max planck institut
informatik

Récolte des Connaissances de la Toile

Gerhard Weikum

Max Planck Institute for Informatics
Saarland University
http://www.mpi-inf.mpg.de/~weikum/

Knowledge Harvesting From the Web

Gerhard Weikum

Max Planck Institute for Informatics
Saarland University
http://www.mpi-inf.mpg.de/~weikum/

Acknowledgements

Google

living
know
ledge

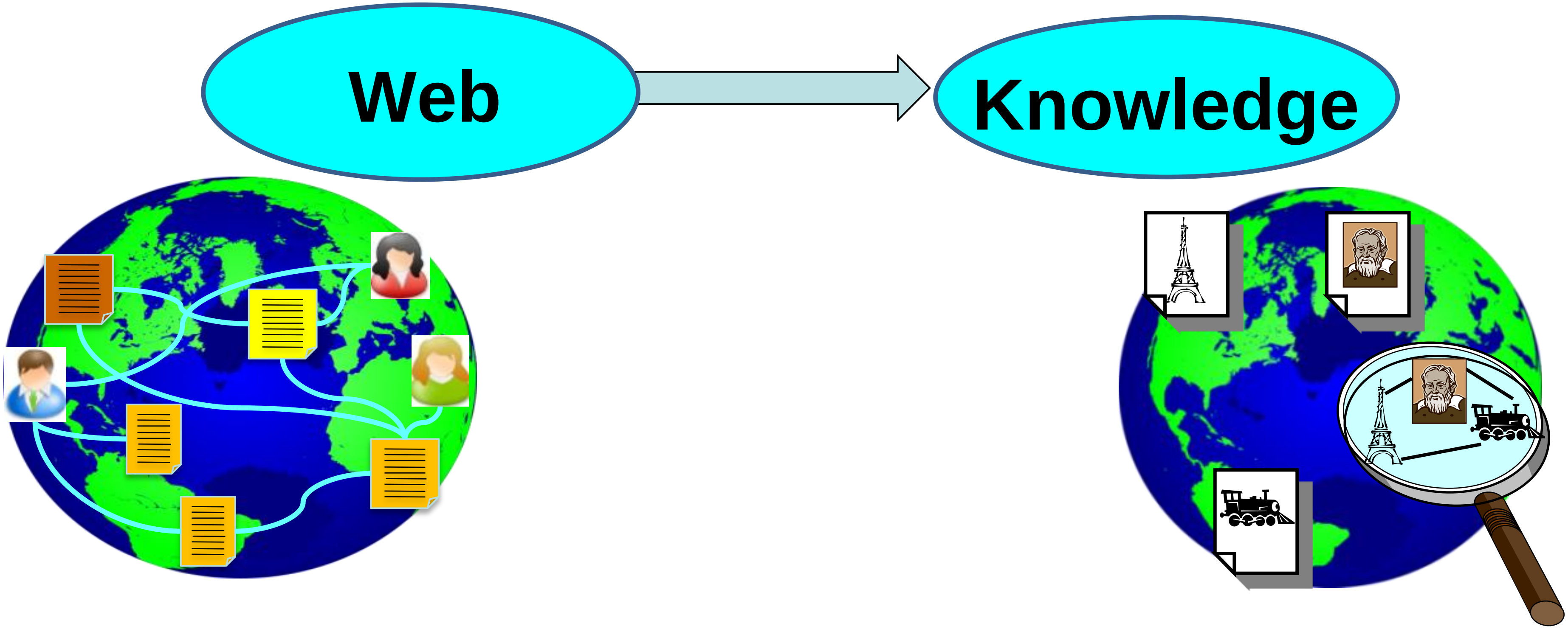
LAWA

SEVENTH FRAMEWORK PROGRAMME

M2CI
CLUSTER OF EXCELLENCE

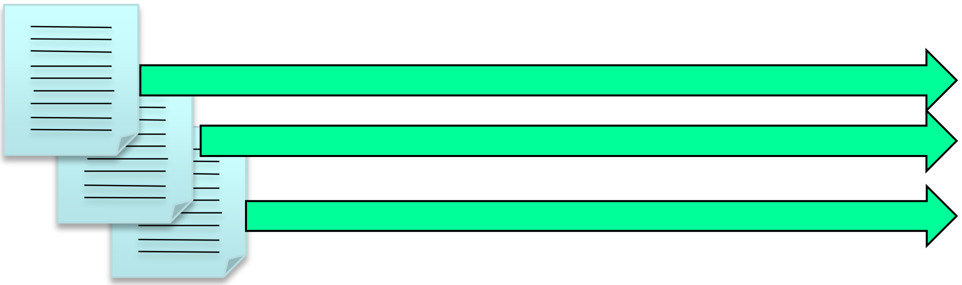
DFG Deutsche Forschungsgemeinschaft

Knowledge from the Web



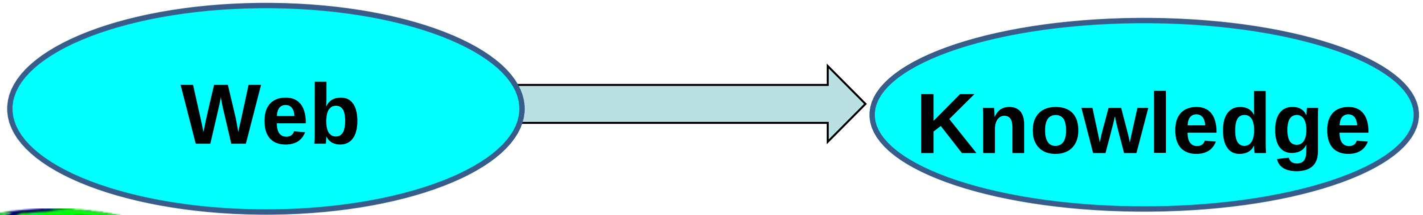
Web

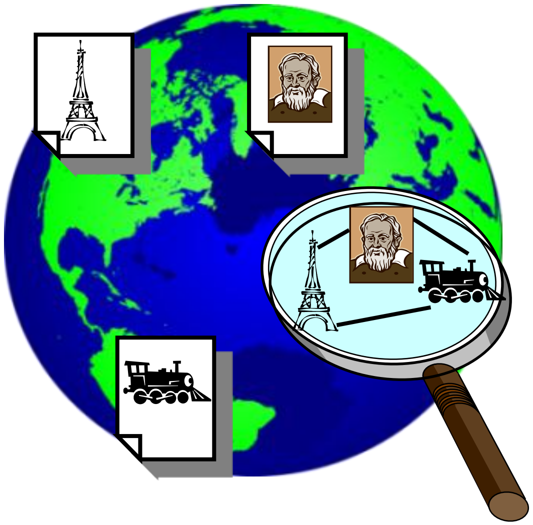
Knowledge

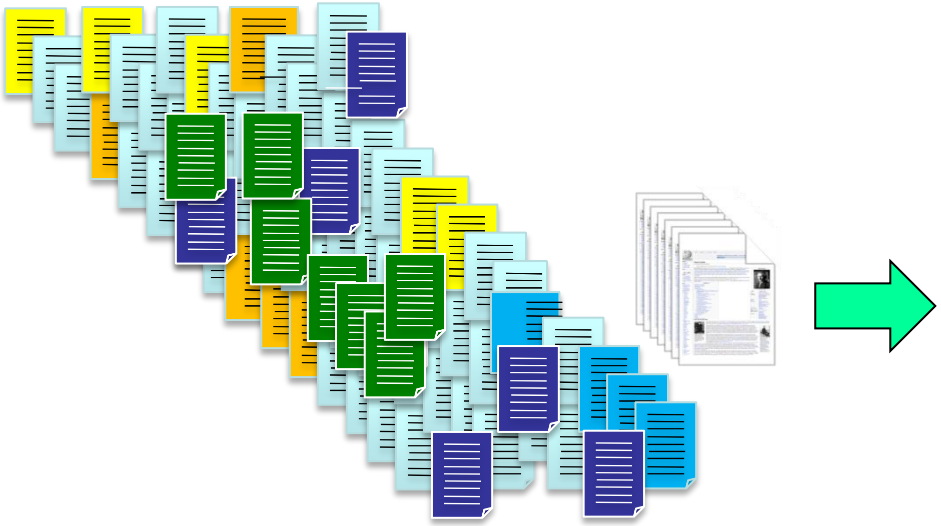


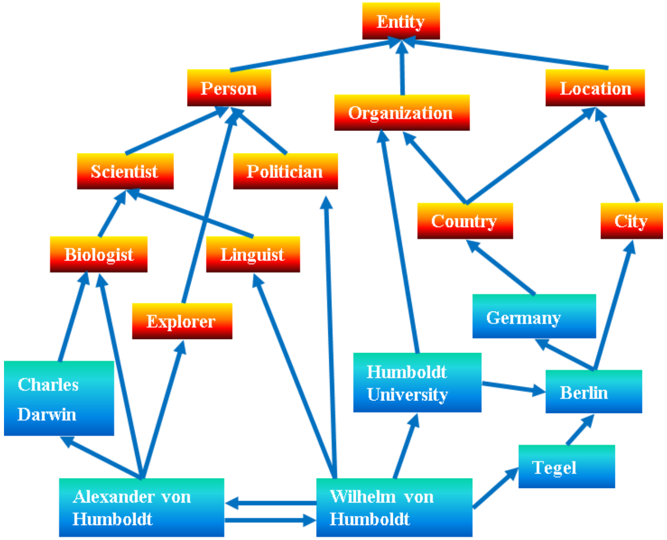
Movie Title	Price	Rating
Sherlock Holmes 2	15.99	7.5
Inglorious Basterds	11.99	8.3
OSS 117: Lost in Rio	7.99	6.7

Knowledge from the Web









Outline

- ★ Knowledge for Machines
- ★ Construction of Knowledge Bases
- ★ KB Population from Text & Web Pages
- ★ Knowledge for Intelligent Applications
- ★ Opportunities & Challenges
- ★ Conclusions

Spectrum of Machine Knowledge

factual:
bornIn (SteveJobs, SanFrancisco), hasFounded (SteveJobs, Pixar),
hasWon (SteveJobs, NationalMedalOfTechnology), livedIn (SteveJobs, PaloAlto)

taxonomic (ontology):
instanceOf (SteveJobs, computerArchitects), instanceOf(SteveJobs, CEOs)
subclassOf (computerArchitects, engineers), subclassOf(CEOs, businesspeople)

lexical (terminology):
means (“Big Apple“, NewYorkCity), means (“Apple“, AppleComputerCorp)
means (“MS“, Microsoft) , means (“MS“, MultipleSclerosis)

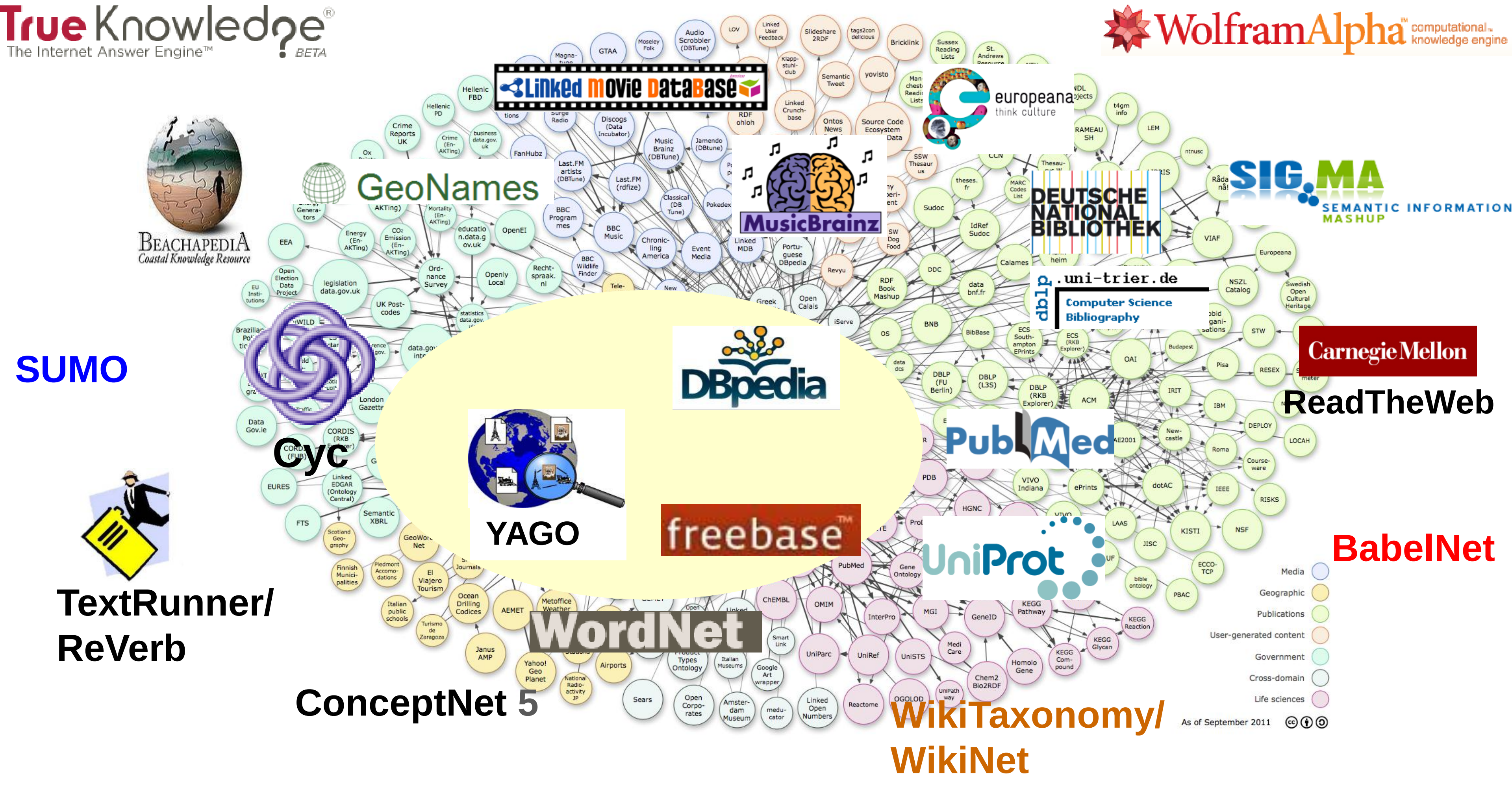
multi-lingual:
meansInChinese („乔戈里峰“, K2), meansInUrdu („کے ٹو“, K2)
meansInFr („école“, school (institution)), meansInFr („banc“, school (of fish))

temporal (fluents):
hasWon (SteveJobs, NationalMedalOfTechnology)@1985
marriedTo (AlbertEinstein, MilevaMaric)@[6-Jan-1903, 14-Feb-1919]
presidentOf (NicolasSarkozy, France)@[16-May-2007, 15-May-2012]

Knowledge Harvesting from Web Sources

(Automatic) Construction of Comprehensive Knowledge Bases

100 Mio's of entities, classes, relationships, common-sense properties



SUMO

Cyc

ReadTheWeb

YAGO

TextRunner/ ReVerb

ConceptNet 5

BabelNet

WikiTaxonomy/ WikiNet

Knowledge for Intelligence

Enabling technology for:

✶ disambiguation in written & spoken natural language

✶ deep reasoning (e.g. QA to win quiz game)

✶ machine reading (e.g. to summarize book or corpus)

✶ semantic search in terms of entities&relations (not keywords&pages)

✶ entity-level linkage for the Web of Data

★ Politicians who are also scientists?

★ European composers who have won film music awards?

★ French professors who founded Internet companies?

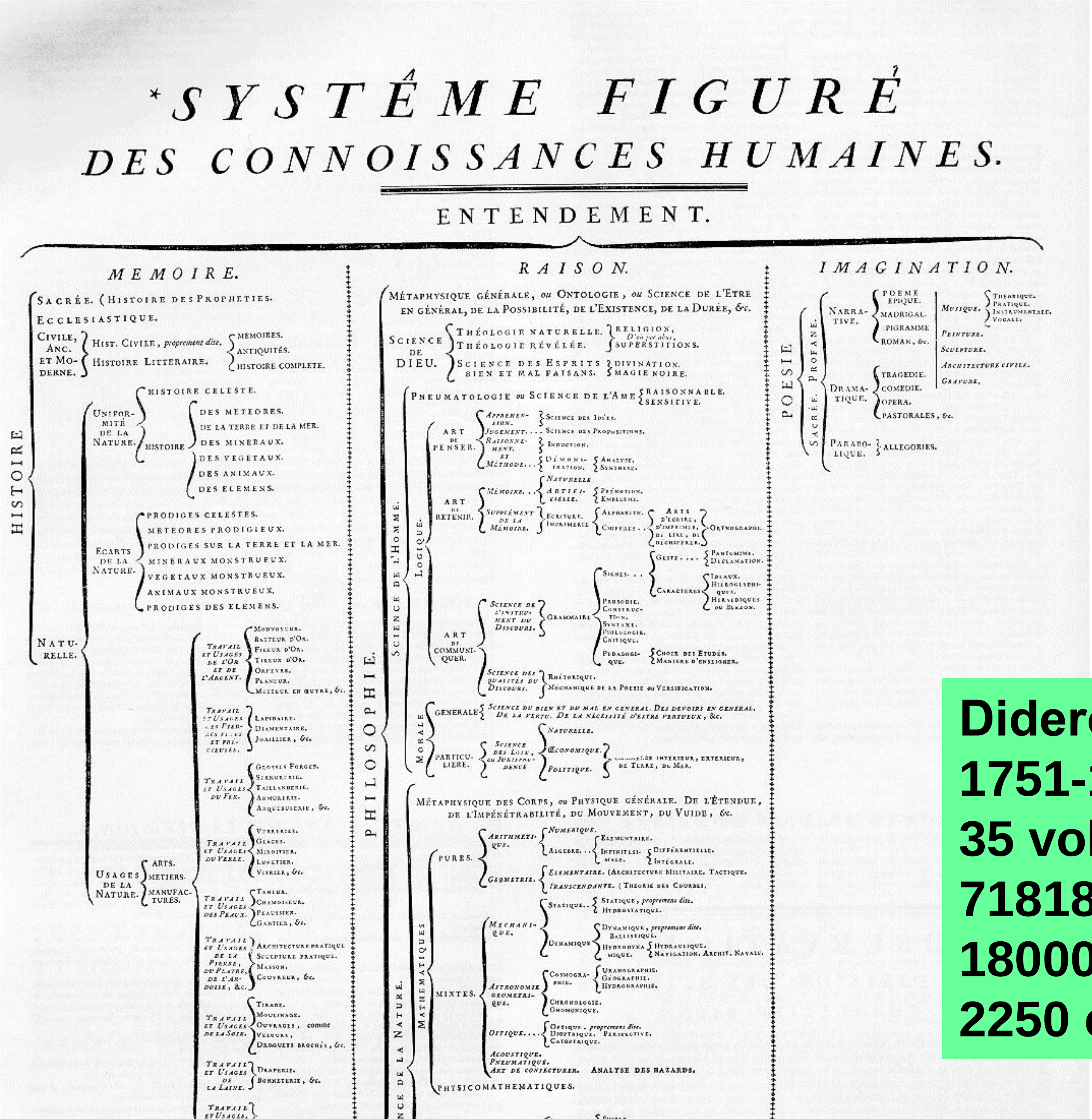
★ Relationships between
Alexander Pushkin, Evariste Galois, Johnny Ringo, and Hamlet?

★ Enzymes that inhibit HIV?
Influenza drugs for teens with high blood pressure?
...

Outline

- ✓ Knowledge for Machines
- ★ Construction of Knowledge Bases
- ★ KB Population from Text & Web Pages
- ★ Knowledge for Intelligent Applications
- ★ Opportunities & Challenges
- ★ Conclusions

The World’s First Knowledge Base

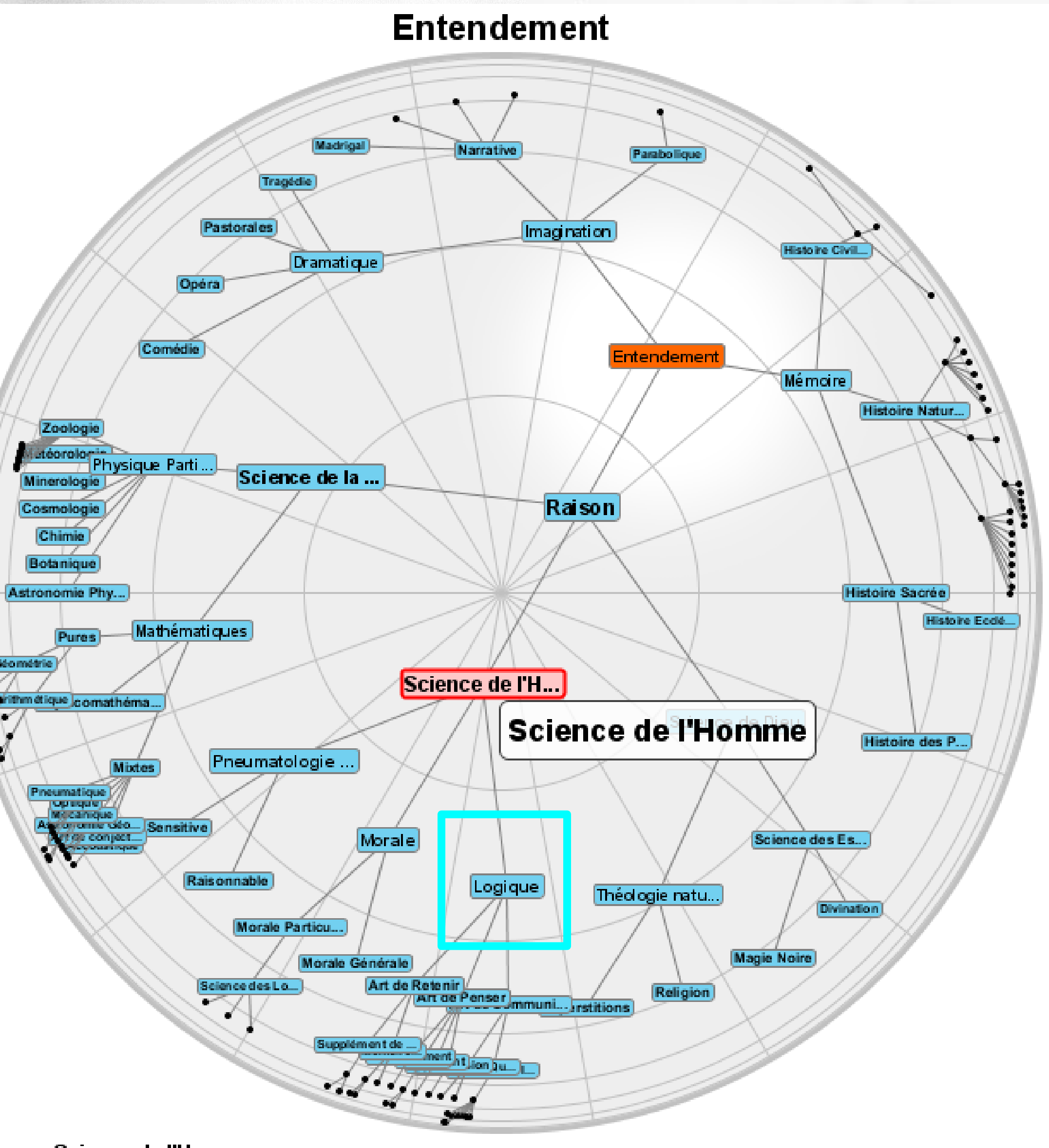


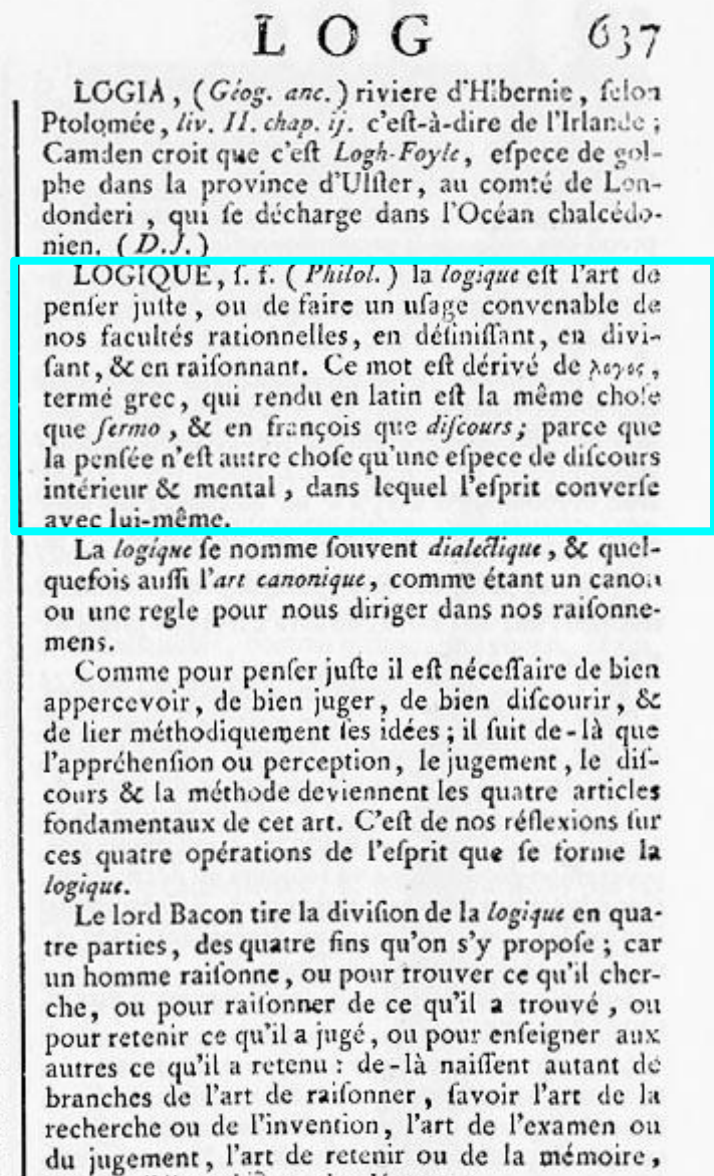
Denis Diderot (1713-1784)

Jean le Rond d‘Alembert (1717-1783)

Diderot & d‘Alembert (Eds.)
1751-1780
35 volumes, 4250 copies
71818 articles, 3129 figures
18000 pages, 20 Mio. words
2250 contributors

The World‘s First Knowledge Base





L O G 637

LOGIA, (*Géog. anc.*) riviere d'Hibernie, ſelon Ptolomée, *liv. II. chap. ij.* c'eſt-à-dire de l'Irlande; Camden croit que c'eſt *Logh-Foyle*, eſpece de golphe dans la province d'Ulſter, au comté de Londonderi, qui ſe décharge dans l'Océan chalcédonien. (*D. J.*)

LOGIQUE, ſ. f. (*Philol.*) la *logique* eſt l'art de penſer juſte, ou de faire un uſage convenable de nos facultés rationnelles, en définiſſant, en diviſant, & en raiſonnant. Ce mot eſt dérivé de λογος, terme grec, qui rendu en latin eſt la même choſe que *ſermo*, & en françois que *diſcours;* parce que la penſée n'eſt autre choſe qu'une eſpece de diſcours intérieur & mental, dans lequel l'eſprit converſe avec lui-même.

La *logique* ſe nomme ſouvent *dialectique*, & quelquefois auſſi l'*art canonique*, comme étant un canon ou une regle pour nous diriger dans nos raiſonnemens.

Comme pour penſer juſte il eſt néceſſaire de bien appercevoir, de bien juger, de bien diſcourir, & de lier méthodiquement ſes idées; il ſuit de-là que l'appréhenſion ou perception, le jugement, le diſcours & la méthode deviennent les quatre articles fondamentaux de cet art. C'eſt de nos réflexions ſur ces quatre opérations de l'eſprit que ſe forme la *logique*.

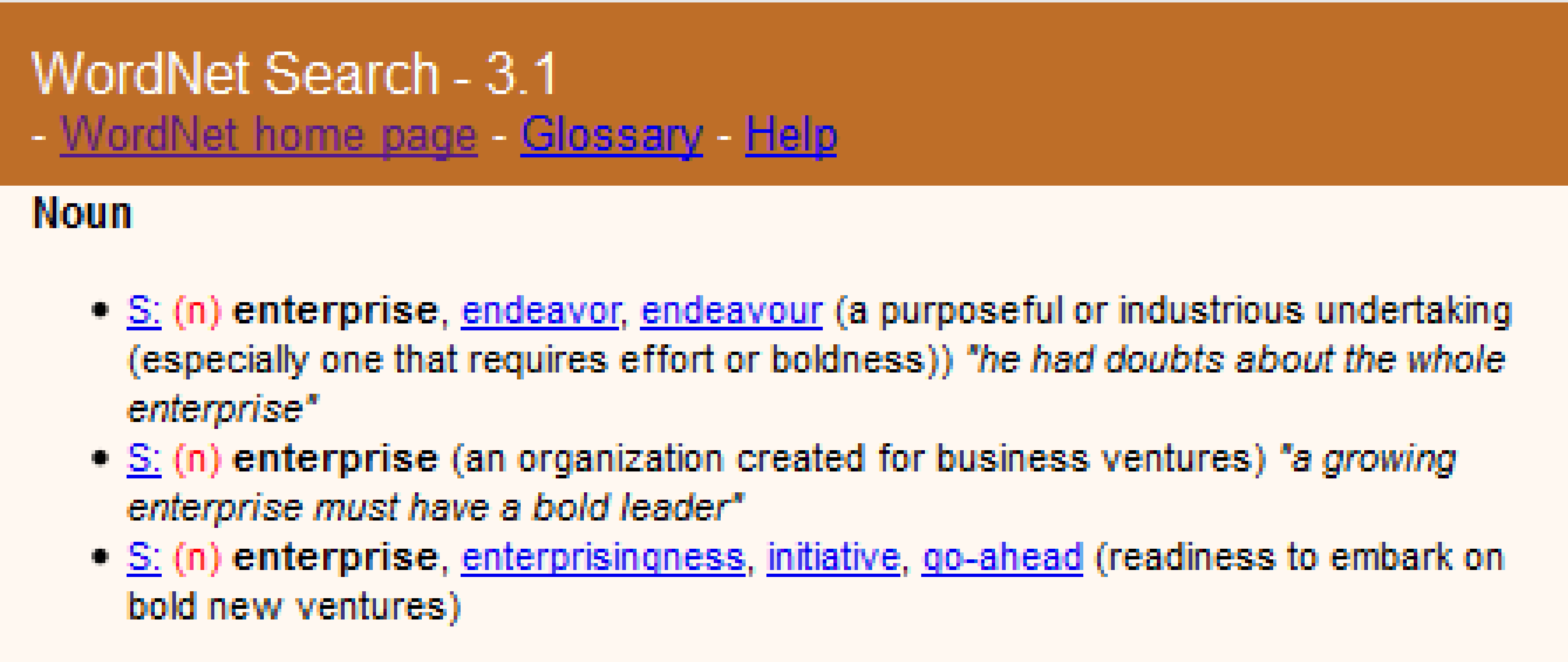
Le lord Bacon tire la diviſion de la *logique* en quatre parties, des quatre fins qu'on s'y propoſe; car un homme raiſonne, ou pour trouver ce qu'il cherche, ou pour raiſonner de ce qu'il a trouvé, ou pour retenir ce qu'il a jugé, ou pour enſeigner aux autres ce qu'il a retenu: de-là naiſſent autant de branches de l'art de raiſonner, ſavoir l'art de la recherche ou de l'invention, l'art de l'examen ou du jugement, l'art de retenir ou de la mémoire,

Modern Knowledge Bases

WordNet project (1985-now)

George Miller

Christiane Fellbaum



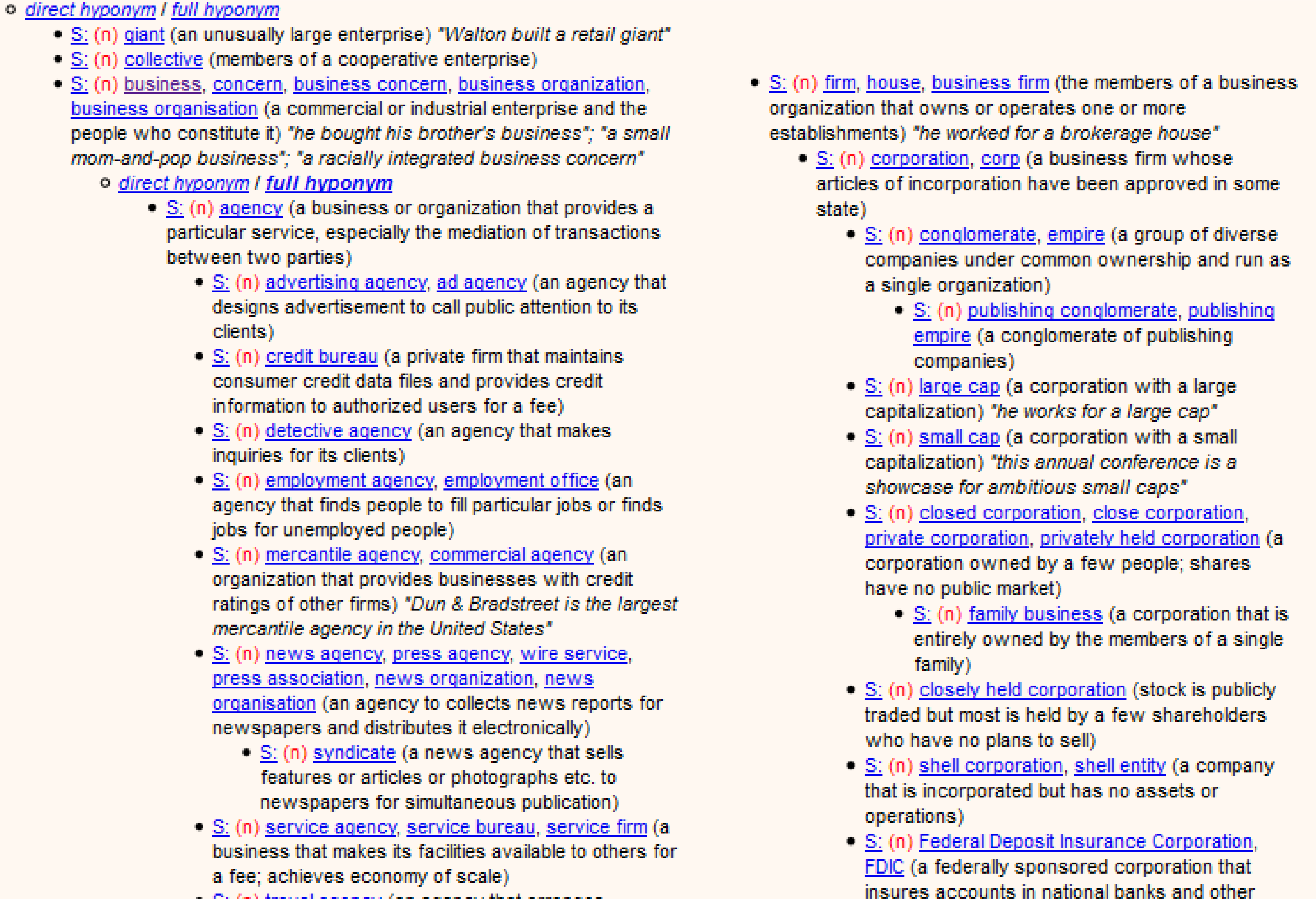
WordNet Search - 3.1

- WordNet home page - Glossary - Help

Noun

- S: (n) **enterprise**, endeavor, endeavour (a purposeful or industrious undertaking (especially one that requires effort or boldness)) *"he had doubts about the whole enterprise"*
- S: (n) **enterprise** (an organization created for business ventures) *"a growing enterprise must have a bold leader"*
- S: (n) **enterprise**, enterprisingness, initiative, go-ahead (readiness to embark on bold new ventures)

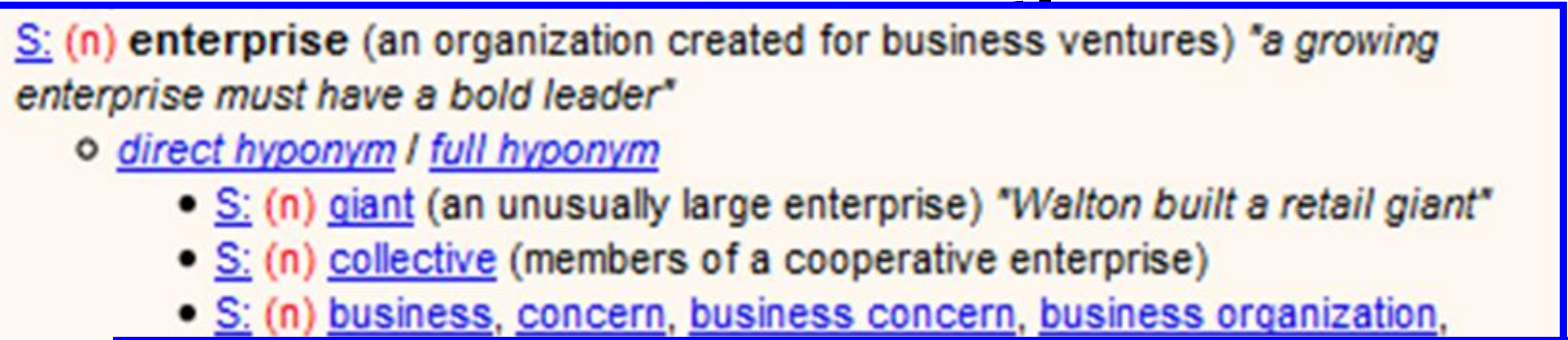
Modern Knowledge Bases



- *direct hyponym* / *full hyponym*
  - S: (n) giant (an unusually large enterprise) *"Walton built a retail giant"*
  - S: (n) collective (members of a cooperative enterprise)
  - S: (n) business, concern, business concern, business organization, business organisation (a commercial or industrial enterprise and the people who constitute it) *"he bought his brother's business"; "a small mom-and-pop business"; "a racially integrated business concern"*
    - *direct hyponym* / ***full hyponym***
      - S: (n) agency (a business or organization that provides a particular service, especially the mediation of transactions between two parties)
        - S: (n) advertising agency, ad agency (an agency that designs advertisement to call public attention to its clients)
        - S: (n) credit bureau (a private firm that maintains consumer credit data files and provides credit information to authorized users for a fee)
        - S: (n) detective agency (an agency that makes inquiries for its clients)
        - S: (n) employment agency, employment office (an agency that finds people to fill particular jobs or finds jobs for unemployed people)
        - S: (n) mercantile agency, commercial agency (an organization that provides businesses with credit ratings of other firms) *"Dun & Bradstreet is the largest mercantile agency in the United States"*
        - S: (n) news agency, press agency, wire service, press association, news organization, news organisation (an agency to collects news reports for newspapers and distributes it electronically)
          - S: (n) syndicate (a news agency that sells features or articles or photographs etc. to newspapers for simultaneous publication)
        - S: (n) service agency, service bureau, service firm (a business that makes its facilities available to others for a fee; achieves economy of scale)
        - S: (n) travel agency (an agency that arranges

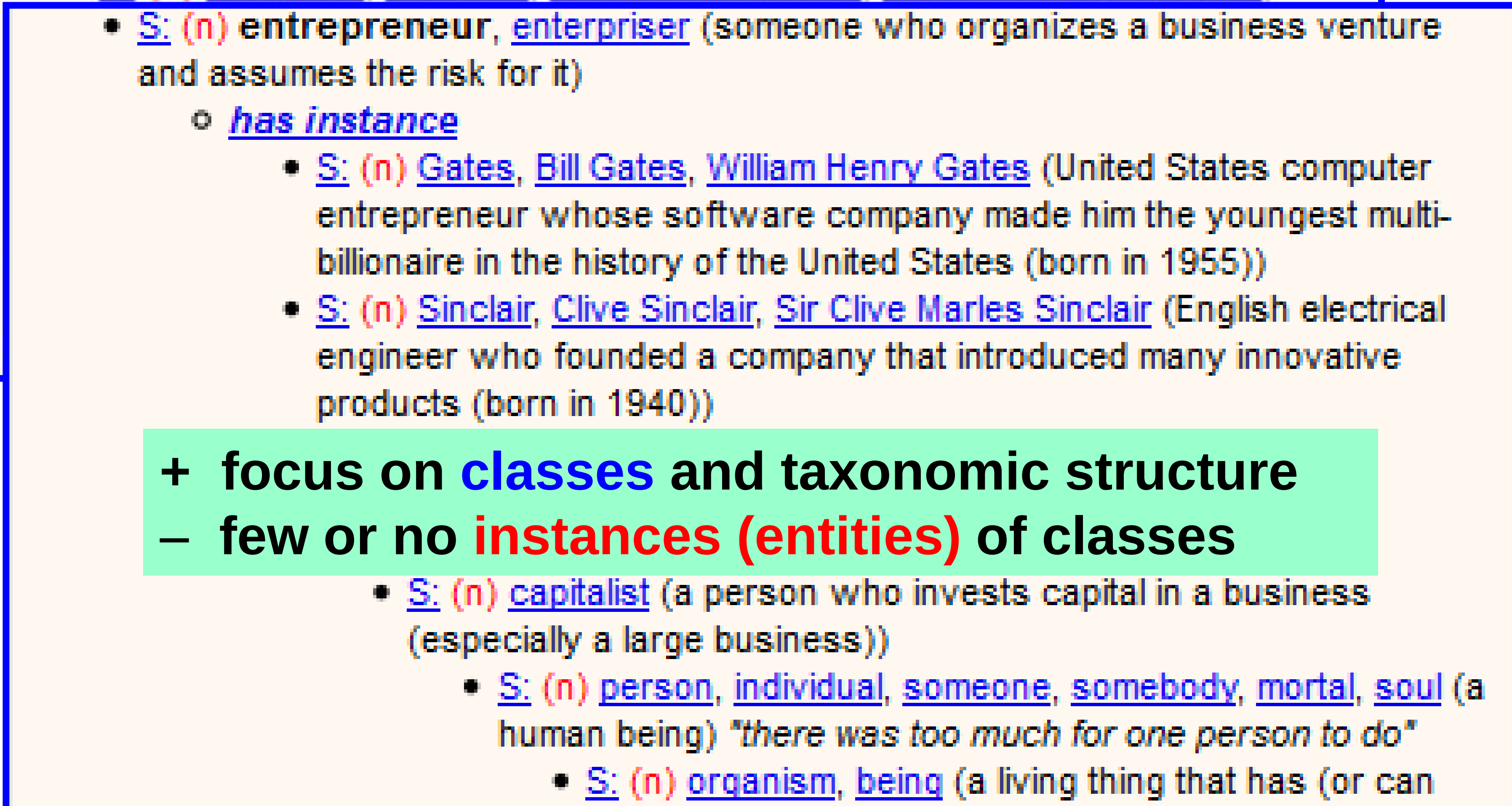
- S: (n) firm, house, business firm (the members of a business organization that owns or operates one or more establishments) *"he worked for a brokerage house"*
  - S: (n) corporation, corp (a business firm whose articles of incorporation have been approved in some state)
    - S: (n) conglomerate, empire (a group of diverse companies under common ownership and run as a single organization)
      - S: (n) publishing conglomerate, publishing empire (a conglomerate of publishing companies)
    - S: (n) large cap (a corporation with a large capitalization) *"he works for a large cap"*
    - S: (n) small cap (a corporation with a small capitalization) *"this annual conference is a showcase for ambitious small caps"*
    - S: (n) closed corporation, close corporation, private corporation, privately held corporation (a corporation owned by a few people; shares have no public market)
      - S: (n) family business (a corporation that is entirely owned by the members of a single family)
    - S: (n) closely held corporation (stock is publicly traded but most is held by a few shareholders who have no plans to sell)
    - S: (n) shell corporation, shell entity (a company that is incorporated but has no assets or operations)
    - S: (n) Federal Deposit Insurance Corporation, FDIC (a federally sponsored corporation that insures accounts in national banks and other

Modern Knowledge Bases



S: (n) **enterprise** (an organization created for business ventures) *"a growing enterprise must have a bold leader"*

- *direct hyponym* / *full hyponym*
  - S: (n) giant (an unusually large enterprise) *"Walton built a retail giant"*
  - S: (n) collective (members of a cooperative enterprise)
  - S: (n) business, concern, business concern, business organization,



- S: (n) **entrepreneur**, enterpriser (someone who organizes a business venture and assumes the risk for it)
  - ***has instance***
    - S: (n) Gates, Bill Gates, William Henry Gates (United States computer entrepreneur whose software company made him the youngest multi-billionaire in the history of the United States (born in 1955))
    - S: (n) Sinclair, Clive Sinclair, Sir Clive Marles Sinclair (English electrical engineer who founded a company that introduced many innovative products (born in 1940))
- S: (n) capitalist (a person who invests capital in a business (especially a large business))
  - S: (n) person, individual, someone, somebody, mortal, soul (a human being) *"there was too much for one person to do"*
    - S: (n) organism, being (a living thing that has (or can

+ focus on classes and taxonomic structure

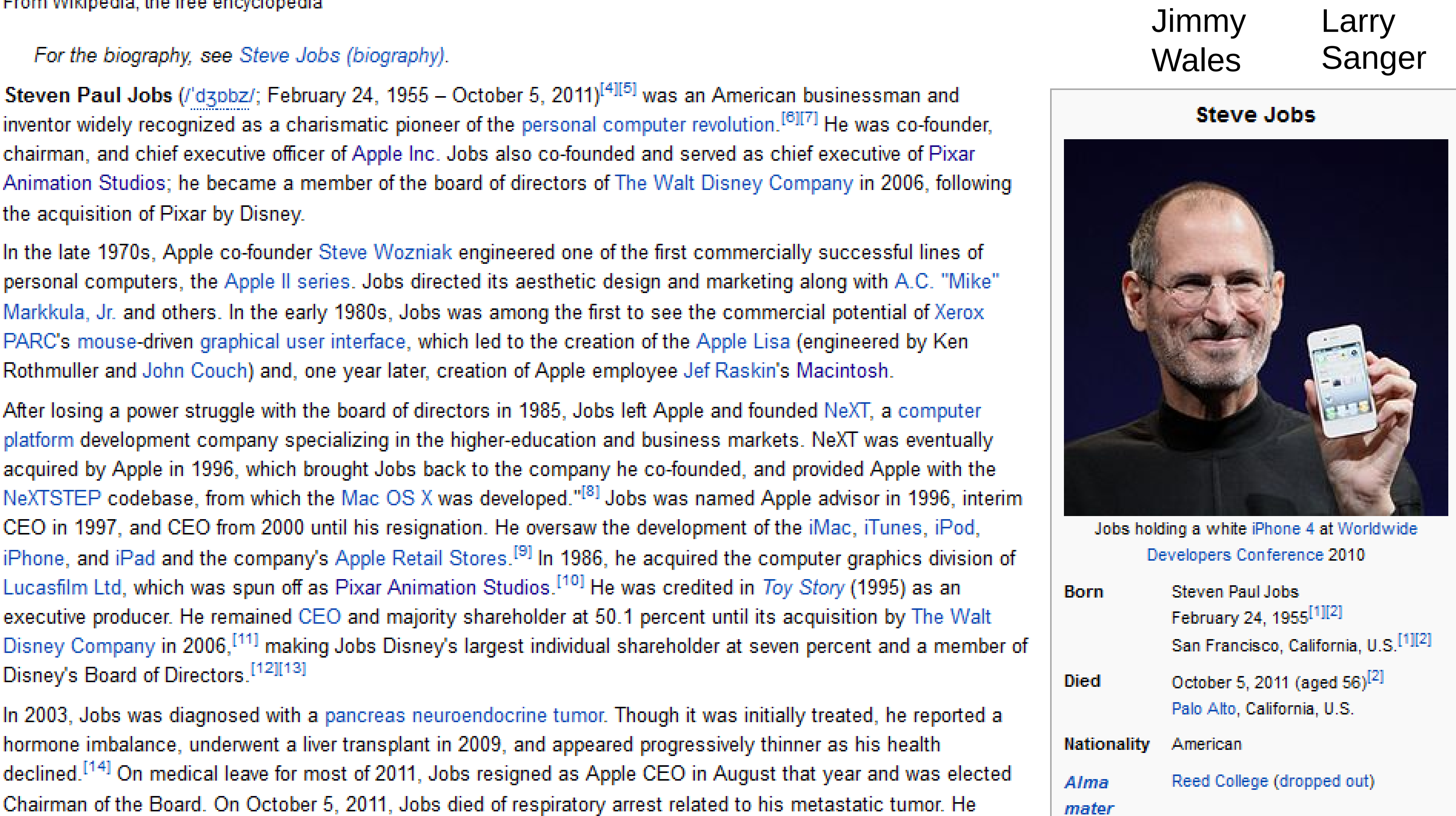
– few or no instances (entities) of classes

Knowledge Communities & New Opportunities

Jimmy Wales

Larry Sanger

Steve Jobs



From Wikipedia, the free encyclopedia

For the biography, see Steve Jobs (biography).

Steven Paul Jobs (/ˈdʒɒbz/; February 24, 1955 – October 5, 2011)[4][5] was an American businessman and inventor widely recognized as a charismatic pioneer of the personal computer revolution.[6][7] He was co-founder, chairman, and chief executive officer of Apple Inc. Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney.

In the late 1970s, Apple co-founder Steve Wozniak engineered one of the first commercially successful lines of personal computers, the Apple II series. Jobs directed its aesthetic design and marketing along with A.C. "Mike" Markkula, Jr. and others. In the early 1980s, Jobs was among the first to see the commercial potential of Xerox PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa (engineered by Ken Rothmuller and John Couch) and, one year later, creation of Apple employee Jef Raskin's Macintosh.

After losing a power struggle with the board of directors in 1985, Jobs left Apple and founded NeXT, a computer platform development company specializing in the higher-education and business markets. NeXT was eventually acquired by Apple in 1996, which brought Jobs back to the company he co-founded, and provided Apple with the NeXTSTEP codebase, from which the Mac OS X was developed."[8] Jobs was named Apple advisor in 1996, interim CEO in 1997, and CEO from 2000 until his resignation. He oversaw the development of the iMac, iTunes, iPod, iPhone, and iPad and the company's Apple Retail Stores.[9] In 1986, he acquired the computer graphics division of Lucasfilm Ltd, which was spun off as Pixar Animation Studios.[10] He was credited in *Toy Story* (1995) as an executive producer. He remained CEO and majority shareholder at 50.1 percent until its acquisition by The Walt Disney Company in 2006,[11] making Jobs Disney's largest individual shareholder at seven percent and a member of Disney's Board of Directors.[12][13]

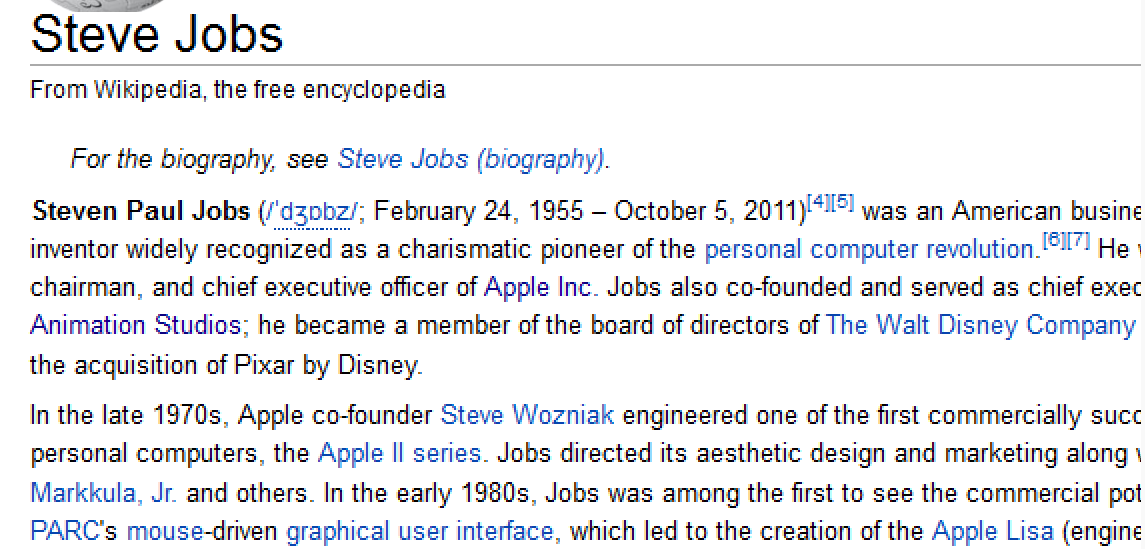
In 2003, Jobs was diagnosed with a pancreas neuroendocrine tumor. Though it was initially treated, he reported a hormone imbalance, underwent a liver transplant in 2009, and appeared progressively thinner as his health declined.[14] On medical leave for most of 2011, Jobs resigned as Apple CEO in August that year and was elected Chairman of the Board. On October 5, 2011, Jobs died of respiratory arrest related to his metastatic tumor. He

Steve Jobs

Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

Born	Steven Paul Jobs February 24, 1955[1][2] San Francisco, California, U.S.[1][2]
Died	October 5, 2011 (aged 56)[2] Palo Alto, California, U.S.
Nationality	American
Alma mater	Reed College (dropped out)

Knowledge Commu
& New Opportunit



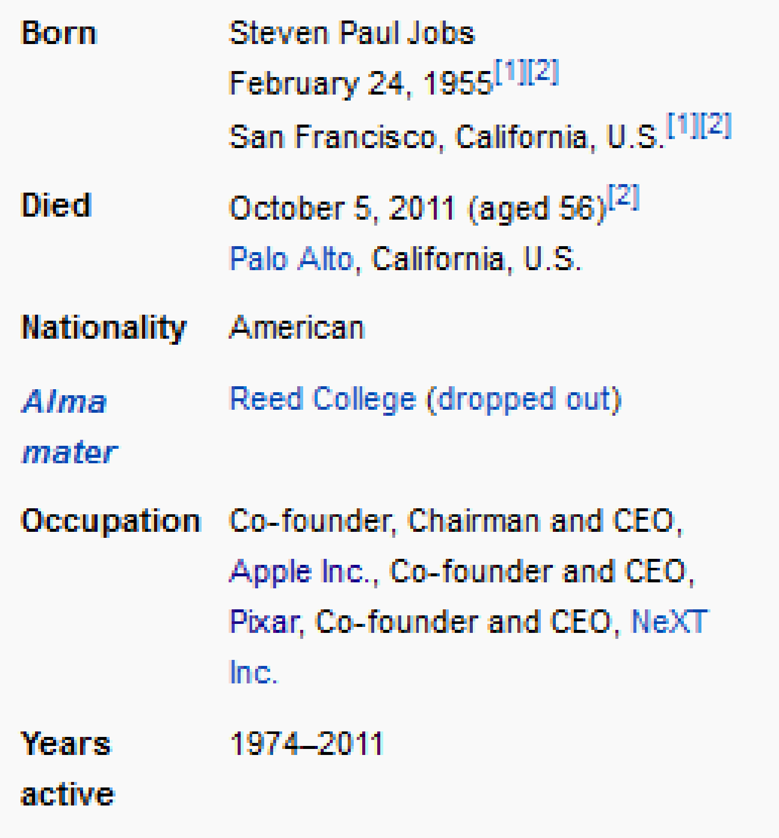
Steve Jobs

From Wikipedia, the free encyclopedia

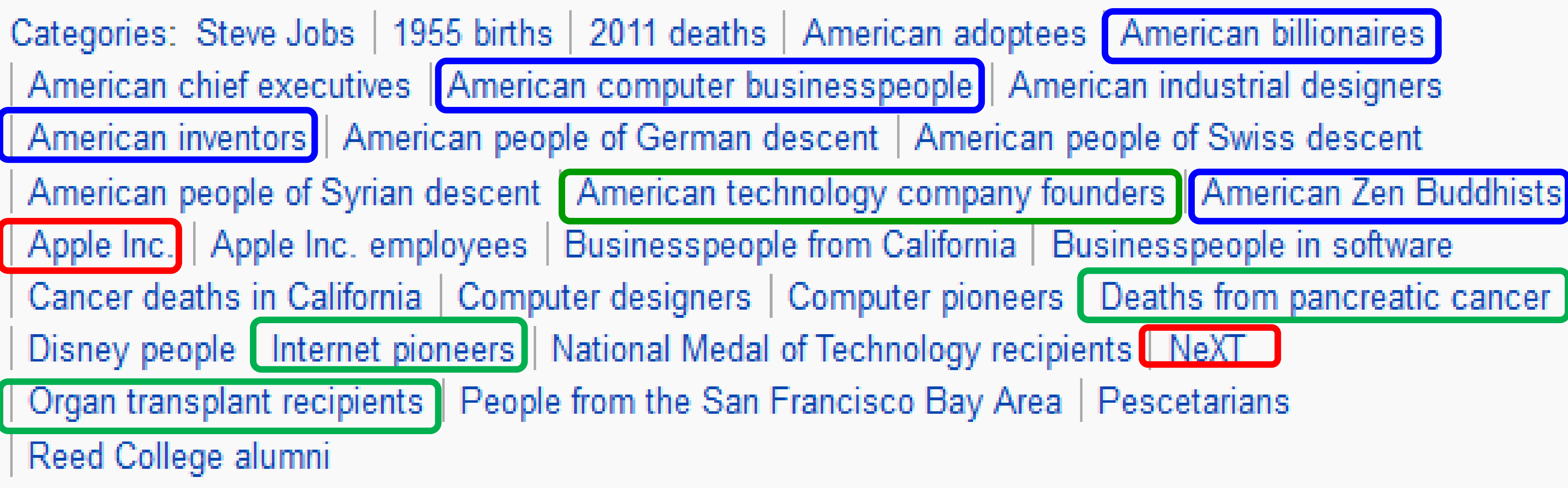
For the biography, see Steve Jobs (biography).

Steven Paul Jobs (/ˈdʒɒbz/; February 24, 1955 – October 5, 2011)[4][5] was an American busine
inventor widely recognized as a charismatic pioneer of the personal computer revolution.[6][7] He
chairman, and chief executive officer of Apple Inc. Jobs also co-founded and served as chief exec
Animation Studios; he became a member of the board of directors of The Walt Disney Company
the acquisition of Pixar by Disney.

In the late 1970s, Apple co-founder Steve Wozniak engineered one of the first commercially succ
personal computers, the Apple II series. Jobs directed its aesthetic design and marketing along
Markkula, Jr. and others. In the early 1980s, Jobs was among the first to see the commercial pot
PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa (engine

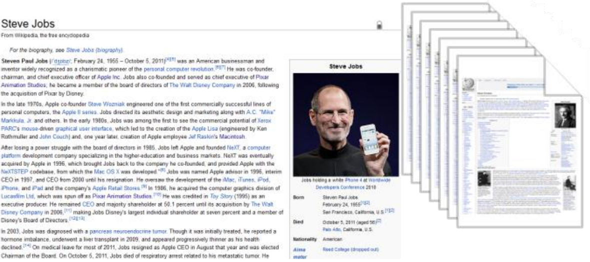


Born	Steven Paul Jobs February 24, 1955[1][2] San Francisco, California, U.S.[1][2]
Died	October 5, 2011 (aged 56)[2] Palo Alto, California, U.S.
Nationality	American
Alma mater	Reed College (dropped out)
Occupation	Co-founder, Chairman and CEO, Apple Inc., Co-founder and CEO, Pixar, Co-founder and CEO, NeXT Inc.
Years active	1974–2011



Categories: Steve Jobs | 1955 births | 2011 deaths | American adoptees | American billionaires | American chief executives | American computer businesspeople | American industrial designers | American inventors | American people of German descent | American people of Swiss descent | American people of Syrian descent | American technology company founders | American Zen Buddhists | Apple Inc. | Apple Inc. employees | Businesspeople from California | Businesspeople in software | Cancer deaths in California | Computer designers | Computer pioneers | Deaths from pancreatic cancer | Disney people | Internet pioneers | National Medal of Technology recipients | NeXT | Organ transplant recipients | People from the San Francisco Bay Area | Pescetarians | Reed College alumni

Automatic Knowledge Base Construction



Steve Jobs

From Wikipedia, the free encyclopedia

For the biography, see Steve Jobs (biography).

Steven Paul Jobs (/ˈdʒɒbz/; February 24, 1955 – October 5, 2011)[4][5] was an American businessman and inventor widely recognized as a charismatic pioneer of the personal computer revolution.[6][7] He was co-founder, chairman, and chief executive officer of Apple Inc. Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney.

In the late 1970s, Apple co-founder Steve Wozniak engineered one of the first commercially successful lines of personal computers, the Apple II series. Jobs directed its aesthetic design and marketing along with A.C. "Mike" Markkula, Jr. and others. In the early 1980s, Jobs was among the first to see the commercial potential of Xerox PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa (engineered by Ken Rothmuller and John Couch) and, one year later, creation of Apple employee Jef Raskin's Macintosh.

After losing a power struggle with the board of directors in 1985, Jobs left Apple and founded NeXT, a computer platform development company specializing in the higher-education and business markets. NeXT was eventually acquired by Apple in 1996, which brought Jobs back to the company he co-founded, and provided Apple with the NeXTSTEP codebase, from which the Mac OS X was developed.[8] Jobs was named Apple advisor in 1996, interim CEO in 1997, and CEO from 2000 until his resignation. He oversaw the development of the iMac, iTunes, iPod, iPhone, and iPad and the company's Apple Retail Stores.[9] In 1986, he acquired the computer graphics division of Lucasfilm Ltd, which was spun off as Pixar Animation Studios.[10] He was credited in *Toy Story* (1995) as an executive producer. He remained CEO and majority shareholder at 50.1 percent until its acquisition by The Walt Disney Company in 2006,[11] making Jobs Disney's largest individual shareholder at seven percent and a member of Disney's Board of Directors.[12][13]

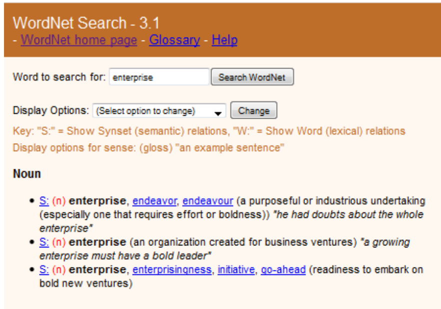
In 2003, Jobs was diagnosed with a pancreas neuroendocrine tumor. Though it was initially treated, he reported a hormone imbalance, underwent a liver transplant in 2009, and appeared progressively thinner as his health declined.[14] On medical leave for most of 2011, Jobs resigned as Apple CEO in August that year and was elected Chairman of the Board. On October 5, 2011, Jobs died of respiratory arrest related to his metastatic tumor. He

Steve Jobs

Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

Born	Steven Paul Jobs February 24, 1955[1][2] San Francisco, California, U.S.[1][2]
Died	October 5, 2011 (aged 56)[2] Palo Alto, California, U.S.
Nationality	American
Alma mater	Reed College (dropped out)

map 300K Wikipedia categories onto 150K WordNet classes



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: enterprise

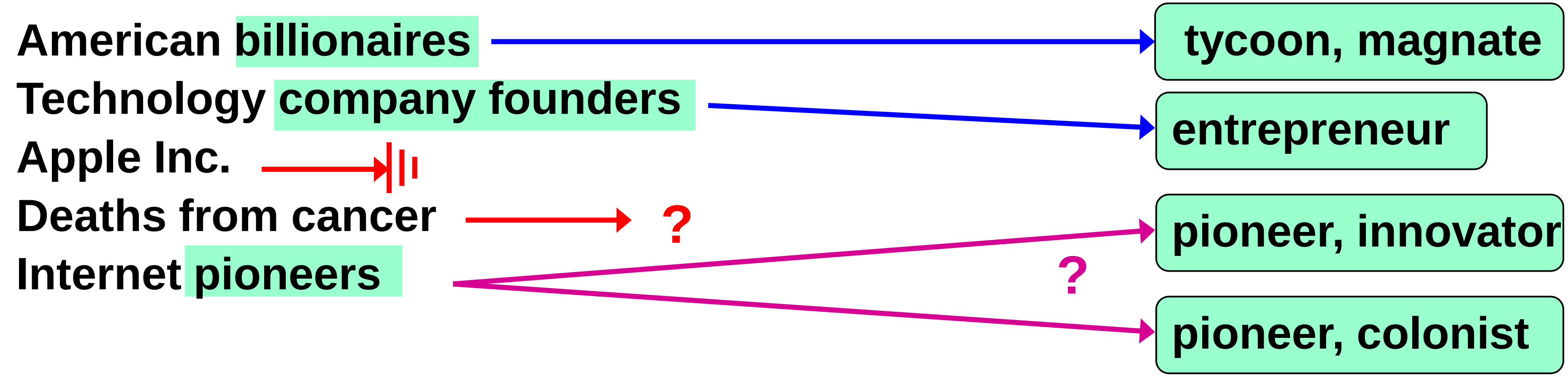
Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **enterprise**, endeavor, endeavour (a purposeful or industrious undertaking (especially one that requires effort or boldness)) *"he had doubts about the whole enterprise"*
- S: (n) **enterprise** (an organization created for business ventures) *"a growing enterprise must have a bold leader"*
- S: (n) **enterprise**, enterprisingness, initiative, go-ahead (readiness to embark on bold new ventures)



American billionaires → **tycoon, magnate**

Technology company founders → **entrepreneur**

Apple Inc. →

Deaths from cancer → **?**

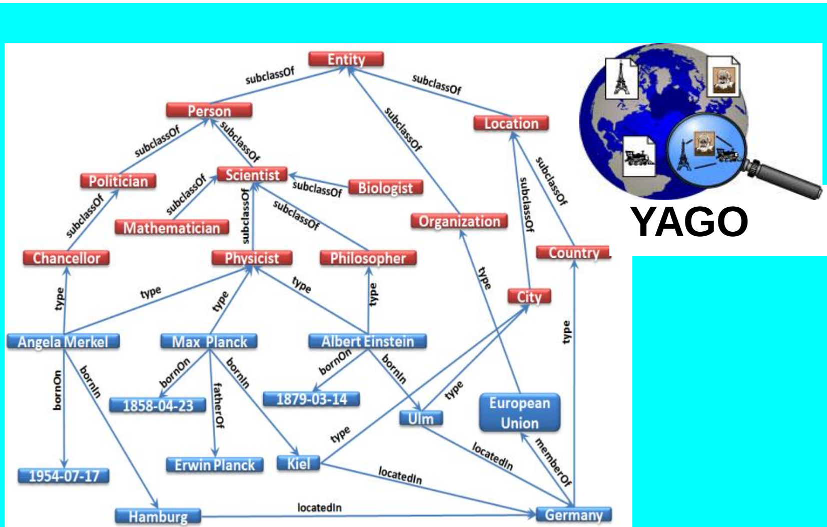
Internet pioneers → **pioneer, innovator** **?**

Internet pioneers → **pioneer, colonist**

Integrating entities & facts from Wikipedia with semantic classes in WordNet

→ YAGO knowledge base:

10 Mio. entities, 350 000 classes,
180 Mio. facts, 100 relations,
100 languages, 2 Bio. triples,
95% accuracy



Large-Scale Universal Knowledge Bases

Yago: 10 Mio. entities, 350 000 classes,
180 Mio. facts, 100 properties, 100 languages
high accuracy, no redundancy, limited coverage
http://yago-knowledge.org

YAGO

Dbpedia: 4 Mio. entities, 250 classes,
500 Mio. facts, 6000 properties
high coverage, live updates
http://dbpedia.org

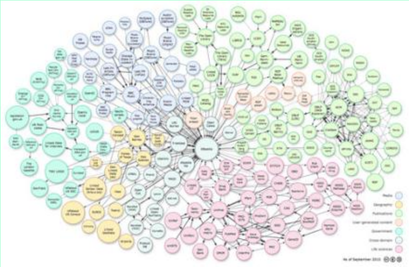
Freebase: 25 Mio. entities, 2000 topics,
100 Mio. facts, 4000 properties
interesting relations (e.g., romantic affairs)
http://freebase.com

NELL: 300 000 entity names, 300 classes, 500 properties,
1 Mio. beliefs, 15 Mio. low-confidence beliefs
learned rules
http://rtw.ml.cmu.edu/rtw/

ReadTheWeb

and more …

plus Linked Data



Outline

✓ Knowledge for Machines

✓ Construction of Knowledge Bases

★ KB Population from Text & Web Pages

★ Knowledge for Intelligent Applications

★ Opportunities & Challenges

★ Conclusions

Knowledge Base Population

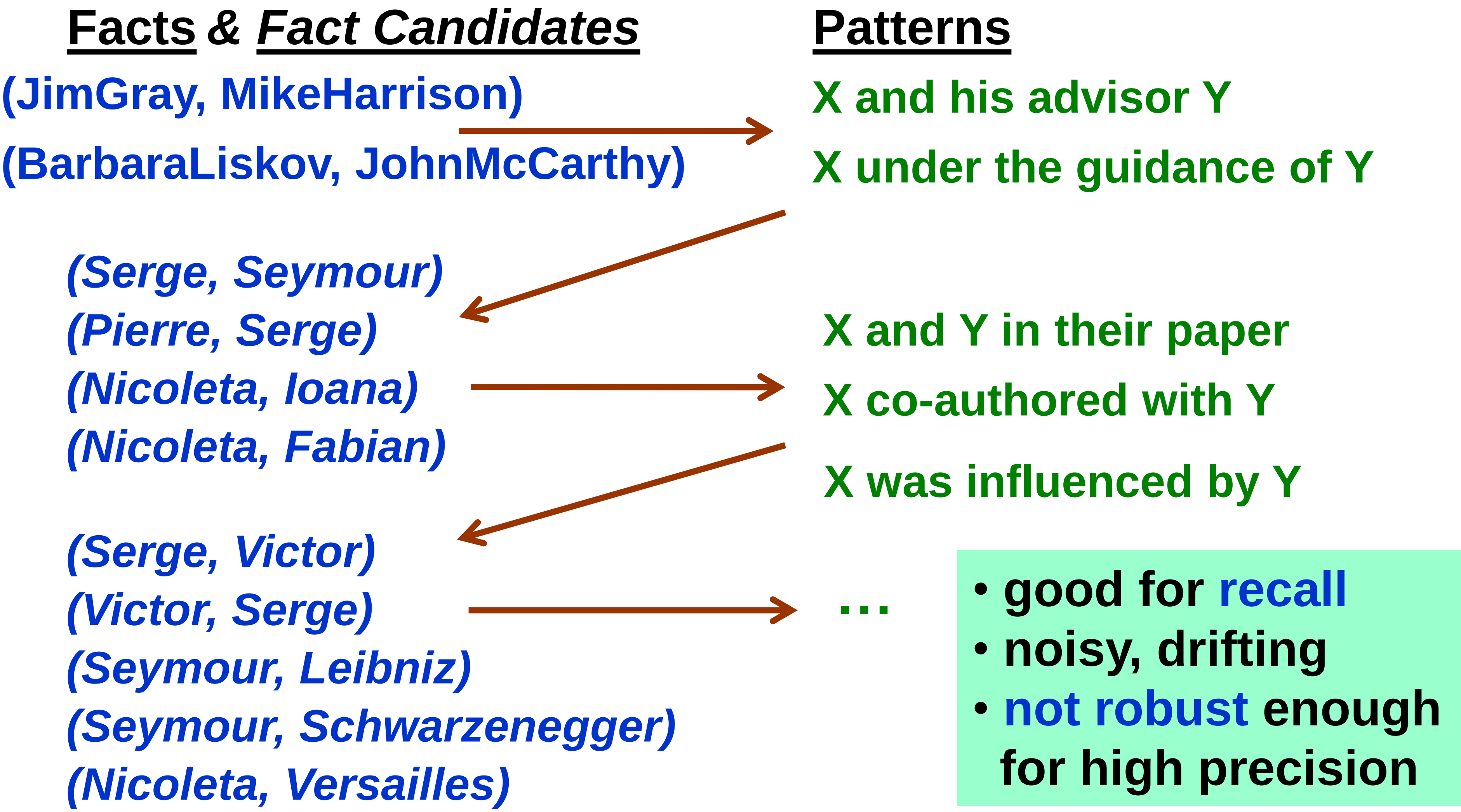
hasAdvisor (<person>, <person>) **AlmaMater (<person>, <university>)**

Jim Gray was one of Harrison's best students
Barbara Liskov wrote her thesis under the guidance of McCarthy
Pierre Senellart was one of Serge's best students
Serge Abiteboul was grateful to his advisor Ginsburg
Seymour Ginsburg was influenced by Leibniz
Seymour and his advisor Ben Dushnik never co-authored any paper
Serge and Victor co-authored many papers

➡ hasAdvisor (JimGray, MikeHarrison)
hasAdvisor (BarbaraLiskov, JohnMcCarthy)
…
AlmaMater (JimGray, Berkeley)
AlmaMater (BarbaraLiskov, Stanford)
…

Pattern-based Gathering (statistical evidence) **+** **Constraint-aware Reasoning (logical consistency)**

Pattern-based Gathering of Fact Candidates (Linguistic Patterns and Statistical Evidence)



Facts & ***Fact Candidates***

(JimGray, MikeHarrison)

(BarbaraLiskov, JohnMcCarthy)

(Serge, Seymour)
(Pierre, Serge)
(Nicoleta, Ioana)
(Nicoleta, Fabian)

(Serge, Victor)
(Victor, Serge)
(Seymour, Leibniz)
(Seymour, Schwarzenegger)
(Nicoleta, Versailles)

Patterns

X and his advisor Y
X under the guidance of Y

X and Y in their paper
X co-authored with Y
X was influenced by Y

…

- good for recall
- noisy, drifting
- not robust enough for high precision

Constrained Reasoning for Logical Consistency

Use knowledge (consistency contraints) for joint reasoning on hypotheses and pruning of false candidates

$\forall\, x, y, z: \text{hasA}(x,y) \wedge \text{hasA}(x,z) \Rightarrow y=z$

$\forall\, x, y: \text{hasA}(x,y) \Rightarrow \neg\, \text{hasA}(y,x)$

$\forall\, x, y, p: \text{occurs}\ (x, y, p) \wedge \text{goodPattern}(p, \text{hasA})$
$\quad \Rightarrow \text{hasA}(x,y)$

$\forall\, x, y, p: \text{occurs}\ (x, y, p) \wedge \text{hasA}(x,y)$
$\quad \Rightarrow \text{goodPattern}(p, \text{hasA})$

$\forall\, x, y: \text{hasA}(x,y) \Rightarrow \text{type}(x)=\text{Person}$

$\forall\, x, y: \text{hasA}(x,y) \Rightarrow \text{type}(y)=\text{Scientist}$

$\forall\, x, y: \text{hasA}(x,y) \Rightarrow \text{hasStudent}(y,x)$

$\forall\, x, y: \text{AlmaMater}(x,y) \Rightarrow \text{type}(y)=\text{University}$

$\forall\, x,y,u: \text{hasA}(x,y) \wedge \text{AlmaMater}(x,u) \Rightarrow \text{FacultyOf}\ (y,u)$

$\forall\, x,y,s,t: \text{hasA}(x,y) \wedge \text{gradYear}(x,s) \wedge \text{gradYear}(y,t) \Rightarrow t < s$

ground atoms:

hasA(Serge, Seymour)
hasA(Nicoleta, Ioana)
hasA(Nicoleta, Fabian)
hasA(Serge, Victor)
hasA (Victor, Serge)
hasA (Nicoleta, Versailles)
hasA (Seymour, Leibniz)

…

Constrained Reasoning for Logical Consistency

Use knowledge (consistency contraints) for joint reasoning on hypotheses and pruning of false candidates

ground atoms:

hasA(Serge, Seymour)
hasA(Nicoleta, Ioana)
~~hasA(Nicoleta, Fabian)~~
~~hasA(Serge, Victor)~~
~~hasA (Victor, Serge)~~
~~hasA (Nicoleta, Versailles)~~
~~hasA (Seymour, Leibniz)~~
…

$\forall\ x, y, z: \text{hasA}(x,y) \wedge \text{hasA}(x,z) \Rightarrow y=z$

$\forall\ x, y: \text{hasA}(x,y) \Rightarrow \neg\, \text{hasA}(y,x)$

$\forall\ x, y, p: \text{occurs}\ (x, y, p) \wedge \text{goodPattern}(p, \text{hasA}) \Rightarrow \text{hasA}(x,y)$

$\forall\ x, y, p: \text{occurs}\ (x, y, p) \wedge \text{hasA}(x,y) \Rightarrow \text{goodPattern}(p, \text{hasA})$

$\forall\ x, y: \text{hasA}(x,y) \Rightarrow \text{type}(x)=\text{Person}$

$\forall\ x, y: \text{hasA}(x,y) \Rightarrow \text{type}(y)=\text{Scientist}$

$\forall\ x, y: \text{hasA}(x,y) \Rightarrow \text{studentOf}(y,x)$

$\forall\ x, y: \text{AlmaMater}(x,y) \Rightarrow \text{type}(y)=\text{University}$

$\forall\ x,y,u: \text{hasA}(x,y) \wedge \text{AlmaMater}(x,u) \Rightarrow \text{FacultyOf}\ (y,u)$

$\forall\ x,y,s,t: \text{hasA}(x,y) \wedge \text{gradYear}(x,s) \wedge \text{gradYear}(y,t) \Rightarrow s < t$

Find consistent subset(s) of atoms (“possible world(s)“, “the truth“)

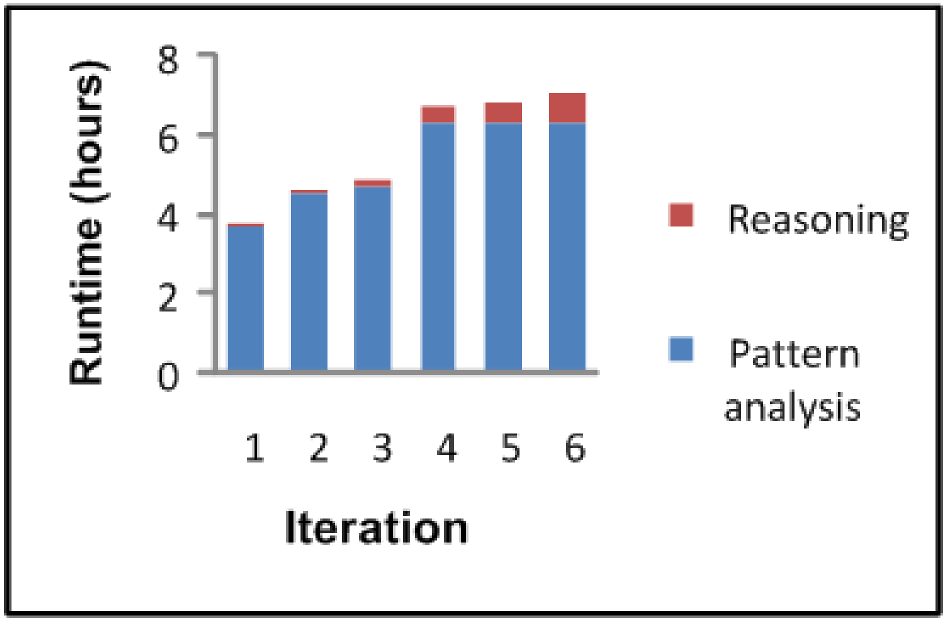
→ customized Weighted MaxSat solver for set of clauses

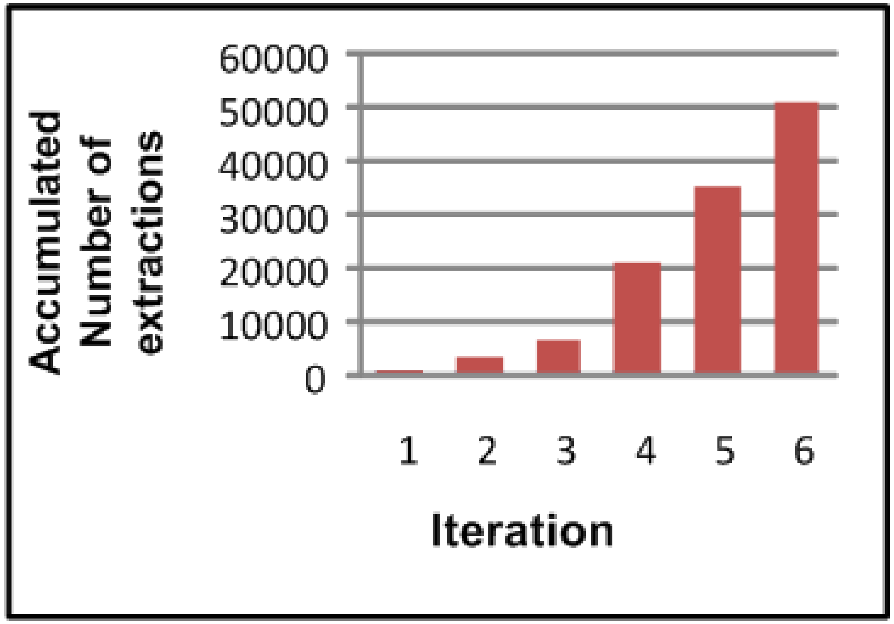
→ max a posteriori for probabilistic factor graph

PROSPERA: Web-Scale Experiments

[N. Nakashole et al.: WSDM‘11]

- on ClueWeb‘09 corpus: 500 Mio. English Web pages
- with Hadoop cluster of 10x16 cores and 10x48 GB memory
- 10 seed examples, 5 counter examples for each of 15 relations on sports and academia





more than 100,000 facts acquired after 6 iterations
with overall precision ≈ 90%
and 99% precision@1000 for each of the relations

www.mpi-inf.mpg.de/yago-naga/prospera/

Outline

✓ Knowledge for Machines

✓ Construction of Knowledge Bases

✓ KB Population from Text & Web Pages

★ Knowledge for Intelligent Applications

★ Opportunities & Challenges

★ Conclusions

Application: Question Answering

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

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

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

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

question classification & decomposition → knowledge back-ends

WIKIPEDIA
The Free Encyclopedia

DBpedia

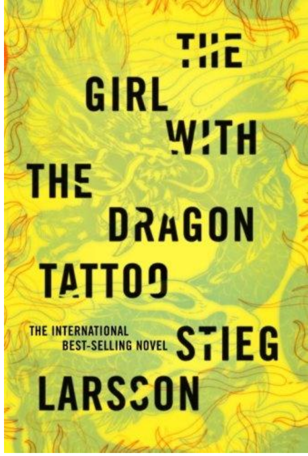
freebase™

YAGO

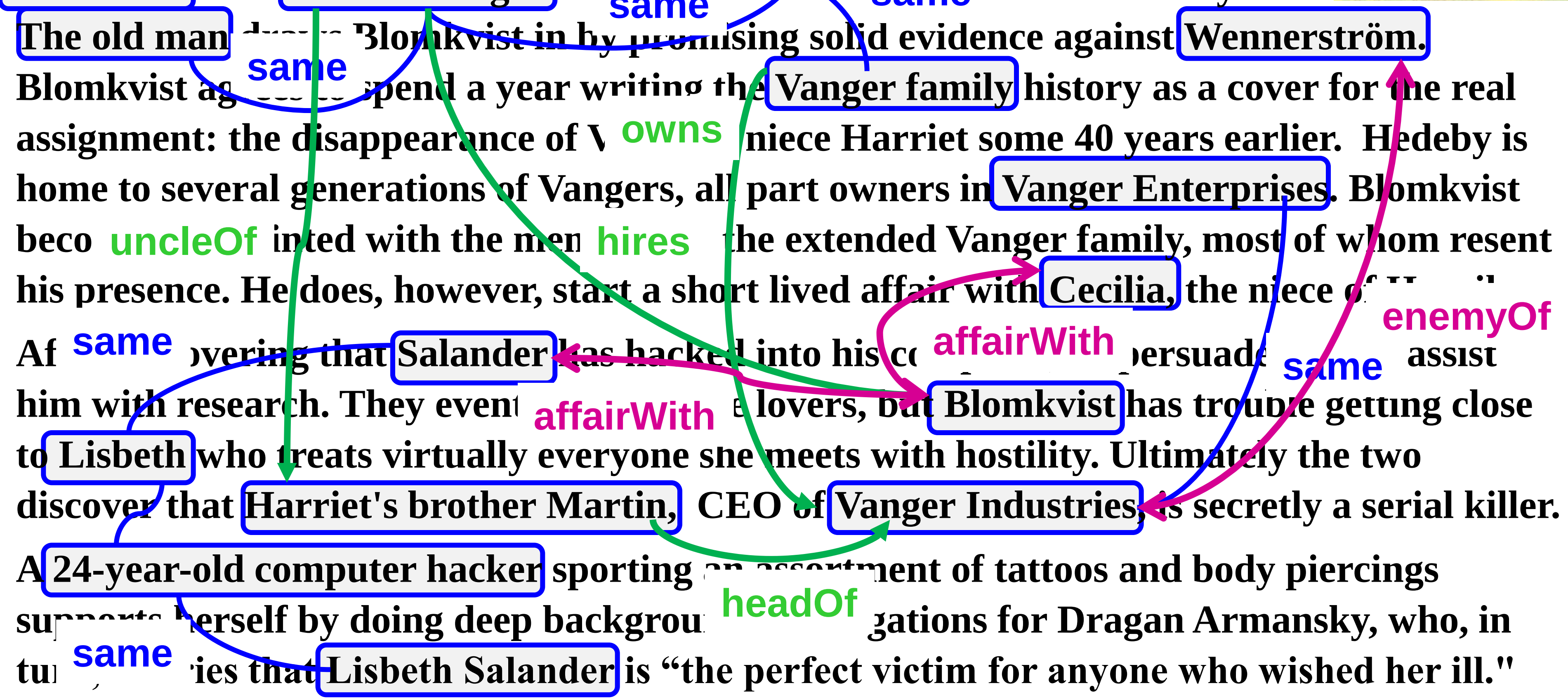
D. Ferrucci et al.: Building Watson: An Overview of t
DeepQA Project. AI Magazine, Fall 2010.

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

Application: Machine Reading



It's about the disappearance forty years ago of Harriet Vanger, a young scion of one of the wealthiest families in Sweden, and about her uncle, determined to know the trut [illegible] hat he believes was her murder.



Blomkvist visits Henrik Vanger at [illegible] te on th [illegible] and of Hedeby. The old man [illegible] Blomkvist in by promising solid evidence against Wennerström. Blomkvist ag [illegible] spend a year writing the Vanger family history as a cover for the real assignment: the disappearance of V [illegible] niece Harriet some 40 years earlier. Hedeby is home to several generations of Vangers, all part owners in Vanger Enterprises. Blomkvist beco [illegible] inted with the mem [illegible] the extended Vanger family, most of whom resent his presence. He does, however, start a short lived affair with Cecilia, the niece of [illegible]

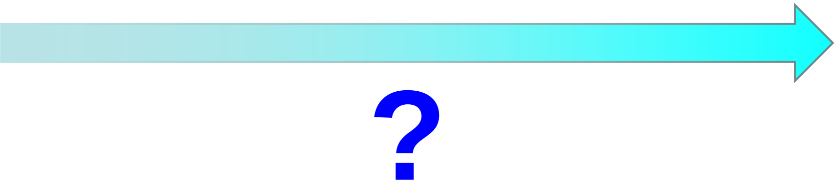
Af [illegible] overing that Salander has hacked into his c [illegible] persuade [illegible] assist him with research. They event [illegible] e lovers, but Blomkvist has trouble getting close to Lisbeth who treats virtually everyone she meets with hostility. Ultimately the two discover that Harriet's brother Martin, CEO of Vanger Industries, is secretly a serial killer.

A 24-year-old computer hacker sporting an assortment of tattoos and body piercings supports herself by doing deep backgrou [illegible] gations for Dragan Armansky, who, in tu [illegible] ies that Lisbeth Salander is "the perfect victim for anyone who wished her ill."

O. Etzioni, M. Banko, M.J. Cafarella: Machine Reading, AAAI ,06

Named Entity Disambiguation

Sergio talked to Ennio about Eli‘s role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio‘s trilogy of western films.

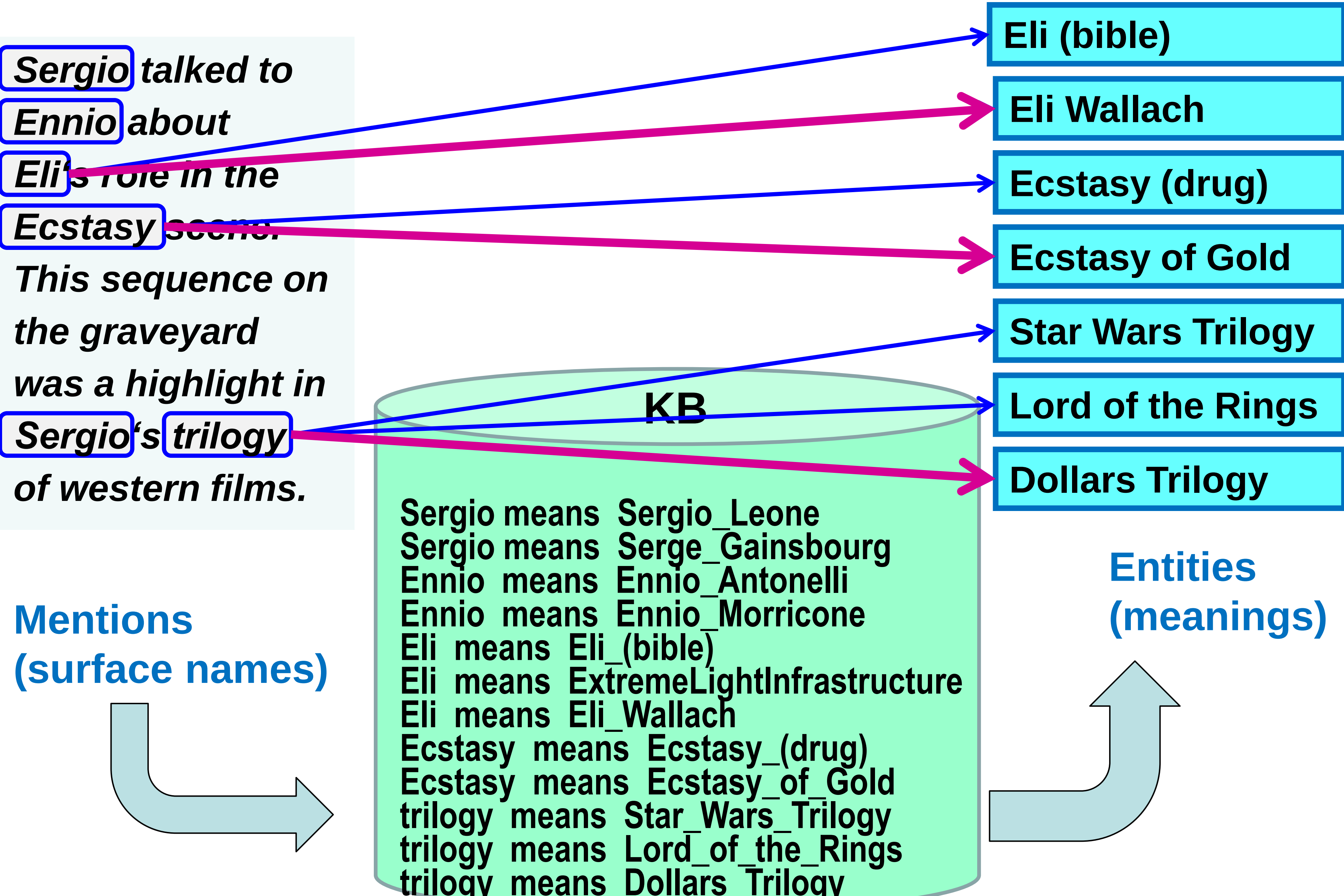


Eli (bible)

Eli Wallach

Mentions (surface names)

Named Entity Disambiguation

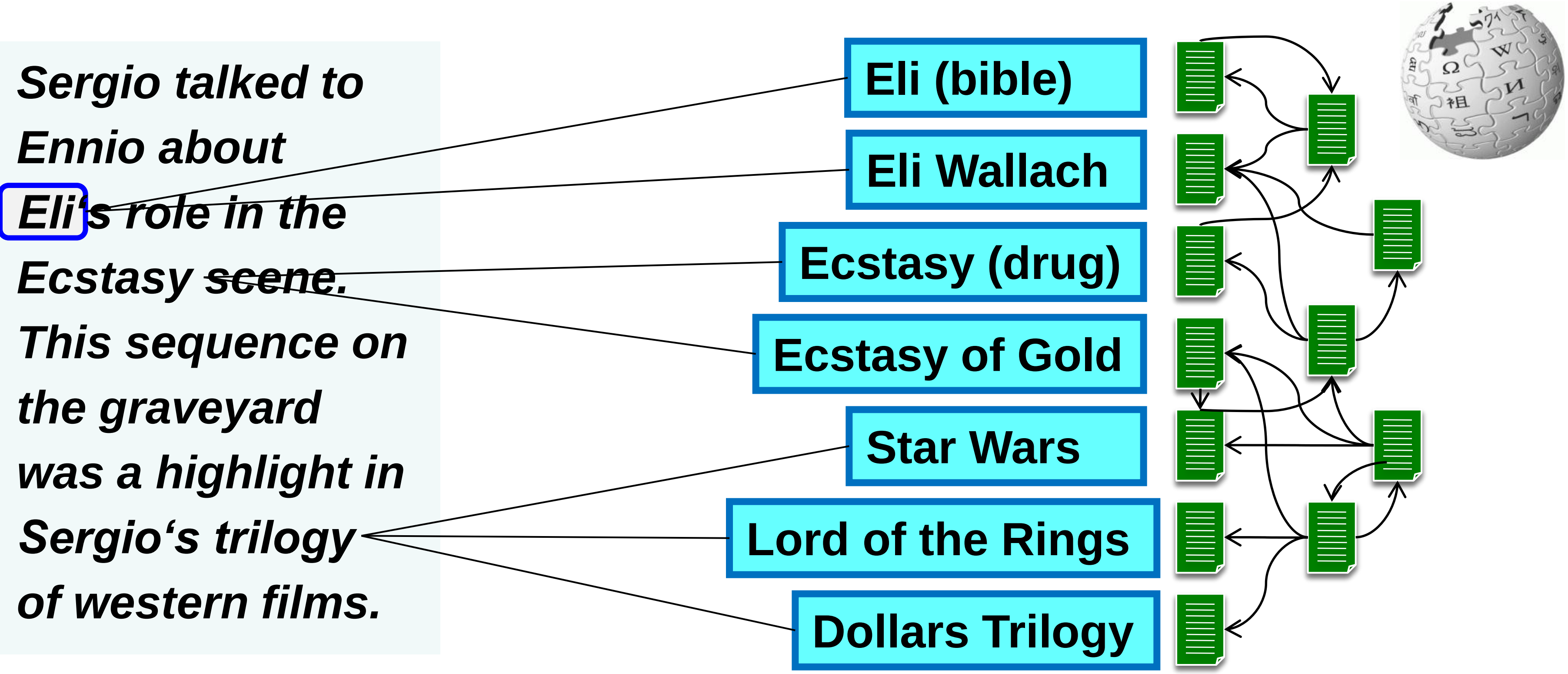


Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

Sergio means Sergio_Leone
Sergio means Serge_Gainsbourg
Ennio means Ennio_Antonelli
Ennio means Ennio_Morricone
Eli means Eli_(bible)
Eli means ExtremeLightInfrastructure
Eli means Eli_Wallach
Ecstasy means Ecstasy_(drug)
Ecstasy means Ecstasy_of_Gold
trilogy means Star_Wars_Trilogy
trilogy means Lord_of_the_Rings
trilogy means Dollars_Trilogy

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

Eli (bible)

Eli Wallach

Ecstasy (drug)

Ecstasy of Gold

Star Wars

Lord of the Rings

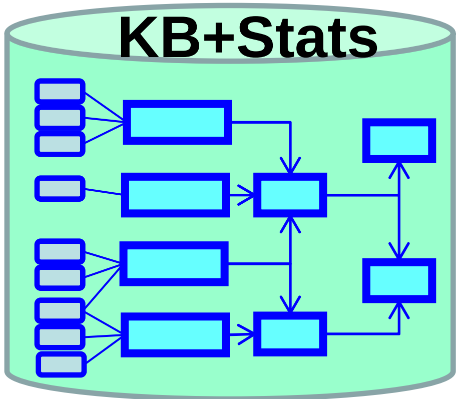
Dollars Trilogy

Popularity (m,e):

- **freq(e|m)**
- **length(e)**
- **#links(e)**

Similarity (m,e):

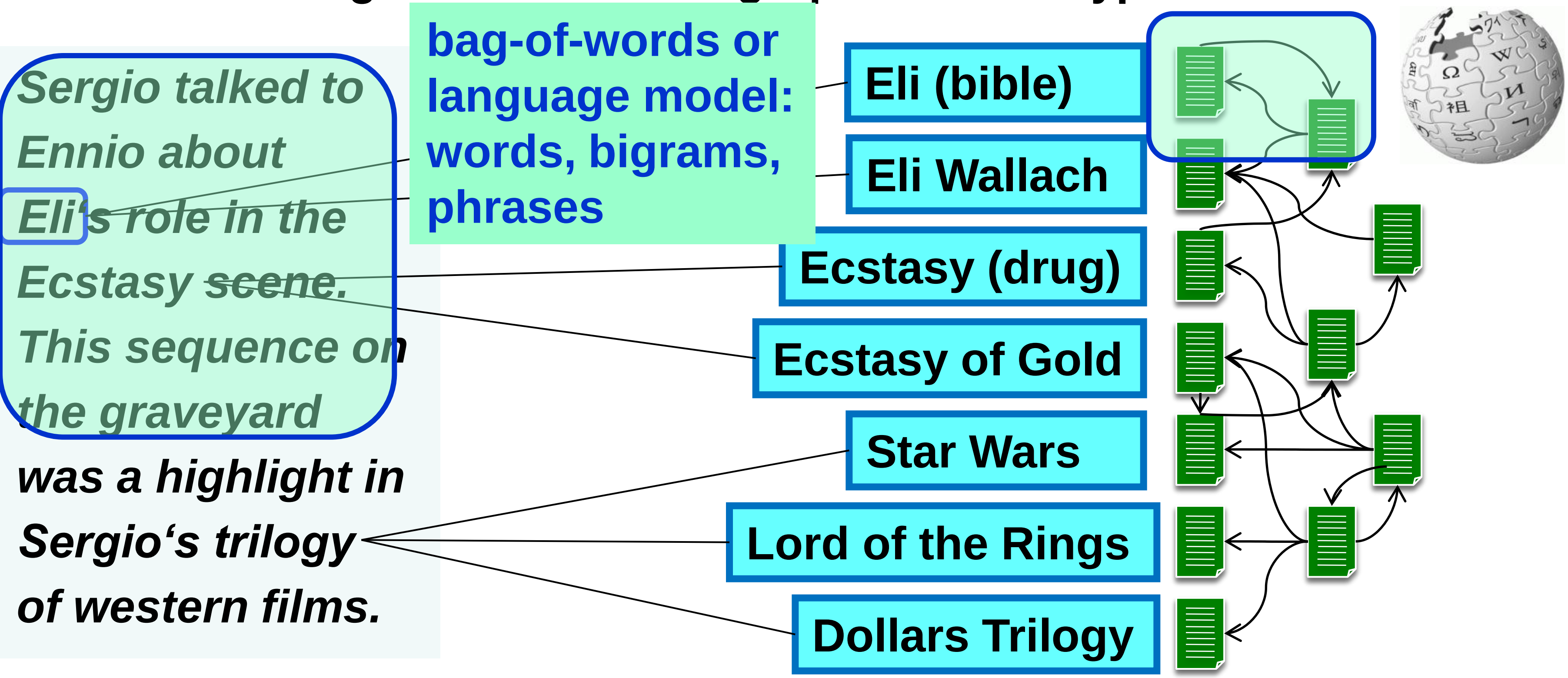
- **cos/Dice/KL (context(m), context(e))**



KB+Stats

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

bag-of-words or language model: words, bigrams, phrases

Eli (bible)

Eli Wallach

Ecstasy (drug)

Ecstasy of Gold

Star Wars

Lord of the Rings

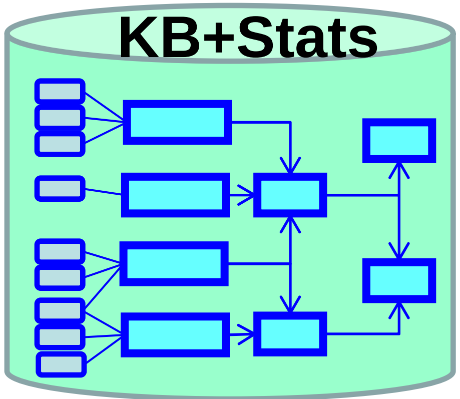
Dollars Trilogy

Popularity (m,e):

- freq(e|m)
- length(e)
- #links(e)

Similarity (m,e):

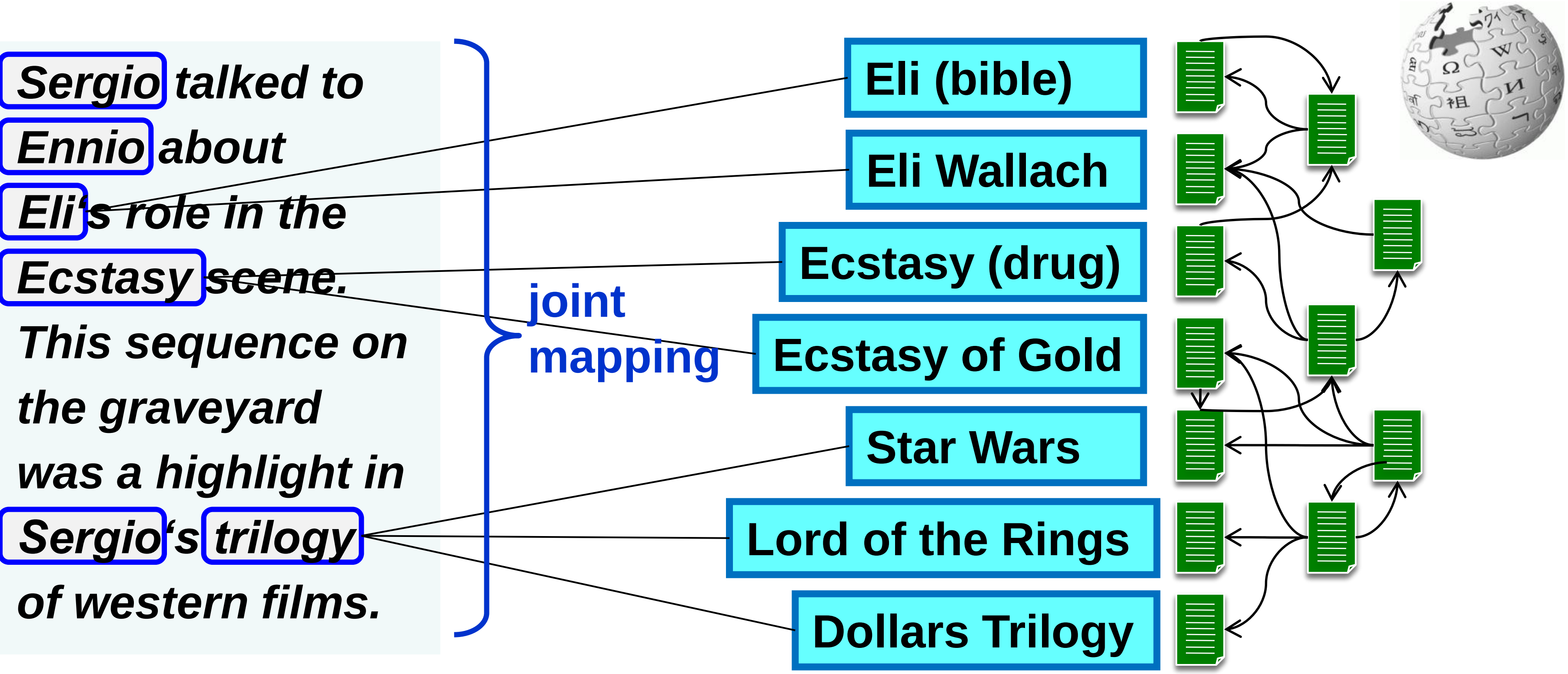
- cos/Dice/KL (context(m), context(e))



KB+Stats

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

joint mapping

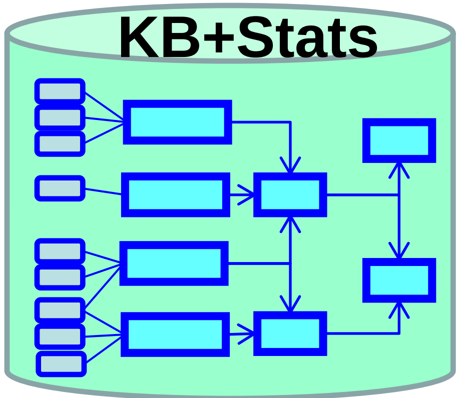
- Eli (bible)
- Eli Wallach
- Ecstasy (drug)
- Ecstasy of Gold
- Star Wars
- Lord of the Rings
- Dollars Trilogy

Popularity (m,e):

- freq(e|m)
- length(e)
- #links(e)

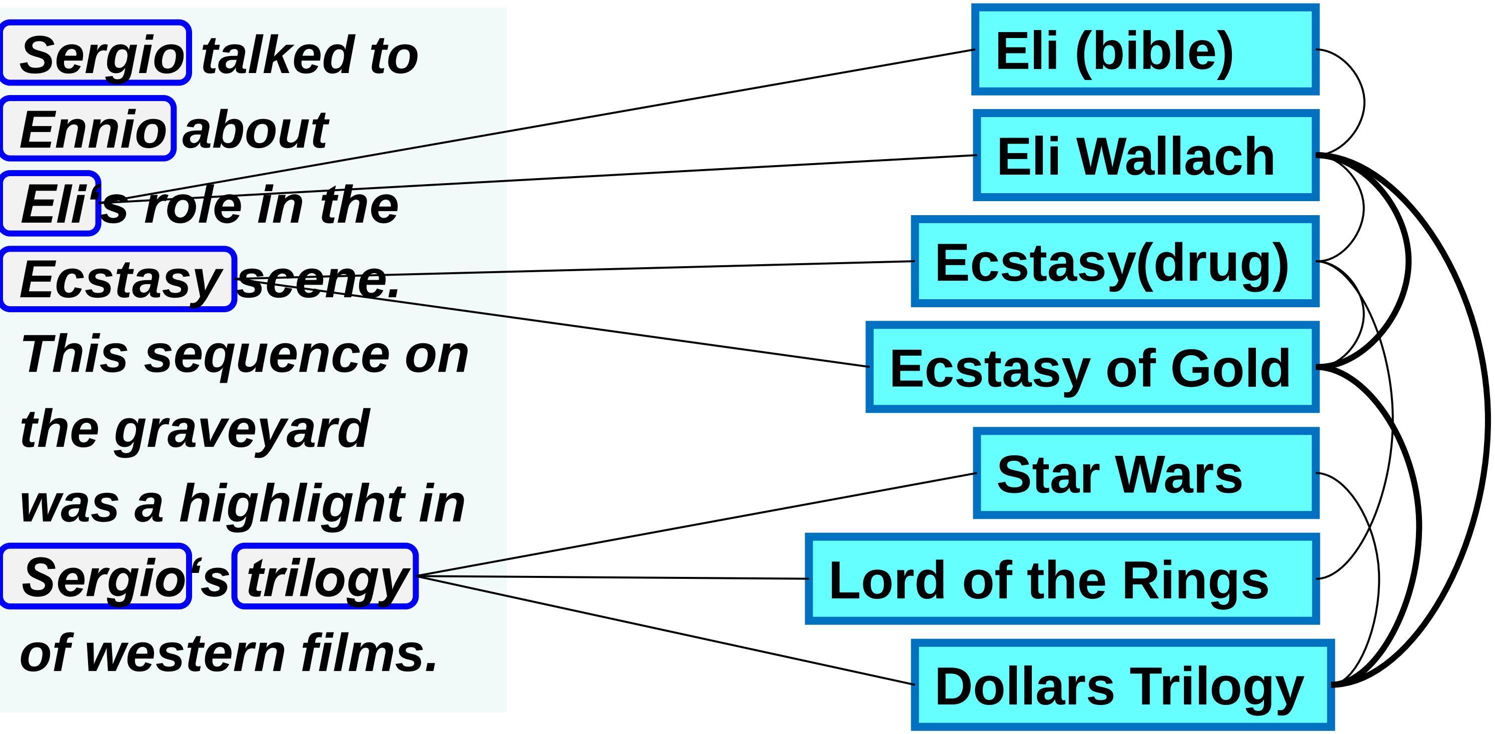
Similarity (m,e):

- cos/Dice/KL (context(m), context(e))



Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

Eli (bible)

Eli Wallach

Ecstasy(drug)

Ecstasy of Gold

Star Wars

Lord of the Rings

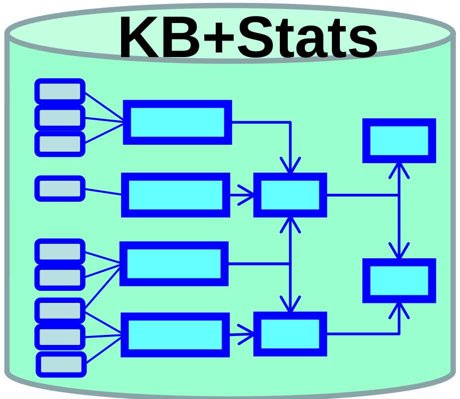
Dollars Trilogy

Popularity (m,e):

- freq(m,e|m)
- length(e)
- #links(e)

Similarity (m,e):

- cos/Dice/KL
  (context(m), context(e))

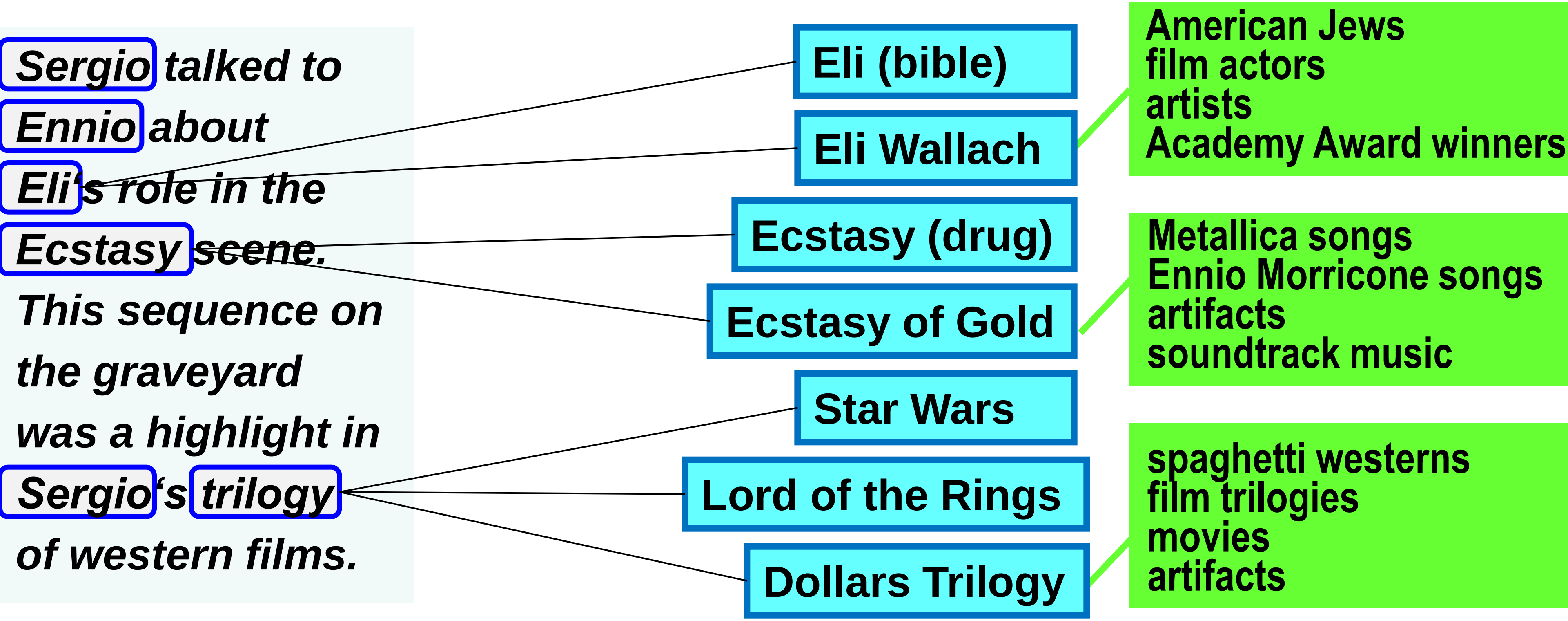


Coherence (e,e'):

- dist(types)
- overlap(links)
- overlap
  (anchor words)

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli‘s role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio‘s trilogy of western films.

Eli (bible)

Eli Wallach

Ecstasy (drug)

Ecstasy of Gold

Star Wars

Lord of the Rings

Dollars Trilogy

American Jews
film actors
artists
Academy Award winners

Metallica songs
Ennio Morricone songs
artifacts
soundtrack music

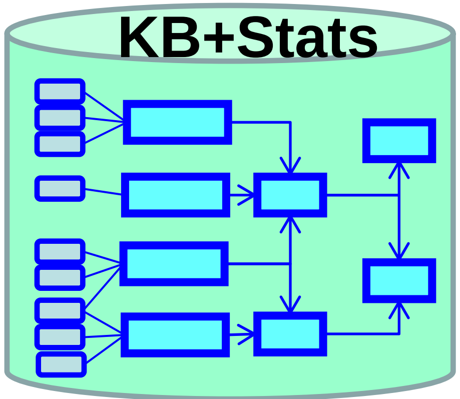
spaghetti westerns
film trilogies
movies
artifacts

Popularity (m,e):

- freq(m,e|m)
- length(e)
- #links(e)

Similarity (m,e):

- cos/Dice/KL (context(m), context(e))

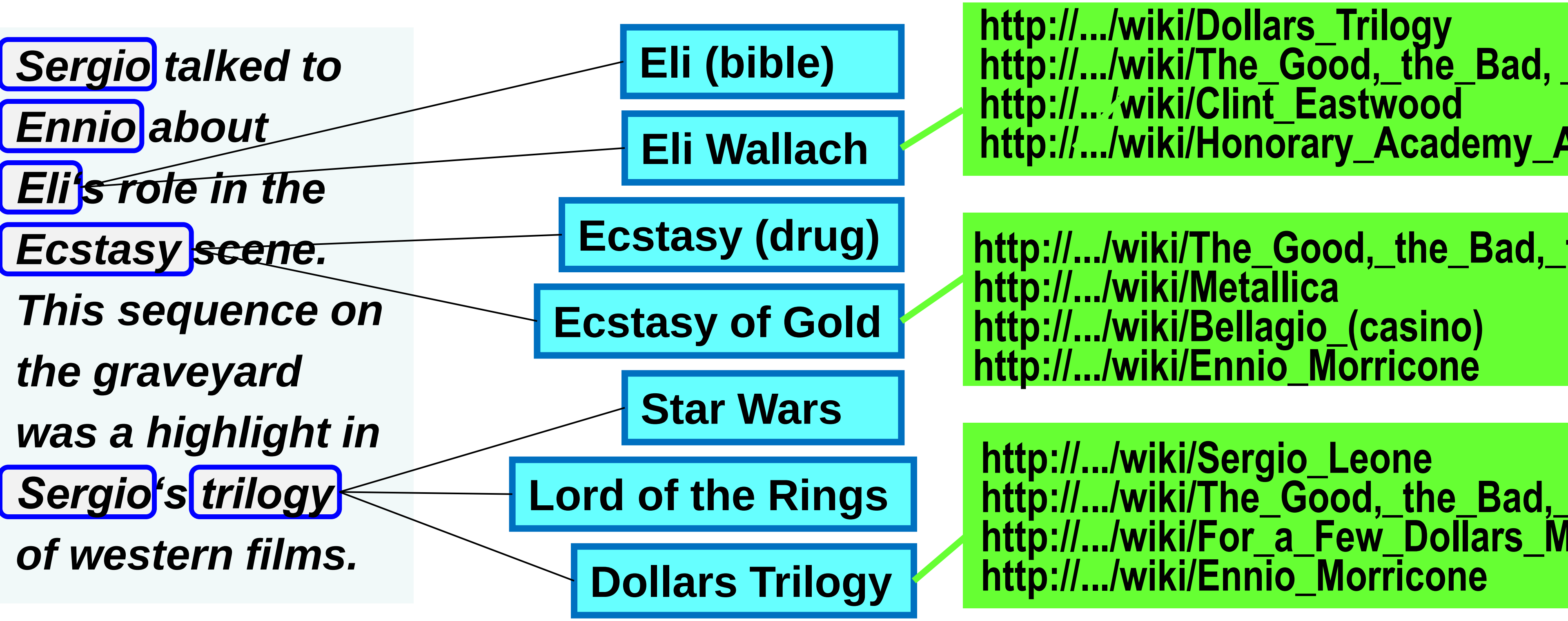


Coherence (e,e‘):

- dist(types)
- overlap(links)
- overlap (anchor words)

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

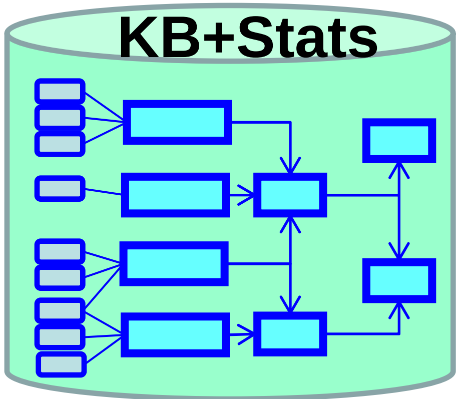
Eli (bible)
Eli Wallach
Ecstasy (drug)
Ecstasy of Gold
Star Wars
Lord of the Rings
Dollars Trilogy
http://.../wiki/Dollars_Trilogy
http://.../wiki/The_Good,_the_Bad,_
http://.../wiki/Clint_Eastwood
http://.../wiki/Honorary_Academy_A
http://.../wiki/The_Good,_the_Bad,_t
http://.../wiki/Metallica
http://.../wiki/Bellagio_(casino)
http://.../wiki/Ennio_Morricone
http://.../wiki/Sergio_Leone
http://.../wiki/The_Good,_the_Bad,_
http://.../wiki/For_a_Few_Dollars_M
http://.../wiki/Ennio_Morricone

Popularity (m,e):

- freq(m,e|m)
- length(e)
- #links(e)

Similarity (m,e):

- cos/Dice/KL (context(m), context(e))

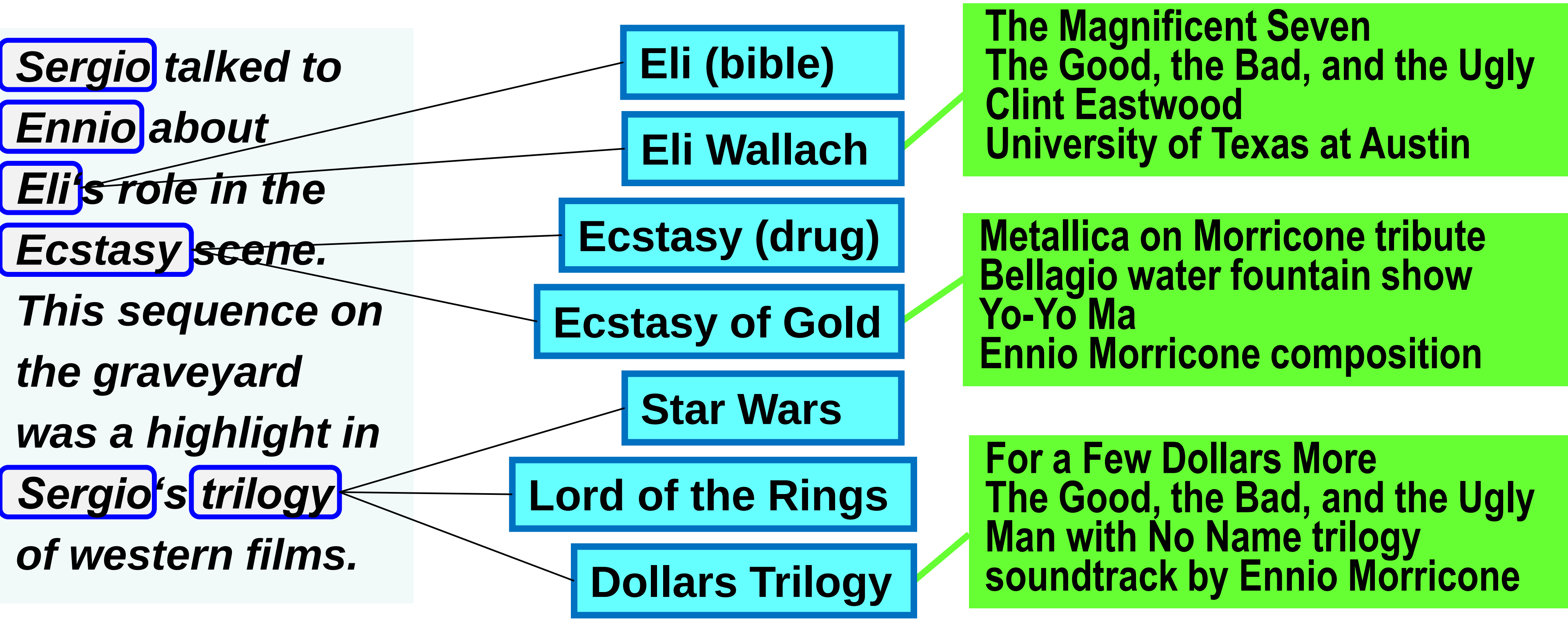


Coherence (e,e'):

- dist(types)
- overlap(links)
- overlap (anchor words)

Mention-Entity Graph

weighted undirected graph with two types of nodes



Sergio talked to Ennio about Eli‘s role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio‘s trilogy of western films.

Eli (bible)

Eli Wallach

Ecstasy (drug)

Ecstasy of Gold

Star Wars

Lord of the Rings

Dollars Trilogy

The Magnificent Seven
The Good, the Bad, and the Ugly
Clint Eastwood
University of Texas at Austin

Metallica on Morricone tribute
Bellagio water fountain show
Yo-Yo Ma
Ennio Morricone composition

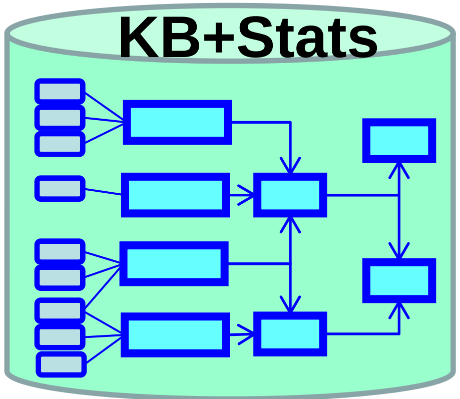
For a Few Dollars More
The Good, the Bad, and the Ugly
Man with No Name trilogy
soundtrack by Ennio Morricone

Popularity (m,e):

- freq(m,e|m)
- length(e)
- #links(e)

Similarity (m,e):

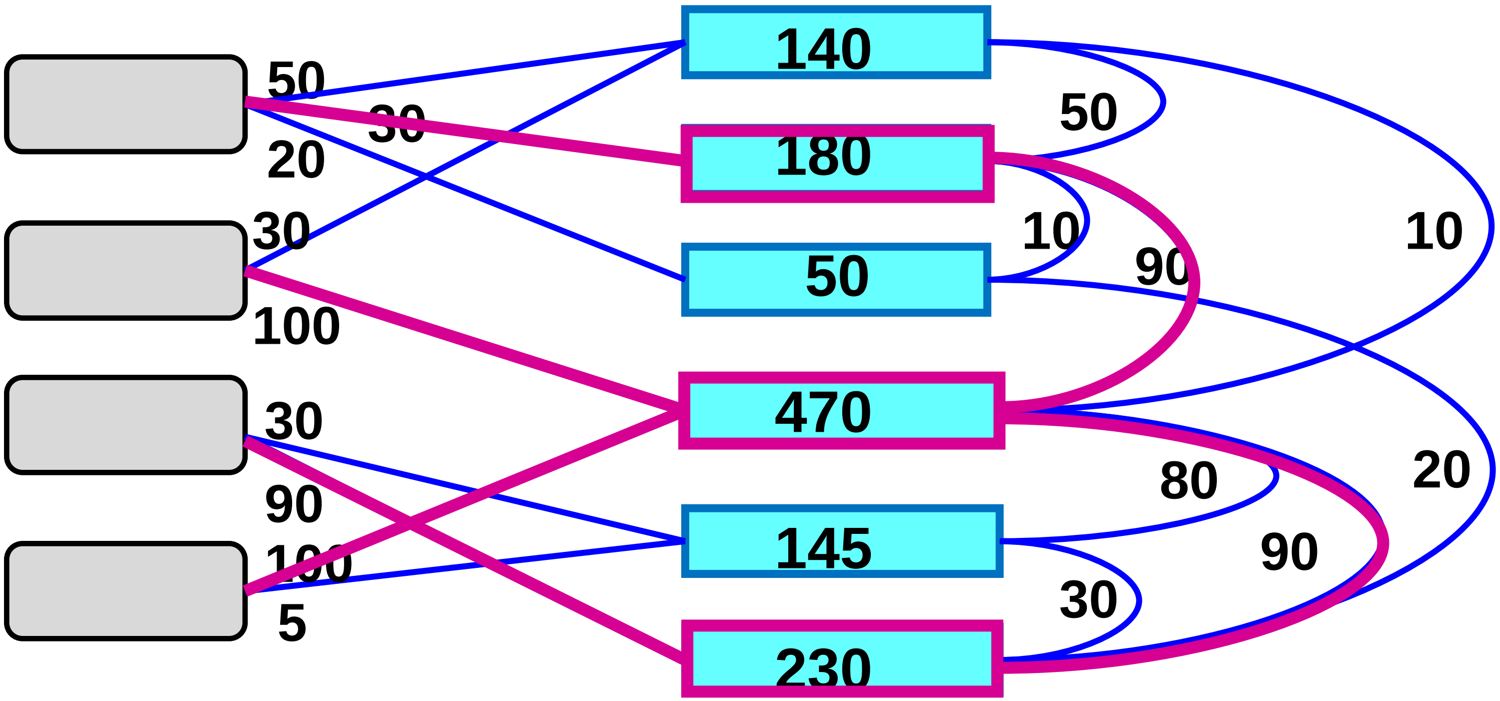
- cos/Dice/KL (context(m), context(e))



Coherence (e,e‘):

- dist(types)
- overlap(links)
- overlap (anchor words)

Graph Algorithm for Joint Mapping



- Compute dense subgraph to

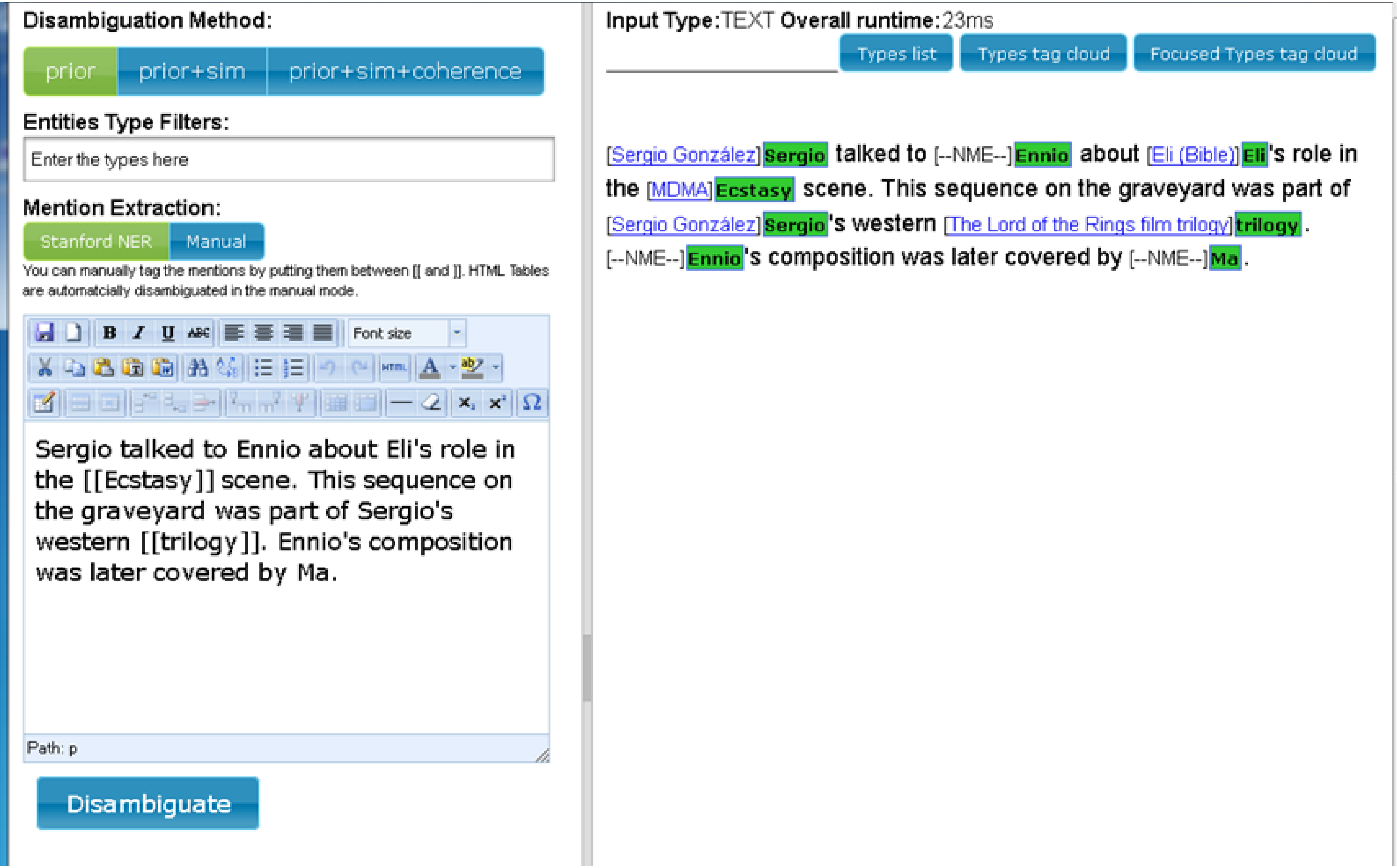
  maximize min weighted degree among entity nodes

  such that:

  each m is connected to exactly one e (or at most one e)
- NP-hard → approximation algorithms

[J. Hoffart et al.: EMNLP‘11, M. Yosef et al.: VLDB‘11, M. Yahya et al.: WWW‘12]

AIDA: Accurate Online Disambiguation



Disambiguation Method:

Entities Type Filters: Enter the types here

Mention Extraction:

You can manually tag the mentions by putting them between [[and]]. HTML Tables are automatcially disambiguated in the manual mode.

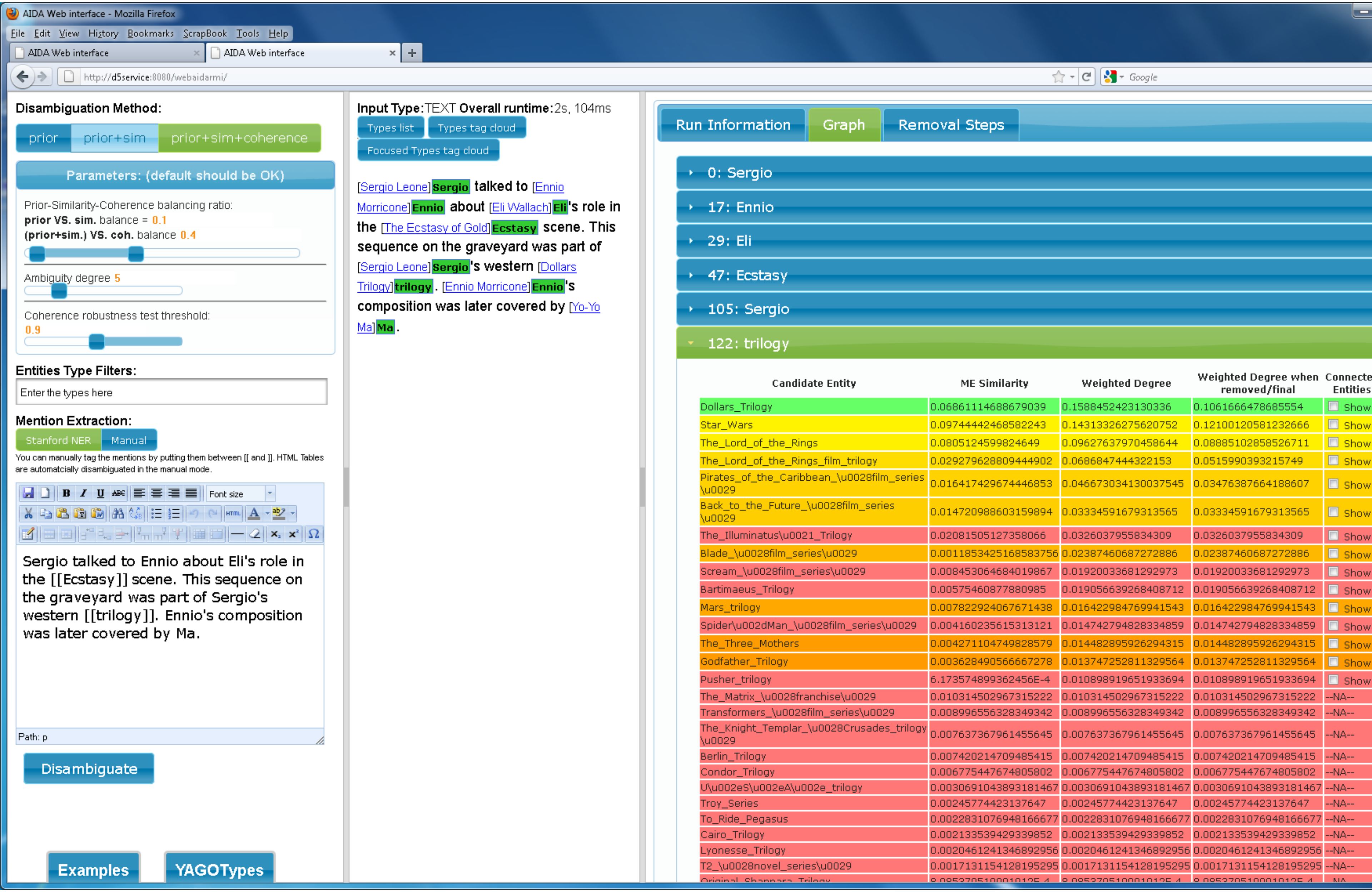
Sergio talked to Ennio about Eli's role in the [[Ecstasy]] scene. This sequence on the graveyard was part of Sergio's western [[trilogy]]. Ennio's composition was later covered by Ma.

Input Type:TEXT **Overall runtime:**23ms

[Sergio González]**Sergio** talked to [--NME--]**Ennio** about [Eli (Bible)]**Eli**'s role in the [MDMA]**Ecstasy** scene. This sequence on the graveyard was part of [Sergio González]**Sergio**'s western [The Lord of the Rings film trilogy]**trilogy**. [--NME--]**Ennio**'s composition was later covered by [--NME--]**Ma**.

http://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation


AIDA Web interface - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA Web interface
AIDA Web interface
http://d5service:8080/webaidarmi/
Google

Disambiguation Method:

prior
prior+sim
prior+sim+coherence
Parameters: (default should be OK)

Prior-Similarity-Coherence balancing ratio:
prior VS. sim. balance = 0.1
(prior+sim.) VS. coh. balance 0.4

Ambiguity degree 5

Coherence robustness test threshold: 0.9

Entities Type Filters:

Enter the types here

Mention Extraction:

You can manually tag the mentions by putting them between [[and]]. HTML Tables are automatcially disambiguated in the manual mode.

Sergio talked to Ennio about Eli's role in the [[Ecstasy]] scene. This sequence on the graveyard was part of Sergio's western [[trilogy]]. Ennio's composition was later covered by Ma.

Path: p

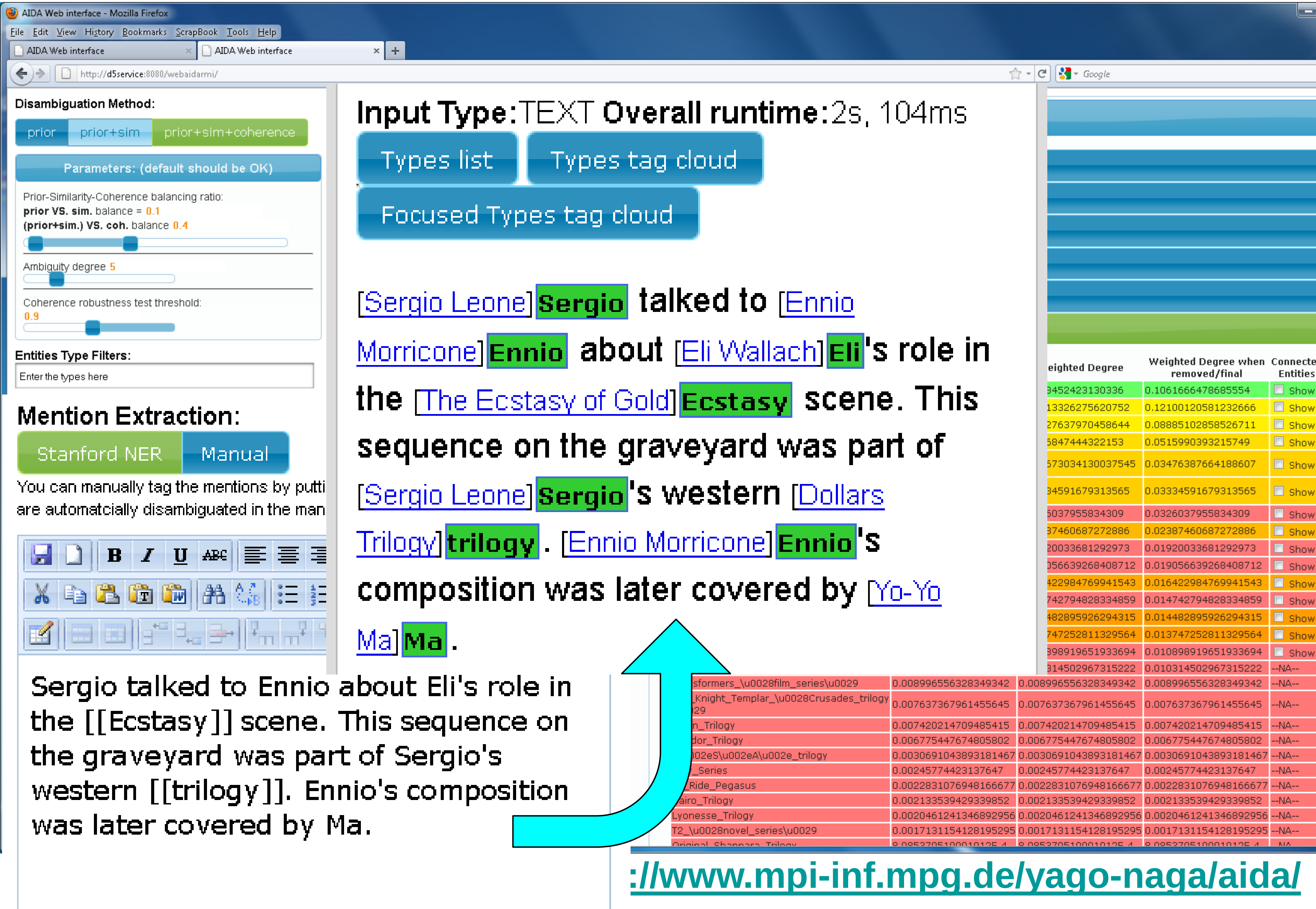
Input Type:TEXT **Overall runtime:**2s, 104ms

[Sergio Leone]**Sergio** talked to [Ennio Morricone]**Ennio** about [Eli Wallach]**Eli**'s role in the [The Ecstasy of Gold]**Ecstasy** scene. This sequence on the graveyard was part of [Sergio Leone]**Sergio**'s western [Dollars Trilogy]**trilogy**. [Ennio Morricone]**Ennio**'s composition was later covered by [Yo-Yo Ma]**Ma**.

Candidate Entity	ME Similarity	Weighted Degree	Weighted Degree when removed/final	Connecte Entities
Dollars_Trilogy	0.06861114688679039	0.1588452423130336	0.106166478685554	□ Show
Star_Wars	0.09744442468582243	0.14313326275620752	0.12100120581232666	□ Show
The_Lord_of_the_Rings	0.0805124599824649	0.09627637970458644	0.08885102858526711	□ Show
The_Lord_of_the_Rings_film_trilogy	0.029279628809444902	0.0686847444322153	0.051599039321 5749	□ Show
Pirates_of_the_Caribbean_\u0028film_series \u0029	0.016417429674446853	0.046673034130037545	0.03476387664188607	□ Show
Back_to_the_Future_\u0028film_series \u0029	0.014720988603159894	0.03334591679313565	0.03334591679313565	□ Show
The_Illuminatus\u0021_Trilogy	0.02081505127358066	0.0326037955834309	0.0326037955834309	□ Show
Blade_\u0028film_series\u0029	0.0011853425168583756	0.02387460687272886	0.02387460687272886	□ Show
Scream_\u0028film_series\u0029	0.008453064684019867	0.01920033681292973	0.01920033681292973	□ Show
Bartimaeus_Trilogy	0.00575460877880985	0.019056639268408712	0.019056639268408712	□ Show
Mars_trilogy	0.007822924067671438	0.016422984769941543	0.016422984769941543	□ Show
Spider\u002dMan_\u0028film_series\u0029	0.004160235615313121	0.014742794828334859	0.014742794828334859	□ Show
The_Three_Mothers	0.004271104749828579	0.014482895926294315	0.014482895926294315	□ Show
Godfather_Trilogy	0.003628490566667278	0.013747252811329564	0.013747252811329564	□ Show
Pusher_trilogy	6.173574899362456E-4	0.010898919651933694	0.010898919651933694	□ Show
The_Matrix_\u0028franchise\u0029	0.010314502967315222	0.010314502967315222	0.010314502967315222	--NA--
Transformers_\u0028film_series\u0029	0.008996556328349342	0.008996556328349342	0.008996556328349342	--NA--
The_Knight_Templar_\u0028Crusades_trilogy \u0029	0.007637367961455645	0.007637367961455645	0.007637367961455645	--NA--
Berlin_Trilogy	0.007420214709485415	0.007420214709485415	0.007420214709485415	--NA--
Condor_Trilogy	0.006775447674805802	0.006775447674805802	0.006775447674805802	--NA--
U\u002eS\u002eA\u002e_trilogy	0.0030691043893181467	0.0030691043893181467	0.0030691043893181467	--NA--
Troy_Series	0.002457744231 37647	0.00245774423137647	0.00245774423137647	--NA--
To_Ride_Pegasus	0.0022831076948166677	0.0022831076948166677	0.0022831076948166677	--NA--
Cairo_Trilogy	0.002133539429339852	0.002133539429339852	0.002133539429339852	--NA--
Lyonesse_Trilogy	0.0020461241346892956	0.0020461241346892956	0.0020461241346892956	--NA--
T2_\u0028novel_series\u0029	0.0017131154128195295	0.0017131154128195295	0.0017131154128195295	--NA--
Original_Shannara_Trilogy	[illegible]	[illegible]	[illegible]	--NA--

http://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation


AIDA Web interface - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA Web interface
AIDA Web interface
http://d5service:8080/webaidarmi/
Google
Types list
Types tag cloud
Focused Types tag cloud

Disambiguation Method:

prior | prior+sim | prior+sim+coherence

Parameters: (default should be OK)

Prior-Similarity-Coherence balancing ratio:
prior VS. sim. balance = 0.1
(prior+sim.) VS. coh. balance 0.4

Ambiguity degree 5

Coherence robustness test threshold:
0.9

Entities Type Filters:

Enter the types here

Mention Extraction:

Stanford NER | Manual

You can manually tag the mentions by putti
are automatcially disambiguated in the man

Sergio talked to Ennio about Eli's role in the [[Ecstasy]] scene. This sequence on the graveyard was part of Sergio's western [[trilogy]]. Ennio's composition was later covered by Ma.

Input Type:TEXT **Overall runtime:**2s, 104ms

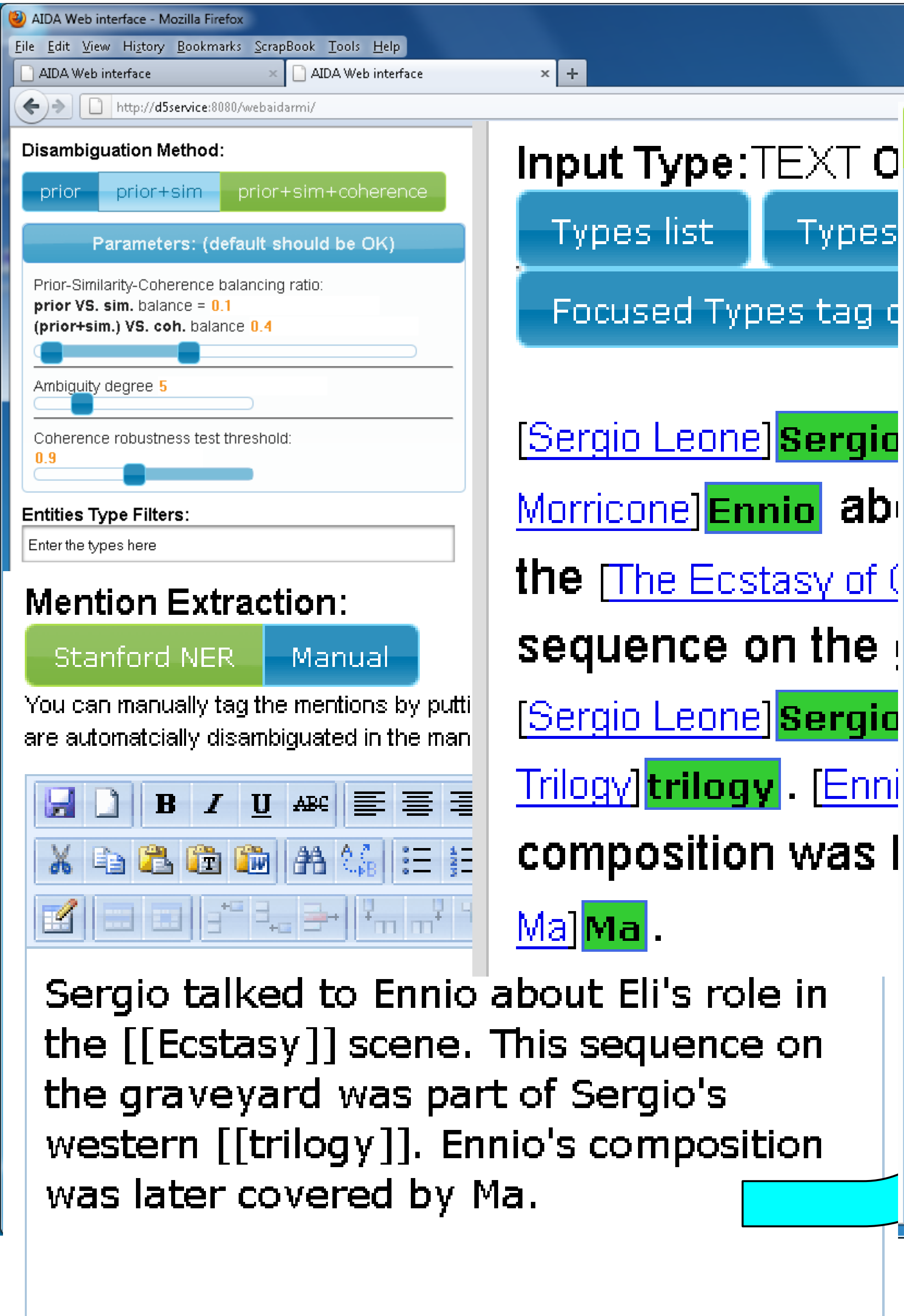
[Sergio Leone] **Sergio** talked to [Ennio Morricone] **Ennio** about [Eli Wallach] **Eli**'s role in the [The Ecstasy of Gold] **Ecstasy** scene. This sequence on the graveyard was part of [Sergio Leone] **Sergio**'s western [Dollars Trilogy] **trilogy**. [Ennio Morricone] **Ennio**'s composition was later covered by [Yo-Yo Ma] **Ma**.

eighted Degree	Weighted Degree when removed/final	Connecte Entities
3452423130336	0.1061666478685554	□ Show
13326275620752	0.12100120581232666	□ Show
27637970458644	0.08885102858526711	□ Show
5847444322153	0.0515990393215749	□ Show
573034130037545	0.03476387664188607	□ Show
34591679313565	0.03334591679313565	□ Show
5037955834309	0.0326037955834309	□ Show
37460687272886	0.02387460687272886	□ Show
20033681292973	0.01920033681292973	□ Show
056639268408712	0.019056639268408712	□ Show
422984769941543	0.016422984769941543	□ Show
742794828334859	0.014742794828334859	□ Show
482895926294315	0.014482895926294315	□ Show
747252811329564	0.013747252811329564	□ Show
398919651933694	0.010898919651933694	□ Show
314502967315222	0.010314502967315222	--NA--

sformers_\u0028film_series\u0029	0.008996556328349342	0.008996556328349342	0.008996556328349342	--NA--
_Knight_Templar_\u0028Crusades_trilogy 29	0.007637367961455645	0.007637367961455645	0.007637367961455645	--NA--
n_Trilogy	0.007420214709485415	0.007420214709485415	0.007420214709485415	--NA--
dor_Trilogy	0.006775447674805802	0.006775447674805802	0.006775447674805802	--NA--
02eS\u002eA\u002e_trilogy	0.0030691043893181467	0.0030691043893181467	0.0030691043893181467	--NA--
_Series	0.00245774423137647	0.00245774423137647	0.00245774423137647	--NA--
Ride_Pegasus	0.0022831076948166677	0.0022831076948166677	0.0022831076948166677	--NA--
airo_Trilogy	0.002133539429339852	0.002133539429339852	0.002133539429339852	--NA--
Lyonesse_Trilogy	0.0020461241346892956	0.0020461241346892956	0.0020461241346892956	--NA--
T2_\u0028novel_series\u0029	0.0017131154128195295	0.0017131154128195295	0.0017131154128195295	--NA--
Original_Shannara_Trilogy	[illegible]	[illegible]	[illegible]	[illegible]

://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation


AIDA Web interface - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA Web interface
AIDA Web interface
http://d5service:8080/webaidarmi/
prior
prior+sim
prior+sim+coherence
Parameters: (default should be OK)
Types list
Types
Focused Types tag c
Stanford NER
Manual

Disambiguation Method:

Prior-Similarity-Coherence balancing ratio:
prior VS. sim. balance = 0.1
(prior+sim.) VS. coh. balance 0.4

Ambiguity degree 5

Coherence robustness test threshold:
0.9

Entities Type Filters:

Enter the types here

Mention Extraction:

You can manually tag the mentions by putti
are automatcially disambiguated in the man

Input Type:TEXT O

[Sergio Leone] **Sergio**

Morricone] **Ennio** ab

the [The Ecstasy of (

sequence on the

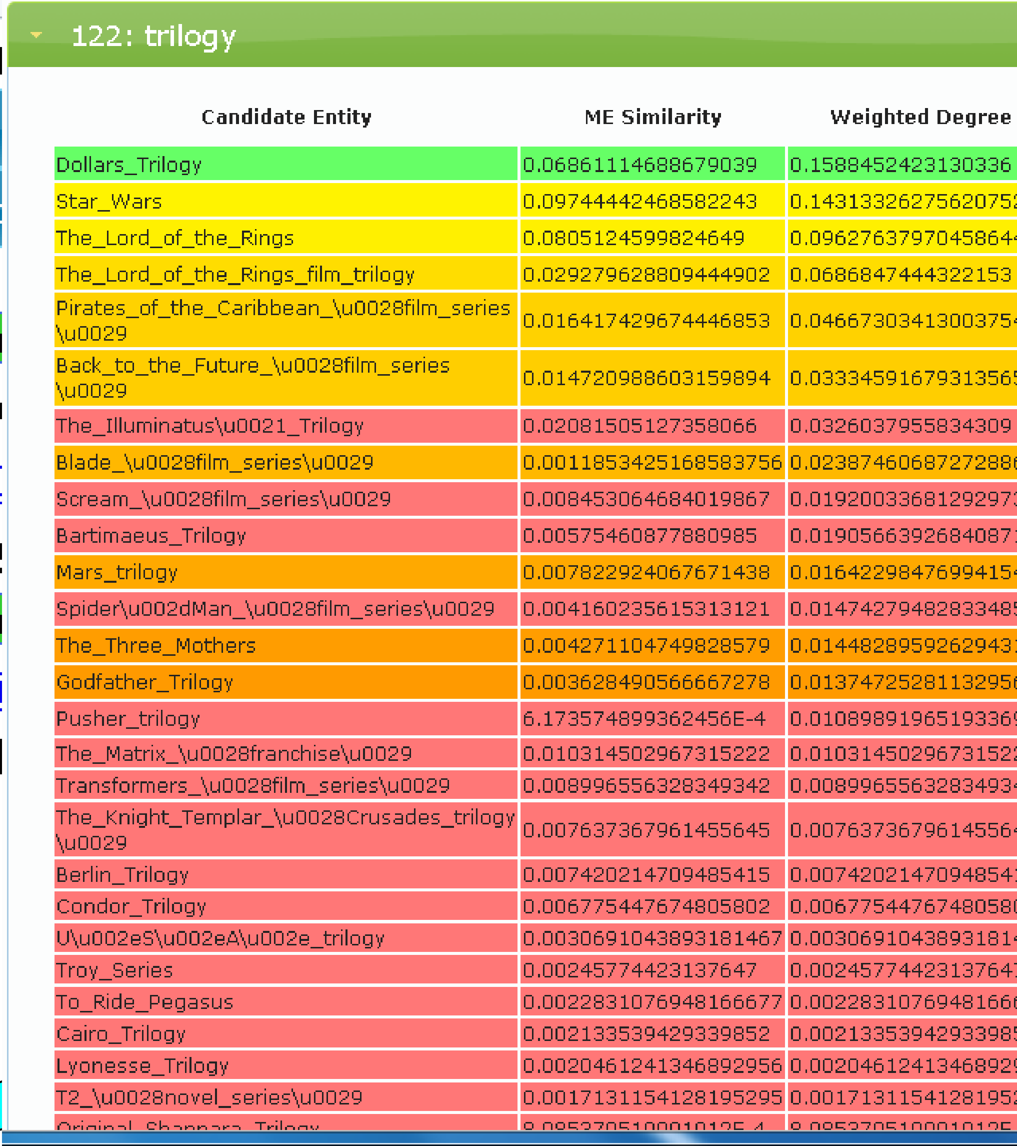
[Sergio Leone] **Sergio**

Trilogy] **trilogy** . [Enni

composition was l

Ma] **Ma** .

Sergio talked to Ennio about Eli's role in the [[Ecstasy]] scene. This sequence on the graveyard was part of Sergio's western [[trilogy]]. Ennio's composition was later covered by Ma.

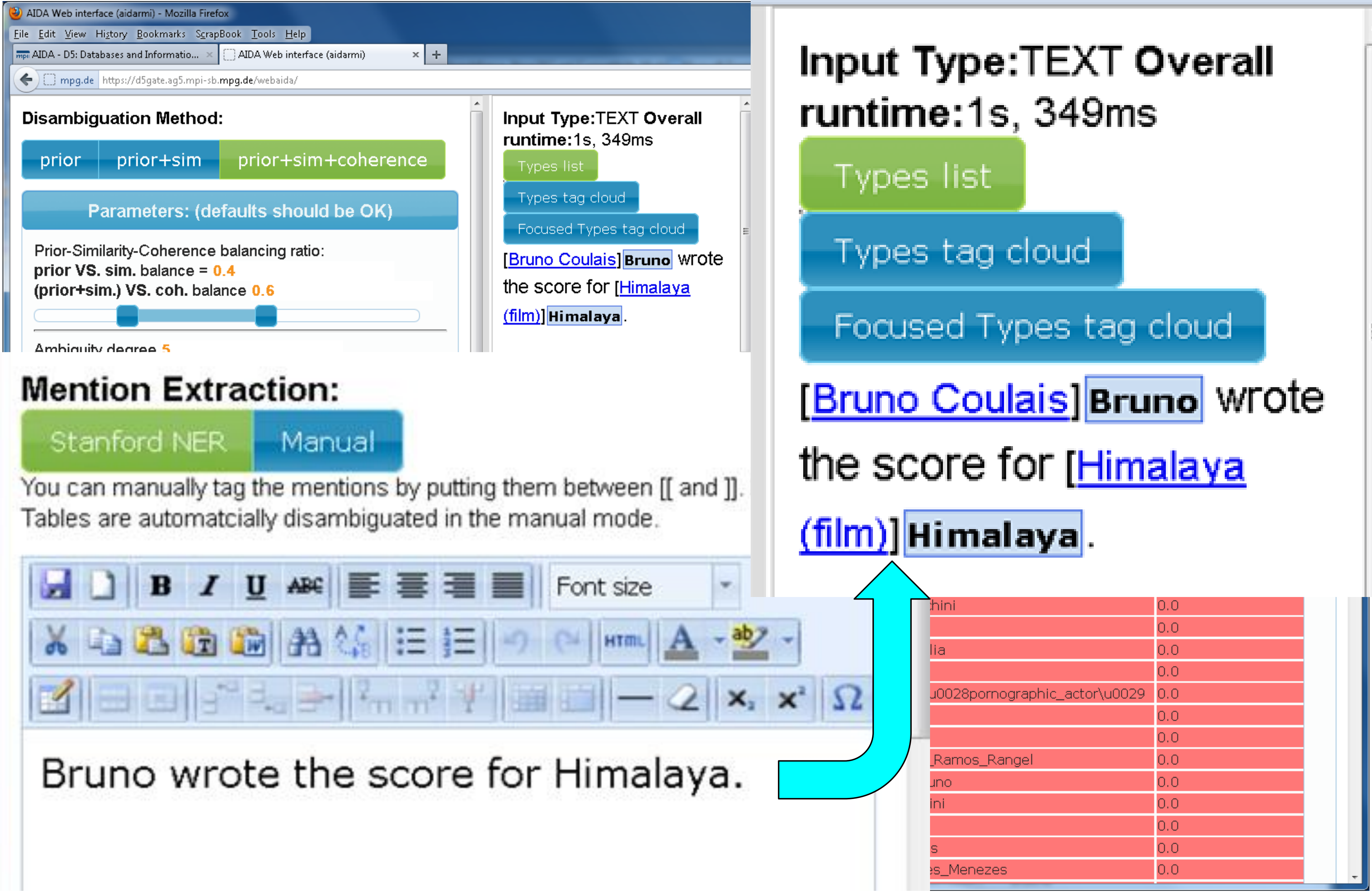


122: trilogy

Candidate Entity	ME Similarity	Weighted Degree
Dollars_Trilogy	0.06861114688679039	0.1588452423130336
Star_Wars	0.09744442468582243	0.1431332627562075
The_Lord_of_the_Rings	0.0805124599824649	0.0962763797045864
The_Lord_of_the_Rings_film_trilogy	0.029279628809444902	0.0686847444322153
Pirates_of_the_Caribbean_\u0028film_series \u0029	0.016417429674446853	0.0466730341300375
Back_to_the_Future_\u0028film_series \u0029	0.014720988603159894	0.0333459167931356
The_Illuminatus\u0021_Trilogy	0.02081505127358066	0.0326037955834309
Blade_\u0028film_series\u0029	0.0011853425168583756	0.0238746068727288
Scream_\u0028film_series\u0029	0.008453064684019867	0.0192003368129297
Bartimaeus_Trilogy	0.00575460877880985	0.0190566392684087
Mars_trilogy	0.007822924067671438	0.0164229847699415
Spider\u002dMan_\u0028film_series\u0029	0.004160235615313121	0.0147427948283348
The_Three_Mothers	0.004271104749828579	0.0144828959262943
Godfather_Trilogy	0.003628490566667278	0.0137472528113295
Pusher_trilogy	6.173574899362456E-4	0.0108989196519336
The_Matrix_\u0028franchise\u0029	0.010314502967315222	0.0103145029673152
Transformers_\u0028film_series\u0029	0.008996556328349342	0.0089965563283493
The_Knight_Templar_\u0028Crusades_trilogy \u0029	0.007637367961455645	0.0076373679614556
Berlin_Trilogy	0.007420214709485415	0.0074202147094854
Condor_Trilogy	0.006775447674805802	0.0067754476748058
U\u002eS\u002eA\u002e_trilogy	0.0030691043893181467	0.0030691043893181
Troy_Series	0.00245774423137647	0.0024577442313764
To_Ride_Pegasus	0.0022831076948166677	0.0022831076948166
Cairo_Trilogy	0.002133539429339852	0.0021335394293398
Lyonesse_Trilogy	0.0020461241346892956	0.0020461241346892
T2_\u0028novel_series\u0029	0.0017131154128195295	0.0017131154128195
Original_Shannara_Trilogy	[illegible]	[illegible]

://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation


AIDA Web interface (aidarmi) - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA - D5: Databases and Informatio...
AIDA Web interface (aidarmi)
mpg.de https://d5gate.ag5.mpi-sb.mpg.de/webaida/
Font size

Disambiguation Method:

prior | prior+sim | prior+sim+coherence

Parameters: (defaults should be OK)

Prior-Similarity-Coherence balancing ratio:
prior VS. sim. balance = 0.4
(prior+sim.) VS. coh. balance 0.6

Ambiguity degree 5

Mention Extraction:

Stanford NER | Manual

You can manually tag the mentions by putting them between [[and]].
Tables are automatcially disambiguated in the manual mode.

Bruno wrote the score for Himalaya.

Input Type:TEXT **Overall runtime:**1s, 349ms

Types list

Types tag cloud

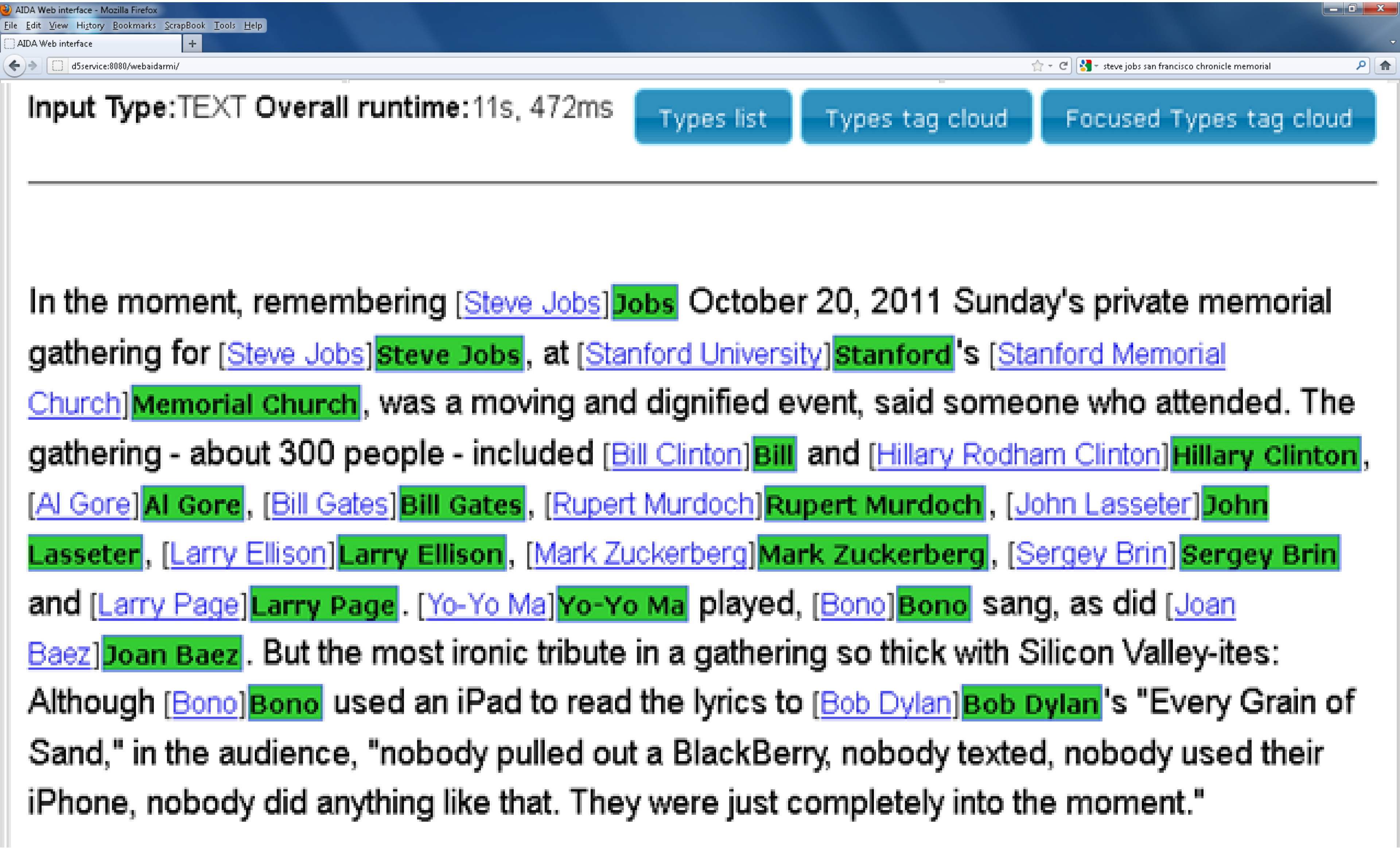
Focused Types tag cloud

[Bruno Coulais] **Bruno** wrote the score for [Himalaya (film)] **Himalaya**.

hini	0.0
	0.0
lia	0.0
	0.0
u0028pornographic_actor\u0029	0.0
	0.0
	0.0
_Ramos_Rangel	0.0
uno	0.0
ini	0.0
	0.0
s	0.0
es_Menezes	0.0

http://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation

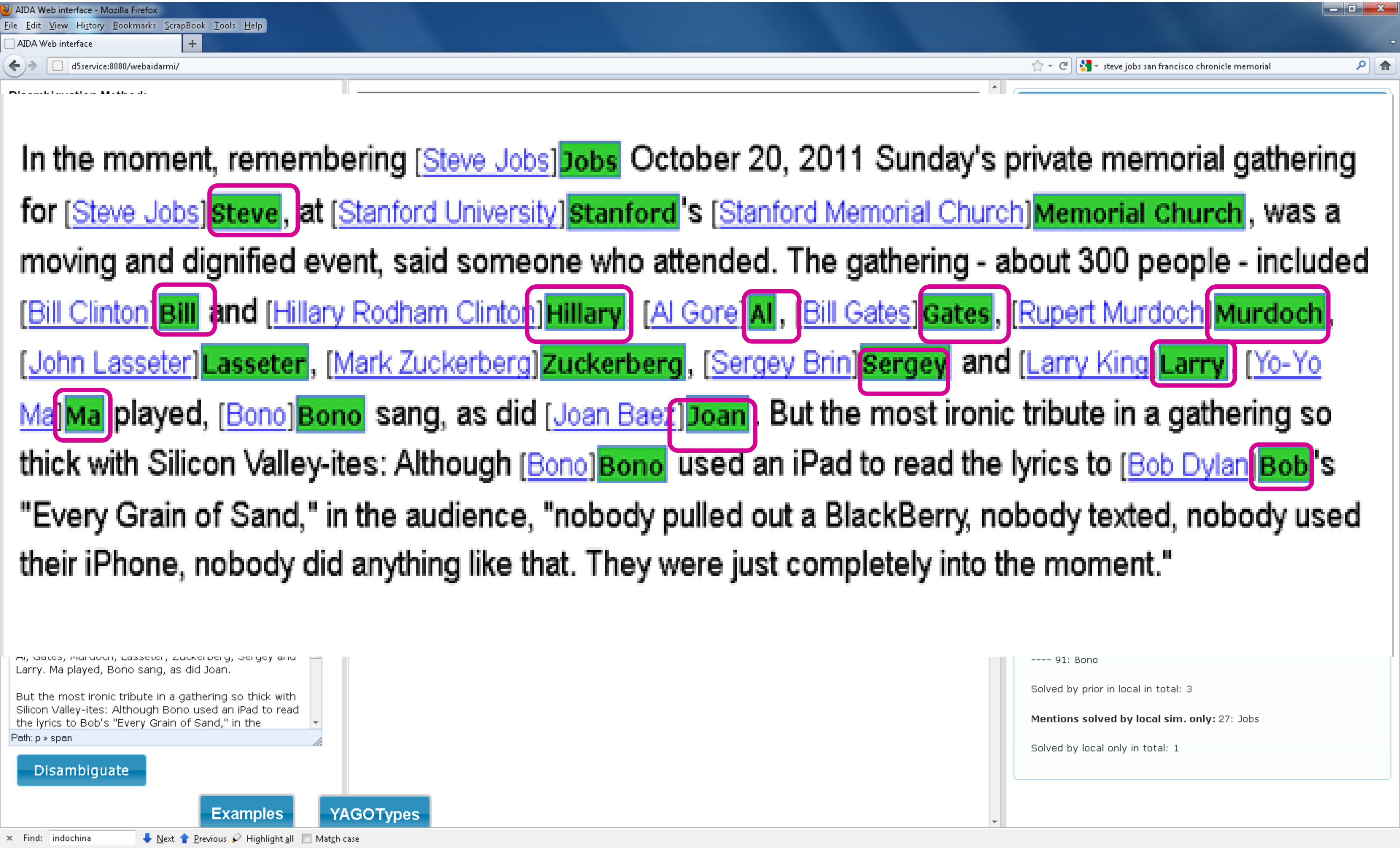

AIDA Web interface - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA Web interface
d5service:8080/webaidarmi/
steve jobs san francisco chronicle memorial
Types list
Types tag cloud
Focused Types tag cloud

Input Type:TEXT **Overall runtime:**11s, 472ms

In the moment, remembering [Steve Jobs]**Jobs** October 20, 2011 Sunday's private memorial gathering for [Steve Jobs]**Steve Jobs**, at [Stanford University]**Stanford**'s [Stanford Memorial Church]**Memorial Church**, was a moving and dignified event, said someone who attended. The gathering - about 300 people - included [Bill Clinton]**Bill** and [Hillary Rodham Clinton]**Hillary Clinton**, [Al Gore]**Al Gore**, [Bill Gates]**Bill Gates**, [Rupert Murdoch]**Rupert Murdoch**, [John Lasseter]**John Lasseter**, [Larry Ellison]**Larry Ellison**, [Mark Zuckerberg]**Mark Zuckerberg**, [Sergey Brin]**Sergey Brin** and [Larry Page]**Larry Page**. [Yo-Yo Ma]**Yo-Yo Ma** played, [Bono]**Bono** sang, as did [Joan Baez]**Joan Baez**. But the most ironic tribute in a gathering so thick with Silicon Valley-ites: Although [Bono]**Bono** used an iPad to read the lyrics to [Bob Dylan]**Bob Dylan**'s "Every Grain of Sand," in the audience, "nobody pulled out a BlackBerry, nobody texted, nobody used their iPhone, nobody did anything like that. They were just completely into the moment."

http://www.mpi-inf.mpg.de/yago-naga/aida/

AIDA: Accurate Online Disambiguation


AIDA Web interface - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help
AIDA Web interface
d5service:8080/webaidarmi/
steve jobs san francisco chronicle memorial
Disambiguation Method:
Disambiguate
Examples
YAGOTypes
Find: indochina
Next
Previous
Highlight all
Match case

In the moment, remembering [Steve Jobs]Jobs October 20, 2011 Sunday's private memorial gathering for [Steve Jobs]Steve, at [Stanford University]Stanford's [Stanford Memorial Church]Memorial Church, was a moving and dignified event, said someone who attended. The gathering - about 300 people - included [Bill Clinton]Bill and [Hillary Rodham Clinton]Hillary, [Al Gore]Al, [Bill Gates]Gates, [Rupert Murdoch]Murdoch, [John Lasseter]Lasseter, [Mark Zuckerberg]Zuckerberg, [Sergey Brin]Sergey and [Larry King]Larry. [Yo-Yo Ma]Ma played, [Bono]Bono sang, as did [Joan Baez]Joan. But the most ironic tribute in a gathering so thick with Silicon Valley-ites: Although [Bono]Bono used an iPad to read the lyrics to [Bob Dylan]Bob's "Every Grain of Sand," in the audience, "nobody pulled out a BlackBerry, nobody texted, nobody used their iPhone, nobody did anything like that. They were just completely into the moment."

Al, Gates, Murdoch, Lasseter, Zuckerberg, Sergey and Larry. Ma played, Bono sang, as did Joan.

But the most ironic tribute in a gathering so thick with Silicon Valley-ites: Although Bono used an iPad to read the lyrics to Bob's "Every Grain of Sand," in the

Path: p » span

---- 91: Bono

Solved by prior in local in total: 3

Mentions solved by local sim. only: 27: Jobs

Solved by local only in total: 1

http://www.mpi-inf.mpg.de/yago-naga/aida/

General Word Sense Disambiguation

Which

song writers

covered

ballads

written by

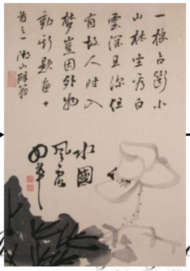
the Stones ?

{songwriter, composer}

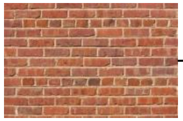
{cover, perform}

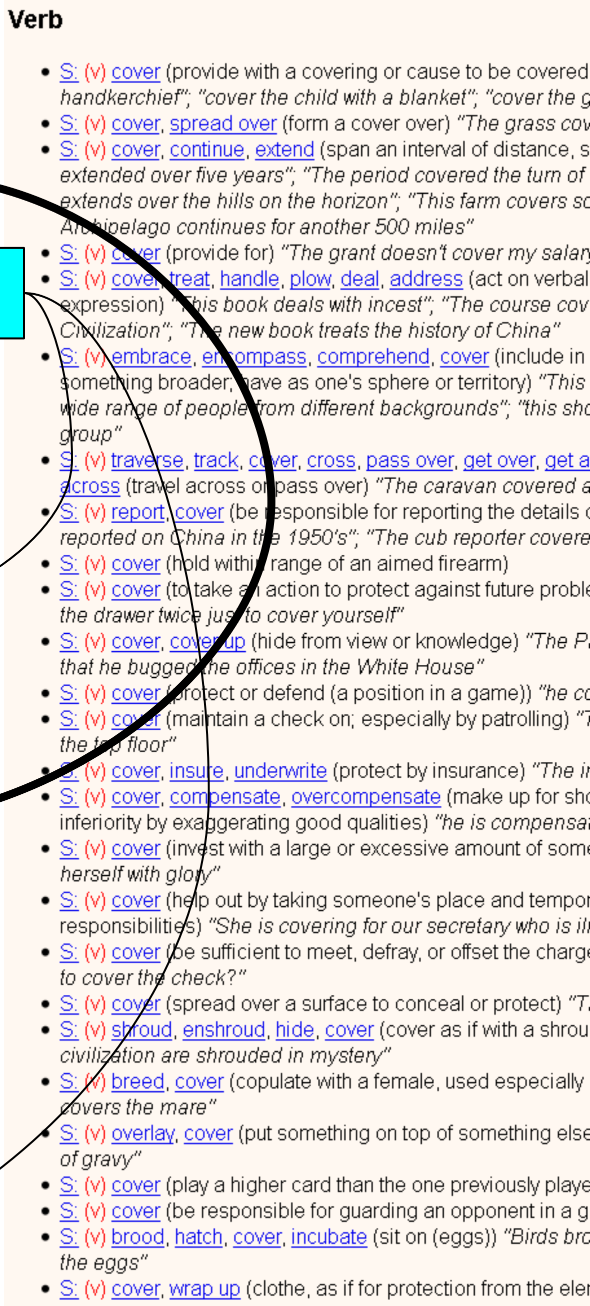
{cover, report, treat}

{cover, help out}



Robert Burns 1759 - 1796
Violin
Vln.
Vln.
Should auld ac-quaint-ance be for-got and
And there's a hand my trus-ty friend and
nev-er brought to mind? Should auld ac-quaint-ance
give a hand of thine! We'll take a right good-
be for-got and days of Auld Lang Syne? For
will draught for Auld Lang Syne For





Verb

- S: (v) cover (provide with a covering or cause to be covered
handkerchief"; "cover the child with a blanket"; "cover the g
- S: (v) cover, spread over (form a cover over) *"The grass cov*
- S: (v) cover, continue, extend (span an interval of distance, s
extended over five years"; "The period covered the turn of
extends over the hills on the horizon"; "This farm covers so
Archipelago continues for another 500 miles"
- S: (v) cover (provide for) *"The grant doesn't cover my salary*
- S: (v) cover, treat, handle, plow, deal, address (act on verbal
expression) *"This book deals with incest"; "The course cov*
Civilization"; "The new book treats the history of China"
- S: (v) embrace, encompass, comprehend, cover (include in
something broader; have as one's sphere or territory) *"This*
wide range of people from different backgrounds"; "this sho
group"
- S: (v) traverse, track, cover, cross, pass over, get over, get a
across (travel across or pass over) *"The caravan covered a*
- S: (v) report, cover (be responsible for reporting the details o
reported on China in the 1950's"; "The cub reporter covere
- S: (v) cover (hold within range of an aimed firearm)
- S: (v) cover (to take an action to protect against future proble
the drawer twice just to cover yourself"
- S: (v) cover, cover up (hide from view or knowledge) *"The Pa*
that he bugged the offices in the White House"
- S: (v) cover (protect or defend (a position in a game)) *"he co*
- S: (v) cover (maintain a check on; especially by patrolling) *"T*
the top floor"
- S: (v) cover, insure, underwrite (protect by insurance) *"The in*
- S: (v) cover, compensate, overcompensate (make up for sho
inferiority by exaggerating good qualities) *"he is compensa*
- S: (v) cover (invest with a large or excessive amount of some
herself with glory"
- S: (v) cover (help out by taking someone's place and tempor
responsibilities) *"She is covering for our secretary who is ill*
- S: (v) cover (be sufficient to meet, defray, or offset the charge
to cover the check?"
- S: (v) cover (spread over a surface to conceal or protect) *"T*
- S: (v) shroud, enshroud, hide, cover (cover as if with a shrou
civilization are shrouded in mystery"
- S: (v) breed, cover (copulate with a female, used especially
covers the mare"
- S: (v) overlay, cover (put something on top of something else
of gravy"
- S: (v) cover (play a higher card than the one previously playe
- S: (v) cover (be responsible for guarding an opponent in a g
- S: (v) brood, hatch, cover, incubate (sit on (eggs)) *"Birds bro*
the eggs"
- S: (v) cover, wrap up (clothe, as if for protection from the ele

Outline

✓ Knowledge for Machines

✓ Construction of Knowledge Bases

✓ KB Population from Text & Web Pages

✓ Knowledge for Intelligent Applications

★ Opportunities & Challenges

★ Conclusions

Discovering “Unknown” Knowledge

so far KB has relations with type signatures
<entity1, relation, entity2>

< CarlaBruni marriedTo NicolasSarkozy> ∈ Person × R × Person
< NataliePortman wonAward AcademyAward > ∈ Person × R × Prize

Open and Dynamic Knowledge Harvesting:
would like to discover new entities and new relation types
<name1, phrase, name2>

Madame Bruni in her happy marriage with Nicolas....
The first lady is in passionate love with the French president ....
Natalie was honored by the Oscar ...
Bonham Carter was disappointed that her nomination for the Oscar ...

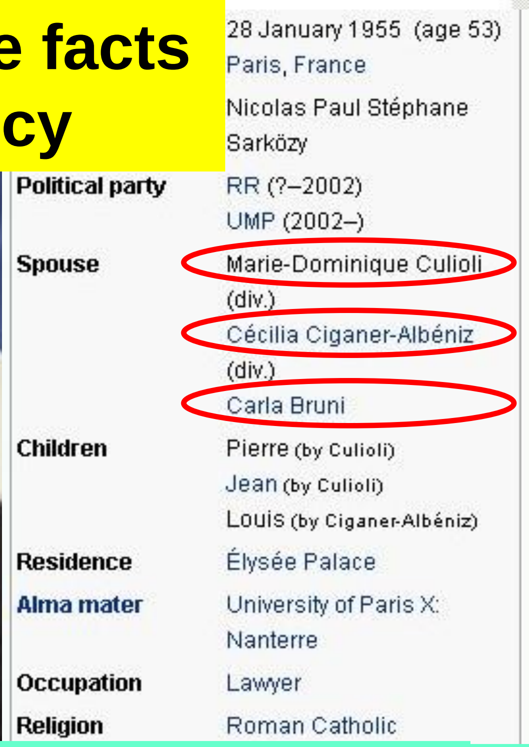
Temporal Knowledge

for all people in Wikipedia (300 000) gather all spouses, incl. divorced & widowed, and corresponding time periods!

>95% accuracy, >95% coverage, in one night

1) **recall: gather temporal scopes for base facts**
2) **precision: reason on mutual consistency**

1. Catherine of Aragon
Divorced
2. Anne Boleyn
Beheaded
3. Jane Seymour
Died



	28 January 1955 (age 53) Paris, France
	Nicolas Paul Stéphane Sarközy
Political party	RR (?–2002) UMP (2002–)
Spouse	Marie-Dominique Culioli (div.) Cécilia Ciganer-Albéniz (div.) Carla Bruni
Children	Pierre (by Culioli) Jean (by Culioli) Louis (by Ciganer-Albéniz)
Residence	Élysée Palace
Alma mater	University of Paris X: Nanterre
Occupation	Lawyer
Religion	Roman Catholic

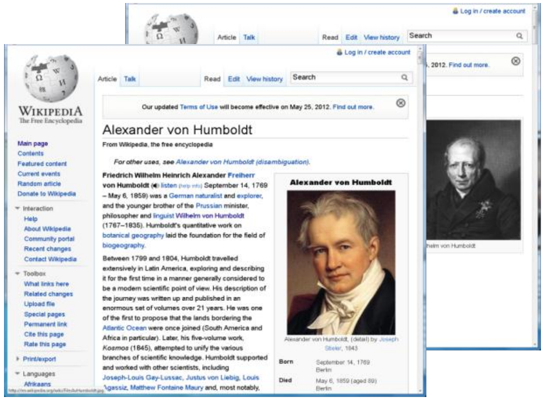
consistency constraints are potentially helpful:

- **functional dependencies:** ***husband, time → wife***
- **inclusion dependencies:** ***marriedPerson ⊆ adultPerson***
- **age/time/gender restrictions:** ***birthdate + Δ < marriage < divorce***

Outline

- ✓ Knowledge for Machines
- ✓ Construction of Knowledge Bases
- ✓ KB Population from Text & Web Pages
- ✓ Knowledge for Intelligent Applications
- ✓ Opportunities & Challenges
- ★ Conclusions

Summary



Article | Talk | Read | Edit | View history | Search

Log in / create account

Our updated Terms of Use will become effective on May 25, 2012. Find out more.

WIKIPEDIA
The Free Encyclopedia

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

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
What links here
Related changes
Upload file
Special pages
Permanent link
Cite this page
Rate this page

Print/export

Languages
Afrikaans

Alexander von Humboldt

From Wikipedia, the free encyclopedia

For other uses, see Alexander von Humboldt (disambiguation).

Friedrich Wilhelm Heinrich Alexander Freiherr von Humboldt (listen (help·info) September 14, 1769 – May 6, 1859) was a German naturalist and explorer, and the younger brother of the Prussian minister, philosopher and linguist Wilhelm von Humboldt (1767–1835). Humboldt's quantitative work on botanical geography laid the foundation for the field of biogeography.

Between 1799 and 1804, Humboldt travelled extensively in Latin America, exploring and describing it for the first time in a manner generally considered to be a modern scientific point of view. His description of the journey was written up and published in an enormous set of volumes over 21 years. He was one of the first to propose that the lands bordering the Atlantic Ocean were once joined (South America and Africa in particular). Later, his five-volume work, *Kosmos* (1845), attempted to unify the various branches of scientific knowledge. Humboldt supported and worked with other scientists, including Joseph-Louis Gay-Lussac, Justus von Liebig, Louis Agassiz, Matthew Fontaine Maury and, most notably,

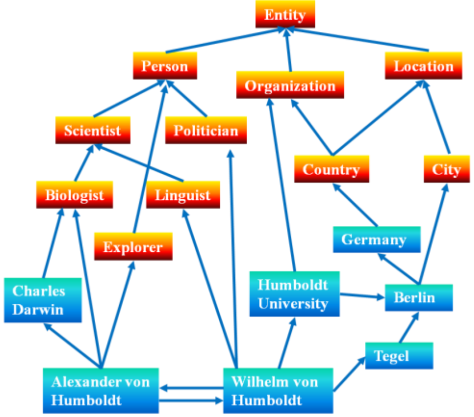
Alexander von Humboldt

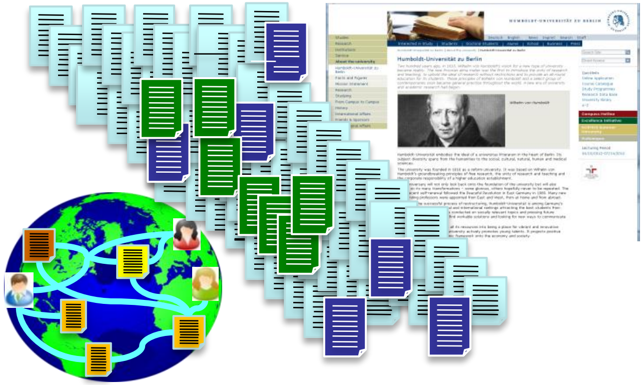
Alexander von Humboldt, (detail) by Joseph Stieler, 1843

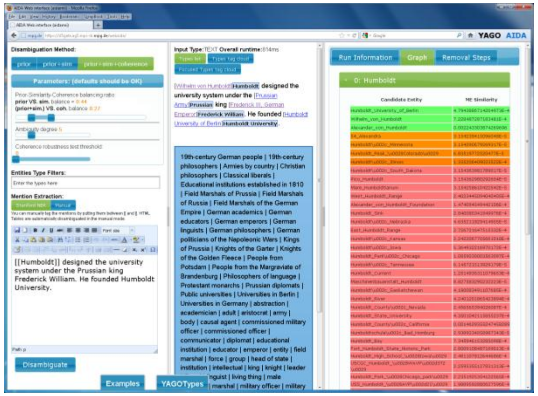
Born	September 14, 1769 Berlin
Died	May 6, 1859 (aged 89) Berlin

, 2012. Find out more.

Wilhelm von Humboldt







AIDA Web interface (aidano) - Mozilla Firefox

Disambiguation Method:

prior | prior+sim | prior+sim+coherence

Parameters: (defaults should be OK)

Prior-Similarity-Coherence balancing ratio
prior VS. sim. balance = 0.44
(prior+sim.) VS. coh. balance 0.27

Ambiguity degree 5

Coherence robustness test threshold

Entities Type Filters:

Enter the types here

Mention Extraction:

Stanford NER | Manual

You can manually tag the mentions by putting them between [[and]]. HTML Tables are automatically disambiguated in the manual mode.

[[Humboldt]] designed the university system under the Prussian king Frederick William. He founded Humboldt University.

Path: p

Disambiguate

Input Type: TEXT **Overall runtime:** 814ms

Types list | Types tag cloud | Focused Types tag cloud

[Wilhelm von Humboldt]**Humboldt** designed the university system under the [Prussian Army]**Prussian** king [Frederick III, German Emperor]**Frederick William**. He founded [Humboldt University of Berlin]**Humboldt University**.

19th-century German people | 19th-century philosophers | Armies by country | Christian philosophers | Classical liberals | Educational institutions established in 1810 | Field Marshals of Prussia | Field Marshals of Russia | Field Marshals of the German Empire | German academics | German educators | German emperors | German linguists | German philosophers | German politicians of the Napoleonic Wars | Kings of Prussia | Knights of the Garter | Knights of the Golden Fleece | People from Potsdam | People from the Margraviate of Brandenburg | Philosophers of language | Protestant monarchs | Prussian diplomats | Public universities | Universities in Berlin | Universities in Germany | abstraction | academician | adult | aristocrat | army | body | causal agent | commissioned military officer | commissioned officer | communicator | diplomat | educational institution | educator | emperor | entity | field marshal | force | group | head of state | institution | intellectual | king | knight | leader | ... nguist | living thing | male ... marshal | military officer | military

Examples | YAGOTypes

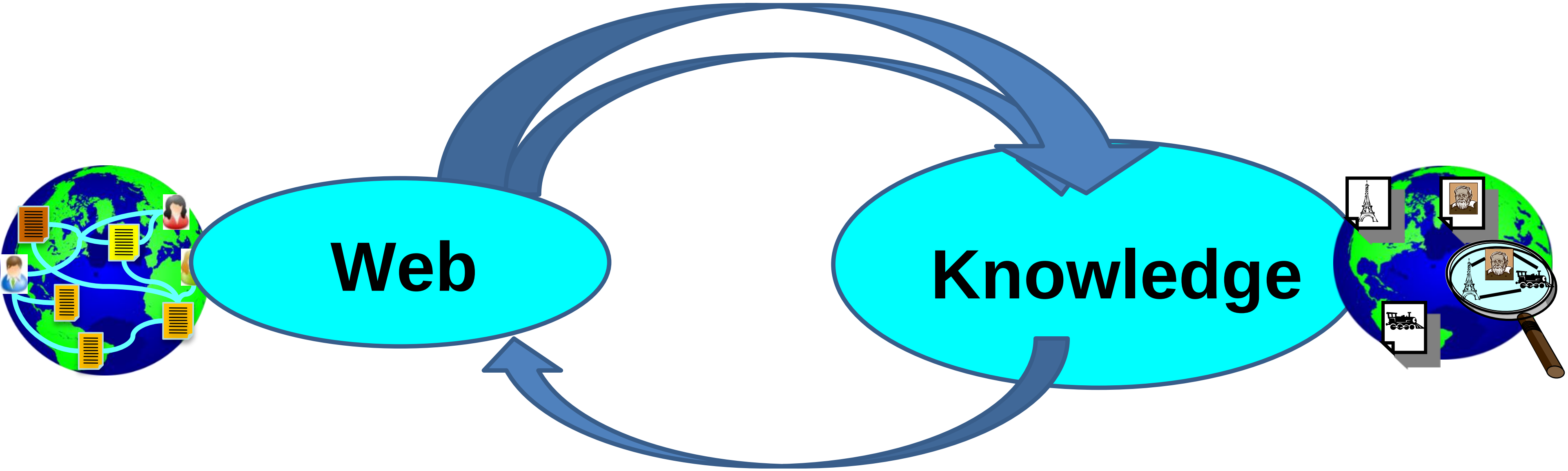
YAGO AIDA

Run Information | Graph | Removal Steps

0: Humboldt

Candidate Entity	ME Similarity
Humboldt_University_of_Berlin	4.794388671426467E-4
Wilhelm_von_Humboldt	7.229487287183481E-4
Alexander_von_Humboldt	0.002243303874269696
54_Alexandra	3.154239419096048E-5
Humboldt\u002c_Minnesota	3.154990679269317E-5
Humboldt_Peak_\u0028Colorado\u0029	6.835197105204773E-5
Humboldt\u002c_Illinois	1.335356408323522E-4
Humboldt\u002c_South_Dakota	3.154363981789981E-5
Pico_Humboldt	3.154362960292694E-5
Mare_Humboldtianum	3.154258610922549E-5
West_Humboldt_Range	2.403344029404040E-4
Alexander_von_Humboldt_Foundation	1.474094049442106E-4
Humboldt_Sink	2.840380341849978E-4
Humboldt\u002c_Nebraska	4.635221182941495E-5
East_Humboldt_Range	2.795721647513332E-4
Humboldt\u002c_Kansas	2.242385770568101E-4
Humboldt\u002c_Iowa	5.364935216970172E-4
Humboldt_Park\u002c_Chicago	1.080903000315620E-4
Humboldt\u002c_Tennessee	6.145721513829179E-5
Humboldt_Current	1.281493531107965E-4
Maschinenbauanstalt_Humboldt	8.827883290232223E-5
Humboldt\u002c_Saskatchewan	4.190082491107588E-4
Humboldt_River	4.240120106542389E-4
Humboldt_County\u002c_Nevada	2.498565384026087E-4
Humboldt_State_University	4.393104211385532E-4
Humboldt_County\u002c_California	0.0014629932247450
Humboldtschule\u002c_Bad_Homburg	2.930993405898734E-5
Humboldt_Bay	7.349946153391598E-4
Fort_Humboldt_State_Historic_Park	2.800310940716801E-4
Humboldt_High_School_\u0028Iowa\u0029	2.481107812844686E-4
USCGC_Humboldt_\u0028WAVP\u002d372\u0029	2.259335512793131E-4
Humboldt_Park_\u0028Chicago_park\u0029	2.235195535413356E-4
USS_Humboldt_\u0028AVP\u002d21\u0029	1.999956898933759E-4

Take-Home Message



- **Machine Knowledge from Web Sources:**
  finally real and big!
- **Enabling Asset for Intelligent Applications:**
  digital humanities, big-data analytics,
  deep machine reading, QA & HCI, etc.
- **Rich Research Opportunities & Challenges:**
  scale & robustness, temporal, multimodal,
  open & real-time knowledge discovery, etc.

Merci Beaucoup!