German Tourism Knowledge Graph

Umutcan Serles[1,2][0000−0001−6459−474X], Elias Kärle[2][0000−0002−2686−3221],
Richard Hunkel[3], and Dieter Fensel[1]

[1] STI Innsbruck, Department of Computer Science, University of Innsbruck
{name.surname}@sti2.at
[2] Onlim GmbH
{name.surname}@onlim.com
[3] German National Tourist Board
richard.hunkel@germany.travel

Abstract. Tourism is one of the most critical sectors of the global economy. Due to its heterogeneous and fragmented nature, it provides one of the most suitable use cases for knowledge graphs. In this poster, we introduce the German Tourism Knowledge Graph that integrates tourism-related data from 16 federal states of Germany and various other sources to provide a curated knowledge source for various applications. It is publicly available through GUIs and an API.

Keywords: tourism, knowledge graphs, data integration

1 Introduction

Tourism is one of the most fragmented sectors, bringing many different services together, such as accommodation, transportation, events, and many more. From a data management perspective, such fragmentation and abundance of service providers bring a large size and heterogeneity in data sources. Most of the data are isolated behind closed systems, preventing digitalization in tourism from reaching its full potential. Using the power of integrated data and benefiting from the synergies between regional touristic offers can help overcome these obstacles.

With that motivation, the German National Tourist Board (GNTB)[1] commissioned the construction of German Tourism Knowledge Graph (GTKG) that integrates data from the regional marketing organizations[2] of 16 states in Germany. It was launched in May 2023 and being updated daily by the regional marketing organizations of German states. To the best of our knowledge, GTKG is the first tourism knowledge graph driven by stakeholders in the tourism industry that is continuously updated and maintained at a large scale.

In the remainder of this paper, we will first briefly describe different tasks for building the GTKG in Section 2. Then, we provide a literature review and comparison with similar knowledge graphs in Section 3. We conclude the paper by summarizing the key points and indications for future work in Section 4.

---

[1] In German, Deutsche Zentrale für Tourismus. https://germany.travel
[2] These are agencies that promote tourism in their regions

2   Building the German Tourism Knowledge Graph

We will present the building process of GTKG in three steps, namely knowledge creation, enrichment and deployment. The necessary tooling is provided by GNTB through its technical provider Onlim GmbH[3]. Onlim provides the technical infrastructure for each step of building the knowledge graph as well as support and training for the data providers.

2.1   Knowledge Creation

The knowledge creation task can be split in two: the creation of the schema (TBox) and the instantiation of the schema (ABox). The schema used in the GTKG is created by the Open Data Tourism Alliance[4]. Open Data Tourism Alliance (ODTA) is an initiative consisting of tourism domain experts representing the DACH region (Germany, Austria and Switzerland). The main goal of the initiative is to create *domain specifications* [4] for tourism-related data. Domain specifications are extended subsets of schema.org. ODTA selects the relevant subsets of types and properties from schema.org and extends them with new types and properties when schema.org is not enough for more detailed description of touristic items. Furthermore, constraints (e.g. cardinality) on the properties can be defined. The domain specifications are encoded as SHACL shapes[5].

The regional marketing organizations of German states create instances of the types in the TBox and import it to GTKG on a daily basis. Each instance is verified against its corresponding SHACL shape (based on target declarations and explicit shape references on instances) to ensure that they satisfy all constraints provided by the domain specifications made by ODTA. The regional marketing organizations typically use RML [2] mappings between the ODTA schemas and the metadata of their internal systems.

The major types[6] represented in GTKG (on 07/03/2024) are schema:Event (239K instances), odta:PointOfInterest (118K instances), odta:Trail (23K instances), schema:LodgingBusiness (25K instances), schema:LocalBusiness (70K instances) and schema:FoodEstablishment (20K instances)[7].

---

[3] https://onlim.com

[4] https://open-data-germany.org/en/open-data-germany/

[5] The list of domain specifications currently used by GTKG can be found here: https://semantify.it/list/LRVOilZZ6. There was a new release of ODTA domain specifications in February 2024 (https://odta.sti2.org/) and the migration process is still ongoing.

[6] See https://semantify.it/list/LRVOilZZ6 for the domain specifications based on these types.

[7] *schema* is the prefix for schema.org namespace. *odta* is the prefix for https://odta.io/voc/ namespace that represents newly created types and properties. Instance counts are result of the query here: https://purl.archive.org/gtkg/queries

2.2   Knowledge Enrichment

The knowledge graph created based on the data provided by 16 regional marketing organizations is further enriched with data from other sources. The enrichment is conceptually a standalone step, and can be initiated outside of the knowledge creation process. However, it is also possible to configure it to run during knowledge creation.

A prominent example of enrichment is geo-linking of e-charging stations[8] and public transportation stops[9] with other points of interest. The enrichment process can be configured via the API by the data providers. The configuration mainly contains the SPARQL queries that select the sets of candidate instances that can be linked and the distance threshold for two POIs to be considered *close enough* to each other. If the distance between two POIs is below the configured threshold, these POIs are linked. The details of the configuration is presented in the API documentation (Section 2.3) under "Guides".

2.3   Knowledge Deployment

The GTKG API[10] enables programmatical access to the GTKG. It allows SPARQL queries against the graph but also many other ways of accessing the knowledge graph (e.g., search by name). The API needs an authentication key, which can be obtained from GNTB free of charge[11]. Listing 1 shows an example request to the API to run the SPARQL query in https://purl.archive.org/gtkg/queries.

```
curl --request POST \
    --url https://proxy.opendatagermany.io/api/ts/v1/kg/sparql \
    --header 'Accept: application/sparql-results+json' \
    --header 'X-API-KEY: fb0758867a9a01fce51546b0e537bb56' \
    --header 'content-type: text/plain' \
    --data 'PREFIX schema: <https://schema.org/>   PREFIX odta:
    ↪  <https://odta.io/voc/> SELECT ?s ?name ?poi (xsd:decimal(?dist)
    ↪  as ?distance) ?unit WHERE  { ?s a odta:PointOfInterest. ?s
    ↪  schema:name ?name. ?s odta:geoLink [odta:linkSource/schema:name
    ↪  ?poi; odta:walkingDistance/schema:value ?dist;
    ↪  odta:walkingDistance/schema:unitCode ?unit] } LIMIT 10'
```

Listing 1: A SPARQL query to return 10 results about POIs and the walking distances to the entities linked via geo-linking to them

Additional to the API, the GTKG provides a graphical user interface. The Search Widget[12] allows text-based search of instances and filtering by type. Different visualization modalities can be accessed for an instance from the search results.

---

[8] https://ladestationen.api.bund.dev/

[9] https://www.opendata-oepnv.de/

[10] See https://purl.archive.org/gtkg/api-doc for a comprehensive documentation

[11] https://purl.archive.org/gtkg/access

[12] Currently being tested on a subset of GTKG https://purl.archive.org/gtkg/search

GTKG allows the regional marketing organizations to use their own instance identifiers, which may not be always dereferencable. To mitigate this drawback, GTKG provides a *a wrapper URI* that takes the instance identifier as query parameter and creates an HTML depiction as well as provides a way of downloading an RDF serialization of the instance. See https://purl.archive.org/gtkg/example for an example.

3   Related Work

The survey from Abu-Salih [1] covers domain-specific knowledge graphs, including tourism-related knowledge graphs. These graphs typically target specific regions and may focus on specific kinds of tourism-related data. For example, Tyrol region in Austria [3], London in England, Madeira in Portugal and Cote d'Azur in France [5].

Although many other knowledge graphs exist in the tourism domain, the German Tourism Knowledge Graph stands out. It is not a purely academic project, but driven and funded by stakeholders in the tourism sector. It contains data about various tourism-related topics and has a national coverage instead of focusing a single region. It also provides various access modalities that can serve different purposes (e.g., machine or human consumption).

4   Conclusion and Future Work

We presented the German Tourism Knowledge Graph. The knowledge graph is a publicly accessible resource and uses a schema.org-based schema to describe the integrated data. It is important to note that the GTKG is built around standard semantic technologies like RDF(S), SPARQL, SHACL and widespread technologies like RML. Each instance in the knowledge graph has its own license, which is typically a Creative Common license like CC-BY-SA.

The German National Tourism Board is currently internally testing a chatbot powered by the GTKG, which will be available for public soon. Many other applications have been developed at a hackathon in November 2023[13]. Currently, individual instances can be downloaded in an RDF serialization. However, periodic releases of the entire knowledge graph are also considered in the future work.

References

1. Abu-Salih,   B.:   Domain-specific   knowledge   graphs: A   survey.   Journal of Network and Computer Applications **185**, 103076 (Jul 2021). https://doi.org/10.1016/j.jnca.2021.103076

---

[13] https://de.linkedin.com/posts/german-national-tourist-board_kollaboration-innovation-dzt-activity-7132391151659016192-k5hW

2. Dimou, A., Sande, M.V., Colpaert, P.: RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In: Proceedings of the Workshop on Linked Data on the Web@WWW 2014. Seoul, South Korea (Apr 2014)
3. Kärle, E., Şimşek, U., Panasiuk, O., Fensel, D.: Building an Ecosystem for the Tyrolean Tourism Knowledge Graph. In: Current Trends in Web Engineering. pp. 260–267. Lecture Notes in Computer Science, Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-03056-8_25
4. Şimşek, U., Angele, K., Kärle, E., Panasiuk, O., Fensel, D.: Domain-Specific Customization of Schema.org Based on SHACL. In: The Proceedings of the 19th International Semantic Web Conference - ISWC 2020. pp. 585–600. Lecture Notes in Computer Science, Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-62466-8_36
5. Troncy, R., Rizzo, G., Jameson, A., et al.: 3cixty: Building comprehensive knowledge bases for city exploration. Journal of Web Semantics **46–47**, 2–13 (Oct 2017). https://doi.org/10.1016/j.websem.2017.07.002