ARTICLE TEMPLATE

An Ontology-based approach to Knowledge-assisted Integration and Visualization of Urban Mobility Data

Thiago Sobral[a,b] , Teresa Galvão[a,b] and José Borges[a,b]

[a]INESC TEC, Porto, Portugal; [b]Faculty of Engineering of the University of Porto, Portugal

ARTICLE HISTORY
Compiled September 2, 2019

ABSTRACT
Visualizing data from multiple heterogeneous sources may be technically demanding for transportation stakeholders, due to technical and human factors. Several visualization techniques have been proposed to support the analysis of urban mobility phenomena, although the relation between data integration and visualization remains an unexplored topic in Transportation studies. This paper proposes an ontology-based approach to support knowledge-assisted integration and visualization of spatio-temporal data from Intelligent Transportation Systems. VUMO (Visualization-oriented Urban Mobility Ontology) can provide a semantic foundation to knowledge-assisted visualization tools (KVTs), which can reduce the technical burden that often limits the use of data visualization in practice. VUMO contains three facets that interrelate the characteristics of spatio-temporal mobility data, visualization techniques and expert knowledge. A built-in ruleset leverages semantic technologies standards to infer which visualization techniques are compatible with analytical tasks, and to discover implicit relations within data. The representation of expert knowledge formalizes qualitative and quantitative information from domain experts that can be exploited by recommendation methods to automate part of the visualization workflow. Data from the city of Porto, Portugal were used to demonstrate practical applications of the ontology for each facet. As a foundational ontology, VUMO can be extended to meet the distinctiveness of a KVT.

KEYWORDS
data integration; data visualization; urban mobility; semantic web; ontology

1. Introduction

Intelligent Transportation Systems (ITS) and ubiquitous computing generate data that, when duly exploited, can provide better understanding of the dynamics of people in cities. Since the last decade, transportation researchers and stakeholders – domain experts – have developed interactive tools featuring novel visual metaphors to aid decision making through exploratory analysis of spatio-temporal (S-T) data.

In practice, technical and human factors still hinder the use of data visualization techniques, in particular when various datasets are considered simultaneously. Technical factors include syntactic and semantic data heterogeneity, as ITSs have diverse data models. The absence of a unifying abstraction layer to represent data hinders the process of data integration, and may lead to ineffective approaches in which datasets are not analyzed in a integrated way, thus contradicting the logic of exploiting the

CONTACT Thiago Sobral. Author. Email: thiago.sobral@fe.up.pt

large amount of data generated by ITSs. Moreover, *ad hoc* visualization tools may have limited scalability and interoperability, in the sense that they require datasets to follow a particular schema or syntax, as identified in related studies. Geographic Information Systems (GIS) provide means for integrating data published according to well-known syntaxes, e.g. JSON or CSV, but fall short on providing novel visualization techniques found in recent Transportation studies.

Human factors include limited knowledge of data visualization and manipulation of massive datasets. The use of popular visualization tools, such as Excel and Tableau, may not suffice to explore multidimensional urban mobility data. Moreover, the choice of appropriate visual metaphors for a given analytical task is not trivial, as it depends on the datasets to be analyzed, the questions that such a task should answer, and the analytical profile of those who carry out such a task.

This article proposes an ontology-based approach to address both technical and human factors that affect integration and visualization of heterogeneous S-T urban mobility data. The approach leverages semantics as a foundation for the development of knowledge-assisted visualization tools (KVTs) that can be used by domain experts with distinct analytical profiles. Besides allowing for data interoperability, KVTs can also automate part of visualization techniques selection by incorporating tailored recommendation method according to the desired analytical task, and available empirical knowledge.

The core contribution is the Visualization-oriented Urban Mobility Ontology (VUMO). The ontology provides a formal knowledge representation model for describing ITS instance data, and annotating visualization techniques and expert knowledge. The ontology provides constructs to represent knowledge that is usually subjective regarding domain experts' preferences, and empirical feedback about visualization techniques. The present paper significantly updates previous studies of ours (Sobral, Galvão, & Borges, 2017) on regards to the ontology structure and the process for interrelating the three facets.

The ontology interrelates three facets: (1) ITS data; (2) visualization techniques; and (3) empirical knowledge retrieved from domain experts. VUMO embeds reasoning rules that infer, for example, the compatibility of visualization techniques with analytical tasks based on their semantic annotation. Rules also infer implicit knowledge that can feed recommendation methods, such as the visual patterns of spatial events based on the constructs that describe them, and the structure of analytical tasks. Heterogeneous instance data can be semantically aligned with VUMO constructs, so that distinct objects in distinct datasets that convey the same semantics can have a unified representation, regardless of their original syntax or schema. The rule set uses *de facto* Semantic Web standards, such as SPIN and SPARQL.

As an illustration, the typical use case of a KVT begins with the definition of analytical tasks, i.e. questions about data, such as "ridership of stops located within downtown area during peak-hour traffic", which are represented by one or more queries. The tool then suggests one or more candidate visualization techniques based on the existing information for the three facets, including the structure of analytical tasks, previous records of experts' preferences, profiles, and previous ratings given to each technique.

In accordance to the rationale of the use of ontologies and Semantic Web technologies, VUMO's modular structure accounts for scalable reusability, i.e. it is possible to extend it to meet distinct requirements for data integration and the development of KVTs. The proposed representation of S-T data aims to allow modeling the majority of spatial events types found in ITS data. The annotation of visualization techniques draws concepts from Information Visualization theory. Moreover, the available constructs for annotating expert knowledge are agnostic with respect to recommendation methods, hence they can feed a broad spectrum of such algorithms.

The remainder of this paper is structured as follows: Section 2 surveys the applications of visualization and Semantic Web technologies to studies on urban mobility. Section 3 lays the theoretical foundation of our approach by proposing a model that formalizes and interrelates the various facets of a KVT for S-T data. Section 4 describes the VUMO ontology and provides the practical foundation by implementing such formal model. Some real-world examples are also described to illustrate its applicability. Section 5 provides a practical demonstration of our approach using data from Porto's public transportation system in Portugal. Section 7 concludes this article and states future research directions.

2. Related work

Since the last decade, interactive visualization tools have been developed to support the analysis of urban mobility phenomena such as traffic congestion (Cheng, Tanaksaranond, Brunsdon, & Haworth, 2013; Wang, Lu, Yuan, Zhang, & Wetering, 2013); vehicle and passenger travel patterns (Andrienko, Andrienko, & Rinzivillo, 2016; Chen et al., 2016; Huang et al., 2016; Lu, Liang, Wang, & Yuan, 2016; Mao, Ji, & Liu, 2016; Nunes, Ribeiro, Prandi, & Nisi, 2017; von Landesberger et al., 2016); travel times and people's behavior (Wu et al., 2012; Zeng, Fu, Arisona, Erath, & Qu, 2014). In comparison to other topics in urban mobility and Transportation in general, the share of visualization studies was still small in 2011 (Zhang et al., 2011), although a new stream of studies gained momentum in the last five years. The earliest applications of data visualization to urban mobility analysis were based on Geographic Information Systems (GIS). Currently, software packages, map services and visualization frameworks facilitate the development of GIS-independent visualization tools with novel visual metaphors to represent multidimensional data, and were influential for the their development. The surveyed studies do not address the issue of data heterogeneity, thus each of the proposed visualization techniques and tools require data in a different shape or form.

In parallel, the availability of spatio-temporal data has increased due to systems like Automated Fare Collection (AFC), Traveler Information Systems (TIS), Automatic Number Plate Recognition (ANPR), and sensing devices, e.g. loop sensors and smartphones. The wide availability of ITSs specifications introduces new challenges for data integration and interoperability across visualization tools. Those issues were already pointed out by Hughes (2005), in a research agenda for the applications of data visualization to Transportation. In one of the agenda's topics, "Development of Visualization Standards for Source Data and Interoperability", the authors discussed the importance of meta-data on ensuring interoperability across systems, including visualization tools, to support heterogeneous data sources. To the best of our knowledge, few studies partially addressed this topic.

Some studies addressed the integration of urban mobility data by leveraging ontologies and Semantic Web technologies. Those technologies have been acknowledged by researchers and practitioners as an effective approach to data integration in various domains, including genomics (The Gene Ontology Consortium, 2019) and enterprise modeling (De Giacomo, Lembo, Lenzerini, Poggi, & Rosati, 2018). In Transportation, ontologies have been applied to other problems, e.g. knowledge management and travel planning (Katsumi & Fox, 2018). The work of Oliveira et al. proposed an ontology to support passenger travel planning, and to automate generation of user interfaces for interactive traveler systems using rule-based inference (Oliveira, Bacha, Mnasser, & Abed, 2013). Lorenz et al. proposed the Ontology of Transportation Networks (OTN) to model the topology of a transportation network, including public transportation systems (Lorenz, Ohlbach, & Yang, 2005). OTN has limited capabilities on regards to modeling spatial events, as it is oriented towards mostly static aspects of a transportation network. The process of semantic integration of heterogeneous data sources has been explored by Psyllidis et al. (Psyllidis, 2016). The authors developed the OSMoSys Knowledge Representation Framework and the ROUTE ontology for applying computing methods and semantic web technologies to city analytics and planning. The SocialGlass prototype was developed to demonstrate the framework. Data analysis is assisted by a pre-defined set of visualization techniques that can be manually selected. Still, few constructs are provided to model spatial events. Recently, ontologies have been applied to translate standard data models, e.g. Transmodel, into their semantic counterparts, and support performance monitoring in public transportation systems (Benvenuti, Diamantini, Potena, & Storti, 2017). The link between heterogeneous urban mobility data and visualization still offers research opportunities, especially on regards to S-T data which is becoming increasingly available by ITSs.

Domain experts may be unacquainted with technical visualization knowledge, thus posing negative implications for data exploration (Mutlu, Veas, & Trattner, 2016). To overcome this problem, studies in Computer Science proposed KVTs, also denominated Visualization Recommender Systems, which also offer research opportunities, due to the impossibility of devising solutions that simultaneously fit all contexts of use and the several domains of knowledge, as we intend to explore in the present article. The following literature addresses the development of KVTs for general domain data.

The work of Voigt et al. is among the earliest in knowledge-assisted visualization using ontologies (Voigt, Franke, & Meissner, 2013). They designed a Collaborative Filtering (CF) based algorithm for suggesting visualizations for semantic data. The algorithm uses information about visualization components, data context, and ratings provided by system users. A formal ontology, VISO, was developed to describe data schema and the components of visualization techniques. The approach assumes that data are semantically integrated, i.e. aligned according to one or more domain ontologies. Mutlu et al. carried out a comprehensive approach by involving lay users in a crowd-sourced study with general public (Mutlu et al., 2016), to use and evaluate the VizRec system. Several dimensions for rating visualizations were defined according to usability factors. The collected empirical knowledge was used to evaluate the accuracy of various recommendation methods.

In summation, novel applications of visualization techniques to urban mobility analysis have emerged, although it is still unclear how such techniques can overcome the issue of data heterogeneity in the context of heterogeneous ITSs specifications. KVTs can assist users with limited knowledge to visually explore data, although no applications of such systems to urban mobility and Transportation were identified. The surveyed KVTs in Computer Science leverage ontologies as a foundation for data integration and knowledge representation. In order to investigate the relevance of KVTs in urban mobility, it is necessary to use domain ontologies as a foundation for such systems. Nonetheless, existing urban mobility ontologies are still insufficient to model spatial events, and do not address the remaining facets that are also relevant for the process of data visualization. The present paper intends to contribute towards closing this gap, and providing a novel application of KVTs to the transportation domain.

3. Modeling Data, Visualization Techniques and Expert Knowledge for KVTs

In this section, we propose a conceptual formalization for S-T urban mobility data, visualization techniques and empirical knowledge collected from domain experts. Such formalization is required to provide common, coherent semantics to the three facets that will form a KVT according to our proposed approach. The resulting model lays the theoretical foundation for building the VUMO ontology.

The formalization of data is built upon acknowledged frameworks for modeling geographic and movement data (Aigner, Miksch, Schumann, & Tominski, 2011; Andrienko & Andrienko, 2006; Chrisman, 1997; Peuquet, 2002), in which the work of Aigner et al. provides important contributions for visualization of time-oriented data. We define every instance of spatio-temporal data as an *event*. An event is an action performed by one or more agents of a transportation network, and occurs at/in one or more spatial and temporal dimensions. Events can also be described in terms of thematic attributes (characteristics). An entity-based perspective was adopted (Andrienko & Andrienko, 2006), thus space and time were regarded as attributes of every conceptual entity represented by the model. The choice for this perspective is justified by the nature of semantic instance data: an entity is an object that exists by itself, which is described in terms of other instances.

The following subsections formalize the fundamental structures for KVTs. Spatial references and distributions (3.1, 3.2) and temporal references (3.3), thematic attributes (3.4) and data transformations (3.5) form the facet that describes ITS instance data. Visualization techniques (3.6) formalize their intrinsic properties and their link to data transformations, thus forming the facet for annotating visualization techniques. The formalization of empirical knowledge (3.7) includes the specification of domain experts and the structure of the feedback they can provide to a KVT in the form of ratings. Throughout the subsections, it is shown how the concepts (classes) of this conceptual model are interrelated. Finally, the UML Class Diagram in Figure 2 shows a schematic representation of the model. The color gray indicates classes that refer to S-T data; blue refers to visualization techniques classes; green refer to expert knowledge classes.

3.1. Spatial References

An entity may have one or more spatial attributes depending on the nature of data. Two entity categories were defined: *Point* and *PointSet*.

Definition 3.1. A *Point* is described by latitude and longitude, following the WGS84 datum. This definition is flexible in the sense that other optional attributes may exist, e.g. elevation.

Points are divided into two disjoint subcategories:

Definition 3.2. A *Generic Point* is a *Point* that does not have any thematic attribute that acts as an identifier, i.e. does not contain any property that assigns a unique identification, name, code, or any other textual attribute in a dataset.

Generic points are references that, in practice, do not require a unique identification, e.g. the location of a citizen at a given time, retrieved from a GPS-assisted device.

Definition 3.3. A *Known Point* is a *Point* that contains at least one identifier attribute.

Known points are those for which the identification is relevant for some practical purpose. For instance, the bus stops of a network are known points, as it is possible to retrieve their spatial characteristics by knowing their identification, e.g., `STCP_AEPT1`.

Definition 3.4. A *Point Set* is a (un)ordered collection of two or more *Point*s.

Point sets are divided into two disjoint subcategories:

Definition 3.5. An *Ordered Point Set* P is well-ordered, i.e. every non-empty subset of P has a least element in this ordering.

Definition 3.6. An *Unordered Point Set* is a set of points that is not well-ordered.

In contrast to a Point, the disjointness assumption was not imposed for two reasons: (1) to account for situations where the notion of order or its absence is semantically irrelevant, and (2) to reduce computational complexity at inference time with VUMO rules.
An example of an ordered point set is a passenger route plan, formed by points which describe the departure and arrival locations. It is possible to infer the rank of any point. In contrast, shapes can be described in terms of unordered point sets, e.g., a set of Points which defines the boundaries of a polygon representing a public transportation system zone.

3.2. Spatial Distributions

Spatial references induce distinct visual arrangements according to their type. Such arrangements are defined as Spatial Distributions. Three primary arrangements are considered: Discrete, Quasi-continuous and Graph. Without loss of generality, spatial distributions are present in both geographic and abstract spaces. Spatial arrangements become apparent as the number of entities increase in visualization space, as shown in Figure 1. We defined the following axioms that link spatial reference types to distributions:

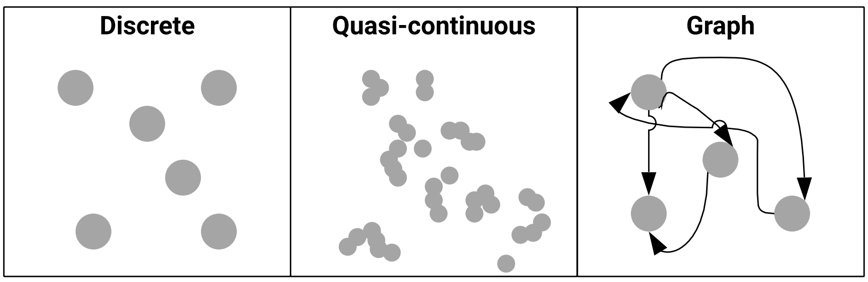


Figure 1. Examples of Discrete, Quasi-continuous and Graph spatial distributions in visualization space, based on various types of spatial reference types. Visual arrangements are present, without loss of generality, in both geographic and abstract spaces.

Definition 3.7. A *Generic Point* induces a *Quasi-Continuous* spatial distribution.

Generic points are related to entities that can be spatially located in any location within a geographic or abstract grid. Hence, the visual pattern depicted by some of those points induces the notion of smoothness (continuity).

Definition 3.8. A *Known Point* induces a *Discrete* spatial distribution.

The requirement of one or more identifier attributes suggests that known points represent entities that have a well-defined location within a region, and are less dense than generic points.

Definition 3.9. An *Unordered Point Set* induces a Discrete spatial distribution. Unordered point sets yield structures such as shapes and clusters.

Definition 3.10. An *Ordered Point Set* induces a Graph spatial distribution. They form arrangements that resemble the notion of trajectory.

Definition 3.11. A *Point Set* inherits the spatial distribution(s) of its points.

The practical purpose of Definition 3.11 is to account for situations in which a visualization provides ways to not only explore point sets as a whole, but to shift the perspective to its intrinsic elements, using interaction mechanisms like semantic zoom. Point sets can be interpreted as paths in the mathematical sense, either ordered or not. Hence, point sets can also induce *Discrete* or *Quasi-Continuous* spatial distributions, depending on the categories of their points.

3.3. Temporal References

An entity may have one or more temporal attributes depending on the nature of data. Two types were defined: *Instant* and *Interval*.

Definition 3.12. An *Instant* is described by a timestamp, i.e. containing date and time, and an optional time zone description.

Definition 3.13. An *Interval* is described by two timestamps, corresponding to start and end.

Although an instant could be considered a zero-length interval in the mathematical sense, we argue that such consideration does not add practical value in the context of data integration and KVTs.

3.4. Thematic attributes

Thematic attributes can have various types, e.g., strings, numerical values, binary file structures or geometric shapes. Due to our orientation towards visualization, we introduce the concept of measures of an entity.

Definition 3.14. A *Measure* is a certain quantity or degree of something, expressed by a quantitative, ordinal or categorical value.

In practice, a spatial event instance depicting an accident may have measures as characteristics, such as the number of injured people and damaged vehicles.

3.5. Transformations

Domain experts carry analytical tasks by performing data transformations (e.g. queries) or more complex operations (e.g. data mining techniques) to extract useful information. We define a *transformation* as a sequence of operations that uses raw data as input, and produces a new dataset as output. This output may contain the same entities found in raw data, and new ones that derive from such operations.

The output of a transformation, herein defined as a set of *output variables*, may contain spatio-temporal references, depending on the output they generate. Likewise, it can generate new measures based on raw data, or use existing ones as output. The existence of spatial references in the output of a transformation implies that there are also visual spatial patterns associated with it. The following proposition demonstrates this statement.

Proposition 3.1. *A transformation may have zero or more spatial distributions.*

Proof. Let T be a transformation that yields $\{o_1, ..., o_n\}$ as output variables, where $O_s \subseteq O$ is a non-empty subset of O containing spatial references, with $1 \leq i \leq n$. By definition, a spatial reference a_i induces one or more spatial distributions, hence the output of T also induces one or more spatial distributions. In the null case, it suffices to consider that all outputs are attributes other than spatial references. $\square$

3.6. Visualization techniques

A visualization technique has one or more features and provides one or more interaction tasks. Visualization techniques features are the intrinsic components for data visualization. The proposed model specifies four features: *input variable*, *reference frame*, *spatial dimensionality* and *temporal arrangement*. The last three features are derived from a classification of visualization techniques from Aigner et al. (2011).

Input variables are responsible for receiving the values from output variables (of a transformation) that will be mapped onto visual variables.

Reference frame describes the ability of a visualization technique to represent *geographic* (georeferenced) data, i.e. map-based visualizations, and *abstract* data.

Spatial dimensionality describes the number of dimensions used by the visualization canvas, e.g. 2D or 3D.

Temporal arrangement describes how the time dimension is represented on the visualization canvas, e.g. linear, cyclic.

3.7. Empirical knowledge

Empirical knowledge allows for the specification of the users of a KVT, herein considered domain experts, and the annotation of feedback collected during its use.

A user has an analytical profile, which is specified according to the system's objectives. For instance, in our former work, we approached experts that belonged to strategical and operational profiles (Sobral et al., 2017). Such characterization is not meant to be neither exhaustive nor unique.

We define two perspectives for user feedback. The first consists of statements related only to a visualization technique, such as ratings about one or more usability factors as in Aigner et al. (2011), e.g. visual complexity, or more subjective statements, e.g. "this visualization is recommended for analyzing ridership from smart card transactions".

The second perspective allows users to provide specialized feedback about visualizations with respect to a transformation, which is defined as a *cross rating*. In cases in which the feedback is provided as a quantitative measure, a technique rating has a rating score related to a rating component. Qualitative feedback is expressed in terms of categorical values. A quantitative rating component is subject to a scale and polarization, i.e. whether its value impacts positively or negatively in the overall rating. Both perspectives for user feedback can feed recommendation methods to provide suggestions of visualization techniques. Section 5 introduces practical modeling examples of such kinds.

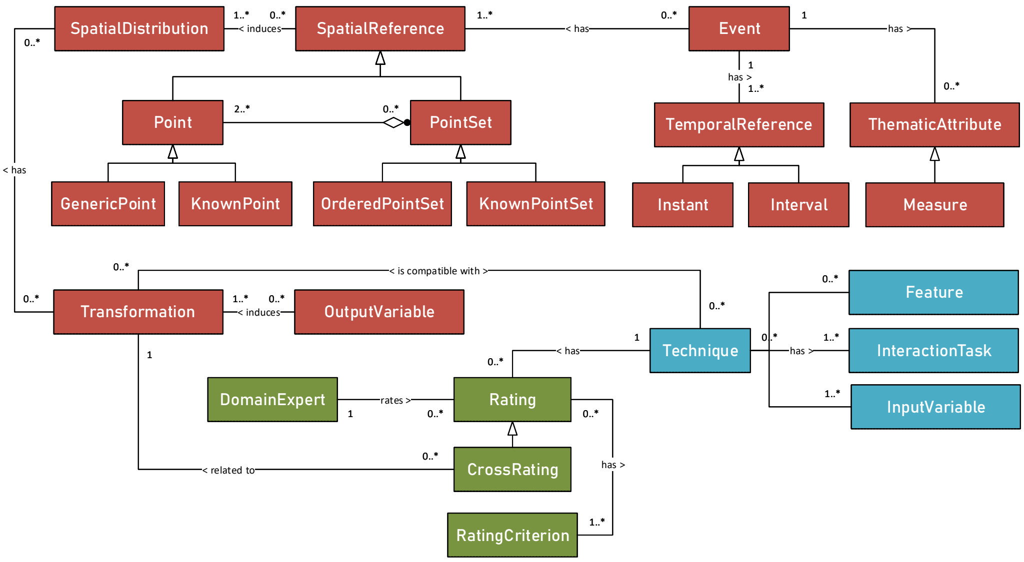


Figure 2. UML Class Diagram for the conceptual model, which serves as the basis for the implementation of the VUMO ontology. The three facets represent ITS instance data (gray), visualization technique (blue), and expert knowledge (green).

4. The VUMO Ontology

This section describes the Visualization-oriented Urban Mobility Ontology (VUMO), which lays the semantic foundation for integrating S-T urban mobility data from heterogeneous sources, and building knowledge-assisted visualization tools. VUMO supports the following tasks:

- Integration of multi-source heterogeneous urban mobility data related to spatial events, and their description in terms of transportation network elements;
- Specification of analytical tasks that domain experts want to carry out in the form of data transformations (queries);
- Annotation of visualization techniques implemented in a visualization tool, using concepts from Information Visualization theory;
- Annotation of empirical domain expert knowledge, e.g. user information and feedback about visualization techniques in the form of qualitative or quantitative ratings, which can feed recommendation methods;

- Rule-based inference of implicit knowledge from instance data that is relevant for the data exploration and visualization process, e.g. implicit links between instances from distinct datasets; characteristics of data transformations; and compatibility between data transformations and visualization techniques.

The structure of VUMO is modular, i.e. classes and properties were thought with the goal of having a well defined role on the development of KVTs, according to the following pipelines we defined:

(1) *Data integration*: visualization tools should allow to integrate data from multiple sources. The data structure not only maintains the original attributes of instance data; the structure is also used to infer visual attributes that can be considered when recommending visualization techniques;
(2) *Visualization technique design and development*: a visualization technique should be characterized in terms of its intrinsic attributes. Such attributes are considered during rule-based inference of compatibility between visualization techniques and data transformations (defined in Subsection 3.5), and to aid users on the process of finding appropriate visualization techniques;
(3) *Visualization technique evaluation and specification of users*: user feedback about visualization techniques should be formalized by using the constructs available in the ontology. The annotation of users (domain experts) can also be exploited by recommendation methods.

VUMO is an OWL ontology that conforms to the OWL 2 RL profile, which is a syntactic subset of OWL 2 that supports rule-based inference by trading some of its logical expressiveness (W3C OWL Working Group, 2012). The implementation of inference rules and functions uses the SPIN vocabulary and modeling language, hence they are expressed as queries in standard SPARQL language, which are *de facto* technologic standards for the Semantic Web. The ontology was built according to a top-down approach, i.e. upper classes and properties were defined and further refined, and reuses existing ontologies for topological spatial relationships (GeoSPARQL) and organization of abstract knowledge (Simple Knowledge Ontology Specification). Given that VUMO is strongly oriented to practical contexts, concepts were modeled after analyzing real-world datasets. In addition, the ontology implements a full semantic counterpart of the Google Transit Feed Specification (GTFS) schema, to facilitate the integration of data published according to this standard.

Subsection 4.1 introduces the upper classes of VUMO. Subsections 4.2, 4.3, 4.4 and 4.5 thoroughly describe the ontology upper classes and their components. Section 4.6 describes the rules and functions embedded into VUMO that support inference of implicit knowledge.

4.1. Upper classes

VUMO is divided into four upper classes:

- *UrbanMobilityConcept (UMC)*: a superclass for all classes that describe urban mobility concepts, including spatial events;
- *DataConcept (DC)*: a superclass for all classes that relate to S-T data and data transformations, as in Subsections 3.1-3.5;
- *VisualizationConcept (VC)*: a superclass for all classes that characterize a visualization technique, its features and interaction tasks, as in Subsection 3.6;

- *DomainExpertConcept (DEC)*: a superclass for all classes that characterize domain experts and their empirical knowledge about visualization techniques, as in Subsection 3.7.

Table 1 shows the role of classes in each pipeline. A main role ($\star$) means that visualization tools developers (and users) will mostly use elements from that class on a specific pipeline. An auxiliary role ($\circ$) means that such a class is mostly referred indirectly (automatically) through rule-based inference, thus not requiring human intervention, unless it is desired to extend the ontology capabilities, e.g. by introducing new rules.

Table 1. The role of each superclass on the pipelines of a visualization system. A main role ($\star$) indicate a strong, direct use of elements of that class on a pipeline. An auxiliary role ($\circ$) indicate that such class mostly support a specific pipeline, through rule-based inference, thus requiring minimal to none manual intervention.

Pipeline	UMC	DC	VC	DUC
Data integration	$\star$	$\circ$		
Visualization design and development	$\circ$	$\circ$	$\star$	$\circ$
Visualization evaluation and system user specification	$\circ$	$\star$	$\star$	$\star$

Upper classes branch out into subclasses that represent more specific concepts. VUMO contains Object and Datatype properties to relate instances from the aforementioned classes. Semantic data is herein represented using the standard form, i.e. subject-predicate-object triples, according to the Resource Description Framework (RDF) data model. Table 2 provides a natural language definition for each first-level subclass.

4.2. UrbanMobilityConcept (UMC)

UMC concepts are fundamental to semantic integration of raw ITS data. The `Agent` and `InfrastructureComponent` subclasses group structural concepts of a transportation system. An `Event` enables to describe distinct types of spatial events, e.g. smart card transactions from AFC systems, trajectories from AVL sensors, among others. Table 2 provides a natural language definition for the subclasses of UMC.

If applicable, an instance can be provided an identification by asserting the property `hasID` or its semantically equivalent subproperties, such as `hasInternalID` or `hasFriendlyID`. To illustrate the utility of multiple ID properties, consider an example of bus stop shown in Figure 3 (left) from the city of Porto. `AEPT1` is a user friendly identification used by passengers to consult schedules using real time services. From the operator's perspective, two internal identifiers are used: `STCP_AEPT1` or `54`. In the AFC dataset, which does not follow a standard schema, the three IDs uniquely determine the stop within their distinct semantic contexts, yet they refer to the same entity. The semantics of the OWL language allows the inference that all given IDs relate to the same entity.

The bus operator (STCP) dataset describes the network according to a different schema: GTFS. The same stop is described in Figure 3 (right). After integration, both stop instances are need to be aligned, so that computers can recognize them as semantically equivalent. The `owl:sameAs` property is used for that purpose. In particular, the latitude and longitude coordinates are slightly different across datasets.

Table 2. Upper classes of VUMO and their respective first-level subclasses

Upper class	Subclass	Definition
Urban Mobility Concept (UMC)	*Agent*	- An entity that is part of the urban mobility network, which can perform actions such as *Events*.
	InfrastructureComponent	- A physical or abstract entity. *Agents* can use it (in)directly trigger *Events*, or to provide them with context information.
	Event	- An action performed by one or more *Agents*, which may take place in multiple space and time dimensions.
Data Concept (DC)	*SpatialReference*	- A type of spatial reference.
	TemporalReference	- A type of temporal reference.
	SpatialDistribution	- A type of visual arrangement of spatial data.
	SpatialDistributionAxiom	- An abstract container for the spatial distribution axioms defined in the conceptual model.
	Transformation	- A sequence of operations (e.g. query) that yields new information from raw data.
Visualization Concept (VC)	*Technique*	- A visualization technique available in a VUMO-based visualization system.
	Feature	- An intrinsic characteristic of a *VTechnique*.
	InteractionTask	- An interaction task available in a *Technique*.
Domain Expert Concept (DEC)	*DomainUser*	- A user of a VUMO-based visualization system.
	DomainUserProfile	- The profile of a *DomainExpert* with respect to the context of his/her activities, e.g. *Strategical*, *Operational*.
	Rating	- An abstract container that stores rating information made by a *DomainUser* with respect to a *Technique*.
	RatingComponent	- An abstract component evaluated in a rating that impacts the user experience, e.g. *Difficulty*, *Visual clutter*. Such components are divided into *Positive/NegativeRatingComponent*.
	CrossRating	- An abstract container that stores rating information about a *Technique* with respect to a *Transformation*. A *CrossRating* provides a specialized rating by relating a *Technique* to a *Transformation*.

Table 2. First-level and further subclasses of UrbanMobilityConcept (UMC)

Subclass	Second-level subclass	Definition and further subclasses (if applicable)
Agent	*Operator*	- A transportation operator, e.g. bus or subway companies, taxi agencies and shared bicycles operators).
	Passenger	- An individual who uses a public transportation system.
	Vehicle	- An action performed by one or more *Agents*, which may take place in multiple space and time dimensions, e.g. buses, trains and bicycles.
Infrastructure Component	*Line*	- A PTS line that may consist of various *Routes*.
	Node	- A node of the transportation network graph, e.g. *BusStop*, *SubwayStation*, *Sensor*, *BicycleStation*.
	Zone	- A pre-defined geographic zone used for a specific purpose, e.g. to define fares.
	Route	- A path followed by a *Line*. Lines may consist of several routes
	RouteSegment	- An elementary part of a *Route*, generally defined by two *Nodes*.
	Ticket	- A ticket that allows a passenger to travel within a public transportation network.
	TicketType	- The type of a ticket.
Event	*TravelEvent*	- A travel made by a passenger within a transportation network triggered by an action, e.g. *TicketValidation*, *BicycleTrip*.
	TravelIntention	- An intention of traveling within a transportation network, based on information requests, e.g. schedule requests or route plans.
	SensorReading	- A reading made by a *Sensor*.
	SocialMediaPost	- A post from social media regarding transportation information.
	TripEvent	- A trip made by a *Vehicle*, e.g. bus or taxi.
	UnexpectedEvent	- An unforeseen *Event*, often capable of (partly) disrupting a transportation network in some way.

The distinct fare zones (N10 and VCD8) are due to the age of datasets, with the GTFS one being the most recent. In Section 5, it is discussed how VUMO supported semantic equivalence of fare zones in the context of a recent change in Porto's public transportation network.

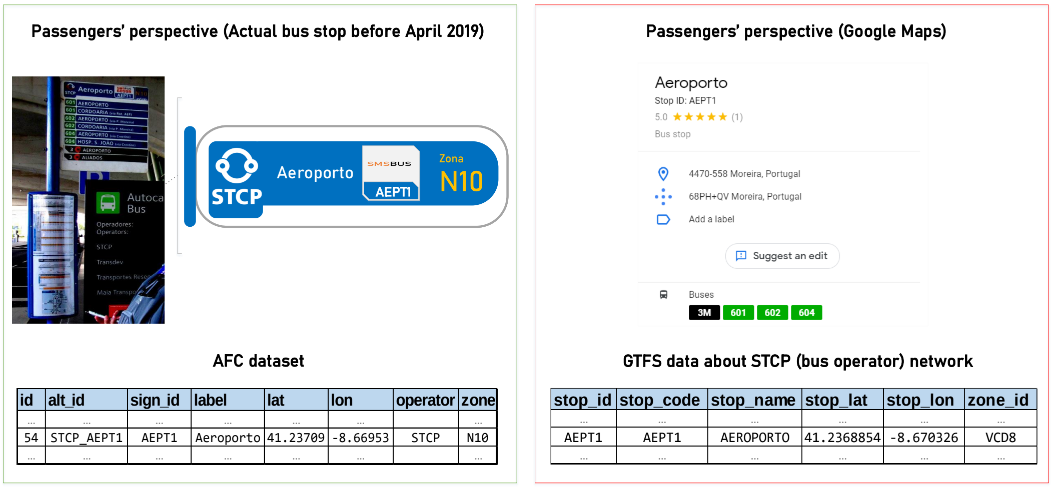


Figure 3. An illustration of a bus stop sign of STCP operator in Porto (left), and its representation with the AFC dataset from the smart card operator. The identification AEPT1 is meant to be used by passengers when checking schedules in a real-time service. From the operator's perspective, multiple internal identifications may exist for the same stop. The GTFS representation of the network (right) has a different schema in comparison to the AFC dataset and different values for each attribute, yet they refer to the same bus stop. VUMO's semantics is able to align heterogeneous instance data while preserving their distinctiveness.

Measures are available as as subproperties of `hasMeasure` and expect literal values. Some of the existing properties include:

- `hasDuration`;
- `hasNumberOfInjuredPassengers`;
- `hasNumberOfAvailableBicycles`.

The latter two properties can be used to describe, for example, an `Accident` or the status of a `BicycleStation`. It is possible to infer all measures related to a spatial event or entity, as they are subproperties of `hasMeasure`.

VUMO facilitates the integration of geometry information for spatial references by supporting WKT literals that describe geometric entites like points, polygons and collections of such kinds. The alternative dual representation uses the WGS84 vocabulary to encode latitude and longitude of points. Polygons are represented as linked lists using RDF properties that semantically encode an ordered relation. In practice, most datasets still do not encode geospatial data as WKT elements beforehand. VUMO provides built-in SPIN rules that automatically infer its dual. Figure 4(a) shows an example in which a dataset. In (b), the polygon that describes a fare zone consists of a linked list of points, from which its WKT counterpart was inferred. The dual representation preserves the semantics of geospatial data, even if a RDF triple store does not support GeoSPARQL. The built-in semantics of VUMO allows the inference of spatial references that are not originally asserted in data, due to the transitivity of the property `location`. For instance, Figure 4(c) demonstrates how a ticket validation instance `TV1` is automatically inferred to have the zone `VCD8`, as it contains the bus stop `AEPT1`.

Temporal references can be expressed in manifold forms depending on the desired type of data granularity, by using the constructs from the Time ontology. For instance, Figure 4(c), the temporal reference is a timestamp. In Figure 4(d), an origin-destination flow measurement `OD1` has a recurrent temporal reference, `Weekdays-AM-Peak`, which was created to depict morning peak traffic period between 7 a.m. and 9 a.m.. The representation of temporal references in such a fashion allows for a finer and reusable representation of temporal entities, and facilitates the construction of data queries, which would otherwise be cumbersome to express in relational databases and SQL dialects.

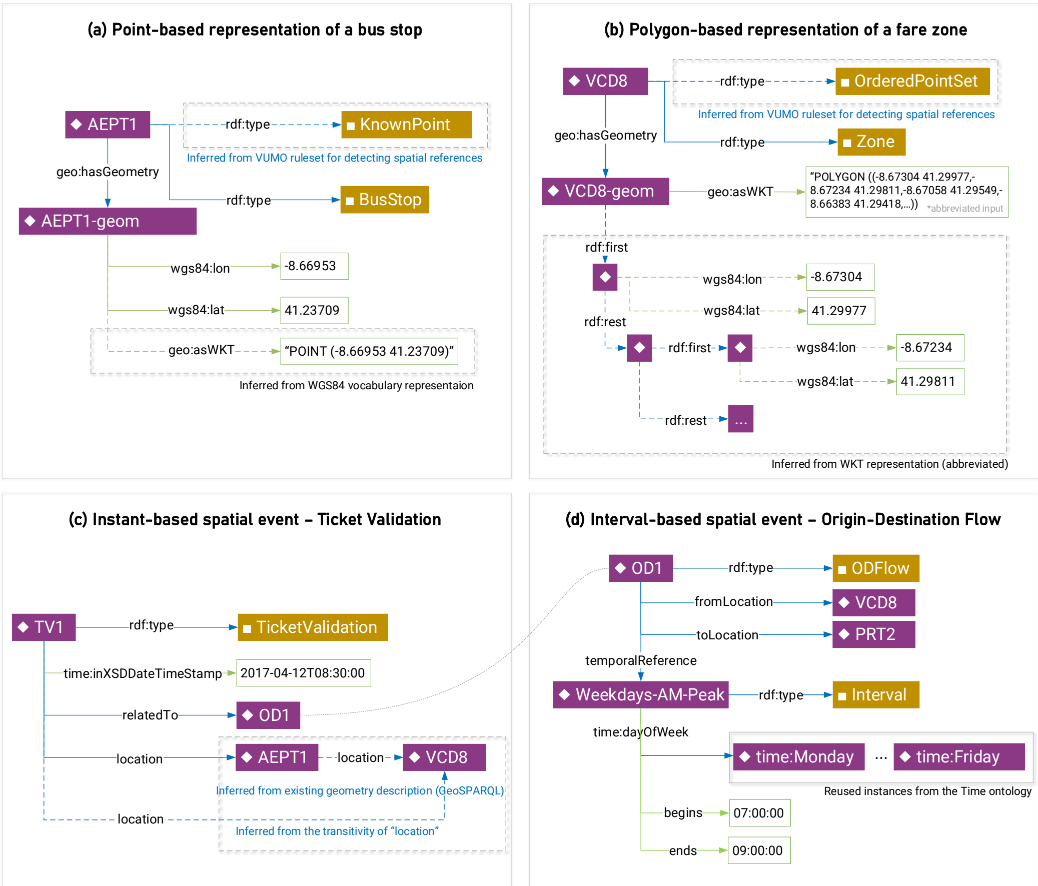


Figure 4. Semantic representation of two spatial references and their geometries: a bus stop (left) and a fare zone (right). The AEPT1 instance contains a point-type geometry originally described with the WGS84 vocabulary. The VCD8 instance contains a polygon-type geometry in WKT. VUMO's built-in logic automatically infers the dual representation of each geometry, and the type of spatial reference.

As the volume of integrated instance data grows, it is possible to visualize the complex interrelation between datasets. Figure 5 shows the interconnection between instance data from various datasets, which is described in Section 5. A symbolic notation was adopted for representing data: classes (□) and their instances (◊). Object and Datatype properties are represented by blue and green edges. Solid and dashed edges indicate asserted (explicit) and inferred (implicit) triples, respectively.

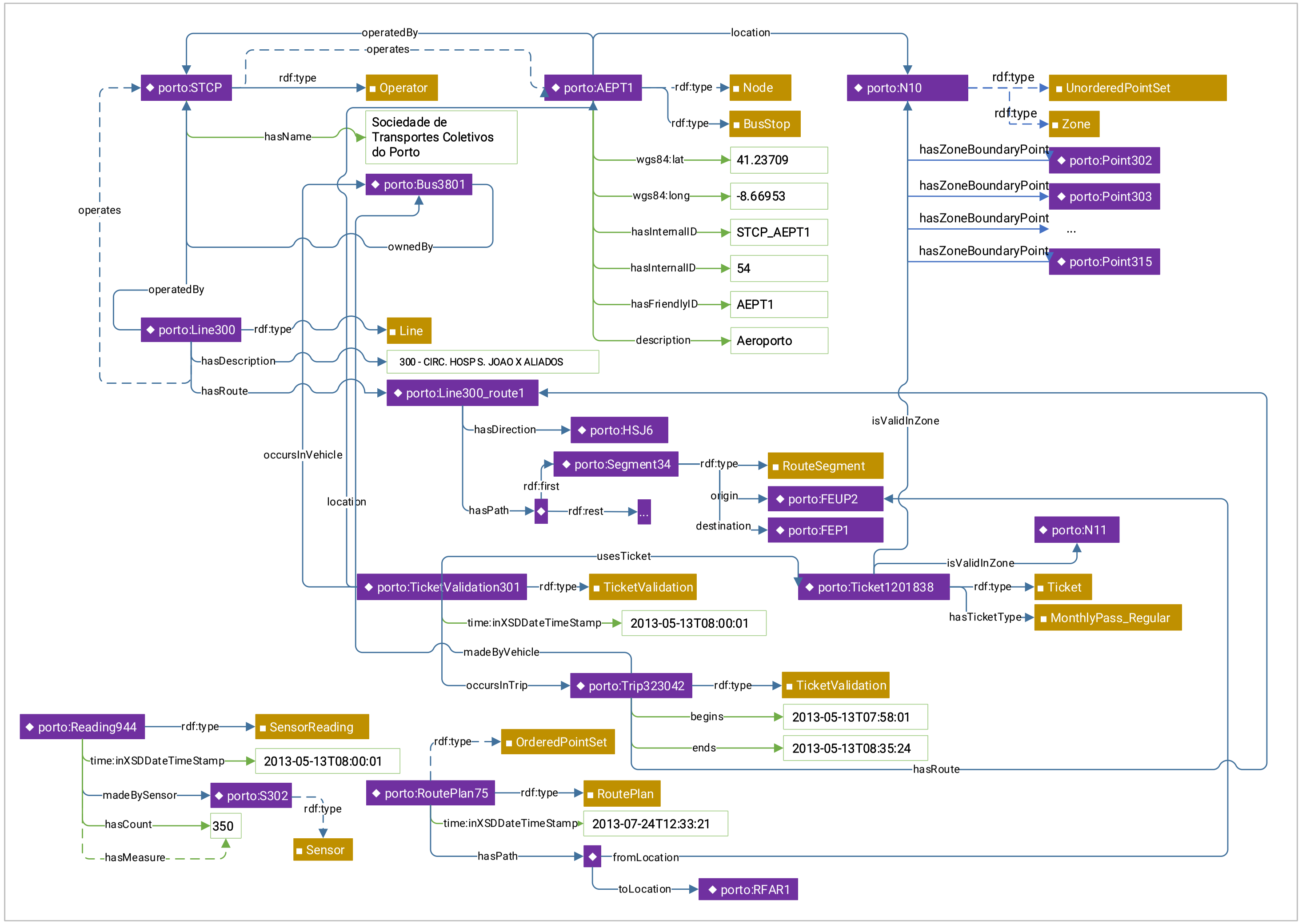


Figure 5. Excerpt of data graph of urban mobility data from Porto, Portugal to be described in Section 5, which makes use of several constructs of UMC, and other ontologies such as GeoSPARQL, WGS84, and Time.

4.3. DataConcept (DC)

The elements of DC describe structural properties of S-T data and transformations. The subclass `SpatialReference` is refined into two subclasses:

- `Point`, with further subclasses `GenericPoint` and `KnownPoint`;
- `PointSet`, with further subclasses `UnorderedPointSet` and `OrderedPointSet`.

`TemporalReference` contains two subclasses: `Instant` and `Interval`. The `SpatialDistribution` class contains three instances: `Discrete`, `Quasi-continuous` and `Graph`.

With the exception of `Transformation`, the remaining classes defined in DC are not expected to be manipulated directly, as the VUMO rules are responsible for inferring data properties from instance data, i.e. data described in terms of UMC subclasses.

In VUMO, data transformations are modeled as instances of `Transformation`, and their queries can be expressed in SPARQL. The SPIN vocabulary translates a plain-text query in to a graph. Such specification is seamless to the user. Moreover, we leverage this capability to infer implicit knowledge from transformations, i.e. spatial distributions, tags (themes), and compatibility with visualization techniques.

4.4. VisualizationConcept (VC)

VC allows for the annotation of visualization techniques. The `Technique` class is used to create instances that represent the visualization techniques implemented in a visualization tool. `InteractionTask` refers to interactive mechanisms that a technique provides. Some instances are already available in the ontology, e.g. `SemanticZoom` or `Filtering`. `Feature` comprises intrinsic components related to the graphical and data aspects of visualizations. The subclasses of `Feature` already have a number of pre-defined instances as shown below:

- `ReferenceFrame`: `Abstract`, `Geographic`;
- `SpatialDimensionality`: `2D`, `3D`;
- `TemporalArrangement`: `Linear`, `Cyclic`;
- `TemporalRepresentation`: `Static`, `Dynamic`;
- `InputVariable`.

Instances for input variables can have any valid Uniform Resource Identifier (URI). In this article, the names `var1`, `var2` and `var3` were used for clarity. The semantics of the property `hasInputVariable` can automatically infer that such instances belong to the class `InputVariable`, as VUMO specifies the `rdfs:range` of this property to `InputVariable`.

The property `hasCompatibleValueType` allows the specification of several datatypes that are accepted by an input variable. The property `isRequired` expects a boolean value. It is used to specify whether an input variable is optional or not. This property is also considered by VUMO when evaluating the compatibility of visualization techniques with data transformations.

4.5. DomainUserConcept (DUC)

DUC allows for the annotation of empirical domain user knowledge. Such knowledge can be used to assess *appropriateness* of visualizations. Users are represented as instances of `DomainUser`, where each user has one or more `DomainUserProfile`. VUMO provides two pre-defined instances of user profiles: `Strategic` or `Operational`.

`TechniqueRating`s are statements made by `DomainUser`s about a `Technique`. A `TechniqueRating` contain one or more statements regarding `RatingComponent`s.

VUMO allows for the annotation of specialized ratings. `CrossRating`s are used to rate a `Technique` with respect to a `Transformation`, according to one or more instances of `RatingComponent`s.

4.6. VUMO rules and functions

Rules and functions extend the capability of the VUMO ontology on regards to inference of new knowledge beyond the intrinsic semantics of OWL and RDF. We developed a set of rules and functions to automatically infer visualization-related properties from instance data, e.g. types of spatial references and spatial distributions, and to infer implicit knowledge from data transformations and visualization techniques. The proposed rules and functions are not meant to be exhaustive, and can be extended to fit the requirements of KVTs.

4.7. Rules

VUMO contains seven rules, labeled from R1 to R7. They are independent in the sense that the execution of a rule during inference is made independently from the others. In general, inference engines execute rules dynamically as new instance data are ingested into a database.

Rules *R1* and *R2* detect spatial references within instance data and infer their type, i.e. points, point sets, and their subtypes. *R3* infers the duration of intervals if their start and finish times are specified. Rules *R4*, *R5* and *R6* infer characteristics of *Transformation* queries based on their structure, namely: spatial distribution (*R4*), themes (*R5*), use of aggregate functions (*R6*). *R7* infers the compatibility between transformations and visualization techniques.

Tables 3 and 4 provides the pseudocode representation of rules R1-3 and R4-8, respectively. Inferred triples are represented with a specific notation. For instance, $s \in \mathit{GenericPoint}$ is equivalent to "s is an instance of *GenericPoint*"; $t\ \mathit{isCompatibleWith}\ v$ denotes a subject-predicate-object triple.

R4 analyzes conditional clauses for predicates containing equivalent subproperties of *hasSpatialReference*. If one or more clauses satisfy that condition, the range of such property is used to retrieve the spatial reference type. The corresponding axiom is then used to retrieve the spatial distribution.

R5 extracts themes, i.e. tags, that describe the urban mobility concepts related to a *Transformation*. The rule finds condition clauses whose properties' ranges are subclasses of *UrbanMobilityConcept*. Themes provide a natural language description of the contents of a *Transformation*.

R6 verifies if a *Transformation* returns aggregate data, i.e. if at least one *OutputVariable* contains an aggregate function. Such verification occurs while evaluating compatibility, as a *Technique* may expect disaggregate instance data to perform aggregations externally.

R7 evaluates the compatibility of a *Transformation* with respect to a *Technique*. Compatibility holds if the aggregate requirements (*R5*) match, and if there exists at least one bijective mapping m such that

Table 3. Pseudocode representation of VUMO rules related to data integration

Rule	Pseudocode

R1	// R1 infers *Point*s and their subtypes
	$s \leftarrow$ instance of *rdfs:Resource* // receives an instance of any class
	if *containsLatitudeLongitude(s)* **then**
	  if *containsIdentification(s)* **then**
	    $s \in$ *KnownPoint* // infers s as an instance of *KnownPoint*
	  else
	    $s \in$ *GenericPoint* // infers s as an instance of *GenericPoint*

R2	// R2 infers *PointSet*s and their subtypes
	$s \leftarrow$ instance of *rdfs:Resource* // receives an instance of any class
	$P \leftarrow \bigcup_s p$ // Points referred by s, if any
	if $\|P\| \geq 2$ **then** // P should have at least two Points
	  if *isOrdered(P)* **then**
	    $P \in$ *OrderedPointSet* // infers P is an *OrderedPointSet*
	  else
	    $P \in$ *UnorderedPointSet* // infers P is an *UnorderedPointSet*

R3	// R3 infers the duration of intervals, when applicable $e \leftarrow$ instance of *Event* $p_i, p_f \leftarrow$ // ordered temporal reference properties (initial and final) **if** $(e \quad p_i \quad t_i) \wedge (e \quad p_f \quad t_f)$ **then** // if triples exist for start and finish times e *hasDuration* $(t_f - t_i)$ // inferred triple

$$
\begin{aligned}
m\colon O' \subseteq O &\to I \\
o_j &\longmapsto i_k
\end{aligned}
$$

and $\theta(o_j) = \theta(i_k)\ \forall(o_j, i_k)$, where $o_j \in O'$ and $i_k \in I$ are the output and input variables, respectively.

O is the set of all output variables returned by a data transformation. I is the set of all input variables of a visualization technique. O' is a subset of O. The function θ represents an operator that returns the type of an output or input variable, e.g. string, integer, resource.

4.8. Functions

Besides compatibility of evaluation of visualization techniques and data transformations, the evaluation of appropriateness is specific to the implementation of recommendation methods of each KVT. VUMO provides embedded functions to assist methods on retrieving asserted empirical knowledge:

- `getTechniqueRating(t)`: returns all ratings given to a visualization technique `t`;
- `getCrossRating(t,v)`: returns all cross ratings assigned to the a transformation `t` and a visualization technique `v`;

Table 4. Pseudocode representation of VUMO rules related to data transformations

Rule	Pseudocode

R4	// R4 infers *SpatialDistributions* of a *Transformation*
	$t \leftarrow$ instance, $q_t \leftarrow$ query within t, such that $t \in$ *Transformation*
	$C \leftarrow \bigcup_{q_t} c$ // condition clauses of q_t
	for each $c \in C$ **do**
	  $p_c \leftarrow$ *property(c)* // receives the property (predicate) of c
	  if $p_c \equiv$ *hasSpatialReference* **then**
	    $r_{p_c} \leftarrow$ *range*(p_c) // receives the range of property p_c
	    $\sigma_{r_{p_c}} \leftarrow$ *getSpatialDistribution*(r_{p_c})
	    t *hasSpatialDistribution* $\sigma_{r_{p_c}}$ // inferred triple

R5	// R5 infers themes (tags) of a *Transformation*
	$t \leftarrow$ instance, $q_t \leftarrow$ query within t, such that $t \in$ *Transformation*
	$C \leftarrow \bigcup_{q_t} c$ // condition clauses in q_t
	for each $c \in C$ **do**
	  $p_c \leftarrow$ *property(c)* // receives the property (predicate) of c
	  if *range*$(p_c) \equiv$ *UrbanMobilityConcept* **then**
	    $r_{p_c} \leftarrow$ *range*(p_c) // receives the range of property p_c
	    t *hasTheme* r_{p_c} // inferred triple

R6	// R6 infers if the query of a *Transformation* returns aggregate results
	$t \leftarrow$ instance, $q_t \leftarrow$ query within t, such that $t \in$ *Transformation*
	$V \leftarrow \bigcup_{q_t} v$ // set of output variables of q_t
	for each $v \in V$ **do**
	  if *isAggregate(v)* **then**
	    t *returnsAggregateResults true* // inferred triple
	    break // one occurrence is sufficient
	  else
	    t *returnsAggregateResults false* // inferred triple

R7	// R7 infers compatibility of a *Transformation-Technique* pair
	$t, v \leftarrow$ instances, such that $t \in$ *Transformation*, $v \in$ *Technique*
	$O, I \leftarrow$ output and input variables sets of t, respectively
	$\xi \leftarrow \varnothing$ // result set of compatible mappings
	if *meetsAggregateRequirements*(t, v) **then**
	  $\xi \leftarrow$ *findMapping*(t, v) // stores compatible mappings in ξ
	  if $\lvert\xi\rvert \geq 1$ **then**
	    t *isCompatibleWith* v // inferred triple

- `getExpertInfo(e)`: returns asserted knowledged related to domain expert `e`.
- `getBroaderConcepts(c)`: returns concepts that are broader than `c`, based on the assertions made with the SKOS vocabulary;
- `getNarrowerConcepts(c)`: returns concepts that are narrower than `c`, based on the assertions made with the SKOS vocabulary;

Functions are also stored as SPARQL queries, and take advantage of the SPIN vocabulary to be executed as such.

5. Practical Applications

This section demonstrates VUMO applied to semantic integration of data, annotation of visualizations and expert knowledge. The supporting data is related to the public transportation system of Porto, Portugal, and provide information about (a) smart card ridership for bus and subway services, description of (b) stops and stations, (c) fare zones, (d) ticket types, and (e) usage data from Move-me schedule consultation service. Move-me is a mobile application that provides real-time information about Porto's public transportation system. Two prototypical visualization techniques were developed to support the demonstration. Expert knowledge was collected through exploratory usability tests with local domain experts, who are involved in strategic and operational decisions for the public transportation system. Two prototypical domain experts were defined based on such knowledge.

5.1. Semantic integration of data

Datasets were provided in various formats: CSV (a,b), Excel spreadsheets (c,d), and SQL dumps (e). The mapping of each dataset schema onto the VUMO classes and properties is supported by an *ad hoc* parser which also returns semantic instance data in RDF. The resulting graph was stored in a triple store engine. Figure 5 illustrates an excerpt of the resulting graph containing elements from original source data.

In particular, dataset (b) is described by the schema shown in Fig. 3, where each row corresponds to a stop. Table 5 shows a possible mapping between the source attributes and VUMO properties.

As the RDF graph is independent of source data and their schemes, it is possible to manipulate data from all datasets simultaneously. For instance, *TicketValidation301* and *SchedReq12* have *BusStop54* as a common spatial reference.

It follows from *R1* and *R2* that *AEPT1* and *N10* are *KnownPoint* and *OrderedPointSet*, respectively. *TicketValidation301* and *N10* are inferred as instances of *TicketValidation* and *Zone*, as they are the domain of properties *hasValidationDateTime* and *hasZoneBoundaryPoint*, respectively. Inferences related to domain and range are native to RDF semantics and do not require the specification of inference rules.

5.2. Visualization Knowledge and Data Transformations

Map-based and abstract visualization techniques were developed: geographic (*GeoHeatMap*) and calendar (*CalHeatMap*) heat maps, respectively. The former depicts the density of instance data in geographic space. The latter depicts density in a daily arrangement. Each day is assigned to a color according to the number of occurrences.

Table 5. Mapping between attributes of (b) and VUMO properties

Source attribute	Target property	Property type
id	*hasID*	Datatype
alt_id	*hasID*	Datatype
sign_id	*hasFriendlyID*	Datatype
label	*hasName*	Datatype
lat	*wgs84:lat*	Datatype
lon	*wgs84:lon*	Datatype
operator	*operatedBy*	Object
zoneID	*location*	Object

GeoHeatMap requires the following input variables: (a) an event instance; (b,c) spatial (latitude and longitude) and (c) temporal references (instant). Variable (a) receives the entity to be plotted, (b,c) provide the instance coordinates on a map, and (d) is used for temporal filtering. *CalHeatMap* requires the same variables but (b,c). Semantic zoom is available in both techniques. By construction, *CalHeatMap* require aggregate data. *GeoHeatMap* does not. Fig. 7 shows the semantic annotation of both visualization techniques.

Let *AT1* be an instance of *AnalyticalTask* described by "analysis of *TicketValidations* and *ScheduleRequests*", and *T1* an instance of *Transformation* that implements the following query with five output variables:

SELECT *?ev ?lat ?lon ?time* COUNT(*?node*) AS *?total*
WHERE {
?ev rdf:type (:TicketValidation OR :ScheduleRequest) .
?ev time:inXSDDateTimeStamp ?time .
location ?node .
?nodeID wgs84:lat ?lat .
wgs84:lon ?lon .
FILTER (*?time* $\geq$ *?arg1* && *?time* $\leq$ *?arg2* **) }**
GROUP BY *?node*

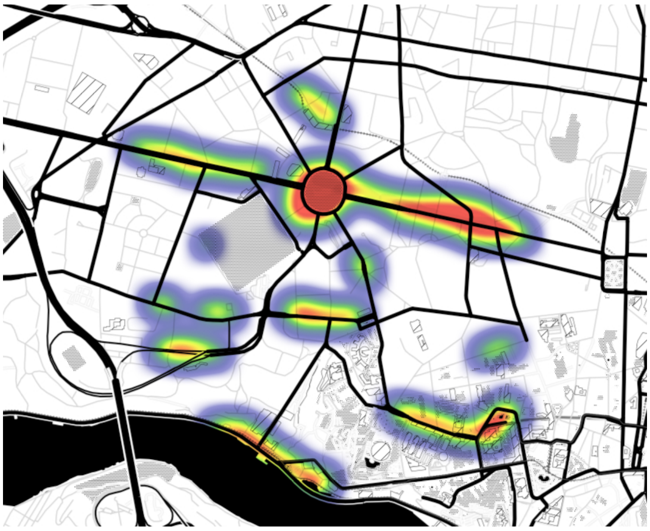
AT1 is linked to *T1* via the property *hasRelatedTransformation*. An *AnalyticalTask* can be linked to one or more *Transformations*. Arguments *arg1* and *arg2* are placeholder values that can be changed by the user.

From R4, it follows that *T1* contains a *Discrete* spatial distribution, as the property *occursAtNode* has *Node* as its range, which is defined as semantically equivalent to a *KnownPoint*. R4 yields the tags *TicketValidation* and *ScheduleRequest*. It follows from R6 that *T1* returns aggregate results due to the COUNT function.

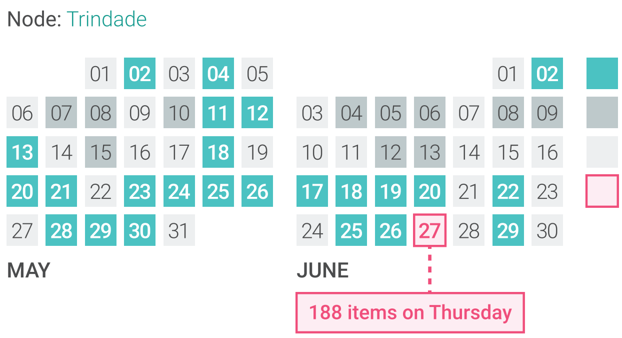
It follows from R7 that *T1* is compatible with *CalHeatMap*, but not *GeoHeatMap*, due to the aggregate results requirement. Consider another instance of *Transformation*, *T2*, that implements the same query as *T1* except for the aggregate function. Hence, *T2* is compatible with *GeoHeatMap*, but not *CalHeatMap*.

5.3. Expert Knowledge and Appropriateness Evaluation

Let *U1* and *U2* be instances of *DomainUser*, which represent two users of a visualization system, with strategic and operational profiles, respectively. *Strategic* and *Operational* are instances of *DomainUserProfile*. Each user can rate a visualization technique with respect to multiple criteria represented by instances of *RatingCompo-*



(a) Geographic heat map



(b) Calendar heat map

Figure 6. Prototypical visualization techniques compliant with VUMO

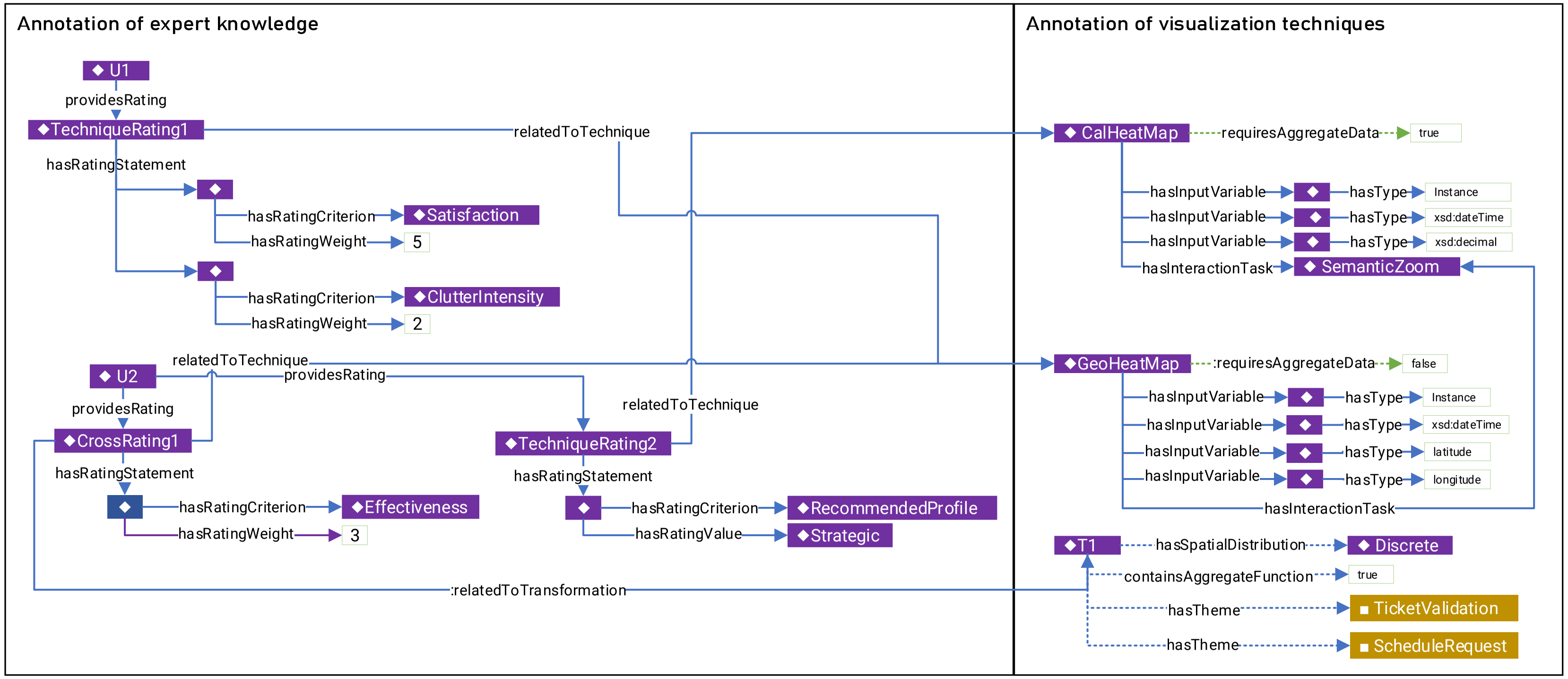


Figure 7. Excerpt from RDF graph containing annotations about experts' knowledge (b) and visualization techniques (c).

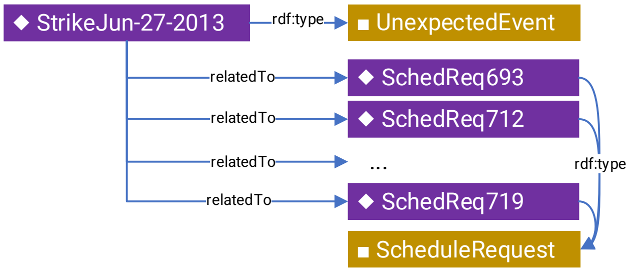


Figure 8. An instance of *UnexpectedEvent* represents a public transportation system strike, which is related to several *ScheduleRequests*

nent. Each criterion can be evaluated according to categorical or numerical values. Ratings are stored as instances of *TechniqueRating*. Users can also provide a specialized rating that relates a visualization technique to a data transformation. Such rating is modeled as an instance of *CrossRating*. Fig. 7 exemplifies a general and specialized rating by *U1* and *U2*, respectively.

U1 stated that the *GeoHeatMap* is appropriate for data transformations that induce a continuous domain. Quantitative ratings were also provided: *Satisfaction* was rated "5" on an arbitrary scale of 0 to 5, and *ClutterIntensity* was rated "2" on the same scale. The latter components are instances of *Positive-* and *NegativeRatingCriteria*, respectively. The type of each component can be used by a recommendation algorithm to define the appropriateness of each visualization technique.

U2 provided a specialized rating to the same technique and transformation *T1*, and rated "3" on regards to the *Effectiveness* criterion. The same user provided a global rating for *GeoHeatMap*, in which it was stated that the technique is appropriate for users with a *Strategic* profile.

5.4. Asserting new knowledge

VUMO can be used to formalize new knowledge derived from the use of visualization techniques. In particular, the *CalHeatmap* technique allowed the identification of an abnormal amount of schedule requests for Trindade, the main subway station in Porto. The number of requests is related to the public transportation strike that occurred on June 27th, 2013. The *CalHeatMap* prototype was used to relate all schedule requrest instances on that date to a new event asserted as *Strike27Jun13*, which is an instance of *UnexpectedEvent*. Fig. 8 shows the resulting graph.

6. Discussion of Results

For semantic integration of mobility data, the mapping of a dataset schema to VUMO classes and properties should be planned in advance. *Ad hoc* parsers are considered an effective approach to convert raw data onto RDF graphs. A dataset may not be entirely described in terms of VUMO components due to the absence of specific classes or properties. To address that limitation, developers can extend or import additional ontologies which declare additional concepts.

The ontology does not require KVTs to be implemented in a specific programming language. To use VUMO, the only technological requirements are the support to RDF data, OWL ontologies and SPIN inference. Visualization techniques implemented in a VRS should be able to digest RDF data and comply to their annotation, especially on regards to the specification of input variables and interaction features.

KVTs may be impaired by cold start, i.e. absence or lack of sufficient domain expert knowledge to perform recommendations during early system use. VUMO addresses this limitation by evaluating compatibility (*R5*), which can guarantee a minimal set of visualizations for experts to start with.

On regards to transformations, VUMO rules are limited to infer characteristics from queries, as the SPIN vocabulary allows them to be stored as a RDF graph. Still, transformations may assume other specialized forms, e.g. data mining algorithms or simulations. VUMO can still be useful to provide semantic annotation, as developers may manually specify such characteristics.

7. Conclusion and Further Work

This article proposed an ontology-based approach to integration and visualization of spatio-temporal urban mobility data from ITS. The core contribution is VUMO, an ontology that provides the semantic foundation for the development of Knowledge-assisted Visualization Systems. Such topic, to the best of our knowledge, has not yet been explored in Transportation studies.

The ontology provides two contributions; it specifies a formal vocabulary for describing spatial events in terms of the components of transportation networks, taking advantage of acknowledged ontologies for representing geospatial (GeoSPARQL, WGS84) and temporal (Time) data. The available constructs. Furthermore, visualization techniques can be described in terms of its features. Expert knowledge can be represented in terms of analytical profile of domain experts, ratings of visualization techniques, and analytical tasks consisting of one or more transformations.

A conceptual model, also introduced by this paper, was implemented into VUMO to derive visual features from such events, according to their semantic description. Such features are exploited by an expandable built-in ruleset to infer, for example, which visualizations are compatible and appropriate to represent a (sub)set of instance data.

A demonstration based on the city of Porto, Portugal, was introduced as a case study. It showed how multiple heterogeneous datasets could be semantically integrated. Prototypical visualization techniques and domain experts were defined to illustrate the implications of their characteristics on recommendation results.

As the ontology converges to its first stable version, an official specification and documentation will be disclosed, which will. Recently, the SPIN rule language has been standardized into SHACL (Shapes Constraint Language) by the W3C; the language is considered to be an evolution of OWL. We plan to translate VUMO into SHACL, so that researchers and practitioners can opt for OWL or SHACL, in accordance to their requirements. Nonetheless, we alert the reader that, by the time of this thesis, few Semantic Web based systems fully support SHACL. We also plan to provide an online shared repository for reuse of empirical knowledge. This would allow researchers and practitioners from various contexts to reuse knowledge made public from other studies or actual practical contexts. This idea is aligned with our orientation towards the involvement of domain experts throughout the visualization process. Moreover, the idea is also aligned with the Semantic Web principle of reusing existing knowledge, to enhance the capability of other systems that also make use of such technologies.

Acknowledgements

We thank Transportes Intermodais do Porto (TIP) and Optimização e Planeamento de Transportes S.A. (OPT) for providing the datasets.

Funding

This work was funded by Fundação para Ciência e Tecnologia (FCT), grant number PD/BD/105910/2014, and partially financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT within project POCI-010145-FEDER-032053.

References

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of Time-Oriented Data* (1st ed.). Springer Publishing Company, Incorporated.

Andrienko, N., & Andrienko, G. (2006). *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach.* Springer Berlin Heidelberg.

Andrienko, N., Andrienko, G., & Rinzivillo, S. (2016). Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. *Information Systems*, *57*, 172–194.

Benvenuti, F., Diamantini, C., Potena, D., & Storti, E. (2017, aug). An ontology-based framework to support performance monitoring in public transport systems. *Transportation Research Part C: Emerging Technologies*, *81*, 188–208. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X17301547`

Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., ... Zhang, J. (2016). Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-tagged Social Media Data. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 270–279.

Cheng, T., Tanaksaranond, G., Brunsdon, C., & Haworth, J. (2013). Exploratory visualisation of congestion evolutions on urban transport networks. *Transportation Research Part C: Emerging Technologies*, *36*(1), 296–306.

Chrisman, N. (1997). *Exploring Geographic Information Systems.* Wiley.

De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., & Rosati, R. (2018). Using Ontologies for Semantic Data Integration. In S. Flesca, S. Greco, E. Masciari, & D. Saccà (Eds.), *A comprehensive guide through the italian database research over the last 25 years* (pp. 187–202). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-61893-7_11`

Huang, X., Zhao, Y., Ma, C., Yang, J., Ye, X., & Zhang, C. (2016). TrajGraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 160–169.

Hughes, R. (2005, jan). Research Agenda for the Application of Visualization to Transportation Systems. *Transportation Research Record: Journal of the Transportation Research Board*, *1937*, 145–151.

Katsumi, M., & Fox, M. (2018, apr). Ontologies for transportation research: A survey. *Transportation Research Part C: Emerging Technologies*, *89*, 53–82. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X18300858#!`

Lorenz, B., Ohlbach, H. J., & Yang, L. (2005). *Ontology of transportation networks* (Tech. Rep.). Reasoning on the Web with Rules and Semantics.

Lu, M., Liang, J., Wang, Z., & Yuan, X. (2016). Exploring OD patterns of interested region based on taxi trajectories. *Journal of Visualization*, *19*(4), 811–821.

Mao, F., Ji, M., & Liu, T. (2016). Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Frontiers of Earth Science*, *10*(2), 205–221.

Mutlu, B., Veas, E., & Trattner, C. (2016, nov). VizRec: Recommending Personalized Visualizations. *ACM Transactions on Interactive Intelligent Systems*, *6*(4), 1:39.

Nunes, N., Ribeiro, M., Prandi, C., & Nisi, V. (2017). Beanstalk: A Community Based Passive Wi-fi Tracking System for Analysing Tourism Dynamics. In *Proceedings of the acm sigchi symposium on engineering interactive computing systems* (pp. 93–98). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/3102113.3102142`

Oliveira, K. M., Bacha, F., Mnasser, H., & Abed, M. (2013). Transportation ontology definition and application for the content personalization of user interfaces. *Expert Systems with Applications*, *40*(8), 3145–3159.

Peuquet, D. (2002). *Representations of Space and Time*. Guilford Press.

Psyllidis, A. (2016). Revisiting Urban Dynamics through Social Urban Data. *Architecture and the Built Environment*, *6*(18), 1–334.

Sobral, T., Galvão, T., & Borges, J. (2017). Semantic integration of urban mobility data for supporting visualization. *Transportation Research Procedia*, *24*, 180–188. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S2352146517303873`

The Gene Ontology Consortium. (2019, jan). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research*, *47*(D1), D330–D338.

Voigt, M., Franke, M., & Meissner, K. (2013). Using expert and empirical knowledge for context-aware recommendation of visualization components. *International Journal on Advances in Life Sciences*, *5*, 27–41.

von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., & Kerren, A. (2016). MobilityGraphs: Visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 11–20.

W3C OWL Working Group. (2012). *OWL 2 Web Ontology Language Document Overview*. Retrieved from `https://www.w3.org/TR/owl2-overview/`

Wang, Z., Lu, M., Yuan, X., Zhang, J., & Wetering, H. V. D. (2013). Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2159–2168.

Wu, J., Fu, Z., Liu, Z., Pan, J., Long, H., Lin, X., ... Tang, J. (2012). City flow: Prototype exploration for visualizing urban traffic conversations. In *Ieee international conference on social computing* (pp. 481–489).

Zeng, W., Fu, C. W., Arisona, S. M., Erath, A., & Qu, H. (2014). Visualizing mobility of public transportation system. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1833–1842.

Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., & Chen, C. (2011). Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1624–1639.