Discovering Implicational Knowledge in Wikidata

Tom Hanika\textsuperscript{1,2,3} Maximilian Marx\textsuperscript{4} Gerd Stumme\textsuperscript{1,2}

\textsuperscript{1}Knowledge & Data Engineering Group, University of Kassel

\textsuperscript{2}ITeG, University of Kassel

\textsuperscript{3}Berlin School of Library and Information Science, Humboldt University of Berlin

\textsuperscript{4}Knowledge-Based Systems Group, TU Dresden

Wikimania 2019
Explicit Knowledge in Wikidata

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

physicsNobel

awarded

time : 2018

for : cpa

strickland

Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; 彭嘉傑, CC BY-SA 3.0
Explicit Knowledge in Wikidata

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

- physicsNobel
- time: 2018
- for: cpa
- awarded
- strickland
- fieldOfWork
- fieldOfWork
- optics
- lasers

Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; 彭嘉傑, CC BY-SA 3.0
Explicit Knowledge in Wikidata

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

physicsNobel

1959-05-27

fieldOfWork: optics

fieldOfWork: lasers

awarded

time: 2018

for: cpa

strickland

dateOfBirth

User: Akorenchkin / @korenchkin (TU Dresden)

Discovering Implicational Knowledge in Wikidata

Image sources: Public Domain; Bengt Nyman, CC BY 2.0; D-Kuru, CC BY-SA 3.0; 彭嘉傑, CC BY-SA 3.0
Implicit Knowledge in Wikidata

strickland \( \xrightarrow{\text{awarded}} \) physicsNobel

Everybody who was awarded a Nobel Prize has a Nobel Prize ID.
Implicit Knowledge in Wikidata

strickland \quad \text{awarded} \quad \text{awarded} \quad \text{curie}

physicsNobel \quad \text{nobelId}

phys/1903/…
Implicit Knowledge in Wikidata

- strickland
- physicsNobel
- physics/1903/…

- curie
- chemistryNobel
- chem/1911/…

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

Image sources: Bengt Nyman, CC BY 2.0; Public Domain;
Implicit Knowledge in Wikidata

- **strickland** awarded **physicsNobel**
  - nobelId: `phys/1903/...`

- **curie** awarded **chemistryNobel**
  - nobelId: `chem/1911/...`

- **arnold** awarded **chemistryNobel**
  - nobelId: `chem/2018/...`

Everybody who was awarded a Nobel Prize has a nobelId.
‘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

- 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’?

\[ \text{awarded(nobelPrize)} \rightarrow \text{nobelId} \]

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

\[
\text{awarded(nobelPrize)} \rightarrow \text{nobelId}
\]

- 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
- \(\sim\) no complex syntax, no quantifiers, no SPARQL, no Description Logics
‘Everybody who was awarded a Nobel Prize has a noble 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**
- no complex syntax, no quantifiers, no SPARQL, no Description Logics
- must be **computable**

**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
- $\leadsto$ 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  
- $\leadsto$ 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 × 6,400 context
- ⊳ 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

fieldOfWork

strickland

lasers

time: 2018, with: mourou

ashkin

fieldOfWork

awarded

physicsNobel

time: 2018

with: mourou

awarded

moureu

ashkin

arnold

curie

awarded

(awarded[with: mourou]

fieldOfWork[lasers]

fieldOfWork[radioactivity]

<table>
<thead>
<tr>
<th>strickland</th>
<th>X</th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>mourou</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ashkin</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arnold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curie</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scaling by Example

Field of Work

<table>
<thead>
<tr>
<th>Name</th>
<th>Field of Work</th>
<th>physics Nobel</th>
<th>awarded (physics Nobel)</th>
<th>awarded (chemistry Nobel)</th>
<th>awarded (nobel Prize)</th>
</tr>
</thead>
<tbody>
<tr>
<td>strickland</td>
<td>lasers</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>mourou</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>ashkin</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time: 2018</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with: mourou</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time: 2018</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arnold</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curie</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Image sources: Bengt Nyman, CC BY 2.0; 彭嘉杰, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0
Scaling by Example

fieldOfWork → fieldOfWork

Time: 2018, with: mourou

Time: 2018, with: mourou

Time: 2018

awarded(physicNobel)

awarded(chemistryNobel)

awarded nobelPrize

awarded@{with: mourou}

fieldOfWork(lasers)

fieldOfWork(radioactivity)

<table>
<thead>
<tr>
<th></th>
<th>award(physicNobel)</th>
<th>award(chemistryNobel)</th>
<th>award nobelPrize</th>
<th>award@{with: mourou}</th>
</tr>
</thead>
<tbody>
<tr>
<td>strickland</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>mourou</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>ashkin</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>arnold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curie</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Image sources: Bengt Nyman, CC BY 2.0; 彭嘉傑, CC BY-SA 3.0; Jérémy Barande, CC BY SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0
Scaling by Example

Image sources: Bengt Nyman, CC BY 2.0; 彭嘉傑, CC BY-SA 3.0; Jérémy Barande, CC BY-SA 2.0; Public Domain; Bengt Nyman, CC BY 2.0
Scaling by Example

- awarded@\{with : mourou\} → awarded(physicsNobel), fieldOfWork(lasers)

### Table

<table>
<thead>
<tr>
<th></th>
<th>awarded(chemistryNobel)</th>
<th>awarded(nobelPrize)</th>
<th>awarded@{with : mourou}</th>
<th>fieldOfWork(lasers)</th>
<th>fieldOfWork(lasers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>strickland</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>mourou</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>ashkin</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>arnold</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curie</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

Image sources: Bengt Nyman, CC BY 2.0; 彭 嘉 杰, 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
- 💡 Ask an expert!

The Exploration Game

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

<table>
<thead>
<tr>
<th>data set</th>
<th>items</th>
<th>properties</th>
<th>canonical base</th>
<th>supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>awards</td>
<td>429,207</td>
<td>27</td>
<td>280</td>
<td>17</td>
</tr>
<tr>
<td>family</td>
<td>306,908</td>
<td>19</td>
<td>593</td>
<td>576</td>
</tr>
<tr>
<td>math</td>
<td>36,913</td>
<td>45</td>
<td>752</td>
<td>71</td>
</tr>
<tr>
<td>space</td>
<td>7,693</td>
<td>20</td>
<td>157</td>
<td>125</td>
</tr>
</tbody>
</table>
## 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 → awarded

**family** ^father, ^relative, spouse → child

... but ^father → child has 1,634 (non-fictional) counter-examples
### Experimental Results

<table>
<thead>
<tr>
<th>data set</th>
<th>items</th>
<th>properties</th>
<th>canonical base</th>
<th>supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>awards</td>
<td>429,207</td>
<td>27</td>
<td>280</td>
<td>17</td>
</tr>
<tr>
<td>family</td>
<td>306,908</td>
<td>19</td>
<td>593</td>
<td>576</td>
</tr>
<tr>
<td>math</td>
<td>36,913</td>
<td>45</td>
<td>752</td>
<td>71</td>
</tr>
<tr>
<td>space</td>
<td>7,693</td>
<td>20</td>
<td>157</td>
<td>125</td>
</tr>
</tbody>
</table>

- **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
- **math** vertexFigure, base $\rightarrow$ facetPolytope
## Experimental Results

<table>
<thead>
<tr>
<th>data set</th>
<th>items</th>
<th>properties</th>
<th>canonical base</th>
<th>supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>awards</td>
<td>429,207</td>
<td>27</td>
<td>280</td>
<td>17</td>
</tr>
<tr>
<td>family</td>
<td>306,908</td>
<td>19</td>
<td>593</td>
<td>576</td>
</tr>
<tr>
<td>math</td>
<td>36,913</td>
<td>45</td>
<td>752</td>
<td>71</td>
</tr>
<tr>
<td>space</td>
<td>7,693</td>
<td>20</td>
<td>157</td>
<td>125</td>
</tr>
</tbody>
</table>

**awards** nobelId → awarded

**family** ^father, ^relative, spouse → child

... but ^father → child has 1,634 (non-fictional) counter-examples

**math** vertexFigure, base → facetPolytope

**space** orbitType, periapsis → apoapsis