knowledge graphs: wat het zijn en hoe je ze maakt

en ook nog iets over rdf, tripels en ontologieën

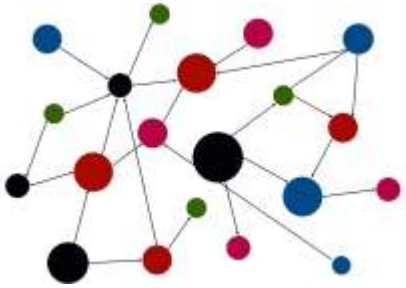
eric sieverts

vogin IP

~~VOGIN-IP-lezing~~
~~19 maart 2020~~

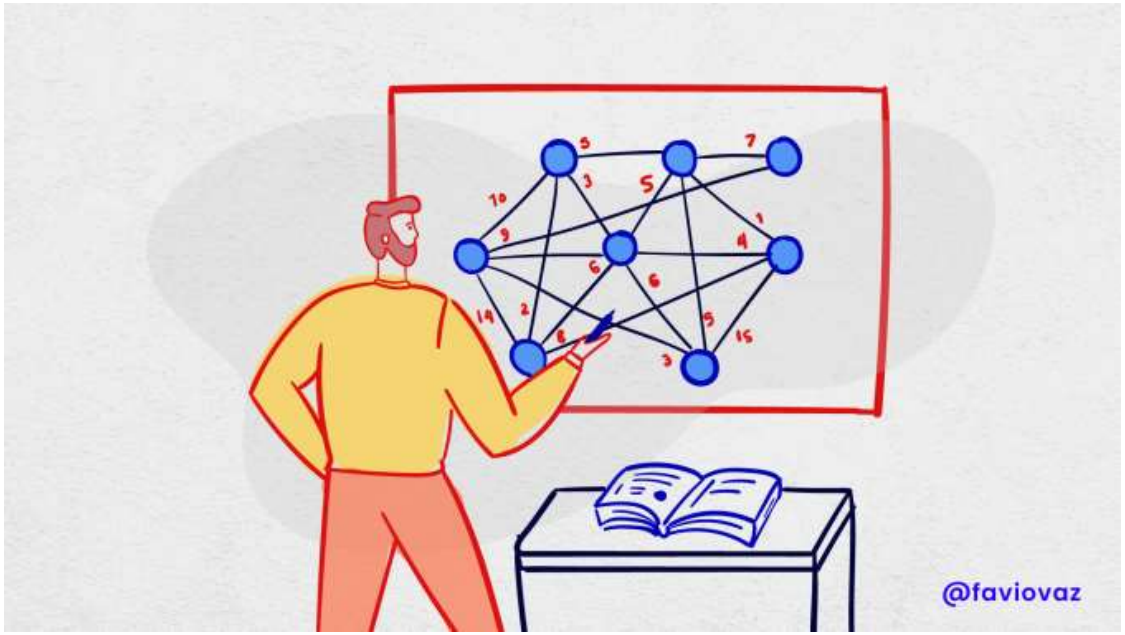
was canceled due to corona

wie weet wat een knowledge graph is?



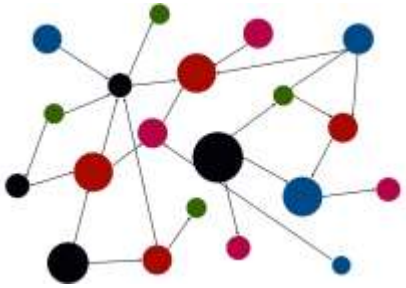
- wat is het?
- wie kent voorbeelden?
- wie heeft die wel eens gebruikt?

links:
- deze presentatie
- artikel van Joyce van Aalten in IP (januari 2020)
- korte handleiding voor de applicatie gra.fo
- materiaal voor oefening 2



People don't think in tables (like in traditional RDBMS), but they do immediately understand graphs. When you draw the structure of a knowledge graph on a whiteboard, it is obvious what it means to most people.

programma



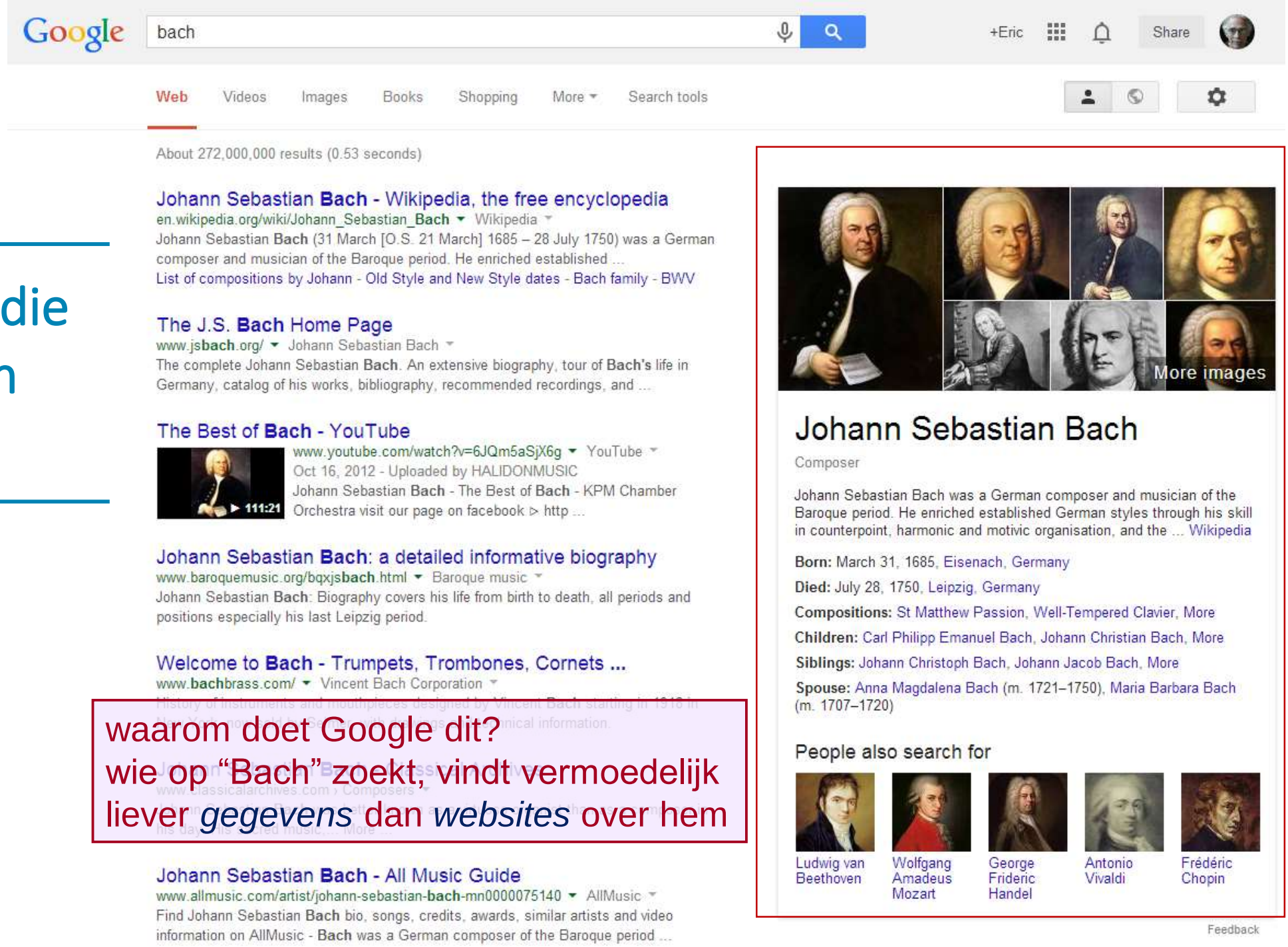
wat we gaan doen	theorie daarbij
• kennismaking met bestaande systemen	tripels, URI's, RDF
• uitwerken van stukje graph (op papier)	
• ontologie voor die graph bedenken en online bouwen daarvan	ontologie en klassen relatie met databasevelden en met thesaurus/taxonomie
• online bouwen van stukje graph	
• bespreken en tonen van resultaten	
• praktische toepassingen	

google knowledge graph

- >1 miljard concepten/entiteiten met >70 miljard relaties en kenmerken
- gegevens komen uit o.a.: Wikipedia (DBpedia), Wikidata, Freebase, CIA World factbook, datasets van toonaangevende organisaties + analyse van gegevens uit webpagina's uit eigen index

4

wat merk je van die knowledge graph van Google?



waarom doet Google dit?
wie op “Bach” zoekt, vindt vermoedelijk liever *gegevens* dan *websites* over hem

wat zie je van die knowledge graph?

- ***niet*** het hele netwerk
- zelfs ***niet*** stukjes netwerk
- ***wel*** o.a. "knowledge cards", waarin gegevens uit de knowledge graph, die horen bij een persoon/object/entiteit, worden gecombineerd
- die knowledge cards verschijnen bij een zoekactie rechts naast het gewone zoekresultaat
- die knowledge cards bestaan uit een heleboel zogenaamde tripels, waarin feitelijke gegevens zijn vastgelegd

Leonardo da Vinci

Painter

Leonardo di ser Piero da Vinci was an Italian polymath, painter, sculptor, architect, musician, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer. Wikipedia

Born: April 15, 1452, Vinci, Italy

Died: May 2, 1519, Amboise, France

Periods: High Renaissance, Pastel Art

Structures: Milan Cathedral

Parents: Piero Fruosino di Antonio da Vinci, Caterina da Vinci

Artwork View 15+ more

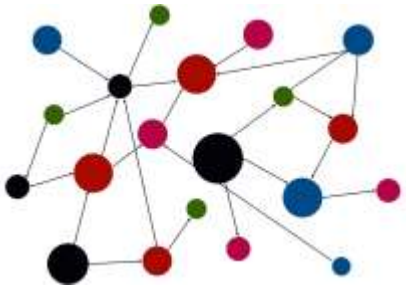
Mona Lisa
1517

The Last Supper
1498

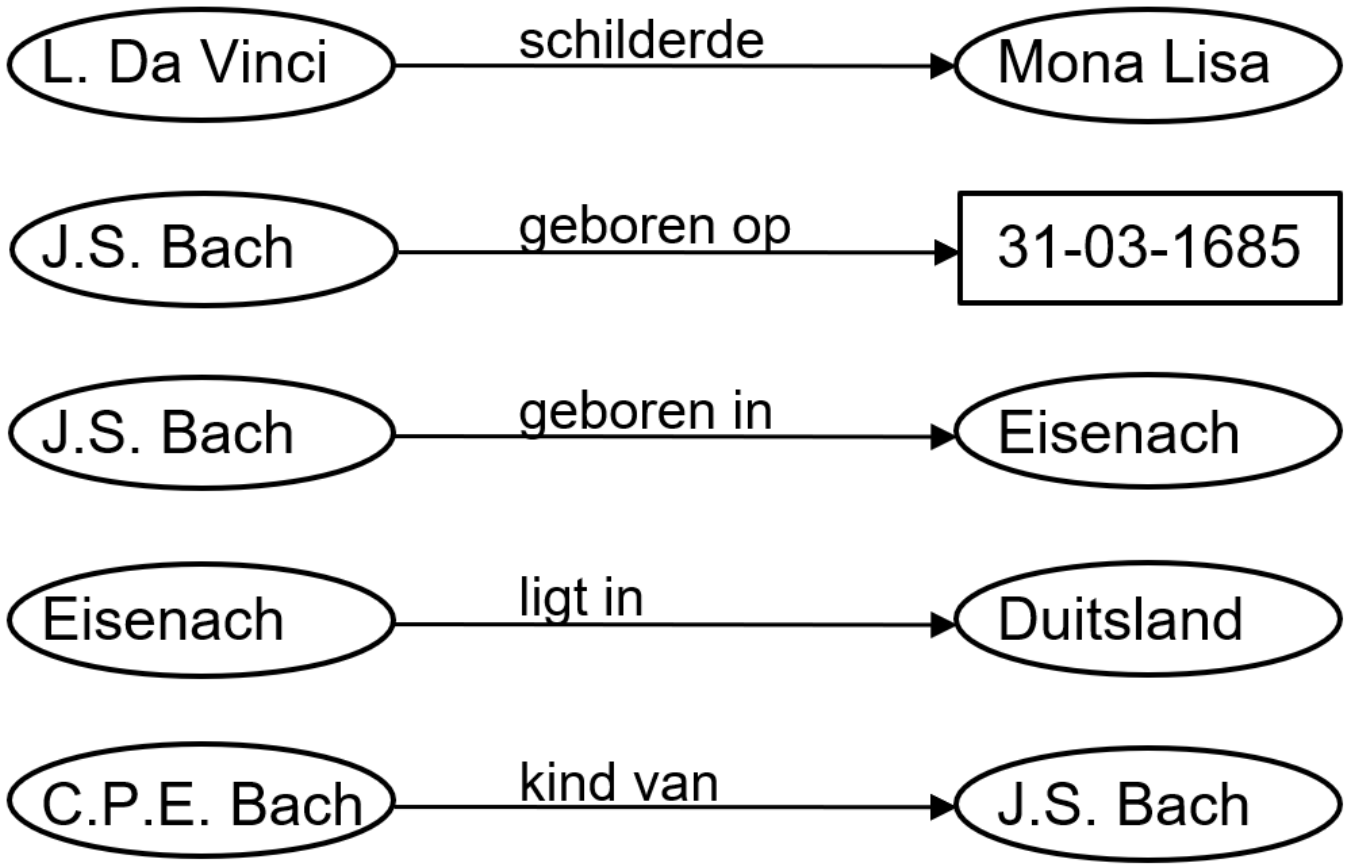
Vitruvian Man
1490

The Battle of Anghiari
1505

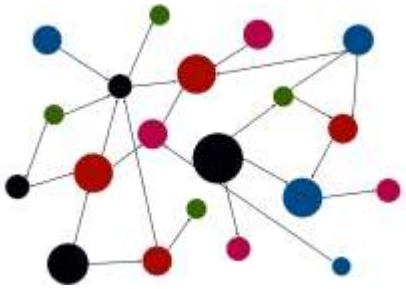
tripels



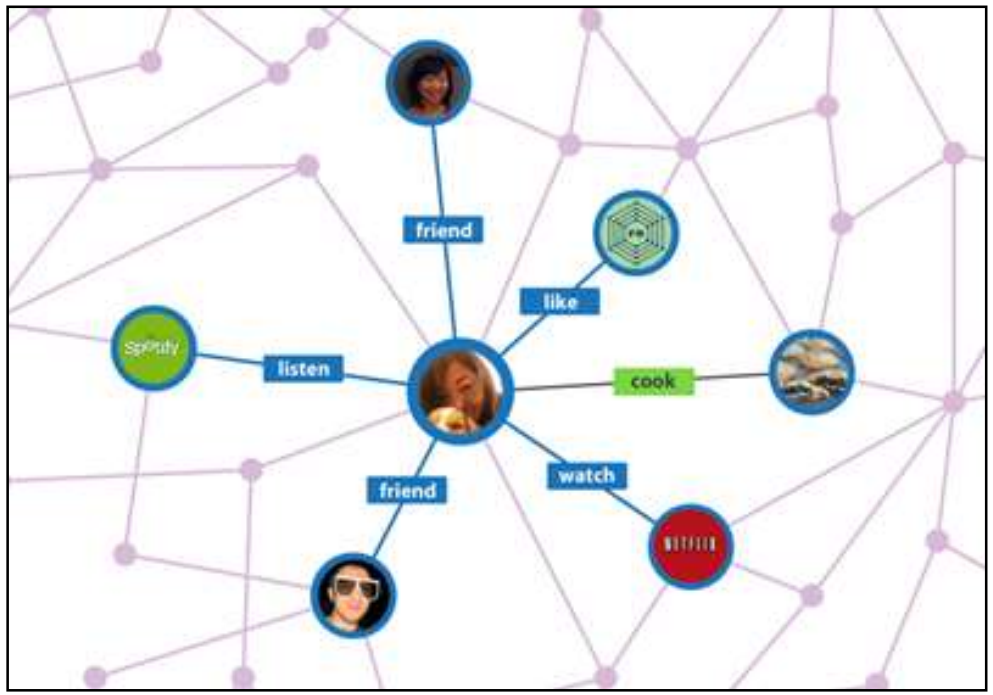
de informatie uit die graph wordt opgeslagen als "tripels"



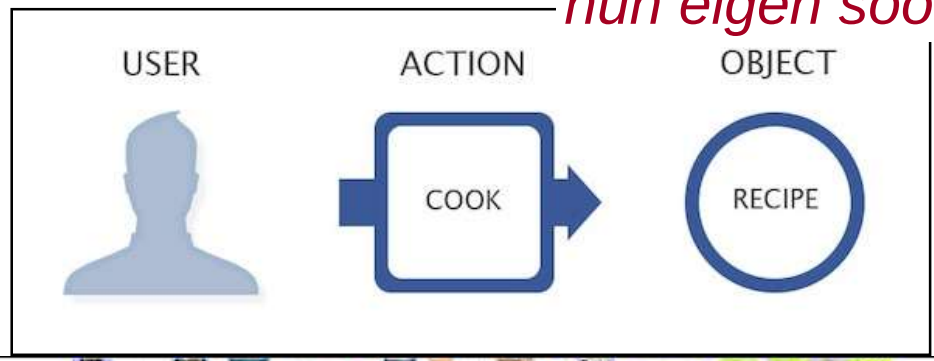
Facebook graph (search)

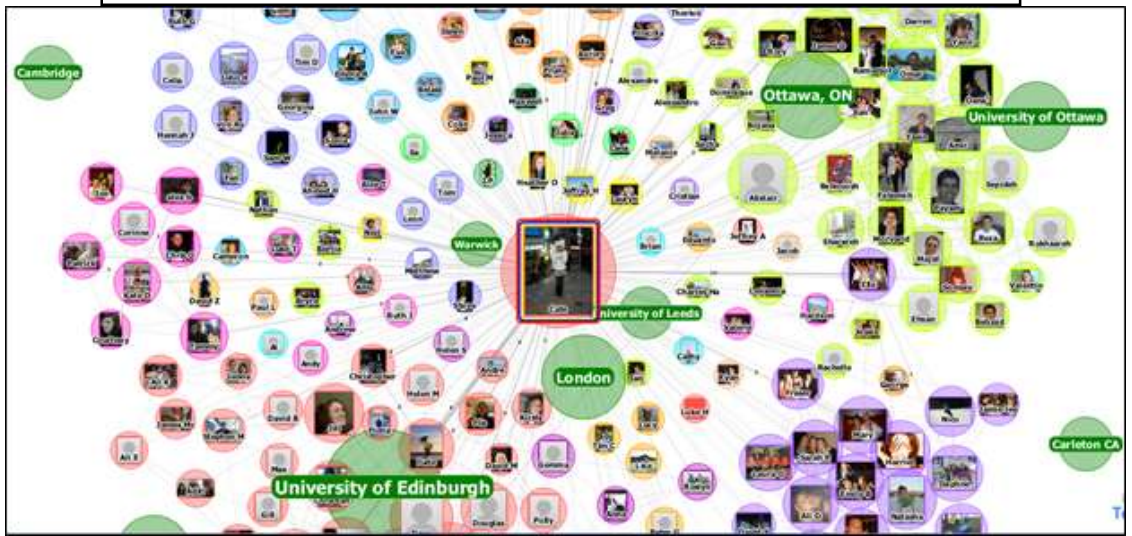


- vaste achterliggende datastructuur en relaties om verbanden te kunnen tonen en er gestructureerd in te kunnen zoeken

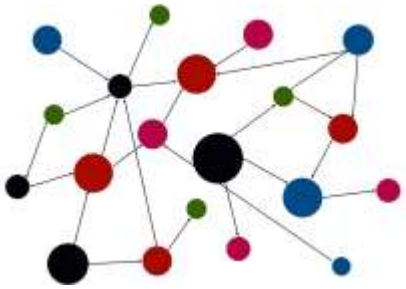


hun eigen soort tripels





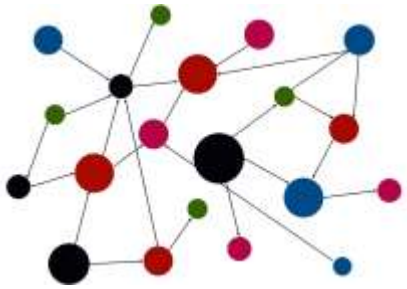
RDF (resource description framework)



- RDF is een standaard methode waarmee je de relatie tussen een resource (of object) en zijn metadata kunt beschrijven
- die relaties zijn vastgelegd in zogenaamde tripels
- resources moeten een *URI* hebben om ernaar te verwijzen
- definities van de relaties tussen object en metadata moeten liefst ook een URI hebben
- RDF is voor computers bedoeld en wordt meestal in XML- of JSON-notatie weergegeven
- je hebt ook nog RDFa *(RDF in attributes)* waarmee metadata in de inhoud van (X)HTML webpagina's verwerkt worden *(als attributen van HTML-codes)*

* *URI (uniform resource identifier) is generalisatie van URL's van het web*

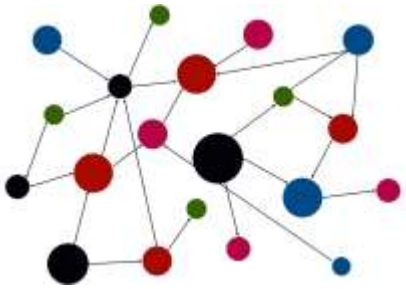
RDF en tripels



- eigenschappen (metadata) worden vastgelegd in zogenaamde tripels:
  subject <predicaat> object
  - van de Engelse woorden voor *onderwerp – gezegde – lijdend voorwerp*
  - je zou het ook kunnen noemen: *ding <eigenschap> waarde*
    of soms ook: *ding <relatie> ding*

  daarbij
  - heeft te beschrijven ding een webadres (URI)
  - heeft definitie van de eigenschap (of relatie) van dat ding liefst ook een URI
  - heeft de "waarde" van die eigenschap (of het ding waarmee een relatie bestaat) liefst ook een URI (want het kan zelf subject in een ander tripel zijn, anders is het een “literal”, bijv. iemands leeftijd)

  voorbeeld:
  - *boek* (heeft een webadres: URI)
  - *heeft auteur* (betekenis van die eigenschap/relatie is ergens gedefinieerd: URI)
  - *persoon* (gegevens van persoon ergens op web te vinden: URI)



tripels

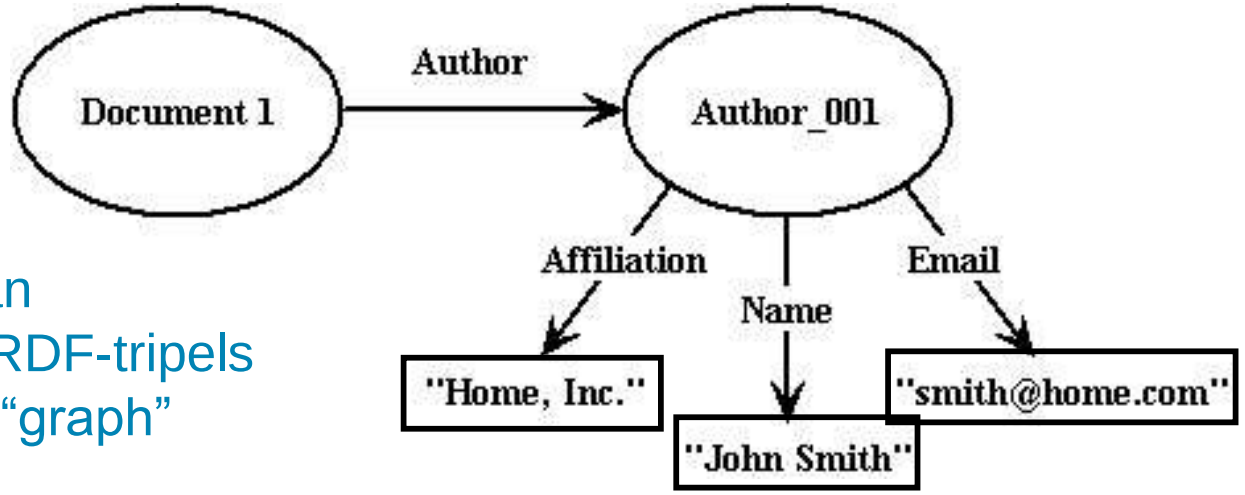
subject <predicaat> object

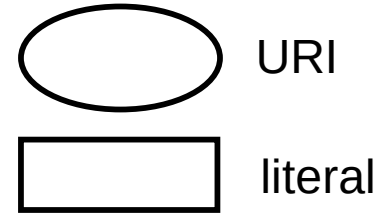
document 1 <heeft auteur> author_001

author_001 <heeft naam> john smith

author_001 <heeft affiliatie> home inc.

author_001 <heeft email> smith@home.com





grafische weergave van
simpel netwerk van 4 RDF-tripels
gecombineerd tot een “graph”

- tripels (kunnen) worden opgeslagen in zogenaamde *triple stores*

intermezzo RDF en tripels

naar idee van
Lucas Koster / IP

http://www.worldcat.org/oclc/10098995

"Uit Berlijn"

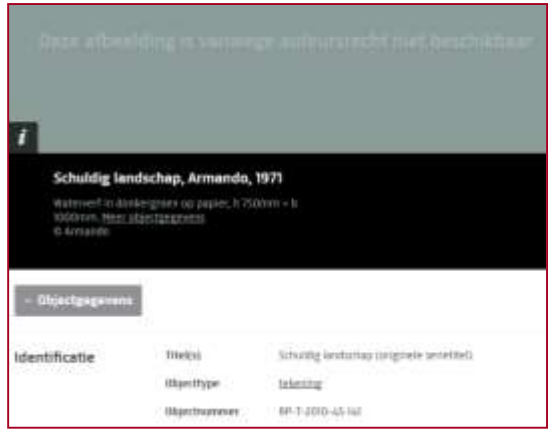
is een

http://schema.org/Book

boek

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

```
<rdf:Property rdf:about="http://www.w3.org/1999/02/22-rdf-
syntax-ns#type">
  <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-
rdf-syntax-ns#"/>
  <rdfs:label>type</rdfs:label>
  <rdfs:comment>The subject is an instance of a class.
</rdfs:comment>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
</rdf:Property>
```



http://hdl.handle.net/10934/RM0001.COLLECT.496040

title

http://purl.org/dc/terms/title

Uit Berlijn

is geschreven door

http://purl.org/dc/terms/creator

"Armando"

is schilder van

Schuldig landschap

http://purl.org/dc/terms/creator

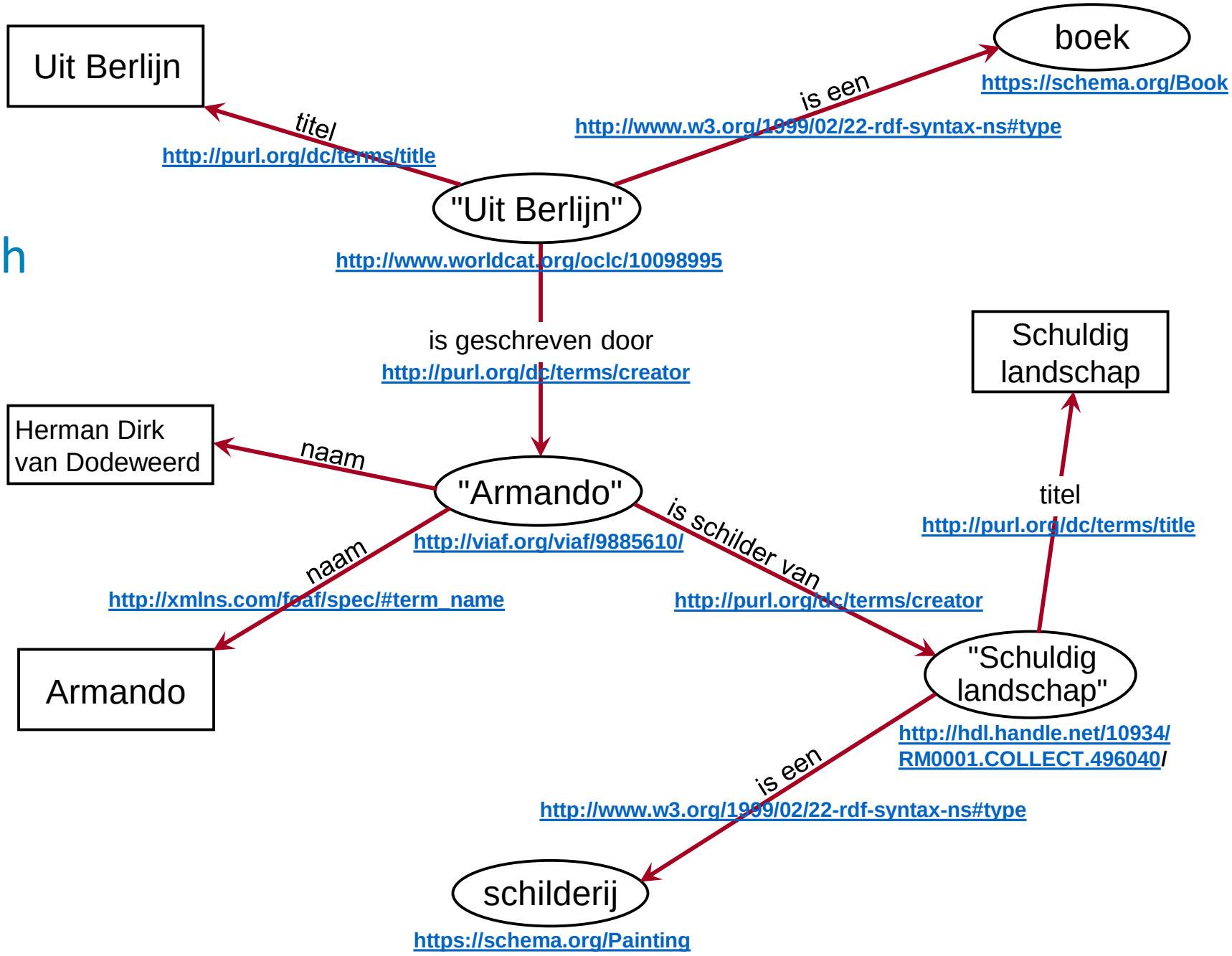
http://viaf.org/viaf/9885610/

http://xmlns.com/foaf/spec/#term_name

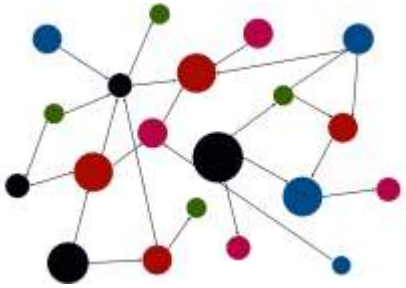
naam

Armando

dit is al een
knowledge graph



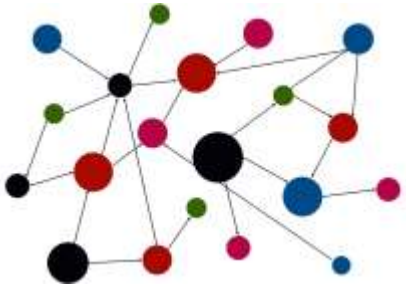
oefening 1:
exploreer een knowledge graph (1)



exploreer wat van de mogelijkheden van een paar van de hierna genoemde voorbeelden (met aanwijzingen)

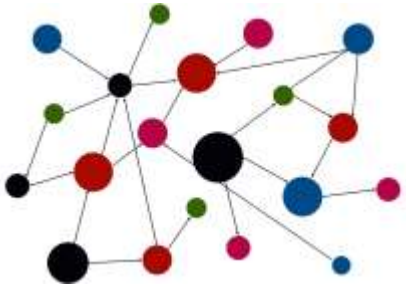
Google	https://carlhendy.com/knowledge-graph-search/ Je kunt hier in Google's Knowledge Graph zoeken en van gevonden entiteiten (met de knop "View in Google") de "Knowledge Card" opvragen. Deze dienst biedt geen visualisatie van (delen van) de "graph".

oefening 1: exploreer een knowledge graph (2)



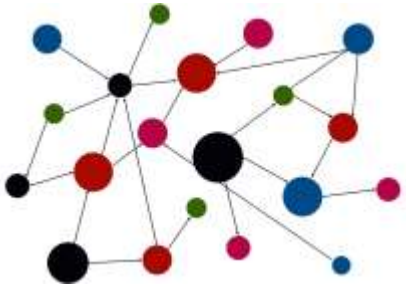
Microsoft Academic Graph *gegevens over > 200 miljoen publicaties, hun auteurs, referenties etc*	http://ma-graph.org/ De graph zelf (of delen ervan) kun je hier niet visualiseren. Je ziet hier wel om hoeveel informatie over wetenschappelijke publicaties het gaat. Als je de link naar "RDF Dump Files" aanklikt, krijg je een lijst met bestanden die elk de beschikbare RDF-tripels van een bepaalde klasse van entiteiten bevatten. Aan de grootte van de bestanden zie je dat dat er bij sommige klassen heel veel zijn. Kies er daarom eentje die niet al te groot is en klik op de [nt]-link (middenkolom). • Welke properties van de betreffende entiteit worden in de tripels beschreven? • Welke "literals" zijn daarbij? • Kun je achterhalen wat de eerste "instance" is waarvan properties worden beschreven?

oefening 1: exploreer een knowledge graph (3)



Factforge *>1 miljard feiten uit DBPedia en andere bronnen*	http://factforge.net • Tik onder "View Resource" een term in waarvan je verwacht dat daar wel gegevens van zullen zijn, bijv. Greenpeace. • Klik op [Text] om een tekstueel overzicht te krijgen van de tripels waarvan Greenpeace subject is. Dat blijken er erg veel te zijn, zodat je misschien de Find-functie van je browser moet gebruiken om te zien of een bepaald kenmerk of relatie aanwezig is, bijvoorbeeld "Headquarters". • Klik op [Visual Graph]. Dit heeft hetzelfde effect als wanneer je al in het vorige scherm op [Visual] geklikt had. Je krijgt nu een veel beperkter aantal tripels te zien in een graph representatie (maximaal 20, want allemaal zou niet passen). • Dubbelklik op een interessante entiteit (object) waarmee Greenpeace verbonden is, zodat de visualisatie van de Knowledge Graph in die richting wordt uitgebreid. Dat kun je nog een paar keer doen, totdat het scherm te onoverzichtelijk wordt (of de responsetijd te traag). • Doe het voorgaande eventueel nog eens voor een andere entiteit.

oefening 1:
exploreer een knowledge graph (4)



Wikidata *collectie van 80 miljoen items met >10 miljard tripels*	https://www.wikidata.org/ • Tik rechtsboven een zoekterm in • Klik uit de resultatenlijst een entiteit aan die met je intentie correspondeert • Kijk welke statements (tripels) bij die entiteit getoond worden en klik eventueel door naar een genoemd "object" • Klik ook eens een "predikaat" (relatie) aan (in de linker kolom). https://query.wikidata.org/ • Hier kun je zelf SPARQL queries intikken, maar uit de bovenbalk kun je ook voorgeprogrammeerde queries kiezen, die een beetje idee geven hoe de SPARQL zoektaal werkt, en waarvan je details eventueel zelf kunt aanpassen. • Run de query door op de blauwe pijl te klikken • Uit de resultaatlijst kun je weer entiteiten aanklikken

How big is the Microsoft Academic Knowledge Graph?

The Microsoft Academic Knowledge Graph as of 2018-11 contai

209,792,741
papers

253,641,783
authors

25,431
affiliations

1,380,196,397
references

146,257,535
citations

48,650
journals

15,704
conference instances

4,337
conference series

229,716
fields of study

RDF DUMPS

We provide the Microsoft Academic Knowledge Graph in RDF in several parts, enabling to use also only some specific

RDF dump from 2018-11-09:

File	Link (comp.)	Compressed file size	Link (uncomp.)	Uncompr.file size	Data sample
Affiliations	nt.bz2	1MB	nt	31MB	sample
Authors	nt.bz2	6GB	nt	192GB	sample
Conference instances	nt.bz2	1MB	nt	22MB	sample
Conference series	nt.bz2	1MB	nt	4MB	sample
Field of study children	nt.bz2	2MB	nt	41MB	sample
Fields of study	nt.bz2	8MB	nt	221MB	sample
Journals	nt.bz2	2MB	nt	40MB	sample
Paper abstracts	nt.bz2	27GB	nt	106GB	sample
Paper author affiliations	nt.bz2	4GB	nt	61GB	sample
Paper citation contexts	nt.bz2	14GB	nt	148GB	sample
Paper fields of study	nt.bz2	5GB	nt	120GB	sample
Paper languages	nt.bz2	505MB	nt	21GB	sample
Paper references	nt.bz2	8GB	nt	149GB	sample
Papers	nt.bz2	18GB	nt	340GB	sample
Paper URLs	nt.bz2	5GB	nt	46GB	sample
Related fields of study	nt.bz2	1MB	nt	17MB	sample
Sum:		86GB		**1.2TB**	

Semantic Scholar zoekmachine maakt hier gebruik van voor automatische onderwerpstagging

voorbeeld van tripels uit de "affiliations" van Microsoft Academic Graph

subject

predicaat

object

welke organisatie is 20455151?

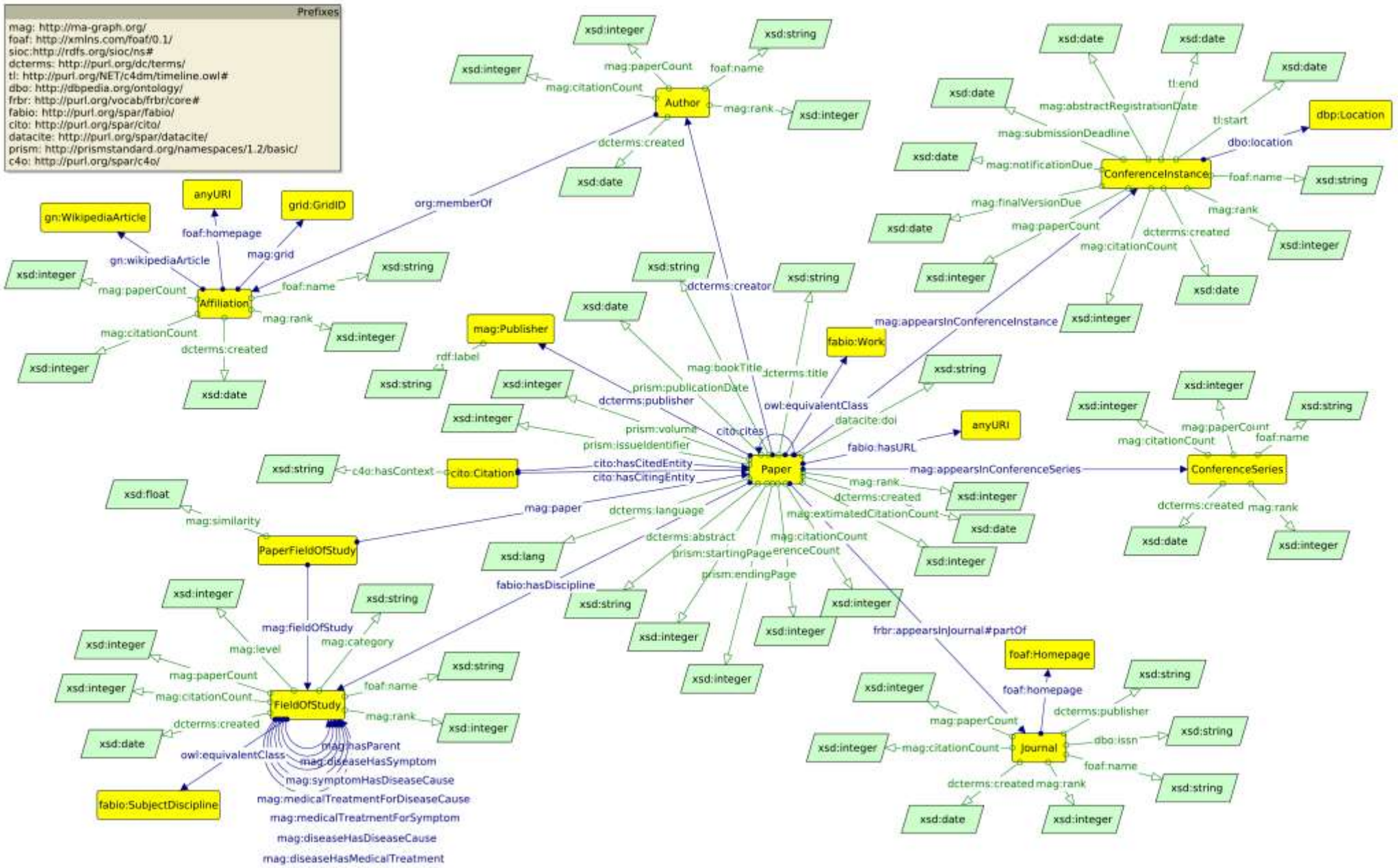
en welke is 24386293?

welke kenmerken van organisaties herken je hier?

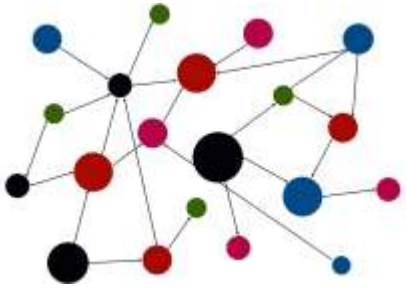
zie je hier ook literals bij?

```
<http://ma-graph.org/entity/20455151> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://ma-graph.org/class/Affiliation> .
<http://ma-graph.org/entity/20455151> <http://ma-graph.org/property/rank> "9761"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/20455151> <http://xmlns.com/foaf/0.1/name> "Air Liquide"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://ma-graph.org/entity/20455151> <http://ma-graph.org/property/grid> <http://www.grid.ac/institutes/grid.476009.c> .
<http://ma-graph.org/entity/20455151> <http://xmlns.com/foaf/0.1/homepage> <https://web.archive.org/web/20100205175402/http://airliquide.com/en/home.html> .
<http://ma-graph.org/entity/20455151> <http://www.w3.org/2000/01/rdf-schema#seeAlso> <http://en.wikipedia.org/wiki/Air_Liquide> .
<http://ma-graph.org/entity/20455151> <http://www.w3.org/2002/07/owl#sameAs> <http://en.wikipedia.org/wiki/Air_Liquide> .
<http://ma-graph.org/entity/20455151> <http://ma-graph.org/property/paperCount> "6355"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/20455151> <http://ma-graph.org/property/citationCount> "45307"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/20455151> <http://purl.org/dc/terms/created> "2016-06-24"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://ma-graph.org/entity/24386293> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://ma-graph.org/class/Affiliation> .
<http://ma-graph.org/entity/24386293> <http://ma-graph.org/property/rank> "13649"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/24386293> <http://xmlns.com/foaf/0.1/name> "Hellenic National Meteorological Service"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://ma-graph.org/entity/24386293> <http://xmlns.com/foaf/0.1/homepage> <http://www.hnms.gr/hnms/english/index_html> .
<http://ma-graph.org/entity/24386293> <http://www.w3.org/2000/01/rdf-schema#seeAlso> <http://en.wikipedia.org/wiki/Hellenic_National_Meteorological_Service> .
<http://ma-graph.org/entity/24386293> <http://www.w3.org/2002/07/owl#sameAs> <http://en.wikipedia.org/wiki/Hellenic_National_Meteorological_Service> .
<http://ma-graph.org/entity/24386293> <http://ma-graph.org/property/paperCount> "82"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/24386293> <http://ma-graph.org/property/citationCount> "1539"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/24386293> <http://purl.org/dc/terms/created> "2016-06-24"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://ma-graph.org/entity/32956416> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://ma-graph.org/class/Affiliation> .
<http://ma-graph.org/entity/32956416> <http://ma-graph.org/property/rank> "13117"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/32956416> <http://xmlns.com/foaf/0.1/name> "Catholic University of the West"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://ma-graph.org/entity/32956416> <http://ma-graph.org/property/grid> <http://www.grid.ac/institutes/grid.448708.7> .
<http://ma-graph.org/entity/32956416> <http://xmlns.com/foaf/0.1/homepage> <http://www.uco.fr/> .
<http://ma-graph.org/entity/32956416> <http://www.w3.org/2000/01/rdf-schema#seeAlso> <http://en.wikipedia.org/wiki/Catholic_University_of_the_West> .
<http://ma-graph.org/entity/32956416> <http://www.w3.org/2002/07/owl#sameAs> <http://en.wikipedia.org/wiki/Catholic_University_of_the_West> .
<http://ma-graph.org/entity/32956416> <http://ma-graph.org/property/paperCount> "239"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/32956416> <http://ma-graph.org/property/citationCount> "2619"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/32956416> <http://purl.org/dc/terms/created> "2016-06-24"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://ma-graph.org/entity/35926432> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://ma-graph.org/class/Affiliation> .
<http://ma-graph.org/entity/35926432> <http://ma-graph.org/property/rank> "12035"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/35926432> <http://xmlns.com/foaf/0.1/name> "Mackay Medical College"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://ma-graph.org/entity/35926432> <http://ma-graph.org/property/grid> <http://www.grid.ac/institutes/grid.476009.c> .
<http://ma-graph.org/entity/35926432> <http://xmlns.com/foaf/0.1/homepage> <http://www.mmc.edu.tw/> .
<http://ma-graph.org/entity/35926432> <http://www.w3.org/2000/01/rdf-schema#seeAlso> <http://en.wikipedia.org/wiki/Mackay_Medical_College> .
<http://ma-graph.org/entity/35926432> <http://www.w3.org/2002/07/owl#sameAs> <http://en.wikipedia.org/wiki/Mackay_Medical_College> .
<http://ma-graph.org/entity/35926432> <http://ma-graph.org/property/paperCount> "840"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/35926432> <http://ma-graph.org/property/citationCount> "6296"^^<http://www.w3.org/2001/XMLSchema#integer> .
<http://ma-graph.org/entity/35926432> <http://purl.org/dc/terms/created> "2016-06-24"^^<http://www.w3.org/2001/XMLSchema#date> .
```

Microsoft Academic Graph: klassen van entiteiten en hun kenmerken en relaties



wikidata



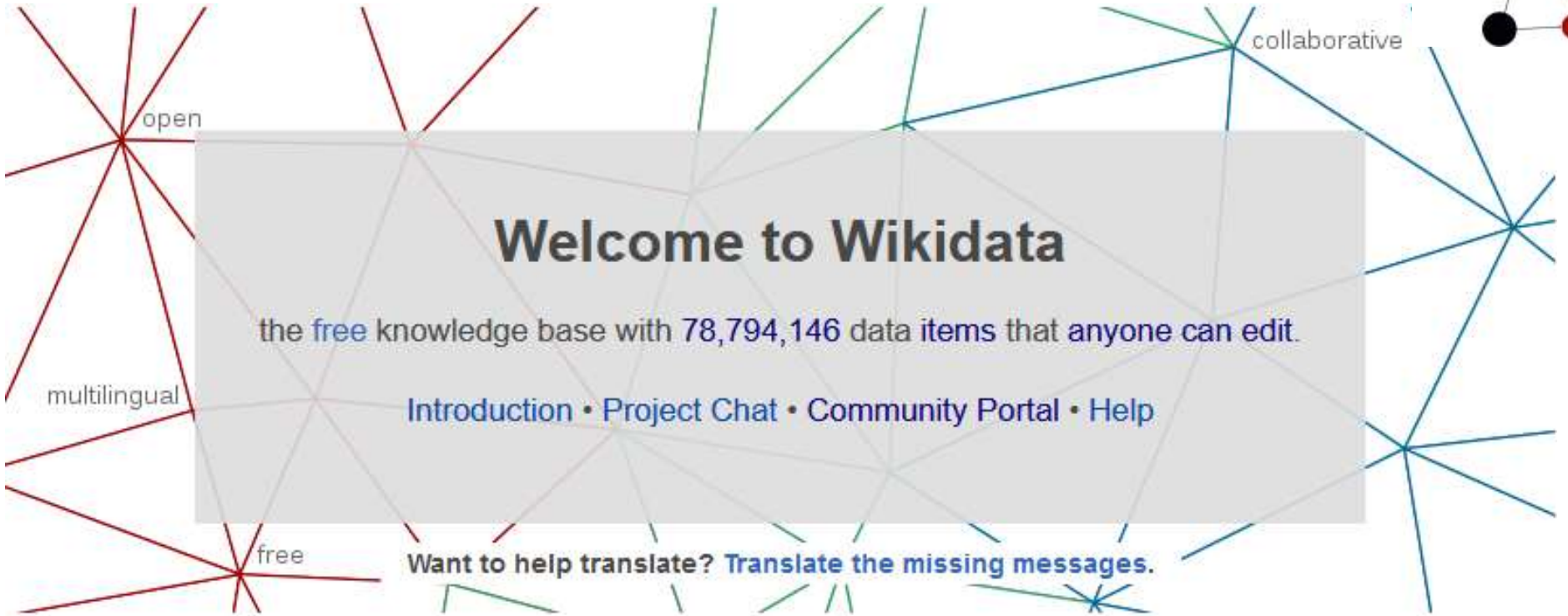
dbpedia is een (linked open data) bron waarvan de tripels - deels automatisch - uit de (tekstuele) inhoud van de Wikipedia zijn afgeleid

wikidata is - omgekeerd - een bron waar feitelijke gegevens rechtstreeks worden ingevoerd en waaraan door andere systemen gegevens ontleend kunnen worden, ook door de Wikipedia

https://www.wikidata.org/wiki/Wikidata:Main_Page

CC BY NC

wikidata



de ruim 20.000 mensen die meewerken aan het vullen van wikidata hebben intussen al meer dan 1 miljard edits gedaan

wikidata heft ook een sparql endpoint: https://query.wikidata.org/

zie voor uitleg: https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service

wikidata items

de items (of entiteiten) zijn genummerd met wikidata identifiers, beginnend met

Q1 = universe

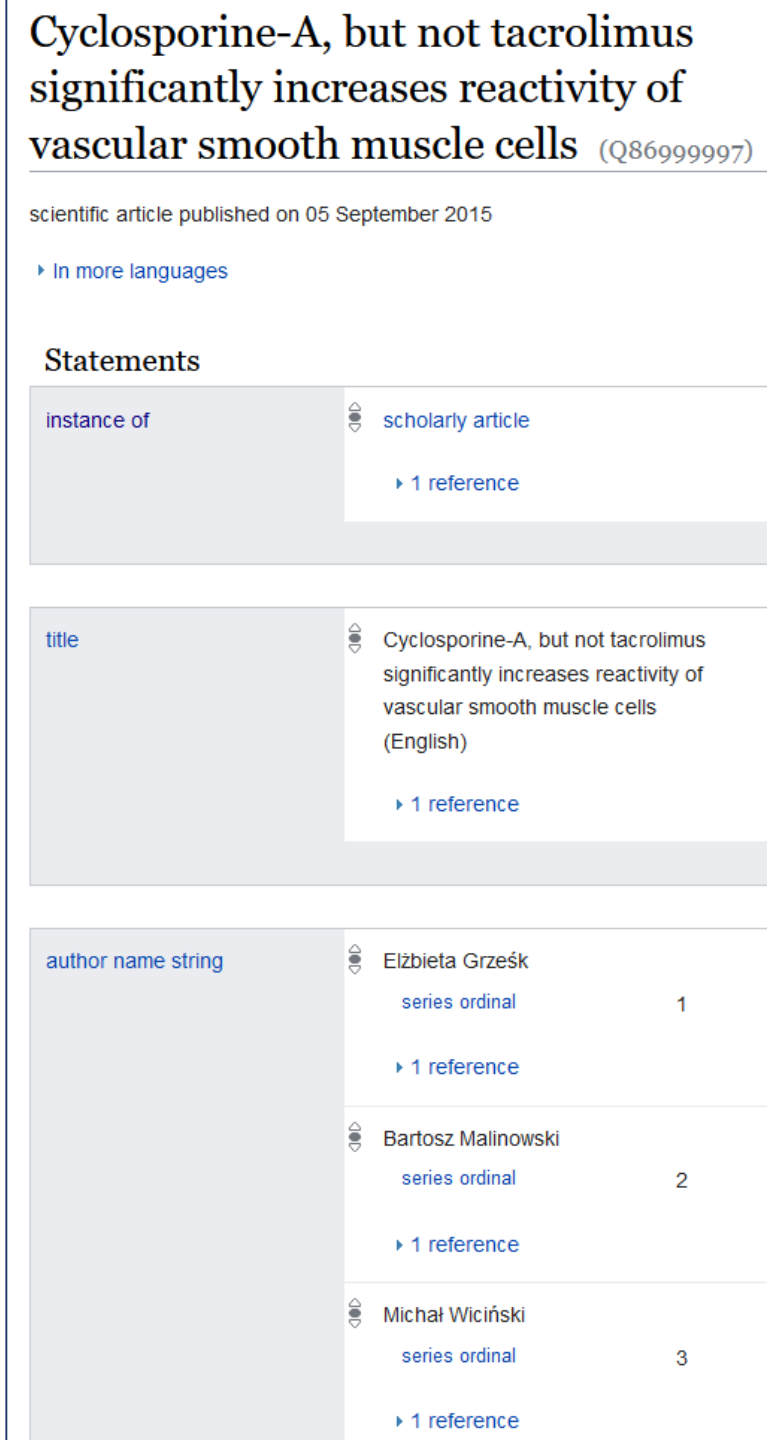
het hoogste nummer is intussen al ver boven 78.794.146, omdat een heleboel nog in eerste bewerking is; zo is er al

Q84263196 = COVID-19
(de aandoening die wordt veroorzaakt door het SARS coronavirus 2)

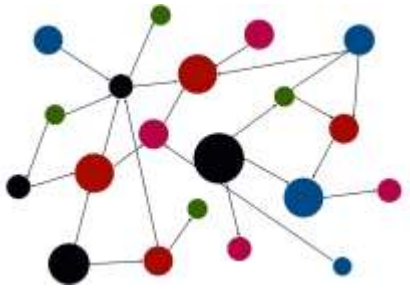
en intussen (29-03-2020)

Q86999997 = Cyclosporine-A, but not tacrolimus signif....
(wetenschappelijk artikel)

zo krijgt elk item een URI, bijv.
https://www.wikidata.org/wiki/Q1 voor universe



wikidata properties



items kunnen eigenschappen (*properties*) hebben, de *predikaten* in RDF

de *statements* in het eerdere voorbeeld koppelen items via zo'n *property* aan een waarde (=*object*);
elk statement representeert dus een *tripel*

de te gebruiken properties zijn ook in wikidata gedefinieerd en krijgen een P-nummer

bijv. P19 voor *place of birth*

of P31 voor *instance of*

place of birth (P19)

most specific known (e.g. city instead of country, or hospital instead of city) birth location of a person, animal or fictional

birthplace | born in | POB | birth place | location born | born at | birth location | location of birth | birth city

▾ In more languages

Configure

Language	Label	Description	Also known as
English	place of birth	most specific known (e.g. city instead of country, or hospital instead of city) birth location of a person, animal or fictional character	birthplace born in POB birth place location born born at birth location location of birth birth city
Dutch	geboorteplaats	de meest specifieke bekende geboorteplaats van een persoon	geboren in plaats van geboorte
German	Geburtsort	Ort, in dem ein Mensch geboren wurde (so genau wie bekannt, z. B. Stadt anstatt von Land)	geboren in
French	lieu de naissance	lieu où la personne est née, le plus spécifique possible (exemple : ville plutôt que pays)	né à naissance née à ville de naissance

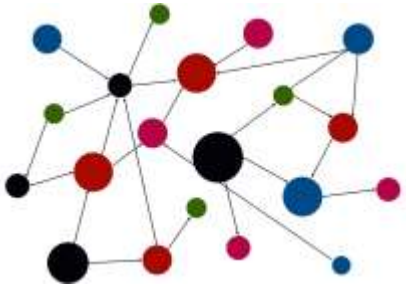
wikidata properties

wikidata bevatte op 2 maart 2020

10 miljard tripels

(statements)

wikidata properties



objecten uit de tripels zijn zelf vaak ook weer wikidata-items, zoals in het eerdere voorbeeld:

Q7000000 - instance of - chemical compound

met URI's zou dit tripel zijn:

<https://www.wikidata.org/wiki/Q70000000> <https://www.wikidata.org/wiki/Property:P31><https://www.wikidata.org/wiki/Q11173>

andere zijn literals, zoals:

Q70000000 - melting point - 108 degree Celsius

of met URI's:

<https://www.wikidata.org/wiki/Q70000000> <https://www.wikidata.org/wiki/Property:P2101> "108_degree_Celsius"

---

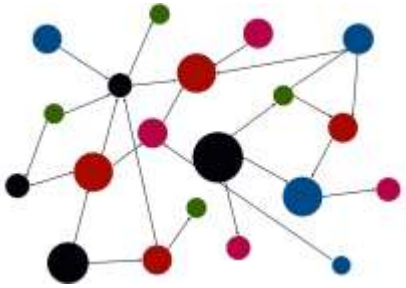
tezamen vormen die meer dan 10 miljard wikidata tripels
in feite een enorme Knowledge Graph

in specifieke disciplines wordt dat soms ook expliciet zo gesteld

artikel (2020): *Wikidata as a knowledge graph for the life sciences*

wikidata properties

de mensen die wikidata vullen, proberen waar mogelijk ook (links naar) vrij beschikbare illustraties toe te voegen, liefst afkomstig uit Wikimedia Commons



Albert Einstein (Q937)

German-born physicist and founder of the theory of relativity

Einstein | Einstein, Albert | A. Einstein

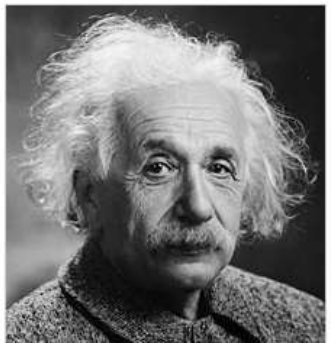
▾ In more languages

Configure

Language	Label	Description
English	Albert Einstein	German-born physicist and founder of the theory of relativity
Dutch	Albert Einstein	Duits-Zwitsers-Amerikaanse theoretisch natuurkundige en uitvinder
German	Albert Einstein	theoretischer Physiker
French	Albert Einstein	physicien d'origine allemande

All entered languages

Statements

instance of	human ▸ 1 reference
image	

https://upload.wikimedia.org/wikipedia/commons/thumb/d/d3/Albert_Einstein_Head.jpg/220px-Albert_Einstein_Head.jpg

Thomson Reu

A linked data feed of Thomson Reuters fi
connections within and across data sets.

the answer company
THOMSON REUTERS

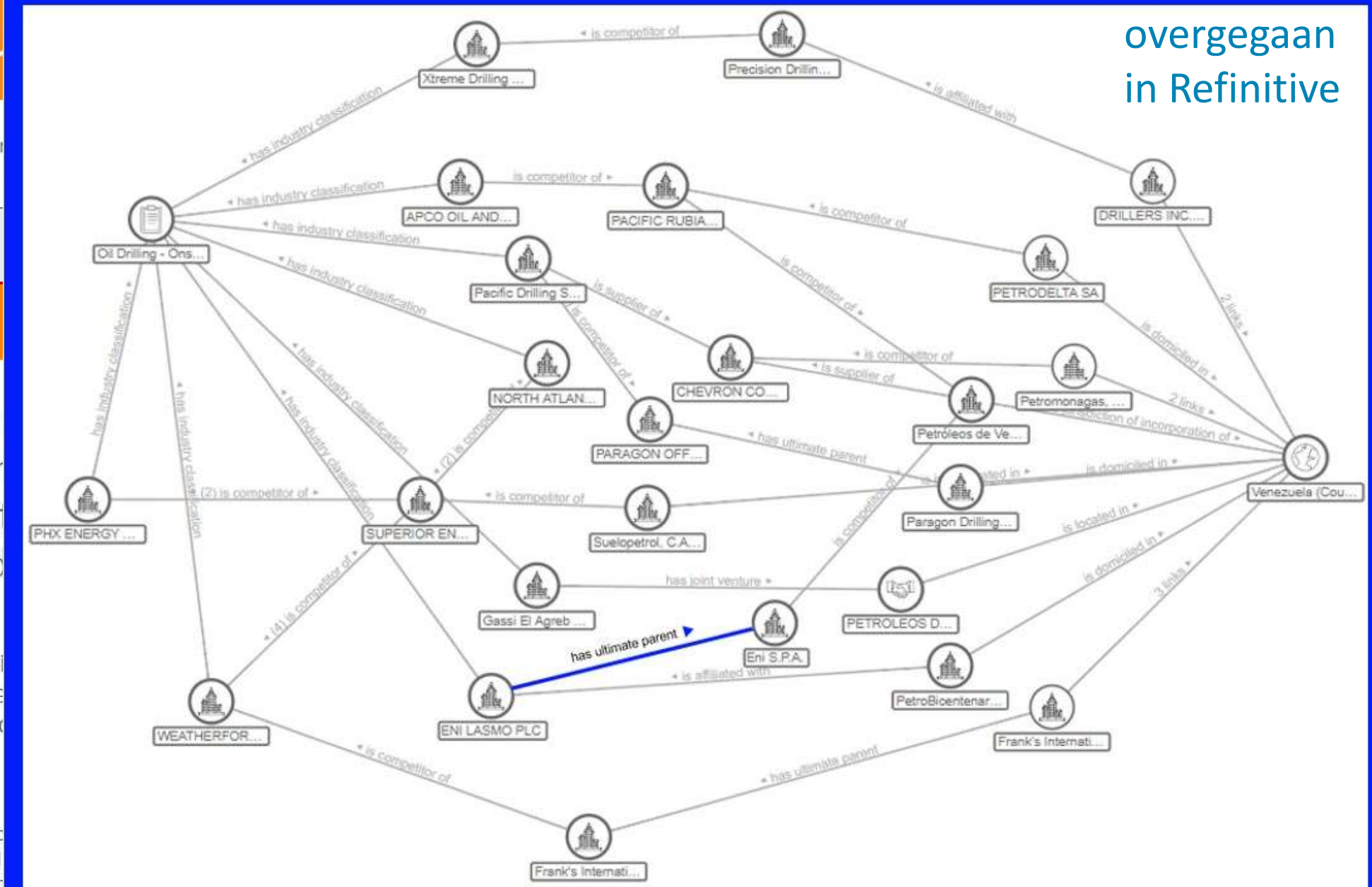
OCT 23, 2017

Thomson Reuters Lau
kind Knowledge Graph
Financial Services cust
accelerate their AI and
Strategies

Graph network of 2 billion relationshi
comprehensive view of the financial e
can uncover new, unexpected or diffi
connected data.

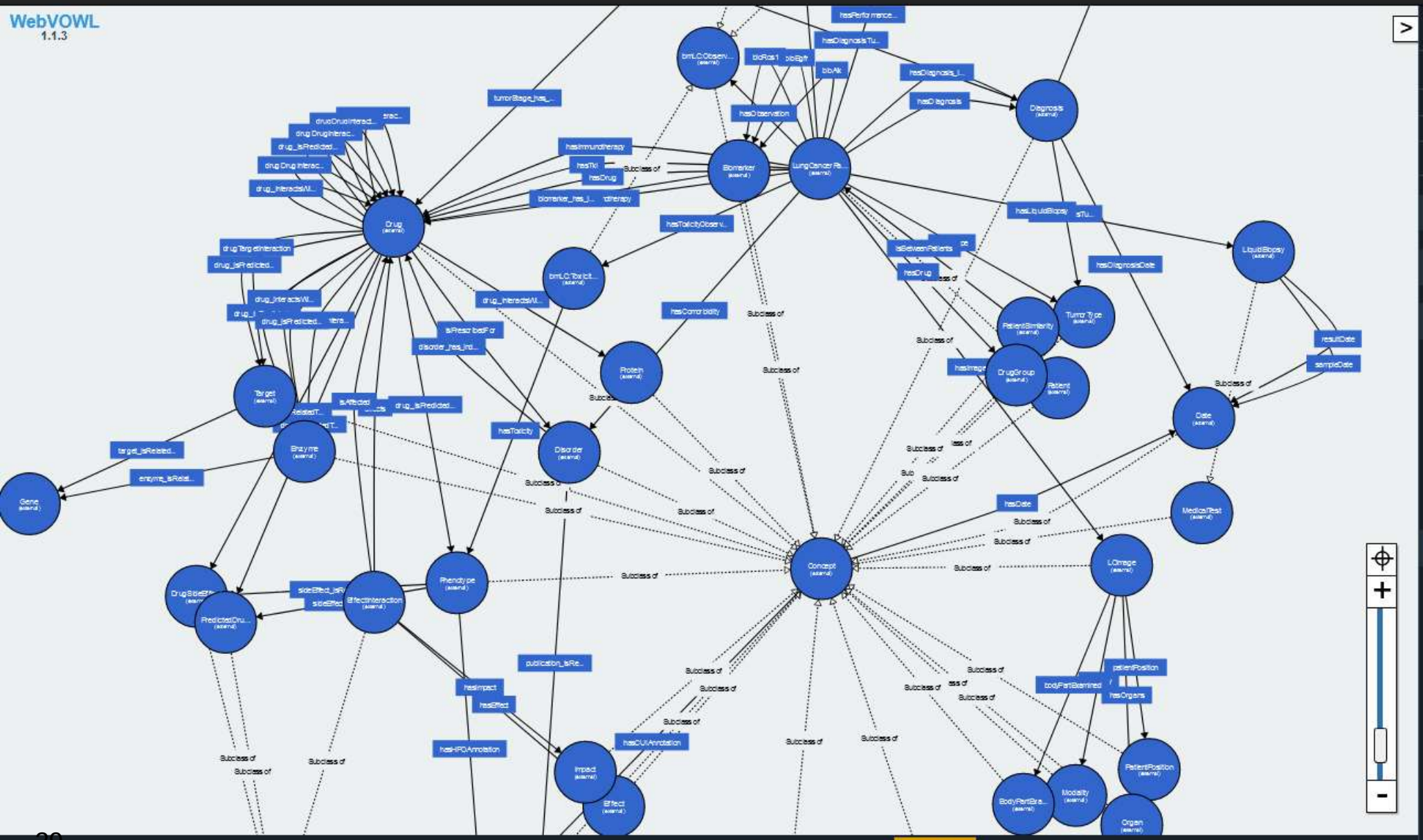
LONDON – Thomson Reuters is the first major info
release a Knowledge Graph Feed for the Financial
verified Big Data management principles and mach
the Thomson Reuters Knowledge Graph provides a
model for customers to leverage, build their enterpr
and benefit in new, connected ways from the bread
Reuters open platform.

overgegaan
in Refinitive



Source: Refinitiv Knowledge Graph

REFINITIV



No title available

No IRI set

Version: --

Author(s): --

Language: undefined

▶ Description

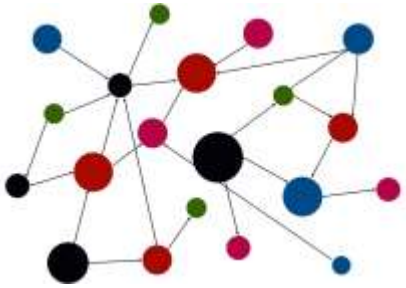
▶ Metadata

▼ Statistics

Classes: *57*
Object prop.: *75*
Datatype prop.: *91*
Individuals: *0*
Nodes: *149*
Edges: *213*

▶ Selection Details

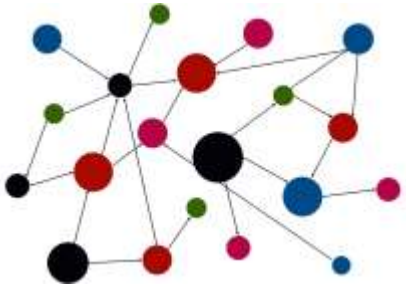
wat is een knowledge graph ?



One common interpretation is that a knowledge graph represents a collection of interlinked descriptions of entities – real-world objects, events, situations or abstract concepts.[3] Unlike ontologies, knowledge graphs, such as Google's Knowledge Graph, often contain large volumes of factual information with less formal semantics. In some contexts, the term *knowledge graph* is used to refer to any knowledge base that is represented as a graph.

https://en.wikipedia.org/wiki/Ontology_(information_science)

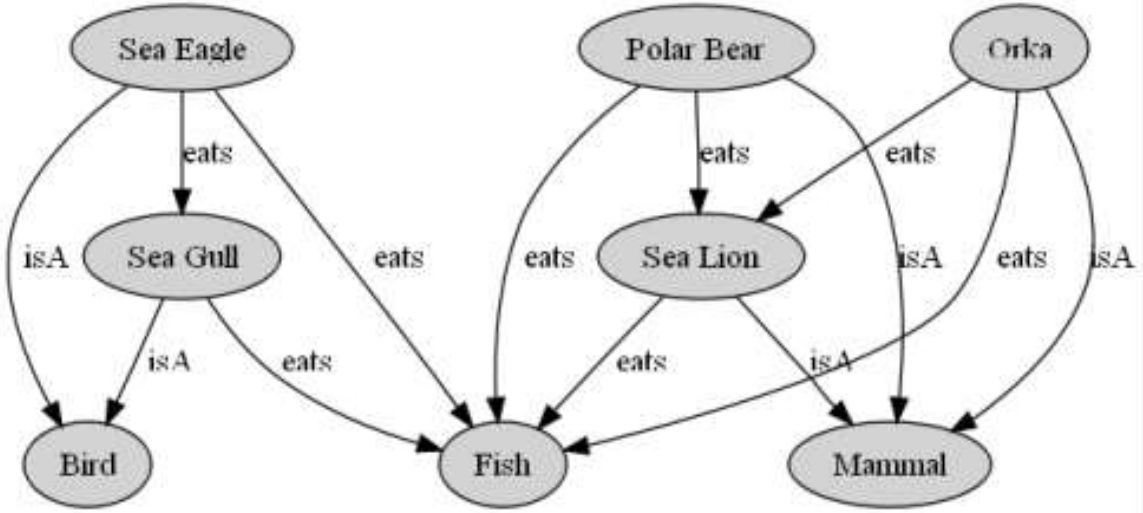
wat is een knowledge graph ?



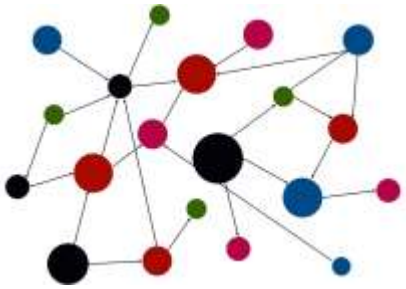
onze werkdefinitie:

- netwerk van concepten waarin hun onderlinge relaties en kenmerken zijn aangegeven
- vastgelegd als RDF-tripels

de structuur van semantische netwerken en ontologieën kan ook op deze manier gepresenteerd / gevisualiseerd worden

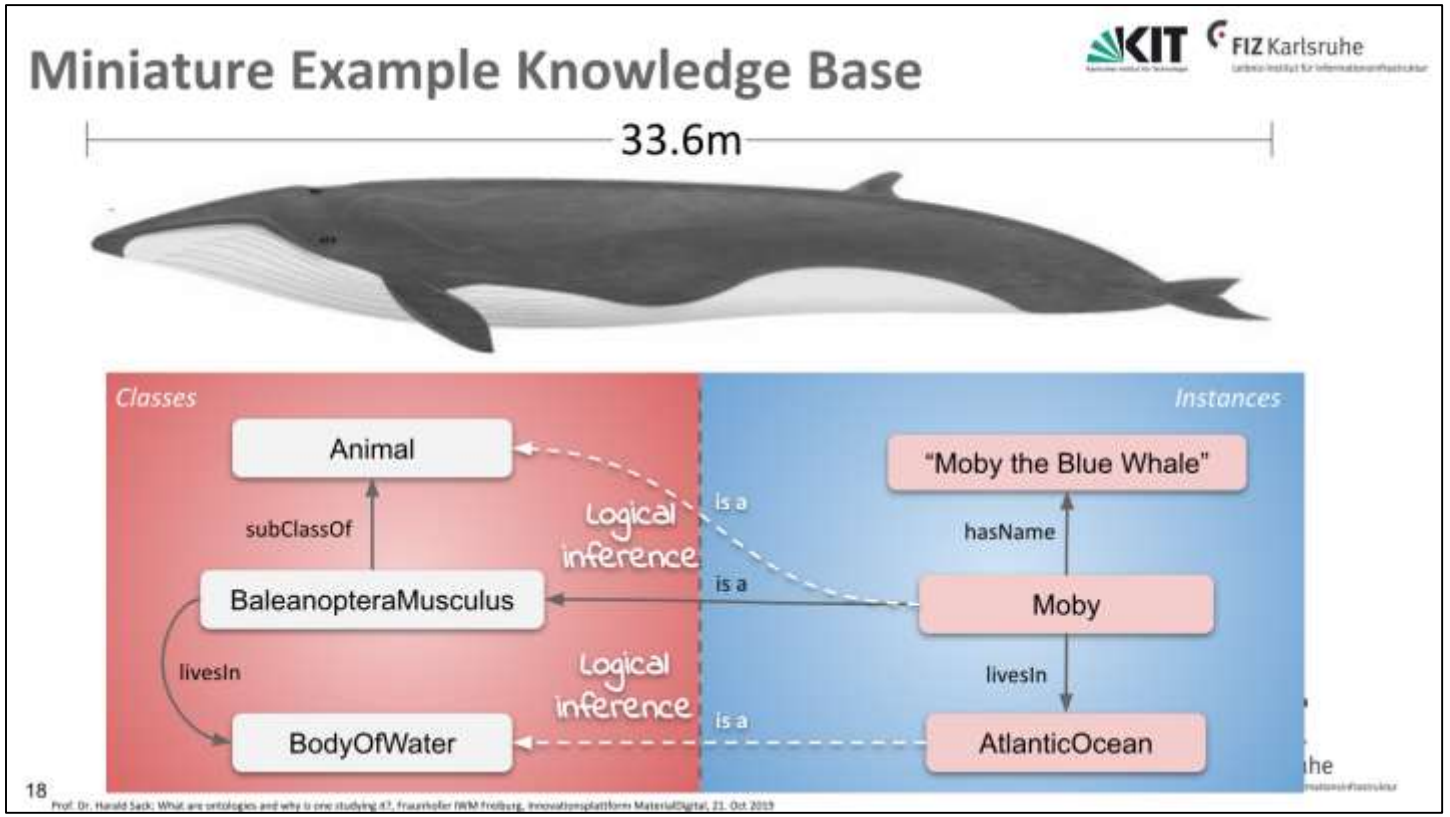


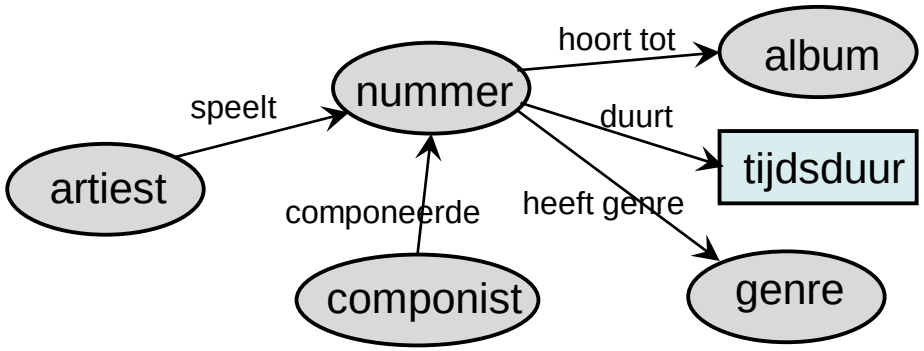
wat is een knowledge graph ?



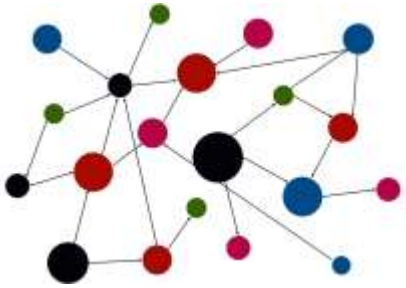
- daarbij liefst ook een vaste datastructuur van *klassen* van concepten en daartussen toegestane relaties *(ontologie)*

instances ↔ *classes*

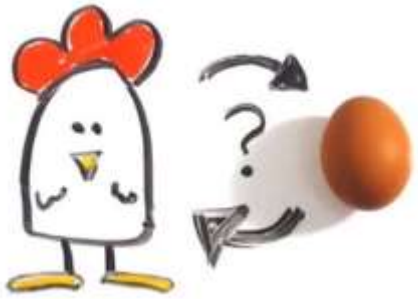




wat is een knowledge graph ?



wat was er eerder? de tripel of the graph?



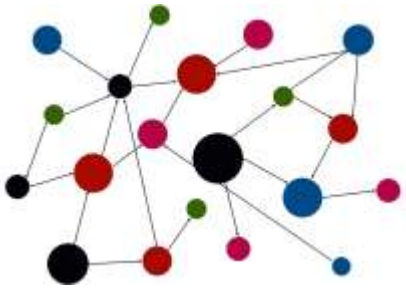
- uit een verzameling tripels kun je een graph maken
- uit een graph kun je tripels afleiden

hoe sla je het op?

- tripels kun je opslaan in zogenaamde *triple stores*
- voor graphs heb je *graph databases*

maar daartussen is eigenlijk alleen technisch verschil

oefening 2:
bouw op papier een knowledge graph



ontwerp in groepjes van 4 of 5

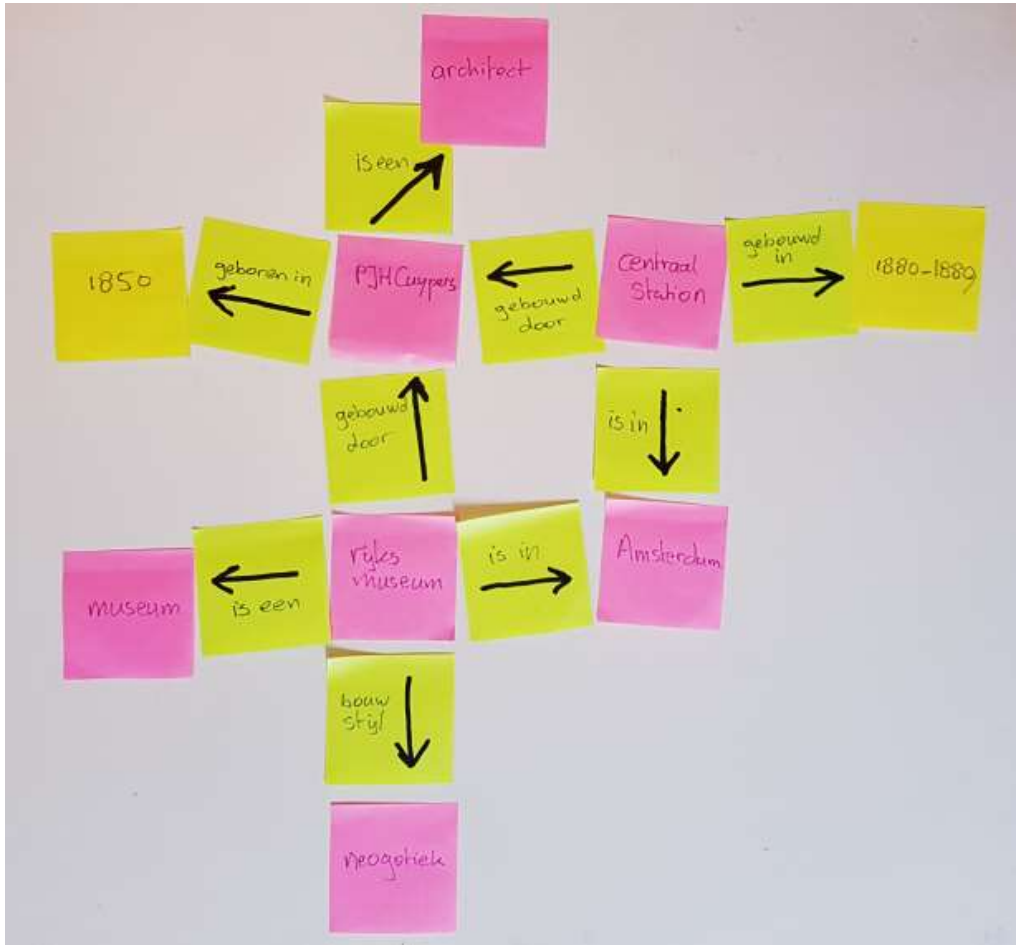
op basis van de informatie

op de uitgedeelde pakjes informatiekaarten

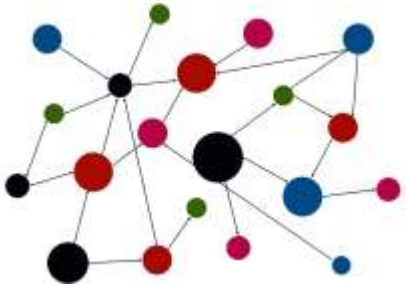
met behulp van post-its

een eerste aanzet tot een knowledge graph

bestand met alle informatieteksten



oefening 3: ontwerp een ontologie voor je knowledge graph



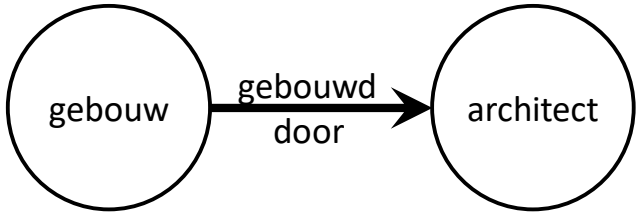
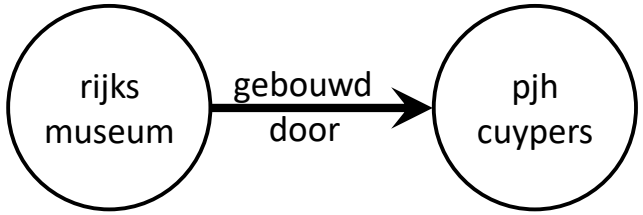
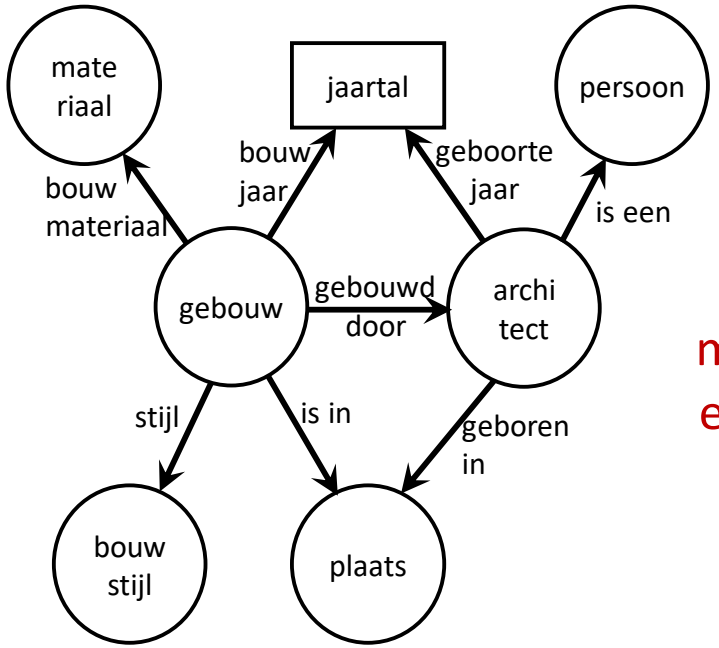
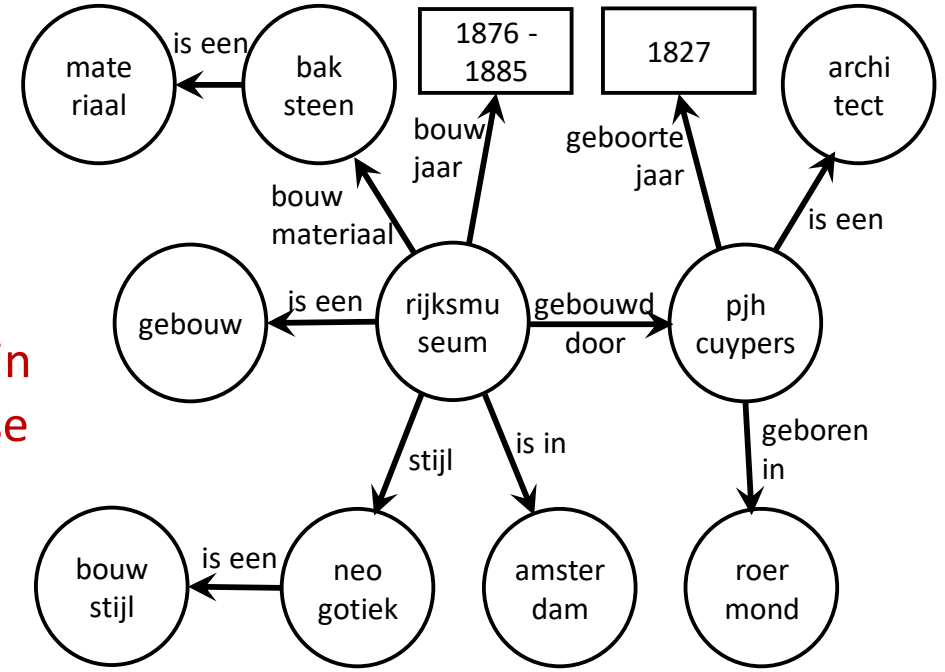
bedenk welke basiselementen (klassen)

en welke relaties daartussen

in je knowledge graph steeds terugkomen

ook als je die verder gaat uitbouwen

en maak daar ook een graph van

	klassen/ontologie	instances
triple		
graph		

vergelijk predikaten met velden in een database

entiteit met URI

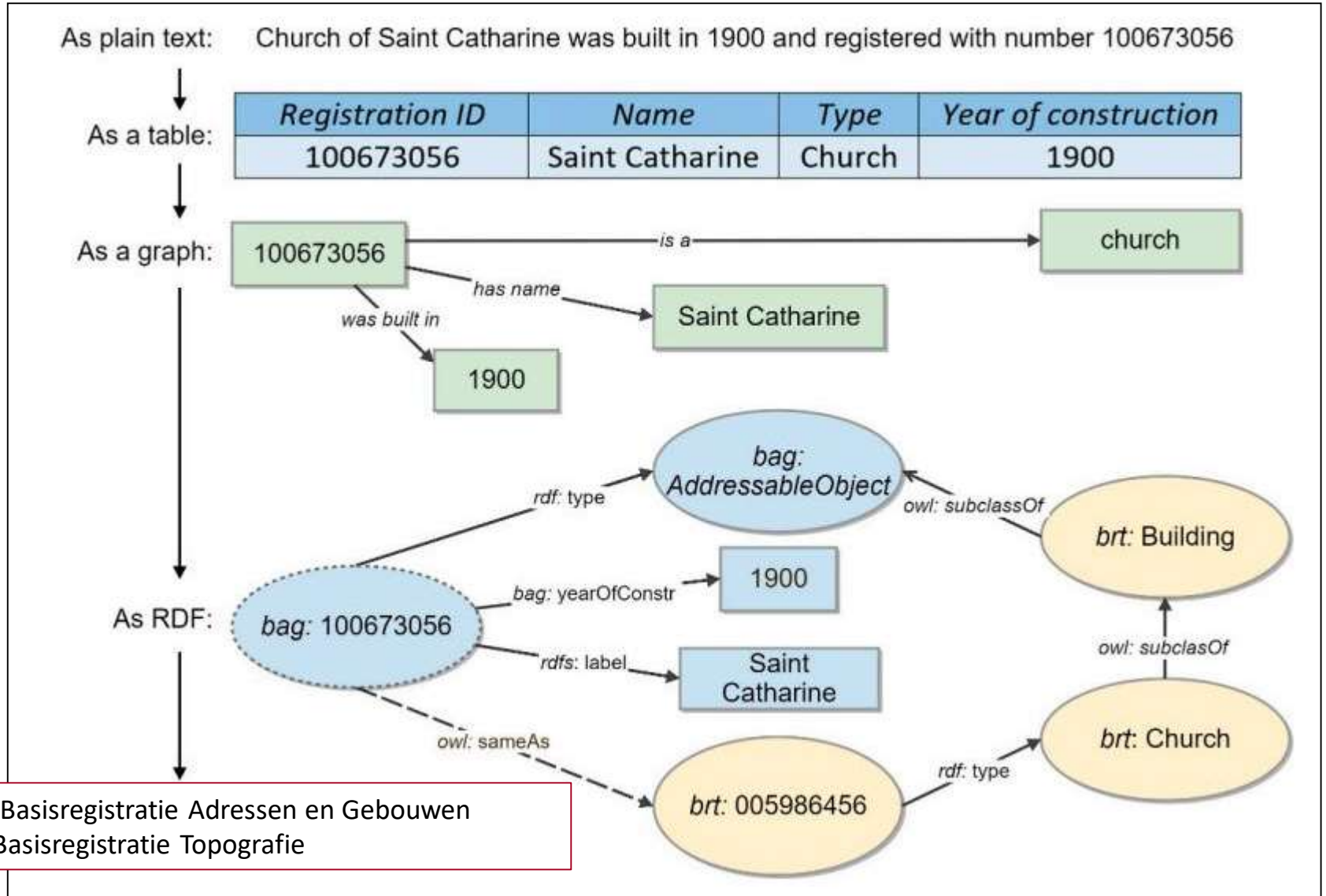
literal

voorbeeld uit:

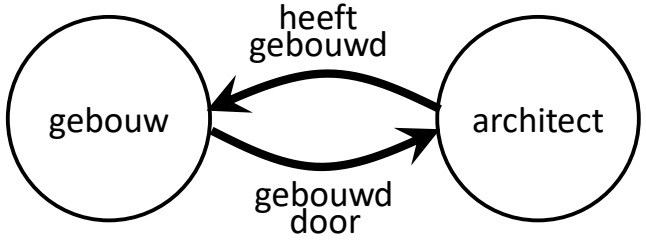
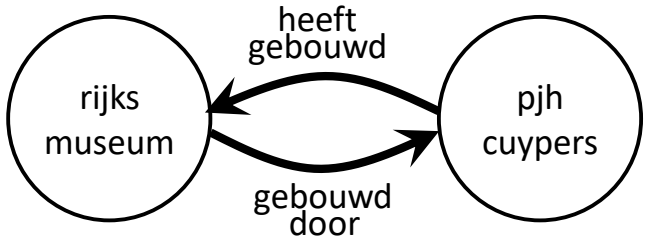
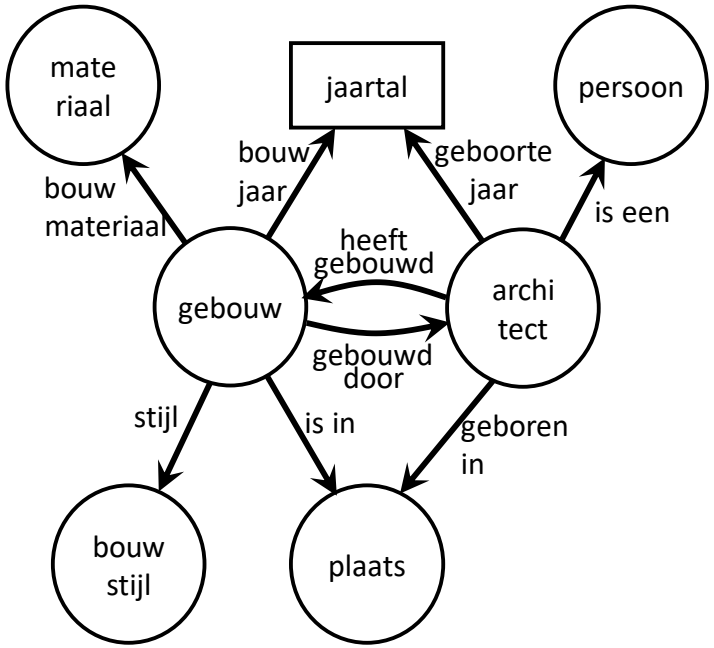
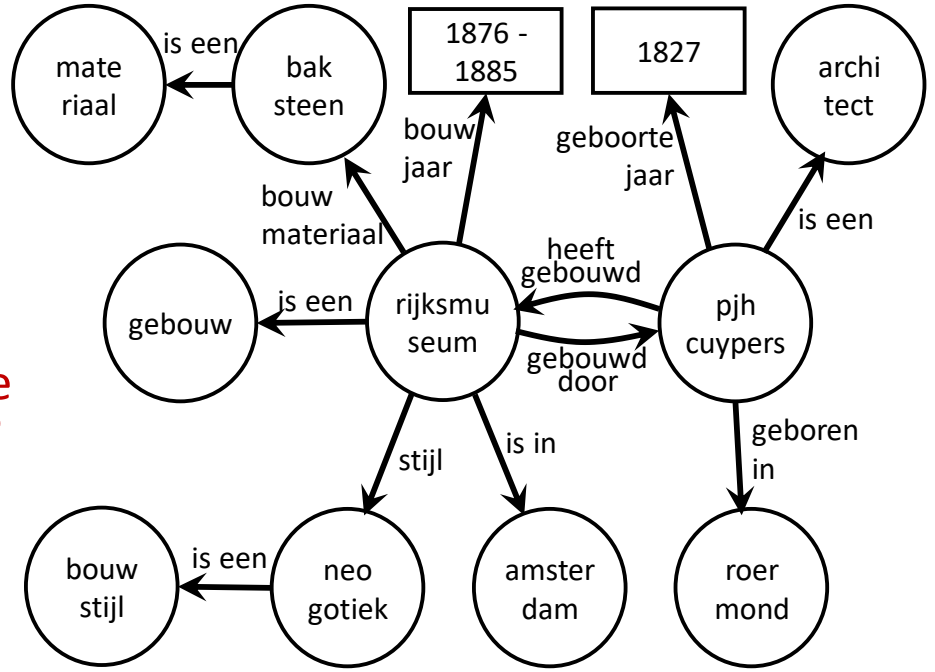
Information **2019**, *10*, 310
doi:10.3390/info10100310

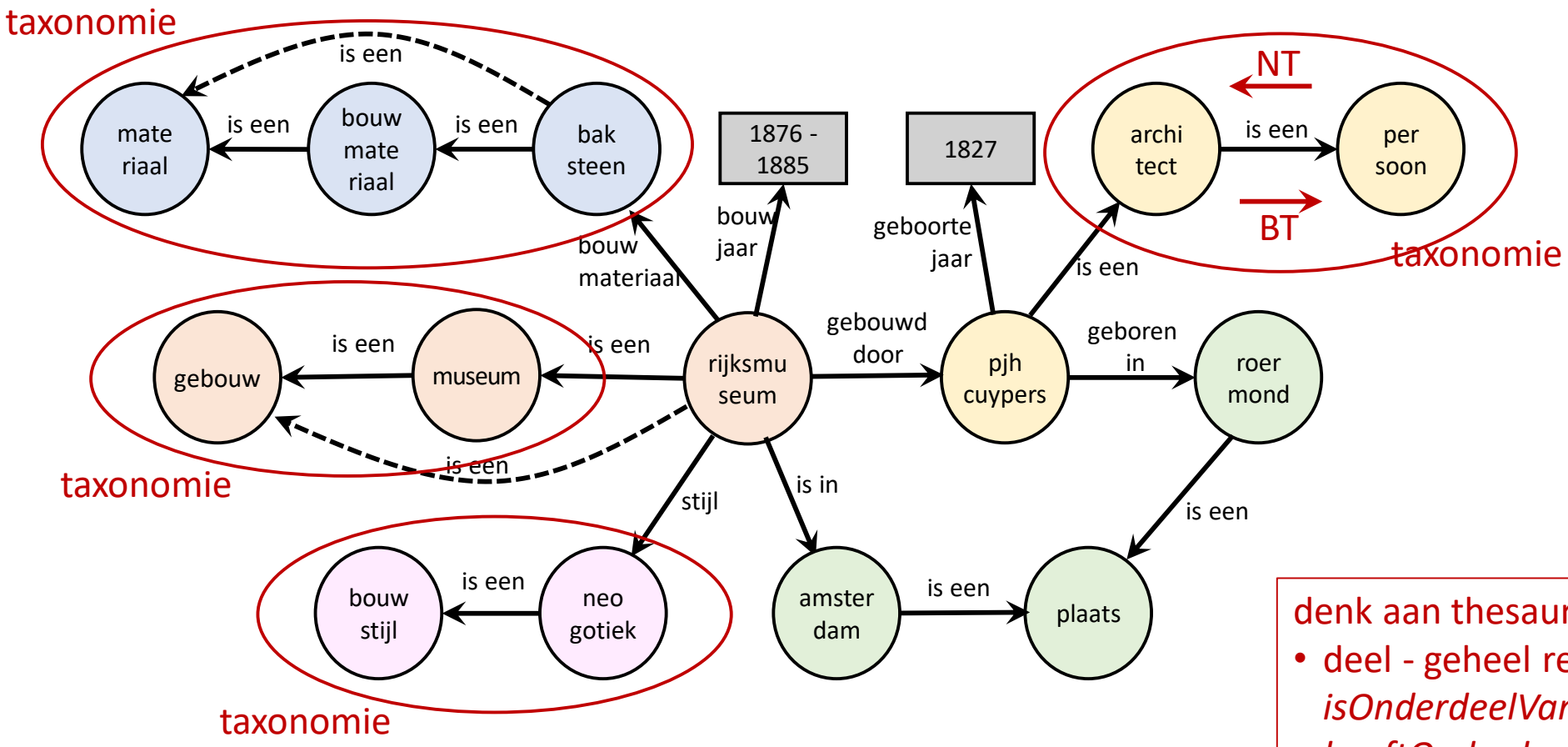
Kadaster Knowledge Graph: Beyond the Fifth Star of Open Data

Stanislav Ronzhin [1,*], Erwin Folmer [2,3], Pano Maria [3], Marco Brattinga [3], Wouter Beek [4], Rob Lemmens [1] and Rein van't Veer [3]



bag = Basisregistratie Adressen en Gebouwen
brt = Basisregistratie Topografie

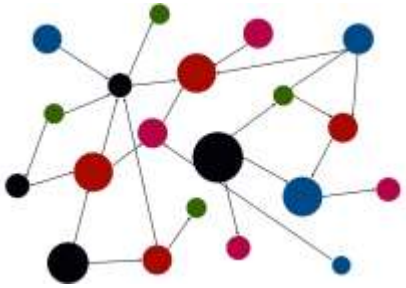
	klassen/ontologie		instances
triple		voorkeur voor een richting?	
graph		zie je nog meer omkeerbare richtingen?	



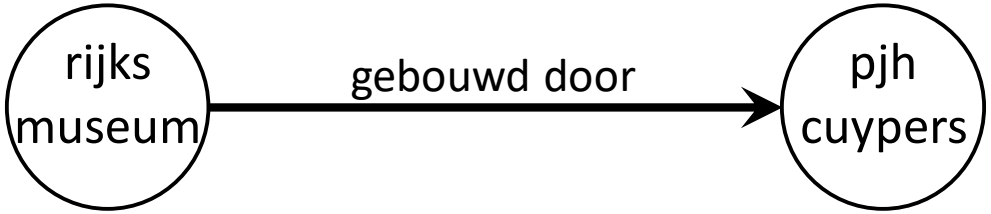
denk aan thesaurusrelaties:

- deel - geheel relaties
  isOnderdeelVan
  heeftOnderdeel
- geslacht - soort relaties
  isEen

wat is een knowledge graph ?

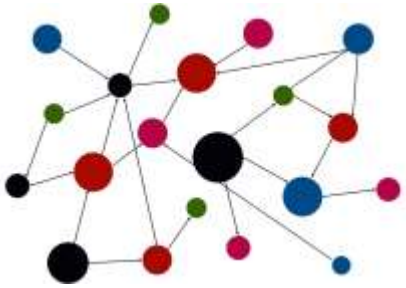


jargon



tripel	subject entity	predicate property relation	object entity
graph	node knoop	edge pijl	node knoop

standaardisatie van predikaten

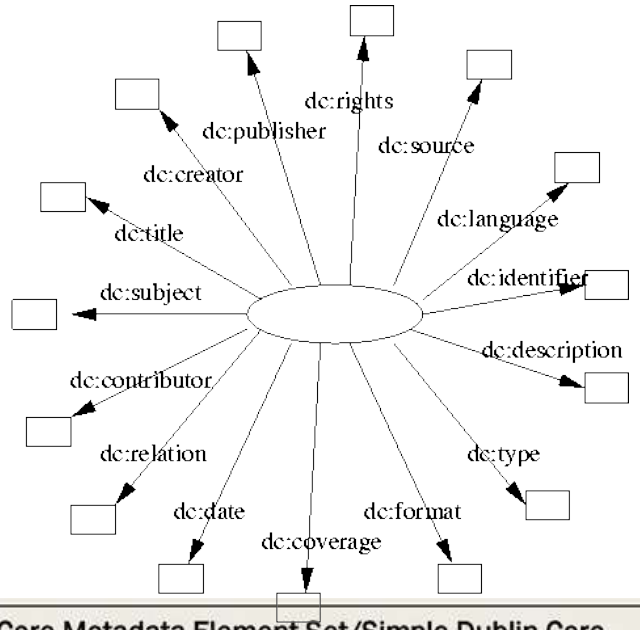


voor interne consistentie en uitwisselbaarheid (linked open data) graag standaardisatie van predikaten (en met URI's verwijzen naar definities)

voorbeelden van standaarden en definities:

- Dublin Core
- schema.org *(web - algemeen)*
- cidoc-crm *(erfgoed)*
- foaf *(personen)*
- phunware ontology *(bedrijven en marketing)*
- ....

je mag "shoppen" om benodigde predikaten te verzamelen en dat aanvullen met voor jouw organisatie essentiële relaties die elders ontbreken



Dublin Core Metadata Element Set/Simple Dublin Core	
Contributor	An entity responsible for making contributions to the resour
Coverage	The spatial or temporal topic of the resource, the spatial app or the jurisdiction under which the resource is relevant.
Creator	An entity primarily responsible for making the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Description	An account of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Language	A language of the resource.
Publisher	An entity responsible for making the resource available.
Relation	A related resource.
Rights	Information about rights held in and over the resource.
Source	A related resource from which the described resource is derived.
Subject	The topic of the resource.
Title	A name given to the resource.
Type	The nature or genre of the resource.

FOAF Core

- Agent
- Person
- name
- title
- img
- depiction (depicts)
- familyName
- givenName
- knows
- based_near
- age
- made (maker)
- primaryTopic (pr
- Project
- Organization
- Group
- member
- Document
- Image

Social Web

- nick
- mbox
- homepage
- weblog
- openid
- jabberID
- mbox_sha1sum
- interest
- topic_interest
- topic (page)
- workplaceHomepage
- workInfoHomepage
- schoolHomepage
- publications
- currentProject
- pastProject
- account
- OnlineAccount
- accountName
- accountServiceHomepage
- PersonalProfileDocument
- tipjar
- sha1
- thumbnail
- logo

schema.org

Person

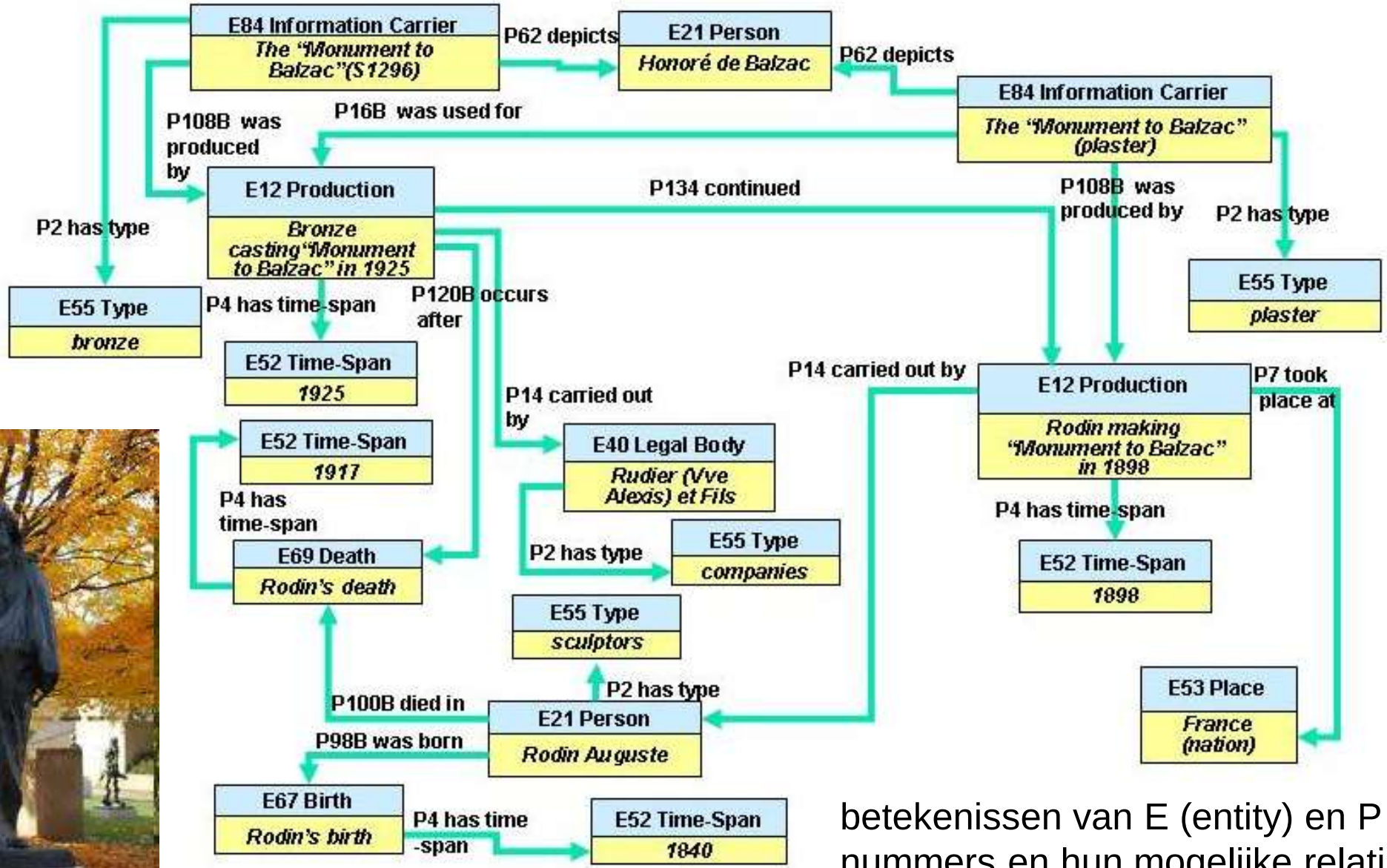
Thing > Person

A person (alive, dead, undead, or fictional).

Property	Expected Type
Properties from Person	
additionalName	Text
address	PostalAddress or Text
affiliation	Organization
alumniOf	EducationalOrganization or Organization
award	Text
birthDate	Date
birthPlace	Place
brand	Brand or Organization
callSign	Text
children	Person
colleague	Person or URL
contactPoint	ContactPoint

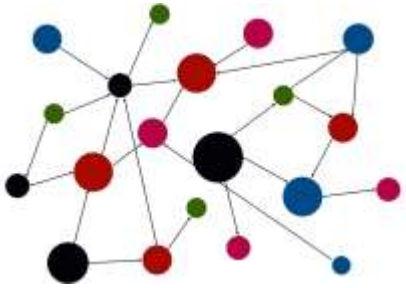
Rodin: Balzac

voorbeeld van de relaties tussen concepten m.b.t. “standbeeld van Balzac” door Rodin *[CIDOC-CRM]*



betekenissen van E (entity) en P (property) nummers en hun mogelijke relaties zijn in CIDOC-CRM vastgelegd

oefening 4:
bouw de ontologie uit oefening 3 in gra.fo

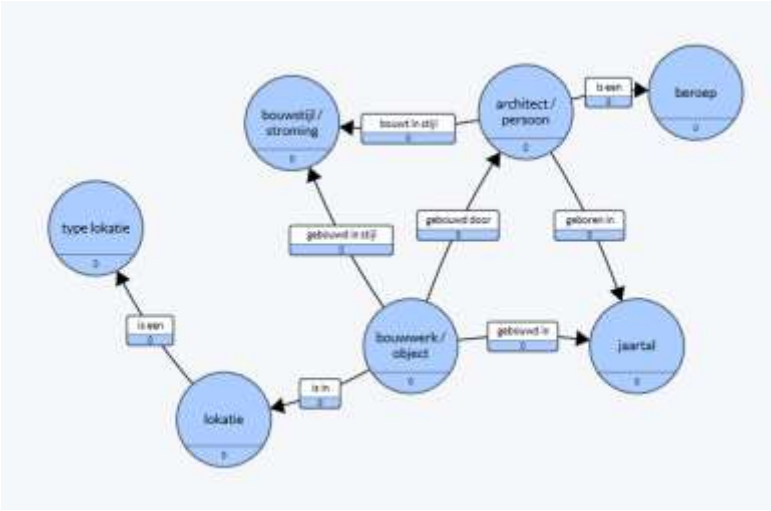


gebruik de online applicatie gra.fo

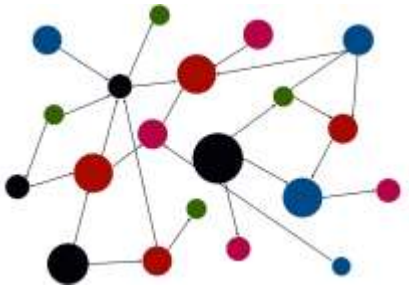
om een digitale versie van de in oefening 3

bedachte ontologie te bouwen

[handleiding]



oefening 5: bouw een stuk van je knowledge graph in gra.fo



GRAFO

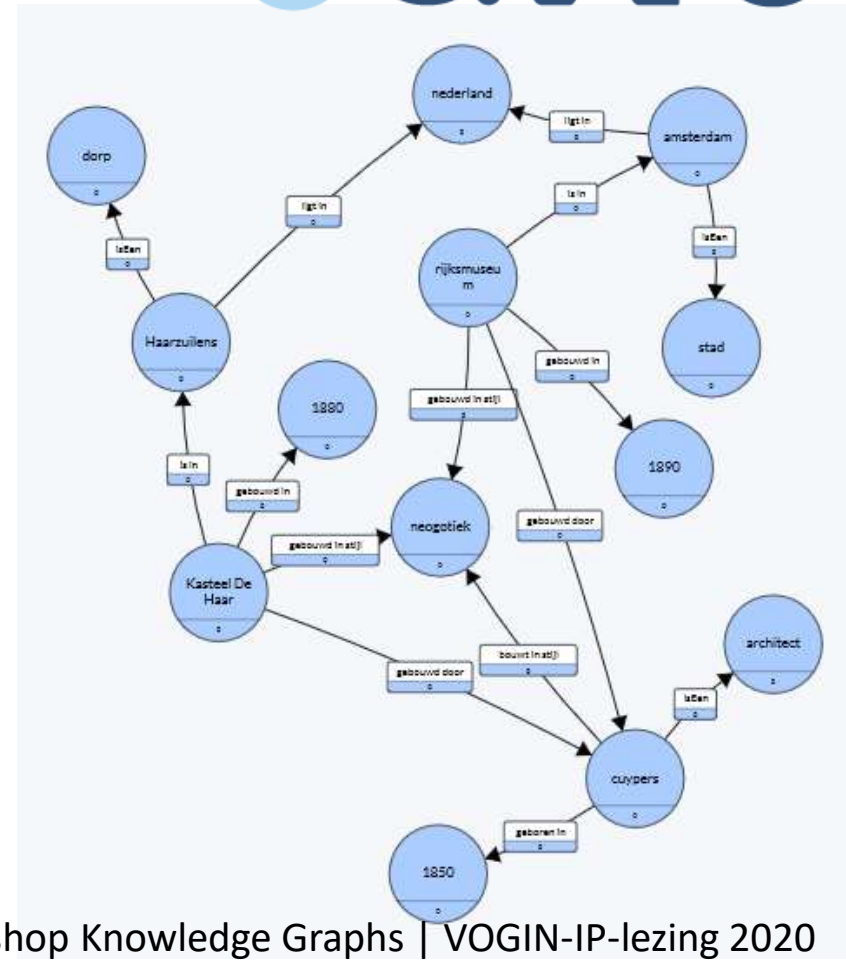
gebruik de online applicatie gra.fo

om een digitale versie te bouwen van

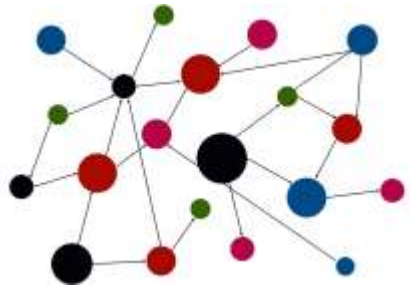
(een deel van) de in oefening 2 bedachte

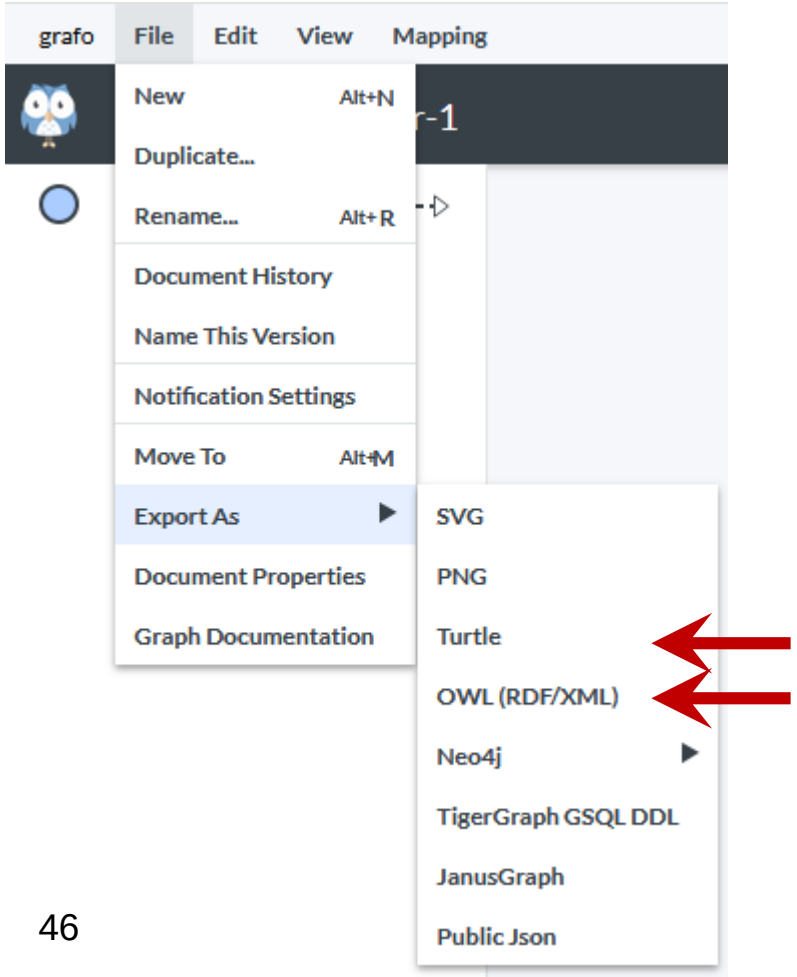
knowledge graph

[handleiding]



oefening 6: exporteren van je knowledge graph uit gra.fo





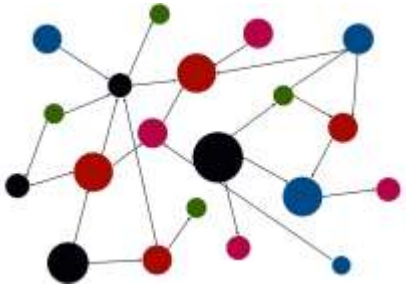
exporteer je knowledge graph successievelijk

- als Turtl
- als OWL (RDF/XML)

en bekijk de gedownloade files in kladblok

zijn die een beetje begrijpelijk?

toepassing van knowledge graphs



een paar voorbeelden bij bedrijven / organisaties

- Zalando (Fashion Knowledge Graph)
- AirBNB (context van bestemmingen) *[video]* *[artikel in IP]*
- Kadaster Knowledge Graph
- FactForge (Data about People, Organizations and Locations)
- Thomson Reuters / Refinitiv (Financial Services)
- Pinterest (taste/interest graph)
- Walmart (retail product graph)

automatisch genereren van tripels door beeldherkenning van elementen uit een foto, teneinde de hele scene in een knowledge graph te vangen

arXiv:2001.02314v1 [cs.CV] 7 Jan 2020

Bridging Knowledge Graphs to Generate Scene Graphs

Alireza Zareian, Svebor Karaman, and Shih-Fu Chang
Columbia University, New York, NY, USA
{az2407,sk4089,sc250}@columbia.edu

Abstract

Scene graphs are powerful representations that encode images into their abstract semantic elements, i.e. objects and their interactions, which facilitates visual comprehension and explainable reasoning. On the other hand, commonsense knowledge graphs are rich repositories that encode how the world is structured, and how general concepts interact. In this paper, we present a unified formulation of these two constructs, where a scene graph is seen as an image-conditioned instantiation of a commonsense knowledge graph. Based on this new perspective, we re-formulate scene graph generation as the inference of a bridge between the scene and commonsense graphs, where each entity or predicate instance in the scene graph has to be linked to its corresponding entity or predicate class in the commonsense graph. To this end, we propose a heterogeneous graph inference framework allowing to exploit the rich structure within the scene and commonsense at the same time. Through extensive experiments, we show the proposed method achieves significant improvement over the state of the art.

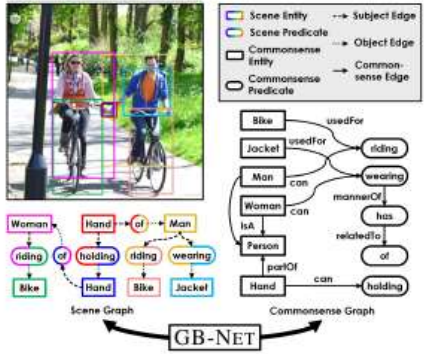
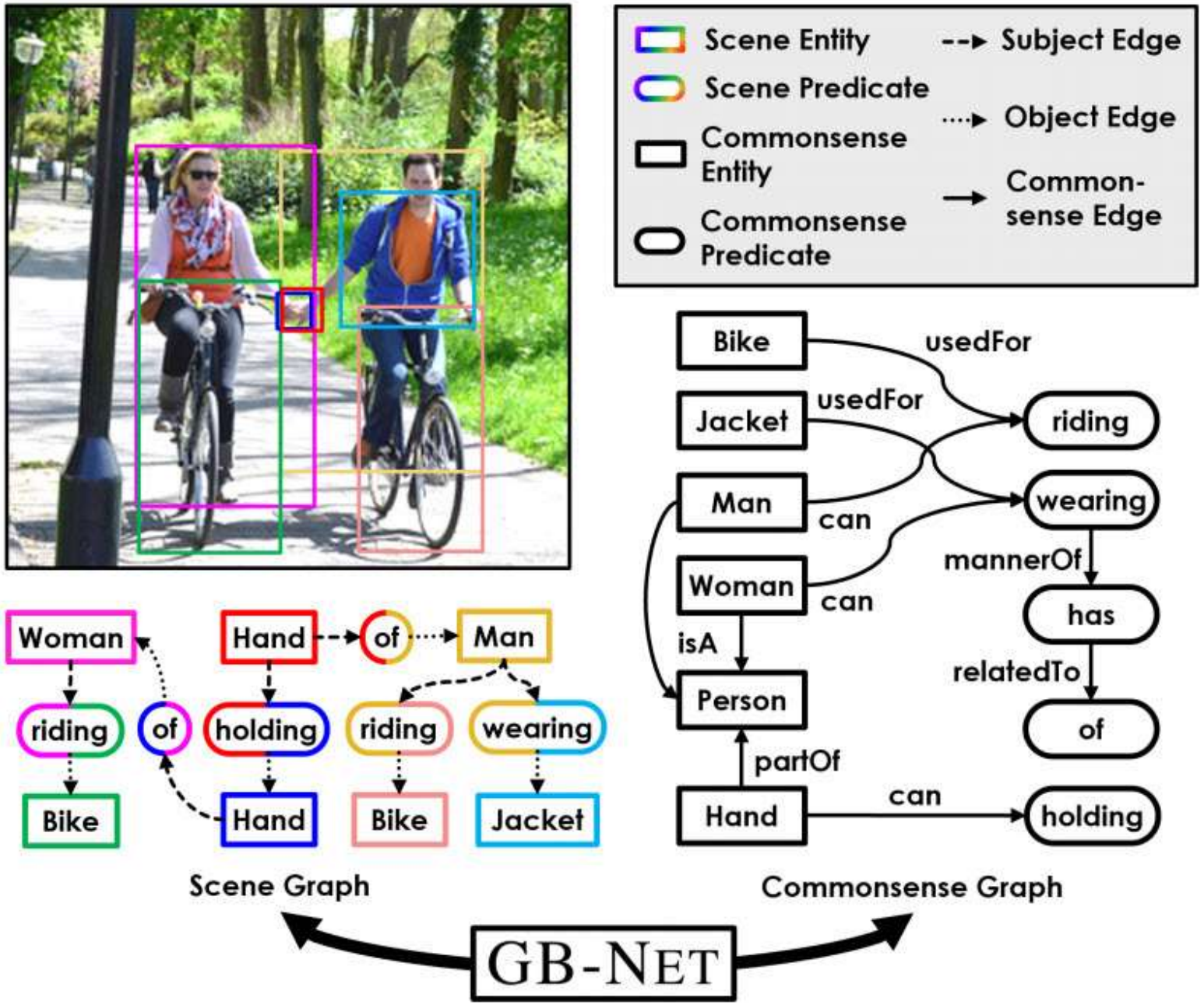


Figure 1. Left: An example of a Visual Genome image and its ground truth scene graph. Right: A relevant portion of the commonsense graph. In this paper we formulate the task of Scene Graph Generation as the problem of creating a bridge between these two graphs. Such bridge not only classifies each scene entity and predicate, but also creates an inter-connected heterogeneous graph whose rich structure is exploited by our method (GB-NET).

1. Introduction

Extracting structured, symbolic, semantic representations from data has a long history in Natural Language Processing (NLP), under the umbrella terms of *semantic parsing* at the sentence level [8, 7] and *knowledge extraction* at the document level [20, 34]. The resulting *semantic graphs* or *knowledge graphs* have many applications such as question answering [6, 15] and information retrieval [5, 41]. In captioning [40, 39], image retrieval [13, 31], and image synthesis [12]. In VQA for instance, SGG not only improves performance, but also promotes interpretability and enables explainable reasoning [32].

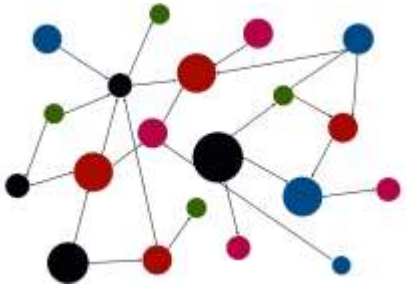
Although several methods have been proposed, the state-of-the-art performance for SGG is still far from accept-



[get the free PDF version]

[view the book launch webinar (22-04-2020)]

nog wat te lezen / bestuderen:



What exactly is a Knowledge Graph: Using Semantic Enrichment to connect the dots
https://www.scibite.com/news/what-exactly-is-a-knowledge-graph-using-semantic-enrichment-to-connect-the-dots/

Knowledge Management, Knowledge Graphs, and Ontologies
https://www.synaptica.com/knowledge-management-knowledge-graphs-and-ontologies/

Interacting with Your Graph (or Why Knowledge Graphs? Part 3)
https://www.synaptica.com/why-knowledge-graphs-part-3/

Connecting the dots in your business using Knowledge Graphs
https://www2.deloitte.com/nl/nl/pages/risk/articles/connecting-the-dots-in-your-business-using-knowledge-graphs.html

How to Build a Knowledge Graph in Four Steps: The Roadmap From Metadata to AI https://enterprise-knowledge.com/how-to-build-a-knowledge-graph-in-four-steps-the-roadmap-from-metadata-to-ai/

Teaching Computers to Answer Complex Questions
https://www.amazon.science/blog/teaching-computers-to-answer-complex-questions

QuickGraph#9 The fashion Knowledge Graph. Inferencing with Ontologies in Neo4j
https://jbarrasa.com/2019/11/25/quickgraph9-the-fashion-knowledge-graph-inferencing-with-ontologies-in-neo4j/

User-Specific Knowledge Graphs to Support Queries and Predictions (Google patent)
http://www.seobythesea.com/2019/11/user-specific-knowledge-graphs/

Knowledge Graphs: How should AI explicitly represent knowledge? (Online cursus van Stanford University)
https://web.stanford.edu/class/cs520/