Information Retrieval & Data Mining

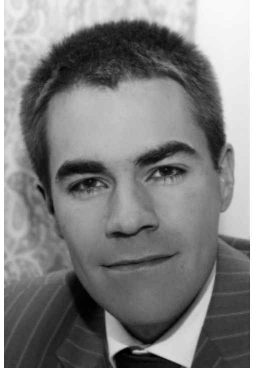
Universität des Saarlandes, Saarbrücken

Winter Semester 2013/14

The Course

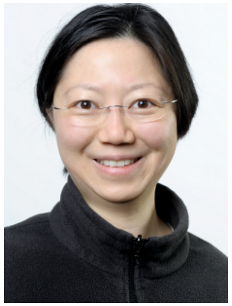
Lecturers

Klaus Berberich
kberberi@mpi-inf.mpg.de



Pauli Miettinen
pmiettin@mpi-inf.mpg.de

Teaching Assistants



Amy Siu
sui@mpi-inf.mpg.de

Arunav Mishra
amishra@mpi-inf.mpg.de

Erdal Kuzey
ekuzey@mpi-inf.mpg.de

Kai Hui
khui@mpi-inf.mpg.de

Kaustubh Beedkar
kbeedkar@mpi-inf.mpg.de

Sourav Dutta
sdutta@mpi-inf.mpg.de

D5: Databases & Information Systems Group
Max Planck Institute for Informatics

Organization

- **Lectures:**
  - **Tuesday 16-18** and **Thursday 14-16**
    in **Building E1.3**, **HS-002**
- **Office hours:**
  - **Tuesday 14-16**
- **Assignments/tutoring groups**
  - **Monday 12-14 / 14-16 / 16-18, R021, E1.4** (MPI-INF building)
  - **Friday 12-14 / 14-16, R021, E1.4** (MPI-INF building)

Assignments given out in Thursday lecture, to be solved until next Thursday

- First assignment sheet given out on **Thursday, Oct 17**
- First meetings of tutoring groups on **Friday, Oct 25**

Requirements for Obtaining 9 Credit Points

- **Pass 2 out of 3 written tests**

  Tentative dates: **Tue, Nov 12**; **Thu, Dec 12**; **Tue, Jan 28**

  (45-60 min each)

- **Pass the final written exam**

  Tentative date: **Tue, Feb 13** (120-180 min)

- Must **present solutions to 3 assignments**, more possible

  (**You must return your assignment sheet and have a correct solution in order to present in the exercise groups.**)

  - **1 bonus point** possible in tutoring groups
  - **Up to 3 bonus points** possible in tests
  - Each bonus point earns one mark in letter grade (0.3 in numerical grade)

Register for Tutoring Groups

http://bit.ly/irdm

- Register for one of the tutoring groups **until Oct 22**
- Check back frequently for updates & announcements

Agenda

I. Introduction

II. Probability theory, statistics, linear algebra

II. Ranking principles

III. Link analysis

IV. Indexing & searching

V. Information extraction

Information Retrieval

VI. Frequent itemsets & association rules

VII. Unsupervised clustering

VIII. (Semi-)supervised classification

IX. Advanced topics in data mining

Data Mining

X. Wrap-up & summary

Literature (I)

- **Information Retrieval**
  - Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze.
    Introduction to Information Retrieval
    Cambridge University Press, 2008.
    Website: http://nlp.stanford.edu/IR-book/
  - R. Baeza-Yates, R. Ribeiro-Neto.
    Modern Information Retrieval: The concepts and technology behind search.
    Addison-Wesley, 2010.
    Website: http://www.mir2ed.org
  - W. Bruce Croft, Donald Metzler, Trevor Strohman.
    Search Engines: Information Retrieval in Practice.
    Addison-Wesley, 2009.
    Website: http://www.pearsonhighered.com/croft1epreview/

Literature (II)

- **Data Mining**

  - Mohammed J. Zaki, Wagner Meira Jr.

    Data Mining and Analysis: Fundamental Concepts and Algorithms
    Manuscript (will be made available during the semester)

  - Pang-Ning Tan, Michael Steinbach, Vipin Kumar.

    Introduction to Data Mining

    Addison-Wesley, 2006.

    Website: http://www-users.cs.umn.edu/%7Ekumar/dmbook/index.php

Literature (III)

- **Background & Further Reading**
  - Jiawei Han, Micheline Kamber, Jian Pei.
    Data Mining - Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011
    Website: http://www.cs.sfu.ca/~han/dmbook
  - Stefan Büttcher, Charles L. A. Clarke, Gordon V. Cormack.
    Information Retrieval: Implementing and Evaluating Search Engines,
    MIT Press, 2010
  - David B. Skillicorn.
    Understanding complex datasets: data mining with matrix decomposition,
    Chapman & Hall/CRC, 2007
  - Christopher M. Bishop.
    Pattern Recognition and Machine Learning, Springer, 2006
  - Larry Wasserman.
    All of Statistics, Springer, 2004
    Website: http://www.stat.cmu.edu/~larry/all-of-statistics/

Quiz Time!

- Please answer the **20 quiz questions** during the rest of the lecture.
- The quiz is completely **anonymous**, but keep your id on the top-right corner. There will be a **prize for the 3 best answer** sheets.

Chapter I: Introduction – Information Retrieval and Data Mining in a Nutshell

Information Retrieval & Data Mining

Universität des Saarlandes, Saarbrücken

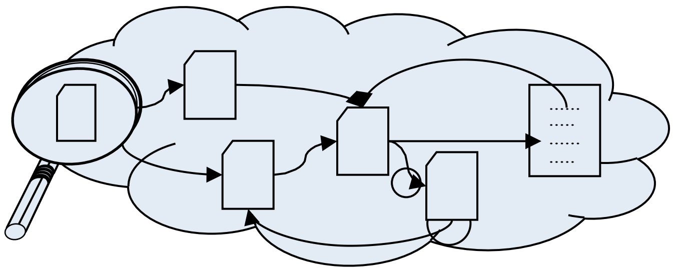
Winter Semester 2013/14

Chapter I: Information Retrieval and Data Mining in a Nutshell

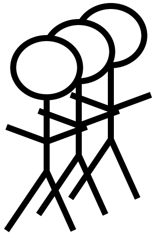
- **1.1 Information Retrieval in a Nutshell**
  - Search & beyond
- **1.2 Data Mining in a Nutshell**
  - Real-world DM applications

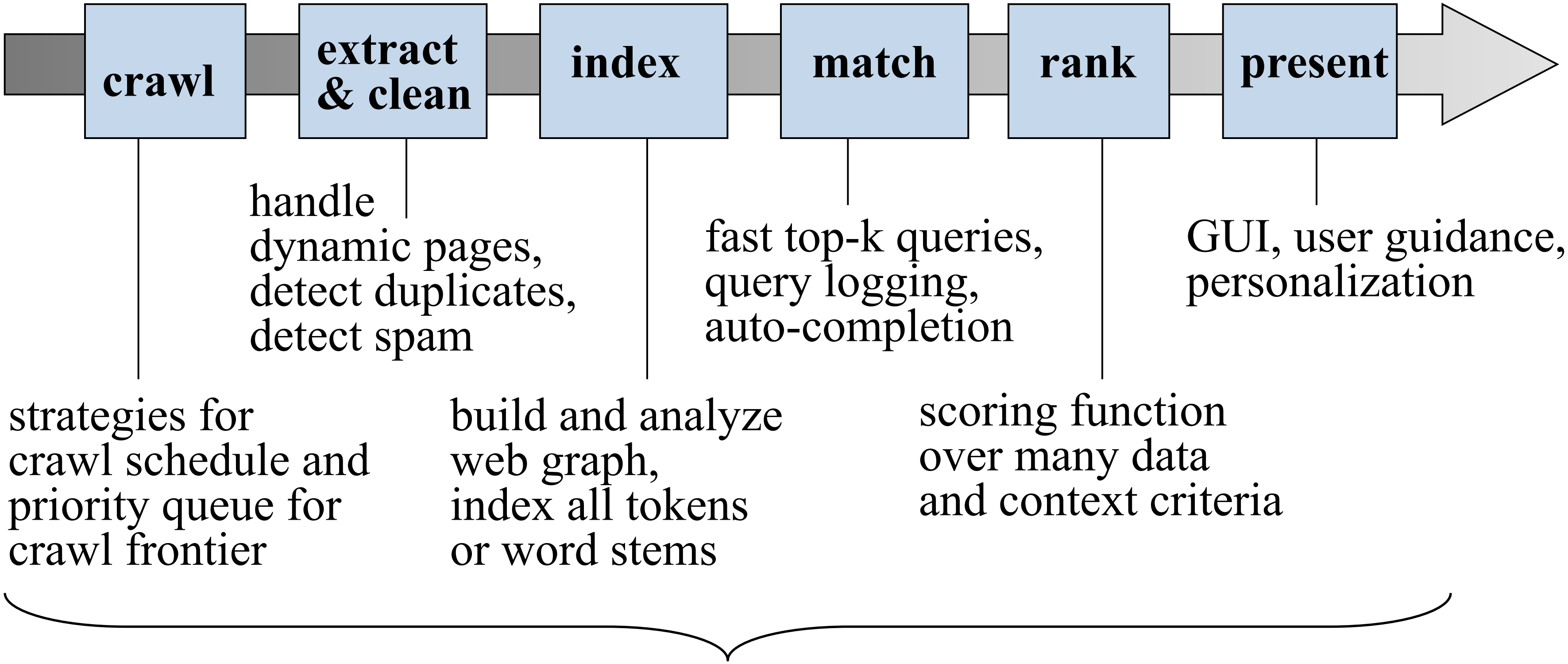
„We are drowning in information, and starved for knowledge.“
-- John Naisbitt

I.1 Information Retrieval in a Nutshell



- Web, intranet, digital libraries, desktop search
- Unstructured/semi-structured data





strategies for crawl schedule and priority queue for crawl frontier

handle dynamic pages, detect duplicates, detect spam

build and analyze web graph, index all tokens or word stems

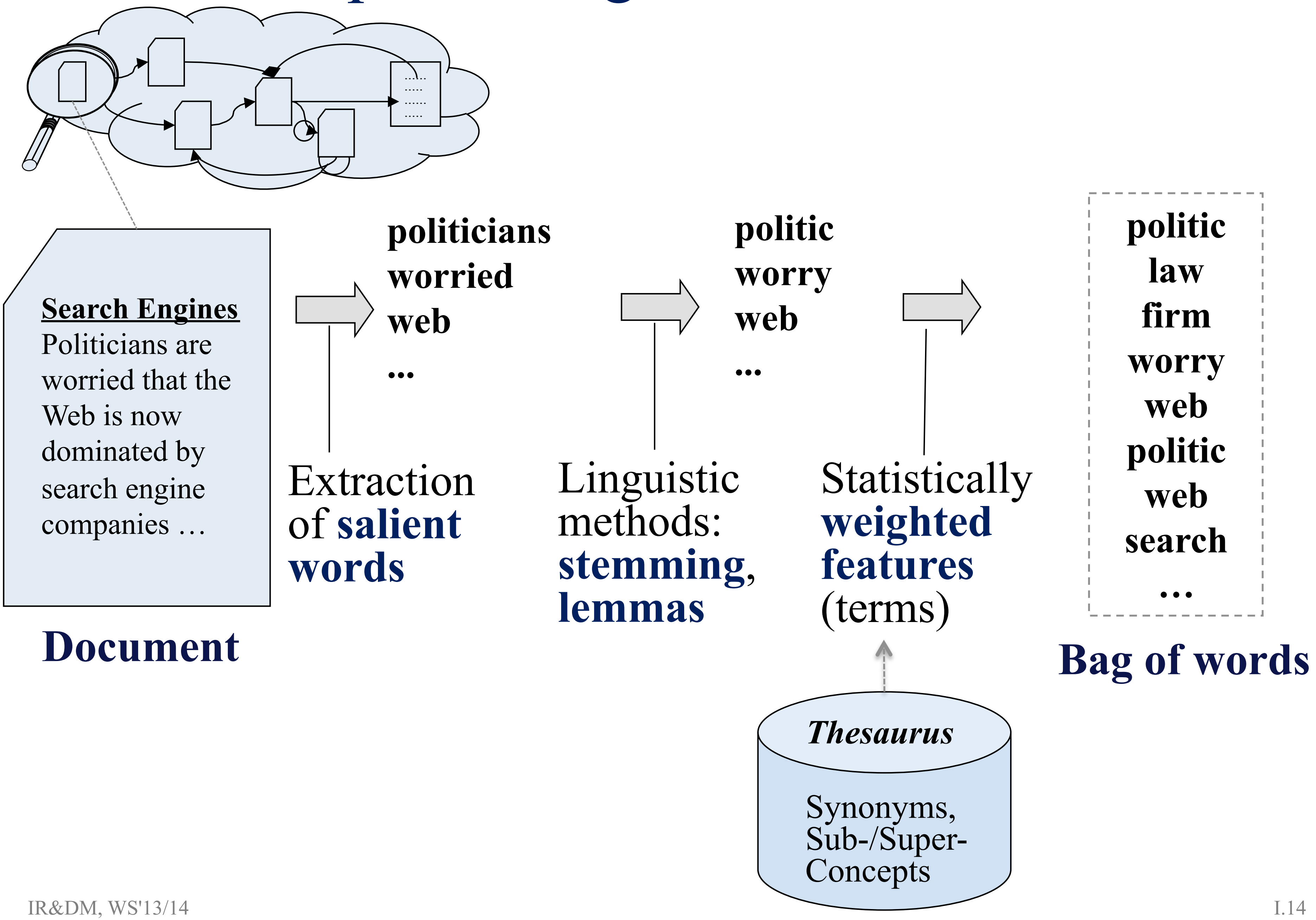
fast top-k queries, query logging, auto-completion

scoring function over many data and context criteria

GUI, user guidance, personalization

Server farms with **10 000‘s** (2002) – **100,000’s** (2010) computers,
distributed/replicated data in high-performance file system (**GFS**, **HDFS**,…),
massive parallelism for query processing (**MapReduce**, **Hadoop**,…)

Content Preprocessing



Search Engines
Politicians are worried that the Web is now dominated by search engine companies …

Document

politicians
worried
web
…

Extraction of **salient words**

politic
worry
web
…

Linguistic methods: **stemming, lemmas**

Statistically **weighted features** (terms)

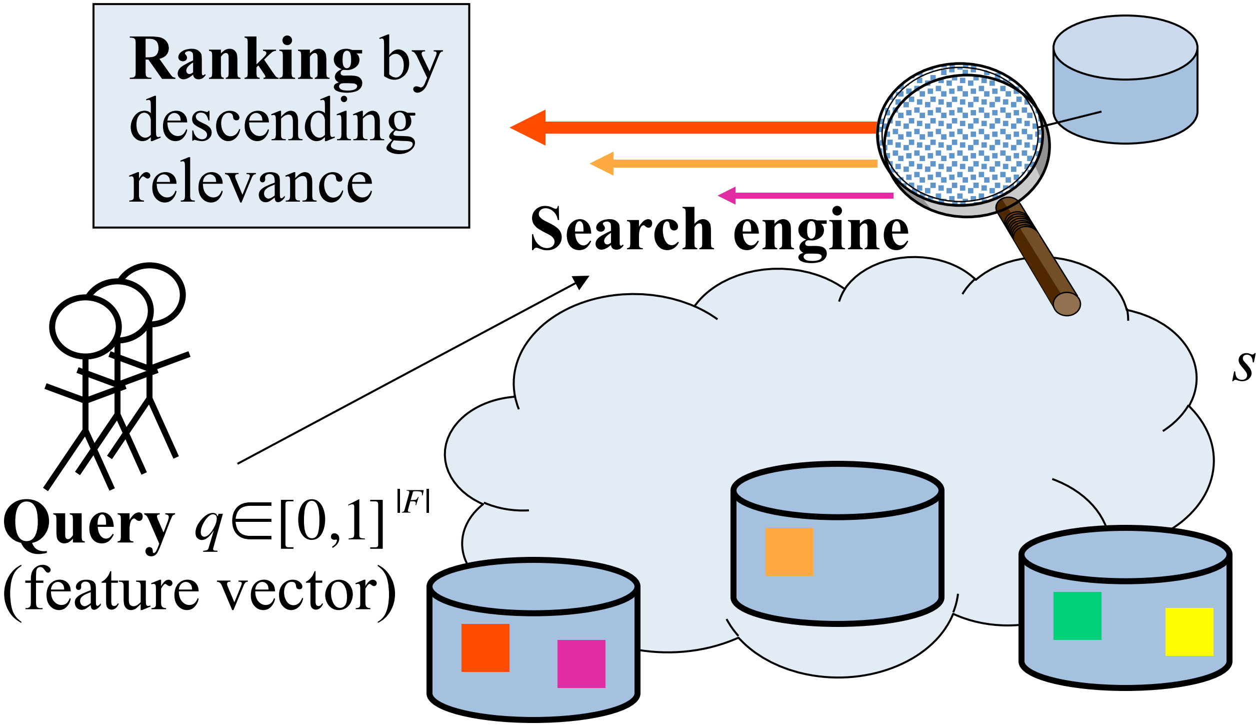
Thesaurus

Synonyms, Sub-/Super-Concepts

politic
law
firm
worry
web
politic
web
search
…

Bag of words

Vector Space Model for Relevance Ranking



Ranking by descending relevance

Search engine

Query $q \in [0,1]^{|F|}$ (feature vector)

Documents are **feature vectors**

Similarity metric:

$$sim(d_i, q) := \frac{\sum_{j=1}^{|F|} d_{ij}\, q_j}{\sqrt{\sum_{j=1}^{|F|} d_{ij}^2 \sum_{j=1}^{|F|} q_j^2}}$$

with $d_i \in [0,1]^{|F|}$

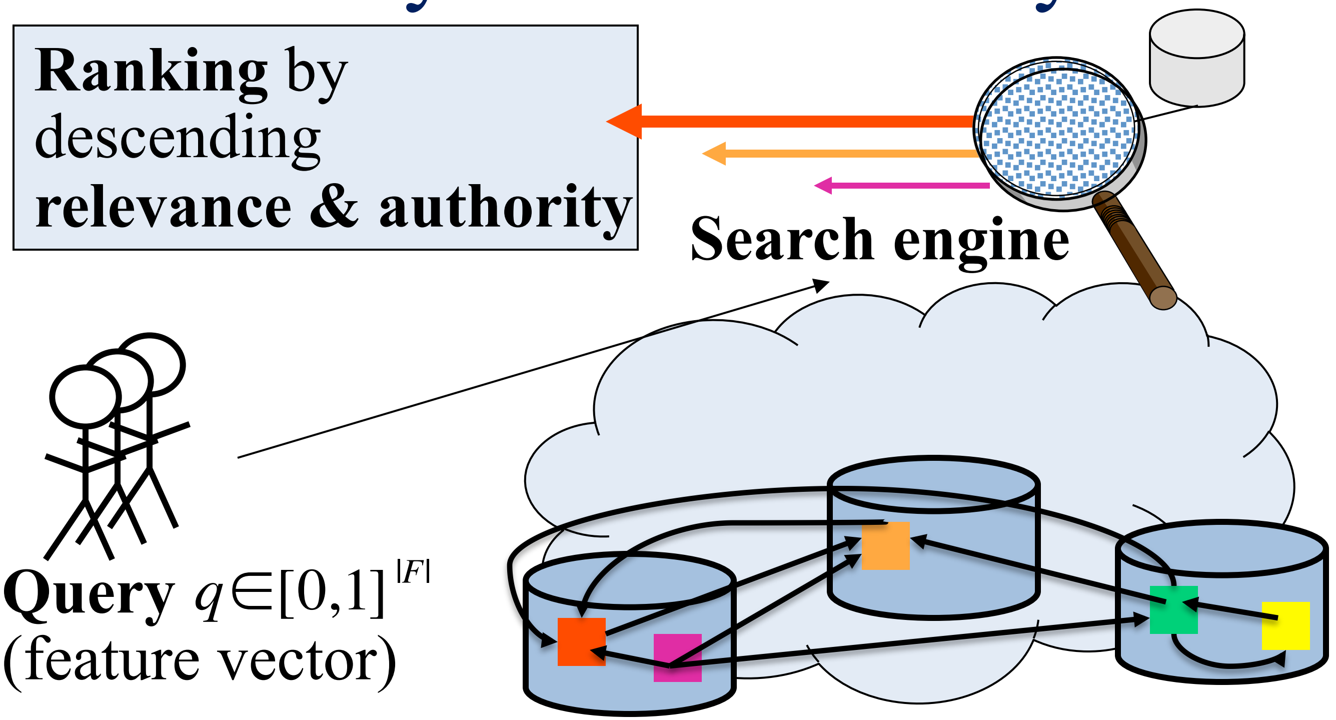
e.g., using: $d_{ij} := w_{ij} / \sqrt{\sum_k w_{ik}^2}$

$$w_{ij} := \log\left(1 + \frac{freq(f_j, d_i)}{\max_k \, freq(f_k, d_i)}\right) \log \frac{\#docs}{\#docs\ with\ f_i}$$

tf*idf formula

Link Analysis for Authority Ranking

Ranking by descending **relevance & authority**



Search engine

Query $q \in [0,1]^{|F|}$
(feature vector)

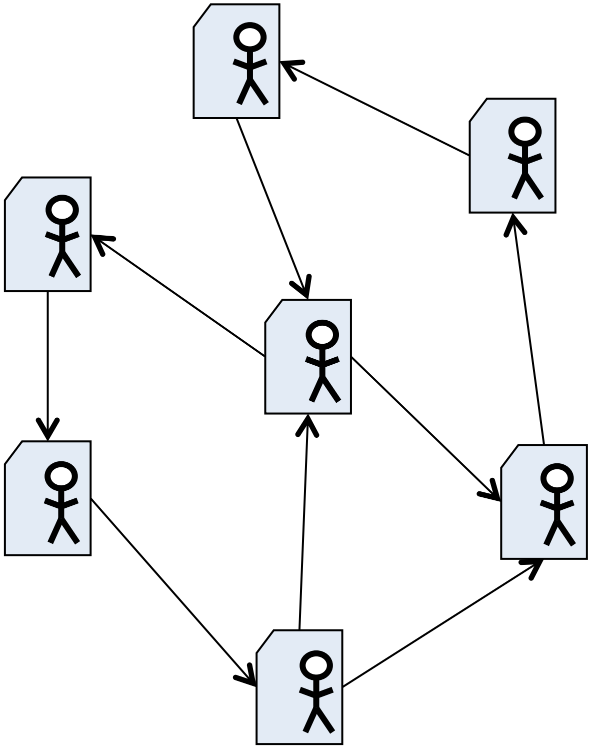
+ Consider in-degree and out-degree of web pages:
Authority (d_i) :=
Stationary visiting probability [d_i]
in random walk on the Web (ergodic Markov Chain)

+ Reconciliation of relevance and authority by ad hoc weighting

Google's PageRank [Page and Brin 1998]

- **Ideas:** (i) Hyperlinks are endorsements
  (ii) Page is important if many important pages link to it
- **Random walk** on web graph *G(V, E)* with random surfer that randomly follows outgoing link or jumps to another random page

$$P(v) = (1-\varepsilon) \sum_{(u,v)\in E} \frac{P(u)}{out(u)} + \frac{\varepsilon}{|V|}$$

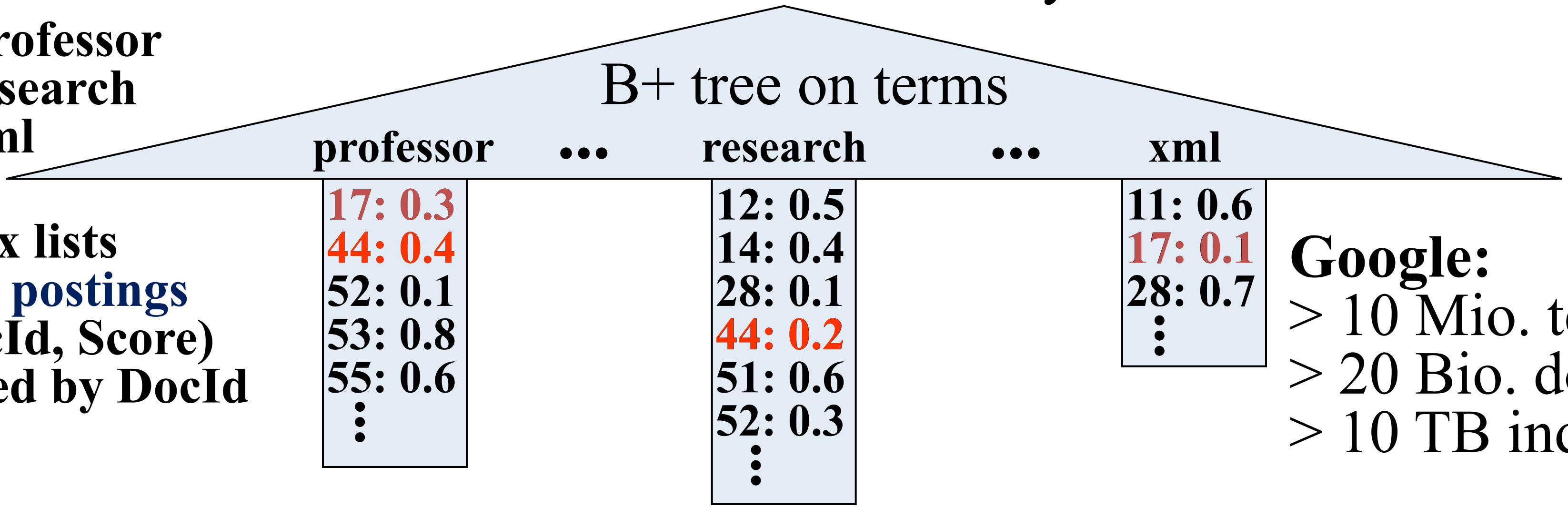


- **PageRank** *P(v)* corresponds to the stationary visiting probability of state *v* in an ergodic Markov chain

Inverted Index

Vector space model suggests **term-document matrix**,
but data is sparse and queries are even very sparse
→ better use **inverted index** with terms as keys for B+ tree

q: professor
research
xml



B+ tree on terms

professor **•••** **research** **•••** **xml**

professor	research	xml
17: 0.3	**12: 0.5**	**11: 0.6**
44: 0.4	**14: 0.4**	**17: 0.1**
52: 0.1	**28: 0.1**	**28: 0.7**
53: 0.8	**44: 0.2**	⋮
55: 0.6	**51: 0.6**	
⋮	**52: 0.3**	
	⋮	

index lists with postings (DocId, Score) sorted by DocId

Google:
> 10 Mio. terms
> 20 Bio. docs
> 10 TB index

terms can be full words, word stems, word pairs, substrings, N-grams, etc.
(whatever "dictionary terms" we prefer for the application)

- index-list entries in **DocId order** for fast Boolean operations
- many techniques for excellent **compression** of index lists
- additional **position index** needed for phrases, proximity, etc.

(or other pre-computed data structures)

Evaluation of Search Result Quality

Ideal measure is "**satisfaction of user's information need**" heuristically approximated by benchmarking measures (on test corpora with query suite and relevance assessment by experts)

Capability to return **only** relevant documents:

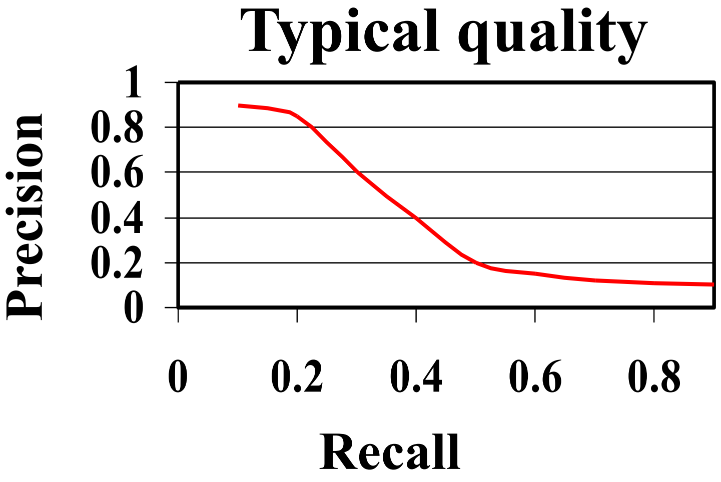
$$\textbf{\textit{Precision}} = \frac{\#\ relevant\ docs\ among\ top\ r}{r}$$

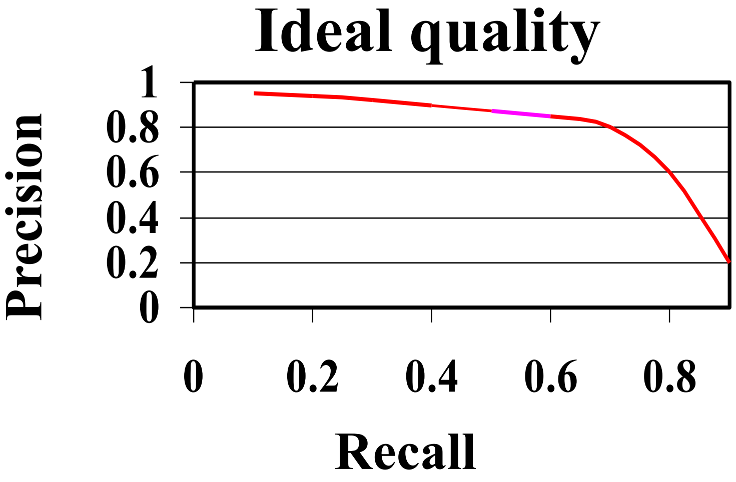
typically for
r = 10, 100, 1000

Capability to return **all** relevant documents:

$$\textbf{\textit{Recall}} = \frac{\#\ relevant\ docs\ among\ top\ r}{\#\ relevant\ docs}$$

typically for
r = corpus size



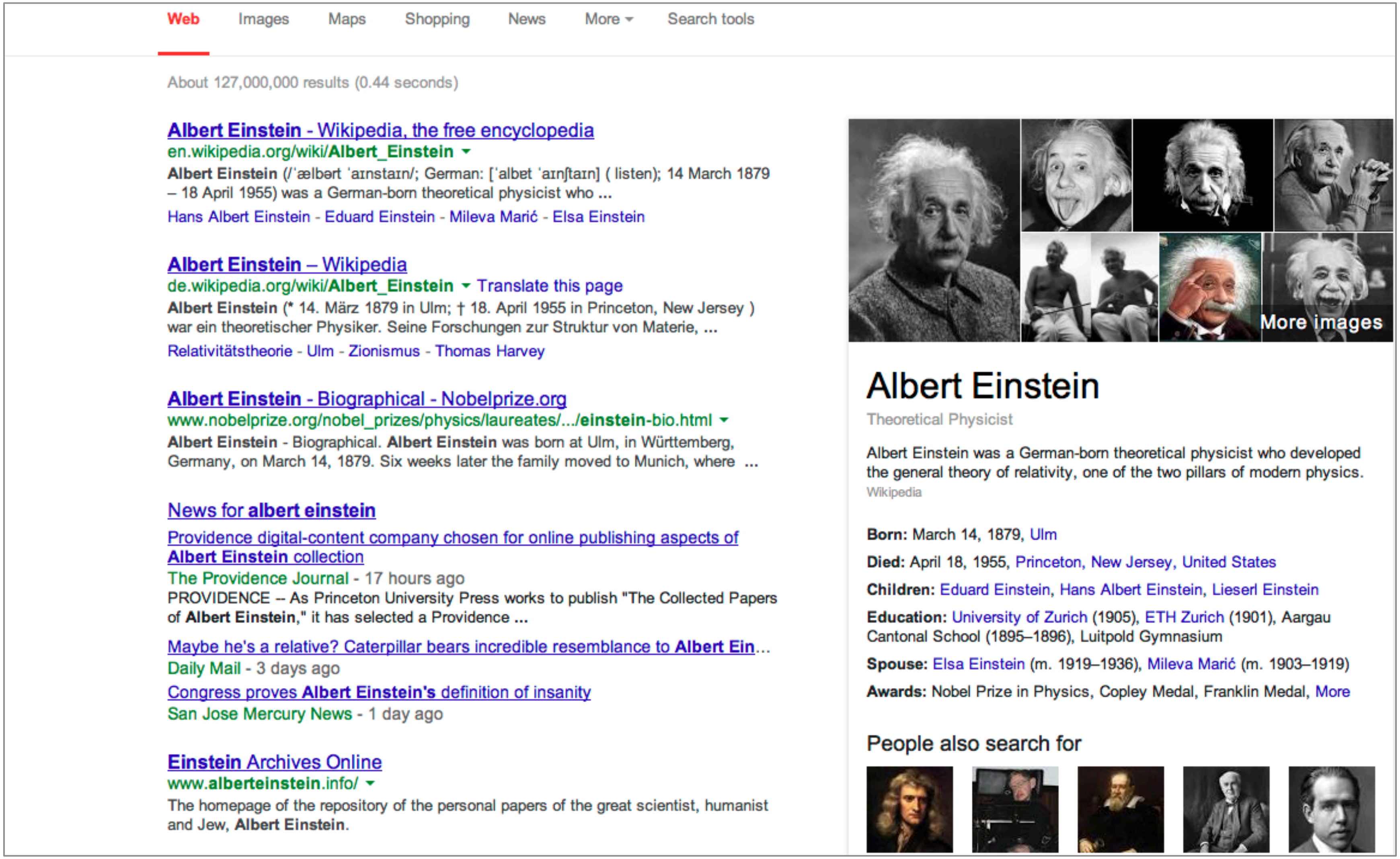


Beyond Web Search…

- Find answers to **“knowledge queries”** and **natural language questions** (e.g., by scientists or journalists)
  - *Who was German chancellor when Angela Merkel was born?*
  - *How are Max Planck, Angela Merkel, and the Dalai Lama related?*
  - *Which politicians are also entrepreneurs?*
  - *What was the population of Munich in 1972?*
  - …

- Knowledge about **entities** (e.g., persons and locations), **classes**, **attributes**, **relationships** between them is required
  - focus on **structured data sources** (e.g., relational, XML, RDF)
  - perform **information extraction** on semi-structured & textual data

Google Knowledge Graph



About 127,000,000 results (0.44 seconds)

Albert Einstein - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Albert_Einstein**
Albert Einstein (/ˈælbərt ˈaɪnstaɪn/; German: [ˈalbɐt ˈaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist who ...
Hans Albert Einstein - Eduard Einstein - Mileva Marić - Elsa Einstein

Albert Einstein – Wikipedia
de.wikipedia.org/wiki/**Albert_Einstein** Translate this page
Albert Einstein (* 14. März 1879 in Ulm; † 18. April 1955 in Princeton, New Jersey) war ein theoretischer Physiker. Seine Forschungen zur Struktur von Materie, ...
Relativitätstheorie - Ulm - Zionismus - Thomas Harvey

Albert Einstein - Biographical - Nobelprize.org
www.nobelprize.org/nobel_prizes/physics/laureates/.../**einstein**-bio.html
Albert Einstein - Biographical. **Albert Einstein** was born at Ulm, in Württemberg, Germany, on March 14, 1879. Six weeks later the family moved to Munich, where ...

News for **albert einstein**

Providence digital-content company chosen for online publishing aspects of **Albert Einstein** collection
The Providence Journal - 17 hours ago
PROVIDENCE -- As Princeton University Press works to publish "The Collected Papers of **Albert Einstein**," it has selected a Providence ...

Maybe he's a relative? Caterpillar bears incredible resemblance to **Albert Ein**...
Daily Mail - 3 days ago

Congress proves **Albert Einstein's** definition of insanity
San Jose Mercury News - 1 day ago

Einstein Archives Online
www.**alberteinstein**.info/
The homepage of the repository of the personal papers of the great scientist, humanist and Jew, **Albert Einstein**.

Albert Einstein
Theoretical Physicist

Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, one of the two pillars of modern physics.
Wikipedia

Born: March 14, 1879, Ulm

Died: April 18, 1955, Princeton, New Jersey, United States

Children: Eduard Einstein, Hans Albert Einstein, Lieserl Einstein

Education: University of Zurich (1905), ETH Zurich (1901), Aargau Cantonal School (1895–1896), Luitpold Gymnasium

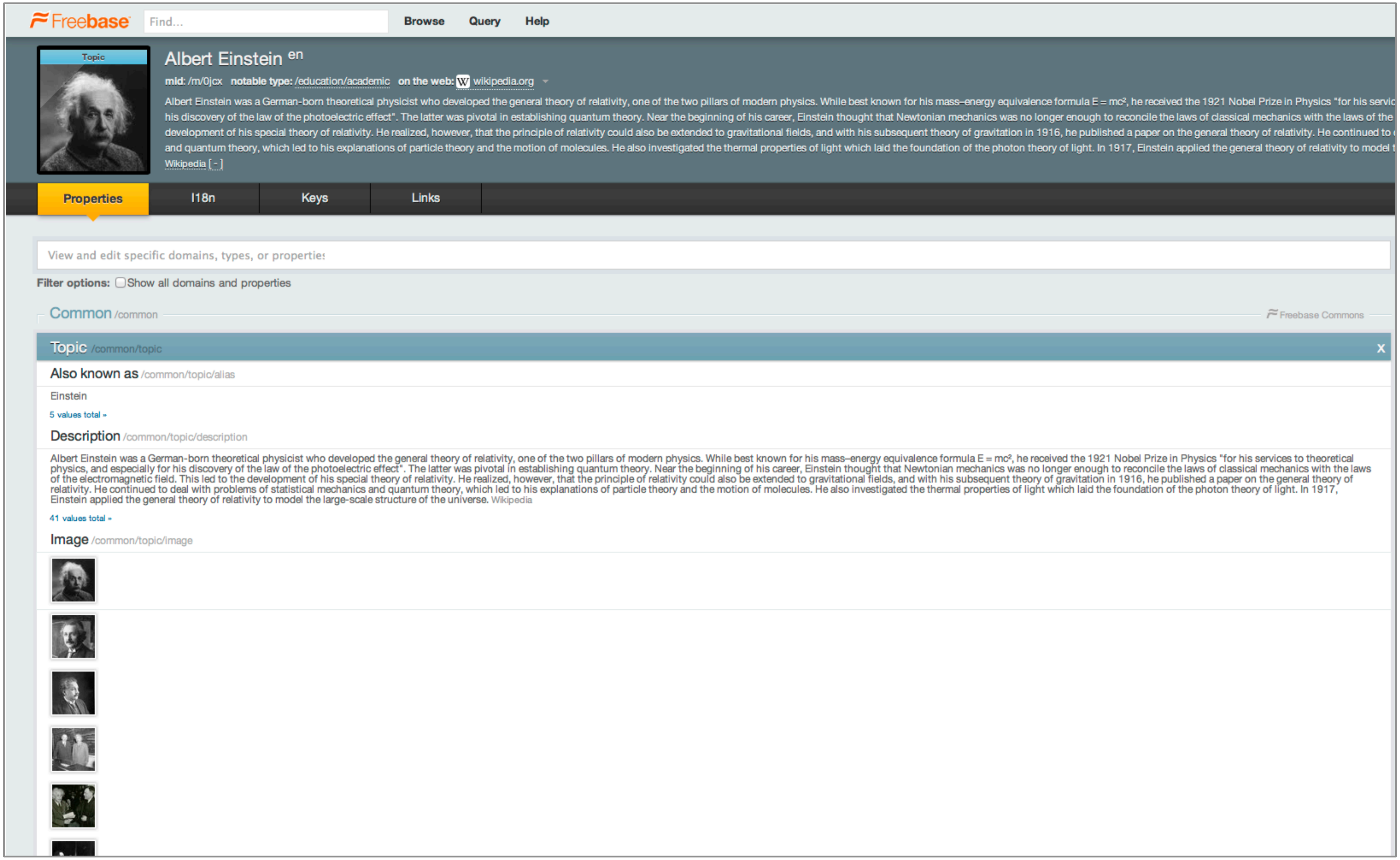
Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

Awards: Nobel Prize in Physics, Copley Medal, Franklin Medal, More

People also search for

http://www.google.com

Freebase



Albert Einstein en

mid: /m/0jcx notable type: /education/academic on the web: wikipedia.org

Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, one of the two pillars of modern physics. While best known for his mass–energy equivalence formula E = mc², he received the 1921 Nobel Prize in Physics "for his servic
his discovery of the law of the photoelectric effect". The latter was pivotal in establishing quantum theory. Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the
development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on the general theory of relativity. He continued to
and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In 1917, Einstein applied the general theory of relativity to model t
Wikipedia [-]

View and edit specific domains, types, or properties

Filter options: □ Show all domains and properties

Common /common

Topic /common/topic

Also known as /common/topic/alias

Einstein

5 values total »

Description /common/topic/description

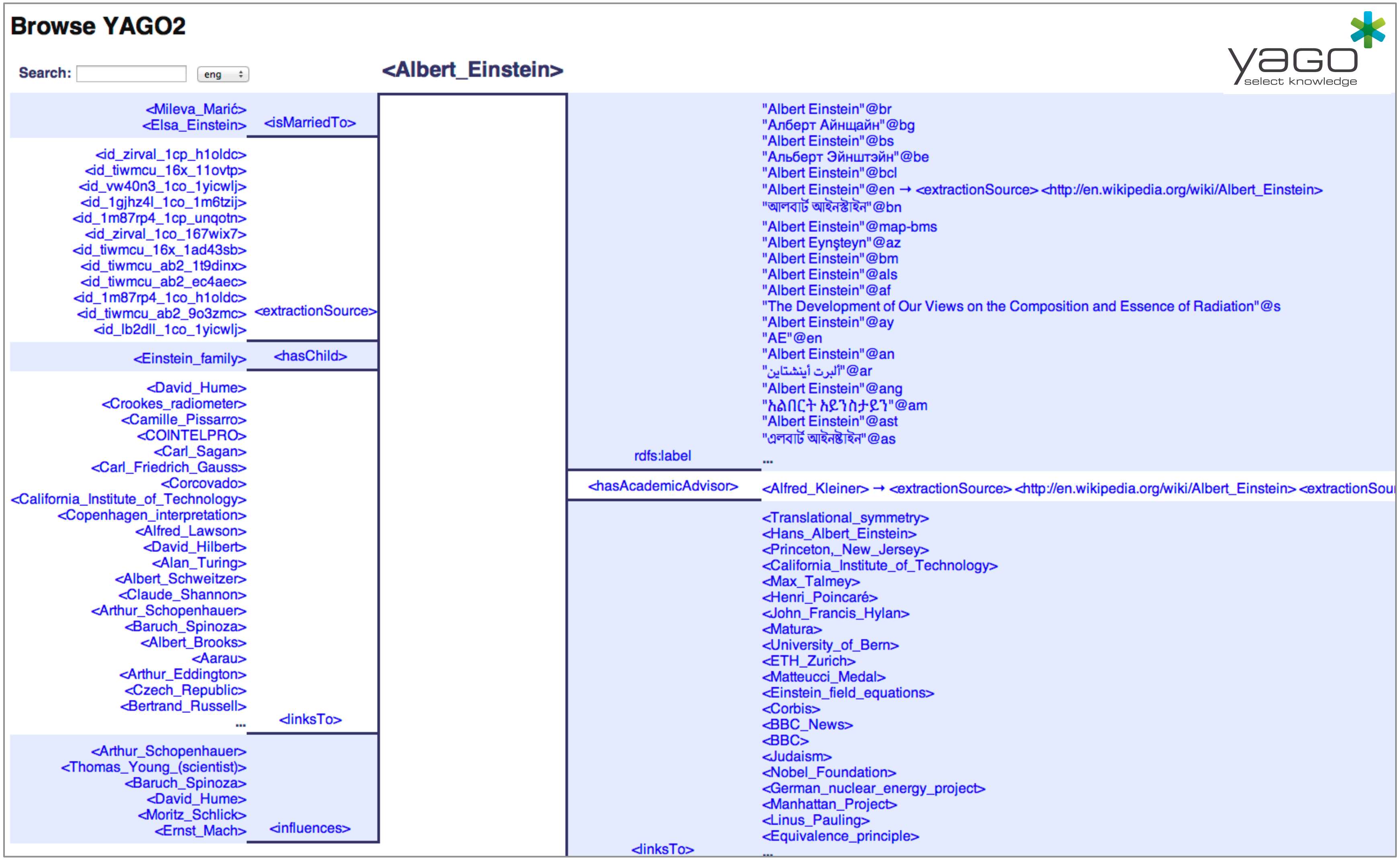
Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, one of the two pillars of modern physics. While best known for his mass–energy equivalence formula E = mc², he received the 1921 Nobel Prize in Physics "for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect". The latter was pivotal in establishing quantum theory. Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the electromagnetic field. This led to the development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on the general theory of relativity. He continued to deal with problems of statistical mechanics and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. Wikipedia

41 values total »

Image /common/topic/image

http://www.freebase.com

YAGO



Browse YAGO2

Search: eng

<Albert_Einstein>

<Mileva_Marić>
<Elsa_Einstein> <isMarriedTo>

<id_zirval_1cp_h1oldc>
<id_tiwmcu_16x_11ovtp>
<id_vw40n3_1co_1yicwlj>
<id_1gjhz4l_1co_1m6tzij>
<id_1m87rp4_1cp_unqotn>
<id_zirval_1co_167wix7>
<id_tiwmcu_16x_1ad43sb>
<id_tiwmcu_ab2_1t9dinx>
<id_tiwmcu_ab2_ec4aec>
<id_1m87rp4_1co_h1oldc>
<id_tiwmcu_ab2_9o3zmc> <extractionSource>
<id_lb2dll_1co_1yicwlj>

<Einstein_family> <hasChild>

<David_Hume>
<Crookes_radiometer>
<Camille_Pissarro>
<COINTELPRO>
<Carl_Sagan>
<Carl_Friedrich_Gauss>
<Corcovado>
<California_Institute_of_Technology>
<Copenhagen_interpretation>
<Alfred_Lawson>
<David_Hilbert>
<Alan_Turing>
<Albert_Schweitzer>
<Claude_Shannon>
<Arthur_Schopenhauer>
<Baruch_Spinoza>
<Albert_Brooks>
<Aarau>
<Arthur_Eddington>
<Czech_Republic>
<Bertrand_Russell>
... <linksTo>

<Arthur_Schopenhauer>
<Thomas_Young_(scientist)>
<Baruch_Spinoza>
<David_Hume>
<Moritz_Schlick>
<Ernst_Mach> <influences>

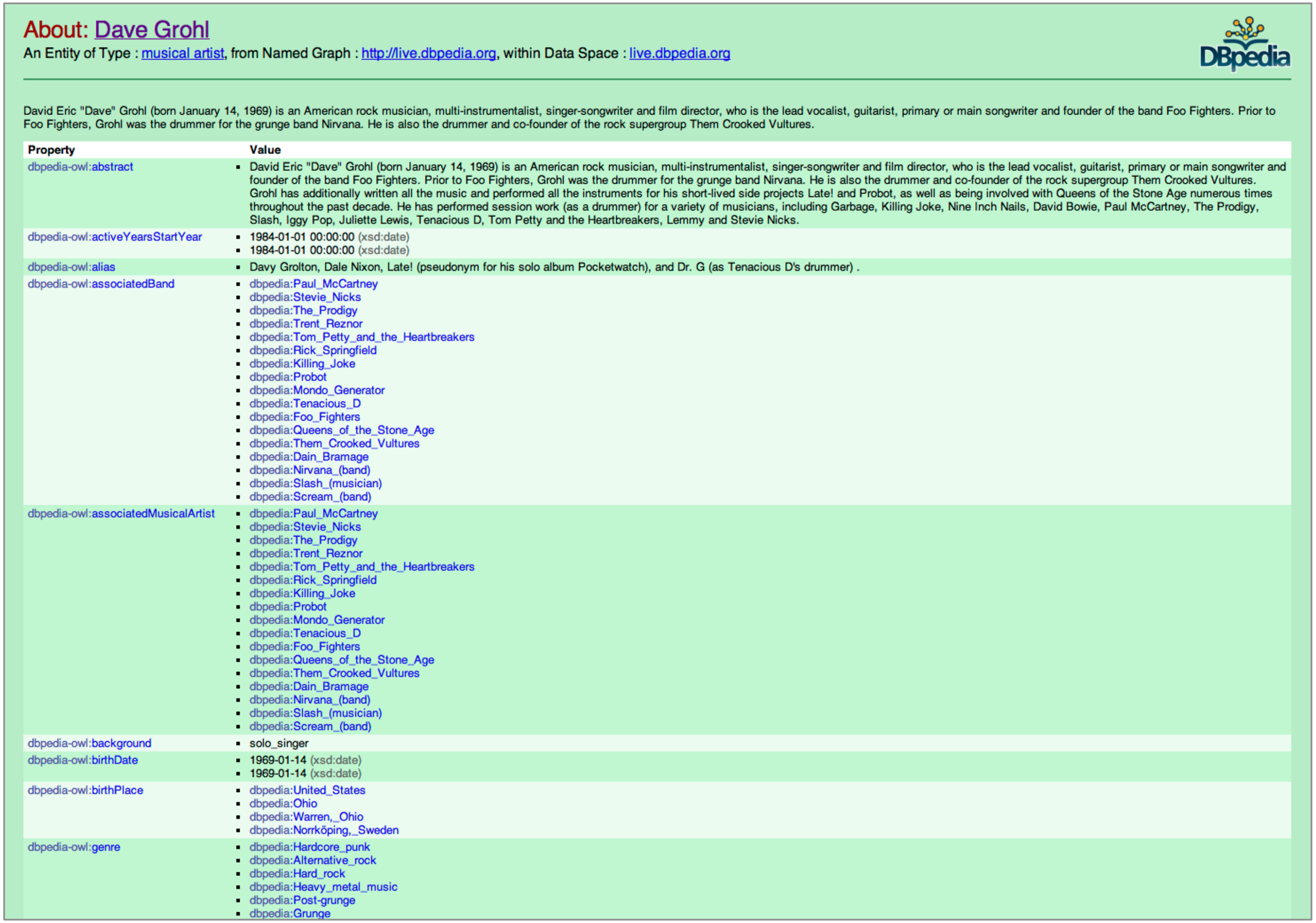
rdfs:label
"Albert Einstein"@br
"Алберт Айнщайн"@bg
"Albert Einstein"@bs
"Альберт Эйнштэйн"@be
"Albert Einstein"@bcl
"Albert Einstein"@en → <extractionSource> <http://en.wikipedia.org/wiki/Albert_Einstein>
"আলবার্ট আইনস্টাইন"@bn
"Albert Einstein"@map-bms
"Albert Eynşteyn"@az
"Albert Einstein"@bm
"Albert Einstein"@als
"Albert Einstein"@af
"The Development of Our Views on the Composition and Essence of Radiation"@s
"Albert Einstein"@ay
"AE"@en
"Albert Einstein"@an
"ألبرت أينشتاين"@ar
"Albert Einstein"@ang
"አልበርት አይንስታይን"@am
"Albert Einstein"@ast
"এলবাৰ্ট আইনষ্টাইন"@as
...

<hasAcademicAdvisor> <Alfred_Kleiner> → <extractionSource> <http://en.wikipedia.org/wiki/Albert_Einstein> <extractionSou

<linksTo>
<Translational_symmetry>
<Hans_Albert_Einstein>
<Princeton,_New_Jersey>
<California_Institute_of_Technology>
<Max_Talmey>
<Henri_Poincaré>
<John_Francis_Hylan>
<Matura>
<University_of_Bern>
<ETH_Zurich>
<Matteucci_Medal>
<Einstein_field_equations>
<Corbis>
<BBC_News>
<BBC>
<Judaism>
<Nobel_Foundation>
<German_nuclear_energy_project>
<Manhattan_Project>
<Linus_Pauling>
<Equivalence_principle>
...

http://www.yago-knowledge.org

DBpedia



About: Dave Grohl

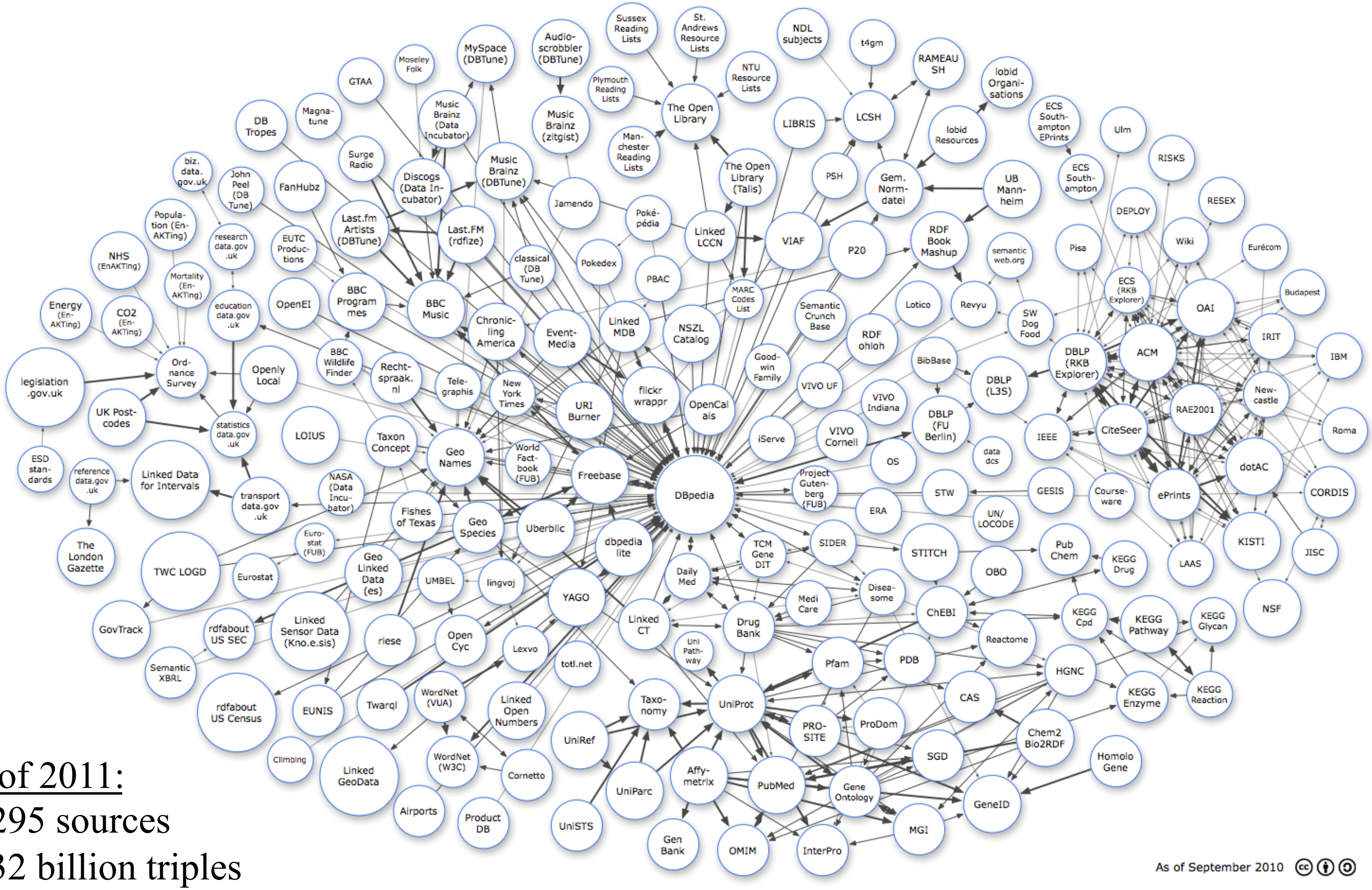
An Entity of Type : musical artist, from Named Graph : http://live.dbpedia.org, within Data Space : live.dbpedia.org

David Eric "Dave" Grohl (born January 14, 1969) is an American rock musician, multi-instrumentalist, singer-songwriter and film director, who is the lead vocalist, guitarist, primary or main songwriter and founder of the band Foo Fighters. Prior to Foo Fighters, Grohl was the drummer for the grunge band Nirvana. He is also the drummer and co-founder of the rock supergroup Them Crooked Vultures.

Property	Value
dbpedia-owl:abstract	▪ David Eric "Dave" Grohl (born January 14, 1969) is an American rock musician, multi-instrumentalist, singer-songwriter and film director, who is the lead vocalist, guitarist, primary or main songwriter and founder of the band Foo Fighters. Prior to Foo Fighters, Grohl was the drummer for the grunge band Nirvana. He is also the drummer and co-founder of the rock supergroup Them Crooked Vultures. Grohl has additionally written all the music and performed all the instruments for his short-lived side projects Late! and Probot, as well as being involved with Queens of the Stone Age numerous times throughout the past decade. He has performed session work (as a drummer) for a variety of musicians, including Garbage, Killing Joke, Nine Inch Nails, David Bowie, Paul McCartney, The Prodigy, Slash, Iggy Pop, Juliette Lewis, Tenacious D, Tom Petty and the Heartbreakers, Lemmy and Stevie Nicks.
dbpedia-owl:activeYearsStartYear	▪ 1984-01-01 00:00:00 (xsd:date) ▪ 1984-01-01 00:00:00 (xsd:date)
dbpedia-owl:alias	▪ Davy Grolton, Dale Nixon, Late! (pseudonym for his solo album Pocketwatch), and Dr. G (as Tenacious D's drummer) .
dbpedia-owl:associatedBand	▪ dbpedia:Paul_McCartney ▪ dbpedia:Stevie_Nicks ▪ dbpedia:The_Prodigy ▪ dbpedia:Trent_Reznor ▪ dbpedia:Tom_Petty_and_the_Heartbreakers ▪ dbpedia:Rick_Springfield ▪ dbpedia:Killing_Joke ▪ dbpedia:Probot ▪ dbpedia:Mondo_Generator ▪ dbpedia:Tenacious_D ▪ dbpedia:Foo_Fighters ▪ dbpedia:Queens_of_the_Stone_Age ▪ dbpedia:Them_Crooked_Vultures ▪ dbpedia:Dain_Bramage ▪ dbpedia:Nirvana_(band) ▪ dbpedia:Slash_(musician) ▪ dbpedia:Scream_(band)
dbpedia-owl:associatedMusicalArtist	▪ dbpedia:Paul_McCartney ▪ dbpedia:Stevie_Nicks ▪ dbpedia:The_Prodigy ▪ dbpedia:Trent_Reznor ▪ dbpedia:Tom_Petty_and_the_Heartbreakers ▪ dbpedia:Rick_Springfield ▪ dbpedia:Killing_Joke ▪ dbpedia:Probot ▪ dbpedia:Mondo_Generator ▪ dbpedia:Tenacious_D ▪ dbpedia:Foo_Fighters ▪ dbpedia:Queens_of_the_Stone_Age ▪ dbpedia:Them_Crooked_Vultures ▪ dbpedia:Dain_Bramage ▪ dbpedia:Nirvana_(band) ▪ dbpedia:Slash_(musician) ▪ dbpedia:Scream_(band)
dbpedia-owl:background	▪ solo_singer
dbpedia-owl:birthDate	▪ 1969-01-14 (xsd:date) ▪ 1969-01-14 (xsd:date)
dbpedia-owl:birthPlace	▪ dbpedia:United_States ▪ dbpedia:Ohio ▪ dbpedia:Warren,_Ohio ▪ dbpedia:Norrköping,_Sweden
dbpedia-owl:genre	▪ dbpedia:Hardcore_punk ▪ dbpedia:Alternative_rock ▪ dbpedia:Hard_rock ▪ dbpedia:Heavy_metal_music ▪ dbpedia:Post-grunge ▪ dbpedia:Grunge

http://dbpedia.org

The Linked Data Project

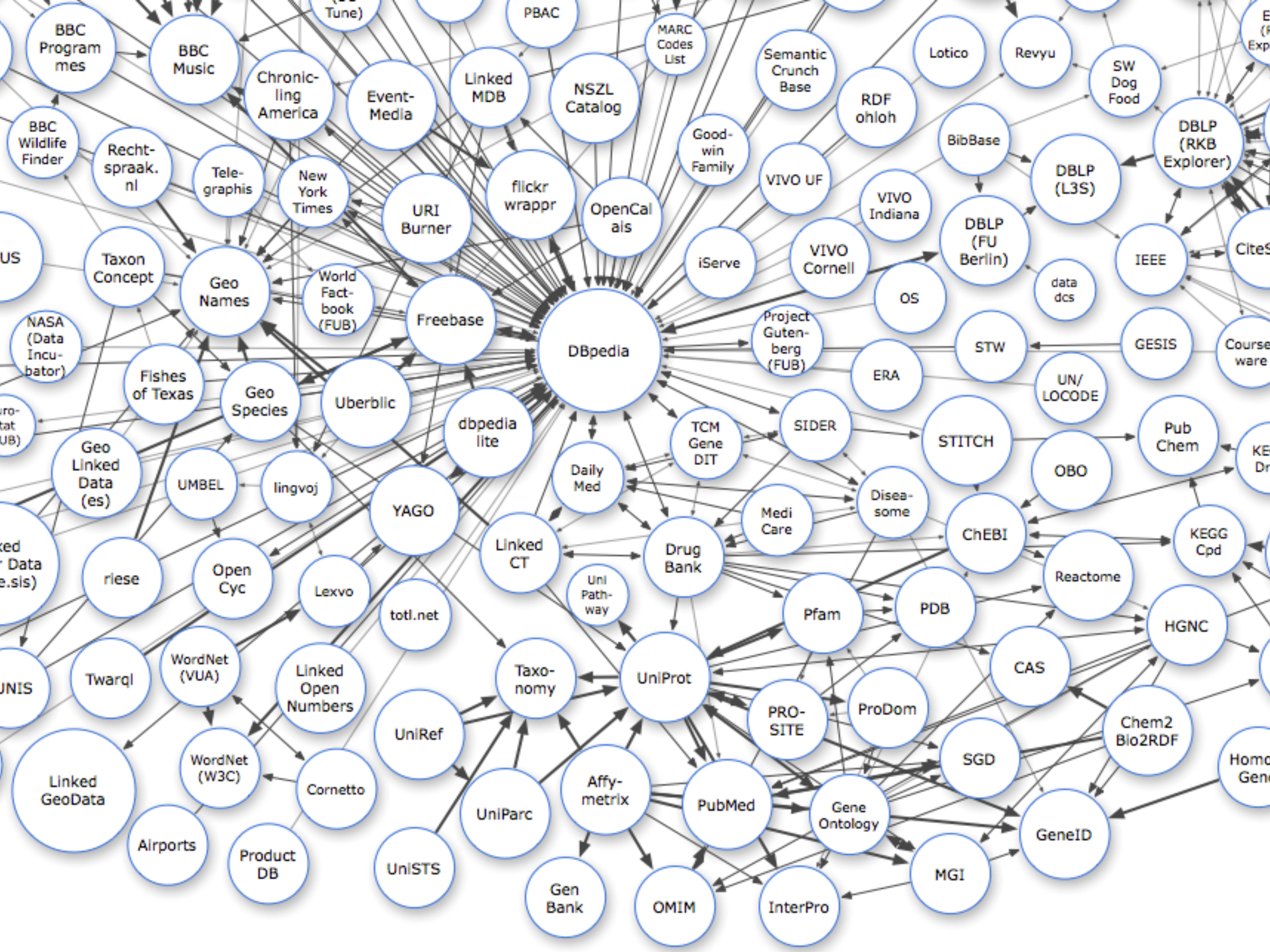


As of September 2010

<u>as of 2011:</u>

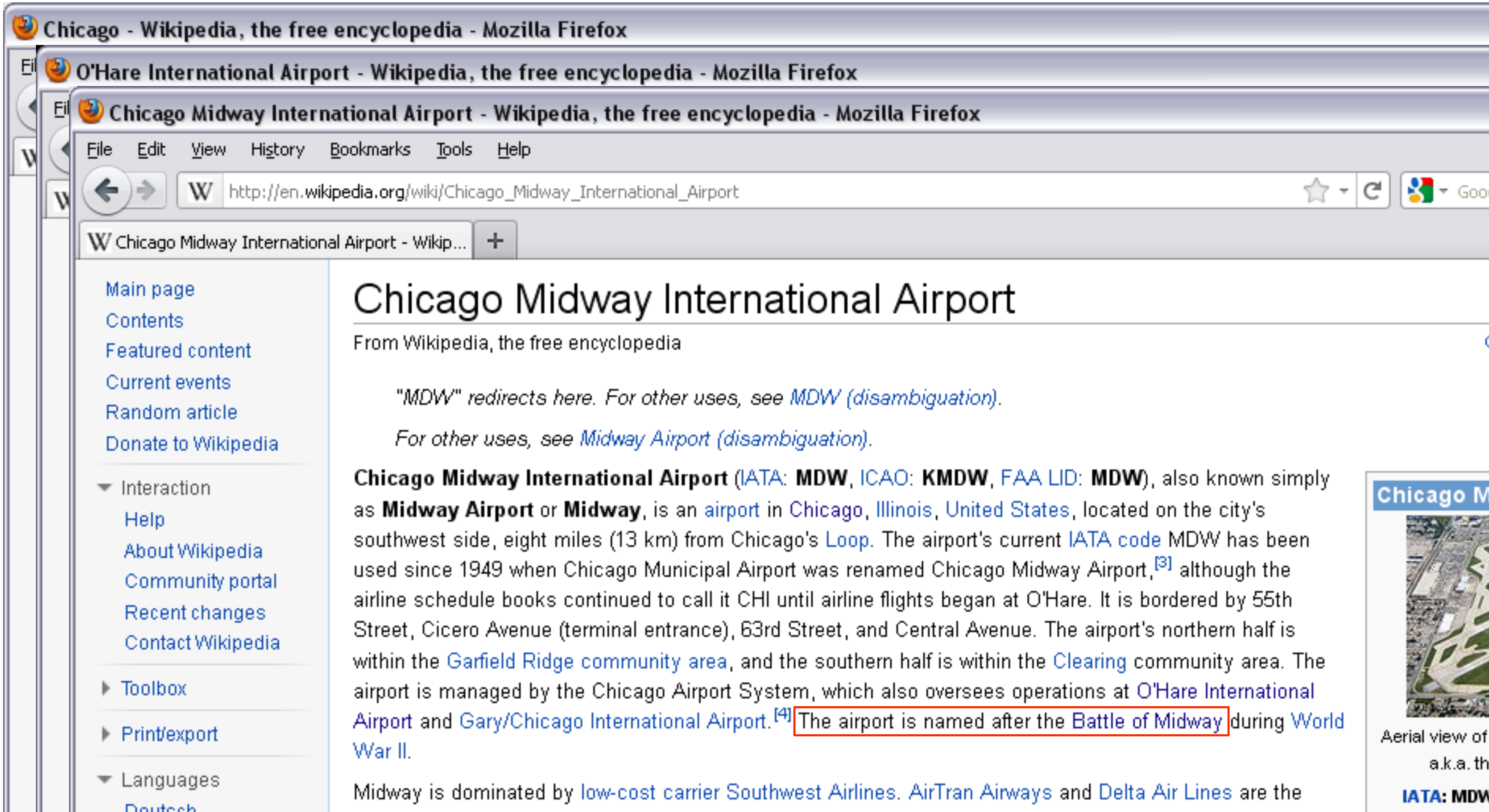
- 295 sources
- 32 billion triples
- 504 million links

http://linkeddata.org


Tune)
PBAC
MARC Codes List
BBC Programmes
BBC Music
Chronicling America
Event-Media
Linked MDB
NSZL Catalog
Semantic Crunch Base
Lotico
Revyu
SW Dog Food
RDF ohloh
Good-win Family
BBC Wildlife Finder
Recht-spraak.nl
Tele-graphis
New York Times
URI Burner
flickr wrappr
OpenCalais
VIVO UF
BibBase
DBLP (RKB Explorer)
DBLP (L3S)
VIVO Indiana
DBLP (FU Berlin)
IEEE
CiteS
US
Taxon Concept
Geo Names
World Fact-book (FUB)
Freebase
DBpedia
iServe
VIVO Cornell
OS
data dcs
NASA (Data Incu-bator)
Project Guten-berg (FUB)
ERA
STW
GESIS
Course ware
UN/LOCODE
Fishes of Texas
Geo Species
Uberblic
dbpedia lite
SIDER
TCM Gene DIT
STITCH
Pub Chem
Geo Linked Data (es)
UMBEL
lingvoj
YAGO
Daily Med
Disea-some
OBO
Medi Care
ChEBI
KEGG Cpd
Linked CT
Drug Bank
riese
Open Cyc
Lexvo
totl.net
Uni Path-way
Reactome
Pfam
PDB
HGNC
CAS
Twarql
WordNet (VUA)
Linked Open Numbers
Taxo-nomy
UniProt
PRO-SITE
ProDom
Chem2 Bio2RDF
UNIS
UniRef
SGD
Homo Gen
WordNet (W3C)
Cornetto
Linked GeoData
Affy-metrix
PubMed
Gene Ontology
GeneID
UniParc
Airports
Product DB
UniSTS
Gen Bank
OMIM
InterPro
MGI

Jeopardy!

A big US city with two airports, one named after a World War II hero, and one named after a World War II battle field?


Chicago - Wikipedia, the free encyclopedia - Mozilla Firefox
O'Hare International Airport - Wikipedia, the free encyclopedia - Mozilla Firefox
Chicago Midway International Airport - Wikipedia, the free encyclopedia - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
http://en.wikipedia.org/wiki/Chicago_Midway_International_Airport
Chicago Midway International Airport - Wikip...
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Toolbox
Print/export
Languages
Chicago M
Aerial view of
a.k.a. th
IATA: MDW

Chicago Midway International Airport

From Wikipedia, the free encyclopedia

"MDW" redirects here. For other uses, see MDW (disambiguation).

For other uses, see Midway Airport (disambiguation).

Chicago Midway International Airport (IATA: **MDW**, ICAO: **KMDW**, FAA LID: **MDW**), also known simply as **Midway Airport** or **Midway**, is an airport in Chicago, Illinois, United States, located on the city's southwest side, eight miles (13 km) from Chicago's Loop. The airport's current IATA code MDW has been used since 1949 when Chicago Municipal Airport was renamed Chicago Midway Airport,[3] although the airline schedule books continued to call it CHI until airline flights began at O'Hare. It is bordered by 55th Street, Cicero Avenue (terminal entrance), 63rd Street, and Central Avenue. The airport's northern half is within the Garfield Ridge community area, and the southern half is within the Clearing community area. The airport is managed by the Chicago Airport System, which also oversees operations at O'Hare International Airport and Gary/Chicago International Airport.[4] The airport is named after the Battle of Midway during World War II.

Midway is dominated by low-cost carrier Southwest Airlines. AirTran Airways and Delta Air Lines are the

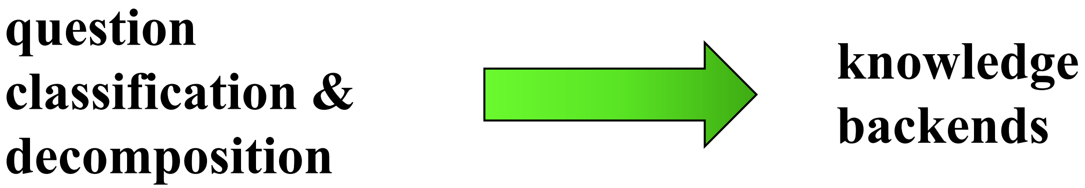
Deep-QA in NL

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel

This town is known as "Sin City" & its downtown is "Glitter Gulch"

As of 2010, this is the only former Yugoslav republic in the EU

99 cents got me a 4-pack of Ytterlig coasters from this Swedish chain


question classification & decomposition → knowledge backends

WIKIPEDIA The Free Encyclopedia

DBpedia

freebase™

yago select knowledge

D. Ferrucci et al.: Building Watson: An Overview of the DeepQA Project. AI Magazine, 2010.

www.ibm.com/innovation/us/watson/index.htm

IRDM Research Literature

Important **conferences** on IR and DM
(see DBLP bibliography for full detail, http://www.dblp.org)
SIGIR, WSDM, ECIR, CIKM, WWW, KDD, ICDM, ICML, ECML

Important **journals** on IR and DM
(see DBLP bibliography for full detail, http://www.dblp.org)
TOIS, TOW, InfRetr, JASIST, InternetMath, TKDD, TODS, VLDBJ

Performance **evaluation/benchmarking** initiatives:

- Text Retrieval Conference (TREC), http://trec.nist.gov
- Cross-Language Evaluation Forum (CLEF), http://www.clef-campaign.org
- Initiative for the Evaluation of XML Retrieval (INEX), http://www.inex.otago.ac.nz/
- KDD Cup, http://www.kdnuggets.com/datasets/kddcup.html & http://www.sigkdd.org/kddcup/index.php