Accessing Tourism data as a Virtual Knowledge Graph

Guohui Xiao, Benjamin Cogrel and Geri Skenderi

Software Developers' Thursday, NOI, Bolzano, 06/02/2020
About us

- Guohui Xiao
  - Chief Scientist and co-founder of Ontopic
  - Assistant Professor at Unibz

- Benjamin Cogrel
  - CTO and co-founder of Ontopic

- Geri Skenderi
  - Freelance
Context
Open Data Hub Südtirol - Alto Adige

Open Data Hub is your access point to South Tyrol's relevant data. You and your service providers can use Open Data Hub for all channels of digital communication. The data is updated on a regular basis and easily accessible.
Open Data Hub (South Tyrol)

- https://opendatahub.bz.it/
- **Tourism** and mobility
- Data about hotels, restaurants, wines, etc.
- **Web API**
  - Accessible to anyone (Open Data)
  - Specific interface (custom)
  - PostgreSQL DB underneath
- Operates at the provincial level
DACH-KG: auf dem Weg zum touristischen Knowledge Graph!


Larger scale: DACH-KG

- Ongoing initiative
- Knowledge Graph (KG) for the German-speaking area in Europe
- Integrating tourism data
  - From many data providers (e.g. local tourism organizations)
  - To the big actors such as Google and Amazon
- Uses schema.org as a foundation
schema.org

URL: https://schema.org

- Designed by Google and others
- Known by everybody in the Search Engine Optimization (SEO) industry
- Graph data model (RDF)
- Set of classes and properties
- Class hierarchy
- Extensible
https://schema.org/Hotel

**Hotel**

A hotel is an establishment that provides lodging paid on a short-term basis (Source: Wikipedia, the free encyclopedia, see http://en.wikipedia.org/wiki/Hotel).

See also the dedicated document on the use of schema.org for marking up hotels and other forms of accommodations.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Properties from LodgingBusiness</strong></td>
<td></td>
</tr>
<tr>
<td>amenityFeature</td>
<td>LocationFeatureSpecification</td>
<td>An amenity feature (e.g. a characteristic or service) of the Accommodation. This generic property does not make a statement about whether the feature is included in an offer for the main accommodation or available at extra costs.</td>
</tr>
<tr>
<td>audience</td>
<td>Audience</td>
<td>An intended audience, i.e. a group for whom something was created. Supersedes serviceAudience.</td>
</tr>
<tr>
<td>availableLanguage</td>
<td>Language or Text</td>
<td>A language someone may use with or at the item, service or place. Please use one of the language codes from the IETF BCP 47 standard. See also inLanguage.</td>
</tr>
<tr>
<td>checkinTime</td>
<td>DateTime or Time</td>
<td>The earliest someone may check into a lodging establishment.</td>
</tr>
<tr>
<td>checkoutTime</td>
<td>DateTime or Time</td>
<td>The latest someone may check out of a lodging establishment.</td>
</tr>
</tbody>
</table>
Virtual Knowledge Graphs
Virtual Knowledge Graphs (VKGs)

- Maps relational DBs into KGs
- Virtual
  - Data stays in the data sources
  - No ETL, no duplicates
  - No RDF graph DB (triplestore)
- Application to ODH
  - Maps the existing PostgreSQL DB...
  - ...into a VKG using schema.org as vocabulary
- Powered by Ontop
Demo: SPARQL endpoint

URL: https://sparql.opendatahub.bz.it

- SPARQL endpoint: Standard Web API based on the SPARQL query language for RDF-based KGs
Mapping

Edit Mapping

Mapping ID: LodgingBusiness

Target (Triples Template):
.data/accommodation/{id} a schema:LodgingBusiness : geo:asWKT "POINT (longitude) (latitude)"
^ geo:wktLiteral ; schema:email < {email} > ;
schema:name {de_name}@de , {t_name}@t , {en_name}@en ;
schema:telephone {de_phone} ; schema:faxNumber {de_fax} .

Source (SQL Query):
SELECT "id" AS id, "Latitude" AS latitude, "Longitude" AS longitude, CONCAT(‘mailto:’, "AccoDetail-de-Email") AS email, "AccoDetail-de-Name" AS de_name, "AccoDetail-en-Name" AS en_name, "AccoDetail-it-Name" AS t_name, "AccoDetail-phone" AS de_phone, "AccoDetail-de-Mobile" AS mobile, "AccoDetail-de-Fax" AS de_fax FROM "v_accommodationsopen"
Query reformulation

- SPARQL queries are reformulated into SQL queries
- A lot of optimizations are applied
- Efficient reasoning
  ➢ Close to zero extra-cost
Ontop

URL: https://ontop-vkg.org

- Open Source VKG engine (Apache 2.0)
- Maintained by
  - In2Data research group at Unibz
  - Ontopic s.r.l.
Applications
Current applications

- Quick visualization of the data
  - Illustrate some data quality issues

- Amazon Alexa skill
  - Voice-based UI
Why using a VKG for the Alexa skill?

❖ Based on schema.org
  ➢ Loosely-coupled to South Tyrol
  ➢ Most queries would work for other regions
❖ Makes prototyping faster
  ➢ New queries can be tested immediately
  ➢ No custom Web API to change
❖ Ongoing experimentation