

# Discovering Implicational Knowledge in Wikidata

Tom Hanika<sup>1,2</sup>   **Maximilian Marx**<sup>3</sup>   Gerd Stumme<sup>1,2</sup>

<sup>1</sup>Knowledge & Data Engineering Group, University of Kassel

<sup>2</sup>ITeG, University of Kassel

<sup>3</sup>Knowledge-Based Systems Group, TU Dresden

15th International Conference on Formal Concept Analysis

Full paper: <https://iccl.inf.tu-dresden.de/web/Inproceedings3202/en>

# Explicit Knowledge in Wikidata



Wikidata: the free and open Knowledge Graph of the Wikimedia Foundation

Image sources:

# Explicit Knowledge in Wikidata



**Wikidata:** the free and open Knowledge Graph of the Wikimedia Foundation  
729,892,295 **statements** about 57,535,817 **items** using 6,249 **properties**

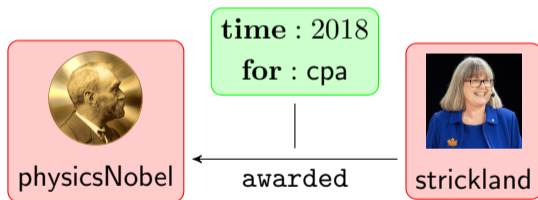


Image sources: Public Domain; Bengt Nyman, CC BY 2.0;

# Explicit Knowledge in Wikidata



**Wikidata:** the free and open Knowledge Graph of the Wikimedia Foundation  
729,892,295 statements about 57,535,817 items using 6,249 properties

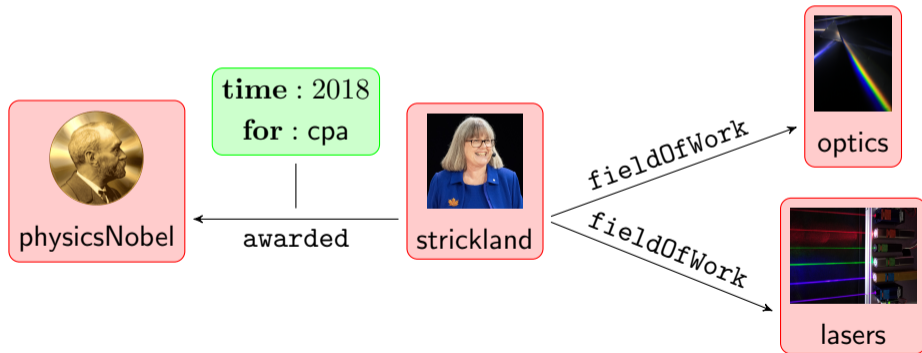


Image sources: Public Domain; Bengt Nyman, CC BY 2.0;  
D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

# Explicit Knowledge in Wikidata



**Wikidata:** the free and open Knowledge Graph of the Wikimedia Foundation  
729,892,295 statements about 57,535,817 items using 6,249 properties

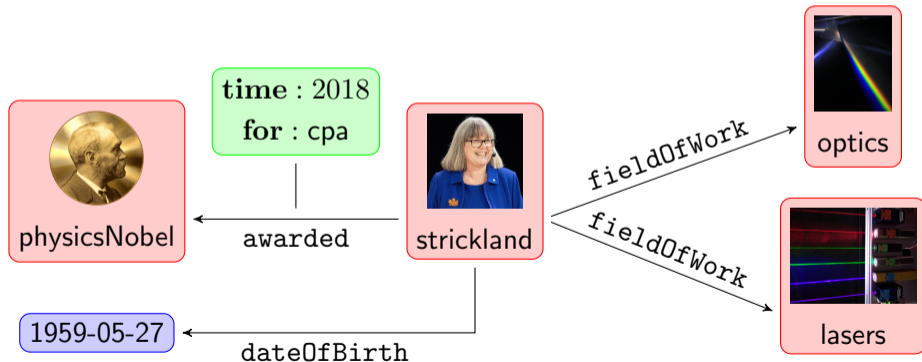


Image sources: Public Domain; Bengt Nyman, CC BY 2.0;  
D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

# Implicit Knowledge in Wikidata

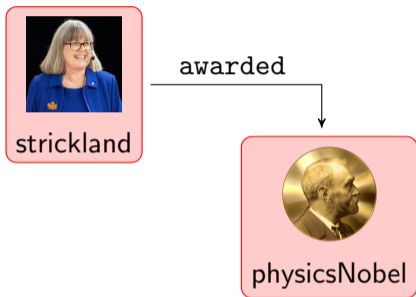


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;

# Implicit Knowledge in Wikidata

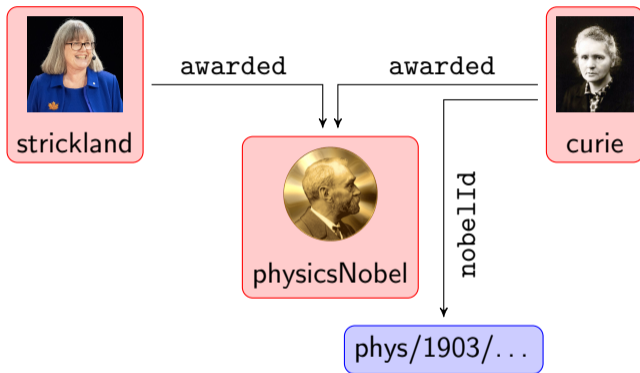


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;  
Public Domain;

# Implicit Knowledge in Wikidata

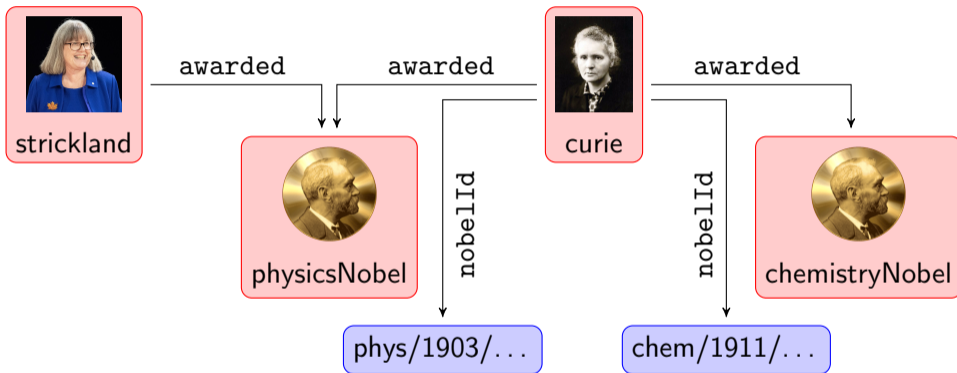


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;  
Public Domain;



# Implicit Knowledge in Wikidata

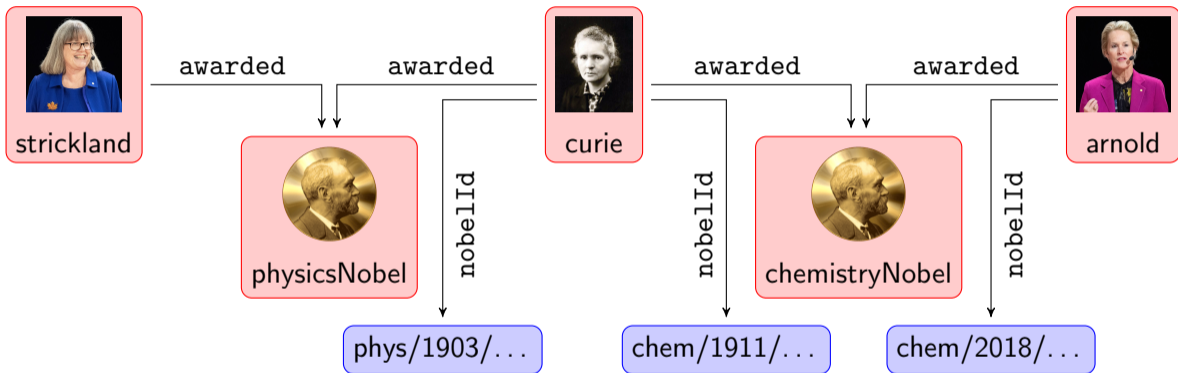
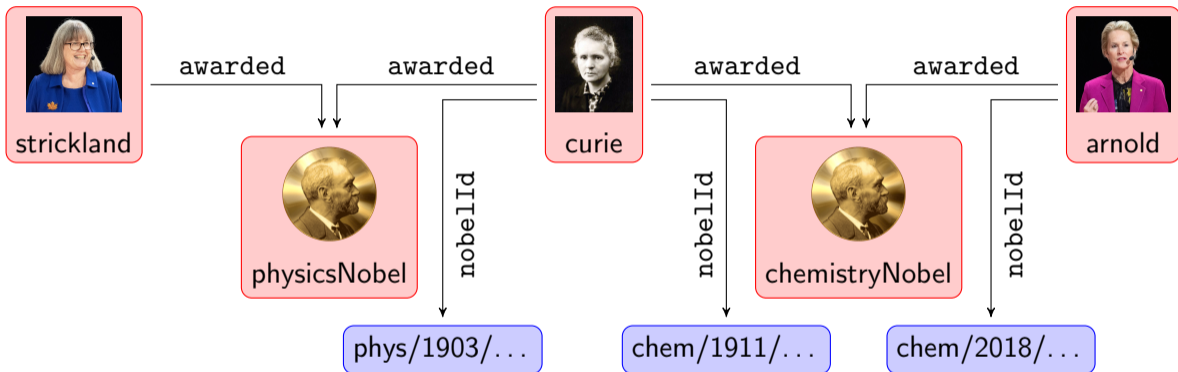


Image sources: Bengt Nyman, CC BY 2.0; Public Domain; Public Domain; Bengt Nyman, CC BY 2.0

# Implicit Knowledge in Wikidata



**'Everybody who was awarded a Nobel Prize has a nobel prize ID'**

`awarded(nobelPrize) → nobelId`

**'Everybody who was awarded a Nobel Prize has a nobel prize ID'?**

`awarded(nobelPrize) → nobelId`

- ▶ 14 counter-examples (e.g., Donna Strickland), all from 2017 and 2018

# Implicational Knowledge in Wikidata

**'Everybody who was awarded a Nobel Prize has a nobel prize ID'?**

`awarded(nobelPrize) → nobelId`

- ▶ 14 counter-examples (e.g., Donna Strickland), all from 2017 and 2018
- ▶ easy to overlook something on the scale of Wikidata

# Implicational Knowledge in Wikidata

**'Everybody who was awarded a Nobel Prize has a nobel prize ID'?**

`awarded(nobelPrize) → nobelId`

- ▶ 14 counter-examples (e.g., Donna Strickland), all from 2017 and 2018
- ▶ easy to overlook something on the scale of Wikidata

**Idea: implicational knowledge can guide editors**

**'Everybody who was awarded a Nobel Prize has a nobel prize ID'?**

`awarded(nobelPrize) → nobelId`

- ▶ 14 counter-examples (e.g., Donna Strickland), all from 2017 and 2018
- ▶ easy to overlook something on the scale of Wikidata

**Idea: implicational knowledge can guide editors**

- ▶ must be **accessible** and **succinct**
- ▶  $\rightsquigarrow$  no complex syntax, no quantifiers, no SPARQL, no Description Logics

# Implicational Knowledge in Wikidata

‘Everybody who was awarded a Nobel Prize has a nobel prize ID’?

`awarded(nobelPrize) → nobelId`

- ▶ 14 counter-examples (e.g., Donna Strickland), all from 2017 and 2018
- ▶ easy to overlook something on the scale of Wikidata

**Idea: implicational knowledge can guide editors**

- ▶ must be **accessible** and **succinct**
- ▶  $\rightsquigarrow$  no complex syntax, no quantifiers, no SPARQL, no Description Logics
- ▶ must be **computable**
- ▶  $\rightsquigarrow$  **Idea:** generate formal context, compute canonical base

## Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence . . .



## Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a  $57,000,000 \times 6,000$  context

## Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a  $57,000,000 \times 6,000$  context
- ▶  $\rightsquigarrow$  choose a **subset** of (related) properties

## Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a  $57,000,000 \times 6,000$  context
- ▶  $\rightsquigarrow$  choose a **subset** of (related) properties

But how to define the incidence?

## Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a  $57,000,000 \times 6,000$  context
- ▶  $\rightsquigarrow$  choose a **subset** of (related) properties

**But how to define the incidence?**

Scaling depends on the properties.

# Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a  $57,000,000 \times 6,000$  context
- ▶  $\rightsquigarrow$  choose a **subset** of (related) properties

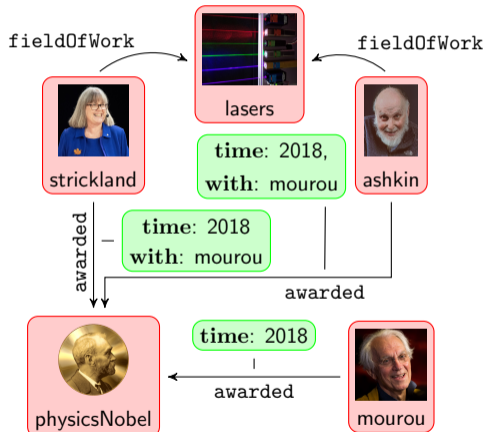
**But how to define the incidence?**

Scaling depends on the properties.

Need to account for

- ▶ direction of statements
- ▶ qualifiers
- ▶ subclass hierarchy
- ▶ ...

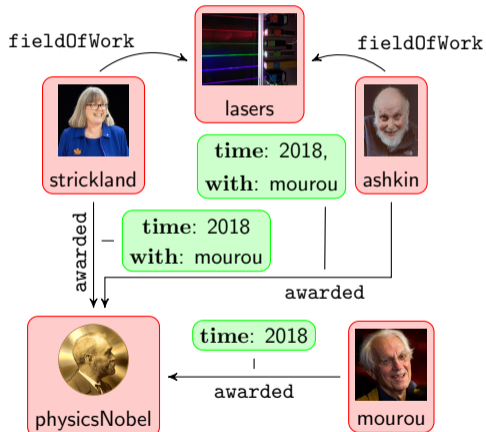
# Scaling by Example



	awarded(physicsNobel)	awarded(chemistryNobel)	awarded(nobelPrize)	awarded@{with: mourou}	fieldOfWork(lasers)	fieldOfWork(radioactivity)
strickland						
mourou						
ashkin						
arnold						
curie						

Image sources: Bengt Nyman, CC BY 2.0; 物理, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

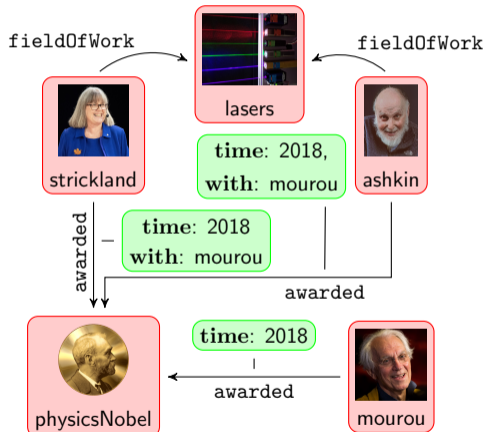
# Scaling by Example



	<code>awarded(physicsNobel)</code>	<code>awarded(chemistryNobel)</code>	<code>awarded(nobelPrize)</code>	<code>awarded@{with: mourou}</code>	<code>fieldOfWork(lasers)</code>	<code>fieldOfWork(radioactivity)</code>
strickland	×		×	×	×	
mourou						
ashkin						
arnold						
curie						

Image sources: Bengt Nyman, CC BY 2.0; 影集, CC BY-SA 3.0;  
 Jérémy Barande, CC BY SA 2.0; Public Domain;  
 Bengt Nyman, CC BY 2.0

# Scaling by Example

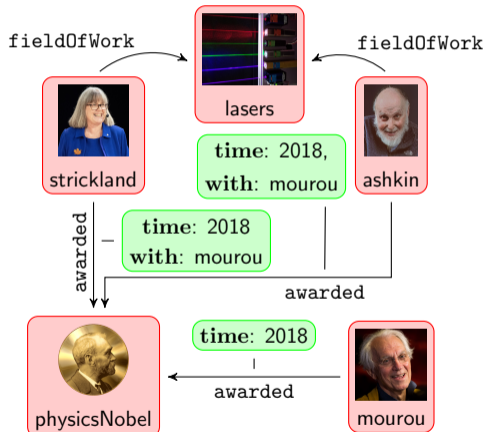


	awarded(physicsNobel)	awarded(chemistryNobel)	awarded(nobelPrize)	awarded@{with: mourou}	fieldOfWork(lasers)	fieldOfWork(radioactivity)
strickland	×		×	×	×	
mourou	×		×			
ashkin						
arnold						
curie						

Image sources: Bengt Nyman, CC BY 2.0; 科学, CC BY-SA 3.0;  
 Jérémy Barande, CC BY SA 2.0; Public Domain;  
 Bengt Nyman, CC BY 2.0



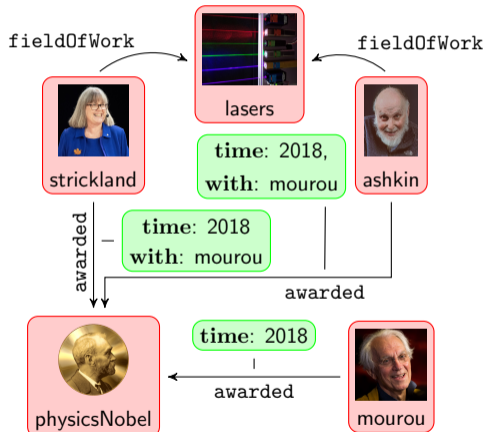
# Scaling by Example



	awarded(physicsNobel)	awarded(chemistryNobel)	awarded(nobelPrize)	awarded@{with: mourou}	fieldOfWork(lasers)	fieldOfWork(radioactivity)
strickland	×		×	×	×	
mourou	×		×			
ashkin	×		×	×	×	
arnold						
curie						

Image sources: Bengt Nyman, CC BY 2.0; 科学, CC BY-SA 3.0;  
 Jérémy Barande, CC BY SA 2.0; Public Domain;  
 Bengt Nyman, CC BY 2.0

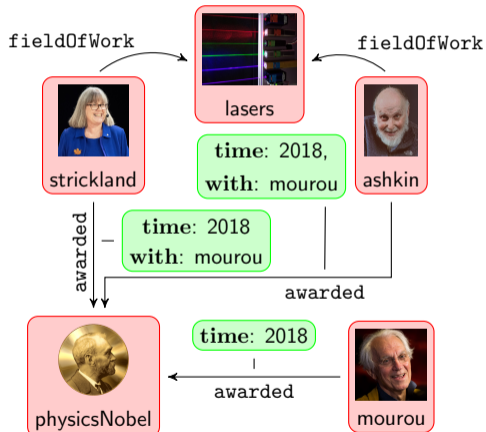
# Scaling by Example



	awarded(physicsNobel)	awarded(chemistryNobel)	awarded(nobelPrize)	awarded@{with: mourou}	fieldOfWork(lasers)	fieldOfWork(radioactivity)
strickland	×		×	×	×	
mourou	×		×			
ashkin	×		×	×	×	
arnold		×	×			
curie						

Image sources: Bengt Nyman, CC BY 2.0; 科学, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

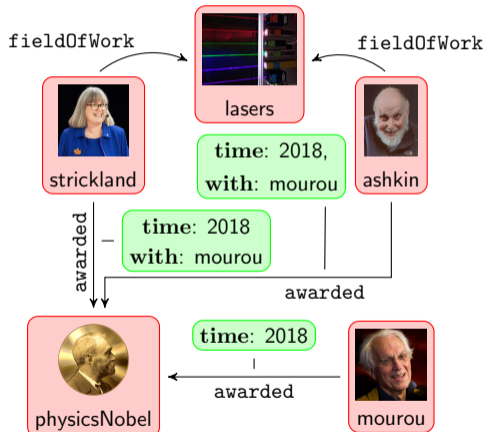
# Scaling by Example



	awarded(physicsNobel)	awarded(chemistryNobel)	awarded(nobelPrize)	awarded@{with: mourou}	fieldOfWork(lasers)	fieldOfWork(radioactivity)
strickland	×		×	×	×	
mourou	×		×			
ashkin	×		×	×	×	
arnold		×	×			
curie	×	×	×			×

Image sources: Bengt Nyman, CC BY 2.0; 科学, CC BY-SA 3.0;  
 Jérémy Barande, CC BY SA 2.0; Public Domain;  
 Bengt Nyman, CC BY 2.0

# Scaling by Example



	<code>awarded(physicsNobel)</code>	<code>awarded(chemistryNobel)</code>	<code>awarded(nobelPrize)</code>	<code>awarded@{with: mourou}</code>	<code>fieldOfWork(lasers)</code>	<code>fieldOfWork(radioactivity)</code>
strickland	×		×	×	×	
mourou	×		×			
ashkin	×		×	×	×	
arnold		×	×			
curie	×	×	×			×

`awarded@{with : mourou}`  $\rightarrow$  `awarded(physicsNobel), fieldOfWork(lasers)`

Image sources: Bengt Nyman, CC BY 2.0; CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

## Experimental Results

data set	items	properties	canonical base	supported
awards	429,207	27	280	17
family	306,908	19	593	576
math	36,913	45	752	71
space	7,693	20	157	125

## Experimental Results

data set	items	properties	canonical base	supported
awards	429,207	27	280	17
family	306,908	19	593	576
math	36,913	45	752	71
space	7,693	20	157	125

awards nobelId → awarded

## Experimental Results

data set	items	properties	canonical base	supported
awards	429,207	27	280	17
family	306,908	19	593	576
math	36,913	45	752	71
space	7,693	20	157	125

awards nobelId  $\rightarrow$  awarded

family  $\hat{\text{father}}, \hat{\text{relative}}, \text{spouse} \rightarrow \text{child}$

... but  $\hat{\text{father}} \rightarrow \text{child}$  has 1,634 (non-fictional) counter-examples

## Experimental Results

data set	items	properties	canonical base	supported
awards	429,207	27	280	17
family	306,908	19	593	576
math	36,913	45	752	71
space	7,693	20	157	125

`awards` `nobelId`  $\rightarrow$  `awarded`

`family`  $\hat{\text{father}}$ ,  $\hat{\text{relative}}$ , `spouse`  $\rightarrow$  `child`

... but  $\hat{\text{father}} \rightarrow \text{child}$  has 1,634 (non-fictional) counter-examples

`math` `vertexFigure`, `base`  $\rightarrow$  `facetPolytope`



## Experimental Results

data set	items	properties	canonical base	supported
awards	429,207	27	280	17
family	306,908	19	593	576
math	36,913	45	752	71
space	7,693	20	157	125

`awards` `nobelId`  $\rightarrow$  `awarded`

`family` `^father`, `^relative`, `spouse`  $\rightarrow$  `child`

... but `^father`  $\rightarrow$  `child` has 1,634 (non-fictional) counter-examples

`math` `vertexFigure`, `base`  $\rightarrow$  `facetPolytope`

`space` `orbitType`, `periapsis`  $\rightarrow$  `apoapsis`

# Conclusions & Outlook

## Conclusions:

- ▶ We find accessible implications that can help editors curating Wikidata
- ▶ Computing implications is feasible for 'small' sets of properties
- ▶ Unwanted/missing implications point to missing/incomplete data
- ▶ Need real-world data? Why not Wikidata?
- ▶ ↪ <https://github.com/mmarx/wikidata-fca>

## Outlook:

- ▶ Add support for semantics of special properties (e.g., temporal intervals)
- ▶ Harness property constraints for automated scaling
- ▶ Integrate readable implications of [Ganter 2019]
- ▶ Exploration: interactively guide editors towards completing Wikidata