

Knowledge Graphs in Machine Learning

Ozer Ozdikis

oyer.ozdikis@ntnu.no

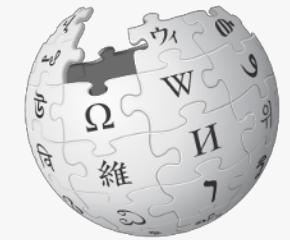
<https://oozdikis.github.io>

Outline and Objectives

- Part 1: Knowledge Graphs – Introduction
- Part 2: Machine Learning and Knowledge Graphs
- Part 3: Special Issue: Entity Linking using Knowledge Graphs

Knowledge Graph: Introduction

Knowledge for Humans



WIKIPEDIA
The Free Encyclopedia



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link

Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

In other projects
Wikimedia Commons

Languages

العربية
Deutsch
Español
Bahasa Indonesia
Latina
Norsk

Article Talk

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)
Read Edit View history Search Wikipedia

University of Bergen

From Wikipedia, the free encyclopedia

Coordinates: 60°23'17.11"N 5°19'22.34"E

The **University of Bergen** (Norwegian: *Universitetet i Bergen*) is a public university located in **Bergen**, Norway. The university today serves approximately 17,000 students, and is one of eight universities in Norway.

Contents [hide]

- 1 History
- 2 Organization
 - 2.1 Ranking
 - 2.2 Tuition
- 3 Faculties and academia at the University of Bergen
 - 3.1 Faculty of Humanities
 - 3.2 Faculty of Law
 - 3.3 Faculty of Mathematics and Natural Sciences
 - 3.4 Faculty of Medicine
 - 3.5 Faculty of Psychology
 - 3.6 Faculty of Social Sciences
- 4 Notable academics and faculty
 - 4.1 Notable alumni
- 5 Other notes
- 6 See also
- 7 References
- 8 External links

University of Bergen	
Universitetet i Bergen	
	Latin: <i>Universitas Bergensis</i>
Type	Public University
Established	1946 (1825)
Rector	Dag Rune Olsen
Administrative staff	3,200 (2007)
Students	17,000 (2015)
Location	Bergen, Norway
Affiliations	Aurora EUA WUN Coimbra Group Utrecht Network www.uib.no
Website	

History [edit]

Although the university was founded as late as 1946, academic activity had taken place in **Bergen** since the founding of **Bergen Cathedral School** in 1153, the **Seminarium Fredericianum** in 1750 and the establishment of the **Royal Norwegian Naval Academy** in 1817. Academia and higher education would also be significantly advanced in the city with the establishment of **Bergen Museum**, later renamed **University Museum of Bergen**, in 1825. Founded by **Wilhelm Frimann Christie** and **Jacob Neumann**, the museum became a venue for both research and education specialized on natural science, and featured prominent researcher like **Michael Sars**, **Daniel Cornelius Danielsen** and **Fridtjof Nansen**.^[1]

Bergen would eventually become a city with several arenas for higher education and research with the **Geophysical Institute** being established in 1917, the **Chr. Michelsen Institute** in 1930, the **Norwegian School of Economics** in 1936 and finally the university in 1946.^[2]



University Museum of Bergen

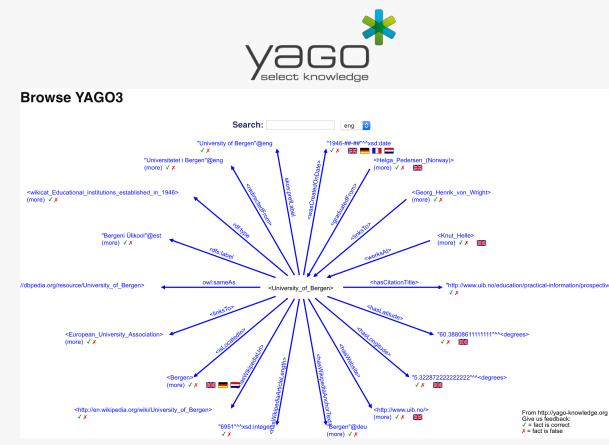
Knowledge Graph: Knowledge for Machines



<p>WIKIDATA</p> <p>Main page</p> <p>Community portal</p> <p>Project chat</p> <p>Create a new item</p> <p>Create a new Lexeme</p> <p>Recent changes</p> <p>Random item</p> <p>Query Service</p> <p>Nearby</p> <p>Help</p> <p>Donate</p> <p>Tools</p> <p>What links here</p> <p>Related changes</p> <p>Special pages</p> <p>Permanent link</p> <p>Page information</p> <p>Concept URI</p> <p>Cite this page</p>	<p>Item Discussion</p> <p>University of Bergen (Q204457)</p> <p>university in Norway</p> <p>Universitet i Bergen Universitas Bergensis UIB Bergen University</p> <p>edit</p> <p>Wikipedia (30 entries) edit</p> <p>ar جامعہ بیرونی az Bergen Universiteti be Бергенскі ўніверсітэт ca Universitat i Bergen da Universitetet i Bergen de Universität Bergen en University of Bergen es Universidad de Bergen et Bergen Ülikool fa دانشگاه برگن fi Bergenin yliopisto fr Université de Bergen fy Universiteit fan Bergen hy Բերգենի համալսարակ id Universitas Bergen it Università di Bergen ja ベルゲン大学 kk Берген университеті ko 베르겐 대학교 la Universitas Bergensis nl Universiteit van Bergen nn Universitetet i Bergen no Universitetet i Bergen pt Universidade de Bergen ru Бергенский университет sh Univerzitet i Bergenu sv Universitetet i Bergen tl Universidad ng Bergen uk Бергенський університет zh 哥尔根大学</p> <p>Statements</p> <p>instance of [*] university 1 reference</p> <p>logo image  Bergen university777.JPG 2,463 x 528; 886 KB 1 reference</p> <p>image  Bergen museum Norway 2009 2.JPG 3,008 x 2,000; 3.46 MB 1 reference</p> <p>inception [*] 1946 1 reference</p> <p>rector [*] Dag Rune Olsen 0 references</p> <p>country [*] Norway 0 references</p> <p>Wikibooks (0 entries) edit</p> <p>Wikinews (0 entries) edit</p> <p>Wikiquote (0 entries) edit</p> <p>Wikisource (0 entries) edit</p>
--	---

 DBpedia  Formats  Faceted Browser  Sparql Endpoint

Property	Value
dbo:abstract	<ul style="list-style-type: none">The University of Bergen (Norwegian: Universitetet i Bergen) is a public university located in Bergen, Norway. Although founded as late as 1946, academic activity had taken place at Bergen Museum as far back as 1825. The university today serves approximately 17,000 students, and is one of eight universities in Norway.
dbo:affiliation	<ul style="list-style-type: none">dbr:European_University_Associationdbr:Worldwide_Universities_Networkdbr:Comisia_Groupdbr:Utrecht_Network
dbo:city	<ul style="list-style-type: none">dbr:Bergen,_Norway
dbo:country	<ul style="list-style-type: none">dbr:Norway
dbo:numberOfStudents	<ul style="list-style-type: none">17000 (xsd:integer)
dbo:rector	<ul style="list-style-type: none">dbr:Dag_Rune_Olsen
dbo:staff	<ul style="list-style-type: none">3200 (xsd:integer)
dbo:type	<ul style="list-style-type: none">dbr:Public_University



Knowledge Graph: Things, not strings!

A graph that describes real-world entities
and their relationships to one another.

Google Announcement in 2012

<http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>

The screenshot shows a Knowledge Graph card for the University of Bergen. At the top left is the university's logo featuring an owl. To the right is a map showing the location of the university in Bergen, Norway. Below the logo and map, the text "University of Bergen" is displayed, followed by three buttons: "Website", "Directions", and "Save". Underneath this, it says "Public university in Bergen, Norway". To the right of the main title, there are two sections: "Events" and "Notable alumni". The "Events" section lists "Tue, Jan 1 PhD-position (4 years)", "Fri, Jan 18 Smart grids + people + energy?", and "Fri, Feb 1 Your Guide to 20th Century Atlanta in the 2...". The "Notable alumni" section shows five thumbnail images of notable graduates: Erna Solberg, Frank Aarebrot, Karl Ove Knausgård, Gunnar Skirbekk, and Lars Gule. Each thumbnail has a name and a "View 45+ more" link below it.

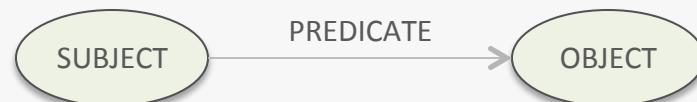
The screenshot shows a Knowledge Graph card for Elon Musk. At the top right is a collage of six images of him in various settings. Below the collage, the text "Elon Musk" is displayed, followed by "CEO of Tesla Motors" and a share icon. The main text block describes him as a technology entrepreneur, investor, and engineer, mentioning his multiple nationalities and various companies he founded or worked for. It includes details about his birth (June 28, 1971), net worth (\$22.1 billion USD), spouse (Talulah Riley and Justine Musk), education (University of Pennsylvania), children (Nevada Alexander Musk, Kai Musk, Xavier Musk, Saxon Musk, Griffin Musk, Damian Musk), and parents (Maye Musk, Errol Musk). Below this, there are "Profiles" for Twitter and Google+, and a "People also search for" section with thumbnails for Talulah Riley, Bill Gates, Justine Musk, Tim Cook, and Jeff Bezos.

The screenshot shows a Google search results page. The search query "How old is the CEO of Tesla?" is entered in the search bar, which includes a microphone and magnifying glass icon. Below the search bar, there are filter buttons for "All", "Images", "News", "Shopping", "Videos", "More", "Settings", and "Tools". The search results page displays the query "Elon Musk / Age" and the answer "47 years" in large text, with the date "June 28, 1971" underneath. To the right of the text is a portrait of Elon Musk. Below this, there is a "People also search for" section with thumbnails for Justine Musk (46 years), Joe Rogan (51 years), and Jeff Bezos (54 years).

Knowledge Graph: Definition

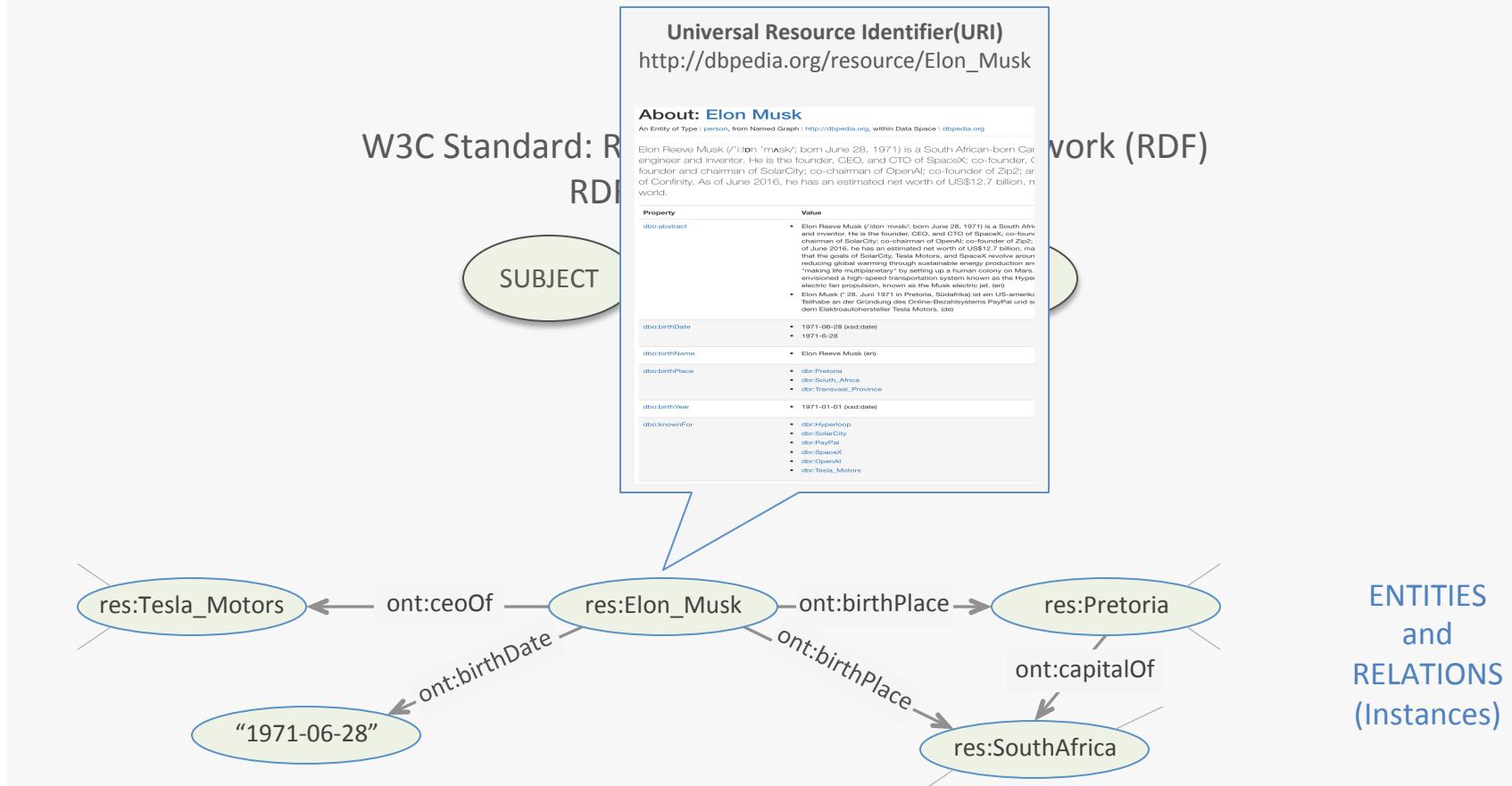
A set of typed **entities** and their **relationships** that are defined according to a **schema** (ontology).

W3C Standard: Resource Description Framework (RDF)
RDF Statements as “Triples”



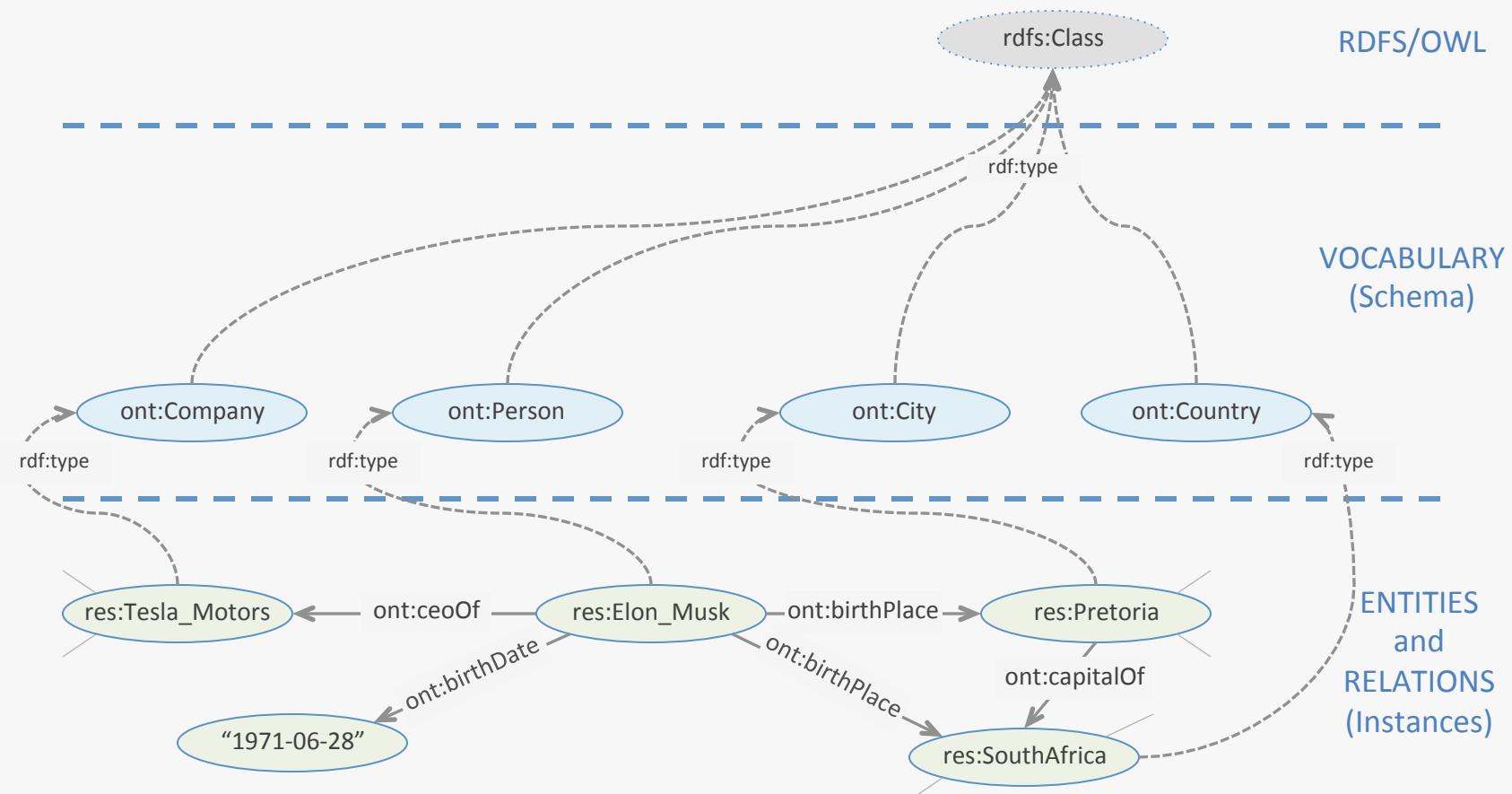
Knowledge Graph: Definition

A set of typed entities and their relationships that are defined according to a schema (ontology).



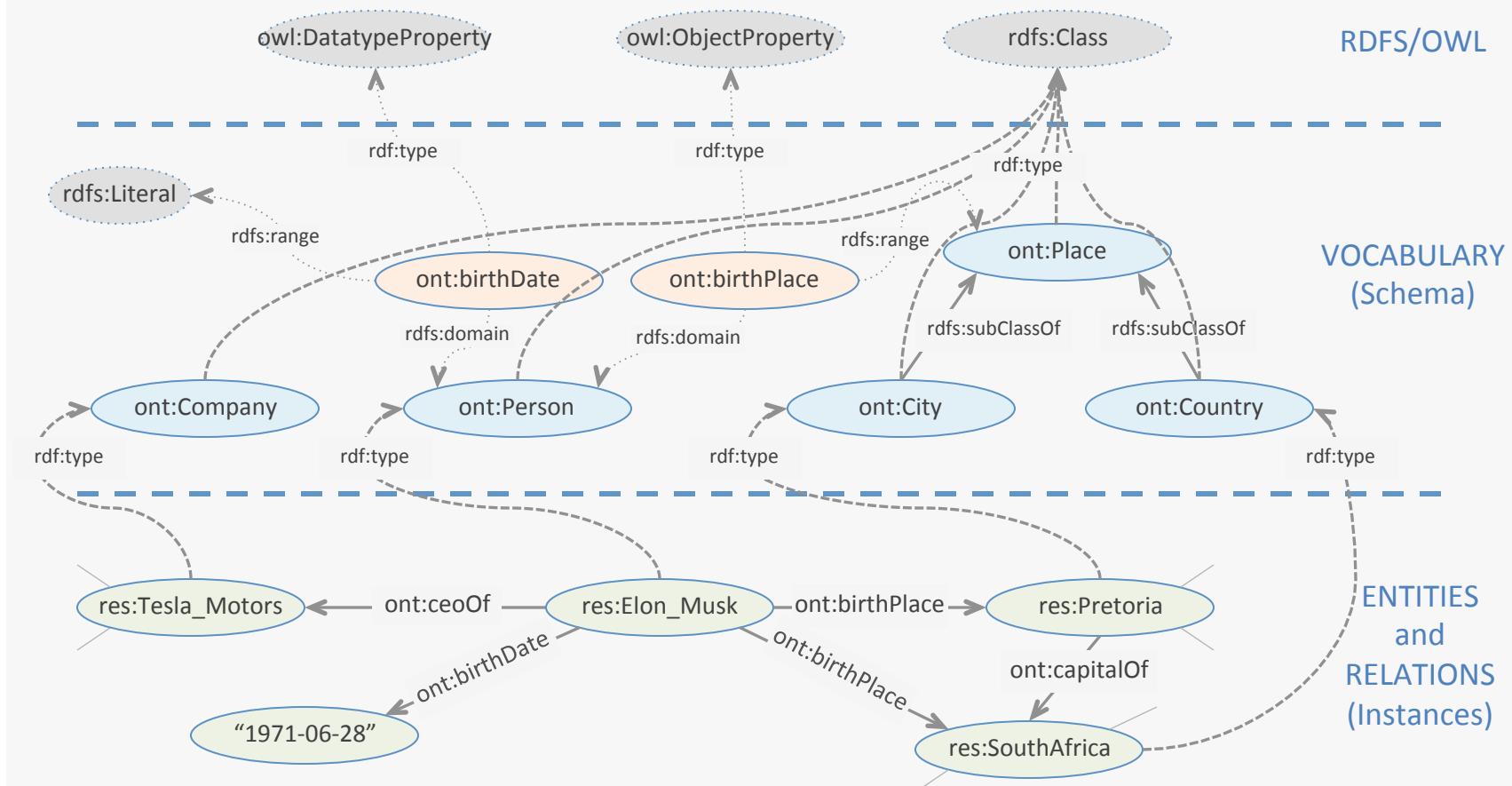
Knowledge Graph: Definition

A set of typed **entities** and their **relationships** that are defined according to a **schema** (ontology).



Knowledge Graph: Definition

A set of typed **entities** and their **relationships** that are defined according to a **schema** (ontology).



Knowledge Graph: Structure

- Vocabulary: RDFS (RDF Schema), OWL (Web Ontology Language)
 - Class hierarchy, equivalence, intersection, disjoint classes...
 - Property hierarchy, equivalence, domain-range classes, symmetry, transitivity...
 - Constraints that define new classes
 - Entity equivalence, difference...
- Formats: XML, N3-Turtle, N-Triples

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:dbo="http://dbpedia.org/ontology/">
    <rdf:Description rdf:about="http://dbpedia.org/resource/Elon_Musk">
        <rdf:type rdf:resource="http://dbpedia.org/ontology/Person" />
        <rdf:type rdf:resource="http://schema.org/Person" />
        <owl:sameAs rdf:resource="http://www.wikidata.org/entity/Q317521" />
        <owl:sameAs rdf:resource="http://yago-knowledge.org/resource/Elon_Musk" />
        <dbo:birthPlace rdf:resource="http://dbpedia.org/resource/Pretoria" />
        <dbo:birthPlace rdf:resource="http://dbpedia.org/resource/South_Africa" />
        <dbo:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1971-06-28</dbo:birthDate>
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/SpaceX" />
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/Tesla_Motors" />
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/OpenAI" />
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/Hyperloop" />
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/PayPal" />
        <dbo:knownFor rdf:resource="http://dbpedia.org/resource/SolarCity" />
    </rdf:Description>
</rdf:RDF>
```

```
N3/Turtle

@prefix dbr: <http://dbpedia.org/resource/> .
@prefix dbp: <http://dbpedia.org/property/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix yago: <http://dbpedia.org/class/yago/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix wikidata: <http://www.wikidata.org/entity/> .
@prefix schema: <http://schema.org/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix yago-res: <http://yago-knowledge.org/resource/> .
dbr:Elon_Musk rdf:type dbo:Person ,
    schema:Person ;
    owl:sameAs yago-res:Elon_Musk ,
    wikidata:Q317521 ;
    dbo:birthPlace dbr:Pretoria,
    dbo:South_Africa ;
    dbo:birthDate "1971-06-28"^^xsd:date ;
    dbo:knownFor dbr:SpaceX ,
        dbr:Tesla_Motors ,
        dbr:OpenAI ,
        dbr:Hyperloop ,
        dbr:PayPal ,
        dbr:SolarCity .
```

Knowledge Graph: Structure

- SPARQL: SPARQL Protocol and RDF Query Language

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)
<http://dbpedia.org>

Query Text

```
prefix dbr: <http://dbpedia.org/resource/>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix dbo: <http://dbpedia.org/ontology/>

SELECT ?p
{
  ?p rdf:type dbo:Person.
  ?p dbo:knownFor dbr:Tesla_Motors
}
```

Try at: <http://dbpedia.org/sparql/>

- Query Results:

p
http://dbpedia.org/resource/JB_Straubel
http://dbpedia.org/resource/Elon_Musk
http://dbpedia.org/resource/Martin_Eberhard

DBpedia Browse using Formats Faceted Browser Sparql Endpoint

About: Tesla Motors

An Entity of Type : [Public company](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Tesla Motors (often shortened to Tesla) is an American automaker and energy storage company co-founded in July 2003 originally by Martin Eberhard and Marc Tarpenning, with JB Straubel, Elon Musk and Ian Wright, and is based in Palo Alto, California. The company specializes in electric cars and their powertrain components and also produces battery charging equipment.

Property	Value
foaf:name	▪ Tesla Motors, Inc. (en)
is dbo:designCompany of	▪ dbr:Tesla_Roadster
is dbo:designer of	▪ dbr:Tesla_Roadster
is dbo:knownFor of	▪ dbr:JB_Straubel ▪ dbr:Elon_Musk ▪ dbr:Martin_Eberhard
is dbo:manufacturer of	▪ dbr:Tesla_Model_S ▪ dbr:Tesla_Roadster ▪ dbr:Toyota_RAV4_EV_Second_generation_1 ▪ dbr:Tesla_Model_3 ▪ dbr:Tesla_Model_X
dbo:industry	▪ dbr:Automotive_industry ▪ dbr:Energy_storage
dbo:keyPerson	▪ dbr:JB_Straubel ▪ dbr:Elon_Musk
dbo:numberOfEmployees	▪ 14 (xsd:integer)
dbo:product	▪ dbr:Electric_car ▪ dbr:Tesla_Powerwall
dbo:thumbnail	▪ wiki-commons:Special:FilePath/Tesla_Motors.svg?width=300
dbo:type	▪ dbr:Public_company
dbo:wikiPageExternalLink	▪ http://tesla.com/ ▪ http://www.mojomotors.com/blog/where-can-tesla-sell-cars/ ▪ https://www.tesla.com/blog/master-plan-part-deux

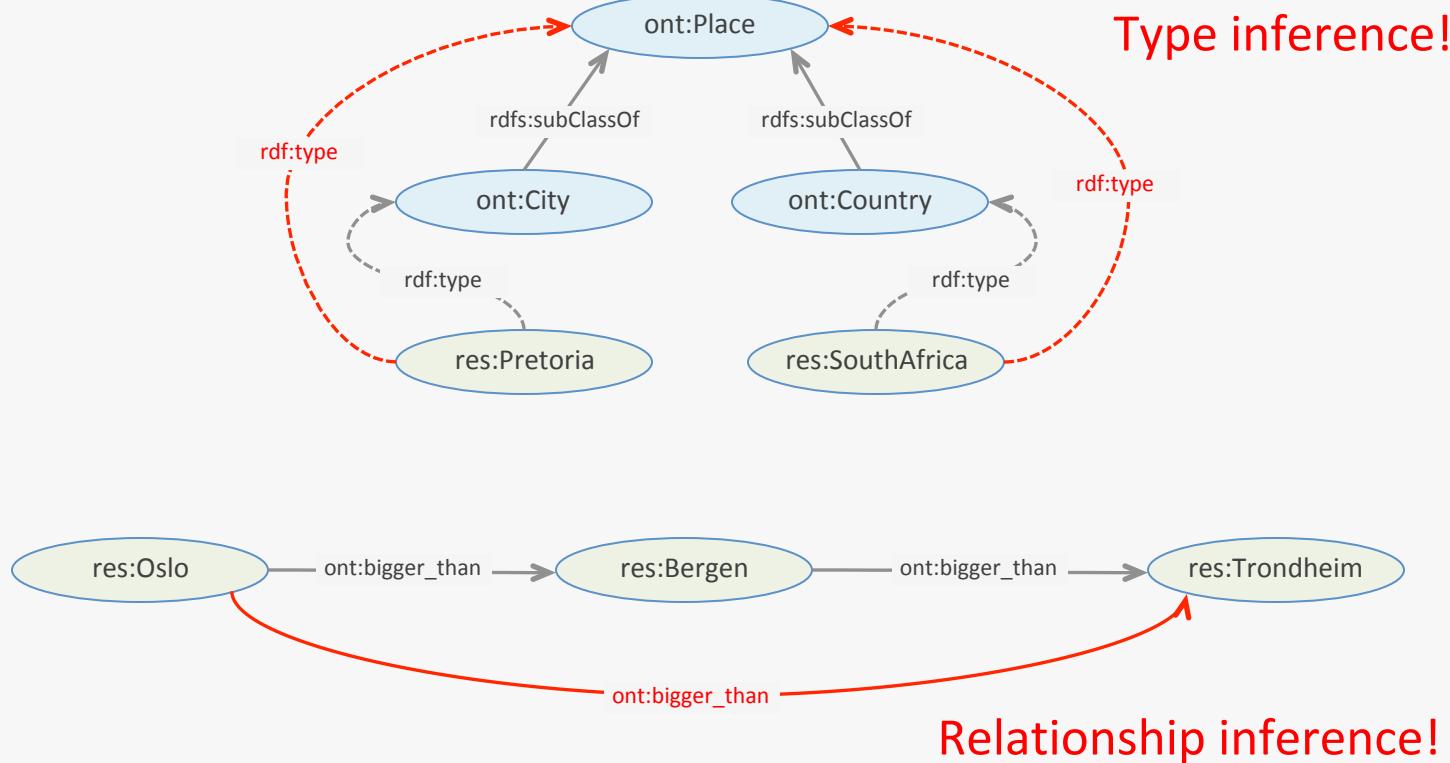
Knowledge Graphs: Definition

“Knowledge Graphs provide semantically structured information that is interpretable by computers.”

*Nickel et al.: A Review of Relational Machine Learning for Knowledge Graphs,
Proceedings of the IEEE, vol. 104, no. 1, pp. 11-33, 2016.*

Knowledge Graph: Reasoning

Reasoning enables making **inferences** for new knowledge that is not directly visible in the graph.



Knowledge Graph: Reasoning

Reasoning enables making **inferences** for new knowledge that is not directly visible in the graph.

Reasoning services	Explanation
Ontology consistency checking	Checking if an ontology contains contradiction
Classification	Computing the inferrable OWL : subClassOf relations between classes
Realization	Computing the inferrable RDF : type relation between an individual and a class
Class satisfiability checking	Checking if a class can have any instance
Axiom entailment checking	Checking if an axiom can be deduced from an ontology
Conjunctive query answering	Answering a query against an ontology

Pan J.Z. et. al, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*.
Springer Publishing Company, 2017

- SPARQL queries go beyond simple graph matching to increase the effectiveness of question answering, semantic search and automatic summarization.
- Require Reasoning Engines (e.g., FaCT++, Pellet, Hermit)

Popular Knowledge Graphs



Name	Instances	Facts	Types	Relations
DBpedia (English)	4,806,150	176,043,129	735	2,813
YAGO	4,595,906	25,946,870	488,469	77
Freebase	49,947,845	3,041,722,635	26,507	37,781
Wikidata	15,602,060	65,993,797	23,157	1,673
NELL	2,006,896	432,845	285	425
OpenCyc	118,499	2,413,894	45,153	18,526
Google's Knowledge Graph	570,000,000	18,000,000,000	1,500	35,000
Google's Knowledge Vault	45,000,000	271,000,000	1,100	4,469
Yahoo! Knowledge Graph	3,443,743	1,391,054,990	250	800

Paulheim H., Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods, Semantic Web Journal, vol. 8, no. 3, pp. 489-508, 2017

- They differ in: public-private, construction, coverage, correctness, freshness...
- They cover various topical domains!
- There can be interlinks between different knowledge graphs.

Recap



Knowledge Graph...

About: University of Bergen	
An Entity of Type : Public_University , from Named Graph : http://dbpedia.org , within Data Space : dbpedia.org	
Property	Value
dbo:abstract	<ul style="list-style-type: none">The University of Bergen (Norwegian: Universitetet i Bergen) is a public university located in Bergen, Norway. Although founded as late as 1946, academic activity had taken place at Bergen Museum as far back as 1825. The university today serves approximately 17,000 students, and is one of eight universities in Norway. (en)Die Universität Bergen (norwegian: Universitetet i Bergen, latinisch: Universitas Bergensis) ist mit etwa 14.500 Studierenden und 3.200 Angestellten im Sommersemester (2008) nach den Universitäten in Oslo und Trondheim die drittgrößte Universität Norwegens. Sie wurde 1946 gegründet. Zum Campus der Universität gehört das Bergen Museum. (de)
dbo:affiliation	<ul style="list-style-type: none">dbr:European_University_Associationdbr:Worldwide_Universities_Networkdbr:Coimbra_Groupdbr:Utrecht_Network
dbo:city	<ul style="list-style-type: none">dbr:Bergen,_Norway
dbo:country	<ul style="list-style-type: none">dbr:Norway
dbo:numberOfStudents	<ul style="list-style-type: none">17000 (xsd:integer)
dbo:rector	<ul style="list-style-type: none">dbr:Dag_Rune_Olsen
dbo:staff	<ul style="list-style-type: none">3200 (xsd:integer)
dbo:type	<ul style="list-style-type: none">dbr:Public_University
dbo:wikiPageExternalLink	<ul style="list-style-type: none">http://www.uib.no/enhttp://www.hit.uib.no/http://www.uib.no/en/http://www.uib.no/fakinsthttp://www.uib.no/fremmedsprakhttp://www.uib.no/research/research-fields/research-groupshttp://www.hit.uib.no/wab/
dbo:wikiPageID	<ul style="list-style-type: none">31799 (xsd:integer)
dbo:wikiPageRevisionID	<ul style="list-style-type: none">703499175 (xsd:integer)
dbp:established	<ul style="list-style-type: none">1946 (xsd:integer)
dbp:wordnet_type	<ul style="list-style-type: none">http://www.w3.org/2006/03/wn/wn20/instances/synset-university-noun-2

- defines a schema,
- describes entities and their relationships,
- covers various topical domains,
- supports reasoning.
- Examples: DBpedia, YAGO, Wikidata

Part 2: Machine Learning and Knowledge Graphs

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

Search and Recommendation:

- E.g., content-based recommendation of music-video according to metadata, semantic search

Other related research areas:

- Question answering, knowledge graph construction, link prediction, summarization

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

Search and Recommendation:

- E.g., content-based recommendation of music-video according to metadata, semantic search.

Other related research areas:

- Question answering, knowledge graph construction, link prediction, summarization

Mostly based on measuring the similarity of texts.

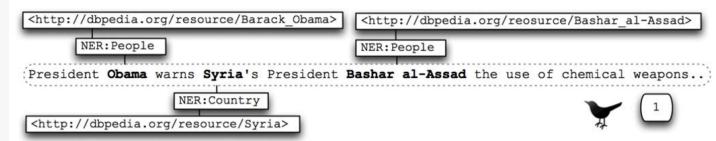
Syntactic vs. Semantic similarity

ID	Document content
Doc 1	Puma, an American Feline Resembling a Lion.
Doc 2	Puma, a Famous Sports Brand from German.
Doc 3	Welcome to Zoo, an Animal World.

Doc1 and Doc3 are about similar topics

Language differences (e.g., *Munich*=*München*), abbreviations (e.g., *UiB*=*University of Bergen*), synonyms...

Bag-of-words features + Entities in Knowledge Graph.



* Berven et. al, *News Hunter: Building and Mining Knowledge Graphs for Newsroom Systems*, NOKOBIT, 2018

* Ozdikis et. al, *Evidential estimation of event locations in microblogs using the Dempster–Shafer theory*, IPM, vol 52-6, pp. 1227-1246, 2016

* Ozdikis et. al, *Incremental clustering with vector expansion for online event detection in microblogs*, *Social Network Analysis and Mining*, vol. 7 (1), pp. 56-73, 2017

* Gabrilovich E., and Markovitch S., *Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge*, AAAI, pp. 1301-1306, 2006

* Wu et. al, *An efficient Wikipedia semantic matching approach to text document classification*, *Journal of Inf. Sci.*, vol. 393-C, pp.15-28, 2017

* Varga et. al, *Linked knowledge sources for topic classification of microposts: A semantic graph-based approach*, *Journal of Web Semantics*, Vol. 26, pp. 36-57, 2014

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

Search and Recommendation:

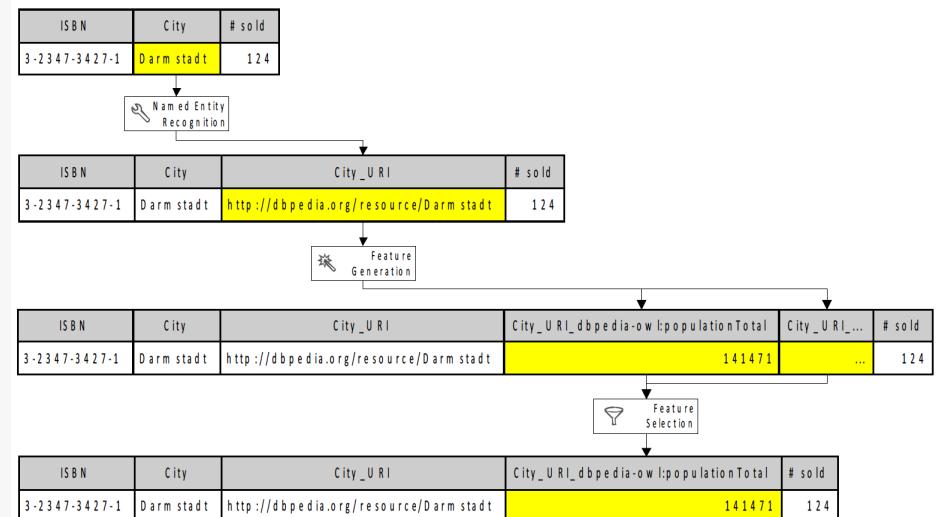
- E.g., content-based recommendation of music-video according to metadata, semantic search.

Other related research areas:

- Question answering, knowledge graph construction, link prediction, summarization

Making predictions based on a set of related features

Feature generation from background knowledge in Knowledge Graph.



Paulheim H., and Fümkranz J., Unsupervised generation of data mining features from linked open data, WIMS, pp. 1-12, 2012

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

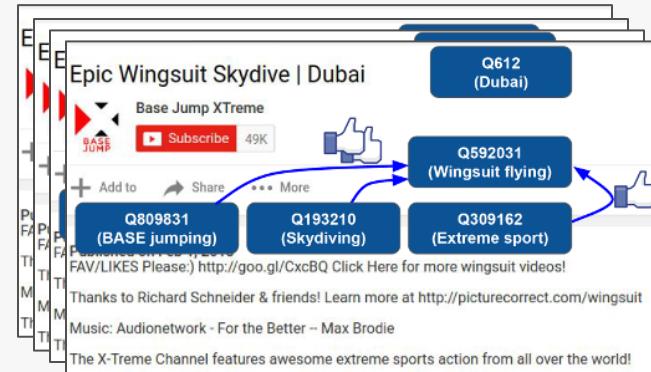
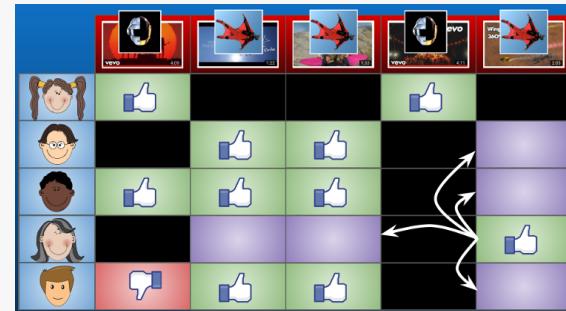
Search and Recommendation:

- E.g., content-based recommendation of music-video according to metadata, semantic search.

Other related research areas:

- Question answering, knowledge graph construction, link prediction, summarization

Content-based recommendation based on topic.
Identify Knowledge Graph entities in description.



* Covington et. al, Deep Neural Networks for YouTube Recommendations, RecSys, pp. 191-198, 2016

** <https://conferences.oreilly.com/strata/strata-eu-2017/public/schedule/detail/57609>

*** <https://www.youtube.com/watch?v=D-bTGefjjoA>

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

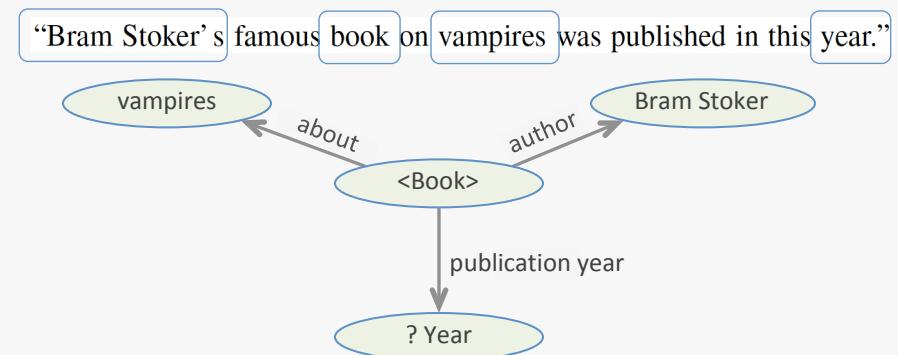
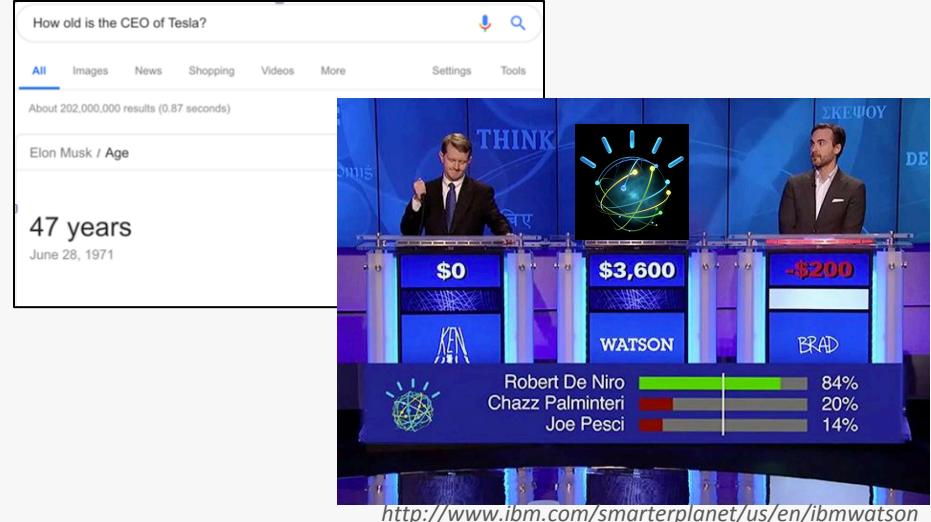
Search and Recommendation:

- E.g., content-based recommendation of music-video according to metadata, semantic search.

Other related research areas:

- Question answering**, knowledge graph construction, link prediction, summarization

Question analysis and hypothesis generation



* Pan J.Z. et. al, Exploiting Linked Data and Knowledge Graphs in Large Organisations, Springer Publishing, 2017

Common Machine Learning Problems

Classification and Clustering:

- E.g., determining the category for a news report, categorization of spam tweets, event detection and summarization, location estimation, News Hunter

Regression:

- E.g., predicting sales of a product in a new store based on observed features

Search and Recommendation:

- E.g., content-based recommendation of music-video according to metadata, semantic search.

Other related research areas:

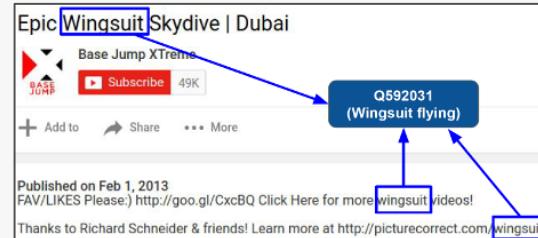
- Question answering, knowledge graph construction, link prediction, summarization

Common challenge?



ISBN	City	# sold
3-2347-3427-1	Darmstadt	124

ISBN	City	City_URI	# sold
3-2347-3427-1	Darmstadt	http://dbpedia.org/resource/Darmstadt	124



“Bram Stoker’s famous book on vampires was published in this year.”

Detect Knowledge Graph entities in a given texts.

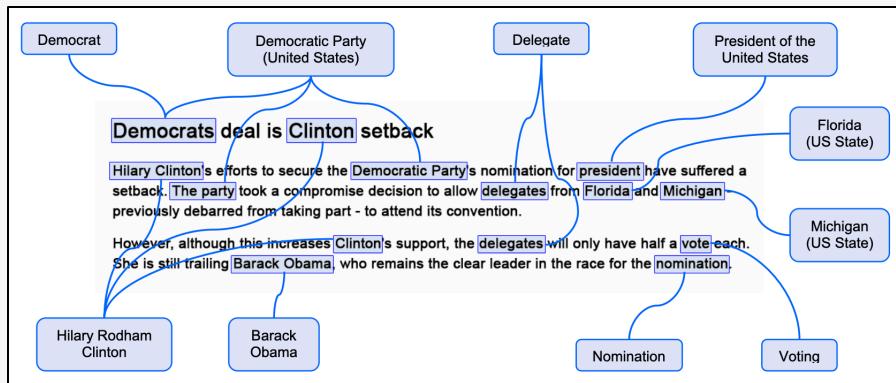
Entity Linking!

Part 3: Entity Linking using Knowledge Graphs

Entity Linking using Knowledge Graphs

- Given a textual passage, identify substrings corresponding to entities and link them to corresponding instances in a given Knowledge Graph.

Ling et. al, Design Challenges for Entity Linking, TACL, vol. 3, pp. 315–328, 2015



The screenshot shows the DBpedia Spotlight interface. At the top, there's a logo with the text "DBpedia Spotlight". Below it is a search bar containing the text "Berlin". Underneath the search bar are several input fields: "Confidence:" with a slider set at 0.5, "Language:" set to English, and a checkbox for "n-best candidates". There are also "SELECT TYPES..." and "ANNOTATE" buttons. The main area displays a paragraph about Berlin's history and division. At the bottom right, there's a "BACK TO TEXT" button and a note stating "This demo uses the statistical DBpedia Spotlight web service at <https://api.dbpedia-spotlight.org/en>".

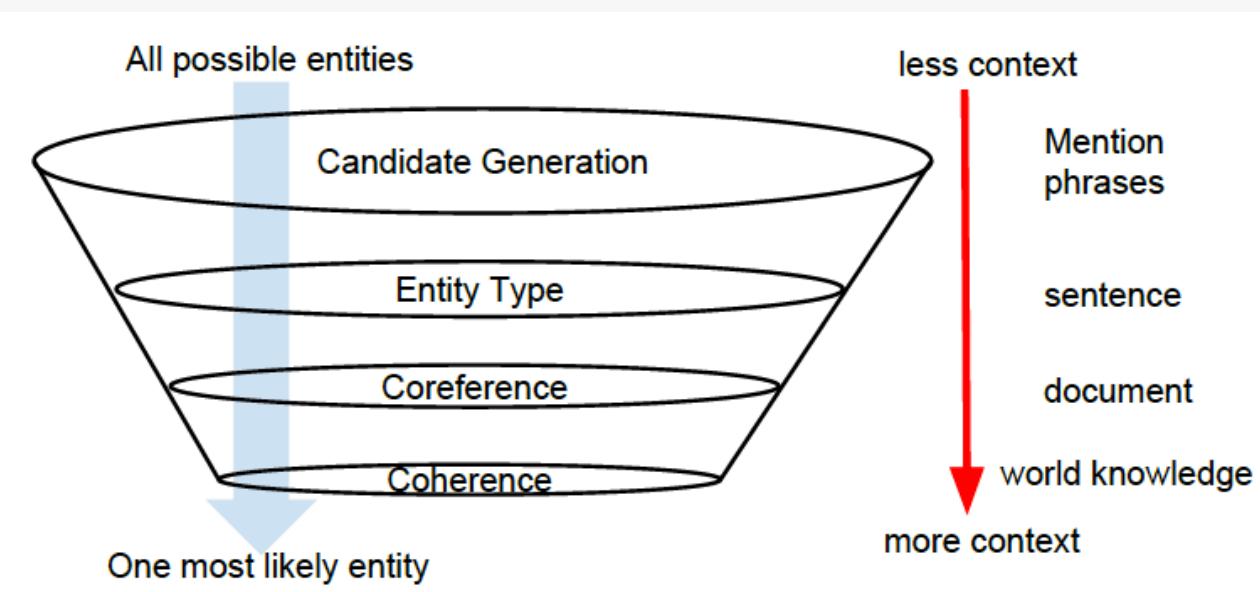
Provides:

- Semantic annotations to human readers,
- Machine-consumable representation of text.

Entity Linking using Knowledge Graphs

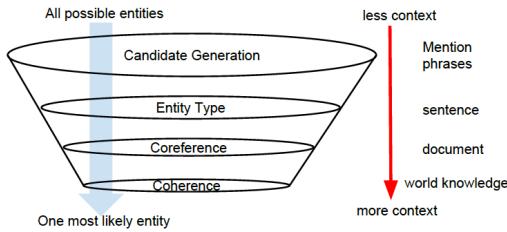
Common approach:

- 1) Find candidate phrases (recognition)
- 2) Resolve ambiguities (resolution)



*VINCULUM: Ling et. al, *Design Challenges for Entity Linking*, TACL, vol. 3, pp. 315–328, 2015

Entity Linking using Knowledge Graphs



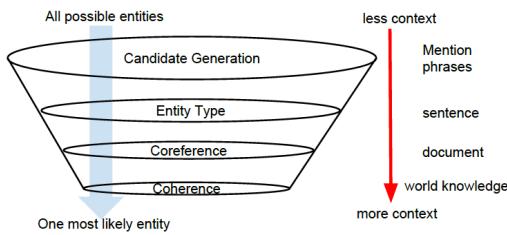
Example:

Find what “Washington” refers to in the following sentence:

Wizards’ two-way player Jordan McRae will remain at Washington during the team’s back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

* DBpedia-Spotlight: Daiber et. al, Improving Efficiency and Accuracy in Multilingual Entity Extraction, ISEM, pp. 121-124, 2013

Entity Linking using Knowledge Graphs



Step1: Candidate Generation

Finding candidate entities in Knowledge Graph

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

* DBpedia-Spotlight: Daiber et. al, Improving Efficiency and Accuracy in Multilingual Entity Extraction, ISEM, pp. 121-124, 2013

Candidate Entities

Washington DC

George Washington

Washington State

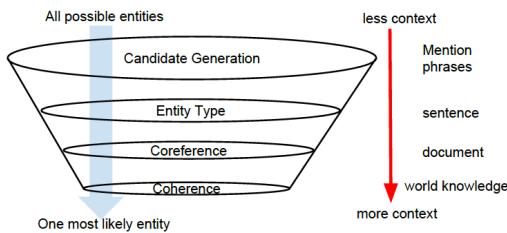
Washington Wizards

Denzel Washington

University of Washington

Washington High School

Entity Linking using Knowledge Graphs



Step2: Entity Types

Compare entity types (e.g., PER, LOC, ORG) in Knowledge Graph

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

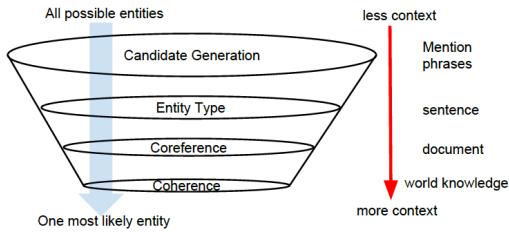
LOC or ORG

Candidate Entities

Washington DC
George Washington
Washington State
Washington Wizards
Denzel Washington
University of Washington
Washington High School

?

Entity Linking using Knowledge Graphs



Step2: Entity Types

Compare entity types (e.g., PER, LOC, ORG) in Knowledge Graph

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

LOC or ORG

Candidate Entities

Washington DC (LOC)

George Washington (PER)

Washington State (LOC)

Washington Wizards (ORG)

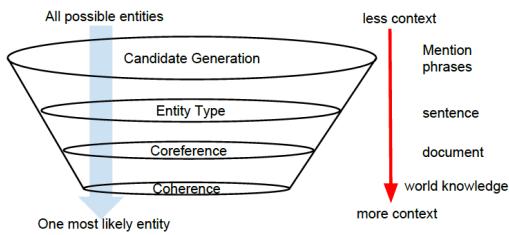
Denzel Washington (PER)

University of Washington (ORG)

Washington High School (ORG)

?

Entity Linking using Knowledge Graphs



Step2: Entity Types

Compare entity types (e.g., PER, LOC, ORG) in Knowledge Graph

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

LOC or ORG

Candidate Entities

Washington DC (LOC)

~~George Washington (PER)~~

Washington State (LOC)

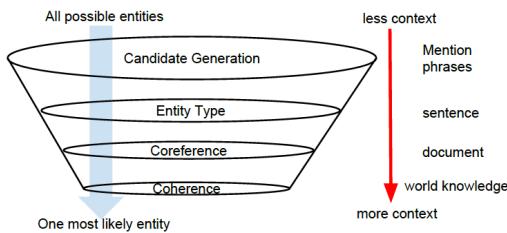
Washington Wizards (ORG)

~~Denzel Washington (PER)~~

University of Washington (ORG)

Washington High School (ORG)

Entity Linking using Knowledge Graphs



Step3: Co-reference resolution

Finding expressions that refer to the same entity.

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

Co-references in KG

Candidate Entities

Washington DC

~~*George Washington*~~

Washington State

Washington Wizards

~~*Denzel Washington*~~

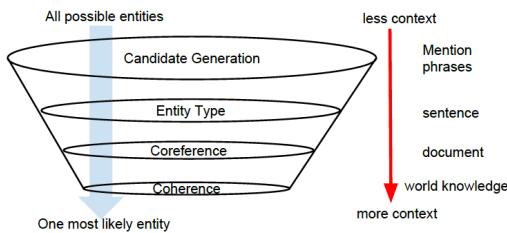
University of Washington

Washington High School

?

* Lee et. al, Deterministic Coreference Resolution
Based on Entity-Centric, Precision-Ranked Rules,
Computational Linguistics, vol. 39-4, pp.885-916,
2013

Entity Linking using Knowledge Graphs



Step3: Co-reference resolution

Finding expressions that refer to the same entity.

Wizards' two-way player Jordan McRae will remain at Washington during the team's back-to-back against Atlanta Hawks and Houston Rockets, reported at UW News.

Co-references in KG

Candidate Entities

~~Washington DC~~

~~George Washington~~

~~Washington State~~

Washington Wizards

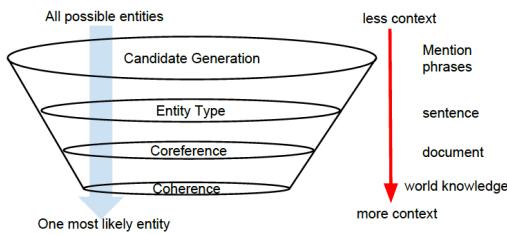
~~Denzel Washington~~

University of Washington

~~Washington High School~~

* Lee et. al, Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules, Computational Linguistics, vol. 39-4, pp.885-916, 2013

Entity Linking using Knowledge Graphs



Step4: Coherence

Resolve based on semantic relatedness with other entities

*Wizards' two-way player **Jordan McRae** will remain at **Washington** during the team's back-to-back against **Atlanta Hawks** and **Houston Rockets**, reported at UW News.*

Check coherence with other entities using KG

Candidate Entities

~~Washington DC~~

~~George Washington~~

~~Washington State~~

Washington Wizards

~~Denzel Washington~~

University of Washington

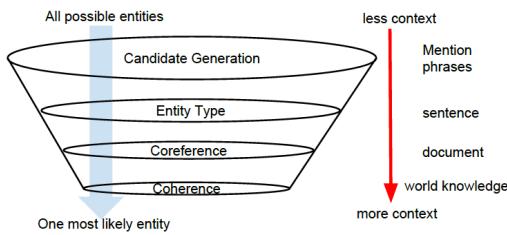
~~Washington High School~~

?

* Milne D. and Witten I.H., Learning to Link with Wikipedia, CIKM, pp.509-518, 2008

** Cilibrasi R.L. and Vitanyi P.M.B., The Google Similarity Distance, TKDE, vol. 19, no 3, pp.370–383, 2007

Entity Linking using Knowledge Graphs



Step4: Coherence

Resolve based on semantic relatedness with other entities

*Wizards' two-way player **Jordan McRae** will remain at **Washington** during the team's back-to-back against **Atlanta Hawks** and **Houston Rockets**, reported at UW News.*

Check coherence with other entities using KG

Candidate Entities

~~Washington DC~~

~~George Washington~~

~~Washington State~~

Washington Wizards

~~Denzel Washington~~

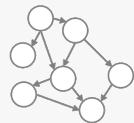
~~University of Washington~~

~~Washington High School~~

* Milne D. and Witten I.H., *Learning to Link with Wikipedia*, CIKM, pp.509-518, 2008

** Cilibrasi R.L. and Vitanyi P.M.B., *The Google Similarity Distance*, TKDE, vol. 19, no 3, pp.370–383, 2007

Recap



Knowledge Graphs in Machine Learning

- Wide range of application areas: Classification, clustering, regression, recommendation, semantic search, question answering...
- Entity linking for text-based analysis and mining.
- Entity linking: recognition and resolution steps.



<https://www.youtube.com/watch?v=C5Xnxjq63Zg>

References

- Berven et. al, *News Hunter: Building and Mining Knowledge Graphs for Newsroom Systems*, NOKOBIT, 2018
- Cilibrasi R.L. and Vitanyi P.M.B., *The Google Similarity Distance*, TKDE, vol. 19, no 3, pp.370–383, 2007
- Covington et. al, *Deep Neural Networks for YouTube Recommendations*, RecSys, pp. 191-198, 2016
- Daiber et. al, *Improving Efficiency and Accuracy in Multilingual Entity Extraction*, ISEM, pp. 121-124, 2013
- Derczynski et. al, *Analysis of named entity recognition and linking for tweets*, IPM, vol. 51, pp.32-49, 2015
- Gabrilovich E., and Markovitch S., *Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge*, AAAI, pp. 1301-1306, 2006
- Lee et. al, *Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules*, Computational Linguistics, vol. 39-4, pp.885-916, 2013
- Ling et. al, *Design Challenges for Entity Linking*, TACL, vol. 3, pp. 315–328, 2015
- Milne D. and Witten I.H., *Learning to Link with Wikipedia*, CIKM, pp.509-518, 2008
- Nickel et. al, *A Review of Relational Machine Learning for Knowledge Graphs*, Proc. of the IEEE, vol. 104, no. 1, pp. 11-33, 2016
- Ozdikis et. al, *Evidential estimation of event locations in microblogs using the Dempster-Shafer theory*, IPM, vol 52-6, pp. 1227-1246, 2016
- Ozdikis et. al, *Incremental clustering with vector expansion for online event detection in microblogs*, Social Network Analysis and Mining, vol. 7 (1), pp. 56-73, 2017
- Pan et. al, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer Publishing Company, 2017
- Paulheim H., *Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods*, Semantic Web Journal, vol. 8, no. 3, pp. 489-508, 2017
- Varga et. al, *Linked knowledge sources for topic classification of microposts: A semantic graph-based approach*, Journal of Web Semantics, Vol. 26, pp. 36-57, 2014
- Wu et. al, *An efficient Wikipedia semantic matching approach to text document classification*, Journal of Inf. Sci., vol. 393-C, pp. 15-28, 2017
- AAAI'18 tutorial <http://usc-isi-i2.github.io/AAAI18Tutorial/>
- KDD'18 tutorial <https://kdd2018tutorial39.azurewebsites.net/KDD%20Tutorial%20T39.pdf>
- WSDM'18 tutorial <https://kgtutorial.github.io/>
- <http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>
- <https://conferences.oreilly.com/strata/strata-eu-2017/public/schedule/detail/57609>

Thank you!

