Decentralized Route Planning across the Web of Data

Abstract. Wheelchair users looking for accessible public transport routes, tourists discovering attractive routes to go around a new city, or bicycle users trying to avoid highly polluted routes are some examples where highly individualized route planning is needed. Current route planning applications lack query flexibility. The types of queries supported by a route planner are only determined at design time and heavily depend on centralized pre-selected data sources. Integrating a new data source such as another transport mode, a different road network or wheelchair accessibility is not straightforward as it generally requires human intervention to extend the subjacent data model and route planning algorithm implementation. I investigate how relevant data sources available on the Web can be dynamically reused for answering custom queries, and thus allow creating more flexible and personalized route planning applications. Semantic Web and Linked Data technologies provide a common framework for data integration. Yet it is still unclear how relevant data can be automatically reused while remaining independent from specific route planning algorithm implementations. Preliminary work (i) tests the feasibility of solving route planning queries over live and static public transport data sources on the Web, (ii) explores the trade-offs of different Web APIs for publishing and consuming live data streams on the Web and (iii) introduces a Linked Data based approach for publishing road networks data.

1. Relevance

Computer-based route planning is nowadays used by millions of people to obtain journey directions [1]. Public transport providers around the world offer route planning services to keep their users better informed. Public authorities also recognize route planning as a mechanism to improve the mobility conditions of their cities and regions, by enabling citizens and visitors to move around more efficiently [2, 3, 4], as it occurs in the case of Portland[1], London[2], Antwerp[3], among others. In recent years, multimodal route planning has received special attention [5, 6, 7]. Applications like Google Maps[4] or Citymapper[5], allow users to plan their journeys using multiple transport modes, e.g. bus, metro, bicycle or walking. They focus on incorporating mobility related data about different transport modes into their applications to offer more diverse route alternatives.

Meanwhile, millions of datasets have been published on the Web [8]. Following Open Data and Smart City initiatives, national and regional governments, private companies and others, publish heterogeneous data from multiple domains. Many of these datasets contain information relevant for route planning purposes, e.g. parking alternatives, road works or street occupancy; which opens the door for creating richer applications capable of addressing more diverse or specific needs. Considering accessibility constraints is a very typical scenario for route planning. Enabling users to specify for example, if they are on a wheelchair [9] or if they are visually impaired [10], are features that are getting included in existing route planners.

Different types of data available on the Web provide virtually unlimited possibilities for extending route planners: Environmental data (*calculate a route that avoids biking through highly polluted areas*), meteorological data (*calculate a route that limits walking outside under bad weather conditions*), security-related data (*calculate a route that avoid areas with high criminality indexes*) or even general knowledge bases such as DBPedia [11] or Wikidata [12] (*calculate a route that pass by the historic monuments of a city*) are some examples of data that could be used to create enhanced route planners.

2. Problem Statement

Currently available route planning applications are built following a *centralized data strategy*. A route planner[6] calculating routes using both the bus and metro public transport services in a certain city:

1. Collects datasets of each transport mode, containing the minimum necessary information for route planning, i.e. station locations and vehicle schedules.
2. Integrates the data following a predefined (usually graph-based) data model in a centralized data store.
3. Calculates available routes using a route planning algorithm tailored to run over the predefined data model.

The route planner service offered by Google Maps for example, follows this approach. By encouraging public transport operators around the world to publish their data using the GTFS [13] format, Google collects these datasets and integrates them into their route planning application.

However, centralized approaches incur in high costs in terms of computational infrastructure. Hosting, integrating and keeping up to date the data of every transport alternative that wants to be supported, and effectively processing route planning queries from a scaling number of users, requires big data storages and powerful processors, specially if aiming on providing a global route planning solution. Furthermore, they also limit the types of queries supported due to the difficulties that are faced when dealing with heterogeneous data. As long as data are homogeneous in terms of structure and format, i.e. dealing only with public transport data using a common format as GTFS, integrating new data sources into a route planning application can be easily automated and the same route planning algorithm implementation will remain useful without requiring many adaptations. Other types of data make necessary to manually adapt the data model and the route planning algorithm that runs on top of it. This could explain why route planning applications such as Google Maps or Open Trip Planner do not allow querying for routes that consider, for example air quality, even though this type of data can be found published on the Web as Open Data[7].

Problem Statement - *Integrating new heterogeneous data sources to extend route planners capabilities requires adapting the subjacent data models and route planning algorithm implementations, resulting in ad-hoc and hence not reusable solutions.*

3. Related Work

Including new data sources to allow more diverse and specific queries in route planning applications can be generalized as a data integration problem. The Semantic Web [14] and Linked Data [15] technologies provide a common environment where data is given a well-defined meaning, better allowing machines to comprehend and work with heterogeneous data from different sources. Moreover, the Web offers a common platform for data sharing where different data publishing strategies may facilitate and influence a decentralized approach for route planning applications.

In this section I present an overview of different aspects related to the problem of dynamically integrating new data sources to enable more diverse and specific queries in route planning applications, such as data publishing and route planning algorithms.

3.1. Data Publishing on the Web

Millions of datasets are published on the Web [8]. Following the right strategy when publishing data is key to increase the interoperability and reusability of data, and foster the creation of new and innovative services. In this direction, Tim Berners-Lee introduced a set of principles[8] for publishing data as Linked Data on the Web. Extending this principles, the W3C published a document of best practices for publishing Linked Data [16], which provides a set of guidelines and design principles aimed at data publishers.

Traditionally, data is often published on the Web either as a data dump or through an API. In the public transport domain, GTFS [13], which is regarded as the de-facto standard for describing and exchanging transit schedules, is a common example of data dump publishing. The European Committee for Standardization released the NeTEx standard [17] as a general purpose format for exchanging public transport schedules and related data. NeTEx is the standard selected by the European Union, under the Directive 2010/40/EU[9], for the provision of an EU-wide multimodal travel information service, where every member state will publish their public transport-related datasets through a National Access Point.

Transport related data is also offered on the Web as route planning APIs. Navitia.io[10], Plannerstack[11], CityMapper or the open source Open Trip Planner are some examples. Geospatial data is also fundamental for route planning applications. Different initiatives exist for publishing geospatial data on the Web. For instance, OpenStreetMap [18](OSM) stands as a community-driven rich source of freely available spatial data. LinkedGeoData [19] publishes a spatial knowledge base derived from OSM that uses RDF as its data model. Also GeoNames [20] publishes over 25 million geographical names using stable URIs and a semantically annotated vocabulary[12].

On the one hand, data dumps offer full query flexibility over the data but are expensive for clients who need to deal with integration of complete datasets. On the other hand, APIs offer simpler access to data, but are expensive for data publishers and limit query flexibility for clients. The Linked Data Fragments (LDF) framework [21] explores the different trade-offs in terms of query flexibility and computational costs of *in between* Web interfaces determined by *fragmenting* datasets and publishing them on the Web. Based on the LDF concept, the Linked Connections specification [22] was introduced as a light-weight data interface for publishing public transit schedules that allows to perform route planning on the client-side. It uses the Linked Connections Ontology[13] and the Linked GTFS vocabulary[14] to describe vehicle departures in public transport networks [23].

Data summarization techniques are also used when publishing data on the Web to facilitate its access and speed up querying processes. Taelman et al. introduced Multidimensional Interfaces [24] as mechanism to semantically describe ordinal ranges within datasets, e.g. time-based or geospatial-based ranges, which are inherent to the nature of commonly used route planning related datasets, such as public transport network topologies (geospatial range) and their schedules (time range). Graph-based summarization techniques such as Contraction Hierarchies [25], K-means [26] or connectivity-based clustering [27] are also commonly used in the route planning domain to reduce the amount of data that needs to be processed when solving route planning queries.

Hypermedia plays also an important role when publishing data on the Web. Providing declarative descriptions about how data can be consumed, helps clients to understand the means of accessing data sources and increases their interoperability. The Hydra vocabulary[15] aims on providing a common terminology for describing hypermedia-driven Web APIs to create generic API clients. Hypermedia is also used to semantically describe responses of Web APIs in Tealman et al. [28]. A shape-based approach using SHACL [29], is used to define a parameterized structure of API responses.

3.2. Route Planning

Route Planning has been extensively studied throughout the years. Multiple algorithms were proposed to calculate routes over different types of networks (e.g. road, public transport, multimodal, etc), and using different criteria (e.g. shortest path, travel time, number of transfers, etc). Bast et al. [30] and Pajor [31] present a comparative analysis of multiple route planning algorithms for road, public transit and multimodal networks. Most algorithms use subjacent graph-based data models to represent the different transportation networks over which they operate, and are defined as extensions of Dijkstra’s algorithm [32]. In recent years, different approaches that disregard Dijkstra-like graph algorithms, have been introduced for public transport and multimodal networks. RAPTOR [33], CSA [34], Transfer Patters [35], Trip-based routing [36] are among the approaches that exploit the basic elements of public transport networks to calculate routes directly on the timetables.

Different solutions that use some of the aforementioned route planning algorithms are available. For instance Open Trip Planner[16] provides an open source implementation of Dijkstra-based and RAPTOR algorithms for planning routes over public transport and road networks. OsmAnd[17] is also an open source route planner for road networks that implements a Dijkstra-based algorithm on top of OSM data.

Being open source allows these solutions to be extended to include new different types of data sources for route planning. However, extending them requires substantial effort to adapt the algorithms and the data models over which they operate.

There is no consensus about the criteria that route planning algorithms should support. The parameters supported by route planning algorithms change from one approach to another depending on the data model, the algorithm implementation and use-case. Kelly et al. [6] presents a list of data requirements for *ideal multimodal route planners*. It identifies 22 different parameters from traditional data sources such as GTFS feeds and OSM. The NeTEx standard defines also a *Trip Plan Query Model*[18] that considers among others, accessibility, via and trip fares parameters. A extensive analysis of formal languages using regular expressions to define constrained shortest path queries is presented in Barret et al. [37]. Furthermore, the Web of Data opens the door for new possibilities and use-cases that need to be supported by route planning algorithms. For instance, the Multi-Criteria RAPTOR algorithm supports route calculation with an arbitrary number of input parameters. However, additional criteria has a significant impact on the algorithm performance, causing it to become too slow for practical use [33].

Given that Linked Data is also described using a graph-based model, the problem of finding routes between nodes have been previously studied. SPARQL 1.1 introduced the Property Paths (PP) syntax that allows query engines to test for path existence between nodes. Savenkov et al. [38] propose an extension for PP syntax to compute the top-k shortest paths of a given path expression between two nodes. An extension of the query semantics of PP is also presented in Hartig et al. [39] to support navigation of PP over the Web of Data instead of over single RDF graphs. A more expressive alternative is later introduced as the Linked Data Query Language (LDQL) [40] that allows expressing query graph patterns and navigation paths independently. Another approach is introduced in Rutgers et al. [41] as an extension of the property graph query language CypherQL[19], which allows for top-k shortest paths queries, calculated path weights and filtering on path edges and nodes. Lastly, multimodal route plans are calculated using a GraphQL interfaces published by the Finish[20] and Norwegian[21] Transit Agencies.

4. Research Questions

Given the related work, the need for a structured and declarative approach to express route planning queries, that allows identifying specific data requirements to resolve them, becomes apparent. A publishing strategy of heterogeneous data on the Web, that allows for decentralized data access and efficiently solving route planning queries using state of the art algorithms is also a need. I thus investigate the following research questions:

Research Question 1 - *How can route planning queries be expressed in a use-case independent and declarative way that allows identifying data requirements for resolving given queries?*

Research Question 2 - *How do different ways of publishing linked open data on the Web affect decentralized route planning query evaluation in terms of performance?*

5. Hypotheses

The first research question is related to the need of having a declarative and use-case independent mechanism to express route planning queries, that allows including arbitrary constraints, while being self-describing about the type of data needed to answer a query. Existing solutions are usually tightly coupled to the subjacent data model and algorithm implementation, limiting the type of queries that can be expressed. Also, implementation independent approaches such as regular expression-based syntaxes do not allow for automated identification of data requirements. Therefore, I define the following Hypothesis related to Research Question 1:

Hypothesis 1 - *A semantic model for expressing route planning queries is at least as expressive as state of the art approaches and allows for complete and automatic identification of query solving data requirements.*

The second research question relates to the problem of how data should be published on the Web to allow for efficient decentralized route planning solutions. Current Linked Data based approaches for publishing data on the Web allow for interoperable data integration. Also, data summarization techniques provide data publishing alternatives for clients to access just the necessary data to solve a certain query, accelerating query evaluation times. Considering this, I define the following Hypothesis related to Research Question 2:

Hypothesis 2 - *A data publishing strategy based on data summarization techniques allows for decentralized route planning query evaluation to be as performant as state of the art centralized approaches such as Open Trip Planner.*

6. Approach

My approach consists of three steps.

1. *Semantic modeling of route planning queries* - A flexible semantic model to describe route planning queries that considers different constraints and parameters for route calculations. By using domain vocabularies in a structured way, automatic data requirements identification for a given query becomes possible.
2. *Hypermedia-driven Linked Data summaries* - Definition of general requirements to create and publish data summaries on the Web, based on spatial and temporal dimensions and including declarative hypermedia controls. By interpreting the hypermedia controls, route planning applications are able to integrate subsets of relevant data sources on the fly.

7. Evaluation Plan

For evaluating my approach I propose the following plan:

1. *Hypothesis 1* - An evaluation that (i) identifies available approaches in the literature to express route planning queries and the different types of queries supported by each approach. Then (ii) a qualitative comparison between the proposed semantic model and the literature to verify that the proposed approach is at least as expressive as the state of the art.
2. *Hypothesis 2* - A performance evaluation in terms of response time and answer completeness will compare different data summarization approaches through an multimodal route planning use-case and compare its results to a centralized solution such as Open Trip Planner.

8. Preliminary Results[22]

Preliminary work related to Hypotheses 2 and 3, has tested the feasibility of performing decentralized route planning over public transport networks including live data. Based on the non Dijkstra-based algorithm CSA [34] and a data publishing strategy extended from the proposal given in [22], clients are able to execute the route planning algorithm on the client-side and use hypermedia controls to discover and integrate the data during query execution.

Other preliminary work, related to Hypothesis 2, evaluates computational cost and query performance for push and pull-based Web interfaces for publishing data streams. Also introduce a modular architecture for publishing data summaries based on the concepts introduced in [24]. Initial results show that push based approaches perform better for high frequency data streams. Also related to Hypothesis 2, a Linked Data based a approach for publishing road networks is introduced. OSM road network data is published in tiles allowing clients to dynamically fetch the correct tiles to calculate a route.

9. Reflections

Route planning applications are often showcased as a success story for the Open Data movement. By publishing public transport schedules as Open Data on the Web, a wave of innovation was triggered, leading to the creation of applications used by millions. Today the Web offers a much more diverse world of data with virtually unlimited possibilities. Bringing the power of the Semantic Web and Linked Data to route planning could open the door for the next innovation wave. Automating data integration in route planners sets the bases for building more flexible and personalized applications, able to consider and learn from the needs of their users to provide more opportune and significant answers.

References

1. Delling, D., Dibbelt, J., Pajor, T.: Fast and Exact Public Transit Routing with Restricted Pareto Sets. In: 2019 Proceedings of the Twenty-First Workshop on Algorithm Engineering and Experiments (ALENEX). pp. 54–65
2. Rode, P., Hoffmann, C., Kandt, J., Graff, A., Smith, D.: Towards new urban mobility: the case of London and Berlin. LSE Cities - The London School of Economics and Political Science, http://eprints.lse.ac.uk/id/eprint/68875 (2015).
3. Berry, R.: London Assembly - Future transport: How is London responding to technological innovation? Mayor of London - Transport Commitee, https://www.london.gov.uk/sites/default/files/future_transport_report_-_final.pdf (2018).
4. Ahlers, D., Akerkar, R., Krogstie, J., Opdahl, A.L., Tessem, B., Zhang, W.: Harnessing Mobility Data in Cities: A Case Study From the Bergen Region. Presented at the (2018).
5. Dibbelt, J., Pajor, T., Wagner, D.: User-Constrained Multimodal Route Planning. J. Exp. Algorithmics. 19, 1–19 (2015).
6. Kelly, T., Miller, R., Datz, A., Hillsman, E., Barbeau, S.: Enabling Cost-Effective Multimodal Trip Planners through Open Transit Data. National Center for Transit Research, https://www.nctr.usf.edu/wp-content/uploads/2011/06/77926.pdf (2011).
7. Kujala, R., Weckström, C., Mladenović, M.N., Saramäki, J.: Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. Computers, Environment and Urban Systems. 67, 41–54 (2018).
8. Noy, N., Burgess, M., Brickley, D.: Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In: 28th Web Conference (WebConf 2019) (2019).
9. Ding, C., Wald, M., Wills, G.: A Survey of Open Accessibility Data. In: Proceedings of the 11th Web for All Conference. pp. 37:1–37:4. ACM, New York, NY, USA (2014).
10. Dornhofer, M., Bischof, W., Krajnc, E.: Comparison of Open Source routing services with OpenStreetMap Data for blind pedestrians. Presented at the (2014).
11. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: The Semantic Web. pp. 722–735. Springer Berlin Heidelberg, Berlin, Heidelberg (2007).
12. Vrandečić, D., Krötzsch, M.: Wikidata: A Free Collaborative Knowledge Base. Communications of the ACM. 57, 78–85 (2014).
13. General Transit Feed Specification. https://developers.google.com/transit/gtfs/ (2006).
14. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American. 284, 34–43 (2001).
15. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. (2011).
16. Hyland, B., Atemezing, G., Villazón-Terrazas, B.: Best Practices for Publishing Linked Data. World Wide Web Consortium, https://www.w3.org/TR/ld-bp/ (2014).
17. Network and Timetable Exchange. http://netex-cen.eu/ (2014).

18. OpenStreetMap. https://www.openstreetmap.org/ (2004).
19. Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In: The Semantic Web - ISWC 2009. pp. 731–746. Springer Berlin Heidelberg (2009).
20. GeoNames. https://www.geonames.org/
21. Verborgh, R., Vander Sande, M., Colpaert, P., Coppens, S., Mannens, E., Van de Walle, R.: Web-Scale Querying through Linked Data Fragments. In: Bizer, C., Heath, T., Auer, S., and Berners-Lee, T. (eds.) Proceedings of the 7th Workshop on Linked Data on the Web (2014).
22. Colpaert, P., Verborgh, R., Mannens, E.: Public Transit Route Planning Through Lightweight Linked Data Interfaces. In: Cabot, J., De Virgilio, R., and Torlone, R. (eds.) Web Engineering. pp. 403–411. Springer International Publishing (2017).
23. Colpaert, P.: Publishing Transport Data for Maximum Reuse. https://phd.pietercolpaert.be (2017).
24. Taelman, R., Colpaert, P., Verborgh, R., Mannens, E.: Multidimensional Interfaces for Selecting Data within Ordinal Ranges. In: Hartig, O., Sequeda, J., and Hogan, A. (eds.) Proceedings of the 7th International Workshop on Consuming Linked Data (2016).
25. Geisberger, R., Sanders, P., Schultes, D., Vetter, C.: Exact Routing in Large Road Networks Using Contraction Hierarchies. Transportation Science. 46, 388–404 (2012).
26. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics). 28, 100–108 (1979).
27. Flinsenberg, I., Van Der Horst, M., Lukkien, J., Verriet, J.: Creating graph partitions for fast optimum route planning. WSEAS. 3, 569–574 (2004).
28. Taelman, R., Verborgh, R.: Declaratively Describing Responses of Hypermedia-Driven Web APIs. In: Proceedings of the 9th International Conference on Knowledge Capture (2017).
29. Knublauch, H., Dimitris, K.: Shapes Constraint Language (SHACL). https://www.w3.org/TR/2017/REC-shacl-20170720/ (2017).
30. Bast, H., Delling, D., Goldberg, A.V., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., Werneck, R.F.: Route Planning in Transportation Networks. CoRR. abs/1504.05140, (2015).
31. Pajor, T.: Algorithm Engineering for Realistic Journey Planning in Transportation Networks. https://d-nb.info/1058165240/34 (2013).
32. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik. 1, 269–271 (1959).
33. Delling, D., Dibbelt, J., Pajor, T.: Fast and Exact Public Transit Routing with Restricted Pareto Sets. In: Proceedings of the Twenty-First Workshop on Algorithm Engineering and Experiments, ALENEX 2019, San Diego, CA, USA, January 7-8, 2019. pp. 54–65 (2019).
34. Dibbelt, J., Pajor, T., Strasser, B., Wagner, D.: Connection Scan Algorithm. J. Exp. Algorithmics. 23, 1.7:1–1.7:56 (2018).

35. Bast, H., Hertel, M., Storandt, S.: Scalable Transfer Patterns. In: 2016 Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX). pp. 15–29
36. Witt, S.: Trip-Based Public Transit Routing Using Condensed Search Trees. In: ATMOS (2016).
37. Barrett, C., Jacob, R., Marathe, M.: Formal-Language-Constrained Path Problems. SIAM Journal on Computing. 30, 809–837 (2000).
38. Savenkov, V., Mehmood, Q., Umbrich, J., Polleres, A.: Counting to K or How SPARQL1.1 Property Paths Can Be Extended to Top-k Path Queries. In: Proceedings of the 13th International Conference on Semantic Systems. pp. 97–103. ACM, New York, NY, USA (2017).
39. Hartig, O., Pirrò, G.: A Context-Based Semantics for SPARQL Property Paths Over the Web. In: The Semantic Web. Latest Advances and New Domains. pp. 71–87. Springer International Publishing, Cham (2015).
40. Hartig, O., Pérez, J.: LDQL: A query language for the Web of Linked Data. Journal of Web Semantics. 41, 9–29 (2016).
41. Rutgers, P., Martella, C., Voulgaris, S., Boncz, P.: Powerful and Efficient Bulk Shortest-path Queries: Cypher Language Extension & Giraph Implementation. In: Proceedings of the Fourth International Workshop on Graph Data Management Experiences and Systems. pp. 6:1–6:7. ACM, New York, NY, USA (2016).

[1] https://trimet.org/#/planner
[2] https://tfl.gov.uk/plan-a-journey/
[3] https://www.slimnaarantwerpen.be/en/home
[4] http://maps.google.com/landing/transit/index.html
[5] https://citymapper.com/
[6] See Open Trip Planner's case, a widely used open source route planning application: https://github.com/opentripplanner/OpenTripPlanner/blob/dev-1.x/docs/Configuration.md
[7] Latest measurements from Europe's air quality monitoring network: https://www.eea.europa.eu/data-and-maps/explore-interactive-maps/up-to-date-air-quality-data
[8] https://www.w3.org/DesignIssues/LinkedData.html
[9] http://data.europa.eu/eli/dir/2010/40/oj
[10] http://navitia.io/
[11] http://www.plannerstack.org/
[12] http://www.geonames.org/ontology/ontology_v3.1.rdf
[13] http://semweb.mmlab.be/ns/linkedconnections#
[14] http://vocab.gtfs.org/gtfs.ttl#
[15] https://www.hydra-cg.com/spec/latest/core/
[16] https://github.com/opentripplanner/OpenTripPlanner
[17] https://github.com/osmandapp/Osmand
[18] http://www.netex-cen.eu/model/conceptual/passenger_info/index.htm
[19] https://neo4j.com/developer/cypher-query-language/
[20] https://digitransit.fi/en/developers/apis/1-routing-api/itinerary-planning/

[21] https://api.entur.io/journey-planner/v2/ide/
[22] To comply with the double-blind review process, references to publications on related preliminary work are not provided.