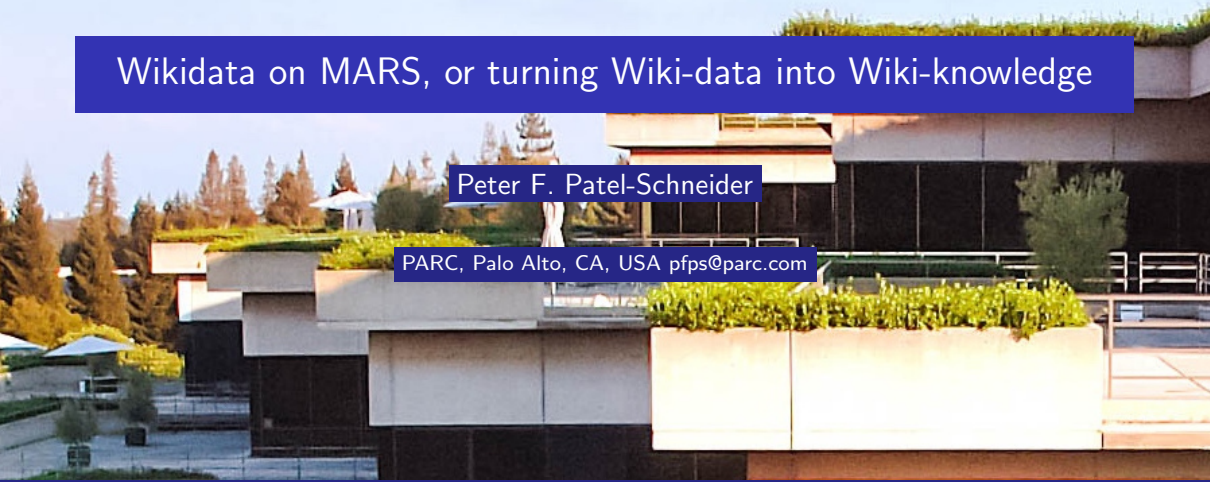
Wikidata on MARS, or turning Wiki-data into Wiki-knowledge

Peter F. Patel-Schneider

PARC, Palo Alto, CA, USA pfps@parc.com



Abstract

Users (particularly programs) need to know far too much to reliably get the information they need from Wikidata. So Wikidata matches its name in that it contains data, not knowledge. An underlying formal basis can be provided for Wikidata, turning Wiki-data into Wiki-knowledge. Getting these formal bases used, so that Wikidata and other Knowledge Graphs contain knowledge, is a separate matter.

Purpose of This Talk

- Show that there is more to knowledge than just Knowledge Graphs
- Show a way of turning Knowledge Graphs into knowledge

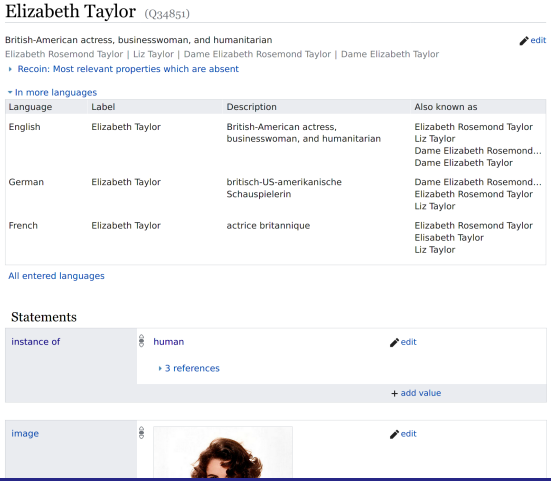
- I'm using Wikidata as my example Knowledge Graph.
  - Wikidata is useful, available, and examinable.
  - Although Wikidata has differences from other Knowledge Graphs, it has much more in common with them, particularly when support for knowledge is considered.
- I'm being a bit harsh on Wikidata.
  - Wikidata is a community effort, which makes organization hard.
  - Wikidata is improving, but not in the way I think it should.

What is Wikidata

- Large community-driven repository
  - Over 93 million entities
  - Anyone can update
  - Anyone can use freely

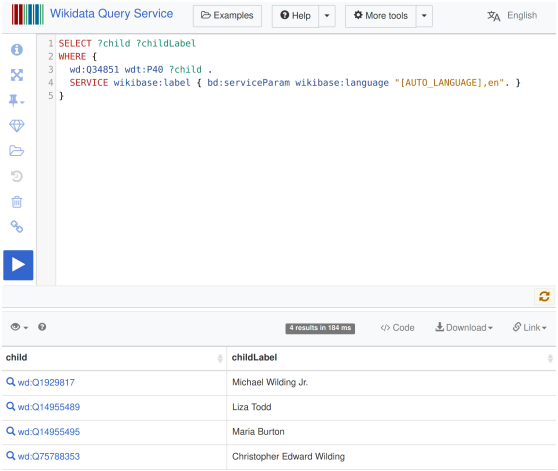
What is Wikidata

- Large community-driven repository
  - Over 93 million entities
  - Anyone can update
  - Anyone can use freely
- Display interface
  Elizabeth Taylor



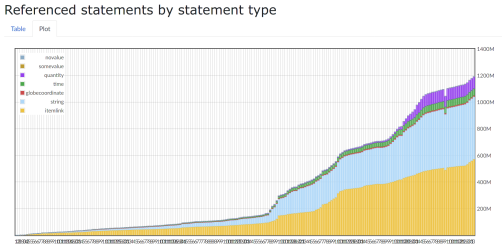
What is Wikidata

- Large community-driven repository
  - Over 93 million entities
  - Anyone can update
  - Anyone can use freely
- Display interface
- SPARQL query interface
  children of Elizabeth Taylor
  - (Annoying to use because of numeric identifiers)



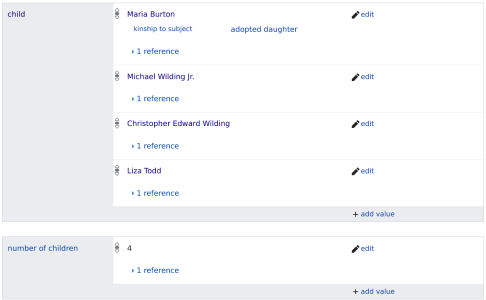
Is Wikidata Useful?

- Absolutely
  - Large
    1.3 billion "facts"



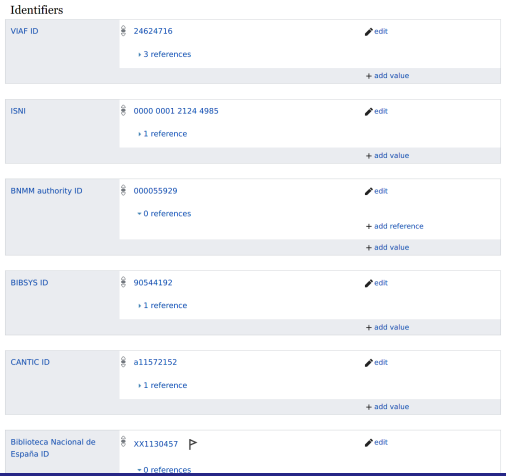
Is Wikidata Useful?

- Absolutely
  - Large
  - High Quality
    both adopted and natural children of Elizabeth Taylor



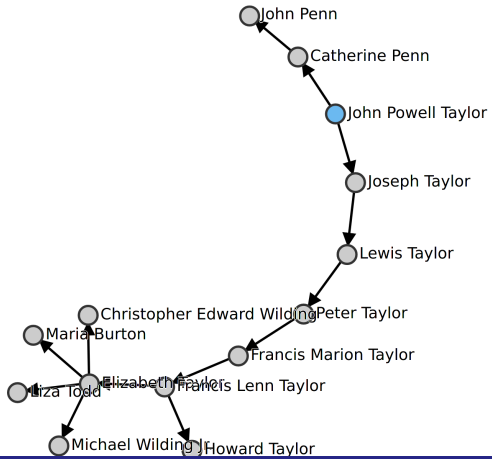
Is Wikidata Useful?

- Absolutely
  - Large
  - High Quality
  - Connections to other sources
    Elizabeth Taylor in other repositories



Is Wikidata A Knowledge Graph?

- It can be considered to be a graph *Descendants of John Powell Taylor*

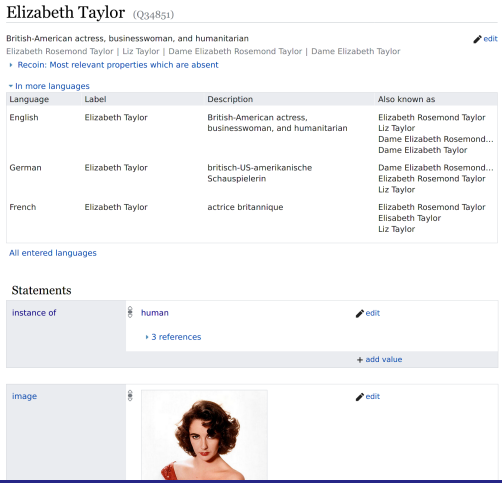


Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
  People, places, etc.

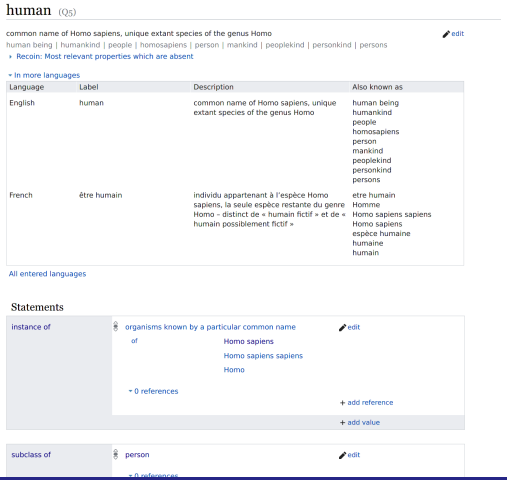
Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Definition of the class human

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



human (Q5)

common name of Homo sapiens, unique extant species of the genus Homo — edit

human being | humankind | people | homosapiens | person | mankind | peoplekind | personkind | persons

▸ Recoin: Most relevant properties which are absent

▾ In more languages

Language	Label	Description	Also known as
English	human	common name of Homo sapiens, unique extant species of the genus Homo	human being humankind people homosapiens person mankind peoplekind personkind persons
French	être humain	individu appartenant à l'espèce Homo sapiens, la seule espèce restante du genre Homo – distinct de « humain fictif » et de « humain possiblement fictif »	etre humain Homme Homo sapiens sapiens Homo sapiens espèce humaine humaine humain

All entered languages

Statements

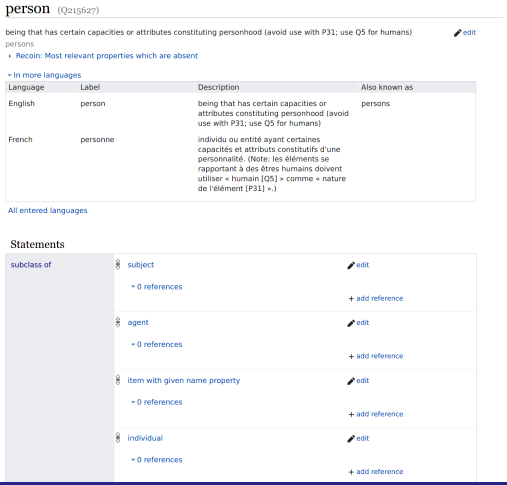
instance of — organisms known by a particular common name — edit
of: Homo sapiens; Homo sapiens sapiens; Homo
▾ 0 references
+ add reference
+ add value

subclass of — person — edit
▾ 0 references

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Definition of the class person, a superclass of human

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



person (Q215627)

being that has certain capacities or attributes constituting personhood (avoid use with P31; use Q5 for humans) edit

persons

Recoin: Most relevant properties which are absent

In more languages

Language	Label	Description	Also known as
English	person	being that has certain capacities or attributes constituting personhood (avoid use with P31; use Q5 for humans)	persons
French	personne	individu ou entité ayant certaines capacités et attributs constitutifs d'une personnalité. (Note: les éléments se rapportant à des êtres humains doivent utiliser « humain (Q5) » comme « nature de l'élément (P31) ».)	

All entered languages

Statements

subclass of

subject edit
0 references
add reference

agent edit
0 references
add reference

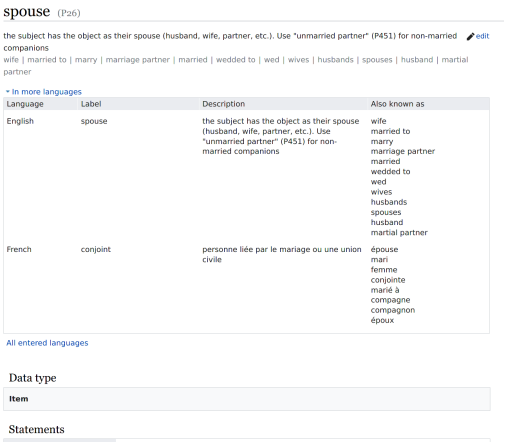
item with given name property edit
0 references
add reference

individual edit
0 references
add reference

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Definition of the relation spouse

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



spouse (P26)

the subject has the object as their spouse (husband, wife, partner, etc.). Use "unmarried partner" (P451) for non-married companions edit

wife | married to | marry | marriage partner | married | wedded to | wed | wives | husbands | spouses | husband | martial partner

In more languages

Language	Label	Description	Also known as
English	spouse	the subject has the object as their spouse (husband, wife, partner, etc.). Use "unmarried partner" (P451) for non-married companions	wife married to marry marriage partner married wedded to wed wives husbands spouses husband martial partner
French	conjoint	personne liée par le mariage ou une union civile	épouse mari femme conjointe marié à compagne compagnon époux

All entered languages

Data type

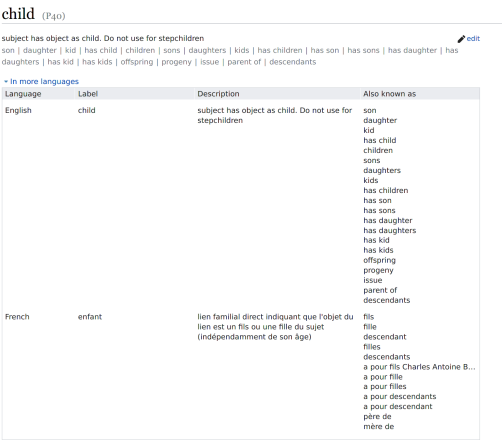
Item

Statements

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Definition of the relation child

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



child (P40)

subject has object as child. Do not use for stepchildren

edit

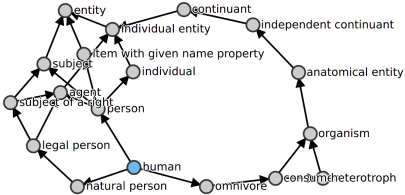
son | daughter | kid | has child | children | sons | daughters | kids | has children | has son | has sons | has daughter | has daughters | has kid | has kids | offspring | progeny | issue | parent of | descendants

▸ In more languages

Language	Label	Description	Also known as
English	child	subject has object as child. Do not use for stepchildren	son daughter kid has child children sons daughters kids has children has son has sons has daughter has daughters has kid has kids offspring progeny issue parent of descendants
French	enfant	lien familial direct indiquant que l'objet du lien est un fils ou une fille du sujet (indépendamment de son âge)	fils fille descendant filles descendants a pour fils Charles Antoine B… a pour fille a pour filles a pour descendants a pour descendant père de mère de

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Generalizations of human

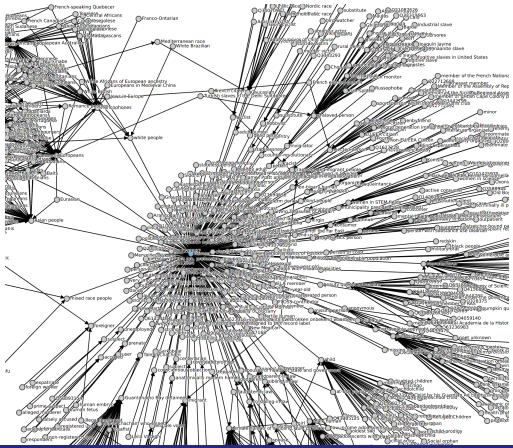


Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
  Some specializations of human

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Abstracting from definitions in "Towards a Definition of Knowledge Graphs", Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

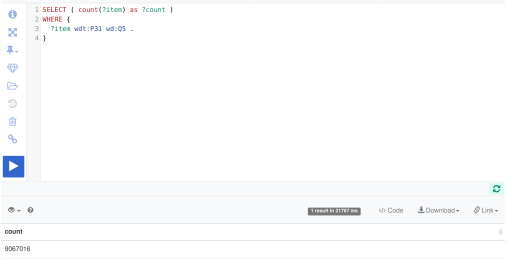
So, even though its name includes *data*, Wikidata meets requirements for Knowledge Graphs.

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*

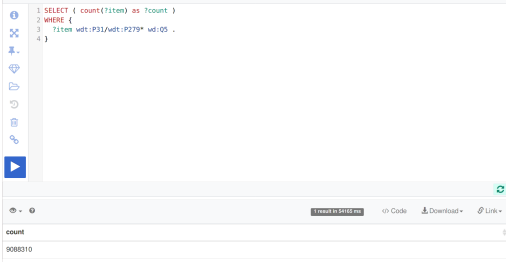
Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
  - Consequences of information are not computed
    How many humans in Wikidata?
    9,067,016



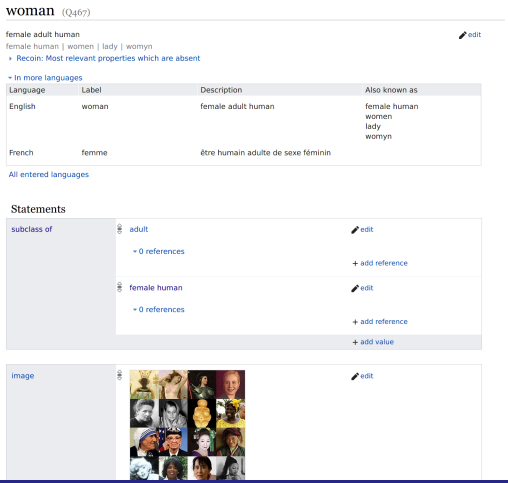
Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
  - Consequences of information are not computed
    How many humans in Wikidata?
    9,067,016 . . . or 9,088,310?
    Need to explicitly count transitive closure of subclasses.
    Also affects properties.



Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct *woman is described as female adult human*



Does Wikidata Have Knowledge?

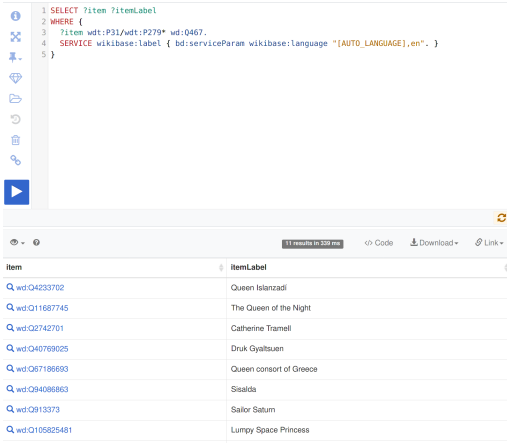
- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
  But no (really 11) women in Wikidata
  (At one time there were 42)

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
  But no (really 11) women in Wikidata
  (At one time there were 42)

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
  But no (really 11) women in Wikidata
  (At one time there were 42)



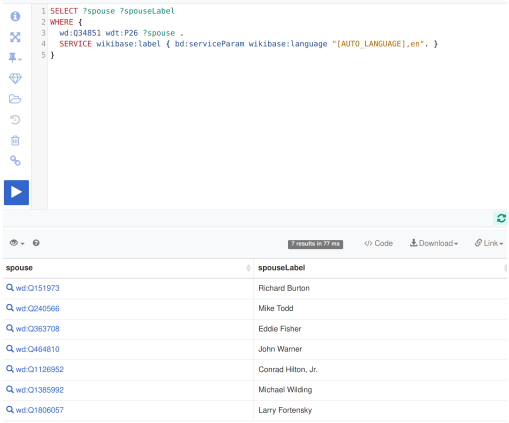
```
SELECT ?item ?itemLabel
WHERE {
  ?item wdt:P31/wdt:P279* wd:Q467.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

11 results in 339 ms Code Download Link

item	itemLabel
wd:Q4233702	Queen Islanzadi
wd:Q11687745	The Queen of the Night
wd:Q2742701	Catherine Tramell
wd:Q40769025	Druk Gyaltsuen
wd:Q67186693	Queen consort of Greece
wd:Q94086863	Sisalda
wd:Q913373	Sailor Saturn
wd:Q105825481	Lumpy Space Princess

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
  - Who are the spouses of Elizabeth Taylor?



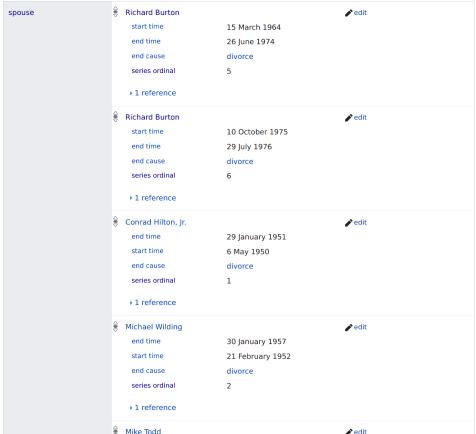
```
SELECT ?spouse ?spouseLabel
WHERE {
  wd:Q34851 wdt:P26 ?spouse .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

7 results in 77 ms | Code | Download | Link

spouse	spouseLabel
wd:Q151973	Richard Burton
wd:Q240566	Mike Todd
wd:Q363708	Eddie Fisher
wd:Q464810	John Warner
wd:Q1126952	Conrad Hilton, Jr.
wd:Q1385992	Michael Wilding
wd:Q1806057	Larry Fortensky

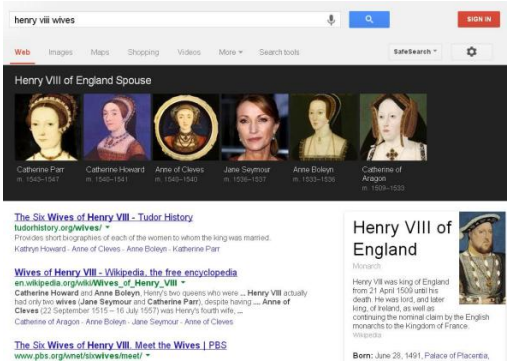
Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
  - What context (qualifiers) needs to be taken into account is unclear



Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
- Few sanity checks are made
  A picture of a wife of Henry VII? (from Google)



What is Knowledge?

Adapted from Russell Ackoff, "From Data to Wisdom" (1989).

- Data: Simple findings and figures
- Information: Collected and analyzed data
- Knowledge: Information organized and ready to put to use
- Wisdom: Proper use of knowledge

See also T. S. Eliot, *The Rock* (1934).

Does Wikidata Have Knowledge

Using Wikidata requires knowledge that is not in the data

- Humans can (often? sometimes?) determine what is needed
- Programs generally can't

So information in Wikidata isn't ready to use.
Thus Wikidata doesn't really have knowledge.

Turning Wikidata Into Wiki-knowledge

- Add in consequences of data (but what, exactly are they?)
  - Design a logic for Wikidata to define consequences
  - Implement a reasoner for the logic
- Incorporate information from descriptions
  - By turning them into expressions in the logic
- Take context into account
  - By providing simple ways to combine qualifiers in the logic
  - By augmenting query interfaces with a current context
    not considered further here

MARS, MARPL, eMARS, eMARPL

- A family of logics for Knowledge Graphs
  - Augment regular logic with qualifiers on facts
  - Add Wikidata datatypes
  - Add rules for handling qualifiers
  - Add constraints
- MARPL and eMARPL can be implemented using forward-chaining rules
  - Complexity of inference is not too bad

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata

$$\text{instance\_of}(\text{Elizabeth\_Taylor}, \text{human})$$

$$\text{spouse}(\text{Elizabeth\_Taylor}, \text{Richard\_Burton})@\{\text{start\_time} : \text{'15 March 1964'}, \ldots\}$$

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
  - Not absolutely necessary, but makes the process easier

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties

$$\text{subclass\_of}(c, d) \wedge \text{subclass\_of}(d, e) \rightarrow \text{subclass\_of}(c, e)$$

$$\text{instance\_of}(y, c)@X \wedge \text{subclass\_of}(c, d) \rightarrow \text{instance\_of}(y, d)@X$$

$$\text{subproperty\_of}(c, d) \wedge \text{subproperty\_of}(d, e) \rightarrow \text{subproperty\_of}(c, e)$$

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties
- Add rules that determine consequences of domain properties

$$\text{spouse}(x, y)@X \rightarrow \text{spouse}(y, x)@X$$

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties
- Add rules that determine consequences of domain properties
- Convert Wikidata constraints to logic constraints

$$\text{property\_constraint}(p, \text{distinct\_values\_constraint}) \wedge p(s1, o1) \wedge p(s2, o2) \wedge s1 \neq s2 \rightarrow o1 \neq o2$$

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers

$$\text{start\_time} : \mathit{max}(\mathit{start\_time})$$

$$\text{end\_time} : \mathit{min}(\mathit{end\_time})$$

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers

$$\text{start\_time} : \mathit{max}(\mathit{start\_time})$$

$$\text{end\_time} : \mathit{min}(\mathit{end\_time})$$

Difficult to determine desired meaning of some combinations

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers
- Provide recognition rules for many classes

$$\text{instance\_of}(\text{human}, h) \wedge \text{sex\_or\_gender}(h, \text{female}) \rightarrow \text{female\_human}(h)$$

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers
- Provide recognition rules for many classes

$$\text{instance\_of}(\text{human}, h) \wedge \text{sex\_or\_gender}(h, \text{female}) \rightarrow \text{female\_human}(h)$$

Very many such classes, some may not have good recognition conditions

Status (Implementation)

- Just a proposal . . .
- There are implementations of property graphs that can store Wikidata statements
- Some graph databases have rules, e.g., Ontotext GraphDB
- No system implements rules as described above.
- Considerable work would be required to implement the logic here
- But it's just forward chaining rules
  - Could implement by encoding in RDF and using RDF rule engine

Status (Community)

- Wikidata community is fragmented
- Portions of Wikidata community not friendly to this kind of effort
  - Seem to want to put more work on consumers of Wikidata
- Support group for Wikidata (Wikimedia Deutschland) is resource constrained

Summary

- Wikidata is a useful knowledge graph
- Like other knowledge graphs, Wikidata doesn't quite qualify as providing knowledge
- A formal account can be given for (most of) Wikidata that adds the missing pieces
  - Needs implementation
  - Not too much pain to implement
  - Computational aspects not horrible
  - Consumers see knowledge, much less pain
- Buy-in for this approach is missing

References

Peter F. Patel-Schneider and David Martin. Wikidata on MARS. DL Workshop, 2020.

David Martin and Peter F. Patel-Schneider. Wikidata Constraints on MARS. Wikidata workshop at ISWC 2020.

Marx, M., Krötzsch, M., Thost, V. Logic on MARS: Ontologies for generalised property graphs. IJCAI 2017.

Krötzsch, M., Marx, M., Ozaki, A., Thost, V. Attributed description logics: Ontologies for knowledge graphs. ISWC 2017.

Krötzsch, M., Marx, M., Ozaki, A., Thost, V. Attributed description logics: Reasoning on knowledge graphs. IJCAI 2018.

R.L. Ackoff. From data to wisdom. *Journal of applied systems analysis*, 1989. https://pdfs.semanticscholar.org/ea36/744970968bc18f9cd89592a86c6d2a366ba7.pdf

T. S. Eliot. *The Rock*. Harcourt, Brace, and Company, 1934. https://books.google.com/books?id=Qmr8AgAAQBAJ