

Interactively Discovering Implicational Knowledge in Wikidata (The Exploration Game)

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

¹Knowledge & Data Engineering Group, University of Kassel

²ITeG, University of Kassel

³Knowledge-Based Systems Group, TU Dresden

WikiPaka WG @ 36C3

Preliminary report: <https://arxiv.org/abs/1902.00916>

Publication: https://doi.org/10.1007/978-3-030-21462-3_21

Explicit Knowledge in Wikidata



890,137,644 statements about 71,505,937 items using 7,043 properties

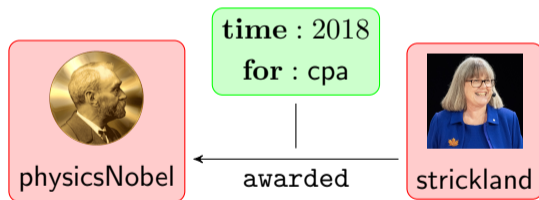


Image sources: Public Domain; Bengt Nyman, CC BY 2.0;

Explicit Knowledge in Wikidata



890,137,644 statements about 71,505,937 items using 7,043 properties

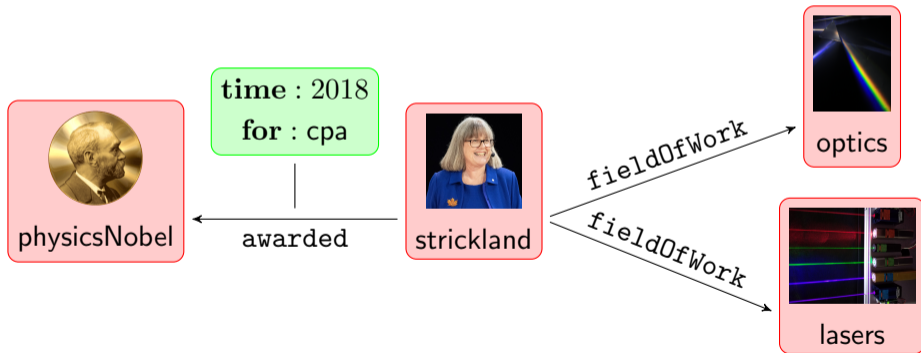


Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

Explicit Knowledge in Wikidata



890,137,644 statements about 71,505,937 items using 7,043 properties

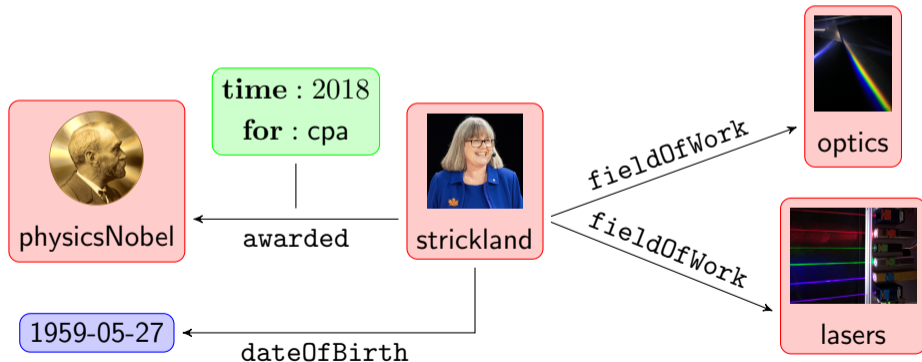


Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; ###, CC BY-SA 3.0

Implicit Knowledge in Wikidata

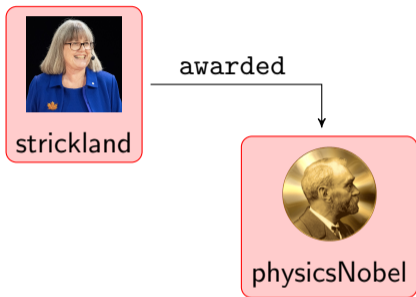


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;

Implicit Knowledge in Wikidata

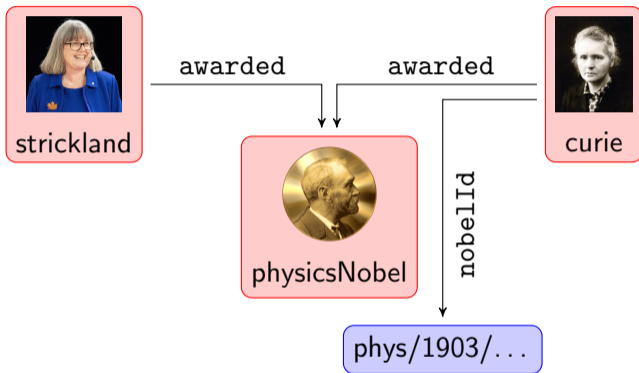


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;
Public Domain;

Implicit Knowledge in Wikidata

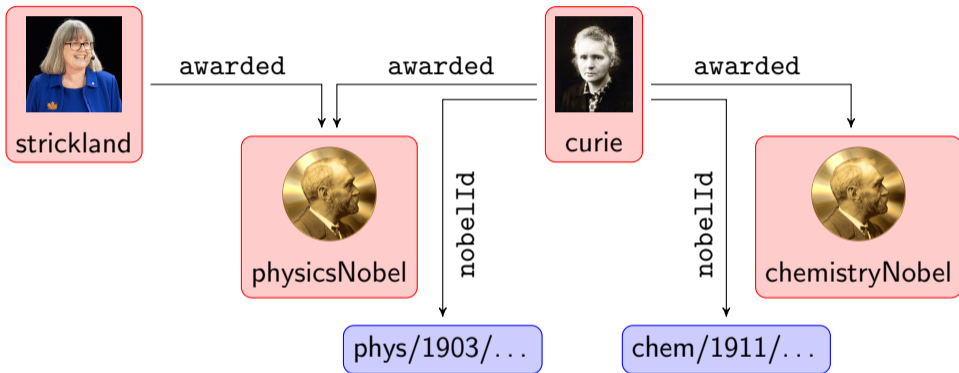


Image sources: Bengt Nyman, CC BY 2.0; Public Domain;
Public Domain;

Implicit Knowledge in Wikidata

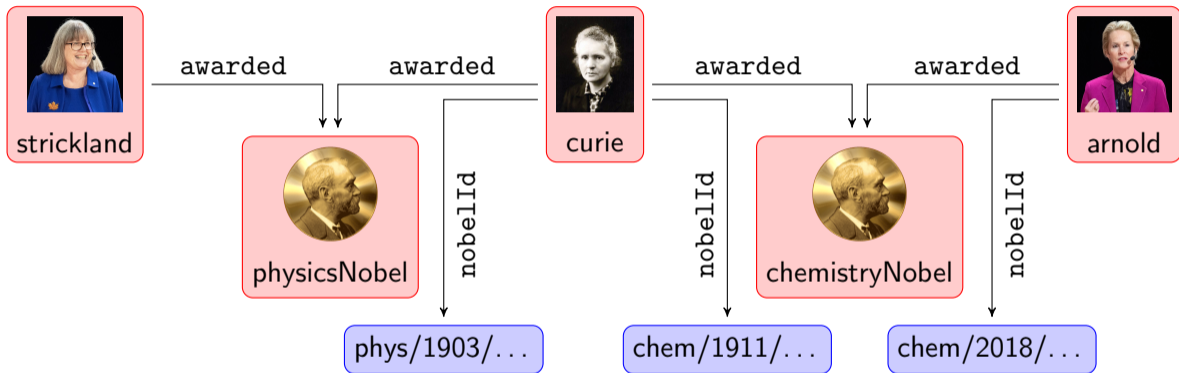
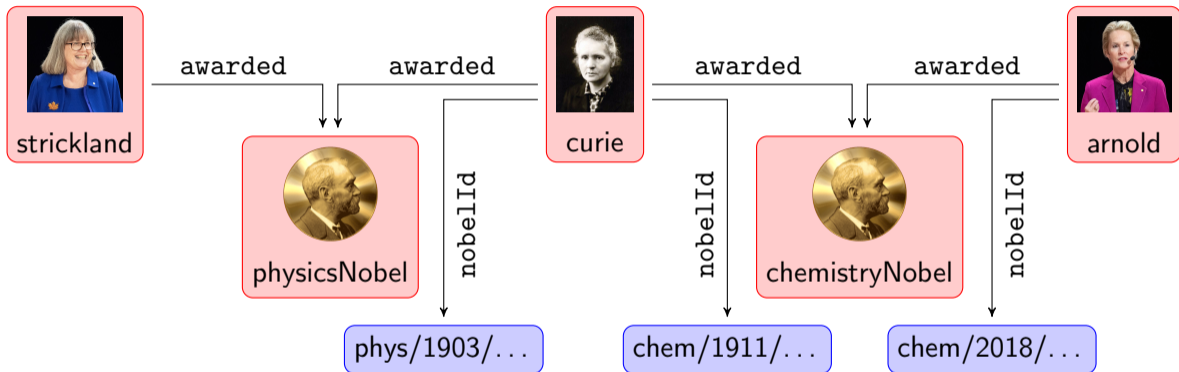


Image sources: Bengt Nyman, CC BY 2.0; Public Domain; Public Domain; Bengt Nyman, CC BY 2.0

Implicit Knowledge in Wikidata



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

`awarded(nobelPrize) → nobelId`

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

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has 25 counter-examples (easy to overlook 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 25 counter-examples (easy to overlook on the scale of Wikidata)

Idea: find implicational knowledge to highlight missing & wrong information

This Knowledge ...

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has 25 counter-examples (easy to overlook on the scale of Wikidata)

Idea: find implicational knowledge to highlight missing & wrong information

This Knowledge ...

- ▶ must be **accessible** and **succinct** to non-experts
 - ⇒ no complex syntax, no quantifiers
 - ⇒ no SPARQL, no Description Logics
 - ⇒ compact representation

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has 25 counter-examples (easy to overlook on the scale of Wikidata)

Idea: find implicational knowledge to highlight missing & wrong information

This Knowledge ...

- ▶ must be **accessible** and **succinct** to non-experts
 - ⇒ no complex syntax, no quantifiers
 - ⇒ no SPARQL, no Description Logics
 - ⇒ compact representation
- ▶ must be **computable** (in short time)

Implicational Knowledge in Wikidata

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

`awarded(nobelPrize) → nobelId`

- ▶ has 25 counter-examples (easy to overlook on the scale of Wikidata)

Idea: find implicational knowledge to highlight missing & wrong information

This Knowledge ...

- ▶ must be **accessible** and **succinct** to non-experts
 - ⇒ no complex syntax, no quantifiers
 - ⇒ no SPARQL, no Description Logics
 - ⇒ compact representation
- ▶ must be **computable** (in short time)

↪ **Approach:** employ **Formal Concept Analysis** to extract propositional implications

Formal Concept Analysis for Explicit Domains

... was developed for restructuring lattice theory in the 1980s by Wille and Ganter.

Definition

$\mathbb{K} := (G, M, I)$ with G, M sets and $I \subseteq G \times M$ is called **formal context**.

- ▶ G called **object set**
- ▶ M called **attribute set**
- ▶ I called **incidence**

Example (Formal Context)

\mathbb{K}	Venomous	Eggs	Mammal
Platypus	×	×	×
Black Widow	×	×	
Duck		×	
Cat			×

Formal Concept Analysis for Explicit Domains

... was developed for restructuring lattice theory in the 1980s by Wille and Ganter.

Definition

$\mathbb{K} := (G, M, I)$ with G, M sets and $I \subseteq G \times M$ is called **formal context**.

- ▶ G called **object set**
- ▶ M called **attribute set**
- ▶ I called **incidence**

Example (Formal Context)

\mathbb{K}	Venomous	Eggs	Mammal
Platypus	×	×	×
Black Widow	×	×	
Duck		×	
Cat			×

Example (Implication)

From the explicit data we can compute:

$$\{\text{Mammal, Venomous}\} \rightarrow \{\text{Eggs}\}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)

Algorithm

Expert (Wikidata)

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)

Algorithm

\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

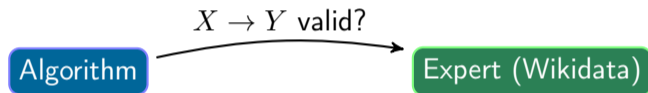
Expert (Wikidata)

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



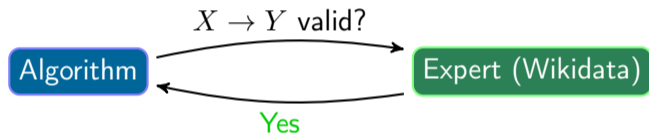
\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



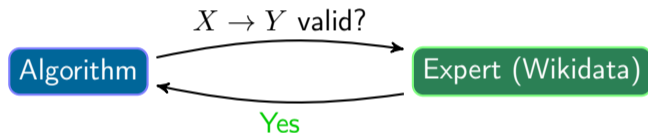
\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



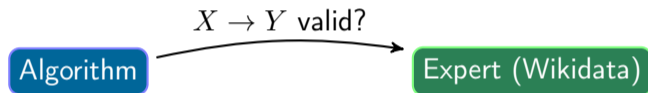
\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell, \\ X \rightarrow Y \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



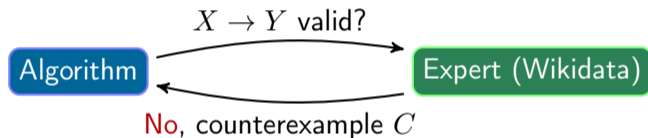
\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



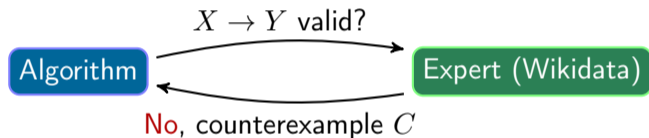
\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



\mathbb{K}	m_1	\dots	m_n
g_1	\dots		
\vdots			
g_k	\dots		
g_{k+1}	<u>C</u>		

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

Interactive extraction of knowledge from domain experts (with minimal # of questions)



\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$$

Formal Concept Analysis for Implicit Knowledge

Attribute Exploration

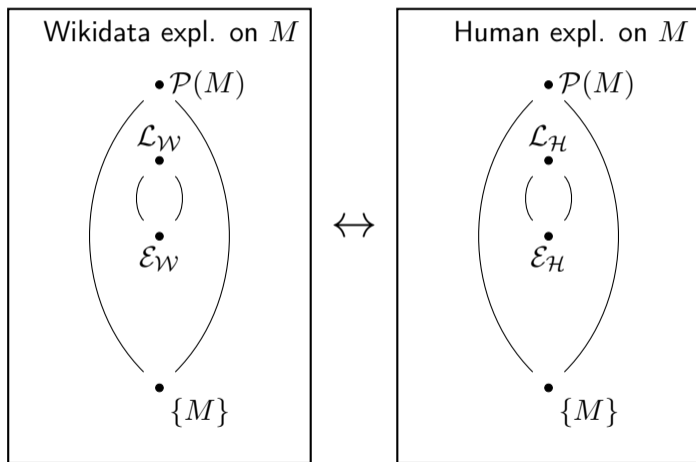
Interactive extraction of knowledge from domain experts (with minimal # of questions)



\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{L} = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$$

Closure System of Knowledge Bases



Viewing Subsets of Wikidata as Formal Contexts

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

Viewing Subsets of Wikidata as Formal Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $71,000,000 \times 7,000$ context

Viewing Subsets of Wikidata as Formal Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $71,000,000 \times 7,000$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

Viewing Subsets of Wikidata as Formal Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $71,000,000 \times 7,000$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

But how to define the incidence?

Viewing Subsets of Wikidata as Formal Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $71,000,000 \times 7,000$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

But how to define the incidence?

Scaling depends on the properties.

Viewing Subsets of Wikidata as Formal Contexts

- ▶ easy: items as **objects**, properties as **attributes**, incidence ...
- ▶ ... and we end up with a $71,000,000 \times 7,000$ context
- ▶ \rightsquigarrow choose a **subset** of (related) properties

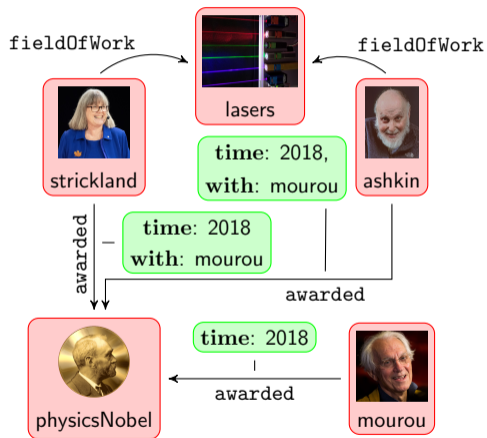
But how to define the incidence?

Scaling depends on the properties.

Possibly want to account for

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

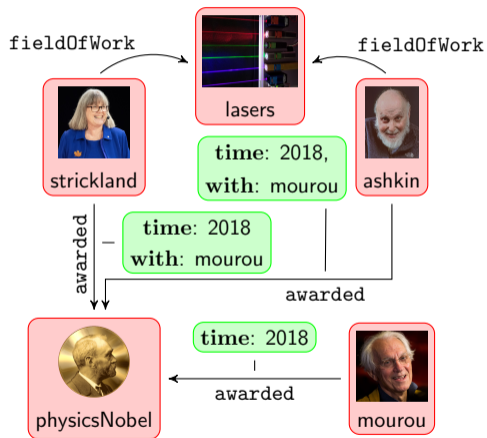
Scaling by Example



	<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>
<code>strickland</code>						
<code>mourou</code>						
<code>ashkin</code>						
<code>arnold</code>						
<code>curie</code>						

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

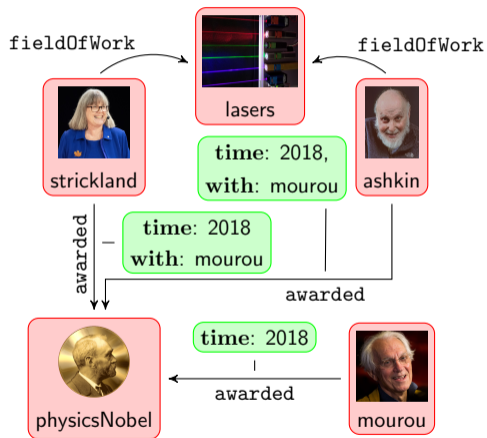
Scaling by Example



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

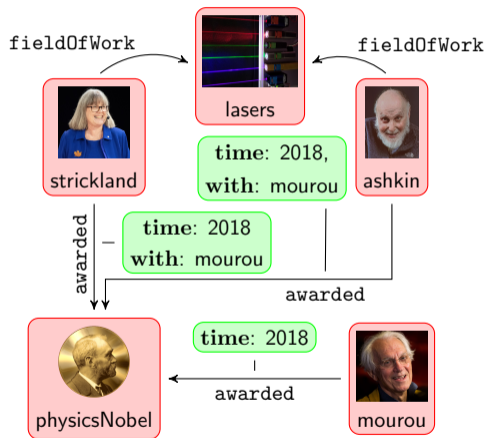
Scaling by Example



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

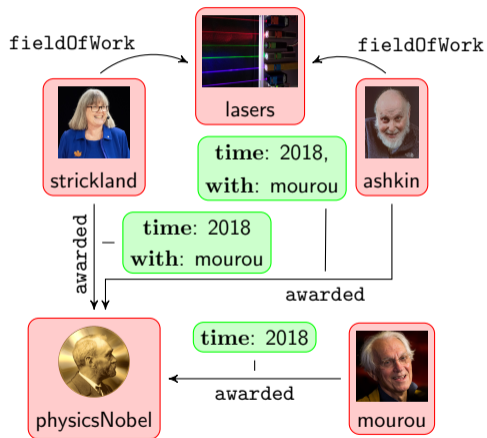
Scaling by Example



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

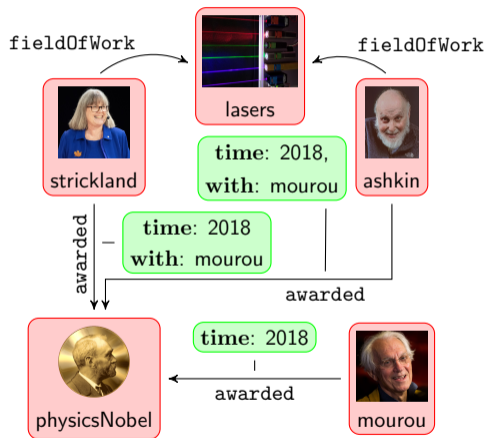
Scaling by Example



	<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>
<code>strickland</code>	×		×	×	×	
<code>mourou</code>	×		×			
<code>ashkin</code>	×		×	×	×	
<code>arnold</code>		×	×			
<code>curie</code>						

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

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

The Exploration Game (<https://tools.wmflabs.org/teg/>)

The Exploration Game

Game Setup

Properties

Presets

- Countries & Borders**
country, continent, shares border with
- Credit Cards**
card network, operator, fee
- Use of Energy**
source of energy, use
- Memory & Computation**
CPU, volatile random-access memory capacity
- Space launches**
time of spacecraft launch, time of spacecraft landing

Custom Properties

Current selection

Game

Number of Counterexamples

start the game! add at least one property to start the game

The Exploration Game

Game Setup

Properties

Presets

- Countries & Borders**
country, continent, shares border with
- Credit Cards**
card network, operator, fee
- Use of Energy**
source of energy, use
- Memory & Computation**
CPU, volatile random-access memory capacity
- Space launches**
time of spacecraft launch, time of spacecraft landing

Custom Properties

Current selection

- card network remove
- operator remove
- fee remove

Game

Number of Counterexamples

start the game!

The Exploration Game: Example

card network(P4443), operator (P137), fee (P2555)

Current Candidate



Is it true that every item should have statements for

1. P2555: [fee](#)
2. P137: [operator](#)
3. P4443: [card network](#)

?

[empty] → [fee](#), [operator](#), [card network](#)

✓accept this implication

✗reject this implication

There are more than 1000 items that do not satisfy this implication. Here are some of them. Are they valid counterexamples?

- [First Mas Government](#): [empty]

The Exploration Game: Example

Current Candidate ▾

Is it true that every item with statements for

1. P4443: [card network](#)

also should have statements for

1. P2555: [fee](#)
2. P137: [operator](#)

?

[card network](#) → [fee](#), [operator](#)

There are 26 items that do not satisfy this implication. Here are some of them. Are they valid counterexamples?

- [Discover Card: card network](#)
- [Palladium Card: card network](#)

The Exploration Game: Example

Current Candidate ▼

Is it true that every item with statements for

1. P137: [operator](#)

also should have statements for

1. P2555: [fee](#)
2. P4443: [card network](#)

?

[operator](#) → [fee](#), [card network](#)

There are more than 1000 items that do not satisfy this implication. Here are some of them. Are they valid counterexamples?

- [Raidió Teilifís Éireann: operator](#)

The Exploration Game: Example

Current Candidate

Is it true that every item with statements for

1. P137: [operator](#)
2. P4443: [card network](#)

also should have statements for

1. P2555: [fee](#)

?

[operator](#), [card network](#) → [fee](#)

✓accept this implication

✗reject this implication

There are 23 items that do not satisfy this implication. Here are some of them. Are they valid counterexamples?

- [Palladium Card](#): [operator](#), [card network](#)
- [American Express Gold Card](#): [operator](#), [card network](#)

The Exploration Game: Example

Current Candidate



Is it true that every item with statements for

1. P2555: [fee](#)
2. P4443: [card network](#)

also should have statements for

1. P137: [operator](#)

?

[fee](#), [card network](#) → [operator](#)

✓ accept this implication

✗ reject this implication

The Exploration Game: Example

Game State ▼

Implications Properties Counterexamples

1. Is it true that every item with statements for

1. P137: [operator](#)
2. P4443: [card network](#)

also should have statements for

1. P2555: [fee](#)

?

[operator, card network](#) → [fee](#)

2. Is it true that every item with statements for

1. P2555: [fee](#)
2. P4443: [card network](#)

also should have statements for

1. P137: [operator](#)

?

[fee, card network](#) → [operator](#)

Conclusions & Outlook

Conclusions:

- ▶ We offer a novel procedure for exploring properties in Wikidata.
- ▶ Unwanted/missing implications point to missing statements & items
- ▶ ... and sometimes you learn that the world is more complicated than you thought
- ▶ \rightsquigarrow **Implications can guide you in improving Wikidata**

Outlook:

- ▶ Configurable and filterable counter-examples
- ▶ Domain-specific scaling of properties
- ▶ Compare your knowledge against Wikidata
- ▶ ... and more? Let us know!

Conclusions & Outlook

Conclusions:

- ▶ We offer a novel procedure for exploring properties in Wikidata.
- ▶ Unwanted/missing implications point to missing statements & items
- ▶ ... and sometimes you learn that the world is more complicated than you thought
- ▶ \rightsquigarrow **Implications can guide you in improving Wikidata**

Outlook:

- ▶ Configurable and filterable counter-examples
- ▶ Domain-specific scaling of properties
- ▶ Compare your knowledge against Wikidata
- ▶ ... and more? Let us know!

The End

Play (a prototype of) **The Exploration Game** (\rightsquigarrow <https://tools.wmflabs.org/teg/>)