An Introduction to Knowledge Graphs

https://www.wikidata.org/wiki/Wikidata:Main_Page

Even a simple graph-based vocabulary could be published as a KG.

Image from Challenges of Knowledge Graphs From Strings to Things — An Introduction By Sebastien Dery

Gary Berg-Cross  Ontolog Board Member
Wikidata Entry Example

highest point (P610)
point with highest elevation in a region, on a path, of a race
highest peak | zenith | summit | extreme point highest

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>highest point</td>
<td>point with highest elevation in a region, on a path, of a race</td>
<td>highest peak</td>
</tr>
<tr>
<td>Spanish</td>
<td>punto más alto</td>
<td>punto con mayor elevación en el país, en el camino, de un evento</td>
<td></td>
</tr>
<tr>
<td>Traditional Chinese</td>
<td>最高點</td>
<td>区域內海拔的最高點</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>最高点</td>
<td>区域内海拔的最高点</td>
<td></td>
</tr>
</tbody>
</table>

All entered languages

| Data type | Item
---|---

Statements

instance of

Wikidata property example
Norway highest point Galdhøpiggen 0 references (it may be an item)
2015 Tour de France, Stage 12 highest point Plateau de Beille 0 references
type constraint class geographic location racing fictional location mythical place geographic region exception to constraint Earth Relation instance of
No single definition of a knowledge graph...but..

- Some kind of graph-based, flexible formal data structures as a knowledge representation so you can easily map this KB to other data formats using generic tools, pipelines for updates.....

- In the preface of the 13th International Semantic Web Conference Proceedings (2014), the following statement was published:
  - Significantly, major companies, such as Google, Yahoo, Microsoft, and Facebook, have created their own “knowledge graphs” that power semantic searches and enable smarter processing and delivery of data:
  - The use of these knowledge graphs is now the norm rather than the exception.

- De Facto - now a popular format for representing knowledge, supporting many applications to semantic search engines, question answering systems, and recommender systems (RDF)

### Various Definitions with Different Features

<table>
<thead>
<tr>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Knowledge graphs could be envisaged as a network of all kind things which are relevant to a specific domain or to an organization. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets.”</td>
<td>A. Blumauer. From Taxonomies over Ontologies to Knowledge Graphs, July 2014. <a href="https://blog.semantic-web.at/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs">https://blog.semantic-web.at/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs</a> [August, 2016].</td>
</tr>
<tr>
<td>“We define a Knowledge Graph as an RDF graph. An RDF graph consists of a set of RDF triples where each RDF triple ((s, p, o)) is an ordered set of the following RDF terms: a subject (s\in\mathbb{U}), a predicate (p\in\mathbb{P}), and an object (o\in\mathbb{L}). An RDF term is either a URI (u\in\mathbb{U}), a blank node (b\in\mathbb{B}), or a literal (l\in\mathbb{L})”</td>
<td>M. Färber, B. Ell, C. Menne, A. Rettinger, and F. Bartelscher. Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web Journal, 2016. <a href="http://www.semantic-web-journal.net/content/linked-data-quality-dbpedia-freebase-opencyc-wikidata-and-yago">http://www.semantic-web-journal.net/content/linked-data-quality-dbpedia-freebase-opencyc-wikidata-and-yago</a> [August, 2016] (revised version, under review)</td>
</tr>
<tr>
<td>“[...] systems exist, [...], which use a variety of techniques to extract new knowledge, in the form of facts, from the web. These facts are interrelated, and hence, recently this extracted knowledge has been referred to as a knowledge graph.”</td>
<td>J. Pujara, H. Miao, L. Getoor, and W. Cohen. Knowledge Graph Identification. In Proceedings of the 12th International Semantic Web Conference - Part I, ISWC ’13, pages 542–557, New York, USA, 2013.</td>
</tr>
</tbody>
</table>

Table from “Towards a Definition of Knowledge Graphs”.
KG use in modern applications is characterized by several, simplifying properties

- Markus Krötzsch argues that together these distinguish KGs from more traditional knowledge management paradigms:

1. (Graph Form) Normalization: Information is decomposed into small units of information, interpreted as edges of some form of graph.
   1. Allows graph-computing techniques & algorithms (eg, shortest path computations, or network analysis)
2. Connectivity: Knowledge is represented by formal (first-class) relationships between these units.
3. Context: Data is enriched with contextual information to record aspects such as temporal validity, provenance, trustworthiness, or other side conditions and details.
   1. But much of this seems also true of ontologies??
4. Others see a KG as a special (trade-off) flexible combination of knowledge representation structures (maybe partitioned), information management processes and supporting search algorithms or easier querying.
   - Functionally the KG form is intended to support and easy interchange of knowledge between various groups of humans as well as KE tools.
There are many ingredients to a quality KG

- Unstructured/Semistructured content
- Data Graphs
- References, Key Concepts, Relations
- Tables
- Enrichment and Encoding via Domain Ontology
- Relational DB
- External Domain Data
- Customer Data
- DL/ML?

Semantics to understand the data

• From https://www.researchgate.net/post/What_is_Knowledge_Graphs
  Aiit kumar Roy’s What is Knowledge Graphs?

Efforts for communities related to Deep Learning, Knowledge Graphs, and NLP join their forces in order to develop more effective algorithms and applications.
What Formal Ontologizing offers KGs

- “Quality”/expressive knowledge (avoid problem of arbitrary items being used as classes) with formal language experience.
  - but even ontologies can’t express all the knowledge people want to put in KGs.
- Better knowledge organization,
- Support for reasoning, (is symmetric property formally expressed for reasoning?) and.
- Knowledge practices - proper acquisition, maintenance, such as redundancy elimination. (after *Ontologies for Knowledge Graphs*? Markus Krötzsch)
  - But how are user-generated assertions checked?
(Some) Issues

- Real-world knowledge graphs are usually incomplete, so how do we update them to increase coverage and keep things fresh, but also accurate and consistent with quality semantics? (embedding methods etc.)
  - Need advanced functions like autonomous knowledge inference and verification
- Expressiveness - “semantic nets of the 1970s were, almost unilaterally, /much/ more expressive than knowledge graphs or RDF, or any of the other ‘graph’-like modern notations. They typically had ways of encoding quantifier scopes, disjunction, negation and sometimes such things as modal operators.” Pat Hayes
- If you start less formally to build a large KG can you incrementally add better semantics?
- Integration - data available on the Web has diverse formats so integrating data from different sources is still an open research area.
- Can KG efforts be integrated or are we building silos at a different level?
- Need common tools...and KR (coordinated)...etc.
Examples of Current knowledge graph applications

- Obviously many by Google, Facebook, Yahoo etc.

- Some application areas:
  - Question Answer (QA) systems,
  - semantic search (e.g. GeoSpatial entities),
  - dynamic risk analysis,
  - content-based recommendation engines,
  - knowledge arbitrage,


- Voice recognition system (Dragon)

- Manufacturing

- Finance/business/thematic investing..
References & Links

- http://iswc2018.semanticweb.org/panel-enterprise-scale-knowledge-graphs/
- https://lists.w3.org/Archives/Public/semantic-web/2019Jun/0051.html
- http://kgb-workshop.org/