Enriching Mobility Data with Linked Open Data

Livia Ruback
Departamento de Informatica
PUC-Rio
Rio de Janeiro, Brazil
lrodrigues@inf.puc-rio.br

Marco Antonio Casanova
Departamento de Informatica
PUC-Rio
Rio de Janeiro, Brazil
casanova@inf.puc-rio.br

Alessandra Raffaetà
Dip. di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari, Venezia, Italy
raffaeta@unive.it

Chiara Renso
HPCLab
ISTI-CNR
Pisa, Italy
chiara.renso@isti.cnr.it

Vania Vidal
Departamento de Informatica
Uiniv. Federal Ceará,
Fortaleza, Brazil
vvidal@lia.ufc.br

ABSTRACT

Recent research has pointed out the needs and advantages of the semantic enrichment of movement data, a process where trajectories are partitioned into homogeneous segments that are annotated with contextual information. However, the lack of a comprehensive and well-defined framework for the enrichment makes this process difficult and error-prone. In this paper, we therefore propose a conceptual framework for the semantic enrichment of movement data, which benefits from the emerging Web of Data (or Linked Open Data) both as a unifying formalism and as the source of contextual data, which can be greatly useful for trajectories enrichment. Moreover, the semantic structure of such sources makes it easier to share and process enriched trajectories. We illustrate the enrichment process by presenting a case study in the tourism domain.

The final authenticated version is available online at http://dx.doi.org/10.1145/2938503.2938550

Categories and Subject Descriptors

H.2.1 [**Database Management**]: Logical Design—*Schema and subschema*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Semantic Web*

Keywords

Semantic trajectory, data interlinking, Linked Open Data

1. INTRODUCTION

The availability of wearable GPS-enabled devices and geolocated social media allows the collection of personal footprints, capturing a huge amount of human traces that are becoming the base of big data analysis in several application fields, from traffic management, to urban development and tourism recommendation, to name a few [17].

Most of the analysis methods investigated so far focused on the pure spatio-temporal facets of these trajectories. However, we have recently witnessed a growing research area where the representation of movement is enriched with external contextual data, thus resulting in the so-called *semantic trajectories* [20, 16, 6, 12].

Not only the amount of mobility data available is increasing fast but, thanks to the Linked Data initiative, another unprecedented global space is also growing fast: the Web of Data. Specifically, the Linked Data principles [13, 5] promote the creation and publication of previously isolated databases as interlinked, reusable data graphs using known Web standards.

In this paper we propose a conceptual framework with the objective of guiding the whole trajectory enrichment process to generate semantic trajectories. The focus of this framework is to take advantage of the Linked Data principles in two aspects. First of all, representing trajectories according to the Linked Data principles offers a strategy to incorporate trajectories into this global data space in a way they can be easily shared and reused, which is the main motivation of the Linked Data initiative. As a second important aspect, the Web of Data can be used as the main source of contextual information to enrich movement data. The proposed framework has been designed to create a publishable movement data repository by incrementally building a *semantic trajectory ontology*. This approach makes the process application-oriented, simple, reusable, flexible and general enough to cover many different domains.

The paper is organized as follows. Section 2 reports some basic concepts and introduces a running example. Section 3 presents the main contribution of the paper describing the process and its two main steps: the segmentation and the enrichment. Section 4 gives examples of queries and illustrates some potential analyses on semantic trajectories. Section 5 discusses related work. Finally, Section 6 contains the conclusions and suggests future work.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

2. PRELIMINARIES AND RUNNING EXAMPLE

Moving objects are entities having a time-varying position, uniquely determined at each time instant. A *trajectory* is a continuous part of the movement of that object. Commonly, in mobility applications, the continuous movement of an object is not completely known since it is often given by means of a finite set of timestamped positions collected by a mobile device, called trajectory *points* or *samples*. We call *raw trajectory* a sequence of samples as collected by the mobile device. A *segmented trajectory* is a partition of these points into homogeneous segments where a given property holds. For example, according to the stop-and-move model [20], a raw trajectory can be split into segments of two kinds: *stop* where the speed of the object is lower than a certain threshold and *move* where the speed is greater than such a threshold. Other segmentation criteria are sometimes used, such as the change of direction [19].

The notion of *semantic trajectory* goes further and enriches a segmented trajectory with contextual information retrieved from an external data source: the points of interest visited, the means of transportation employed, or the goal of the movement [6, 16].

The Linked Data Principles [13] recommend the use of ontologies and RDF (Resource Description Framework) to publish databases on the Web, thereby minimizing the problem of schema alignment, a difficult and error-prone task. Intuitively, following the Linked Data Principles facilitates the task of linking trajectories with external data sources. Due to the huge and heterogeneous amount of Linked Open Data available, it is critical to be able to properly select and integrate the relevant entities. Applications that combine, aggregate and transform data available on the Web of Data are known as *Linked Data Mashups* (LDM) applications. Vidal et al. in [21] introduced an approach where the multiple Linked Data sources are provided by the mashup through homogenized views, the so-called *Linked Data Mashup views*.

We clarify the whole enrichment process by introducing a running example from the tourism domain. We assume the use of position-enabled devices that track and collect the movement of tourists visiting Florence in Italy with the objective to offer personalized services. Let us consider the typical trajectory of a tourist that starts at her hotel, combines sightseeing and lunch during the day, before going to the train station to depart. Figure 1 shows the first part of the trajectory until the bridge "Ponte Vecchio". The trajectory is represented as a sequence of samples, i.e., timestamped coordinates *(x,y,t)* as collected by the GPS device. The objective of the analysis is to have a better understanding of the tourists' behaviour, such as characterizing them based on the features of the attractions visited, transportation means used to identify the visiting profiles (e.g. the tourist spending profile). It is clear that, to reach this objective, we need to build a new kind of trajectory data by augmenting the pure location data with a large amount of contextual, semantically rich data.

We describe in the next section the framework to build these semantically rich trajectory data by identifying the parts to be enriched and then exploiting Linked Open Data to give the actual meaning to the trajectory parts.

3. THE CONCEPTUAL FRAMEWORK

3.1 Overview of the Conceptual Framework

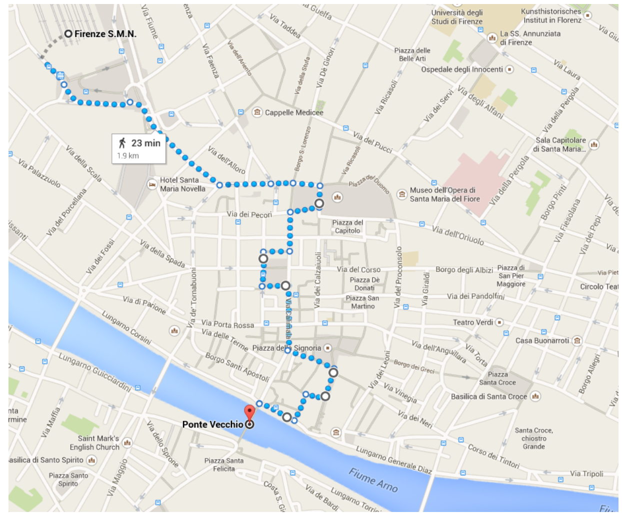

Figure 1: A raw trajectory of a tourist in Florence collected as GPS samples.

In this section we first give a general overview of the proposed process, that covers the trajectory segmentation and the Linked Open Data enrichment up to the construction of a semantic trajectory repository to be used for the analysis, as illustrated in Figure 2.

The semantic enrichment process takes as input a raw trajectory and a number of Linked Open Data sources and builds a semantic trajectory repository. This process is driven by the use of ontologies and it is structured into two main steps: segmentation and enrichment. The *segmentation step* splits a raw trajectory into homogeneous segments, specifying the entities that will be enriched. This step is driven by a *Segmented Trajectory Ontology (STO)*, which identifies the different types of segmentation required by a specific application. The *enrichment step* matches the segments with the most appropriate semantic entities made available as a *linked data mashup view*, which provides cleaned and integrated data from selected linked data sources. In this step, the *mashup view ontology* specifies the concepts of the mashup view (i.e., the *conceptual model*), which is the common vocabulary for integrating data exported by the selected Linked Data sources.

The outcome of the enrichment step is the *semantic trajectory ontology (SemTO)* and an RDF repository containing semantically enriched trajectories. The semantic trajectory ontology contains all concepts and properties of the segmented trajectory and mashup ontologies, and also the definition of the semantic links between them. Research in semantic data mining [9] attested that the domain knowledge formally encoded in an ontology is very helpful in all stages of the analysis process. In the proposed framework each step is driven by the incremental use of ontologies that can be easily adapted to meet the application needs. We recall that the Linked Data principles provide the unifying formalism where the entities to be enriched (trajectories) and the enriching contextual information (Linked Open data sources) are homogeneously modeled.

3.2 The Segmentation Step

The first step of our process is called *segmentation*: it

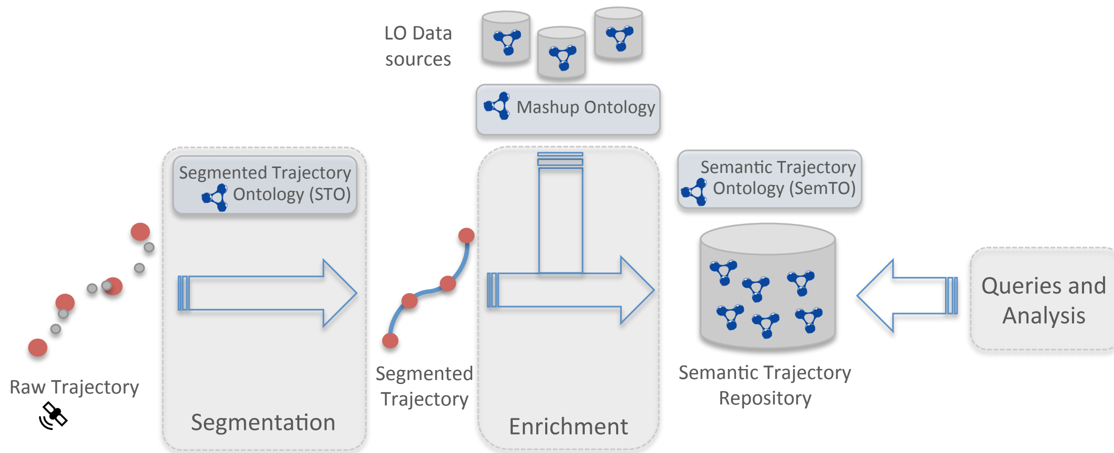

Figure 2: The trajectory enrichment process

takes as input a raw trajectory and some particular segmentation criterion and splits the trajectory into segments, which are the entities to be enriched. This step is driven by the Segmented Trajectory Ontology (STO). This ontology aims at representing the trajectory entities featuring the parts to be enriched. There are many works in the literature [24, 18, 14] proposing trajectory ontologies that can be easily adapted to the application needs by adding specializations of classes and properties.

Figure 3 shows the segmented trajectory ontology related to our running example. It has been inspired by [14]: by choosing a particular splitting criterion, a raw trajectory is transformed into a segmented trajectory which is composed of segments, each satisfying the criterion. Each segment is associated with two spatio-temporal points, the begin and the end of the segment, and possibly with a textual label that describes a property of the segment.

We notice that, here, the class *Segment* is specialized in *Stop* and *Move* to represent the two specific kinds of segments to be enriched. It is worth observing that a stop is characterized also by a spatial location which could be the centroid of the segment. Naturally the specific ontology can be tailored to the application needs and other specializations are possible, like the transport mode segmentation [27] or the activity segmentation [26]. We remark how this incremental ontology building is aimed at representing the entities to be enriched with the Linked Open Data.

Going back to our running example, we segmented the tourist raw trajectory following the "stop-and-move" model previously mentioned. The resulting segmentation is shown in Figure 4 where the tourist trajectory has been segmented into a "begin", seven "stops", eight "moves" and an "end".

During the move the tourist uses different transportation means: first she walks in the city center (from the begin to stop 6), then she uses a vehicle from Ponte Vecchio to Piazzale Michelangelo (e.g. a taxi) and another vehicle to go back to the railway station (e.g. a bus). The means of transportation (Walking or Vehicle) is the label associated with the move, whereas the arrival time to the stop (e.g. 10:30 am) and its duration (e.g. 30') are the labels related to the stop.

Being able to distinguish segments into subclasses such as stops and moves allows us to differentiate the enrichment of these two kinds of segments, as illustrated in the next section.

3.3 The Enrichment Step

The second step of the process shown in Figure 2 is called *enrichment* and it matches (or enriches) trajectory segments (e.g. stops or moves) to the most appropriate semantic entities made available by the external contextual information through the Linked Data Mashup [10]. A linked data mashup tuned to the application domain is therefore fundamental since it simplifies the step of enriching trajectories bringing all necessary data tailored to the actual needs [23].

The enrichment step involves two main tasks:

1. Creation or selection (if it already exists) of the Linked Data Mashup (LDM), which integrates data from the relevant selected linked data sources.

2. Linking trajectory segments with Linked Data Mashup entities.

3.3.1 Creation or Selection of the Linked Data Mashup

The creation of the Linked Data Mashup is a complex and time-consuming task. We simply reuse a mashup built according to the previous mentioned views [21, 7].

Aimed at enriching both the stops and the moves, in Figures 5 and 6 we present two mashup fragments in the tourism domain. The entities available by the mashup brings semantic to the places the tourists can visit during a trip, and the transportation means used to move between these places.

The mashup aggregates entities from two different data sources:

1. the DBPedia dataset [3], which constitutes a major part of the semantic data on the Web; and

2. the OpeNER Linked Dataset (OLD), part of the OpeNER project [4], which consists of a repository for the tourism domain that covers the Tuscany region.

The mashup reuses terms from widely used vocabularies:

1. DBPedia vocabulary[1];

[1]http://wiki.dbpedia.org

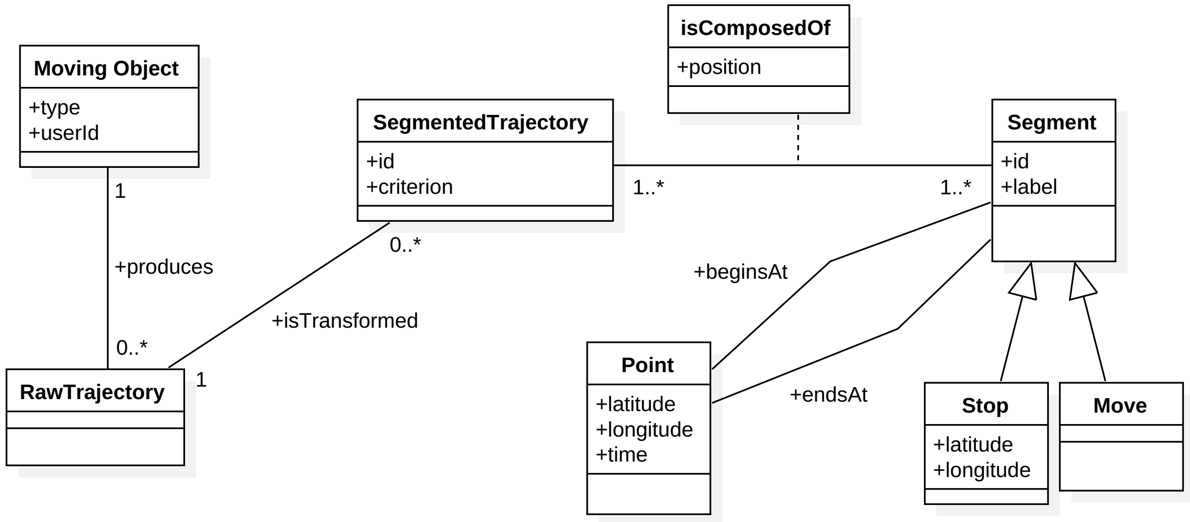

Figure 3: The Segmented Trajectory Ontology

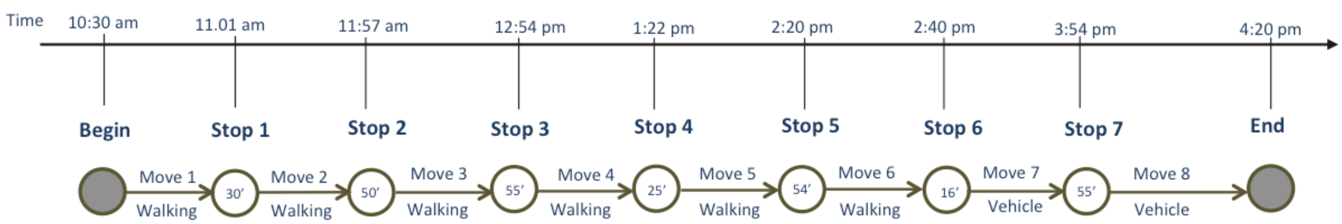

Figure 4: The running example segmented trajectory

2. FOAF (Friend of a Friend)[2];
3. RDFs (RDF schema)[3] and
4. vCard vocabulary[4].

For the accommodation domain, the mashup reuses both HOntology [8] and Accommodation Ontology[5].

The mashup view fragment, shown in Figure 5, contains the places features. The entities that represent museums, religious buildings, artists and art works come from DBPedia and are represented respectively by the classes `dbo:Museum`, `dbo:Religious Building` (both are also points of interest), `dbo:Artist` and `dbo:Artwork`. Some attributes of these entities are provided by the mashup itself, for example, the list of categories which a museum is related to (e.g.: Modern art museum and History museum). This part of the view can intuitively model, for instance, an art (category) museum that exhibits paintings (art work) of Botticelli (artist), belonging to the High Renaissance period (movement). Also a church (religious building) can be related to the Gothic period (architecture style) and it can also have works of some artists.

[2]http://xmlns.com/foaf/spec/
[3]https://www.w3.org/TR/rdf-schema/
[4]https://www.w3.org/TR/vcard-rdf/
[5]http://ontologies.sti-innsbruck.at/acco/ns.html

In turn, the entities that represent points of interests, accommodations and restaurants come from the OpeNER Linked Dataset (OLD) and are represented respectively by the classes `hont:PointsOfInterest`, `acco:Accommodation`, `dbo:Restaurant`.

The mashup view fragment, shown in Figure 6, contains the transportation features. The vocabulary extended by the mashup is the GTFS (General Transit Feed Specification)[6], a common format for public transportation schedules and associated geographic information, used by Google Maps, which also provides open transportation data.

The instances of `gtfs:Agency` represent the companies that provide the routes. The instances of `gtfs:Route` represent the entire journey made by an agency. The property `gtfs:route_type` holds the type of vehicle used on the route (tram, subway, rail, bus, ferry, cable car, or funicular). The class `gtfs:Trip` is a part of a route related to the direction (ex.: from the airport to the train station or vice-versa). Each trip is composed by `gtfs:Stop` instances, that have their respective latitude and longitude.

The white boxes are part of the GTFS data model and comprise the scheduled transport part of the mashup. Besides the scheduled transportation, travelers can also move by using other types of transportation. Classes `tmo:ByTaxi`,

[6]https://developers.google.com/transit/gtfs/reference

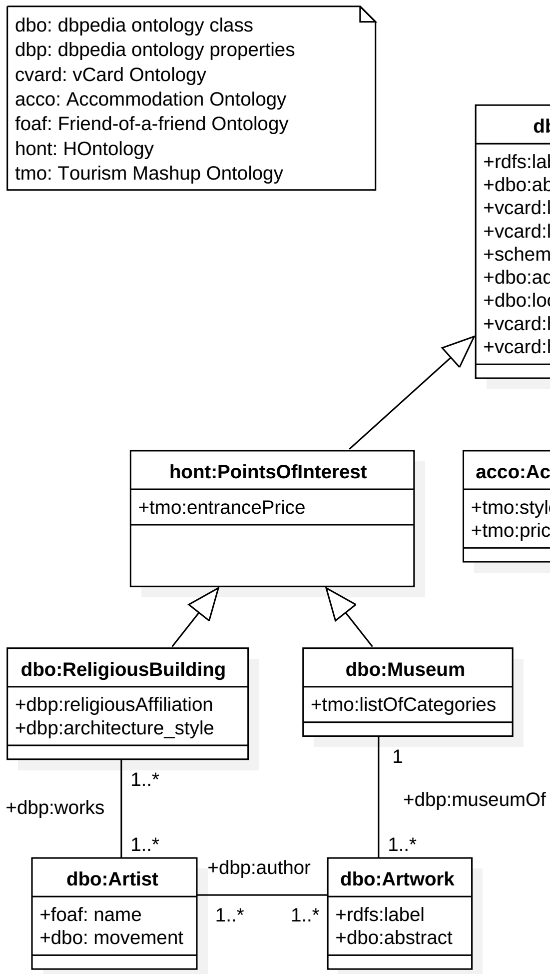

Figure 5: The tourism mashup view fragment representing places.

`tmo:ByCar` and `tmo:ByBike` respectively represent information about taxis companies, car and bicycle rentals/sharing, somehow available as Linked Open Data. In the case of bicycles, for example, one can use data available by a bike sharing system like Bicincittà , widely used in Italy.

These two mashup views (Figure 5 and Figure 6) are the contextual sources used for the enrichment of the trajectory segments, presented in what follows.

3.3.2 Linking Trajectory Segments and Linked Data Mashups

The linking between trajectory segments and Linked Data mashups is specified as a view [21] and it is actually performed as a matching between entities to be enriched (segments expressed by the Segmented Trajectory Ontology) and entities that enrich (classes of the Mashup Ontology). A typical example is the matching of stops with Points Of Interest (POIs) where the match predicate might be based on the distance between the stop and the POI [17, 25]. Considering our example, we use the places mashup fragment to match the stops (Figure 5) and the transportation mashup fragment to enrich the moves (Figure 6).

The execution of this matching produces one or more links between segments of a trajectory (stops and moves in our case) and mashup entities. In fact, some LOD tools help automate the matching process, such as Silk [22] and Limes [15].

We now discuss in detail the stop enrichment process for our example. First, as shown in Figure 7, we introduce the property `foaf:based_near`, which is part of FOAF and relates two "spatial things" being close to each other. This property is part of the Semantic Trajectory Ontology (Sem TO) that reuses it from FOAF, following the Linked Data principles. Linking (i.e., enriching) instances of `sto:Stop` with instances of `dbo:Place` using this property we are stating that the stop is near a place.

The following triples (described in turtle notation) represent two instances for stop and place entities: (a) Stop 1 is an instance of Stop of the Segmented Trajectory Ontology (Figure 3); and (b) the church *Basilica di Santa Maria Novella* is a mashup instance of the `dbo:ReligiousBuilding` class (Figure 5).

```
1. sto-resource:stop1
2.     rdf:type sto:Stop;
3.     vcard:latitude  "43.774836";
4.     vcard:longitude "11.249375";
```

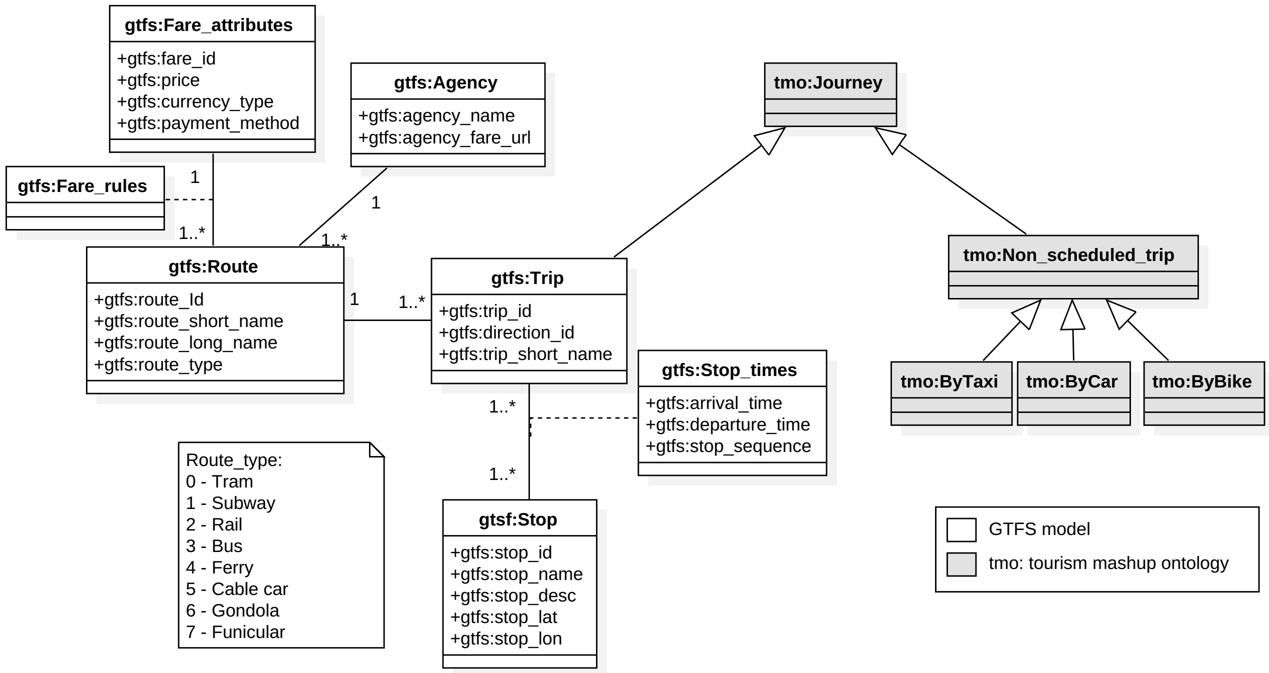

Figure 6: The Tourism Mashup view fragment about transportation

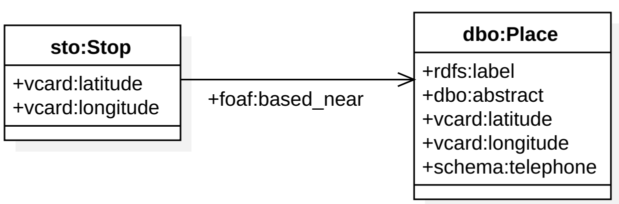

Figure 7: The Stop Enrichment

```
5.     foaf:based_near
            dbr:Basilica_of_Santa_Maria_Novella.
6. dbr:Basilica_of_Santa_Maria_Novella
7.     rdf:type dbo:ReligiousBuilding;
8.     vcard:latitude  "43.774601";
9.     vcard:longitude "11.249300";
10.    rdfs:label "Basilica of Santa Maria Novella";
11.    dbo:abstract "Santa Maria Novella is a
       church in Florence, Italy, situated just
       across from the main railway station which
       shares its name".
```

Lines 1 to 4 describe the `sto-resource:stop1`, Stop 1. Likewise, lines 5 to 9 correspond to the church *Basilica di Santa Maria Novella* (other triples relating to the church were omitted). We note that `sto-resource` is the prefix of the Semantic Trajectory Repository resources (entities) and the `dbr` is the prefix of the `http://dbpedia.org/resource/` namespace for the DBPedia entities.

Continuing our example, the result of the matching process will be a set of triples of the form `(s, foaf:based_near, o)`, where `s` is the subject denoting a stop of the trajectory, `foaf:based_near` is the linking predicate introduced in Figure 7 and `o` is the object denoting a POI. Similarly, we have the mapping of the other trajectory stops:

```
1. sto-resource:stop2 foaf:based_near
            dbr:Piazza_del_Duomo,_Florence
2. sto-resource:stop3 foaf:based_near
            dbr:OsteriaDellOlio
3. sto-resource:stop4 foaf:based_near
            dbr:Piazza_della_Repubblica,_Florence
4. sto-resource:stop5 foaf:based_near
            dbr:Palazzo_Vecchio
5. sto-resource:stop6 foaf:based_near
            dbr:Ponte_Vecchio
6.  sto-resource:stop7 foaf:based_near
             dbr:Piazzale_Michelangelo
```

Besides the stops, we can also enrich the move segments with the transportation features taken from the mashup transportation fragment as illustrated in Figure 8.

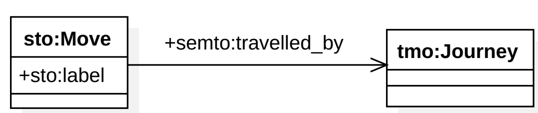

Figure 8: The Move Enrichment

The property `semto:travelled_by` is part of the *semantic trajectory ontology (SemTO)* and links a move in the Segmented Trajectory Ontology to `tmo:Journey` (or to one of its subclasses `gtfs:Trip`, `tmo:ByCar`, `tmo:ByTaxi` or `tmo:ByBike`). The moving parts can be traveled by *scheduled journeys* (represented by the GTFS classes), which are trips made by tram, subway, rail, bus, ferry, cable car or funicular, or *not scheduled journeys*, such as cars, taxis and bikes (see Figure 6).

The following triples represent the output of the move enrichment:

1. `Move8` is an instance of a move of the Segmented Trajectory Ontology (see Figure 3);
2. `trip13/1` is an instance of the `gtfs:Trip` class; and
3. `route13` is an instance of the `gtfs:Route` class.

```
1. sto-resource:move8
2.     rdf:type sto:Move;
3.     sto:label "Vehicle";
4.     semto:travelled_by gtfs-resource:trip13/1.

5. gtfs-resource:trip13/1
6.     rdf:type gtfs:Trip;
7.     gtfs:direction_id "1";
8.     gtfs:trip_short_name
9.        "ATAF Linea 13/Direzione Stazione FS SMN";
10.    gtfs:route gtfs-resource:route13.

11. gtfs-resource:route13
12.    rdf:type gtfs:Route;
13.    gtfs:route_short_name "ATAF Linea 13";
14.    gtfs:route_long_name
15.        "ATAF Linea 13 /
            Stazione Palazzo dei Congressi -
            Cartoleria Il Gatto e La Volpe <=>
                     Piazzale Michelangiolo";
16.    gtfs:route_type "3".
```

The triples above provide the semantics to `move8` (Lines 1-4), meaning that this segment of the trajectory was traveled with the line `ATAF Linea 13 / Direzione Stazione FS SMN` (lines 5-10) having as direction '1' (one way), that is part of the route with short name `ATAF Linea 13` (lines 11-15) made by bus `(route_type 3)` (line 16).

Figure 9 illustrates the output of the stop and the move mapping phases for the running example, i.e., the trajectory of the tourist shown in Figure 4, having all its segments - stops and moves - enriched with the external data provided by the Linked Data mashups.

4. QUERYING AND ANALYZING SEMANTIC TRAJECTORIES

The output of the enrichment step is stored in the semantic trajectory repository where movement data and the associated semantics are represented in a uniform formalism. We notice that this repository is represented by the SEMantic Trajectory Ontology (SemTO), which is the union of the Segmented Trajectory Ontology (STO), the Mashup Ontology (MO) and the additional properties introduced during the enrichment step (e.g. `semto:based_near` and `semto:travelled_by` in our example).

The semantic trajectory repository enables a number of interesting analyses as we illustrate in the following SPARQL queries. Here we present explicative examples querying enriched trajectories:

1. a single trajectory (Q1);
2. a set of trajectories from the same traveler (Q2) and
3. all trajectories of the repository (Q3) and (Q4).

We remark that the following queries return entities made available by the Semantic Trajectory Repository, combining the movement data and the semantic data through the mashup view, as described in Section 3.

The first query characterizes the cultural tastes of the tourist during her path through the city center by retrieving the artists and artworks information of the museums the traveler visited. We can notice the use of the matching property `foaf:based_near`, described in Section 3.3.2, and the use of property `tmo:listOfCategories` from the mashup view.

Q1) Which are the categories of the museums visited during the trip with id "tripFlorenceJune16" and the art movements related to them?

```
SELECT
    ?museumName ?museumCategories ?artmovement

WHERE{
    ?segmentedTrajectory
        sto:id "tripFlorenceJune16";
        sto:isComposedOf ?stop.
    ?stop foaf:based_near ?museum.
    ?museum rdf:type dbo:Museum;
        rdfs:label ?museumName;
        tmo:listOfCategories ?museumCategories;
        dbp:museumOf ?artwork.
    ?artwork dbo:author ?artist.
    ?artist dbo:movement ?artmovement
}
```

The above query returns the categories of all the museums that were visited in trip "tripFlorenceJune16". For our running example, the query result is *Palazzo Vecchio* as museum, *Art museums and galleries* as category and *High Renaissance* as art movement. In fact, that museum contains the marble sculpture *The Genious Of Victory* (the artwork), that was made by *Michelangelo* (the artist), that in turn is associated with *High Renaissance* (art movement).

Q2) Which was the average price spent in museums and churches by the traveler with id 257?

```
SELECT
    (AVG(?entrancePrice) AS ?avgEntrancePrice)
WHERE{
    ?movingObject sto:userId "257";
                sto:produces ?rawTrajectory.
    ?rawTrajectory sto:isTransformed
                         ?segmentedTrajectory.
    ?segmentedTrajectory sto:isComposedOf ?stop.
    ?stop foaf:based_near ?attraction.
    ?attraction tmo:entrancePrice ?entrancePrice.
    {?attraction rdf:type dbo:Museum}
    UNION {
      ?attraction rdf:type dbo:ReligiousBuilding;
        dbp:architecture_style dbr:Church_(building)
    }.
}
```

This query adds up all money spent in museums and churches during all trips made by the traveler with id '257'.

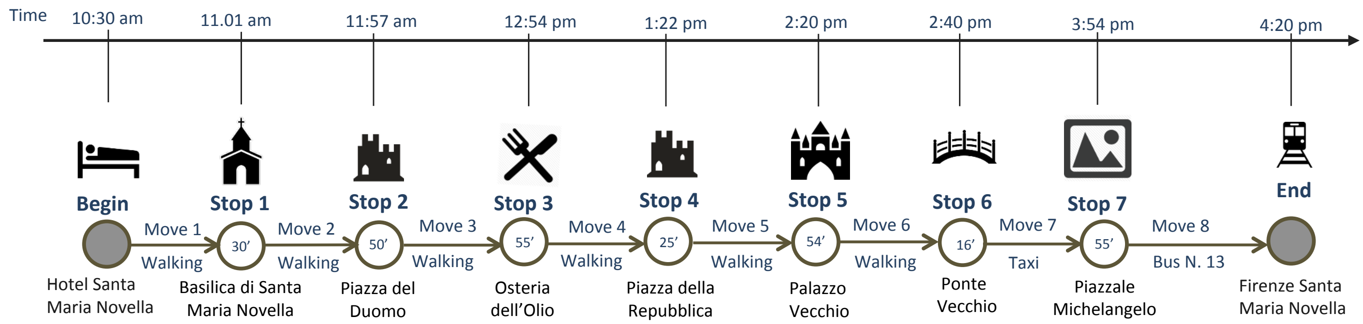

Figure 9: The running example trajectory Enrichment

The UNION clause combines graph patterns thus allowing that alternative possibilities match the same variable `?entrancePrice` refers to money spent in attractions that can be a `dbo:Museum` or can also be a `dbo:ReligiousBuilding` of the type church.

The next query spans on all the tourists' trajectories to retrieve those that used transportation from the local bus company named ATAF. Notice the use of the property `semto:travelled_by` (described in Section 3.3.2) of the semantic trajectory ontology that links the move to the actual transportation means used in that trajectory segment.

Q3) Which tourists used buses provided by the ATAF company?

```
SELECT
    DISTINCT ?travellerId
WHERE{
    ?movingObject sto:userId ?travellerId;
                  sto:produces ?rawTrajectory.
    ?rawTrajectory sto:isTransformed
            ?segmentedTrajectory.
    ?segmentedTrajectory sto:isComposedOf ?move.
    ?move semto:travelled_by ?trip.
    ?trip rdf:type gtfs:Trip;
          gtfs:route ?route.
    ?route rdf:type gtfs:Route;
           gtfs:agency ?agency;
           gtfs:route_type "3".
    FILTER(?agency gtfs:agency_name "ATAF")
}
```

This query returns the identifiers of all tourists that traveled using buses of the ATAF Company, the local bus company of Florence.

Q4: Which travelers are interested in High Renaissance?

```
SELECT
    ?userId
WHERE{
    ?movingObject sto:userId ?userId;
                  sto:produces ?rawTrajectory.
    ?rawTrajectory sto:isTransformed
            ?segmentedTrajectory.
    ?segmentedTrajectory sto:isComposedOf ?stop.
    ?stop foaf:based_near ?museum.
    ?museum rdf:type dbo:Museum;
            dbp:museumOf ?artwork.
    ?artwork rdf:type dbo:Artwork;
             dbp:author ?author.
    ?author rdf:type dbo:Artist;
                    dbo:movement
    <http://dbpedia.org/resource/High_Renaissance>.
}
```

The above query filters out only travelers that may have interests in the period of the Italian Renaissance art production called High Renaissance.

Other examples of queries that can be easily answered by the semantic trajectory repository involve restaurant reviews and temporal filtering such as:

1. *How many travelers stayed in 5 stars hotels and ate at "very good" restaurants?*
2. *How many travelers went to catholic churches that contains Baroque artwork in Florence in the last 2 months?*

The SPARQL queries provide many answers to the requirements such as characterizing the visited venues or the cultural level of tourists. However, some more sophisticated questions can only be answered after analyzing data contained in the repository.

Going back to our running example, consider the discovery of the spending profile of the tourists. We may wish to distinguish between high spending and low spending tourist profiles. They define a visiting behavior based, for example, on the entry price of the visited venues, combined with the identification of prestigious accommodations and restaurants, properly joined with additional information such as the average life cost in the city.

The most natural way to perform these kinds of analyses is to exploit the structured format provided by the RDF model to perform inferences using OWL reasoning capabilities [1]. In this case we can combine the semantic trajectory ontology with a larger application ontology containing concepts related to the specific application representing the users' behaviors we want to infer from data in the style of [18].

Another challenging analysis is to identify groups of semantically similar trajectories. In the mobility field, similarity usually relies on the spatio-temporal characteristics of the raw trajectories, while here we can take advantage of the rich semantics coming from Linked Open Data that can be conveniently combined with the spatio-temporal component. This offers a new opportunity to define innovative similarity measures, such as finding groups of tourists with similar cultural preferences, lifestyles and spending profiles, all useful information to design novel sophisticated recommendation systems.

5. RELATED WORK

The formalization of the trajectory semantic enrichment process has been firstly outlined in [11], where Baquara has been proposed as a general all-inclusive ontology representing both the trajectory and the enriching concepts. This approach has been pioneering and traced the way to the use of ontology to support the enrichment process with Linked Open Data. However, it has some limitations. For example, the proposed Baquara ontology, in trying to be general enough to cover a large range of applications, has become a "monolithic" approach, complex and difficult to personalize to different needs. Furthermore, the Linked Open Data sources that enrich the trajectory are predefined. Our approach tries to complement this vision offering a process to build a semantic trajectory ontology "step-by-step" driven by the application domain. This is done by using incremental ontologies that can be properly extended based on the application needs and a mashup ontology built from selected data sources tailored to the application domain.

One of the first approaches trying to conceptualize movement data as a trajectory ontology was proposed by Yan et al. in [24]. The conceptual framework was aimed at combining in a unique top-level ontology the different aspects of the movement embedded into three main ontologies representing application-related information, the trajectory as a sequence of stops and moves and the geography information. However, it offers a conceptual and top-level vision of trajectories without explicitly dealing with the problem of the enrichment process or adopting the Linked Open Data formalism. Another approach trying to represent semantic trajectories based on ontologies was proposed by [2] and was built on Ontology Engineering techniques to connect Generic Places Ontologies with POI instances. Different from our approach, they focus on the enrichment of POIs using the proper Ontologies terms, while our approach faces all steps involved in the trajectory enrichment process and analysis.

A few years later, [18] made a step towards employing the Athena ontology (structured into application ontology giving the application domain analysis concepts and core ontology representing the segmented trajectories) into a reasoning process based on OWL to support meaningful pattern interpretations of human behavior combining inductive reasoning and deductive reasoning. However, this paper still does not deal with the modeling of the enrichment step and it can be seen as complementary to our proposal since it is focused on the analysis part instead of the enrichment part.

The approach proposed by Yingjie et al. in [14] introduces a geo-ontology design pattern for semantic trajectories that is very similar to our Segmented Trajectory Ontology. A formal encoding of the classes together with their properties is given by using OWL. Authors also define a number of interfaces to integrate related geographic information, domain knowledge and device data. We go a step beyond with respect to the approach proposed by Yingjie et al. in [14] since we face also the issues of how implementing the enrichment step by using Linked Data Mashups.

6. CONCLUSIONS

We introduced a conceptual framework for the semantic enrichment of movement data based on Linked Open Data. Our proposal offers a flexible, reusable, application-oriented process based on ontologies that support the transformation of movement data into a Linked Open Data semantically enriched trajectory repository. We highlighted the different steps and how the availability of such repository improves the ability to formulate application analysis questions, thanks to the richness of the linked contextual data. We discussed the process with the help of a running example in the tourism domain.

It is important to stress that this process is meant to be general and support the semantic enrichment of several kinds of movement data in different domains, thus including not only GPS data but also social networks geo-located photos from Flickr, Twitter posts, FourSquare check-ins, and others. For this reason, as future work, we are instantiating this process to a real dataset containing tourist trajectories by developing a prototype implementation having Flickr photos together with its metadata, taken in Florence and in London. We are also investigating different techniques to compute the similarity between enriched trajectories using its POIs and segments features, such as the categories of the places related to the POIs and the means of transportation used to move between them, with the main objective to improve the accuracy of recommendations systems of the tourism domain.

7. ACKNOWLEDGMENTS

This work was partially supported by the Italian Ministry of University and Research (MIUR) Projects PON ADAPT (no. SCN00447) and by CNPq under grants 557128/2009-9, 442338-2014-7, 303332/2013-1, 201690/2015-2, and FAPERJ under grants E-26-170028/2008 and E-26/201.337/2014.

8. REFERENCES

[1] W3C OWL Working group, 2012. OWL 2 web ontology language. document overview (second edition). W3C recommendation 11 december 2012. available at https://www.w3.org/tr/owl2-overview/.

[2] Ana Alves, Bruno Antunes, Francisco C Pereira, and Carlos Bento. Semantic enrichment of places: Ontology learning from web. *International Journal of Knowledge-based and Intelligent Engineering Systems*, 13(1):19–30, 2009.

[3] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. *Dbpedia: A nucleus for a web of open data*. Springer, 2007.

[4] Clara Bacciu, Angelica Lo Duca, Andrea Marchetti, and Maurizio Tesconi. Accommodations in Tuscany as Linked Data. In *LREC*, pages 3542–3545, 2014.

[5] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.

[6] Vania Bogorny, Chiara Renso, Artur Ribeiro Aquino, Fernando Lucca Siqueira, and Luis Otavio Alvares. Constant–a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18(1):66–88, 2014.

[7] Marco A Casanova, Vânia MP Vidal, Giseli Rabello Lopes, Luiz André P Paes Leme, and Livia Ruback. On materialized sameas linksets. In *Database and Expert Systems Applications*, pages 377–384. Springer, 2014.

[8] Marcirio Silveira Chaves, Larissa A. Freitas, and Renata Vieira. Hontology: a multilingual ontology for the accommodation sector in the tourism industry. In *Proceedings of the 4th International Conference on Knowledge Engineering and Ontology Development*, 2012.

[9] Dejing Dou, Hao Wang, and Haishan Liu. Semantic data mining: A survey of ontology-based approaches. In Mohan S. Kankanhalli, Tao Li, and Wei Wang, editors, *ICSC*, pages 244–251. IEEE Computer Society, 2015.

[10] Brigitte Endres-Niggemeyer. *Semantic Mashups: Intelligent Reuse of Web Resources*. Springer Publishing Company, Incorporated, 2013.

[11] Renato Fileto, Marcelo Krüger, Nikos Pelekis, Yannis Theodoridis, and Chiara Renso. Baquara: A holistic ontological framework for movement analysis using linked data. In *Conceptual Modeling*, pages 342–355. Springer, 2013.

[12] Renato Fileto, Alessandra Raffaetà, Alessandro Roncato, Juarez AP Sacenti, Cleto May, and Douglas Klein. A semantic model for movement data warehouses. In *Proceedings of the 17th International Workshop on Data Warehousing and OLAP*, pages 47–56. ACM, 2014.

[13] Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.

[14] Yingjie Hu, Krzysztof Janowicz, David Carral, Simon Scheider, Werner Kuhn, Gary Berg-Cross, Pascal Hitzler, Mike Dean, and Dave Kolas. *Spatial Information Theory: 11th International Conference, COSIT 2013, Scarborough, UK, September 2-6, 2013. Proceedings*, chapter A Geo-ontology Design Pattern for Semantic Trajectories, pages 438–456. Springer International Publishing, Cham, 2013.

[15] Axel-Cyrille Ngonga Ngomo and Sören Auer. LIMES - A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data. *integration*, 15:3, 2011.

[16] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady Andrienko, Natalia Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, Jose Antonio de Macedo, Nikos Pelekis, et al. Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, 45(4):42, 2013.

[17] Nikos Pelekis and Yannis Theodoridis. Semantic aspects on mobility data. In *Mobility Data Management and Exploration*, pages 189–209. Springer, 2014.

[18] Chiara Renso, Miriam Baglioni, Jose Antonio de Macedo, Roberto Trasarti, and Monica Wachowicz. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and information systems*, 37(2):331–362, 2013.

[19] Jose Antonio M. R. Rocha, Valéria Cesário Times, Gabriel Oliveira, Luis Otávio Alvares, and Vania Bogorny. Db-smot: A direction-based spatio-temporal clustering method. In *IEEE Conf. of Intelligent Systems*, pages 114–119. IEEE, 2010.

[20] Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, and Christelle Vangenot. A conceptual view on trajectories. *Data & knowledge engineering*, 65(1):126–146, 2008.

[21] Vânia MP Vidal, Marco A Casanova, Narciso Arruda, Mariano Roberval, Luiz Paes Leme, Giseli Rabello Lopes, and Chiara Renso. Specification and incremental maintenance of linked data mashup views. In *Advanced Information Systems Engineering*, pages 214–229. Springer, 2015.

[22] Julius Volz, Christian Bizer, Martin Gaedke, and Georgi Kobilarov. Silk-a link discovery framework for the web of data. *LDOW*, 538, 2009.

[23] Wei Wang, Guosun Zeng, and Daizhong Tang. Bayesian intelligent semantic mashup for tourism. *Concurrency and Computation: Practice and Experience*, 23(8):850–862, 2011.

[24] Zhixian Yan, Jose Antonio de Macedo, Christine Parent, and Stefano Spaccapietra. Trajectory ontologies and queries. *Transactions in GIS*, 12(s1):75–91, 2008.

[25] Ji Yuan, Xudong Liu, Richong Zhang, Hailong Sun, Xiaohui Guo, and Yanghao Wang. Discovering semantic mobility pattern from check-in data. In *Web Information Systems Engineering–WISE 2014*, pages 464–479. Springer, 2014.

[26] Vincent W. Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 1029–1038, New York, NY, USA, 2010. ACM.

[27] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. Understanding transportation modes based on gps data for web applications. *ACM Trans. Web*, 4(1):1:1–1:36, January 2010.