

SEMANTIC ENRICHMENT OF ROUTING ENGINES USING LINKED DATA: A CASE STUDY USING GRAPHHOPPER

J. A. Jabbar¹, R. Bulbul^{2*}

¹ Department of Space Science, Institute of Space Technology, 44000 Islamabad, Pakistan - junaid.jabar_rsgs5@ist.edu.pk

² Graz University of Technology, Steyrergasse 30/I, 8010 Graz, Austria - bulbul@tugraz.at

Commission IV, WG IV/4

KEY WORDS: Open Source, Vehicle Routing, Linked Data, Semantic Enrichment, Semantic Web, GIS

ABSTRACT:

Traveling is a basic part of our daily life, whenever a person wants to travel e.g. from home to workplace, the essential question that rises is which route to follow. The choice of a route also varies based on traveler's interest e.g. visiting hospital on way back to home or traveling on a greener route. This varied route planning may be easy for any person in his local neighborhood, however in a new neighborhood and increasing number of options e.g. possible restaurant options to visit, a guiding system is required that suggests an optimal route according to traveler's interests i.e. answering semantic queries. Most of the existing routing engines only answer geometric queries e.g. shortest route due to lack of data semantics and adding semantics to a routing graph requires a semantic data source. Geo-semantics can be added through combination of GIS and semantic web. Semantic web is an extension of World Wide Web (WWW) where the content is maintained and structured in a standard way that is understandable by machines; hence providing linked data as a way for semantic enrichment, in this study the semantic enrichment of routing dataset. To use this semantically enriched routing network a routing application needs to be developed that can answer the semantic queries. This research serves as a proof of concept for how linked data can be used for semantic enrichment of routing networks and proposes a prototype routing framework and application designed using open source technologies along with use cases where semantic routing queries are addressed. It also highlights the challenges of this approach and future research perspectives.

1. INTRODUCTION

Traveling is a basic part of our daily life, whenever a person wants to travel e.g. from home to workplace, the essential question that rises is which route to follow. The choice of a route also varies based on traveler's interest e.g. visiting hospital on way back to home i.e. having a hospital as a waypoint. This varied route planning may be easy for any person in his local neighborhood, however in a new neighborhood and increasing number of options e.g. restaurants to visit it can be challenging and time consuming (Gu et al., 2018), a guiding system is required that suggests an optimal route according to traveler's interest. Incorporation of traveler's interest requires a data set which can be used to add semantics to geometric routing graphs and hence answer semantic queries. Adding this information to a spatial dataset leads us to semantic enrichment (Karastoyanova et al., 2007) which would add meaning to routing edges and nodes. GeoSpatial semantics is build upon the combination of GIS, Spatial Databases, Artificial Intelligence, cognitive science and the Semantic Web (Janowicz et al., 2012).

Semantic enrichment can be also done by statically binding the information of Points of Interest (PoI) to respective edges (Fileto et al., 2015) based on proximity, however this results in a static relationship between PoI and routing network, where the reflection of updated relations becomes very difficult. Semantic enrichment can now be done through linked data. (Heath, Bizer, 2011) describe linked data as publishing paradigm in which not only documents, but also data, can be a first class citizen of the Web, thereby enabling the extension of the Web with a global data space based on open standards - the Web of Data.

This study is based on Islamabad and Rawalpindi cities of Pakistan and addresses to the following research objectives:

1. Semantic enrichment of routing network using linked data.
2. Answering semantic queries.
3. Developing a prototype routing application.

In related route planning and semantic enrichment studies (Horvitz, Krumm, 2012) investigated the routing scenarios involving unplanned waypoints along the route. An automated routing assistant has been implemented and tested over a GPS data corpus of over 100 real - world trips. The methods implemented work for identifying ideal waypoints on the way to primary destinations, with minimal expected additional cost of driving.

(Fileto et al., 2015) worked on enrichment of movement data through linked data. Semantic enrichment in this study refers to the associate of locational data of objects with the data having well-defined semantics, that would provide the understanding of what is going on. The researchers have used the PoIs taken from LinkedGeoData in associations with tweets to address a use case that the user who made the tweet also visited that PoI. The dataset used is the movement data of different objects that is semantically enriched to support semantic analysis and also allow semantic mining. The results obtained from mining can be then reused for enhanced semantic enrichment, analysis and mining. Research results conclude that parameters like geographical proximity and textual similarity should be observed and fine-tuned carefully between movement segment and PoI to associate with them.

1.1 Semantic Web

There is huge volume of content available on the World Wide Web (WWW), however according to (Berners-Lee et al., 2001)

*Corresponding author

the large amount of this content is maintained in a way that is only readable to human but not machines, based on the structure; machines may be able to communicate between pages but they cannot understand the actual meaning of the content. Semantic Web is seen by (Berners-Lee et al., 2001) as an extension of WWW so that the content is maintained and structured in a standard way, which is understandable by the machines. Hence allowing the machines to do lot of additional work for humans through processing and understanding the content.

Two of the important technologies that are in place for the development of semantic web include eXtensible Markup Language (XML) and the Resource Description Framework (RDF). XML built upon W3C standards has been addressing the challenges of publishing huge volumes of electronic data on web. It enables the user to format documents in a way that is both human and machine readable. In the context of semantic web, XML yet does not add the meaning, it is just a structure design, the meaning is added by RDF.

1.2 Linked Data

The realization of semantic web is done through linked data. It refers to the data published on the web such that is readable by the machines, is and can be linked to and from external data sets; and has defined meaning (Heath and Bizer, 2011). Linked data does not rely on the hyperlinks to move through web data rather it uses the data formatted in RDF. Datasets linked may be between multiple organizations within same country, between countries or may be between multiple departments of the same organization.

This data web can be accessed using the linked data browser, where the browser uses RDF links between multiple data sources to navigate through the web of data allowing the user to start from a particular dataset and moving through the links to reach a different dataset (Bizer et al., 2008).

The linked data principles were first introduced by Tim Berners-Lee (Bizer et al., 2008) These principles include the following:

1. Use URIs as names for thing.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

2. SEMANTIC ENRICHMENT AND APPLICATION DESIGN

Based on the requirements at hand a solution was required that is totally based on open source technologies. This solution should be able to use the existing routing network and answer semantic queries through a mechanism of semantic enrichment. For the purpose of semantic enrichment linked data has been used and routing scenarios including waypoints along route, close to origin, close to destination for different route weightings along with greener route scenario.

2.1 Architecture Diagram

Figure 1 shows the solution architecture. The proposed architecture is a three tier architecture where each tier is developed using open source technologies. The communication between the client tier and application tier is implemented using java web services.

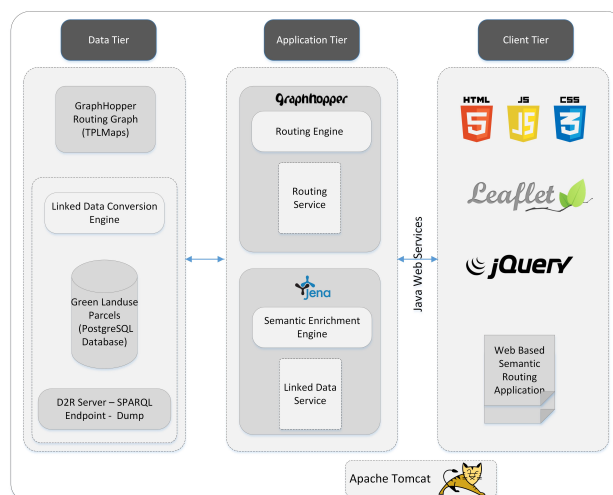


Figure 1. Architecture Diagram

2.2 Datasets

Graphhopper routing dataset: this dataset was provided by TPLMaps. TPLMaps is the largest location-based data provider in Pakistan. The dataset comprises of over 400+ cities mapped, approximately 4.4 million geocoded addresses and over 300,000+ kilometers of road network mapped across the country. Graphhopper uses a road network graph, stored in several different files, to calculate the distances and times.

LinkedGeoData (LGD): is a step towards adding spatial dimension to the Semantic Web. It used the data collected by OpenStreetMap (OSM), converts it to RDF format and serves as linked data. OSM is an example of Volunteered Geographic Information where the individuals have contributed to generate free map information throughout most of the world. LGD links the spatial dimension across datasets in Linking Open Data Initiative. (Stadler et al., 2012) documents in detail the process of publishing OSM data as LGD onto semantic web. This project has three components as given below:

1. Implementation of live-sync module that uses the OSM data extracted in form of OSM database dumps and syncs it with the SPARQL endpoint, so that its up to date
2. A LGD browser that displays LGD using SPARQL endpoints. Its based on javascript widgets
3. A REST API and server that provides access to the OSM relational data as RDF i.e. LGD

Green landuse dataset: the green area polygons were obtained from OSM and further digitized in QGIS using satellite imagery and then exported to the linked data format using technologies discussed in the coming sections. OSM landuse parcels are filtered on the fclass attribute to include the following; "fclass" in ('recreation ground', 'farm', 'forest', 'meadow', 'grass', 'orchard', 'park', 'scrub', 'nature reserve').

3. SOLUTION IMPLEMENTATION

3.1 LGD Categories in Study Area

To implement semantic routing scenarios, the PoI categories from LGD are required. Since there are lot of categories and are distributed across world i.e. data of all categories is not present

within the study area i.e. Islamabad and Rawalpindi. Hence a mechanism needed to implemented that can filter out all the categories, the data for which exist within study area bounding box. The SPARQL query for these results is executed using Jena ARQ is given as follows:

```
String bounding_polygon = "POLYGON((33.4566
72.8147, 33.7926 72.7950, 33.811473.2436,
33.4632 73.2962, 33.4566 72.8147))";
String qs = "Prefix lgdo: <http://linkedgeodata.
org/ontology/>" +
"Prefix geom: <http://geovocab.org/geometry\#>" +
"Prefix ogc:<http://www.opengis.net/ont/geosparql
\#>" +
"Select distinct ?l " +
"From <http://linkedgeodata.org> { ?s " +
"a lgdo:Amenity ; " +
"rdf:type ?l ; " +
"geom:geometry [ ogc:asWKT ?g ] . " +
"Filter(bif:st_intersects(?g , bif:
st_geomFromText("+bounding_polygon+"))}";
```

3.2 GraphHopper Routing

GraphHopper routing was used without any changes of the core routing algorithms, however the function inputs and parameters were modified at the run time based on user interest e.g. route weighting i.e. shortest, fastest, specification of waypoints etc.

3.3 Single Waypoint Routing Scenario

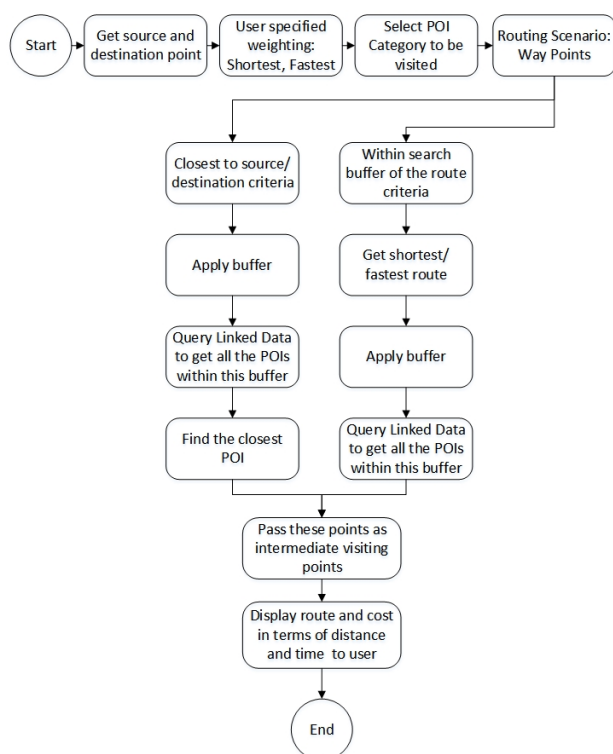


Figure 2. Routing scenario flowchart

In the flowchart given in Figure 2, user specifies the origin and destination point of his journey. To specify these the user can either select the route and destination on the map by clicking or by using the search location option, where the user enters the

name of place to be visited and it is selected as the origin or destination. The next input is to specify the route weighting i.e. the shortest route is required or the fastest is required. Now for the waypoints routing scenario user needs to select whether the places to be visited should be within the search buffer of the origin/destination and closest to them or the should be within the search buffer of the shortest/fastest route. If the user selects the origin/destination criteria, then a buffer of user specified distance is applied on the point and linked is queried to find the available PoIs from the linked data. After fetching these PoIs from the linked data, the are converted to json format for the ease of handling and further processing. Now distance is calculated between the origin/destination and the PoIs returned from the Linked Data, the PoI at the minimum distance from the origin/destination is used as the intermediate point for the journey and the route is displayed to the user with the information of total traveling time and distance.

In case the user selects the within search buffer of the shortest/fastest route option, then first of all the shortest/fastest route is calculated for the origin/destination specified, search buffer is applied on it to get the buffer polygon which is then passed to the linked data query, shortest/fastest routes are calculated for all the PoIs returned by the query and are displayed to the user for selection, display on map and routing.

3.4 Multiple Waypoint Routing Scenario

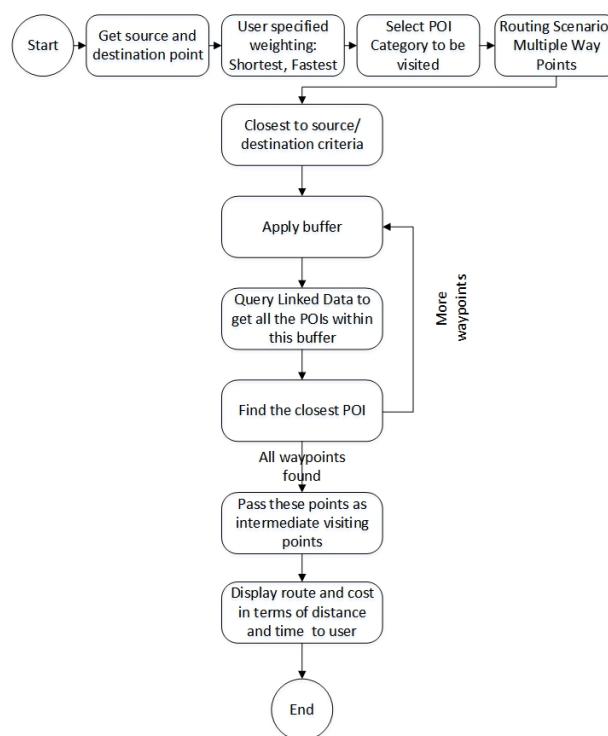


Figure 3. Routing scenario flowchart

In the flowchart given in Figure 3 user specifies the origin and destination point of his journey. The next input is to specify the route weighting i.e. the shortest route is required or the fastest is required. Now for the multiple waypoints routing scenario user needs to select whether the places to be visited should be within the search buffer of the origin/destination and closest to them. The user has to input the desired categories in the same order as required to visit. Now a buffer of user specified distance is applied on the point and linked is queried to find the available

PoIs from the linked data, this continues till all the required category waypoints are identified. The logic implemented for this is that first buffer is applied on start or end based on the user criteria and the next buffer is applied on the linked data PoI fetched as per criteria.

3.5 Conversion of Green Landuse Data to Linked Data

For the implementation of greener route comparison, the green landuse data was managed in PostgreSQL. To convert this data into linked data format, D2R server was used. The command given below allows the generation of mapping file for conversion to linked data.

```
"generate-mapping -o landuse.ttl -u postgres -p
***** jdbc:postgresql: //localhost:5432/
rdf_generation"
```

The resulting mapping file is checked thoroughly and required changes are made. D2R server allows query the data through a local SPARQL Endpoint and allows to create a RDF dump using the following command:

```
"dump-rdf -f RDF/XML-ABBREV -b http://localhost
:2020/ landuse.ttl > landuse.rdf"
```

The flowchart in Figure 4 shows the linked data generation for green landuse data.

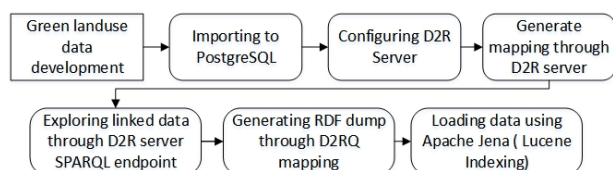


Figure 4. Generating green landuse linked data

3.6 Greener Route Comparison

Once the data is finalized it is loaded in Apache Jena framework, where its is indexed using the Lucene indexing so that it can be queried effectively for the greener route scenario. The flowchart in Figure 5 shows how a greener route is computed.

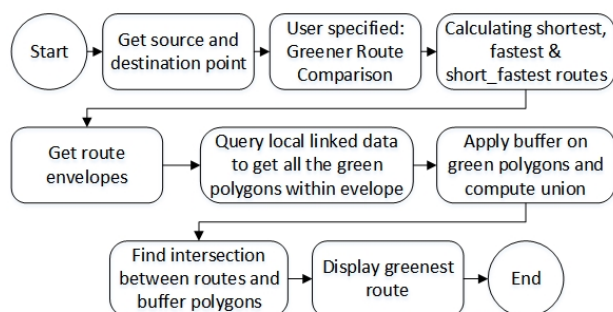


Figure 5. Greener route comparison flowchart

As shown in Figure 5 the user specifies the origin and destination point of his journey and selects the greener route from weighting, now the application computes shortest, fastest and short_fastest routes for the origin and destination and provided.

Once the routes are obtained, route envelopes are calculated due to the Jena Spatial constraint where it does not provide spatial function to query data using a irregular polygon, rather it needs a bounding box. When the route envelope is obtained it is passed to the spatial query that queries all the green parcels within that envelope and returns. Now a buffer is applied on the green polygons to replicate the effect of green areas on surroundings and intersection is computed for the shortest, fastest and short-fastest route. The route summary is shown to user and displayed on the map.

4. RESULTS

The results and discussion of this research are presented in this section. The working of prototype semantic routing application is discussed in detail with figures. Multiple scenarios for semantic routing are created and the results are visualized. The results show how the application is answering semantic queries through enrichment from linked data at run time. Different scenarios presented below show how the system performs multiple routing analysis and suggest the best routes to user based on the specified criteria. The successful working of semantic routing scenarios show the effective usage of linked data and Apache Jena framework for semantic enrichment.

4.1 Application User Interface

The user interface of semantic routing application is given in Figure 6.

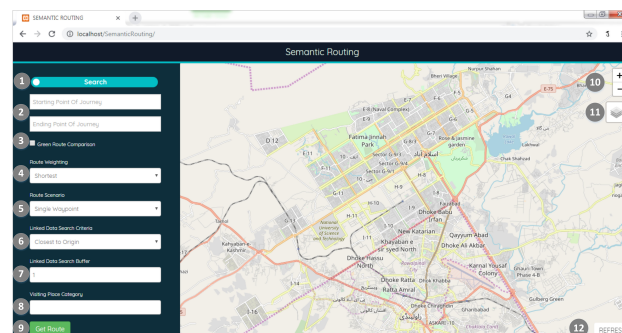


Figure 6. Semantic Routing Application User Interface

1. This is the toggle button for allowing user to switch between 'Search' option to fetch the route origin and destination using TPLMaps search services or toggling it to use the 'DropPin' option to mark origin and destination on the map.
2. Based on the toggle button the components are changed to input fields or button.
3. This is the checkbox button to enable the green route comparison functionality or not.
4. This is a drop down field that shows available routing weightage options . The available options are 'Shortest', 'Fastest', 'Short.Fastest'. This options defines the route calculation cost, whether it should be shortest, fastest or short_fastest.
5. This is a drop down field that shows the available route scenario options. The available options are 'Single Waypoint' and 'Mutiple Waypoints'.
6. This is a drop down field that shows available linked data search criteria options available. The available options

are 'Closest to Origin', 'Closest to Destination', 'Within Route Buffer'. This options specifies the input for feature for buffer to query the linked data. e.g. route start point in case of 'Closest to Origin', route end point in case of 'Closest to Destination' and initial route geometry in case of 'Within Route Buffer'.

7. This is a input field that allows the user to input linked data search buffer in kilometers, so that the buffer polygon can be generated and linked data can be queried.
8. This is a input field that allows user to type and then select desired waypoints category.
9. This is the Get Route button that sends the user specified criteria to routing logic implemented and the response is displayed to user.
10. These are the zoom in and zoom out buttons. Clicking plus sign zooms in one level and clicking minus sign zooms out one level.
11. This is the layer toggle button. It allows the user to toggle between Open Street Basemap and satellite imagery layer.
12. This is the Refresh button, when clicked it clears all the displayed layers and previous criteria values.

4.2 Within Route Buffer Waypoint

First scenario is discussed where the user wants to travel along shortest route between an origin and destination and on his way wants to visit a single landmark i.e. waypoint e.g. a hospital or any other place of interest within the user specified buffer so that there is not much deviation from the original origin-destination route. For this given scenario the semantic routing application would provide the user with a number of possible options within the user specified category alongwith route summary including time and distance. When the user selects anyone of the waypoint from the suggested list, the application would then display the route on map along with selected waypoint. The scenario is summarized as given below:

Origin: 'User Specified'; Destination: 'User Specified'; Route Weighting: 'Shortest'; Route Scenario: 'Single Waypoint'; Linked Data Search Criteria: 'Within Route Buffer'; Linked Data Search Buffer: 'User Specified: 2'; Visiting Place Category: 'Hospital'.

To accomplish this the user will have a specify the origin and destination point, this can be done using 'DropPin' or 'Search' options and specify the criteria given above. This is shown in the Figure 7:

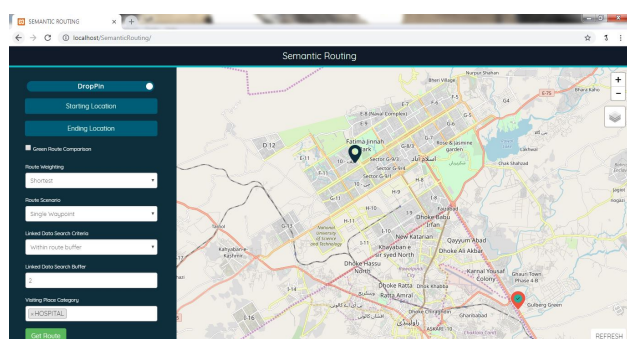


Figure 7. Marking origin and destination using the 'DropPin' option and setting up the routing criteria

Figure 8 shows the possible waypoints to user with route summary, so that the user can choose the waypoint, display on map

and travel accordingly.

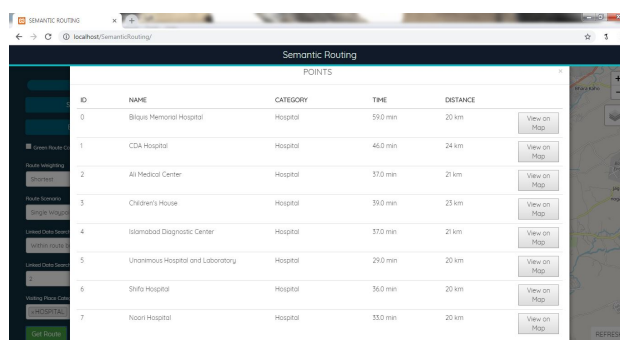


Figure 8. Display of Linked Data PoIs with route summary

When the user decides the waypoint and clicks on 'View On Map'. Figure 9 shows how the route is finally displayed along with the location of Hospital waypoint to the user.

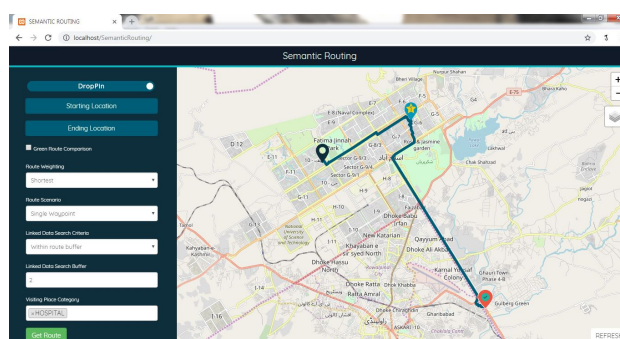


Figure 9. Display of the route for the selected criteria and waypoint

4.3 Closest to Origin Waypoint

Second scenario is discussed where the user wants to travel along fastest route between an origin and destination and on his way wants to visit a closet to origin landmark i.e. waypoint e.g. a place of workshop, or any other place of interest within the user specified buffer. For this given scenario the semantic routing application would provide the user with a single route suggestion that goes through the closest to origin waypoint and displays the route on map.

Origin: 'User Specified'; Destination: 'User Specified'; Route Weighting: 'Fastest'; Route Scenario: 'Single Waypoint'; Linked Data Search Criteria: 'Closest to Origin', Linked Data Search Buffer: 'User Specified: 1'; Visiting Place Category: 'Place of worship'.

The results in Figure 10 show the application automatically suggest 'Lal Masjid' as the closest place of worship and provides fastest route through this waypoint to the destination.

4.4 Closest to Destination Waypoint

Third scenario is discussed where the user wants to travel along short fastest route between an origin and destination and on the way wants to visit a closet to origin landmark i.e. waypoint e.g. a restaurant or any other place of interest within the user specified buffer. For this given scenario the semantic routing application would provide the user with a single route suggestion

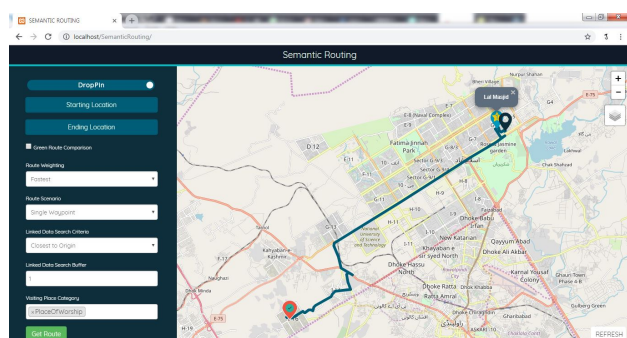


Figure 10. Routing through the waypoint closest to route origin

that goes through the closest to origin waypoint and displays the route on map.

Origin: 'User Specified'; Destination: 'User Specified'; Route Weighting: 'Short_Fastest'; Route Scenario: 'Single Waypoint'; Linked Data Search Criteria: 'Closest to Destination'; Linked Data Search Buffer: 'User Specified: 2'; Visiting Place Category: 'Restaurant'.

The results in Figure 11 show the application automatically suggest 'Mehraaj Hotel' as the closest to destination restaurant and provides short_fastest route through this waypoint to the destination.

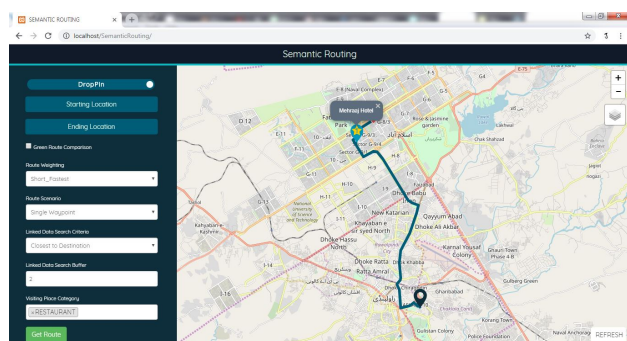


Figure 11. Routing through the waypoint closest to route destination

4.5 Multiple Waypoints

Fourth scenario is discussed where the user wants to travel along fastest route between an origin and destination and on the way wants to visit multiple way points of different categories, with the first landmark closest to origin i.e. multiple waypoint e.g. first visiting a restaurant and then visiting a fuel station or any other place of interest within the user specified buffer.. For this given scenario the semantic routing application would provide the user with a single route suggestion that goes through the multiple waypoints and displays the route on map.

Origin: 'User Specified'; Destination: 'User Specified'; Route Weighting: 'Fastest'; Route Scenario: 'Multiple Waypoint'; Linked Data Search Criteria: 'Closest to Origin', Linked Data Search Buffer: 'User Specified: 2'; Visiting Place Category: 'Restaurant, Fuel Station'.

The results in Figure 12 show the application orders the waypoints as per user suggestion and labels the markers. It is suggesting the user to visit 'Des Perdas' restaurant first as the closest to destination and then go to closest 'Shell' fuel station for refueling the vehicle and move towards destination.

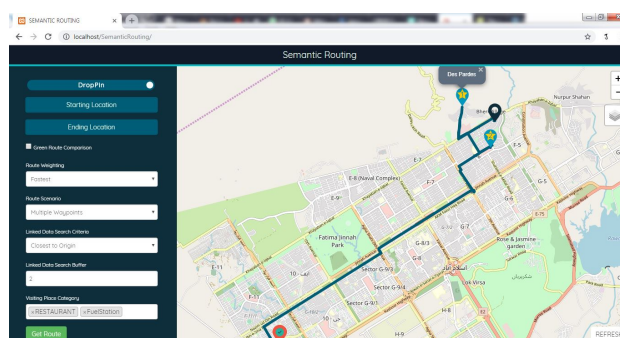


Figure 12. Routing through multiple waypoints, with first being closest to route origin

4.6 Green Route Comparison

Fifth scenario is discussed where the user wants to get the measure of greenery for the three possible route weightings i.e. shortest, fastest and short_fastest between an origin and destination and on the way wants to visit a closet to origin landmark i.e. waypoint e.g. a restaurant or any other place of interest within the user specified buffer. For this given scenario the semantic routing application would provide the user with a three route and displays the route on map. The greenest route is highlighted by green color.

Origin: 'User Specified'; Destination: 'User Specified'; Green Route Comparison: 'Checked' Route Scenario: 'Single Waypoint'; Linked Data Search Criteria: 'Closest to Destination'; Linked Data Search Buffer: 'User Specified: 2'; Visiting Place Category: 'Bank'

The results in the Figure 13 show three routes, shortest, fastest and short_fastest. The application performs semantic enrichment using the linked data green polygons and compute green percentage for each route. The route with most green percentage is then automatically displayed in green color. The popup shows route summary and green percentage of the route.

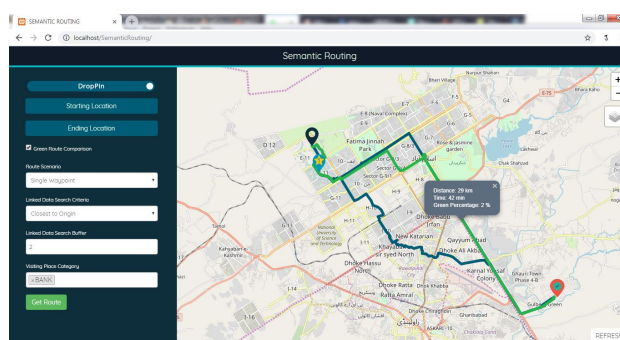


Figure 13. Green route comparison results

5. CONCLUSION

In this research, we successfully presented how linked data can be used for semantic enrichment of vehicle routing networks and have provided multiple scenarios to support the study. The prototype routing application was successfully designed using Apache Jena framework and GraphHopper. User specifies the semantic routing criteria using the application front end from where the parameters are passed to the algorithms through Java

Web Services. Sequentially, series of steps are executed including querying linked data cloud and performing routing analysis. The results are returned by the web services and displayed to the user. Based on the user specified criteria, the implemented solution performs analysis and suggests multiple possibilities or single optimal routing output to the user which are very useful. This research opens more venues in the usage of linked data for semantic enrichment of routing engines. One of the challenges of using linked datasets for Pakistan region is the less coverage/availability, however since it is continuously growing so this gap would be bridged over time.

This research can be further extended through addition of more semantic routing scenario and exploring more linked datasets available in the linked data cloud. It can be further extended to include enriched multi-modal routing scenarios. Another possible area of work is the enrichment of driving directions e.g. instead of informing the user to take right after 100m, the system can additionally show the PoI popup along with a photograph to give driver a better idea and added confidence while traveling in a new area. This work can also be a motivating factor for data industries across Pakistan for publishing their data on semantic web, hence contributing to the increasing linked data cloud.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of TPLMaps (Pvt.) Ltd. for providing the vehicle routing network for this study. It is envisioned as a step towards bridging the gap between industry and academia in Pakistan.

REFERENCES

- Berners-Lee, T., Hendler, J., Lassila, O. et al., 2001. The semantic web. *Scientific american*, 284(5), 28–37.
- Bizer, C., Heath, T., Berners-Lee, T., 2008. Linked data: Principles and state of the art. *World wide web conference*, 1–40.
- Fileto, R., Bogorny, V., May, C., Klein, D., 2015. Semantic enrichment and analysis of movement data: probably it is just starting! *SIGSPATIAL Special*, 7(1), 11–18.
- Gu, Y., Zhou, J., Feng, H., Chen, A., Liu, S., 2018. A recommender for personalized travel itineraries. *International Conference on Cloud Computing and Security*, Springer, 277–288.
- Heath, T., Bizer, C., 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1–136.
- Horvitz, E., Krumm, J., 2012. Some help on the way: Opportunistic routing under uncertainty. *Proceedings of the 2012 ACM conference on Ubiquitous Computing*, ACM, 371–380.
- Janowicz, K., Scheider, S., Pehle, T., Hart, G., 2012. Geospatial semantics and linked spatiotemporal data—Past, present, and future. *Semantic Web*, 3(4), 321–332.
- Karastoyanova, D., Wetzstein, B., Van Lessen, T., Wutke, D., Nitzsche, J., Leymann, F., 2007. Semantic service bus: Architecture and implementation of a next generation middleware. *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, IEEE, 347–354.
- Stadler, C., Lehmann, J., Höffner, K., Auer, S., 2012. Linkedgeodata: A core for a web of spatial open data. *Semantic Web*, 3(4), 333–354.