Wikidata Challenges

In the Semantic Web Community

Andrea Westerinen
Graph Consultant

Lydia Pintscher
Portfolio Lead for Wikidata

DATA MODELLING DAYS

November 30
December 2
2023

WIKIDATA

Discussion Topics

Session explores the challenges of using Wikidata in Semantic Web and Linked Data applications by:

- Examining use cases
- Discussing model design
- Proposing a top-level, conceptual model to address challenges

Use Cases

Semantic Web and Linked Data Usage, 1

To expand and extend research and business models and knowledge graphs (both internal and open-source data models) with Wikidata details, at scale (e.g., *broad*, not specific, queries and usage, *created and interpreted programmatically*)

- Synonyms, alternate names, … improve natural language parse and encoding
- Identifiers from other datasets aid in data fusion and integration
- Relationships to other Wikidata classes and instances expand existing models
- Historical and cultural context
- And more

*** Much of this list is based on personal experience**

Semantic Web and Linked Data Usage, 2

To perform reasoning, inference and semantic similarity search for:

- Knowledge expansion and discovery
- Data validation and consistency/correctness analysis

To improve LLM results by:

- Creating training data and fine-tuning an LLM
- Defining prompts and prompt chains
- Validating the information returned by an LLM
    - Removing inaccuracies/"hallucinations"

Challenges, 1

See https://www.wikidata.org/wiki/Wikidata_talk:Ontology_issues_prioritization

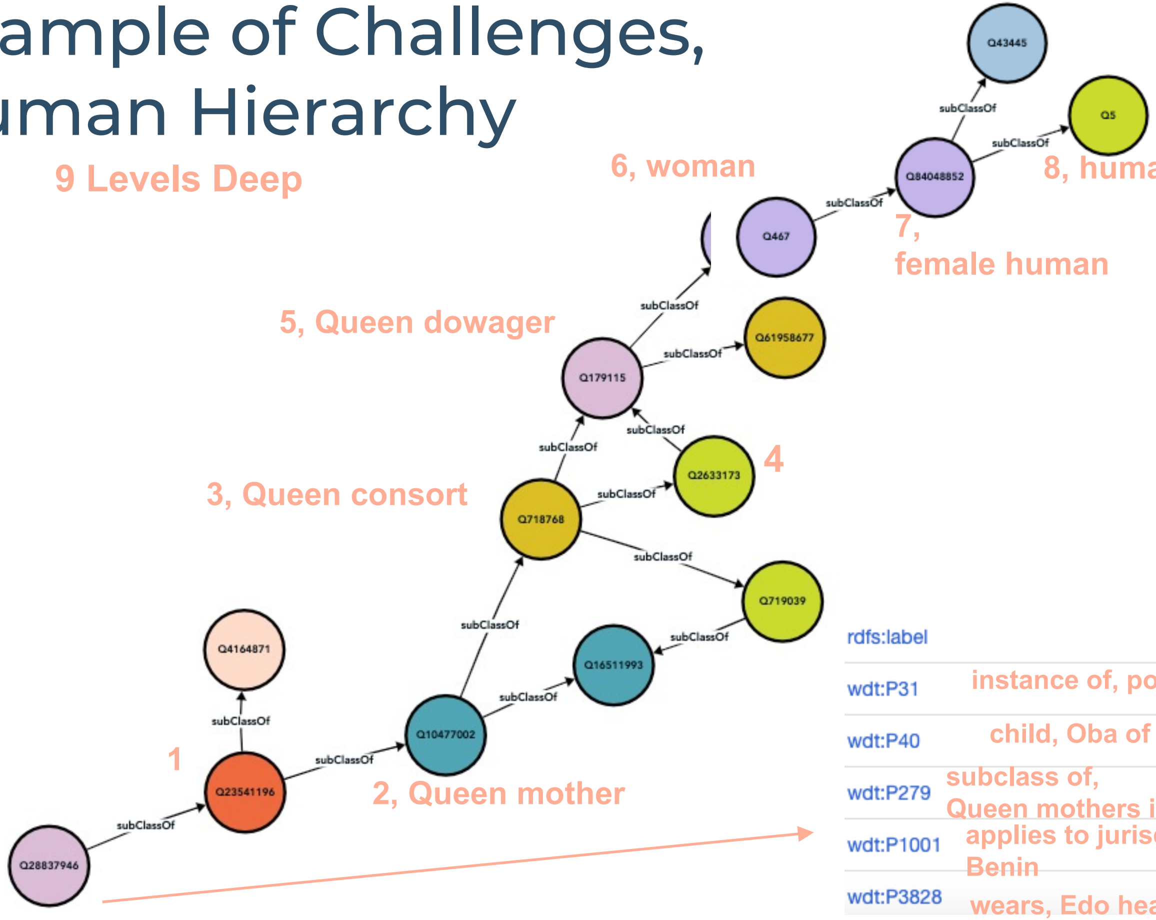
- Overlap and ambiguity of concepts, coupled with inconsistent use of subclass of (P279) and instance of (P31)
    - Example: Scientist as a profession vs a person who is a scientist at some time in their career => Scientist Q901 is both a subclass of person Q215627 and an instance of profession Q28640 (Also note that a person Q215627 is not a subclass of human Q5)
    - Ambiguities: role Q214339 (social role with rights, obligations, ...) is subclass of role Q4897819 (identity relative to another entity), and locations can be geographical location Q2221906, location Q115095765, geographic region Q82794, physical location Q17334923, and geographical area Q3622002
    - Property example: Author P50, Composer P86, Creator P170, Developer P178

Challenges, 2

- Inconsistent level of detail and overspecialization of hierarchy
  - Human (Q5) subclass max depth 9
  - Geographic location (Q2221906) recursive at depth of 16 (Q108800505 subClassOf Q110571169, which is subClassOf Q108800505)
  - Tool (Q39546) recursive at depth of 22
- Structural issues
  - Example: A – subclass of - B, B – subclass of - C, C – subclass of – A (A->B->C->A)
  - Insufficient use of P1647 (subproperty of) – Only 898 triples use wdt:P1647

Example of Challenges, Human Hierarchy

9 Levels Deep



1

2, Queen mother

3, Queen consort

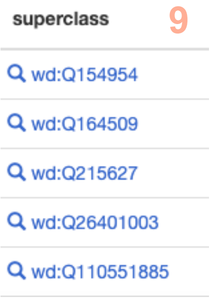
4

5, Queen dowager

6, woman

7, female human

8, human



superclass	9
wd:Q154954	
wd:Q164509	
wd:Q215627	
wd:Q26401003	
wd:Q110551885	

rdfs:label		Iyoba
wdt:P31	instance of, position	wd:Q4164871
wdt:P40	child, Oba of Benin	wd:Q887176
wdt:P279	subclass of, Queen mothers in Africa	wd:Q23541196
wdt:P1001	applies to jurisdiction, Benin	wd:Q171203
wdt:P3828	wears, Edo headdress	wd:Q28837871

Challenges, 3

- Number of properties (11, 324 as of 12 November 2023) and abstract identifier naming
    - 8300+ properties are identifiers (with no subproperty hierarchy) => ~3000 properties to search

      Property details
    - Example: Researcher interested in using locations in a query (13 related to events, 89 related to geography, with overlap) – P159, headquarters; P20, place of death; P654 direction relative to location; P5248 medical evacuation to; P706 located in/on physical feature; …
    - Are these 'OR'ed together in a queries' property path?
    - How does a researcher know that they found most relevant?

DATA MODELLING DAYS

Challenges, 4

See Wikidata: Making of (2023)

- Data uniformity and coherency are challenges; Need to address how to share semantics – “within Wikidata itself, with other Wikimedia projects, and with the world in general”
  - Intended modeling of data communicated via documentation on wikidata.org, shared SPARQL query patterns, and Entity Schemas in ShEx - ”Informal, vague and ambiguous”
  - “As Krötzsch argued in his ISWC 2022 keynote … some sufficiently formal, unambiguous, and declarative way of sharing intended interpretations is … needed. A variety of powerful knowledge representation languages could be used for this purpose, but we still lack both infrastructure and best practices”
- Need support for editors to maintain data quality and increase coherence

Challenges, 5

- Autoregressive Entity Retrieval (2021)
    - “Entities are at the center of how we represent and aggregate knowledge … [Need] components to detect and disambiguate entity mentions in open text, in order to isolate relevant concepts from non-meaningful data”
- Relevant for information extraction – Creating “structured data” from text
    - Relating the entities, locations, times, mechanisms, etc. that are discussed
    - “… consider a group of newswire articles on Latin American terrorism …. define for any given IE task a template … [which, in this case] would have slots corresponding to the perpetrator, victim, and weapon of the terrorist act, and the date on which the event happened”
    - Understanding different terrorist groups, types of weapons and the properties that relate these (as found in Wikidata) improves extraction
    - Require well-documented entity hierarchies and basic properties for classification

Example Issues in Information Extraction

- Complicated and inconsistent class hierarchy
  - Example: some fictional subclasses defined (such as fictional human Q15632617, fictional city Q1964689, …) VERSUS use of a property (e.g., from narrative universe P1080) VERSUS some “entity” subclass (such as imaginary Q2518716 or non-existent entity Q64728693)
- Conflation of concepts such as organizations and locations
  - Example: worldwide organizations are more than their headquarters, and not all businesses are synonymous with location; Problem worsened with the complexity of the organization Q43229 hierarchy tree (max depth = 50+ and instance tree = 137K+ entries)
- Unnecessary creation of discrete subclasses
  - Example: big city Q1549591 is subclass of city Q515, labelled as a “city with a population of more than 100,000 inhabitants”) BUT querying a property could suffice (population P1082)

Modeling Classes and Instances

DATA MODELLING DAYS

November 30
December 2
2023

WIKIDATA

Thinking about the Overall Model, 1

Important to consider the implications of a modeling choice

- Is it based on aspect or perspective? If yes, a subclass is likely problematic
    - Consider a candy bar – when sharing with friends, size matters most; when dealing with an allergy, ingredients matter most
    - Consider time – if thinking in geological time, a minute is nothing, but if thinking in computer time, it is an eternity (and, your Blazegraph query times out)
- Does another domain already model these types of relations, and how do they do it?
    - Create “design patterns” (building on EntitySchemas)

Components of a Modeling “Design Pattern”

- Overview with a short description of the “pattern”, what it is used for *and* not used for
- Diagram or table showing the main concepts and relationships/properties
    - Coupled with the Entity Schema ShEx definition of mandatory/optional properties, and expectation of single or multiple values
- Examples (at least 2 usage scenarios) with queries
- Instructions on extending the pattern
    - Especially when subclassing may versus should not be done

Thinking about the Overall Model, 2

- Is a subclass warranted at this time?
  - 2.7M+ classes defined but 3% are instantiated (how do you know the subclass hierarchy is correct?)

    From Integrating the Wikidata Taxonomy into YAGO (2023)
  - Example: 1M subclasses of chemical entity Q43460564 have no instance
- Should a new subclass be added or might a property suffice? Should a new property be added or could an existing one be reused?
  - Example: Reusing "part of" relation P361 versus P131 (administrative territorial entity), P179 (series), P7938 (electoral district), P8138 (statistical territorial entity), …
  - For properties, could these distinctions be handled by defining the type of entity (in the query) that is related?
- Defining "subproperty of" P1647 where appropriate (currently only used 900 times)

*** Some modeling problems exist due to Wikipedia not using SPARQL queries**

DATA MODELLING DAYS

Example of Query with a Type Restriction

Part of

select ?part where {?part wdt:P361 ?whole . ?whole wdt:P31 wd:Q56061}

- List all *parts of* an administrative division Q56061 (aka administrative territorial entity)
- Query returns 1790 results

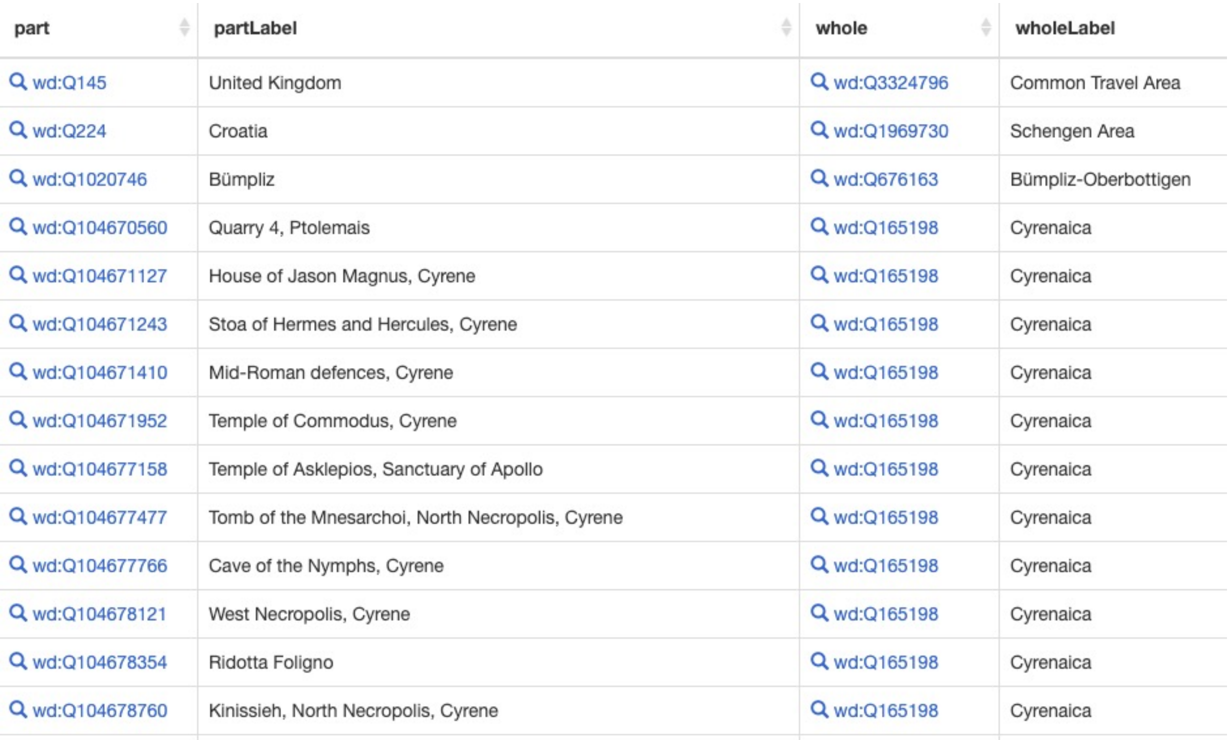
Located in admin territorial entity

select ?part where {?part wdt:P131 ?whole . ?whole wdt:P31 wd:Q56061}

- List all administrative territorial entities that are administrative divisions Q56061
- Query returns 37600 results

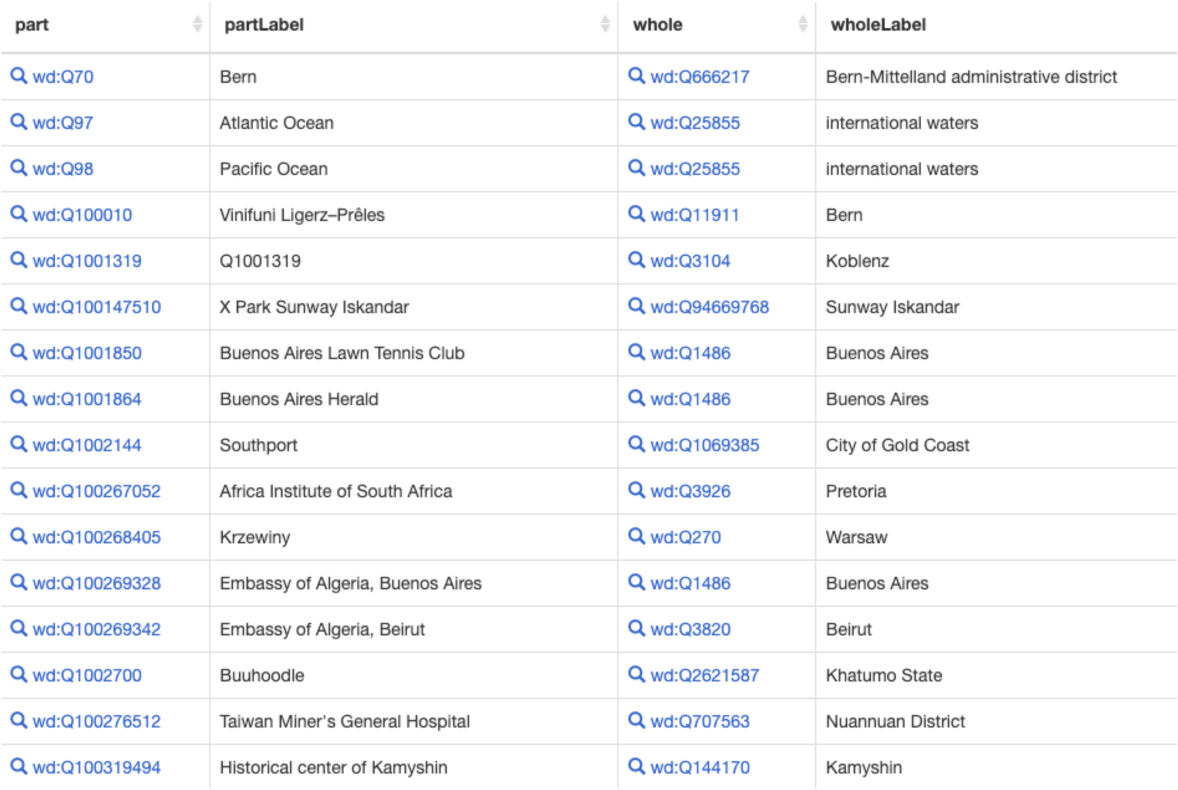
Both results are needed
Only 138 results overlap

Example of wdt:P361 Instances



part	partLabel	whole	wholeLabel
wd:Q145	United Kingdom	wd:Q3324796	Common Travel Area
wd:Q224	Croatia	wd:Q1969730	Schengen Area
wd:Q1020746	Bümpliz	wd:Q676163	Bümpliz-Oberbottigen
wd:Q104670560	Quarry 4, Ptolemais	wd:Q165198	Cyrenaica
wd:Q104671127	House of Jason Magnus, Cyrene	wd:Q165198	Cyrenaica
wd:Q104671243	Stoa of Hermes and Hercules, Cyrene	wd:Q165198	Cyrenaica
wd:Q104671410	Mid-Roman defences, Cyrene	wd:Q165198	Cyrenaica
wd:Q104671952	Temple of Commodus, Cyrene	wd:Q165198	Cyrenaica
wd:Q104677158	Temple of Asklepios, Sanctuary of Apollo	wd:Q165198	Cyrenaica
wd:Q104677477	Tomb of the Mnesarchoi, North Necropolis, Cyrene	wd:Q165198	Cyrenaica
wd:Q104677766	Cave of the Nymphs, Cyrene	wd:Q165198	Cyrenaica
wd:Q104678121	West Necropolis, Cyrene	wd:Q165198	Cyrenaica
wd:Q104678354	Ridotta Foligno	wd:Q165198	Cyrenaica
wd:Q104678760	Kinissieh, North Necropolis, Cyrene	wd:Q165198	Cyrenaica

Example of wdt:P131 Instances



part	partLabel	whole	wholeLabel
wd:Q70	Bern	wd:Q666217	Bern-Mittelland administrative district
wd:Q97	Atlantic Ocean	wd:Q25855	international waters
wd:Q98	Pacific Ocean	wd:Q25855	international waters
wd:Q100010	Vinifuni Ligerz–Prêles	wd:Q11911	Bern
wd:Q1001319	Q1001319	wd:Q3104	Koblenz
wd:Q100147510	X Park Sunway Iskandar	wd:Q94669768	Sunway Iskandar
wd:Q1001850	Buenos Aires Lawn Tennis Club	wd:Q1486	Buenos Aires
wd:Q1001864	Buenos Aires Herald	wd:Q1486	Buenos Aires
wd:Q1002144	Southport	wd:Q1069385	City of Gold Coast
wd:Q100267052	Africa Institute of South Africa	wd:Q3926	Pretoria
wd:Q100268405	Krzewiny	wd:Q270	Warsaw
wd:Q100269328	Embassy of Algeria, Buenos Aires	wd:Q1486	Buenos Aires
wd:Q100269342	Embassy of Algeria, Beirut	wd:Q3820	Beirut
wd:Q1002700	Buuhoodle	wd:Q2621587	Khatumo State
wd:Q100276512	Taiwan Miner's General Hospital	wd:Q707563	Nuannuan District
wd:Q100319494	Historical center of Kamyshin	wd:Q144170	Kamyshin

Using "Systems Thinking"

First consider the big picture and then drill into specifics

- Basic concepts
    - Form and function
    - Entities and relationships
    - Boundaries between entities
    - Wholes vs parts (whole is more than its parts; Parts are more than a piece of the whole)

”Systems Thinking” Design Questions, 1

- What concepts are within and what are external to the system? (What are the *boundaries* of the system?)
- Can a hierarchy of concepts be created?
- Does the system need to be decomposed into ”component sub-systems”? And, how do those components interact?
  - “Interactions” = behavior and events
- How are the systems and components similar to or different from each other?

”Systems Thinking” Design Questions, 2

- How are the systems and components related to each other?
- What is constant (invariant and stable) about a system and its parts?
  - What changes and how is this tracked over time?
  - What properties are used for identity?
  - What is "typical" about the system? What is "exceptional" (but important) about the system?
  - What properties define the system?
- Can we look at a system/problem differently to gain insights?

Top-Level Concepts and Properties in Wikidata

DATA MODELLING DAYS

November 30
December 2
2023

WIKIDATA

Wikidata’s Current “Top-Level”

Building on what exists today

- Recommend building on concepts that are already used for instances (versus trying to incorporate philosophy-based ontologies)
- Current Wikidata ”ontological” hierarchy deep (1099 items subclass from “entity” Q35120, WikiProject Ontology/Top-level ontology list) but mostly un-used for instances
    - Unexpected, overlapping, incompletely defined (example: “something” subclasses from “entity” and is different than “being”, “copy” and “part”)
    - Combination of very broad and very specific concepts (example: “hammock hook” subclasses “object”, which subclasses “entity”)
- Also should include a property hierarchy

What Does a Consistent ”Top-Level” Achieve?

Defines hierarchy to guide and validate new classes and properties, and allows creation of repeatable “design patterns”

- Clarity of definition, coupled with structural and semantic similarity measurements, can reduce or help to align overlapping concepts
  - Also improves ability to query “broadly” across concepts and properties
- Can exist alongside the current Wikidata declarations
  - Can immediately be used for new concept and property definitions
- Reasoning and consistency tools can locate inconsistencies and errors

Sources for Defining “Top-Level” Concepts

- Wikipedia subjects
- Wikidata subgraph hierarchies
- Other sources:
    - Previous effort on gitHub (discontinued 2017) to map to schema.org
    - DBPedia Ontology (DBO)
    - YAGO (also maps to schema.org) with well-defined mapping to/from Wikidata described in Integrating the Wikidata Taxonomy into YAGO

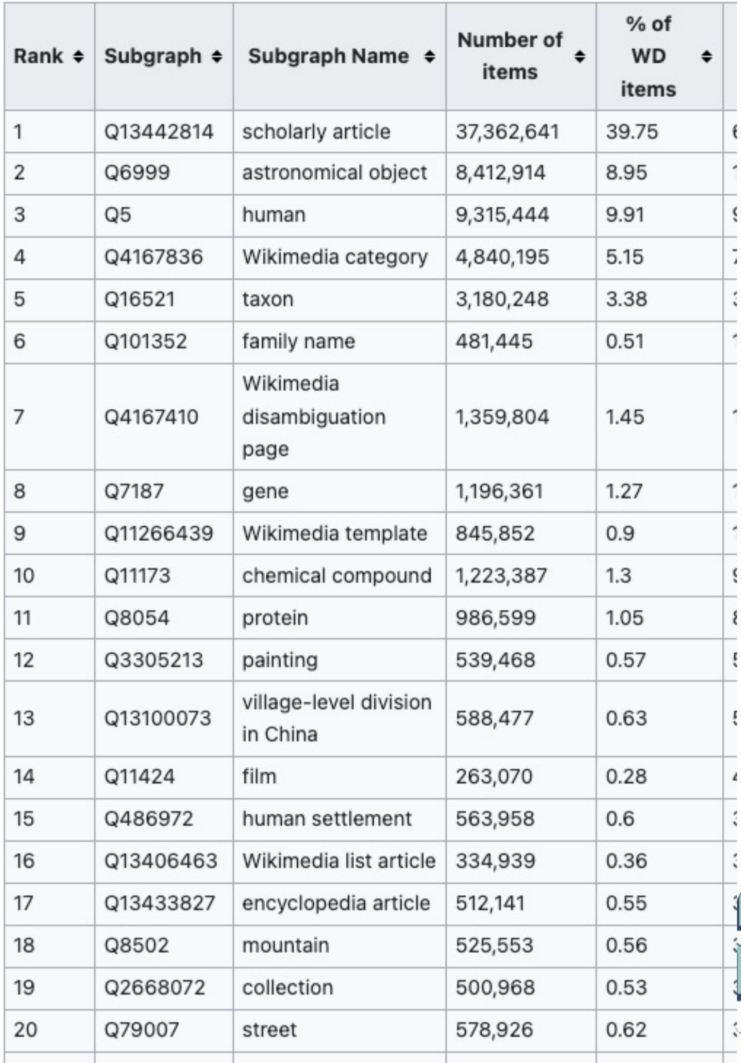
Definition Using Wikipedia Subjects

https://en.wikipedia.org/wiki/Wikipedia:Contents

- Culture
- Geography
- Health
- History and philosophy
- People and human activities
- Math, science and technology
- Religion and society
- Reference works

Definition Using Wikidata Subgraphs

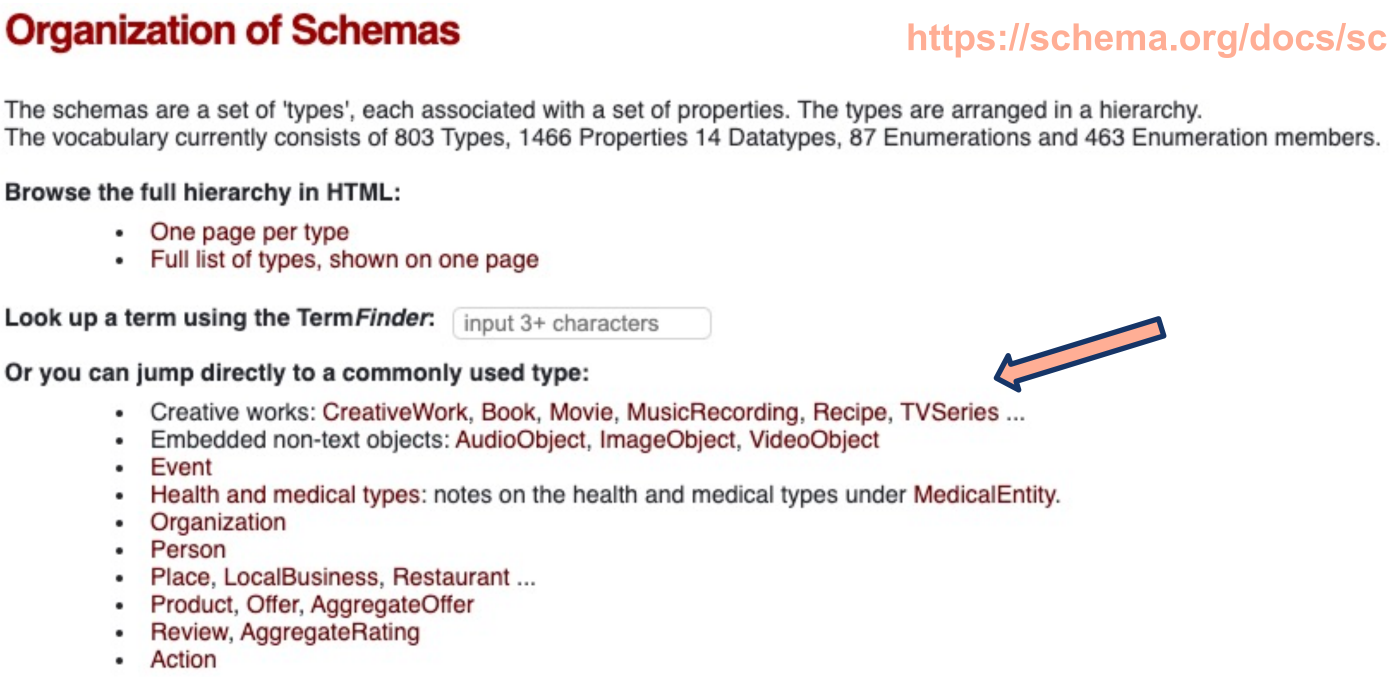
https://wikitech.wikimedia.org/wiki/User:AKhatun/Wikidata_Subgraph_Analysis



Rank	Subgraph	Subgraph Name	Number of items	% of WD items
1	Q13442814	scholarly article	37,362,641	39.75
2	Q6999	astronomical object	8,412,914	8.95
3	Q5	human	9,315,444	9.91
4	Q4167836	Wikimedia category	4,840,195	5.15
5	Q16521	taxon	3,180,248	3.38
6	Q101352	family name	481,445	0.51
7	Q4167410	Wikimedia disambiguation page	1,359,804	1.45
8	Q7187	gene	1,196,361	1.27
9	Q11266439	Wikimedia template	845,852	0.9
10	Q11173	chemical compound	1,223,387	1.3
11	Q8054	protein	986,599	1.05
12	Q3305213	painting	539,468	0.57
13	Q13100073	village-level division in China	588,477	0.63
14	Q11424	film	263,070	0.28
15	Q486972	human settlement	563,958	0.6
16	Q13406463	Wikimedia list article	334,939	0.36
17	Q13433827	encyclopedia article	512,141	0.55
18	Q8502	mountain	525,553	0.56
19	Q2668072	collection	500,968	0.53
20	Q79007	street	578,926	0.62

Definition Using schema.org

https://schema.org/docs/schemas.html



Organization of Schemas

The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy.
The vocabulary currently consists of 803 Types, 1466 Properties 14 Datatypes, 87 Enumerations and 463 Enumeration members.

Browse the full hierarchy in HTML:

- One page per type
- Full list of types, shown on one page

Look up a term using the Term*Finder*: input 3+ characters

Or you can jump directly to a commonly used type:

- Creative works: CreativeWork, Book, Movie, MusicRecording, Recipe, TVSeries ...
- Embedded non-text objects: AudioObject, ImageObject, VideoObject
- Event
- Health and medical types: notes on the health and medical types under MedicalEntity.
- Organization
- Person
- Place, LocalBusiness, Restaurant ...
- Product, Offer, AggregateOffer
- Review, AggregateRating
- Action

Thoughts on a Wikidata “Top-Level”, 1

- Mapping can exist alongside current definitions
- Entity (Q35120) ~ schema.org Thing ~ owl:Thing
    - Physical or legal agent => Human, Organization, … (e.g., combining Human and Legal Person)
    - Physical resource => Food, Tool, …
    - Biochemical entity (=> Taxon, Gene, Protein, Chemical compound, …)
    - Intangible => Language, Structured value (=> Money/currency, …), Culture, Religion, Philosophy, Mathematics, Science and technology (=> Health and medicine, Information processing, …), Occupation, Template, …
    - Creative work => Book, Film, Painting, Scholarly article, Encyclopedia article, …
    - Location => Astronomical entity, Architectural feature, Geographic feature Administrative division, …

Continued

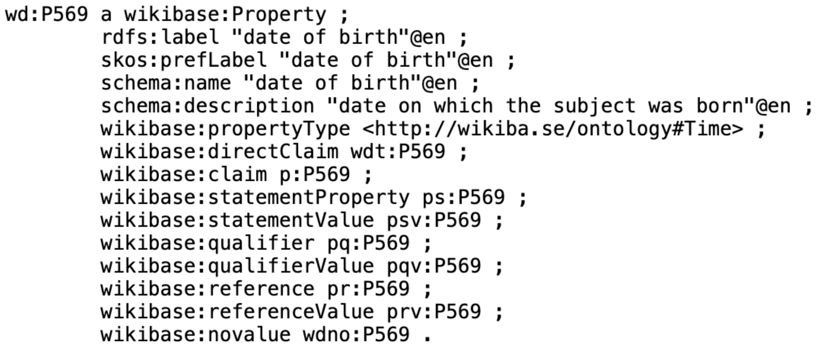
Thoughts on a Wikidata “Top-Level”, 2

- Entity (Q35120) ~ schema.org Thing ~ owl:Thing (continued)
  - Time => Unit of time, Season, …
  - Event/occurrence (including historical) with properties representing the entities involved (e.g., the active, affected or theme entities), location, time, …
  - Multiple inheritance usage => Fictional entity, Collection(=> Series), …
- Map “significant” top-level entities to proposed concepts
- Similarly, map “significant” properties
  - Perhaps starting from User:Rtnf/Properties
  - Include new properties for events to clarify “participant” P710 (versus using a qualifier and “object has role” P3831 property)
- Create “design patterns”/EntitySchemas for each high level concept

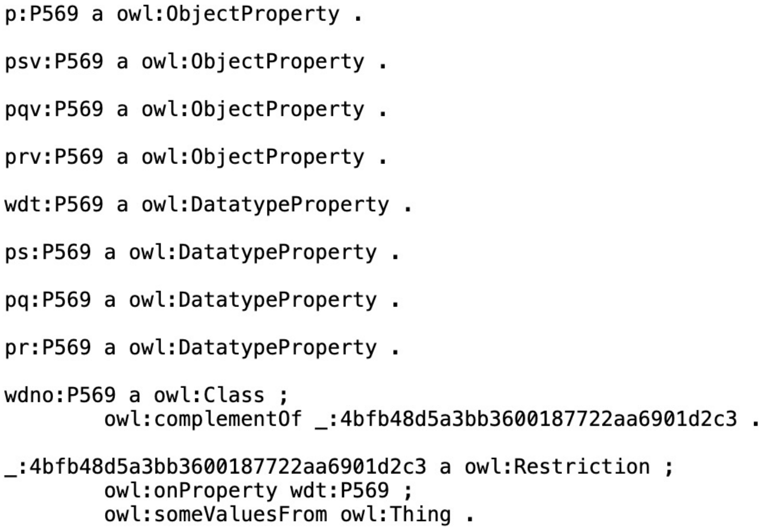
DATA MODELLING DAYS

Thoughts on Wikidata “Top-Level” Properties

- RDF/OWL distinguishes data/scalar and object/reference properties
  - Detail already captured in .ttl (Turtle) output for an entity



```
wd:P569 a wikibase:Property ;
        rdfs:label "date of birth"@en ;
        skos:prefLabel "date of birth"@en ;
        schema:name "date of birth"@en ;
        schema:description "date on which the subject was born"@en ;
        wikibase:propertyType <http://wikiba.se/ontology#Time> ;
        wikibase:directClaim wdt:P569 ;
        wikibase:claim p:P569 ;
        wikibase:statementProperty ps:P569 ;
        wikibase:statementValue psv:P569 ;
        wikibase:qualifier pq:P569 ;
        wikibase:qualifierValue pqv:P569 ;
        wikibase:reference pr:P569 ;
        wikibase:referenceValue prv:P569 ;
        wikibase:novalue wdno:P569 .
```



```
p:P569 a owl:ObjectProperty .

psv:P569 a owl:ObjectProperty .

pqv:P569 a owl:ObjectProperty .

prv:P569 a owl:ObjectProperty .

wdt:P569 a owl:DatatypeProperty .

ps:P569 a owl:DatatypeProperty .

pq:P569 a owl:DatatypeProperty .

pr:P569 a owl:DatatypeProperty .

wdno:P569 a owl:Class ;
        owl:complementOf _:4bfb48d5a3bb3600187722aa6901d2c3 .

_:4bfb48d5a3bb3600187722aa6901d2c3 a owl:Restriction ;
        owl:onProperty wdt:P569 ;
        owl:someValuesFrom owl:Thing .
```

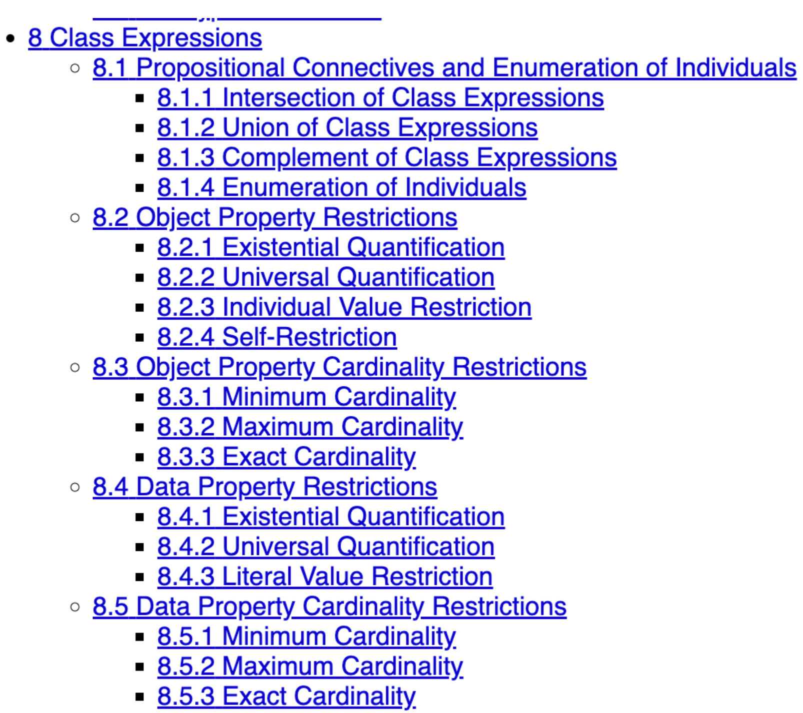
From (for example)
https://www.wikidata.org/wiki/Special:EntityData/Q42.ttl

More Formal Semantics and Constraints, 1

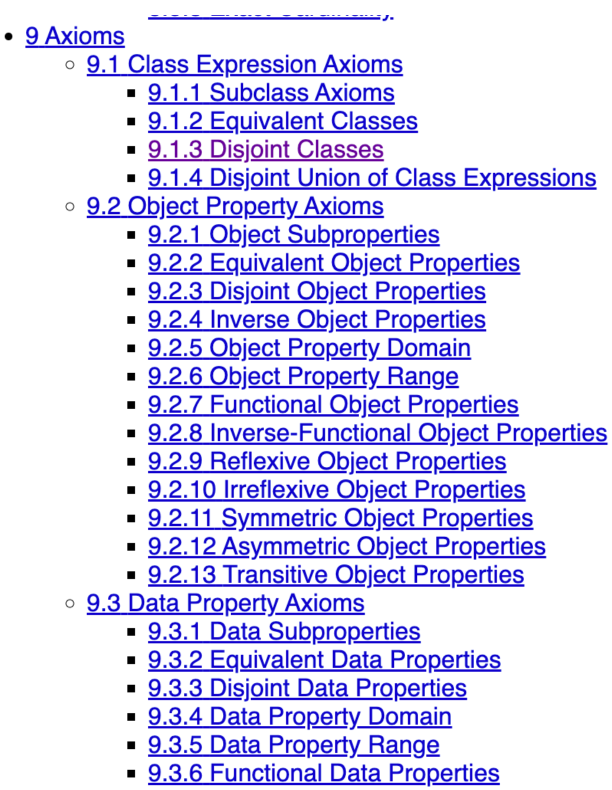
For classes and properties

- Can be formally defined in RDF/OWL via specific constructs
  - Single-valued properties defined as owl:FunctionalProperty (e.g., for identifiers); Inverse properties defined as owl:inverseOf; Properties that can be applied from both directions defined as owl:SymmetricProperty (e.g., A knows B means that B knows A); ...
  - Disjoint-ness of concepts – For example: Physical or legal agent – disjoint with – Intangible, Creative work, ...; Location – disjoint with – Time
- And/or using restrictions

More Formal Semantics and Constraints, 2



- 8 Class Expressions
  - 8.1 Propositional Connectives and Enumeration of Individuals
    - 8.1.1 Intersection of Class Expressions
    - 8.1.2 Union of Class Expressions
    - 8.1.3 Complement of Class Expressions
    - 8.1.4 Enumeration of Individuals
  - 8.2 Object Property Restrictions
    - 8.2.1 Existential Quantification
    - 8.2.2 Universal Quantification
    - 8.2.3 Individual Value Restriction
    - 8.2.4 Self-Restriction
  - 8.3 Object Property Cardinality Restrictions
    - 8.3.1 Minimum Cardinality
    - 8.3.2 Maximum Cardinality
    - 8.3.3 Exact Cardinality
  - 8.4 Data Property Restrictions
    - 8.4.1 Existential Quantification
    - 8.4.2 Universal Quantification
    - 8.4.3 Literal Value Restriction
  - 8.5 Data Property Cardinality Restrictions
    - 8.5.1 Minimum Cardinality
    - 8.5.2 Maximum Cardinality
    - 8.5.3 Exact Cardinality

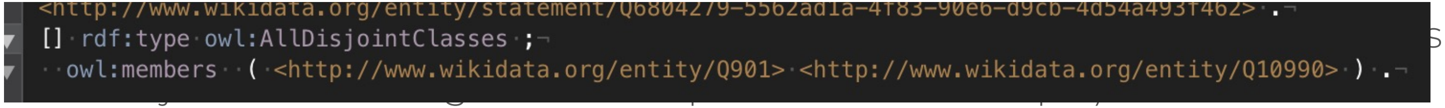


- 9 Axioms
  - 9.1 Class Expression Axioms
    - 9.1.1 Subclass Axioms
    - 9.1.2 Equivalent Classes
    - 9.1.3 Disjoint Classes
    - 9.1.4 Disjoint Union of Class Expressions
  - 9.2 Object Property Axioms
    - 9.2.1 Object Subproperties
    - 9.2.2 Equivalent Object Properties
    - 9.2.3 Disjoint Object Properties
    - 9.2.4 Inverse Object Properties
    - 9.2.5 Object Property Domain
    - 9.2.6 Object Property Range
    - 9.2.7 Functional Object Properties
    - 9.2.8 Inverse-Functional Object Properties
    - 9.2.9 Reflexive Object Properties
    - 9.2.10 Irreflexive Object Properties
    - 9.2.11 Symmetric Object Properties
    - 9.2.12 Asymmetric Object Properties
    - 9.2.13 Transitive Object Properties
  - 9.3 Data Property Axioms
    - 9.3.1 Data Subproperties
    - 9.3.2 Equivalent Data Properties
    - 9.3.3 Disjoint Data Properties
    - 9.3.4 Data Property Domain
    - 9.3.5 Data Property Range
    - 9.3.6 Functional Data Properties

From OWL 2 Web Ontology Language
Structural Specification and Functional-Style Syntax

Inconsistency Checking Example, 1

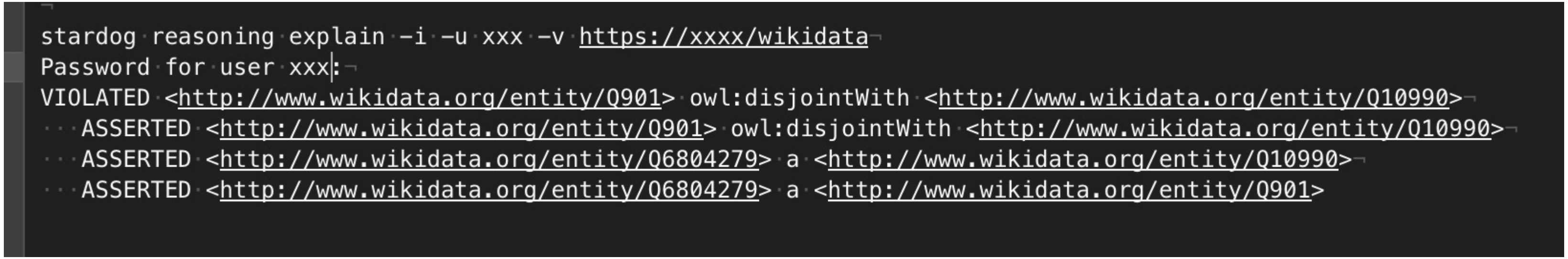
- Meat scientist Q6804279 is instance of Meat Q10990 and Scientist Q901
  - Created based on a ”field of occupation” P425 of meat science
  - Mapping to new concept top-level results in: Meat sub-class of Physical_Resource, Scientist is sub-class of Intangible
- Using WDQS, output SPARQL results for all the triples involving “meat scientist”, “meat” and “scientist”
  - Output as .tsv verbose file and converted via Python to RDF



```
<http://www.wikidata.org/entity/statement/Q6804279-5562ad1a-4f83-90e6-d9cb-4d54a493f462> .
[] rdf:type owl:AllDisjointClasses ;
  owl:members  ( <http://www.wikidata.org/entity/Q901> <http://www.wikidata.org/entity/Q10990> ) .
```

Inconsistency Checking Example, 2

- Conversion defined each superclass as an “owl:Class” and defined P279 as “rdfs:subClassOf” to enable reasoning/consistency check
- Loaded the RDF into Stardog Free Cloud
- Ran a reasoning consistency check with explanation



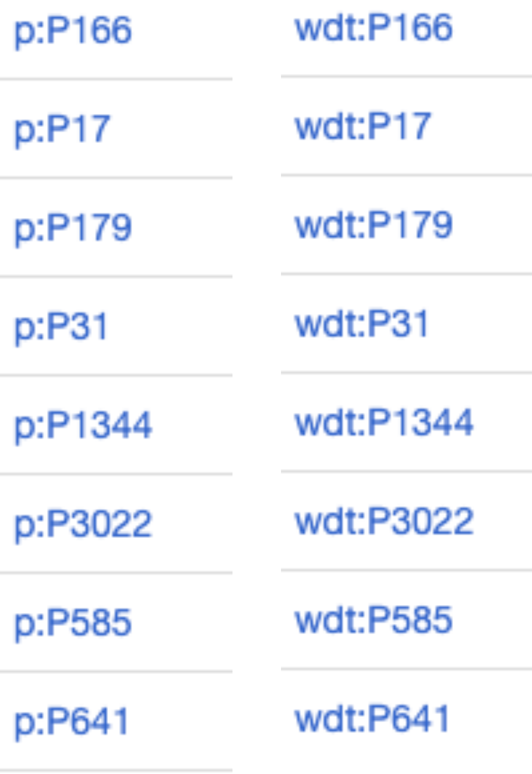
```
stardog reasoning explain -i -u xxx -v https://xxxx/wikidata
Password for user xxx:
VIOLATED <http://www.wikidata.org/entity/Q901> owl:disjointWith <http://www.wikidata.org/entity/Q10990>
   ASSERTED <http://www.wikidata.org/entity/Q901> owl:disjointWith <http://www.wikidata.org/entity/Q10990>
   ASSERTED <http://www.wikidata.org/entity/Q6804279> a <http://www.wikidata.org/entity/Q10990>
   ASSERTED <http://www.wikidata.org/entity/Q6804279> a <http://www.wikidata.org/entity/Q901>
```

What about Instances such as Q143842?

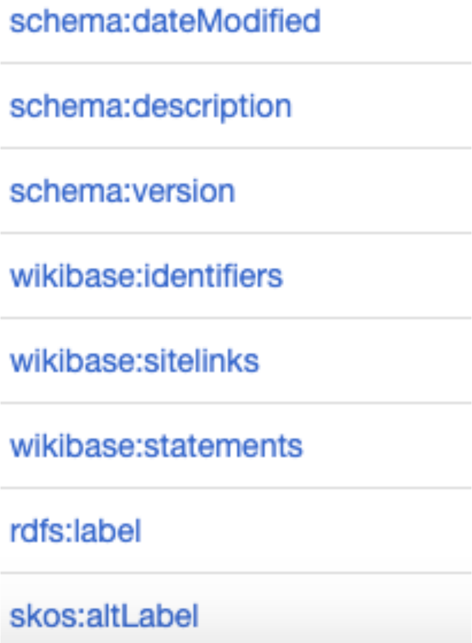
- “Contextual entity”
    - Defines that the specific entity type (for example, an Olympic delegation) makes sense/is relevant in a specific context (in this case, 1968 Summer Olympics)
- “Uganda at the 1968 Summer Olympics”
    - Instance of Olympic delegation Q26213387, a subclass of “physical or legal entity”
    - Add instance of “contextual entity”
    - Context: 1968 Summer Olympics Q8429 already modeled as an “event”
    - Relationship: “participant in” property P1344 exists + new “context” property

*** What else could be problematic?**

Current Properties of Q143842



p:P166	wdt:P166	award received (gold, silver, bronze)
p:P17	wdt:P17	country (Uganda)
p:P179	wdt:P179	part of the series*
p:P31	wdt:P31	instance of (Olympic delegation)
p:P1344	wdt:P1344	**participant in (1968 Summer Olympics)**
p:P3022	wdt:P3022	flag bearer
p:P585	wdt:P585	point in time (1968)
p:P641	wdt:P641	sport (Olympic sport)



schema:dateModified

schema:description

schema:version

wikibase:identifiers

wikibase:sitelinks

wikibase:statements

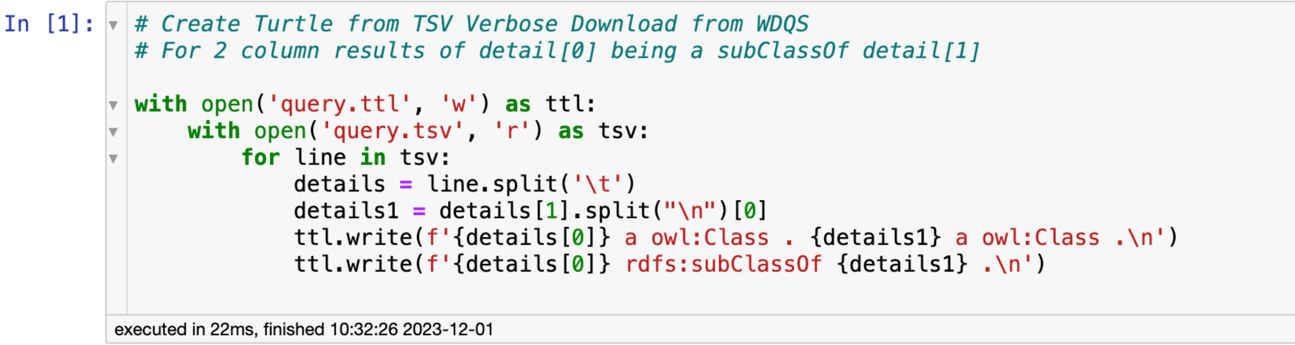
rdfs:label

skos:altLabel

*** Series is “Uganda at the Olympics” Q510152**

Backup

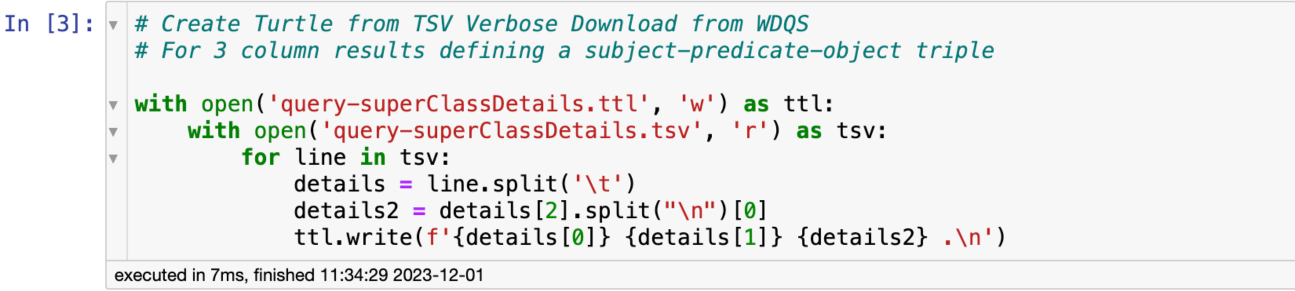
Python Code to Convert TSV to RDF



```
In [1]: # Create Turtle from TSV Verbose Download from WDQS
        # For 2 column results of detail[0] being a subClassOf detail[1]

        with open('query.ttl', 'w') as ttl:
            with open('query.tsv', 'r') as tsv:
                for line in tsv:
                    details = line.split('\t')
                    details1 = details[1].split("\n")[0]
                    ttl.write(f'{details[0]} a owl:Class . {details1} a owl:Class .\n')
                    ttl.write(f'{details[0]} rdfs:subClassOf {details1} .\n')

executed in 22ms, finished 10:32:26 2023-12-01
```



```
In [3]: # Create Turtle from TSV Verbose Download from WDQS
        # For 3 column results defining a subject-predicate-object triple

        with open('query-superClassDetails.ttl', 'w') as ttl:
            with open('query-superClassDetails.tsv', 'r') as tsv:
                for line in tsv:
                    details = line.split('\t')
                    details2 = details[2].split("\n")[0]
                    ttl.write(f'{details[0]} {details[1]} {details2} .\n')

executed in 7ms, finished 11:34:29 2023-12-01
```

Thanks for your attention!

Get in touch with us:

Andrea Westerinen
arwesterinen@gmail.com
https://www.linkedin.com/in/andreawesterinen/

Lydia Pintscher
lydia.pintscher@Wikimedia.de

DATA MODELLING DAYS
November 30
December 2
2023
WIKIDATA