

Ontologi dalam Wikidata

Adila A. Krisnadhi
Fakultas Ilmu Komputer, Universitas Indonesia

Lokakarya *Wikidata for AI* (WD4AI) 2018
Universitas Indonesia

Pengakuan

- Salindia-salindia ini sebagian besar merupakan hasil adaptasi dan terjemahan dari salindia-salindia milik Markus Kroetzsch berjudul “Ontological Modelling in Wikida” yang disajikan dalam paparan kunci di WOP 2018.



Besar
Terhubung
Berbeda
Berguna
Menarik

place of
birth

Parepare

From Wikipedia, the free encyclopedia

Parepare is a city (*kota*) in [South Sulawesi](#), [Indonesia](#), located on the southwest coast of [Sulawesi](#), about 155 km (96 mi) north of the provincial capital of [Makassar](#). A port town, it is one of the major population centers of the [Bugis](#) people. The city has a population of 129,542 people, according to the 2010 Census.

country

Indonesia

From Wikipedia, the free encyclopedia

Indonesia (/ˌɪndəˈniːʒə/ [ⓘ] listen) *IN-də-NEE-zhə*, /-ziə/ *-zee-ə*; Indonesian: [ɪndoneˈsia]), officially the **Republic of Indonesia** (Indonesian: *Republik Indonesia* [rɛpublik ɪndoneˈsia]),^[*lacks stress*] is a sovereign state in Southeast Asia, between the [Indian](#) and [Pacific](#) oceans. It is the world's largest island country, with more than [thirteen thousand islands](#),^[11] and at 1,904,569 square kilometres (735,358 square miles), the 14th largest by land area and the 7th largest in combined sea and land area.^[12] With over 261 million people, it is the world's 4th most populous country as well as the most populous Muslim-majority country. [Java](#), the world's most populous island,^[13] contains more than half of the country's population.

B. J. Habibie

From Wikipedia, the free encyclopedia
(Redirected from [Habibie](#))

"Habibie" redirects here. For other uses, see [Habib](#)

Bacharuddin Jusuf Habibie (Malay: [baxaruddin jusuf habibi] [ⓘ] listen); born 25 June 1936) is an Indonesian engineer who was [President of Indonesia](#) from 1998 to 1999. He succeeded [Suharto](#), who [resigned](#) in 1998. His presidency is seen as a transition to the [post-Suharto era](#). Upon becoming president, he

liberalize
and held
resulted
was the

Personal details

Born	Bacharuddin Jusuf Habibie 25 June 1936 (age 82) Parepare, South Sulawesi, Dutch East Indies (now Indonesia)
-------------	---

Ide awal: Anotasi pranala Wikipedia

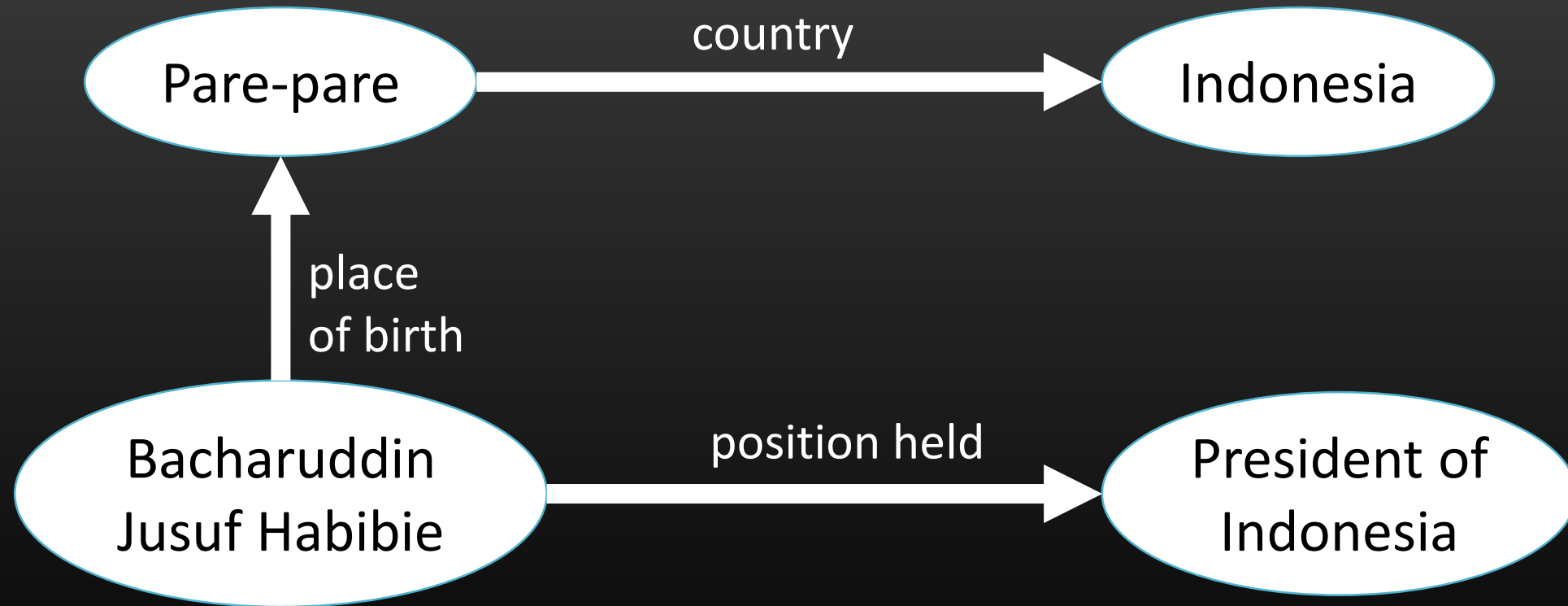
position held

President of Indonesia

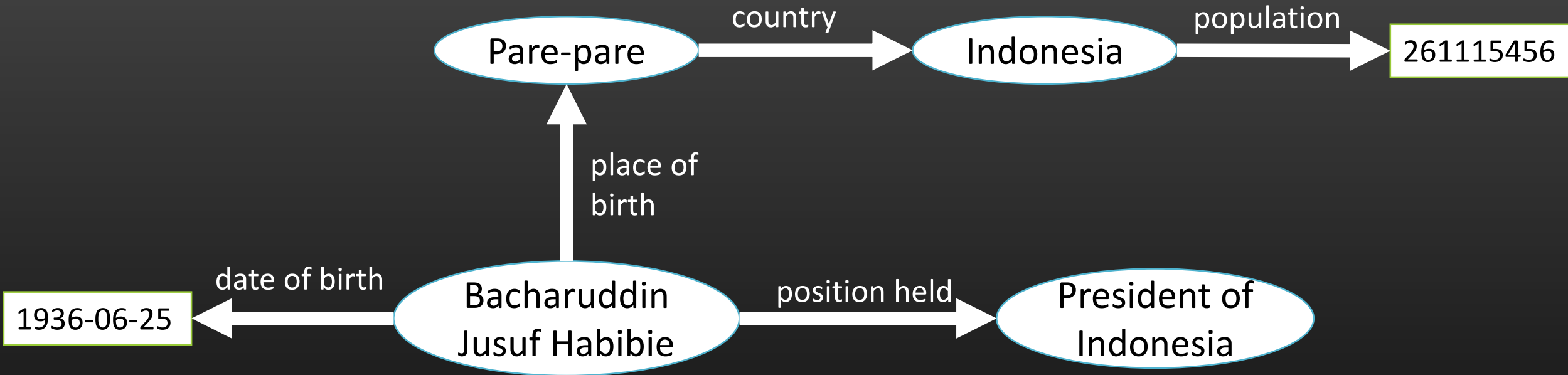
From Wikipedia, the free encyclopedia

The **President of the Republic of Indonesia** (Indonesian: *Presiden Republik Indonesia*) is the head of state and also head of government of the [Republic of Indonesia](#). The president leads the executive branch of the Indonesian government and is the commander-in-chief of the Indonesian National Armed Forces.

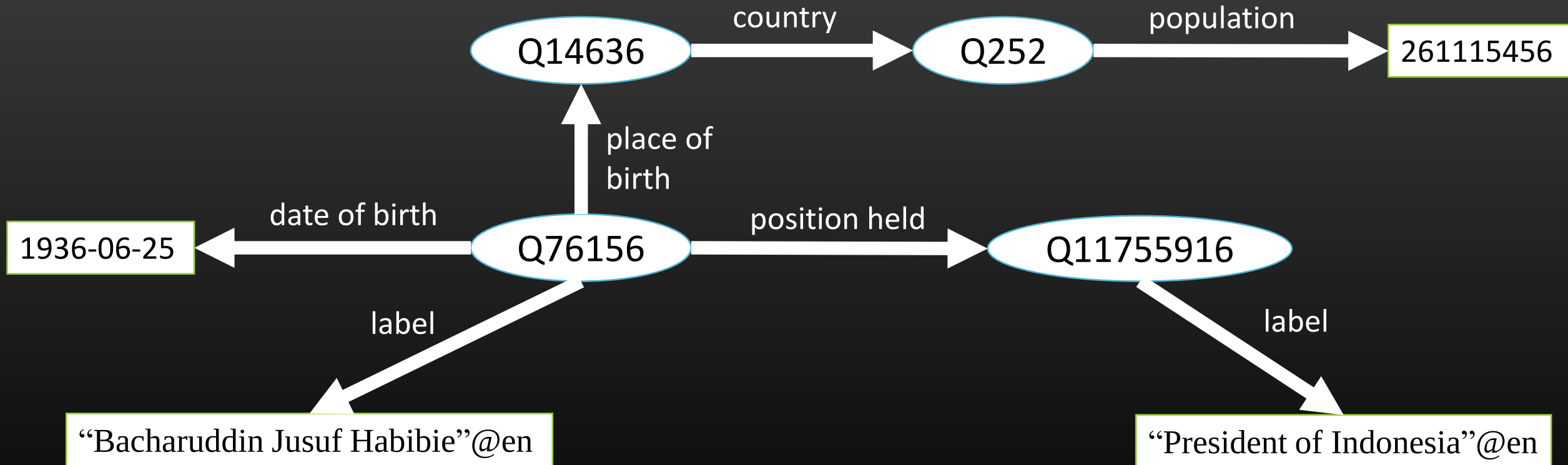
Dari pranala menuju graf (Semantic MediaWiki, 2005)



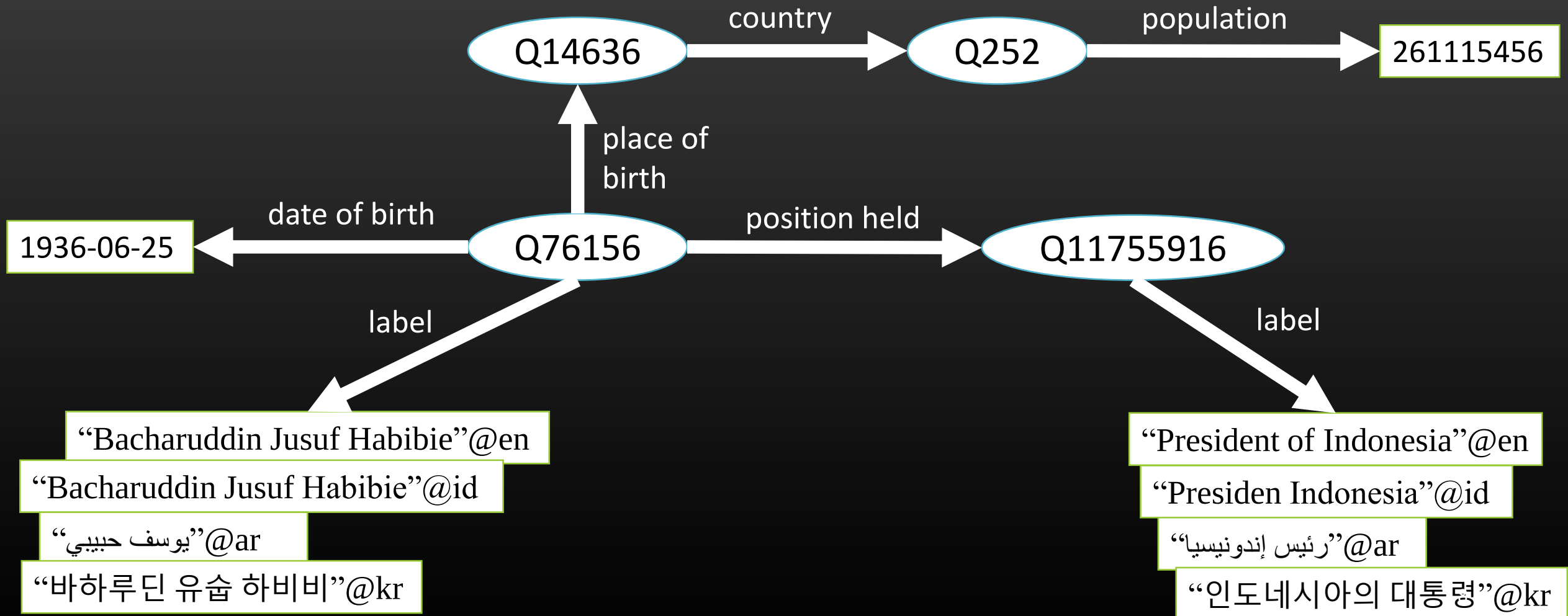
Tambahkan data bertipe primitif



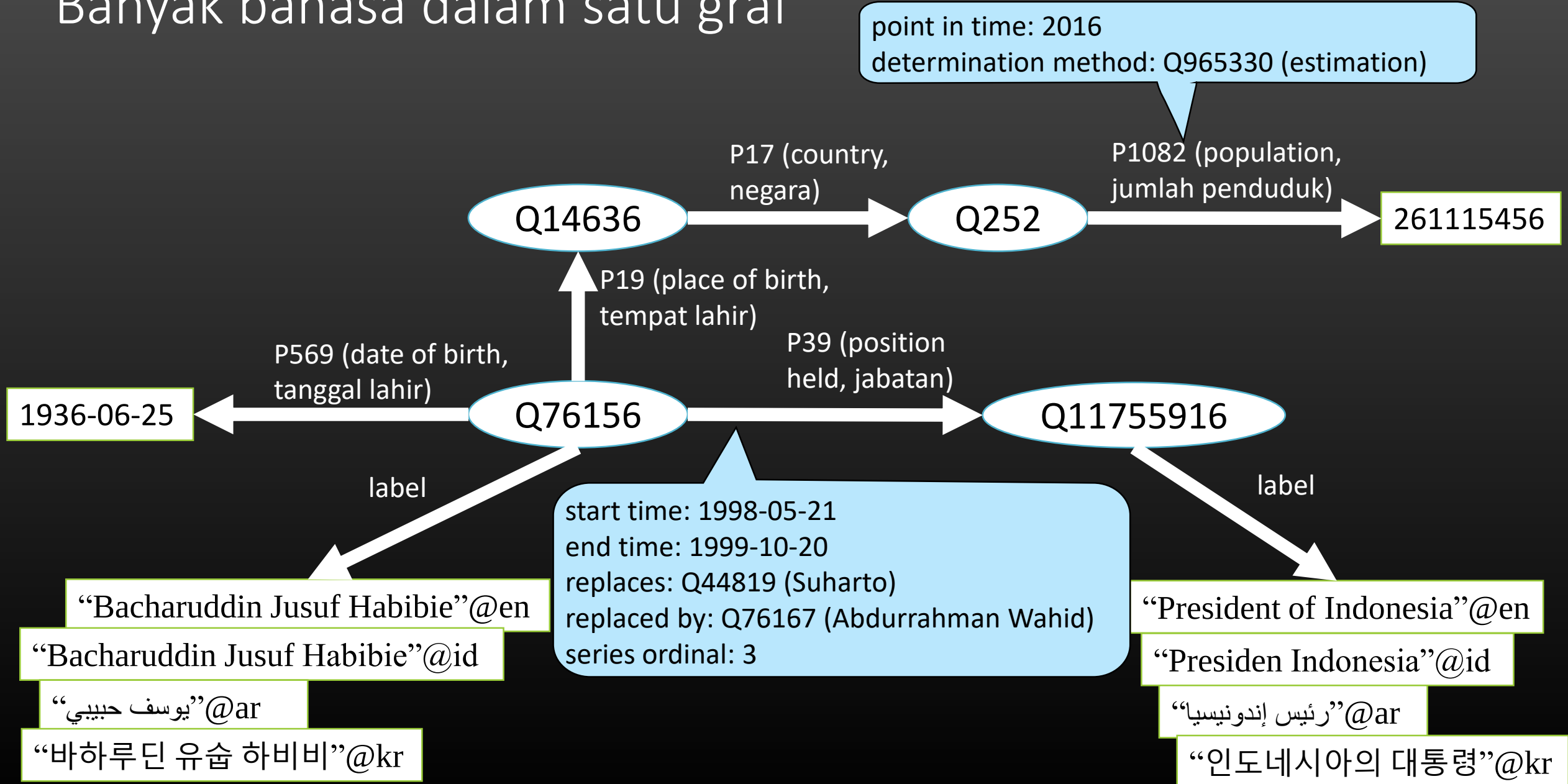
Banyak bahasa dalam satu graf



Banyak bahasa dalam satu graf



Banyak bahasa dalam satu graf



Hasilnya tidak lagi sederhana ...

Bacharuddin Jusuf Habibie (Q76156)

The third President of the Republic of Indonesia

 [edit](#)

BJ. Habibie

position held



President of Indonesia

 [edit](#)

start time

21 May 1998

end time

20 October 1999

replaces

Suharto

replaced by

Abdurrahman Wahid

series ordinal

3

► 1 reference



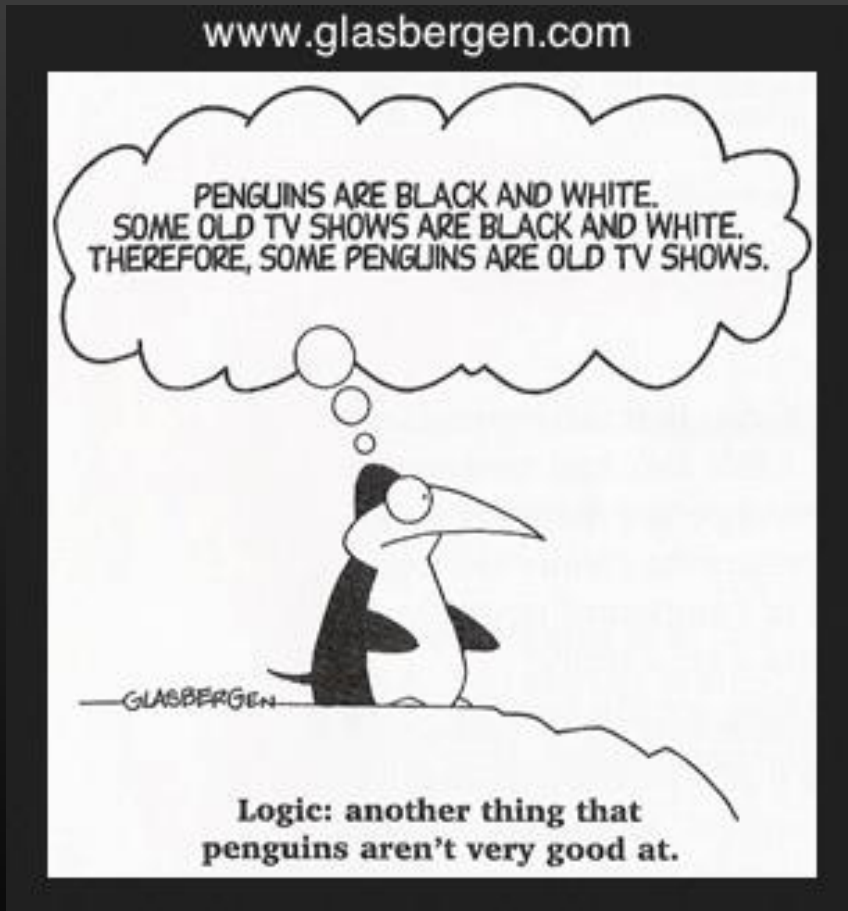
Besar dan terhubung

- >570 juta pernyataan pada >50 juta entitas
- >65 juta pranala ke halaman Wiki(p|m)edia
- >200 juta label dan alias
- >1200 juta deskripsi pemecah ambiguitas
- >200 ribu kontributor terdaftar, +- 19 ribu aktif

Pengelolaan data akan terbantu jika ada semacam prinsip sistematis yang berlaku

Ontologi: Pemodelan Konseptual di Wikidata

Apa itu ontologi?



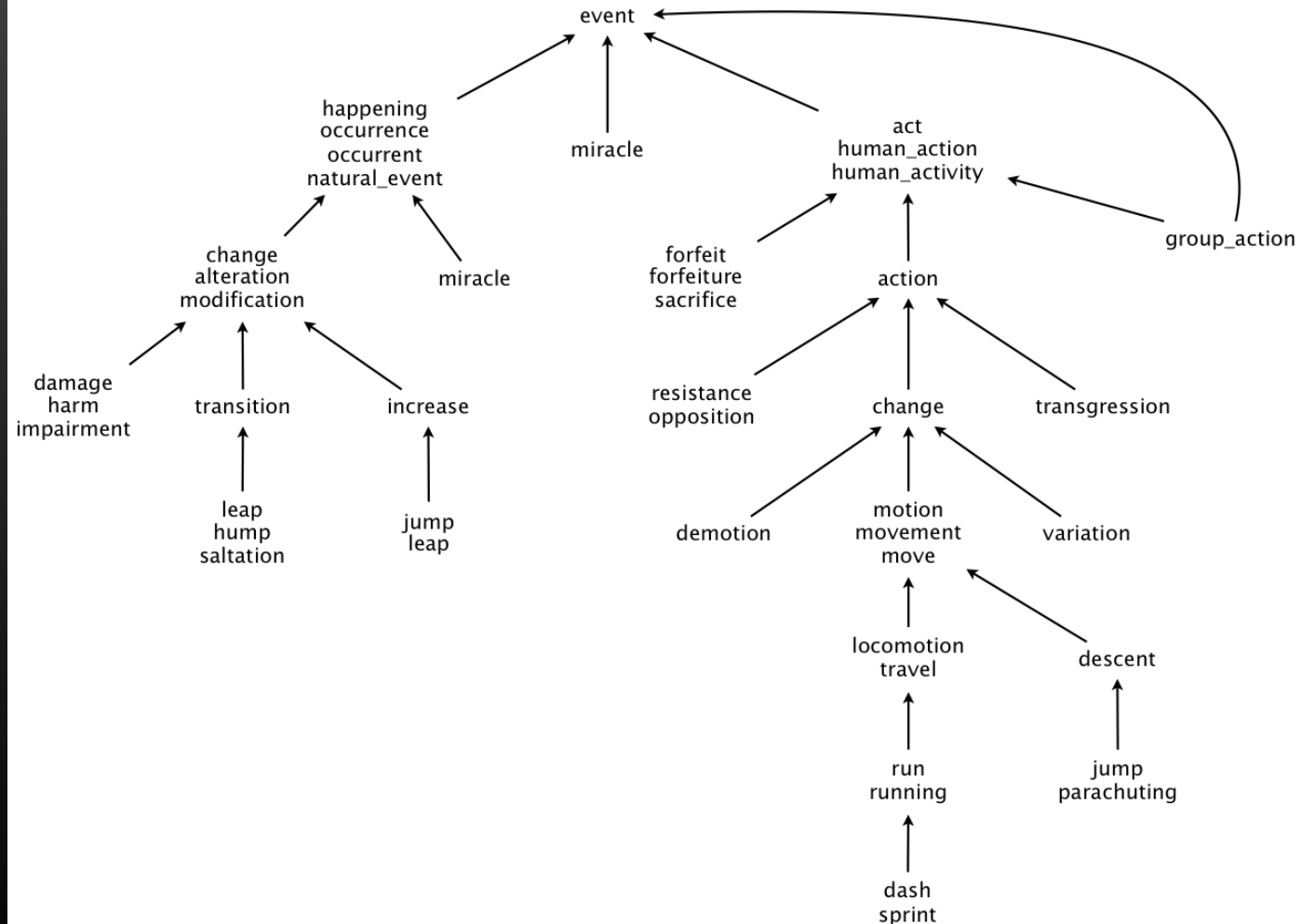
- Spesifikasi eksplisit dari suatu konseptualisasi (Gruber)
- Suatu himpunan istilah (*term*) yang tersusun secara hierarkis untuk menggambarkan suatu ranah tertentu yang dapat digunakan sebagai rangka dasar sebuah basis pengetahuan (Swartout, Patil, Knight, Russ)
- Sarana untuk mendeskripsikan konseptualisasi dari pengetahuan secara eksplisit yang berasal dari sebuah basis pengetahuan (Bernaras, Lasergoiti, Correra)
- Suatu spesifikasi eksplisit dan formal dari suatu konseptualisasi bersama (Studer, Benjamins, Fensel)

Contoh

GENE ONTOLOGY (GO)

immune response

- i- acute-phase response
- i- anaphylaxis
- i- antigen presentation
- i- antigen processing
- i- cellular defense response
- i- cytokine metabolism
 - i- cytokine biosynthesis synonym cytokine production
- ...
- p- regulation of cytokine biosynthesis
- ...
- i- B-cell activation
 - i- B-cell differentiation
 - i- B-cell proliferation
- i- cellular defense response
- ...
- i- T-cell activation
 - i- activation of natural killer cell activity
- ...



Wikidata4AI: Di mana kecerdasan artifisial-nya?

Kecerdasan artifisial (AI) dapat dibagi menjadi dua fokus:

- Fokus pada pengembangan **mesin cerdas**
 - *Machine learning, search and optimization, constraint satisfaction, neural network and deep learning, evolutionary computation*
- Fokus pada pengembangan **data cerdas**
 - *Logic, knowledge representation and reasoning, ontology, semantic web, multiagent systems*

Lalu, ada pula subbidang kecerdasan artifisial yang sering menggabungkan kedua fokus di atas:

- Robotika dan kendali, pemrosesan bahasa alami, penglihatan artifisial, pengenalan suara.

Mengapa ontologi?

- Pemodelan secara ontologis merupakan bagian penting dari pengetahuan → penting bagi Wikidata
- Membantu pengorganisasian pengetahuan dan memudahkan pencarian informasi
- Informasi skema memudahkan pemeriksaan kualitas data dan memungkinkan derivasi fakta baru
- Sarana komunikasi antara pemilik dan pengguna data
- Mempermudah gunakembali (*reuse*) pengetahuan
- Memfasilitasi interoperabilitas antar sistem

Komponen-komponen ontologi

Konsep/Kelas

- Himpunan atau kelas entitas dalam ranah tertentu
- *manusia, negara, organisasi, kota di Indonesia*

Instans

- Entitas individual yang dapat menjadi anggota suatu konsep
- *Habibie, Indonesia, Pare-pare*

Relasi

- Relasi; bisa melibatkan lebih atau kurang dari 2 pihak
- *tempat lahir, jabatan, negara*

Aksioma

- Fakta atau pernyataan yang dianggap benar dalam ranah pembicaraan
- *Setiap negara (pasti) memiliki sebuah ibu kota, jabatan publik adalah sebuah posisi*

Jenis-jenis ontologi

- Kosa kata (*vocabulary*) terkontrol
 - Hanya berupa daftar konsep
- Taksonomi
 - Daftar konsep dan hierarki konsep berdasarkan relasi *is-a*
- Kamus/*thesaurus*
 - Daftar konsep dan relasi-relasi antar konsep (tidak hanya *is-a*)
- Ontologi berbasis logika formal
 - Daftar konsep, relasi, serta aksioma

Inti Wikidata: pemodelan secara ontologis

- Satu (instans) jembatan (yang sama) di dua lokasi yang berbeda harusnya diwakili oleh satu atau dua butir data
- Q181409 vs. Q16336079 vs. Q1302994
 - Sama atau beda?
- Q18530 -- subclass of (P279) -- Q1789452
 - Benar atau salah?
- Bagaimana memodelkan bahwa suatu karakter terbunuh di suatu episode/buku/film spesifik dari suatu serial?
 - Pakai P1196 dengan P1343 sebagai penjelas atautkah sebagai referensi?
 - Contoh: Albus Dumbledore (Q712548)
- Perselisihan dalam penggunaan atribut ayah (P22) pada Yesus (Q302)
 - Tanpa bermaksud memancing perdebatan, lihat riwayat sunting di sekitar akhir Juni dan awal Juli 2018.

“Kelas” dalam Wikidata

- Secara formal, Wikidata tidak mengandung memiliki konsep “kelas” (himpunan objek)
 - ... walaupun banyak butir yang mewakili suatu kelas
 - Satu butir data dapat menjadi kelas dan juga instans sekaligus (metamodel)
- Atribut ontologis dapat memiliki penjelas (*qualifiers*).
- Beberapa atribut sering dipakai untuk pemodelan secara ontologis:
 - P31 – instance of
 - P279 – subclass of

Lihat penggunaan di <https://tools.wmflabs.org/sqid/>

Semantiknya di mana?

- Semantik = deskripsi dari interpretasi yang benar/diinginkan
 - Lebih dari sekedar mengikuti sintaks (format yang benar)
- Tidak ada semantik berbasis logika formal
- Makna yang diinginkan ditulis di dokumentasi
 - Makna P31: anggota himpunan
 - Makna P279: relasi subset/himpunan bagian
- Banyak kasus khusus yang tidak jelas semantikanya:
 - P279 (subclass of) boleh diberi penjelasan?
 - Bagaimana menginterpretasikan butir data yang menjadi kelas dan instans sekaligus?
- Secara praktis, interpretasi tergantung konteks.

Jadi bagaimana cara kerjanya?

astronomical body (Q6999)

celestial body

physical body of astronomically-significant size, mass, or role, naturally occurring in a universe; single, tightly bound contiguous entity (while an astronomical/celestial object is a complex, less cohesively bound structure, may consist of multiple bodies)

subclass of: every astronomical body is also a(n) physical object ⓘ, location, space object

instance of: astronomical body is a(n) astronomical object type

Instances	
Direct instances	954 Hourglass Nebula , NGC 554A , NGC 6778 , Pi ¹ Ursae Majoris , HD 17925 , NGC 6589 , Q617138 , Hanny's Voorwerp , Pillars of Creation , IC 4605 , Abell 31 , NGC 744 , NGC 6857 , GCIRS 13E , Q222717 , 2MASS 1047+2124 , Q222364 , 2MASS J03552337+1133437 , Mice Galaxies , 1 Lupi , ... further results
All instances	154045
Typical Properties	galaxy morphological type , Minor Planet Center body ID , provisional designation , stellar rotational velocity , SIMBAD ID , longitude of ascending node , mean anomaly , argument of periapsis , orbital eccentricity , minor planet group , apparent magnitude , DAMIT asteroid ID , site of astronomical discovery , constellation , absolute magnitude , spectral class , Lagrangian point , JPL Small-Body Database ID , redshift , semi-major axis ,
Classification	
Direct superclasses	physical object ⓘ 36356978 , location 7321659 , space object 154045
Direct subclasses	With instances 40 With subclasses 31 All 60 substellar object 110043 , star 24162 , galaxy 11120 , satellite 4217 , extrasolar object 3001 , minor-planet group 1687 , gravitationally bound system 988 , star cluster 971 , interstellar cloud 844 , star system 696 , massive compact halo object 650 , asterism 644 , emission nebula 632 , constellation 464 , unknown astronomical object 293 , reflection nebula 160 , active galactic nucleus 134 , supercluster 60 , gamma-ray burst 39 , Group of Stars 36 , pair of stars 23 , nakshatra 21 ,

- Petunjuk (*guideline*) di masing-masing subkomunitas topik
- Butir-butir untuk topik yang spesifik dapat disepakati secara lebih mudah.
- Tidak ada koordinasi, konsultasi (QA) dan petunjuk yang bersifat global.

Masalah pemodelan ontologis di Wikidata

- *Semantic drift*: makna entitas bergeser dengan berjalannya waktu
- Relasi sirkular melibatkan *subclass of* dan/atau *instance of*
 - Switzerland (Q39) *instance of* Federal Treaty (Q687554)
 - Federal Treaty *subclass of* Switzerland
 - Lihat juga binder dan thickener
- Meta-model yang tercampur
 - “Noodle” subclass of “pasta”
 - “Noodle” instance of “type of pasta”
 - “Type of pasta” subclass of “pasta”
- Konsep di level “atas”
 - Lihat *subclass* dari entity (Q35120) yang tercampur-campur

entity (Q35120)

thing | something

something that exists in the identified universe

subclass of:

instance of: entity is a(n) variable-order metaclass

Instances



Classification



Direct superclasses

Direct subclasses

With instances 52

With subclasses 59

All 61

object 45154952 , structure 37656580 , artefact 31962097 , agent 5964040 , subject 5039393 ,
temporal entity 4633276 , individual 4602279 , class 4359527 , phenomenon 3330581 , property 2004302 ,
quality 1990451 , geographic entity 1979615 , part 1447645 , problem 1302033 , substance 797093 ,
chemical entity 795971 , political entity 334692 , resource 101157 , former entity 80390 , incentive 41106 ,
type 38748 , creature 21827 , organism 21827 , disposition 17138 , information source 15362 ,
existence 15254 , assumed entity 12160 , (meta)class 7089 , spacetime volume 4570 ,
anatomical entity 2779 , independent continuant 2772 , stakeholder 2518 , Norse entity 1535 ,
observable entity 206 , psychological phenomenon 172 , unknown 118 , liminal being 105 , space 31 ,
cause 25 , interpreter 17 , sensation 14 , converter 11 , being 8 , untitled entity 7 , blue 3 ,
substance 2 , white 2 , green 2 , contact point 2 , perception set 1 , orange 1 , red 1

All subclasses

1458051

Masalah pemodelan ontologis Wikidata (2)

- Penambahan makna bisa terjadi secara konsep.

embassy (Q3917681)

ambassadorial delegation

permanent diplomatic mission of higher level, representing its operator in the country the embassy is in

subclass of: every embassy is also a(n) [building of public administration](#), [diplomatic mission](#)

instance of: embassy is not an instance of any other class

Instances



Classification



Direct superclasses [building of public administration](#) **19624**, [diplomatic mission](#) **3075**

Direct subclasses

With instances **1**

With subclasses **0**

All **1**

[Apostolic Nunciature](#) **191**

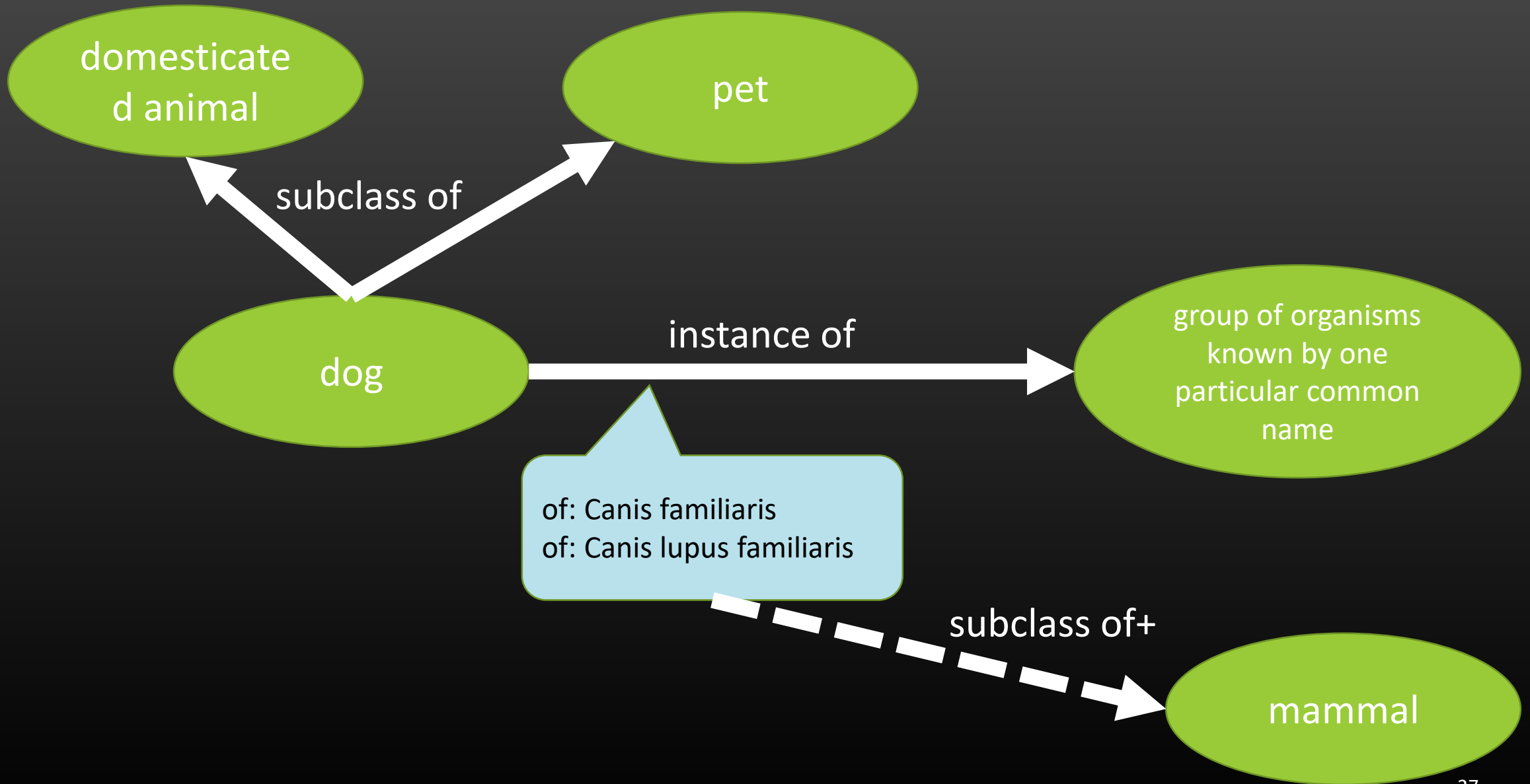
All subclasses

1

Masalah pemodelan ontologis Wikidata (3)

Perbedaan keputusan pemodelan (*modeling decision*) untuk ranah-
ranah yang berbeda karena diambil secara lokal:

- Khususnya pada pertanyaan:
“Aspek mana yang akan dimodelkan sebagai kelas?”
- Gejalanya (silakan dicek):
 - Granularitas yang tidak koheren:
 - “mammal” hampir tidak punya subkelas
 - “building” punya hierarki kompleks di bawahnya
 - “human” punya banyak subkelas yang tidak untuk dipakai dengan *instance of*
- Relasi *subclass of* tidak dapat dipakai untuk menelusuri hierarki
 - Apa hubungan antara dog (Q144) dengan mammal (Q7377)?



Tantangan pemodelan ontologis di Wikidata

- Penyuntingan local tapi mengharap definisi semantik yang bersifat global
- Kebanyakan penyunting tidak akan berpikir global (terlalu rumit)
- Tidak ada *use case* yang umum untuk keseluruhan Wikidata; tapi banyak *use case* yang bersifat local
- Internasionalisasi: konsep belum tentu dapat diterjemahkan dengan tepat.
- Pemodelan ontologi sendiri memang tidak mudah: para ahli saja bisa bertengkar/berdebat keras.

Penggunaan *constraint* (batasan)

sex or gender (P21)

gender identity | gender expression | gender | biological sex | man | woman | male | female | intersex | sex

sexual identity of subject: male (Q6581097), female (Q6581072), intersex (Q1097630), transgender female (Q1052281), transgender male (Q2449503). Animals: male animal (Q44148), female animal (Q43445). Groups of same gender use "subclass of" (P279)

property constraint

16 statements >

conflicts-with constraint (type of constraint for Wikidata properties: used to specify that an item must not have a given statement) >

property : instance of

item of property constraint : [Wikimedia category](#) (use with 'instance of' (P31) for Wikimedia category)

item of property constraint : [Wikimedia list article](#) (page of a Wikimedia project with a list of something)

item of property constraint : [position](#) (office with a set of powers and responsibilities within a private or public organization or the state)

item of property constraint : [novel](#) (narrative text, normally of a substantial length and in the form of prose describing a fictional and sequential story)

item of property constraint : [work](#) (physical or virtual object made by humans)

item of property constraint : [filmography](#) (list of films related by some criteria)

item of property constraint : [discography](#) (study and cataloging of published sound recordings)

item of property constraint : [family name](#) (part of a naming scheme for individuals, used in many cultures worldwide)

item of property constraint : [murder](#) (unlawful killing of a human with malice aforethought)

item of property constraint : [death](#) (permanent cessation of vital functions)

constraint status : [mandatory constraint](#) (status of a Wikidata property constraint: indicates that the specified constraint applies to the subject property without exception and must not be violated)

one-of constraint (type of constraint for Wikidata properties: used to specify that the value for this property has to be one of a given set of items) >

item of property constraint : [male](#) (human who is male)

item of property constraint : [female](#) (human who is female (use with P21))

item of property constraint : [intersex](#) (innate variations in sex characteristics such that individuals differ from norms for male or female bodies)

item of property constraint : [hermaphrodite](#) (organism with both 'male' and 'female' reproductive organs)

item of property constraint : [transgender](#) (state of one's gender identity not matching one's assigned sex)

Constraint – pemodelan melebihi kelas

- Dinyatakan sebagai pernyataan pada atribut
- Mendefinisikan prasyarat kebutuhan (*requirement*) pada penggunaan atribut
 - *functionality, symmetry, transivity, ...*
 - *disjointness, value restrictions, format*
 - *schema information* seperti *allowed qualifiers*
- Bermanfaat untuk kendali mutu dan dokumentasi

property constraint (Q21502402)

Wikidata property constraint

constraint of a Wikidata property

subclass of: every property constraint is also a(n) [Wikidata property definition](#)

instance of: property constraint is not an instance of any other class

Instances

Direct instances 0

All instances 12

Typical Properties [instance of](#) , [subclass of](#) , [partially coincident with](#) , [used by](#) , [Wikidata property](#) , [different from](#)

Classification

Direct superclasses [Wikidata property definition](#) 14

Direct subclasses

[With instances](#) 3

[With subclasses](#) 1

[All](#) 35

[Commons link constraint](#) , [difference within range constraint](#) , [allowed qualifiers constraint](#) , [relation of type constraint](#) , [type constraint](#) , [item requires statement constraint](#) , [conflicts-with constraint](#) , [mandatory constraint](#) , [distinct values constraint](#) , [format constraint](#) , ... [further results](#)

All subclasses 36

Semantiknya?

- Tidak ada semantik formal
 - Dokumentasi terkadang tidak jelas
 - Pemeriksaan kondisi dapat diotomatisasi dengan SPARQL
- Dipakai saat penyuntingan dalam bentuk pengingat (*warning*)
- Juga dipakai untuk merekomendasikan penambahan informasi yang absen dari butir yang disunting

Airbus A380 (Q5830)

double-deck aircraft made by Airbus

A380 | Airbus Jumbo Jet | A380 Jumbo Jet

powerplant



Rolls-Royce Trent 700

edit

quantity

4

▼ 0 references

+ add reference



Engine Alliance GP7000

edit

quantity

4

▼ 0 references

Yang berikut mungkin berlaku umum?

“Setiap pesawat Airbus memiliki mesin Rolls-Royce”

Motivasi untuk semantik

- SEMUA data perlu diinterpretasikan
 - Hierarki kelas dan batasan atribut HANYA dua contoh saja → Ada lagi yang lain
- Perlu cara untuk memberikan makna pada struktur data yang ada
 - *Customizable rules of inference*
 - Tidak unik, dapat didefinisikan pengguna (*user-definable*), bergantung pada konteks (*context-dependent*)
 - Tapi tetap tajam (*crisp*, tidak *fuzzy*) dan presisi (*precise*)

Rules list

spouse is symmetric

materialisable consequences

```
(?x.spouseP26 = ?y)@?S  
→ (?y.spouseP26 = ?x)@?S
```

A male parent is a father

materialisable consequences

```
(?father.childP40 = ?child)@?X,  
(?father.sex or genderP21 = maleQ6581097)@?Y  
→ (?child.fatherP22 = ?father)
```

A female parent is a mother

materialisable consequences

```
(?mother.childP40 = ?child)@?X,  
(?mother.sex or genderP21 = femaleQ6581072)@?Y  
→ (?child.motherP25 = ?mother)
```

a male parent of a parent is a grandfather

informational consequences

```
(?grandfather.sex or genderP21 = maleQ6581097)@?X,  
(?grandfather.childP40 = ?parent)@?Y,  
(?parent.childP40 = ?child)@?Z  
→ (?child.relativeP1038 = ?grandfather)@{type of kinshipP1039 =  
grandfatherQ9238344}
```

Apa yang Wikidata butuhkan?

- Permasalahan yang ada berskala besar. Mulai dari mana?
- Adakah pola atau anti-pola sederhana yang dapat dijadikan fokus?
- Bagaimana melakukan pemodelan yang berbeda secara berdampingan (tanpa merusak)?
- Adakah konsep/atribut pemodelan yang sudah digunakan secara implisit? Bagaimana menangani pendekatan-pendekatan berbeda yang saling bertentangan?

Apa yang ditawarkan Wikidata?

- Struktur dan pola pengetahuan baru
- Pengguna yang terbuka dengan perkakas dan metode baru
- Himpunan data besar untuk analisis
 - *knowledge graph*
 - rekaman aktivitas pengguna
 - kueri
- Relevansi praktis secara langsung

Kesimpulan

- Wikidata berkembang sangat cepat dengan **komunitas** yang **terbuka** dan **ramah** terhadap **ontologi**
- Ontologi yang bagus memerlukan **kerja keras** dan **tidak muncul begitu saja**.
- Banyak yang sudah dilakukan, tapi lebih banyak lagi yang masih dapat dilakukan.



Terima kasih

Ada pertanyaan?