GPS Trajectory Linked Open Data based on Open POI Information- Through an Experiment in ISWC2016-

Kouji Kozaki[1], Teruaki Yokoyama[2] and Fukami Yoshiaki[3]

[1] The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
kozaki@ei.sanken.osaka-u.ac.jp
[2] Kobe Institute of Computing, 2-2-7 Kano-cho, Chuo-ku, Kobe 650-0001, Japan
ytel@kic.ac.jp
[3] Rikkyo University, 3-34-1 Nishi-Ikebukuro,Toshima-ku, Tokyo, 171-8501, Japan
yoshiaki@rikkyo.ac.jp

Abstract. We tried an experiment in ISWC 2016, Kobe, as a practical example of data integration of sensor data through IoT devices and semantic information. We provided small GPS devices to some volunteers from participants of ISWC 2016 and collected trajectory data during the conference. Then, we integrated the data with Points of Interest (POI) information collected through existing open data such as open government data by Kobe city, DBpedia and Wikidata. This paper presents the method to integrate GPS trajectory data and existing open data with consideration on usefulness of them as sources of POI information for practical analysis.

Keywords: GPS trajectory, open data, IoT, Point of Interest.

1 Introduction

Integration of sensor data through IoT devices and semantic information is an important technology in various fields. We tried an experiment in ISWC 2016, Kobe, as a practical example of such a semantic data integration. In the examination, we provided small GPS devices to some volunteers from participants of ISWC 2016 and collected trajectory data during the conference. Then, we integrated the data with Points of Interest (POI) information collected through some existing open data. As the result, we published the integrated trajectory data as Linked Open Data with SPARQL endpoint. It enables us to analysis trajectory information from various aspects and granularities.

This paper presents the method to integrate GPS trajectory data and existing open data from various aspects with multiple granularities. Then, we consider how much useful are existing open data to analysis real GPS trajectory.

2 Experiments for collecting GPS trajectory data in ISWC2016

There are some methods for collecting trajectory data such as Global Positioning System (GPS) [1], radio frequency identifier (RDFID), location recognitions from image

[2], Monitoring of WiFi for smartphones [3] etc. In our experiment, we used GPS because it does not need any special equipment for target area, and it can measure location more accurately than WiFi monitoring and recognitions from image.

We conducted an experiment for collecting GPS trajectory in International Semantic Web Conference 2016 (ISWC2016), Kobe, Japan, October 17-21, 2016. The purpose of this experiment is to demonstrate of the possibility how we can combine GPS trajectory data and existing open data.

We ask some volunteers to receive and bring a small GPS device during the stay in the ISWC2016 conference. Their movements are recorded on the GPS devices. We used i-gotU GT-600 which are GPS devices offered commercially for the experiment. The devises recorded their location per 1 minute. The collected information includes its position (latitude, longitude), the height above sea level, movement speed and moving distance. As the result, we collected trajectory data from 11 volunteers during the ISWC2016. The average of recoded period was 91.8 hours (3.83 days).

3 GPS trajectory LOD

3. 1 Open data for obtaining POI information

The collected GPS trajectory data is converted into RDF using POI information obtained through existing open data. We collected POI information in Kobe city since moving range of the collected all GPS trajectory data, which is shown in Section 2, was limited in Kobe city. As the result, we collected six datasets. They are classified into two kinds as follows.

1. **Open data published by Kobe city (local government).**
   We collected four datasets related to sightseeing from open data portal[1] by Kobe city because our experiments aim to analyze interests of participants for the international conference. They are dataset about (1) Sightseeing facility, (2) Night view spot, (3) Location site and (4) Outdoor sculpture.
2. **POI information extracted from (5) DBpedia Japanese and (6) Wikidata.**
   We extracted data which has position information (latitude and longitude) in Kobe city using SPARQL query.

We supposed that the former contain POIs closely related to the local area because they are provided by the local government. On the other hands, we supposed that the latter cover wide broad kinds of POIs because they are general open data which everyone can edit on the web. In addition to them, we prepared two kinds of merged datasets; merged dataset of open data by Kobe city (1)-(4) and merged dataset of all datasets (1)-(6). When we merged these datasets, we considered that if a distance between POIs in different datasets is less than 50m, they are treated as one merged POI. It is because different datasets may contain the same POI information.

Table 1 shows POI information which we collected. The merged dataset of (1)–(6) contains 730 POIs. Distances between POI in the dataset are 234.3m in average and 139.1m in median.

[1] https://data.city.kobe.lg.jp/

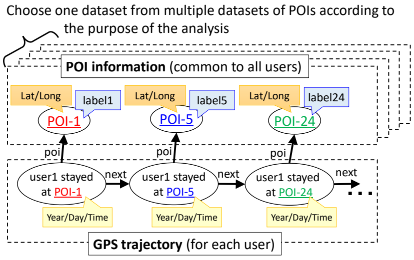

Fig. 1. An overview of RDF data model for POI information and GPS trajectory.

Table 1. POI obtained from open datasets.

Dataset ID	Dataset name	# of POI	Distance between POIs in each dataset (m)			
			Min.	Max.	Ave.	Med.
P1	(1) Sightseeing facility	103	2.4	4,744	546.8	210.3
P2	(2) Night view spot	22	75.7	14,104	1,469.0	531.2
P3	(3) Location site	88	37.9	4,690	504.0	233.7
P4	(4) Outdoor sculpture	35	13.0	157	46.4	35.5
MP1	**Merged dataset of (1)-(4)**	**223**	**2.4**	**4,744**	**332.5**	**125.6**
P5	(5)DBpedia Japanese	178	11.1	2,306	419.4	281.1
P6	(6)Wikidata	544	2.2	3,222	256.3	175.3
MP2	**Merged dataset of (1)-(6)**	**730**	**2.4**	**4,744**	**234.3**	**139.1**

3.2 RDF data model for GPS trajectory

Fig. 1 shows the overview of RDF data model for GPS trajectory. This model consists two data models for POI information and trajectory of each user.

POI information is represented by two classes *POI* class and *MergedPOI* class. *POI* class represents a primitive POI extracted original six datasets, and MergedPOI class represents a merged POI discussed in Section 3.1. Each POI has properties such as *rdfs:label*, position information (*geo:lat*, *geo:long*), data source. Links to other LOD are represented using *rdfs:seeAlso* property. In addition to these properties, a merged POI has *contains* properties which represent primitive POI merged into it. One of these POI datasets is chosen and commonly referred by GPS trajectory data according to the purpose of the analysis.

The GPS trajectory is represented by a series of RDF resources of *StayPOI* class. Each StayPOI resource represents information about a stay that a user stayed a POI during a time interval. It is described using properties such as the *username*, *POI*, *starting/ending time point of the stay*, a link to its *next stay information*. The *POI* is a reference to an instance of POI class discussed the above. Based on the data model, a GPS trajectory is represented as a directed graph which looks like the bottom of Fig. 1.

3.3 Translation from GPS trajectory data to RDF

We translated GPS trajectory data discussed in Section 2 into RDF based on data models shown in the Section 3.2. Position information are compared with each POI information in a selected POI dataset and judged which POI the user stayed. The following steps show how we judge each staying.

1. Each position data in GPS data and each POI information in the selected POI datasets are compared. If the distance between the position in GPS and the POI is less than the threshold ***d***, and the distance is minimum among the position in GPS between each POI, then it is considered that the user stayed the POI.
2. When a series of GPS data records are judged that the user stayed the same POI,
   - The earliest date-time from these records is considered as the time which the user entered (started to stay) in the POI

- The latest date-time from these records is considered as the time which the user leaved (stop to stay) in the POI
3. On the series of GPS data records, when the POI the user stayed changes, the new POI is considered as the new POI that the user stayed in the next.

We developed a Java client program for the translation.

We translated GPS trajectory data collected from the eleven users using eight kinds of POI datasets. We used threshold ***d*** as 100m. In the case that we used the merged dataset (1) – (6), the obtained number of instances of StayPOI class is 1,462 and the number of trajectory data in RDF is 13,158 (generated at May 5, 2017). These dataset is available at http://lodosaka.jp/iswc2016gtl-exp/ with source data, SPARQL endpoint and sample web application to visualize the trajectory data.

4 Discussions and Conclusion

Table 2 shows how many POI obtained existing open datasets and used to transform GPS trajectory into RDF. It shows that 68.58% of trajectory by all users are covered by open data provided Kobe city and 82.88% are covered when Wikidata and DBpedia are used with them. That is, combination of open data by local governments and social open data such as Wikidata and DBpedia is good to obtain practical POI information in this experiment. However, subjects in this experiments did not move so wide area, which is most of them stayed only city are, because we collect their GPS trajectory only during the conference. We are planning to analysis the proposed method in different area and settings.

Table 2. POI obtained from open datasets.

Data set ID	Dataset name	# of POI in dataset	# of POI some user stayed	# of GPS data which can be judged its stay
P1	(1) Sightseeing facility	103	24	1,585 (39.39%)
P2	(2) Night view spot	22	12	71 (1.76%)
P3	(3) Location site	88	36	2,516 (62.52%)
P4	(4) Outdoor sculpture	35	26	201 (5.00%)
MP1	**Merged dataset (1)-(4)**	**223**	**77**	**2,759 (68.56%)**
P5	(5)DBpedia Japanese	178	53	691 (17.17%)
P6	(6)Wikidata	544	97	2,697 (67.02%)
MP2	**Merged dataset (1)-(6)**	**730**	**146**	**3,335 (82.88%)**

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17H01789. The authors are deeply grateful to volunteers who join the experiments in ISWC2016.

References

1. N. Bulusu, J. Heidemann, and D. Estrin. Gps-less low-cost outdoor localization for very small devices. Personal Communications, IEEE, 7(5), (2000).
2. C.-N.E. Anagnostopoulos , I. E. Anagnostopoulos , I. D. Psoroulas , V. Loumos , E. Kayafas, License Plate Recognition From Still Images and Video Sequences: A Survey, IEEE Transactions on Intelligent Transportation Systems, v.9 n.3, p.377-391 (2008).
3. A.B.M.Musa, Jakob Eriksson, Tracking Unmodified Smartphones Using Wi-Fi Monitors, In Proc. of SenSys '12 Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pp. 281-294, (2012) .