

# An introduction to Conceptual Exploration

Exploring Knowledge Graphs: An Example with Wikidata

Bernhard Ganter, Tom Hanika, Johannes Hirth, Sergei Obiedkov

HAI24

10. 6. 2024

# What is Wikidata? – A free Knowledgegraph!

Vrandečić and Krötzsch 2014

## *Wikidata*

- Launched in 29th October 2012 ([wikidata.org](http://wikidata.org))
- Wikipedia's knowledge graph
- Free, community-built database
- Large graph  
(June 2023: > 1.54B statements on >110M entities)
- Large, active community  
(June 2023: > 25,391 active human editors)
- Many applications

# What is Wikidata? – A free Knowledgegraph!

Vrandečić and Krötzsch 2014

## *Wikidata*

- Launched in 29th October 2012 ([wikidata.org](http://wikidata.org))
- Wikipedia's knowledge graph
- Free, community-built database
- Large graph  
(June 2023: > 1.54B statements on >110M entities)
- Large, active community  
(June 2023: > 25,391 active human editors)
- Many applications

*Freely available, relevant, and active knowledge graph*

# The content of Wikidata entity documents

... as seen on the last slide

*Entity ID:* Entities are identified by **language-independent ids** (e.g., “Q80”)

*Terms header:*

- **label**
- **description**
- list of **aliases** in the user’s language.

*Statements:*

- **sourced claims** for several **properties** (Property IDs)
- statements may have a **rank** (normal, preferred, deprecated)

*Site links:* Connections to pages on other **Wikimedia projects** realise entity-level information integration

# WD Example: Malmö

## Malmö (Q2211)

city in Skåne County, Sweden

Malmö




[▼ In more languages](#)

[Configure](#)

Language	Label	Description	Also known as
English	Malmö	city in Skåne County, Sweden	Malmö
German	Malmö	Hauptstadt der Provinz Skåne län in Südschweden	Malmö
French	Malmö	capitale de la province de Scanie, dans le Sud de la Suède	Malmö
Bavarian	Malmö	No description defined	

[All entered languages](#)

### Statements

<a href="#">instance of</a>	 <a href="#">urban area in Sweden</a> <a href="#">▶ 1 reference</a>
	 <a href="#">seaport</a> <a href="#">▼ 0 references</a>
	 <a href="#">city</a> <a href="#">▼ 0 references</a>

<a href="#">inception</a>	 <a href="#">1353</a> <a href="#">▶ 0 references</a>
---------------------------	--

# WD Example: Malmö (cont.)

country	<span>Q1188</span> Sweden ▶ 1 reference
capital of	<span>Q1188</span> Skåne County start time 1 January 1997 ▼ 0 references
	<span>Q1188</span> Malmö Municipality ▼ 0 references
	<span>Q1188</span> Malmöhus County ▼ 0 references
located in the administrative territorial entity	<span>Q1188</span> Malmö Municipality ▶ 1 reference
	<span>Q1188</span> Burlöv Municipality ▶ 1 reference
	<span>Q1188</span> Lomma Municipality start time 31 December 2015 ▶ 1 reference

# Mathematical View on Wikidata

Hanika, Marx and Stumme 2019

## Definition (Wikidata Knowledge Graph)

Let

- $\mathcal{Q}$  be the set of Wikidata entity items,
- $\mathcal{P}$  be the set of Wikidata properties, and
- let  $\mathcal{V}$  be the set of all possible data values.

We denote by  $\mathcal{E} := \mathcal{Q} \cup \mathcal{P}$  the set of all entities, and define  $\Delta := \mathcal{E} \cup \mathcal{V}$ . Now, the *Wikidata Knowledge Graph* is a map

$$\mathcal{W}: \mathcal{P} \rightarrow \mathfrak{F}(\mathcal{E} \times \Delta \times \mathfrak{F}(\mathcal{P} \times \Delta))$$

assigning to each property  $p$  a ternary relation  $\mathcal{W}(p)$ , where a tuple  $\langle s, v, a \rangle \in \mathcal{W}(p)$  corresponds to a  $p$ -statement on  $s$  with value  $v$  and annotation  $a$ .

# Data sets

---

data set	properties in class (“Wikidata property for ...”)
awards	Q56150830 (?... awards, prizes and honours?)
family	Q22964231 (?... human relationships?)
math	Q22988631 (?... mathematics?)
space	Q28104992 (?... spacecraft?)
time	Q51077473 (?... time and duration?)

---

---

data set	items	properties	statements
awards	429,207	27	892,723
family	307,330	10	728,669
math	36,913	45	84,255
space	7,693	20	30,212
time	216,865	9	219,803
wiki44k	45,021	101	295,352

---



## Data sets (cont.)

data set	density	$ CanBase(\cdot) $	# supported
awards	0.039	280	17
family	0.163	46	46
math	0.040	752	71
space	0.195	157	125
time	0.112	27	0
wiki44k	0.045	7,040	3,556

# Ergebnisse Awards / Family

## Plain Incidence Awards

$\{\text{Nobel prize ID}_{P3188}\} \rightarrow \{\text{award received}_{P166}\}$

# Ergebnisse Awards / Family

## Plain Incidence Awards

$\{\text{Nobel prize ID}_{P3188}\} \rightarrow \{\text{award received}_{P166}\}$

## Plain Incidence Family

$\{\text{godparent}_{P1290}, \text{partner}_{P451}\} \rightarrow \{\text{sibling}_{P3373}\}$

# Ergebnisse Awards / Family

## Plain Incidence Awards

$$\{\text{Nobel prize ID}_{P3188}\} \rightarrow \{\text{award received}_{P166}\}$$

## Plain Incidence Family

$$\{\text{godparent}_{P1290}, \text{partner}_{P451}\} \rightarrow \{\text{sibling}_{P3373}\}$$

## Directed Incidence Family

$$\{\hat{\text{father}}_{P22}, \hat{\text{relative}}_{P1038}, \text{spouse}_{P26}\} \rightarrow \{\text{child}_{P40}\}$$

but not true is

$$\{\hat{\text{father}}_{P22}\} \rightarrow \{\text{child}_{P40}\}$$

because of 1634 non-fictional counterexamples

# Ergebnisse Math / Space

## Plain Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{base}_{P_{3263}}\} \rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

# Ergebnisse Math / Space

## Plain Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{base}_{P_{3263}}\} \rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

## Directed Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{MathWorld identifier}_{P_{2812}}, \text{dual to } P_{1322}\}$   
 $\rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

# Ergebnisse Math / Space

## Plain Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{base}_{P_{3263}}\} \rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

## Directed Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{MathWorld identifier}_{P_{2812}}, \hat{\text{dual to}}_{P_{1322}}\}$   
 $\rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

## Plain Incidence Space

$\{\text{type of orbit}_{P_{522}}, \text{periapsis}_{P_{2244}}\} \rightarrow \{\text{apoapsis}_{P_{2243}}\}$

# Ergebnisse Math / Space

## Plain Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{base}_{P_{3263}}\} \rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

## Directed Incidence Math

$\{\text{has vertex figure}_{P_{1678}}, \text{MathWorld identifier}_{P_{2812}}, \hat{\text{dual to}}_{P_{1322}}\}$   
 $\rightarrow \{\text{has facet polytope}_{P_{1678}}\}$

## Plain Incidence Space

$\{\text{type of orbit}_{P_{522}}, \text{periapsis}_{P_{2244}}\} \rightarrow \{\text{apoapsis}_{P_{2243}}\}$

## Directed Incidence Space

$\{\text{apoapsis}_{P_{2243}}, \hat{\text{type of orbit}}_{P_{522}}\}$   
 $\rightarrow \{\text{orbital period}_{P_{2146}}, \text{type of orbit}_{P_{522}}, \text{periapsis}_{P_{2244}}\}$



# References

- [HMS19] Tom Hanika, Maximilian Marx **and** Gerd Stumme. **?**Discovering implicational knowledge in Wikidata? **in** *International Conference on Formal Concept Analysis*: Springer. 2019, **pages** 315–323.
- [VK14] Denny Vrandečić **and** Markus Krötzsch. **?**Wikidata: a free collaborative knowledgebase? **in** *Communications of the ACM*: 57.10 (2014), **pages** 78–85.