

Knowledge Graphs

Evangelos Kanoulas

e.kanoulas@uva.nl

Semantic Search



Semantic Search

- Improve search accuracy by understanding user intent
- Improve relevance assessment by document understanding, including context
- Move towards providing answers (and not just result links)

About 180,000,000 results (0.28 seconds)

Athens - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Athens - Wikipedia

Athens is the capital and largest city of Greece. **Athens** dominates the Attica region and is one of the world's oldest cities, with its recorded history spanning ...

[History of Athens](#) - [Acropolis of Athens](#) - [Classical Athens](#) - [Athens \(disambiguation\)](#)

OpenAthens: Eduserv

www.openathens.net/

Related links. OpenAthens resources All OpenAthens enabled resources. More about Identity and Access Management Eduserv provides a range of products ...

[OpenAthens - Login](#) - [OpenAthens Documentation](#) - [Service Status](#) - [Support](#)

Greece-Athens.com - The Guide To Athens city, Greece

www.greece-athens.com/

The essential guide to the capital city of **Athens**, Greece. Our site offers a virtual tour in **athens** city & tourist information such as accommodation, restaraunts, ...

Visit Greece | Athens

www.visitgreece.gr/en/main.../athe... - Greek National Tourism Organization

Athens is the historical capital of Europe, with a long history, dating from the first settlement in the Neolithic age. In the 5th Century BC (the "Golden Age of ...

Images for Athens

Report images

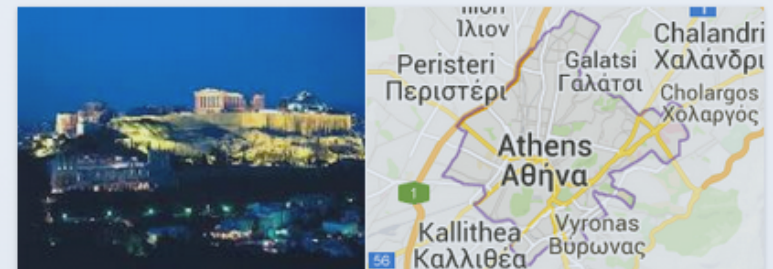


More images for Athens

Athens Georgia travel, vacation, events, hotels, music ...

www.visitathensga.com/

Take a hip trip to **Athens**, Georgia! A vibrant, restored Victorian-era downtown teems with art galleries, trendy shops, world-renowned nightclubs and distinctive ...



Athens

Capital of Greece

Athens is the capital and largest city of Greece. Athens dominates the Attica region and is one of the world's oldest cities, with its recorded history spanning around 3,400 years and the earliest human presence around 11th–7th millennium BC.,. [Wikipedia](#)

Area: 15.04 sq miles (38.96 km²)

Weather: 55°F (13°C), Wind 0 mph (0 km/h), 84% Humidity

Local time: Saturday 4:10 PM

Population: 789,166 (2001) UNdata

Points of interest

View 15+ more



Acropolis of Athens



Parthenon



Acropolis Museum



National Archaeol... Museum...



Plaka

Colleges and Universities

View 15+ more



National and...



National Technical



Panteion University



Athens University of



Technolo... Educational



barack obama

Search

Sign In Mail Settings

Web

Images

Video

News

Local

Answers

Shopping

More

Anytime

Past day

Past week

Past month

Barack Obama News



Obama, Republicans rejoin fight over closing Guantanamo

AFP via Yahoo! News 3 hours ago

US President Barack Obama this week renewed his pledge to shut down the Guantanamo Bay military prison, despite strong objections from Republicans who fear inmates could join jihadist battles...

Obama begins landmark visit to India

AFP via Yahoo! News 2 hours ago

Obama cuts short India visit to meet new Saudi king

AFP via Yahoo! News 12 hours ago

More Barack Obama Headlines

Barack Obama - Image Results



More Barack Obama images

Barack Obama - Official Site

www.barackobama.com Cached

Let's give this everything we've got. President Obama is fighting hard to make change in his final two years in office. It's never been more important to take a stand ...

Barack Obama - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Barack_Obama Cached

Barack Hussein Obama II (j / b e ' r a : k h u : ' s e i n e ' b a : m e /; born August 4, 1961) is



Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. wikipedia.org

Born: August 4, 1961 (age 53), Honolulu, Hawaii, USA

Nationality: United States of America

Height: 6' 1" (1.85m)

Spouse: Michelle Obama (m. 1992-present)

Parents: Barack Obama, Sr., Ann Dunham

Children: Natasha Obama, Malia Ann Obama

Feedback

PART 1: KNOWLEDGE GRAPHS

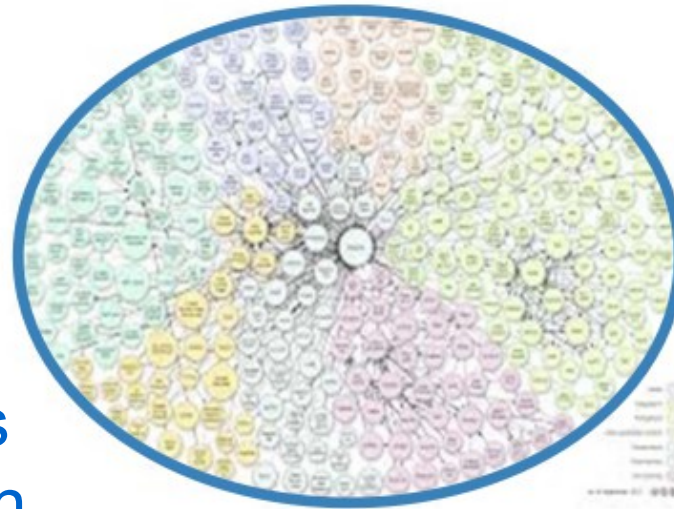
Why (knowledge) graphs?

- We're surrounded by **entities**, which are connected by **relations**
 - We need to store them somehow, e.g., using a **DB** or a **graph**
 - **Graphs** can be processed **efficiently** and offer a convenient **abstraction**
-

Knowledge graphs



Facebook's
Entity Graph



Microsoft's
Satori



OpenIE
(Reverb, OLLIE)

Google's
Knowledge Graph

A sampler of research problems

- **Growth: knowledge graphs are incomplete!**
 - *Link prediction*: add relations
 - *Ontology matching*: connect graphs
 - *Knowledge extraction*: extract new entities and relations from web/text
- **Validation: knowledge graphs are not always correct!**
 - *Entity resolution*: merge duplicate entities, split wrongly merged ones
 - *Error detection*: remove false assertions
- **Interface: how to make it easier to access knowledge?**
 - *Semantic parsing*: interpret the meaning of queries
 - *Question answering*: compute answers using the knowledge graph
- **Intelligence: can AI emerge from knowledge graphs?**
 - *Automatic reasoning* and planning
 - Generalization and abstraction

A SAMPLER OF APPLICATIONS OF KNOWLEDGE GRAPHS

Surfacing structured results in web search

The screenshot shows a Google search for "New York". The search bar at the top contains "New York" and the search button is a magnifying glass icon. Below the search bar, there are navigation tabs for "Web", "Images", "News", "Maps", "Videos", and "More". The search results are displayed in two columns. The left column contains several search results, including Wikipedia entries for "New York" and "New York City", and links to "The Official New York City Guide" and "The New York Times". The right column features a structured knowledge panel for "New York". This panel includes a map of New York state, a list of nearby cities (Toronto, Vermont, Massachusetts, Connecticut, Rhode Island, Pennsylvania, Philadelphia, Ohio), and a detailed description of the state. Below the description, there are sections for "Destinations" (New York City, Buffalo, Long Island, Albany, Finger Lakes) and "Points of interest" (Statue of Liberty, Niagara Falls, Adirondack Mountains, Empire State Building, Metropolitan Museum of Art). A green speech bubble points to the structured results on the right, containing the text "Augmenting the presentation with relevant facts".

Google New York

Web Images News Maps Videos More Search tools

About 716,000,000 results (0.95 seconds)

New York - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/New_York - Wikipedia -
New York is a state in the Northeastern and Mid-Atlantic regions of the United States. New York is the 27th-most extensive, the third-most populous, and the ...
New York City - Albany - List of cities in New York - New York metropolitan area


New York City - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/New_York_City - Wikipedia -
For other uses, see NYC (disambiguation) and New York, New ...
Neighborhoods - History of New York City - Nicknames - Borough

The Official New York City Guide to NYC Attractions, Dining ...
www.nycgo.com/ - New York City -
Visit NYCgo for official NYC information on travel, hotels, deals and offers like Restaurant Week, and the best restaurants, shops, clubs and cultural events.
Must-See NYC - Broadway Shows & Tickets - Events - Tours and Attractions

The New York Times - Breaking News, World News ...
www.nytimes.com/ - The New York Times -
There were no reports of survivors on a Malaysia Airlines flight that crashed on Thursday in eastern Ukraine near the Russian border, the scene of fighting ...
Natalie Glance and one other person +1'd this

New York Magazine -- NYC Guide to Restaurants, Fashion ...
nymag.com/ - New York Magazine -
Daily coverage of New York's restaurants, nightlife, shopping, fashion, politics, and culture. NYMag.com is the online counterpart to New York Magazine.

News for new york

 **New York, Responding to Surge of Child Migrants, Forms ...**
New York Times - by Kirk Semple - 1 hour ago
Opposition to sheltering a wave of young migrants has mounted in many communities across the country, but in New York City, the reaction has ...

More news for new york






NewYork.com - Your Official Site for Travelling To and Living ...
www.newyork.com/ -

New York
US State






New York is a state in the Northeastern and Mid-Atlantic regions of the United States. New York is the 27th-most extensive, the third-most populous, and the seventh-most densely populated of the 50 United States. Wikipedia

Capital: Albany
Secretary of State: Cesar A. Perales
Minimum wage: 8.00 USD per hour (December 31, 2013)
Governor: Andrew Cuomo
Colleges and Universities: Cornell University, More

Destinations View 45+ more

 New York City
 Buffalo
 Long Island
 Albany
 Finger Lakes

Points of interest View 40+ more

 Statue of Liberty
 Niagara Falls
 Adirondack Mountains
 Empire State Building
 Metropolitan Museum of Art

Feedback

Augmenting the presentation with relevant facts

Surfacing facts proactively

The image displays two screenshots of Google search results. The top screenshot shows a search for "san francisco population" with a red circle around the search term. The results include a line graph showing population trends from 1990 to 2011, with a peak of 812,826 in July 2011, and a map of San Francisco. The bottom screenshot shows a search for "san francisco" with a red circle around the search term. The results include a list of links, a map of San Francisco, and a summary of facts about the city. A yellow arrow points from the search term in the top screenshot to the population fact in the bottom screenshot. A red arrow points from the search term in the bottom screenshot to the population fact in the bottom screenshot.

San Francisco Population

Year	Population
1990	~750,000
2000	~770,000
2010	~800,000
Jul 2011	812,826

San Francisco Facts

- Area:** 231.9 sq miles (600.6 km²)
- Founded:** June 29, 1776
- Weather:** 59°F (15°C), Wind NE at 4 mph (6 km/h), 46% Humidity
- Local time:** Sunday 3:55 PM PT
- Population:** 812,826 (2011)

Exploratory search



The screenshot shows a Google search for "new york sightseeing". The search results include a carousel of sightseeing options near New York, NY, and a list of search results. The carousel features items like "Gray Line New York", "City Sightseeing New York", "Circle Line Sightseeing Cruises", "CitySights NY", "NYC & Co", "The New York Pass", "Rockefeller Center", "The Metropolitan Museum of Art", and "Empire State Building". The search results list include:

- Gray Line New York Sightseeing Tours, Cruises & Attractions** - www.newyorksightseeing.com - Gray Line New York - New York's famous Empire State Building, a New York City and a National Historic ... to Top of the Rock AND 1-hour Statue of Liberty New York Harbor Cruise! Nyc double decker tours - Loops Tour Map - All Loops Tour Plus - Contact Us
- New York attractions: The 50 best sights and attractions in ...** - www.timeout.com/newyork/attractions - new-york-attractions - Time Out - by Amy Platt - Apr 25, 2013 - Sights like the Empire State Building and the Statue of Liberty are perennial favorites, but we've also highlighted newcomers and lesser-known ...
- BLDG 92 - Free attractions in New York - Brooklyn Flea - Chrysler Building**
- New York City Tours and Attractions - NYC Sightseeing ...** - www.nycgo.com/toursandattractions/ - New York City - ITINERARIES. Novel New York. by Jessica Allen. NYC's novelistic life comes alive with self-guided tours to landmarks from books like The Catcher in the Rye. ...
- New York: Sightseeing in NYC - TripAdvisor** - www.tripadvisor.com - New York (NY) - Before You Go - TripAdvisor - Inside New York: Sightseeing in NYC - Before you visit New York, visit ... the main sights of the New York City harbor, including the Statue of Liberty, Ellis Island, ...
- NYC Sightseeing Tour | New York City Double Decker Tou...** - skylinesightseeing.com - See NYC's landmarks your way with our Hop-on, ... The most historic neighborhood in New York City, downtown Manhattan features the Empire State Building. ...
- City Sightseeing New York, Hop On - Hop Off Bus Tours** - www.city-sightseeing.com/tours/united-states-of-...new-york.htm - Choose your own way in which you use your ticket according to your own itinerary. There are three tour routes to choose from, allowing you to explore.
- Top 25 New York City Tours - New York Magazine** - nymag.com/visitorsguide/sightseeing/citytours.htm - New York Magazine - Aug 16, 2013 - Views From the Top The Big Apple Tour See the city streets from a ...

A map titled "Map for new york sightseeing" is also visible, showing the New York City area with various sightseeing locations marked.

Connecting people, places and things

The image shows a Facebook page for Harvard University. At the top, there's a search bar and navigation links for Home and Post. The main content area features a large photo of a Harvard campus building, the Harvard crest, and the text "Harvard University" with "2,703,624 likes - 47,280 talking about this - 430,473 were here". Below this, there are tabs for "College & University", "Photos", "Likes", and "Harvard University". A central "Highlights" section shows a "1 Friend" who likes Harvard University.

Two large blue arrows point from the Harvard University page to two separate panels below. The left panel is titled "People who like Harvard University" and lists several users:

- Paul McDonald**: Engineer at Facebook. Likes Harvard University, iRunFar.com and 182 others. Studied Computer Science at Harvard University '03. 3 mutual friends including Clodagh Chloe Takeuchi and Serkan ...
- Ekaterina Skorobogatova**: Works at Facebook. Likes Harvard University, Loves Company and 3,455 others. Studied Interactive Multimedia at New York University. 2 mutual friends: Amina Beighiti and Alexey Spiridonov.
- Gary Johnson**: Corporate Development at Facebook... Likes Harvard University and 278 others. Studied at Wharton School, University of Pennsylvania '08. 5 mutual friends including Jen Holmstrom and Clodagh Chloe T...
- Greg Marra (马格雷)**: Product Manager at Facebook. Likes Harvard University, Emmy's Spaghetti Shack and 693 others. Studied Electrical and Computer Engineering at Franklin W. Olin ... 1 mutual friend: Ledell Wu.

The right panel is titled "People who visited Harvard University" and lists several users:

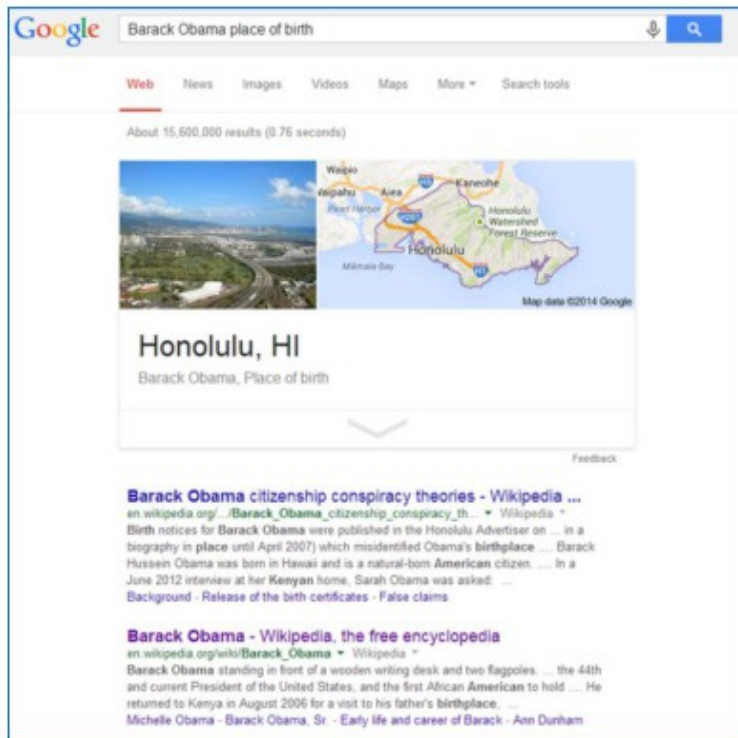
- Florence Trouche**: Global Client Partner at Facebook. Visited Harvard University, Marché Poncelet and 317 other places. Studied at Rouen Business School '90. 35 mutual friends including Michelle Gilbert and Lisa Carucci.
- Andrew Tulloch**: Machine Learning at Facebook. Visited Harvard University, City Beer Store and 90 other places. Studied Machine Learning at University of Cambridge. 21 mutual friends including Jason Weston and Nicolas Vasilache.
- Joseph Barillari (joeb)**: Software Engineer at Facebook. Visited Harvard University, Philz Coffee At Facebook and 990 other places. Studied Computer Science at Harvard University '07. 10 mutual friends including Tudor Bosman and Jessica Traynor.
- Sheryl Sandberg**: Chief Operating Officer at Facebook. Visited Harvard University and 423 other places. Studied at Harvard Business School. 14 mutual friends including Laurent Solly and Catalina Fries Sa...

Connecting people, places and things

Structured search within the graph



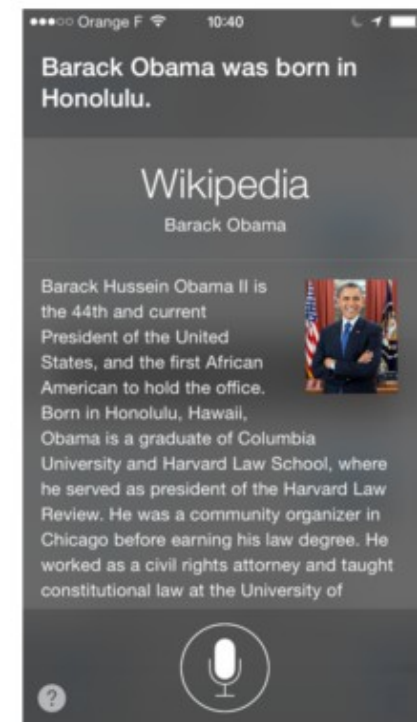
Question answering



Google

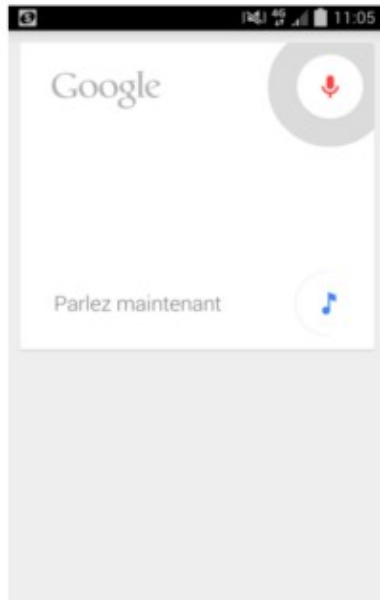


EVI
(Amazon)

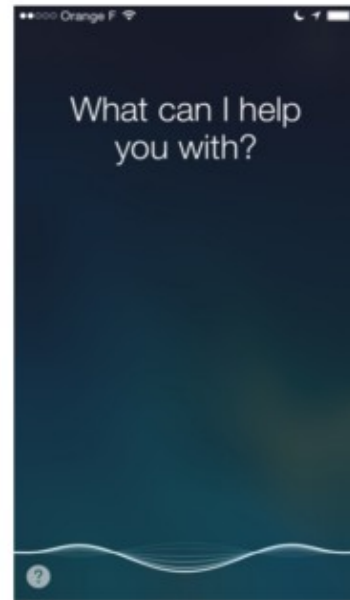


Siri
(Apple)

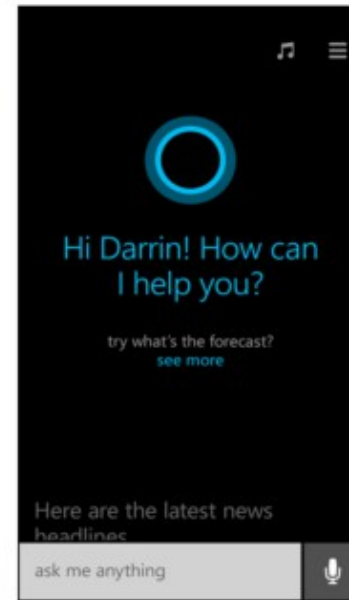
Towards a knowledge-powered digital assistant



OK Google



Siri
(Apple)



Cortana
(Microsoft)

- Natural way of accessing/storing knowledge
- Dialogue system
- Personalization
- Emotion

Interface revolution ◇



**FREEBASE AS AN EXAMPLE
OF A LARGE SCALE
KNOWLEDGE REPOSITORY**

Different approaches to knowledge representation

- Structured (e.g., Freebase or YAGO)
 - Both entities and relations come from a fixed lexicon
 - Semi-structured
 - Predicates come from a fixed lexicon, but entities are strings
 - Unstructured (Open IE)
-



- **Freebase** is an open, [Creative Commons](#) licensed repository of [structured data](#)
- **Typed entities** rather than **strings**

Person Type

Key: /people/person Includes: Topic

A person is a human being (man, woman or child) known to have actually existed. Living persons, celebrities and politicians are persons

Table Diagram


Properties

Property	ID	Expected Type
Date of birth	/people/person/date_of_birth	/type/datetime
Place of birth	/people/person/place_of_birth	/location/location
Country of nationality	/people/person/nationality	/location/country
Gender	/people/person/gender	/people/gender enumerated
Profession	/people/person/profession	/people/profession
Religion	/people/person/religion	/religion/religion
Ethnicity	/people/person/ethnicity	/people/ethnicity
Parents	/people/person/parents	/people/person
Children	/people/person/children	/people/person
Siblings	/people/person/sibling_s	/people/sibling_relationship
Spouse (or domestic partner)	/people/person/spouse_s	/people/marriage mediated
Employment history	/people/person/employment_history	/business/employment_tenu
Education	/people/person/education	/education/education mediated

Relations are typed too!

Barack Obama - Freebase

https://www.freebase.com/m/02mjmr



Topic

Barack Obama^{en}

mid: /m/02mjmr notable type: /government/us_pre

Barack Hussein Obama II is the 44th and current President of the United States. He was born in Honolulu, Hawaii, Obama is a graduate of Columbia University. He was a community organizer in Chicago before running for the U.S. Senate in 2004. He served in the Illinois Senate from 1997 to 2004, running unsuccessfully for the U.S. House of Representatives. He received national attention during his campaign to represent Illinois in the 2004 Democratic Party primary, his keynote address at the Democratic National Convention during his presidential campaign in 2007 and, after a close race, he was elected as the 44th President of the United States in the 2008 United States presidential election. He was re-elected in 2012.

Properties | 118n | Keys

View and edit specific domains, types, or properties..

Filter options: Show all domains and properties

Common /common

Topic /common/topic

Also known as /common/topic/alias

Also known as ▾

- Barack Hussein Obama, Jr.
- Barack Hussein Obama
- Obama
- President Obama
- Barack H. Obama II
- Barack Hussein Obama II
- Barack Obama II
- President Barack Hussein Obama II
- Sen. Barack Obama
- Barak Obama

90 values total ▸

Description /common/topic/description

Barack Obama - Freebase

https://www.freebase.com/m/02mjmr

Evangelos

Person /people/person

Date of birth /people/person/date_of_birth ✓ 2013-10-15

8/4/1961

Place of birth /people/person/place_of_birth ✓ 2014-03-12

Honolulu

Country of nationality /people/person/nationality ✓ 2013-10-15

United States of America

Gender /people/person/gender ✓ 2013-10-15

Male

Profession /people/person/profession

Profession ▾

- Politician
- Lawyer
- Writer
- Author
- Law professor

Religion /people/person/religion

Christianity

Ethnicity /people/person/ethnicity

Ethnicity ▾

- Kenyan American
- English American
- Irish American
- Multiracial American

Parents /people/person/parents ✓ 2013-10-15

Parents ▾

- Barack Obama, Sr.
- Ann Dunham

Children /people/person/children ✓ 2013-10-15

Children ▾

- Natasha Obama

The world changes, but we don't retract facts

We just add more facts!

Marriage Mediator Type

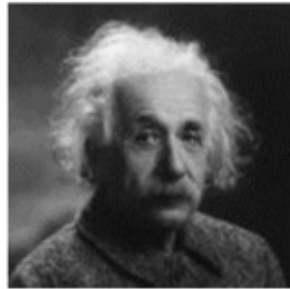
Key: /people/marriage

'Marriage' defines a relationship between two people. The person type uses it to store the two people in the relationship as well as a beginning and end date
[More](#)

Property	ID	Expected Type
Spouse	/people/marriage/spouse	/people/person ↕
From	/people/marriage/from	/type/datetime
To	/people/marriage/to	/type/datetime
Type of union	/people/marriage/type_of_union	/people/marriage_union_type ↕ <i>enumerated</i>
Location of ceremony	/people/marriage/location_of_ceremony	/location/location

Subject-Predicate-Object (SPO) triples

</m/0jcx, /m/04m8, /m/019xz9>



/en/albert_einstein

Albert Einstein

/en/ulm

Ulm

/people/person/place_of_birth

Place of birth

*YAGO2 uses
SPOTL tuples
(SPO + Time
and Location)*

A graph of inter-related objects



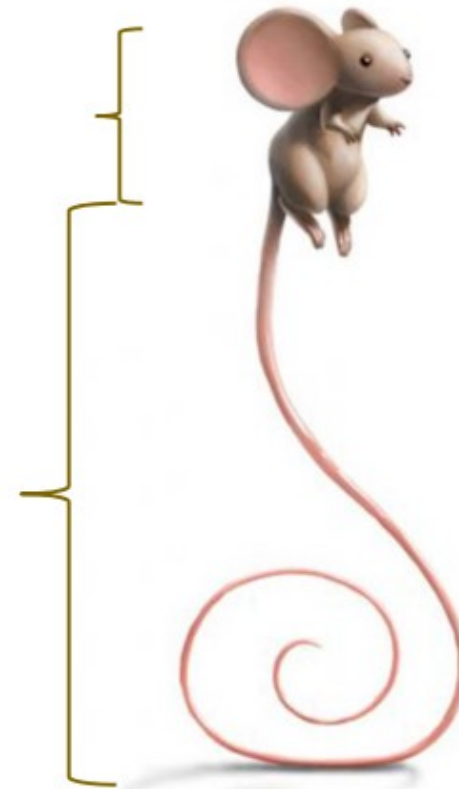
KNOWLEDGE ACQUISITION FROM TEXT

Knowledge acquisition

- Relying on **humans**
 - Volunteer contributions at Freebase.com
 - Import of large datasets (e.g., IMDB)
 - **Head + torso**
- **Automatic** extraction
 - Extraction from web pages
 - **The long tail**
 - Learning patterns using known facts

“... jumped from X into Y ...”

```
</en/tower_bridge,  
/transportation/bridge/body_of_water_spanned,  
/en/river_thames>
```



<http://www.flickr.com/photos/sandreli/4691045841/>

Limits of automatic extraction

- Freebase: **637M** (non-redundant) facts
 - **Knowledge Vault** (automatically extracted):
 - **302M** confident facts with **Prob(true) > 0.9**
 - Of those, 223M are in Freebase (~ **35%**)
-

Relations that are rarely expressed in text

Relation	% entity pairs not found	Notes
/people/person/gender	30%	Pronouns
/people/person/profession	18%	
/people/person/children and /people/person/parents	36%	
/medicine/drug_formulation/ manufactured_forms	99.9%	Sample object: "Biaxin 250 film coated tablet" (/m/0jxc5vb)
/medicine/manufactured_drug _form/available_in	99.4%	Sample subject: "Fluocinolone Acetonide 0.25 cream" (/m/0jxlbx9)
/book/author/works_written and /book/written_work/author	37%	Sample book title: "The birth day: a brief narrative of Eliza Reynolds, who died on Sunday, Oct 19, 1834" (/m/0ydpbtq)

Implicitly stated information



People /people

Person /people/person

Date of birth /people/person/date_of_birth
4004 BCE

Place of birth /people/person/place_of_birth
Garden of Eden

Country of nationality /people/person/nationality
This property has been flagged as having values, but those values are unknown. Remove this flag to add specific values.

Gender /people/person/gender
Male

Profession /people/person/profession
-

Religion /people/person/religion
-

Ethnicity /people/person/ethnicity
-

Parents /people/person/parents
This property has been flagged as having no values. Remove this flag to add new values.

Children /people/person/children
Azura
Seth
Cain
Abel
Awan

Siblings /people/person/sibling_s
Sibling
This property has been flagged as having no values. Remove this flag to add new values.

Spouse (or domestic partner) /people/person/spouse
Spouse - From -
Eve - -

Employment history /people/person/employment_history
Employer -

20 And **Adam** called his wife's name **Eve**; because she was the mother of all living. (*Genesis 3:20*)

Implicitly stated information



People /people	
Person /people/person	
Date of birth /people/person/date_of_birth	4004 BCE
Place of birth /people/person/place_of_birth	Garden of Eden
Country of nationality /people/person/nationality	
Gender /people/person/gender	Male
Profession /people/person/profession	
Religion /people/person/religion	
Ethnicity /people/person/ethnicity	
Parents /people/person/parents	
Children /people/person/children	Azura Seth Cain Abel Awan
Siblings /people/person/siblings	
Spouse (or domestic partner) /people/person/spouse	Eve
Employment history /people/person/employment_history	
Employer /people/person/employer	

(Genesis 1)

1 **In the beginning** God created the heaven and the earth.
 2 And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.
 3 And God said, Let there be light: and there was light.
 ...

(Genesis 2)

7 And the LORD God **formed man** of the dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul.
 8 And the LORD God **planted a garden eastward in Eden; and there he put the man whom he had formed.**
 ...

19 And out of the ground the LORD God formed every beast of the field, and every fowl of the air; and brought them unto
 20 And **Adam** to see what he would call them: and whatsoever **Adam** called every living creature, that was the name thereof.

Knowledge discovery: the long tail of challenges

- Errors in extraction (e.g., parsing errors, overly general patterns)
- Noisy / unreliable / conflicting information
- Disparity of opinion (*Who invented the radio ?*)
- Quantifying completeness of coverage



A screenshot of the Wikipedia article titled "Invention of radio". The page includes a navigation bar with "Article", "Talk", "Read", "Edit", and "View history" options. The main content area features a table of contents with sections such as "1 Wireless signalling methods" and "2 History of the invention of radio". The left sidebar contains various Wikipedia navigation links like "Main page", "Contents", and "Help".

Knowledge discovery: the long tail of challenges

- Errors in extraction (e.g., parsing errors, overly general patterns)
- Noisy / unreliable / conflicting information
- Disparity of opinion (*Who invented the radio ?*)
- Quantifying completeness of coverage

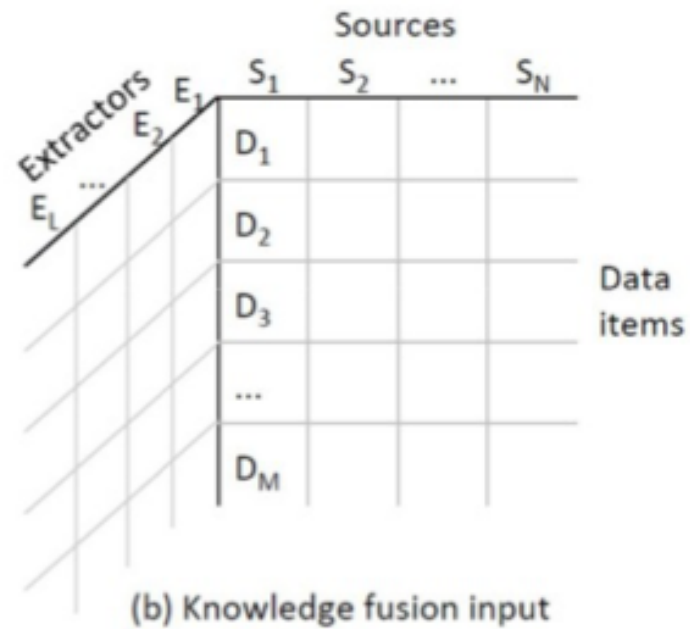


- Fictional contexts

`</en/abraham_lincoln,
/people/person/profession,
/en/vampire_hunter> ?`



Knowledge fusion



[Dong et al., VLDB '14]

Should we trust all sources equally ?

The image shows a screenshot of the Wikipedia article for Barack Obama. The article title is "Barack Obama" and it includes a sub-header "From Wikipedia, the free encyclopedia". The text begins with "Obama" redirects here. For other uses, see Obama (disambiguation). This article is about the 44th president of the United States. For his father, see Barack Obama, Sr. Barack Hussein Obama II (né/ba-rok hu-'seyn ou-ba-mah; born August 4, 1961) is the 44th and current President of the United States, the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney in Chicago and taught constitutional law at the University of Chicago Law School from 1992 to 2004. He served three terms representing the 13th District in the Illinois Senate from 1997 to 2004, running unsuccessfully for the United States House of Representatives in 2000. In 2004, Obama received national attention during his campaign to represent Illinois in the United States Senate with his victory in the March Democratic Party primary, his keynote address at the Democratic National Convention in July, and his election to the Senate in November. He began his presidential campaign in 2007, and in 2008, after a close primary campaign against Hillary Rodham Clinton, he won sufficient delegates in the Democratic Party primaries to receive the presidential nomination. He then defeated Republican nominee John McCain in the general election, and was inaugurated as president on January 20, 2009. Nine months after his election, Obama was named the 2009 Nobel Peace Prize laureate. During his first two years in office, Obama signed into law economic stimulus legislation in response to the Great Recession in the form of the American Recovery and Reinvestment Act of 2009 and the Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010. Other major domestic initiatives in his first term include the Patient Protection and Affordable Care Act, often referred to as "Obamacare", the Dodd-Frank Wall Street Reform and Consumer Protection Act, and the Don't Ask, Don't Tell Repeal Act of 2010. In foreign policy, Obama ended U.S. military involvement in the Iraq War, increased U.S. troop levels in Afghanistan, signed the New START arms control treaty with Russia, ordered U.S. military involvement in Libya, and ordered the military operation that resulted in the death of Osama bin Laden. He later became the first sitting U.S. president to publicly support same-sex marriage. In November 2010, the Republicans regained control of the House of Representatives.

The image shows the header of the website "The Western Center For Journalism". The header includes the organization's name, a tagline "Investigating and Equipping Journalists Who Love Freedom", and a navigation menu with links for Home, Categories, Blogging Tools, About, Pubs and Programs, and Contact Us.

Proof Obama Born in Kenya? Obama Literary Agent Says Yes

Barack.com has introduced some suggestive evidence showing that Obama claimed he was born in Kenya years before he became a presidential candidate. Interestingly, the editors of Barack.com think that new Obama is telling the truth.

The image shows the header of the website "INFOWARS.COM". The header includes the website name, a tagline "BECAUSE THERE IS A WAR ON", and a navigation menu with links for Home, About, News, Radio, Magazine, Blog, and Contact.

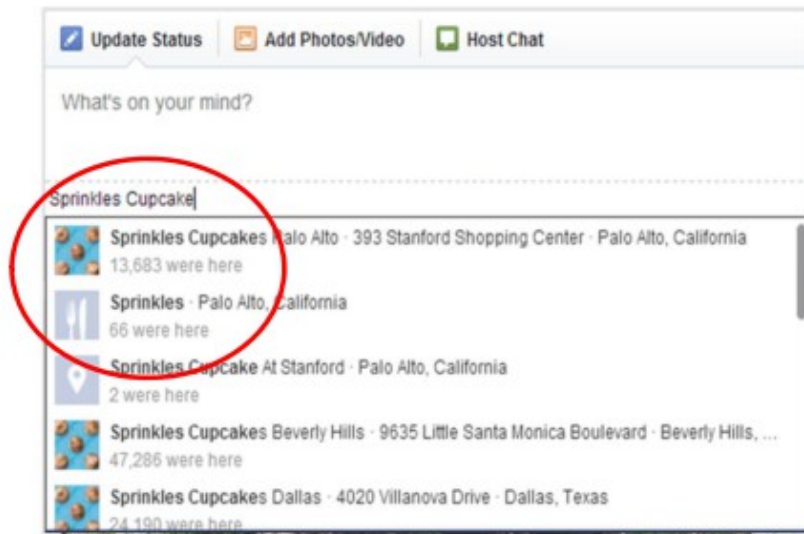
Evidence Obama Born In Kenya Goes Beyond 1991 Brochure

Establishment media pulls must to effort to diffuse further controversy. Paul Joseph Watson, Infowars.com, Friday, May 28, 2010. The establishment media hardly raised an eyebrow over yesterday's explosive story about a literary publication listing Barack Obama's birthplace as Kenya in an effort to claim that the eggs brochure was the "origin" of the entire "birther" issue. In reality, evidence that Obama was born in the African country is abundant. A literary agent's promotional text for a eggs brochure released a correction.

The image shows a screenshot of a Yahoo! News article. The article title is "YAHOO! NEWS SAYS OBAMA WAS BORN IN...KENYA!". The article text includes: "Yahoo! News had to issue a correction Friday after publishing an article about President Barack Obama that called Kenya 'the country of his birth.' The article, about Obama's upcoming trip to Africa, stated: President Barack Obama makes the first extended trip to Africa of his presidency next week—but he won't be stopping at the country of his birth." Below the article text, there is a section titled "White House doesn't have 'figure on costs' of Africa trip" with a sub-header "by Rachel/News Editor, Yahoo! News | The Talk | 1:14:10 mins ago".

Entity resolution / de-duplication

- Multiple mentions of the same entity is wrong and confusing.



Entity resolution / de-duplication

- Multiple mentions of the same entity is wrong and confusing.

The screenshot shows a Yelp search for 'vego' near Montreal, QC, Canada. The results list three restaurants:

- 1. Restaurant Végo** (circled in red): 1720 Saint-Denis Rue, Montreal, QC H2X 3K6, Canada. Phone: (514) 845-2627. 16 reviews. Vegetarian. Price: \$\$.
- 2. Resto Végo McGill**: 1204 Avenue McGill College, Montreal, QC H3B 4J8, Canada. Phone: (514) 871-1480. 5 reviews. Buffets, Vegan, Vegetarian. Price: \$\$.
- 3. Resto Végo St-Denis** (circled in red): 1720 Rue Saint-denis, Montreal, QC H2X 3K6, Canada. Phone: (514) 845-2627. 3 reviews. Vegetarian. Price: \$\$.

The map on the right shows the locations of these three restaurants marked with red pins numbered 1, 2, and 3. Pins 1 and 3 are at the same location (1720 Saint-Denis Rue), while pin 2 is at a different location (1204 Avenue McGill College).

Entity resolution / de-duplication

- Multiple mentions of the same entity is

Find vego Near Montréal, Quebec, Canada

vego Montreal, QC, Canada

Show Filters

- 1. Restaurant Végo**
16 reviews
\$\$ - Vegetarian
1720 Saint-Denis Rue
Montreal, QC H2X 3K6
Canada
[\(514\) 845-2627](tel:5148452627)
- 2. Resto Végo McGill**
5 reviews
\$\$ - Buffets, Vegan, Vegetarian
1204 Avenue McGill College
Montréal, QC H3B 4J8
Canada
[\(514\) 871-1480](tel:5148711480)
- 3. Resto Végo St-Denis**
3 reviews
\$\$ - Vegetarian
1720 Rue Saint-denis
Montreal, QC H2X 3K6
Canada
[\(514\) 845-2627](tel:5148452627)

100 10:32

Mandala Tea House Filter

- 1. Mandala Tea House**
4.3 mi
14 Reviews
14107-H Winchester Blvd, Los Gatos
Gluten-Free, Vegan, Coffee & Tea
\$\$
- 2. Breathe Los Gatos**
4.3 mi
61 Reviews
14107-H Winchester Blvd, Los Gatos
Yoga

Save Search

Can't find what you're looking for?
Add a Business.

MAP

Commonsense knowledge

“Bananas are yellow.”



“Jasmine flowers smell good.”

“Balls bounce.”



- Commonsense information is hard to collect (*too obvious*)
 - Yet commonsense reasoning is often crucial
-

PART 2: METHODS AND **TECHNIQUES**

A sampler of research problems

- **Growth: knowledge graphs are incomplete!**
 - *Link prediction*: add relations
 - *Ontology matching*: connect graphs
 - *Knowledge extraction*: extract new entities and relations from web/text
 - **Validation: knowledge graphs are not always correct!**
 - *Entity resolution*: merge duplicate entities, split wrongly merged ones
 - *Error detection*: remove false assertions
 - **Interface: how to make it easier to access knowledge?**
 - *Semantic parsing*: interpret the meaning of queries
 - *Question answering*: compute answers using the knowledge graph
 - **Intelligence: can AI emerge from knowledge graphs?**
 - *Automatic reasoning* and planning
 - Generalization and abstraction
-

Methods and techniques

1. Relation extraction:

- Supervised models
- Semi-supervised models
- Distant supervision

2. Entity resolution

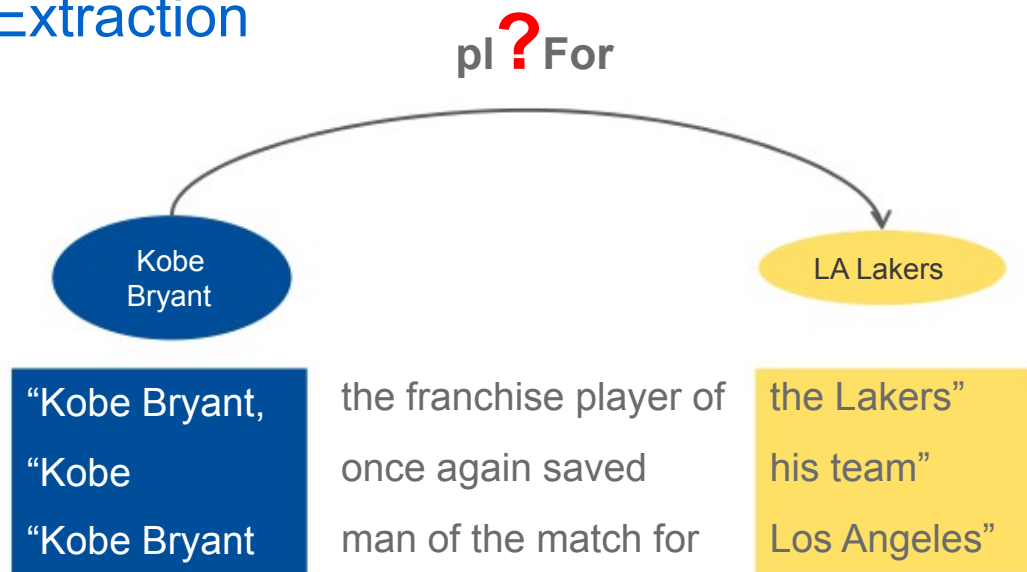
- Single entity methods
- Relational methods

3. Link prediction

- Rule-based methods
- Probabilistic models
- Factorization methods
- [Embedding models](#)

RELATION EXTRACTION

Relation Extraction

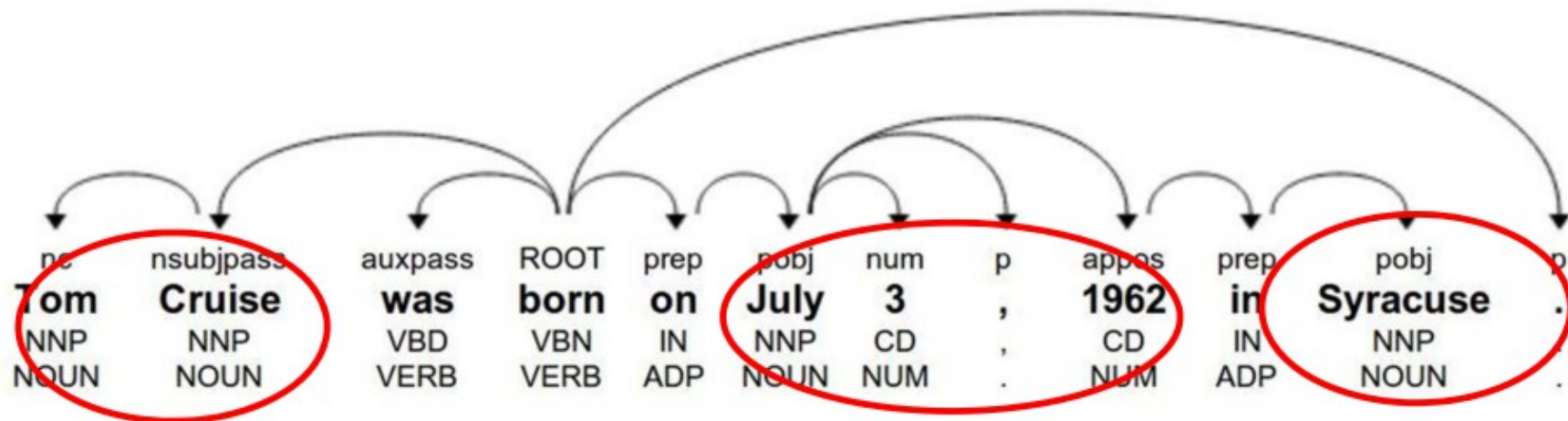


- Extracting semantic relations between sets of [grounded] entities
 - Numerous variants:
 - Undefined vs **pre-determined set of relations**
 - **Binary** vs n-ary relations, facet discovery
 - Extracting temporal information
 - Supervision: {**fully**, un, **semi**, **distant**}-supervision
 - Cues used: only lexical vs **full linguistic features**
-

Supervised relation extraction

- Sentence-level labels of relation mentions
 - "Apple CEO Steve Jobs said.." => (SteveJobs, CEO, Apple)
 - "Steve Jobs said that Apple will.." => NIL
 - Traditional relation extraction datasets
 - ACE 2004
 - MUC-7
 - Biomedical datasets (e.g BioNLP challenges)
 - Learn classifiers from +/- examples
 - Typical features: context words + POS, dependency path between entities, named entity tags, token/parse-path/entity distance
-

Examples of features



X was born on DDDD in Y

- **DEP**: X <nsubjpass / born prep> on pobj> DATE prep> in pobj> Y
- **NER**: X = PER, Y = LOC
- **POS**: X = NOUN, NNP; Y = NOUN, NNP
- **Context**: born, on, in , "born on"

Supervised relation extraction

- Used to be the “traditional” setting [Riloff et al., 06; Soderland et al., 99]
 - **Pros**
 - High quality supervision
 - Explicit negative examples
 - **Cons**
 - **Very expensive to generate supervision**
 - Not easy to add more relations
 - Cannot generalize to text from different domains
-

Semi-supervised relation extraction

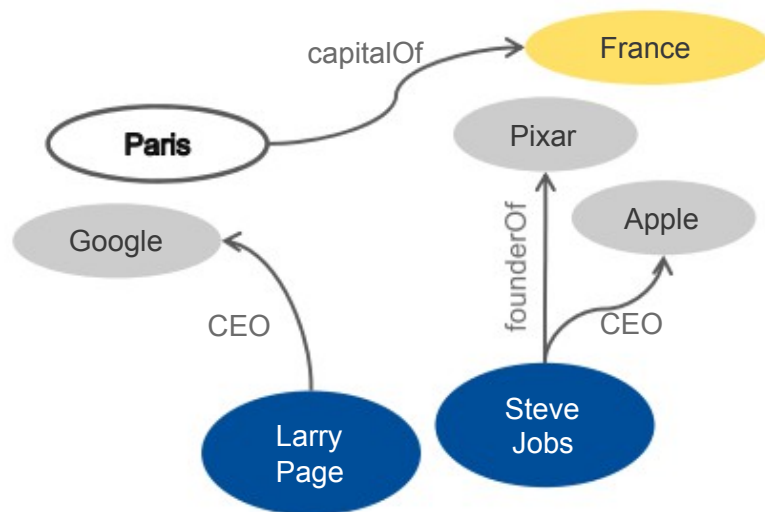
- **Generic algorithm**
 1. Start with **seed triples / golden seed patterns**
 2. **Extract patterns that match** seed triples/patterns
 3. Take the **top-k** extracted patterns/triples
 4. **Add to seed** patterns/triples
 5. Go to 2
 - Many published approaches in this category:
 - Dual Iterative Pattern Relation Extractor [Brin, 98]
 - Snowball [Agichtein & Gravano, 00]
 - **TextRunner [Banko et al., 07] - almost unsupervised**
 - Differ in pattern definition and selection
-

TextRunner [Banko et al., 07]

- **Almost unsupervised**
 - **Relations not fixed**: does not follow Knowledge Graph schema (**growing**)
 - No labeled data
 - Mostly unlabeled text
 - Uses heuristics to **self-label** a starting corpora (using a parser), such as
 - Path length $< k$
 - Path does not cross sentence-like boundaries like relative clauses
 - Neither entity is a pronoun
 - **Self-training**
 - Generate +/- examples \diamond learn classifier
 - **Extract new relation mentions** using this classifier
 - **Generate triples from aggregated mentions**, assign probabilistic score using [Downey et. al., 2005]
 - Later improved in Reverb [Fader et al., 11]
-

Distantly-supervised relation extraction

- Existing knowledge base + unlabeled text \diamond generate examples
 - Locate pairs of related entities in text
 - Hypothesizes that the relation is expressed



Google CEO Larry Page announced that...

Steve Jobs has been Apple for a while...

Pixar lost its co-founder Steve Jobs...

I went to Paris, France for the summer...

Distant supervision

- **Pros**

- Can scale to the web, as no supervision required
- Generalizes to text from different domains
- Generates a lot more supervision in one iteration

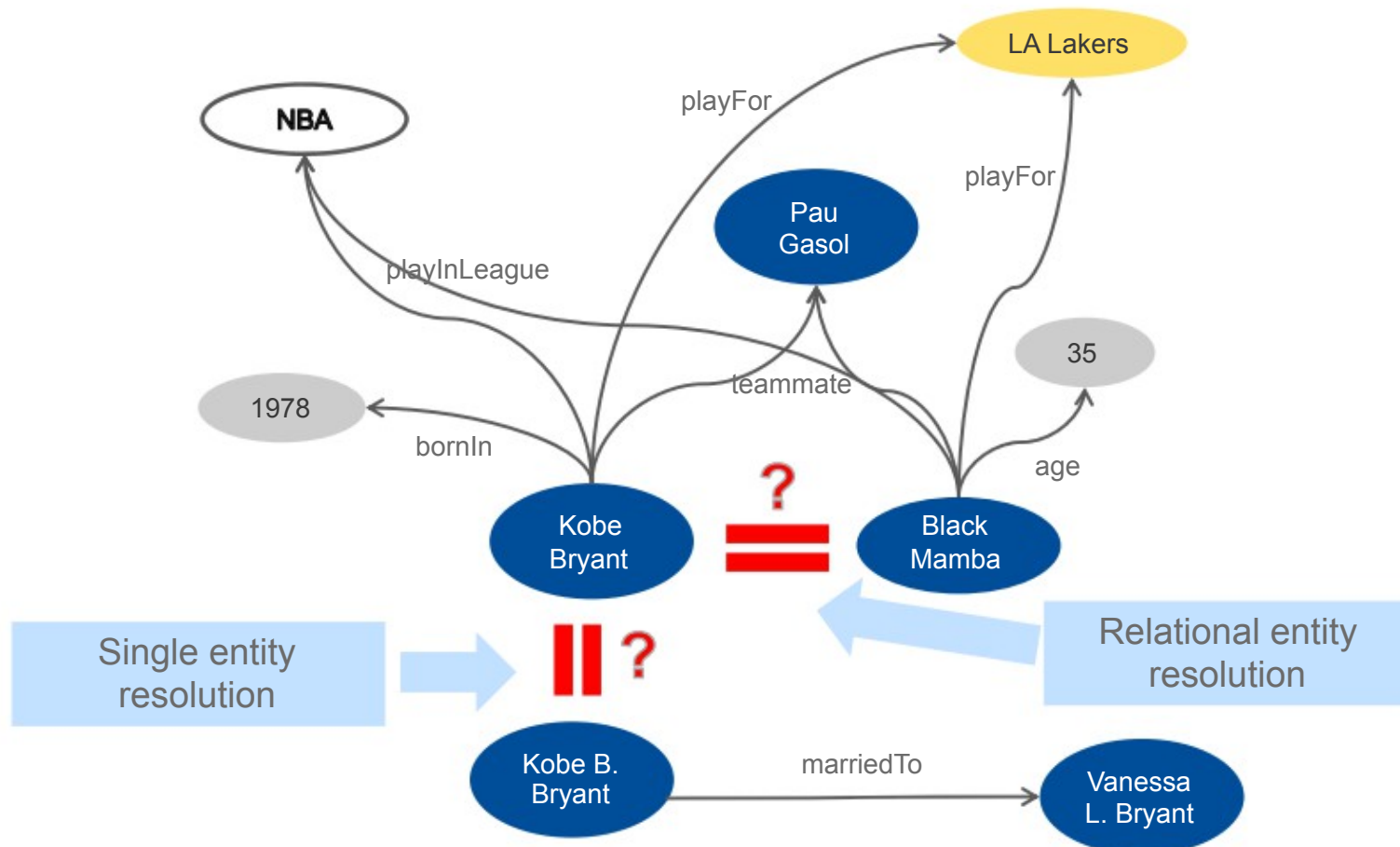
- **Cons**

- Needs high quality entity-matching
 - Relation-expression hypothesis can be wrong
 - ♣ Can be compensated by the extraction model, redundancy, language model
 - Does not generate negative examples
 - ♣ Partially tackled by matching unrelated entities
-

ENTITY RESOLUTION



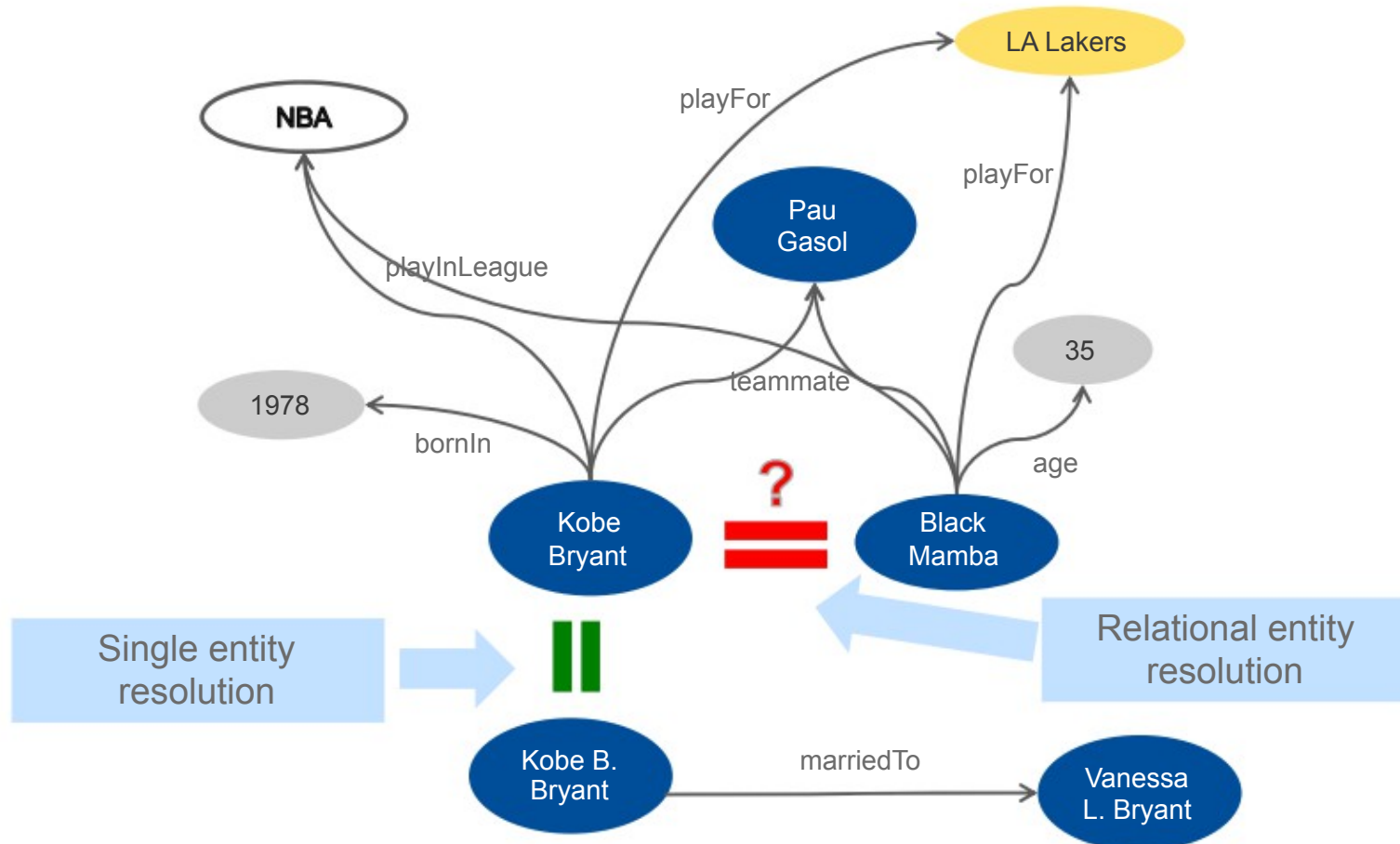
Entity resolution



Single entity resolution

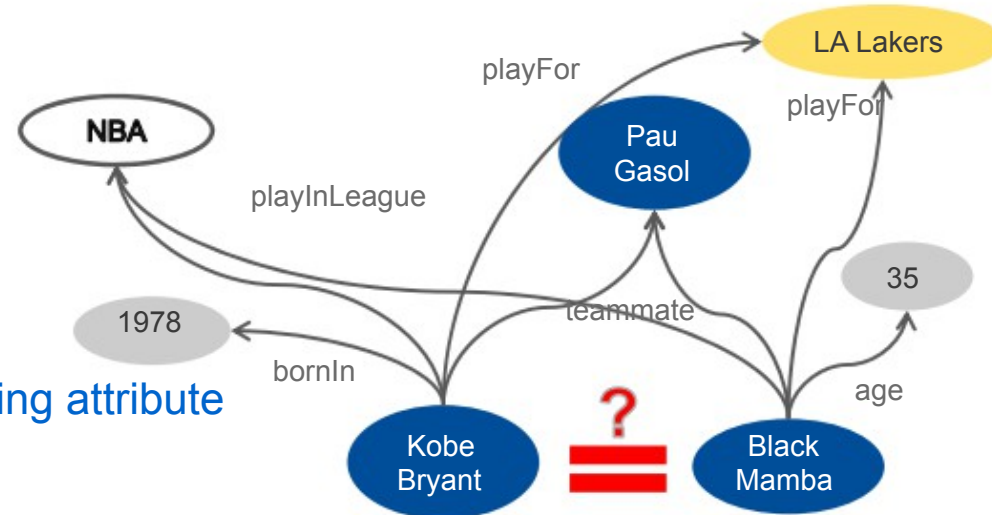
- Entity resolution **without using the relational context** of entities
 - Many **distances/similarities** for single-entity entity resolution:
 - Edit distance (Levenshtein, etc.)
 - Set similarity (TF-IDF, etc.)
 - Alignment-based
 - Numeric distance between values
 - Phonetic Similarity
 - Equality on a boolean predicate
 - Translation-based
 - Domain-specific
-

Entity resolution



Relational entity resolution - Simple strategies

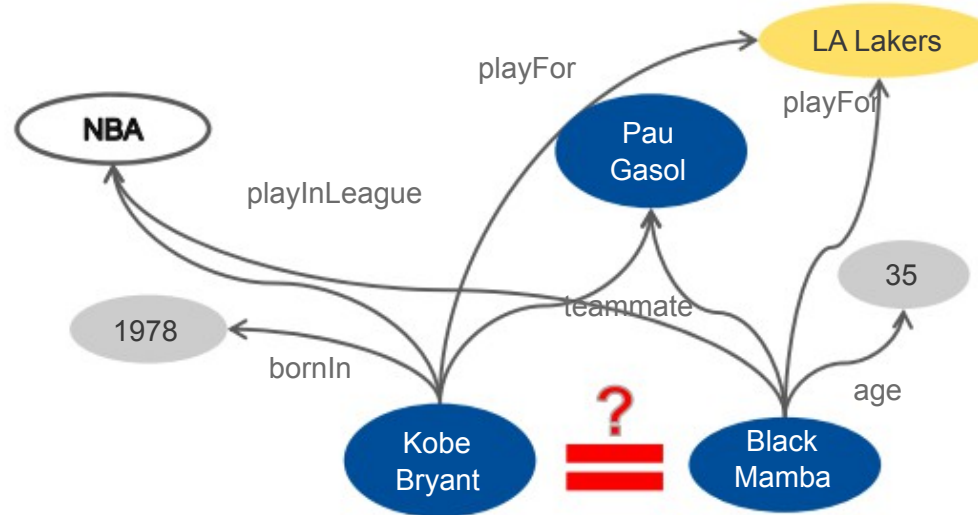
- Enrich model with **relational features** \diamond **richer context** for matching



- **Relational features:**

- Value of **edge or neighboring attribute**
- Set **similarity measures**
 - *Overlap/Jaccard*
 - *Average similarity* between set members
 - *Adamic/Adar*: two entities are more similar if they share more items that are overall less frequent
 - *SimRank*: two entities are similar if they are related to similar objects
 - *Katz score*: two entities are similar if they are connected by shorter paths

Relational entity resolution - Advanced strategies

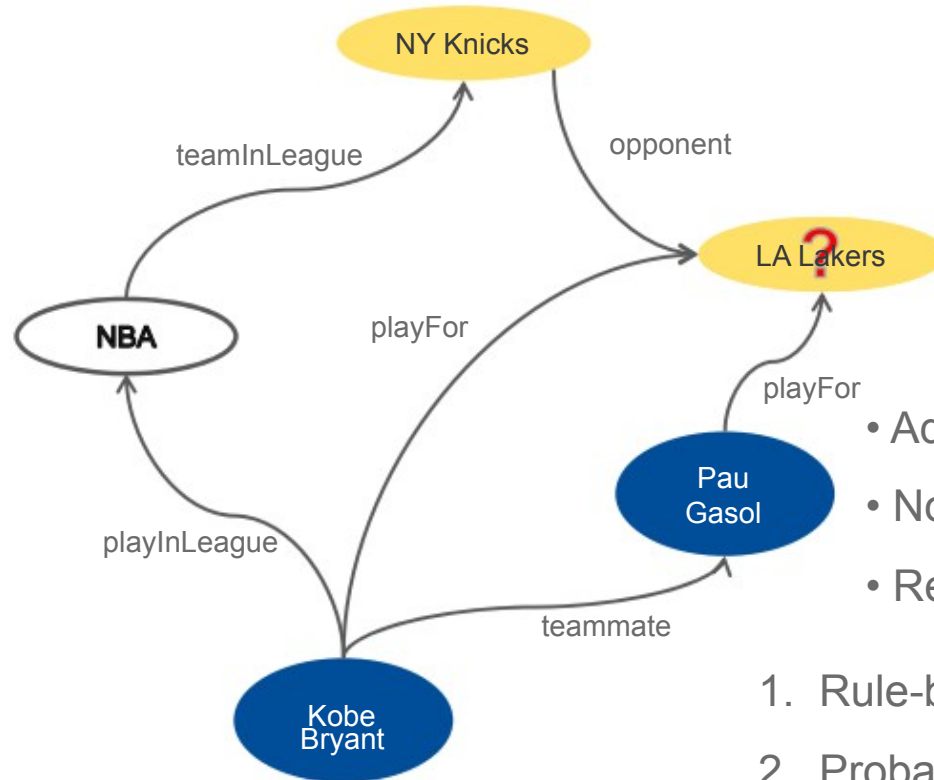


- Dependency graph approaches [Dong et al., 05]
- Relational clustering [Bhattacharya & Getoor, 07]
- **Probabilistic Relational Models** [Pasula et al., 03]
- **Markov Logic Networks** [Singla & Domingos, 06]
- **Probabilistic Soft Logic** [Broecheler & Getoor, 10]

LINK PREDICTION



Link prediction



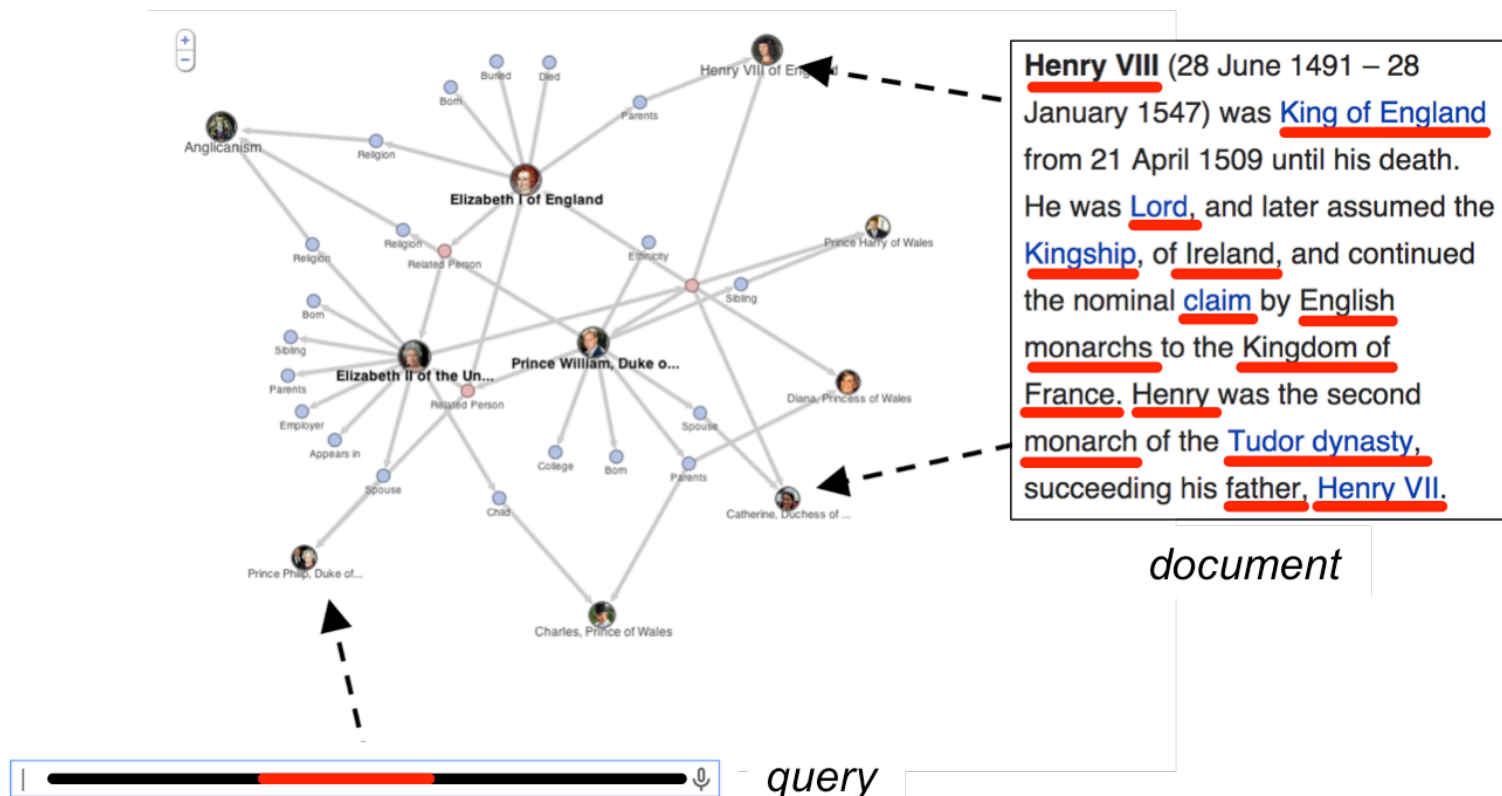
- Add knowledge from existing graph
- No external source
- Reasoning within the graph

1. Rule-based methods
2. Probabilistic models
3. Factorization models
4. **Embedding models**

PART 3: ACCESS KNOWLEDGE
BY ENTITY LINKING

Entity Linking Documents

- Disambiguate every mention (=entity reference) on the Web
- Link token spans in documents to knowledge graph (e.g. Freebase, Wikipedia) entities



Cognitive Computation Group ▶ Demos ▶ Wikifier


Wikifier Demo

fewer concepts more concepts

wikify! x clear

* If you wish to cite this work, please cite the following publications:

The Chiefs **Alex Smith** game manager who would win their offer they needed a **quarterback** who knows how what he's done for most of this season: three necessary and use his **Smith** the chair. These days it **Smith** lost to San Diego -- that he can elevate his



Quarterback of the Kansas City Chief

Cognitive Computation Group ▶ Demos ▶ Wikifier

Wikifier Demo

fewer concepts more concepts

the following publications: (1) Retinov et. al. and (2) Cheng and Roth.

als tight **Alex Smith** for the season. The wrist injury in the third quarter during the regular season Bengals head coach Marvin Lewis described the injury as a "wrist the post game interview on 700 WLW with Dave **Smith** an **Smith** eventuality. More will be on declaring him done. On the other hand, Lewis confirmed the tight end, Tyler Eifert and Jermaine Gresham for the next against the Ravens.




Tight End of the Cincinnati Bengals

San Diego: The San Diego Chargers (A Football team)

Contextual decision on **what is meant** by a given entity or concept. **WSD** with Wikipedia titles as categories.

Ravens: The Baltimore Ravens (A Football team)

Cognitive Computation Group ▶ Demos ▶ Wikifier

 Wikifier Demo

fewer concepts more concepts

* If you wish to cite this work, please cite the following publications: (1) Retinov et. al. and (2) Cheng and Roth.

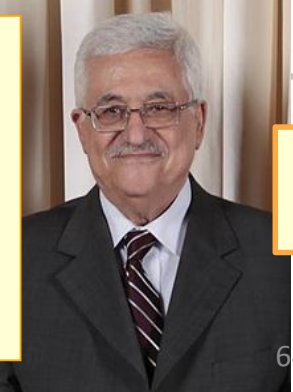
Over and over again I'd hear these **perorations** from **Mahmoud Abbas** that there is no difference between **Fatah** and **Hamas**, or between **Maashal** and **Khaled**. I would cringe at such comments, while knowing full well that he is hardly the perfect **interlocutor**. I'm a strong believer in identifying people without pulling any **punches**. But I believe that it is important to give peace a chance, to search for signs that the **Israelis** are open to change from the destructive and self-destructive path that they have **pursued** for decades. **Hamas** was and is a **useless proposition**. It is based on **extremist religious grounds** and is **anti-Semitic** to the core. It is the "Protocols of the Learned Elders of Zion," blaming the **French Revolution**. Its leadership **denied the Holocaust** and blamed the **financial crisis** on Jewish control.

Mahmoud Abbas

Abu Mazen

Mahmoud Abbas: http://en.wikipedia.org/wiki/Mahmoud_Abbas

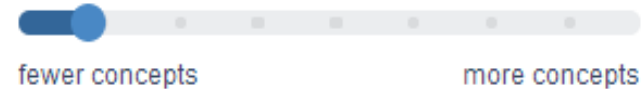
Getting away from **surface representations**.
Co-reference resolution within and across documents, **with grounding**



Abu Mazen: http://en.wikipedia.org/wiki/Mahmoud_Abbas



Wikifier Demo



wikify!

clear

** If you wish to cite this work, please cite the following publications: (1) Retinov et. al. and (2) Cheng and Roth.*

Human immunodeficiency virus (HIV) is the primary etiologic agent responsible for the AIDS pandemic. We constructed a fusion of the gp41 membrane-proximal external region (MPER) peptide along with a variable-length (Gly4Ser)x linker (where x is 4 or 8) between the C terminus of the former and N terminus of the latter. The His-tagged recombinant proteins, expressed in BL21(DE3)pLysS cells and purified by immobilized metal affinity chromatography followed by gel filtration, were found to display a nanomolar efficacy in blocking BaL-pseudotyped HIV-1 infection of HOS.T4.R5 cells. This antiviral activity was HIV-1 specific, since it did not inhibit cell infection by vesicular stomatitis virus (VSV). The chimeric proteins were found to release intraviral p24 protein from both BaL-pseudotyped HIV-1 and fully infectious BaL HIV-1 in a dose-dependent manner in the absence of host cells. The addition of either MPER or CVN was found to outcompete this virolytic effect, indicating that both components of the chimera are required for virolysis. The finding that engaging the Env protein spike and membrane using a chimeric ligand can destabilize the virus and lead to inactivation opens up a means to investigate virus particle metastability and to evaluate this approach for inactivation at the earliest stages of exposure to virus and before host cell encounter.

Chimeric Cyano Engineered Pro HIV-1

Mark Contarino^a, Arang Ramalingam Venkat Ka Vamshi Gangupomu^d, I

+ Author Affiliations

ABSTRACT

Human immunodeficient the AIDS pandemic. In t to test the possibility of that simultaneously bin fusion of the lectin cyan region (MPER) peptide w between the C terminus recombinant proteins, e immobilized metal affin to display a nanomolar HOS.T4.R5 cells. This ar cell infection by vesicul virus. Importantly, the c protein from both BaL- dose-dependent manne or CVN was found to ou components of the chin the Env protein spike ar virus and lead to inactiv metastability and to eva exposure to virus and b



Cognitive Computation G

Wikifier De

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content

wikify! clear

* If you wish to cite this work, please

Human immunodeficiency AIDS pandemic. We const (MPER) peptide along wit between the C terminus of recombinant proteins, exp metal affinity chromatogra nanomolar efficacy in bloc antiviral activity was HIV- stomatitis virus (VSV). The from both BaL-pseudotyp manner in the absence of h outcompete this virolytic e required for virolysis. The using a chimeric ligand car to investigate virus particl the earliest stages of expos

Article Talk

Fusion protein

From Wikipedia, the free encyclopedia

This article is about chimeric fusion proteins. For proteins involved in memb

Fusion proteins or **chimeric proteins** (literally, made of parts from different so originally coded for separate proteins. Translation of this *fusion gene* results in a

Article Talk

Gp41

From Wikipedia, the free encyclopedia

Gp41 also known as **glycoprotein 41** is a subunit of the envelope protein complex of retroviruses *Human immunodeficiency virus* (HIV). Gp41 is a **transmembrane protein** that contains several site ectodomain that are required for infection of host cells.

Article Talk

Affinity chromatography

From Wikipedia, the free encyclopedia

Affinity chromatography is a method of separating biochemical mixtures based on antibody, enzyme and substrate, or receptor and ligand.

Educational Applications: Unfamiliar domains may contain terms unknown to a reader. The Wikifier can supply the necessary background knowledge even when the relevant article titles are not identical to what appears in the text, dealing with both ambiguity and variability.

clinical presentation is identical to the very important foot and mouth disease virus.¹

Wikify System

1. Determine "linkable" phrases
 - mention detection
2. Rank/select candidate entity links
 - link generation
3. Use "context" to disambiguate/filter/improve
 - disambiguation

A sampler of research problems

- **Growth:** knowledge graphs are incomplete!
 - *Link prediction:* add relations
 - *Ontology matching:* connect graphs
 - *Knowledge extraction:* extract new entities and relations from web/text
 - **Validation:** knowledge graphs are not always correct!
 - *Entity resolution:* merge duplicate entities, split wrongly merged ones
 - *Error detection:* remove false assertions
 - **Interface:** how to make it easier to access knowledge?
 - *Semantic parsing:* interpret the meaning of queries
 - *Question answering:* compute answers using the knowledge graph
 - **Intelligence:** can AI emerge from knowledge graphs?
 - *Automatic reasoning* and planning
 - Generalization and abstraction
-

SUMMARY

Knowledge is crucial yet difficult to acquire

- Knowledge is crucial for many AI tasks
 - Knowledge acquisition
 - From **experts**: slow and mostly reliable
 - From **non-experts**: faster and not always reliable
 - **Automated**: fastest and most scalable, yet noisiest
 - Knowledge availability
 - **A lot** can be found online
 - **A lot** cannot be found
 - **A lot** cannot be extracted using today's methods
-

Where we are today

- We can **extract a lot of knowledge from text** and model its correctness
 - **Enforcing structure** makes the extraction problem easier yet imposes limitations
 - **Leveraging existing knowledge repositories** helps a lot
-

Next steps

- We need **new extraction methods**, from **new sources**
- Extracting from **modalities other than text** appears promising yet mostly unexplored

Plenty to be learned, problems are far from solved!

- Vibrant research area
- Numerous open research questions