03/14. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Andrienko, Natalia, Gennady Andrienko, Nikos Pelekis, and Stefano Spaccapietra (2008). “Basic Concepts of Movement Data”. In: *Mobility, Data Mining and Privacy*. Ed. by Fosca Giannotti and Dino Pedreschi. Springer Berlin Heidelberg. Chap. 2, pp. 15–38. ISBN: 978-3-540-75176-2. DOI: [10.1007/978-3-540-75177-9](https://doi.org/10.1007/978-3-540-75177-9){_}2.
- Ballatore, Andrea, Michela Bertolotto, and David C. Wilson (2012). “Geographic Knowledge Extraction and Semantic Similarity in OpenStreetMap”. In: *Knowledge and Information Systems* 37.1, pp. 61–81. DOI: <http://dx.doi.org/10.1007/s10115-012-0571-0>.
- Bennett, Jonathan (2010). *OpenStreetMap*. Birmingham, UK: Packt Publishing, p. 252. ISBN: 9781847197504.
- Cocchia, Annalisa (2014). “Smart and Digital City: A Systematic Literature Review”. In: *Smart City: How to Create Public and Economic Value with High Technology in Urban Space*. Ed. by Renata Paola Dameri and Camille Rosenthal-Sabroux. Cham: Springer International Publishing, pp. 13–43. ISBN: 978-3-319-06160-3. DOI: [10.1007/978-3-319-06160-3_2](https://doi.org/10.1007/978-3-319-06160-3_2). URL: http://link.springer.com/10.1007/978-3-319-06160-3%7B%5C_%7D2.

¹²<http://www.loa.istc.cnr.it/old/DOLCE.html>

¹³<http://sw.opencyc.org>

- Dodge, Somayeh, Robert Weibel, and Anna-Katharina Lautenschütz (2008). “Towards a taxonomy of movement patterns”. In: *Information Visualization 7.3-4*, pp. 240–252. ISSN: 1473-8716. DOI: [10.1057/palgrave.ivs.9500182](https://doi.org/10.1057/palgrave.ivs.9500182). URL: <http://ivi.sagepub.com/lookup/doi/10.1057/palgrave.ivs.9500182>.
- Furletti, Barbara, Paolo Cintia, Chiara Renso, and Laura Spinsanti (2013). “Inferring human activities from GPS tracks”. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13*. New York, New York, USA: ACM Press, p. 1. ISBN: 9781450323314. DOI: [10.1145/2505821.2505830](https://doi.org/10.1145/2505821.2505830). URL: <http://dl.acm.org/citation.cfm?doid=2505821.2505830>.
- Nogueira, Tales P., Reinaldo B. Braga, Carina T. de Oliveira, and Hervé Martin (Feb. 2018). “FrameSTEP: A framework for annotating semantic trajectories based on episodes”. In: *Expert Systems with Applications 92*, pp. 533–545. ISSN: 09574174. DOI: [10.1016/j.eswa.2017.10.004](https://doi.org/10.1016/j.eswa.2017.10.004). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0957417417306796>.
- Ragone, Azzurra, Tommaso Di Noia, Vito Walter Anelli, Andrea Calì, and Matteo Palmonari (2016). “Exposing Open Street Map in the Linked Data cloud”. In: *Proceedings of the 29th International Conference on Industrial, Engineering & other Applications of Applied Intelligent Systems*. URL: <http://www-ictserv.poliba.it/sisinflab/publication%20s/2016/RDACP16>.
- Spaccapietra, Stefano, Christine Parent, and Laura Spinsanti (2013). “Trajectories and Their Representations”. In: *Mobility Data: Modeling, Management, and Understanding*. Ed. by Chiara Renso, Stefano Spaccapietra, and Esteban Zimányi. Cambridge University Press, pp. 3–22. ISBN: 9781139128926.
- Stadler, Claus, Jens Lehmann, Konrad Höffner, and Sören Auer (2012). “LinkedGeoData: A Core for a Web of Spatial Open Data”. In: *Semantic Web 3.4*, pp. 333–354. DOI: [10.3233/SW-2011-0052](https://doi.org/10.3233/SW-2011-0052). URL: <http://iospress.metapress.com/content/141W054666871326>.
- Yan, Zhixian, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer (2011). “SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories”. In: *Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT '11*. New York, New York, USA: ACM Press, p. 259. ISBN: 9781450305280. DOI: [10.1145/1951365.1951398](https://doi.org/10.1145/1951365.1951398). URL: <http://portal.acm.org/citation.cfm?doid=1951365.1951398>.
- (2012). “Semantic Trajectories: Mobility Data Computation and Annotation”. In: *ACM Transactions on Intelligent Systems and Technology (TIST) 4.2*, 39:1–39:38. DOI: [10.1145/2483669.2483682](https://doi.org/10.1145/2483669.2483682).
- Zheng, Yu (2015). “Trajectory Data Mining: An Overview”. In: *ACM Trans. On Intelligent Systems and Technology 6.3*. DOI: [10.1145/2743025](https://doi.org/10.1145/2743025).