

Linked Connections: the story so far

David Chaves-Fraga^a, Julian Rojas^b, Pieter Colpaert^b and Oscar Corcho^a

^a *Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

^b *IDLab, Department of Electronics and Information Systems, Ghent University-imec, Belgium*
E-mails: dchaves@fi.upm.es, julianandres.rojasmelendez@ugent.be, pieter.colpaert@ugent.be,
ocorcho@fi.upm.es

Abstract. Deal with the access and the management of static, but also historical and real-time data, is one of the most relevant problems in several domains which want to take advantage of the benefits of Linked Data. The transport domain is one of these domains but, nowadays, their data are exposing in a way that only ad-hoc systems are able to process them. In previous work, the Linked Connections(LC) framework was introduced as a cost-efficient publishing alternative to the *de-facto* GTFS standard and route planning APIs. At the moment that the route planners take into account real-time and historic data, new issues appear, like for example how we should expose the data on the web, how to improve the query performance or how to ensuring data consistence.

In this paper we present the advances made in the LC framework to deal with real-time and historical data in the transport domain providing a light HTTP interface that allows smart clients to create their own route planners. Our main contribution is a LC server, based on the principles of Linked Data Fragments, that is able to specify the size of each fragment, improving the query evaluation performance and providing a robust access to the data. We also implement the memento protocol in the top of LC and create a new metadata model for improving the discoverability of transport datasets. We evaluate and compare our approach with other possibilities that are also able to deal with real time and historic data. We discover that take into account the size of the fragments has a relevant impact in the performance on the route planning algorithms.

Keywords: Linked Connections, Linked Data Fragments, Transport, Historical Data, Real-time Data, Route Planning

1. Introduction

As th

2. Related Work

On the current state of the Web, huge amount of data are exposed following the principles of Linked Data. We describe the main contributions on this topic, and also the relevance of the ones made in querying data on the web and the current solutions on planning transport routes using semantics.

2.1. Linked Data

One of the most well-known alternatives to publish data on the Web is Linked Data [1]. Linked Data allows to identify in a unique way resources on the Web using identifiers, or HTTP URIs. It is a method to

distribute and scale data over large organizations such as the Web. When looking up this identifier by using the HTTP protocol or using a Web browser, a definition must be returned, including links towards potential other interesting resources, a practice called *dereferencing*. The triple format to be used in combination with URIs is standardized within RDF. The URIs used for these triples already existed in other data sources, and we thus favored using the same identifiers. It is up to a data publisher to make a choice on which data sources can provide the identifiers for a certain type of entities.

There are multiple benefits of exposing the data on the Web as Linked data: i) the data is linked to related data so the information can be combined across different interfaces, ii) the data is queryable using the query language for RDF, SPARQL [2], or other approaches like API-Rest interfaces [3] or [4] iii) the data is represented following a reference model, so the interoper-

ability between different data sources is at the semantic level.

In transport domain

2.2. Querying data on the Web

2.3. Route planners

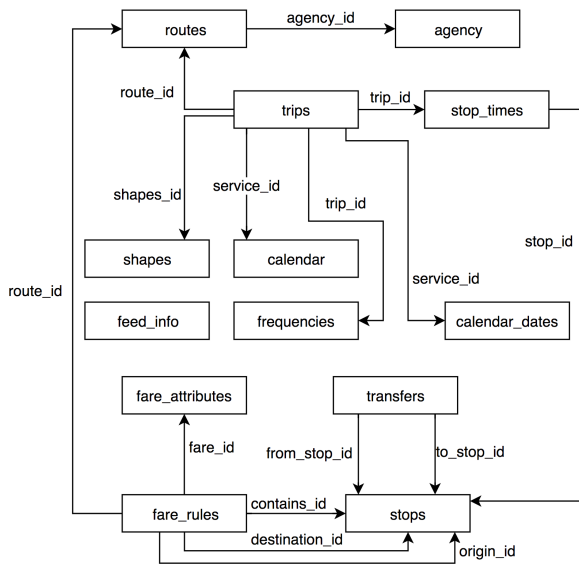


Fig. 1. Underlying GTFS schema

3. Linked Connections

3.1. Linked Connections Server

3.2. The Memento Protocol

3.3. TransportDCAT-AP

4. Evaluation Design

5. Results

6. Discussion

7. Conclusions and Future work

Acknowledgements

References

[1] C. Bizer, T. Heath and T. Berners-Lee, Linked data-the story

so far, *International journal on semantic web and information systems* **5**(3) (2009), 1–22.

[2] E. Prud, A. Seaborne et al., SPARQL query language for RDF (2006).

[3] W.W.W. Consortium et al., JSON-LD 1.0: a JSON-based serialization for linked data (2014).

[4] M. Lanthaler, Creating 3rd generation web APIs with hydra, in: *Proceedings of the 22nd International Conference on World Wide Web*, ACM, 2013, pp. 35–38.

[5] P. Colpaert, A. Llaves, R. Verborgh, O. Corcho, E. Mannens and R. Van de Walle, Intermodal public transit routing using Liked Connections, in: *International Semantic Web Conference: Posters and Demos*, 2015, pp. 1–5.

[6] D. Chaves-Fraga, J. Rojas, P.-J. Vandenberghe, P. Colpaert and O. Corcho, The tripscore Linked Data client: calculating specific summaries over large time series, in: *Proceedings of the Workshop on Decentralizing the Semantic Web (DeSemWeb)*, 2017.

[7] J.A. Rojas Melendez, D. Chaves, P. Colpaert, R. Verborgh and E. Mannens, Providing reliable access to real-time and historic public transport data using linked v-connections, in: *ISWC2017, the 16e International Semantic Web Conference*, Vol. 1931, 2017, pp. 1–4.

[8] H. Van de Sompel, R. Sanderson, M.L. Nelson, L.L. Balakireva, H. Shankar and S. Ainsworth, An HTTP-based versioning mechanism for linked data, *arXiv preprint arXiv:1003.3661* (2010).

[9] R. Verborgh, O. Hartig, B. De Meester, G. Haesendonck, L. De Vocht, M. Vander Sande, R. Cyganiak, P. Colpaert, E. Mannens and R. Van de Walle, Querying datasets on the web with high availability, in: *International Semantic Web Conference*, Springer, 2014, pp. 180–196.

[10] R. Verborgh, M. Vander Sande, P. Colpaert, S. Coppens, E. Mannens and R. Van de Walle, Web-Scale Querying through Linked Data Fragments., in: *LDOW*, Citeseer, 2014.