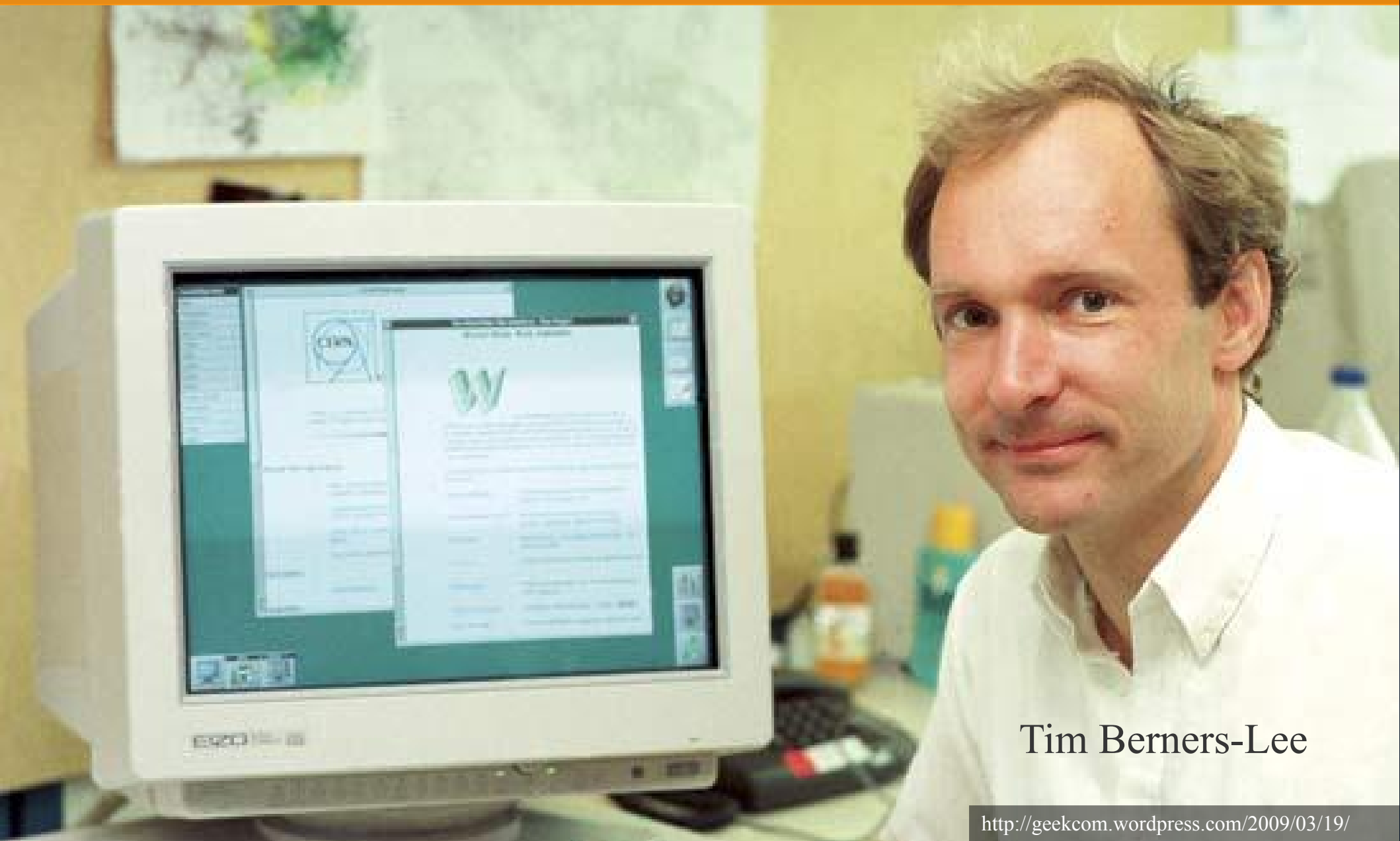
The Future of Knowledge Graphs: Moving beyond Simple Collections of Facts

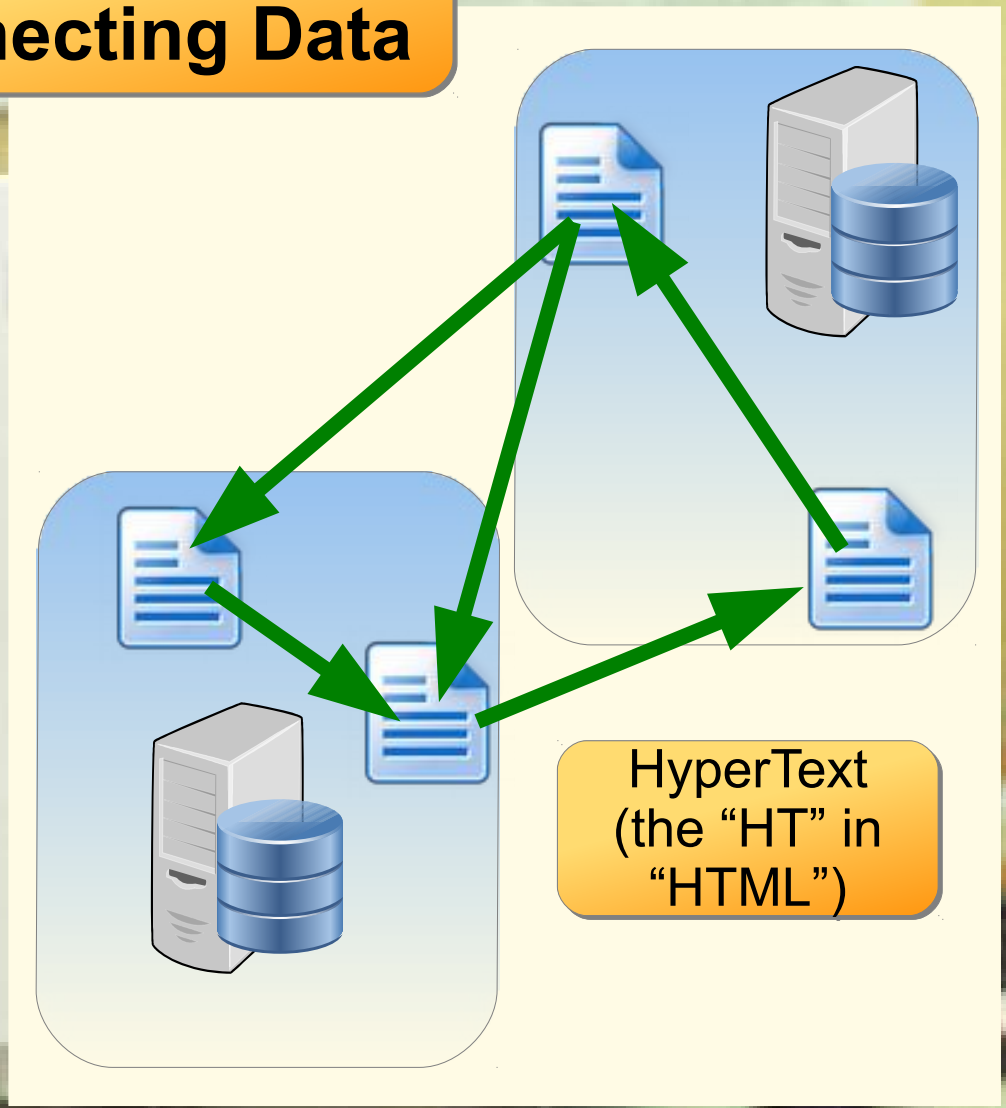
Gerard de Melo
Assistant Professor, Tsinghua University
http://gerard.demelo.org

25 Years of the World Wide Web: 1989–2014



25 Years of the World Wide Web: 1989–2014

Basic Idea: Connecting Data



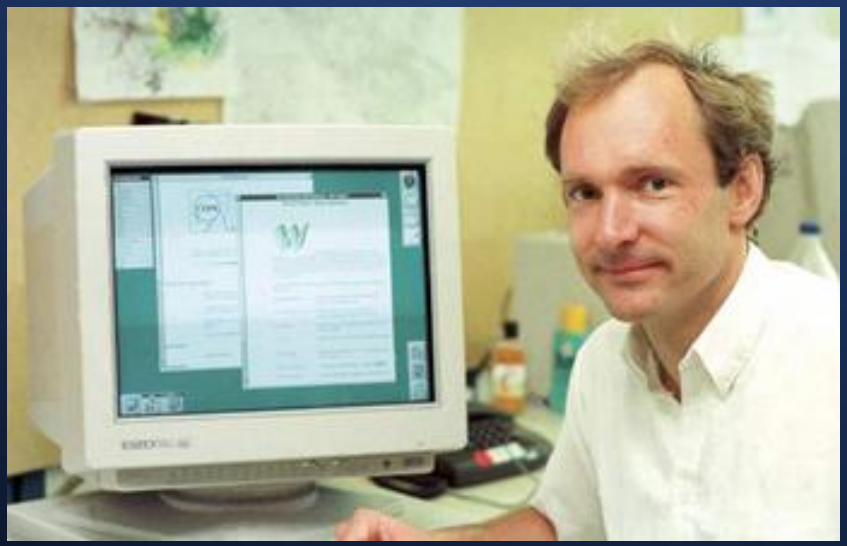
HyperText (the “HT” in “HTML”)

Tim Berners-Lee

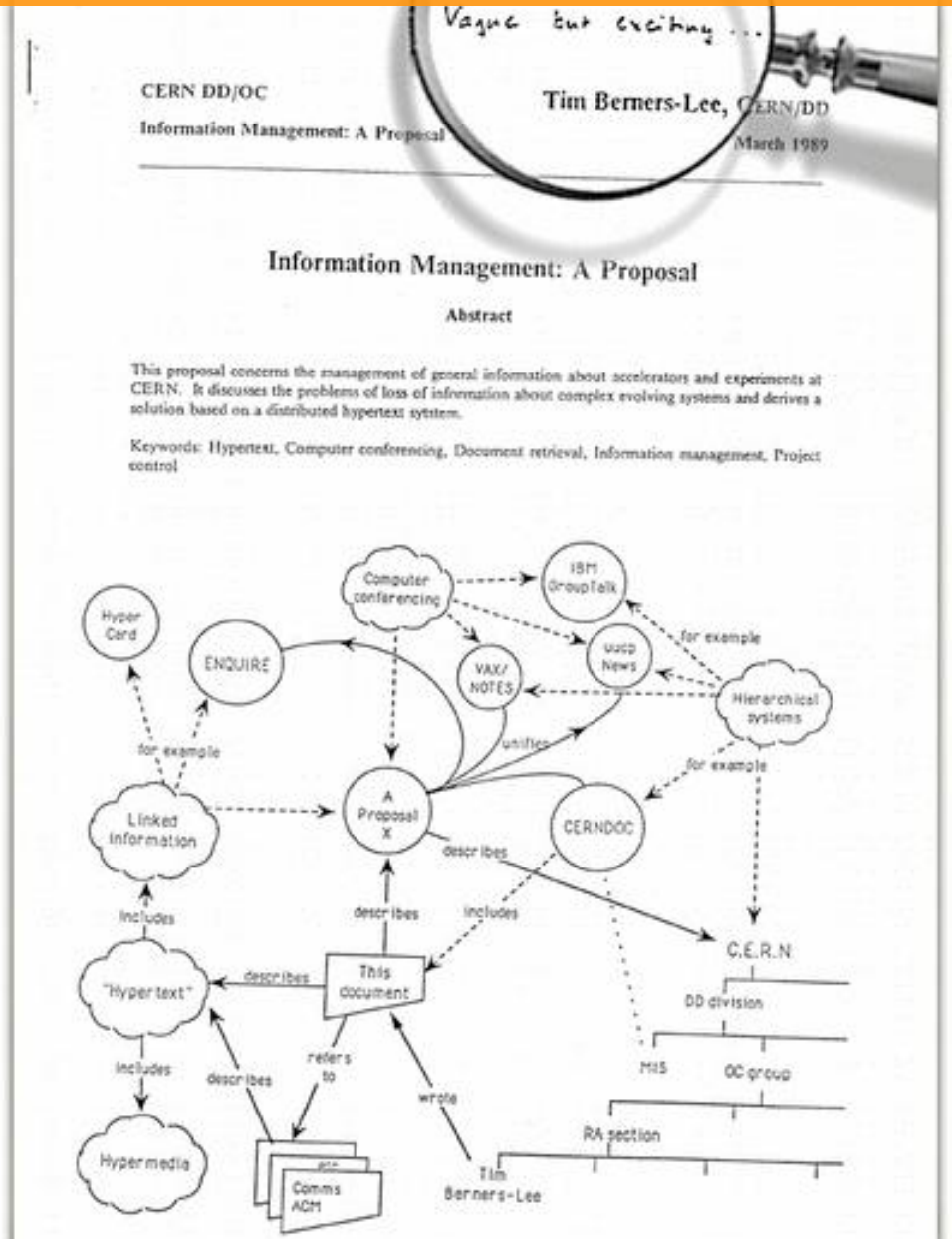
http://geekcom.wordpress.com/2009/03/19/

25 Years of the World Wide Web: 1989–2014

Tim Berners-Lee



http://geekcom.wordpress.com/2009/03/19/



Vague but exciting ...

CERN DD/OC

Information Management: A Proposal

Tim Berners-Lee, CERN/DD

March 1989

Information Management: A Proposal

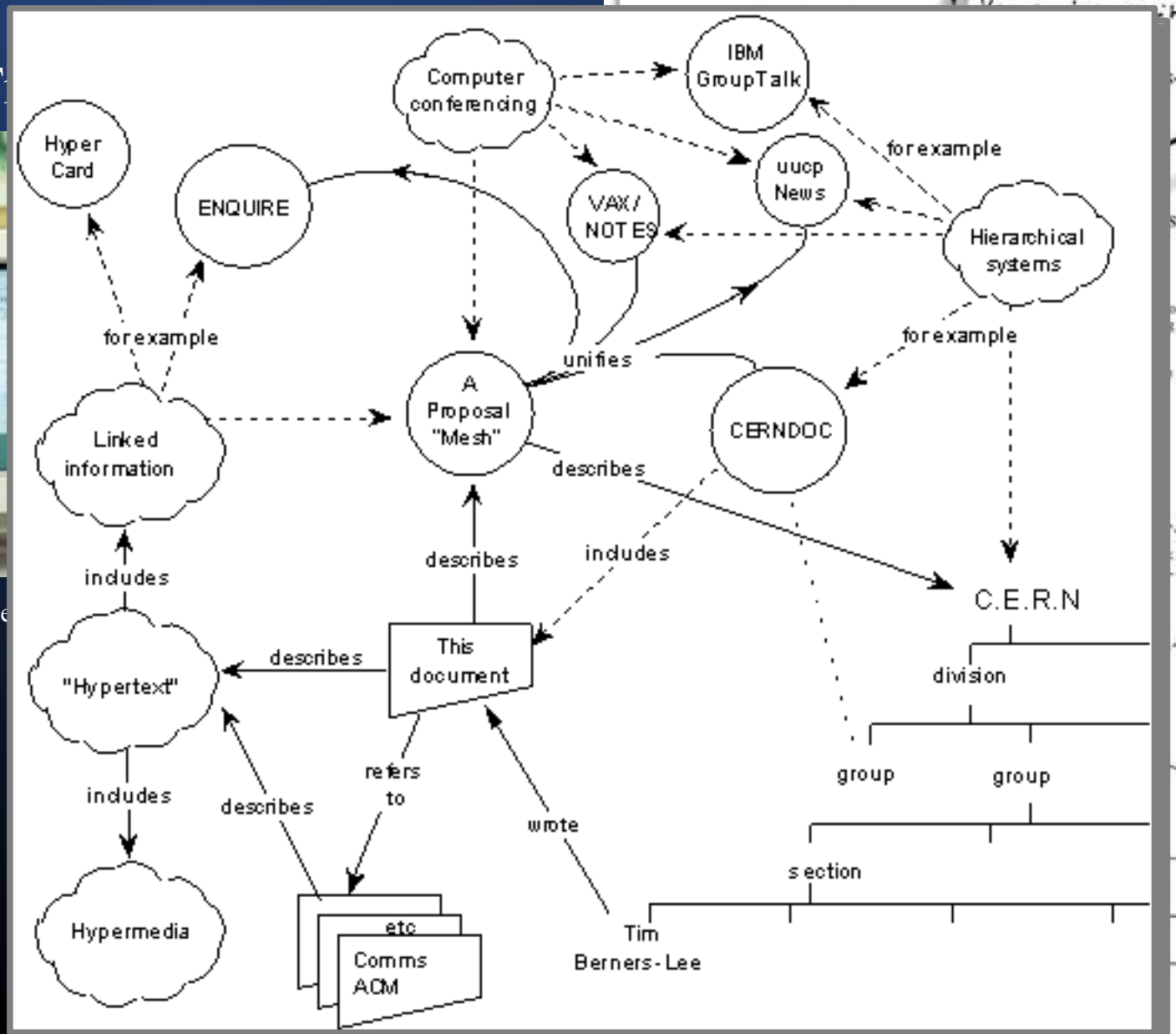
Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

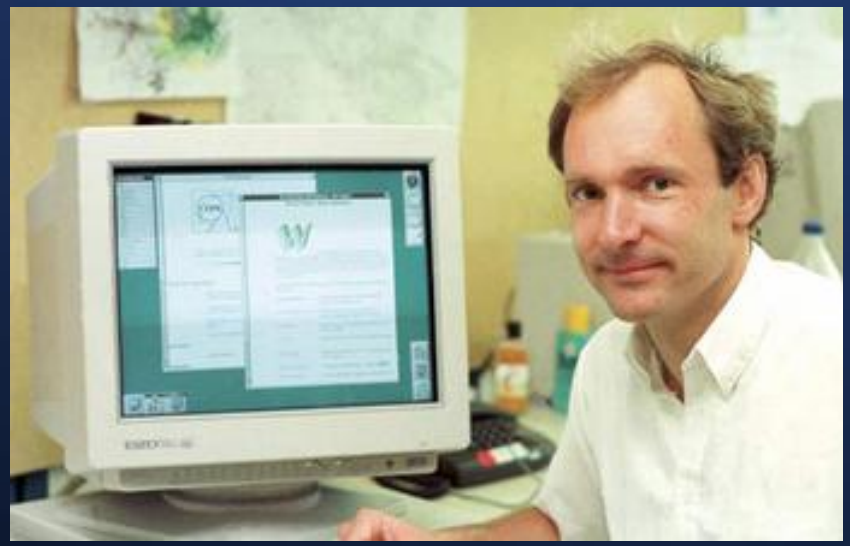
25 Years of the World Wide Web: 1989–2014

http://ge

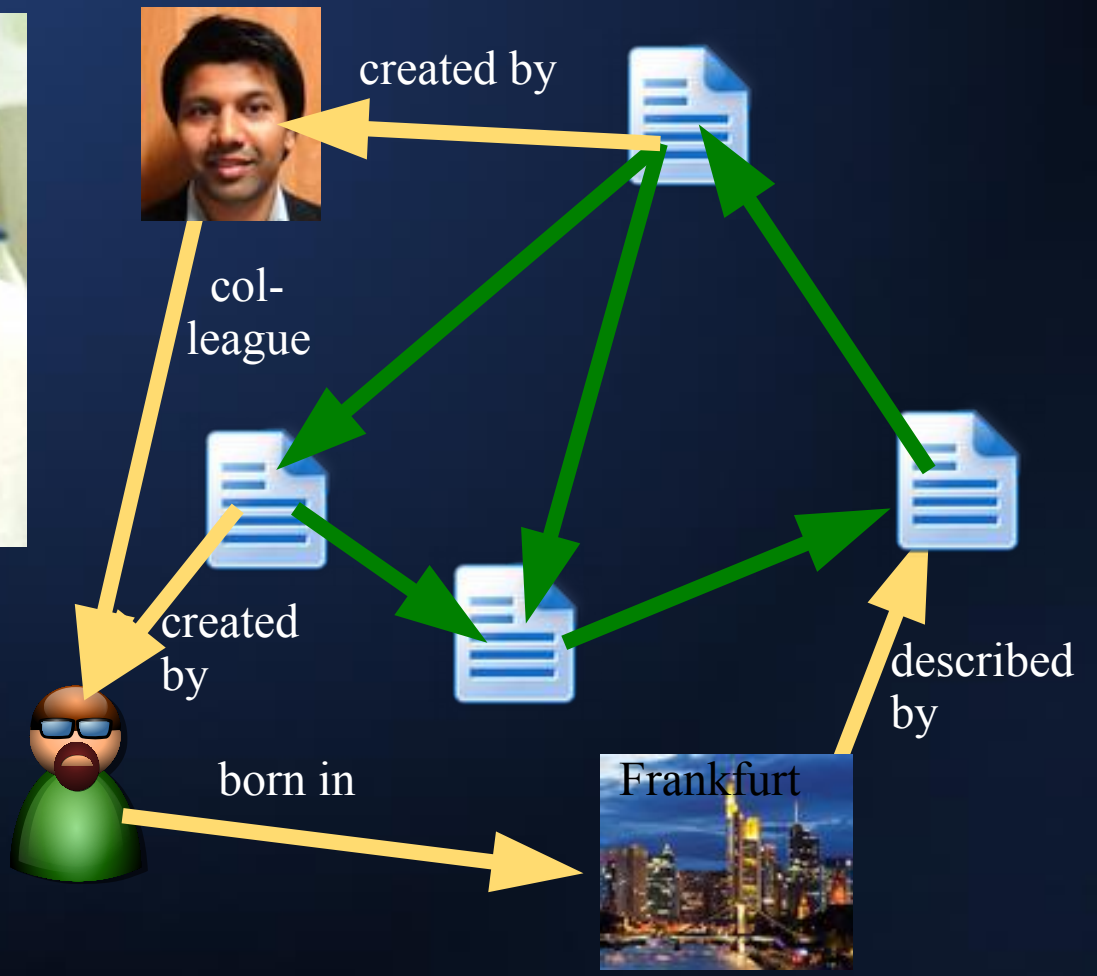


The Semantic Web

Tim Berners-Lee



http://geekcom.wordpress.com/2009/03/19/

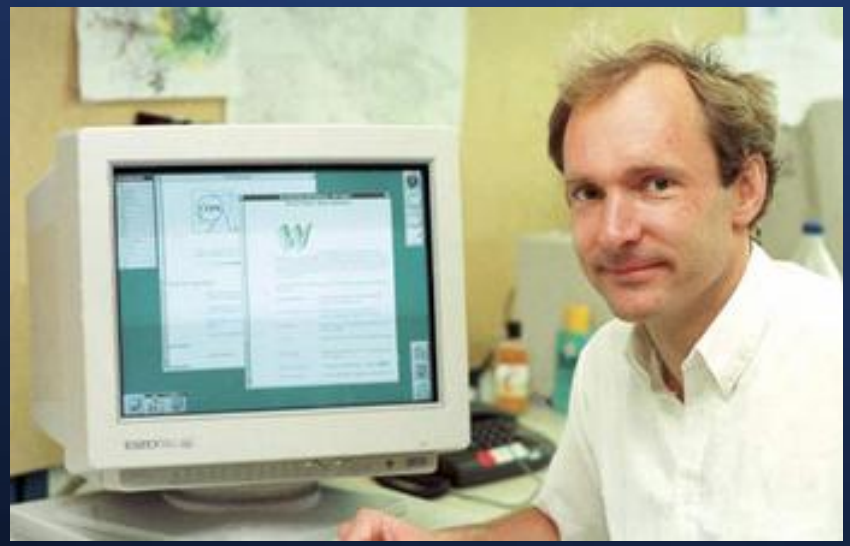


A Web of machine-readable entity-relationship data

Frankfurt image: https://en.wikipedia.org/wiki/File:Frankfurt_Am_Main-Stadtansicht_von_der_Deutschherrnbruecke_am_fruehen_Abend-20110808.jpg

The Semantic Web

Tim Berners-Lee



http://geekcom.wordpress.com/2009/03/19/

A Web of machine-readable entity-relationship data

Problem:
Just an academic idea, not used in practice!

Frankfurt image: https://en.wikipedia.org/wiki/File:Frankfurt_Am_Main-Stadtansicht_von_der_Deutschherrnbruecke_am_fruehen_Abend-20110808.jpg

Big Knowledge Collections

Create account Not logged in Talk Contributions Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read Edit View history Search

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

Languages
Afrikaans
Alemannisch

Max Planck

From Wikipedia, the free encyclopedia

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

Max Karl Ernst Ludwig Planck, FRS[2] (/plɑːŋk/;[3] German: [plaŋk]; April 23, 1858 – October 4, 1947) was a German theoretical physicist who originated quantum theory, which won him the Nobel Prize in Physics in 1918.[4]

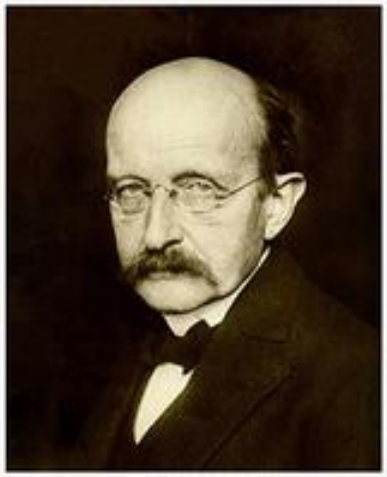
Planck made many contributions to theoretical physics, but his fame as a physicist rests primarily on his role as an originator of the quantum theory. However, his name is also known on a broader academic basis, through the renaming in 1948 of the German scientific institution, the Kaiser Wilhelm Society (of which he was twice president), as the Max Planck Society (MPS). The MPS now includes 83 institutions representing a wide range of scientific directions.

Max Planck's quantum theory revolutionized human understanding of atomic and subatomic processes.

Contents [hide]

1 Early life and career
1.1 Academic career
1.2 Family
1.3 Professor at Berlin University
1.4 Black-body radiation
1.5 Einstein and the theory of relativity
1.6 World War I
1.7 Post War and Weimar Republic
1.8 Quantum mechanics
1.9 Nazi dictatorship and World War II
2 Religious views
3 Publications
4 See also
5 References
6 Sources
7 External links

Max Planck



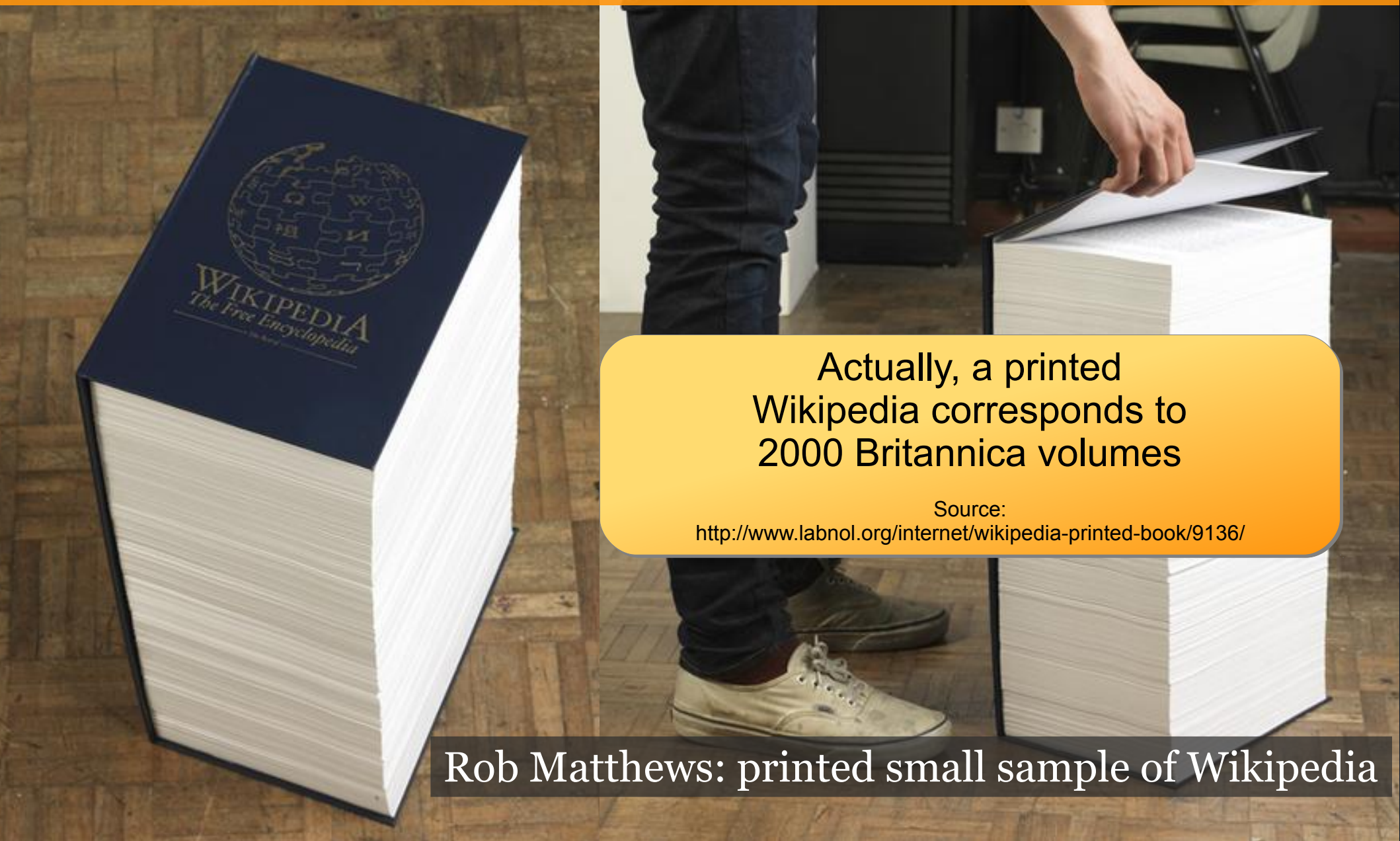
Planck in 1933

Born	Max Karl Ernst Ludwig Planck April 23, 1858 Kiel, Duchy of Holstein
Died	October 4, 1947 (aged 89) Göttingen, Lower Saxony, Germany
Nationality	German
Fields	Physics
Institutions	University of Kiel University of Berlin

Big Knowledge Collections

https://commons.wikimedia.org/wiki/File:Encyclopedia_Britannica_in_the_library_of_The_Kings_School,_Goa.jpg

Big Knowledge Collections

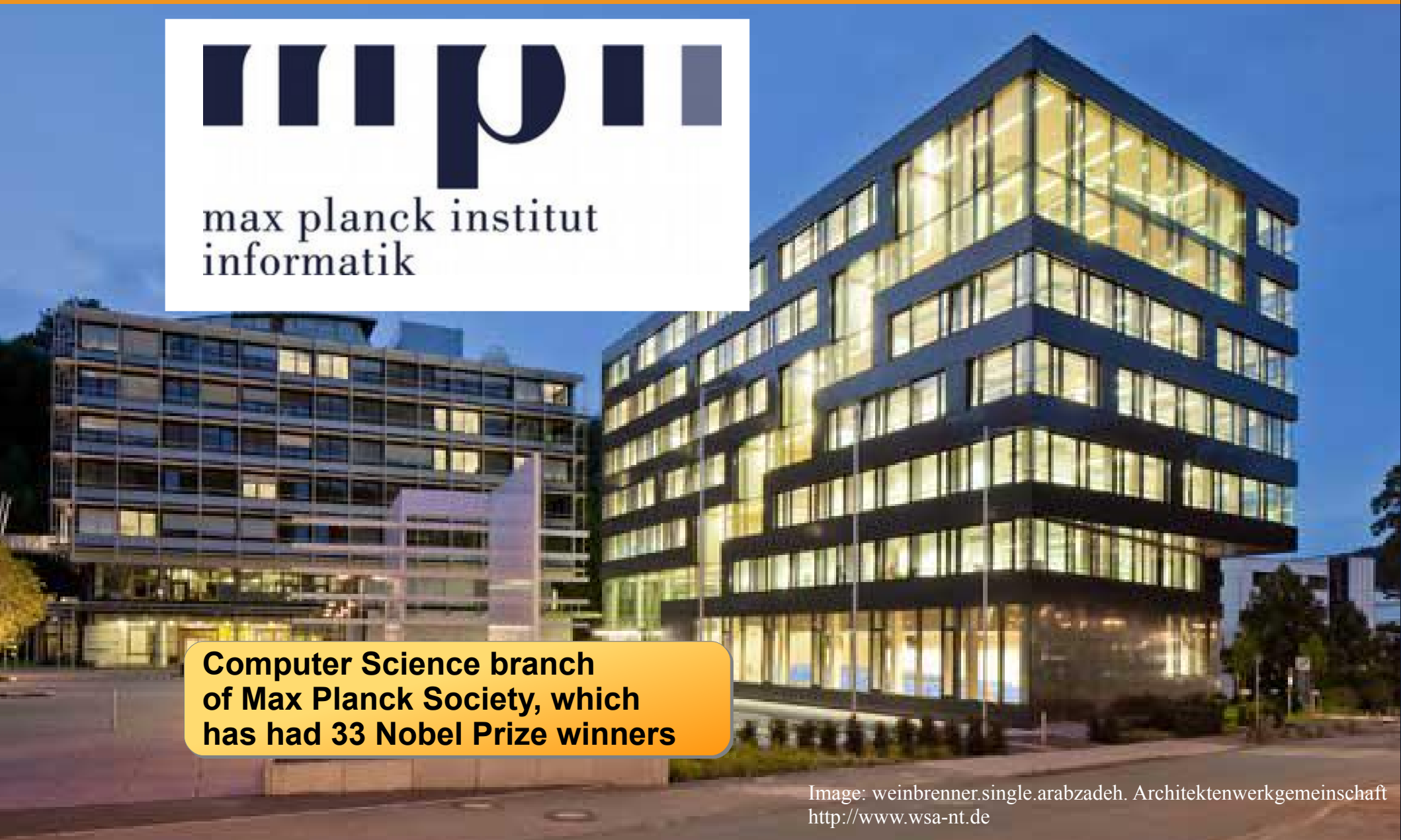


Actually, a printed
Wikipedia corresponds to
2000 Britannica volumes

Source:
http://www.labnol.org/internet/wikipedia-printed-book/9136/

Rob Matthews: printed small sample of Wikipedia

Background: 2006/7



Computer Science branch of Max Planck Society, which has had 33 Nobel Prize winners

Image: weinbrenner.single.arabzadeh. Architektenwerkgemeinschaft
http://www.wsa-nt.de

Big Knowledge Collections

Create account Not logged in Talk Contributions Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read Edit View history Search

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

Languages
Afrikaans
Alemannisch

Max Planck

From Wikipedia, the free encyclopedia

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

Max Karl Ernst Ludwig Planck, FRS[2] (/plɑːŋk/;[3] German: [plaŋk]; April 23, 1858 – October 4, 1947) was a German theoretical physicist who originated quantum theory, which won him the Nobel Prize in Physics in 1918.[4]

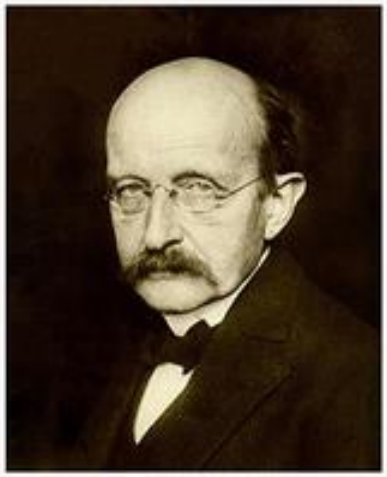
Planck made many contributions to theoretical physics, but his fame as a physicist rests primarily on his role as an originator of the quantum theory. However, his name is also known on a broader academic basis, through the renaming in 1948 of the German scientific institution, the Kaiser Wilhelm Society (of which he was twice president), as the Max Planck Society (MPS). The MPS now includes 83 institutions representing a wide range of scientific directions.

Max Planck's quantum theory revolutionized human understanding of atomic and subatomic processes.

Contents [hide]

1 Early life and career
1.1 Academic career
1.2 Family
1.3 Professor at Berlin University
1.4 Black-body radiation
1.5 Einstein and the theory of relativity
1.6 World War I
1.7 Post War and Weimar Republic
1.8 Quantum mechanics
1.9 Nazi dictatorship and World War II
2 Religious views
3 Publications
4 See also
5 References
6 Sources
7 External links

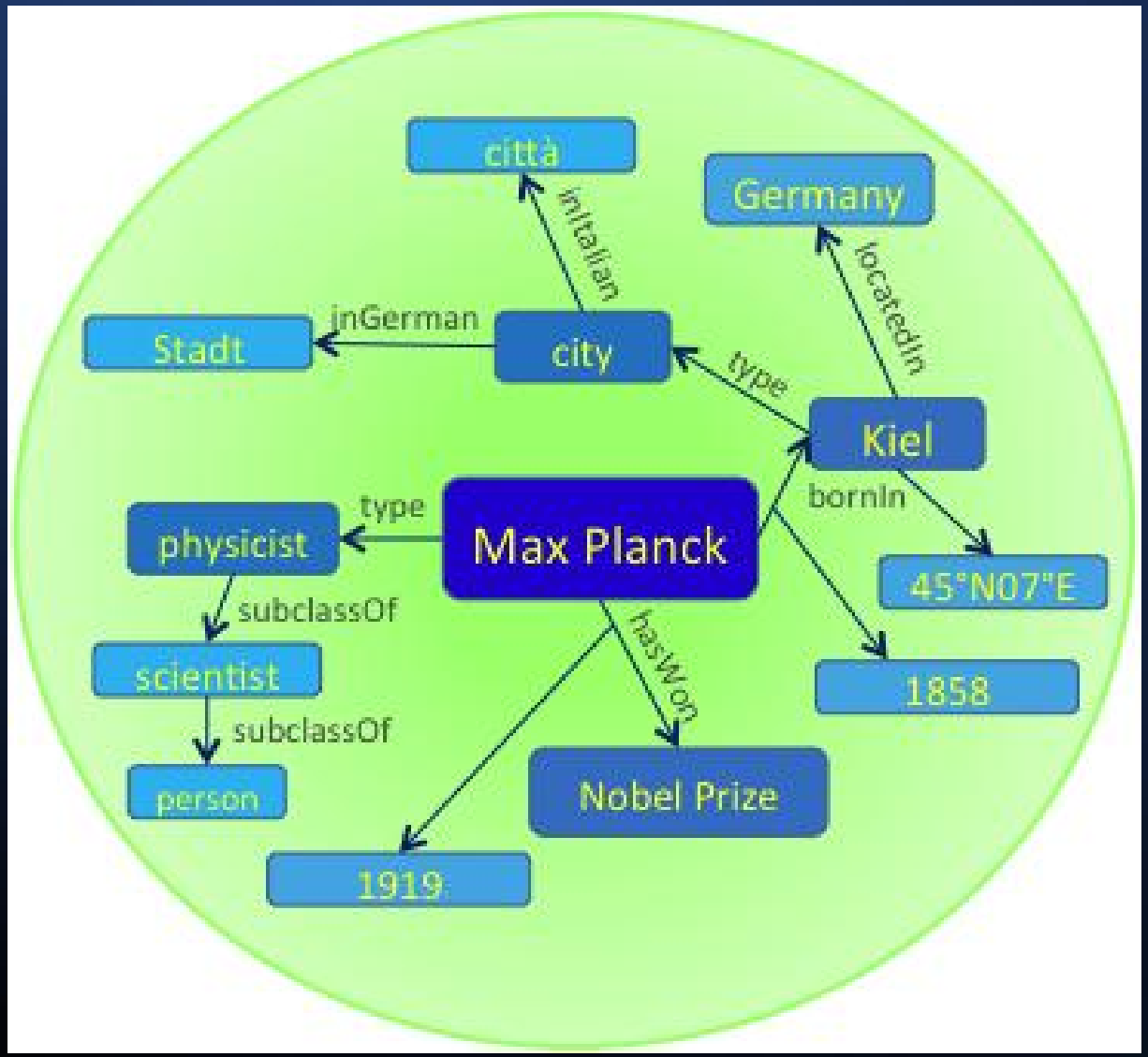
Max Planck



Planck in 1933

Born	Max Karl Ernst Ludwig Planck April 23, 1858 Kiel, Duchy of Holstein
Died	October 4, 1947 (aged 89) Göttingen, Lower Saxony, Germany
Nationality	German
Fields	Physics
Institutions	University of Kiel University of Berlin

Big Knowledge Graphs



YAGO 1: 2007

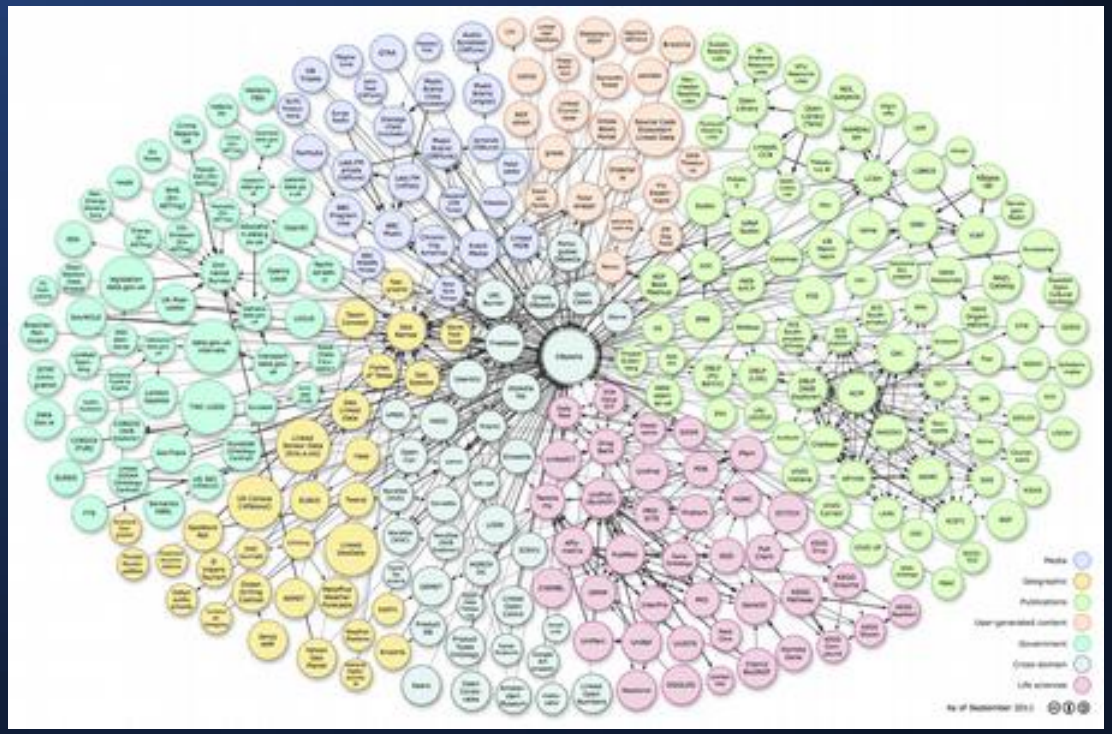
YAGO2.
Hoffart et al.
WWW 2011.

Big Knowledge Graphs

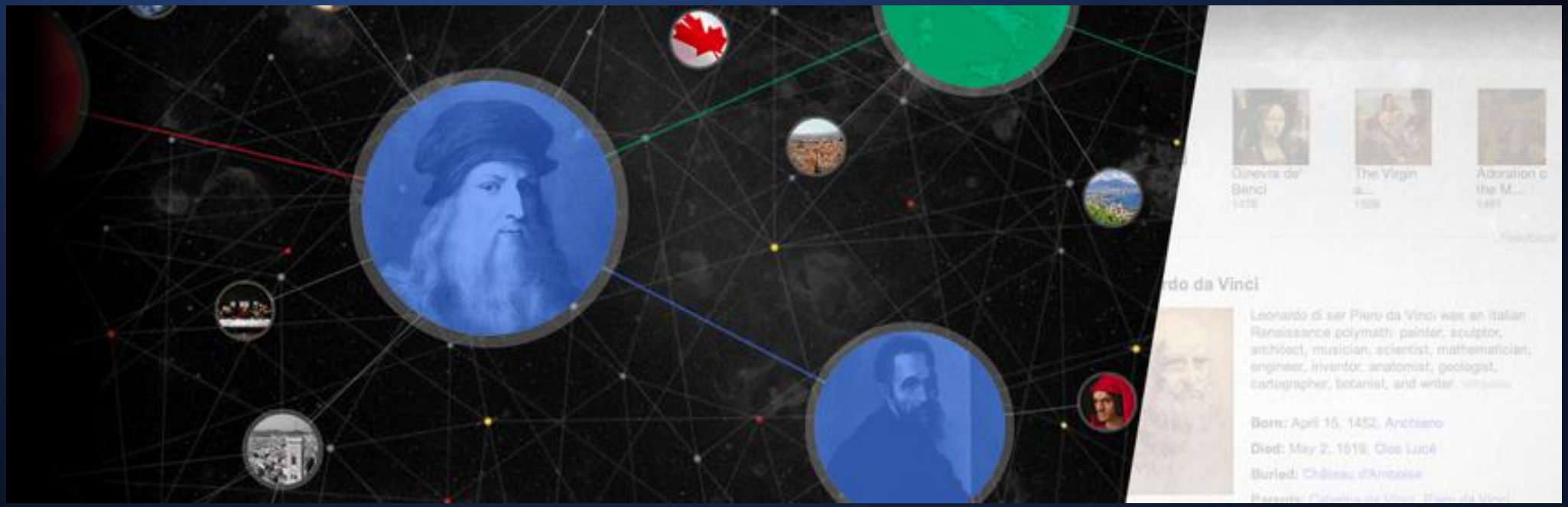
yago
select knowledge

DBpedia

Freebase

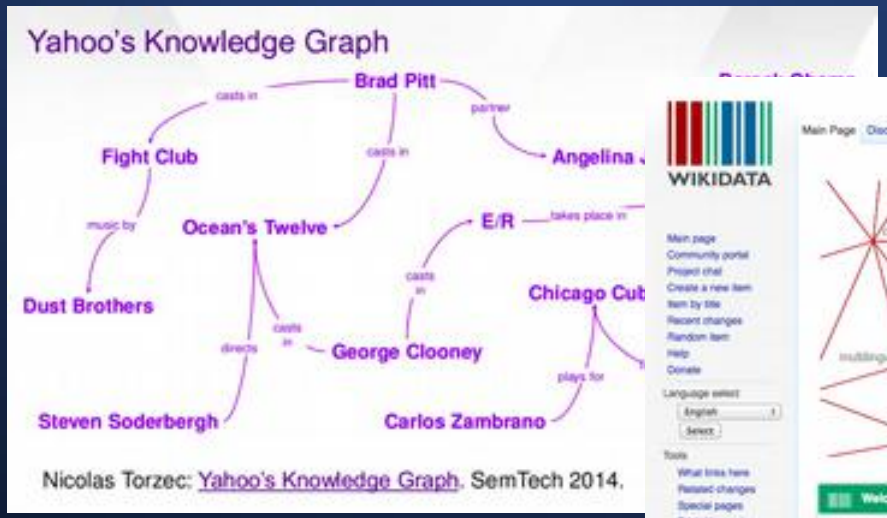


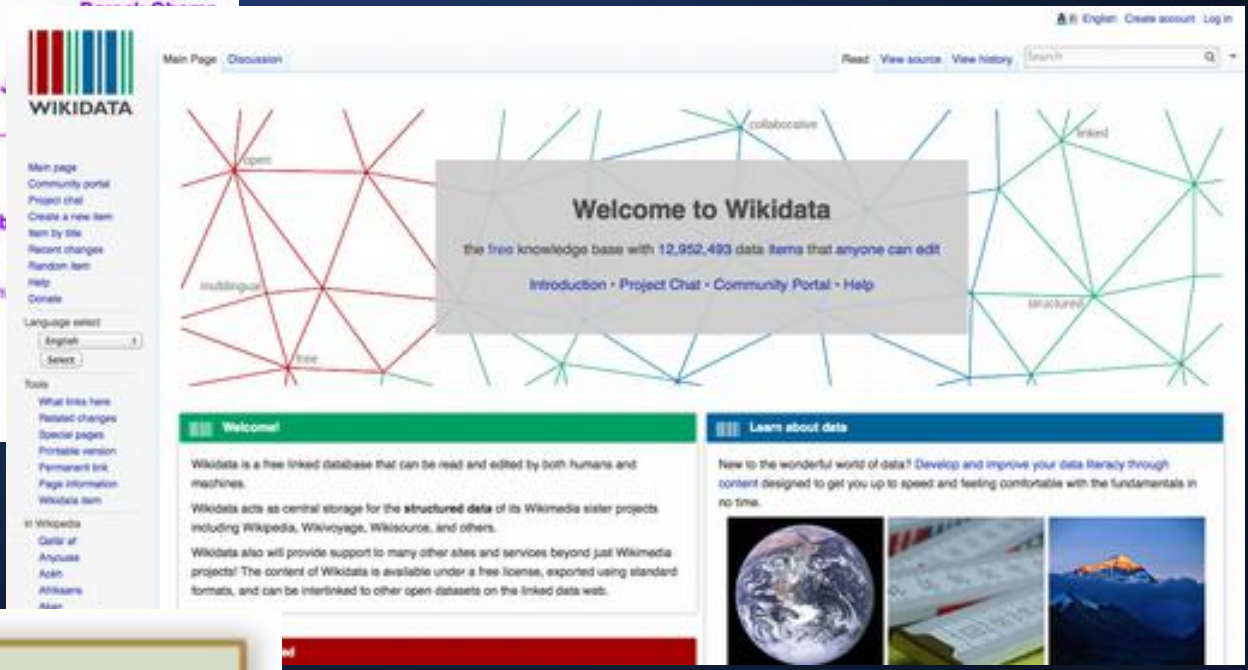
Google Knowledge Graph



Gerard de Melo

Mainstream Adoption





Mainstream Adoption

Mainstream Adoption

Google

capital of new jersey

Web Maps Images Shopping News More Search tools

About 216,000,000 results (0.27 seconds)

Trenton

New Jersey, Capital

Trenton is the capital city of the State of New Jersey and the county seat of Mercer County. The city is officially considered to be within the Greater New York City Combined Statistical Area. Wikipedia

Area: 8.155 sq miles (21.12 km²)

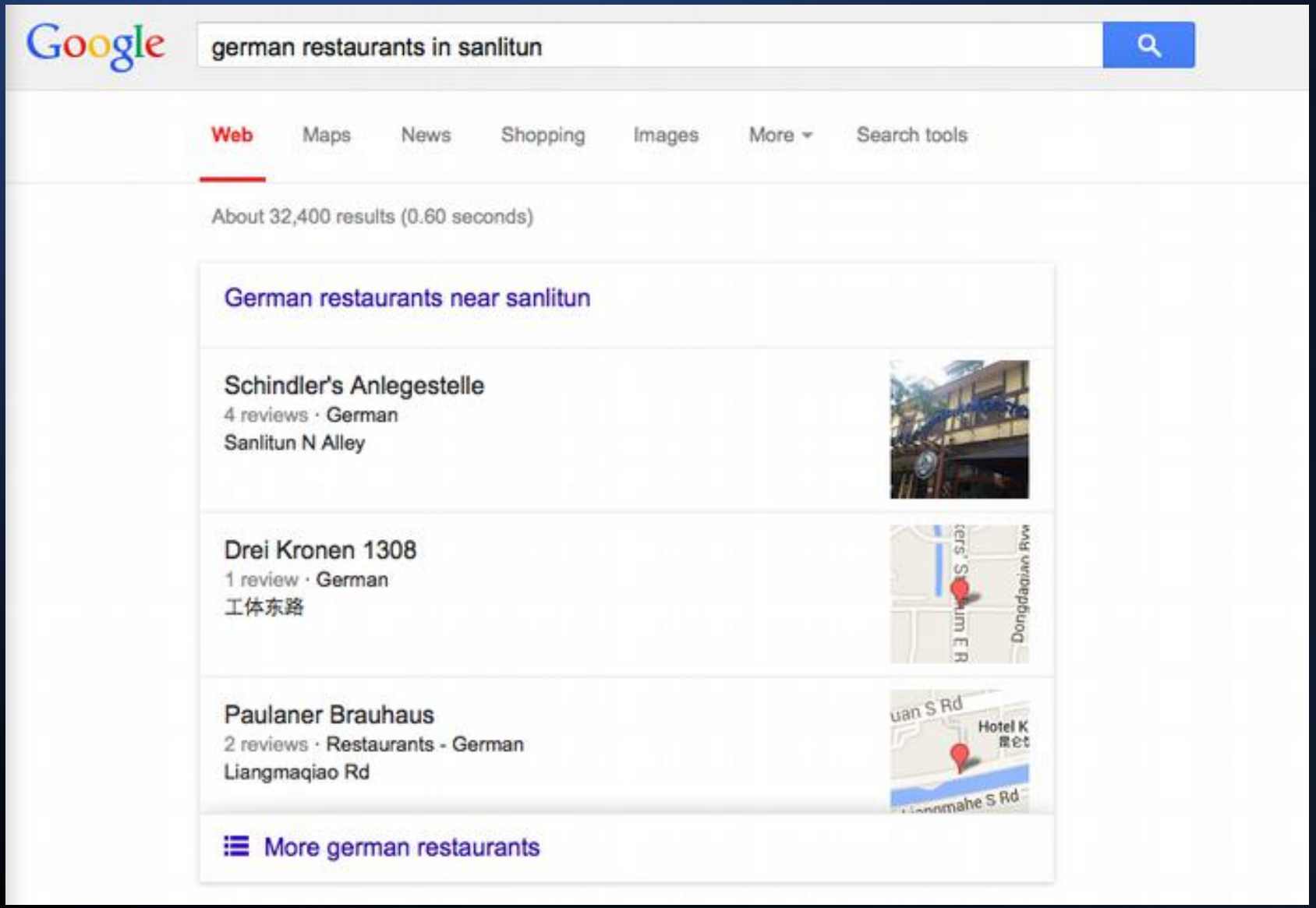
Weather: 65°F (18°C), Wind NE at 13 mph (21 km/h), 56% Humidity

Local time: Tuesday 10:23 AM

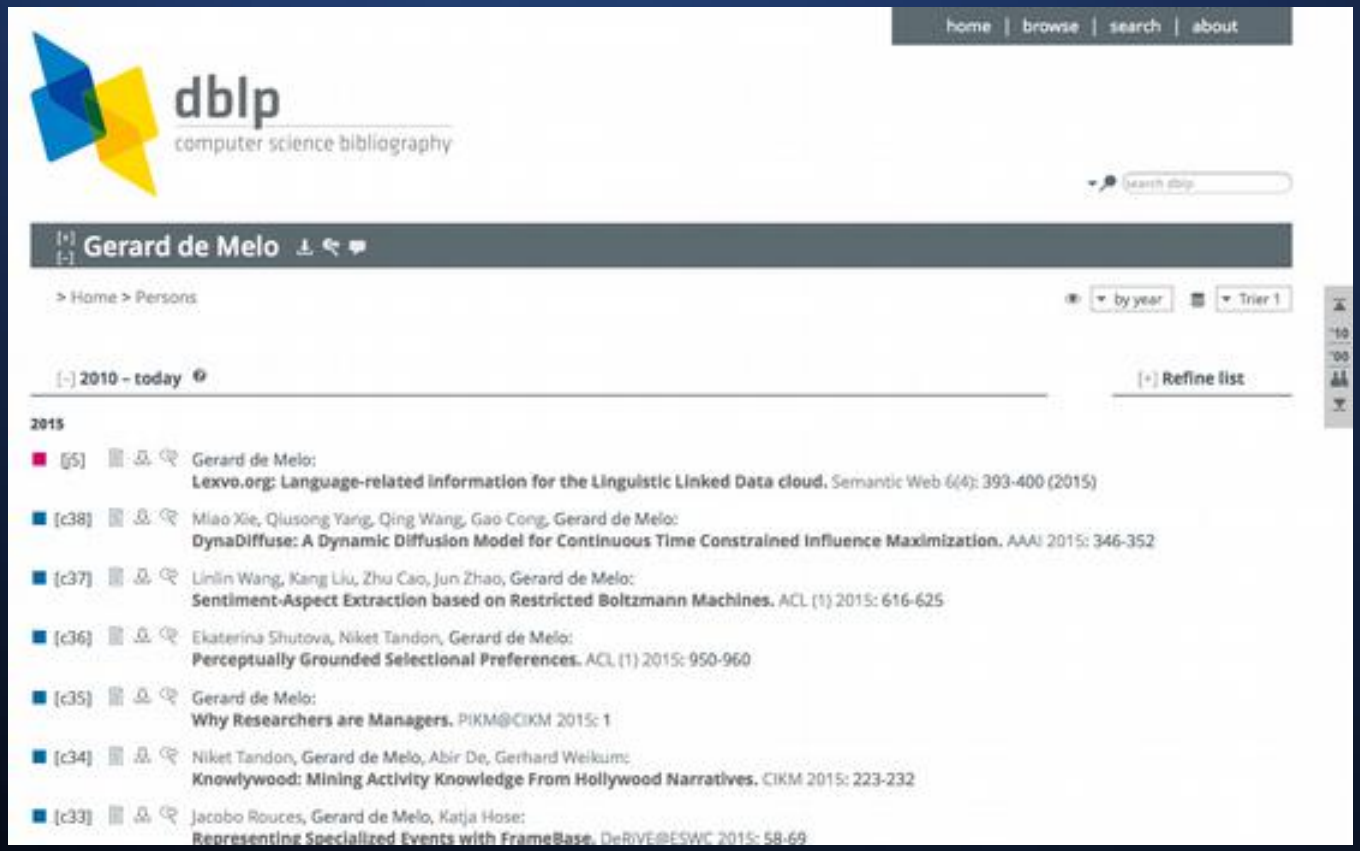
Population: 84,477 (2012)

Unemployment rate: 11.6% (Aug 2013)

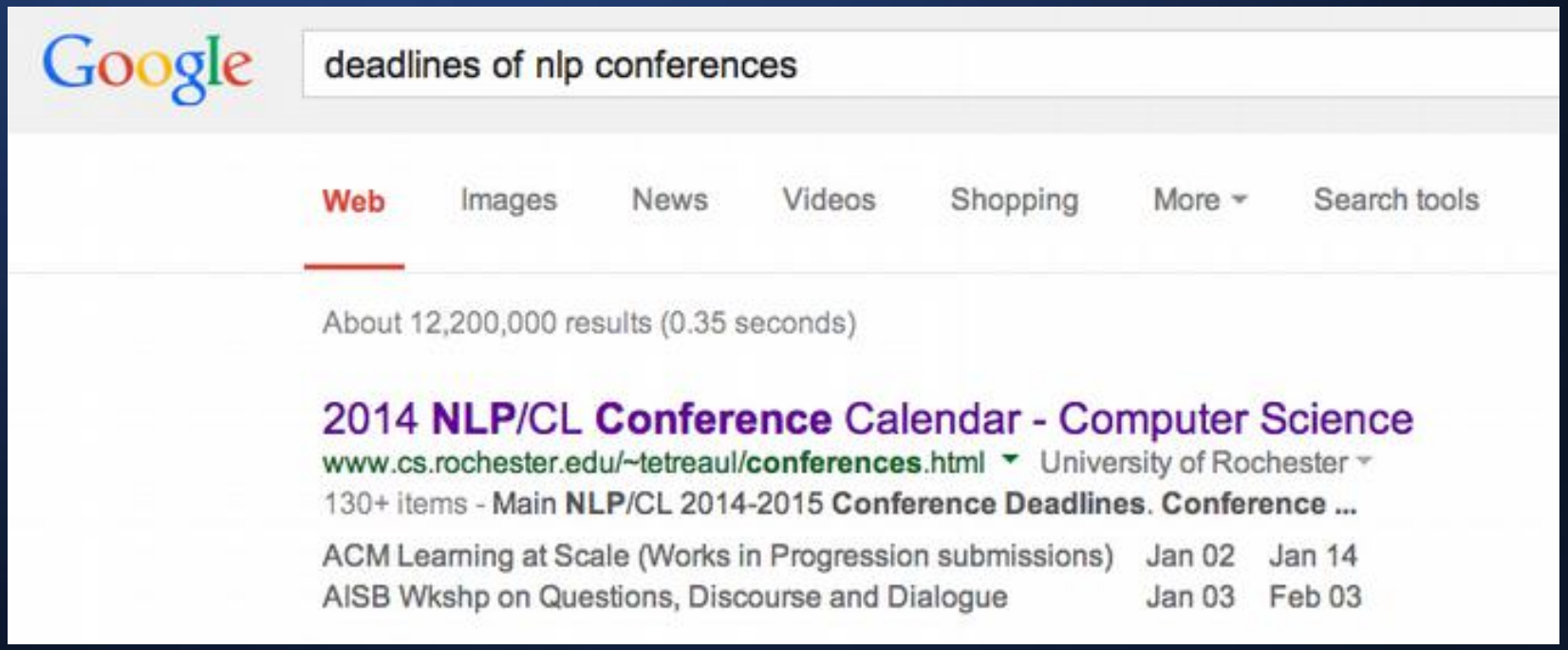
Mainstream Adoption



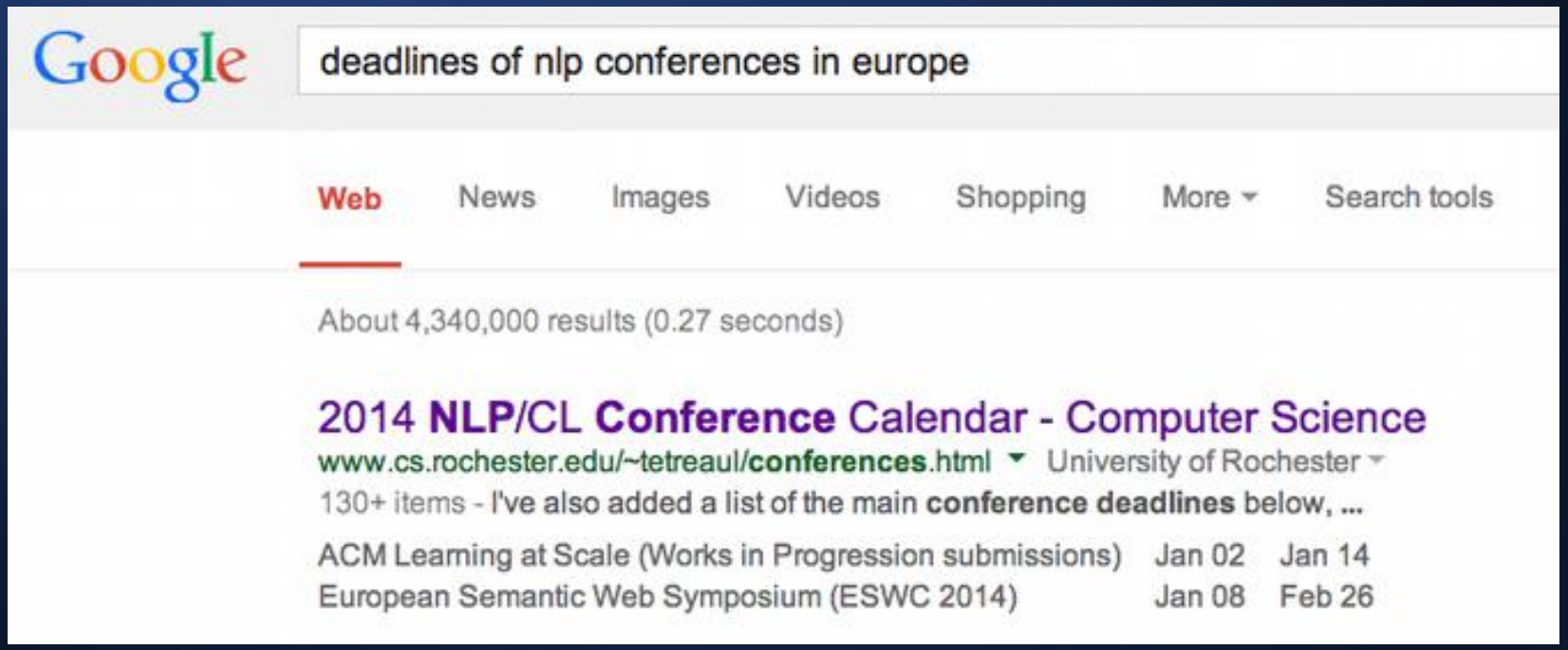
Solved:
The Low-Hanging Fruit



What we should be able to ask

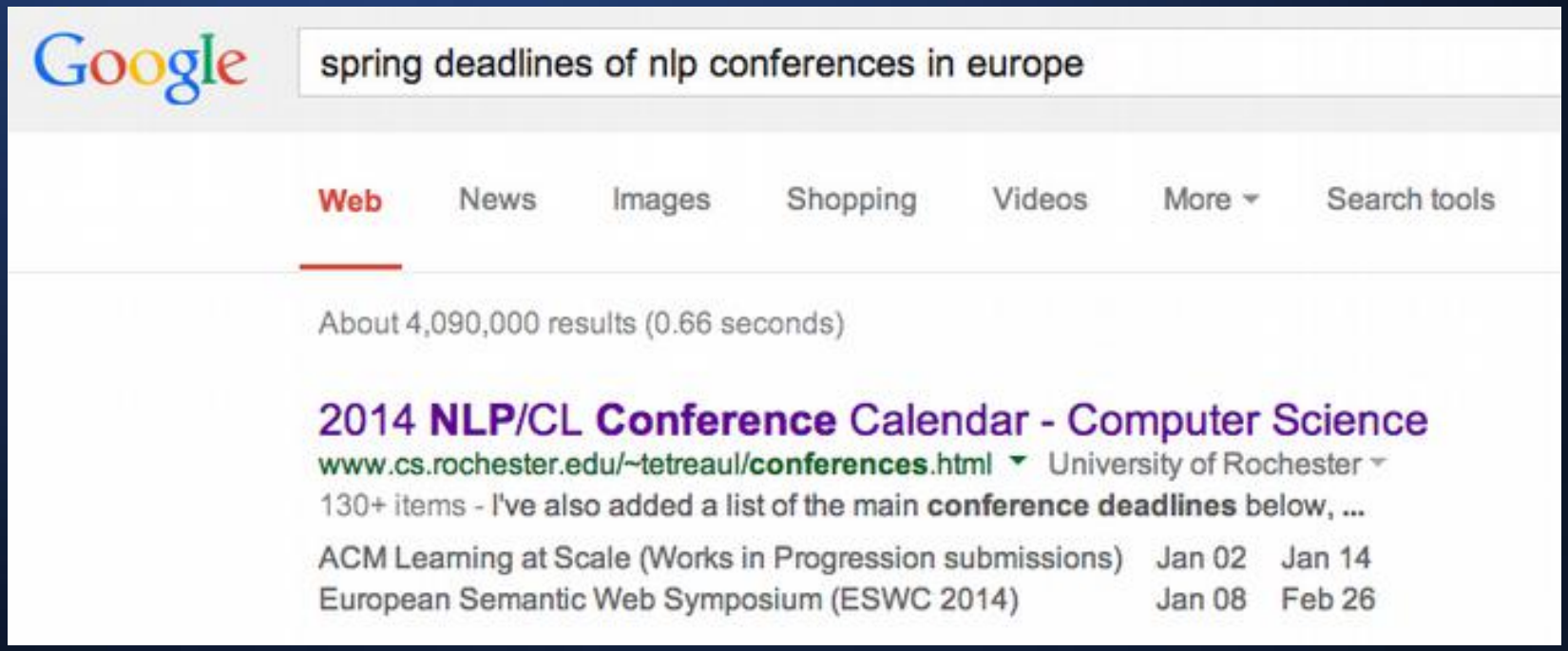


What we should be able to ask



Europe: Could be UK, France, Germany, etc.!

What we should be able to ask



Spring: March, April, etc.

What we should be able to ask

Are any talks at KG 2015 about social media?

The following talks at KG 2015 seem related to social media:
...

Image: Microsoft Cortana

What we should be able to ask

Which talks at KG 2015 are most related to my current projects?

The following talks at KG 2015 seem most related to your work: ...

Image: Microsoft Cortana

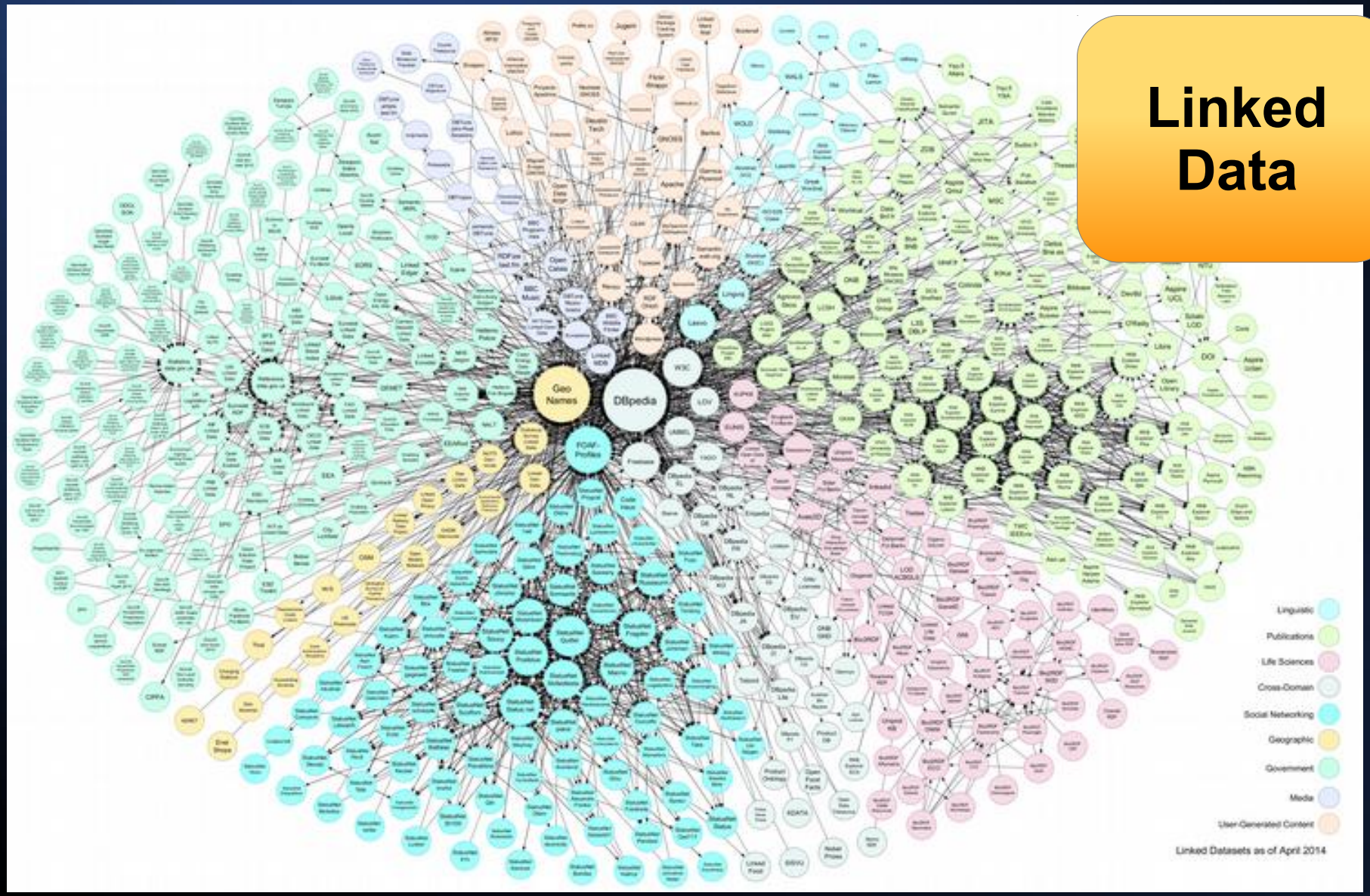
Beyond Collections of Facts

- 1. Knowledge Integration and Organization
- 2. Natural Language Coupling
- 3. Multimodal and Cognitive Knowledge

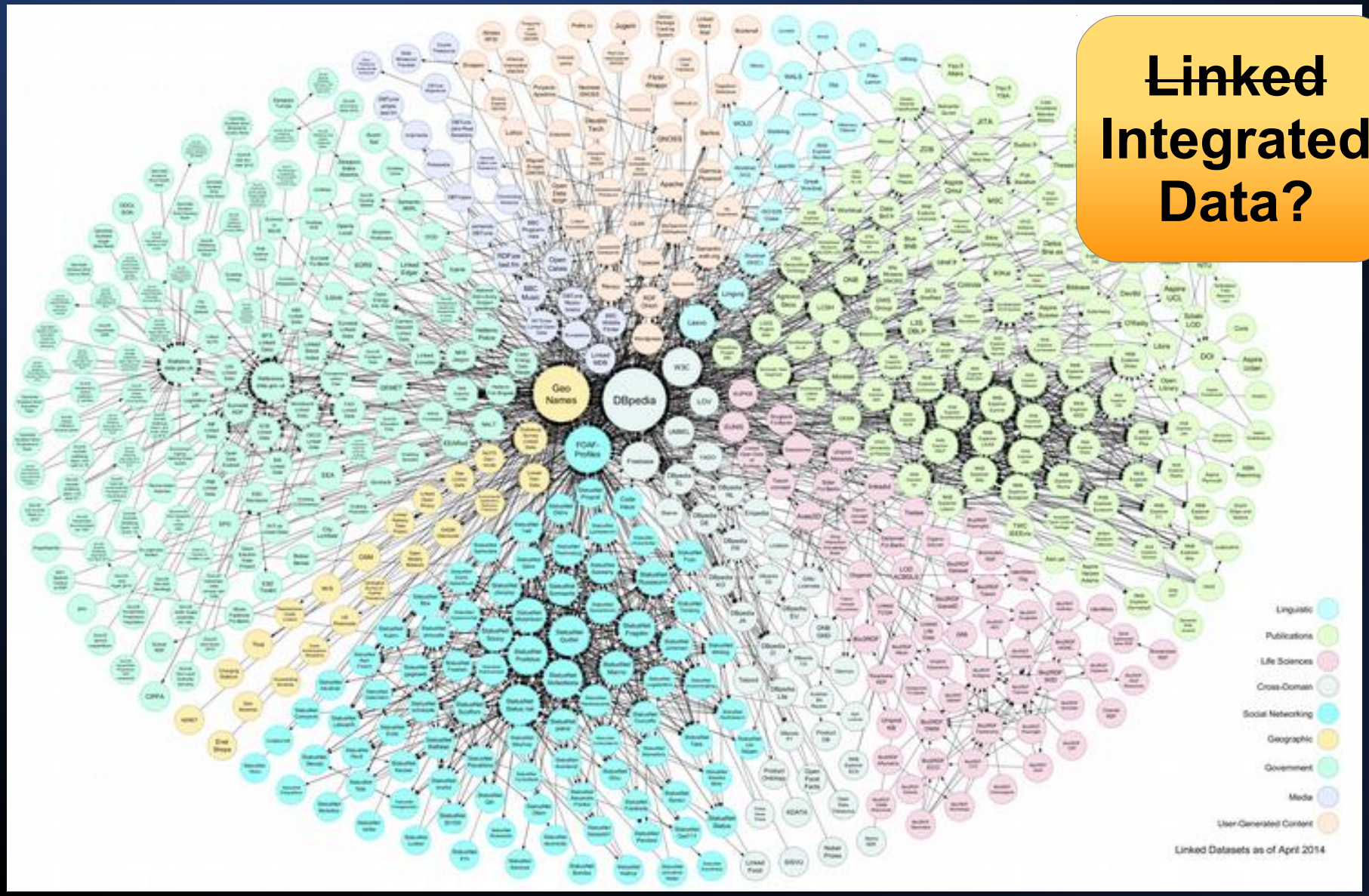
Beyond Collections of Facts

- 1. Knowledge Integration and Organization
- 2. Natural Language Coupling
- 3. Multimodal and Cognitive Knowledge

The Web of Data



The Web of Data



Linked Data Is Still Not Integrated Data

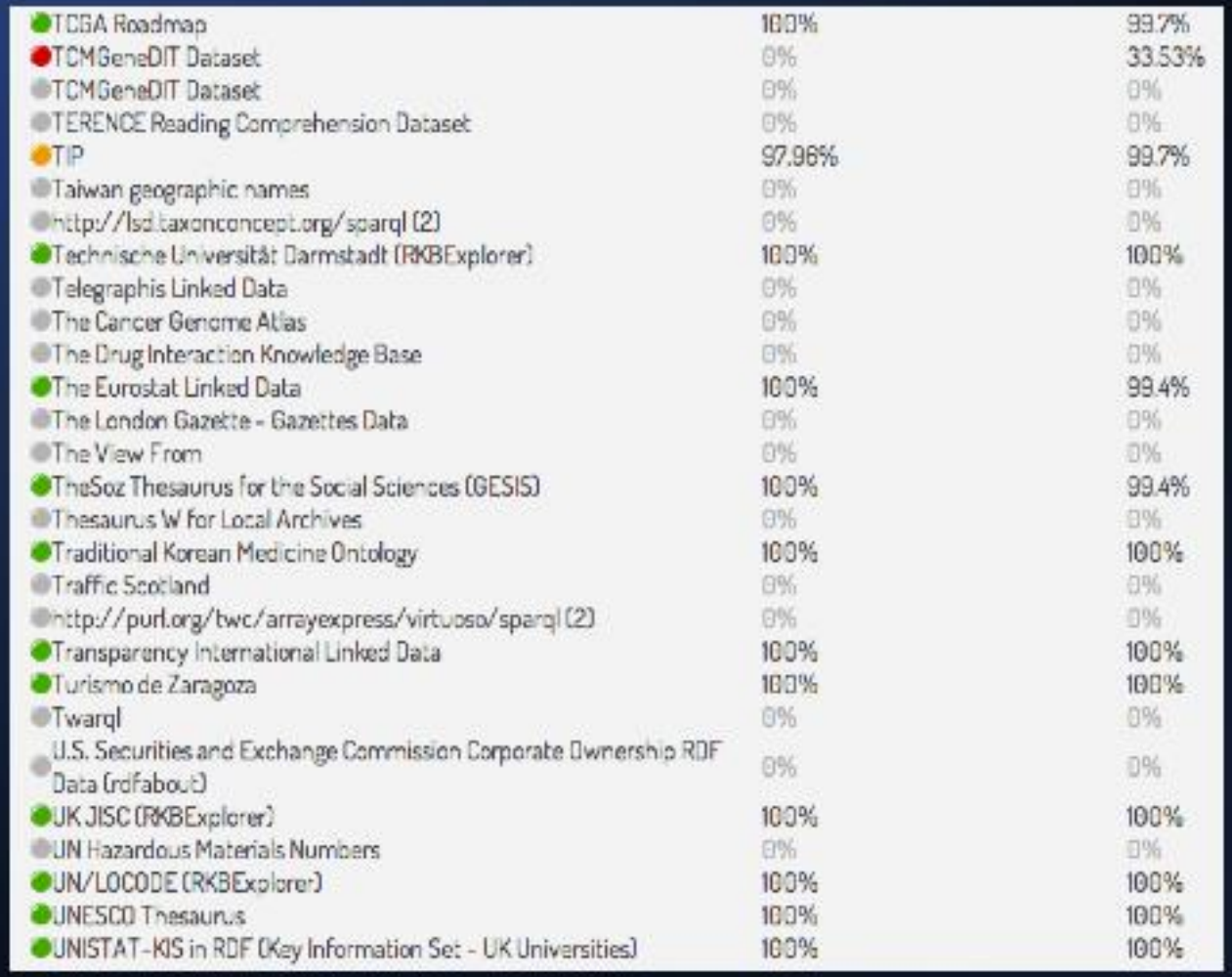
Technical Challenges

- Different access mechanisms:
  Individual HTTP requests, SPARQL endpoints, Downloads
- Downloads often not available
- Servers offline

Linked Data Is Still Not Integrated Data

Servers offline
Poor archiving

Dumps need to be archived and integrated.



TCGA Roadmap	100%	99.7%
TCMGeneDIT Dataset	0%	33.53%
TCMGeneDIT Dataset	0%	0%
TERENCE Reading Comprehension Dataset	0%	0%
TIP	97.96%	99.7%
Taiwan geographic names	0%	0%
http://lsd.taxonconcept.org/sparql (2)	0%	0%
Technische Universität Darmstadt (RKBExplorer)	100%	100%
Telegraphis Linked Data	0%	0%
The Cancer Genome Atlas	0%	0%
The Drug Interaction Knowledge Base	0%	0%
The Eurostat Linked Data	100%	99.4%
The London Gazette - Gazettes Data	0%	0%
The View From	0%	0%
TheSoz Thesaurus for the Social Sciences (GESIS)	100%	99.4%
Thesaurus W for Local Archives	0%	0%
Traditional Korean Medicine Ontology	100%	100%
Traffic Scotland	0%	0%
http://purl.org/twc/arrayexpress/virtuoso/sparql (2)	0%	0%
Transparency International Linked Data	100%	100%
Turismo de Zaragoza	100%	100%
Twarql	0%	0%
U.S. Securities and Exchange Commission Corporate Ownership RDF Data (rdfabout)	0%	0%
UK JISC (RKBExplorer)	100%	100%
UN Hazardous Materials Numbers	0%	0%
UN/LOCODE (RKBExplorer)	100%	100%
UNESCO Thesaurus	100%	100%
UNISTAT-KIS in RDF (Key Information Set - UK Universities)	100%	100%

http://sparqles.okfn.org/

Linked Data Is Still Not Integrated Data

https://commons.wikimedia.org/wiki/File:Gov%27t_Shutdown,_Locked.._-_Flickr_-_daveynin.jpg

Linked Data Is Still Not Integrated Data

Semantic Challenges

- Different ontologies/schemas/vocabularies
- Links are incomplete
- Links are bad
- Data is sometimes bad or outdated

Disconnected Information

Albert Einstein

bornIn

HU Berlin

WIKIPEDIA

Ulm

Rudolf Carnap

IsA

Philosopher

Carnap — Studied at → University of Berlin

Einstein — Was Professor at → University of Berlin

GeoNames

• • •

Current Situation

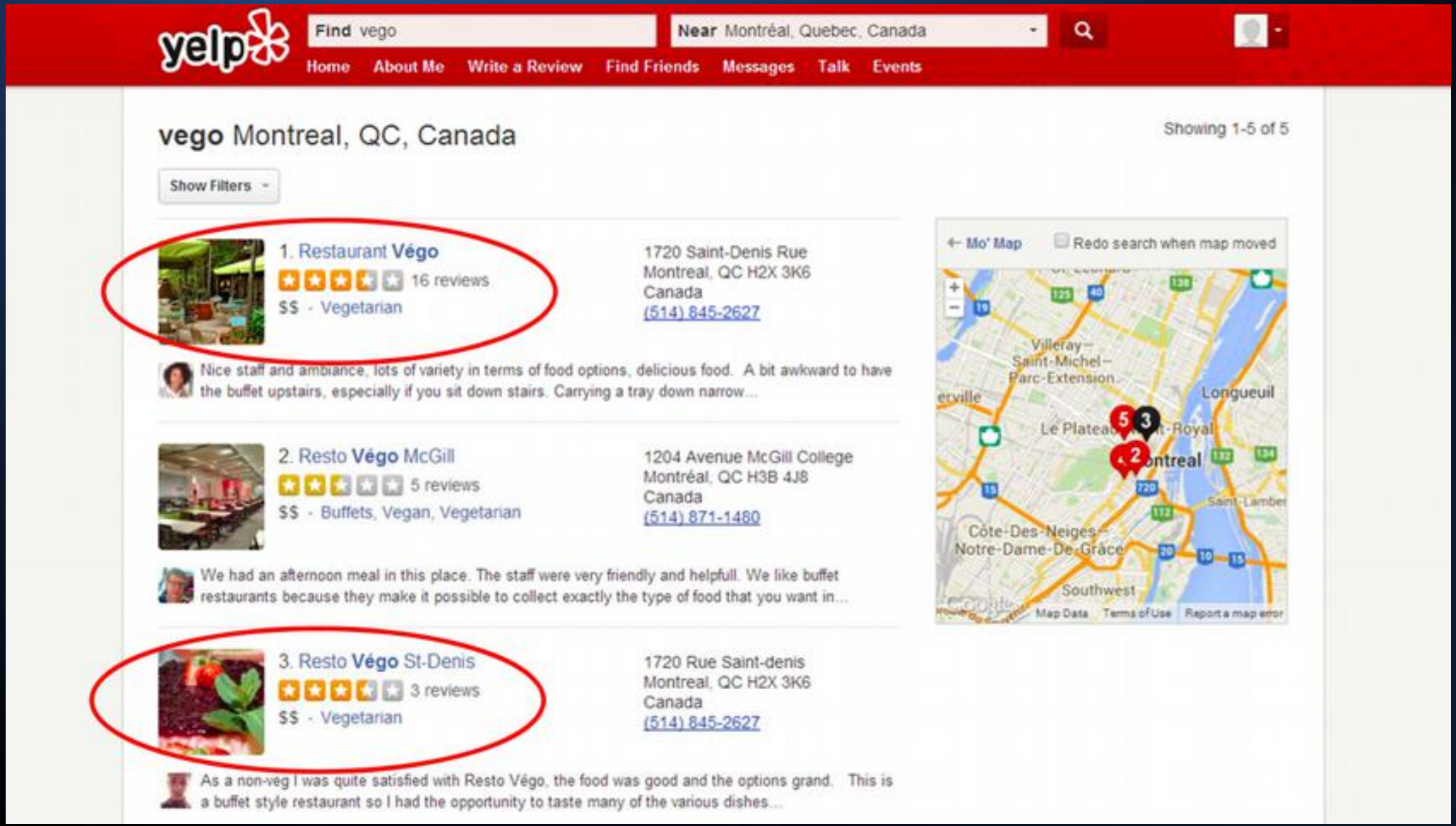
Alexandre Duret-Lutz
https://www.flickr.com/photos/gadl/110845690/

Goal: Integrated and Organized Knowledge

Theological Hall, Strahov Monastery Library, Prague

Entity Integration

Entity Integration



Source: Bordes & Gabrilovich. KDD 2014 Tutorial

Entity Integration: Challenges

Michelangelo

Artist

Michelangelo di Lodovico Buonarroti Simoni , commonly known as Michelangelo, was an Italian sculptor, painter, architect, poet, and engineer of the High Renaissance who exerted an unparalleled influence on the... wikipedia.org

Born: March 6, 1475, Caprese Michelangelo

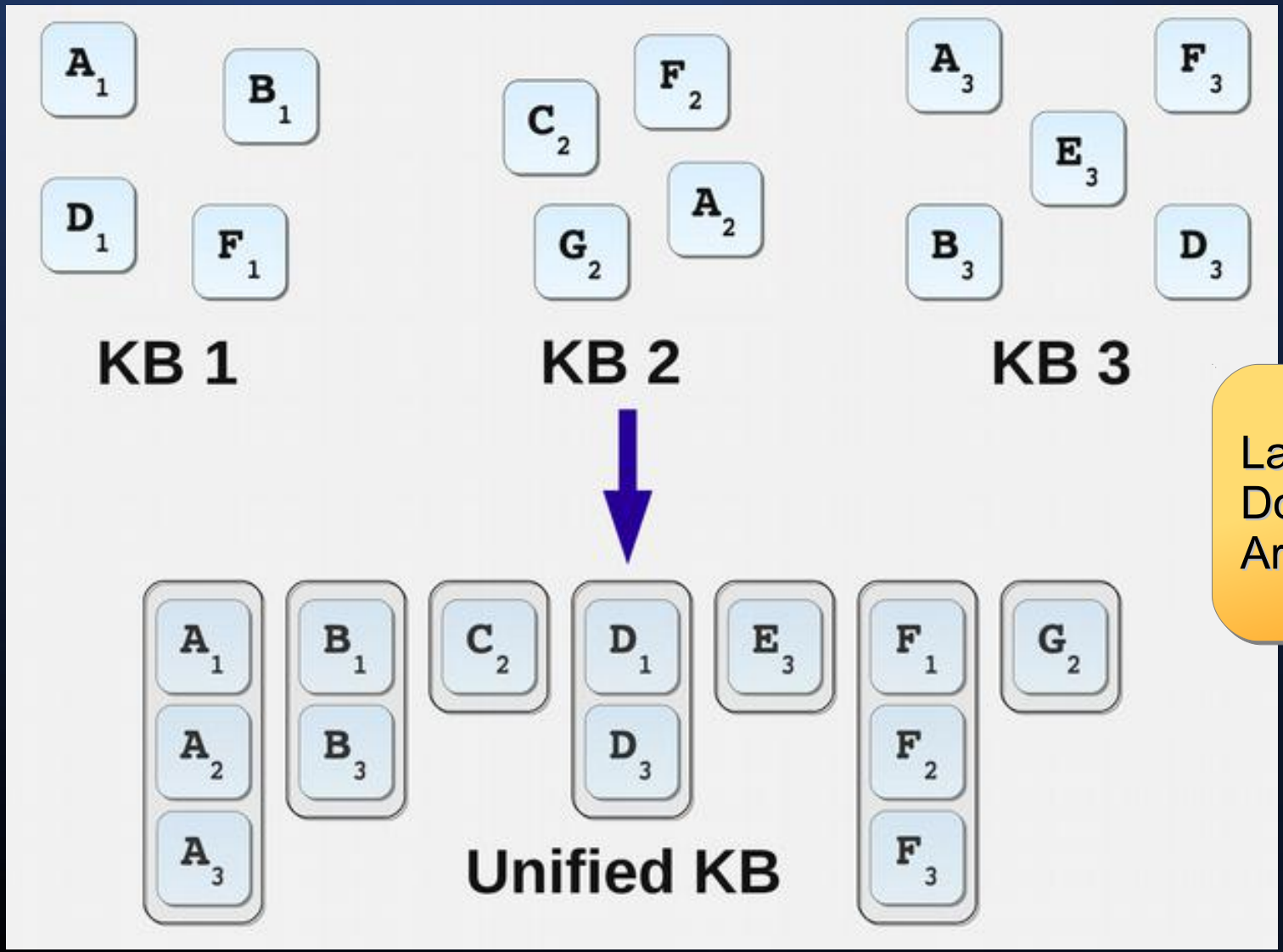
Died: February 18, 1564, Rome

Parents: Ludovico di Leonardo di Buonarotto Simoni, Francesca di Neri del Miniato di Siena

Feedback

Source: Peter Mika

Entity-Level Integration



Language-specific,
Domain-specific,
Arbitrary Databases

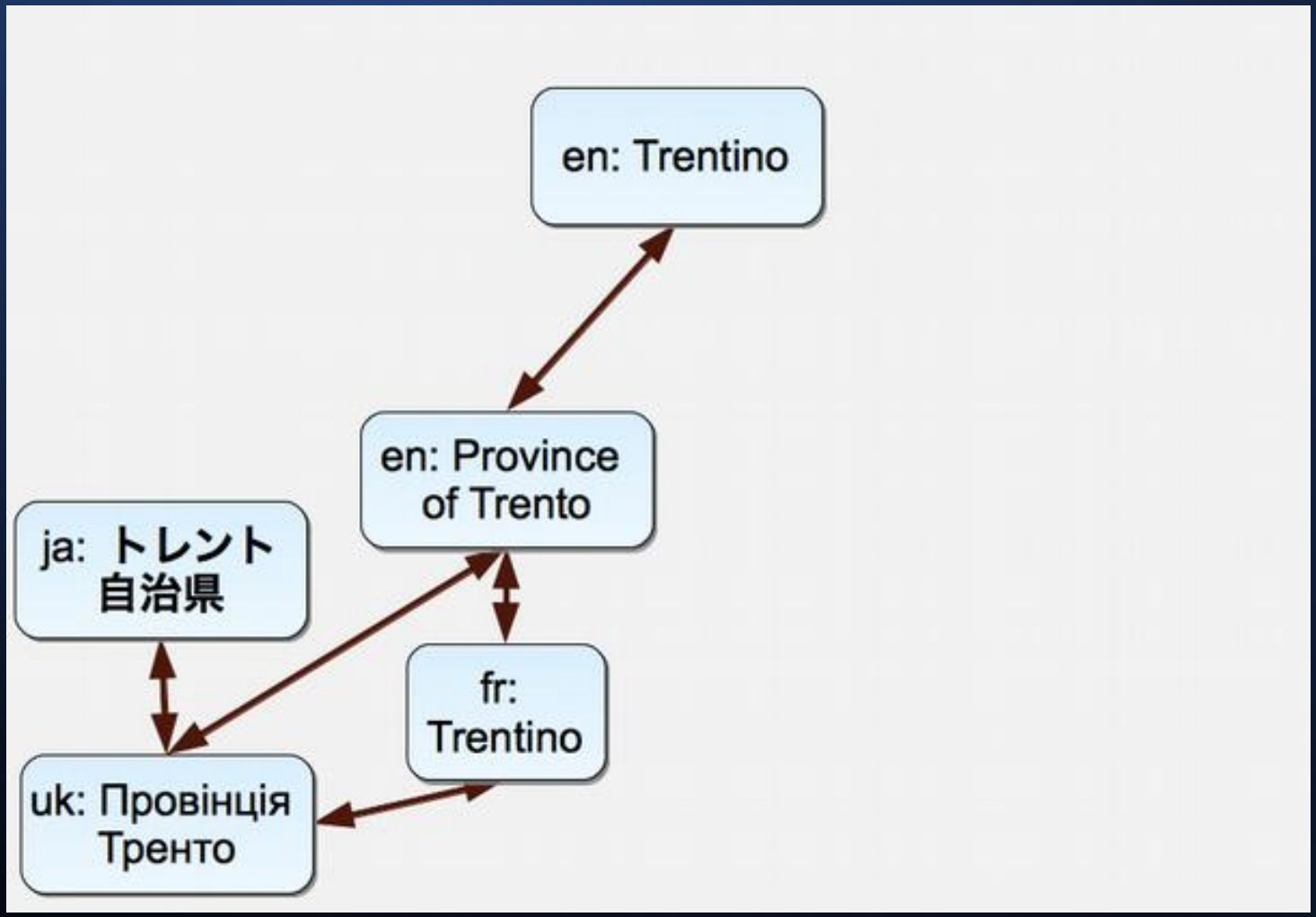
Entity Integration: Computing Similarity Scores

dbpedia:The_Godfather	linkedmdb:film/43338
dbpedia-owl:director dbpedia:Francis_Ford_Coppola	movie:actor linkedmdb:actor/10107
dbpedia-owl:starring dbpedia:Al_Pacino	movie:actor linkedmdb:actor/10486
dbpedia-owl:starring dbpedia:Marlon_Brando	dc:date 1972
dbpedia-owl:starring dbpedia:Marlon_Brando	movie:director linkedmdb:director/8405
dcterms:subject category:Mafia_films	movie:editor linkedmdb:editor/2902
...	...

Are these two the same?

Compare attributes, relations, text
→ Similarity score (weighted link)

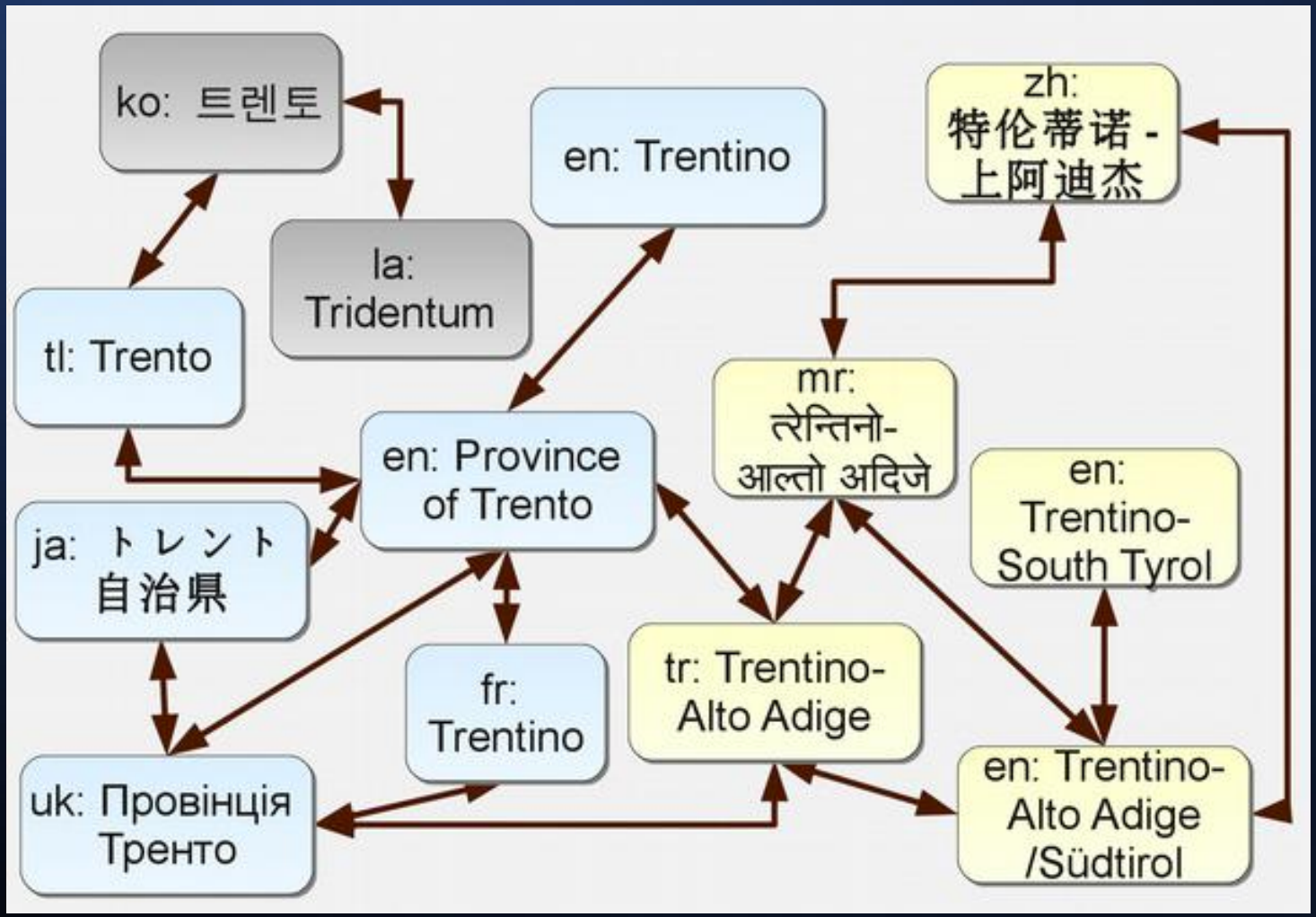
Merging Structured Data



Use Identity Links to connect What is equivalent

ACL 2010
AAAI 2013

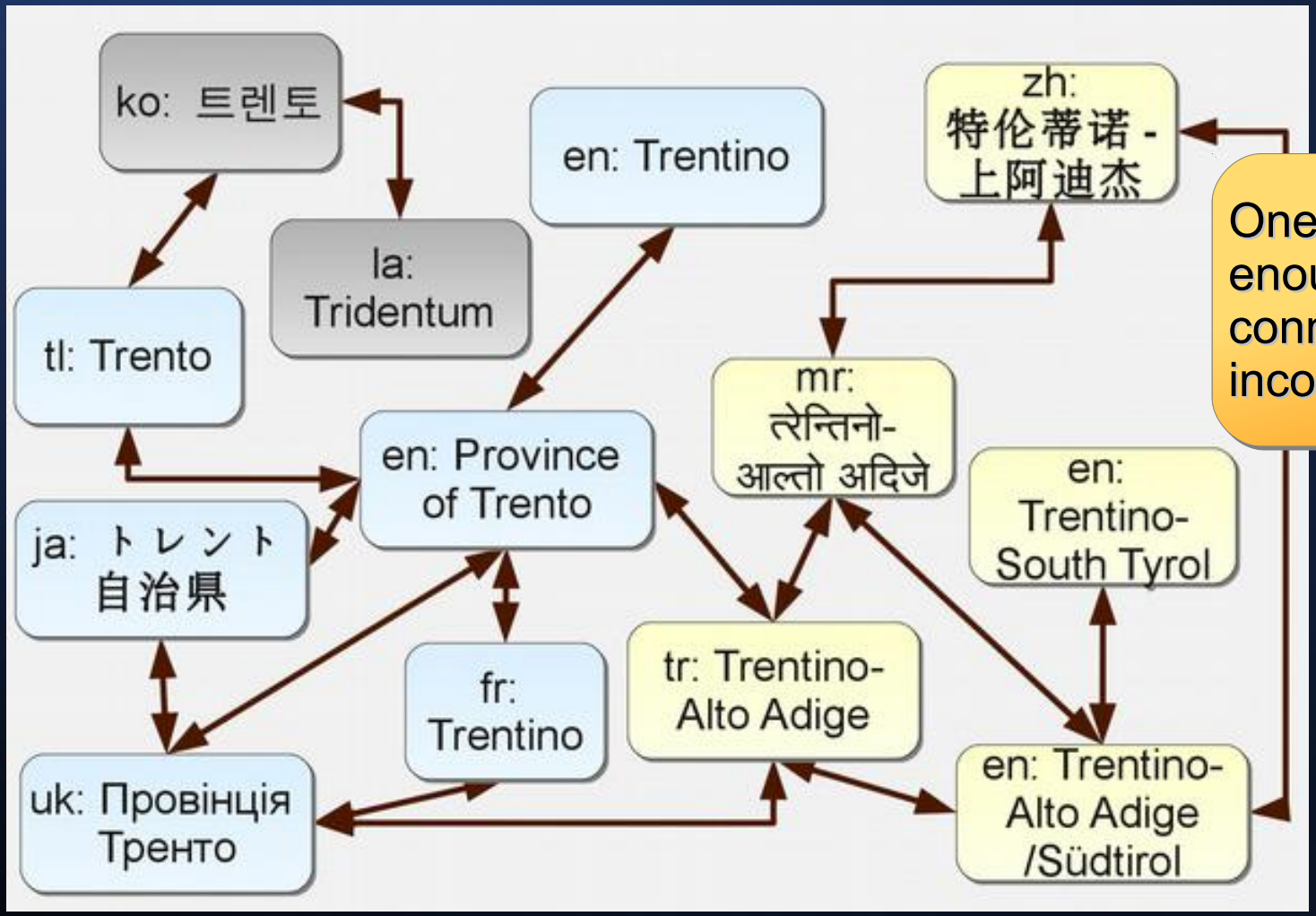
Merging Structured Data



ACL 2010
AAAI 2013

Merging Structured Data

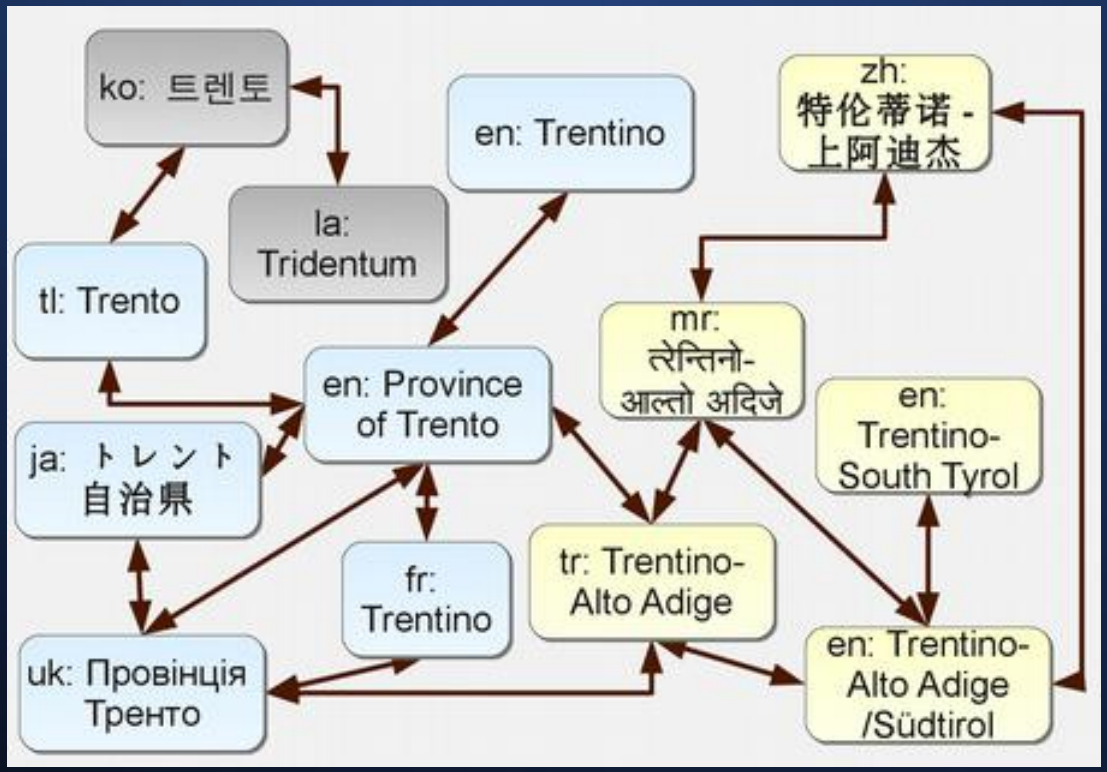
Merging Structured Data



One bad link is enough to make a connected component inconsistent

ACL 2010
AAAI 2013

Merging Structured Data

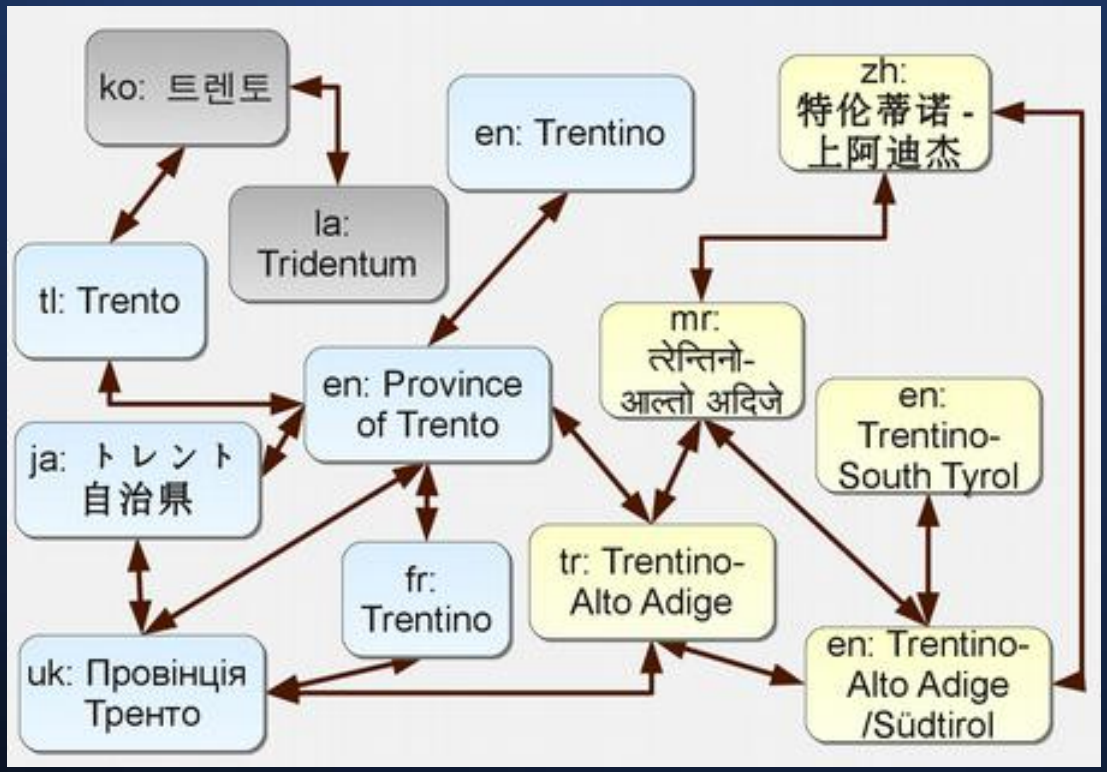


Optimization Problem:
Make clusters consistent

Either by deleting edges or merging concepts

ACL 2010
AAAI 2013

Merging Structured Data



Min. cost solution:
NP-hard
APX-hard

ACL 2010
AAAI 2013

Merging Structured Data

Linear Program Relaxation

minimize

$$\sum_{e \in E} d_e w(e) + \sum_{i=1}^{n} \sum_{j=1}^{l_i} \sum_{v \in D_{i,j}} u_{i,v} w(D_i)$$

subject to

$$p_{i,j,v} = u_{i,v} \qquad \forall i,\ j<l_i,\ v \in D_{i,j}$$

$$p_{i,j,v} + u_{i,v} \geq 1 \qquad \forall i,\ j<l_i,\ v \in \bigcup_{k>j} D_{i,k}$$

$$p_{i,j,v} \leq p_{i,j,u} + d_e \qquad \forall i,\ j<l_i,\ e=(u,v) \in E$$

$$d_e \geq 0 \qquad \forall e \in E$$

$$u_{i,v} \geq 0 \qquad \forall i,\ v \in \bigcup_{j=1}^{l_i} D_{i,j}$$

$$p_{i,j,v} \geq 0 \qquad \forall i,\ j<l_i,\ v \in V$$

- minimize costs
- remove edges from graph or ignore distinctness of nodes
- such that there is no path between node pairs still considered distinct

Finally, use region growing algorithm in the spirit of Leighton & Rao 1988

Approximation Guarantee:
$4\ln(nq+1)$
for n distinctness assertions,
$q=\max |D_{i,j}|$
but independent of $|D_i|$!

Merging Structured Data

Linear Program Relaxation

minimize

$$\sum_{e \in E} d_e w(e) + \sum_{i=1}^{n} \sum_{j=1}^{l_i} \sum_{v \in D_{i,j}} u_{i,v} w(D_i)$$

subject to

$$p_{i,j,v} = u_{i,v} \qquad \forall i,\ j<l_i,\ v \in D_{i,j}$$

$$p_{i,j,v} + u_{i,v} \geq 1 \qquad \forall i,\ j<l_i,\ v \in \bigcup_{k>j} D_{i,k}$$

$$p_{i,j,v} \leq p_{i,j,u} + d_e \qquad \forall i,\ j<l_i,\ e=(u,v) \in E$$

$$d_e \geq 0 \qquad \forall e \in E$$

$$u_{i,v} \geq 0 \qquad \forall i,\ v \in \bigcup_{j=1}^{l_i} D_{i,j}$$

$$p_{i,j,v} \geq 0 \qquad \forall i,\ j<l_i,\ v \in V$$

- minimize costs
- remove edges from graph or ignore distinctness of nodes
- such that there is no path between node pairs still considered distinct

Nice:

This generalizes the Hungarian Algorithm from bipartite matchings to many non-standard kinds of matchings

(cf. de Melo. AAAI 2013)

Taxonomic Integration

Question Answering

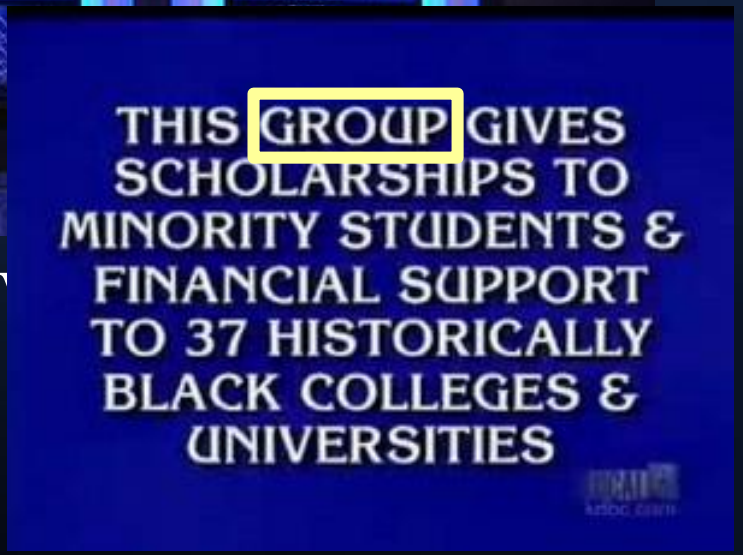
IBM's Jeopardy!-winning Watson system

Question Answering

IBM's Jeopardy!-winning Watson system

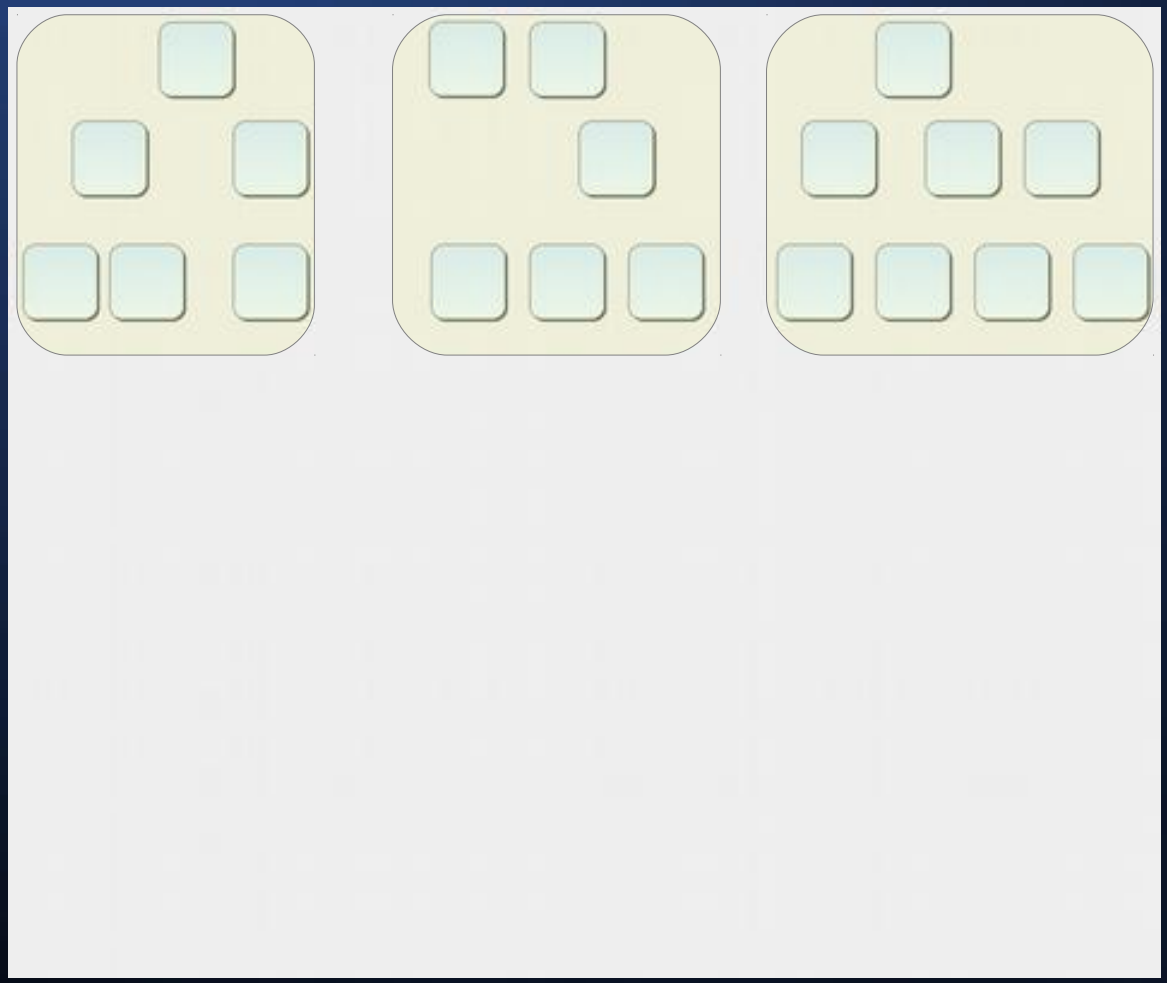
yago
select knowledge

Question Answering



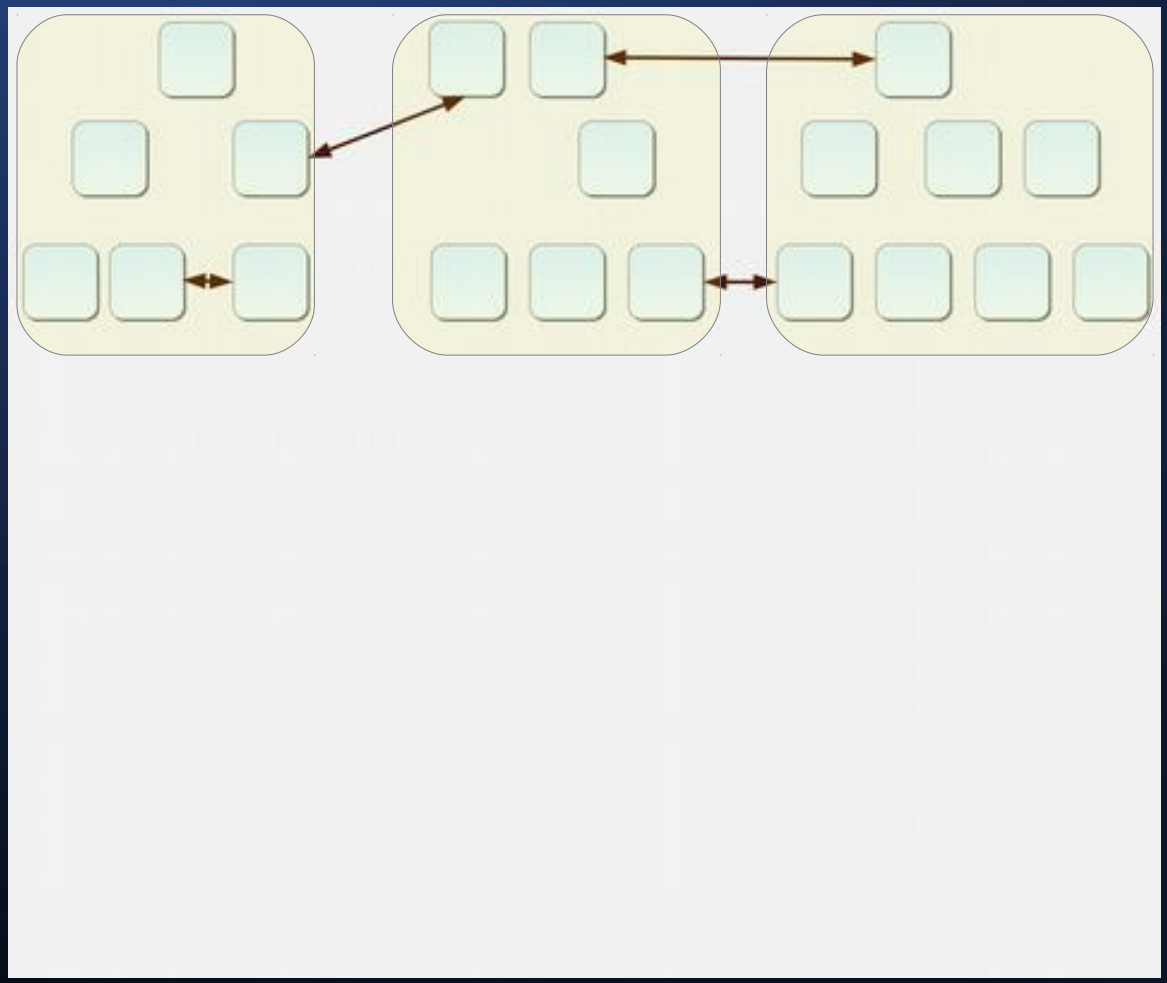
IBM's Jeopardy!-
system

Taxonomic Integration: MENTA Approach



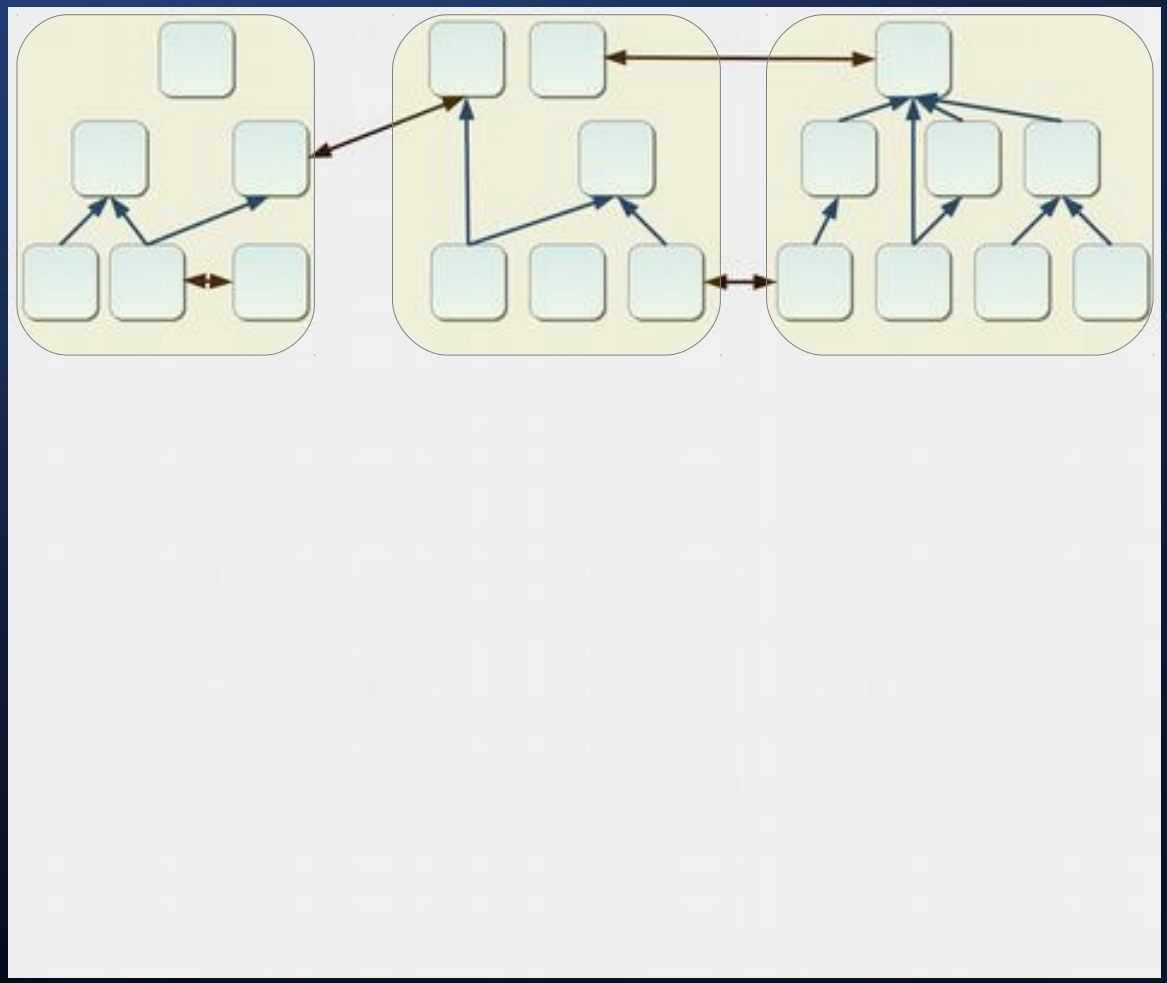
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



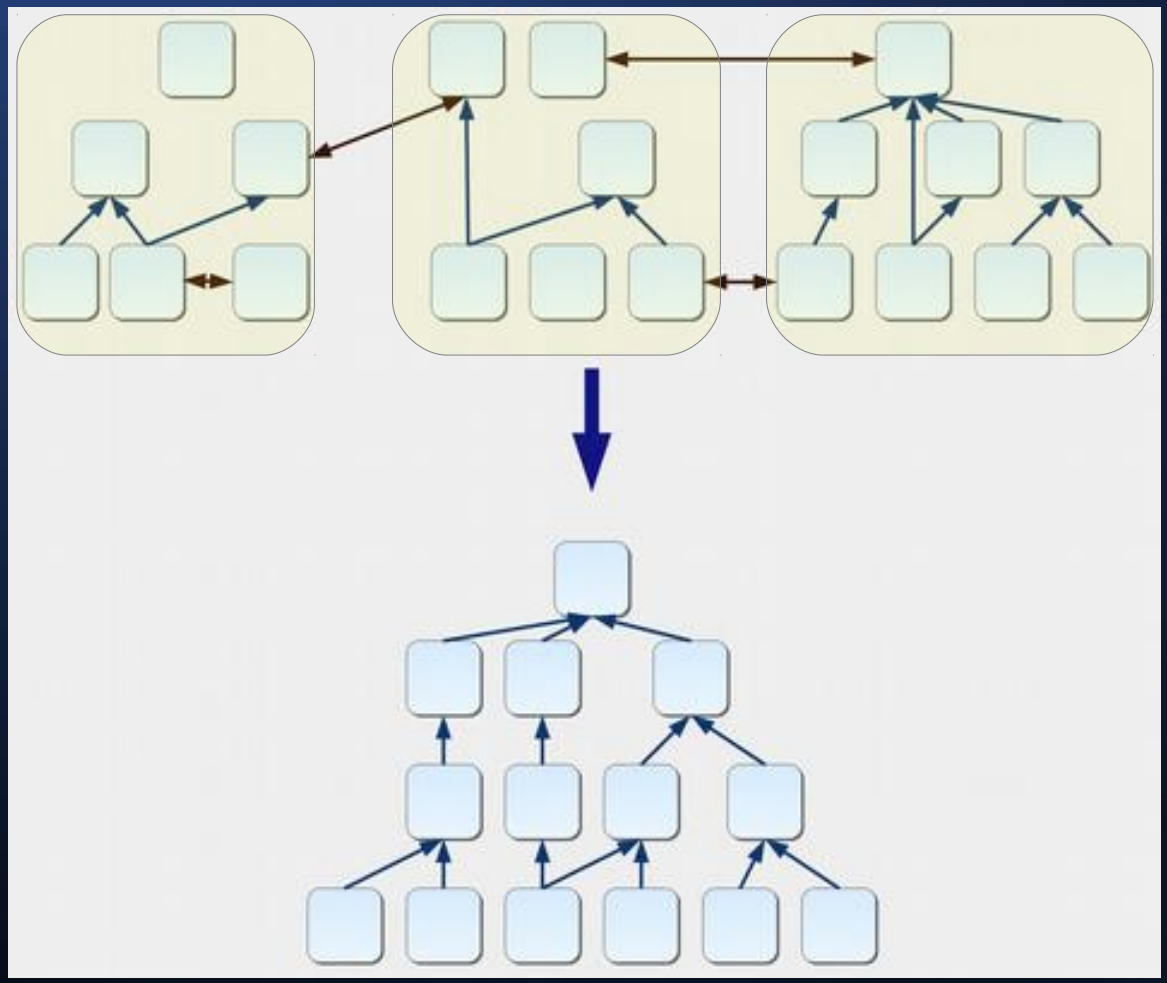
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



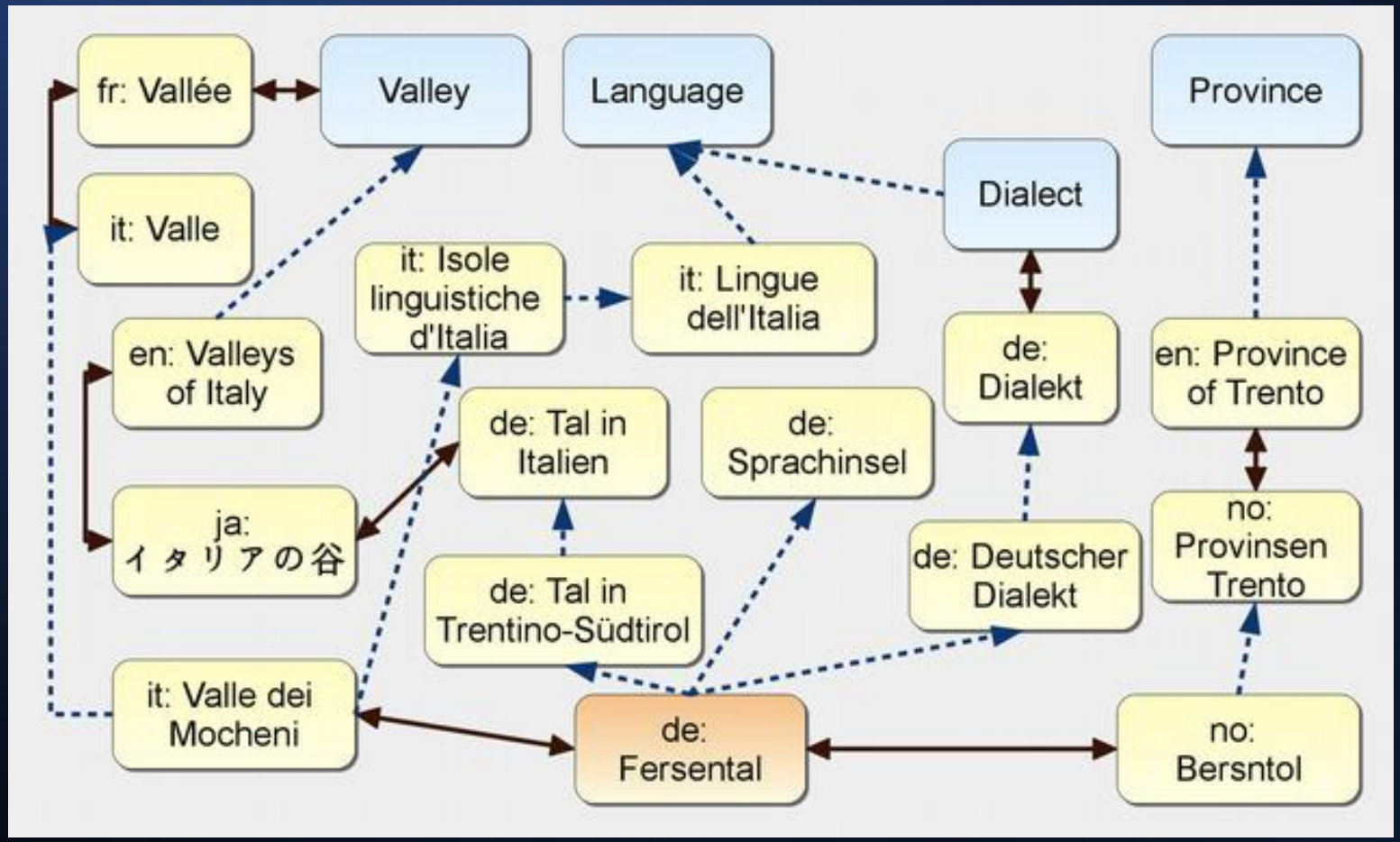
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

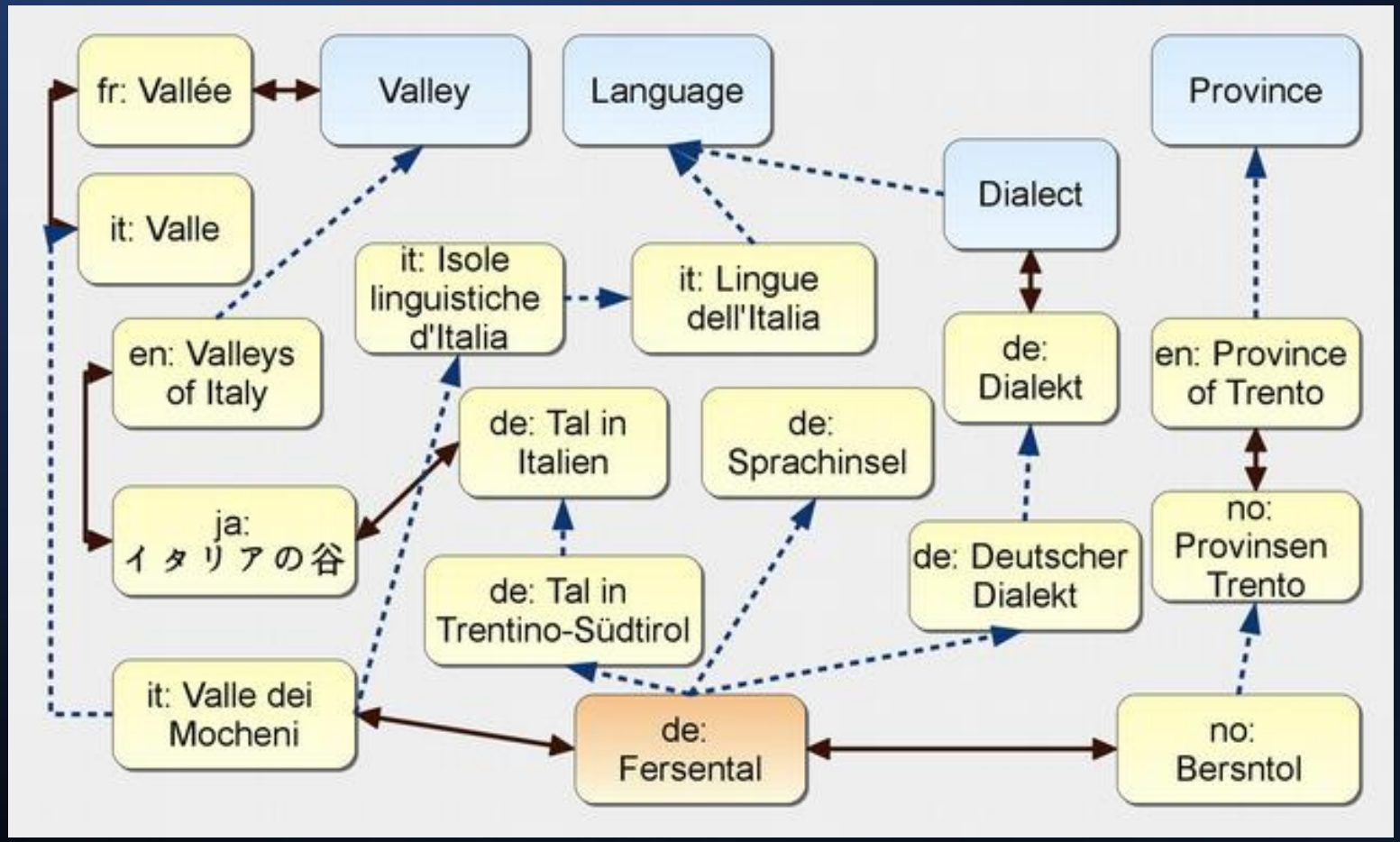
Taxonomic Integration: MENTA



Taxonomic Integration: MENTA

Image: https://de.wikipedia.org/wiki/Datei:Bersntol_palae.jpg

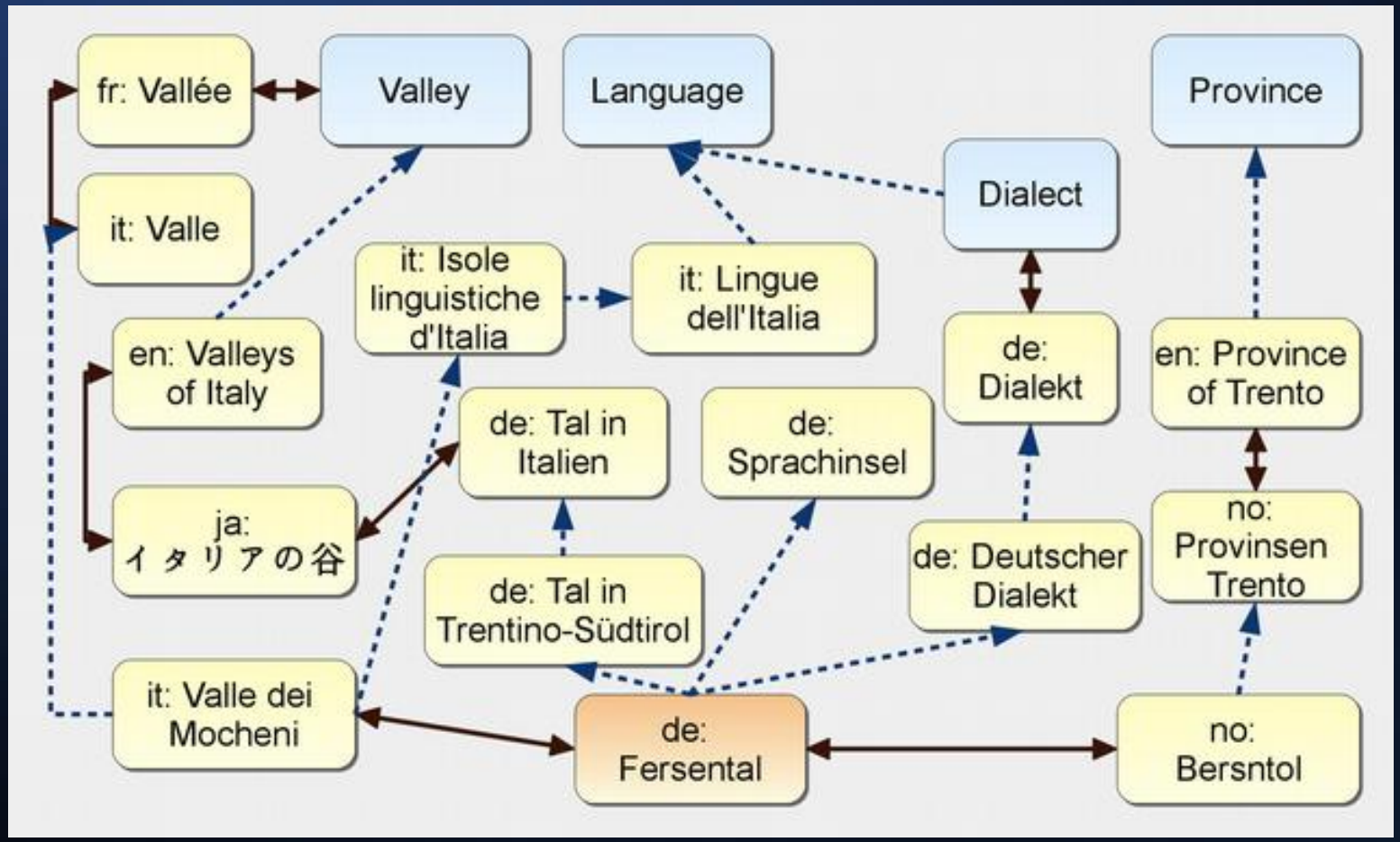
Taxonomic Integration: MENTA



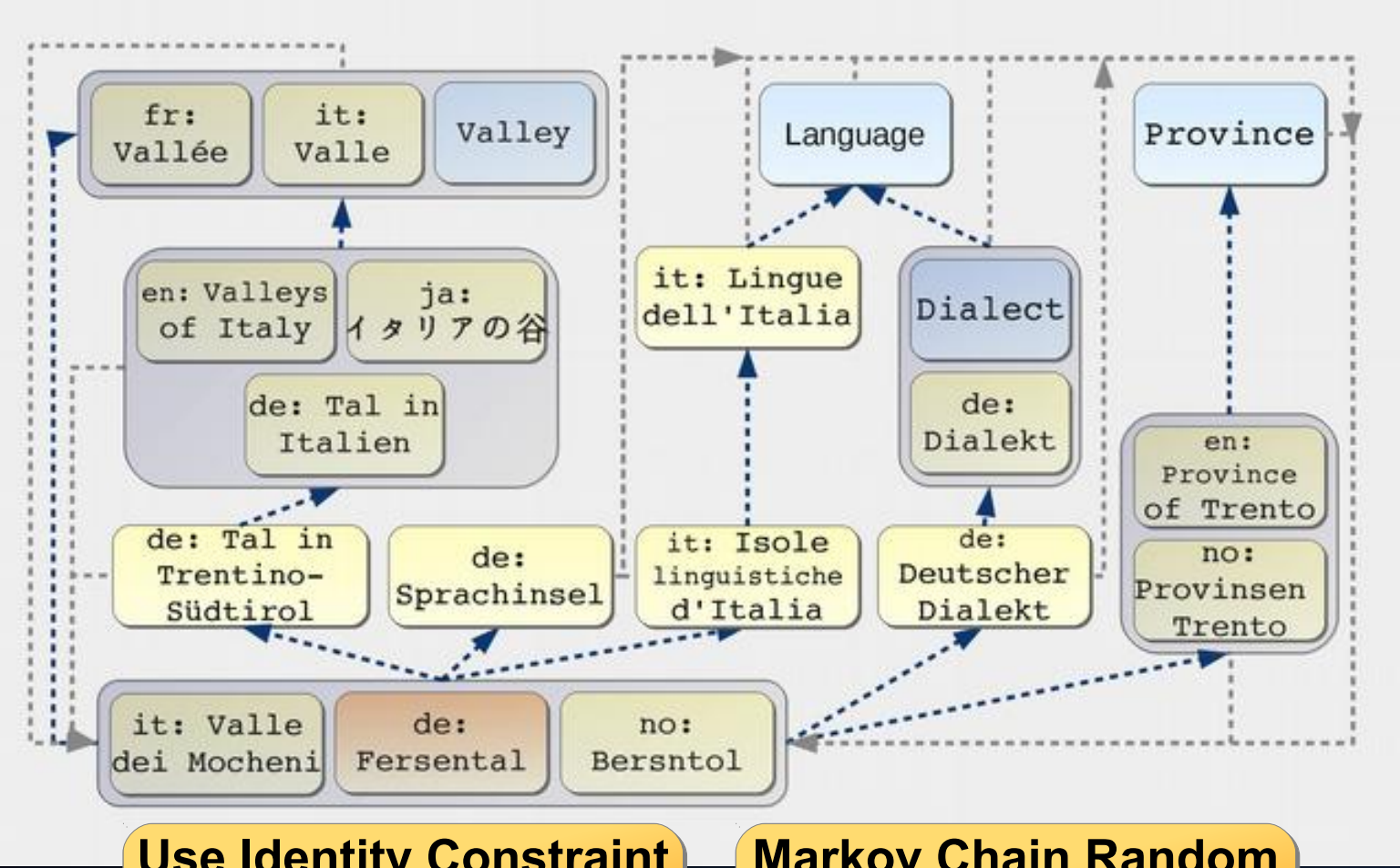
Taxonomic Integration: MENTA

https://de.wikipedia.org/wiki/Datei:Language_distribution_Trentino_2011.png

Taxonomic Integration: MENTA



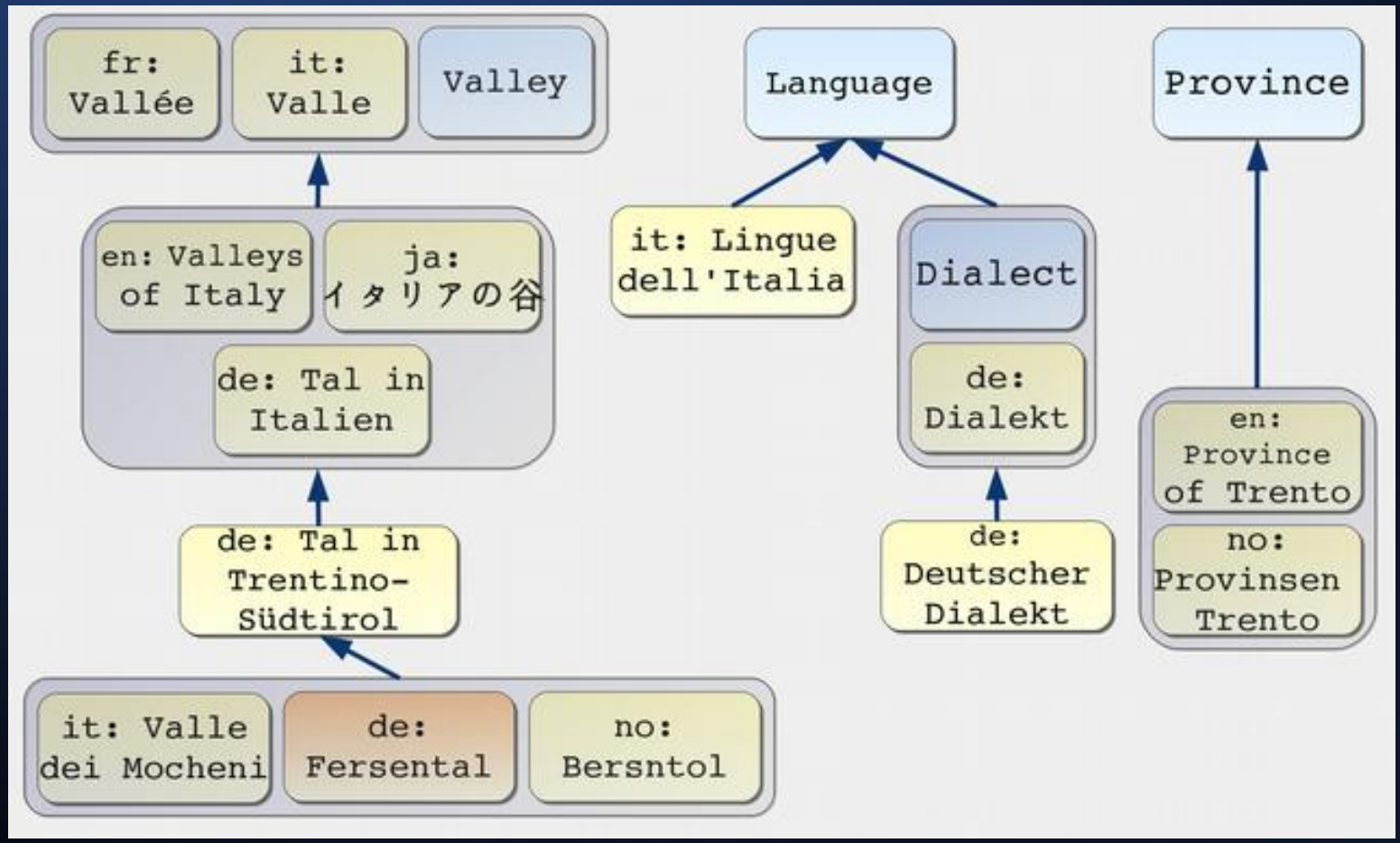
Taxonomic Integration: MENTA



Use Identity Constraint Algorithm to form equivalence classes

Markov Chain Random Walk with Restarts to Rank Parents

Taxonomic Integration: MENTA



Taxonomic Integration: MENTA

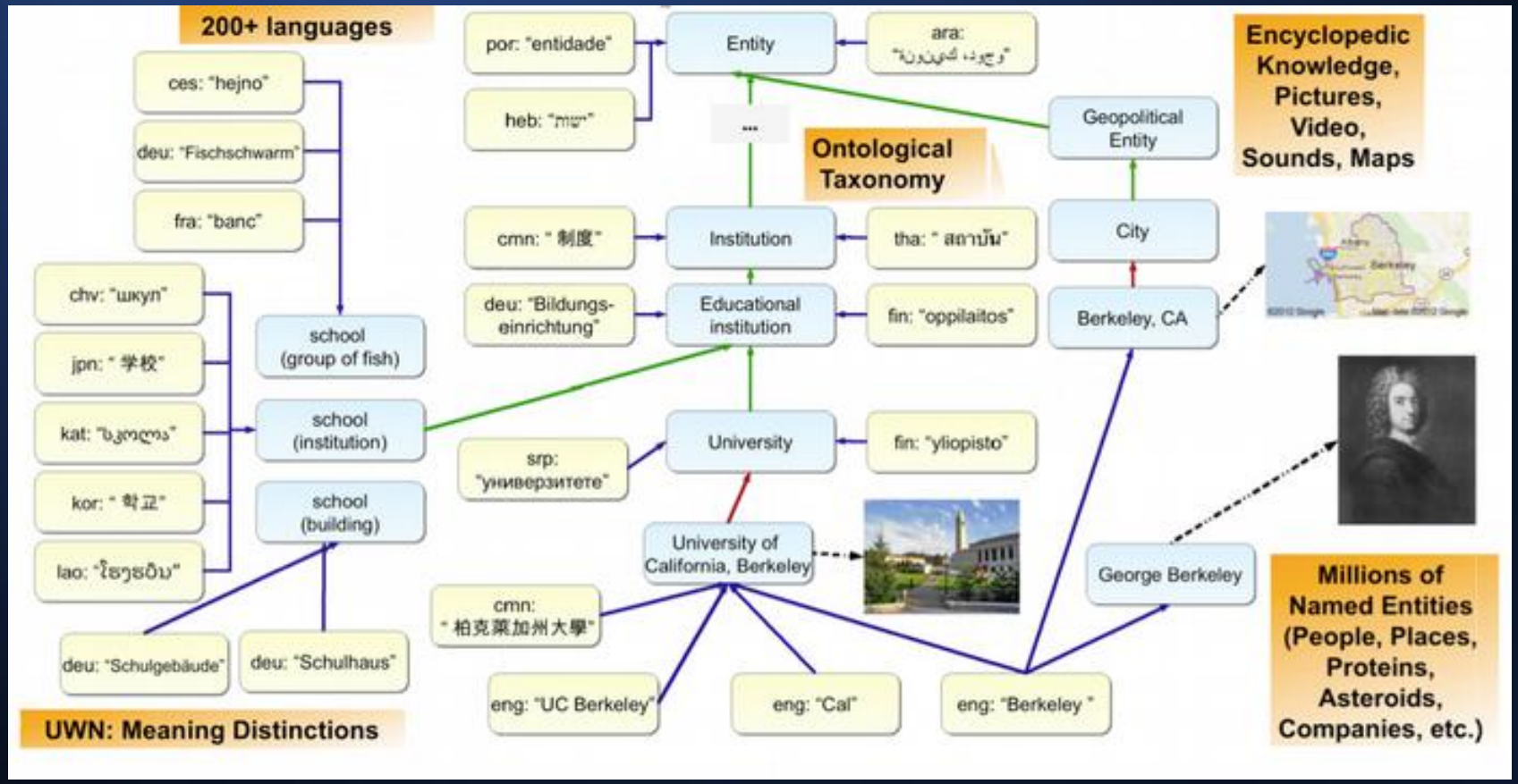
Entity		WordNet parent	Wikipedia parent
Fersental	1.	valley	Valleys
	2.	natural depression	Valleys of Italy
	3.	geological formation	Valleys of Trentino / Alto Adige
Cagayan National High School	1.	secondary school	Secondary school
	2.	school	School
	3.	educational institution	High schools in the Philippines
The Spanish Tragedy	1.	book	Book
	2.	publication	British plays
	3.	piece of work	Plays

Knowledge Integration

Knowledge Integration

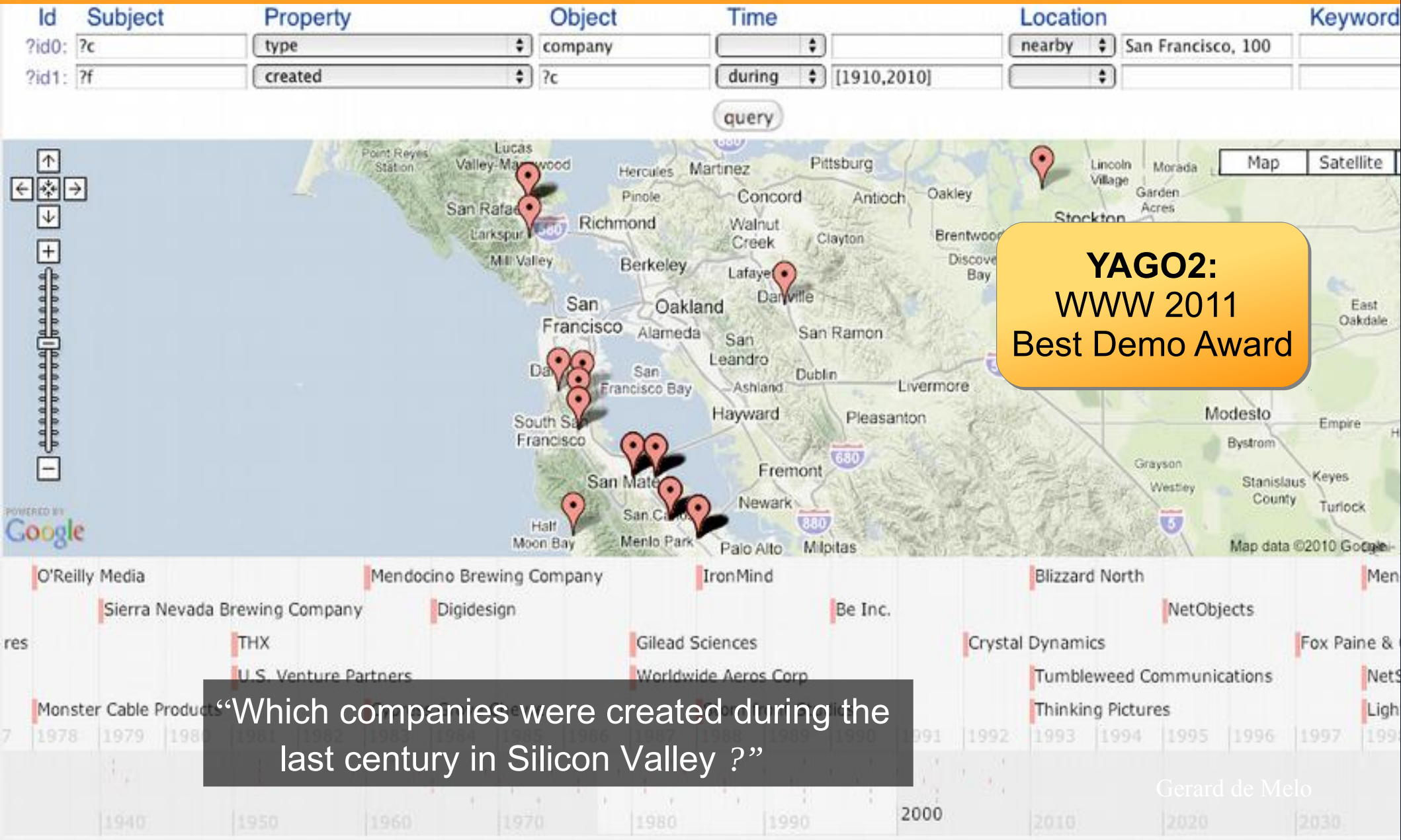
X isAuthorOf Y
Y writtenBy X
X wrote Y
Y writtenInYear Z

UWN/MENTA



multilingual extension of WordNet for
word senses and taxonomical information over 200 languages

Search Interfaces

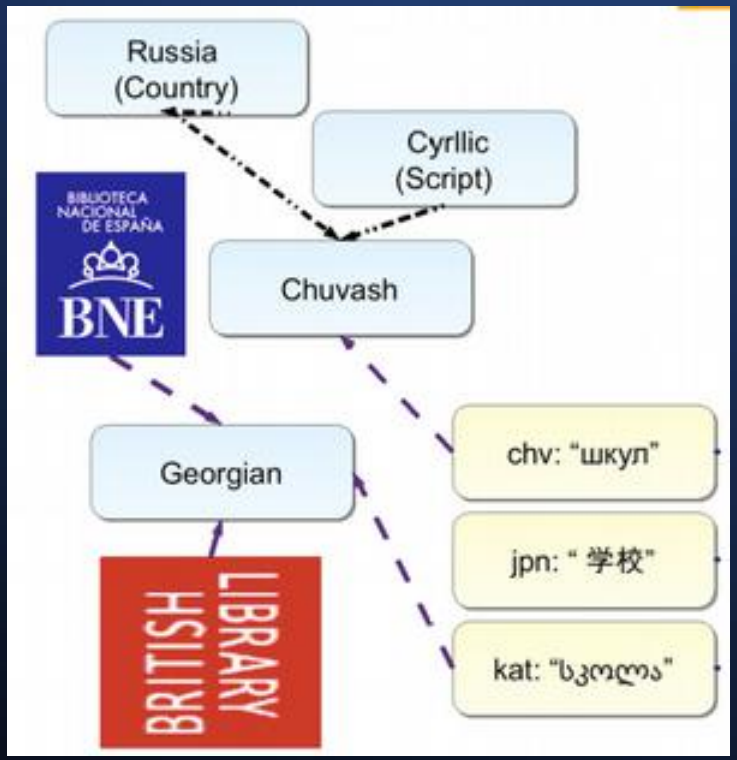


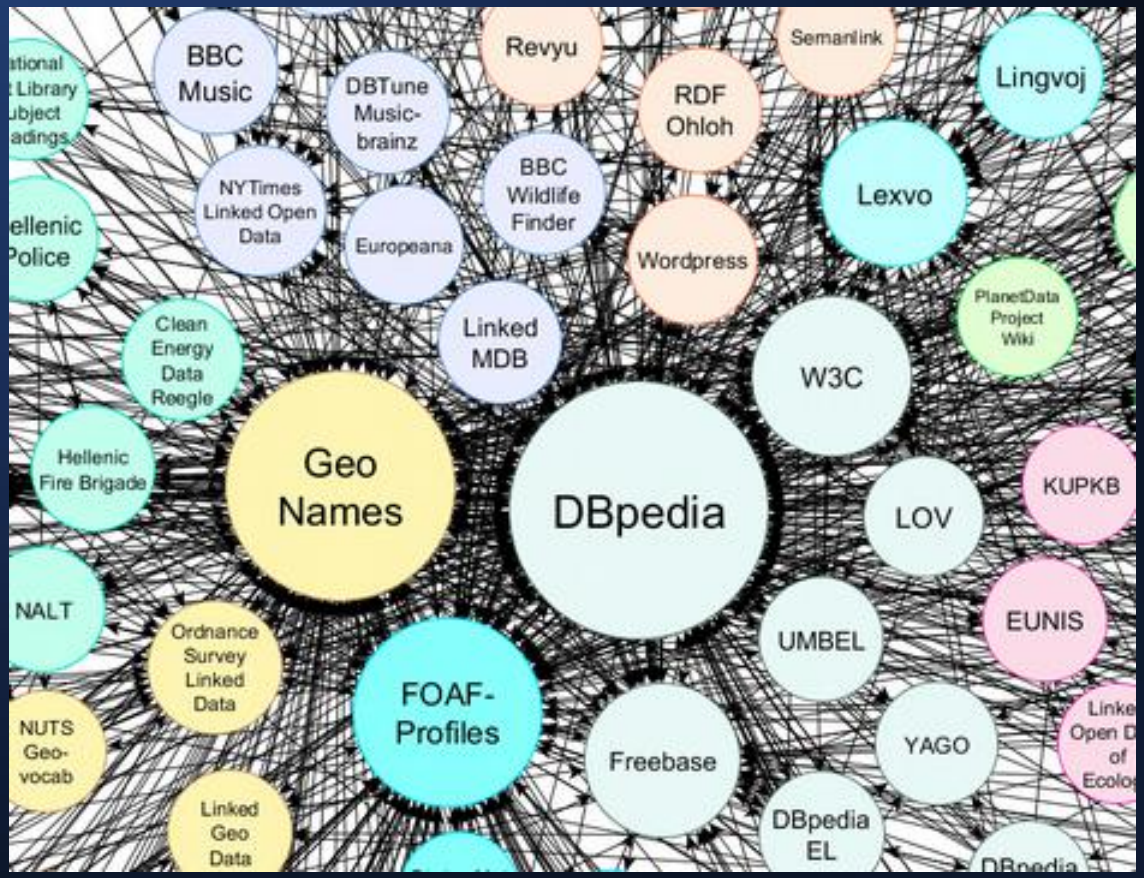
"Which companies were created during the last century in Silicon Valley ?"

Example: Lexvo.org

- **languages:** WordNet, ISO 639-3, Ethnologue, Linguist List, Wikipedia
- **language families/collections:** WordNet, ISO 639-5, Wikipedia
- **writing systems:** ISO 15924, Unicode CLDR
- **geographical regions:** Ethnologue, Unicode CLDR

Lexvo.org





Interdisciplinary Work, e.g. in Digital Humanities

Semantic Web Journal 2014

YAGO-SUMO: Knowledge Integration for Reasoning

SUMO

70000 axioms
open source

extracted fact:
X hasCapital Y
becomes:
(capitalCity Y X)

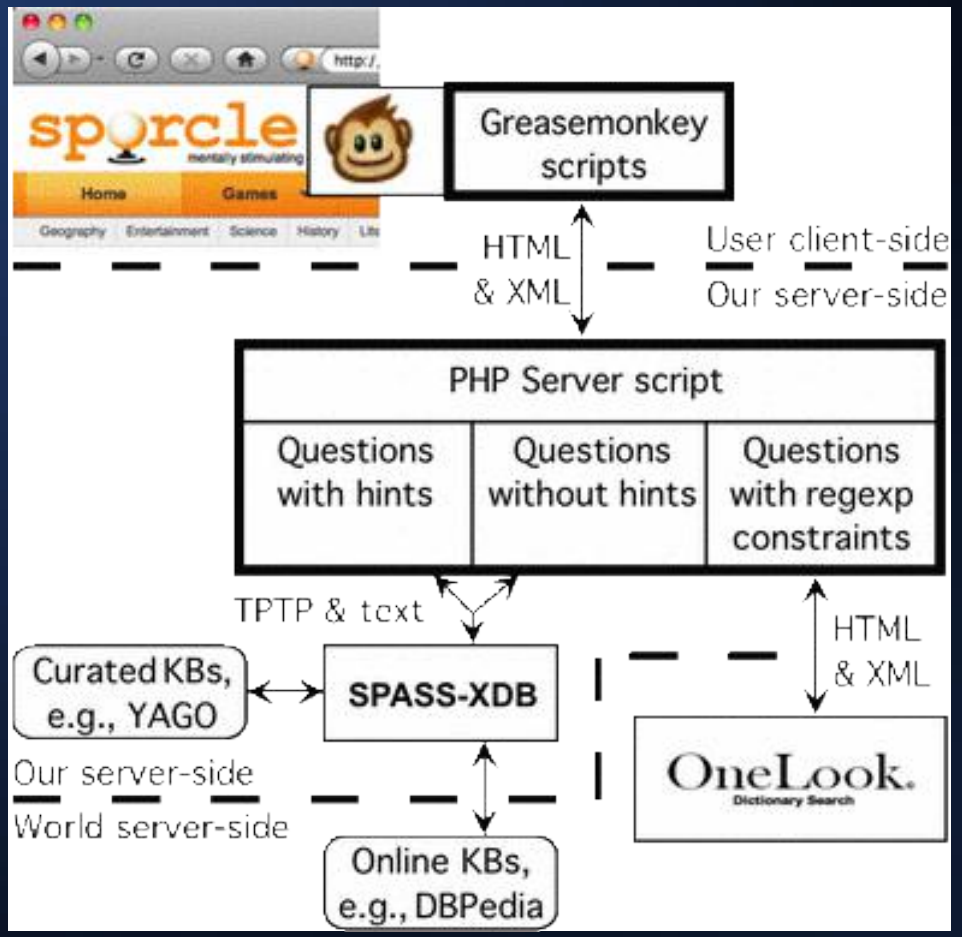
YAGO-SUMO

SUMO's axioms attached to real-world instances,

overall many millions of axioms

```
(exists ?YEARNO ?MONTHNO ?YEARNO
     (and
          (birthdate HerveyDeStanton
               (DayFn ?DAYNO
                    (MonthFn ?MONTHNO
                    (YearFn ?YEARNO))))
          (greaterThanOrEqualTo ?YEARNO 1270)
          (lessThanOrEqualTo ?YEARNO 1279)))
```

Smarter Question Answering



Sutcliffe et al.

Beyond Collections of Facts

- 1. Knowledge Integration and Organization
- 2. Natural Language Coupling
- 3. Multimodal and Cognitive Knowledge

Question Answering

?

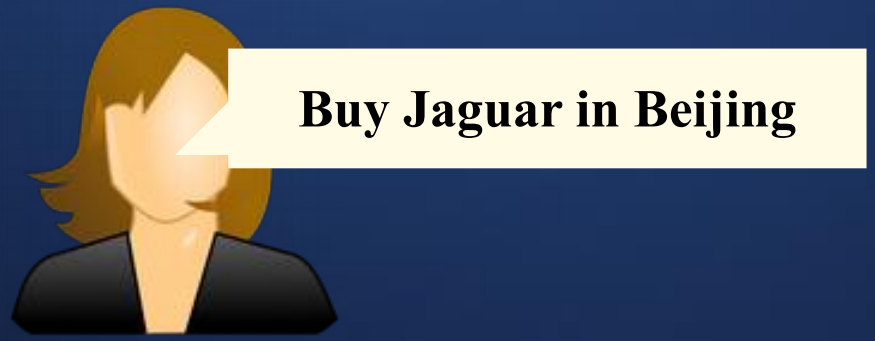
?

What is the capital of New Jersey?

?

?

Question Answering



Question Answering

Buy Jaguar in Beijing

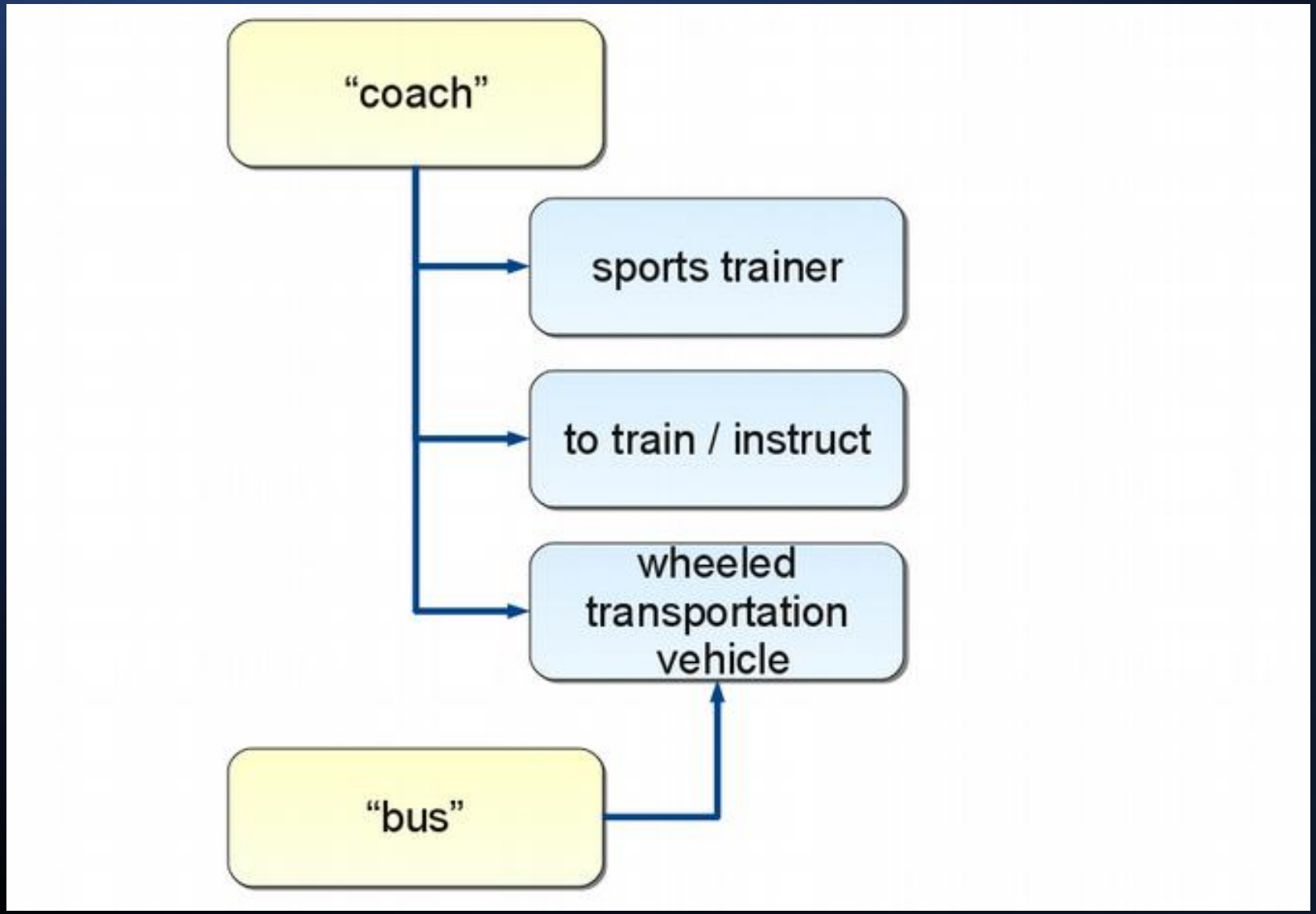
Jaguar

Which meaning?

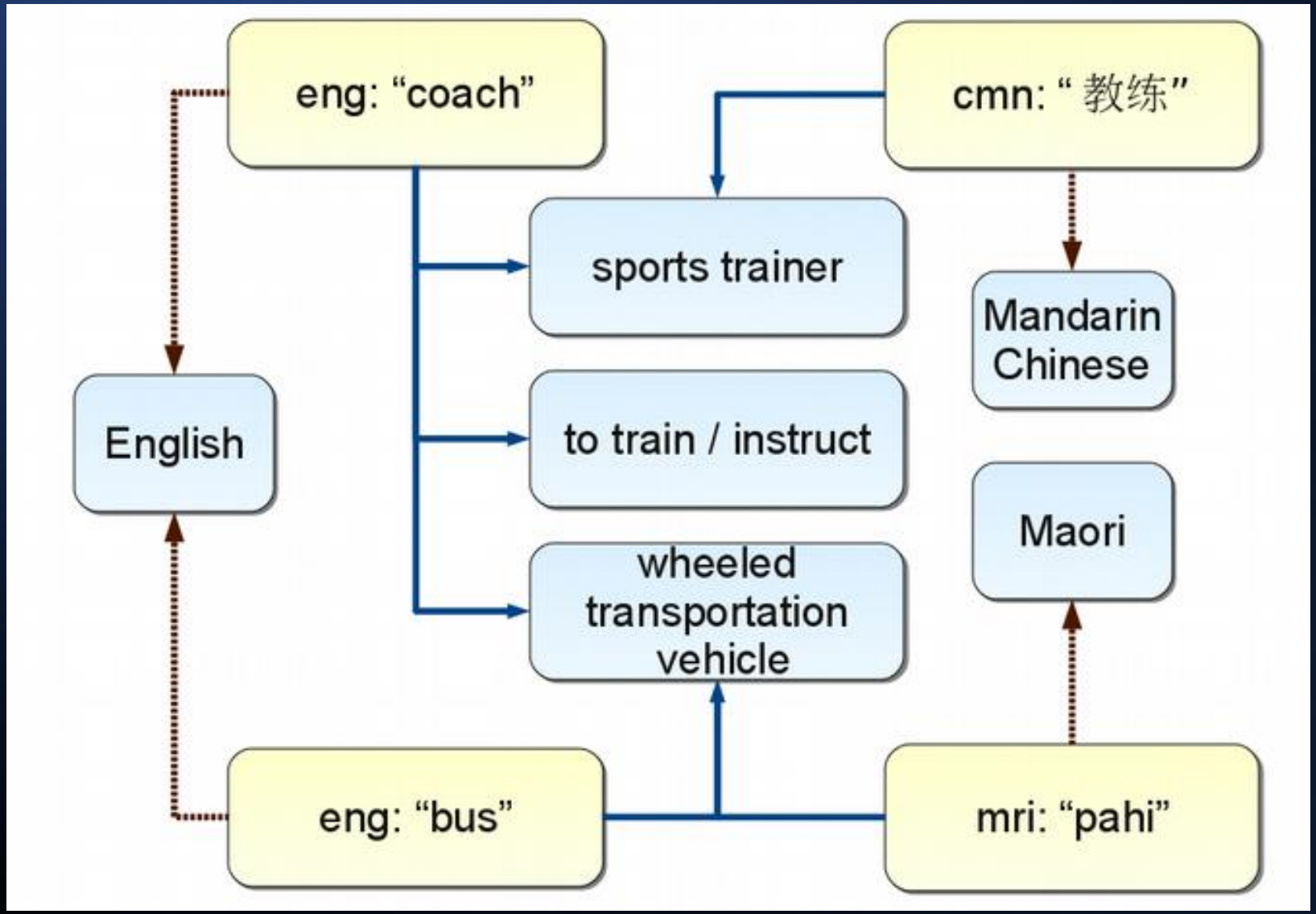
Question Answering

"Do Macs now support Windows?"

Lexical Knowledge Bases

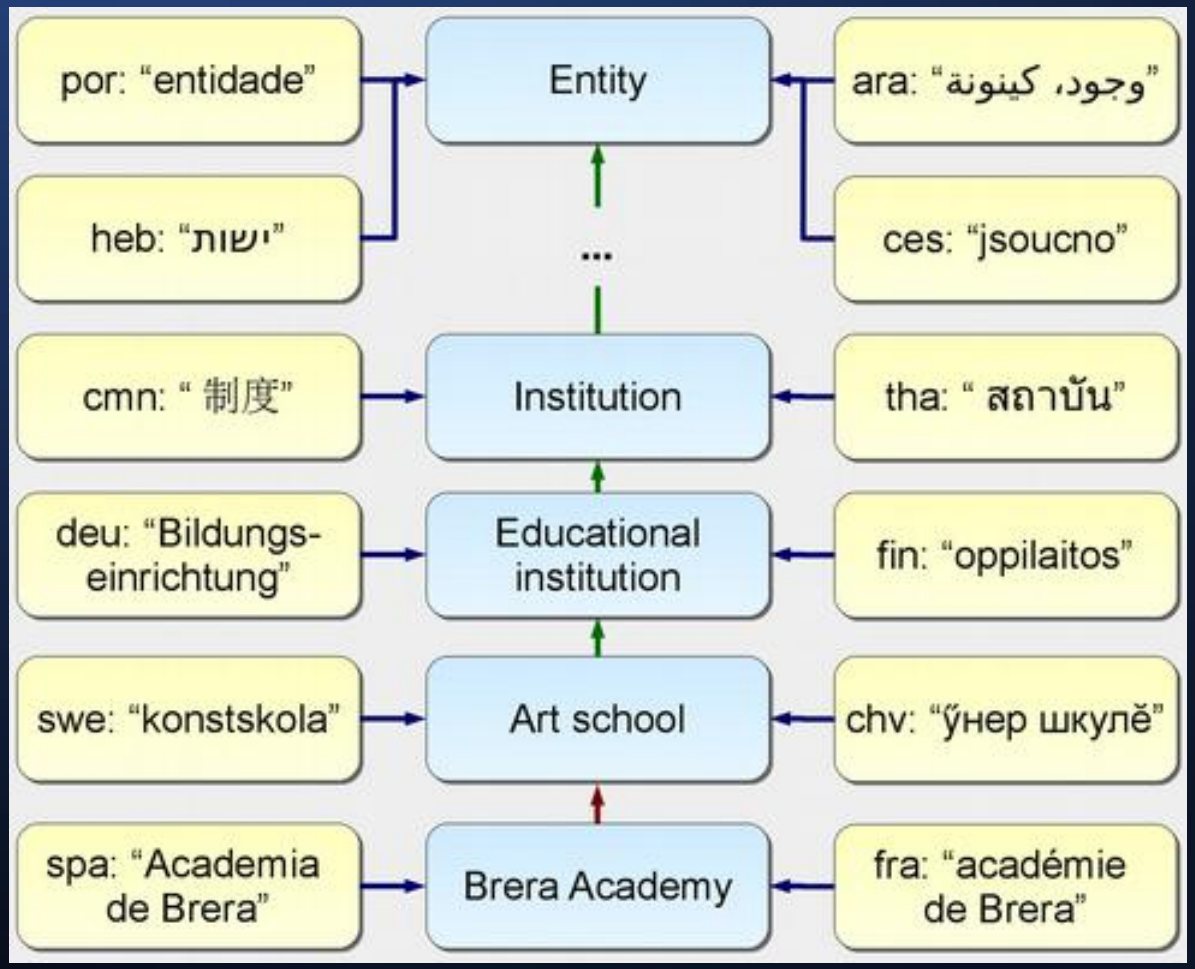


Multilingual Lexical Knowledge

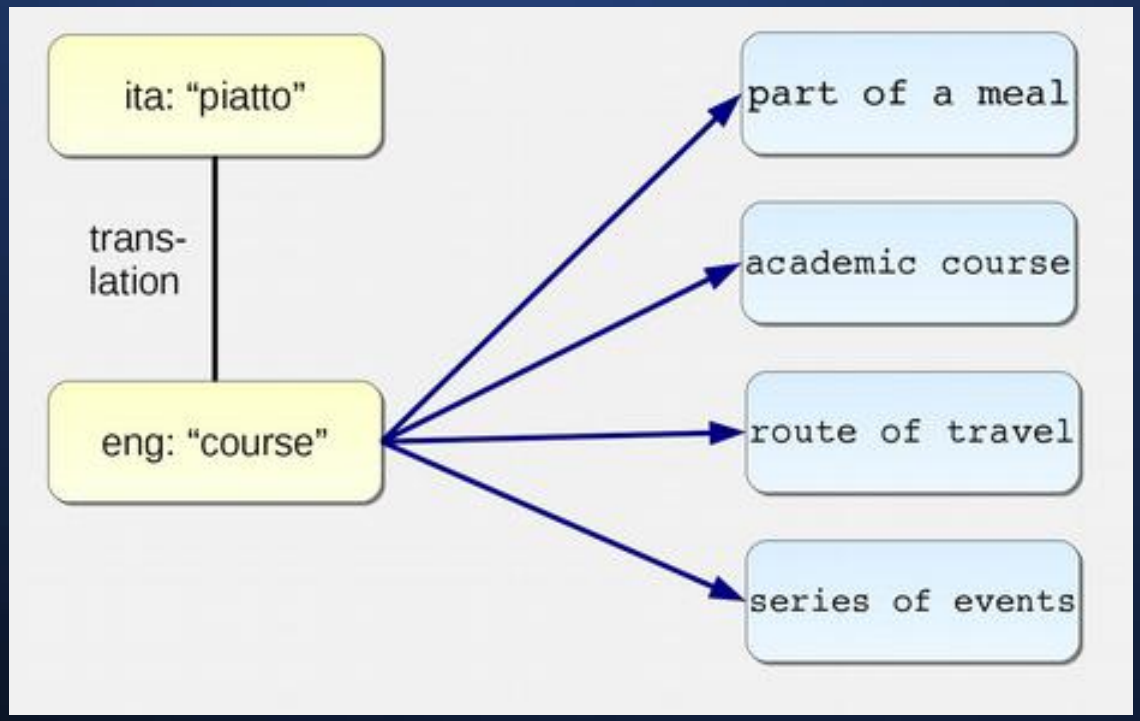


UWN (de Melo & Weikum 2009)

MENTA

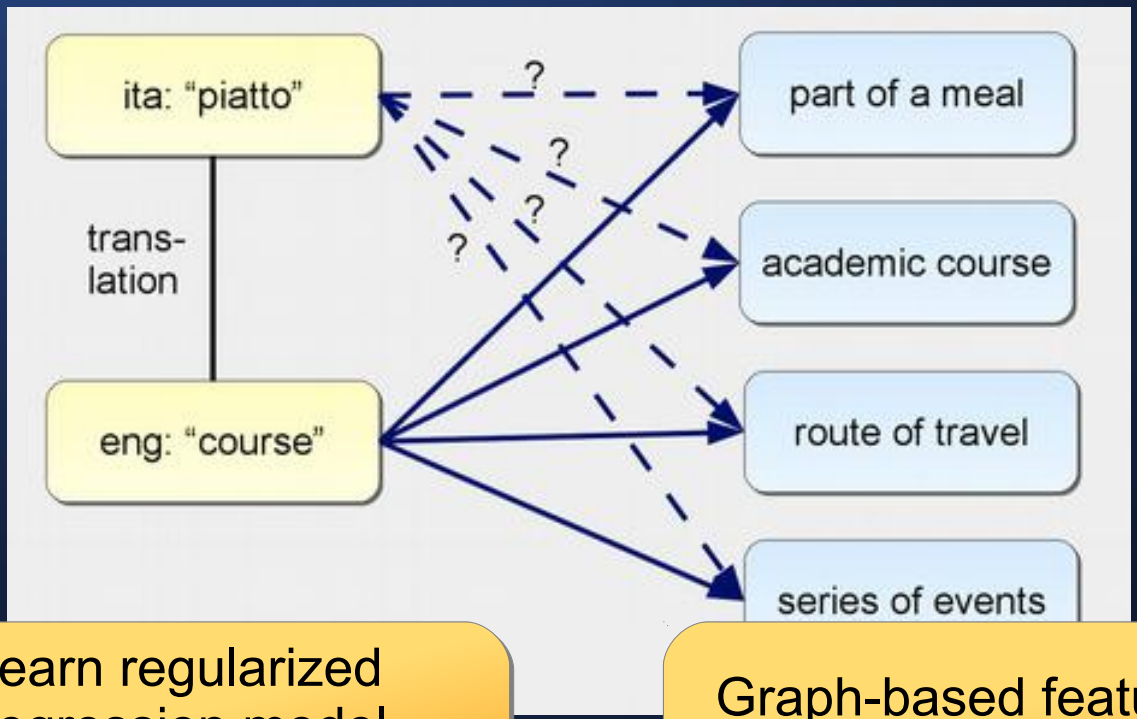

por: “entidade”
heb: “ישות”
Entity
...
ara: “وجود، كينونة”
ces: “jsoucno”
cmn: “ 制度”
Institution
tha: “ สถาบัน”
deu: “Bildungs-einrichtung”
Educational institution
fin: “oppilaitos”
swe: “konstskola”
Art school
chv: “ӳнер шкулĕ”
spa: “Academia de Brera”
Brera Academy
fra: “académie de Brera”

UWN: Universal Wordnet



Exploit millions of translations from the Web

UWN: “Universal Wordnet”

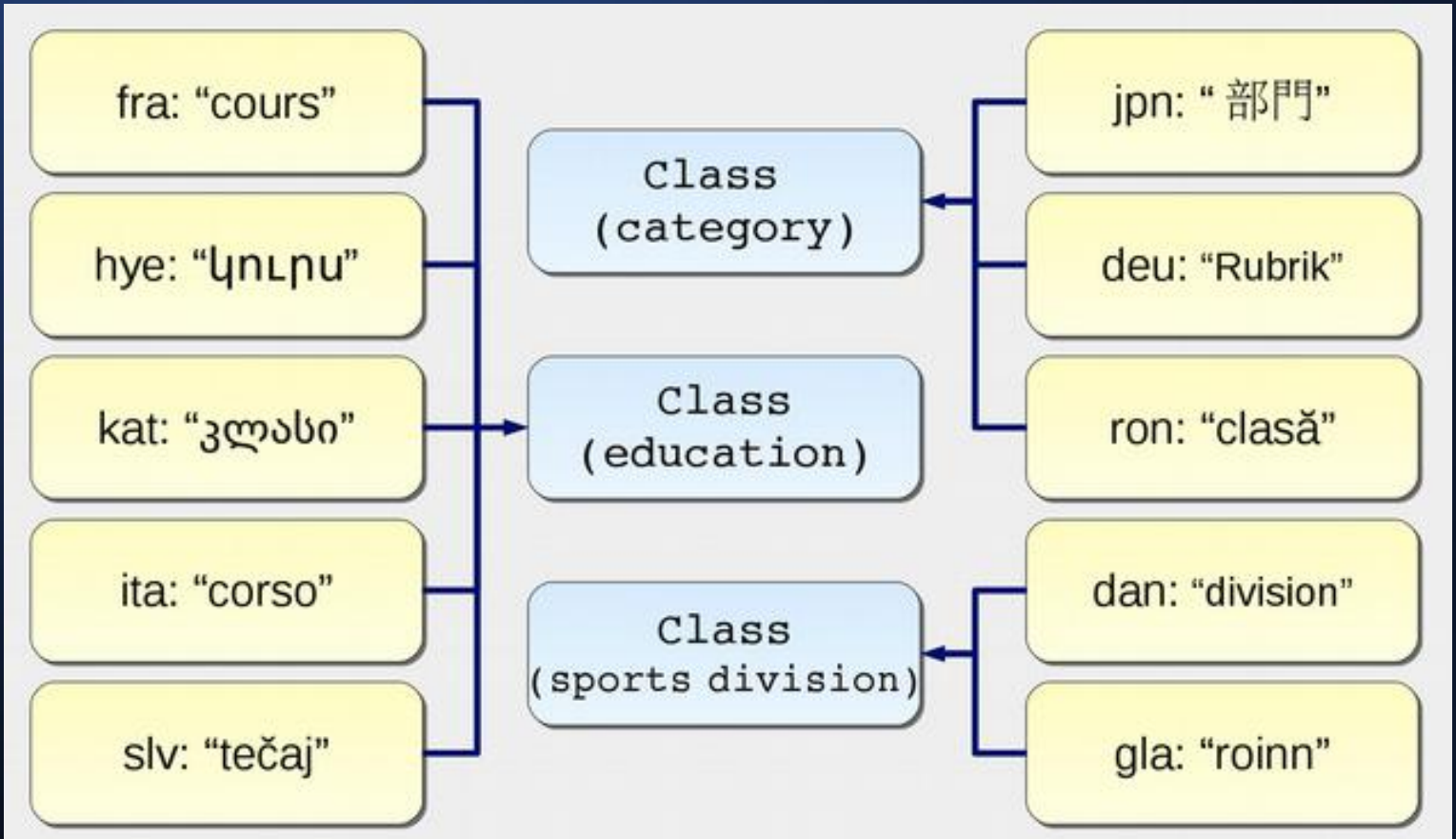


Learn regularized regression model to predict edge weights for a given pair of nodes

Graph-based features, weighted by various edge weight metrics

CIKM 2009

UWN: Universal Wordnet



over 1,000,000 words in over 100 languages

CIKM 2009

ICGL 2008
Best Paper Award

UWN: Getting Started

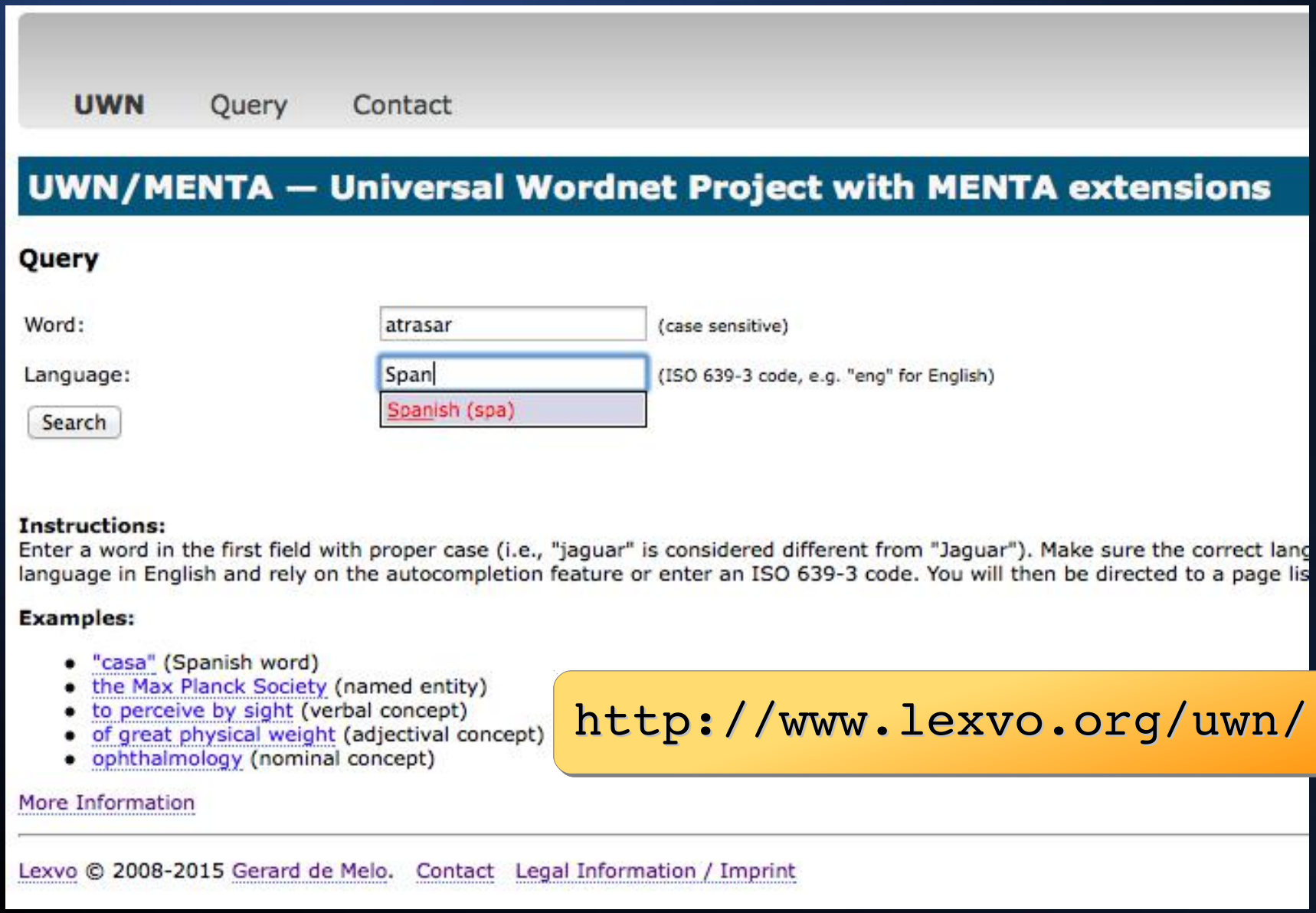
http://lexvo.org/uwn/

Simple API for JVM Languages

```
val uwn = new UWN(new File("plugins/"))
for (m <- uwn.getMeanings("souris", "fra"))
   println(m)
```

Or Just Download the TSV File

UWN

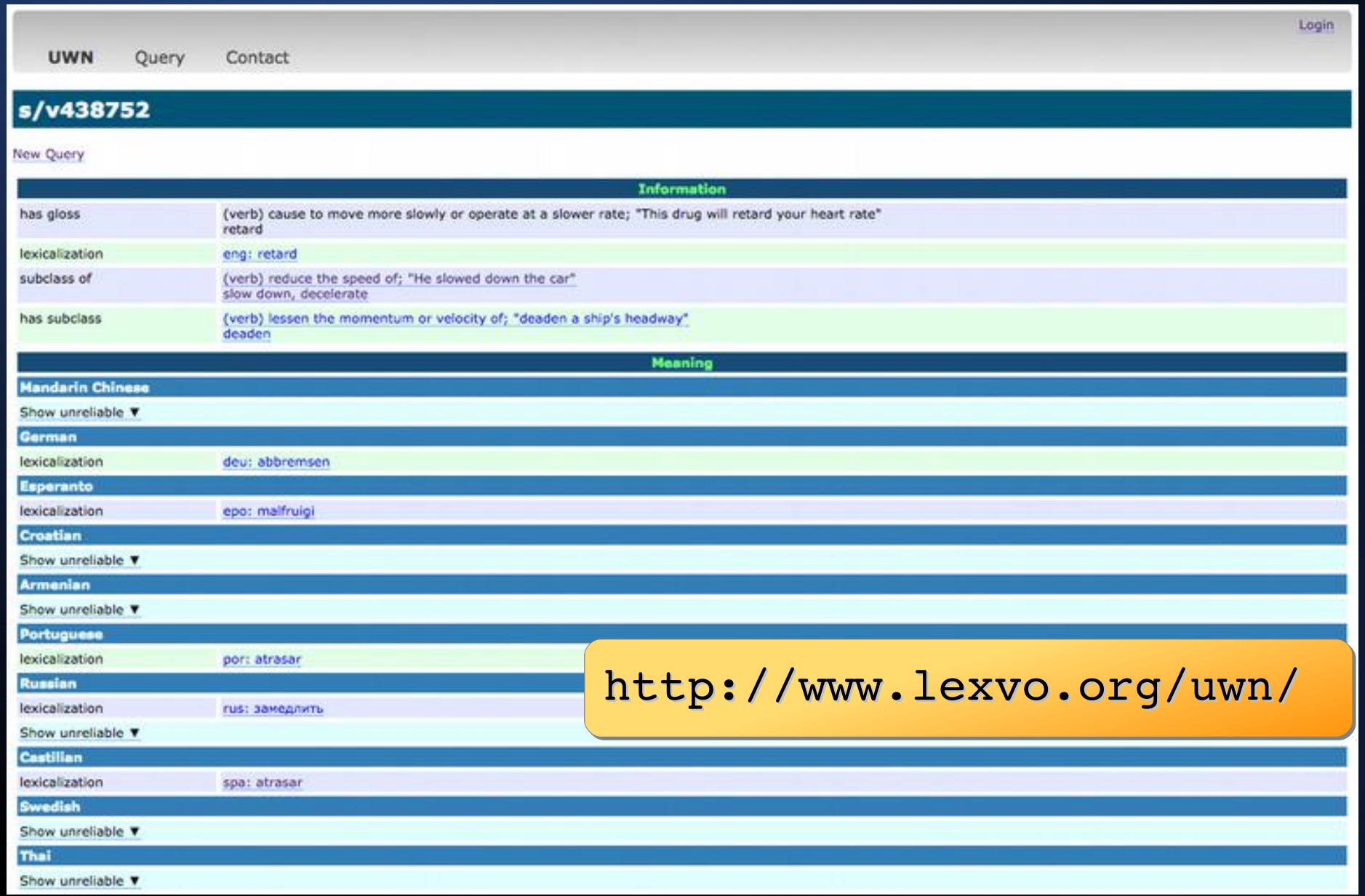


http://www.lexvo.org/uwn/

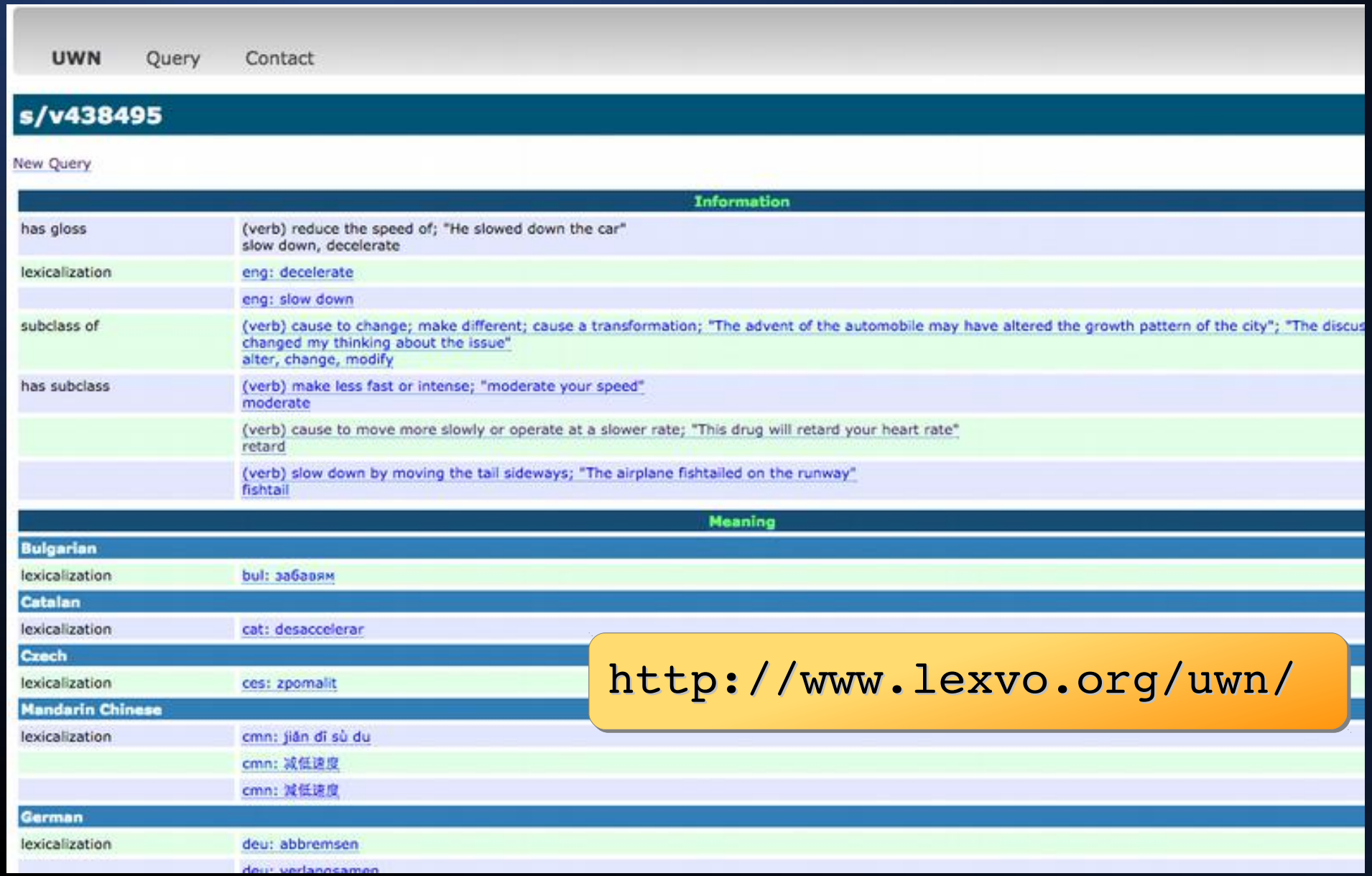
UWN

http://www.lexvo.org/uwn/

UWN


Login
UWN
Query
Contact
s/v438752
New Query
Information
has gloss
(verb) cause to move more slowly or operate at a slower rate; "This drug will retard your heart rate"
retard
lexicalization
eng: retard
subclass of
(verb) reduce the speed of; "He slowed down the car"
slow down, decelerate
has subclass
(verb) lessen the momentum or velocity of; "deaden a ship's headway"
deaden
Meaning
Mandarin Chinese
Show unreliable ▼
German
lexicalization
deu: abbremsen
Esperanto
lexicalization
epo: malfruigi
Croatian
Show unreliable ▼
Armenian
Show unreliable ▼
Portuguese
lexicalization
por: atrasar
Russian
lexicalization
rus: замедлить
Show unreliable ▼
Castilian
lexicalization
spa: atrasar
Swedish
Show unreliable ▼
Thai
Show unreliable ▼
http://www.lexvo.org/uwn/

UWN



UWN Query Contact

s/v438495

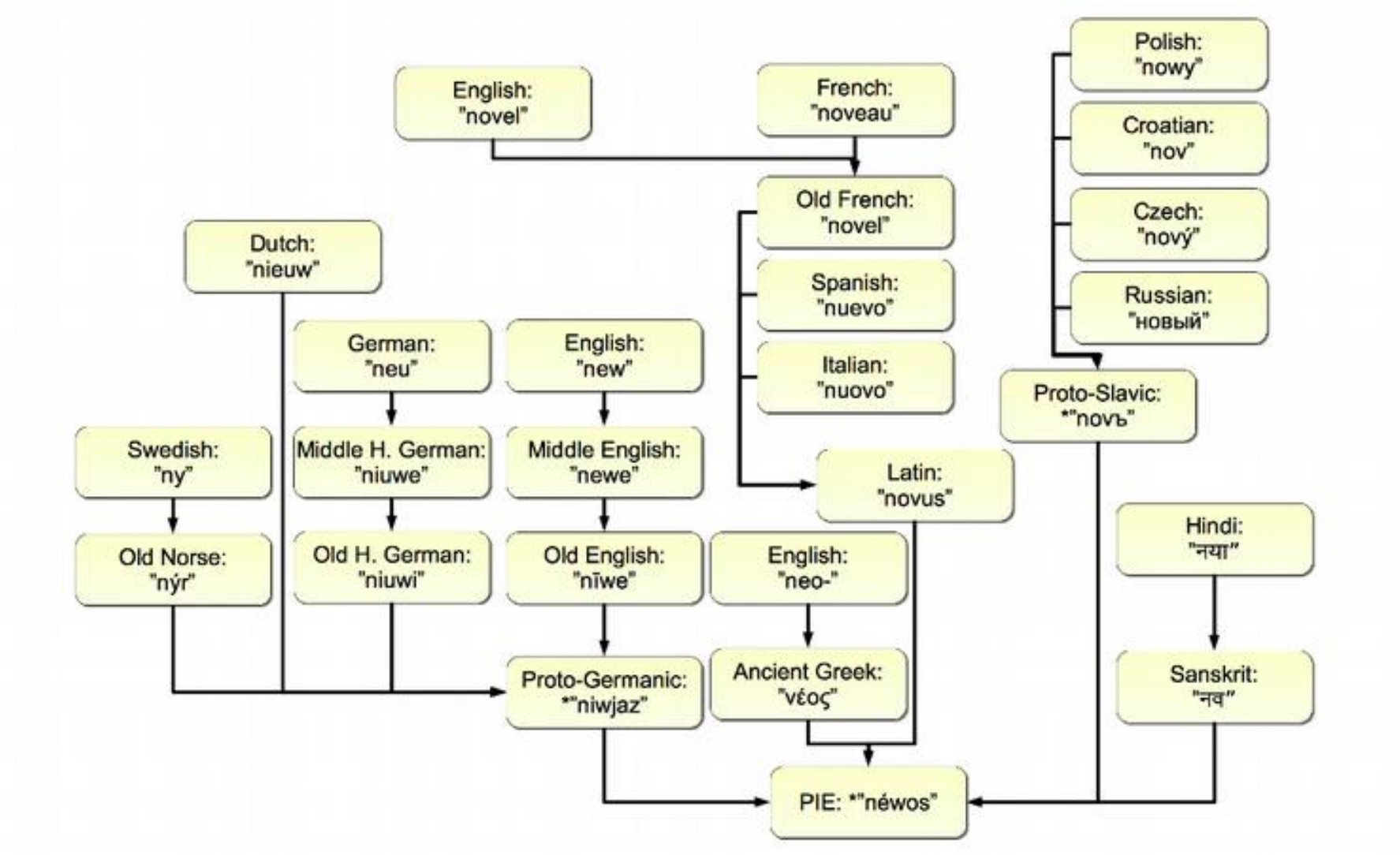
New Query

Information	
has gloss	(verb) reduce the speed of; "He slowed down the car" slow down, decelerate
lexicalization	eng: decelerate
	eng: slow down
subclass of	(verb) cause to change; make different; cause a transformation; "The advent of the automobile may have altered the growth pattern of the city"; "The discus changed my thinking about the issue" alter, change, modify
has subclass	(verb) make less fast or intense; "moderate your speed" moderate
	(verb) cause to move more slowly or operate at a slower rate; "This drug will retard your heart rate" retard
	(verb) slow down by moving the tail sideways; "The airplane fishtailed on the runway" fishtail

Meaning	
Bulgarian	
lexicalization	bul: забавям
Catalan	
lexicalization	cat: desaccelerar
Czech	
lexicalization	ces: zpomalit
Mandarin Chinese	
lexicalization	cmn: jiǎn dī sù du
	cmn: 减低速度
	cmn: 減低速度
German	
lexicalization	deu: abbremsen
	deu: verlangsamen

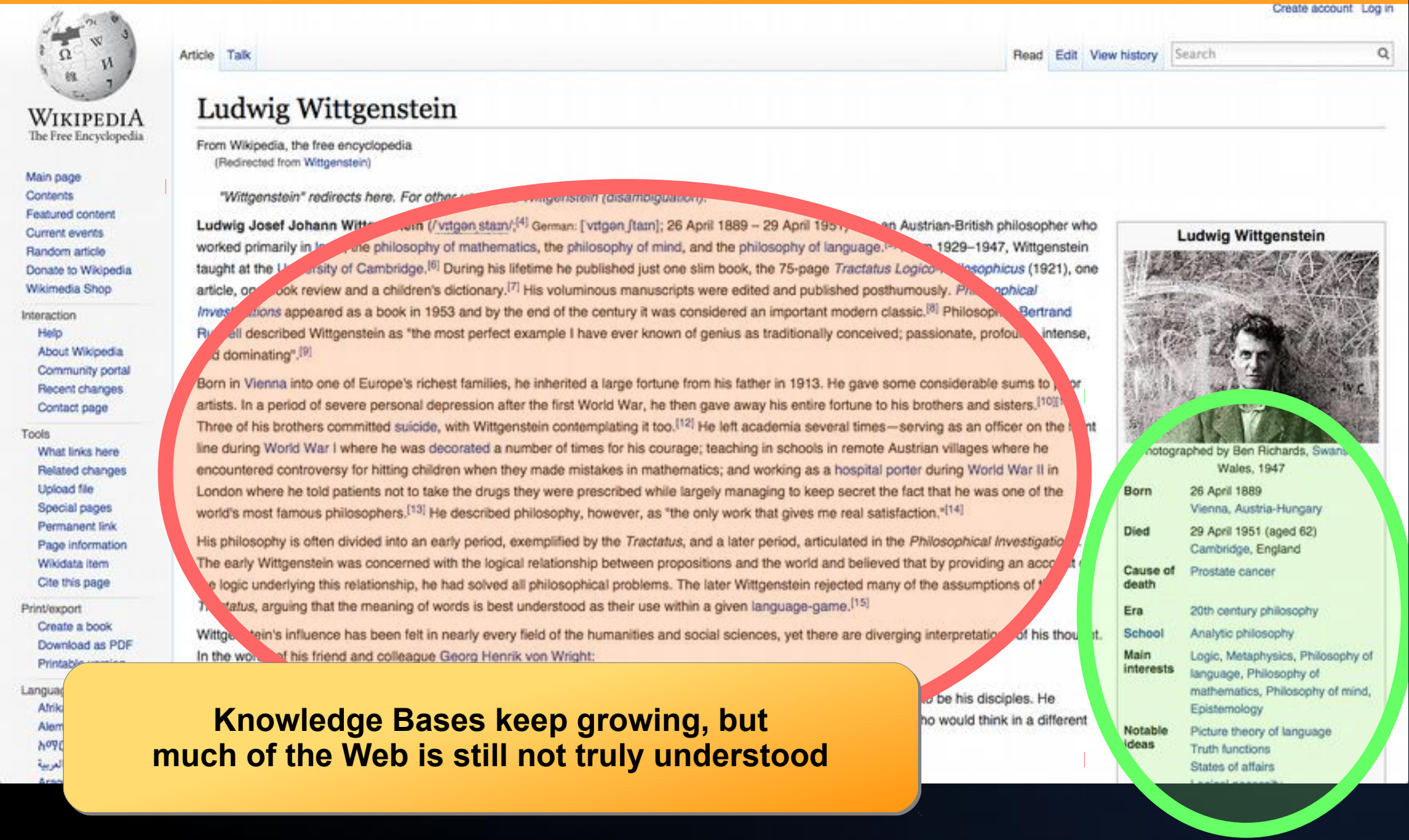
http://www.lexvo.org/uwn/

Etymological Wordnet



LREC 2014

Real Understanding?



Create account Log in

Article Talk Read Edit View history Search

WIKIPEDIA
The Free Encyclopedia

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

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

Ludwig Wittgenstein

From Wikipedia, the free encyclopedia
(Redirected from Wittgenstein)

"Wittgenstein" redirects here. For other ... (disambiguation).

Ludwig Josef Johann Wit... (/ˈvɪtgən staɪn/;[4] German: [ˈvɪtgən ʃtaɪn]; 26 April 1889 – 29 April 195...) an Austrian-British philosopher who worked primarily in ... the philosophy of mathematics, the philosophy of mind, and the philosophy of language.[...] 1929–1947, Wittgenstein taught at the U...ersity of Cambridge.[6] During his lifetime he published just one slim book, the 75-page *Tractatus Logico-...osophicus* (1921), one article, on... book review and a children's dictionary.[7] His voluminous manuscripts were edited and published posthumously. *Phi...ophical Investi...tions* appeared as a book in 1953 and by the end of the century it was considered an important modern classic.[8] Philosop... Bertrand Ru...ell described Wittgenstein as "the most perfect example I have ever known of genius as traditionally conceived; passionate, profou... intense, ...d dominating".[9]

Born in Vienna into one of Europe's richest families, he inherited a large fortune from his father in 1913. He gave some considerable sums to ...or artists. In a period of severe personal depression after the first World War, he then gave away his entire fortune to his brothers and sisters.[10][...] Three of his brothers committed suicide, with Wittgenstein contemplating it too.[12] He left academia several times—serving as an officer on the ...nt line during World War I where he was decorated a number of times for his courage; teaching in schools in remote Austrian villages where he encountered controversy for hitting children when they made mistakes in mathematics; and working as a hospital porter during World War II in London where he told patients not to take the drugs they were prescribed while largely managing to keep secret the fact that he was one of the world's most famous philosophers.[13] He described philosophy, however, as "the only work that gives me real satisfaction."[14]

His philosophy is often divided into an early period, exemplified by the *Tractatus*, and a later period, articulated in the *Philosophical Investigatio...* The early Wittgenstein was concerned with the logical relationship between propositions and the world and believed that by providing an acco...t ...e logic underlying this relationship, he had solved all philosophical problems. The later Wittgenstein rejected many of the assumptions of t... *...tatus*, arguing that the meaning of words is best understood as their use within a given language-game.[15]

Wittg...tein's influence has been felt in nearly every field of the humanities and social sciences, yet there are diverging interpretati... of his thou...t. In the wo... of his friend and colleague Georg Henrik von Wright:

...o be his disciples. He ...ho would think in a different

Ludwig Wittgenstein

...hotographed by Ben Richards, Swans... Wales, 1947

Born	26 April 1889 Vienna, Austria-Hungary
Died	29 April 1951 (aged 62) Cambridge, England
Cause of death	Prostate cancer
Era	20th century philosophy
School	Analytic philosophy
Main interests	Logic, Metaphysics, Philosophy of language, Philosophy of mathematics, Philosophy of mind, Epistemology
Notable ideas	Picture theory of language Truth functions States of affairs

Knowledge Bases keep growing, but much of the Web is still not truly understood

Relation Understanding

Equivalent:

MetaWeb was acquired by Google.
MetaWeb was just recently acquired by Google.
MetaWeb, surprisingly, was acquired by Google.

MetaWeb was bought out by Google.
Google bought MetaWeb.
Google acquired MetaWeb.
MetaWeb was sold to Google.
Google's acquisition of MetaWeb.
Google's MetaWeb acquisition.
and so on...

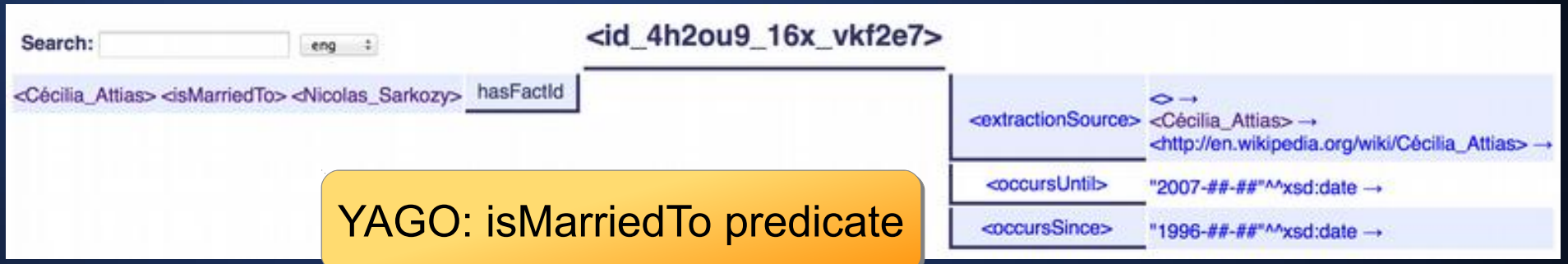
Relation Understanding: N-Ary Relations

- Capture the “who-did-what-to-whom”
- Microsoft bought the patent from Nokia.
  Nokia sold the patent to Microsoft.
  The patent was acquired by Microsoft [from Nokia].
  The patent was sold [by Nokia] to Microsoft.

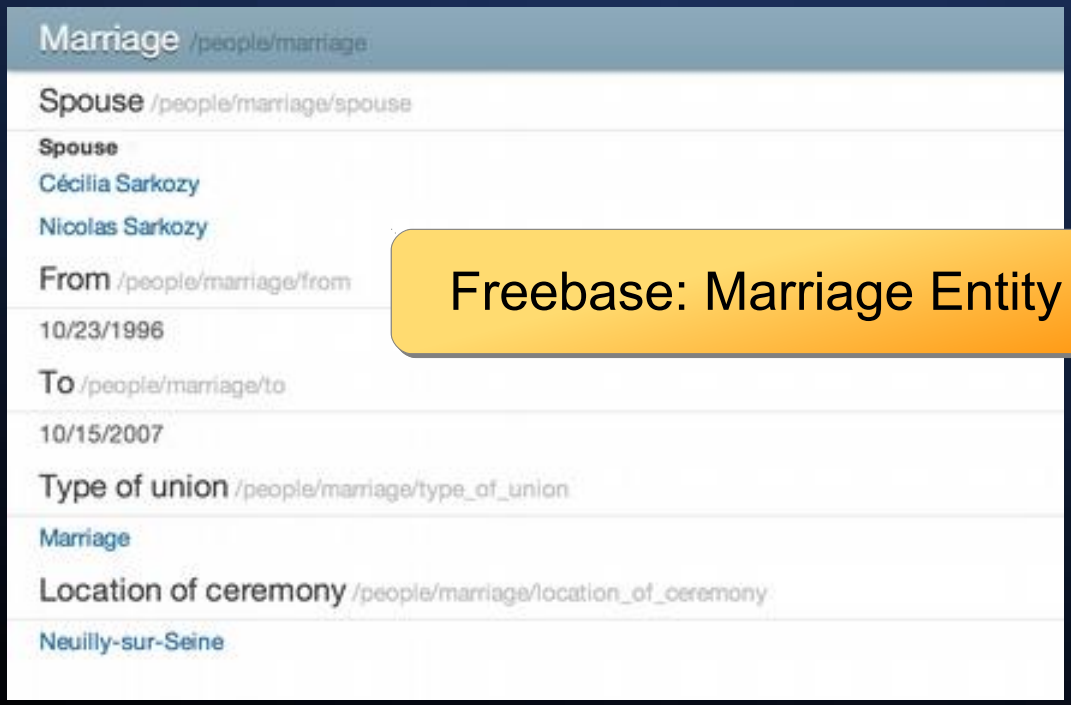
Underlying frame:
Commercial transfer

Buyer:	Microsoft
Seller:	Nokia
Product:	The patent

Relation Integration



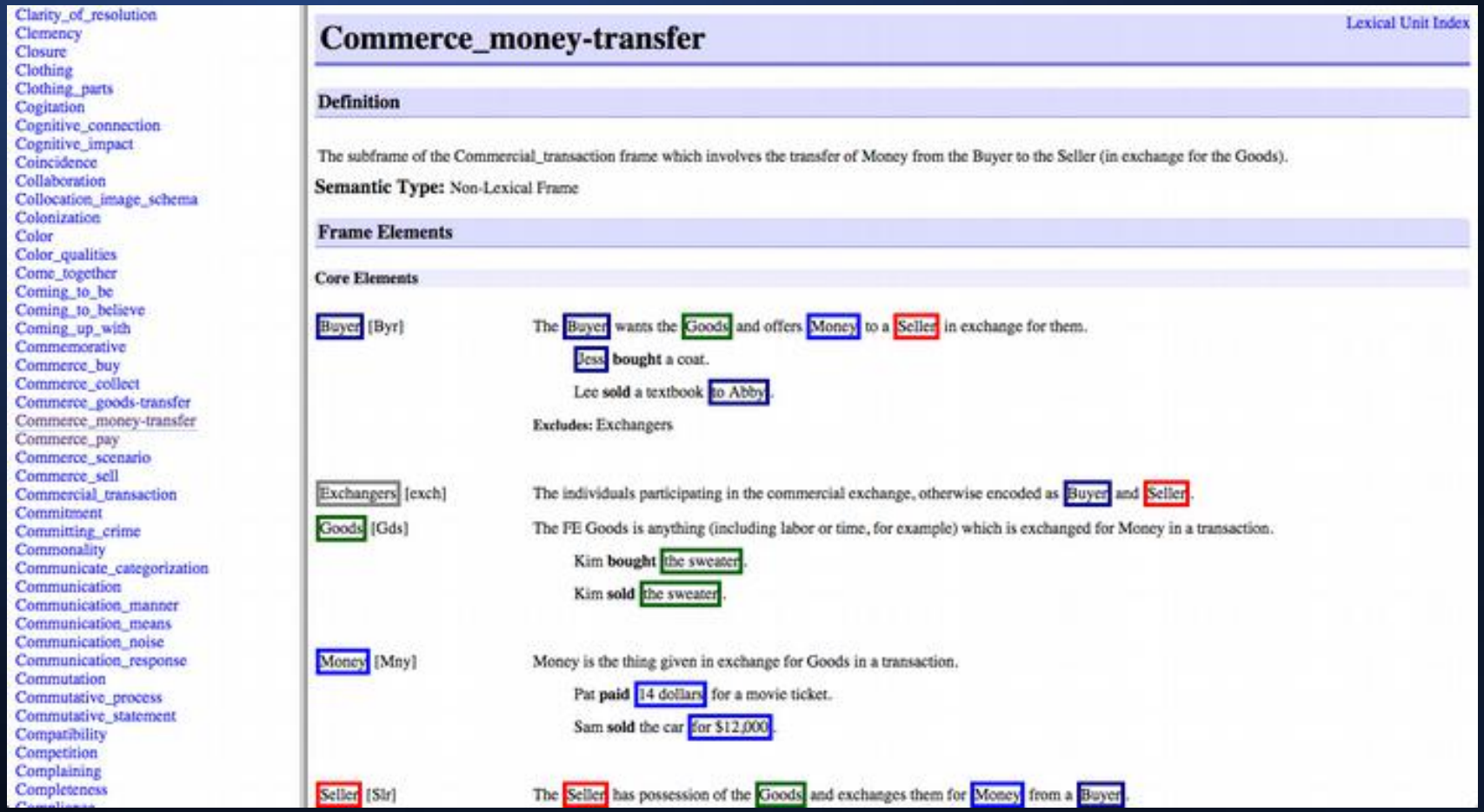
YAGO: isMarriedTo predicate



Freebase: Marriage Entity

Challenge: Modelling Differences

Relation Integration: FrameBase.org



Bringing knowledge into a standard form based on natural language (FrameNet)

Background: Berkeley FrameNet Project

Beyond Collections of Facts

- 1. Knowledge Integration and Organization
- 2. Natural Language Coupling
- 3. Multimodal and Cognitive Knowledge

MENTA: Media

The pools of Pamukkale

More details

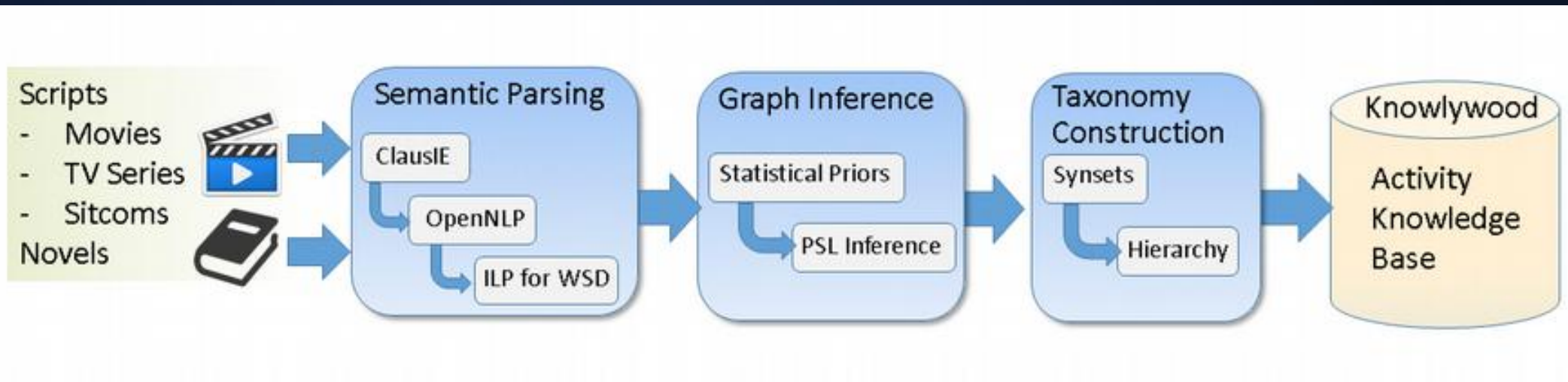
Knowlywood: Human Activities

EXT. SMALL MOUNTAIN--DAY

Wichita charges up the rockage of a small mountain-hill-type thing. The image repeats itself over and over--each time Wichita is more sweaty, gasping, sneering.

Wichita (V.O.)
The rules forbid anyone from the climbing the camp's mountain.

align via subtitles



CIKM 2015

Knowlywood: Human Activities

Go up an	elevation
..	..

Parent activity

Previous activity

Next activity

{Climb up a mountain	, Hike up a hill}
Participating Agent	climber, boy, rope
Location	camp, forest, sea shore
Time	daylight, holiday
Visuals	

Get to	village
..	..

Drink	water
..	..

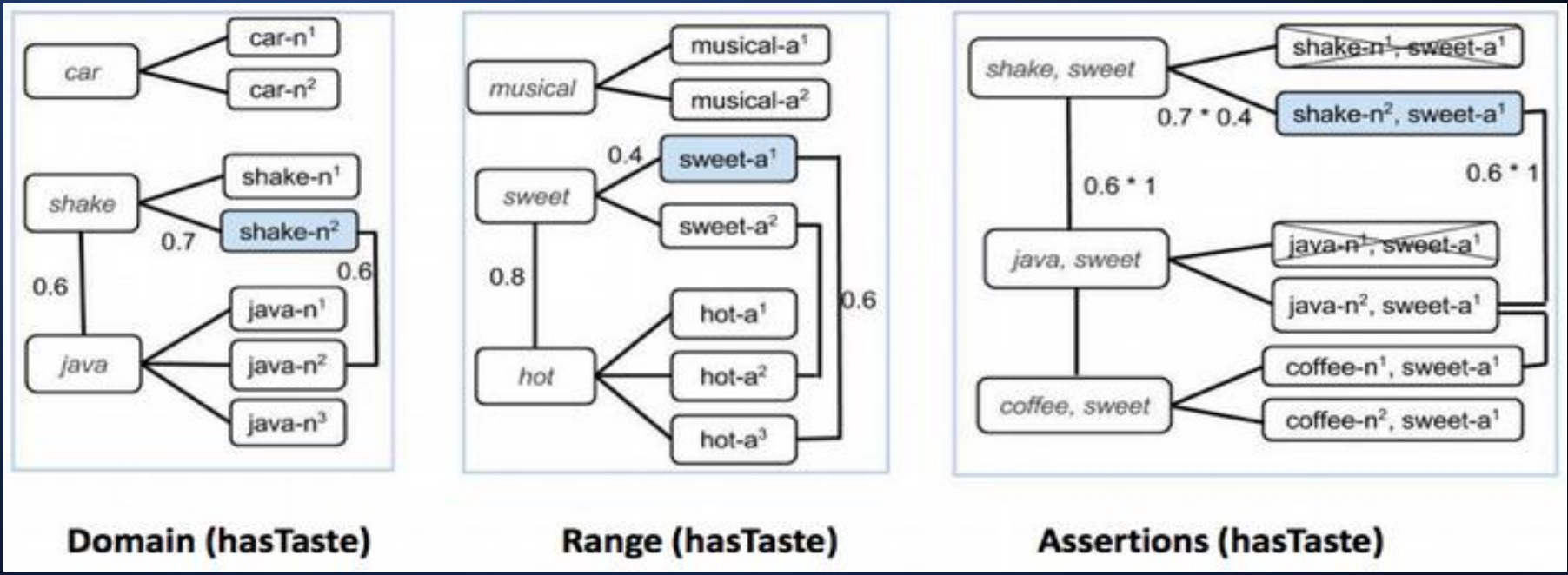
Learning Common-Sense

WebChild

AAAI 2014
WSDM 2014
AAAI 2011

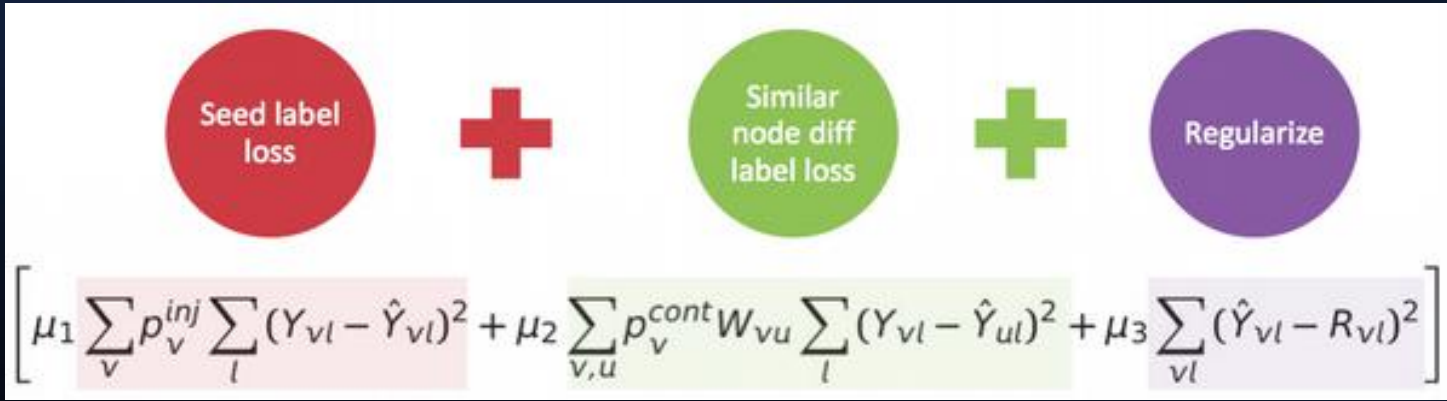
Online: http://bit.ly/webchild

Learning Common-Sense



WebChild

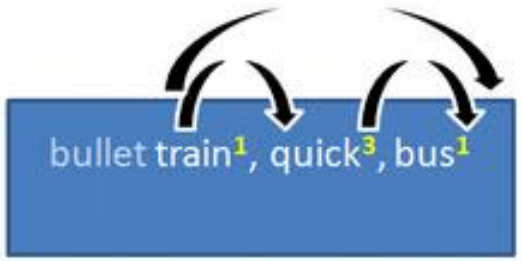
(Tandon et al. WSDM 2014):



Online: http://bit.ly/webchild

Comparative Knowledge

Joint Disambiguation:
<bullet train, quicker than, bus>
<train, slower than, plane>
<plane, faster than, train>
<bus, slower than, plane>



plane[2] ,slow[1] , bus[1]
has a neighbor
bus[1] , slow[1] , plane[2]

slow:
penalize
>1 senses

bus:
penalize
>1 senses

maximize

$$\sum_i \sum_j (\text{coh}_{ij} + \tau_{ij}) x_{ij} + \sum_i \sum_j \sum_k \sum_l \text{sim}_{ij,kl} B_{ij,kl} - \sum_{m \in \text{adjectives}} \sum_s a_{ms} - \sum_{m \in \text{nouns}} \sum_s n_{ms}$$

subject to

$$\ldots$$

$$x_{ij} = x_{kl} \quad \forall i,j,k,l : \mu_{ij,kl} = 1$$

train[1] ,slow[1] , plane[2]
≡
plane[2] , fast[1] , train[1]

Online: http://bit.ly/webchild

AAAI 2014

ImpLex: Implicative Commitments

Netflix **admitted** that they had developed a prototype.

VSoft **pretended** that they had developed a prototype.

Did they develop a prototype?

de Melo & de Paiva (2014) based on Lauri Kartunnen's work

ImpLex: Implicative Commitments

de Melo & de Paiva 2014

It is **confusing** that Apple has acquired the technology.

It is **unlikely** that Apple has acquired the technology.

Have they acquired it?

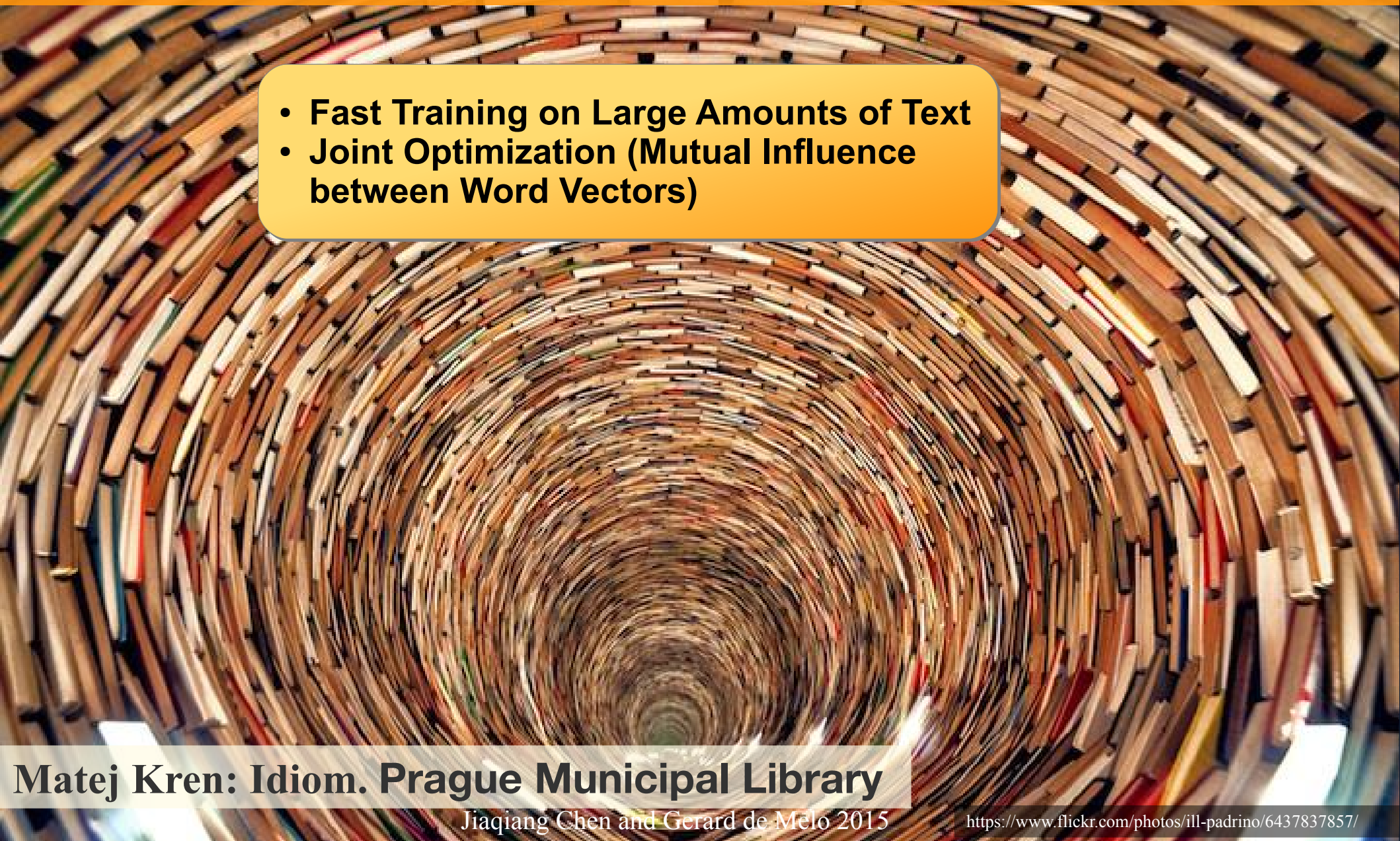
?

Word Vectors

http://www.wikihow.com/Read-a-Book-to-a-Baby-or-Infant#/Image:Read-a-Book-to-a-Baby-or-Infant-Step-5.jpg

Why are Word Vectors Successful?

- **Fast Training on Large Amounts of Text**
- **Joint Optimization (Mutual Influence between Word Vectors)**



Matej Kren: Idiom. Prague Municipal Library

Jiaqiang Chen and Gerard de Melo 2015

https://www.flickr.com/photos/ill-padrino/6437837857/

What Goes into Word Vectors?

The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

What Goes into Word Vectors?

The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

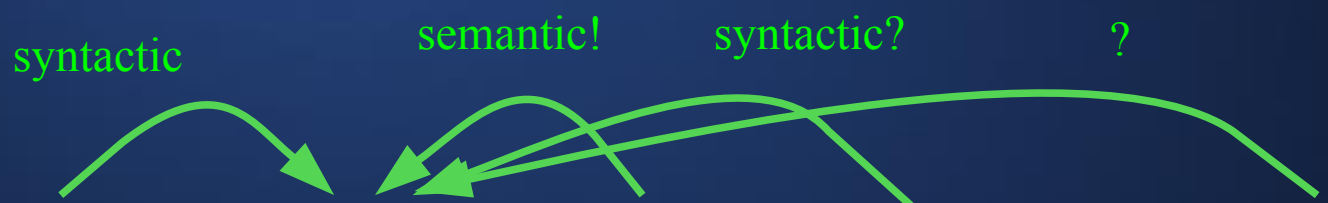
What Goes into Word Vectors?

The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

What Goes into Word Vectors?

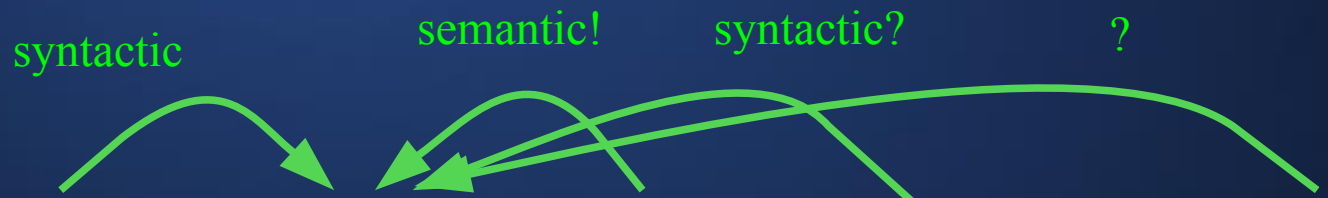
The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

What Goes into Word Vectors?



The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

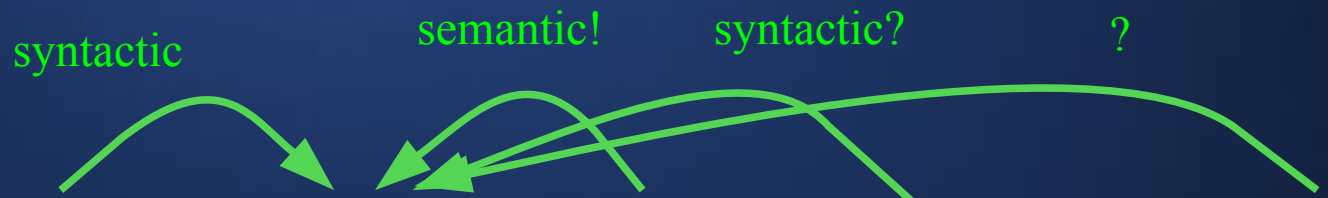
What Goes into Word Vectors?



The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

Word2Vec Solution:
Subsampling

What Goes into Word Vectors?

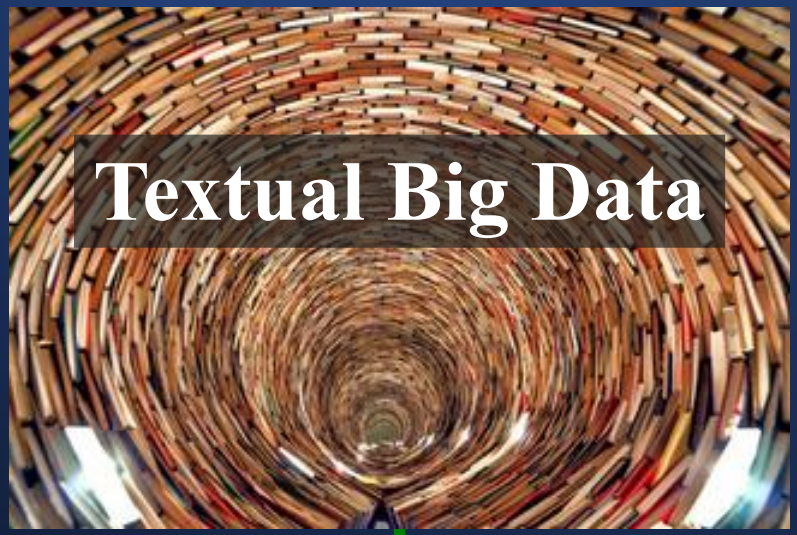


The Roman Empire was remarkably multicultural, with ”a rather astonishing cohesive capacity” to create a sense of shared identity while encompassing diverse peoples within its political system over a long span of time.

Textual Big Data

Image: https://500px.com/photo/96738645/street-bookstore-by-pablo-tarrero

Better Word Vectors by Exploiting Knowledge



Joint Training

Better Word Embeddings

Best Paper Award at NAACL 2015 Vector Space Modeling Workshop

Jiaqiang Chen and Gerard de Melo 2015

Multilingual Word Vectors

Combine word2vec/GloVe Vectors
with multilingual lexical data from Wiktionary

RG65	Chandar A P et al. (2014) En-Es Vectors	0.629 @ 55.4%
	Ours (word2vec)	0.834 @ 100.0%
	Ours (Glove)	0.809 @ 100.0%
MC30	Chandar A P et al. (2014) En-Es Vectors	0.430 @ 60.0%
	Faruqui et al. (2015)	0.591 @ N/A
	Ours (word2vec)	0.811 @ 76.7%
	Ours (Glove)	0.848 @ 76.7%
WS353	Chandar A P et al. (2014) En-Es Vectors	0.256 @ 65.1%
	Ours (word2vec)	0.548 @ 65.6%
	Ours (Glove)	0.591 @ 65.6%

Available soon at
http://gerard.demelo.org/representations/
Contact me!

Multilingual Word Vectors

Combine word2vec/GloVe Vectors
with multilingual lexical data from Wiktionary

JI65	Chandar A P et al. (2014) En-Fr Vectors	0.586 @ 49.2%
	Faruqui et al. (2015)	0.606 @ N/A
	Ours (word2vec)	0.793 @ 96.9%
	Ours (Glove)	0.824 @ 96.9%

Available soon at
http://gerard.demelo.org/representations/
Contact me!

Multilingual Analogical Reasoning

Coffee is to Starbucks
as ... is to Lipton?

System: Tea

Source vectors: GLOVE

What we should be able to get in the future

What is this all about?

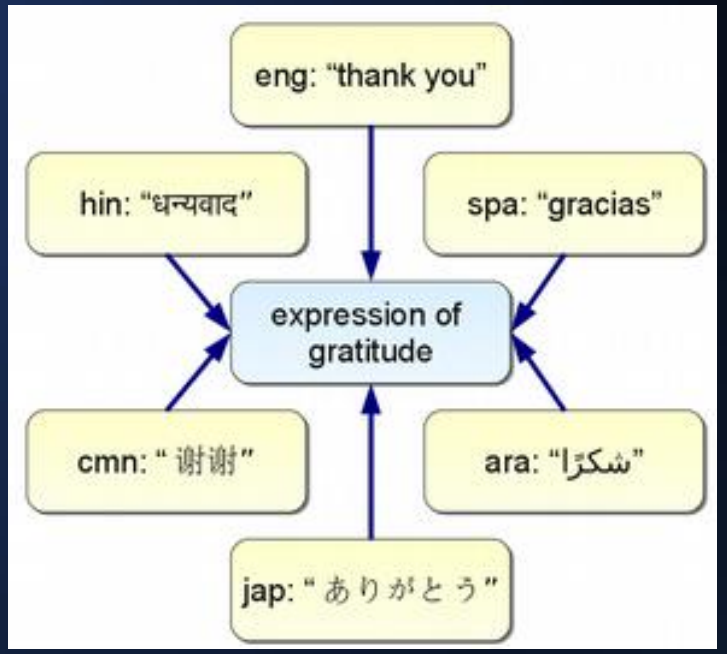
Huffington Post

This is about the US budget debt ceiling, which needs to be extended. Republican Senate majority leader Mitch McConell would like to...

Image: Microsoft Cortana

Summary

- **Knowledge Integration and Organization**
  - ► MENTA taxonomy
  - ► Lexvo.org Linked Data
- **Natural Language Coupling**
  - ► Universal WordNet
  - ► FrameBase.org
- **Multimodal and Cognitive Knowledge**
  - ► Knowlywood
  - ► WebChild



More Information:

www.demelo.org
gdm@demelo.org