ARTICLE IN PRESS

Web Semantics: Science, Services and Agents on the World Wide Web ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Linking open data and the crowd for real-time passenger information

David Corsar [a,*], Peter Edwards [a], John Nelson [b], Chris Baillie [a], Konstantinos Papangelis [c], Nagendra Velaga [d]

[a] Computing Science, University of Aberdeen, Aberdeen, UK
[b] Centre for Transport Research, University of Aberdeen, Aberdeen, UK
[c] Computer Science, Xi'an Jiaotong-Liverpool University, Xi'an, China
[d] Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India

ARTICLE INFO

Article history:
Received 26 July 2016
Received in revised form 13 January 2017
Accepted 21 February 2017
Available online xxxx

Keywords:
Semantic web
Ontology
Quality
Provenance
Transport
Citizen-sensing

ABSTRACT

The availability of real-time passenger information (RTPI) is a key factor in making public transport both accessible and attractive to users. Unfortunately, rural areas often lack the infrastructure necessary to provide such information, and the cost of deploying and maintaining the required technologies outside of urban areas is seen as prohibitive. In this paper we present the *GetThere* system developed to overcome such issues and to provide public transport users in rural areas with RTPI. An ontological framework for representing mobility information is described, along with the Linked Data approach used to integrate heterogeneous data from multiple sources including government, transport operators, and the public. To mitigate possible issues with the veracity of this data, a quality assessment framework was developed that utilises data provenance. We also discuss our experiences working with Semantic Web technologies in this domain, and present results from both a user trial and a performance evaluation of the system.

1. Introduction

Real-time passenger information (RTPI) systems play a key role in the perceived quality of public transport services, influencing their attractiveness and accessibility [1,2]. By providing passengers with real-time vehicle locations and estimates of arrival and departure times, RTPI systems allow people to plan and make decisions regarding their journeys. To achieve this, such systems must integrate heterogeneous information such as service timetables, details of routes, and GPS-based vehicle locations, all of which may be provided by different agencies [3].

This paper reports on the work of the Informed Rural Passenger[1] project which investigated the suitability of a Semantic Web approach to data integration and use [4,5] for smart mobility applications. RTPI systems are rare in rural areas, due to a lack of supporting infrastructure and the cost of maintaining technologies such as vehicle tracking in rural environments [6]. To overcome these issues we developed a mobility information ecosystem that used Linked Data to integrate open transport data with data received via citizen sensing [7]. The latter refers to use of humans as data providers using web-connected mobile devices [7]. In our work this involved asking bus passengers in rural areas to use the *GetThere* smartphone app to share the location of buses on which they were travelling. The app could also be used to access information about bus services stored in the ecosystem.

Insights into the information requirements of rural public transport users were obtained through a series of interviews and structured focus groups conducted in the Scottish Borders area of the UK [8]. Subsequent analysis of the interview and focus group transcripts, and a review of existing RTPI systems identified the following desirable capabilities for the *GetThere* smartphone app: (C1) list available public transport services; (C2) provide timetable (schedule) information for those services; (C3) provide (real-time) vehicle locations; and (C4) provide individuals with information that is timely, accurate, and personalised to them, particularly during disruptions to the transport network.

Based on these requirements we defined the following specification for the computational infrastructure to support the *GetThere* app: (I1) model public transport services and timetables, the transport infrastructure (e.g. roads, public transport access points), and vehicle locations; (I2) use these models to integrate heterogeneous mobility data from multiple providers, including, where possible, open data sources to reduce data acquisition costs; (I3) record and use data provenance to reason about quality issues arising from the use of data from external providers [9,10]; and (I4) use the integrated data to provide the desired RTPI capabilities.

This paper describes the semantic infrastructure developed to fulfil these requirements and its application in user trials. Semantic

* Corresponding author.
E-mail addresses: dcorsar@abdn.ac.uk (D. Corsar), p.edwards@abdn.ac.uk (P. Edwards), j.d.nelson@abdn.ac.uk (J. Nelson), c.baillie@abdn.ac.uk (C. Baillie), k.papangelis@xjtlu.edu.cn (K. Papangelis), n.r.velaga@iitb.ac.uk (N. Velaga).
[1] http://www.dotrural.ac.uk/irp.

http://dx.doi.org/10.1016/j.websem.2017.02.002
1570-8268/

Web technologies are required to support the integration of heterogeneous mobility data that is annotated with meta-data providing schema and provenance information. Our contributions are: an ontological framework for representing mobility information suitable for multiple transport domains and geographic areas; use of W3C recommended and best practice ontologies in a system available to the general public; an approach for integrating highly dynamic data from citizen sensing with other static data sources; experience of reusing linked open data in a deployed system; a framework capable of assessing the quality of transport data; and an evaluation considering the effects of the information provided by the system on a sample population.

The remainder of this paper is structured as follows: Section 2 discusses related work; Section 3 describes the mobility information ecosystem; Section 4 describes the *GetThere* smartphone app; Section 5 discusses a system performance evaluation; Section 6 discusses a user study; Section 7 reflects on our experiences working with Semantic Web technologies in this domain, while Section 8 summarises conclusions and plans for future work.

2. Related work

Garrigos & Zapater [11] describe a Semantic Web infrastructure for managing real-time traffic information, in which RDF describing vehicle entry and exit times is obtained from sensors on a section of road. This is used to calculate congestion levels and control vehicular access (for example, if the congestion level is high, large vehicles are prohibited). Few details are provided describing the ontologies used and the implementation of the vehicular access control reasoning. While this application of semantic technologies differs from our work, they do identify a key requirement relevant to our work: the importance of providing accurate information to users.

Samper et al. [12] use ontologies to integrate traffic information, and support users with search and visualisation tasks. They define a road traffic ontology, covering road and vehicle classifications, location, geography, events, people, and routes. However, application of the ontology is limited to representing information during a multi-agent based traffic simulation, and it is not publicly available.

Plu & Scharffe [13] describe the publication of the Passim and NEPTUNE datasets as linked data. Passim lists French passenger transport services, and NEPTUNE describes French transport lines (stops, timetables, etc.). The publication process involved defining ontologies to model each dataset, and use of the DataLift platform[2] to convert data from CSV (Passim) and XML (NEPTUNE) to RDF, which is published as Linked Data; however, no uses of the data are described.

The Tiramisu system [14] uses a citizen sensing approach to acquire transport information. Bus passengers in three urban areas of the USA use a smartphone app to continuously provide their location and details such as vehicle occupancy. This information is, in-turn, used to provide others with estimated bus arrival times. However, Tiramisu does not consider issues arising from the quality of contributions from the crowd, increasing the risk of imperfect (incorrect, incomplete, or erroneous) information being provided to users.

3. Mobility information ecosystem

Fig. 1 outlines our mobility information ecosystem. Several ontologies are combined to describe datasets, which are used by web services to support client applications, such as the *GetThere* Android app. Semantic Web technologies are used to support data integration, and distributed data storage and access.

[2] http://datalift.org/.

Table 1
Prefixes and namespaces of reused and defined ontologies.

Prefix	Namespace
trn	http://vocab.org/transit
ldg	http://linkedgeodata.org
irpt	https://w3id.org/abdn/irp/transport
irpu	https://w3id.org/abdn/irp/user
sioc	http://rdfs.org/sioc/spec
foaf	http://xmlns.com/foaf/spec
ssn	http://purl.oclc.org/NET/ssnx/ssn
irps	https://w3id.org/abdn/irp/sensors
geo	http://www.w3.org/2003/01/geo/wgs84_pos
qo	http://purl.org/qual/qual-o
prov	http://www.w3.org/ns/prov#

3.1. Ontological framework

The ontological framework models the data required by the *GetThere* app, including details of public transport services, the transport infrastructure, and vehicle locations. When developing the framework, we followed best practise guidelines and reused existing ontologies where possible [15], using sources such as Linked Open Vocabularies[3] and the Linked Open Data cloud[4]; new ontologies were created only for concepts that were not previously defined. Sub-class relationships between the ontologies were defined by systematically reviewing each concept definition to ensure semantic consistency in the alignment; additional relationships were defined as necessary to link the ontologies. The ontologies are listed in the "Ontologies" layer of Fig. 1; Fig. 2 expands this to outline the main ontological classes and their interrelationships; Table 1 lists the ontology prefixes and namespaces.

Public transport services and timetable information are represented using the transit ontology, which was selected as it was based on the General Transit Feed Specification,[5] a common format for representing this type of information. The ontology describes public transport *trn:Services* that operate on *trn:Routes*. Services are described as a list of *trn:ServiceStops* with associated arrival and departure times. The LinkedGeoData ontology which represents OpenStreetMap[6] data, is used to represent the transport network. Roads are represented as a list of *lgd:Ways*, each with a start and end *lgd:Node* that is associated with a geo-location. The Transport ontology uses this to model the *irpt:BusServiceMap* of roads travelled by buses on each *trn:Service*.

The User ontology extends SIOC and FOAF to describe user profiles and their *irpu:Journeys* on public transport services. The citizen sensing aspects (including vehicle locations) are represented by integrating user details with extensions to the W3C Semantic Sensor Network ontology (SSNO), the de facto ontology standard for sensing applications [16]. SSNO describes *ssn:Sensors*, their capabilities, and *ssn:Observations* where a sensor has produced a value for a property of a *ssn:FeatureOfInterest* (thing being observed).

The Sensors ontology extends SSNO to model users as sensors, mobile devices as sensor platforms with attached sensors, the sensing methods they implement, and the types of observations they produce [17]. Example observation types include vehicle location reported by the phone's GPS sensor (captured using the Geo ontology) and occupancy level as observed by the passenger. This highly dynamic data is integrated with other data by modelling the user's *irpu:Journey* as the feature of interest for such observations. The journey refers to both the user (*foaf:Agent*) and the *trn:Route*

[3] http://lov.okfn.org/dataset/lov/.

[4] http://lod-cloud.net/.

[5] https://developers.google.com/transit/gtfs/.

[6] http://www.openstreetmap.org.

Please cite this article in press as: D. Corsar, et al., Linking open data and the crowd for real-time passenger information, Web Semantics: Science, Services and Agents on the World Wide Web (2017), http://dx.doi.org/10.1016/j.websem.2017.02.002.