

# Knowledge Graph on University Campus Issues

Yuto Tsukagoshi<sup>1</sup>, Takahiro Kawamura<sup>1,2</sup>, and Akihiko Ohsuga<sup>1</sup>

<sup>1</sup> University of Electro-Communications, Tokyo, Japan

<sup>2</sup> Japan Science and Technology Agency, Tokyo, Japan

**Abstract.** This paper aims to build a knowledge graph as Linked Open Data (LOD) on a university campus and to support solving campus issues by completing the missing data in this knowledge graph. We first designed LOD schema for the campus issue. Then, we extracted the data including daily parking status and the related data on campus and built the knowledge graph. We then complemented it by TransE and its derivatives, and finally propose the improvement of them for the campus issue.

**Keywords:** Knowledge graph completion · Campus issues · TransE.

## 1 Introduction

Nowadays there are many urban issues, including littering, scribbling, and illegally parked bicycles. Toward the solution of these issues, it is requested not only for the government but also for corporations and individuals to disclose and share their related statistical and sensory data. The Ministry of Internal Affairs and Communications in Japan promotes a plan to use such “open data”, and the local governments are also promoting to solve social issues by using the data. Then, Linked Open Data (LOD) is recommended to increase the usefulness of the open data.

We thus constructed a knowledge graph as LOD regarding illegally parked bicycles in Tokyo and made it publicly available in cooperation with the Tokyo Metropolitan Bureau in 2017 [1]. However, as more immediate concern a university campus is an epitome of a society and has many social issues. Concerning the bicycle parking, for example, a specific area is overflowing on a specific date and time; thus, there can be a risk in an emergency. In our university, there is a student organization called Student Assistant, which observes daily parking status and records the number of bicycles parked in each parking area. However, there are areas that are still not observed because of the shortage of time and student volunteers.

In this paper, we build a knowledge graph containing the number of bicycles in each parking area with other related data on our campus. We then aim to complement the knowledge graph to include the missing number of bicycles. We first disclose the number of bicycles observed as the open data to make it available for everyone to develop the related services on a campus. We then *linked* other related data on the campus to promote the innovative services in Section 3. Secondly, we try to complement the knowledge graph with knowledge graph completion techniques in related literature in Section 4. Finally, we conclude this paper with the future work in Section 5.

## 2 Related Work

Bordes et al.[2] proposed the method called TransE, which embeds a knowledge graph into a vector space and estimates the similarity of entities and relations based on an energy-based model with the set of triples where the head or tail replaced by a possible entity (but not both at the same time). In this paper, we complement the knowledge graph using the TransE and its derivatives, such as TransH[3], TransR[4] and TransD[5]. We then propose the improvement of them based on the domain knowledge on the campus issue.

## 3 Building of knowledge graph on campus issues

In the knowledge graph, we integrated the data about the number of bicycles, the information obtained from a university website, such as buildings and class schedules, and weather data obtained from the Japan Meteorological Agency after extracting values and descriptions on time and places. Prior to this integration, we retrieved requirements on the possible usage of the data from several students and defined the schema, in which individual issues have interactive relations. However, we used existing ontology alone to make it interoperable with LOD cloud as much as possible. We then stored all the data in an RDF database, whose SPARQL endpoint can be found at <http://www.ohsuga.lab.uec.ac.jp/sparql>, and the Graph IRI is [http://www.ohsuga.lab.uec.ac.jp/campus\\_2017](http://www.ohsuga.lab.uec.ac.jp/campus_2017).

### 3.1 Dataset

We retrieved the following information for the 2017 fiscal year from the data collected by the Student Assistant, the university website, Google Maps, and the Japan Meteorological Agency.

- Time, e.g., fiscal year, semester, month, day, date, time zone, etc.
- Parking areas and the number of bicycles
- Course titles, and classrooms
- Name of rooms in every building
- Event titles and venues
- Temperature and precipitation
- Latitude and longitude of every place

### 3.2 Schema design

We then define the following RDF schema as shown in Fig. 1 to make unique entities mutually connecting via specific relations. Specifically, we used 13 ontologies, 25 properties, and 10 classes, including `aiiso`<sup>3</sup>, `event`<sup>4</sup>, `geo`<sup>5</sup>, `gn`<sup>6</sup>, `ical`<sup>7</sup>,

<sup>3</sup> <http://purl.org/vocab/aiiso/schema#>

<sup>4</sup> <http://purl.org/NET/c4dm/event.owl#>

<sup>5</sup> [http://www.w3.org/2003/01/geo/wgs84\\_pos#](http://www.w3.org/2003/01/geo/wgs84_pos#)

<sup>6</sup> <http://www.geonames.org/ontology#>

<sup>7</sup> <http://www.w3.org/2002/12/cal/icaltzd#>

owl<sup>8</sup>, rdf<sup>9</sup>, rdfs<sup>10</sup>, time<sup>11</sup>, teach<sup>12</sup>, wo<sup>13</sup>, and the IPBLOD in our previous work [1].

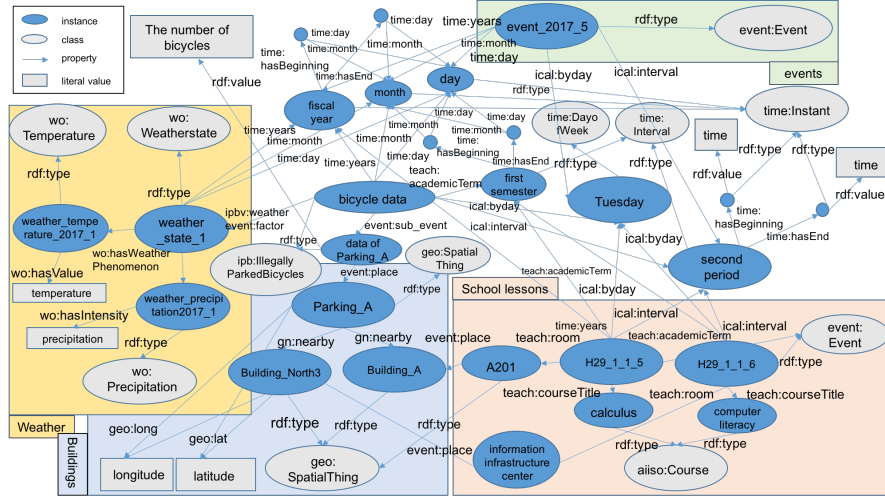


Fig. 1. The schema for campus issues.

We converted the collected dataset to RDF data with the defined schema giving unique entities to URIs. Currently, we have 20,820 triples in the graph. As an RDF database, we used Open Link Virtuoso 7.

#### 4 Completion of knowledge graph

Secondly, we complemented the knowledge graph by embedding it into a vector space using several completion methods, since the knowledge graph includes missing triples such as <bicycle parking, rdf:value, - >. We extracted all the triples from the graph and estimated the missing triples related to the number of bicycles by using TransE[2], TransH[3], TransR[4] and TransD[5]. In the experiment, we trained on 20,448 triples except for 372 missing triples with 1,000 epochs. The hyperparameters were: learning rate = 0.001, hidden layers = 100, minibatch = 100, and margin = 1.0. Table 1 shows the average results of the experiments in five times, where a half of the dataset is randomly selected for the training, and the other half of the dataset is used for the testing. In this

<sup>8</sup> <http://www.w3.org/2002/07/owl#>  
<sup>9</sup> <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
<sup>10</sup> <http://www.w3.org/2000/01/rdf-schema#>  
<sup>11</sup> <http://www.w3.org/2006/time#>  
<sup>12</sup> <http://linkedscience.org/teach/ns#>  
<sup>13</sup> <http://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl>

experiment, a lower MeanRank is better while a higher Hits@n is better. Comparing with the random estimation, we can find that TransE and TransH were superior at least in Hits@3 and Hits@1.

**Table 1.** Results of each method.

	MeanRank	Hits@10	Hits@3	Hits@1
TransE	89.747	0.351	<b>0.284</b>	0.208
TransH	<b>70.479</b>	<b>0.390</b>	0.274	<b>0.219</b>
TransR	289.414	0.205	0.086	0.023
TransD	148.131	0.288	0.208	0.051
Random	1298	0.385	0.116	0.039

However, we are now trying to get more significant results by adjusting the model to this problem domain, since this knowledge graph is not a highly-simplified test data, such as FB15k, but is specialized for a real campus issue. For example, we can add as the domain knowledge the parameter corresponding the numbers of classes and events around bicycle parking areas, because the bicycles parked in the morning will not be moved when the next class is in a short distance.

## 5 Conclusion and Future Work

This paper described a knowledge graph about the numbers of bicycles and other related issues on university campus and complemented by estimating missing triples. As of now, we used existing methods to estimate only the number of bicycles; however, the original model will be required, since the knowledge graph is specialized for campus issues. Besides, we will get more data other than the 2017 fiscal year to improve accuracy of the estimation. Furthermore, through the visualization of parking status and other campus issues obtained from the complemented knowledge graph, we will improve student awareness on their campus through recommendation and visualization of parking status.

**Acknowledgements.** This work was supported by JSPS KAKENHI Grant Numbers JP16K00419, JP16K12411, JP17H04705, JP18H03229, JP18H03340, JP18K19835.

## References

1. S. Egami, T. Kawamura, Y. Sei, A. Ohsuga : Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo, Proc. of ISWC 2016. LNCS, vol.9982, pp.291-307(2016).
2. A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, L. Yakhnenko : Translating Embeddings for Modeling Multi-relational Data, Proc. of NIPS, pp.2787-2795(2013).
3. Z. Wang, J. Zhang, J. Feng, Z. chen : Knowledge graph embedding by translating on hyperplanes, Proc. of AAAI, pp.1112-1119(2014).
4. Y. Lin, J. Zhang, Z. Liu, M. Sun, Y. Liu, X. Zhu : Learning Entity and Relation Embeddings for Knowledge Graph Completion, Proc. of AAAI, pp.2181-2187(2015).
5. J. Guoliang, H. Shizhu, X. Liheng, L. Kang, Z. Jun : Knowledge Graph Embedding via Dynamic Mapping Matrix, Proc. of ACL, pp.687-696(2015).