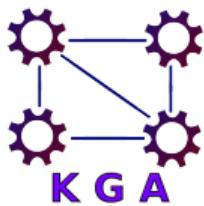




Knowledge Graph Analysis



2018-10-01

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann



Introduction of Lecturers

- ▷ Since 2015:
Full Professor for Data Engineering at University of Bonn
- ▷ 2013-2015:
Leader of AKSW Research Group at University of Leipzig
- ▷ Since 2010:
Leader of Machine Learning and Ontology Engineering Group in Leipzig
- ▷ 2010:
PhD University of Leipzig
- ▷ (Co-)Founder of DBpedia, DL-Learner and LinkedGeoData Projects
- ▷ Research interests: Machine Learning, Semantic Web, Logic



- ▷ 2016 - Now:
Leader of the Knowledge Graphs and Artificial Intelligence Team and Senior Researcher at University of Bonn
- ▷ 2016:
Prize of the University of Siegen for the Promotion of Young Scientists
- ▷ 2015:
PhD of Computer Science at University of Siegen
(*Grade: Summa Cum Laude mit Auszeichnung*)
- ▷ Research interests:
Machine Learning, Knowledge Graphs, Artificial Intelligence





Organisational Matters

- ▷ Knowledge Graph Analysis consists of two modules:
Lecture/Exercises + Seminar
- ▷ Mailing list: sign-up sheet
- ▷ Lecture notes will be provided at
<https://sewiki.iai.uni-bonn.de/teaching/lectures/kga/2018/start>
- ▷ Changes will be announced at the website and via mailing list



Organisational Matters: Exercises and Seminar

Exercises

- ▷ Start date: **Will be announced soon!**
- ▷ Time: **Wednesdays** 14:00-16:00, **Thursdays** 12:00-14:00 & 14:00-16:00 & 16:00-18:00,
Fridays 10:00-12:00 & 12:00-14:00
- ▷ Location: Seminar Room 1.047, Informatik III
- ▷ Address: Endenicher Allee 19a, 53115 Bonn

Seminar

- ▷ Optional in addition to lecture and exercises
- ▷ Deepens understanding of topics discussed in the lecture
- ▷ Based on recently published articles
- ▷ 5 or 6 slots: initial distribution of topics, sprint presentations, final presentations
- ▷ Start date: 9 October 2018
- ▷ Time: **Tuesdays** 10:00-12:00
- ▷ Location: Seminar Room 1.047, Informatik III
- ▷ Address: Endenicher Allee 19a, 53115 Bonn



- ▷ Written exam (90 minutes)
- ▷ Date: Not Fixed Yet!



Organisational Matters: Overview

Tuesday 12:00 - 14:00

- ▷ October 09: Motivation (Prof. Jens Lehmann, Dr. Hamed Shariat Yazdi)
- ▷ October 16: RDF Databases
- ▷ October 23: Property Graph Databases
- ▷ October 30: Statistical Relational Learning (SRL) for KGs
- ▷ November 06: Tensor-Factorization Methods
- ▷ November 13: - No Lecture -
- ▷ November 20: Alternating Least Squares and Stochastic Gradient Descent
- ▷ November 27: Introducing Neural Networks
- ▷ December 04: Neural Networks for KG analysis
- ▷ December 11: Latent Distance and Graph Feature Models
- ▷ January 15: Training SRL Models on KG
- ▷ January 22: Markov Logic Networks
- ▷ January 29: Summary



- ▷ A Review of Relational Machine Learning for Knowledge Graphs by Maximilian Nickel, Kevin Murphy, Volker Tresp, Evgeniy Gabrilovich
<https://arxiv.org/pdf/1503.00759.pdf>
- ▷ O'Reilly's *Graph Databases* by Ian Robinson, Jim Webber and Emil Eifrem
<https://neo4j.com/book-graph-databases/>



Background



Douglas B. Lenat

- ▷ “Knowledge is Power” Hypothesis (the Knowledge Principle): “If a program is to perform a complex task well, **it must know a great deal about the world** in which it operates.”



Edward A. Feigenbaum

- ▷ The Breadth Hypothesis: “To behave intelligently in unexpected situations, an agent must be capable of falling back on **increasingly general knowledge**.¹

¹Taken from: Lenat & Feigenbaum, Artificial Intelligence 47 (1991), “On the Threshold of Knowledge”.



Information vs. Knowledge

- ▷ Information is a precursor for knowledge: necessary, but not sufficient
 - ▷ Knowledge requires cognitive and analytical processing; information does not.
 - ▷ Knowledge = information processed and organized so as to be useful, e.g., a map of immediate environment
- ⇒ **information alone is not enough!**



Why Graphs?

- ▷ Data is increasingly connected (graph data)
- ▷ Relationships within data are often themselves an integral part of the data
- ▷ The world is structured: we are surrounded by **entities** connected by **relations**
- ▷ Graphs are a natural abstraction that captures the relationships between entities
- ▷ Graphs are well-understood mathematical objects



Why Graphs?

A picture is worth a thousand words...



- ▷ 5th Ave intersects E 42nd St
- ▷ 5th Ave intersects E 41st St
- ▷ Madison Ave intersects E 39th St
- ▷ Lexington Ave intersects E 38th St
- ▷ Tunnel Exit St intersects E 38th St
- ▷ Park Ave Tunnel intersects E 39th St
- ▷ etc...

Goal: Get from point A to point B.
Which description would you use?

A picture is worth a thousand words...



- ▷ It is natural for human beings to represent information in graphs
- ▷ Information represented in a graph is also easier to understand and process (for humans)
- ▷ Example: we can think of a map as a graph structure:
 - Nodes = intersections
 - Edges = roads connecting the intersections

Goal: Get from point A to point B.
Which description would you use?

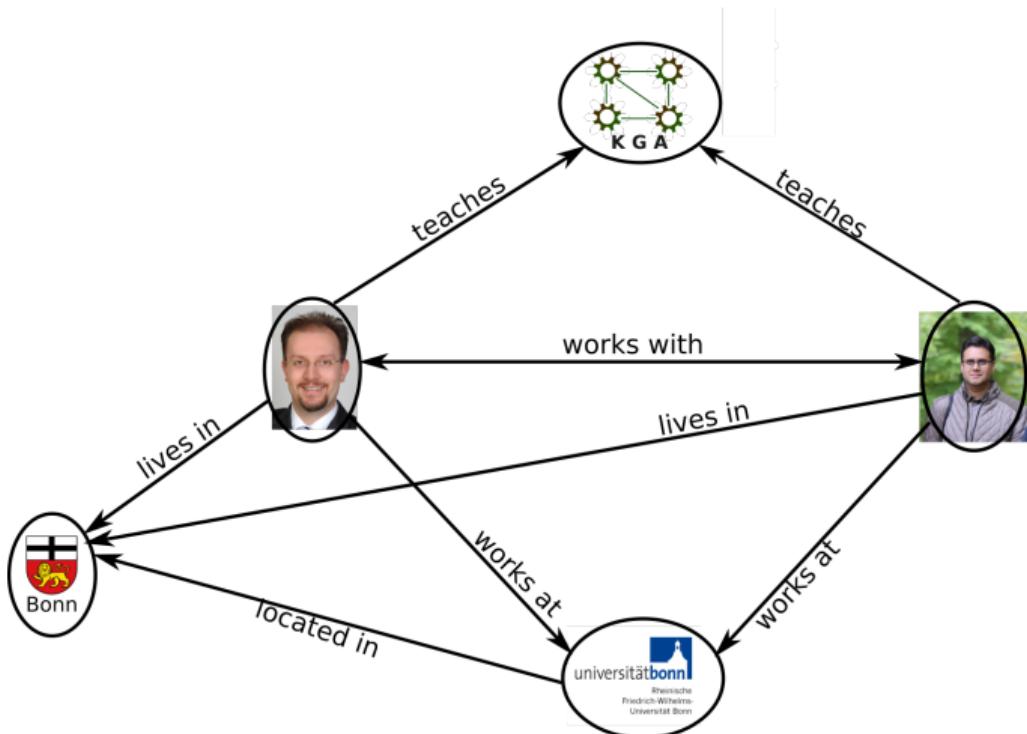


Knowledge Graph: Intuition

- ▷ Models *entities* and *relationships* between these entities
- ▷ Data stored in a graph
- ▷ Nodes correspond to *entities*
- ▷ Directed labeled edges between entities keep track of *relationships*

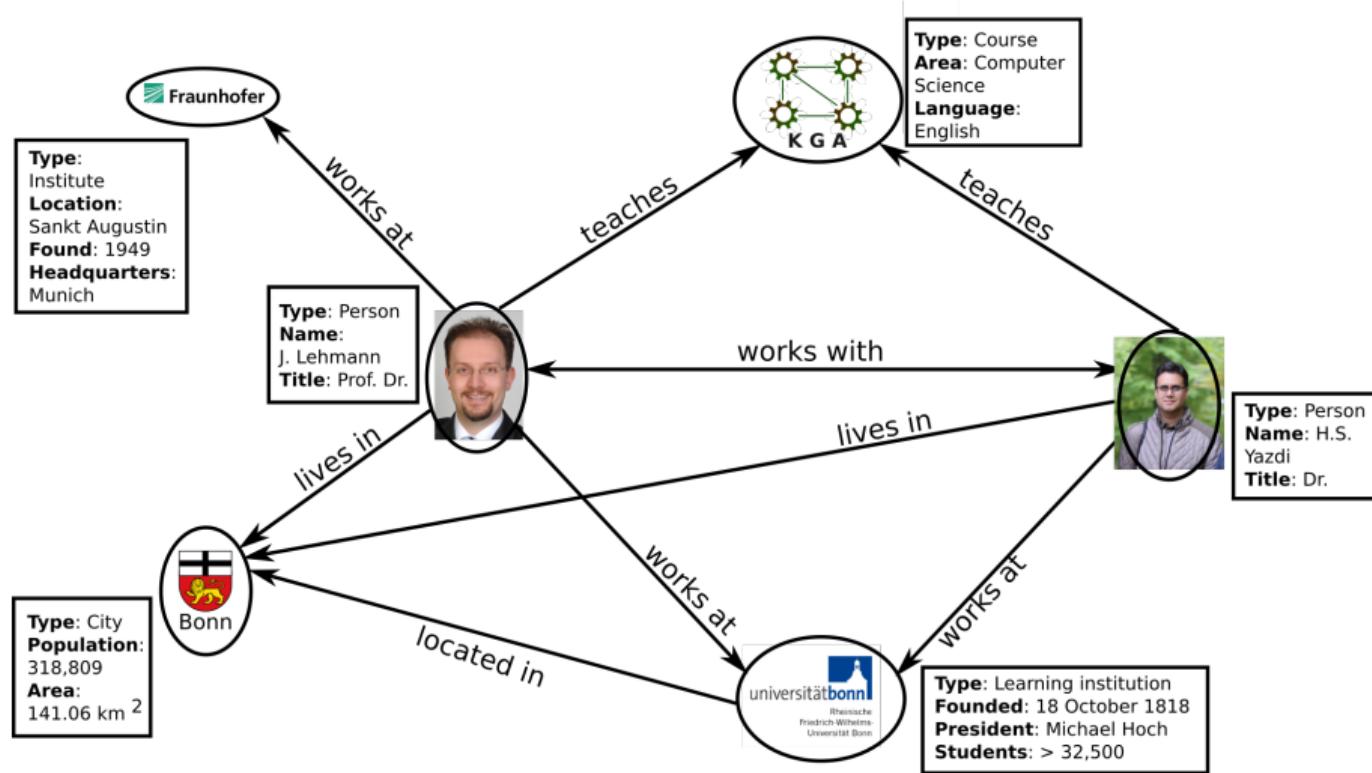


Knowledge Graph: Intuition





Knowledge Graph: Things, not Strings



Draws on a multitude of fields:

- ▷ Information retrieval
- ▷ Natural language processing
- ▷ Databases
- ▷ Machine learning
- ▷ Artificial intelligence





2000s have seen a rapid rise of knowledge graphs to prominence



2007: DBpedia is released by collaborated effort of Leipzig University, University of Mannheim, and OpenLink Software

- ▷ Structured information is mined from Wikipedia
- ▷ Allows semantic queries
- ▷ Collected information, as well as code, are freely available



2007: Freebase



2007: A California-based startup Metaweb releases Freebase

- ▷ Described by Metaweb as “an open shared database of the world’s knowledge”
- ▷ Provided interface to enable users to fill in structured data, to categorize or to connect data
- ▷ No programming experience required
- ▷ Structured data automatically harvested from the web



2008: Max-Planck-Institute in Saarbrücken releases YAGO

- ▷ Stands for: Yet Another Great Ontology
- ▷ Extracts information automatically from Wikipedia, WordNet, and GeoNames
- ▷ Was used by the artificial intelligence system Watson (together with DBpedia and Freebase)



2010: Carnegie Mellon proposes a design for a language learning system NELL

- ▷ Stands for Never-Ending Language Learner
- ▷ Builds a knowledge base by reading the web
- ▷ Continuously improves its ability to process information and to build the knowledge base



2012: Google



2012: Google announces addition of knowledge graph to their search engine

- ▷ Goal: make search easier for the user
- ▷ Step toward enabling search engines to understand the world similarly to how humans understand it
- ▷ Facilitates a broader and deeper search



2012: Microsoft introduces Satori into their Bing search engine

- ▷ Goal: improve search by making it the underlying architecture more intelligent
- ▷ Provide most relevant information and key facts automatically
- ▷ In 2010 Microsoft releases Graph Engine 1.0: a general purpose distributed graph system over a memory cloud. The goal of the engine is to enable processing of large graphs.



2013: Facebook introduces “Entity Graph”

- ▷ Each “entity” (e.g. persons, institutions, restaurants, etc.) is treated as an node in a graph
- ▷ Relations between entities are modeled by their relationships in real-life (e.g. friendships, alumni, checked_in_at, etc)

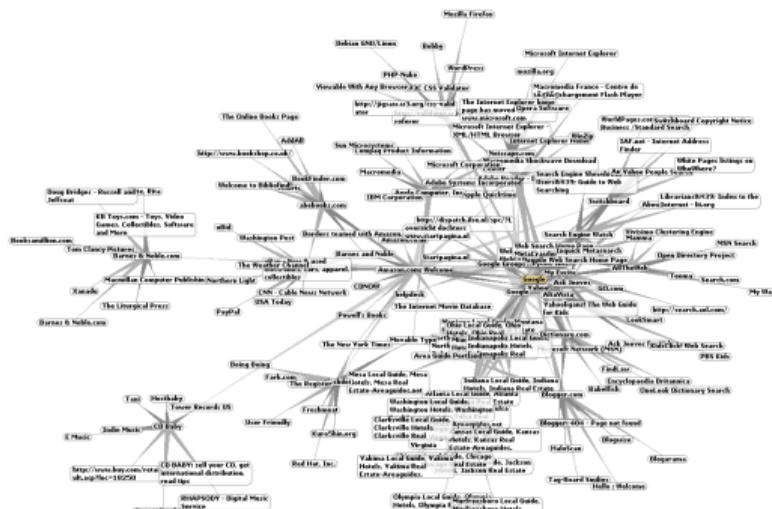
2016: LinkedIn introduces their own knowledge graph

- ▷ A large knowledge base based on LinkedIn members, jobs, titles, companies, skills, etc.
- ▷ Goal: Create a large economic graph - a map of the world economy



In Summary, Knowledge Graphs...

- ▷ are applied both in academia and industry
 - ▷ are rapidly adopted by the most popular search engines and social networks
 - ▷ present a natural way of accessing/storing knowledge in a multitude of domains
 - ▷ promote better interface and access to knowledge: dialog system, personalization, emotion, question answering



²This Wikipedia and Wikimedia Commons image is from the user Chris 73 and is freely available at [//commons.wikimedia.org/wiki/File:WorldWideWebAroundGoogle.png](https://commons.wikimedia.org/wiki/File:WorldWideWebAroundGoogle.png) under the creative commons cc-by-sa 3.0 license.





Things, not Strings!

<https://www.youtube.com/watch?v=mmQl6VGvX-c>



Structured Results

Google Alan Turing

All Images Videos News Books More Search tools

About 499.000 results (0,65 seconds)

Alan Turing - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Alan_Turing • Translate this page
Alan Mathison Turing CBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He was highly ...
John Clarke · Turing machine · Enigma machine · Chemical castration

Alan Turing – Wikipedia
https://de.wikipedia.org/wiki/Alan_Turing • Translate this page
Alan Mathison Turing CBE, FRS [alan matheison tʊərɪŋ] (* 23. Juni 1912 in London; † 7. Juni 1954 in Wilmslow, Cheshire) war ein britischer Logiker, ...
Enigma · Turing-Test · Turing-Maschine · Turing-Bombe

Alan Turing: the enigma
www.turing.org.uk/ •
Alan Turing (1912-1954). Large website by Andrew Hodges, biographer.

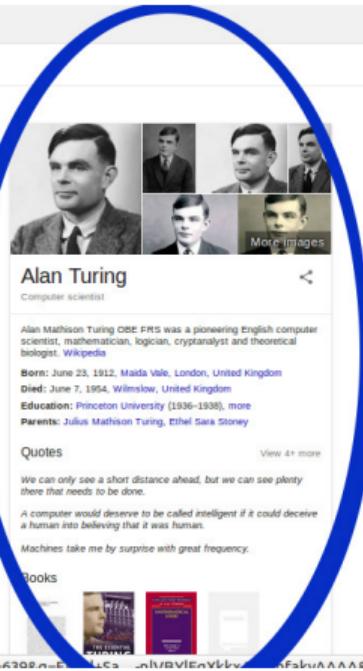
Alan Turing - a short biography
www.turing.org.uk/publications/rib.html •
This short biography, based on the entry for the writer in 1995 for the Oxford Dictionary of Scientific Biography , gives an overview of Alan Turing's life and work.

Alan Turing - Educator, Mathematician - Biography.com
www.biography.com/people/alan-turing-9512017 •
Jul 5, 2016 - Find out more about logician and mathematician Alan Turing, including what he proved in his paper "On Computable Numbers," at ...

Enigma codebreaker Alan Turing receives royal pardon | Science ...
www.theguardian.com/science/alan-turing •
Dec 24, 2013 - Alan Turing, the second world war codebreaker who took his own life after undergoing chemical castration following a conviction for ...

Alan Turing | Science | The Guardian
<http://www.theguardian.com/science/alan-turing> •
Apr 16, 2016 - GCHQ chief apologises for 'horrifying' treatment of Alan Turing Alan Turing notebook sells for more than \$1m at New York auction.

8 things you didn't know about Alan Turing | PBS NewsHour
www.pbs.org/newsHourUpdate/8-things-didnt-know-alan-turing/ •
June 28, 2014 - Alan Turing is hailed as the father of the computer, since he built the first ...

A screenshot of a Google search results page for the query "Alan Turing". The top result is a structured knowledge graph card for Alan Turing. It features a large portrait photo of him on the left and a grid of five smaller images below it. To the right of the images is a summary box containing basic facts: birth date (June 23, 1912), birth place (Maida Vale, London, United Kingdom), death date (June 7, 1954, Wilmslow, United Kingdom), education (Princeton University 1936–1938), and parents (Julius Mathison Turing, Ethel Sara Stoney). Below this summary are sections for "Quotes" and "Books", each with a few examples. A "View 4+ more" link is also present. The entire knowledge graph card is circled in blue.

Augmenting presentation with relevant facts



Google's Knowledge Graph: Surfacing Facts Proactively

Google **alan turing date of birth**

All Images News Videos Shopping More Search tools

About 565.000 results (0,79 seconds)

Alan Turing / Date of birth

June 23, 1912

Charles Babbage December 26, 1791

People also ask

Who is Alan Turing?
Why is Alan Turing famous?
Who cracked the Enigma?

Alan Turing

All Images News Books More Search tools

About 499.000 results (0,65 seconds)

Alan Turing - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Alan_Turing
Alan Mathison Turing OBE FRS was a pioneer English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He was highly regarded as a mathematical genius by his contemporaries, and is widely considered to be the father of modern computing.

Alan Turing – Wikipedia
https://de.wikipedia.org/w/index.php?title=Alan_Turing&oldid=192333709 • Translate this page
Alan Mathison Turing OBE, FRS [ˈælən mæθɪsən ˈtʊərɪŋ] (* 23. Juni 1912 in London; † 7. Juli 1954 in Wiltshire, Cheshire) war ein britischer Logiker, ...
Enigma · Turing-Test · Turing-maschine · Turing-bombe

Alan Turing: the enigma
www.turing.org.uk/ •
Alan Turing (1912–1954). Large website by Andrew Hodges, biographer.

Alan Turing - a short biography
www.turing.org.uk/publications/dbb.html •
The short biography, based on the entry for the written in 1995 for the Oxford Dictionary of Scientific Biography, ... gives an overview of Alan Turing's life and work.

Alan Turing - Educator, Mathematician - Biography.com
www.biography.com/people/alan-turing-9512017 •
Jul 5, 2016 - Find out more about logician and mathematician Alan Turing, including what he proved in his paper, 'On Computable Numbers,' at ...

Enigma codebreaker Alan Turing receives royal pardon | Science ...
www.theguardian.com/science/alan-turing
Dec 24, 2013 - Alan Turing, the second world war codebreaker who took his own life after undergoing chemical castration following a conviction for ...

Alan Turing | Science | The Guardian

Alan Turing

Computer scientist

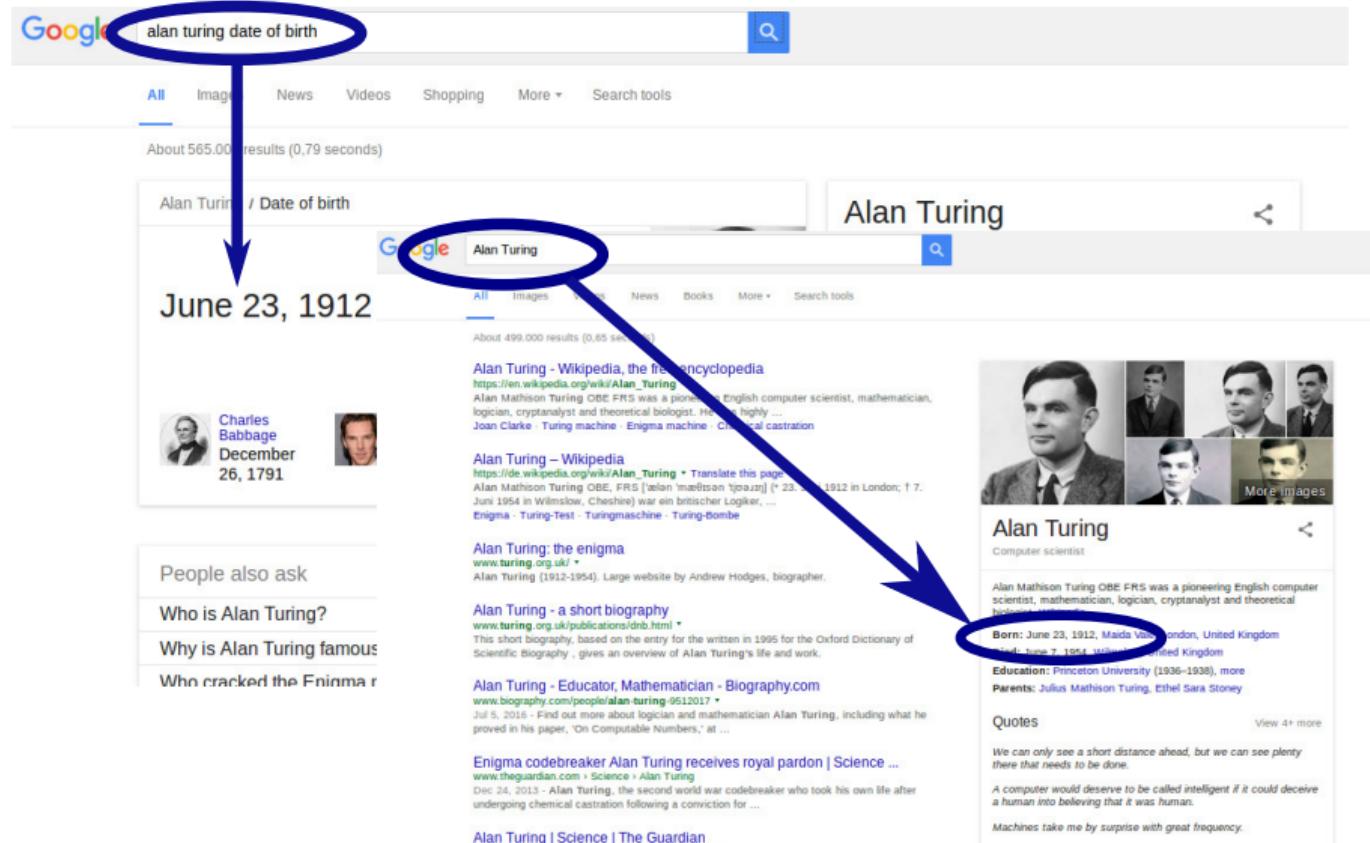
Born: June 23, 1912, Maida Vale, London, United Kingdom
Died: July 7, 1954, London, United Kingdom
Education: Princeton University (1936–1938), more
Parents: Julius Mathison Turing, Ethel Sara Stoney

Quotes

We can only see a short distance ahead, but we can see plenty there that needs to be done.

A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.

Machines take me by surprise with great frequency.





Google's Knowledge Graph: Exploratory Search

A screenshot of a Google search results page. The search query is "famous computer scientists". The results include a "Computer Scientists" knowledge graph card featuring a grid of 30 portraits of influential figures like Alan Turing, Grace Hopper, Donald Knuth, Tim Berners-Lee, Ada Lovelace, Larry Page, Sergey Brin, John von Neumann, Edsger W. Dijkstra, Ken Thompson, Niklaus Wirth, and Dennis Ritchie. A blue arrow points to this card. Below the card is a snippet from a website about the 30 most influential living computer scientists.

The 30 Most Influential Computer Scientists Alive Today
www.computersciencedegreehub.com/30-most-influential-computer-scientists-alive-to...
There are surely more than thirty influential computer scientists who could have been included on this list, but the following men and women were chosen based ...

List of pioneers in computer science - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/List_of_pioneers_in_computer_science

This article presents a list of individuals who helped in the creation, development and imagining of what computers and electronics could do.

List of computer scientists - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/List_of_computer_scientists

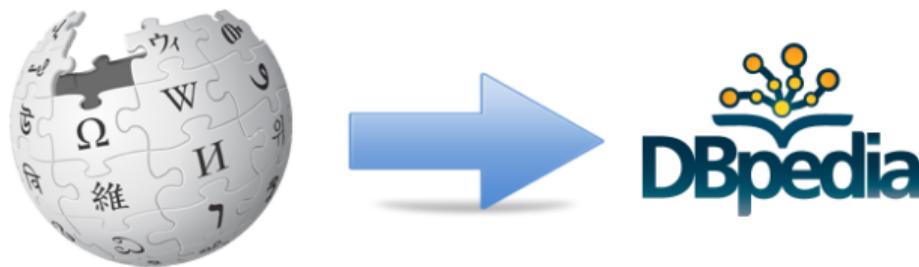
This is a list of computer scientists, people who do work in computer science, in particular researchers and authors. Some persons notable as programmers are ...



github.com/dbpedia/

- ▷ Crowd-sourced community effort to extract structured information from Wikipedia and Wikidata
- ▷ Currently describes 4.58 million things - 4.22 million are classified in a consistent ontology
- ▷ Available in 125 languages
- ▷ Enriches the extracted information with semantic layer
- ▷ Allows semantic query of relationships and properties + many other tools

- ▷ Community effort to extract structured information from Wikipedia and to make this information available on the Web
- ▷ Allows to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data



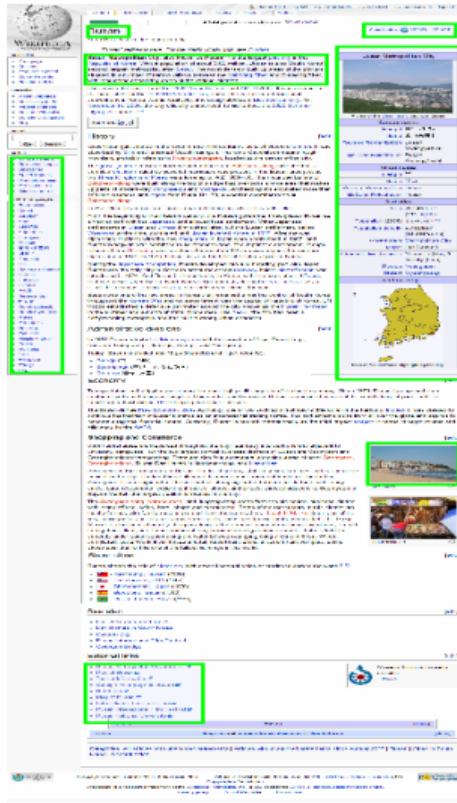
- ▷ Semi-structured Wiki markup → structured information



Simple Questions are hard to answer with Wikipedia:

- ▷ What do Innsbruck and Leipzig have in common?
- ▷ Who are mayors of central European towns elevated more than 1000m?
- ▷ Which movies are starring both Brad Pitt and Angelina Jolie?
- ▷ All soccer players, who played as goalkeeper for a club that has a stadium with more than 40.000 seats and who are born in a country with more than 10 million inhabitants

- ▷ Title
- ▷ Abstract
- ▷ Infoboxes
- ▷ Geo-coordinates
- ▷ Categories
- ▷ Images
- ▷ Links
 - other language versions
 - other Wikipedia pages
 - To the Web
 - Redirects
 - Disambiguation
- ▷ ...



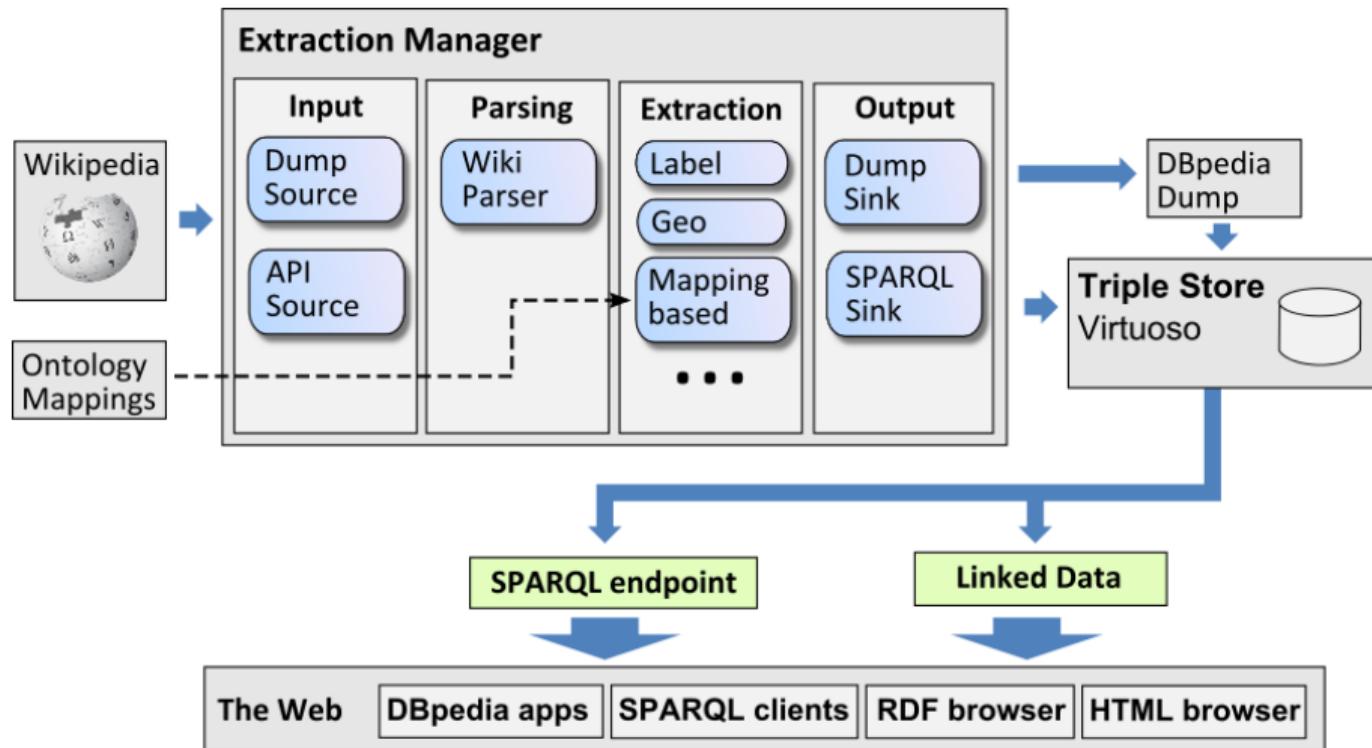


DBpedia Information Extraction Framework (DIEF)

- ▷ Started in 2007
- ▷ Hosted on Sourceforge and Github
- ▷ Initially written in PHP but fully re-written Written in Scala and Java
- ▷ Around 40 Contributors
- ▷ See <https://www.ohloh.net/p/dbpedia> for detailed overview

Can potentially be adapted to other MediaWikis

- ▷ Currently Wiktionary <http://wiktionary.dbpedia.org>



WikiText syntax

```

{{Infobox Korean settlement
|title = Busan Metropolitan City
...
|area_km2 = 763.46
|pop = 3635389
|region = [[Yeongnam]]
}}

```

RDF serialization

```

dbp:Busan dbp:title "Busan Metropolitan City"
dbp:Busan dbp:area_km2 "763.46" ^xsd:float
dbp:Busan dbp:pop "3635389" ^xsd:int
dbp:Busan dbp:region dbp:Yeongnam

```

Busan Metropolitan City	
	
<p>A view of the Geumjeong district in Busan</p>	
Korean name	
Hangul	부산광역시
Hanja	釜山廣域市
Revised Romanization	Busan
	Gwangyeoksi
McCune-Reischauer	Pusan
	Kwangyōkshī
Short name	
Hangul	부산
Hanja	釜山
Revised Romanization	Busan
McCune-Reischauer	Pusan
Statistics	
Area	763.46 km ² (295 sq mi)
Population (2005)	3,635,389 [1]
Population density	4,762/km ² (12,334/sq mi)
Government	Metropolitan City
Mayor	Hur Nam-sik
Administrative divisions	15 wards (Gu), 1 county (Gun)
Region	Yeongnam
Dialect	Gyeongsang

George A. Romero



Romero in Venice, 2009

Born George Andrew Romero
February 4, 1940 (age 69)
New York, NY, U.S.

Occupation film director
screenwriter
editor
actor

Spouse(s) Christine Forrest (1981–present)

Björk



Björk performance at Rock en Seine, Paris 2007

Background information

Birth name Björk Guðmundsdóttir

Born 21 November 1965
(age 43)
Reykjavík, Iceland

Genres Electronica, alternative rock, trip hop, IDM

Occupations Musician, songwriter, producer, actress, fashion model

**The Right Honourable
Gordon Brown
MP**



**Prime Minister of the United Kingdom
Incumbent**

Assumed office 27 June 2007

Monarch Elizabeth II

Preceded by Tony Blair

Chancellor of the Exchequer

In office 2 May 1997 – 27 June 2007

Prime Minister Tony Blair

Preceded by Kenneth Clarke

```

{{Infobox musical artist
| Name          = Björk
| Img           = Björk Rock en Seine 2007 (2).jpg
| Img_capt     = Björk performance at Rock en Sei
| Img_size      =
| Landscape    =
| Background   =
| Birth_name   = Björk Guðmundsdóttir
| Alias        =
| Born         = {{Birth date and age|1965-12-01|1965-12-01}}
| Died         =
| Origin       =
| Instrument  = [[Singer|Vocals]], [[Keyboard instrument|Keyboard in
| Voice_type   = [[Soprano]]<ref>http://www.soprani
| Genre        = [[Electronical]], [[alternative r
| Occupation   = [[Musician]], [[songwriter]], [[record producer|producer]], [[actor|actress]], [[fashion model]]
| Years active =
| Label        = 1977-present
| Snakklaysall =
| Associated_acts = [[One Little Indian]], [[Elektra Records|Elektra]], [[Atlantic Records Group|Atlantic]], [[Polydor Records|Polydor]], [[Mother Re
| URL          = [[The Sugarcubes]], [[Thom Yorke]], [[Tappi Tikkara]], [[KUKL (group)|KUKL]]
}}
```

Björk (Musician)

Occupation = Musician, Actor

Born = 21.12.1965, Reykjavík

[[harp]]

```

{{Infobox Prime Minister
|honorific-prefix = <small>
|name            = Gordon Brown
|honorific-suffix = <br /><small>
|image           = Gordon Brown official portrait.jpg
|imagesize       = 200px
|alt             = Gordon Brown
|office          = [[Prime Minister of the United Kingdom|Prime Minister of the UK]]
|monarch         = [[Elizabeth II|Elizabeth II]]
|term_start      = 27 June 1997
|term_end        =
|predecessor     = [[Tony Blair|Tony Blair]]
|successor       =
|office2         = [[Chancellor of the Exchequer|Chancellor of the Exchequer]]
|term_start2     = 2 May 1997
|term_end2       = 27 June 2007
|prime_minister2 = [[Tony Blair|Tony Blair]]
|predecessor2    = [[Kenneth Clarke|Kenneth Clarke]]
|successor2      = [[Alistair Darling|Alistair Darling]]
```

Brown (Prime Minister)

office = Prime Minister of the UK

birth_date = 20.4.1951

birth_place = Govan

Romero (Actor)

occupation = Actor, Editor

birthdate = 4.2.1940

birthplace = New York

```

{{Infobox actor
| image = George Romero. 66
| imagesize = 200px
| caption = Romero in Venice
| birthname = George Andrew Romero
| birthdate = {{birth date|1940-02-04|1940-02-04}}
| birthplace = [[New York City|New York, United States]]
| spouse = Christine Forrester
| occupation = [[film director|film director]]}}
```



Cleaner data:

- ▷ Combine what belongs together (birth_place, birthplace)
- ▷ Separate what is different (bornIn, birthplace)
- ▷ Correct handling of datatypes

Mappings Wiki:

- ▷ <http://mappings.dbpedia.org>
- ▷ Everybody can contribute to new mappings or improve existing ones
- ▷ ≈ 170 editors

Mapping en:Infobox book

Template Mapping (help)	
map to class	Book
Mappings	
Property Mapping (help)	
template property	author
ontology property	author
Property Mapping (help)	
template property	illustrator
ontology property	illustrator

Class Book:	
Properties	
author	
coverArtist	
firstPublicationDate	
illustrator	
isbn	
lastPublicationDate	
...	

Mapping el:Βιβλίο

Template Mapping (help)	
map to class	Book
Mappings	
Property Mapping (help)	
template property	συγγραφέας
ontology property	author
Property Mapping (help)	
template property	εικονογράφηση
ontology property	illustrator

```
 {{Infobox book
 | author      =
 | title_orig  =
 | translator   =
 | illustrator  =
 | subject     =
 | genre       =
 }}
```

```
 {{Βιβλίο
 | συγγραφέας =
 | ειδος      =
 | εκδότης   =
 | πρώτη_έκδοση =
 | ISBN        =
 | εικονογράφηση =
 }}
```

- ▷ <http://dbpedia.org/resource/Dresden> - URI of the city of Dresden
- ▷ <http://dbpedia.org/page/Dresden> - information resource describing the city of Dresden in HTML format
- ▷ <http://dbpedia.org/data/Dresden> - information resource describing the city of Dresden in RDF/XML format
- ▷ further formats supported, e.g. <http://dbpedia.org/data/Dresden.n3> for N3

About: Dresden

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

Dresden ist die Landeshauptstadt des Freistaates Sachsen. Sie liegt in der Dresdner Elbtalweitung an den Übergängen vom Ober- zum Mittellauf der Elbe und von der Mittelgebirgsschwelle zum Norddeutschen Tiefland im Süden Ostdeutschlands. Archäologische Spuren auf dem späteren Stadtgebiet deuten auf eine Besiedlung schon in der Steinzeit hin.

Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	▪ 328.8
dbpedia-owl:abstract	▪ Dresden on Saksan liittotasavallan osav. on 495 181 (2005) ja suur-Dresdenin asu



DBpedia Links

<i>Data set</i>	<i>Predicate</i>	<i>Count</i>	<i>Tool</i>
Amsterdam Museum	owl:sameAs	627	S
BBC Wildlife Finder	owl:sameAs	444	S
Book Mashup	rdf:type	9 100	
	owl:sameAs		
Bricklink	dc:publisher	10 100	
CORDIS	owl:sameAs	314	S
Dailymed	owl:sameAs	894	S
DBLP Bibliography	owl:sameAs	196	S
DBTune	owl:sameAs	838	S
Diseasome	owl:sameAs	2 300	S
Drugbank	owl:sameAs	4 800	S
EUNIS	owl:sameAs	3 100	S
Eurostat (Linked Stats)	owl:sameAs	253	S
Eurostat (WBSG)	owl:sameAs	137	
CIA World Factbook	owl:sameAs	545	S

<i>Data set</i>	<i>Predicate</i>	<i>Count</i>	<i>Tool</i>
flickr wrappr	dbp:hasPhoto-Collection	3 800 000	C
Freebase	owl:sameAs	3 600 000	C
GADM	owl:sameAs	1 900	
GeoNames	owl:sameAs	86 500	S
GeoSpecies	owl:sameAs	16 000	S
GHO	owl:sameAs	196	L
Project Gutenberg	owl:sameAs	2 500	S
Italian Public Schools	owl:sameAs	5 800	S
LinkedGeoData	owl:sameAs	103 600	S
LinkedMDB	owl:sameAs	13 800	S
MusicBrainz	owl:sameAs	23 000	
New York Times	owl:sameAs	9 700	
OpenCyc	owl:sameAs	27 100	C
OpenEI (Open Energy)	owl:sameAs	678	S

<i>Data set</i>	<i>Predicate</i>	<i>Count</i>	<i>Tool</i>
Revyu	owl:sameAs	6	
Sider	owl:sameAs	2 000	S
TCMGeneDIT	owl:sameAs	904	
UMBEL	rdf:type	896 400	
US Census	owl:sameAs	12 600	
WikiCompany	owl:sameAs	8 300	
WordNet	dbp:wordnet_type	467 100	
YAGO2	rdf:type	18 100 000	
Sum		27 211 732	

Abbreviations:

S: Silk (A Link Discovery Framework), L: LIMES (A Link Discovery Framework),

C: Custom Script, Missing: no regeneration



DBpedia Links - A Query Example

Getting the list of all musicians who were born before 1900 in Germany ordered by their birth date.

```
SELECT ?name ?birth ?death ?person ?countryOfBirth
WHERE {
?person rdf:type dbo:MusicalArtist .
?person foaf:name ?name .
?person dbo:birthPlace/dbo:country* ?countryOfBirth .
?person dbo:birthDate ?birth .
?person dbo:deathDate ?death .
FILTER ( ?birth < "1900-01-01"^^xsd:date ) .
FILTER ( ?countryOfBirth = :Germany ) .
}
ORDER BY ?birth
```

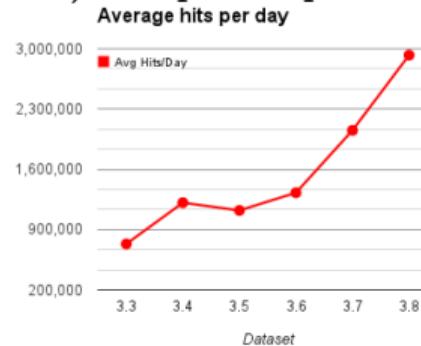
DBpedia has two extraction modes:

- ▷ Wikipedia-database-dump-based extraction
- ▷ DBpedia Live synchronisation (more later)

DBpedia Dumps:

- ▷ The DBpedia Dump archive is located in:
<http://downloads.dbpedia.org/>
- ▷ Latest downloads is described in: <http://dbpedia.org/Downloads>

Official Endpoint (by OpenLink): <http://dbpedia.org/sparql>



Back to our Wikipedia questions:

- ▷ What do Innsbruck and Leipzig have in common?
- ▷ Who are mayors of central European towns elevated more than 1000m?
- ▷ Which movies are starring both Brad Pitt and Angelina Jolie?
- ▷ All soccer players, who played as goalkeeper for a club that has a stadium with more than 40.000 seats and who are born in a country with more than 10 million inhabitants

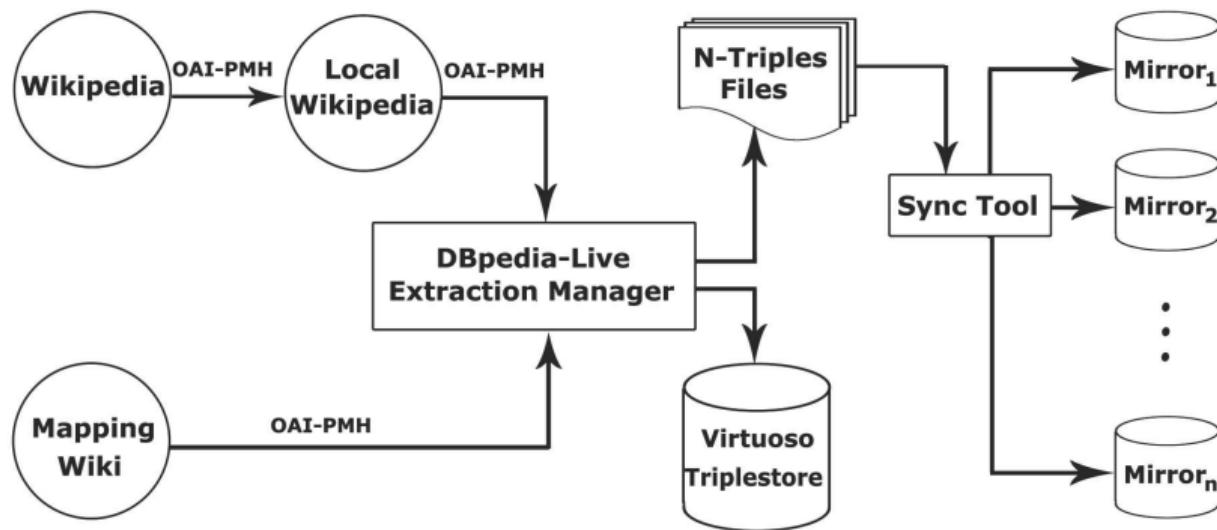
Using the data extracted from Wikipedia and the public SPARQL endpoint DBpedia can answer these questions.



- ▷ DBpedia dumps are generated on a bi-annual basis
- ▷ Wikipedia has around 100,000-150,000 page edits per day
- ▷ DBpedia Live pulls page updates in real-time and extraction results update the triple store
- ▷ In practice, a 5 minute update delay increases performance by 15%

Links

- ▷ SPARQL Endpoint: <http://live.dbpedia.org/sparql>
- ▷ Documentation: <http://wiki.dbpedia.org/DBpediaLive>
- ▷ Statistics: <http://live.dbpedia.org/LiveStats/>





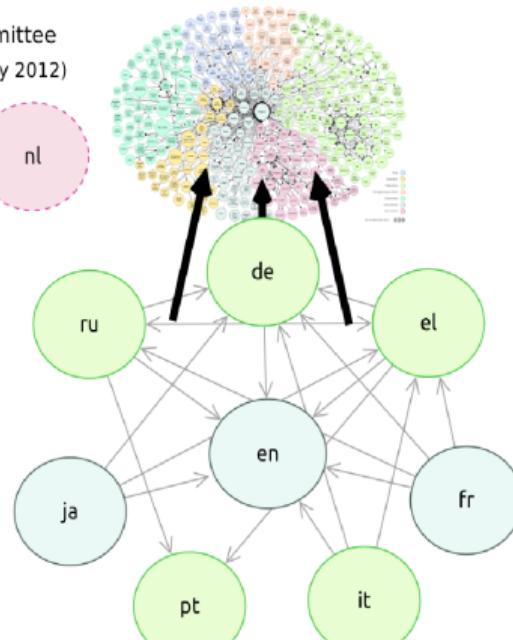
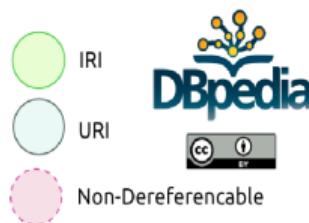
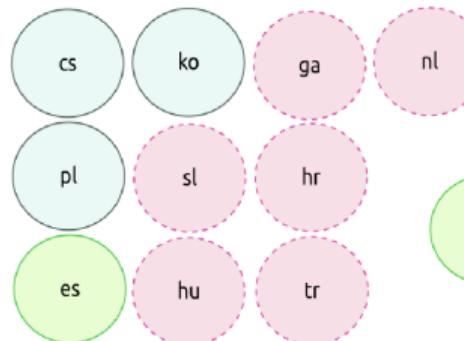
DBpedia Internationalization (I18n)

- ▷ DBpedia Internationalization Committee founded:
 - <http://wiki.dbpedia.org/Internationalization>
- ▷ Available DBpedia language editions in:
 - Korean, Greek, German, Polish, Russian, Dutch, Portuguese, Spanish, Italian, Japanese, French
 - Use the corresponding Wikipedia language edition for input
- ▷ Mappings available for 23 languages



DBpedia I18n - Overview

DBpedia Internationalization Committee
List of DBpedia language Chapters (May 2012)



<http://wiki.dbpedia.org/Internationalization>

Named entity recognition and disambiguation Tools such as: DBpedia Spotlight, AlchemyAPI, Semantic API, Open Calais, Zemanta and Apache Stanbol





Applications: Question Answering

3 elements of a successful question answering system:

- ▷ Questions are asked in natural language
- ▷ Allows people to use their own terminology
- ▷ Provides a concise answer

Example

- ▷ Which books did Dan Brown write?
- ▷ Which books have Dan Brown as their author?
- ▷ Which novels did Dan Brown write?

Applications: Question Answering

- ▷ DBpedia is the primary target for several QA systems in the Question Answering over Linked Data (QALD) workshop series
- ▷ IBM Watson relied also on DBpedia



- ▷ Faceted browsing: allows finding items by restricting the overall set of items along multiple criteria (facets)
- ▷ It is a common feature on websites that deal with structured data e.g. e-commerce sites
- ▷ Some projects:
 - Neofonie Browser
 - gFacet
 - OpenLink faceted browser (fct)



- ▷ Sparklis: A query builder in natural language that allows people to explore and query SPARQL endpoints with all the power of SPARQL and without any knowledge of SPARQL.
- ▷ ExConQuer: Works on top of DBpedia and enables users to construct a SPARQL query without requiring any knowledge of SPARQL or the datasets' underlying schema.
Users are then able to download the data they require in a number of different formats.





- ▷ Virtual International Authority Files (VIAF) project as Linked Data
 - VIAF added a total of 250,000 reciprocal authority links to Wikipedia.
- ▷ DBpedia can also provide:
 - Context information for bibliographic and archive records (e.g. an author's demographics, a film's homepage, an image etc.)
 - Stable and curated identifiers for linking.
 - The broad range of Wikipedia topics can form the basis for a thesaurus for subject indexing.

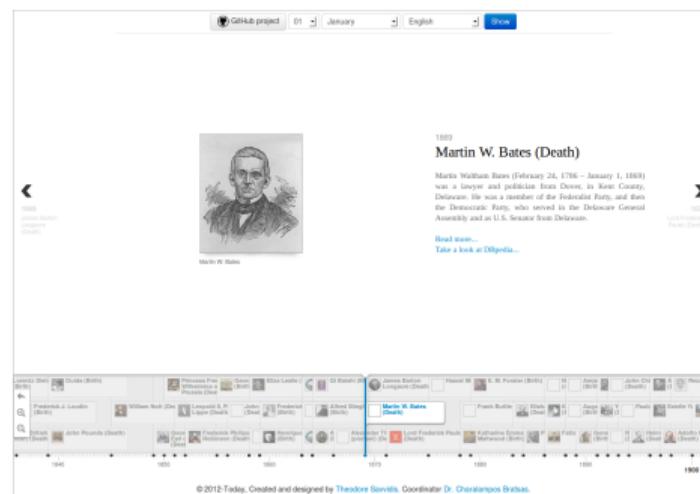
- ▷ DBpedia Mobile is a location-centric DBpedia client application for mobile devices consisting of a map view, the Marbles Linked Data Browser and a GPS-enabled launcher application.
- ▷ Its initial view is a browser-based area map that indicates the user's position and nearby DBpedia resources with appropriate labels and icons.





Other Applications

See <http://wiki.dbpedia.org/projects> for a more complete list





Statistical Relational Learning



KG Completion: Link Prediction

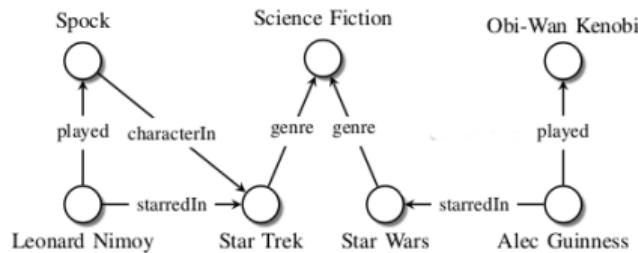
- ▷ **Problem:** KGs are often automatically extracted from text and thus may be incomplete.



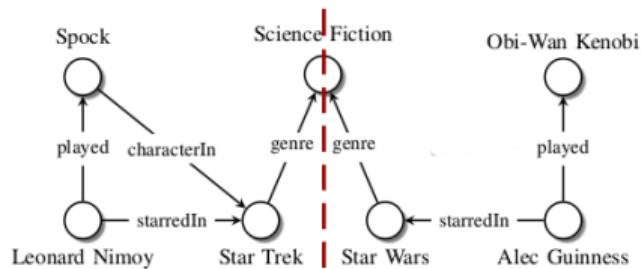
KG Completion: Link Prediction

- ▷ **Problem:** KGs are often automatically extracted from text and thus may be incomplete.
- ▷ **Goal:** use existing graph structure to predict missing edges/links.

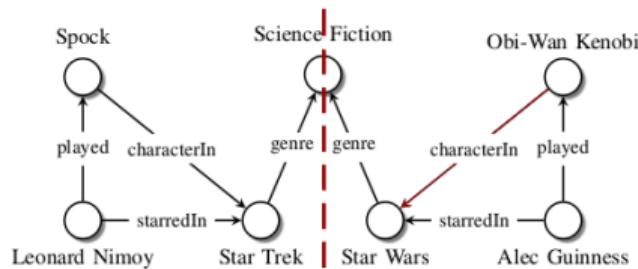
- ▷ **Problem:** KGs are often automatically extracted from text and thus may be incomplete.
- ▷ **Goal:** use existing graph structure to predict missing edges/links.



- ▷ **Problem:** KGs are often automatically extracted from text and thus may be incomplete.
- ▷ **Goal:** use existing graph structure to predict missing edges/links.



- ▷ **Problem:** KGs are often automatically extracted from text and thus may be incomplete.
- ▷ **Goal:** use existing graph structure to predict missing edges/links.



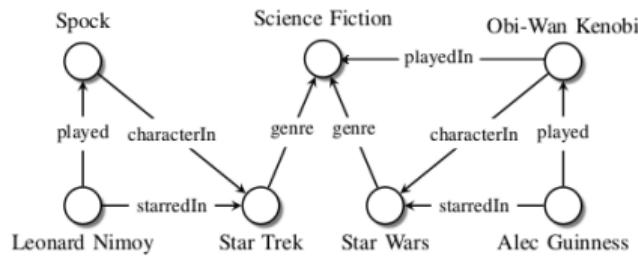


KG Correction: Link correction

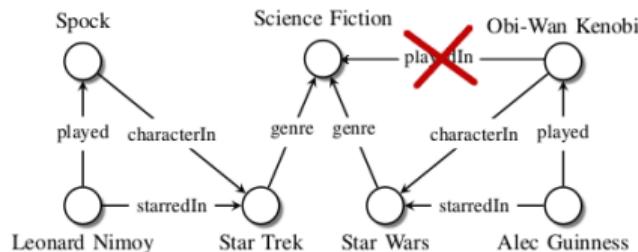
- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.

- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.
- ▷ **Goal:** use existing graph structure to estimate probability of correctness of an edge/link.

- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.
- ▷ **Goal:** use existing graph structure to estimate probability of correctness of an edge/link.



- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.
- ▷ **Goal:** use existing graph structure to estimate probability of correctness of an edge/link.



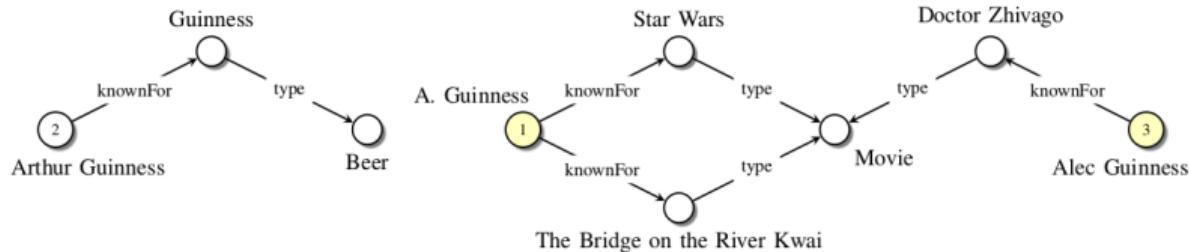


- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.

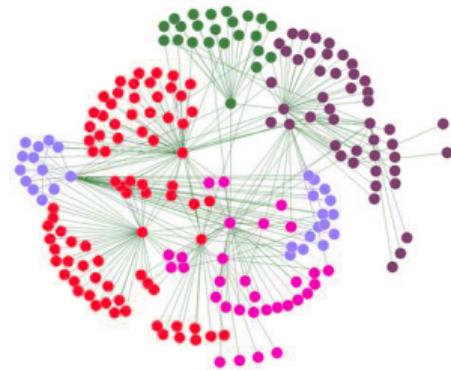


- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.
- ▷ **Goal:** use existing graph structure and other information to merge duplicated entities.

- ▷ **Problem:** KGs are often automatically extracted from text and thus may be wrong.
- ▷ **Goal:** use existing graph structure and other information to merge duplicated entities.



- ▷ **Goal:** Group entities based on their similarity.
- ▷ In contrast to feature-based clustering, in link-based clustering entities are not only grouped by the similarity of their features but also by the similarity of their links.
- ▷ E.g. community detection in social networks.



²This Wikipedia and Wikimedia Commons image is from the user Daniel Tenerife and is freely available at [//commons.wikimedia.org/wiki/File:Social_Red.jpg](https://commons.wikimedia.org/wiki/File:Social_Red.jpg) under the creative commons license.



Related problems

- ▷ You will learn to solve the problems we just saw in this lecture.
- ▷ They also play a role for:
 - merging of KGs
 - similarity estimation of entities
 - suggestion of friends in social networks
 - ...



Methods that play a role for statistical relational learning techniques covered in this class :

- ▷ tensor factorization
- ▷ neural networks
- ▷ probabilistic approaches
- ▷ optimization techniques (like stochastic gradient descent)
- ▷ ...

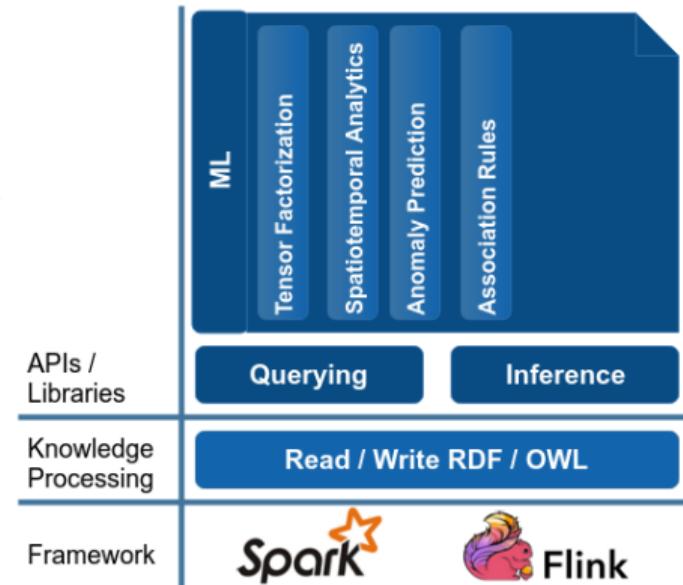
...which can not be addressed by methods learned in the lecture:

- ▷ **Growth:** knowledge graphs are incomplete!
 - *Ontology matching:* connect graphs
 - *Knowledge extraction:* extract new entities and relations from web/text
- ▷ **Validation:** knowledge graphs are not always correct!
 - *Entity resolution:* split wrongly merged entities
 - *Error detection:* remove false assertions
 - *Fact validation:* find evidence for / against facts being true
- ▷ **Interface:** how to make it easier to access knowledge?
 - *Semantic parsing:* interpret the meaning of queries
 - *Question answering:* compute answers using the knowledge graph
- ▷ **Intelligence:** can AI emerge from knowledge graphs?
 - *Automatic reasoning* and planning
 - Generalization and abstraction
 - Clustering and community detection

- ▷ Master thesis on most topics discussed in this lecture
- ▷ Machine Learning on knowledge graphs (and in general)
- ▷ Semantic technologies
- ▷ Spatiotemporal prediction
- ▷ Question answering on knowledge graphs
- ▷ **We are looking for students and offering PhD opportunities!**

- ▷ SANSA will be a suite of APIs for distributed reading, querying, inferencing and analyzing of RDF knowledge graphs

SANSA (Semantic Analytics Stack)



- ▷ KGs facilitate mapping and understanding of the world in a way that is similar to how humans understand it
- ▷ KGs utilize not just information, but knowledge, and thus are an integral part toward knowledge-powered assistance
 - ⇒ a better, more interactive user interface
 - ⇒ *understanding* of user queries instead of simple string-matching
- ▷ KGs naturally allow for efficient pattern matching: numerous applications from fraud detection to spam filters to computational molecular biology



Dr. Hamed Shariat Yazdi
shariat@cs.uni-bonn.de
University of Bonn



Prof. Dr. Jens Lehmann
<http://jens-lehmann.org>
lehmann@uni-bonn.de
jens.lehmann@iais.fraunhofer.de
University of Bonn
Fraunhofer IAIS

Thanks for your attention!

-  S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives.
Dbpedia: A nucleus for a web of open data.
In *The semantic web*, pages 722–735. Springer, 2007.
-  A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell.
Toward an architecture for never-ending language learning.
In *AAAI*, volume 5, page 3, 2010.
-  V. I. Eric Sun.
Under the hood: The entities graph.
[https://www.facebook.com/notes/facebook-engineering/
under-the-hood-the-entities-graph/10151490531588920/](https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-entities-graph/10151490531588920/).
-  Q. He.
Building the linkedin knowledge graph.
https://en.wikipedia.org/wiki/Knowledge_Graph.

References II

-  M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich.
A review of relational machine learning for knowledge graphs.
arXiv preprint arXiv:1503.00759, 2015.
-  D. R. Qian.
Understand your world with bing.
<http://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/>.
-  I. Robinson, J. Webber, and E. Eifrem.
Graph Databases: New Opportunities for Connected Data.
" O'Reilly Media, Inc.", 2015.
-  A. Singhal.
Introducing the knowledge graph: things, not strings.
<https://googleblog.blogspot.de/2012/05/introducing-knowledge-graph-things-not.html>.