

Discovering Implicational Knowledge in Wikidata

Tom Hanika^{1,2,3} **Maximilian Marx**⁴ Gerd Stumme^{1,2}

¹Knowledge & Data Engineering Group, University of Kassel

²ITeG, University of Kassel

³Berlin School of Library and Information Science, Humboldt University of Berlin

⁴Knowledge-Based Systems Group, TU Dresden

Wikimania 2019

Explicit Knowledge in Wikidata



733,638,271 statements about 58,708,331 items using 6,490 properties

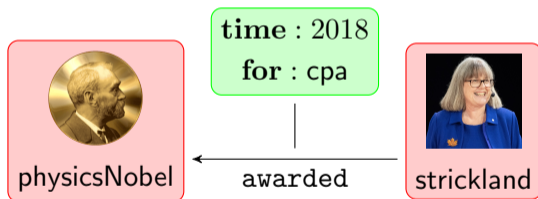


Image sources: Public Domain; Bengt Nyman, CC BY 2.0;

Explicit Knowledge in Wikidata



733,638,271 statements about 58,708,331 items using 6,490 properties

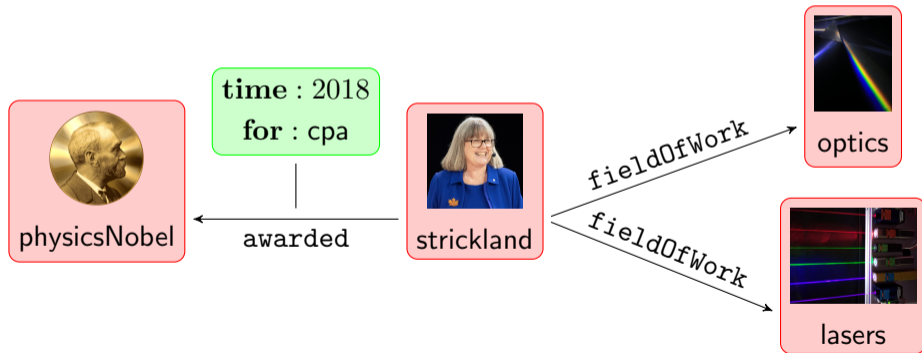


Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

Explicit Knowledge in Wikidata



733,638,271 statements about 58,708,331 items using 6,490 properties

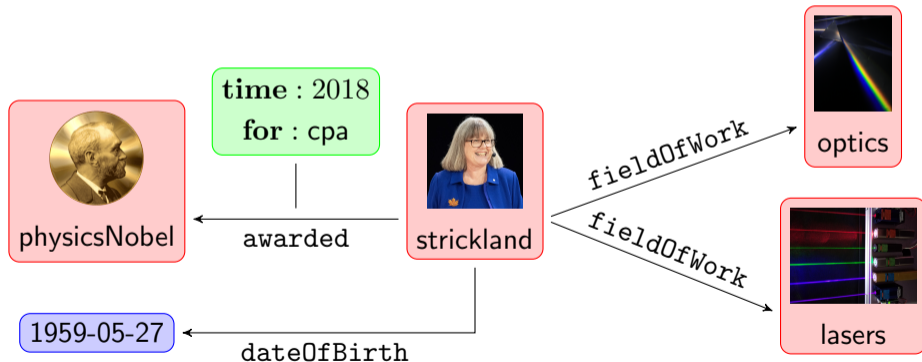


Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

Implicit Knowledge in Wikidata

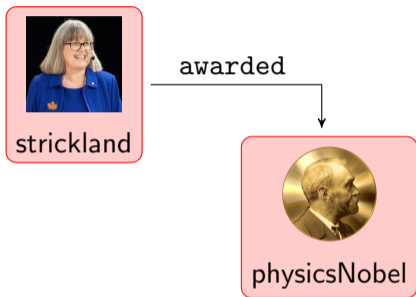


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;

Implicit Knowledge in Wikidata

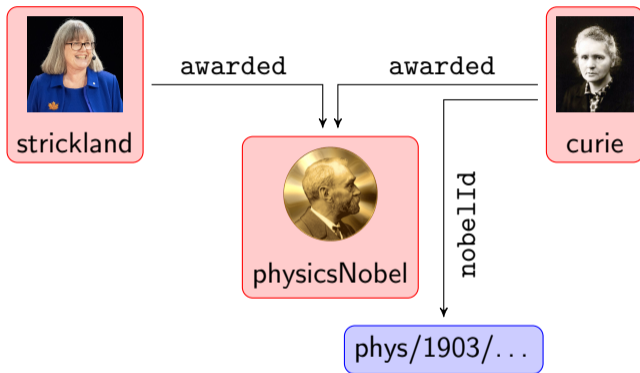


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;
Public Domain;

Implicit Knowledge in Wikidata

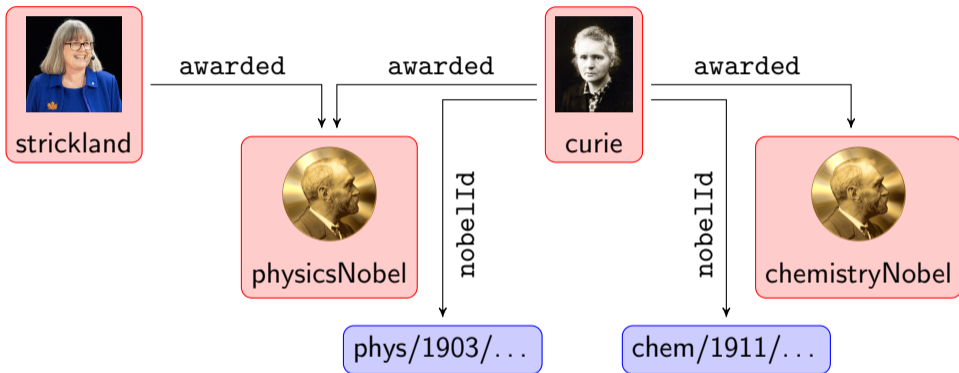


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;
Public Domain;

Implicit Knowledge in Wikidata

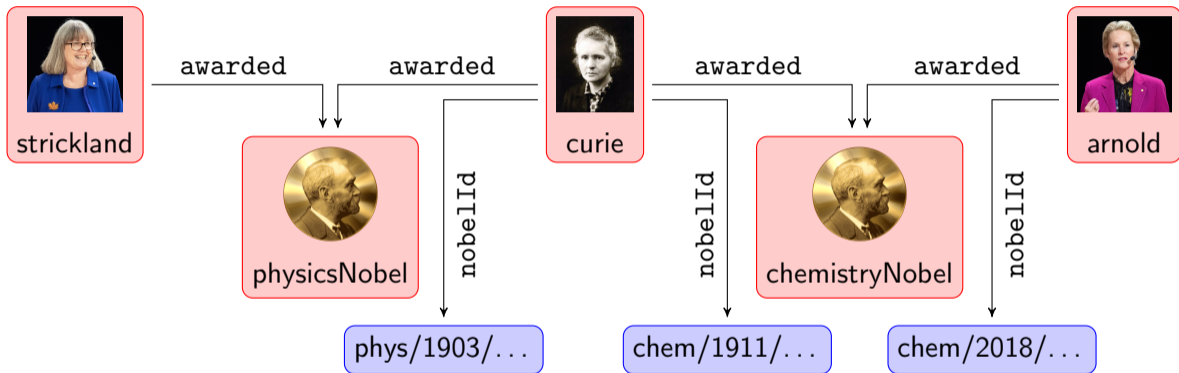
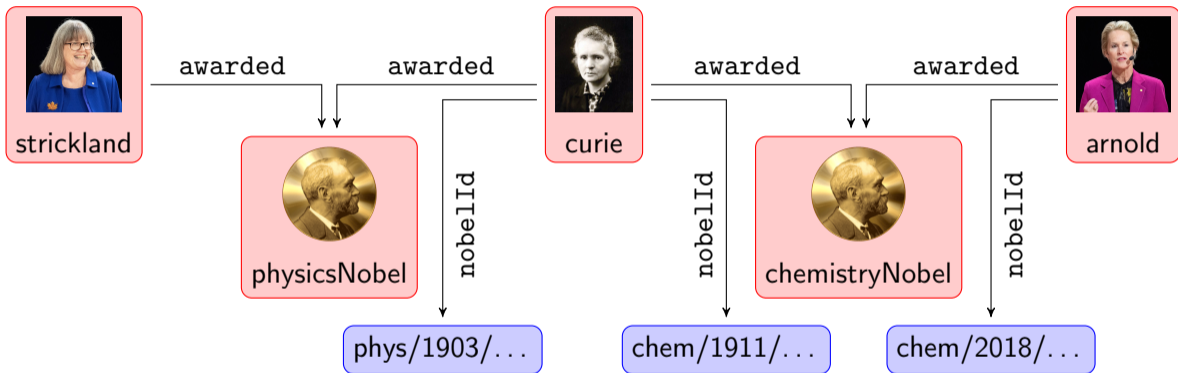


Image sources: Bengt Nyman, CC BY 2.0; Public Domain; Public Domain; Bengt Nyman, CC BY 2.0

Implicit Knowledge in Wikidata



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

`awarded(nobelPrize) → nobelId`

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has a single counter-example (Q509262, Paul M. Romer)

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has a single counter-example (Q509262, Paul M. Romer)
- ▶ 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`

- ▶ has a single counter-example (Q509262, Paul M. Romer)
- ▶ easy to overlook something on the scale of Wikidata

Idea: implicational knowledge can highlight missing & wrong information

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has a single counter-example (Q509262, Paul M. Romer)
- ▶ easy to overlook something on the scale of Wikidata

Idea: implicational knowledge can highlight missing & wrong information

- ▶ 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`

- ▶ has a single counter-example (Q509262, Paul M. Romer)
- ▶ easy to overlook something on the scale of Wikidata

Idea: implicational knowledge can highlight missing & wrong information

- ▶ must be **accessible** and **succinct**
- ▶ \rightsquigarrow no complex syntax, no quantifiers, no SPARQL, no Description Logics
- ▶ must be **computable**
- ▶ \rightsquigarrow **Idea:** employ Formal Concept Analysis to compute propositional implications

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 $58,000,000 \times 6,400$ context

Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $58,000,000 \times 6,400$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

Generating Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $58,000,000 \times 6,400$ 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 $58,000,000 \times 6,400$ 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 $58,000,000 \times 6,400$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

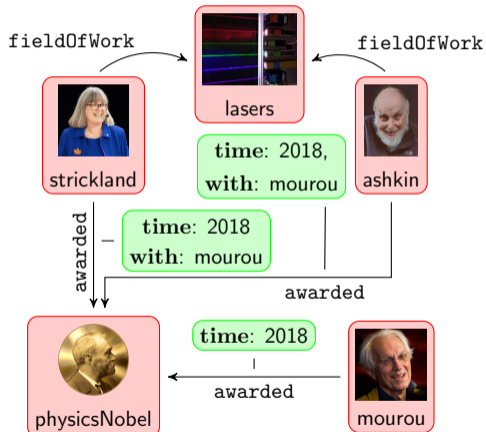
But how to define the incidence?

Scaling depends on the properties.

Need to account for

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

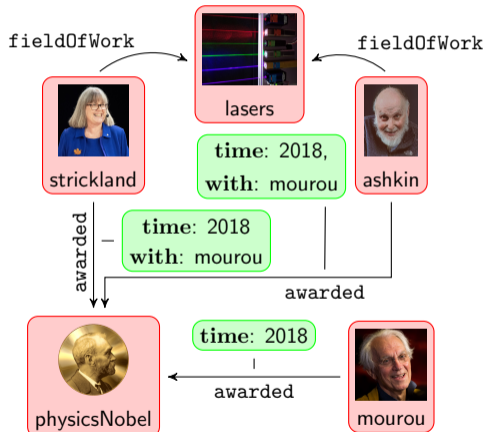
Scaling by Example



	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; 影, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

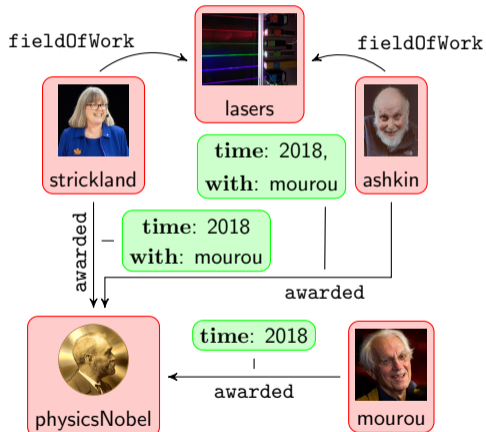
Scaling by Example



	<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, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

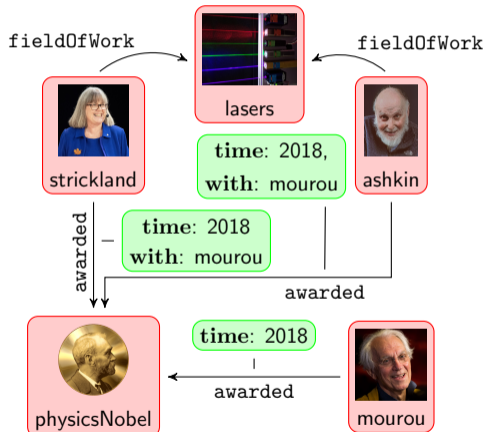
Scaling by Example



	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; 影, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0

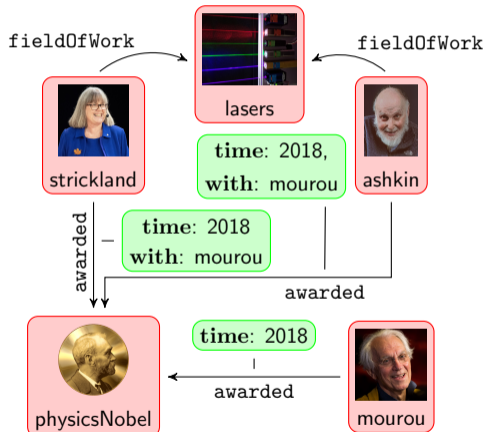
Scaling by Example



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

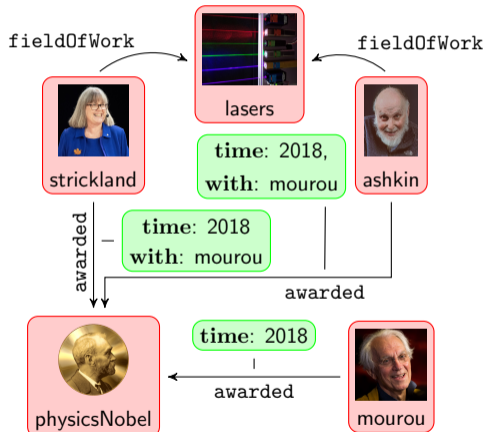
Scaling by Example



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

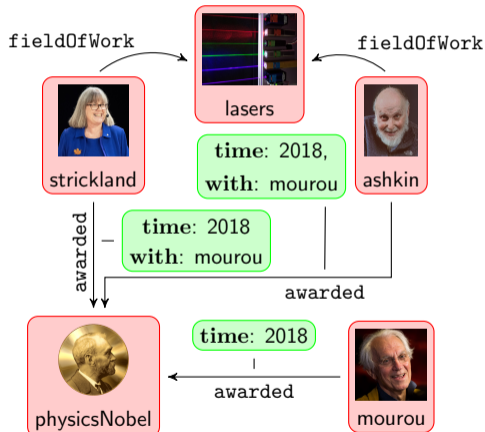
Scaling by Example



	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; 影, 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	×	×	×			×

`awarded@{with : mourou}` → `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

The Exploration Game

- ▶ Computing implications from dumps is possible . . .
- ▶ . . . but many (spurious) counter-examples exclude interesting implications

The Exploration Game

- ▶ Computing implications from dumps is possible . . .
- ▶ . . . but many (spurious) counter-examples exclude interesting implications
- ▶ \rightsquigarrow **Ask an expert!**

The Exploration Game

Current Candidate ▼

Is it true that every item with statements for

1. P364: [original language of film or TV show](#)
2. P495: [country of origin](#)

also has statements for

1. P50: [author](#)
2. P136: [genre](#)

?

[original language of film or TV show, country of origin](#) → [author, genre](#)

Prototype: <https://tools.wmflabs.org/teg/>

Conclusions & Outlook

Conclusions:

- ▶ **Implications can guide you in improving Wikidata**
- ▶ Unwanted/missing implications point to missing statements & items
- ▶ ... and sometimes you learn that the world is more complicated than you thought
- ▶ **Play (a prototype of) The Exploration Game**
- ▶ ↪ <https://tools.wmflabs.org/teg/>

Outlook:

- ▶ Configurable sets of properties
- ▶ Filtering possible counter-examples
- ▶ Scaled properties
- ▶ ... and more?

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

... but `^father` \rightarrow `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`