

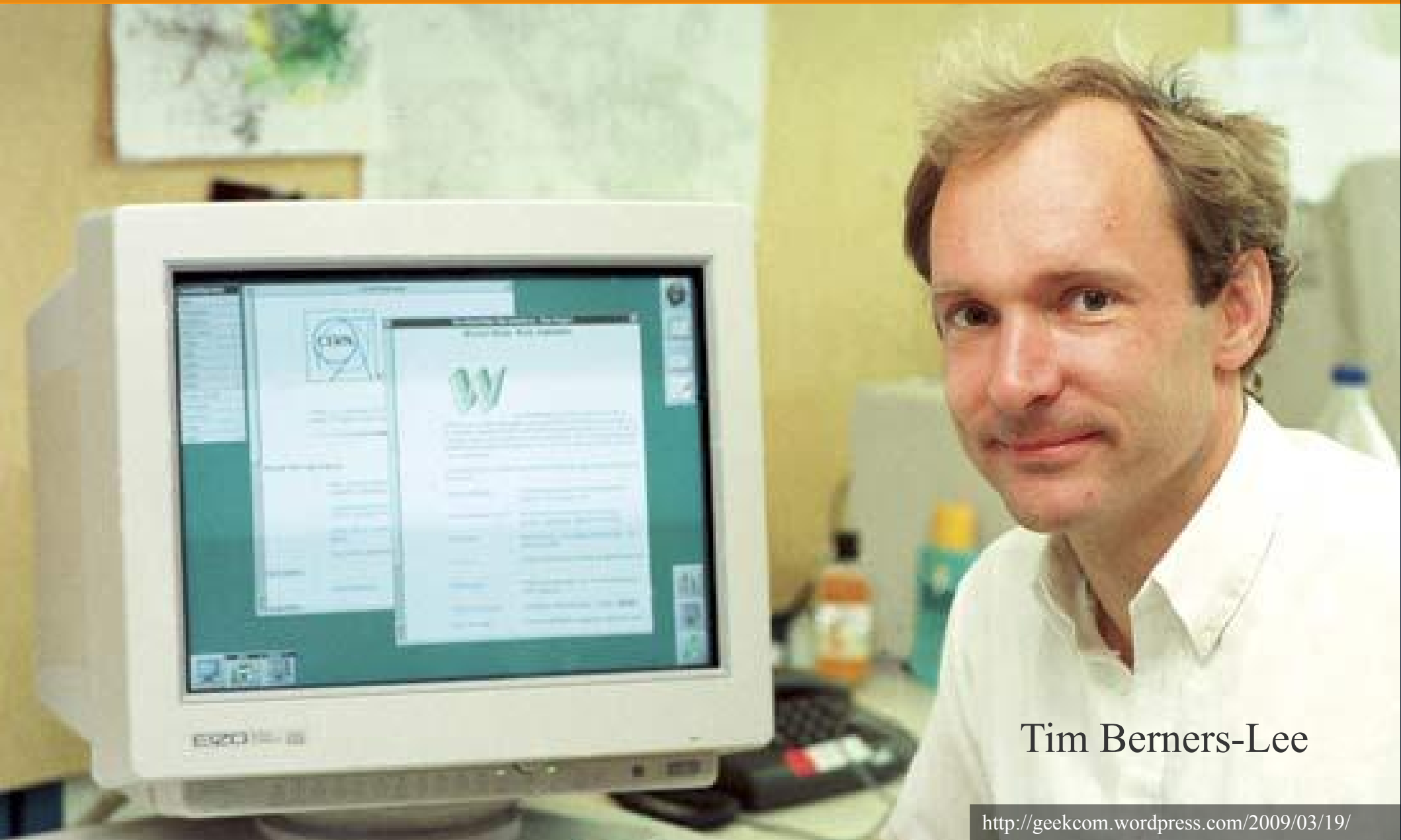
The Future of Knowledge Graphs: Moving beyond Simple Collections of Facts

Gerard de Melo

Assistant Professor, Tsinghua University

<http://gerard.demelo.org>

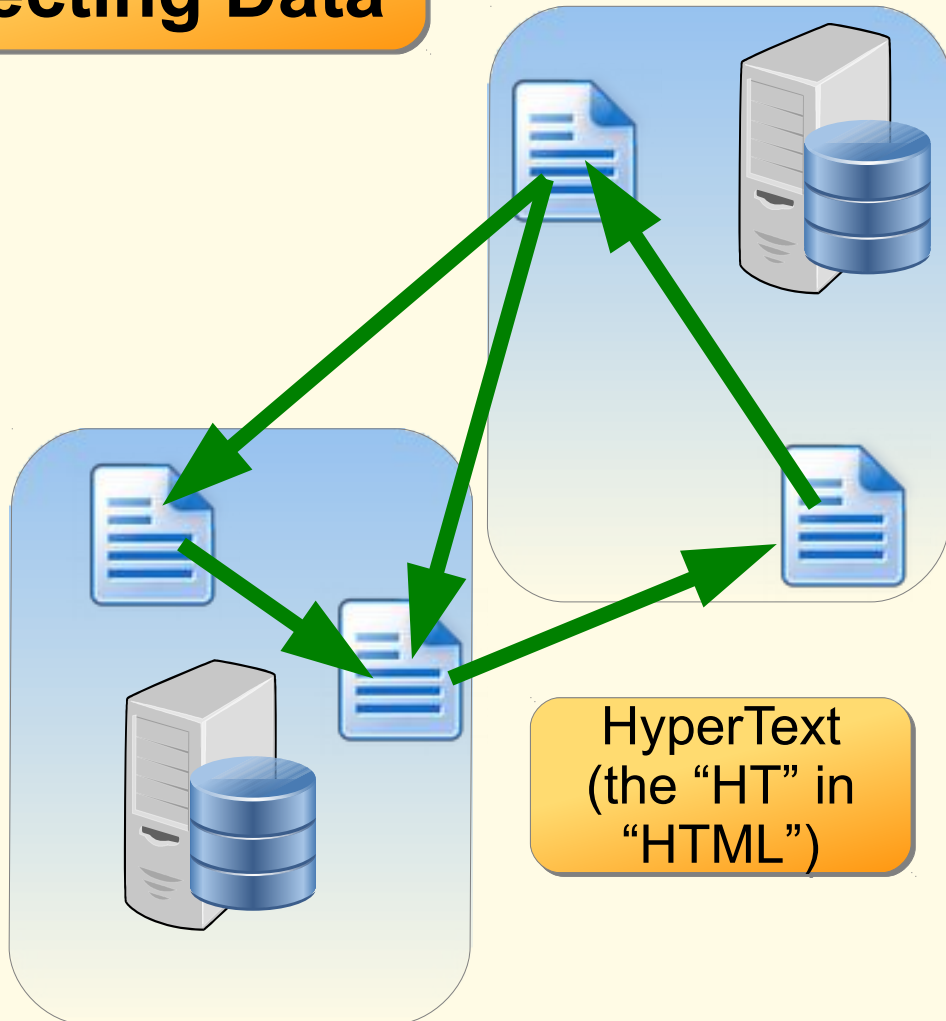
25 Years of the World Wide Web: 1989–2014



Tim Berners-Lee

25 Years of the World Wide Web: 1989–2014

**Basic Idea:
Connecting Data**



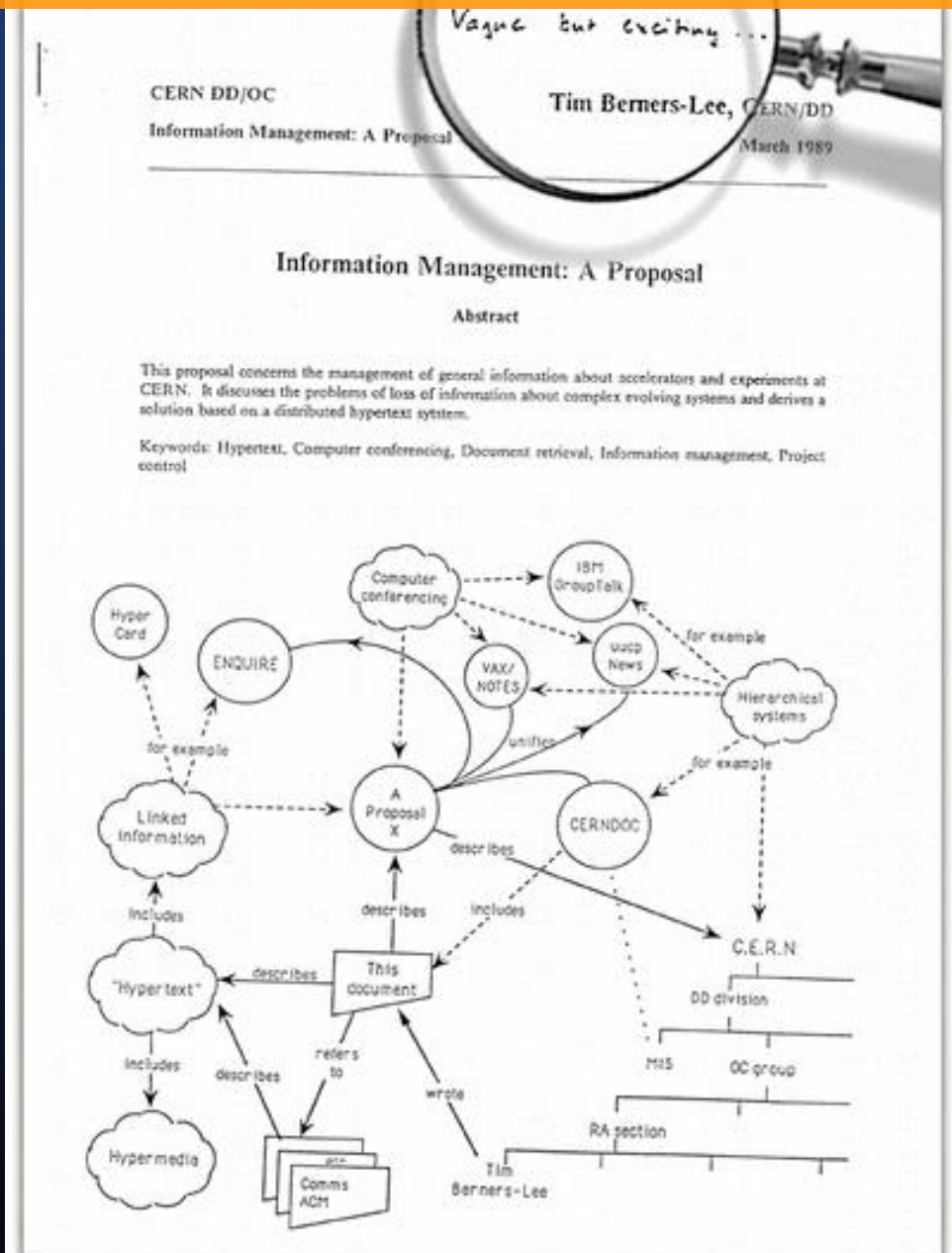
Tim Berners-Lee

25 Years of the World Wide Web: 1989–2014

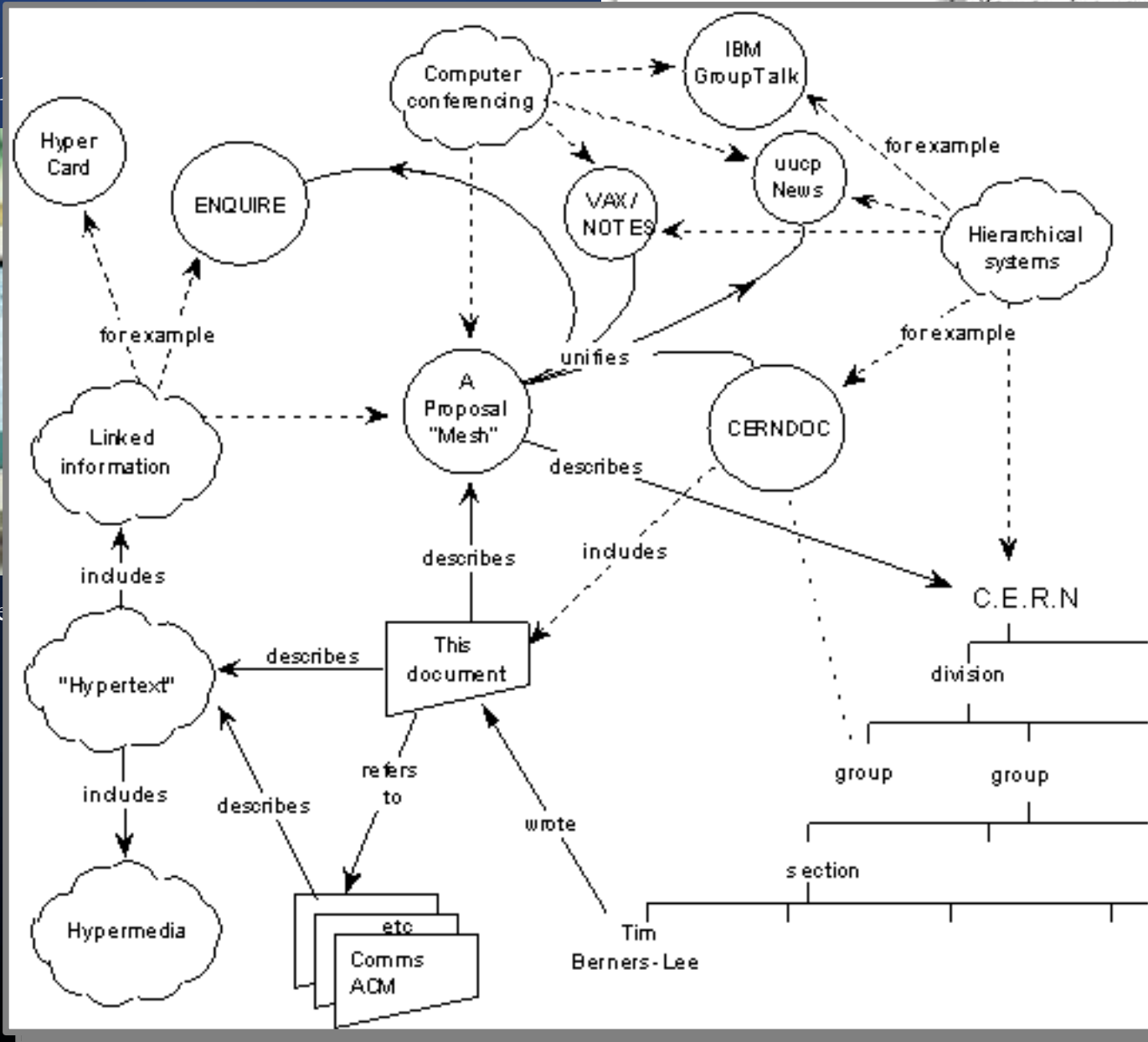
Tim Berners-Lee



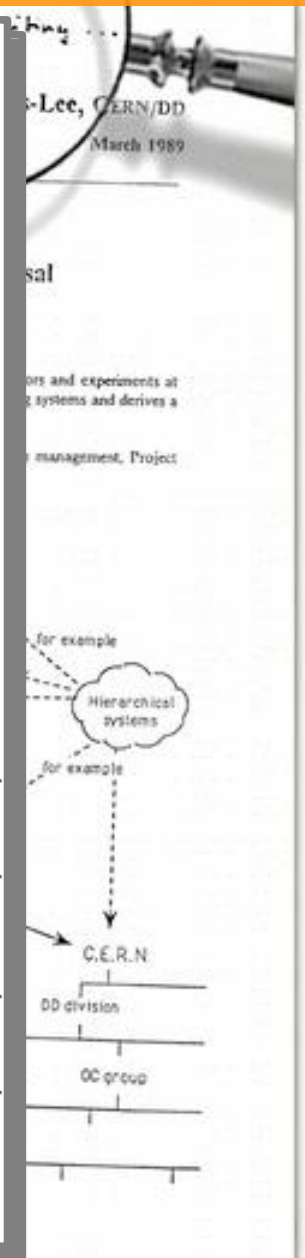
<http://geekcom.wordpress.com/2009/03/19/>



25 Years of the World Wide Web: 1989-2014



<http://ge>



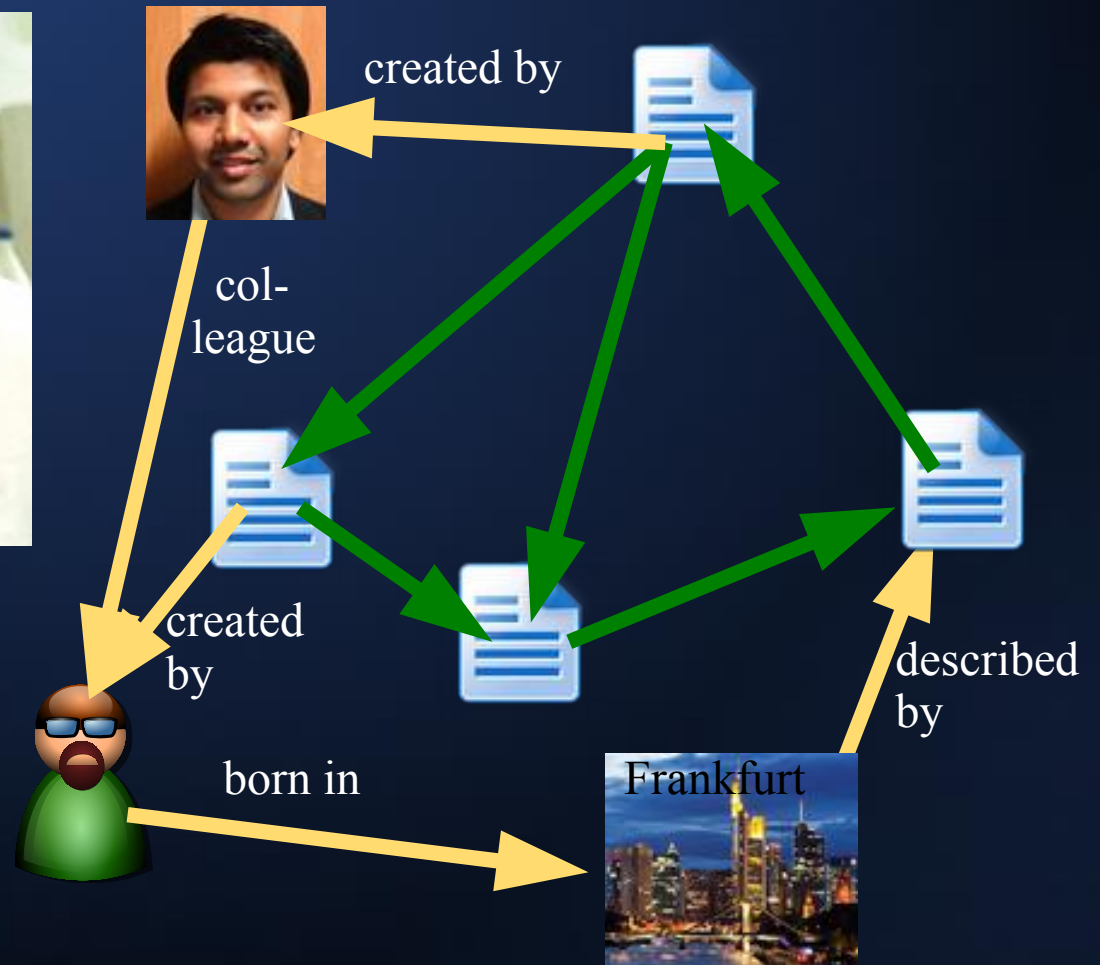
The Semantic Web

Tim Berners-Lee



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

A Web of
machine-readable
entity-relationship
data



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!

Big Knowledge Collections



WIKIPEDIA
The Free Encyclopedia

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

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page


Print/export
Create a book
Download as PDF
Printable version

Languages 
Afrikaans
Alemannisch

Create account Not logged in Talk Contributions Log in

Article Talk

Read Edit View history

Search 

Max Planck

From Wikipedia, the free encyclopedia

"Planck" redirects here. For other uses, see [Planck \(disambiguation\)](#).

Max Karl Ernst Ludwig Planck, FRS^[2] (/ˈplɑːnk/^[3] German: [ˈplɔŋk]; 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]

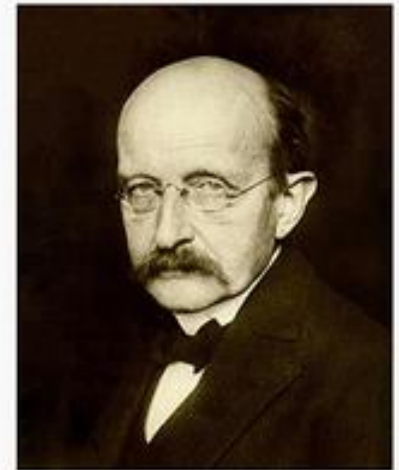
Planck made many contributions to theoretical physics, but his fame as a physicist rests primarily on his role as an originator of the quantum theory. However, his name is also known on a broader academic basis, through the renaming in 1948 of the German scientific institution, the Kaiser Wilhelm Society (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]

- Early life and career
 - Academic career
 - Family
 - Professor at Berlin University
 - Black-body radiation
 - Einstein and the theory of relativity
 - World War I
 - Post War and Weimar Republic
 - Quantum mechanics
 - Nazi dictatorship and World War II
- Religious views
- Publications
- See also
- References
- Sources
- External links

Max Planck



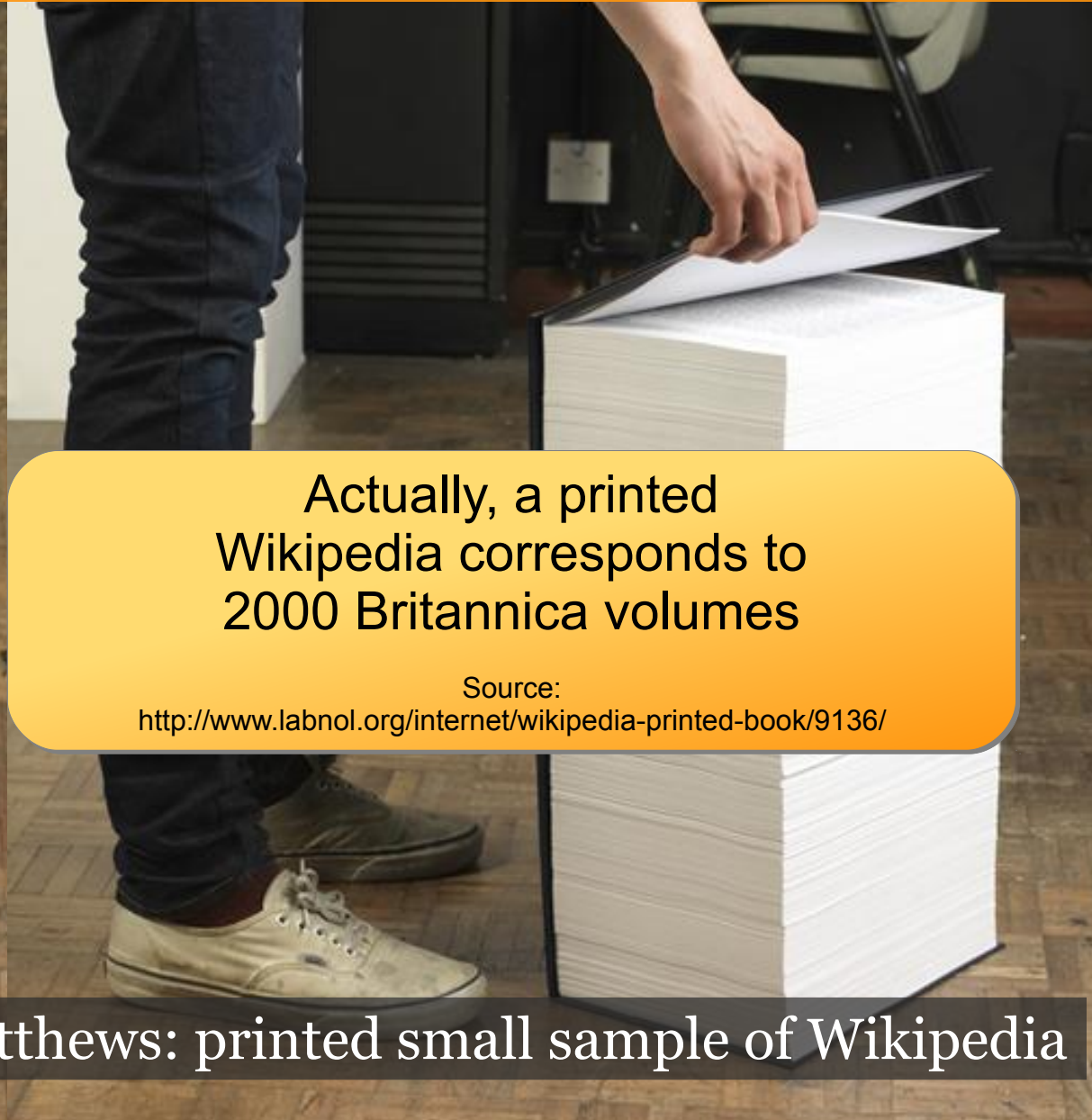
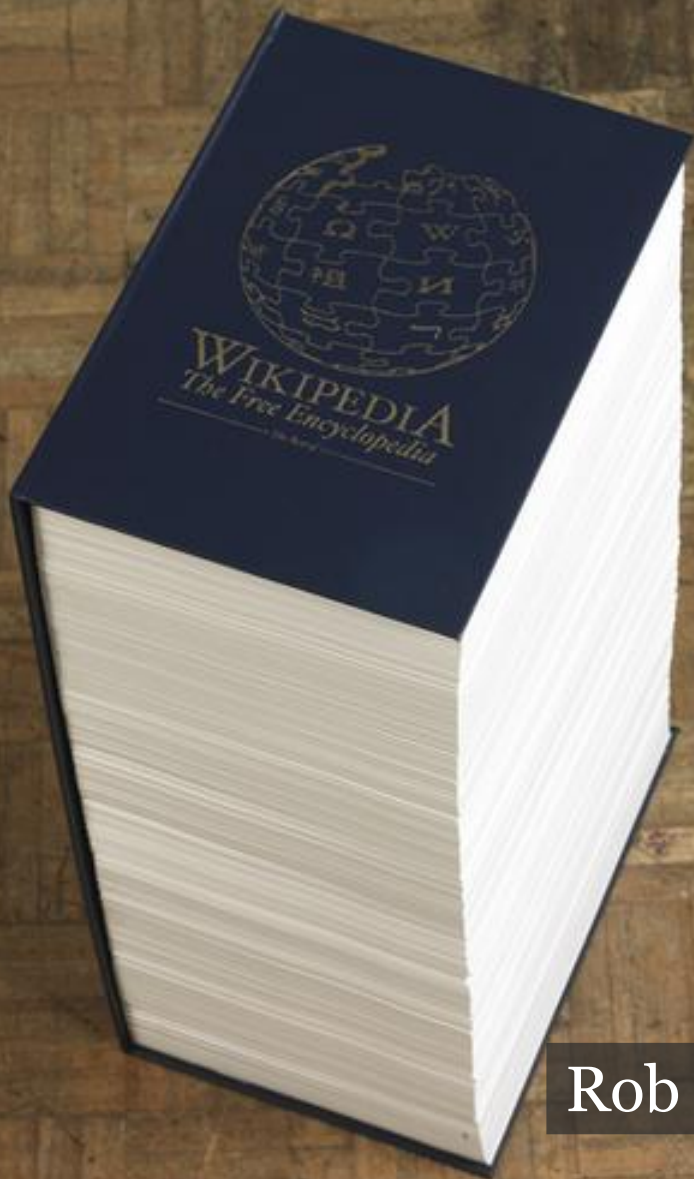
Planck in 1933

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 University of Göttingen

Big Knowledge Collections



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



max planck institut
informatik

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

Big Knowledge Collections



WIKIPEDIA
The Free Encyclopedia

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

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page


Print/export
Create a book
Download as PDF
Printable version

Languages 
Afrikaans
Alemannisch

Create account Not logged in Talk Contributions Log in

Article Talk

Read Edit View history

Search 

Max Planck

From Wikipedia, the free encyclopedia

"Planck" redirects here. For other uses, see [Planck \(disambiguation\)](#).

Max Karl Ernst Ludwig Planck, FRS^[2] (/ˈplɑːnk/^[3] German: [ˈplɔŋk]; 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]

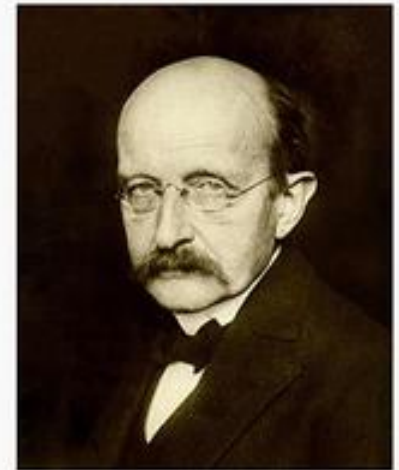
Planck made many contributions to theoretical physics, but his fame as a physicist rests primarily on his role as an originator of the quantum theory. However, his name is also known on a broader academic basis, through the renaming in 1948 of the German scientific institution, the Kaiser Wilhelm Society (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]

- Early life and career
 - Academic career
 - Family
 - Professor at Berlin University
 - Black-body radiation
 - Einstein and the theory of relativity
 - World War I
 - Post War and Weimar Republic
 - Quantum mechanics
 - Nazi dictatorship and World War II
- Religious views
- Publications
- See also
- References
- Sources
- External links

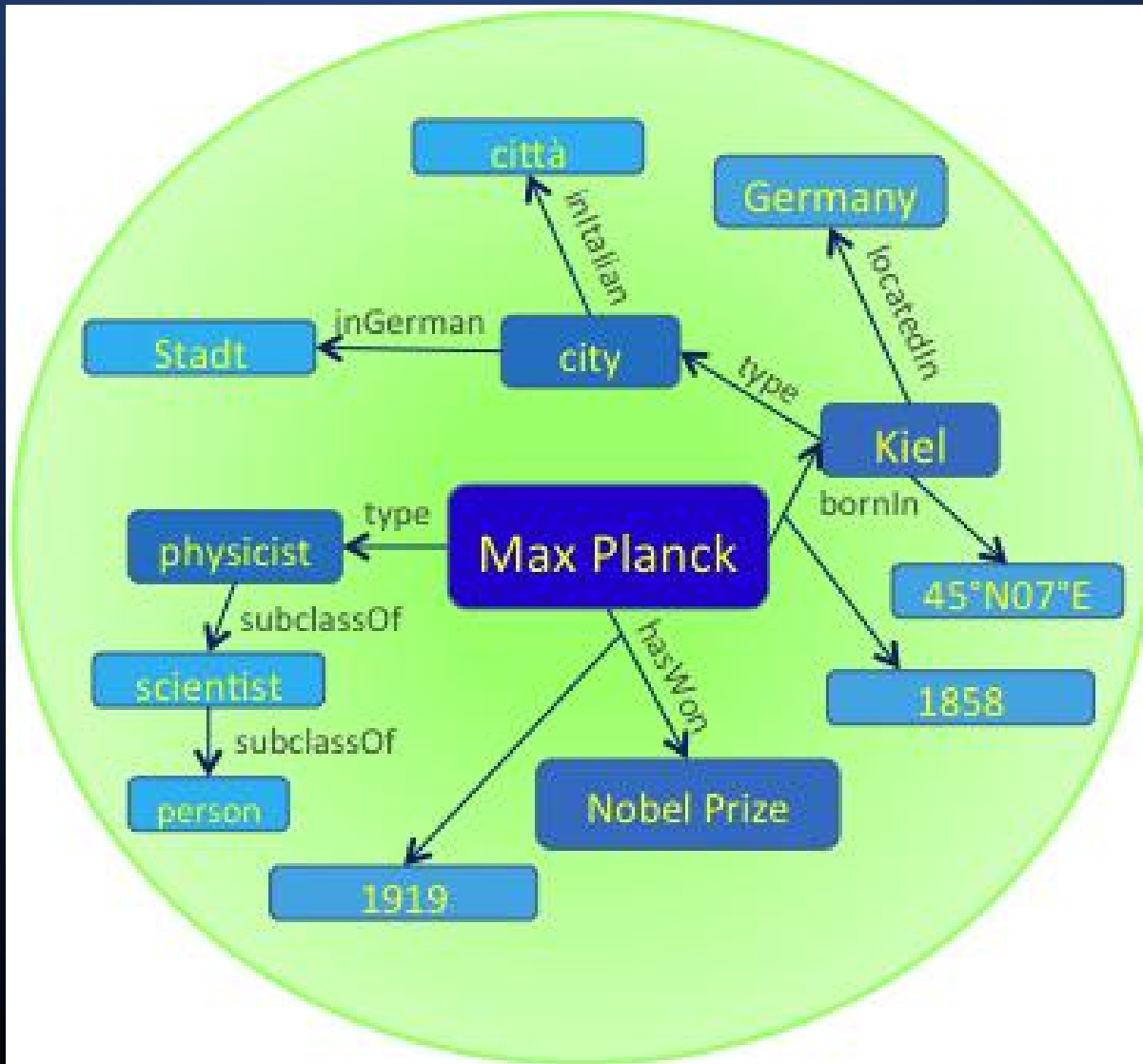
Max Planck



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 University of Göttingen

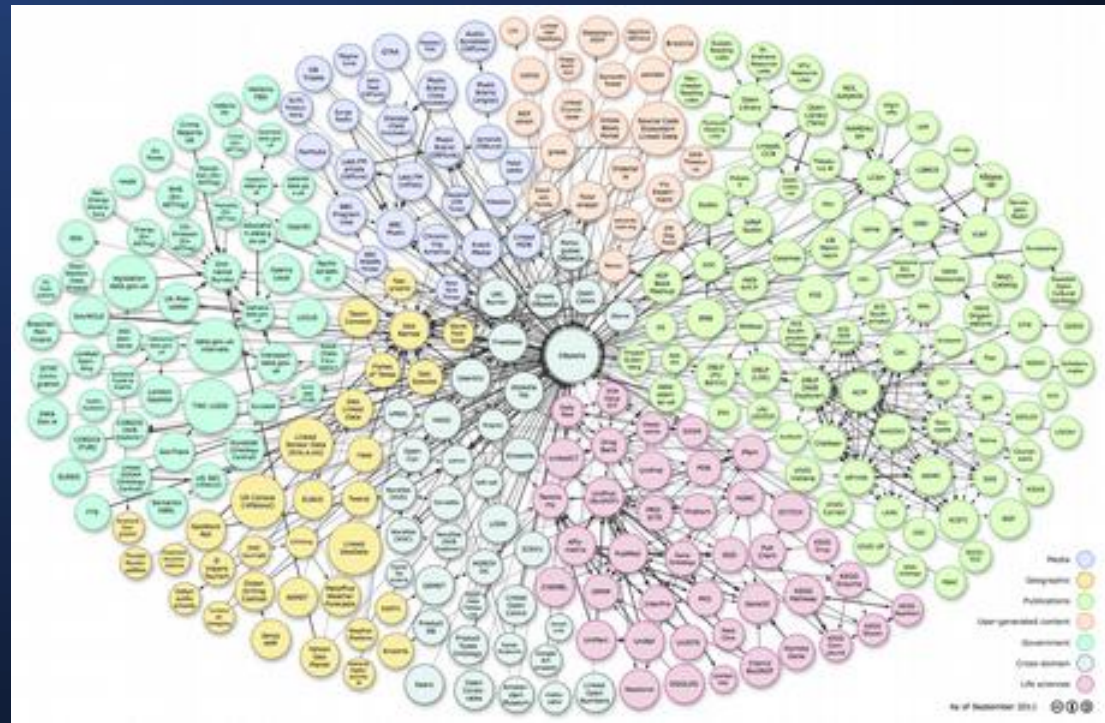
Big Knowledge Graphs



YAGO 1: 2007

YAGO2.
Hoffart et al.
WWW 2011.

Big Knowledge Graphs



Google Knowledge Graph




Cinevra do' Benci
c. 1470


The Virgin
c. 1508

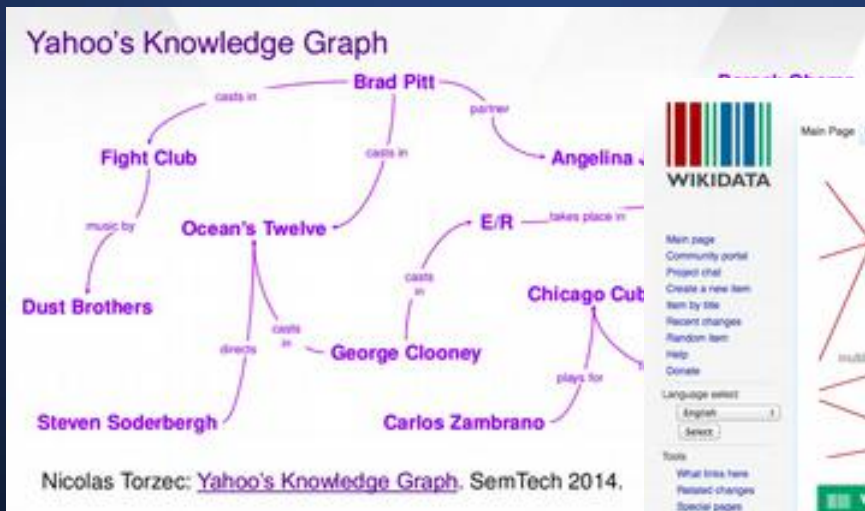

Adoration of the M...
1488

Leonardo da Vinci
Leonardo di ser Piero da Vinci was an Italian Renaissance polymath: painter, sculptor, architect, musician, scientist, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer.
Born: April 15, 1452, Anchiano
Died: May 2, 1519, Clos Lucé
Buried: Chateau d'Amboise



The screenshot shows the Facebook profile page for 'Beijing'. The page header includes the Facebook logo, navigation links (Home, Search, Post), and a user profile picture. The main content area displays the profile name 'Beijing' and a cover photo. Below the cover photo, there are sections for 'About', 'Photos', 'Posts', and 'Likes'. The 'About' section is expanded, showing various categories such as 'Location', 'Website', and 'Phone'. The 'Location' field is filled with 'Beijing, China'. The 'Website' field is empty. The 'Phone' field is also empty. The page footer includes the text '© 2015 Facebook' and 'Google'.

Mainstream Adoption



WIKIDATA

the free knowledge base with 12,952,493 data items that anyone can edit

Introduction • Project Chat • Community Portal • Help

Wikidata is a free linked database that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

Wikidata also will provide support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is available under a free license, exported using standard formats, and can be interlinked to other open datasets on the linked data web.



Mainstream Adoption



capital of new jersey

Mainstream Adoption



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

Google

[Web](#) [Maps](#) [News](#) [Shopping](#) [Images](#) [More ▾](#) [Search tools](#)

About 32,400 results (0.60 seconds)

German restaurants near sanlitun

Schindler's Anlegestelle 4 reviews · German Sanlitun N Alley	
Drei Kronen 1308 1 review · German 工体东路	
Paulaner Brauhaus 2 reviews · Restaurants - German Liangmaqiao Rd	

[☰ More german restaurants](#)

Solved: The Low-Hanging Fruit

home | browse | search | about

dblp
computer science bibliography

Search dblp

Gerard de Melo

> Home > Persons

by year Trier 1

2010 - today Refine list

2015

- [5] Gerard de Melo: Lexvo.org: Language-related information for the Linguistic Linked Data cloud. Semantic Web 6(4): 393-400 (2015)
- [c38] Miao Xie, Qiusong Yang, Qing Wang, Gao Cong, Gerard de Melo: DynDiffuse: A Dynamic Diffusion Model for Continuous Time Constrained Influence Maximization. AAAI 2015: 346-352
- [c37] Linlin Wang, Kang Liu, Zhu Cao, Jun Zhao, Gerard de Melo: Sentiment-Aspect Extraction based on Restricted Boltzmann Machines. ACL (1) 2015: 616-625
- [c36] Ekaterina Shutova, Niket Tandon, Gerard de Melo: Perceptually Grounded Selectional Preferences. ACL (1) 2015: 950-960
- [c35] Gerard de Melo: Why Researchers are Managers. PIKM@CIKM 2015: 1
- [c34] Niket Tandon, Gerard de Melo, Abir De, Gerhard Weikum: Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM 2015: 223-232
- [c33] Jacobo Rouses, Gerard de Melo, Katja Hose: Representing Specialized Events with FrameBase. DeRIVE@FSWC 2015: 58-69

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
University of Göttingen
Kaiser Wilhelm Society

Alma mater Ludwig Maximilian University of Munich

Doctoral adviser Alexander von Brill
Gustav Kirchhoff^[1]
Hermann von Helmholtz

Doctoral students Erich Kretschmann
Gustav Ludwig Hertz
Julius Edgar Lilienfeld
Max Abraham
Max von Laue
Moritz Schlick
Walter Schottky
Walther Bothe
Walther Meissner

Other notable students Lise Meitner

Known for Planck constant
Planck postulate
Planck's law of black body radiation

Notable awards Nobel Prize in Physics (1918)
Lorentz Medal (1927)
Copley Medal (1929)
Max Planck Medal (1929)
Goethe Prize (1945)

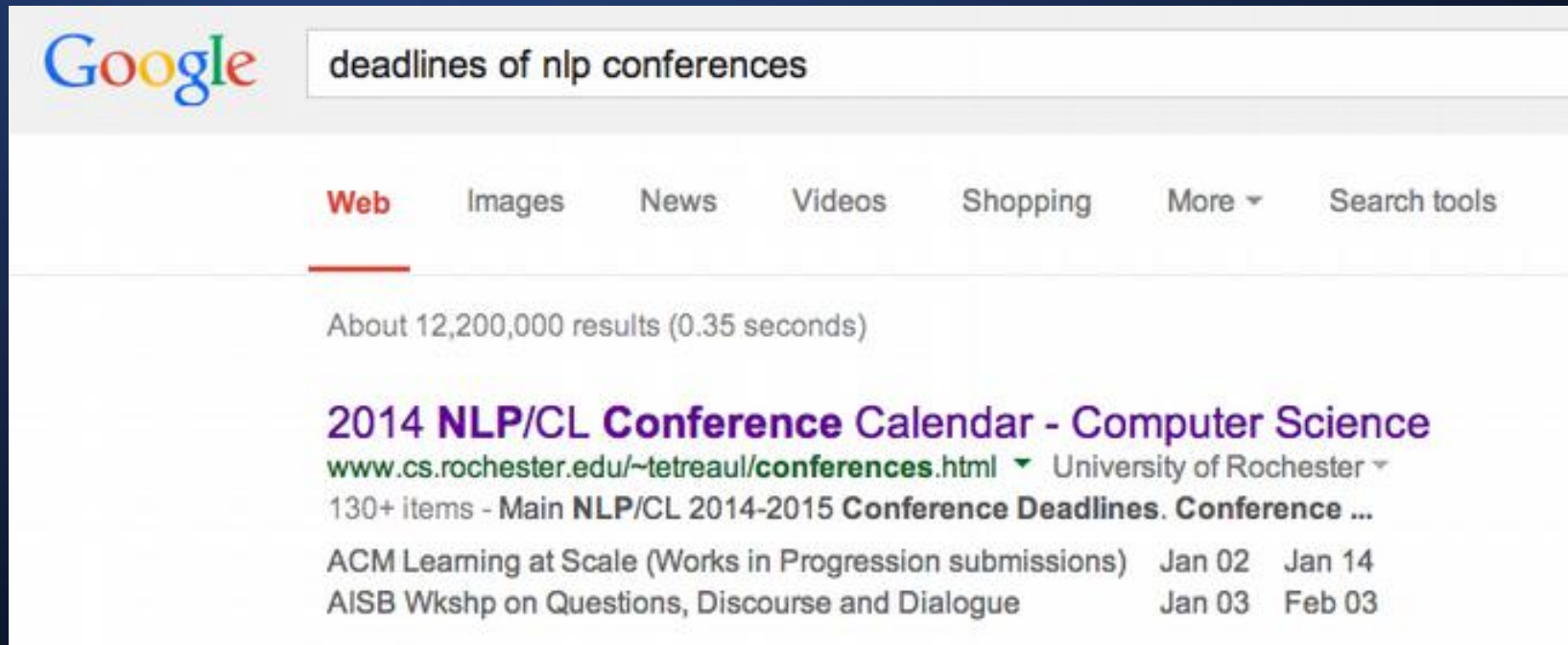
Spouse Marie Merck (1887–1909)
Marga von Hilse (1911–1947)

Signature

Notes

His son Erwin Planck was executed in 1945 by the Gestapo for his part in the assassination attempt on Adolf Hitler July 20 plot.

What we should be able to ask



Google

deadlines of nlp conferences

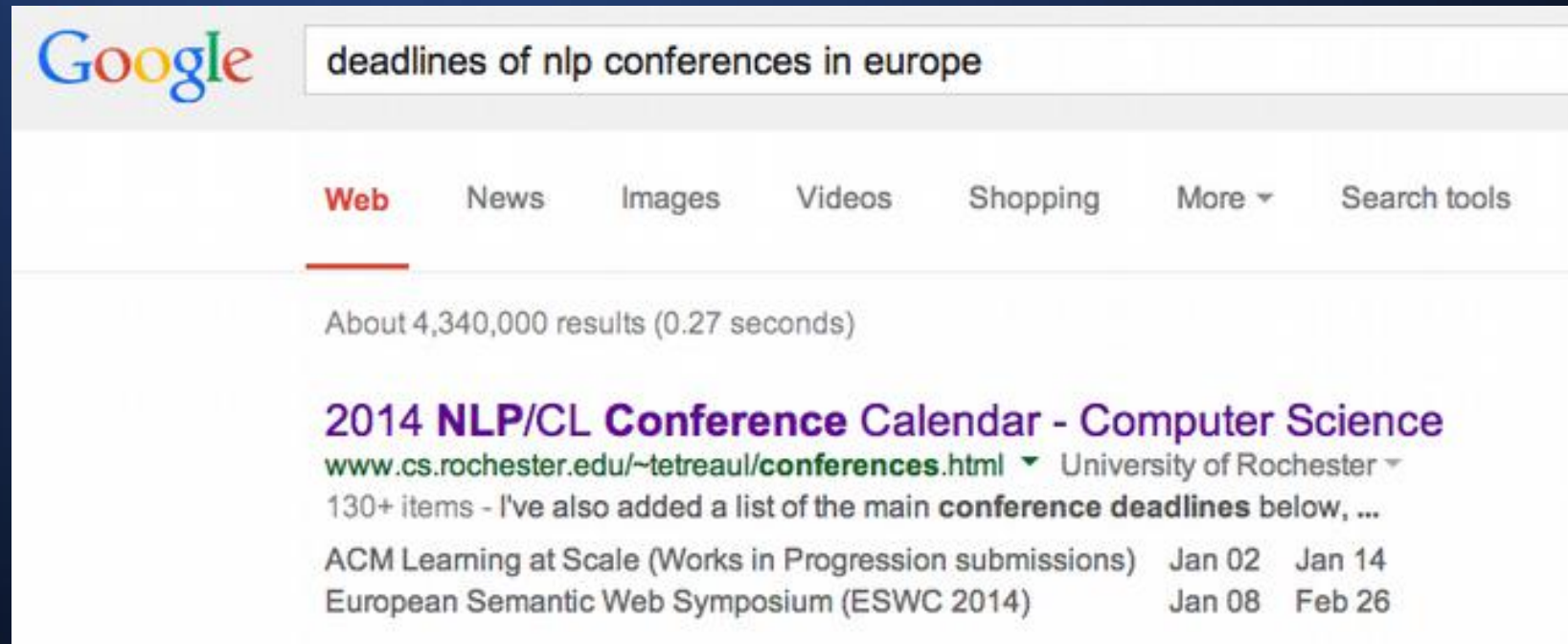
Web Images News Videos Shopping More Search tools

About 12,200,000 results (0.35 seconds)

2014 NLP/CL Conference Calendar - Computer Science
www.cs.rochester.edu/~tetreaul/conferences.html University of Rochester
130+ items - Main NLP/CL 2014-2015 Conference Deadlines. Conference ...

ACM Learning at Scale (Works in Progression submissions)	Jan 02	Jan 14
AISB Wkshp on Questions, Discourse and Dialogue	Jan 03	Feb 03

What we should be able to ask



The screenshot shows a Google search interface. The search bar contains the text "deadlines of nlp conferences in europe". Below the search bar, there are navigation tabs for "Web", "News", "Images", "Videos", "Shopping", "More", and "Search tools". The "Web" tab is selected. Below the tabs, it says "About 4,340,000 results (0.27 seconds)". The first search result is titled "2014 NLP/CL Conference Calendar - Computer Science" and is from the University of Rochester. The result snippet includes the URL "www.cs.rochester.edu/~tetreaul/conferences.html" and mentions "130+ items - I've also added a list of the main conference deadlines below, ...". Below the snippet, there is a table listing conference deadlines:

ACM Learning at Scale (Works in Progression submissions)	Jan 02	Jan 14
European Semantic Web Symposium (ESWC 2014)	Jan 08	Feb 26

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

What we should be able to ask

The screenshot shows a Google search interface. The search bar contains the text "spring deadlines of nlp conferences in europe". Below the search bar, the "Web" tab is selected. The search results show "About 4,090,000 results (0.66 seconds)". The top result is titled "2014 NLP/CL Conference Calendar - Computer Science" and is from the University of Rochester. The result snippet includes a link to "www.cs.rochester.edu/~tetreaul/conferences.html" and mentions "130+ items - I've also added a list of the main conference deadlines below, ...". Below the snippet, there is a table of conference deadlines:

ACM Learning at Scale (Works in Progression submissions)	Jan 02	Jan 14
European Semantic Web Symposium (ESWC 2014)	Jan 08	Feb 26

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:

...

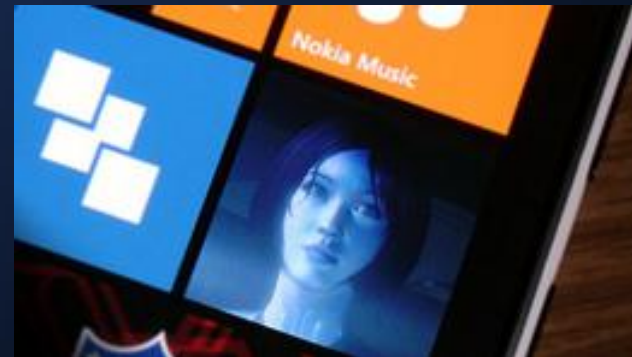


Image: Microsoft Cortana

Beyond Collections of Facts

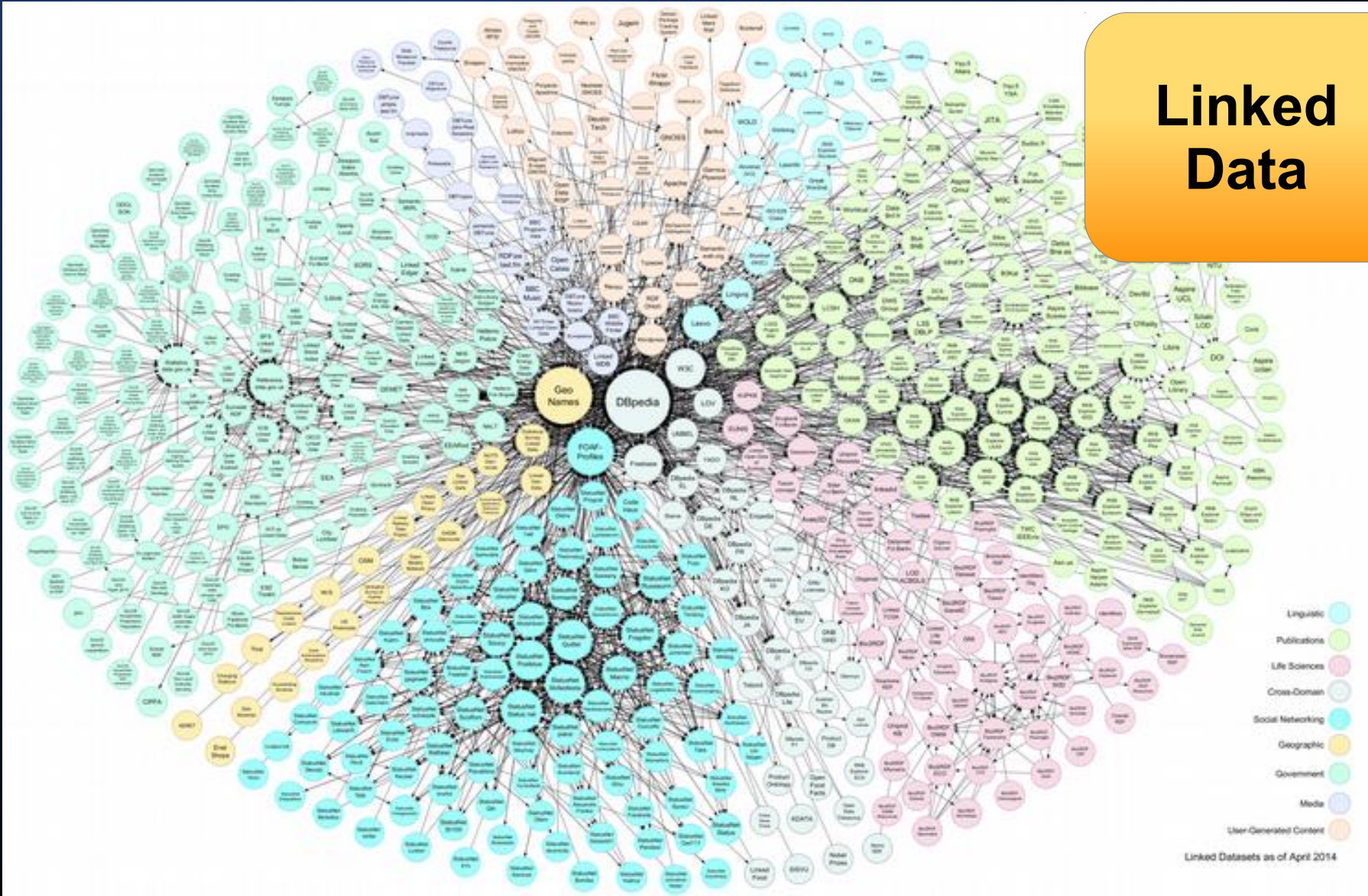
- **1. Knowledge Integration and Organization**
- **2. Natural Language Coupling**
- **3. Multimodal and Cognitive Knowledge**

Beyond Collections of Facts

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

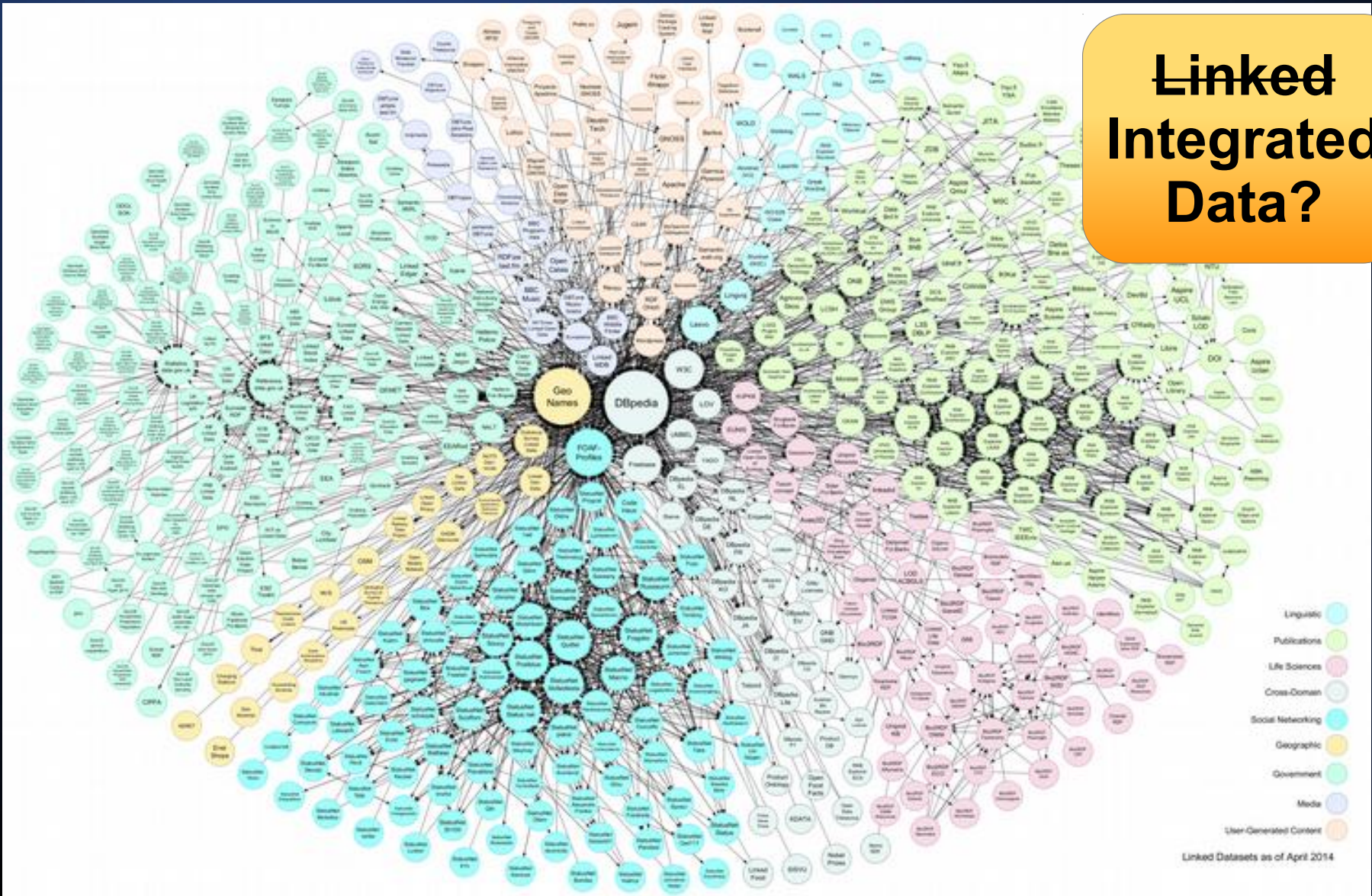
The Web of Data

**Linked
Data**



The Web of Data

**Linked
Integrated
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/bwc/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://sparqls.okfn.org/>

Linked Data Is Still Not Integrated Data




A lot of data is locked away to some extent

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




HU Berlin

bornIn




Ulm




WIKIPEDIA

This block illustrates a disconnected information graph. It features three Wikipedia article screenshots: Albert Einstein, Humboldt University of Berlin, and Ulm. A green arrow labeled 'bornIn' points from the Einstein article to the HU Berlin article. The Wikipedia logo is also present.



Rudolf Carnap

IsA



Philosopher

This block shows a simple relationship: a portrait of Rudolf Carnap is connected to a red Phi symbol (representing a philosopher) via a green arrow labeled 'IsA'.

Carnap


Studied at

University of Berlin

Einstein

Was Professor at

This block shows two relationships: 'Carnap' is connected to 'University of Berlin' via a green arrow labeled 'Studied at', and 'Einstein' is connected to 'University of Berlin' via a green arrow labeled 'Was Professor at'.



GeoNames

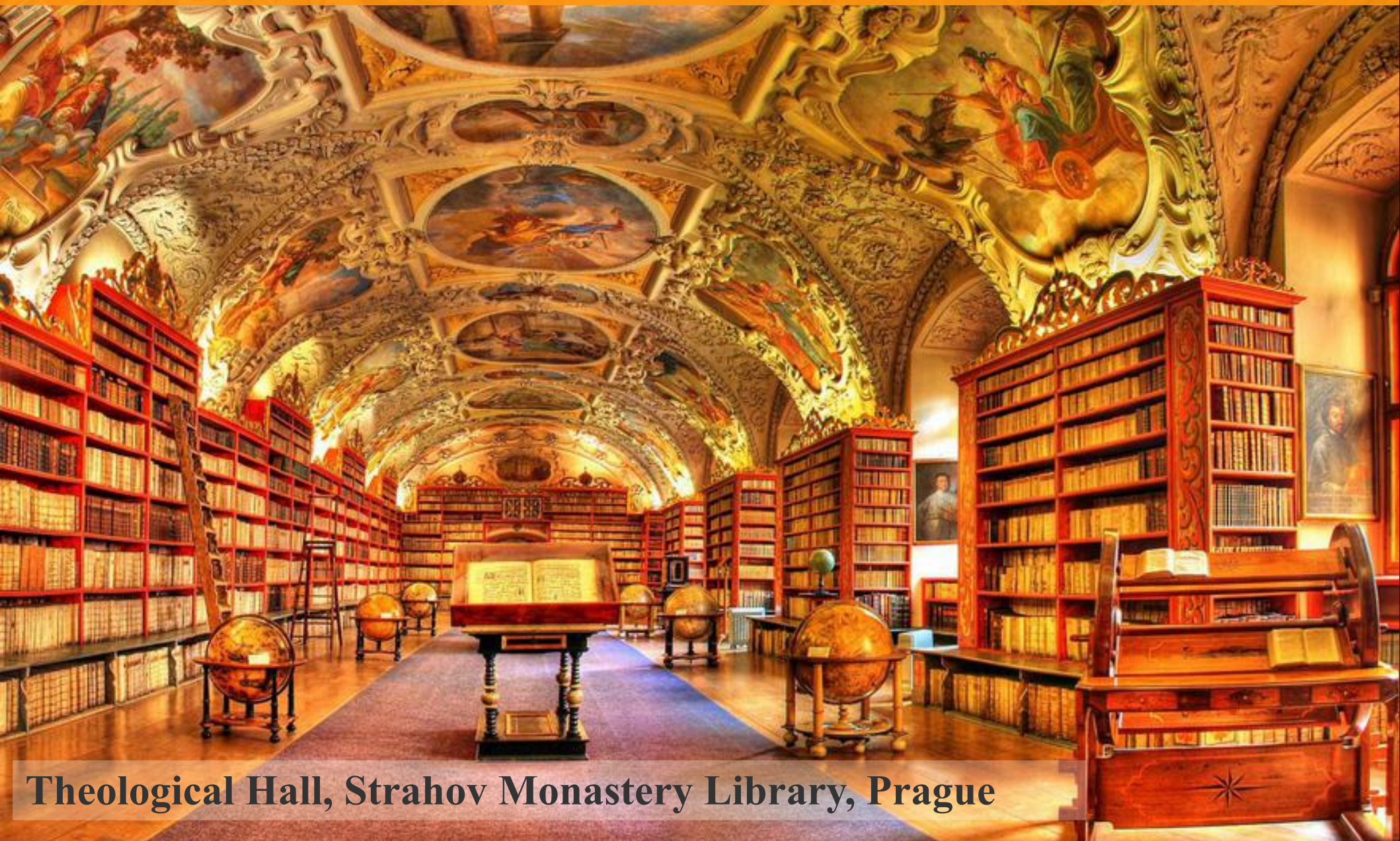
...

This block features the GeoNames logo, which consists of a globe icon and the text 'GeoNames'. Below it are three dots indicating further information.

Current Situation



Goal: Integrated and Organized Knowledge



Theological Hall, Strahov Monastery Library, Prague

Entity Integration



Entity Integration



Find vego

Near Montréal, Quebec, Canada



Home About Me Write a Review Find Friends Messages Talk Events

vego Montreal, QC, Canada

Showing 1-5 of 5

Show Filters -



1. Restaurant **Végo**

★★★★☆ 16 reviews

\$\$ - Vegetarian

1720 Saint-Denis Rue
Montréal, QC H2X 3K6
Canada
[\(514\) 845-2627](tel:(514)845-2627)



Nice staff and ambiance, lots of variety in terms of food options, delicious food. A bit awkward to have the buffet upstairs, especially if you sit down stairs. Carrying a tray down narrow...



2. Resto **Végo** McGill

★★★★☆ 5 reviews

\$\$ - Buffets, Vegan, Vegetarian

1204 Avenue McGill College
Montréal, QC H3B 4J8
Canada
[\(514\) 871-1480](tel:(514)871-1480)



We had an afternoon meal in this place. The staff were very friendly and helpful. We like buffet restaurants because they make it possible to collect exactly the type of food that you want in...



3. Resto **Végo** St-Denis

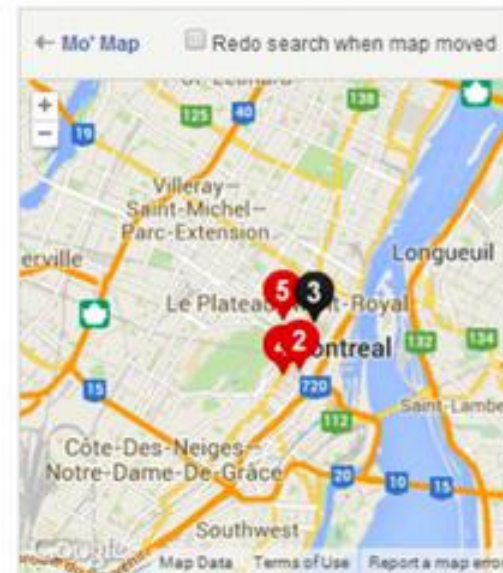
★★★★☆ 3 reviews

\$\$ - Vegetarian

1720 Rue Saint-denis
Montréal, QC H2X 3K6
Canada
[\(514\) 845-2627](tel:(514)845-2627)



As a non-veg I was quite satisfied with Resto Végo, the food was good and the options grand. This is a buffet style restaurant so I had the opportunity to taste many of the various dishes...



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](https://en.wikipedia.org/wiki/Michelangelo)

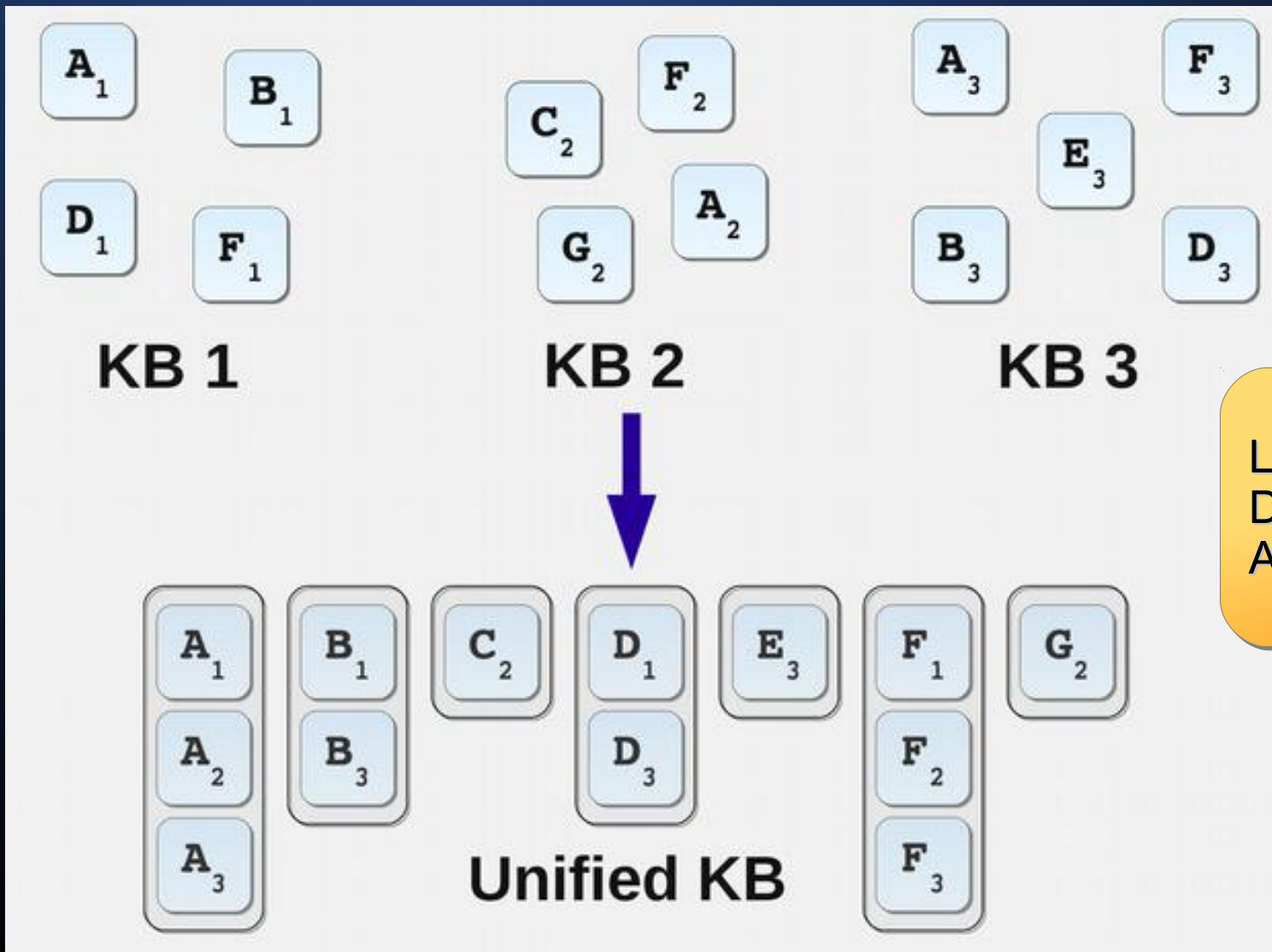
Born: March 6, 1475, [Caprese Michelangelo](#)

Died: February 18, 1564, [Rome](#)

Parents: [Ludovico di Leonardo di Buonarotto Simoni](#), [Francesca di Neri del Miniato di Siena](#)

Feedback

Entity-Level Integration



Language-specific,
Domain-specific,
Arbitrary Databases

Entity Integration: Computing Similarity Scores

dbpedia:The_Godfather

dbpedia-owl:director dbpedia:Francis_Ford_Coppola

dbpedia-owl:starring dbpedia:Al_Pacino

dbpedia-owl:starring dbpedia:Marlon_Brando

dbpedia-owl:starring dbpedia:Marlon_Brando

dcterms:subject category:Mafia_films

...

linkedmdb:film/43338

movie:actor linkedmdb:actor/10107

movie:actor linkedmdb:actor/10486

dc:date 1972

movie:director linkedmdb:director/8405

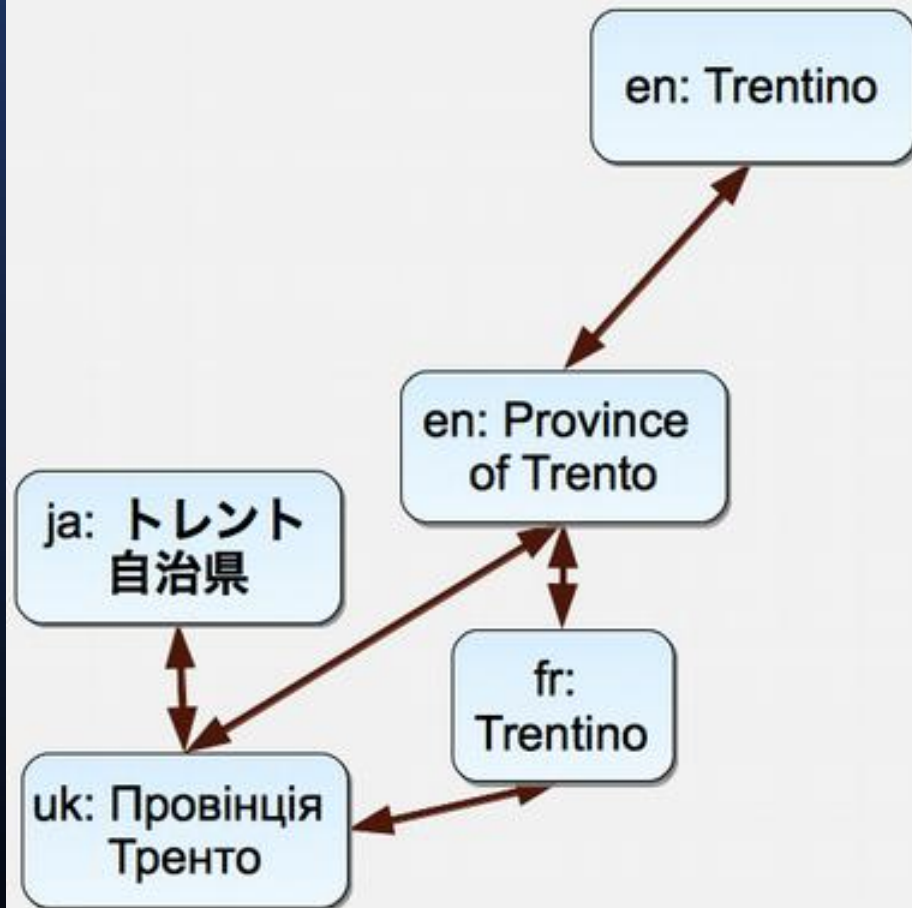
movie:editor linkedmdb:editor/2902

...

Are these two the same?

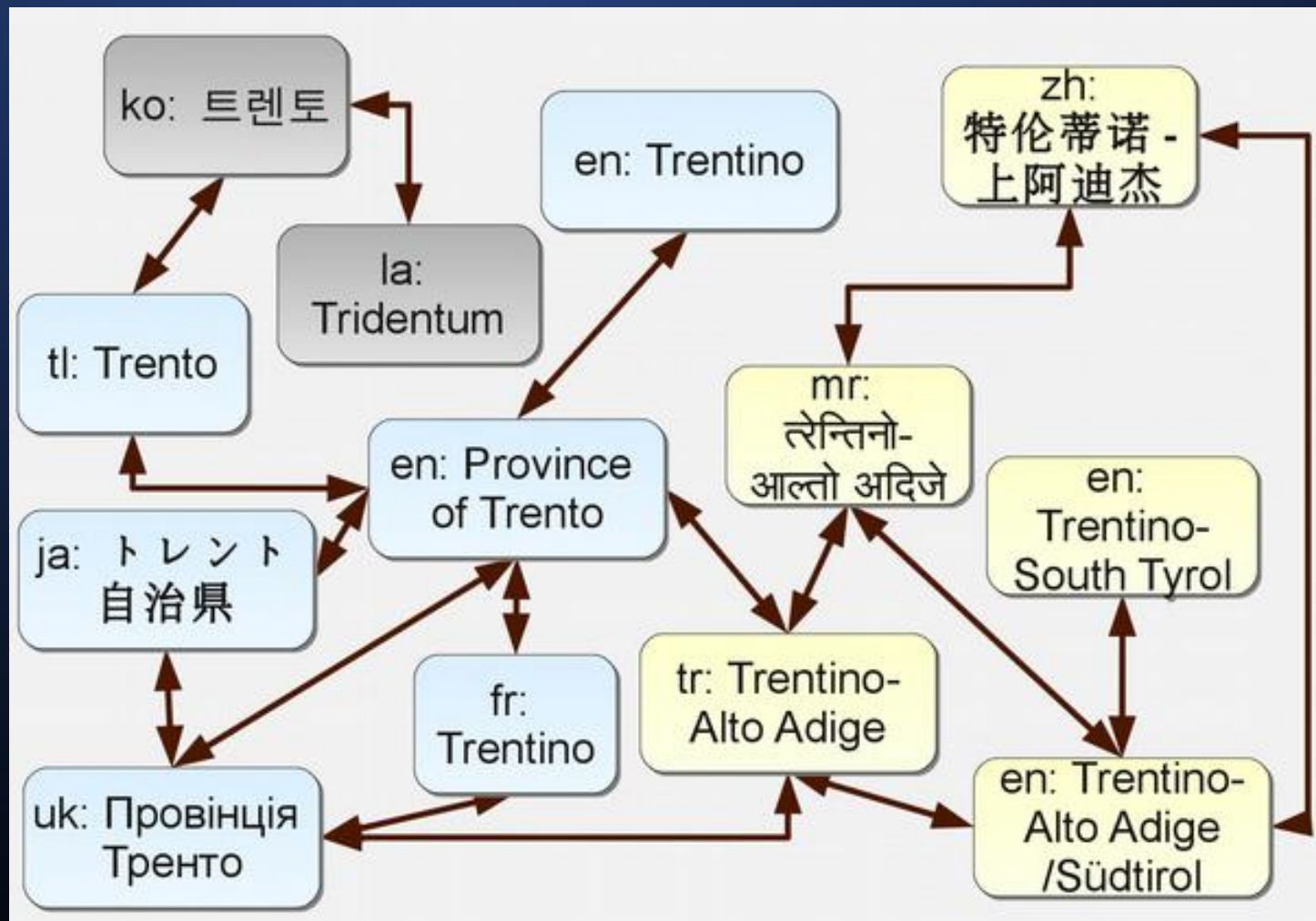
Compare attributes, relations, text
→ Similarity score (weighted link)

Merging Structured Data



Use
Identity Links
to connect
What is
equivalent

Merging Structured Data

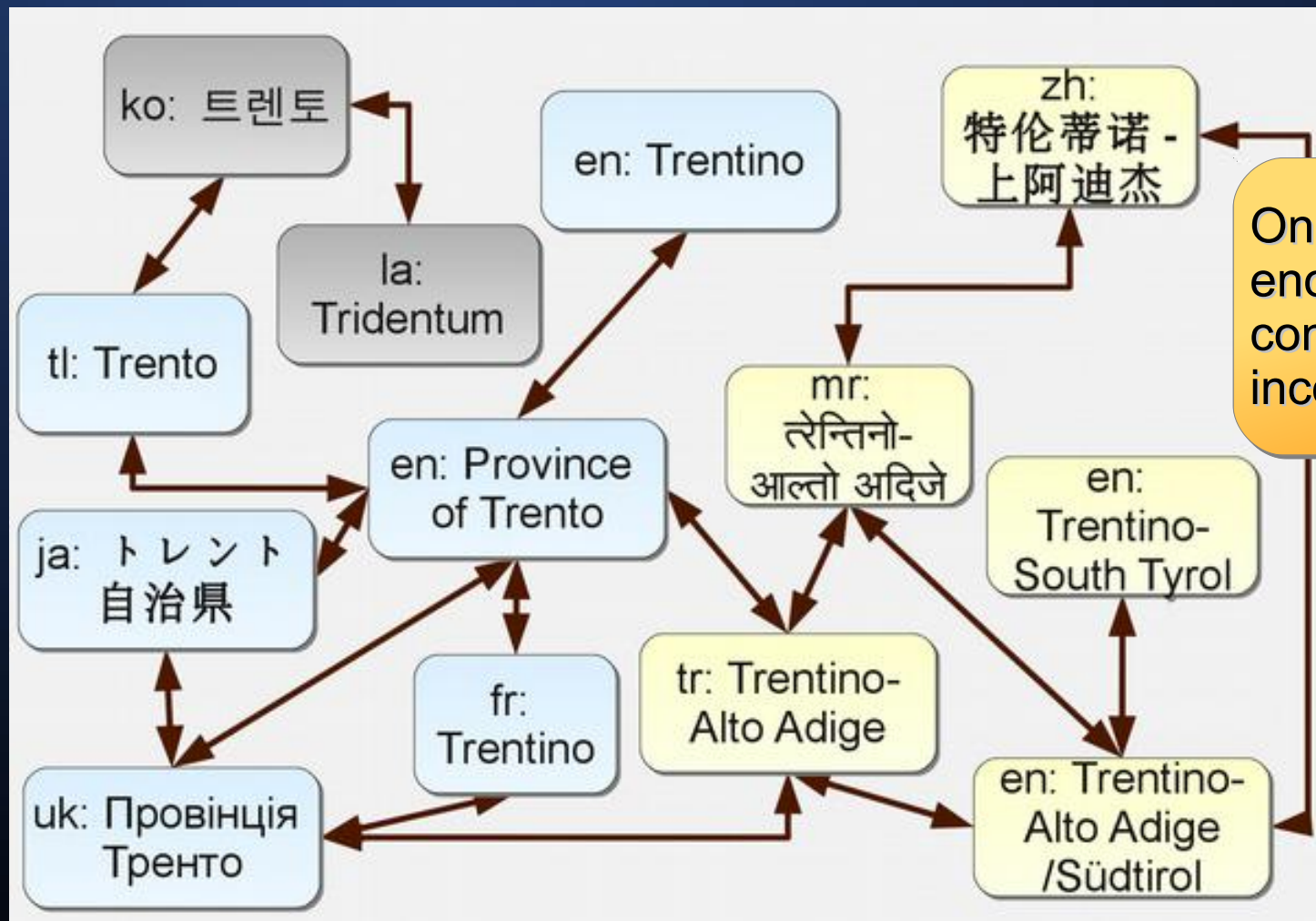


ACL 2010
AAAI 2013

Merging Structured Data

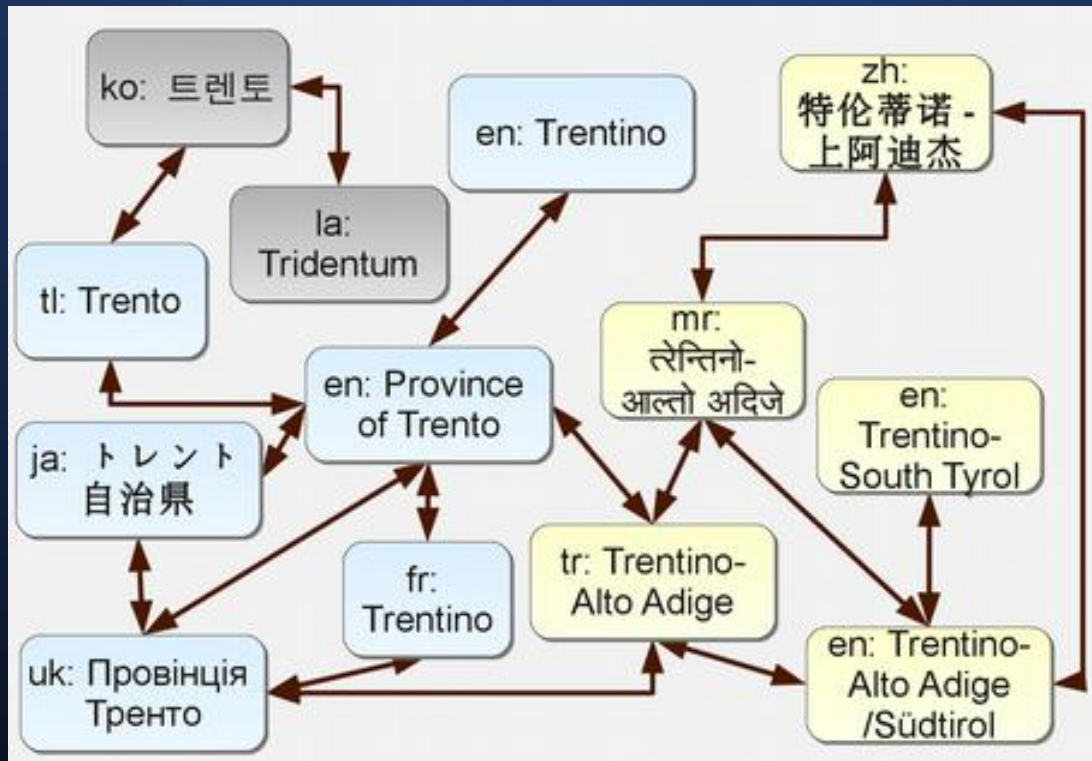


Merging Structured Data



One bad link is enough to make a connected component inconsistent

Merging Structured Data

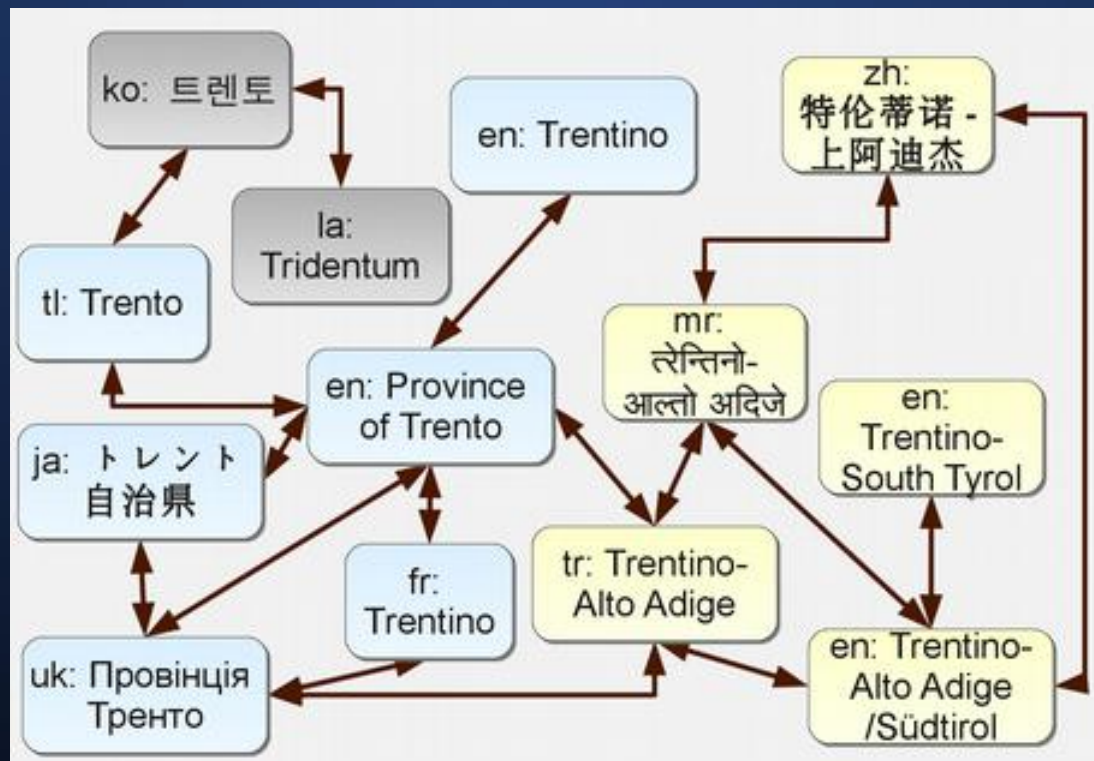


Optimization Problem:
Make clusters consistent

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

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

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

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

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

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

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

- 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

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

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

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

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

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

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

- 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

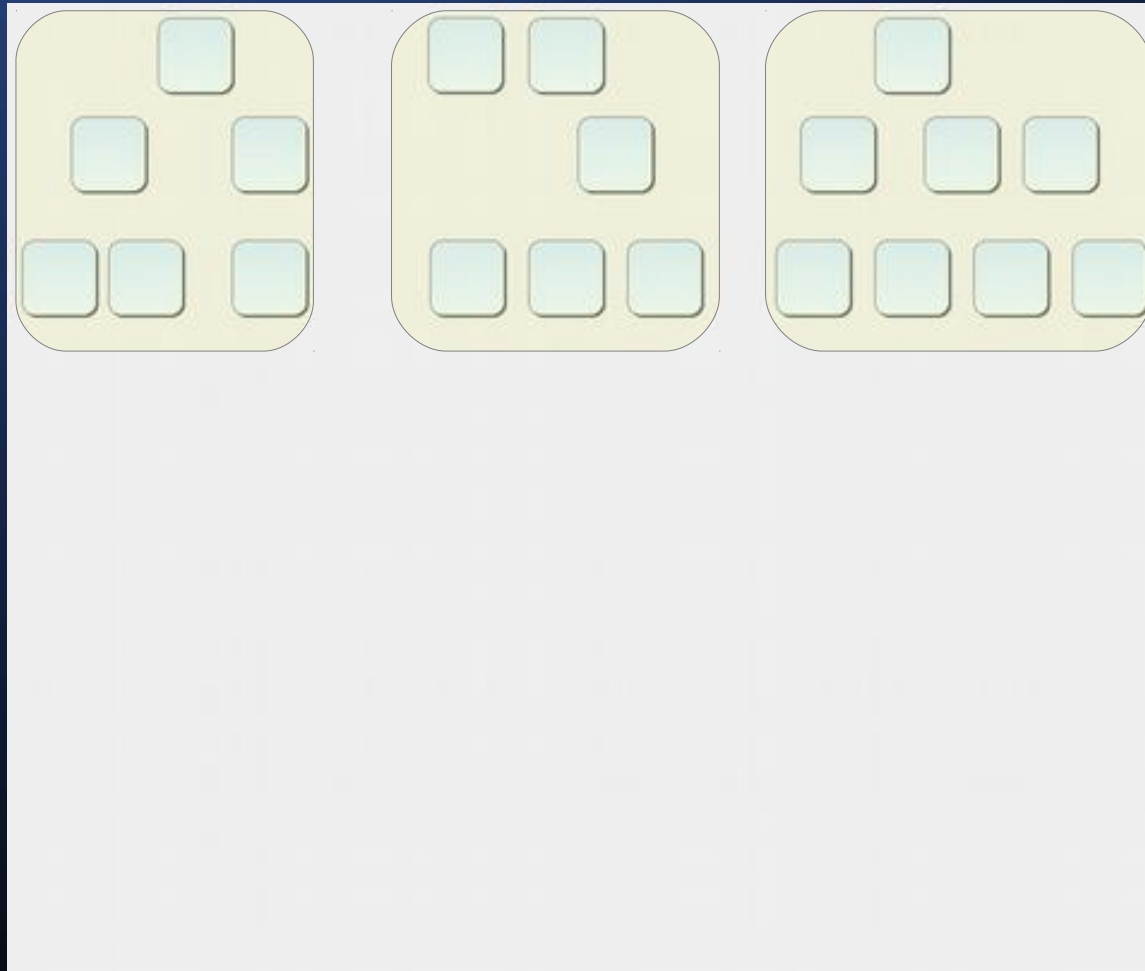
Question Answering



IBM's Jeopardy!-
system

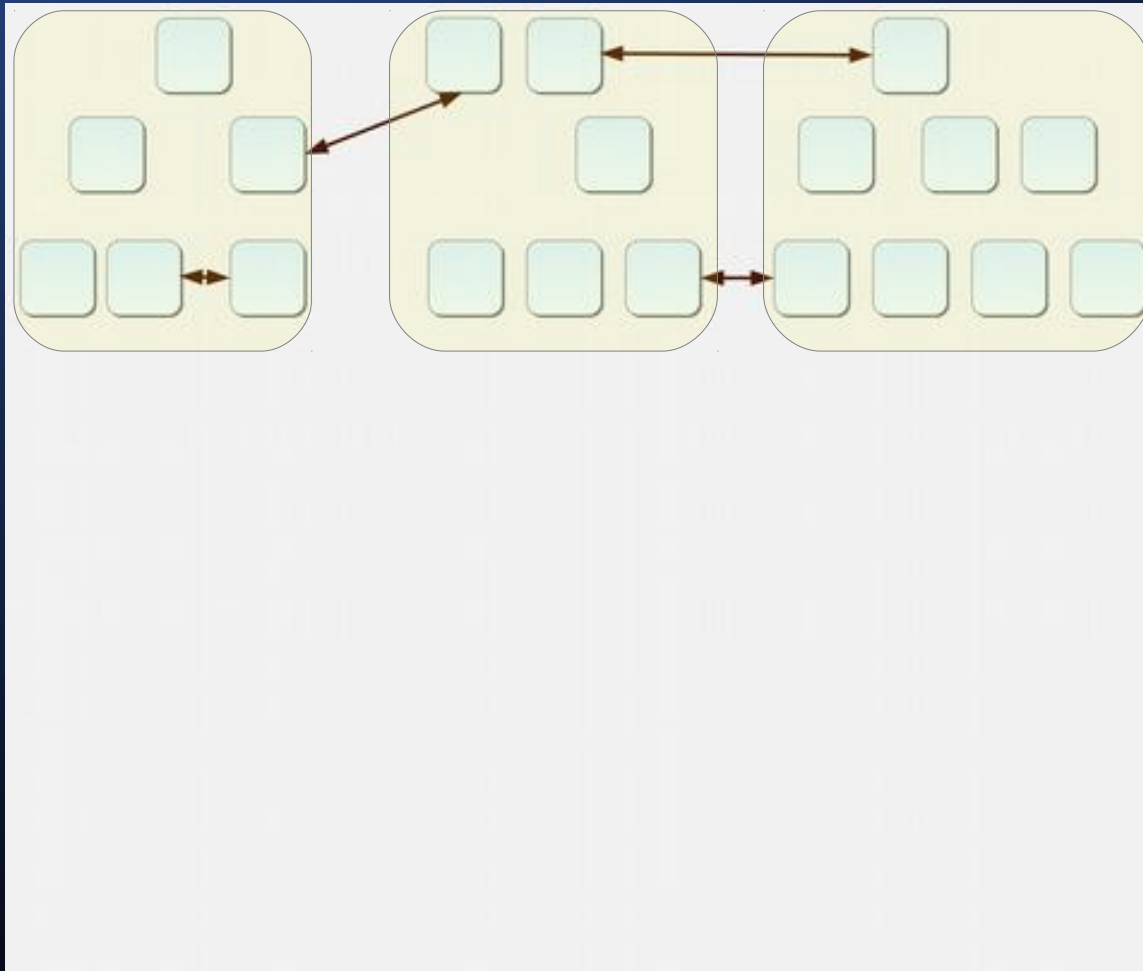
THIS **GROUP** GIVES
SCHOLARSHIPS TO
MINORITY STUDENTS &
FINANCIAL SUPPORT
TO 37 HISTORICALLY
BLACK COLLEGES &
UNIVERSITIES

Taxonomic Integration: MENTA Approach



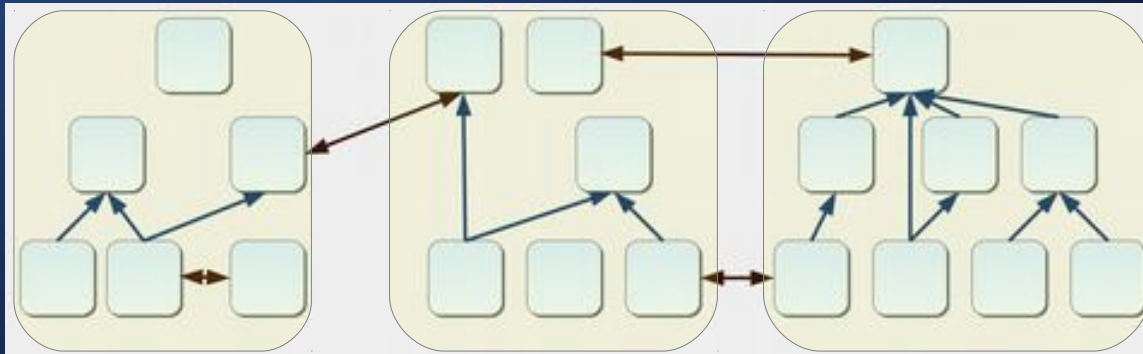
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



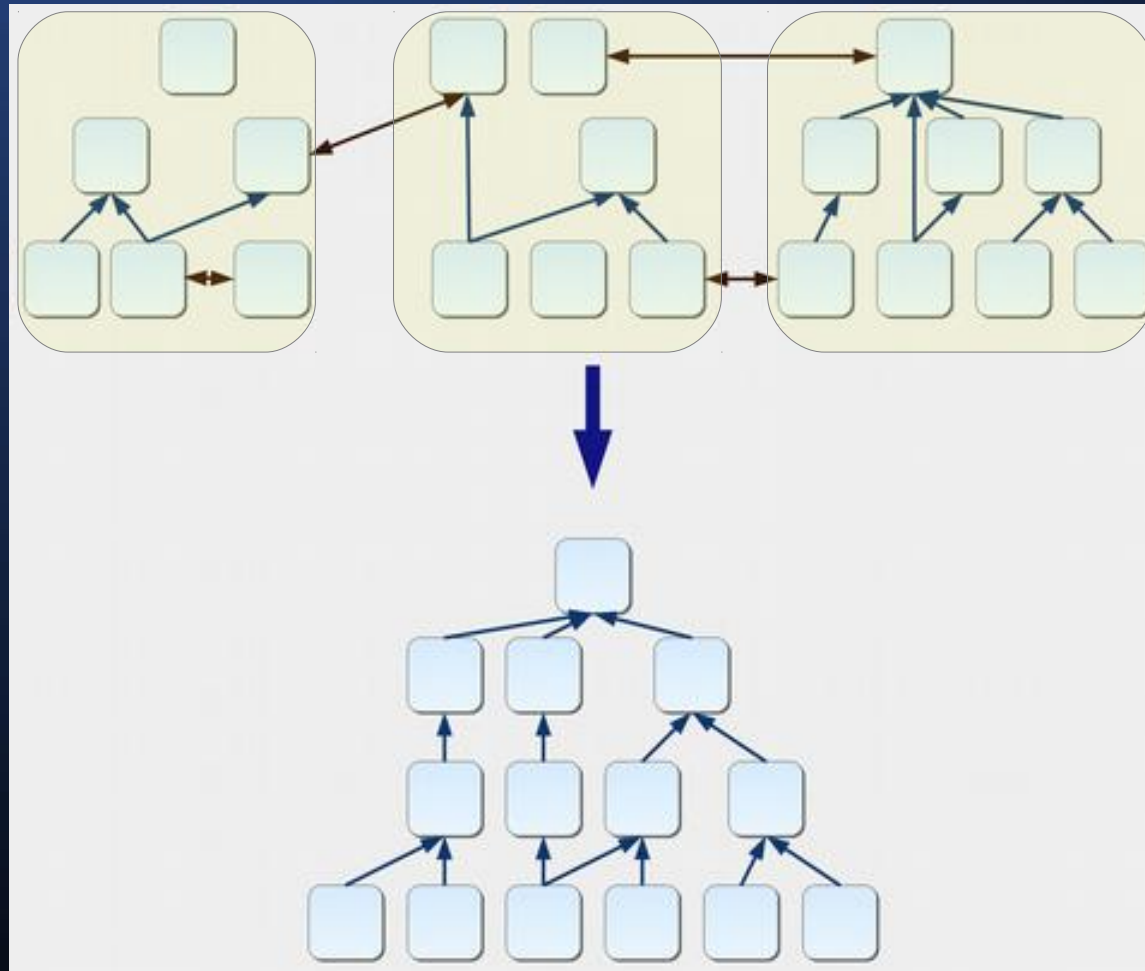
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



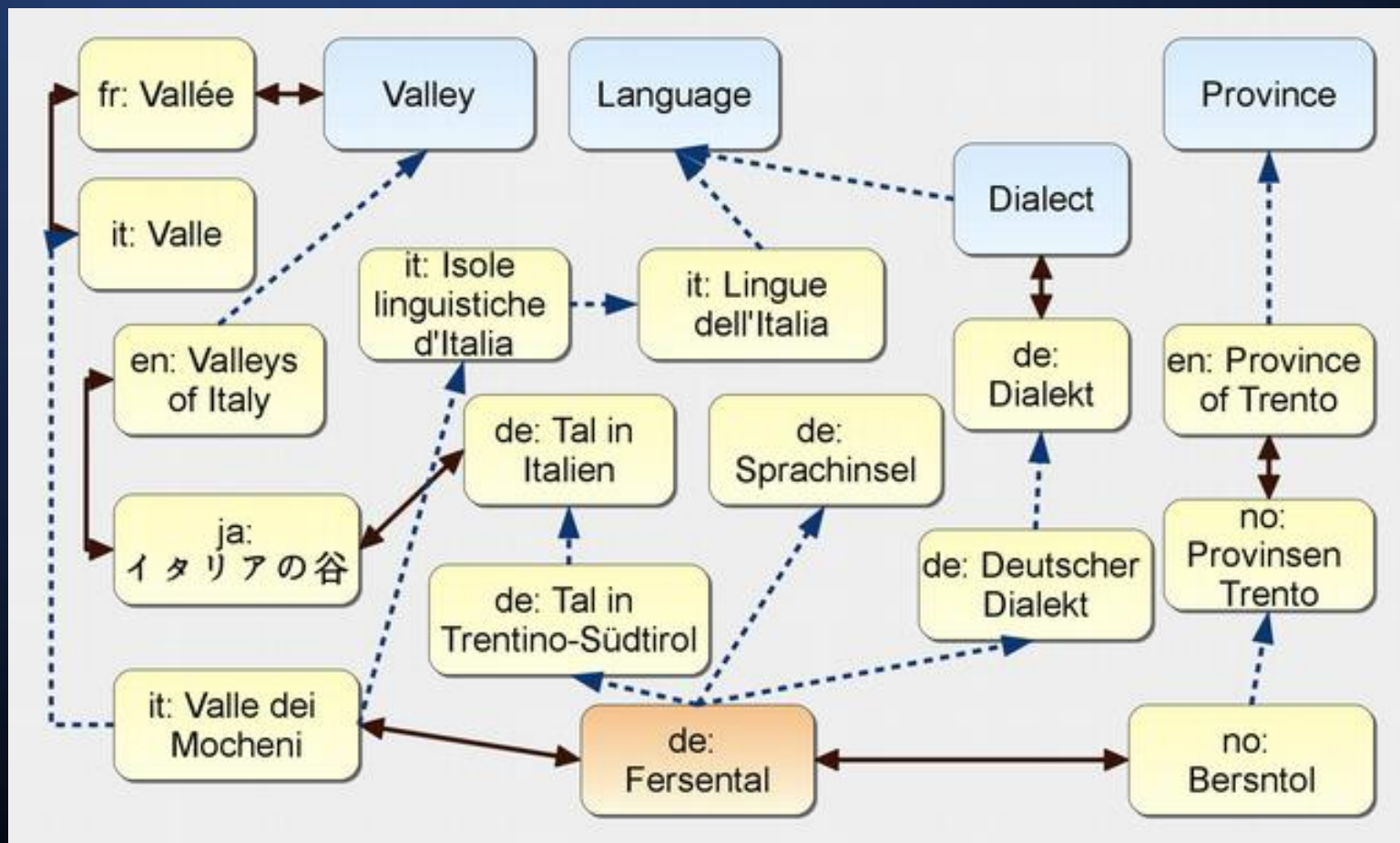
De Melo & Weikum (2010).
CIKM Best Interdisciplinary Paper Award

Taxonomic Integration: MENTA Approach



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

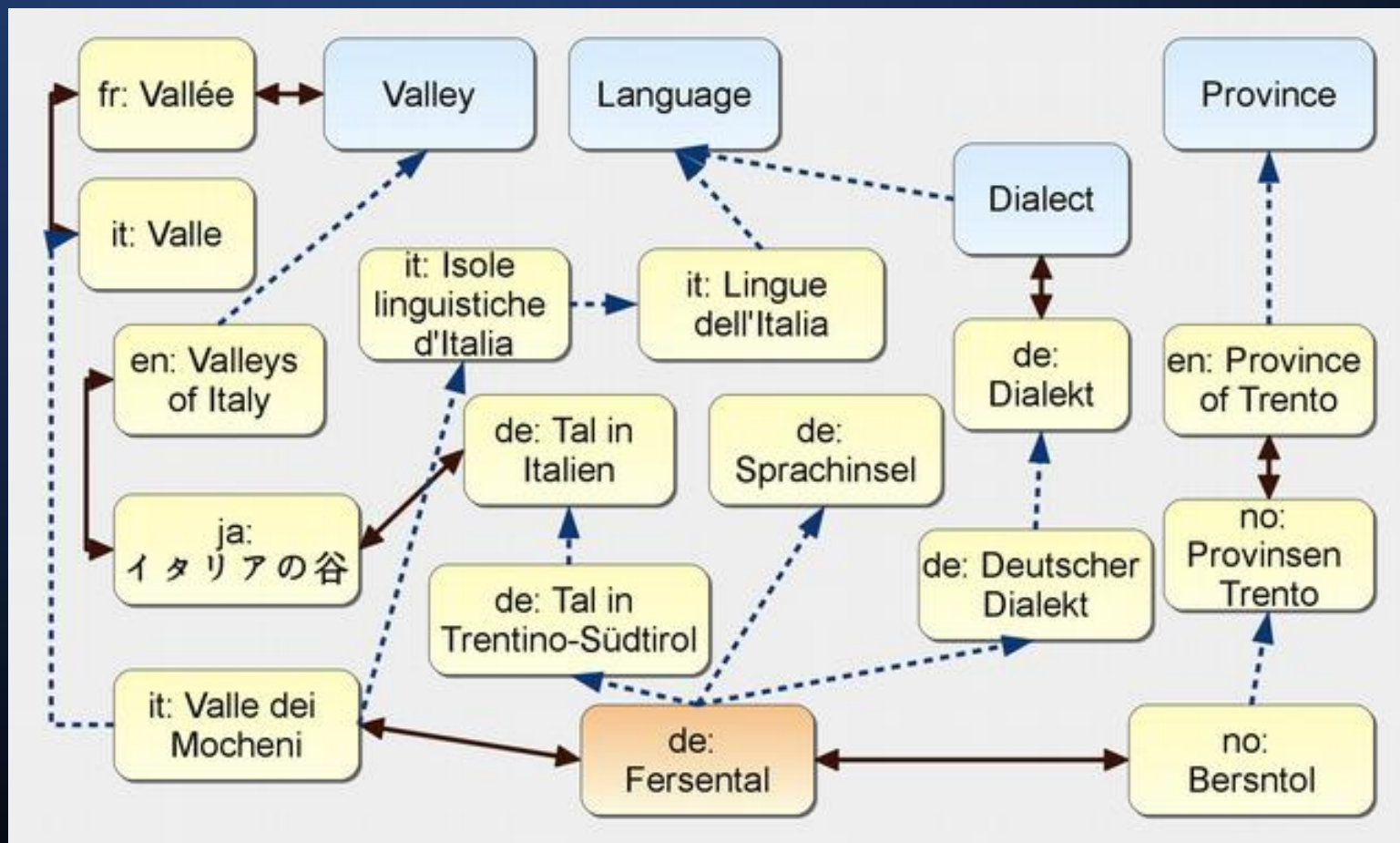
Taxonomic Integration: MENTA



Taxonomic Integration: MENTA

**Fersental
(Bersntol, Valle dei Mòcheni)**

Taxonomic Integration: MENTA

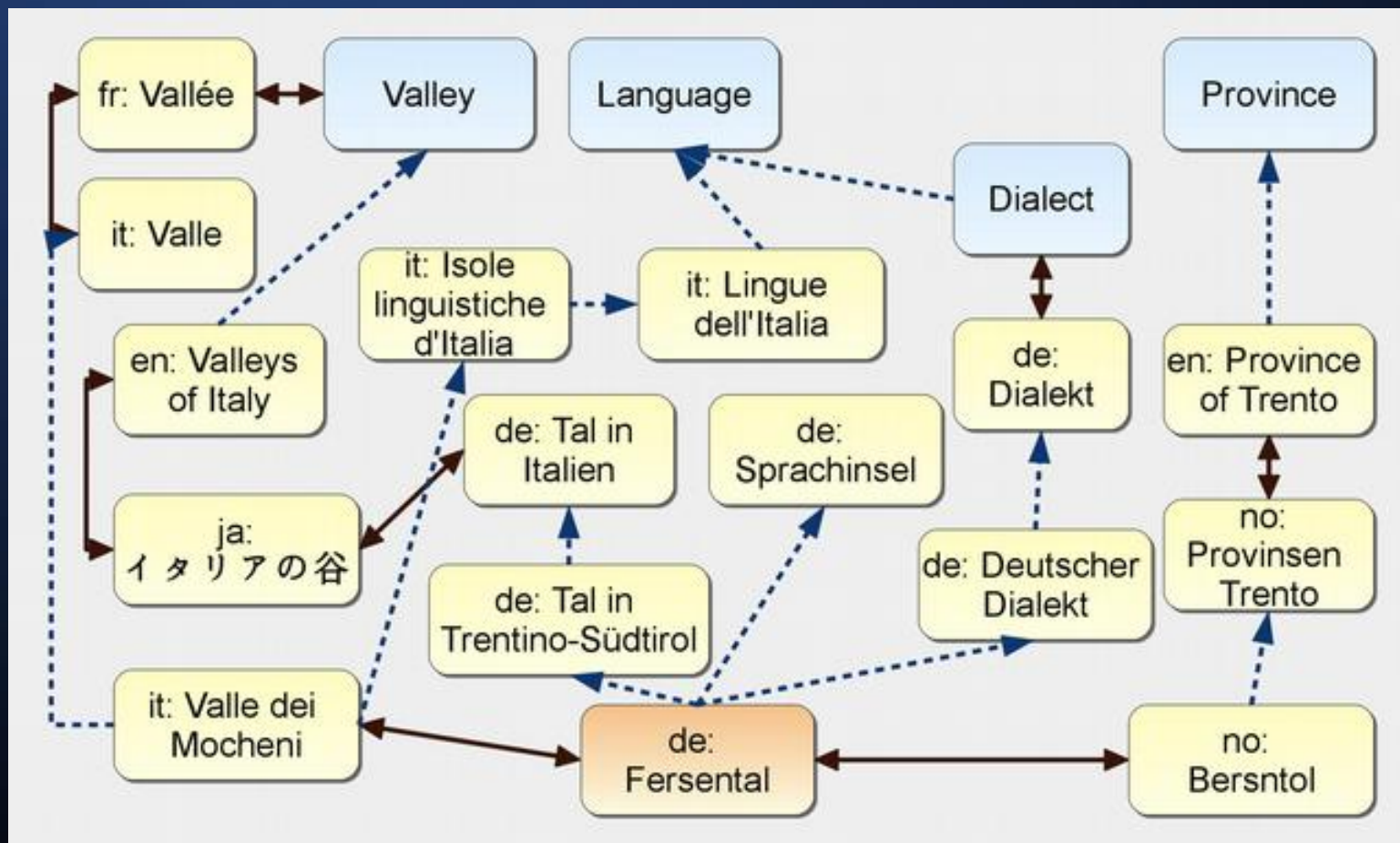


Taxonomic Integration: MENTA

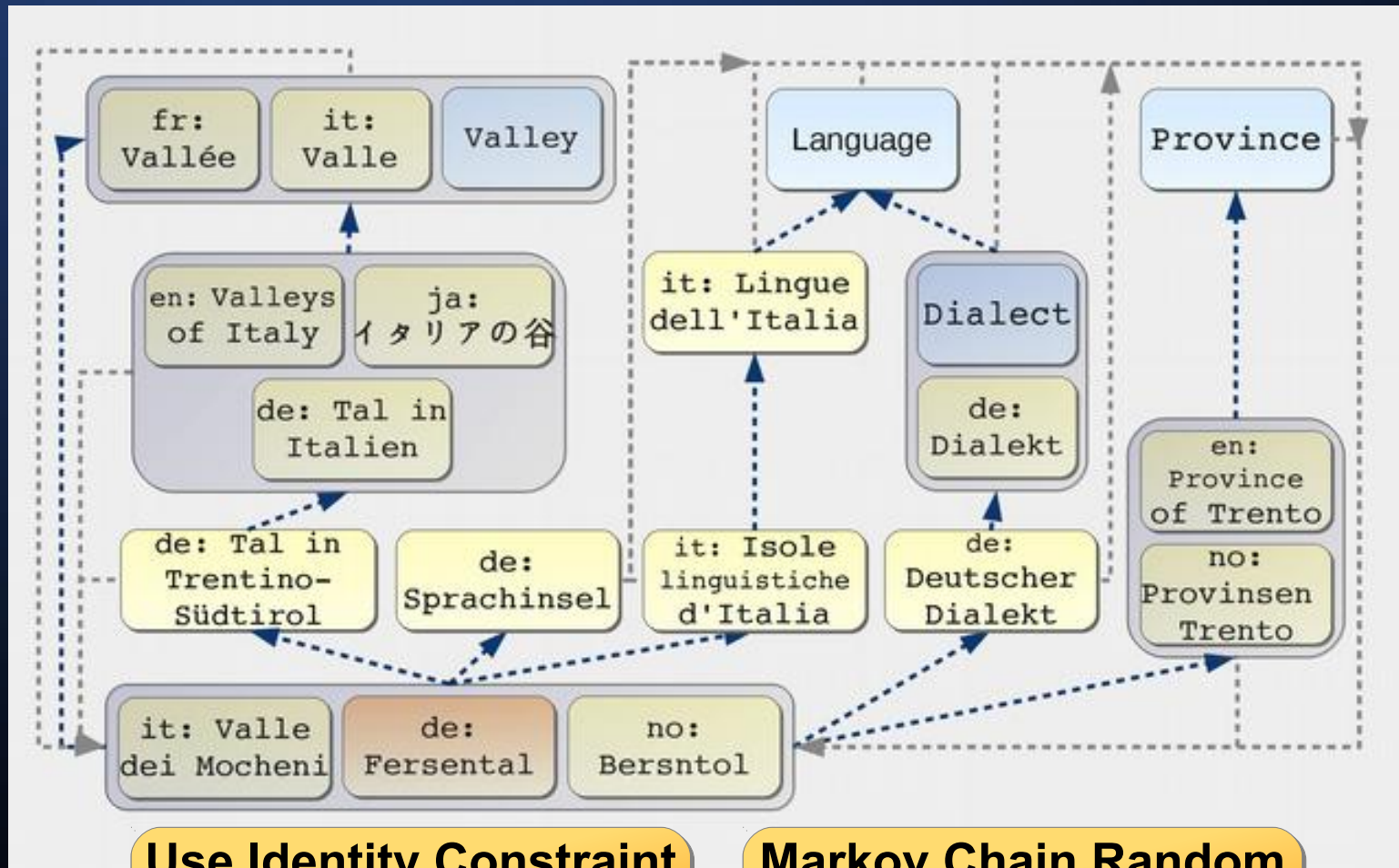
Minority Languages in Trentino - Census 2011



Taxonomic Integration: MENTA



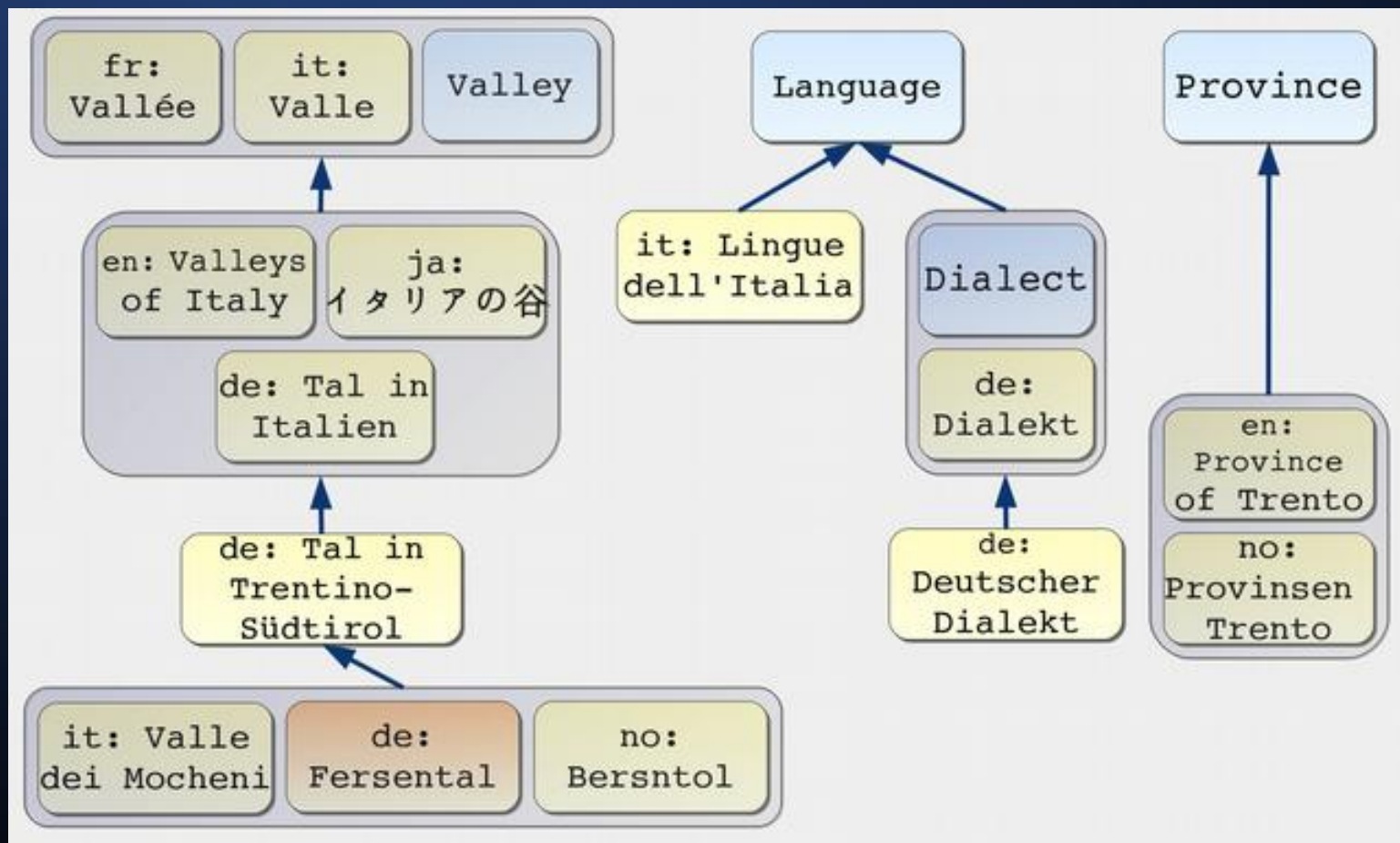
Taxonomic Integration: MENTA



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	<ol style="list-style-type: none">1. valley2. natural depression3. geological formation	Valleys Valleys of Italy Valleys of Trentino / Alto Adige
Cagayan National High School	<ol style="list-style-type: none">1. secondary school2. school3. educational institution	Secondary school School High schools in the Philippines
The Spanish Tragedy	<ol style="list-style-type: none">1. book2. publication3. piece of work	Book British plays Plays

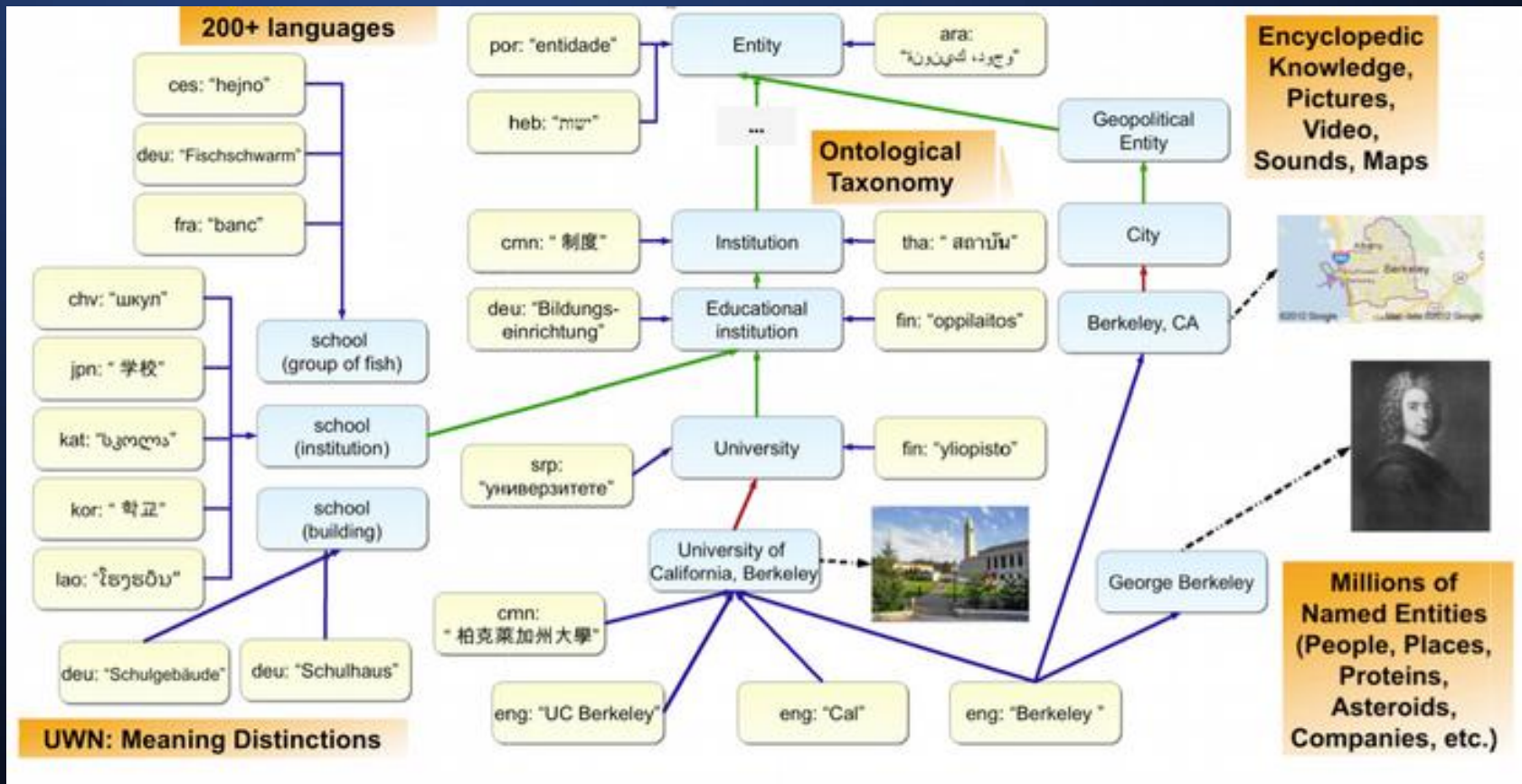
Knowledge Integration



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

Id	Subject	Property	Object	Time	Location	Keyword
?id0:	?c	type	company		nearby	San Francisco, 100
?id1:	?f	created	?c	during	[1910,2010]	

query



YAGO2:
WWW 2011
Best Demo Award

O'Reilly Media	Mendocino Brewing Company	IronMind	Blizzard North	Men
Sierra Nevada Brewing Company	Digidesign	Be Inc.	NetObjects	
THX	Gilead Sciences	Crystal Dynamics	Fox Paine &	
U.S. Venture Partners	Worldwide Aeros Corp	Tumbleweed Communications	NetS	
Monster Cable Products		Thinking Pictures	Ligh	

1978 1979 1980 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2010 2020 2030

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

Example: Lexvo.org

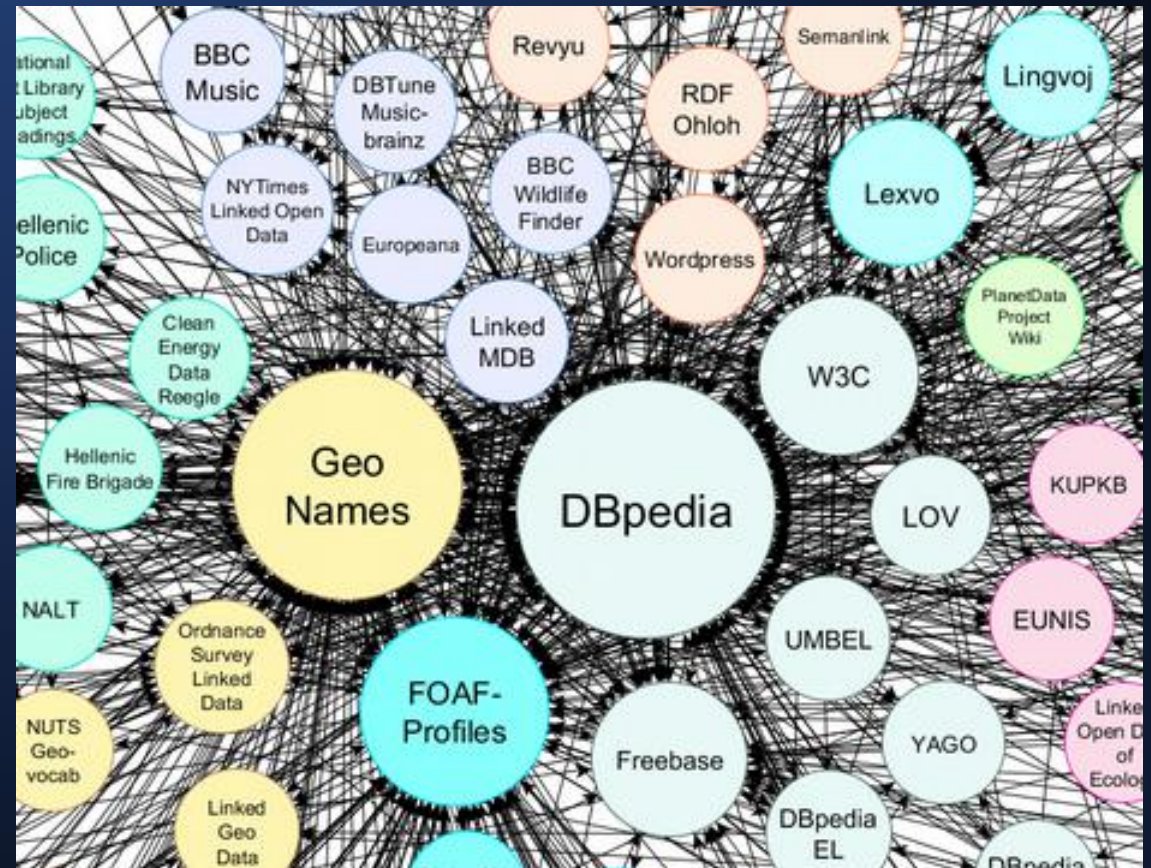
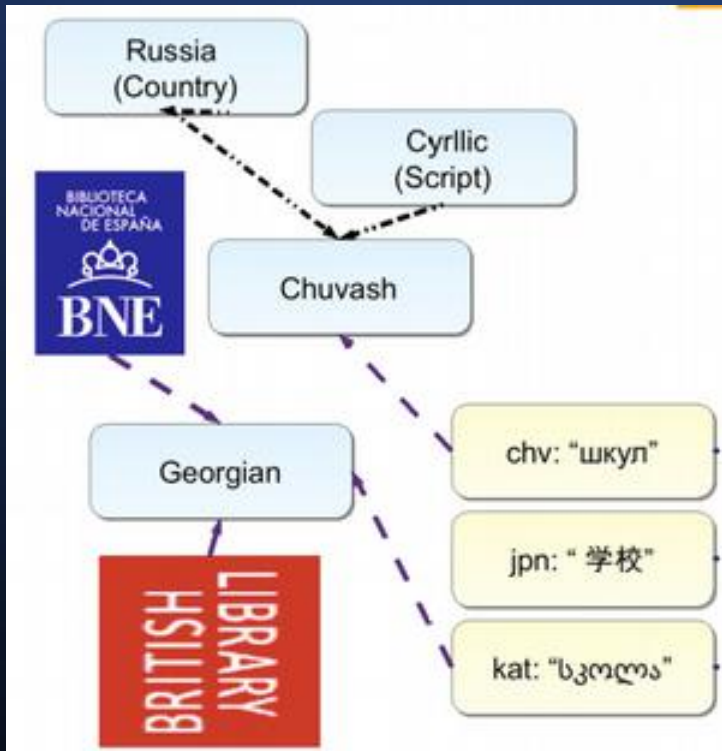
- **languages:** WordNet, ISO 639-3, Ethnologue, Linguist List, Wikipedia
- **language families/collections:** WordNet, ISO 639-5, Wikipedia
- **writing systems:** ISO 15924, Unicode CLDR
- **geographical regions:** Ethnologue, Unicode CLDR

The screenshot shows the Wikipedia article for 'Votic language'. The article text is partially visible, discussing the language's status and its relation to other languages in the region. On the right side, there are two tables: 'Words' and 'Language codes'. The 'Words' table lists 'Votic', 'Votian', and 'Votisch'. The 'Language codes' table lists ISO 639-1 (None), ISO 639-2 (vot), and ISO 639-3 (vot).

Words	
Wiktionary, glossary	
Speaks in	Russia
Region	VLC
Total speakers	20,000 (2014)
Language family	Votic <ul style="list-style-type: none">Finnic<ul style="list-style-type: none">Finno-Ugric<ul style="list-style-type: none">Uralic<ul style="list-style-type: none">Votic

Language codes	
ISO 639-1	None
ISO 639-2	vot
ISO 639-3	vot

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:

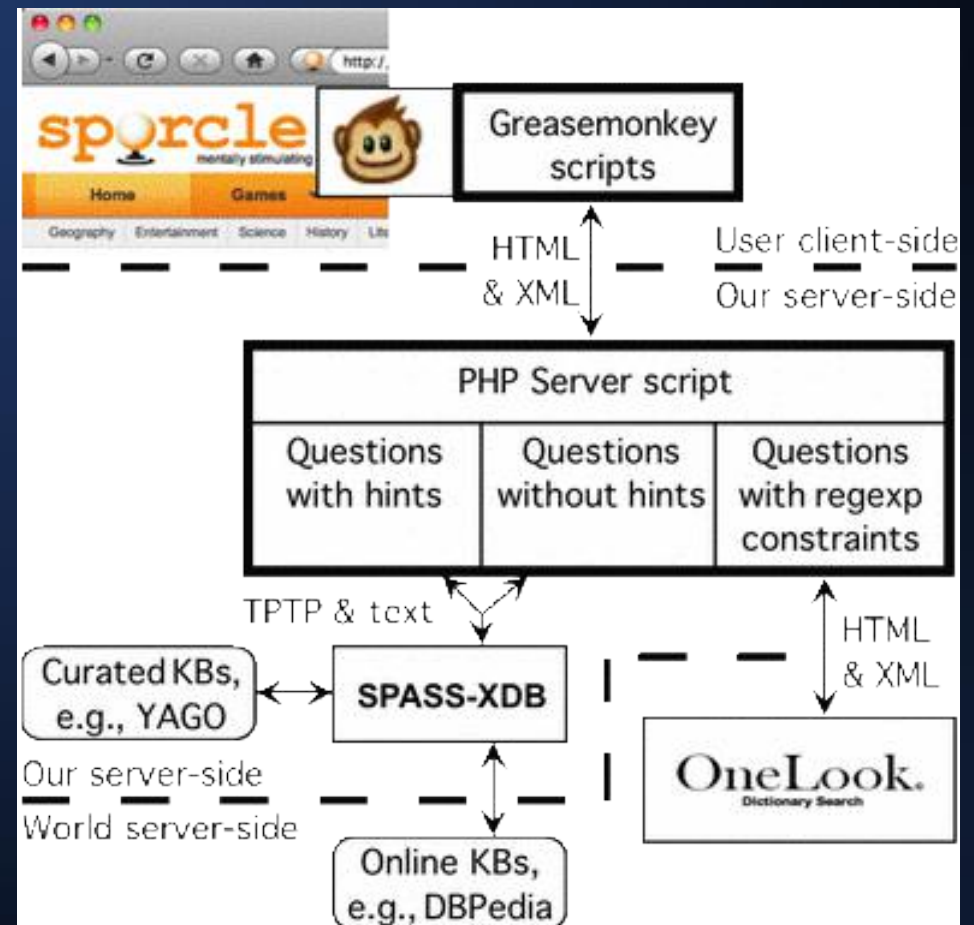
```
(capitalCity Y X)
```

YAGO-SUMO

SUMO's axioms
attached to
real-world instances,
overall many millions
of axioms

```
(exists ?YEARNO ?MONTHNO ?DAYNO  
  (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?**

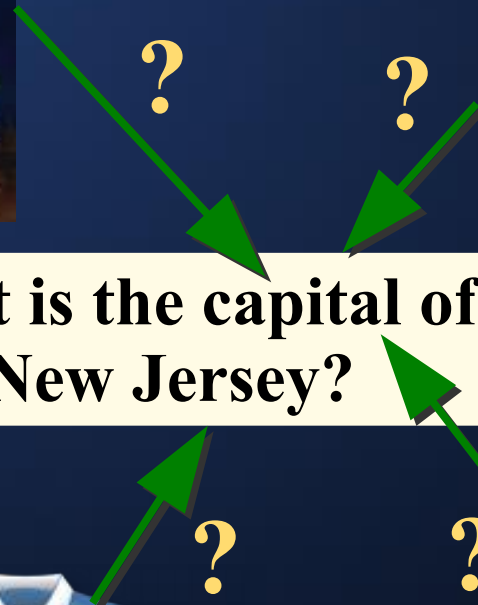


?

?

?

?



Question Answering



Buy Jaguar in Beijing

Question Answering



Buy Jaguar in Beijing



Jaguar

Which meaning?

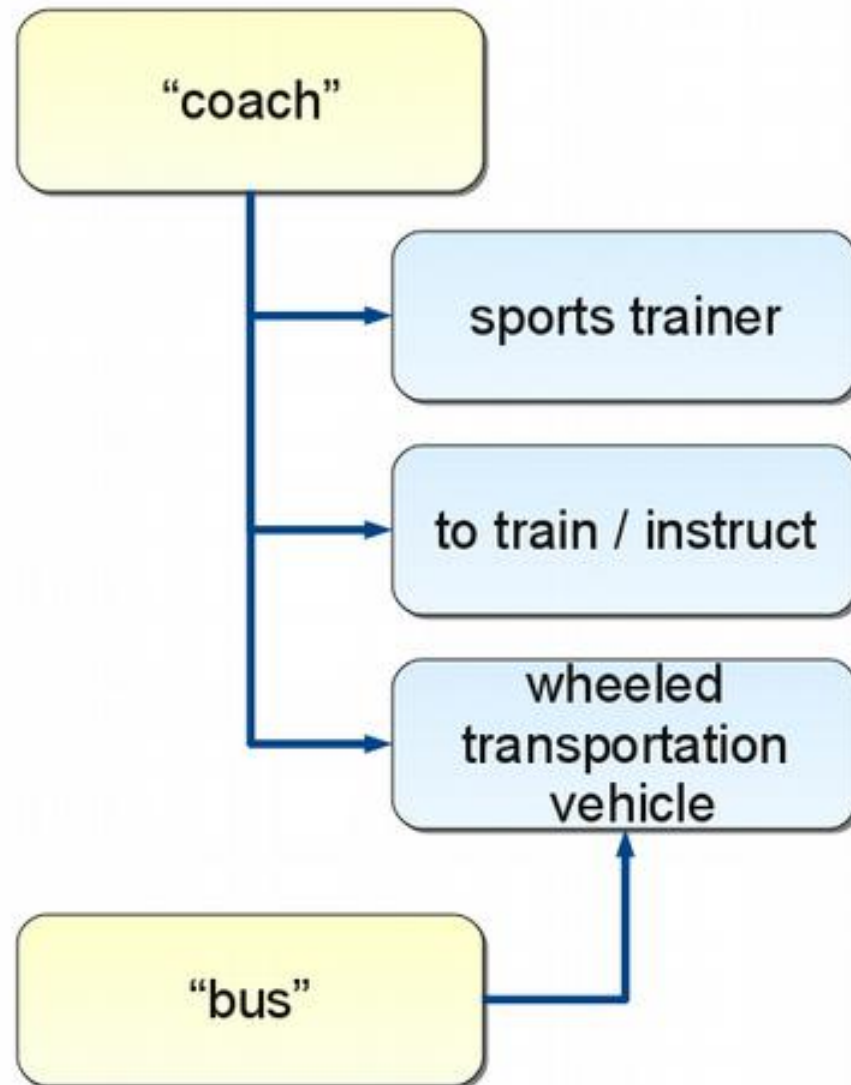


Question Answering

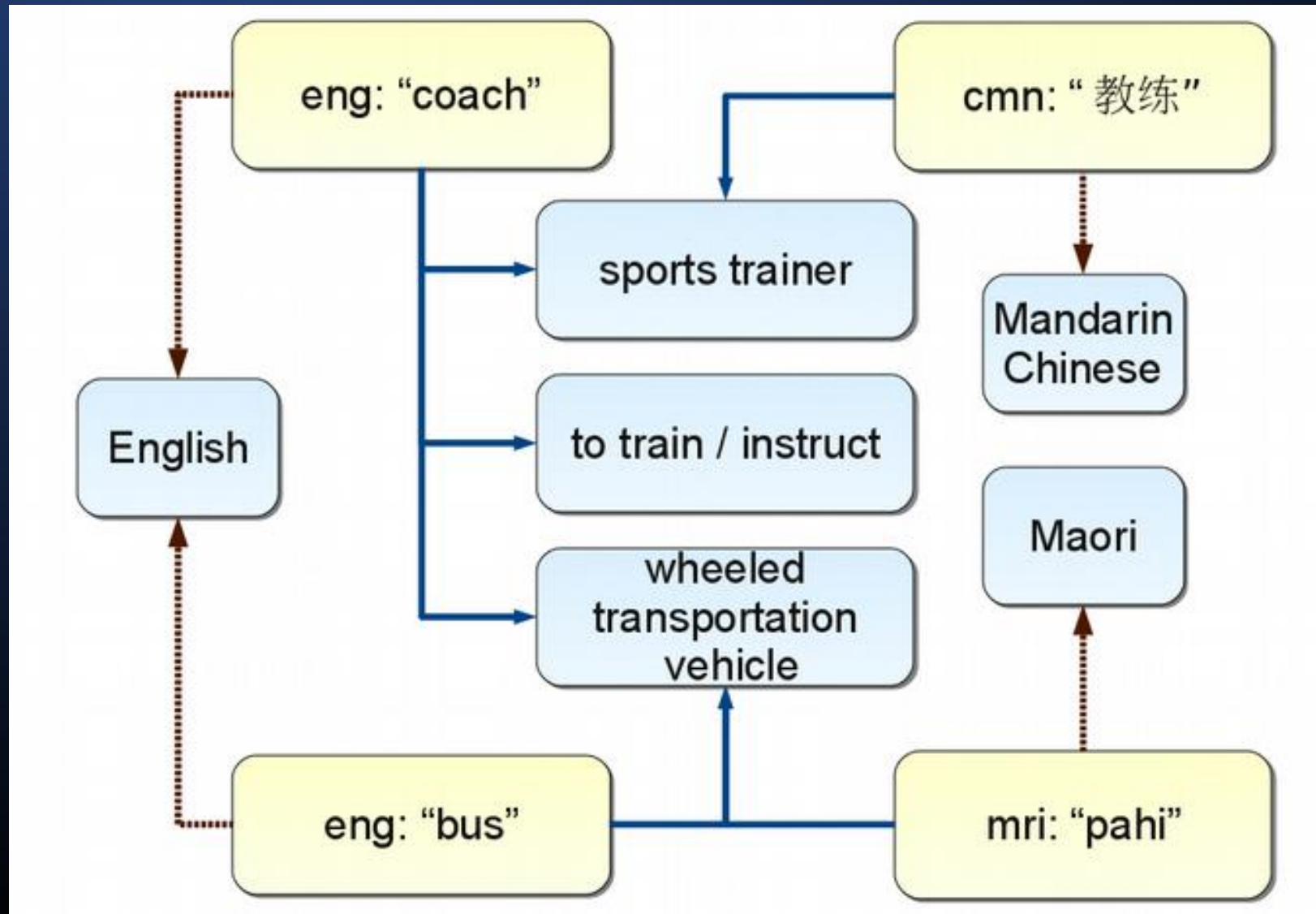


“Do Macs
now support
Windows?”

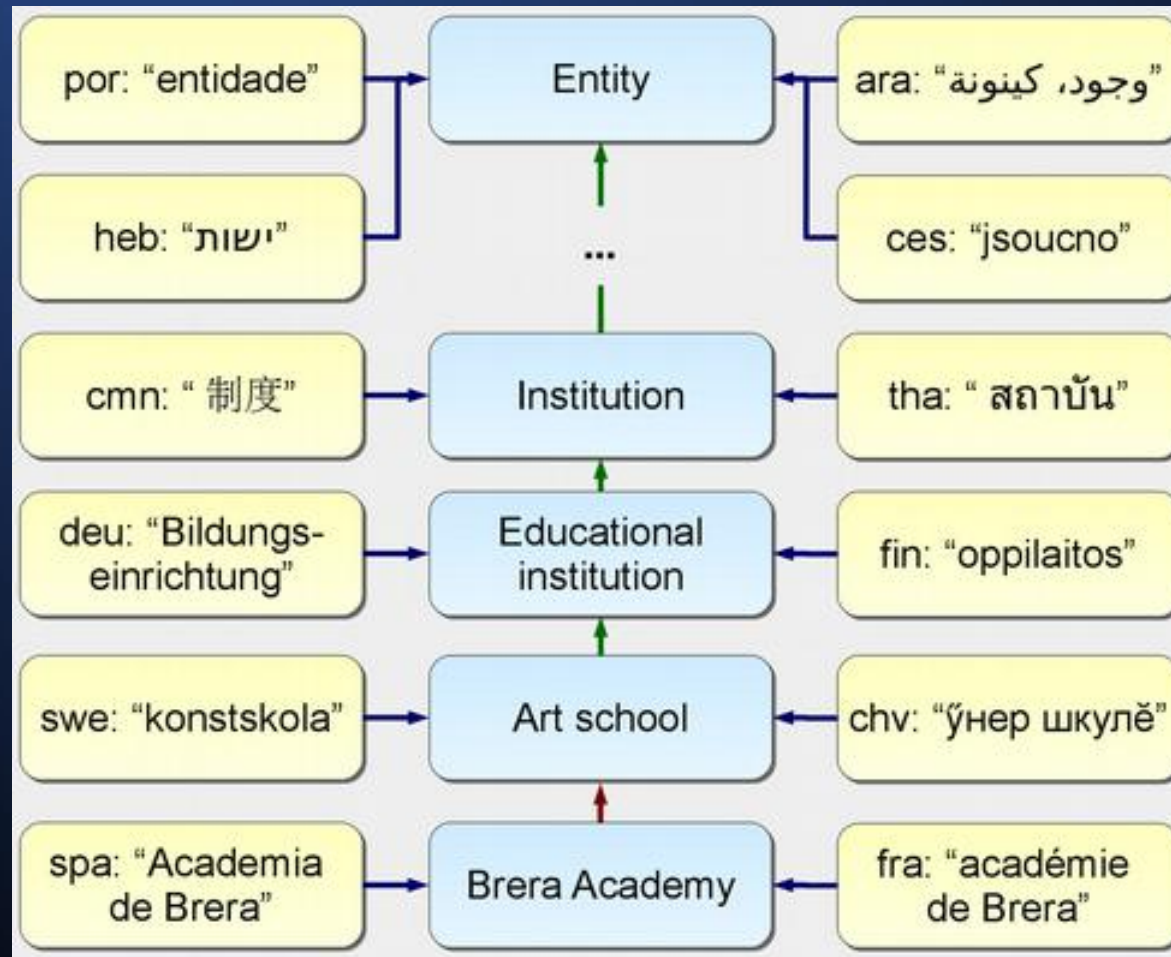
Lexical Knowledge Bases



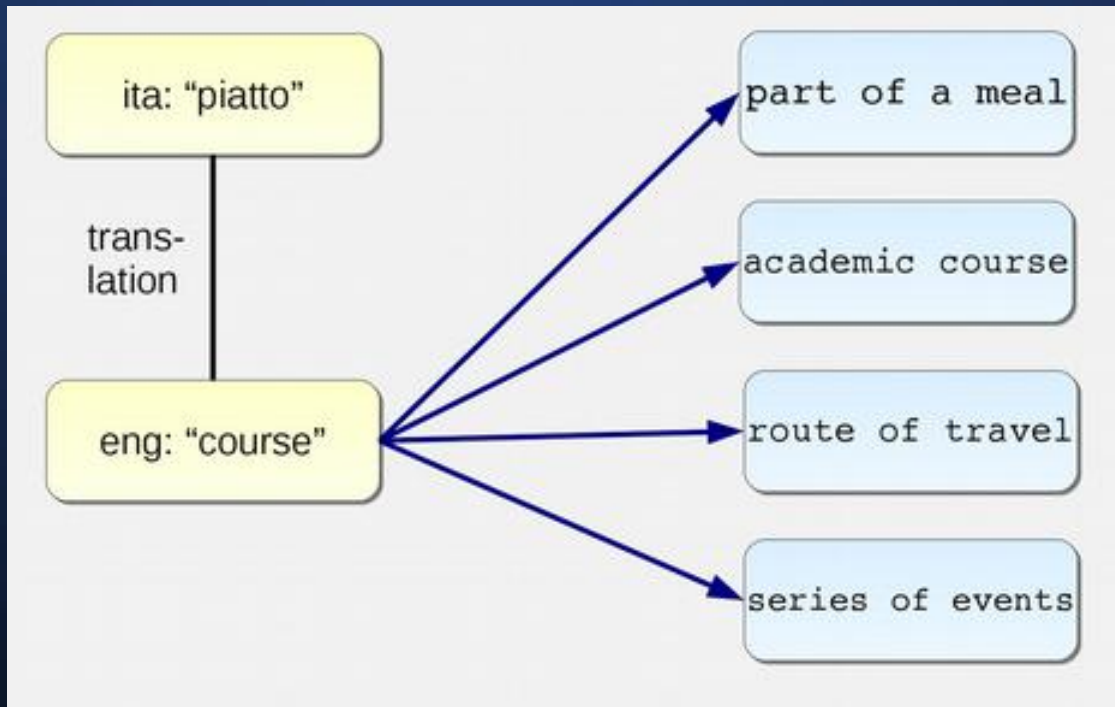
Multilingual Lexical Knowledge



MENTA

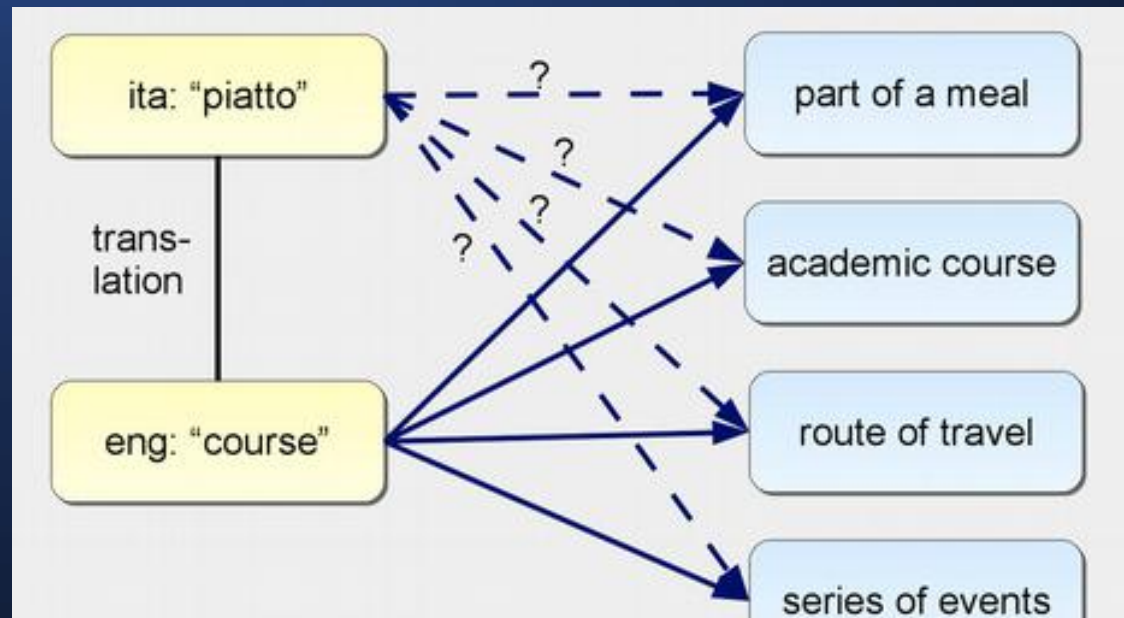


UWN: Universal Wordnet



Exploit millions of translations from the Web

UWN: “Universal Wordnet”

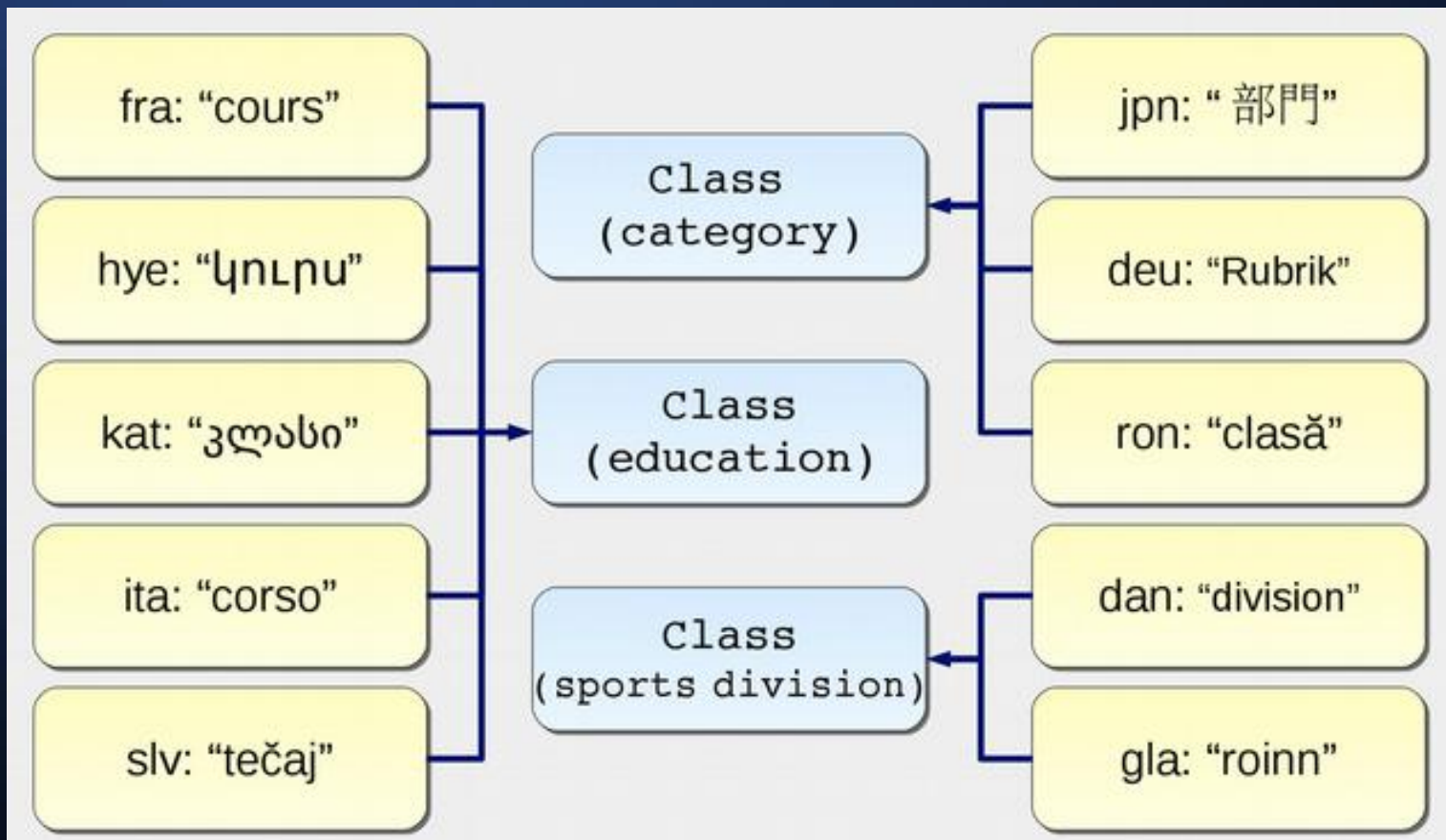


Learn regularized regression model to predict edge weights for a given pair of nodes

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

UWN

Query

Contact

UWN/MENTA — Universal Wordnet Project with MENTA extensions

Query

Word:

(case sensitive)

Language:

(ISO 639-3 code, e.g. "eng" for English)

Spanish (spa)

Instructions:

Enter a word in the first field with proper case (i.e., "jaguar" is considered different from "Jaguar"). Make sure the correct language in English and rely on the autocompletion feature or enter an ISO 639-3 code. You will then be directed to a page list

Examples:

- ["casa"](#) (Spanish word)
- [the Max Planck Society](#) (named entity)
- [to perceive by sight](#) (verbal concept)
- [of great physical weight](#) (adjectival concept)
- [ophthalmology](#) (nominal concept)

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

[More Information](#)

UWN

UWN Query Contact [Login](#)

spa: atrasar

[New Query](#)

Information	
has URI	http://lexvo.org/id/term/spa/atrasar

Meaning	
means	(verb) cause to move more slowly or operate at a slower rate; "This drug will retard your heart rate" retard

Query

Word: (case sensitive)

Language: (ISO 639-3 code, e.g. "eng" for English)

Lexvo © 2008-2015 [Gerard de Melo](#). [Contact](#) [Legal Information / Imprint](#)

<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 discussion 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 dù

cmn: 减低速度

cmn: 减低速度

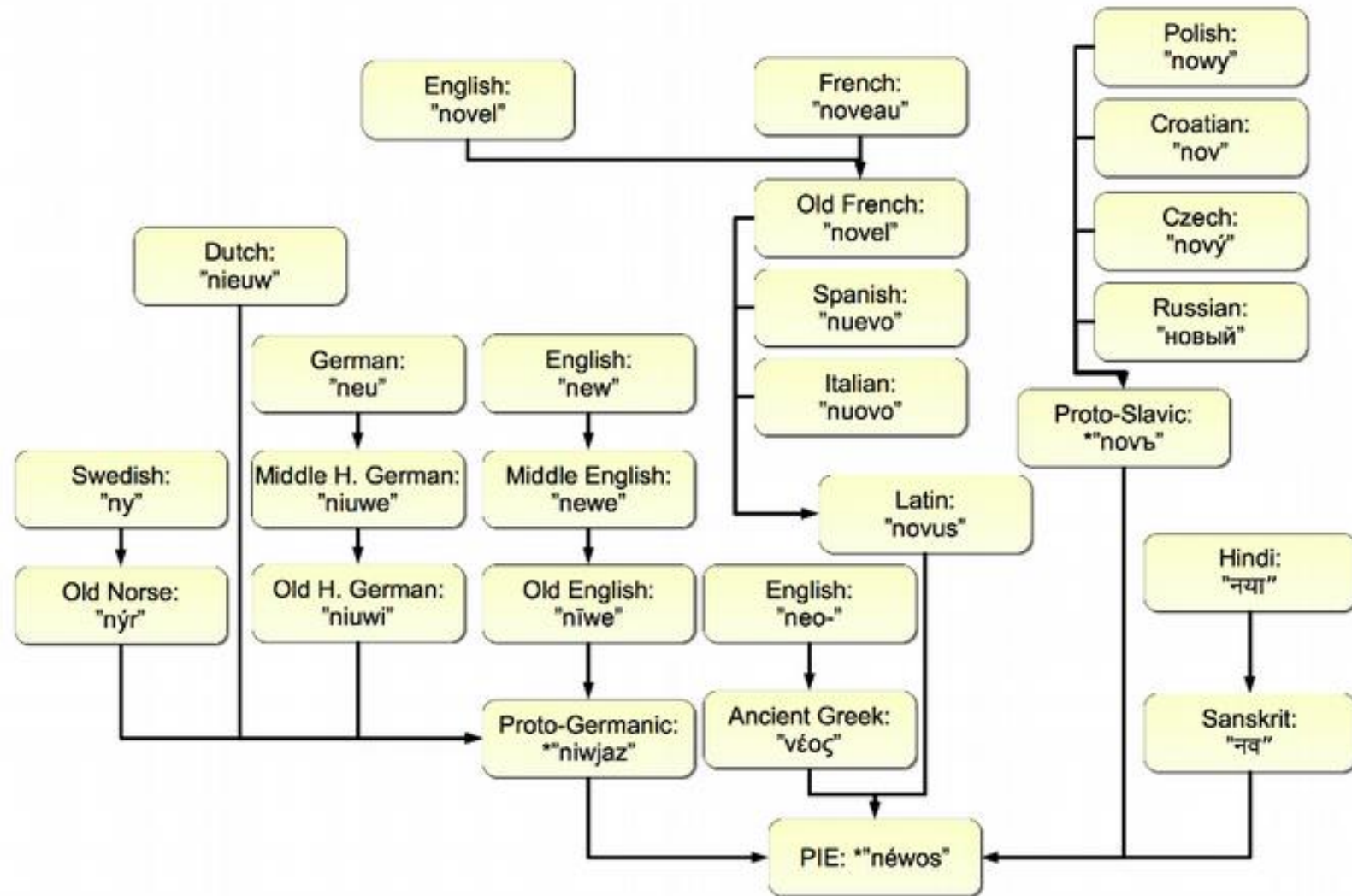
German

lexicalization deu: abbremsen

deu: verlangsamen

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

Etymological Wordnet



Real Understanding?



- 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
 - Printable version
- Language
 - Afrikaans
 - Alemannisch
 - العربية
 - Български

Article Talk

Read Edit View history

Search

Ludwig Wittgenstein

From Wikipedia, the free encyclopedia
(Redirected from Wittgenstein)

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

Ludwig Josef Johann Wittgenstein (/ˈwɪtʃənˈstaɪn/^[4] German: [ˈvʏtʃənˌʃtaɪn]; 26 April 1889 – 29 April 1951) was an Austrian-British philosopher who worked primarily in logic, the philosophy of mathematics, the philosophy of mind, and the philosophy of language. From 1929–1947, Wittgenstein taught at the University of Cambridge.^[6] During his lifetime he published just one slim book, the 75-page *Tractatus Logico-Philosophicus* (1921), one article, one book review and a children's dictionary.^[7] His voluminous manuscripts were edited and published posthumously. *Philosophical Investigations* appeared as a book in 1953 and by the end of the century it was considered an important modern classic.^[8] Philosopher Bertrand Russell described Wittgenstein as "the most perfect example I have ever known of genius as traditionally conceived; passionate, profound, intense, and 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 poor 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 front 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 Investigations*. The early Wittgenstein was concerned with the logical relationship between propositions and the world and believed that by providing an account of the logic underlying this relationship, he had solved all philosophical problems. The later Wittgenstein rejected many of the assumptions of the *Tractatus*, arguing that the meaning of words is best understood as their use within a given language-game.^[15]

Wittgenstein's influence has been felt in nearly every field of the humanities and social sciences, yet there are diverging interpretations of his thought. In the words of his friend and colleague Georg Henrik von Wright:

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

Ludwig Wittgenstein



Photographed by Ben Richards, Swansea, 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 Logical necessity

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

Search: eng ⌵ <id_4h2ou9_16x_vkf2e7>

<Cécilia_Attias> <isMarriedTo> <Nicolas_Sarkozy> hasFactId

<extractionSource>	⊠ → <Cécilia_Attias> → <http://en.wikipedia.org/wiki/Cécilia_Attias> →
<occursUntil>	"2007-##-##"^^xsd:date →
<occursSince>	"1996-##-##"^^xsd:date →

YAGO: isMarriedTo predicate

Marriage /people/marriage

Spouse /people/marriage/spouse

Spouse
Cécilia Sarkozy
Nicolas Sarkozy

From /people/marriage/from
10/23/1996

To /people/marriage/to
10/15/2007

Type of union /people/marriage/type_of_union
Marriage

Location of ceremony /people/marriage/location_of_ceremony
Neuilly-sur-Seine

Freebase: Marriage Entity

Challenge:
Modelling
Differences

Relation Integration: FrameBase.org

Clarity_of_resolution
Clemency
Closure
Clothing
Clothing_parts
Cogitation
Cognitive_connection
Cognitive_impact
Coincidence
Collaboration
Collocation_image_schema
Colonization
Color
Color_qualities
Come_together
Coming_to_be
Coming_to_believe
Coming_up_with
Commemorative
Commerce_buy
Commerce_collect
Commerce_goods-transfer
Commerce_money-transfer
Commerce_pay
Commerce_scenario
Commerce_sell
Commercial_transaction
Commitment
Committing_crime
Commonality
Communicate_categorization
Communication
Communication_manner
Communication_means
Communication_noise
Communication_response
Commutation
Commutative_process
Commutative_statement
Compatibility
Competition
Complaining
Completeness

Commerce_money-transfer Lexical Unit Index

Definition

The subframe of the Commercial_transaction frame which involves the transfer of Money from the Buyer to the Seller (in exchange for the Goods).

Semantic Type: Non-Lexical Frame

Frame Elements

Core Elements

Buyer [Byr] The **Buyer** wants the **Goods** and offers **Money** to a **Seller** in exchange for them.
 Jess bought a coat.
 Lee sold a textbook **to Abby**.

Excludes: Exchangers

Exchangers [exch] The individuals participating in the commercial exchange, otherwise encoded as **Buyer** and **Seller**.

Goods [Gds] The FE Goods is anything (including labor or time, for example) which is exchanged for Money in a transaction.
 Kim bought **the sweater**.
 Kim sold **the sweater**.

Money [Mny] Money is the thing given in exchange for Goods in a transaction.
 Pat paid **14 dollars** for a movie ticket.
 Sam sold the car **for \$12,000**.

Seller [Slr] The **Seller** has possession of the **Goods** and exchanges them for **Money** from a **Buyer**.

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



Scripts
- Movies
- TV Series
- Sitcoms
Novels



Semantic Parsing

ClausIE
OpenNLP
ILP for WSD

Graph Inference

Statistical Priors
PSL Inference

Taxonomy Construction

Synsets
Hierarchy

Knowlywood

Activity Knowledge Base

CIKM 2015

Knowlywood: Human Activities



Learning Common-Sense

WebChild

AAAI 2014
WSDM 2014
AAAI 2011



pop-singer-n¹
hasAppearance
hot-a³

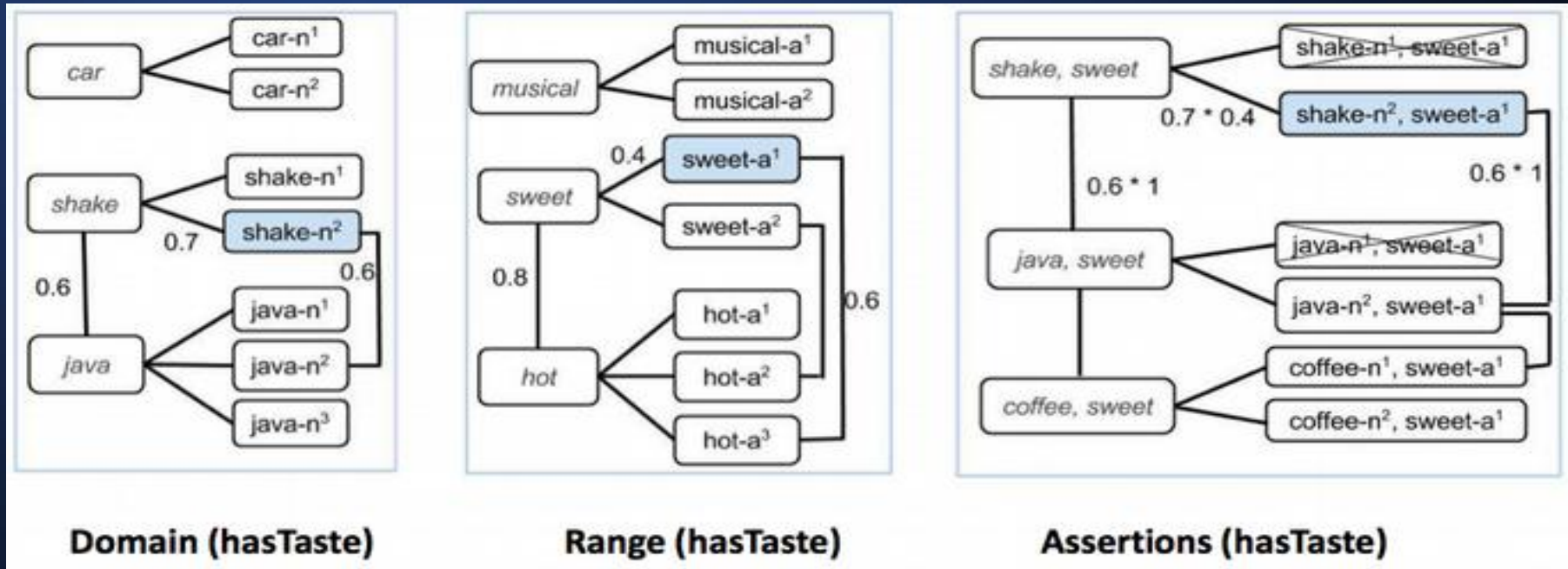


chili-n¹
hasTaste
hot-a⁹



volcano-n¹
hasTemperature
hot-a¹

Learning Common-Sense



WebChild

(Tandon et al. WSDM 2014):

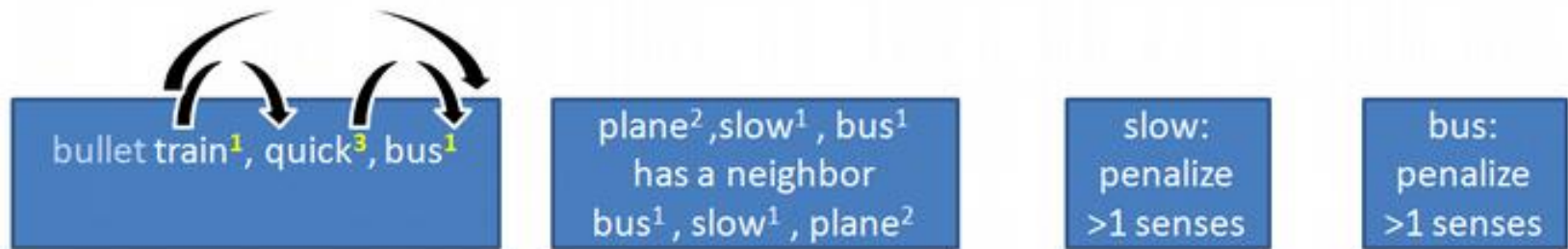
Seed label loss + Similar node diff label loss + Regularize

$$\left[\mu_1 \sum_v p_v^{inj} \sum_l (Y_{vl} - \hat{Y}_{vl})^2 + \mu_2 \sum_{v,u} p_v^{cont} W_{vu} \sum_l (Y_{vl} - \hat{Y}_{ul})^2 + \mu_3 \sum_{vl} (\hat{Y}_{vl} - R_{vl})^2 \right]$$

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>



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

subject to

$$\dots$$

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

train¹, slow¹, plane²
 ≡
 plane², fast¹, train¹

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 Karttunen's work

Gerard de Melo

ImpLex: Implicative Commitments

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

de Melo &
de Paiva 2014

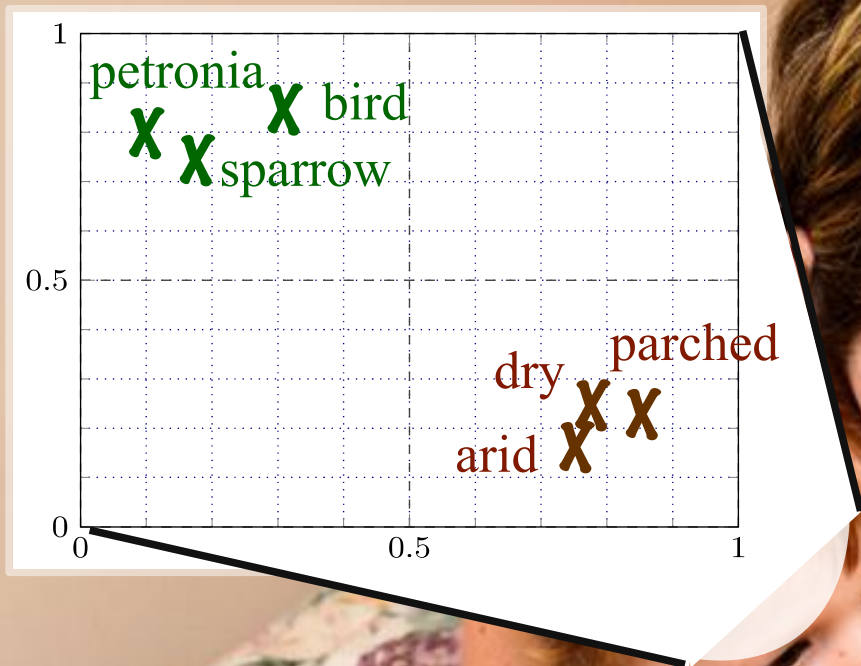
It is **unlikely** that
Apple has acquired
the technology.



Have they
acquired it?

?

Word Vectors



Why are Word Vectors Successful?

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

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?

syntactic




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?

syntactic

semantic!




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?

syntactic

semantic!

syntactic?



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?

syntactic

semantic!

syntactic?

?



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?

syntactic

semantic!

syntactic?

?

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?

syntactic

semantic!

syntactic?

?



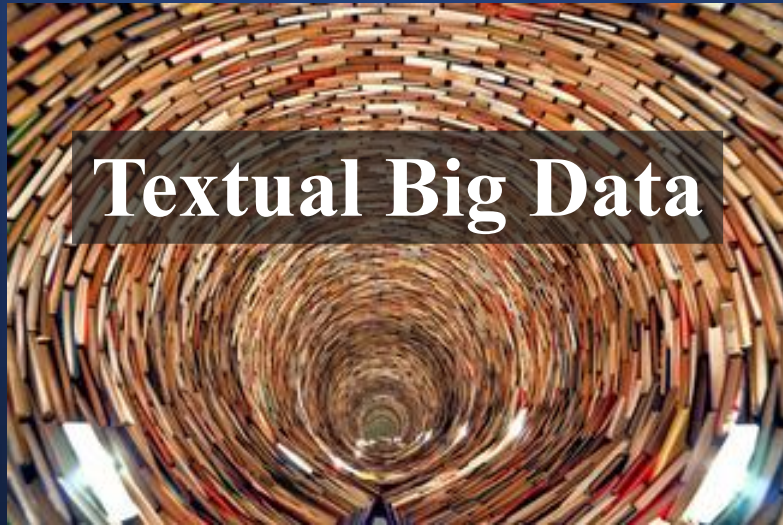
The **Roman Empire** was remarkably **multicultural**, with "a rather astonishing cohesive capacity" to create a sense of shared identity while encompassing **diverse peoples** within its **political** system over a long span of time.

The diagram consists of four green labels at the top: 'syntactic', 'semantic!', 'syntactic?', and '?'. Below these labels are four green curved arrows pointing downwards to specific words in the sentence below. The first arrow points from 'syntactic' to 'Roman'. The second arrow points from 'semantic!' to 'Empire'. The third arrow points from 'syntactic?' to 'multicultural'. The fourth arrow points from '?' to 'political'.

Textual Big Data



Better Word Vectors by Exploiting Knowledge



Joint Training

**Better
Word Embeddings**

Best Paper Award
at NAACL 2015
Vector Space
Modeling Workshop

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

Jl65	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 McConnell would like to...



Image: Microsoft Cortana

Summary

■ Knowledge Integration and Organization

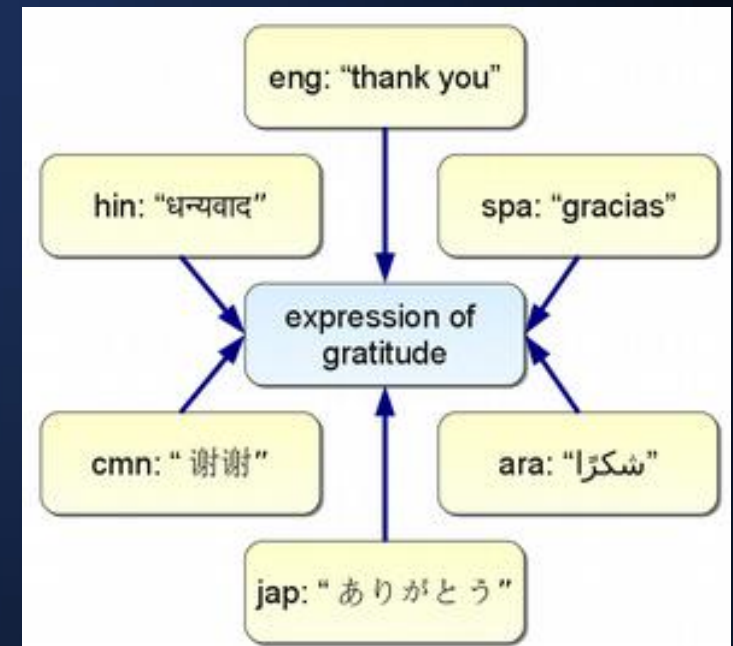
- ▶ MENTA taxonomy
- ▶ Lexvo.org Linked Data

■ Natural Language Coupling

- ▶ Universal WordNet
- ▶ FrameBase.org

■ Multimodal and Cognitive Knowledge

- ▶ Knowlywood
- ▶ WebChild



More Information:

www.demelo.org
gdm@demelo.org