Cross-lingual Entity Discovery and Linking

Heng Ji

(RPI → UIUC in Fall 2019; I’m Hiring)

Some materials from joint ACL2018 tutorial with Avi Sil (IBM Research AI), Dan Roth (UPenn), Silviu-Petru Cucerzan (MSR)

Tutorial: http://nlp.cs.rpi.edu/ccl.pptx
Reading List: http://nlp.cs.rpi.edu/kbp/2018/elreading.html

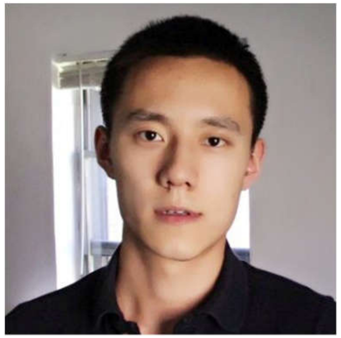
Thank You – Our Brilliant EDLers!!

Not shown: earlier generation of students

Lifu Huang

Ying Lin

Xiaoman Pan



Boliang Zhang

Nitish Gupta

Chen-Tse Tsai

Shyam Upadhyay

A few classroom rules for today!

- Homework: talk to Kevin Knight @ DiDi AI Labs for the best NLP job opportunities
- I’ll speak Mandarin upon popular requests, but my Mandarin is bad, especially on terminologies, so I’ll be doing code-switch like a 假洋鬼子 or a banana
    - But research does not have national boundaries!
- Ask me questions, comment, talk to me, please
    - Why don’t we do Zumba with Youku?
    - I’ll treat the first student who asks me questions 口水虾 for 夜宵
- Call me with my first name or full name instead of Professor / Dr. / Sister unless I’m more than 50 years older than you or your last name is Liu; Use 你 instead of 您
- Warning: the content of this tutorial will be out-of-date within five years!
    - Why should you like this area: useful, challenging, good starting point
    - My goal is to stimulate your interests in this area and let you take it over
    - 小猫在秋天的土地上种下果实，到了春天，远方的田野上姹紫嫣红，开满了美丽的鱼和诗歌

Outline

➡ Motivation, Task, Application	30 min
• Traditional “Ancient” Approaches	15 min
• Modern Approaches	
○ Language Universal EDL	15 min
○ Multi-lingual Common Space Construction	30 min
Coffee Break	
○ Cross-lingual Transfer Learning	20 min
○ Cross-lingual Neural Entity Linking	20 min
• Remaining Challenges and New Directions	30 min
• Demos, Resources and QA	20 min

What are “entities”?

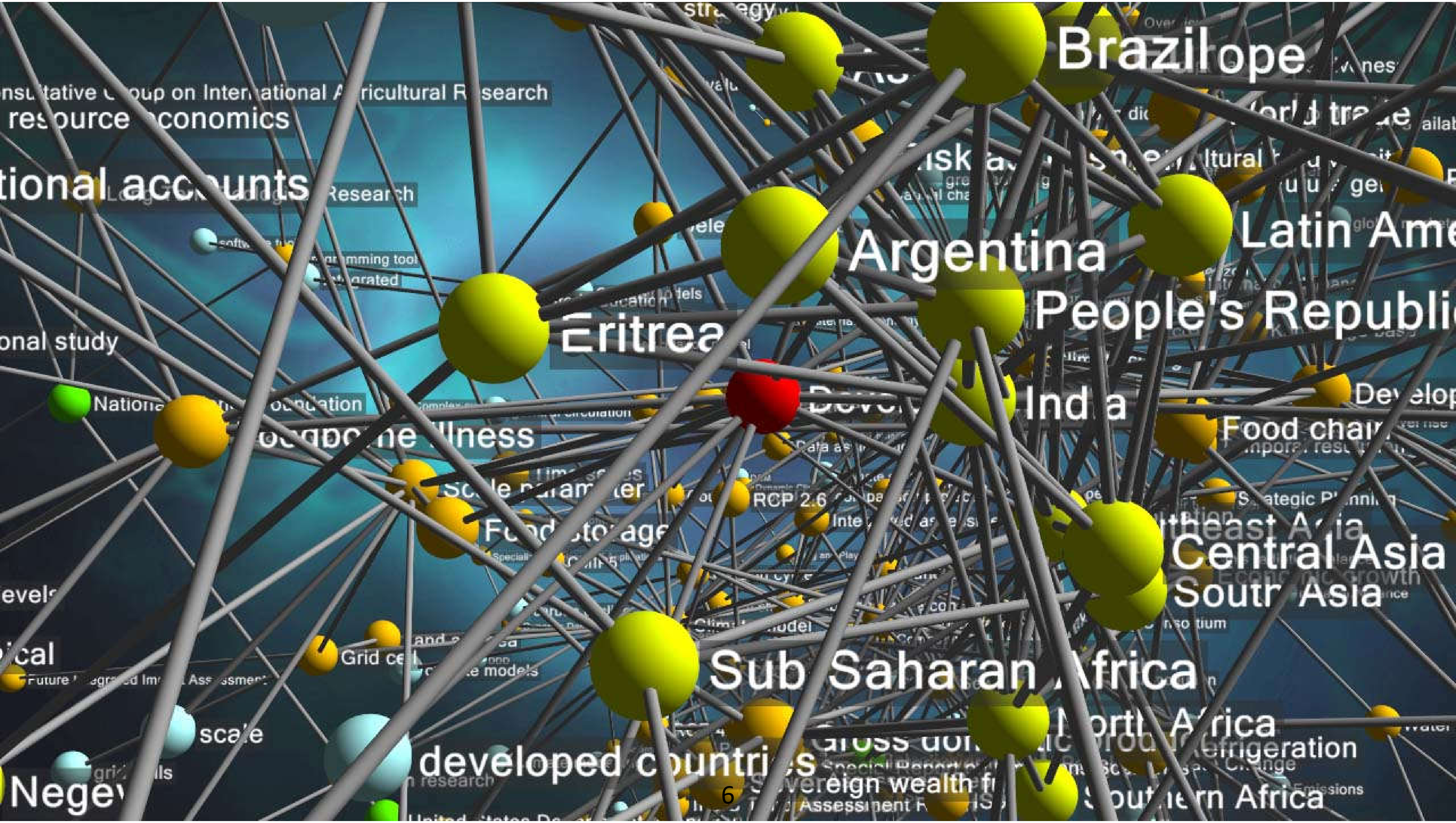
- **[Main meaning]**

- unique world bodies with (non-unique) names, such as people, organizations, locations
  e.g. ***Washington County***

- **[Extended meaning – information extraction]**

- unique identifiers, such as URLs, email addresses, tracking numbers, hashtags
- expressions of time, quantities, monetary values
- Concepts (e.g. ***county***)

- **But, everything is ambiguous**


Brazil
Argentina
Latin Ame
People's Republi
Eritrea
India
Central Asia
South Asia
Sub Saharan Africa
North Africa
Southern Africa
developed countries
Negev
tional accounts
resource economics
nsultative Group on International Agricultural Research
onal study
Food chain
Food storage
Scale parameter
RCP 2.6
Refrigeration
Sovereign wealth fu
World trade
Research
scale
Grid cell
Strategic Planning
evels
ical
Future Integrated Impact Assessment
Emissions

How Many Entities Are There?

- **Named places on Earth:**

  ~ 10 million

- **Books written:**

  ~ 130 million

- **People:**

  ~ 8 billion

- **Species on Earth:**

  ~ 1 trillion (of which ~10 million catalogued)

- **Stars in the Universe:**

  ~ 1 billion trillion (1,000,000,000,000,000,000,000)

How Many Attributes?

How Many Relationships?

How Many Mentions?

1. Mentions

- A mention: a phrase used to refer to something in the world
  - Named entity (person, organization), object, substance, event, philosophy, mental state, rule …

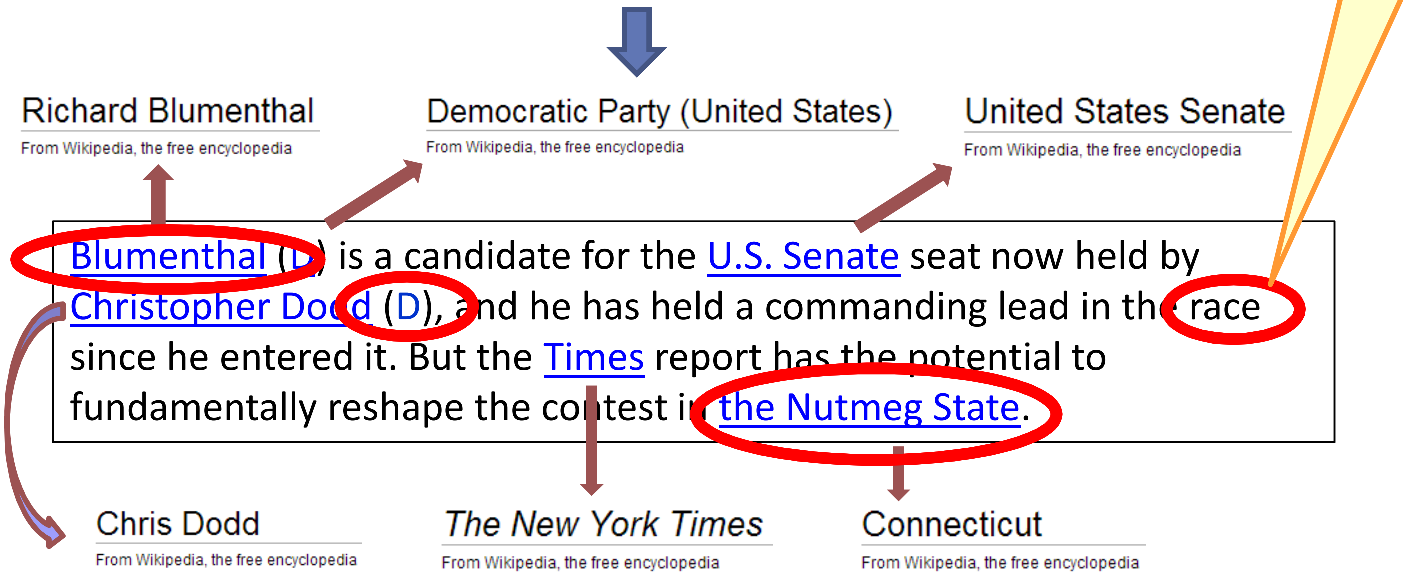
- Task definitions vary across the definition of mentions
  - All N-grams (up to a certain size); Dictionary-based selection; Data-driven controlled vocabulary (e.g., all Wikipedia titles); **only named entities.**

- Ideally, one would like to have a mention definition that adapts to the application/user

Examples of Mentions

This specific senate race (coreference)?
Or: the general concept of a political race?

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.


Richard Blumenthal
From Wikipedia, the free encyclopedia
Democratic Party (United States)
From Wikipedia, the free encyclopedia
United States Senate
From Wikipedia, the free encyclopedia

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

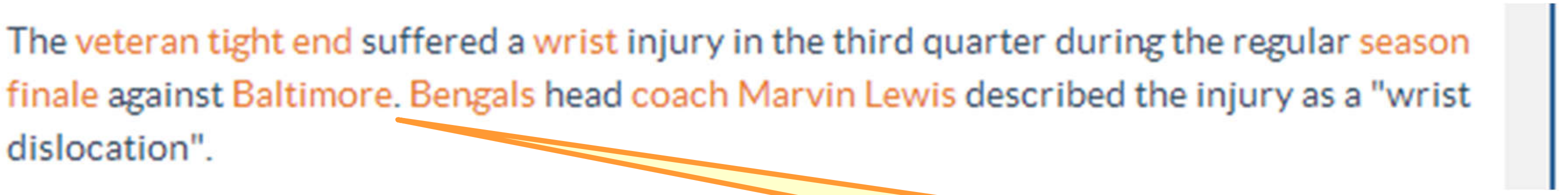
Chris Dodd
From Wikipedia, the free encyclopedia
The New York Times
From Wikipedia, the free encyclopedia
Connecticut
From Wikipedia, the free encyclopedia

2. Concept Inventory (KB)

- Multiple KBs can be used, in principle, as the target KB.
- Wikipedia has the advantage of a broad coverage, regularly maintained KB, with significant amount of text associated with each title.
  - All type of pages?
    - Content pages
    - Disambiguation pages
    - List pages
- What should happened to mentions that do not have entries in the target KB?

3. What to Link to?

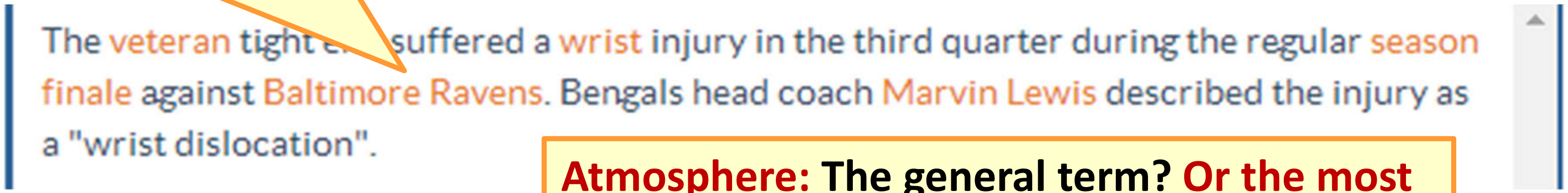
- Often, there are multiple sensible links.



The veteran tight end suffered a wrist injury in the third quarter during the regular season finale against Baltimore. Bengals head coach Marvin Lewis described the injury as a "wrist dislocation".

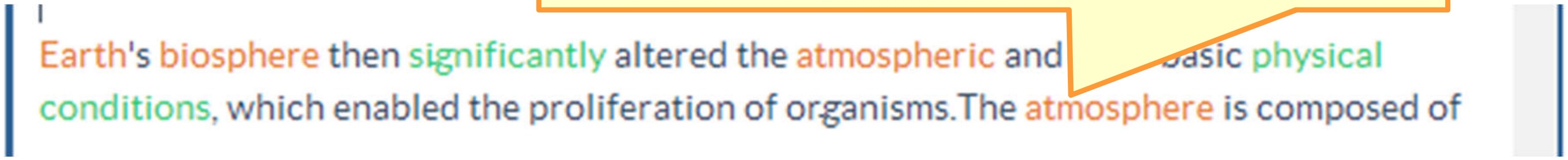
Baltimore Ravens: Should the link be any different? Both?

Baltimore: The city? Baltimore Ravens, the Football team? Both?



The veteran tight end suffered a wrist injury in the third quarter during the regular season finale against Baltimore Ravens. Bengals head coach Marvin Lewis described the injury as a "wrist dislocation".

Atmosphere: The general term? Or the most specific one “Earth Atmosphere?



Earth's biosphere then significantly altered the atmospheric and basic physical conditions, which enabled the proliferation of organisms.The atmosphere is composed of

3. Null Links

There is a Dorothy Byrne Wikipedia page; but it’s not this Dorothy…

- Often, there are multiple sensible links.

Dorothy Byrne, a state coordinator for the Florida Green Party,…

- How to capture the fact that Dorothy Byrne does not refer to any concept in Wikipedia?

- Wikification: Simply map Dorothy Byrne → NIL
- Entity Linking: If multiple mentions in the given document(s) correspond to the same concept, which is outside KB
  - First cluster relevant mentions as representing a single concept
  - Map the cluster to NIL

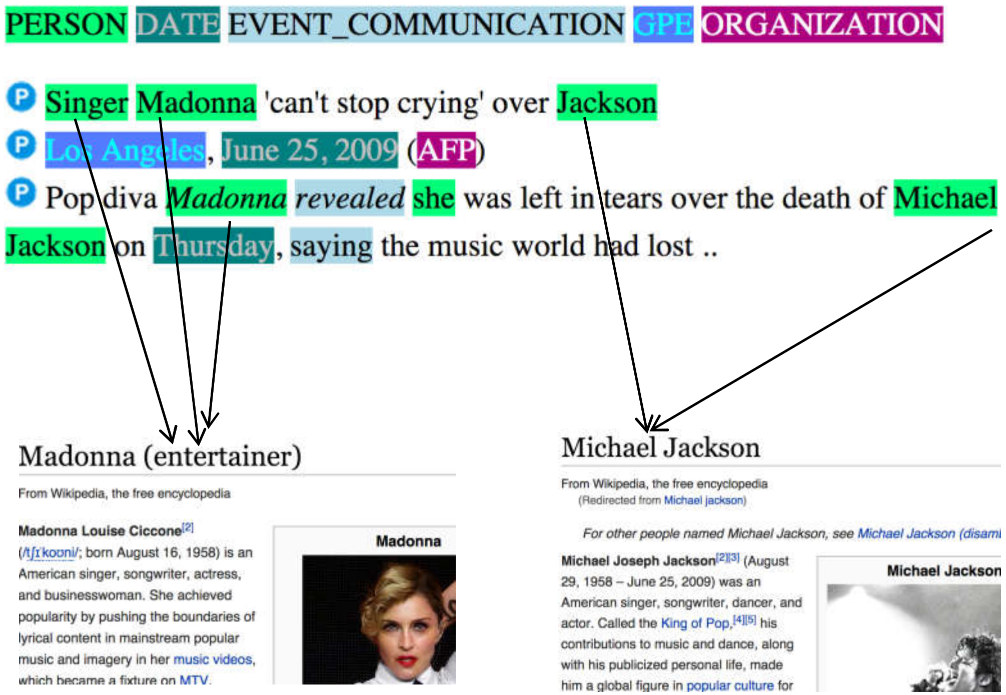
Evaluation

- In principle, evaluation on an application is possible, but hasn't been pursued
- Factors in Entity Discovery and Linking Evaluation:
  - Mention Selection (*)
    - Are the mentions chosen for linking correct (R/P)
  - Linking accuracy
    - Evaluate quality of links chosen per-mention
      - Ranking
      - Accuracy (including NIL) (*)
    - End-to-End
  - NIL clustering (*)
    - Entity Linking: evaluate out-of-KB clustering (co-reference)
  - Evaluating nominal as co-reference
  - Other (including IR-inspired) metrics
    - E.g. MRR, MAP, R-Precision, Recall, accuracy

TAC-KBP Cross-lingual Entity Discovery and Linking Track

- 2009 – Present
  - 2009 - 2013: Entity Linking
  - 2014– Present: Entity Discovery and Linking

```
<DOC id="AFP_ENG_20090626.0737"
type="story" >
<HEADLINE>Singer Madonna 'can't stop
crying' over Jackson</HEADLINE>
<DATELINE>Los Angeles, June 25, 2009
(AFP)</DATELINE>
<TEXT><P>Pop diva Madonna revealed
she was left in tears over the death
of
Michael Jackson on Thursday, saying
the music world had lost ..</P>
</TEXT>
</DOC>
```



PERSON DATE EVENT_COMMUNICATION GPE ORGANIZATION

Singer Madonna 'can't stop crying' over Jackson

Los Angeles, June 25, 2009 (AFP)

Pop diva Madonna revealed she was left in tears over the death of Michael Jackson on Thursday, saying the music world had lost ..

Madonna (entertainer)

From Wikipedia, the free encyclopedia

Madonna Louise Ciccone[2] (/tʃɪˈkoʊni/; born August 16, 1958) is an American singer, songwriter, actress, and businesswoman. She achieved popularity by pushing the boundaries of lyrical content in mainstream popular music and imagery in her music videos, which became a fixture on MTV.

Madonna

Michael Jackson

From Wikipedia, the free encyclopedia
(Redirected from Michael jackson)

For other people named Michael Jackson, see Michael Jackson (disam

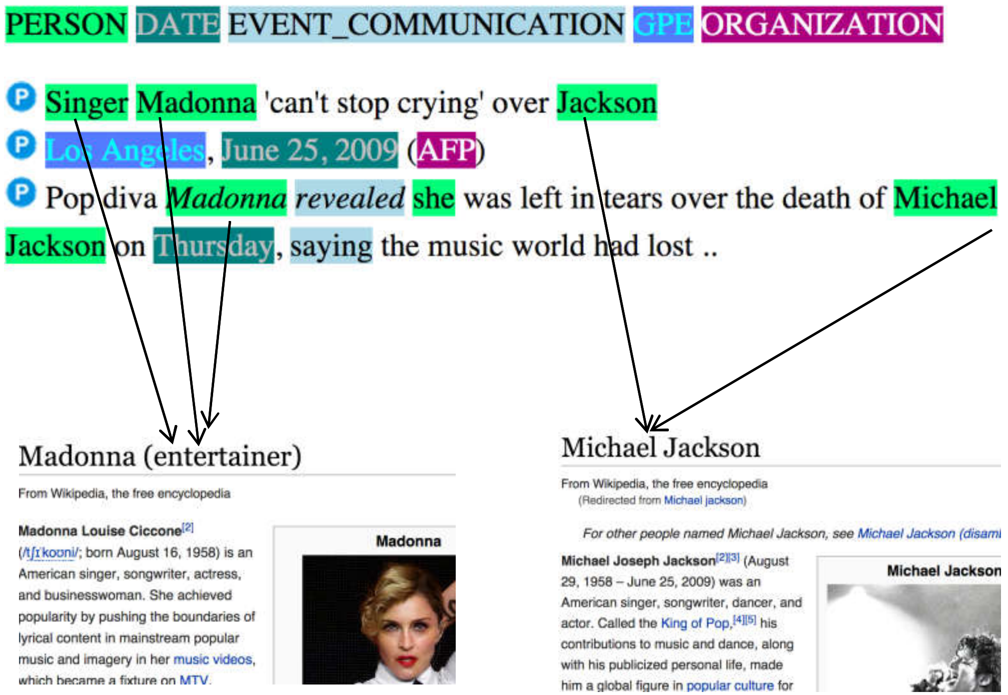
Michael Joseph Jackson[2][3] (August 29, 1958 – June 25, 2009) was an American singer, songwriter, dancer, and actor. Called the King of Pop,[4][5] his contributions to music and dance, along with his publicized personal life, made him a global figure in popular culture for

Michael Jackson

TAC-KBP Cross-lingual Entity Discovery and Linking Track

- 2009 – Present
  - ○ 2009 - 2014: Entity Linking
  - ○ 2014– Present: Entity Discovery and Linking

Link entities in English documents to an English KB



PERSON DATE EVENT_COMMUNICATION GPE ORGANIZATION

Singer Madonna 'can't stop crying' over Jackson

Los Angeles, June 25, 2009 (AFP)

Pop diva Madonna revealed she was left in tears over the death of Michael Jackson on Thursday, saying the music world had lost ..

Madonna (entertainer)

From Wikipedia, the free encyclopedia

Madonna Louise Ciccone[2] (/tʃɪˈkoʊni/; born August 16, 1958) is an American singer, songwriter, actress, and businesswoman. She achieved popularity by pushing the boundaries of lyrical content in mainstream popular music and imagery in her music videos, which became a fixture on MTV.

Madonna

Michael Jackson

From Wikipedia, the free encyclopedia

(Redirected from Michael jackson)

For other people named Michael Jackson, see Michael Jackson (disam

Michael Joseph Jackson[2][3] (August 29, 1958 – June 25, 2009) was an American singer, songwriter, dancer, and actor. Called the King of Pop,[4][5] his contributions to music and dance, along with his publicized personal life, made him a global figure in popular culture for

Michael Jackson

Evaluation metrics:

Linking accuracy (A),

Known-entity linking accuracy (A_{Wiki}),

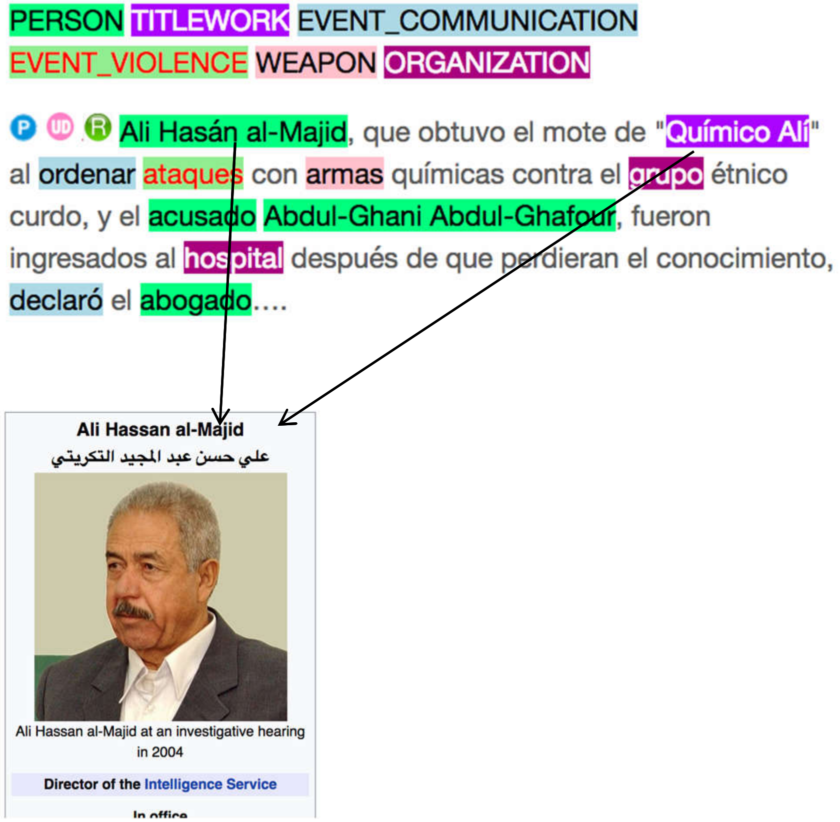
NIL accuracy (A_{NIL}),

B-cubed precision and recall with equal element weighting

TAC-KBP Cross-lingual Entity Discovery and Linking Track

- 2009 – Present
  - 2009 - 2013: Entity Linking
  - 2014– Present: Entity Discovery and Linking

```
<DOC id=”XIN_SPA_20090726.0345" type="story" >
<TEXT><P>  Ali Hasán al-Majid, que obtuvo el mote de "Químico Alí" al
ordenar ataques con armas químicas contra el grupo étnico curdo, y el acusado
Abdul-Ghani Abdul-Ghafour, fueron ingresados al hospital después de que
perdieran el conocimiento, declaró el abogado…..</P>
</TEXT>
</DOC>
```



PERSON TITLEWORK EVENT_COMMUNICATION
EVENT_VIOLENCE WEAPON ORGANIZATION

P UD R Ali Hasán al-Majid, que obtuvo el mote de "Químico Alí" al ordenar ataques con armas químicas contra el grupo étnico curdo, y el acusado Abdul-Ghani Abdul-Ghafour, fueron ingresados al hospital después de que perdieran el conocimiento, declaró el abogado….

Ali Hassan al-Majid
علي حسن عبد المجيد التكريتي

Ali Hassan al-Majid at an investigative hearing in 2004

Director of the Intelligence Service

In office

- Spanish and Chinese documents
- Link mentions to English KB!!

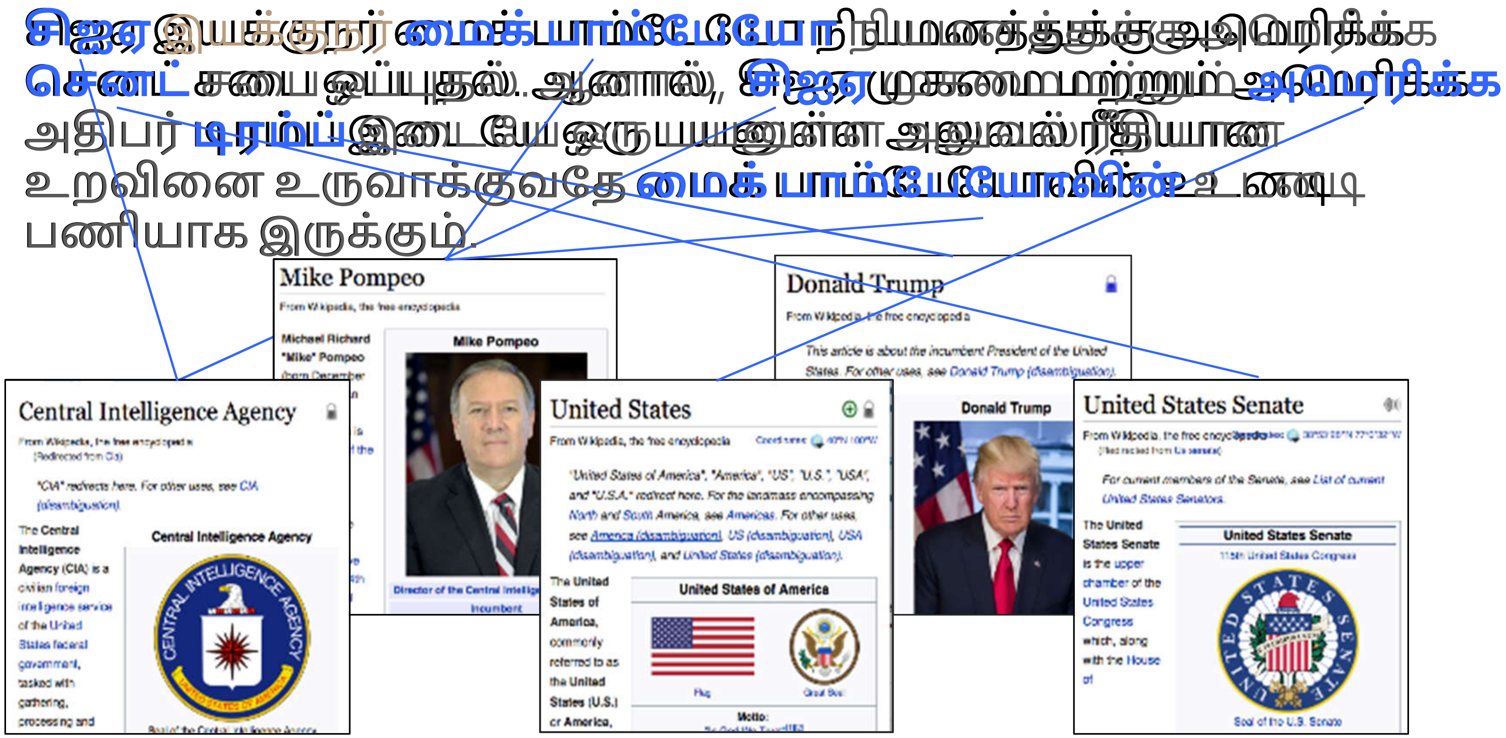
Recently new languages have been added:

- Norther Sotho, Nepali, Kikuyu…

Some with very small or no Wikipedia!!!

TAC-KBP Cross-lingual Entity Discovery and Linking Track

- Given a non-English document, extract named entities and disambiguate into the English Wikipedia



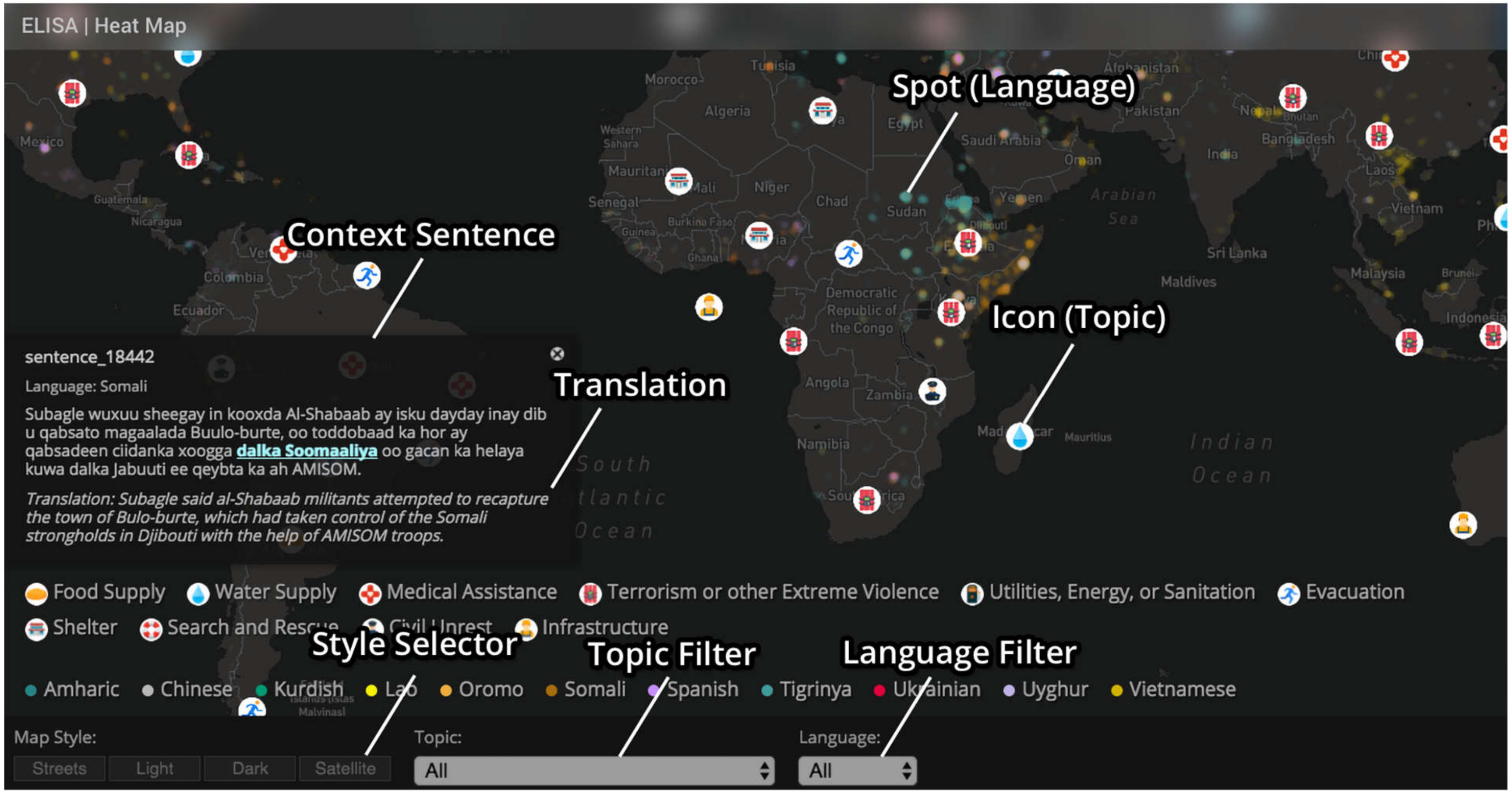
சிஐஏ இயக்குநர் மைக் பாம்பேயோ நியமனத்துக்கு அமெரிக்க செனட் சபை ஒப்புதல். ஆனால், சிஐஏ முகமை மற்றும் அமெரிக்க அதிபர் டிரம்ப் இடையே ஒரு பயனுள்ள அலுவல் ரீதியான உறவினை உருவாக்குவதே மைக் பாம்பேயோவின் உண்மைப் பணியாக இருக்கும்.

Mike Pompeo
From Wikipedia, the free encyclopedia
Mike Pompeo
Donald Trump
From Wikipedia, the free encyclopedia
Donald Trump
Central Intelligence Agency
From Wikipedia, the free encyclopedia
"CIA" redirects here. For other uses, see CIA (disambiguation).
Central Intelligence Agency
United States
From Wikipedia, the free encyclopedia
United States of America
United States Senate
From Wikipedia, the free encyclopedia
United States Senate

Applications of Entity Linking

- **Not Enough!**
- Used as an intermediate task in other NLP tasks:
  - Knowledge Acquisition (via grounding)
    - Still remains open: how to organize the knowledge in a useful way?
    - **Fine entity typing** [an emerging application]
  - Co-reference resolution (Ratinov & Roth, 2012)
    - "After the vessel suffered a catastrophic torpedo detonation, Kursk sank in the waters of Barents Sea…"
    - Knowing Kursk → Russian submarine K-141 Kursk helps system to co-ref "Kursk" and "vessel"
  - Document classification
    - Tweets labeled World, US, Science & Technology, Sports, Business, Health, Entertainment (Vitale et. al., 2012)
    - Zero-shot/Datalesss classification (ESA-based representations; Song & Roth' 14)
    - Document and concepts are represented via Wikipedia titles
  - Visualization: Geo- visualization of News (Gao et. al. CHI'14)
- Products

Application: Disaster Relief



sentence_18442

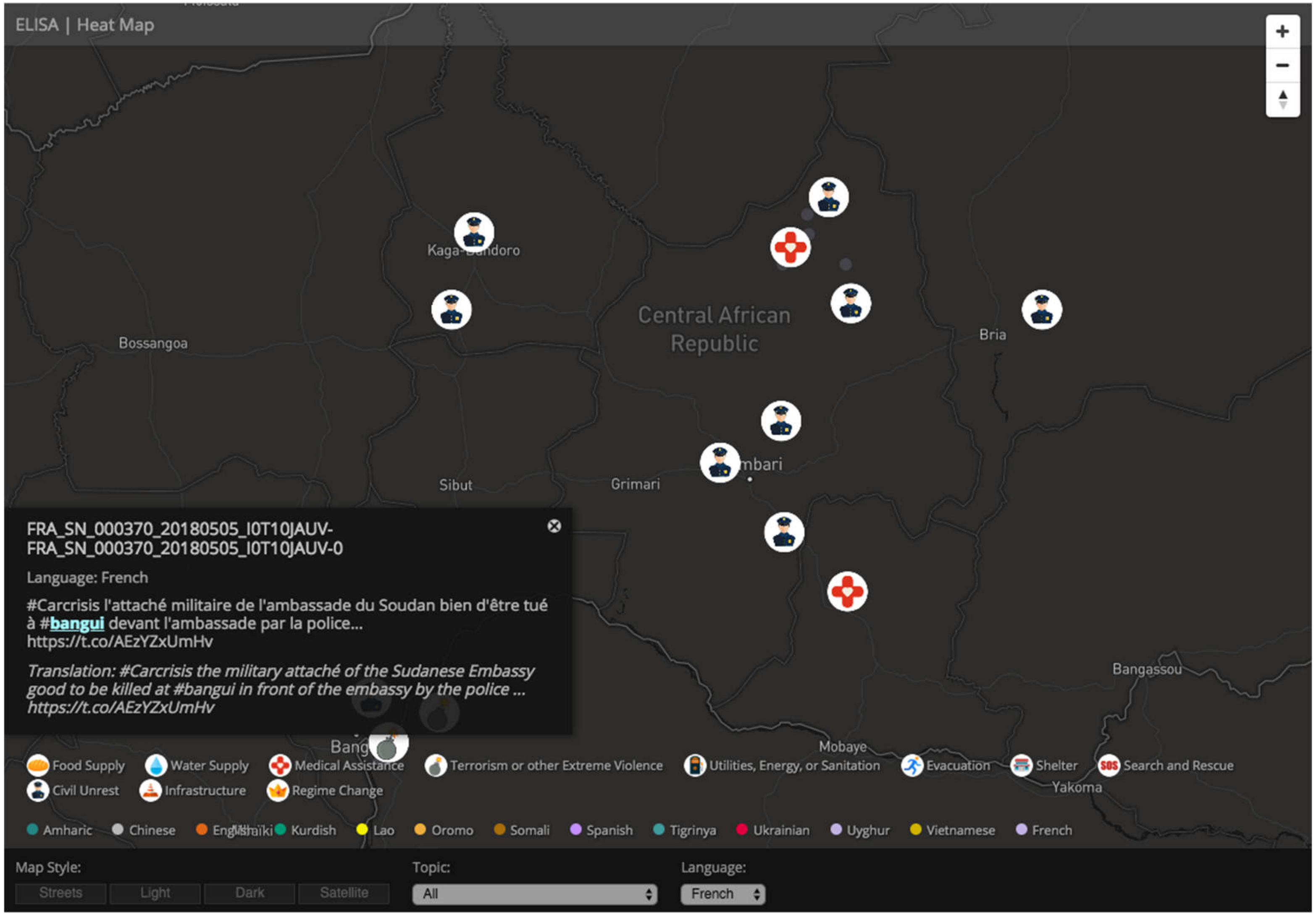
Language: Somali

Subagle wuxuu sheegay in kooxda Al-Shabaab ay isku dayday inay dib u qabsato magaalada Buulo-burte, oo toddobaad ka hor ay qabsadeen ciidanka xoogga **dalka Soomaaliya** oo gacan ka helaya kuwa dalka Jabuuti ee qeybta ka ah AMISOM.

Translation: Subagle said al-Shabaab militants attempted to recapture the town of Bulo-burte, which had taken control of the Somali strongholds in Djibouti with the help of AMISOM troops.

- Selected for the US DARPA 60 Anniversary

Central Africa Republic Exercise



FRA_SN_000370_20180505_I0T10JAUV-FRA_SN_000370_20180505_I0T10JAUV-0

Language: French

#Carcrisis l'attaché militaire de l'ambassade du Soudan bien d'être tué à #**bangui** devant l'ambassade par la police... https://t.co/AEzYZxUmHv

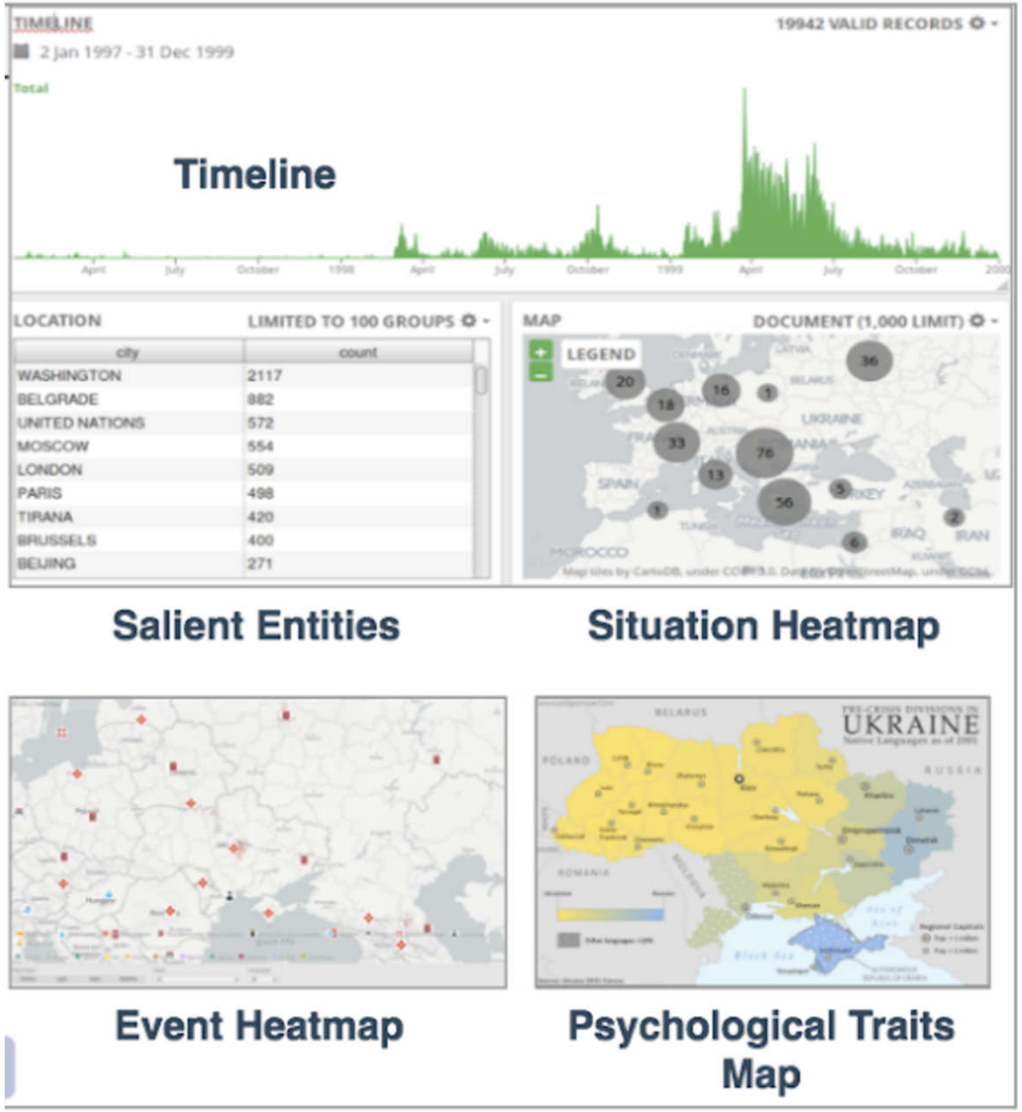
Translation: #Carcrisis the military attaché of the Sudanese Embassy good to be killed at #bangui in front of the embassy by the police ... https://t.co/AEzYZxUmHv

Map Style:

Topic: All

Language: French

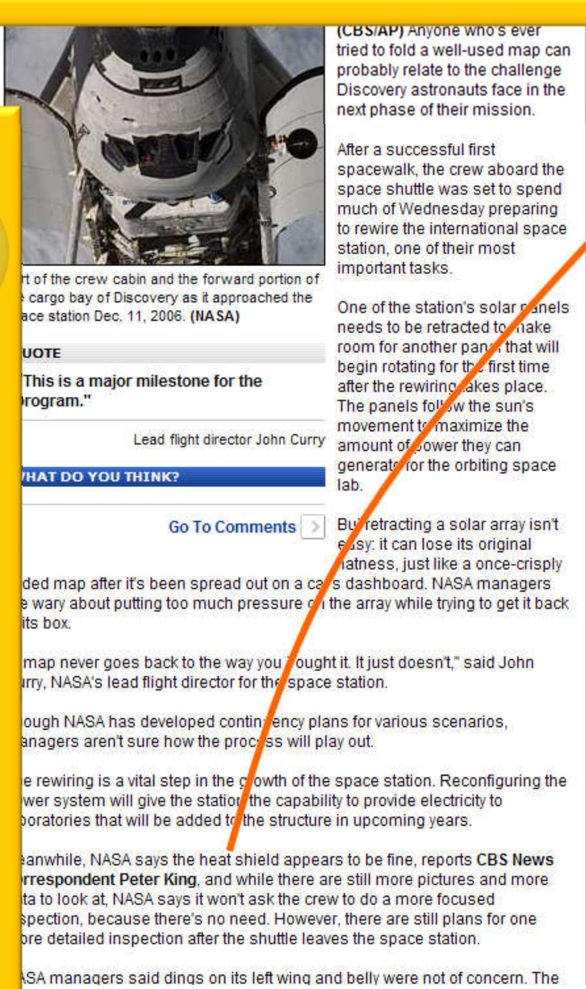
Macedonian Exercise



city	count
WASHINGTON	2117
BELGRADE	882
UNITED NATIONS	572
MOSCOW	554
LONDON	509
PARIS	498
TIRANA	420
BRUSSELS	400
BEIJING	271

Application (and Motivation): Context-aware Search

NASA
Space station
Solar panels
Discovery
Space shuttle
Space lab
John Curry
Atmospheric reentry
Power system
Peter King
CBS News
Associated Press



rt of the crew cabin and the forward portion of
 cargo bay of Discovery as it approached the
ace station Dec. 11, 2006. **(NASA)**

UOTE

**This is a major milestone for the
rogram."**

Lead flight director John Curry

HAT DO YOU THINK?

Go To Comments

(CBS/AP) Anyone who's ever tried to fold a well-used map can probably relate to the challenge Discovery astronauts face in the next phase of their mission.

After a successful first spacewalk, the crew aboard the space shuttle was set to spend much of Wednesday preparing to rewire the international space station, one of their most important tasks.

One of the station's solar panels needs to be retracted to make room for another panel that will begin rotating for the first time after the rewiring takes place. The panels follow the sun's movement to maximize the amount of power they can generate for the orbiting space lab.

But retracting a solar array isn't easy: it can lose its original flatness, just like a once-crisply
ded map after it's been spread out on a car's dashboard. NASA managers
e wary about putting too much pressure on the array while trying to get it back
its box.

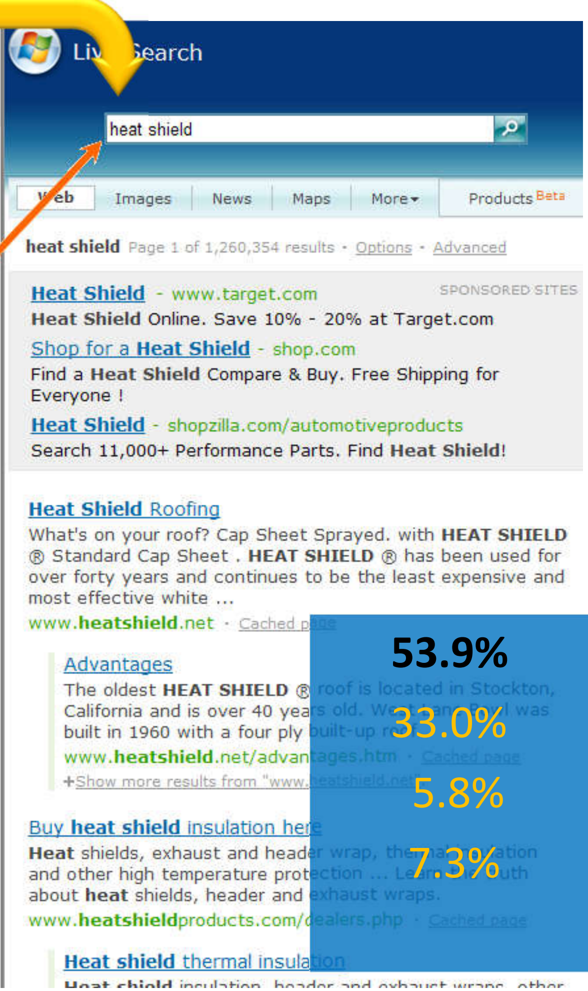
map never goes back to the way you bought it. It just doesn't," said John
rry, NASA's lead flight director for the space station.

ough NASA has developed contingency plans for various scenarios,
anagers aren't sure how the process will play out.

e rewiring is a vital step in the growth of the space station. Reconfiguring the
wer system will give the station the capability to provide electricity to
oratories that will be added to the structure in upcoming years.

anwhile, NASA says the heat shield appears to be fine, reports **CBS News
rrespondent Peter King**, and while there are still more pictures and more
ta to look at, NASA says it won't ask the crew to do a more focused
spection, because there's no need. However, there are still plans for one
ore detailed inspection after the shuttle leaves the space station.

SA managers said dings on its left wing and belly were not of concern. The



heat shield Page 1 of 1,260,354 results · Options · Advanced

SPONSORED SITES

Heat Shield - www.target.com
Heat Shield Online. Save 10% - 20% at Target.com

Shop for a **Heat Shield** - shop.com
Find a **Heat Shield** Compare & Buy. Free Shipping for Everyone !

Heat Shield - shopzilla.com/automotiveproducts
Search 11,000+ Performance Parts. Find **Heat Shield**!

Heat Shield Roofing
What's on your roof? Cap Sheet Sprayed. with **HEAT SHIELD** ® Standard Cap Sheet . **HEAT SHIELD** ® has been used for over forty years and continues to be the least expensive and most effective white ...
www.**heatshield**.net · Cached page

Advantages
The oldest **HEAT SHIELD** ® roof is located in Stockton, California and is over 40 years old. West Lane Bowl was built in 1960 with a four ply built-up roof.
www.**heatshield**.net/advantages.htm · Cached page
+Show more results from "www.heatshield.net"

Buy **heat shield** insulation here
Heat shields, exhaust and header wrap, thermal insulation and other high temperature protection ... Learn the truth about **heat** shields, header and exhaust wraps.
www.**heatshield**products.com/dealers.php · Cached page

Heat shield thermal insulation

53.9%	relevant
33.0%	irrelevant
5.8%	cannot say
7.3%	navigational for news

S. Cucerzan and E. Brill. Context-based Search and Document Retrieval. US Patent 7,974,964. 1/17/2007
M. Rahurkar and S. Cucerzan. Predicting when Browsing Context Is Relevant to Search. SIGIR 2008

Application: Ad Matching, Query Suggestion

Bing: 2008

- Normalization of keyphrases and queries:

e.g. big lots sales → big lots, sale

on sale at big lots → big lots, sale

big sale of lots → big, lot, sale

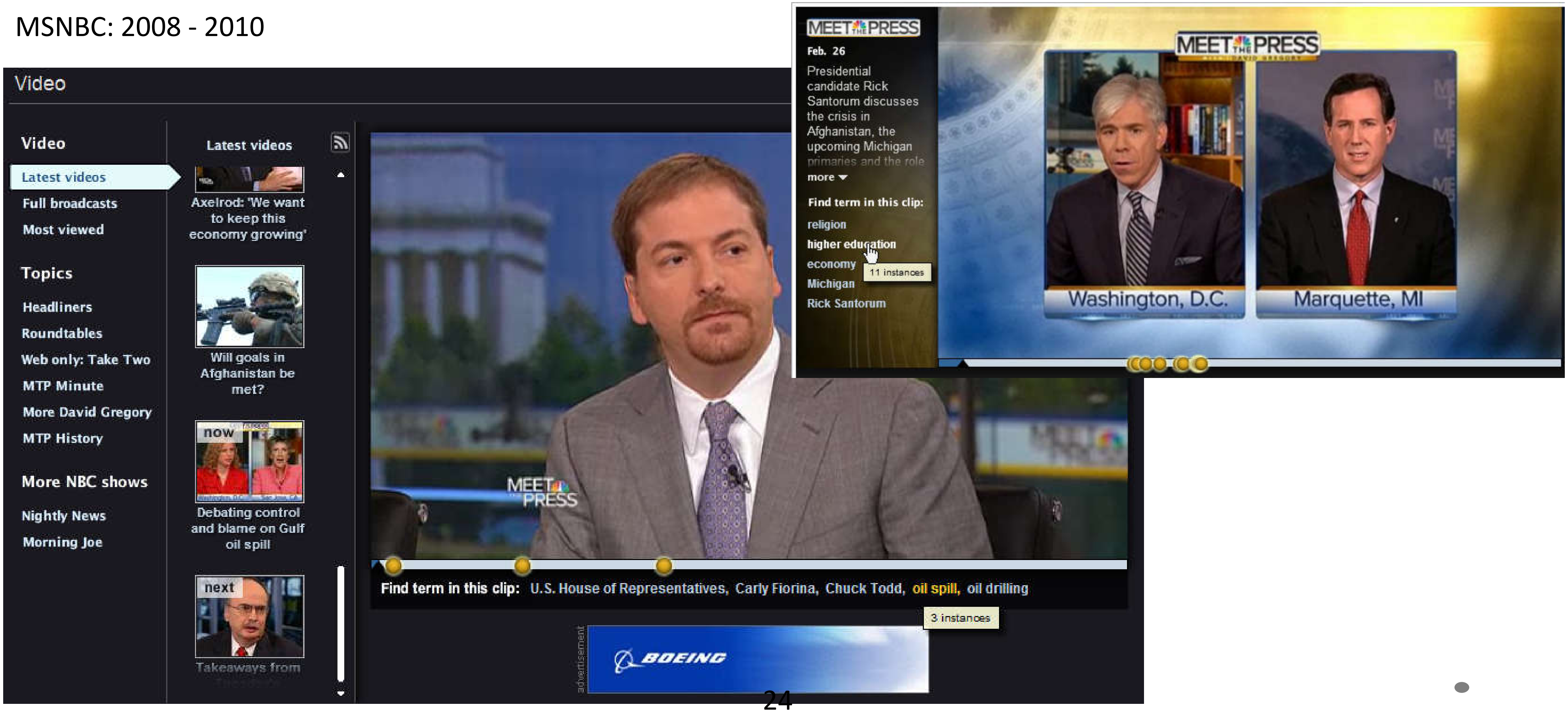
- Broad matching / Search query suggestion:

replace entities in a query with related entities

e.g. kia sorento mpg → hyundai santa fe mpg

Application: News Indexing and Video Annotation

MSNBC: 2008 - 2010



Axelrod: 'We want to keep this economy growing'

Will goals in Afghanistan be met?

Debating control and blame on Gulf oil spill

Takeaways from

Find term in this clip: U.S. House of Representatives, Carly Fiorina, Chuck Todd, oil spill, oil drilling

3 instances

MEET THE PRESS

Feb. 26

Presidential candidate Rick Santorum discusses the crisis in Afghanistan, the upcoming Michigan primaries and the role

Find term in this clip:

religion

higher education

economy

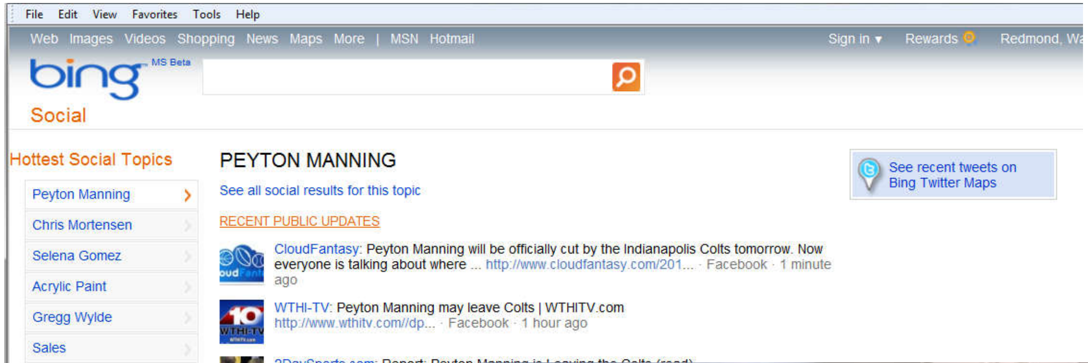
Michigan

Rick Santorum

11 instances

Applications: Social-media-driven News, Trending Topics

Real-time analysis
of social media



Bing: 2011

File Edit View Favorites Tools Help

Web Images Videos Shopping News Maps More | MSN Hotmail Sign in ▾ Rewards Redmond, Wa

Hottest Social Topics

- Peyton Manning
- Chris Mortensen
- Selena Gomez
- Acrylic Paint
- Gregg Wylde
- Sales
- Super Tuesday
- Beast Boy

PEYTON MANNING

See all social results for this topic

See recent tweets on Bing Twitter Maps

RECENT PUBLIC UPDATES

CloudFantasy: Peyton Manning will be officially cut by the Indianapolis Colts tomorrow. Now everyone is talking about where ... http://www.cloudfantasy.com/201... · Facebook · 1 minute ago

WTHI-TV: Peyton Manning may leave Colts | WTHITV.com http://www.wthitv.com//dp... · Facebook · 1 hour ago

2DaySports.com: Report: Peyton Manning is Leaving the Colts (read) http://www.2daysports.com/pey...

Yahoo Search: Say it ain't so! Pey Where do you think he'll end up? ago

ESPN 1420: It's official, the Colts w http://espn1420.com/pey... · Face

wrapupp.com: Colts to Cut Peyton http://www.wrapupp.com/nfl... · Fa

100.9 The Eagle: Peyton Manning http://www.huffingtonpost.com/201

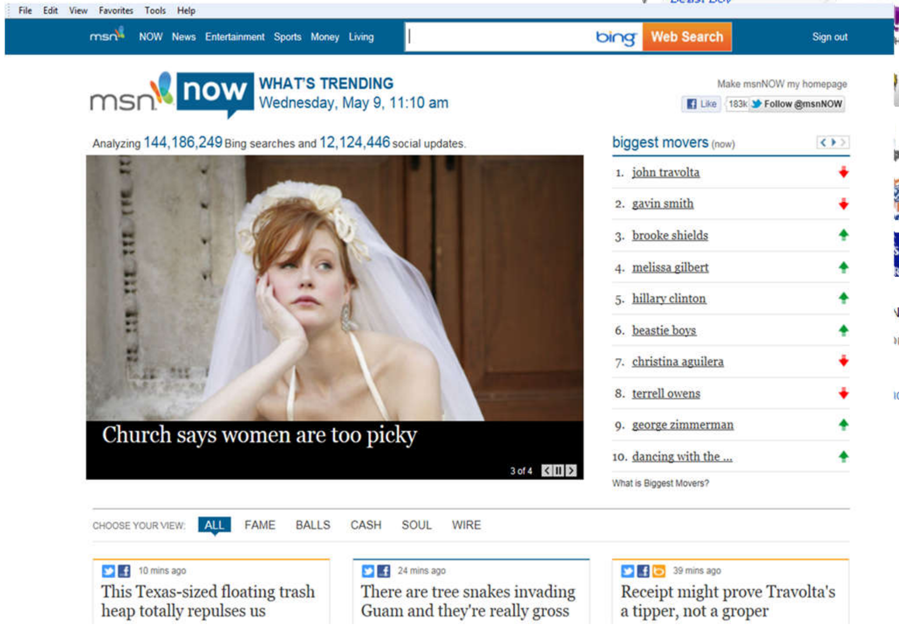
Fanatic Sports and Cards: Colts to http://www.fanaticsportsandcards.c

THIS TOPIC

rds Indianapolis Colts, ESPN, Chris

ore social results for Peyton Manning

MSN: 2012



File Edit View Favorites Tools Help

msn NOW News Entertainment Sports Money Living bing Web Search Sign out

WHAT'S TRENDING
Wednesday, May 9, 11:10 am

Make msnNOW my homepage
Like 183k Follow @msnNOW

Analyzing 144,186,249 Bing searches and 12,124,446 social updates.

Church says women are too picky

3 of 4

biggest movers (now)

1. john travolta
2. gavin smith
3. brooke shields
4. melissa gilbert
5. hillary clinton
6. beastie boys
7. christina aguilera
8. terrell owens
9. george zimmerman
10. dancing with the ...

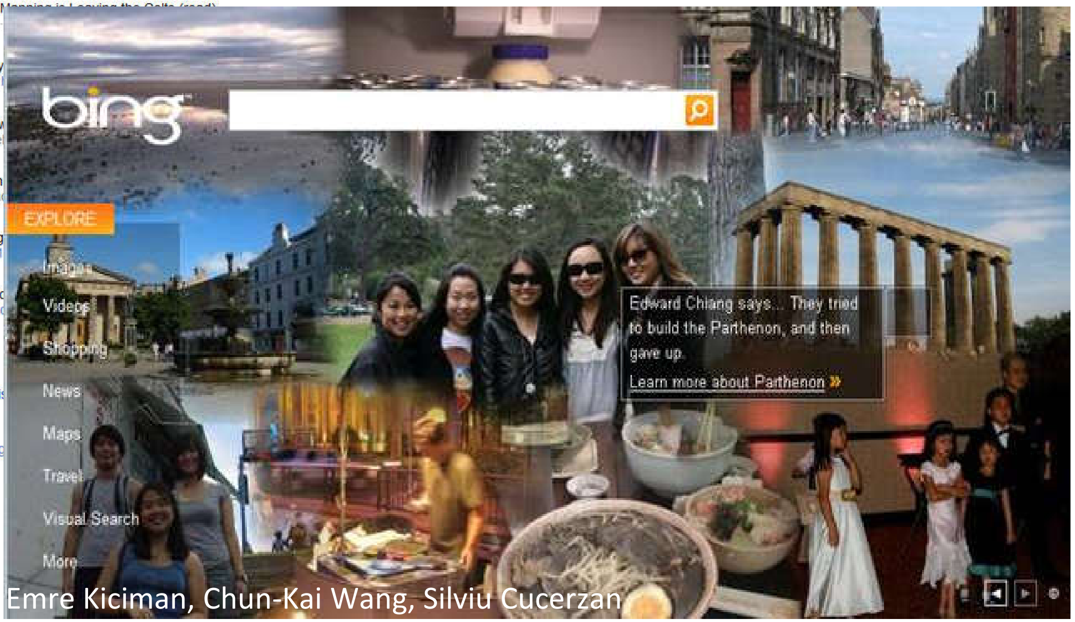
What is Biggest Movers?

CHOOSE YOUR VIEW: ALL FAME BALLS CASH SOUL WIRE

10 mins ago
This Texas-sized floating trash heap totally repulses us

24 mins ago
There are tree snakes invading Guam and they're really gross

39 mins ago
Receipt might prove Travolta's a tipper, not a groper

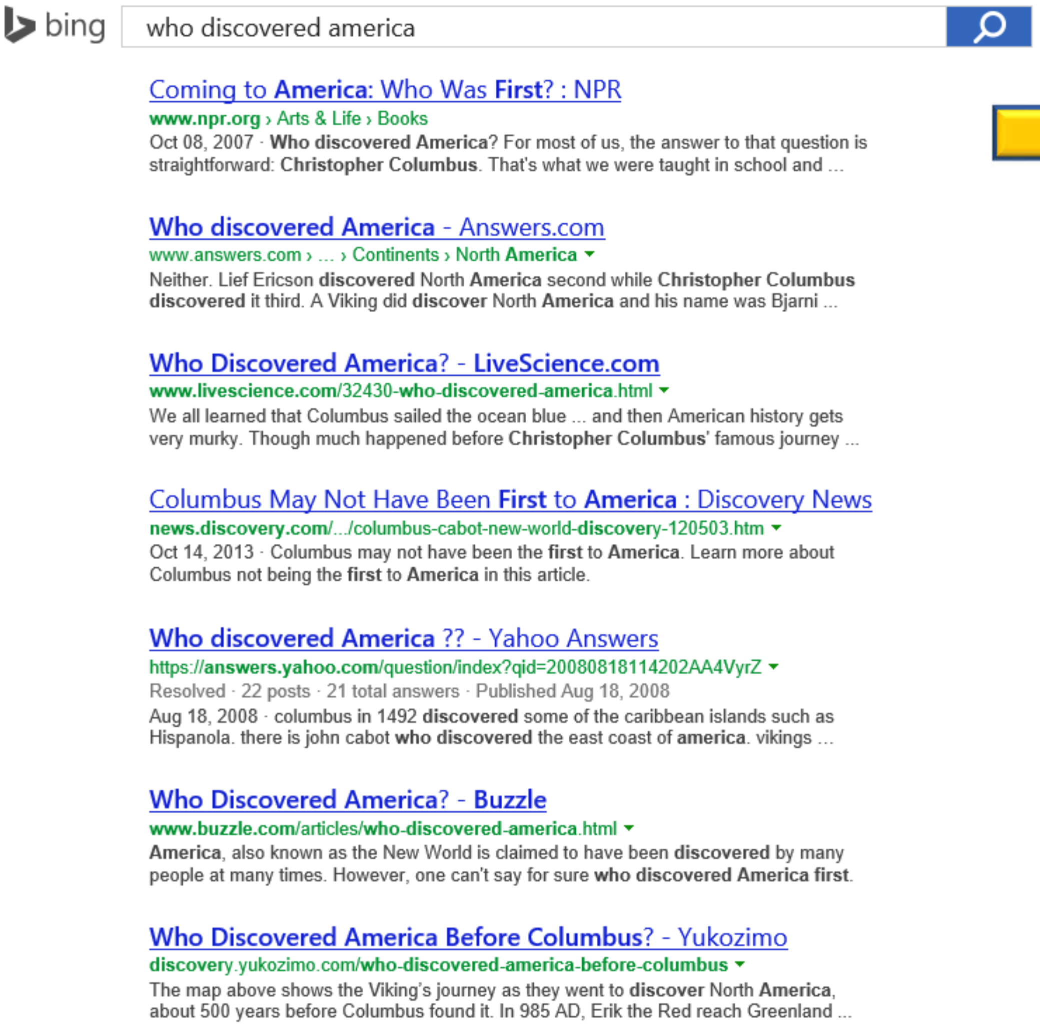


Edward Chiang says... They tried to build the Parthenon, and then gave up.
Learn more about Parthenon »

Emre Kiciman, Chun-Kai Wang, Silviu Cucerzan

http://apps.facebook.com/msrcollage/

Application: Web-based Factoid Question Answering



who discovered america

Coming to **America**: Who Was **First**? : NPR
www.npr.org › Arts & Life › Books
Oct 08, 2007 · **Who discovered America**? For most of us, the answer to that question is straightforward: **Christopher Columbus**. That's what we were taught in school and ...

Who discovered America - Answers.com
www.answers.com › ... › Continents › North **America**
Neither. Lief Ericson **discovered** North **America** second while **Christopher Columbus** **discovered** it third. A Viking did **discover** North **America** and his name was Bjarni ...

Who Discovered America? - **LiveScience.com**
www.livescience.com/32430-who-discovered-america.html
We all learned that Columbus sailed the ocean blue ... and then American history gets very murky. Though much happened before **Christopher Columbus**' famous journey ...

Columbus May Not Have Been **First** to **America** : Discovery News
news.discovery.com/.../columbus-cabot-new-world-**discovery**-120503.htm
Oct 14, 2013 · Columbus may not have been the **first** to **America**. Learn more about Columbus not being the **first** to **America** in this article.

Who discovered America ?? - Yahoo Answers
https://**answers.yahoo.com**/question/index?qid=20080818114202AA4VyrZ
Resolved · 22 posts · 21 total answers · Published Aug 18, 2008
Aug 18, 2008 · columbus in 1492 **discovered** some of the caribbean islands such as Hispanola. there is john cabot **who discovered** the east coast of **america**. vikings ...

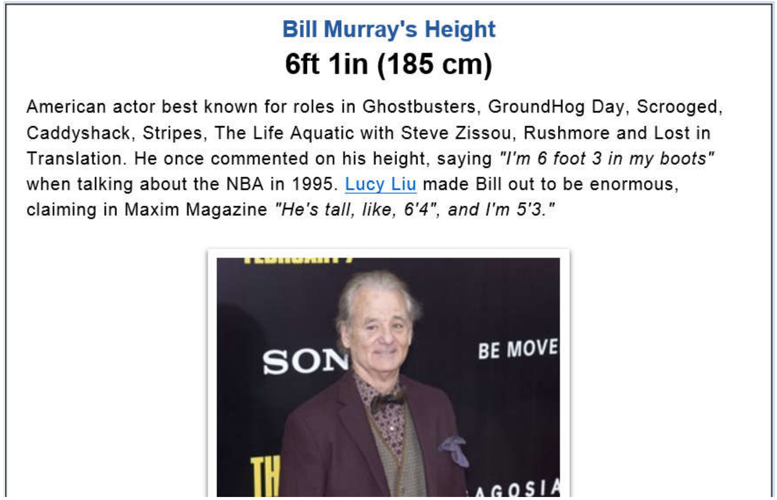
Who Discovered America? - **Buzzle**
www.buzzle.com/articles/**who-discovered-america**.html
America, also known as the New World is claimed to have been **discovered** by many people at many times. However, one can't say for sure **who discovered America first**.

Who Discovered America Before Columbus? - Yukozimo
discovery.yukozimo.com/**who-discovered-america-before-columbus**
The map above shows the Viking's journey as they went to **discover** North **America**, about 500 years before Columbus found it. In 985 AD, Erik the Red reach Greenland ...

Christopher Columbus

Bing: 2014

Application: Fact Extraction from the Web



Bill Murray's Height

6ft 1in (185 cm)

American actor best known for roles in Ghostbusters, GroundHog Day, Scrooged, Caddyshack, Stripes, The Life Aquatic with Steve Zissou, Rushmore and Lost in Translation. He once commented on his height, saying *"I'm 6 foot 3 in my boots"* when talking about the NBA in 1995. Lucy Liu made Bill out to be enormous, claiming in Maxim Magazine *"He's tall, like, 6'4", and I'm 5'3."*

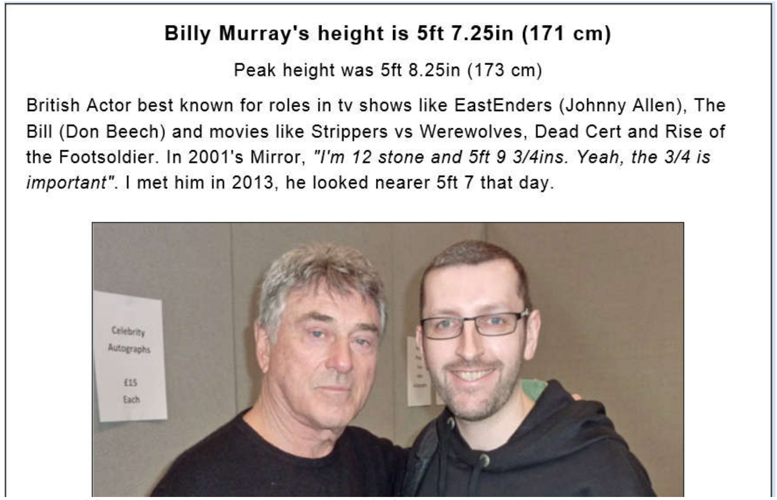
http://www.celebheights.com/s/Bill-Murray-32.html

Bill Murray's Height

NEMO → Satori Id
fe22ca15-6baf-0479-a40b-aab2298398e4

6ft 1in (185 cm)

American actor best known for roles in Ghostbusters, GroundHog Day, Scrooged, Caddyshack, Stripes, The Life Aquatic with Steve Zissou, Rushmore and Lost in Translation. He once commented on his height, saying "I'm 6 foot 3 in my boots" when talking about the NBA in 1995. Lucy Liu made Bill out to be enormous, claiming in Maxim Magazine "He's tall, like, 6'4", and I'm 5'3."



Billy Murray's height is 5ft 7.25in (171 cm)

Peak height was 5ft 8.25in (173 cm)

British Actor best known for roles in tv shows like EastEnders (Johnny Allen), The Bill (Don Beech) and movies like Strippers vs Werewolves, Dead Cert and Rise of the Footsoldier. In 2001's Mirror, *"I'm 12 stone and 5ft 9 3/4ins. Yeah, the 3/4 is important"*. I met him in 2013, he looked nearer 5ft 7 that day.

http://www.celebheights.com/s/Billy-Murray-4501.html

NEMO → Satori Id
de332f72-a4ac-6846-babd-d0f42db4371c

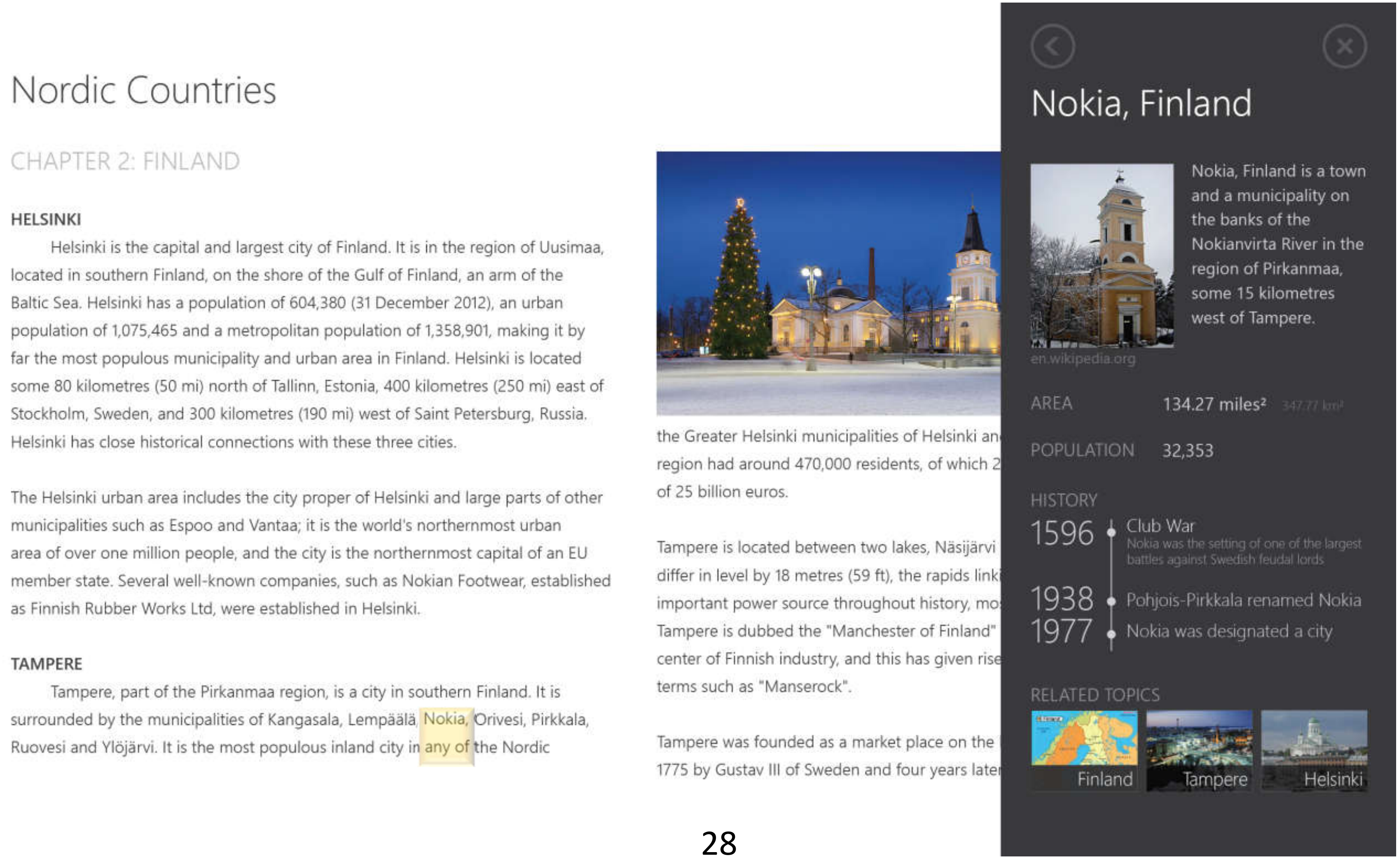
Billy Murray's height is 5ft 7.25in (171 cm)

Peak height was 5ft 8.25in (173 cm)

British Actor best known for roles in tv shows like EastEnders (Johnny Allen), The Bill (Don Beech) and movies like Strippers vs Werewolves, Dead Cert and Rise of the Footsoldier. In 2001's Mirror, "I'm 12 stone and 5ft 9 3/4ins. Yeah, the 3/4 is important". I met him in 2013, he looked nearer 5ft 7 that day.

Smart Lookup

Word: 2014



Nordic Countries

CHAPTER 2: FINLAND

HELSINKI

Helsinki is the capital and largest city of Finland. It is in the region of Uusimaa, located in southern Finland, on the shore of the Gulf of Finland, an arm of the Baltic Sea. Helsinki has a population of 604,380 (31 December 2012), an urban population of 1,075,465 and a metropolitan population of 1,358,901, making it by far the most populous municipality and urban area in Finland. Helsinki is located some 80 kilometres (50 mi) north of Tallinn, Estonia, 400 kilometres (250 mi) east of Stockholm, Sweden, and 300 kilometres (190 mi) west of Saint Petersburg, Russia. Helsinki has close historical connections with these three cities.

The Helsinki urban area includes the city proper of Helsinki and large parts of other municipalities such as Espoo and Vantaa; it is the world's northernmost urban area of over one million people, and the city is the northernmost capital of an EU member state. Several well-known companies, such as Nokian Footwear, established as Finnish Rubber Works Ltd, were established in Helsinki.

TAMPERE

Tampere, part of the Pirkanmaa region, is a city in southern Finland. It is surrounded by the municipalities of Kangasala, Lempäälä, Nokia, Orivesi, Pirkkala, Ruovesi and Ylöjärvi. It is the most populous inland city in any of the Nordic

the Greater Helsinki municipalities of Helsinki an
region had around 470,000 residents, of which 2
of 25 billion euros.

Tampere is located between two lakes, Näsijärvi
differ in level by 18 metres (59 ft), the rapids link
important power source throughout history, mo
Tampere is dubbed the "Manchester of Finland"
center of Finnish industry, and this has given rise
terms such as "Manserock".

Tampere was founded as a market place on the
1775 by Gustav III of Sweden and four years late

Nokia, Finland

Nokia, Finland is a town and a municipality on the banks of the Nokianvirta River in the region of Pirkanmaa, some 15 kilometres west of Tampere.

en.wikipedia.org

AREA	134.27 miles^2 347.77 km^2
POPULATION	32,353

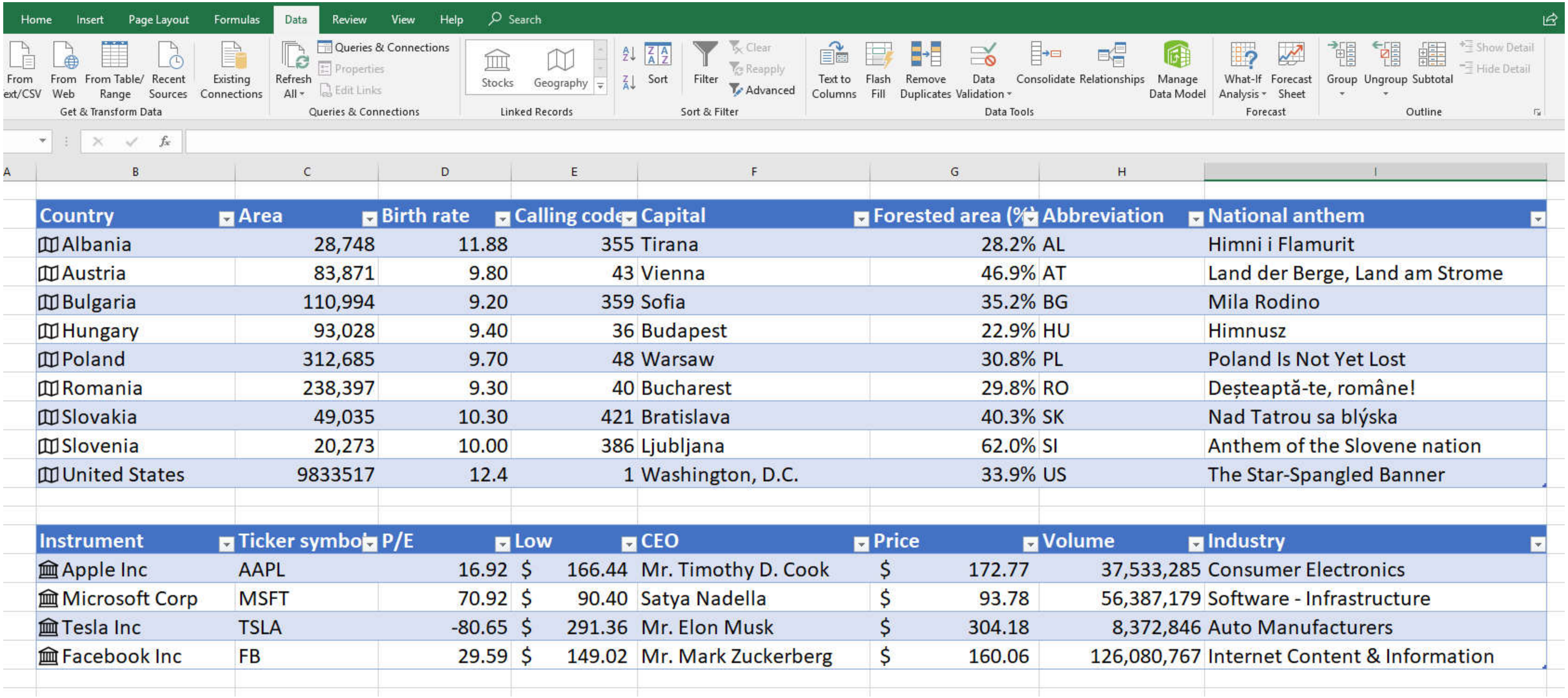
HISTORY

1596	Club War Nokia was the setting of one of the largest battles against Swedish feudal lords
1938	Pohjois-Pirkkala renamed Nokia
1977	Nokia was designated a city

RELATED TOPICS

Entities and Attributes in Spreadsheets

Excel: 2018

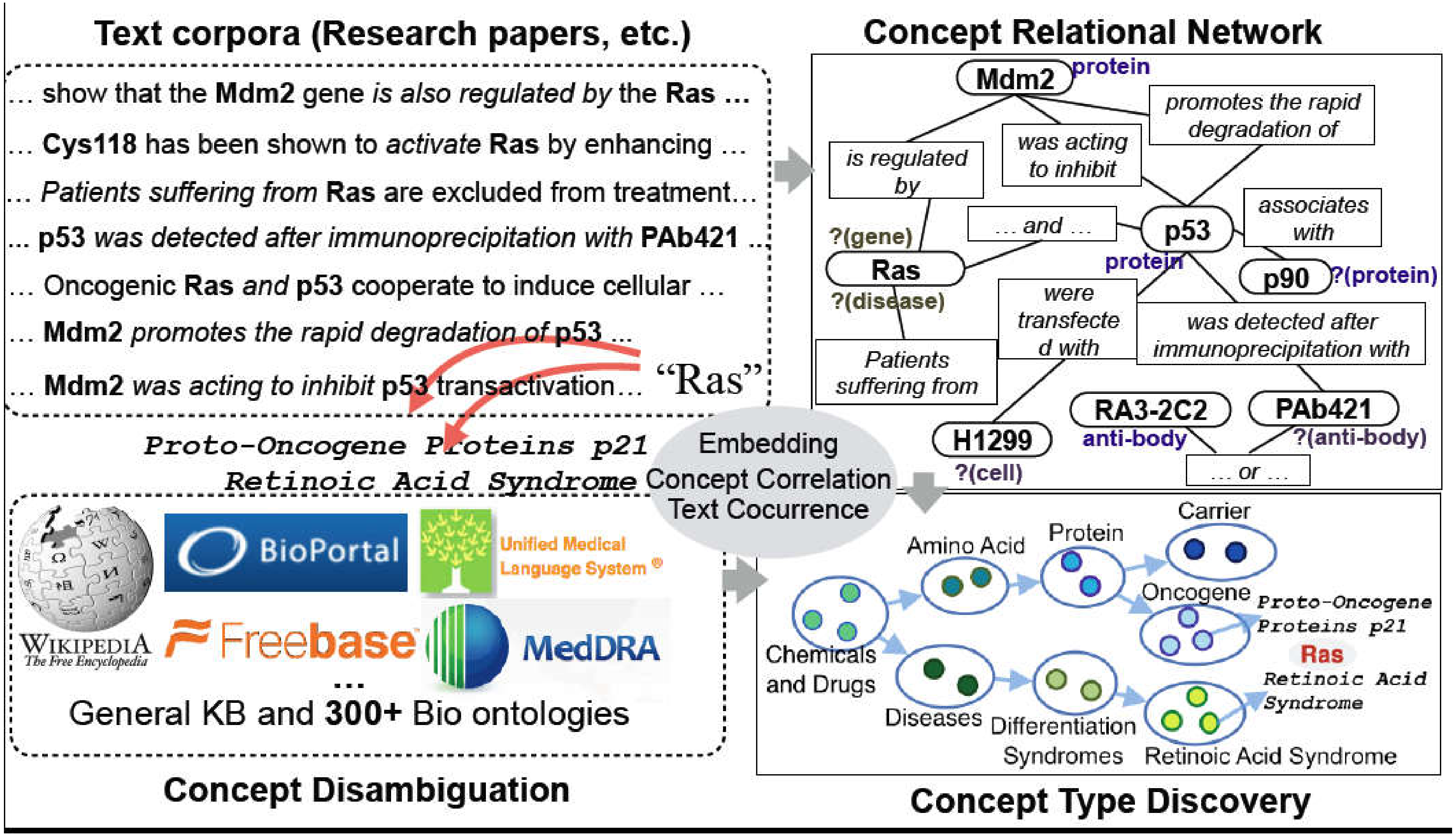


Country	Area	Birth rate	Calling code	Capital	Forested area (%	Abbreviation	National anthem
Albania	28,748	11.88	355	Tirana	28.2%	AL	Himni i Flamurit
Austria	83,871	9.80	43	Vienna	46.9%	AT	Land der Berge, Land am Strome
Bulgaria	110,994	9.20	359	Sofia	35.2%	BG	Mila Rodino
Hungary	93,028	9.40	36	Budapest	22.9%	HU	Himnusz
Poland	312,685	9.70	48	Warsaw	30.8%	PL	Poland Is Not Yet Lost
Romania	238,397	9.30	40	Bucharest	29.8%	RO	Deșteaptă-te, române!
Slovakia	49,035	10.30	421	Bratislava	40.3%	SK	Nad Tatrou sa blýska
Slovenia	20,273	10.00	386	Ljubljana	62.0%	SI	Anthem of the Slovene nation
United States	9833517	12.4	1	Washington, D.C.	33.9%	US	The Star-Spangled Banner

Instrument	Ticker symbo	P/E	Low	CEO	Price	Volume	Industry
Apple Inc	AAPL	16.92	$ 166.44	Mr. Timothy D. Cook	$ 172.77	37,533,285	Consumer Electronics
Microsoft Corp	MSFT	70.92	$ 90.40	Satya Nadella	$ 93.78	56,387,179	Software - Infrastructure
Tesla Inc	TSLA	-80.65	$ 291.36	Mr. Elon Musk	$ 304.18	8,372,846	Auto Manufacturers
Facebook Inc	FB	29.59	$ 149.02	Mr. Mark Zuckerberg	$ 160.06	126,080,767	Internet Content & Information

Microsoft

Biomedical EDL



Text corpora (Research papers, etc.)

… show that the **Mdm2** gene *is also regulated by* the **Ras** …

… **Cys118** has been shown to *activate* **Ras** by enhancing …

… *Patients suffering from* **Ras** are excluded from treatment…

... **p53** *was detected after immunoprecipitation with* **PAb421** ...

… Oncogenic **Ras** *and* **p53** cooperate to induce cellular …

… **Mdm2** *promotes the rapid degradation of* **p53** ...

… **Mdm2** *was acting to inhibit* **p53** transactivation… "Ras"

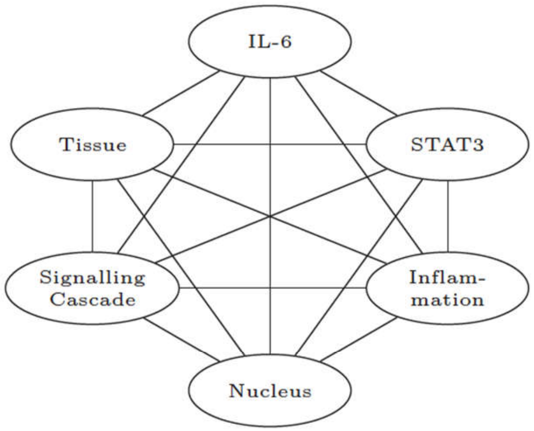
General KB and **300+** Bio ontologies

Concept Disambiguation

Concept Type Discovery

Linking to 300+ Biomedical Ontologies: Biomedical Domain

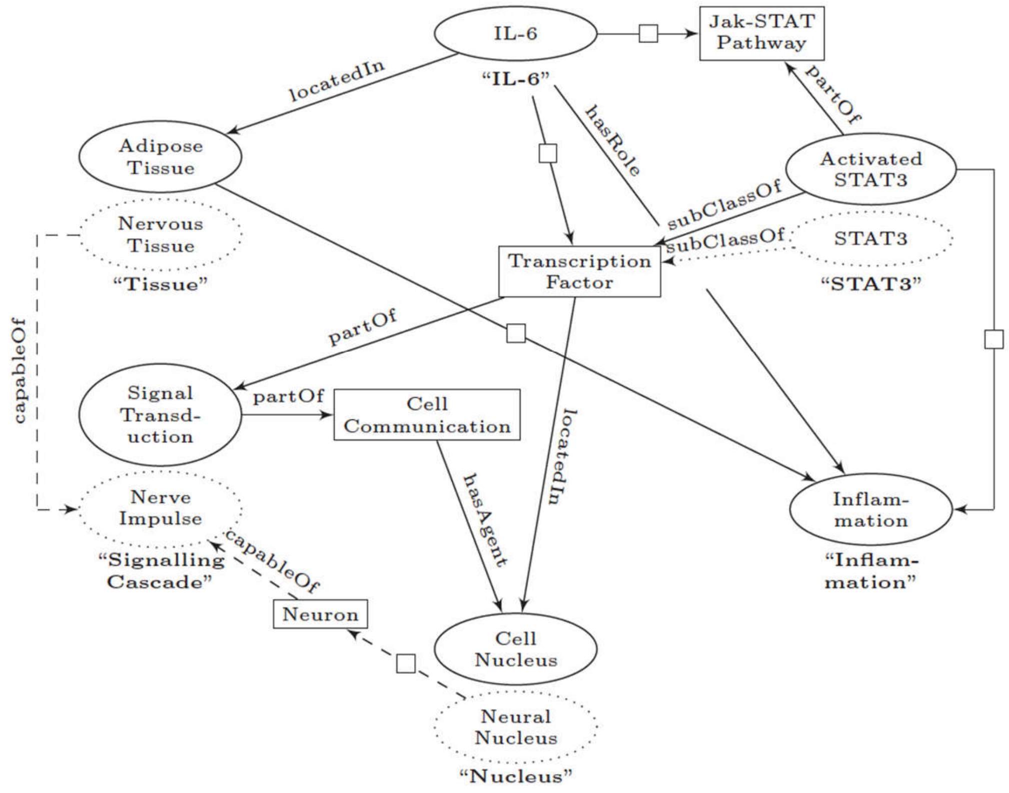
Document Subgraph



Example Sentence

The binding of **IL-6** stimulates a **signalling cascade** where activated **STAT3** enters the **nucleus**, ultimately increasing **inflammation** in the surrounding **tissue**.

Knowledge Subgraph



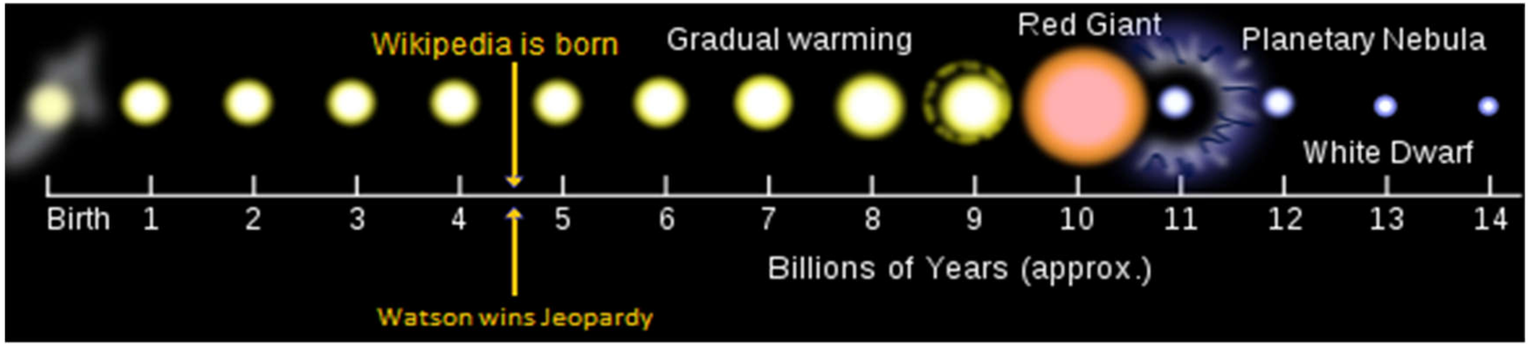
Biomedical Paper Writing

- Input Title: Remote Lifestyle Counseling Influences Cardiovascular Health Outcomes in Youth with Overweight or Obesity and Congenital_Heart_Disease
- Predicted Terms: Children; heart_disease; obesity; CHD
- Output:
- Background: Children with CHD are associated with increased morbidity and mortality in children with obesity. However, the impact of lifestyle interventions on health outcomes in children with CHD is scarce. We aimed to examine the impact of lifestyle modification on health outcomes in children with CHD. Methods: A total of 60 children aged 18 years and older were recruited from the national health care program. The primary outcome measures were the risk of obesity and CHD. The prevalence of CHD was estimated using the Short Form Health Survey (Physical Activity Survey). Results: The prevalence of CHD was estimated by using the bivariate model of the Health and Nutrition Examination Survey (“Short-term home”) .

Where have we been? State-of-the-art

- We’re thriving
  - ○ Entity Linking: 90% accuracy for English news
- We’re making slow but consistent progress: 30%-65% F-score
  - ○ Entity Coreference Resolution
- We’re running around in circles
  - ○ Name Tagging: 90% for English news but drop dramatically to any surprise domain, language, genre

Where We Are, and A Brief History of Entity Work



- Game changer: Wikipedia + hardware advancements

  small number of entity classes → large entity collection

- Current state: good when there is a lot of annotated data

  Within domain: NER performance for coarse entity type [PER/LOC/ORG/MISC] is good (low 90-ies F1)

  Performance drops significantly across domains and for low resource languages.

- Next game changer: deep domain/contextual analysis

  move from one large entity collection to many thousands of domain specific collections, business collections, and personal collections

Outline

- Motivation, Task, Application 30 min
- Traditional “Ancient” Approaches 15 min
- Modern Approaches
  - ○ Language Universal EDL 15 min
  - ○ Multi-lingual Common Space Construction 30 min

Coffee Break

  - ○ Cross-lingual Transfer Learning 20 min
  - ○ Cross-lingual Neural Entity Linking 20 min
- Remaining Challenges and New Directions 30 min
- Demos, Resources and QA 20 min

Entity Discovery

- Also called Mention detection or Named Entity Recognition

- Task :
  - ○ Identify the boundaries of mentions (of entities) in a document
  - ○ Predict the types of the extracted mentions
    - E.g. PER, ORG, LOC
  - ○ A lot of work in this area.
    - Works well when we have supervised data
    - Does not work well across domains – still a challenge
  - ○ We'll talk about it more in the multilingual context

EDL Subtasks

- Identifying Target Mentions
  - Mentions in the input text that should be Wikified
- Name Translation
  - Translate mentions to English
- Identifying Candidate Titles
  - Candidate Wikipedia titles that could correspond to each mention
- Candidate Title Ranking
  - Rank the candidate titles for a given mention
- NIL Detection and Clustering
  - Identify mentions that do not correspond to a Wikipedia title
  - Entity Linking: cluster NIL mentions that represent the same entity

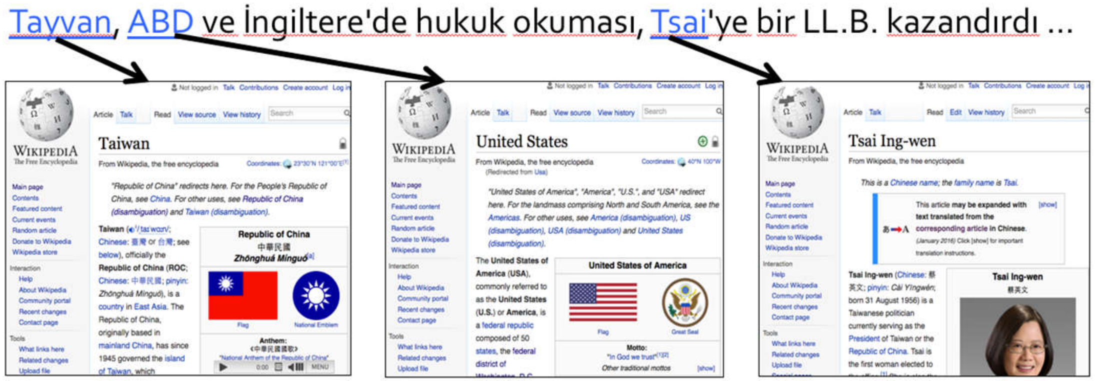
“Old” Days Name Tagging: Supervised Learning with Hand-crafted Features

- Typical Name Tagging Features:
- N-gram: Unigram, bigram and trigram token sequences in the context window of the current token
- Part-of-Speech: POS tags of the context words
- Gazetteers: person names, organizations, countries and cities, titles, idioms, etc.
- Word clusters: to reduce sparsity, using word clusters such as Brown clusters (Brown et al., 1992)
- Case and Shape: Capitalization and morphology analysis based features
- Chunking: NP and VP Chunking tags
- Global feature: Sentence level and document level features. For example, whether the token is in the first sentence of a document
- Conjunction: Conjunctions of various features

- (Ji and Grishman, 2006; Li et al., 2013)

“Traditional” Cross-lingual EDL techniques

- Match context from non-English docs with English Wikipedia



Tayvan, ABD ve İngiltere'de hukuk okuması, Tsai'ye bir LL.B. kazandırdı ...

Taiwan

From Wikipedia, the free encyclopedia Coordinates: 23°30′N 121°00′E

"Republic of China" redirects here. For the People's Republic of China, see China. For other uses, see Republic of China (disambiguation) and Taiwan (disambiguation).

Taiwan (/taɪˈwɑːn/; Chinese: 臺灣 or 台灣; see below), officially the **Republic of China** (ROC; Chinese: 中華民國; pinyin: *Zhōnghuá Mínguó*), is a country in East Asia. The Republic of China, originally based in mainland China, has since 1945 governed the island of Taiwan, which

Republic of China
中華民國
Zhōnghuá Mínguó[a]

United States

From Wikipedia, the free encyclopedia Coordinates: 40°N 100°W
(Redirected from Usa)

"United States of America", "America", "U.S.", and "USA" redirect here. For the landmass comprising North and South America, see the Americas. For other uses, see America (disambiguation), US (disambiguation), USA (disambiguation) and United States (disambiguation).

The **United States of America** (**USA**), commonly referred to as the **United States** (**U.S.**) or **America**, is a federal republic composed of 50 states, the federal district of Washington, D.C.

United States of America

Motto:
"In God we trust"[1][2]
Other traditional mottos [show]

Tsai Ing-wen

From Wikipedia, the free encyclopedia

This is a Chinese name; the family name is Tsai.

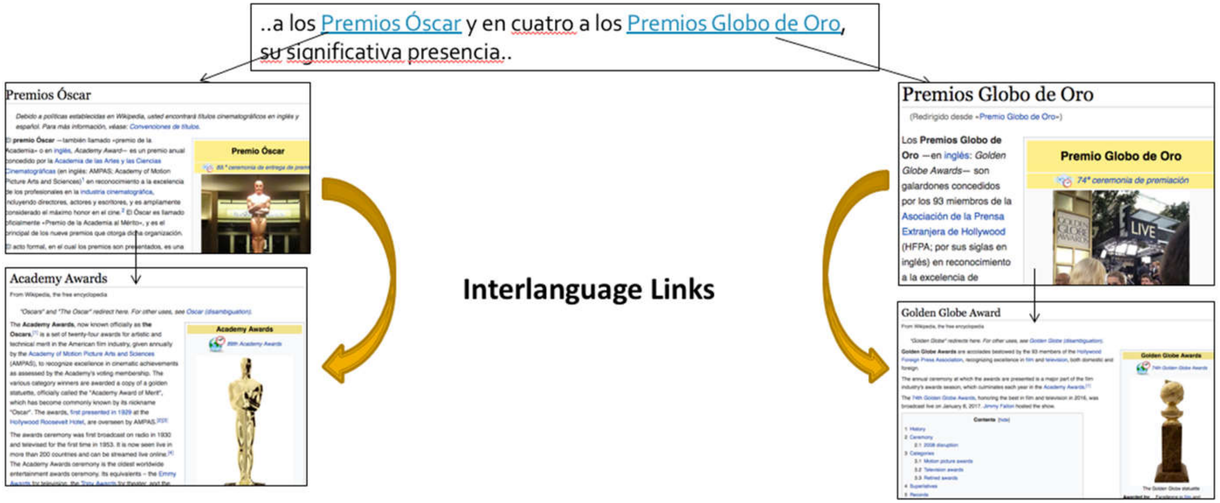
あ→A This article **may be expanded with text translated from the corresponding article in Chinese.** *(January 2016)* Click [show] for important translation instructions. [show]

Tsai Ing-wen (Chinese: 蔡英文; pinyin: *Cài Yīngwén*; born 31 August 1956) is a Taiwanese politician currently serving as the President of Taiwan or the Republic of China. Tsai is the first woman elected to

Tsai Ing-wen
蔡英文

- Use SOTA MT systems

- Use inter-language links



..a los Premios Óscar y en cuatro a los Premios Globo de Oro, su significativa presencia..

Premios Óscar

Debido a políticas establecidas en Wikipedia, usted encontrará títulos cinematográficos en inglés y español. Para más información, véase: Convenciones de títulos.

El **premio Óscar** —también llamado «premio de la Academia» o en inglés, *Academy Award*— es un premio anual concedido por la Academia de las Artes y las Ciencias Cinematográficas (en inglés: AMPAS; Academy of Motion Picture Arts and Sciences)[1] en reconocimiento a la excelencia de los profesionales en la industria cinematográfica, incluyendo directores, actores y escritores, y es ampliamente considerado el máximo honor en el cine.[2] El Óscar es llamado oficialmente «Premio de la Academia al Mérito», y es el principal de los nueve premios que otorga dicha organización.

El acto formal, en el cual los premios son presentados, es una

Premio Óscar
89.ª ceremonia de entrega de premios

Academy Awards

From Wikipedia, the free encyclopedia

"Oscars" and "The Oscar" redirect here. For other uses, see Oscar (disambiguation).

The **Academy Awards**, now known officially as **the Oscars**,[1] is a set of twenty-four awards for artistic and technical merit in the American film industry, given annually by the Academy of Motion Picture Arts and Sciences (AMPAS), to recognize excellence in cinematic achievements as assessed by the Academy's voting membership. The various category winners are awarded a copy of a golden statuette, officially called the "Academy Award of Merit", which has become commonly known by its nickname "Oscar". The awards, first presented in 1929 at the Hollywood Roosevelt Hotel, are overseen by AMPAS.[2][3]

The awards ceremony was first broadcast on radio in 1930 and televised for the first time in 1953. It is now seen live in more than 200 countries and can be streamed live online.[4] The Academy Awards ceremony is the oldest worldwide entertainment awards ceremony. Its equivalents – the Emmy Awards for television, the Tony Awards for theater, and the

Academy Awards
89th Academy Awards

Interlanguage Links

Premios Globo de Oro

(Redirigido desde «Premio Globo de Oro»)

Los **Premios Globo de Oro** —en inglés: *Golden Globe Awards*— son galardones concedidos por los 93 miembros de la Asociación de la Prensa Extranjera de Hollywood (HFPA; por sus siglas en inglés) en reconocimiento a la excelencia de

Premio Globo de Oro
74° ceremonia de premiación

Golden Globe Award

From Wikipedia, the free encyclopedia

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

Golden Globe Awards are accolades bestowed by the 93 members of the Hollywood Foreign Press Association, recognizing excellence in film and television, both domestic and foreign.

The annual ceremony at which the awards are presented is a major part of the film industry's awards season, which culminates each year in the Academy Awards.[1]

The 74th Golden Globe Awards, honoring the best in film and television in 2016, was broadcast live on January 8, 2017. Jimmy Fallon hosted the show.

Contents [hide]

1 History
2 Ceremony
  2.1 2008 disruption
3 Categories
  3.1 Motion picture awards
  3.2 Television awards
  3.3 Retired awards
4 Superlatives
5 Records

Golden Globe Awards
74th Golden Globe Awards

The Golden Globe statuette
Awarded for Excellence in film and

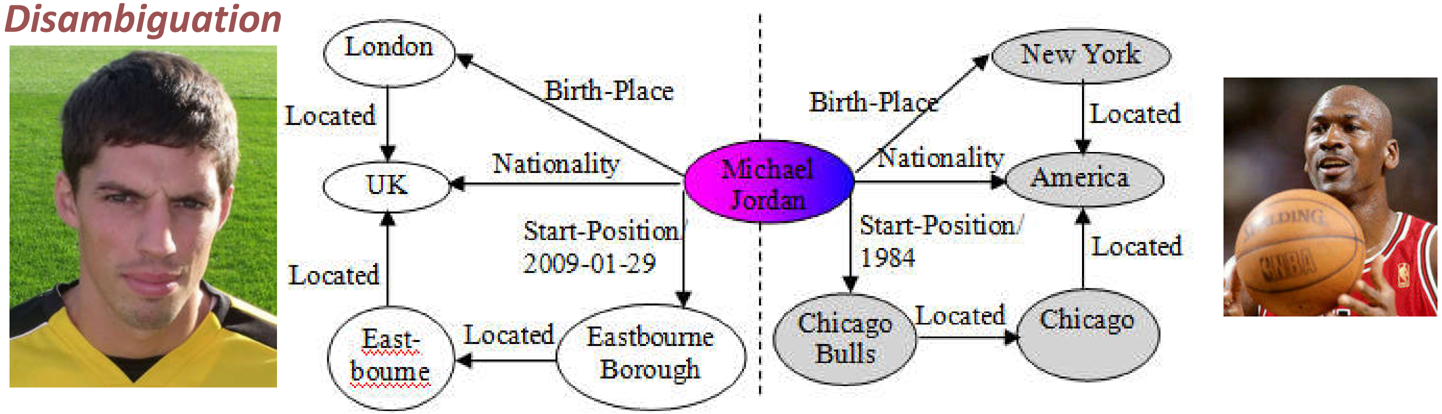
Typical Features for Entity Linking

Mention/Concept Attribute		Description
Name	Spelling match	Exact string match, acronym match, alias match, string matching…
	KB link mining	Name pairs mined from KB text redirect and disambiguation pages
	Name Gazetteer	Organization and geo-political entity abbreviation gazetteers
Document surface	Lexical	Words in KB facts, KB text, mention name, mention text.
		Tf.idf of words and ngrams
	Position	Mention name appears early in KB text
	Genre	Genre of the mention text (newswire, blog, …)
	Local Context	Lexical and part-of-speech tags of context words
Entity Context	Type	Mention concept type, subtype
	Relation/Event	Concepts co-occurred, attributes/relations/events with mention
	Coreference	Co-reference links between the source document and the KB text
Profiling		Slot fills of the mention, concept attributes stored in KB infobox
Concept		Ontology extracted from KB text
Topic		Topics (identity and lexical similarity) for the mention text and KB text
KB Link Mining		Attributes extracted from hyperlink graphs of the KB text
Popularity	Web	Top KB text ranked by search engine and its length
	Frequency	Frequency in KB texts

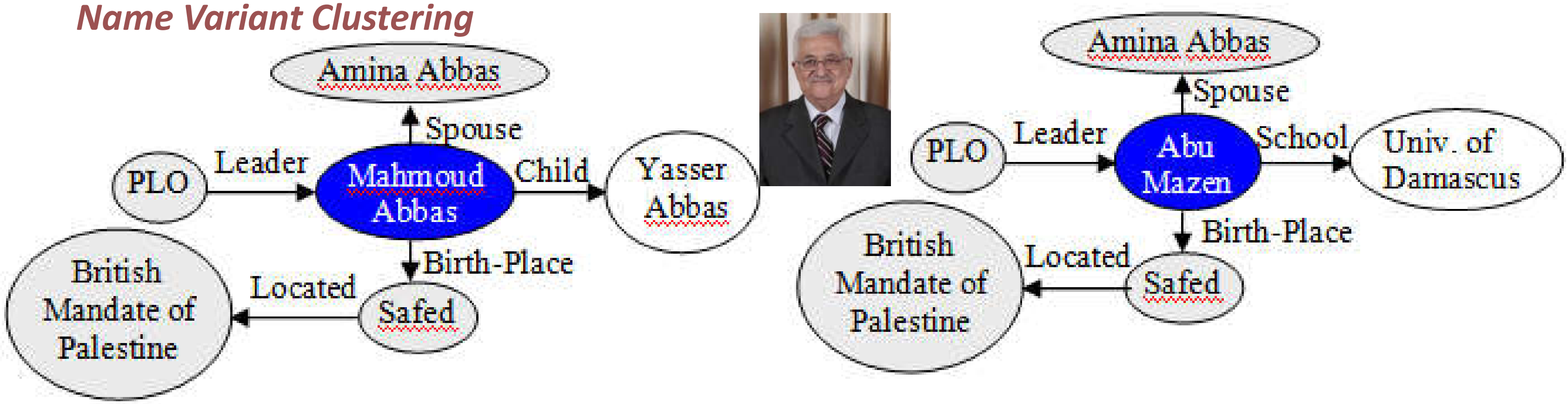
- (Ji et al., 2011; Zheng et al., 2010; Dredze et al., 2010; Anastacio et al., 2011)

Explicit Entity Profiling

Disambiguation



Name Variant Clustering

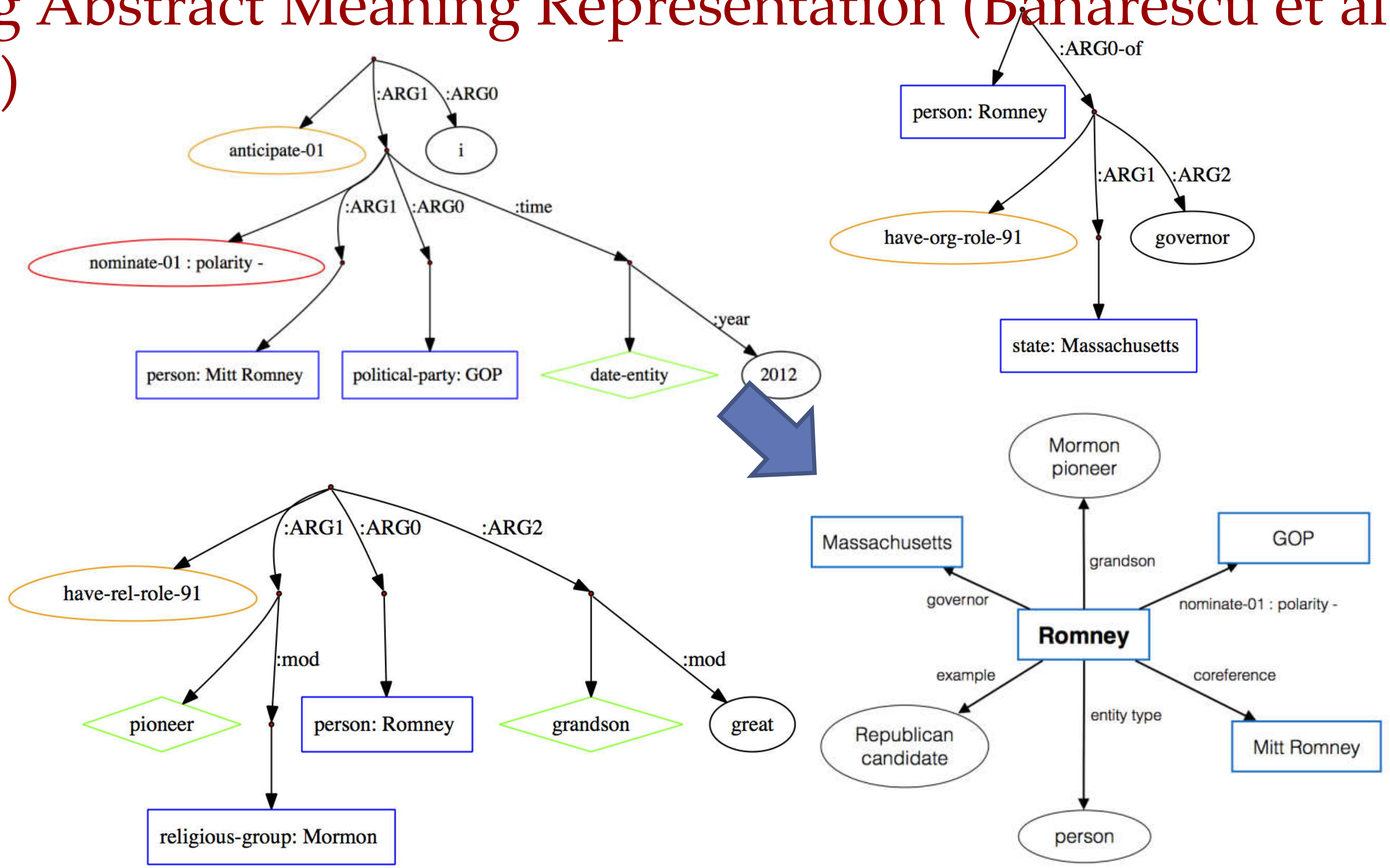


- Deep semantic context exploration and indicative context selection (Gao et al., 2010; Chen et al., 2010; Chen and Ji, 2011; Cassidy et al., 2012)
- Exploit name tagging, Wikipedia infoboxes, synonyms, variants and abbreviations, slot filling results and semantic categories

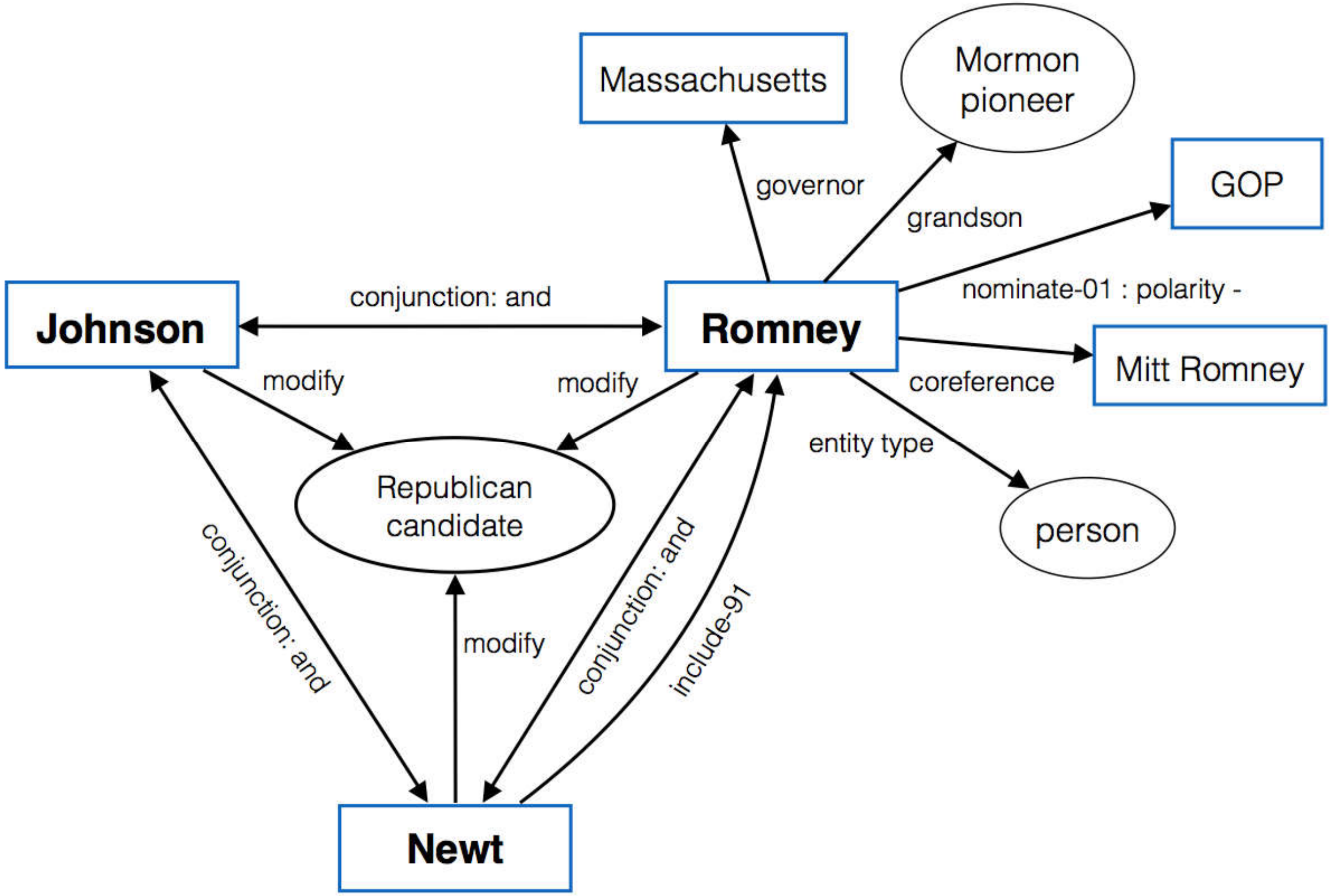
Unsupervised Entity Linking: Source Text Example

- I am cautiously anticipating the **GOP** nominee in 2012 not to be **Mitt Romney.**
- When **Romney** was the Governor of **Massachusetts**, he helped develop health care law. I appreciate his work.
- I think **Newt** is going to be the last of the "Not **Romneys**".
- **Romney** is the great-great-grandson of a **Mormon pioneer**, from being a Mormon to having taken so many positions in the past that annoy conservatives.
- I don't think **Republican candidates** like **Romney**, **Newt**, and **Johnson** have a real chance for the election.

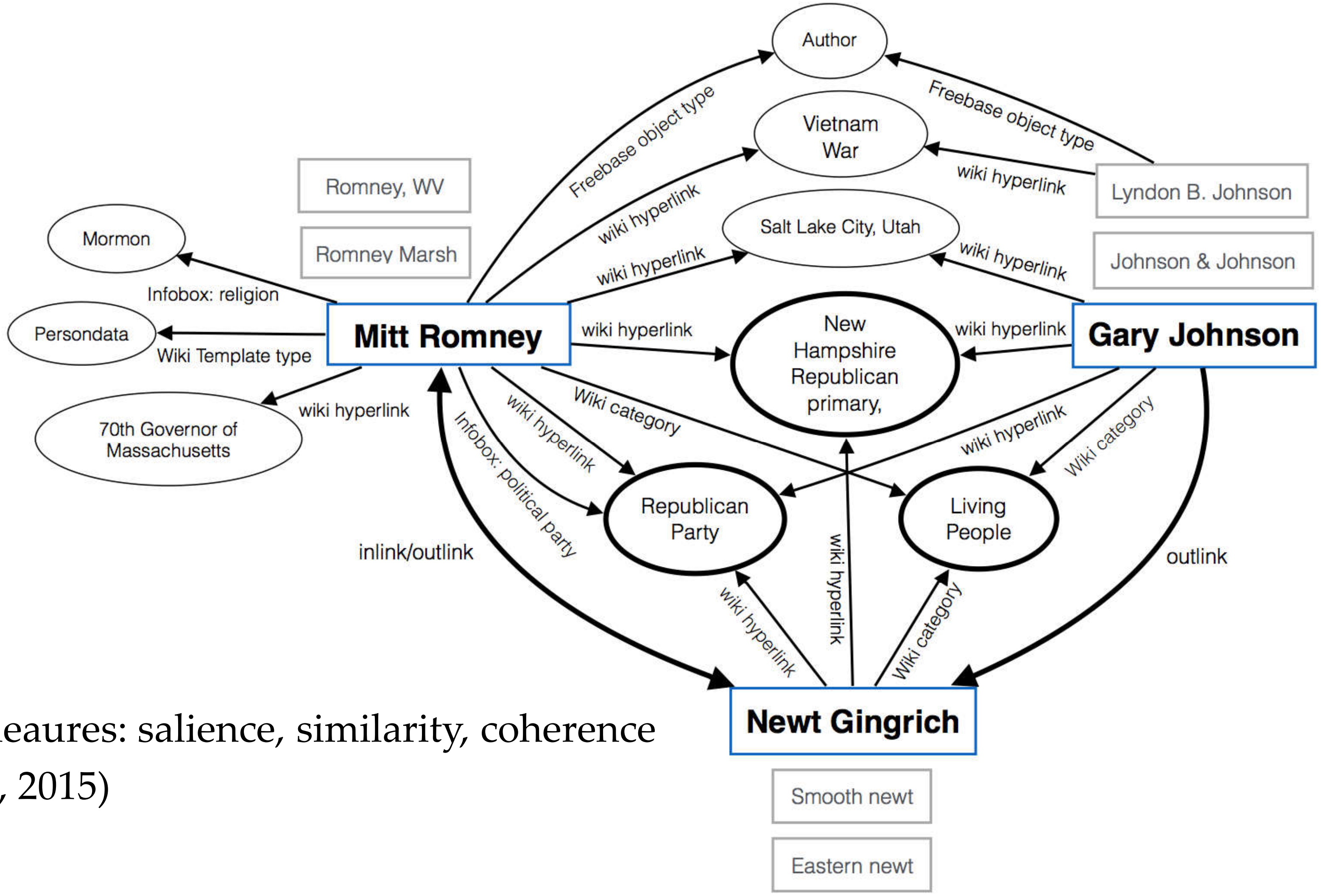
Or Build a Knowledge Graph for Mentions using Abstract Meaning Representation (Banarescu et al., 2013))



Construct Knowledge Graph of Concept Mentions and their Collaborators



Construct Corresponding Knowledge Graph of Concept Candidates and their Collaborators

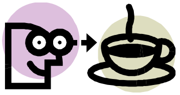


- Typical meaures: salience, similarity, coherence (Pan et al., 2015)

Outline

- Motivation, Task, Application 30 min
- Traditional “Ancient” Approaches 15 min
- Modern Approaches
  - ○ Language Universal EDL 15 min
  - ○ Multi-lingual Common Space Construction 30 min

Coffee Break



  - ○ Cross-lingual Transfer Learning 20 min
  - ○ Cross-lingual Neural Entity Linking 20 min
- Remaining Challenges and New Directions 30 min
- Demos, Resources and QA 20 min

1. Identifying Named Entity Mentions

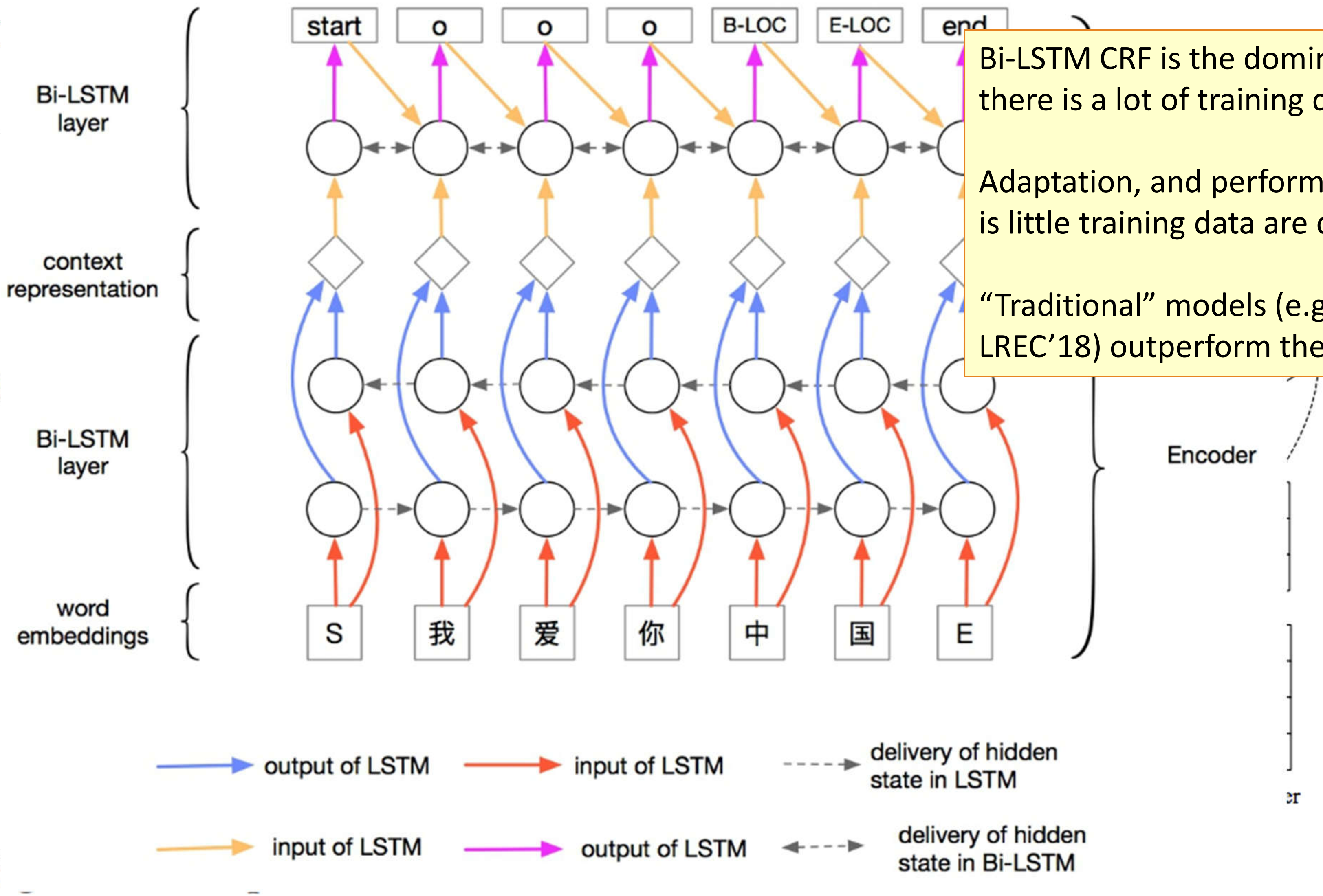
சிஐஏ இயக்குநர் மைக் பாம்பேயோ நியமனத்துக்கு அமெரிக்க செனட் சபை ஒப்புதல். ஆனால், சிஐஏ முகமை மற்றும் அமெரிக்க அதிபர் டிரம்ப் இடையே ஒரு பயனுள்ள அலுவல் ரீதியான உறவினை உருவாக்குவதே மைக் பாம்பேயோவின் உடனடி பணியாக இருக்கும்.

↓

சிஐஏ இயக்குநர் **மைக் பாம்பேயோ** நியமனத்துக்கு அமெரிக்க **செனட்** சபை ஒப்புதல். ஆனால், **சிஐஏ** முகமை மற்றும் **அமெரிக்க** அதிபர் **டிரம்ப்** இடையே ஒரு பயனுள்ள அலுவல் ரீதியான உறவினை உருவாக்குவதே **மைக் பாம்பேயோவின்** உடனடி பணியாக இருக்கும்.

Neu

- BLS

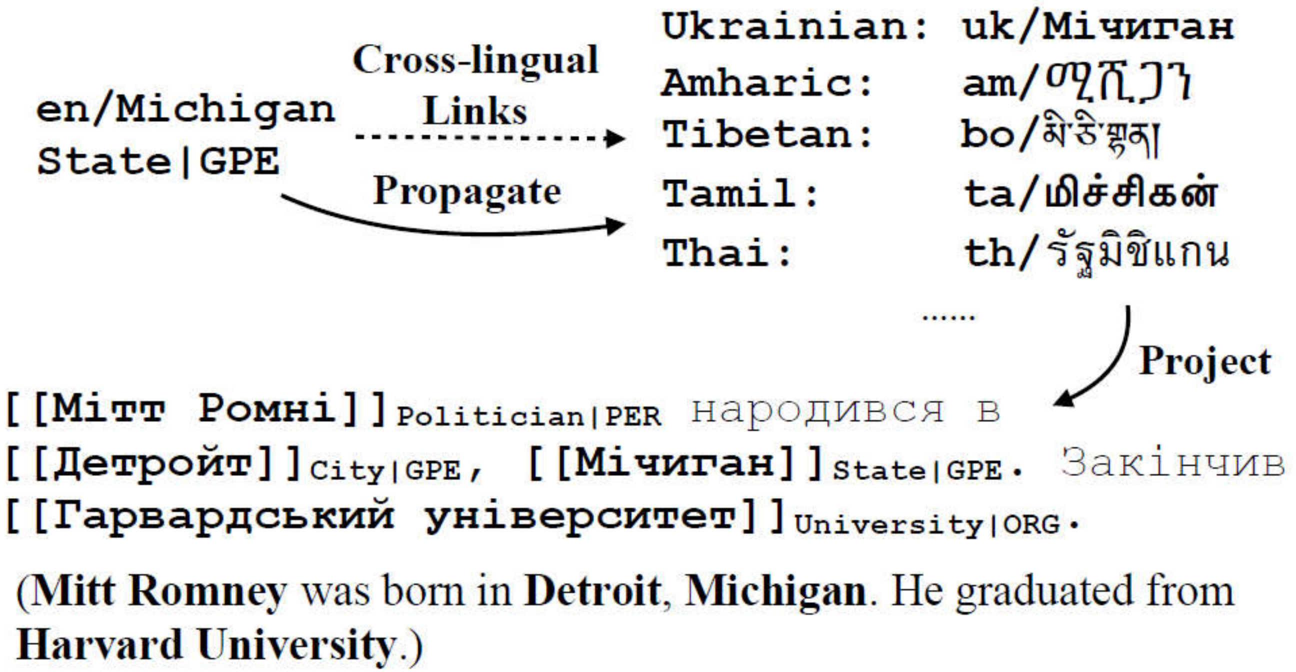


Bi-LSTM CRF is the dominant model when there is a lot of training data.

Adaptation, and performance when there is little training data are questionable.

“Traditional” models (e.g., Illinois NER; LREC’18) outperform them in these cases.

Obtain Silver-Standard Training Data from Wikipedia Markups



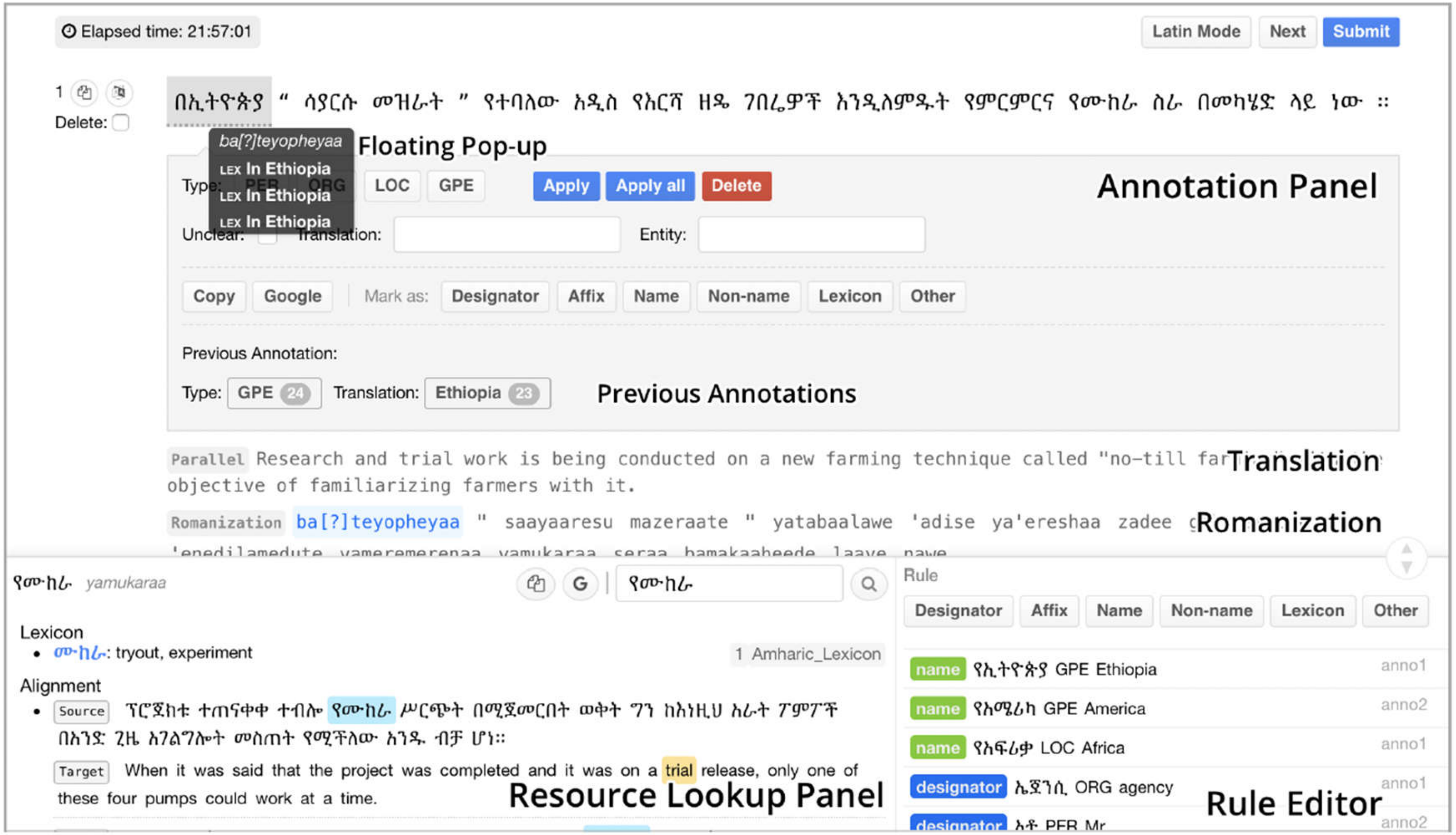
Ukrainian:	uk/Мічиган
Amharic:	am/ሚሺጋን
Tibetan:	bo/མི་ཅི་ཀྲན།
Tamil:	ta/மிச்சிகன்
Thai:	th/รัฐมิชิแกน

[[Мітт Ромні]]Politician|PER народився в [[Детройт]]City|GPE, [[Мічиган]]State|GPE. Закінчив [[Гарвардський університет]]University|ORG.

(**Mitt Romney** was born in **Detroit**, **Michigan**. He graduated from **Harvard University**.)

- Derive “silver-standard” training data automatically from Wikipedia markups and apply self-training
- But DNN is very sensitive to noise...

Obtain Silver-Standard Training Data from Chinese Room



በኢትዮጵያ “ ሳያርሱ መዝራት ” የተባለው አዲስ የአርሻ ዘዴ ገበሬዎች እንዲለምዱት የምርምርና የሙከራ ስራ በመካሄድ ላይ ነው ።

Parallel Research and trial work is being conducted on a new farming technique called "no-till far objective of familiarizing farmers with it.

Romanization ba[?]teyopheyaa " saayaaresu mazeraate " yatabaalawe 'adise ya'ereshaa zadee ç 'enedilamedute yameremerenaa yamukaraa seraa bamakaaheede laaye nawe

የሙከራ yamukaraa

የሙከራ

Lexicon

- ሙከራ: tryout, experiment 1 Amharic_Lexicon

Alignment

- Source ፕሮጀክቱ ተጠናቀቀ ተብሎ የሙከራ ሥርጭት በሚጀመርበት ወቅት ግን ከእነዚህ አራት ፖምፖች በአንድ ጊዜ አገልግሎት መስጠት የሚችለው አንዱ ብቻ ሆነ።

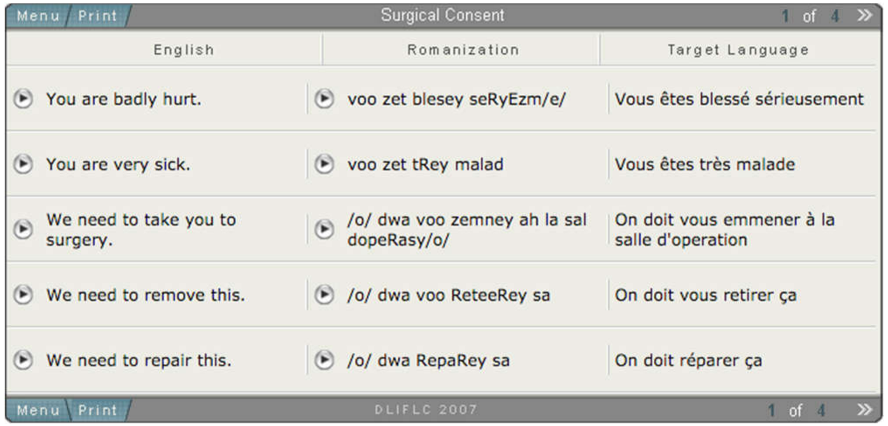
  Target When it was said that the project was completed and it was on a trial release, only one of these four pumps could work at a time.

name	የኢትዮጵያ GPE Ethiopia		anno1
name	የአሜሪካ GPE America		anno2
name	የአፍሪቃ LOC Africa		anno1
designator	ኤጀንሲ ORG agency		anno1
designator	አቶ PER Mr		anno2

- (Lin et al., ACL2018 Best Demo Nomination)

Exploit Non-traditional Universal Linguistic Resources

- Grammar books from Lori Levin’s bookshelf and CIA Names from DARPA PM’s bookshelf
- Unicode Common Locale Data Repository, Wikitionary, Panlex, Multilingual WordNet, GeoNames, JRC Names, phrase pairs mined from Wikipedia
- Phrase Books from Language Survival Kits and Elicitation Corpus

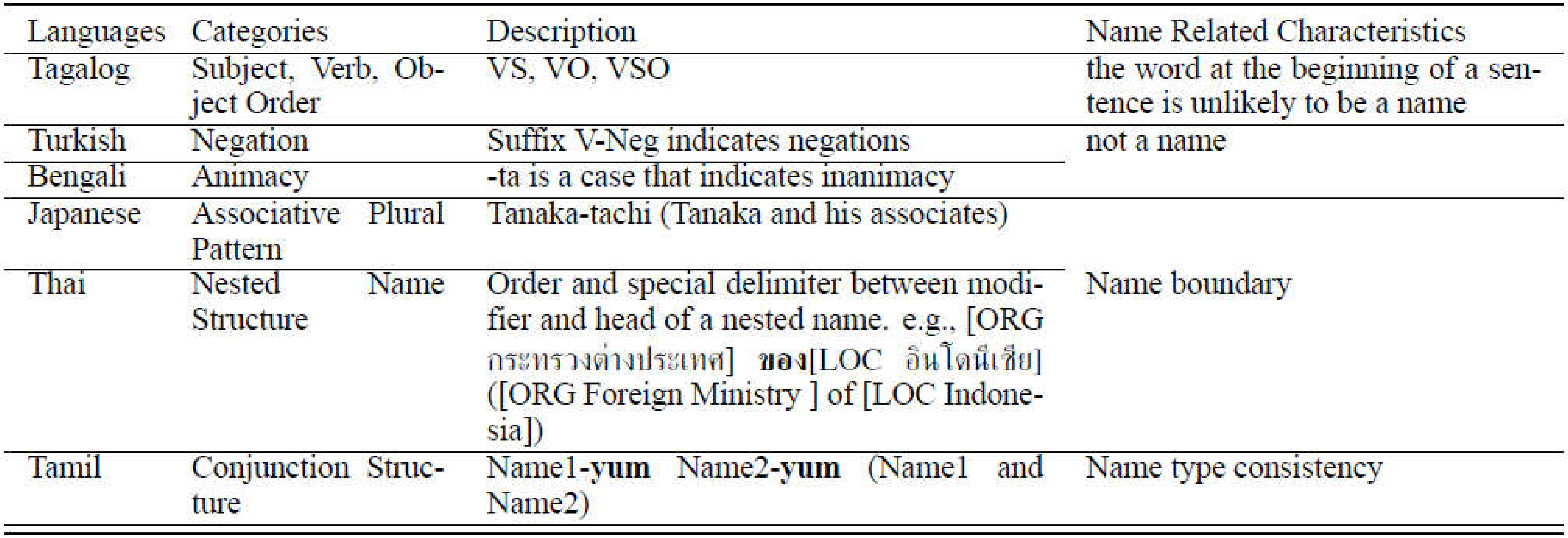


Surgical Consent

English	Romanization	Target Language
You are badly hurt.	voo zet blesey seRyEzm/e/	Vous êtes blessé sérieusement
You are very sick.	voo zet tRey malad	Vous êtes très malade
We need to take you to surgery.	/o/ dwa voo zemney ah la sal dopeRasy/o/	On doit vous emmener à la salle d'operation
We need to remove this.	/o/ dwa voo ReteeRey sa	On doit vous retirer ça
We need to repair this.	/o/ dwa RepaRey sa	On doit réparer ça

DALRYMPLE AND NIKOLAEVA Objects and Information Structure CAMBRIDGE
Comrie Tense Cambridge
Comrie Aspect
Kittilä et al. Case, Animacy and Semantic Roles
INDEFINITE PRONOUNS
Guéron · Lecarme Eds. Time and Modality
De Hoop · De Swart Eds. Differential Subject Marking
Malchukov & Spencer The Oxford Handbook of CASE OXFORD
Lyons Definiteness
MASSAM Count and Mass Across Languages OXFORD
THE EXPRESSION OF TIME
Comorovski von Heusinger Eds. Existence: Semantics and Syntax
THE EXPRESSION OF MODALITY
THE EXPRESSION OF NEGATION
THE EXPRESSION OF POSSESSION

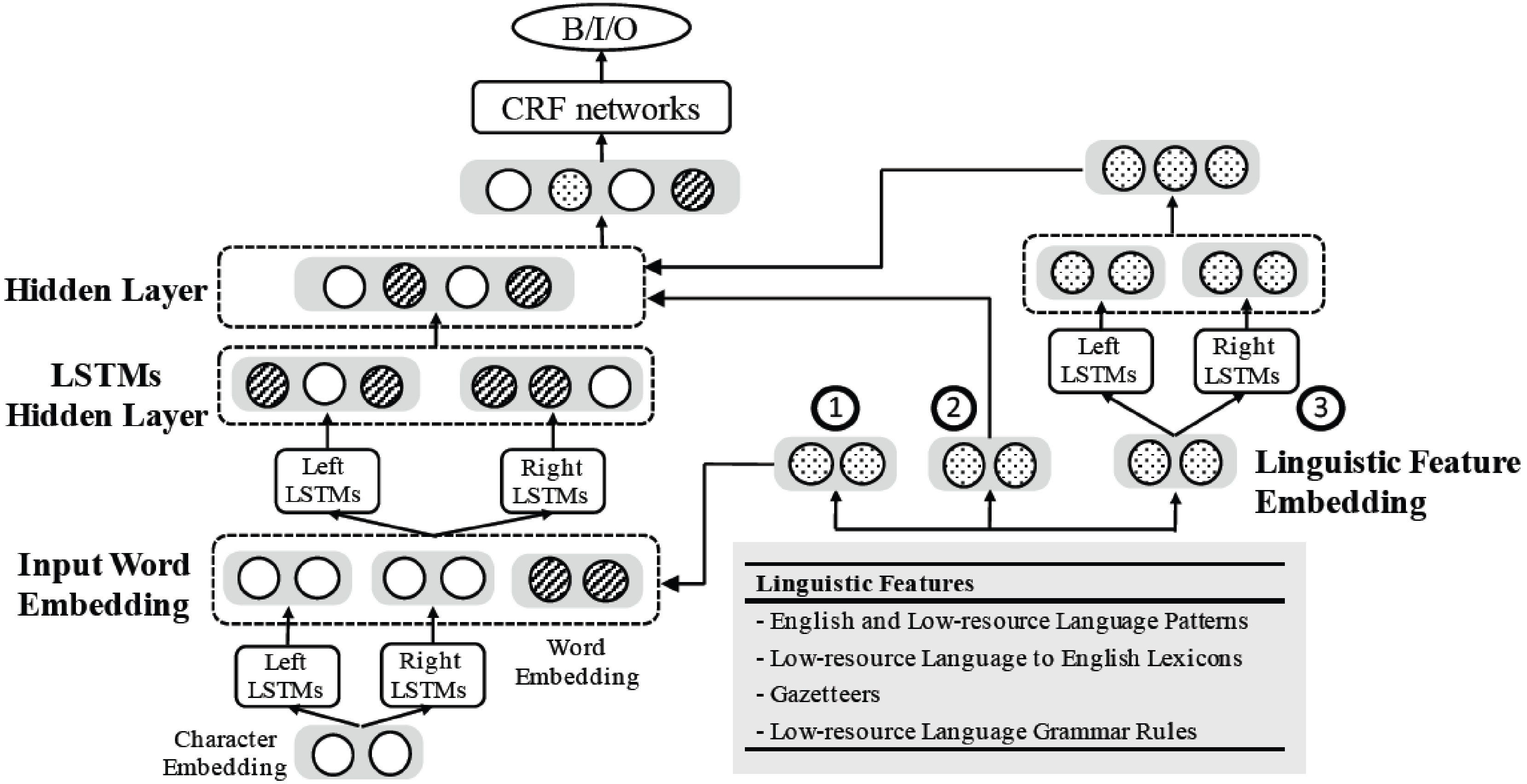
Linguistic Structure from WALS database and Syntactic Structures of the World's Languages



Languages	Categories	Description	Name Related Characteristics
Tagalog	Subject, Verb, Object Order	VS, VO, VSO	the word at the beginning of a sentence is unlikely to be a name
Turkish	Negation	Suffix V-Neg indicates negations	not a name
Bengali	Animacy	-ta is a case that indicates inanimacy	
Japanese	Associative Plural Pattern	Tanaka-tachi (Tanaka and his associates)	
Thai	Nested Name Structure	Order and special delimiter between modifier and head of a nested name. e.g., [ORG กระทรวงต่างประเทศ] **ของ**[LOC อินโดนีเซีย] ([ORG Foreign Ministry] of [LOC Indonesia])	Name boundary
Tamil	Conjunction Structure	Name1-**yum** Name2-**yum** (Name1 and Name2)	Name type consistency

- Universal Morphology Analyzer based on Wikipedia Markups
  - ○ *Kıta Fransası, güneyde [[Akdeniz]]**den** kuzeyde [[Manş Denizi]]**ve** [[Kuzey Denizi]]**ne**, doğuda [[Ren Nehri]]**nden** batıda [[Atlas Okyanusu]]**na** kadar yayılan topraklarda yer alır. (Continental France is located in the south [[Mediterranean Sea]] in the north [[English Sea]] and [[North Sea]] in the east [[Rhine River]] to the west [[Atlantic Ocean]].)*

Feed Non-traditional Linguistic Resources into DNN

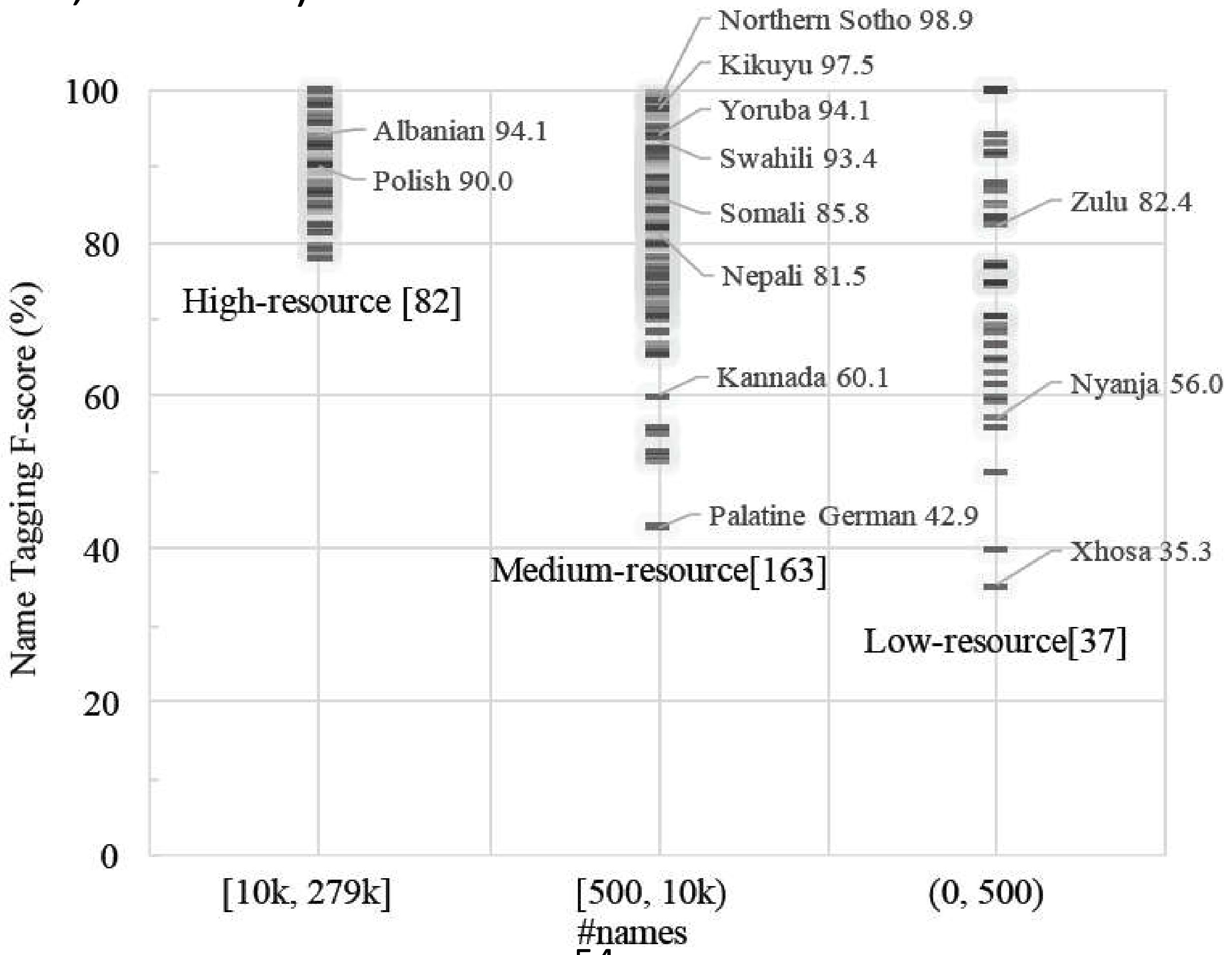


Linguistic Features

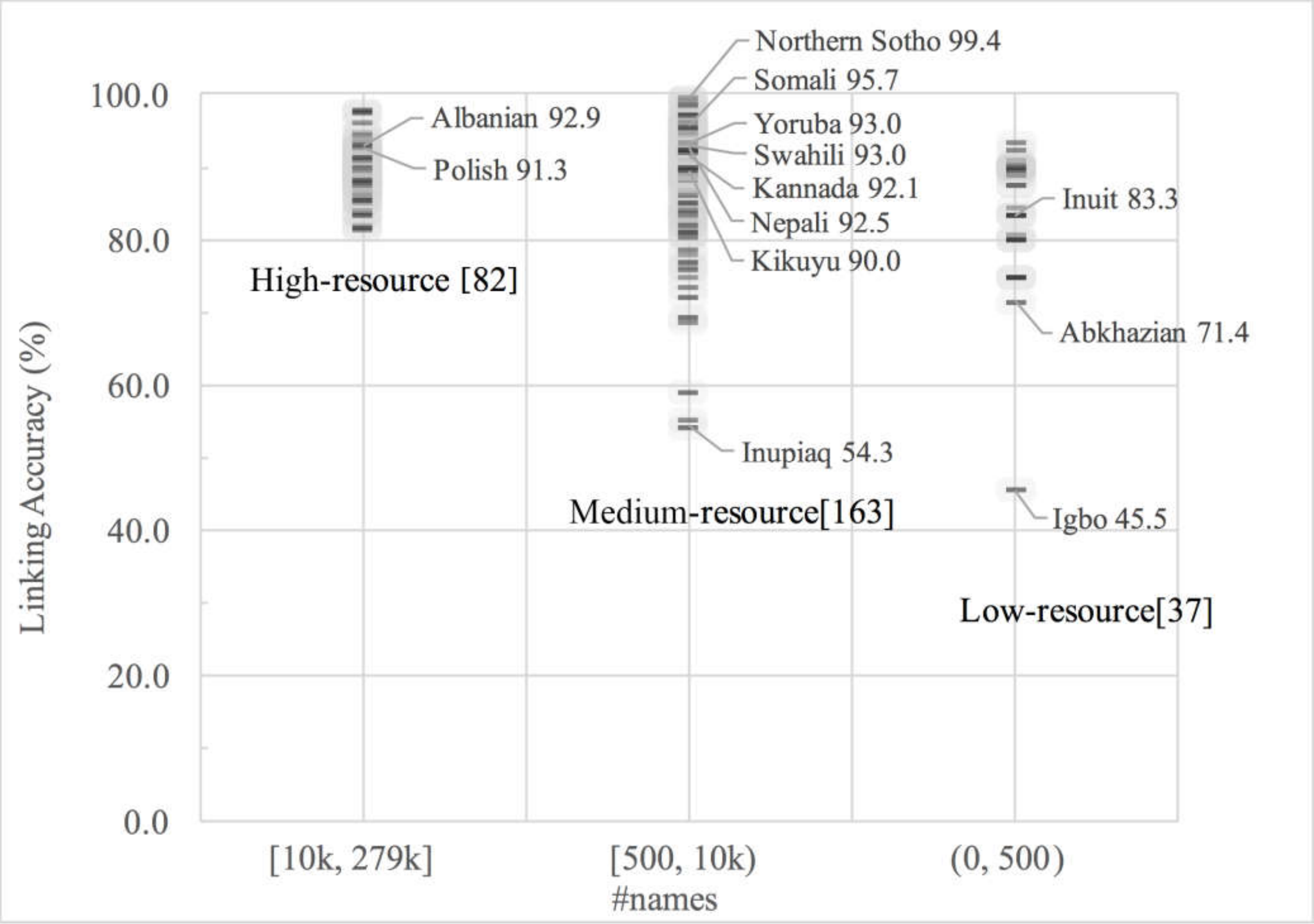
- English and Low-resource Language Patterns
- Low-resource Language to English Lexicons
- Gazetteers
- Low-resource Language Grammar Rules

Overall Name Tagging Results

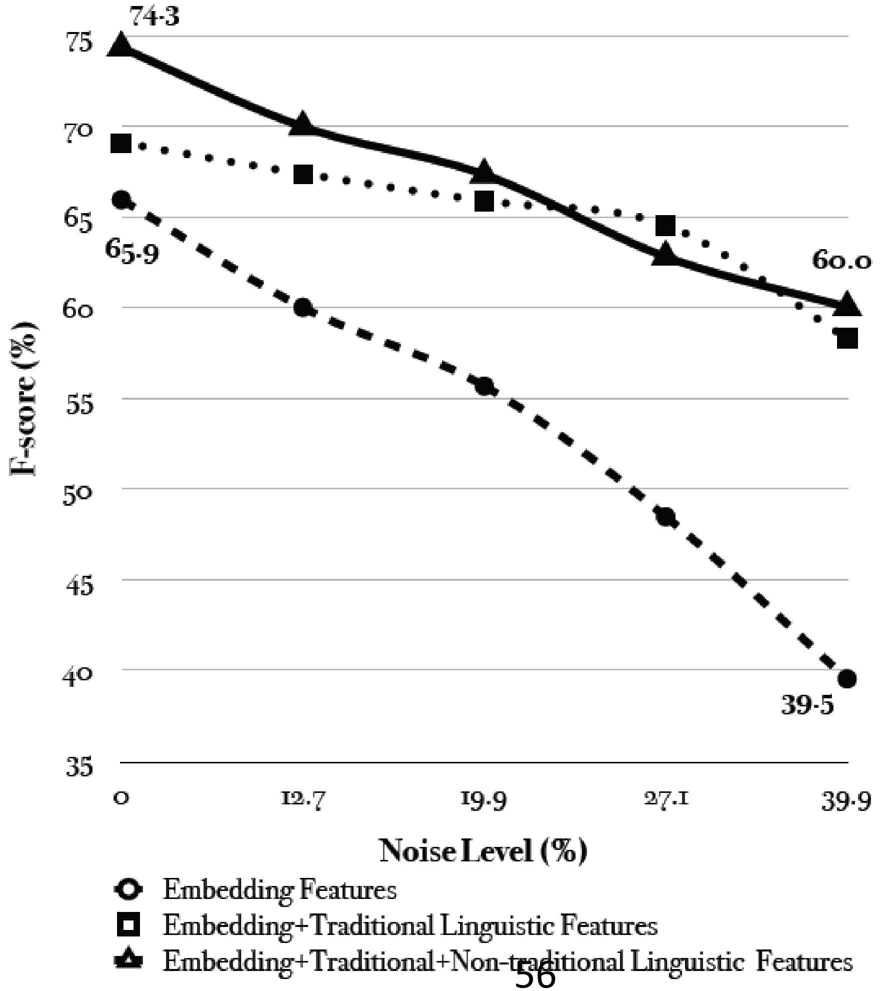
- (Pan et al., ACL2017)



Overall Cross-lingual Entity Linking Results



Impact of Non-Traditional Linguistic Resources: More Robust to Noise



(Zhang et al., 2017)

Outline

• Motivation, Task, Application	30 min
• Traditional “Ancient” Approaches	15 min
• Modern Approaches	
○ Language Universal EDL	15 min
○ Multi-lingual Common Space Construction	30 min
Coffee Break	
○ Cross-lingual Transfer Learning	20 min
○ Cross-lingual Neural Entity Linking	20 min
• Remaining Challenges and New Directions	30 min
• Demos, Resources and QA	20 min

Modern/Recent approaches: Entity Representation

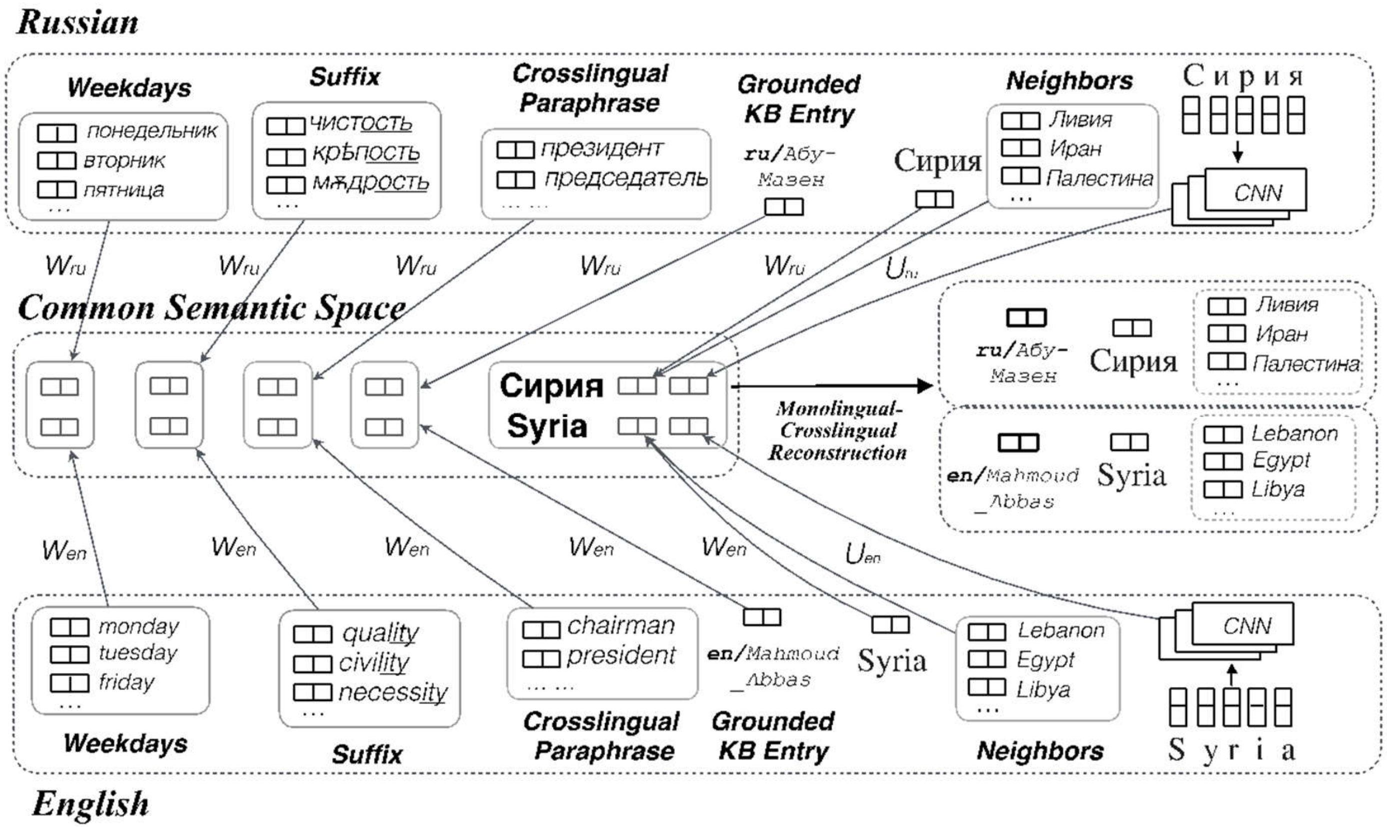
- Until 2016-2017, key EDL approaches were “similarity based”
  - Exceptions: Cucerzan’12
    - Topic based representations of entities
  - Entity Representation: tokens from the Wikipedia Page
  - Linking decision = Similarity between context and description
  - **Sim**(Wikipedia(mention), Text(mention))

- The move to “entity based”: representing additional information about an entity
  - Context in linked data
  - Types (possibly fine-grained)
- Better sentence-level reasoning
- Context at different granularity
- Accomplished by acquiring (dense) representations for entities.

Construct a Common Semantic Space for Thousands of Languages

- Motivations
  - There are 3000+ languages with electronic record
  - NLP training data only available for several dominant languages
- Goals
  - Build a common semantic space across thousands of languages for resource sharing and richer semantic continuous representation for words, concepts and entities
- Limitations of Previous Attempts (e.g., Upadhyay et al., 2016, Cho et al., 2017)
  - Mostly English-anchored, cannot capture all linguistic phenomena
  - Heavily relied on bilingual dictionaries and parallel data which are not always available
  - Only limited to dozens of languages

Multi-lingual Common Space Construction



- Limitations of previous work: rely on bi-lingual dictionaries and treat entity names merely as a combination of words
- Our new hypothesis: Cluster distribution tends to be consistent across languages (Huang et al., EMNLP2018)
- Cross-lingual paraphrase discovery follows (Callison-Burch et all, 2005)

Neighboring Consistent CorrNet

Monolingual Projection for Words and Contexts:

$$H_{l_1} = \sigma(M_{l_1} \cdot W_{l_1} + C_{l_1} \cdot U_{l_1} + b_{l_1}),$$
$$H_{l_2} = \sigma(M_{l_2} \cdot W_{l_2} + C_{l_2} \cdot U_{l_2} + b_{l_2}),$$

Monolingual and Cross-lingual Reconstruction:

$$M'_{l_1} = \sigma(H_{l_1} \cdot W_{l_1}^{\top} + b'_{l_1})\ , \qquad C'_{l_1} = \sigma(H_{l_1} \cdot U_{l_1}^{\top} + b^*_{l_1})\ ,$$

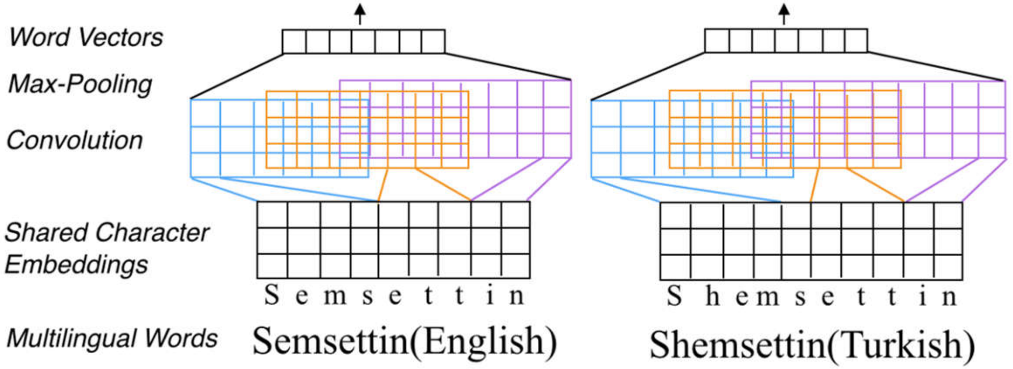
$$M^*_{l_1} = \sigma(H_{l_2} \cdot W_{l_1}^{\top} + b'_{l_1})\ , \qquad C^*_{l_1} = \sigma(H_{l_2} \cdot U_{l_1}^{\top} + b^*_{l_1})\ ,$$

$$M'_{l_2} = \sigma(H_{l_2} \cdot W_{l_2}^{\top} + b'_{l_2})\ , \qquad C'_{l_2} = \sigma(H_{l_2} \cdot U_{l_2}^{\top} + b^*_{l_2})\ ,$$

$$M^*_{l_2} = \sigma(H_{l_1} \cdot W_{l_2}^{\top} + b'_{l_2})\ , \qquad C^*_{l_2} = \sigma(H_{l_1} \cdot U_{l_2}^{\top} + b^*_{l_2})\ ,$$

Character-Aware Word Embedding

- Motivation: mentions of the same concept across languages may share a set of similar characters, e.g., Semsettin Gunaltay (English) = Şemsettin Günaltay (Turkish) = Semsetin Ganoltey (Somali)
- Compose word embeddings from shared character embeddings using Convolutional Neural networks

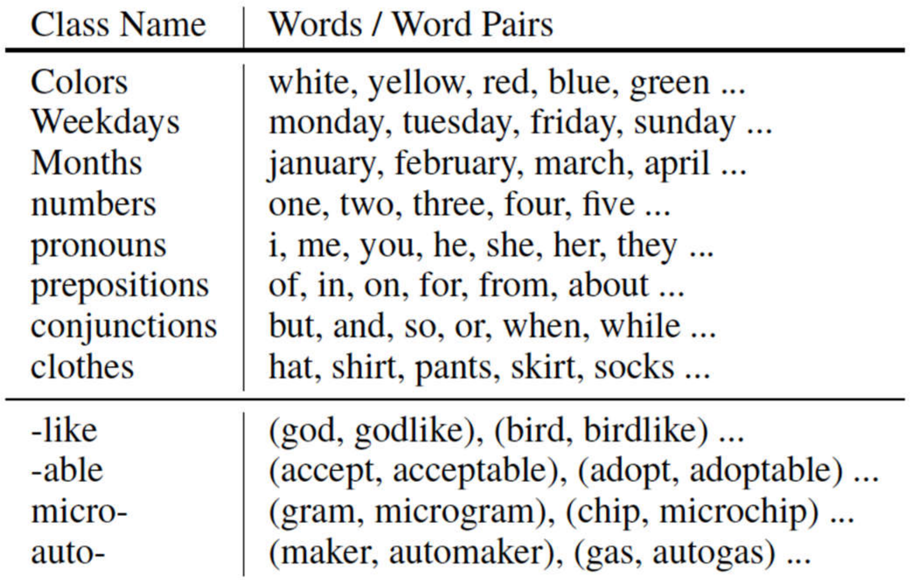


- Cross-lingual Mapping

$$O_{char} = \sum_{\{l_i, l_j\} \in A} L(\hat{W}_{l_i}^{char}, \hat{W}_{l_j}^{char})$$

Linguistic Property Alignment

- Language-Universal Linguistic Knowledge Bases:
  - CLDR (Unicode Common Locale Data Repository)
  - Wiktionary
  - Panlex
- Word Clusters:
  - closed word classes
  - affix



Class Name	Words / Word Pairs
Colors	white, yellow, red, blue, green ...
Weekdays	monday, tuesday, friday, sunday ...
Months	january, february, march, april ...
numbers	one, two, three, four, five ...
pronouns	i, me, you, he, she, her, they ...
prepositions	of, in, on, for, from, about ...
conjunctions	but, and, so, or, when, while ...
clothes	hat, shirt, pants, skirt, socks ...
-like	(god, godlike), (bird, birdlike) ...
-able	(accept, acceptable), (adopt, adoptable) ...
micro-	(gram, microgram), (chip, microchip) ...
auto-	(maker, automaker), (gas, autogas) ...

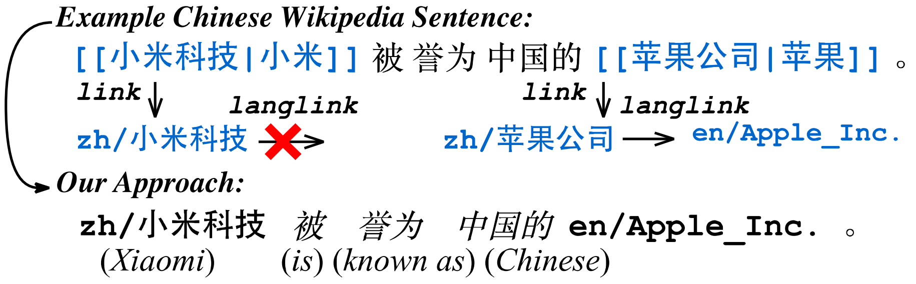
$$H_{l_i}^R = \sigma(M_{l_i}^R \cdot W_{l_i} + b_{l_i}^R) ,$$

$$H_{l_j}^R = \sigma(M_{l_j}^R \cdot W_{l_j} + b_{l_j}^R) ,$$

$$O_R = \sum_{\{l_i,l_j\}\in A} L(H_{l_i}^R, H_{l_j}^R) ,$$

Cross-lingual Joint Entity and Word Embedding Learning

- Code-switch cross-lingual entity/word data generation



Example Chinese Wikipedia Sentence:

[[小米科技|小米]] 被 誉为 中国的 [[苹果公司|苹果]] 。

link ↓ langlink link ↓ langlink

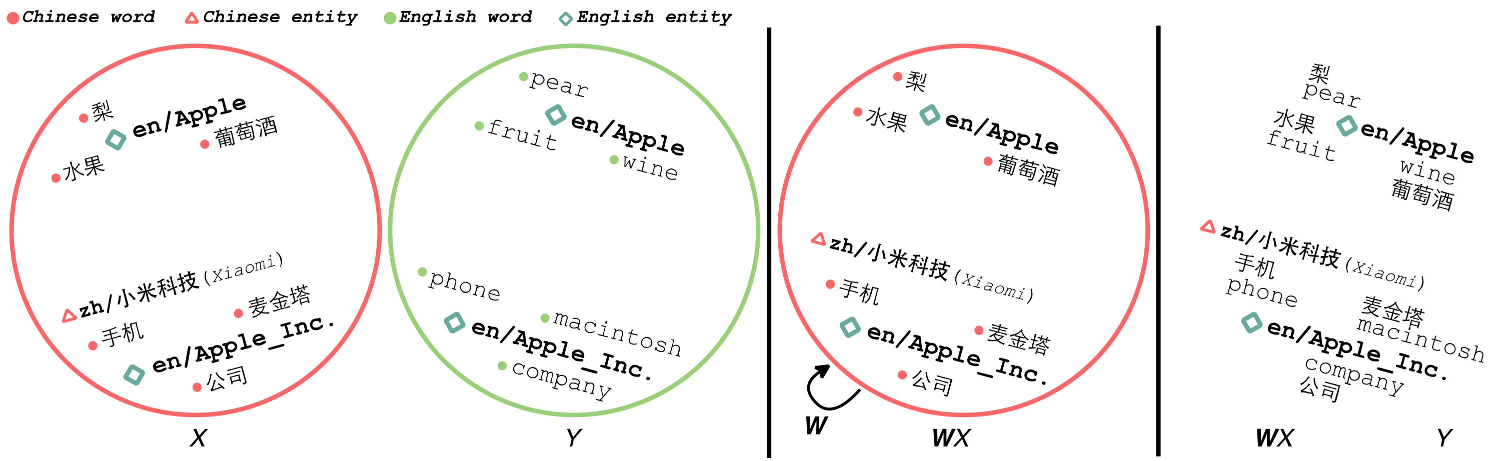
zh/小米科技 ✗→ zh/苹果公司 → en/Apple_Inc.

Our Approach:

zh/小米科技 被 誉为 中国的 en/Apple_Inc. 。

(*Xiaomi*) (*is*) (*known as*) (*Chinese*)

- Use English entities as anchor points to learn a mapping (rotation matrix) W which aligns distributions in IL and English



Cross-Lingual Wikification Using Multilingual Embeddings

- Given mentions in a non-English document, find the corresponding titles in the English Wikipedia

cuarto y actual presidente de los Estados Unidos de América

Amerika Birleşik Devletleri'nin devlet başkanıdır.

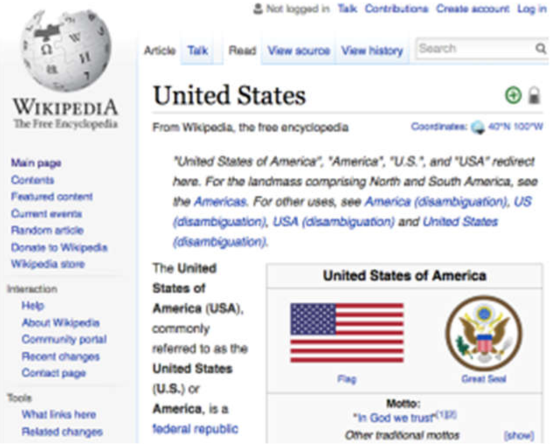
ஐக்கிய அமெரிக்காவின் தற்போதைய குடியரசுத் தலைவர்

นประธานาธิบดีคนที่ 44 คนปัจจุบันของสหรัฐอเมริกา

也是第44任美國總統

dake yankin Hawai a ƙasar Amurika

- Key Challenge
  - Matching words in a foreign language to English Wikipedia titles
- Jointly embedding words and titles in different languages



United States

From Wikipedia, the free encyclopedia

Coordinates: 40°N 100°W

"United States of America", "America", "U.S.", and "USA" redirect here. For the landmass comprising North and South America, see the Americas. For other uses, see America (disambiguation), US (disambiguation), USA (disambiguation) and United States (disambiguation).

The **United States of America** (**USA**), commonly referred to as the **United States** (**U.S.**) or **America**, is a federal republic

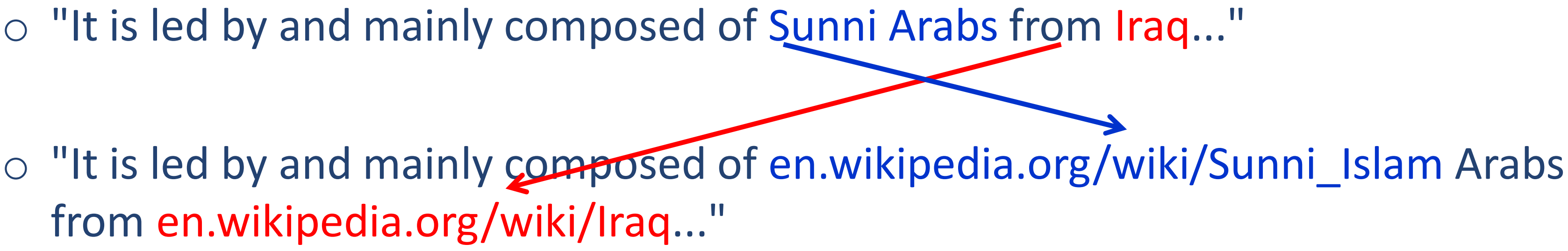
United States of America

Motto: "In God we trust"[12]

Other traditional mottos

Multilingual Word and Title Embeddings

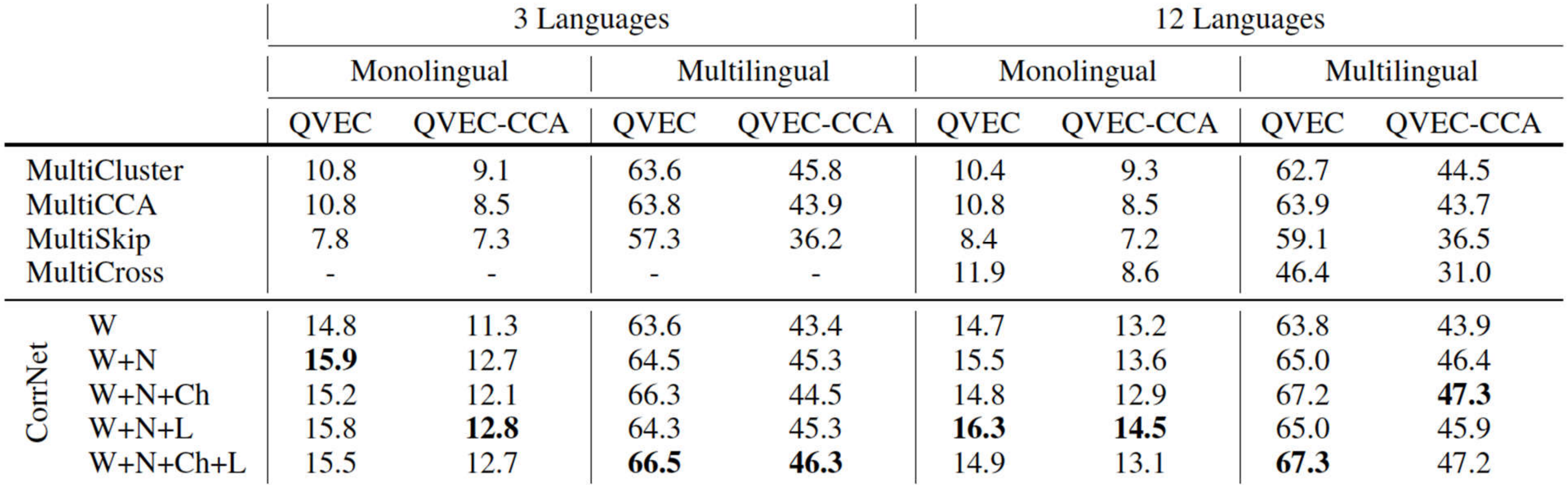
- Learn joint mono-lingual word and title embeddings [Wang et. al. 14]



  - "It is led by and mainly composed of Sunni Arabs from Iraq..."
  - "It is led by and mainly composed of en.wikipedia.org/wiki/Sunni_Islam Arabs from en.wikipedia.org/wiki/Iraq..."
  - Skip-gram with negative sampling [Mikolov et al., 2013]
- Done for English and target Language L.
- Since titles appear as a token in the transformed text, we will obtain an embedding for each word and title from the model.

Intrinsic Evaluation

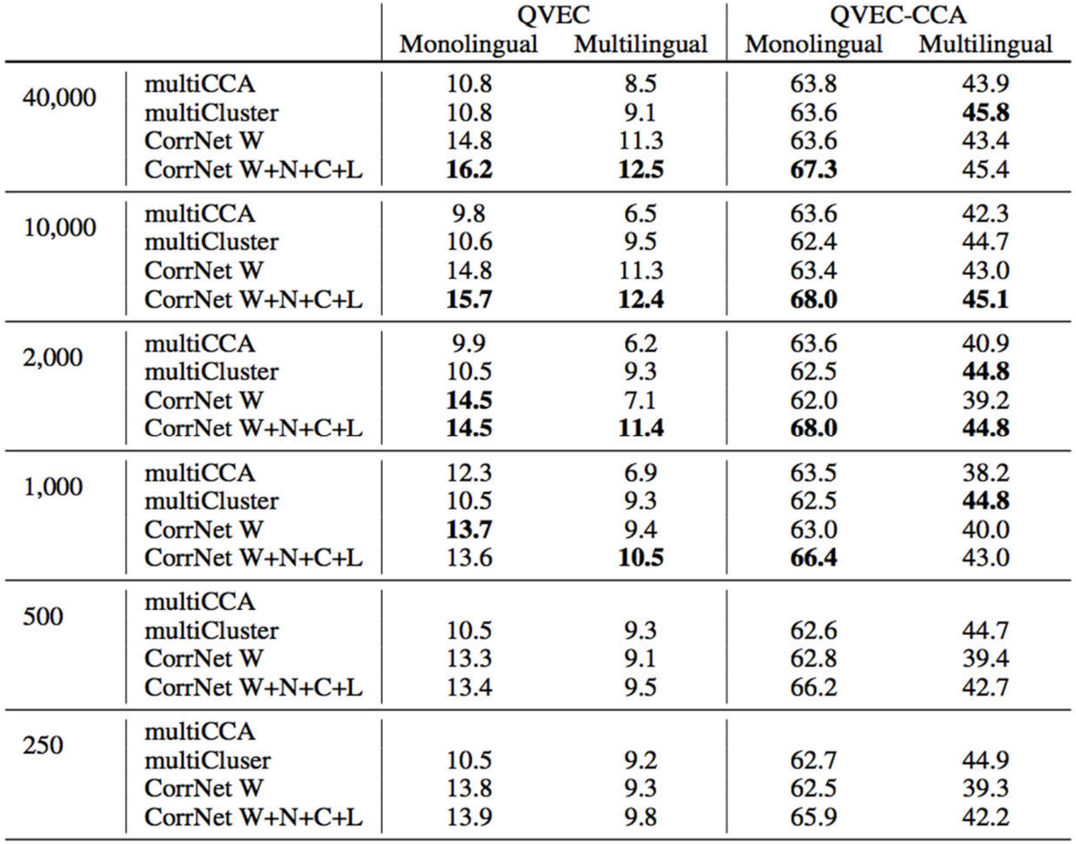
- QVEC: correlation between vectors compared to human created ground truth
- QVEC-CCA: correlation between matrices
- 3 languages: English, Danish, Italian
- 12 languages: Bulgarian, Czech, Danish, German, Greek, English, Spanish, Finnish, French, Hungarian, Italian, Swedish



		3 Languages				12 Languages			
		Monolingual		Multilingual		Monolingual		Multilingual	
		QVEC	QVEC-CCA	QVEC	QVEC-CCA	QVEC	QVEC-CCA	QVEC	QVEC-CCA
MultiCluster		10.8	9.1	63.6	45.8	10.4	9.3	62.7	44.5
MultiCCA		10.8	8.5	63.8	43.9	10.8	8.5	63.9	43.7
MultiSkip		7.8	7.3	57.3	36.2	8.4	7.2	59.1	36.5
MultiCross		-	-	-	-	11.9	8.6	46.4	31.0
CorrNet	W	14.8	11.3	63.6	43.4	14.7	13.2	63.8	43.9
	W+N	**15.9**	12.7	64.5	45.3	15.5	13.6	65.0	46.4
	W+N+Ch	15.2	12.1	66.3	44.5	14.8	12.9	67.2	**47.3**
	W+N+L	15.8	**12.8**	64.3	45.3	**16.3**	**14.5**	65.0	45.9
	W+N+Ch+L	15.5	12.7	**66.5**	**46.3**	14.9	13.1	**67.3**	47.2

Intrinsic Evaluation

- QVEC: correlation between vectors compared to human created ground truth
- QVEC-CCA: correlation between matrices
- 12 languages: Bulgarian, Czech, Danish, German, Greek, English, Spanish, Finnish, French, Hungarian, Italian, Swedish



		QVEC		QVEC-CCA	
		Monolingual	Multilingual	Monolingual	Multilingual
40,000	multiCCA	10.8	8.5	63.8	43.9
	multiCluster	10.8	9.1	63.6	**45.8**
	CorrNet W	14.8	11.3	63.6	43.4
	CorrNet W+N+C+L	**16.2**	**12.5**	**67.3**	45.4
10,000	multiCCA	9.8	6.5	63.6	42.3
	multiCluster	10.6	9.5	62.4	44.7
	CorrNet W	14.8	11.3	63.4	43.0
	CorrNet W+N+C+L	**15.7**	**12.4**	**68.0**	**45.1**
2,000	multiCCA	9.9	6.2	63.6	40.9
	multiCluster	10.5	9.3	62.5	**44.8**
	CorrNet W	**14.5**	7.1	62.0	39.2
	CorrNet W+N+C+L	**14.5**	**11.4**	**68.0**	**44.8**
1,000	multiCCA	12.3	6.9	63.5	38.2
	multiCluster	10.5	9.3	62.5	**44.8**
	CorrNet W	**13.7**	9.4	63.0	40.0
	CorrNet W+N+C+L	13.6	**10.5**	**66.4**	43.0
500	multiCCA				
	multiCluster	10.5	9.3	62.6	44.7
	CorrNet W	13.3	9.1	62.8	39.4
	CorrNet W+N+C+L	13.4	9.5	66.2	42.7
250	multiCCA				
	multiCluser	10.5	9.2	62.7	44.9
	CorrNet W	13.8	9.3	62.5	39.3
	CorrNet W+N+C+L	13.9	9.8	65.9	42.2

Outline

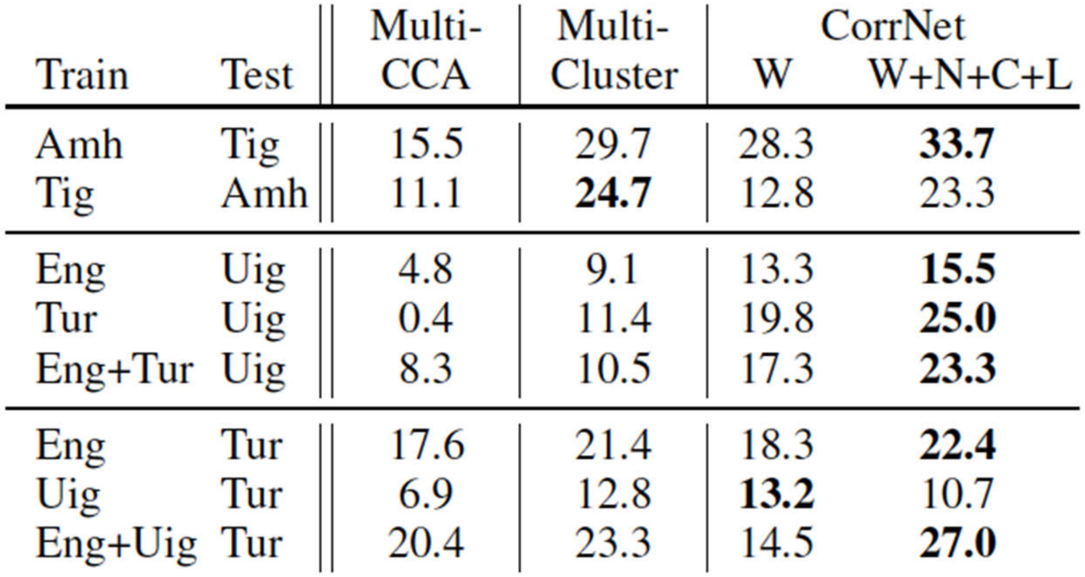
- Motivation, Task, Application 30 min
- Traditional "Ancient" Approaches 15 min
- Modern Approaches
  - Language Universal EDL 15 min
  - Multi-lingual Common Space Construction 30 min

  Coffee Break

  - Cross-lingual Transfer Learning 20 min
  - Cross-lingual Neural Entity Linking 20 min
- Remaining Challenges and New Directions 30 min
- Demos, Resources and QA 20 min

Extrinsic Evaluation on Name Tagging

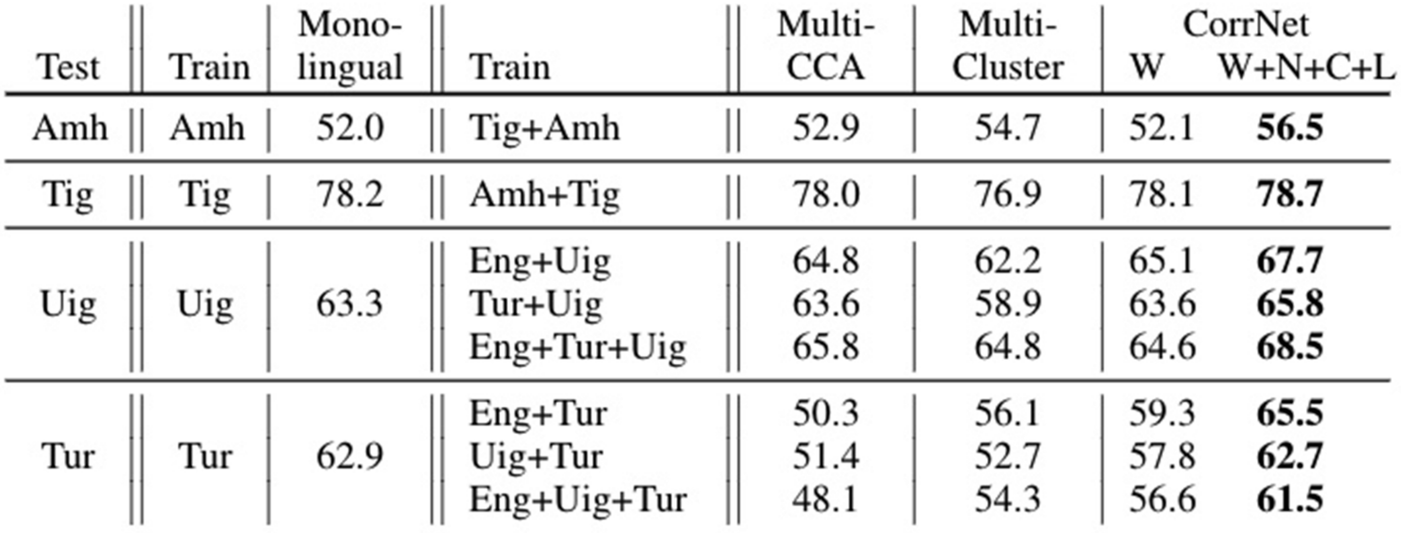
- Cross-lingual direct transfer: trained on a related language and tested on a target language



		Multi-	Multi-	CorrNet	
Train	Test	CCA	Cluster	W	W+N+C+L
Amh	Tig	15.5	29.7	28.3	**33.7**
Tig	Amh	11.1	**24.7**	12.8	23.3
Eng	Uig	4.8	9.1	13.3	**15.5**
Tur	Uig	0.4	11.4	19.8	**25.0**
Eng+Tur	Uig	8.3	10.5	17.3	**23.3**
Eng	Tur	17.6	21.4	18.3	**22.4**
Uig	Tur	6.9	12.8	**13.2**	10.7
Eng+Uig	Tur	20.4	23.3	14.5	**27.0**

Extrinsic Evaluation on Name Tagging

- Mutual enhancement: training set expanded by annotated instances in related languages



		Mono-		Multi-	Multi-	CorrNet	
Test	Train	lingual	Train	CCA	Cluster	W	W+N+C+L
Amh	Amh	52.0	Tig+Amh	52.9	54.7	52.1	**56.5**
Tig	Tig	78.2	Amh+Tig	78.0	76.9	78.1	**78.7**
Uig	Uig	63.3	Eng+Uig	64.8	62.2	65.1	**67.7**
			Tur+Uig	63.6	58.9	63.6	**65.8**
			Eng+Tur+Uig	65.8	64.8	64.6	**68.5**
Tur	Tur	62.9	Eng+Tur	50.3	56.1	59.3	**65.5**
			Uig+Tur	51.4	52.7	57.8	**62.7**
			Eng+Uig+Tur	48.1	54.3	56.6	**61.5**

What is being Transferred?

- Chechen Name Tagging, common space constructed from Russian and Chechen non-parallel documents, bridged by PanLex (4K word entries) and 26K overlapped/reused words

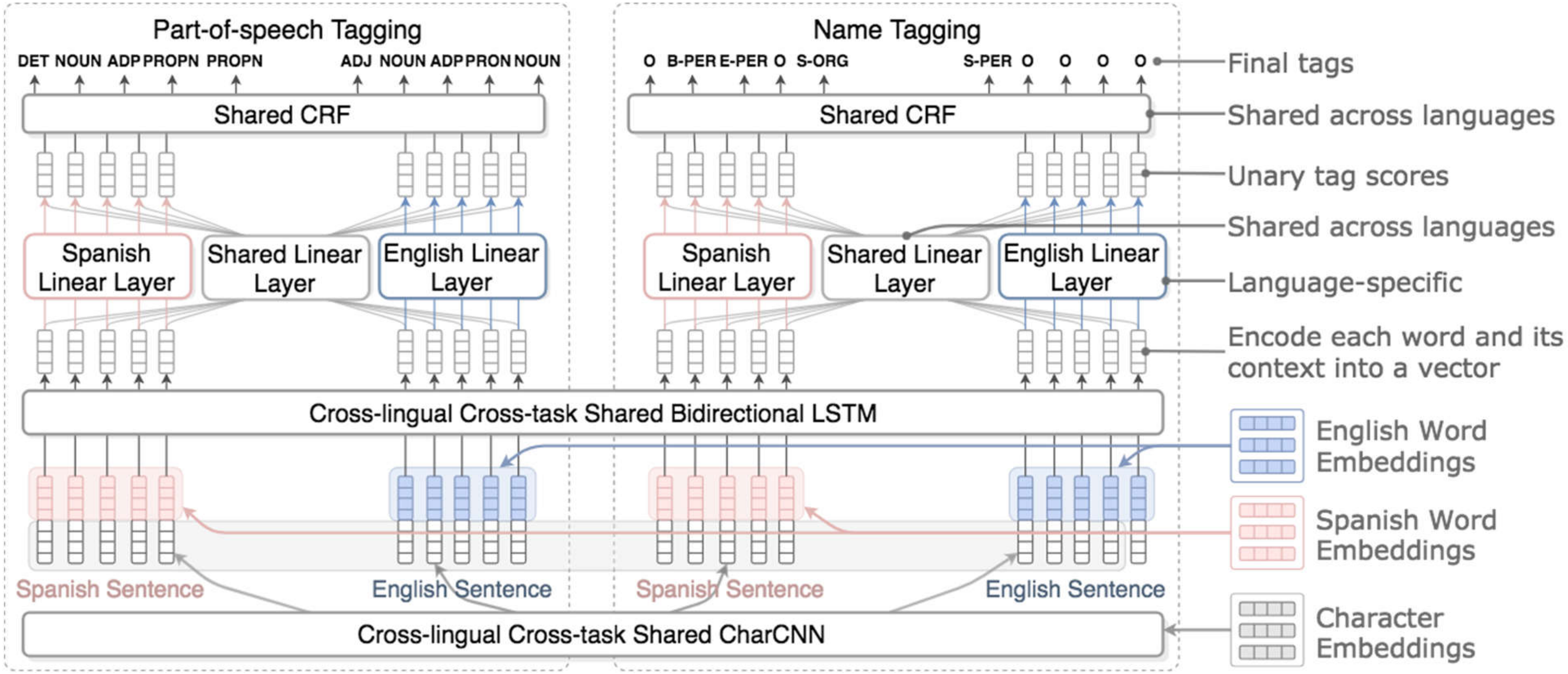
Models	P (%)	R (%)	F (%)
Randomly initialized monolingual embedding	46.3	45.3	45.8
Pre-trained monolingual embedding	67.5	55.6	60.9
+ Common semantic space word embedding	**76.6**	**62.4**	**68.8**

- Common space learned better quality embeddings from related languages and identified many OOV new names in low-resource languages

Type	Chechen Names found by common space & missed by monolingual baseline	English Translation	Frequency in Chechen	Frequency in Russian
GPE	Ши шо даьлча, тхо **Кыргызстане**, тхайн да волчу дIадахара	Kyrgyzstan	2	420
PER	дIа ца. . .ма гIолахь . . . со . . . яда, **Марина** , вайша . . . хьо чу а йигна	Marina	25	20,735
ORG	А |. Халикова |, **ЧГУ** |- н студентка |. ДАЙМОХК № 54, 2005	CSU (Chuvash State University)	1	183

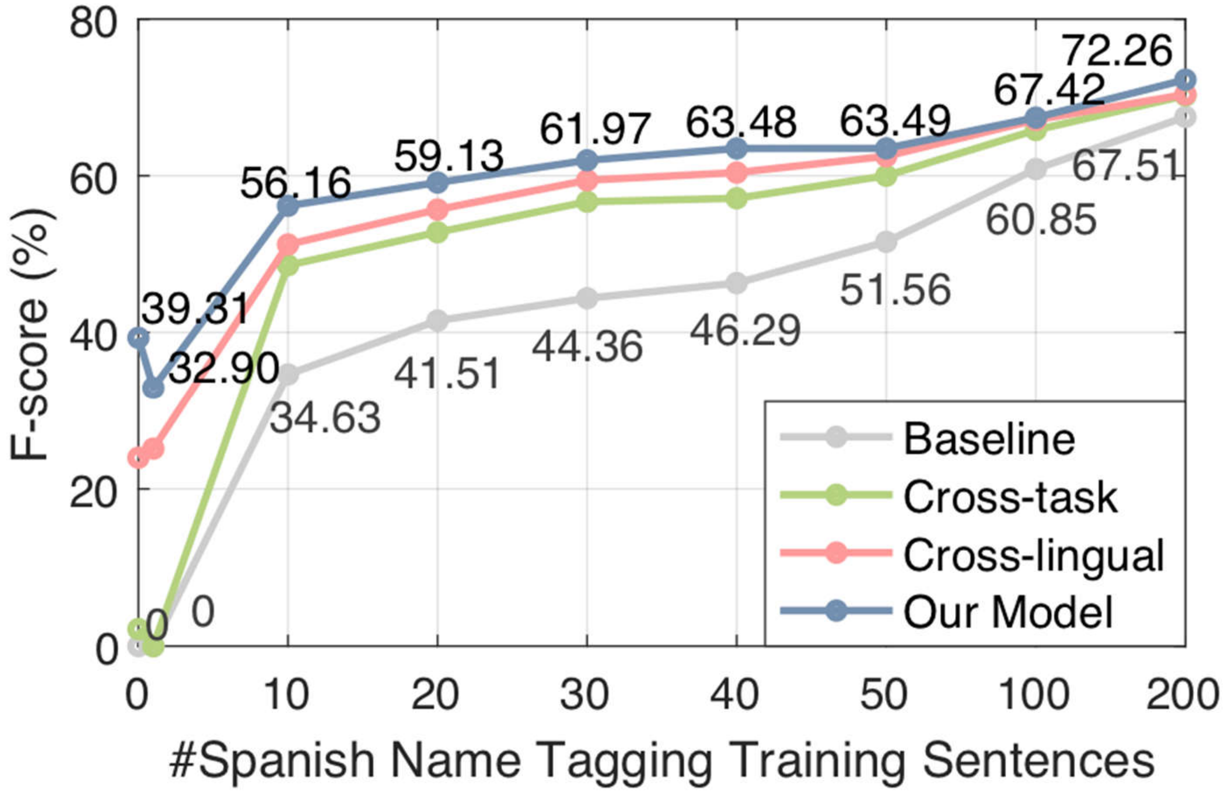
What if the related language is also low-resource for the target task?

Multi-task Multi-lingual Transfer Learning (Lin et al., ACL2018)



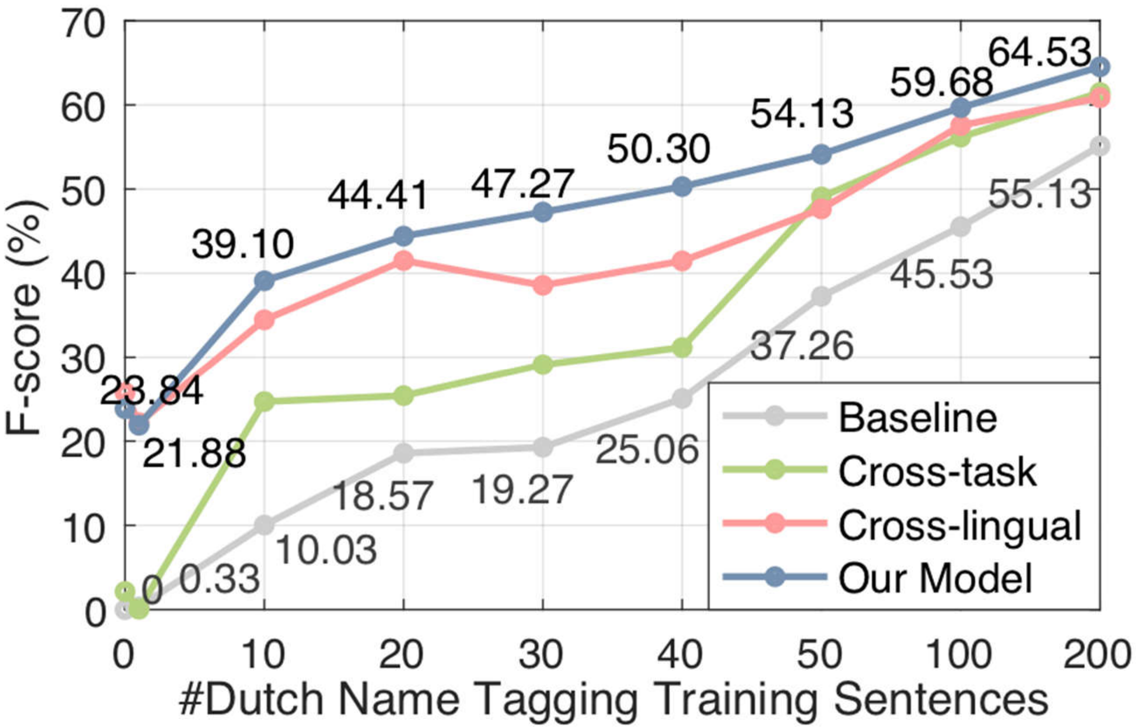
Impact of Cross-lingual Cross-task Transfer Learning

English POS Tagging + English Name Tagging +Spanish POS Tagging + Spanish Name Tagging → Spanish Name Tagging



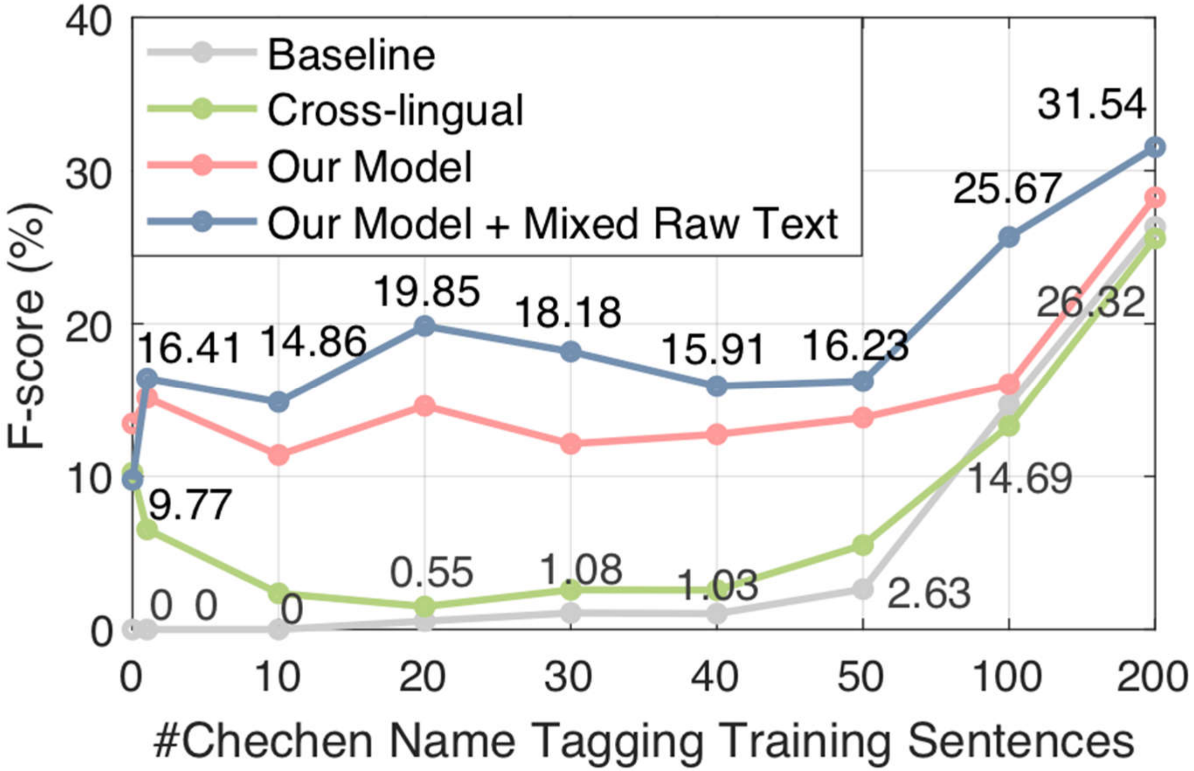
Impact of Cross-lingual Cross-task Transfer Learning

English POS Tagging + English Name Tagging +Dutch POS Tagging + Dutch Name Tagging → Dutch Name Tagging

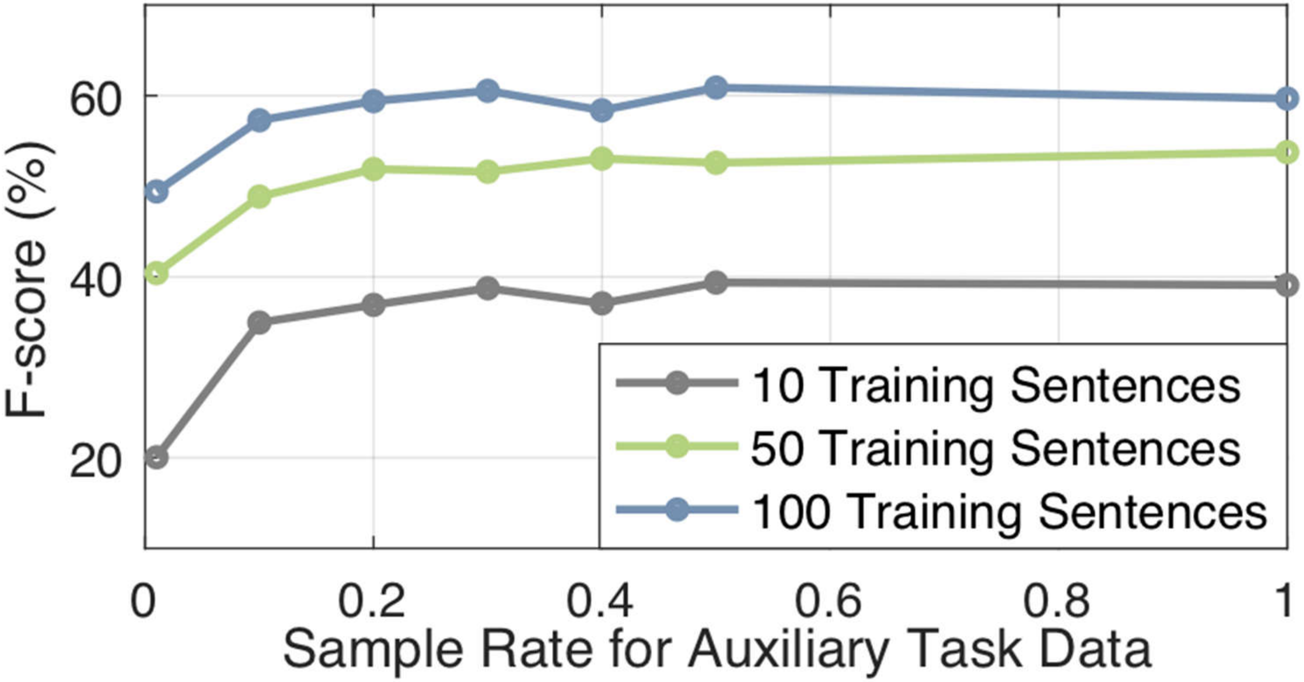


Impact of Cross-lingual Cross-task Transfer Learning

Russian POS tagging + Russian Name Tagging + Chechen Name Tagging → Chechen Name Tagging

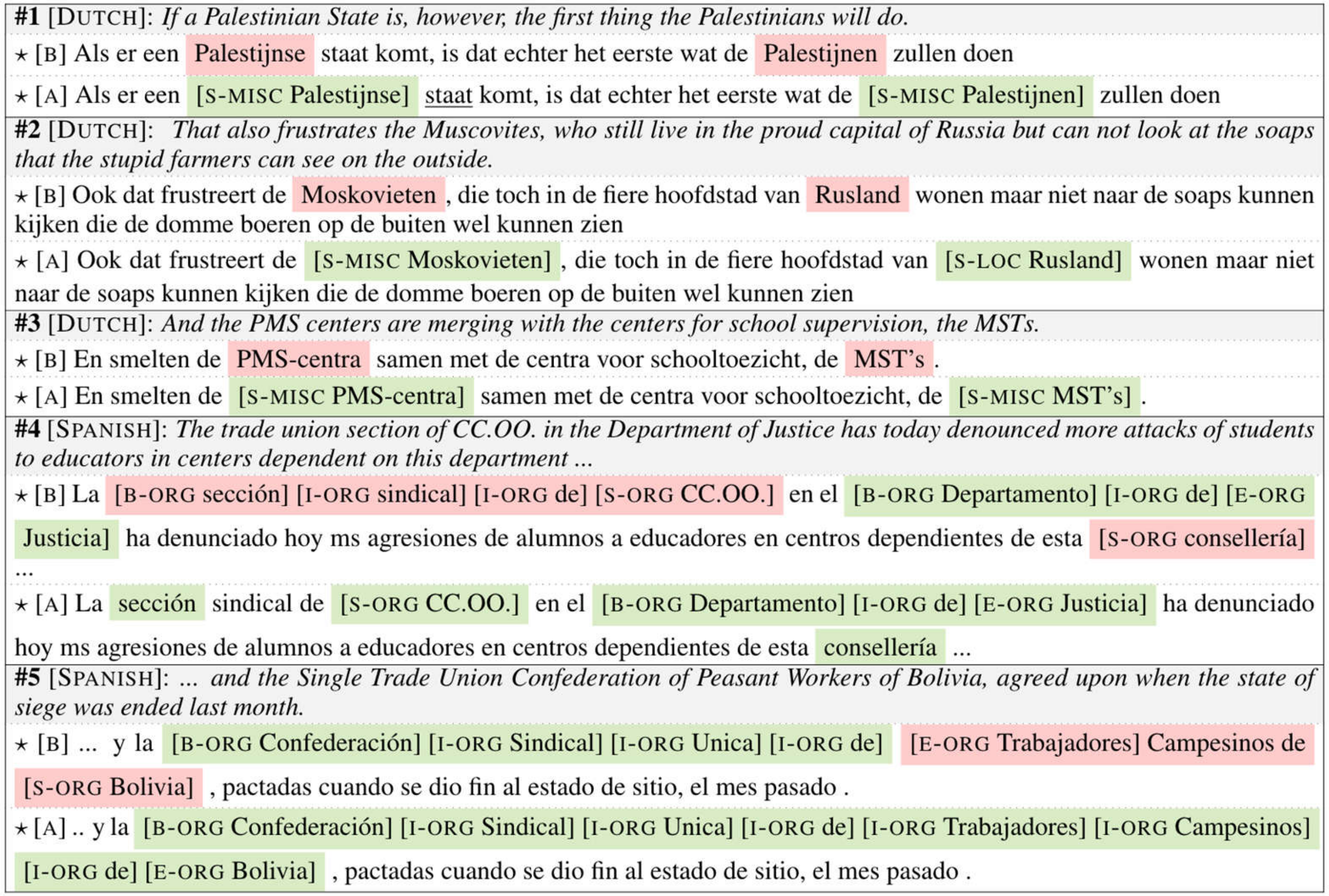


Impact of Cross-lingual Cross-task Transfer Learning



- With only 1% auxiliary task data, our model already obtains 3.9% to 10.5% absolute gains in F-score.

Impact of Cross-lingual Cross-task Transfer Learning



#1 [DUTCH]: *If a Palestinian State is, however, the first thing the Palestinians will do.*

⋆ [B] Als er een Palestijnse staat komt, is dat echter het eerste wat de Palestijnen zullen doen

⋆ [A] Als er een [S-MISC Palestijnse] staat komt, is dat echter het eerste wat de [S-MISC Palestijnen] zullen doen

#2 [DUTCH]: *That also frustrates the Muscovites, who still live in the proud capital of Russia but can not look at the soaps that the stupid farmers can see on the outside.*

⋆ [B] Ook dat frustreert de Moskovieten , die toch in de fiere hoofdstad van Rusland wonen maar niet naar de soaps kunnen kijken die de domme boeren op de buiten wel kunnen zien

⋆ [A] Ook dat frustreert de [S-MISC Moskovieten] , die toch in de fiere hoofdstad van [S-LOC Rusland] wonen maar niet naar de soaps kunnen kijken die de domme boeren op de buiten wel kunnen zien

#3 [DUTCH]: *And the PMS centers are merging with the centers for school supervision, the MSTs.*

⋆ [B] En smelten de PMS-centra samen met de centra voor schooltoezicht, de MST's .

⋆ [A] En smelten de [S-MISC PMS-centra] samen met de centra voor schooltoezicht, de [S-MISC MST's] .

#4 [SPANISH]: *The trade union section of CC.OO. in the Department of Justice has today denounced more attacks of students to educators in centers dependent on this department ...*

⋆ [B] La [B-ORG sección] [I-ORG sindical] [I-ORG de] [S-ORG CC.OO.] en el [B-ORG Departamento] [I-ORG de] [E-ORG Justicia] ha denunciado hoy ms agresiones de alumnos a educadores en centros dependientes de esta [S-ORG consellería] ...

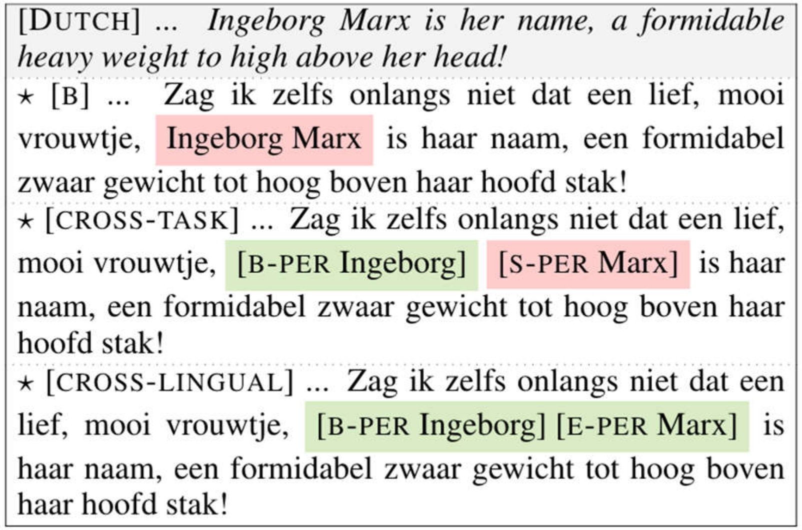
⋆ [A] La sección sindical de [S-ORG CC.OO.] en el [B-ORG Departamento] [I-ORG de] [E-ORG Justicia] ha denunciado hoy ms agresiones de alumnos a educadores en centros dependientes de esta consellería ...

#5 [SPANISH]: ... *and the Single Trade Union Confederation of Peasant Workers of Bolivia, agreed upon when the state of siege was ended last month.*

⋆ [B] ... y la [B-ORG Confederación] [I-ORG Sindical] [I-ORG Unica] [I-ORG de] [E-ORG Trabajadores] Campesinos de [S-ORG Bolivia] , pactadas cuando se dio fin al estado de sitio, el mes pasado .

⋆ [A] .. y la [B-ORG Confederación] [I-ORG Sindical] [I-ORG Unica] [I-ORG de] [I-ORG Trabajadores] [I-ORG Campesinos] [I-ORG de] [E-ORG Bolivia] , pactadas cuando se dio fin al estado de sitio, el mes pasado .

Cross-lingual Vs. Cross-task



[DUTCH] ... *Ingeborg Marx is her name, a formidable heavy weight to high above her head!*

⋆ [B] ... Zag ik zelfs onlangs niet dat een lief, mooi vrouwtje, Ingeborg Marx is haar naam, een formidabel zwaar gewicht tot hoog boven haar hoofd stak!

⋆ [CROSS-TASK] ... Zag ik zelfs onlangs niet dat een lief, mooi vrouwtje, [B-PER Ingeborg] [S-PER Marx] is haar naam, een formidabel zwaar gewicht tot hoog boven haar hoofd stak!

⋆ [CROSS-LINGUAL] ... Zag ik zelfs onlangs niet dat een lief, mooi vrouwtje, [B-PER Ingeborg] [E-PER Marx] is haar naam, een formidabel zwaar gewicht tot hoog boven haar hoofd stak!

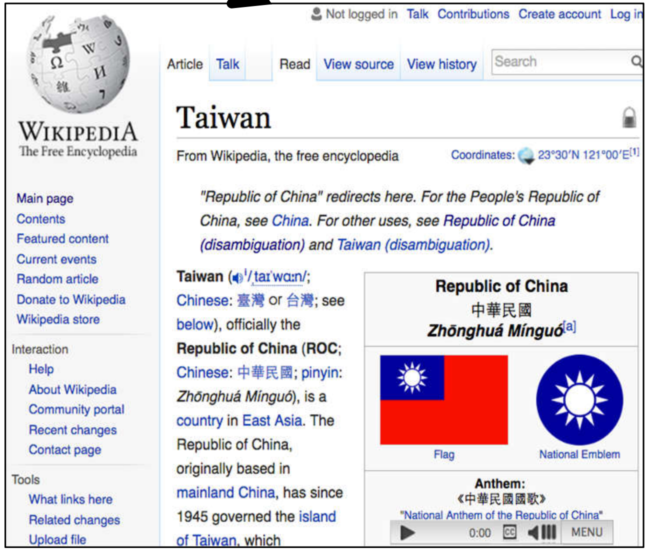
- Both transfer models can identify “*Ingeborg Marx*”.
- In the **cross-lingual** transfer setting, the CRFs layer is shared between Dutch and English models. Invalid transitions, such as B-PER → S-PER, will be punished by the CRFs layer fully trained on high-resource English data.

Outline

• Motivation, Task, Application	30 min
• Traditional “Ancient” Approaches	15 min
• Modern Approaches	
○ Language Universal EDL	15 min
○ Multi-lingual Common Space Construction	30 min
Coffee Break	
○ Cross-lingual Transfer Learning	20 min
○ Cross-lingual Neural Entity Linking	20 min
• Remaining Challenges and New Directions	30 min
• Demos, Resources and QA	20 min

Cross-Lingual Entity Linking

Tayvan, ABD ve İngiltere'de hukuk okuması, Tsai'ye bir LL.B. kazandırdı …



Taiwan

From Wikipedia, the free encyclopedia

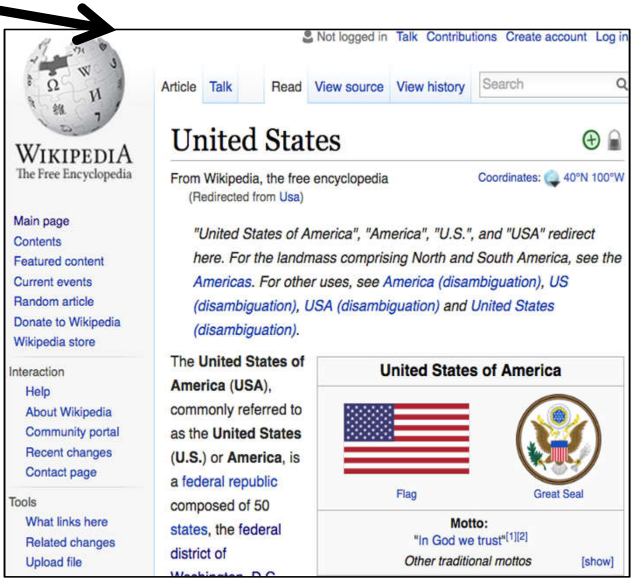
Coordinates: 23°30′N 121°00′E[1]

"Republic of China" redirects here. For the People's Republic of China, see China. For other uses, see Republic of China (disambiguation) and Taiwan (disambiguation).

Taiwan (/ˈtaɪˈwɑːn/; Chinese: 臺灣 or 台灣; see below), officially the **Republic of China** (**ROC**; Chinese: 中華民國; pinyin: *Zhōnghuá Mínguó*), is a country in East Asia. The Republic of China, originally based in mainland China, has since 1945 governed the island of Taiwan, which

Republic of China
中華民國
Zhōnghuá Mínguó[a]

Anthem:
《中華民國國歌》
"National Anthem of the Republic of China"



United States

From Wikipedia, the free encyclopedia
(Redirected from Usa)

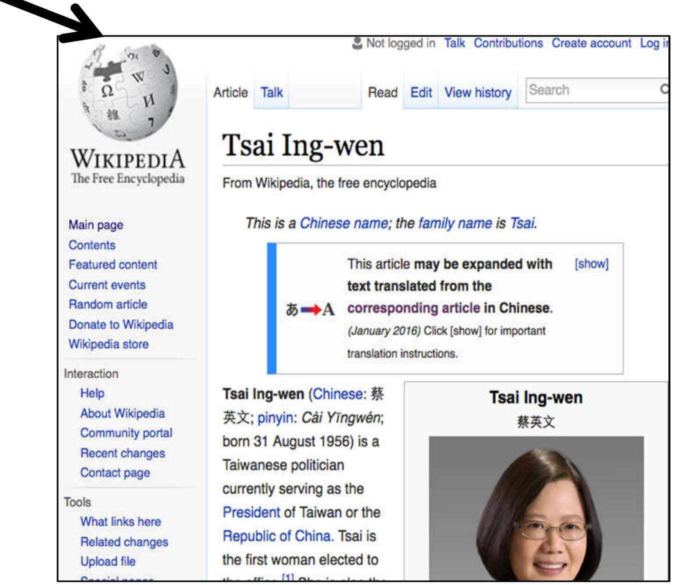
Coordinates: 40°N 100°W

"United States of America", "America", "U.S.", and "USA" redirect here. For the landmass comprising North and South America, see the Americas. For other uses, see America (disambiguation), US (disambiguation), USA (disambiguation) and United States (disambiguation).

The **United States of America** (**USA**), commonly referred to as the **United States** (**U.S.**) or **America**, is a federal republic composed of 50 states, the federal district of Washington, D.C.

United States of America

Motto:
"In God we trust"[1][2]
Other traditional mottos



Tsai Ing-wen

From Wikipedia, the free encyclopedia

This is a Chinese name; the family name is Tsai.

This article **may be expanded with text translated from the corresponding article** in Chinese. *(January 2016)* Click [show] for important translation instructions.

Tsai Ing-wen (Chinese: 蔡英文; pinyin: *Cài Yīngwén*; born 31 August 1956) is a Taiwanese politician currently serving as the President of Taiwan or the Republic of China. Tsai is the first woman elected to

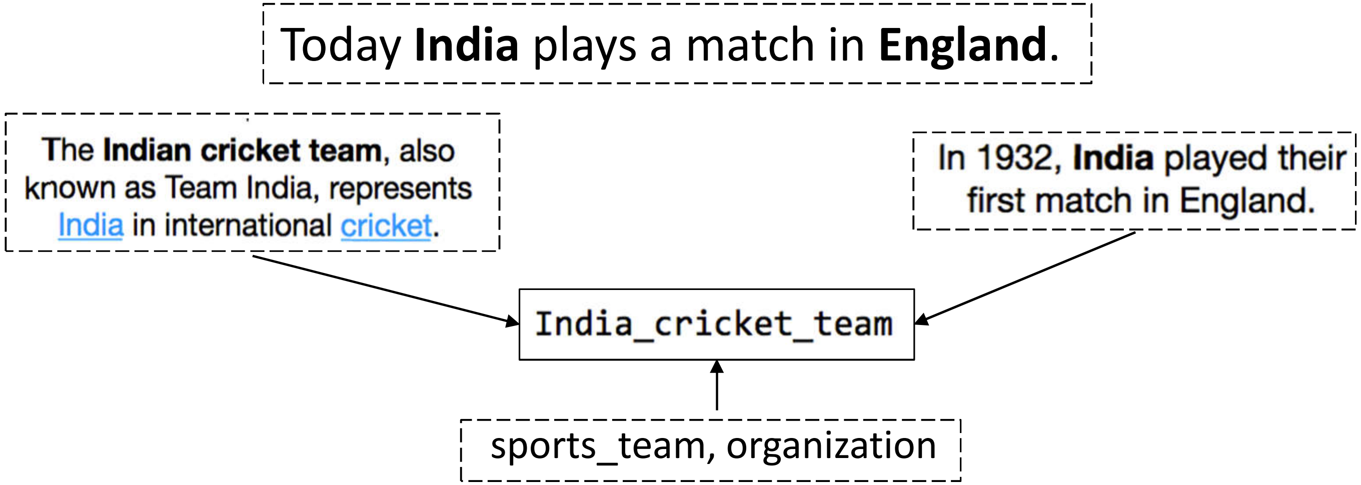
Tsai Ing-wen
蔡英文

Example by Tsai & Roth'16

- **Challenges:**
  - Link to the English Wikipedia
  - Comparing non-English words to English Wikipedia titles

New Approach to Entity Linking [Gupta et al'17]

- Learn Entity Representations
  - ○ Semantic Space of entities
  - ○ Encode information from heterogeneous sources
    - Description of entity from KB (Wikipedia)
    - Context in which entity is mentioned (Wikipedia Linked Data)
    - Fine-Grained Types (Freebase)
  - ○ Hence, learn encoders from observed data to new representation space



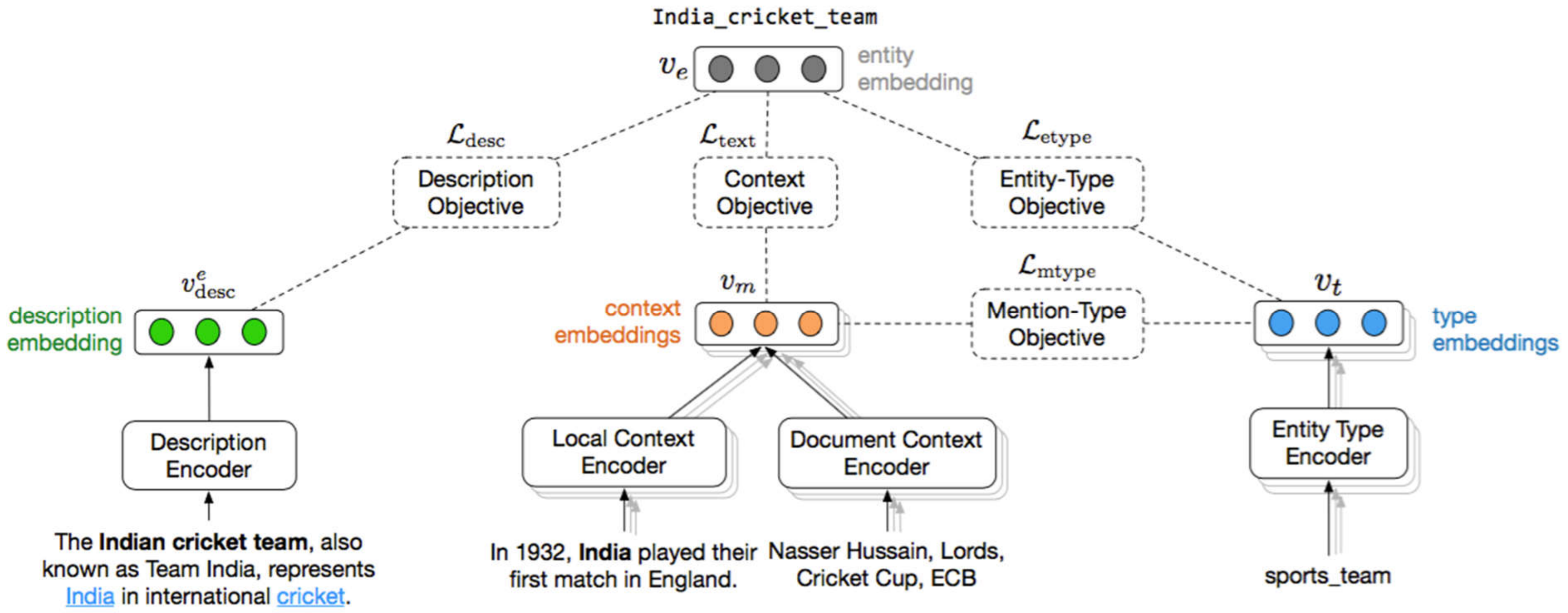
Today **India** plays a match in **England**.

The **Indian cricket team**, also known as Team India, represents India in international cricket.

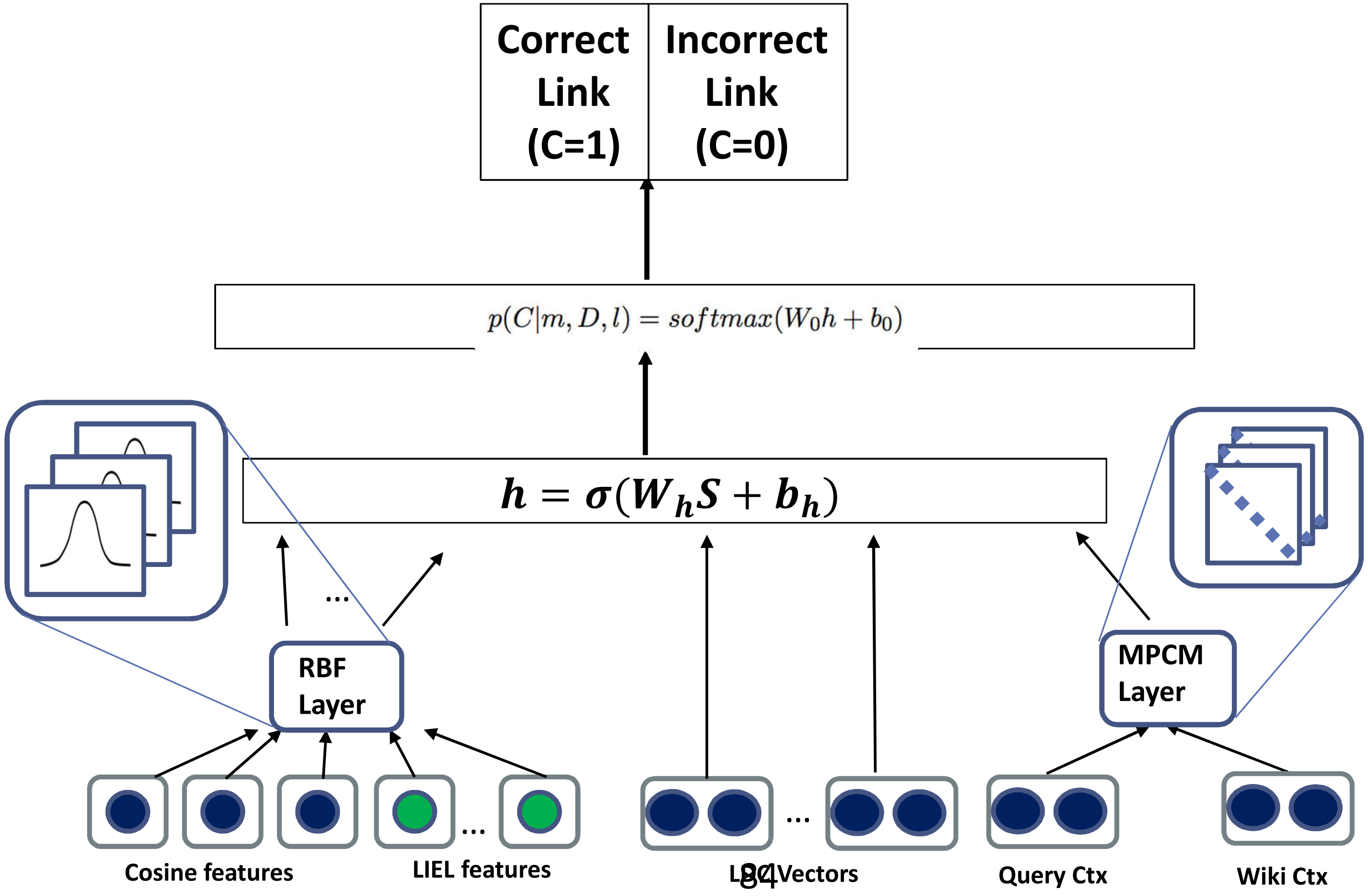
In 1932, **India** played their first match in England.

sports_team, organization

New Approach to Entity Linking [Gupta et al’17]


India_cricket_team
v_e
entity embedding
$\mathcal{L}_{\text{desc}}$
$\mathcal{L}_{\text{text}}$
$\mathcal{L}_{\text{etype}}$
Description Objective
Context Objective
Entity-Type Objective
v^e_{desc}
v_m
$\mathcal{L}_{\text{mtype}}$
v_t
description embedding
context embeddings
Mention-Type Objective
type embeddings
Description Encoder
Local Context Encoder
Document Context Encoder
Entity Type Encoder
The **Indian cricket team**, also known as Team India, represents India in international cricket.
In 1932, **India** played their first match in England.
Nasser Hussain, Lords, Cricket Cup, ECB
sports_team

Zero-shot Cross-lingual Entity Linking (Sil et.al. AAAI18)



Correct Link (C=1)

Incorrect Link (C=0)

$p(C|m, D, l) = softmax(W_0 h + b_0)$

$h = \sigma(W_h S + b_h)$

RBF Layer

MPCM Layer

Cosine features

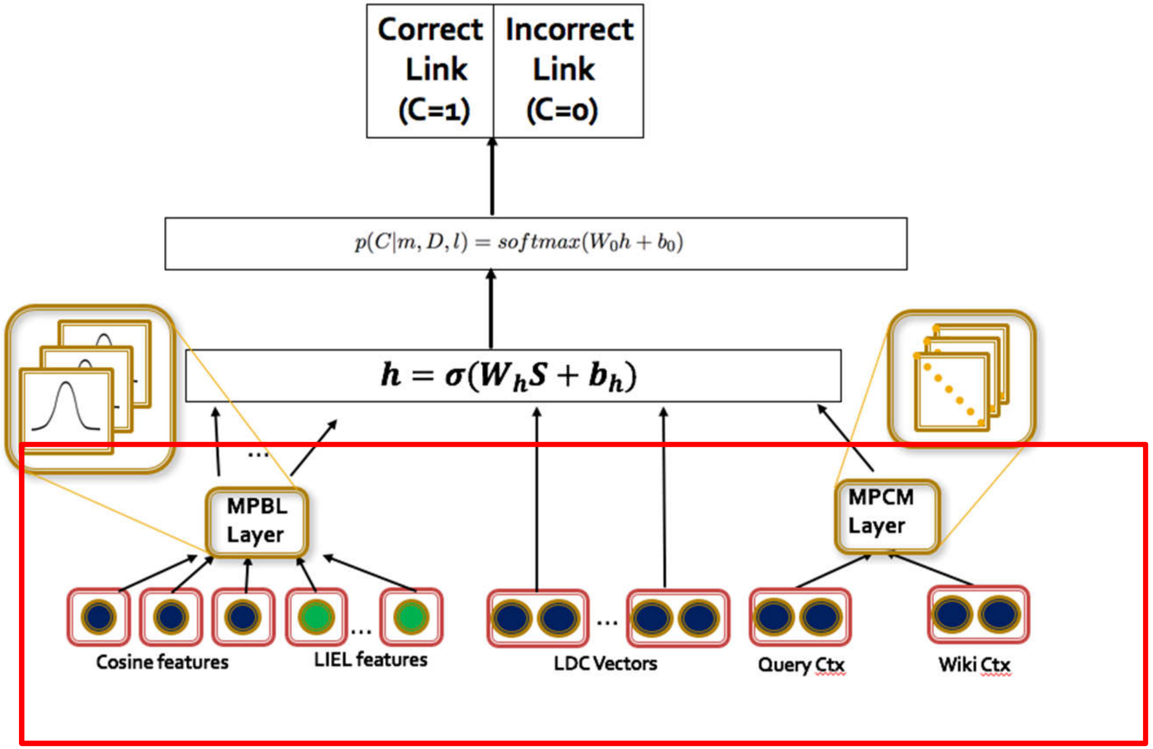
LIEL features

LSC Vectors

Query Ctx

Wiki Ctx

Neural Model Architecture



Feature Abstraction Layer

Impact of Cross-lingual Joint Entity and Word Embedding

- Data: KBP2015 Entity Linking with perfect mentions
- Metric: NERLC (mention boundary + type + KBID)

Source Language	Method	Precision	Recall	F-score
English	State-of-the-art (Pan et al., 2015)	66.1%	68.1%	67.1%
	+ Cross-lingual Joint Entity and Word Embedding (Pan et al., 2019submission)	**69.3%**	**70.8%**	**70%**
Chinese	State-of-the-art (Pan et al., 2015)	78.1%	78.1%	78.1%
	+ Cross-lingual Joint Entity and Word Embedding (Pan et al., 2019submission)	**80.5%**	**81.9%**	**81.2%**

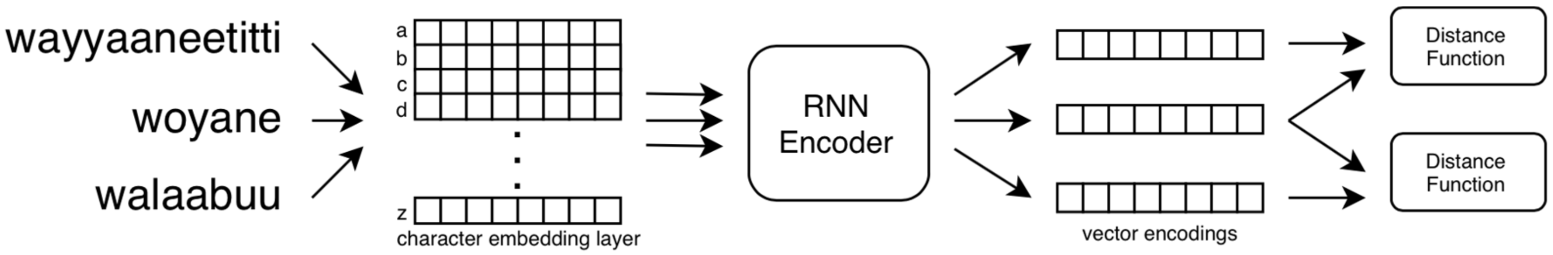
NIL Clustering

(woyane, wayyaaneetitti, walaabuu)

Anchor Positive Negative

$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Loss function from Schroff, et. al. 2015



Mention Clustering F-score on Perfect Mentions	Oromo	Tigrinya
Rule-based	74.3%	77.0%
Character embedding based RNN (Blissett et al., 2019submission)	**83.8%**	**85.0%**

Outline

- Motivation, Task, Application 30 min
- Traditional “Ancient” Approaches 15 min
- Modern Approaches
  - Language Universal EDL 15 min
  - Multi-lingual Common Space Construction 30 min

Coffee Break

  - Cross-lingual Transfer Learning 20 min
  - Cross-lingual Neural Entity Linking 20 min
- Remaining Challenges and New Directions 30 min
- Demos, Resources and QA 20 min

Lack of Domain Knowledge

- The final Perahera of the **Ruhunu Kataragama Maha Devalaya** will be held today.

- In the communiqué the education ministry has cited as a cases in point several instances like the application by a doctor transferred to Bemmulla in Gampaha for admission of his child to the Colombo **D . S . Senanayake Vidyalaya.**

- The navy media unit stated that they suspect that the **Kerala Ganja Cannabis** was brought from India via the mainland

- **IOC**'s fuel prices will again rise again in the light of the increase in fuel prices in Ceylon Petroleum Corporation.

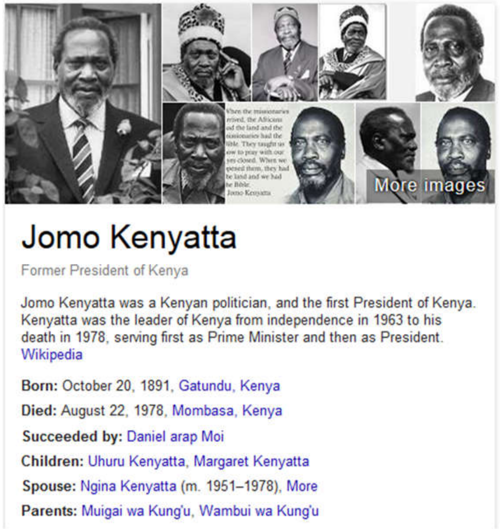
Lack of Deeper Semantic Analysis

- It was a pool **report** typo. Here is exact **Rhodes** quote: ”this is not gonna be a couple of weeks. It will be a period of days.”
- At a **WH briefing** here in Santiago, **NSA** spox **Rhodes** came with a litany of pushback on idea **WH** didn’t consult with **Congress**.
- **Rhodes** **singled out** a **Senate** resolution that passed on March 1st which denounced **Khaddafy’s** atrocities. **WH** says **UN** rez incorporates it

Ben Rhodes
(Speech Writer)

Lack of Commonsense Knowledge

In his first televised address since the attack ended on Thursday, Kenyatta condemned the "barbaric slaughter" and asked help from the Muslim community in rooting out radical elements.



Jomo Kenyatta

Former President of Kenya

Jomo Kenyatta was a Kenyan politician, and the first President of Kenya. Kenyatta was the leader of Kenya from independence in 1963 to his death in 1978, serving first as Prime Minister and then as President. Wikipedia

Born: October 20, 1891, Gatundu, Kenya

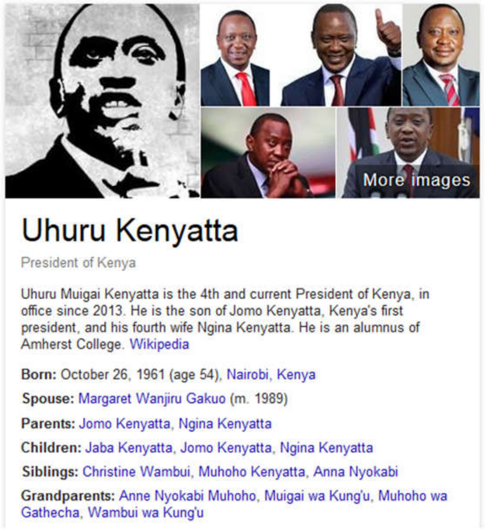
Died: August 22, 1978, Mombasa, Kenya

Succeeded by: Daniel arap Moi

Children: Uhuru Kenyatta, Margaret Kenyatta

Spouse: Ngina Kenyatta (m. 1951–1978), More

Parents: Muigai wa Kung'u, Wambui wa Kung'u



Uhuru Kenyatta

President of Kenya

Uhuru Muigai Kenyatta is the 4th and current President of Kenya, in office since 2013. He is the son of Jomo Kenyatta, Kenya's first president, and his fourth wife Ngina Kenyatta. He is an alumnus of Amherst College. Wikipedia

Born: October 26, 1961 (age 54), Nairobi, Kenya

Spouse: Margaret Wanjiru Gakuo (m. 1989)

Parents: Jomo Kenyatta, Ngina Kenyatta

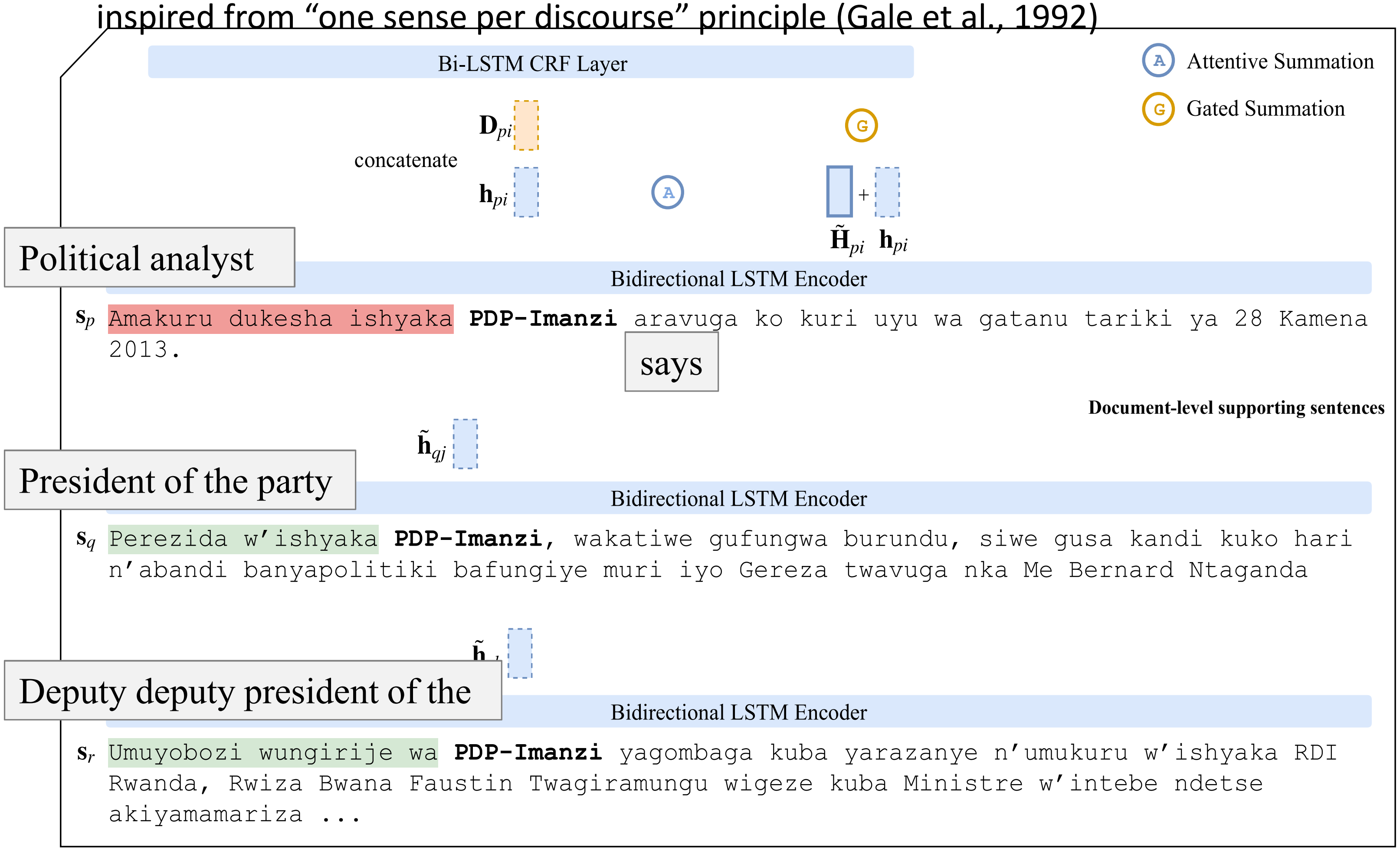
Children: Jaba Kenyatta, Jomo Kenyatta, Ngina Kenyatta

Siblings: Christine Wambui, Muhoho Kenyatta, Anna Nyokabi

Grandparents: Anne Nyokabi Muhoho, Muigai wa Kung'u, Muhoho wa Gathecha, Wambui wa Kung'u

Go beyond sentence-level, sequence labeling, embedding…

- State-of-the-art on CONLL and KBP data sets (Zhang et al., CONLL2018) inspired from “one sense per discourse” principle (Gale et al., 1992)



Political analyst

$\mathbf{s}_p$ Amakuru dukesha ishyaka **PDP-Imanzi** aravuga ko kuri uyu wa gatanu tariki ya 28 Kamena 2013.

says

President of the party

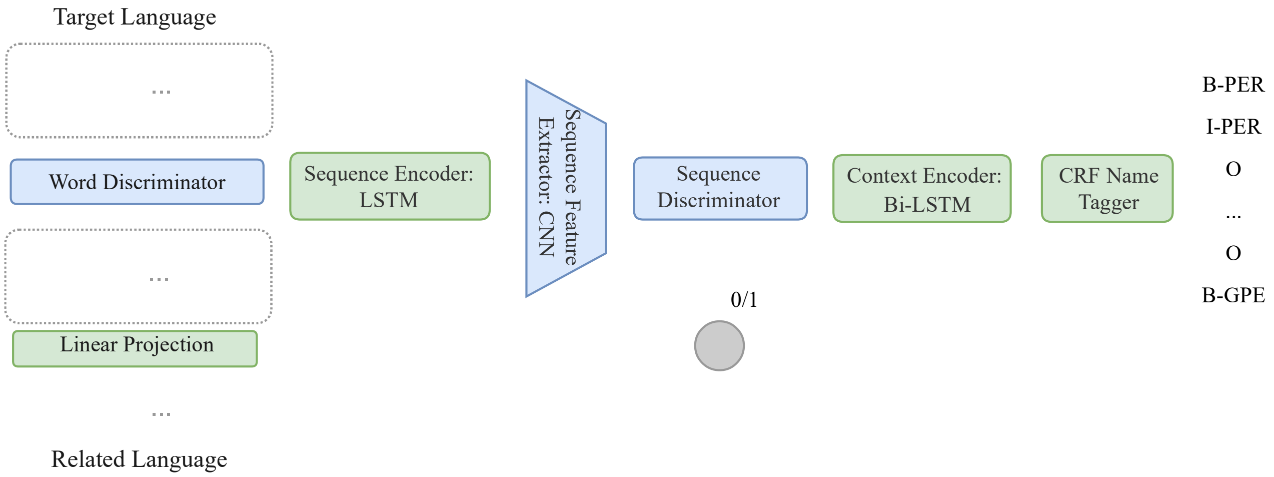
$\mathbf{s}_q$ Perezida w’ishyaka **PDP-Imanzi**, wakatiwe gufungwa burundu, siwe gusa kandi kuko hari n’abandi banyapolitiki bafungiye muri iyo Gereza twavuga nka Me Bernard Ntaganda

Deputy deputy president of the

$\mathbf{s}_r$ Umuyobozi wungirije wa **PDP-Imanzi** yagombaga kuba yarazanye n’umukuru w’ishyaka RDI Rwanda, Rwiza Bwana Faustin Twagiramungu wigeze kuba Ministre w’intebe ndetse akiyamamariza ...

Cross-lingual Adversarial Transfer Learning

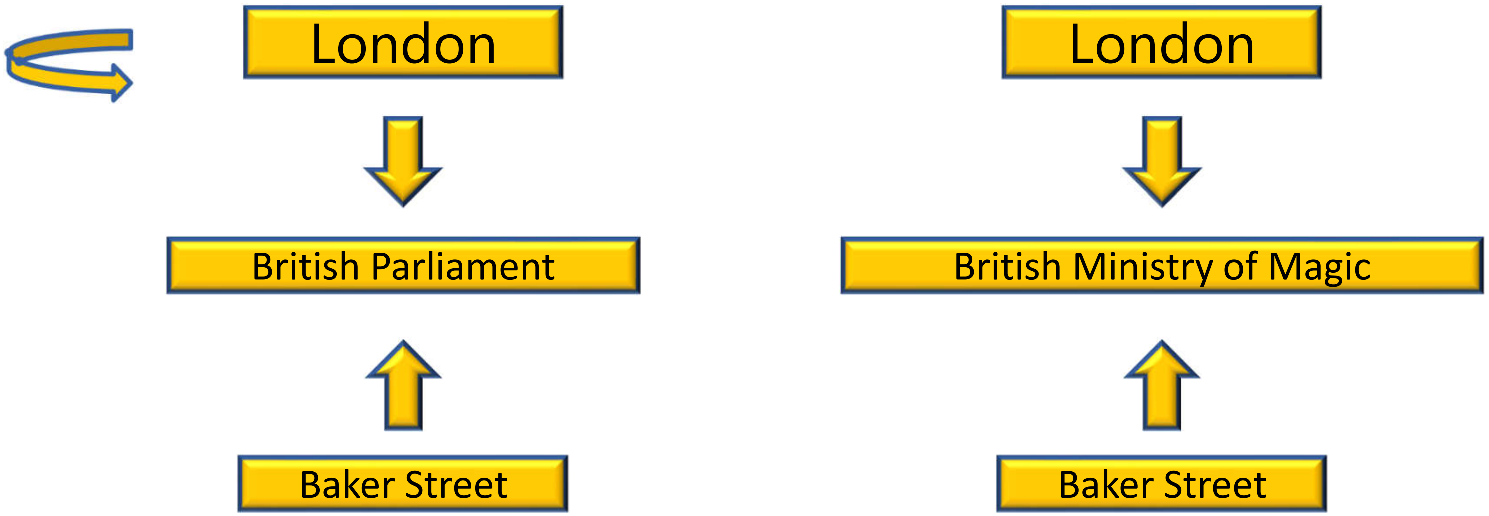
- Goals (Huang et al., 2019submission)
  - Dynamically select domain-related data from related languages
  - Distinguish language-specific features and sharable features
- Multi-level Adversarial Transfer, extension from (Lin et al., ACL2018)
  - Related languages: Swahili for IL9 and Bengali for IL10, and English for both
  - Word-level: A linear projection function is trained so that the word discriminator is unable to predict each word's origin
  - Sequence-level:
    - The LSTM sequence encoder is trained so that the sequence discriminator cannot predict each sequence's origin
    - CNN: generate one unified sequence vector with the same size of dimensions
    - Sequence discriminator: optimize encoder to select sequences from the related language which share similar features with the target language



How Deep ~~Should~~ One Single Knowledge Repository Go

Can

e.g.:



Earth, year 2016:

London

British Parliament

Baker Street

Harry Potter world:

London

British Ministry of Magic

Baker Street

How Deep ~~Should~~ One Single Knowledge Repository Go

Can

- Legal Issues
  - IP: how do we merge a company's knowledge into a universal KB and insure no other company takes advantage of it?
  - Privacy: are we willing to make it easy for everybody to retrieve our personal data (even if some is already public)?

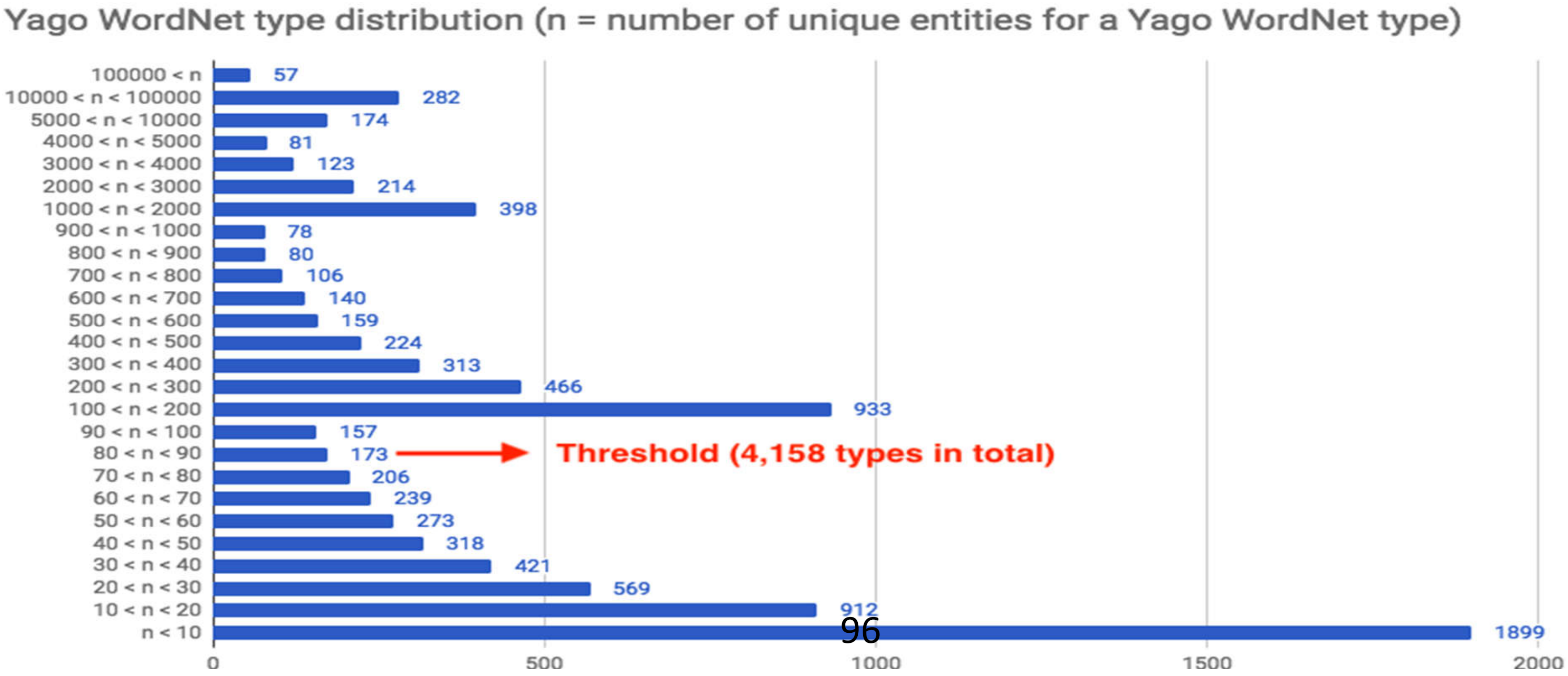
- Expertise

https://en.wikipedia.org/wiki/Airbus_A380

https://en.wikipedia.org/wiki/Boeing_747

Extend to Thousands of Entity Types (TAC-KBP2019 EDL Task)

- 7,297 hierarchical entity types defined in YAGO, derived from WordNet synsets
  - Top level: thing
  - Second level: person, organization, building, artifact, abstraction, physical entity, geographical entity
  - Each type has at least 10 Wikipedia entries
- KBP2018 EDL task: 7,297 hierarchical entity types defined in YAGO and WordNet, each type has at least 10 Wikipedia entries



Entity Extraction Results & Examples

- Mention extraction F-score 70% for Chinese and 75% for English
- Hormone
  - ***Briain-derived neuotrophic factor*** (“***BDNF***”), another important gene in neural plasticity, has also been shown to have reduced methylation and increased transcription in animals that have undergone learning.
- Infectious Disease
  - Notable exceptions include the ***Large Pine Weevil*** (“***Hylobius abietis***”), which can kill young conifers.
- Dumpling
  - ***Shengjian mantou*** is a type of small , pan - fried " baozi " (steamed buns) which is a specialty of Shanghai .
- Fairy
  - The background story of the game starts somewhere in the desert where Anwar , a pure hearted young man finds a rusty oil lamp from what he releases a very powerful and evil ***djinn*** the Nadir .

English Results & Examples (Cont’)

- Lawsuit
  - The landmark ***Brown v. Board of Education*** decision paved they way for PARC v. Commonwealth of Pennsylvania and Mills vs. Board of Education of District of Columbia, which challenged the segregation of students with special needs.
- Mental Disorder
  - Many of these veterans suffer from post ***traumatic stress disorder***, an anxiety disorder that often occurs after extreme emotional trauma involving threat or injury.
- Military Academy
  - The year after, the prince went back to France,[2] where he eventually entered the prestigious academy of ***École spéciale militaire de Saint-Cyr-Coëtquidan***.
- Military Uniform
  - The following below depicted gallery of mounting loops are practically in use in conjunction with the ***5- or 3 color flectarn*** fighting suit.

English Results & Examples (Cont')

- Fundraiser
  - The U.S. Fund administers the long-running ***Trick-or-Treat for UNICEF*** compaign which began as a local fundraising event in Pennsylvania in 1950 and has since raised more than US $170 million to support UNICEF's work.
- Investigator
  - Samuel Hume was born in San Francisco, California in 1885, the son of ***James B. Hume***, a famous Wells Fargo detective.
- Lobbyist
  - Represented by ***Lanny Davis***, the CES lobbied for changes to the "gainful employment rule".
- Medical Scientist
  - Pillemer was born on October 15, 1954, to Jean Burrell Pillemer and ***Louis Pillemer***, and early pioneer in the filed of immunology at Case Western Reserve University.

English Results & Examples (Cont')

- Molecular Biologist
  - Meanwhile an overlapping class of transposable element was described under the name " polintons", derived from the key proteins polymerase and integrase, by ***Vladimir Kapitonov*** and ***Jerzy Jurka***.
- Natural Language
  - The ***Vai*** language , also called ***Vy*** or ***Gallinas*** , is a Mande language spoken by the Vai people , roughly 104,000 in Liberia , and by smaller populations , some 15,500 , in Sierra Leone
- Naval Commander
  - His ship drifting dangerously inshore , at 14:30 Captain ***Thomas Frederick*** gave control to a sailor on board who claimed to have navigated the region and knew a safe anchorage .
- Naval Gun
  - She carried one***15 cm SK L/45 gun***, four ***10.5 cm SK L/45 guns***, four ***SK L/45 gun***, four ***8.8 cm SK L/35 guns***, five ***8.8 cm SK L/30 guns***, and one ***8.8 cm SK L/30 gun*** in a U- boat mounting.

English Results & Examples (Cont’)

- Poisoner
  - It began to be used for murderers who used poisons after the Bishop of Rochester 's cook , ***Richard Rice*** , gave a number of people poisoned porridge , resulting in two deaths in February 1532 .
- President
  - Founder 's Day is national public holiday observed in Ghana to mark the birthday of Ghana 's first president , Dr. ***Kwame Nkrumah*** the key founding father of Ghana .
- Queen
  - He later became the King of Spain and married twice to ***Marie Louise of Savoy*** and then ***Elisabeth Farnese*** .
- Religion
  - The ***Mu'tazila*** tradition of tafsir has received little attention in modern scholarship , owing to several reasons .
- Salad
  - ***Texas caviar*** is a salad of black - eyed peas lightly pickled in a vinaigrette - style dressing , often eaten as a dip accompaniment to tortilla chips .

English Results & Examples (Cont')

- Seafood
  - ***Lauriea siagiani***, is a species of ***squat lobster*** in the family Galatheidae, genus " ***Lauriea***" .
- Sign Language
  - Following this , he has been at many festivals , including Ferstival Clin d'Œil throughout Europe as an actor , performer in various sign languages like ***DGS*** , ***BSL*** , ***LIS*** and ***LSF*** .
- Appetizer
  - This invention of a faux Polynesian experience is heavily influenced by Don the Beachcomber , who is credited for the creation of the ***"pūpū" platter*** and the drink named the " Zombie " for his Hollywood restaurant .
- Bomber
  - The carburetor intake was much larger , a long duct like that on the Nakajima B6N Tenzan was added , and a large spinner — like that on the ***Yokosuka D4Y Suisei*** with the Kinsei 62—was mounted .

Chinese Results & Examples (Cont’)

- Vector
  - ○ 一般 的， 令 D 是 作用 于 黎曼流 形 M 上 的 ***向量丛*** V 的 一阶 微分 算子 。
    (In general, let D be the first-order differential operator of ***the vector bundle*** V acting on the Riemannian manifold M.)
  - ○ 柯西 - 施瓦茨 不等式 叙述， 对于 一个 ***内积空间*** 所有 向量 " x " 和 " y "
    (Cauchy - Schwarz inequality description, for ***an inner product space*** of all vectors "x" and "y”)
- Footbridge
  - ○ 而 较 高 的 一座 哥特式 塔楼 于 1357 年 与 ***查理大桥*** 一起 由 彼得 帕尔 莱勒 兴建， 直到 1464 年 才 完 成 。
    (The taller Gothic tower was built in 1357 by Peter Parleler with the ***Charles Bridge*** until 1464.)
  - ○ 而 中国 最着 名 的 铁索 吊桥 是 四川省 甘孜 的 ***泸定桥*** 。
    (The most famous iron suspension bridge in China is ***Luding Bridge*** in Garze, Sichuan Province.)

Chinese Results & Examples (Cont’)

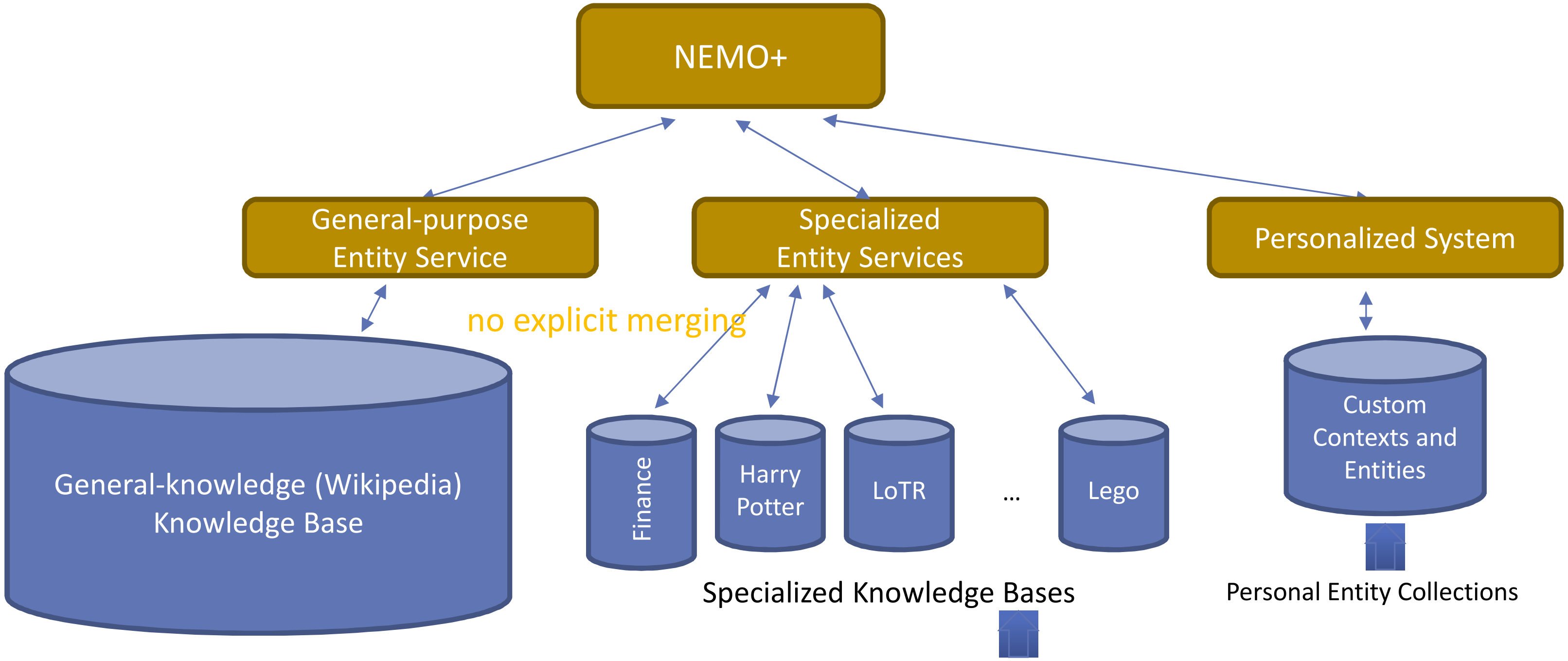
- AutomotiveTechnology
  - 同时，奥迪 也 在 这 一代 A4 中 引入 了 当时 全新 开发 的 Tiptronic ***手自一体变速箱***

    (At the same time, Audi also introduced a newly developed ***Tiptronic tiptronic transmission*** to this generation of A4)
  - ***电子稳定程序***，亦 称 ***车身动态稳定系统***（常 缩写 为 ***ESP®***），又称 ***电子稳定控制系统***（缩写：***ESC***），是 对 旨 在 提升 车辆 的 操控 表现 的 同时 、有效 地 防止 汽车 达到 其 动态 极限 时 失控 的 系统 或 程序 的 通称。

    (***Electronic stability program***, also known as ***dynamic body stability system*** (often abbreviated as ***ESP®***), also known as electronic stability control system (abbreviation: ***ESC***), is designed to improve vehicle handling performance, while effectively preventing the car to reach The generic term for a system or program out of control at its dynamic limits.)
- DataInputDevice
  - 多数 ***键盘布局*** 及 输入法 皆 可 用于 输入 拉丁文 字 或 汉字 。

    (Most ***keyboard layouts*** and input methods are available for entering Latin text or Chinese characters.)

Entity Linking with Multiple Knowledge Bases


NEMO+
General-purpose Entity Service
Specialized Entity Services
Personalized System
no explicit merging
General-knowledge (Wikipedia) Knowledge Base
Finance
Harry Potter
LoTR
...
Lego
Custom Contexts and Entities
Specialized Knowledge Bases
Personal Entity Collections

Two-step approach to analyze an input document:
(1) detect appropriate specialized KB
(2) extract and resolve the entity mentions in the document to the merged-on-the-fly specialized KB with the general-knowledge KB

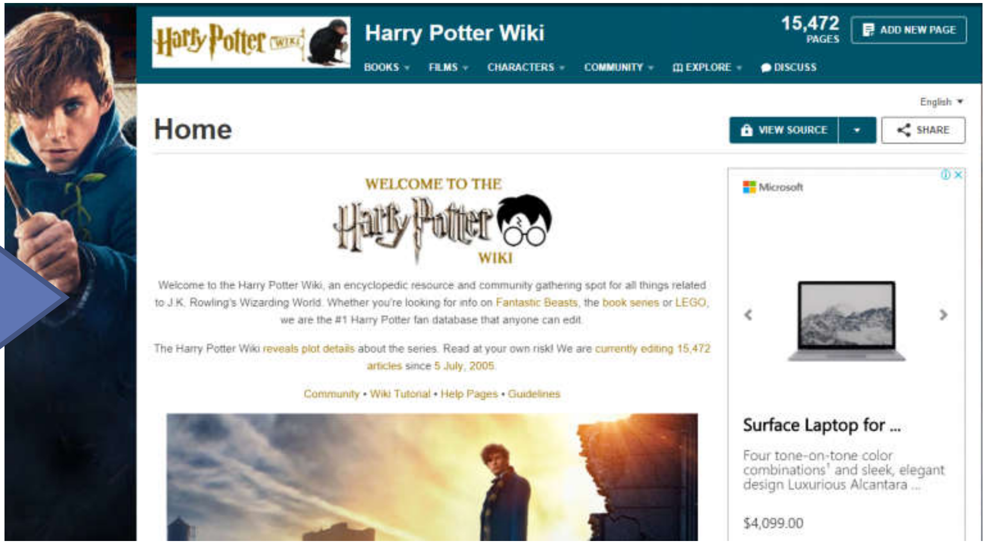
(Microsoft) Silviu Cucerzan. *Entity platform and entity store.*
US Patent App 20150269612 A1, March 18, 2014

Domain/KB Detection

(Gao and Cucerzan, 2017): Entity Linking to One Thousand Knowledge Bases.

Given a large set of entity repositories, identify the one that is the most appropriate for a given document to perform entity linking.

Although he initially laid low, Voldemort was soon forced into the open, and began his bloody conquest of the wizarding world anew. After two years of constant warfare, Voldemort finally gained control of the MOM, and ruled relatively unopposed, save for a few pockets of resistance. Despite his hold over the country, Voldemort was still unsatisfied, as he had yet to remove the danger the prophecy presented to him. After learning of Potter's location, Voldemort set out to destroy the boy once and for all launching his entire amassed force against Hogwarts. Upon arriving at the school, Voldemort was met by a full scale rebellion of Hogwarts staff and students, along with the members of the Order of the Phoenix and the residents of Hogsmeade. As the battle progressed the Death Eaters were driven into the Great Hall, where Voldemort engaged Harry Potter in a duel, and, because all of his Horcruxes were destroyed, Tom Marvolo Riddle was finally killed once and for all.



Home

Welcome to the Harry Potter Wiki, an encyclopedic resource and community gathering spot for all things related to J.K. Rowling's Wizarding World. Whether you're looking for info on Fantastic Beasts, the book series or LEGO, we are the #1 Harry Potter fan database that anyone can edit.

The Harry Potter Wiki reveals plot details about the series. Read at your own risk! We are currently editing 15,472 articles since 5 July, 2005.

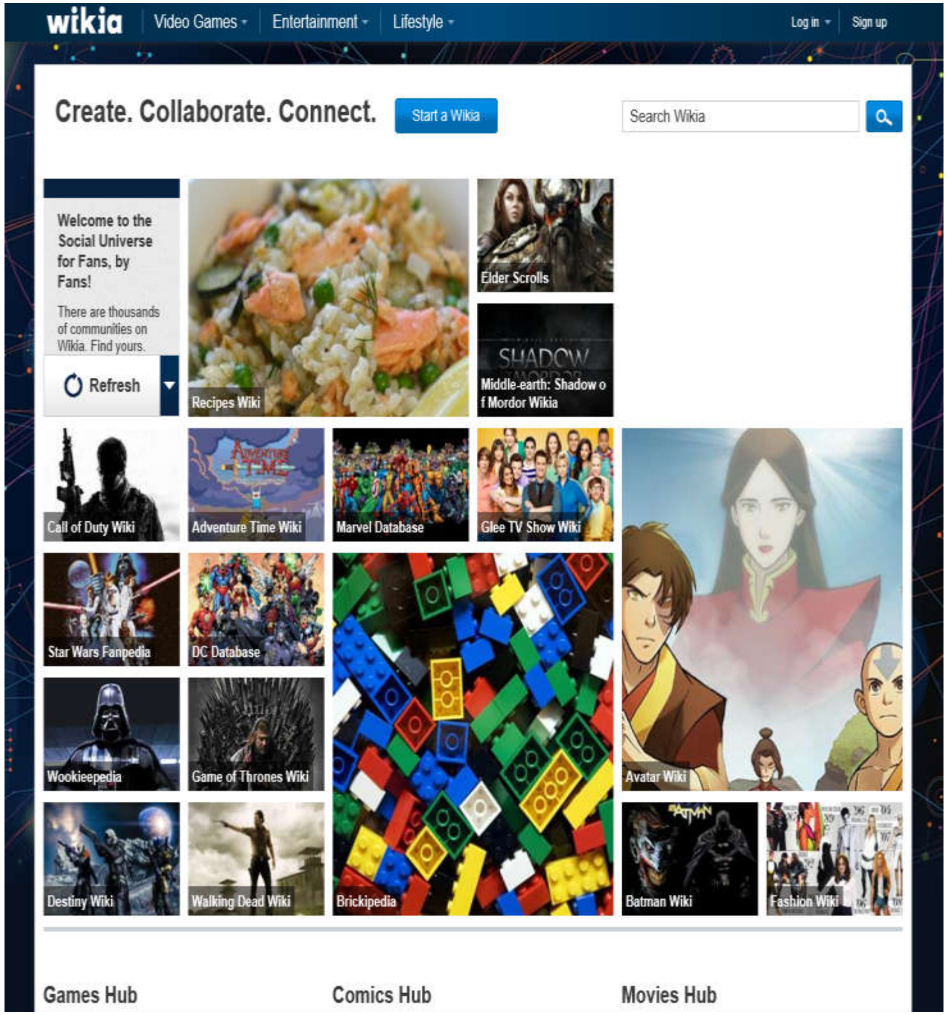
Community • Wiki Tutorial • Help Pages • Guidelines

Surface Laptop for ...

Four tone-on-tone color combinations[1] and sleek, elegant design Luxurious Alcantara ...

$4,099.00

One Wikipedia, 325,000 Wikias



Create. Collaborate. Connect.

Welcome to the Social Universe for Fans, by Fans!

There are thousands of communities on Wikia. Find yours.

Rank	WAM Score	Wikia URL
1	99.84	elderscrolls.wikia.com
2	99.76	disney.wikia.com
3	99.75	runescape.wikia.com
4	99.73	cardfight.wikia.com
5	99.71	starwars.wikia.com
6	99.71	marvel.wikia.com
7	99.67	warframe.wikia.com
8	99.65	avatar.wikia.com
9	99.64	leagueoflegends.wikia.com
10	99.63	yugioh.wikia.com

animals.wikia.com

howtotrainyourdragon.wikia.com

Statistics (summer 2014)

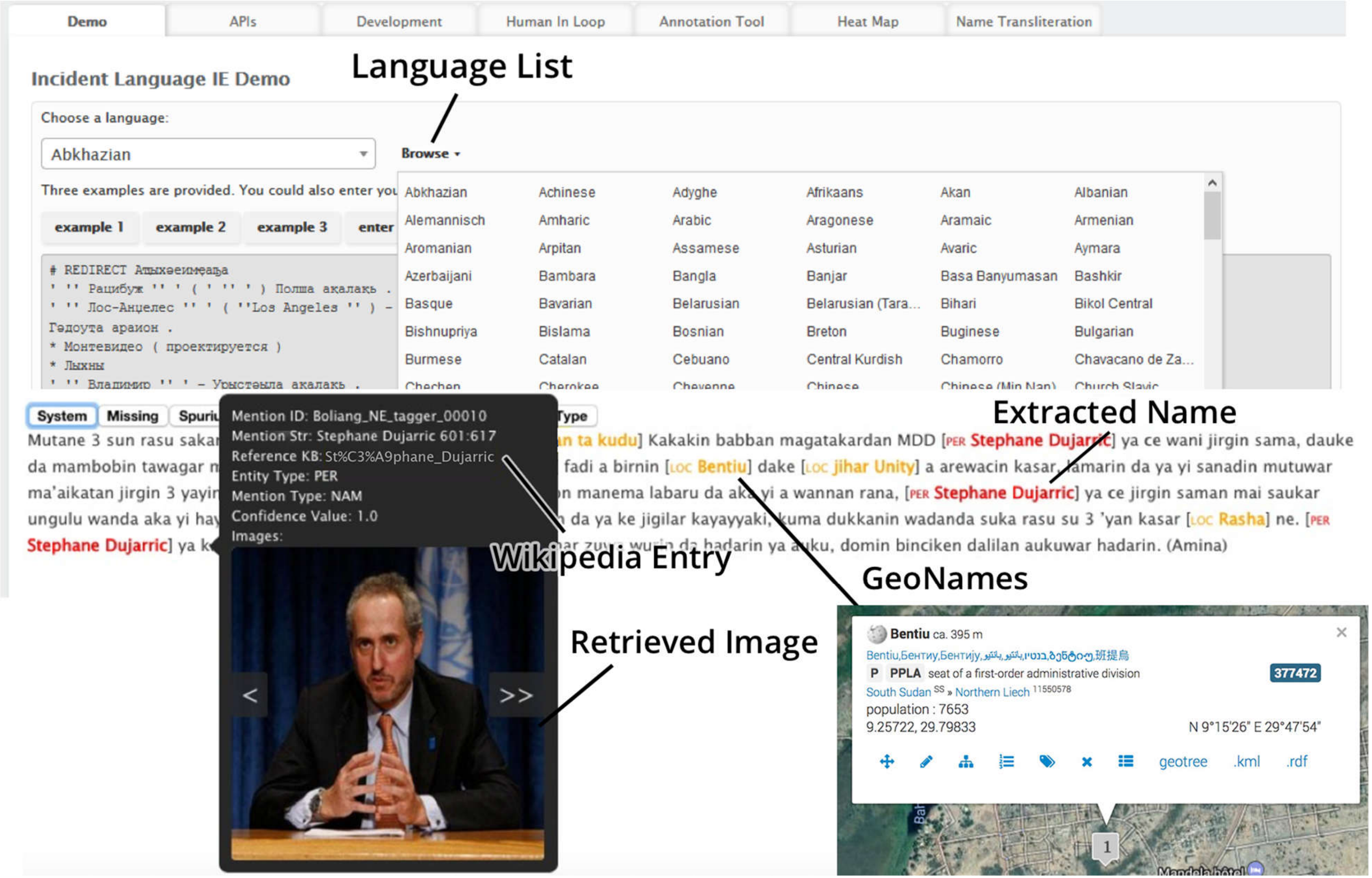
Wikipedia

Entity Pages:	4.5 million
Page Length:	4284 chars
Links to Wikipedia Per Page:	32
Links to Wikias Per Page	0

Wikia (1,163 Collections)

Entity Pages:	3 million
Page Length:	2573 chars
New Entity Names:	2.5 million
Links to the Same Wikia Per Page:	8
Links to Wikipedia Per Page:	0.14
Links to Other Wikias Per Page:	0.03

Linking to Geoname



Incident Language IE Demo

Choose a language:

Abkhazian

Three examples are provided. You could also enter you

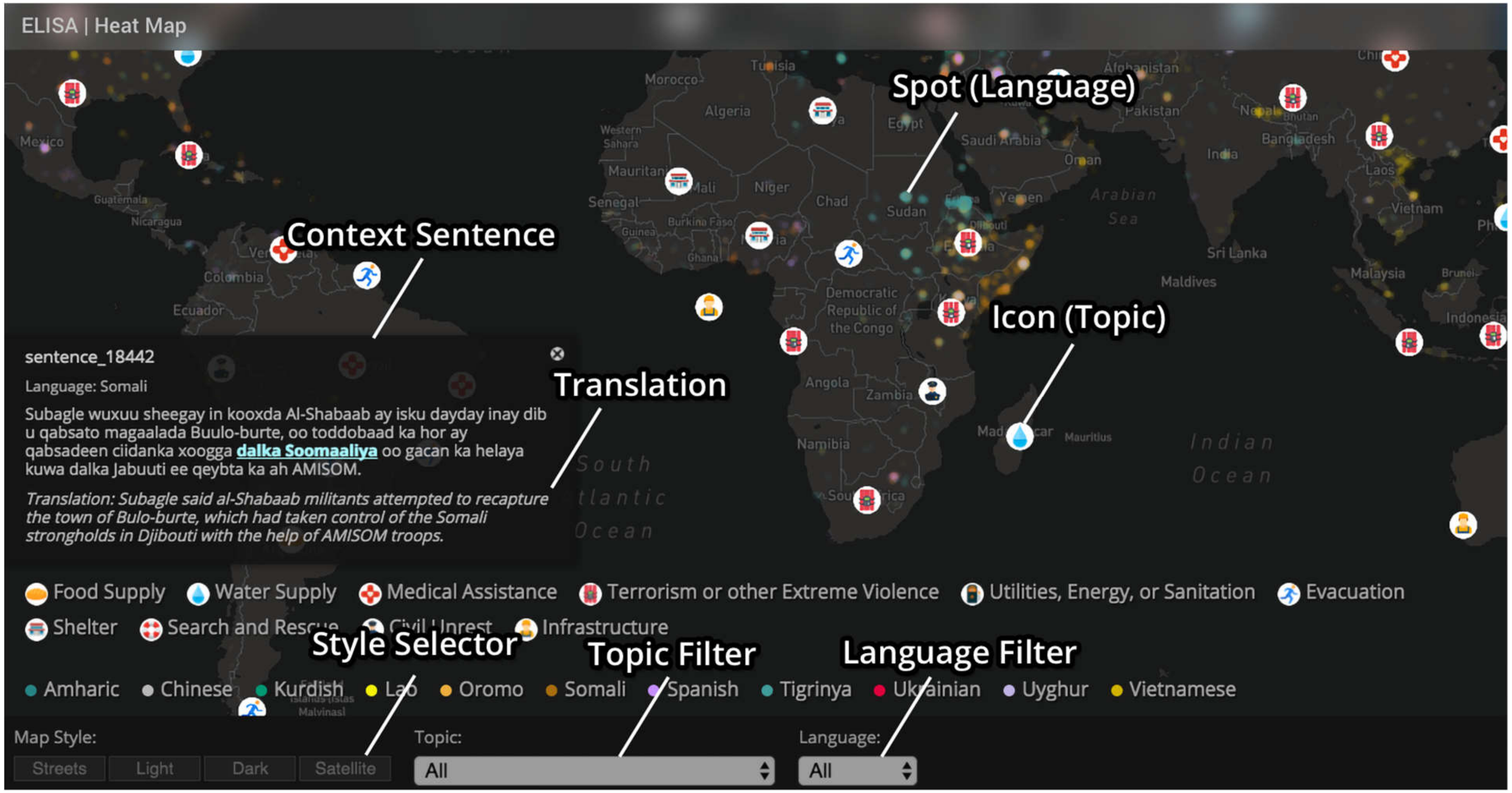
```
# REDIRECT Атыхәеимҿаҕа
' '' Рацибуж '' ' ( ' '' ' ) Полша ақалақь .
' '' Лос-Анцелес '' ' ( ''Los Angeles '' ) –
Гәдоута араион .
* Монтевидео ( проектируется )
* Лыхны
' '' Владимир '' ' – Урыстәыла ақалақь .
```

Mutane 3 sun rasu sakar … an ta kudu] Kakakin babban magatakardan MDD [PER Stephane Dujarric] ya ce wani jirgin sama, dauke da mambobin tawagar m … fadi a birnin [LOC Bentiu] dake [LOC jihar Unity] a arewacin kasar, lamarin da ya yi sanadin mutuwar ma'aikatan jirgin 3 yayin … on manema labaru da aka yi a wannan rana, [PER Stephane Dujarric] ya ce jirgin saman mai saukar ungulu wanda aka yi hay … n da ya ke jigilar kayayyaki, kuma dukkanin wadanda suka rasu su 3 'yan kasar [LOC Rasha] ne. [PER Stephane Dujarric] ya k … har zuwa wurin da hadarin ya auku, domin binciken dalilan aukuwar hadarin. (Amina)

Mention ID: Boliang_NE_tagger_00010
Mention Str: Stephane Dujarric 601:617
Reference KB: St%C3%A9phane_Dujarric
Entity Type: PER
Mention Type: NAM
Confidence Value: 1.0
Images:

Bentiu ca. 395 m
Bentiu,Бентиу,Бентиjу,بانتيو,بانتيو,בנטיו,ბენტიუ,班提烏
P PPLA seat of a first-order administrative division 377472
South Sudan SS » Northern Liech 11550578
population : 7653
9.25722, 29.79833 N 9°15'26" E 29°47'54"

Linking to Geoname: Heatmap



sentence_18442

Language: Somali

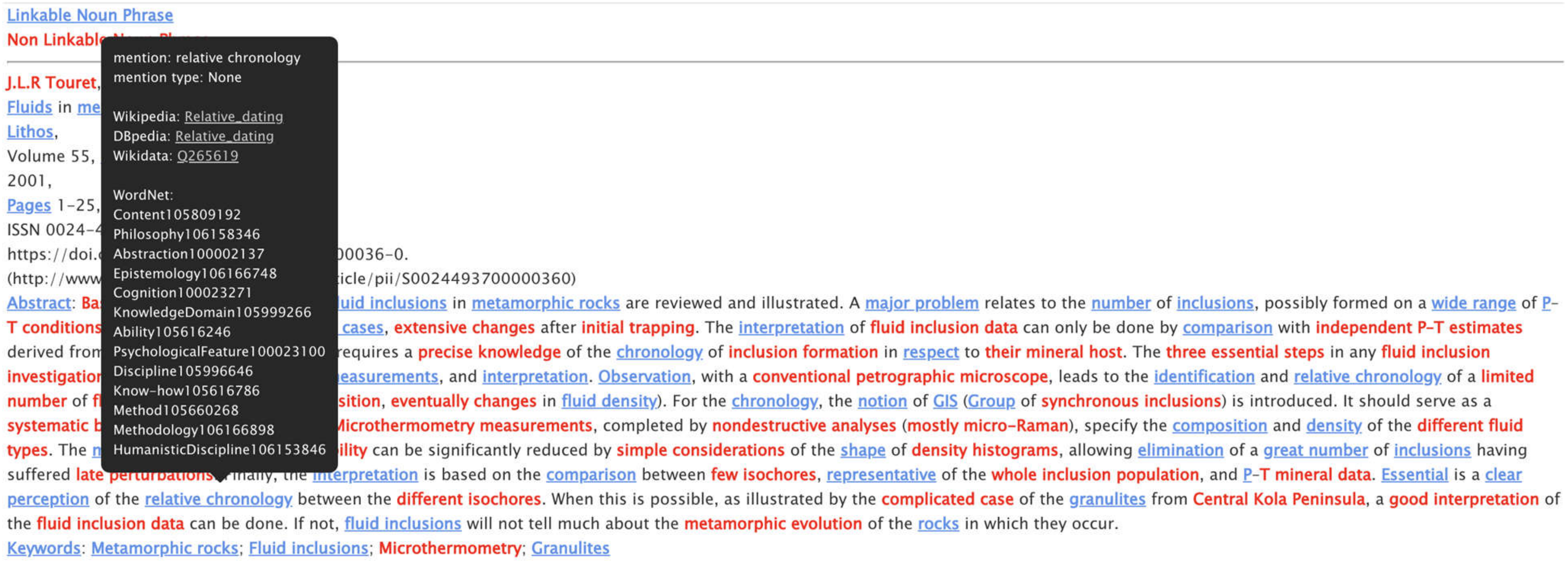
Subagle wuxuu sheegay in kooxda Al-Shabaab ay isku dayday inay dib u qabsato magaalada Buulo-burte, oo toddobaad ka hor ay qabsadeen ciidanka xoogga **dalka Soomaaliya** oo gacan ka helaya kuwa dalka Jabuuti ee qeybta ka ah AMISOM.

Translation: Subagle said al-Shabaab militants attempted to recapture the town of Bulo-burte, which had taken control of the Somali strongholds in Djibouti with the help of AMISOM troops.

- Re-trainable Systems: http://blender02.cs.rpi.edu:3300/elisa_ie/api
- Data and Resources: http://nlp.cs.rpi.edu/wikiann/
- Demos: http://blender02.cs.rpi.edu:3300/elisa_ie http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap

Linking to Wikipedia/DBPedia/YAGO/WikiData/ WordNet: Earth Science Domain

- Earth Science
  - ○ https://blender04.cs.rpi.edu/~panx2/tmp/phrase_linking/20180419/



Linkable Noun Phrase

Non Linkable

mention: relative chronology
mention type: None

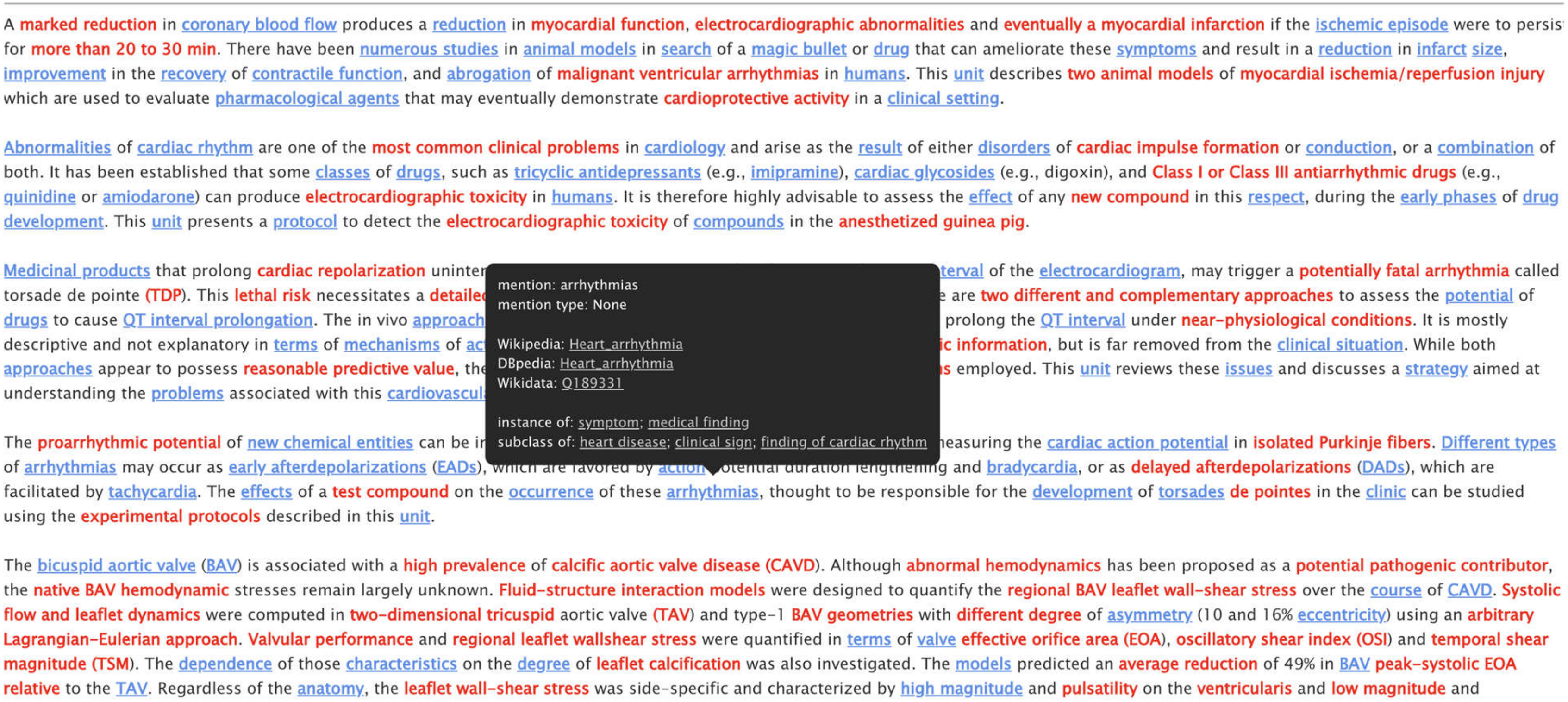
Wikipedia: Relative_dating
DBpedia: Relative_dating
Wikidata: Q265619

WordNet:
Content105809192
Philosophy106158346
Abstraction100002137
Epistemology106166748
Cognition100023271
KnowledgeDomain105999266
Ability105616246
PsychologicalFeature100023100
Discipline105996646
Know-how105616786
Method105660268
Methodology106166898
HumanisticDiscipline106153846

J.L.R Touret,
Fluids in me
Lithos,
Volume 55,
2001,
Pages 1–25,
ISSN 0024–4
https://doi. 00036–0.
(http://www icle/pii/S0024493700000360)
Abstract: Ba luid inclusions in metamorphic rocks are reviewed and illustrated. A major problem relates to the number of inclusions, possibly formed on a wide range of P–
T conditions cases, extensive changes after initial trapping. The interpretation of fluid inclusion data can only be done by comparison with independent P–T estimates
derived from requires a precise knowledge of the chronology of inclusion formation in respect to their mineral host. The three essential steps in any fluid inclusion
investigation easurements, and interpretation. Observation, with a conventional petrographic microscope, leads to the identification and relative chronology of a limited
number of fl sition, eventually changes in fluid density). For the chronology, the notion of GIS (Group of synchronous inclusions) is introduced. It should serve as a
systematic b Microthermometry measurements, completed by nondestructive analyses (mostly micro–Raman), specify the composition and density of the different fluid
types. The n ility can be significantly reduced by simple considerations of the shape of density histograms, allowing elimination of a great number of inclusions having
suffered late perturbations inally, the interpretation is based on the comparison between few isochores, representative of the whole inclusion population, and P–T mineral data. Essential is a clear
perception of the relative chronology between the different isochores. When this is possible, as illustrated by the complicated case of the granulites from Central Kola Peninsula, a good interpretation of
the fluid inclusion data can be done. If not, fluid inclusions will not tell much about the metamorphic evolution of the rocks in which they occur.
Keywords: Metamorphic rocks; Fluid inclusions; Microthermometry; Granulites

Linking to 300+ Bioportal Biomedical Ontologies

- Biomedical Science
  - https://blender04.cs.rpi.edu/~panx2/tmp/phrase_linking/20180419_bio/head10.html



A marked reduction in coronary blood flow produces a reduction in myocardial function, electrocardiographic abnormalities and eventually a myocardial infarction if the ischemic episode were to persis
for more than 20 to 30 min. There have been numerous studies in animal models in search of a magic bullet or drug that can ameliorate these symptoms and result in a reduction in infarct size, improvement in the recovery of contractile function, and abrogation of malignant ventricular arrhythmias in humans. This unit describes two animal models of myocardial ischemia/reperfusion injury which are used to evaluate pharmacological agents that may eventually demonstrate cardioprotective activity in a clinical setting.

Abnormalities of cardiac rhythm are one of the most common clinical problems in cardiology and arise as the result of either disorders of cardiac impulse formation or conduction, or a combination of both. It has been established that some classes of drugs, such as tricyclic antidepressants (e.g., imipramine), cardiac glycosides (e.g., digoxin), and Class I or Class III antiarrhythmic drugs (e.g., quinidine or amiodarone) can produce electrocardiographic toxicity in humans. It is therefore highly advisable to assess the effect of any new compound in this respect, during the early phases of drug development. This unit presents a protocol to detect the electrocardiographic toxicity of compounds in the anesthetized guinea pig.

Medicinal products that prolong cardiac repolarization uninter
torsade de pointe (TDP). This lethal risk necessitates a detailed
drugs to cause QT interval prolongation. The in vivo approach
descriptive and not explanatory in terms of mechanisms of act
approaches appear to possess reasonable predictive value, the
understanding the problems associated with this cardiovascul

terval of the electrocardiogram, may trigger a potentially fatal arrhythmia called
e are two different and complementary approaches to assess the potential of
prolong the QT interval under near-physiological conditions. It is mostly
ic information, but is far removed from the clinical situation. While both
s employed. This unit reviews these issues and discusses a strategy aimed at

mention: arrhythmias
mention type: None

Wikipedia: Heart_arrhythmia
DBpedia: Heart_arrhythmia
Wikidata: Q189331

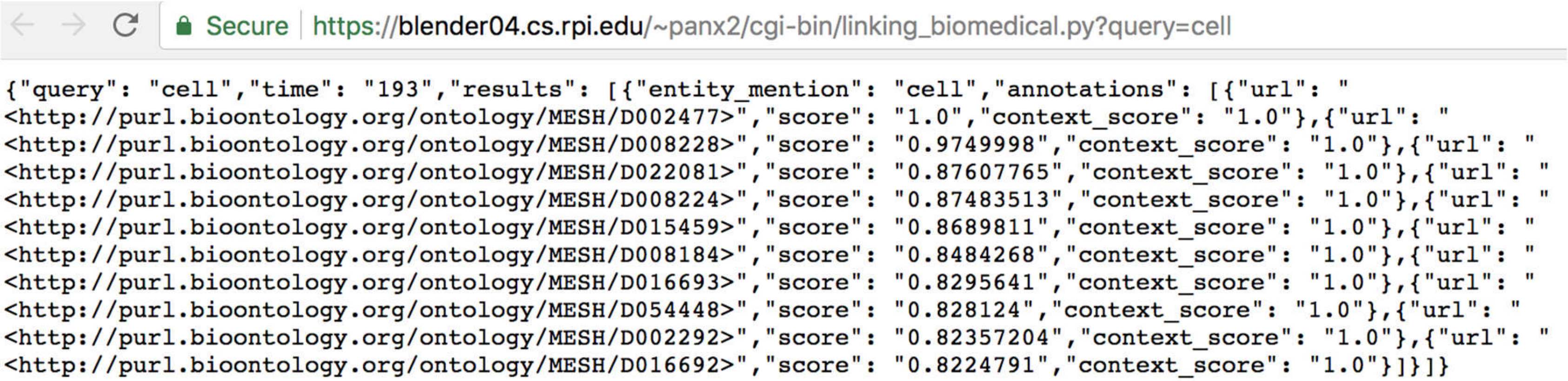
instance of: symptom; medical finding
subclass of: heart disease; clinical sign; finding of cardiac rhythm

The proarrhythmic potential of new chemical entities can be ir
easuring the cardiac action potential in isolated Purkinje fibers. Different types of arrhythmias may occur as early afterdepolarizations (EADs), which are favored by action otential duration lengthening and bradycardia, or as delayed afterdepolarizations (DADs), which are facilitated by tachycardia. The effects of a test compound on the occurrence of these arrhythmias, thought to be responsible for the development of torsades de pointes in the clinic can be studied using the experimental protocols described in this unit.

The bicuspid aortic valve (BAV) is associated with a high prevalence of calcific aortic valve disease (CAVD). Although abnormal hemodynamics has been proposed as a potential pathogenic contributor, the native BAV hemodynamic stresses remain largely unknown. Fluid-structure interaction models were designed to quantify the regional BAV leaflet wall-shear stress over the course of CAVD. Systolic flow and leaflet dynamics were computed in two-dimensional tricuspid aortic valve (TAV) and type-1 BAV geometries with different degree of asymmetry (10 and 16% eccentricity) using an arbitrary Lagrangian-Eulerian approach. Valvular performance and regional leaflet wallshear stress were quantified in terms of valve effective orifice area (EOA), oscillatory shear index (OSI) and temporal shear magnitude (TSM). The dependence of those characteristics on the degree of leaflet calcification was also investigated. The models predicted an average reduction of 49% in BAV peak-systolic EOA relative to the TAV. Regardless of the anatomy, the leaflet wall-shear stress was side-specific and characterized by high magnitude and pulsatility on the ventricularis and low magnitude and

Linking to 300+ Biomedical Ontologies (Zheng et al., 2014)

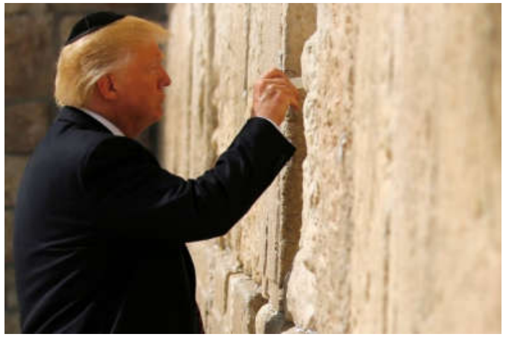
- Biomedical Science
  - https://blender04.cs.rpi.edu/~panx2/cgi-bin/linking_biomedical.py?query=cell


Secure
https://blender04.cs.rpi.edu/~panx2/cgi-bin/linking_biomedical.py?query=cell

```
{"query": "cell","time": "193","results": [{"entity_mention": "cell","annotations": [{"url": "
<http://purl.bioontology.org/ontology/MESH/D002477>","score": "1.0","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D008228>","score": "0.9749998","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D022081>","score": "0.87607765","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D008224>","score": "0.87483513","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D015459>","score": "0.8689811","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D008184>","score": "0.8484268","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D016693>","score": "0.8295641","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D054448>","score": "0.828124","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D002292>","score": "0.82357204","context_score": "1.0"},{"url": "
<http://purl.bioontology.org/ontology/MESH/D016692>","score": "0.8224791","context_score": "1.0"}]}]}
```

The Future of EDL: A broader notion of Coherence

- A long tail of relational information impacts accurate EDL
  - Beyond (Mubarak, wife, Mubarak)
  - Geographic Constraints
  - Temporal Constraints



- Multimodal Coherence
  - Identifying People, Locations, Objects
  - Concrete Objects?
  - **Trump's controversial visit to the Western Wall**
    - Time of article would also help
- Requires Reasoning

The Future of EDL: Mentions

- Ground Everything
  - Other units of text; into multiple resources
  - Once we expand the notion of a mention, a few problems arise
  - **Manchester Bomber Believed Muslims Were Mistreated, Sought Revenge**

- The name of the person?
- The concept of a Bomber?
- The city?
- The Event?

  - **The US has now managed to upset two of its closest allies by allowing the disclosure of sensitive information -**

- The concept
- The allies?
  - **Harder**

  - But not everything needs to be grounded every time
    - Notion of importance
    - Ideally, adaptive to the user
    - Interaction?

The Future of EDL: Mentions

- Ground Everything
  - Other units of text; into multiple resources
  - Even **Events**
    - Requires good definitions; requires reasoning
- Ground Email & Other Social Media
  - Into KB, Encyclopedic Resources; other E-mails and Posts
    - Creative names: “Hellary” (probably meaningful beyond grounding)
  - Better integration between cross-document coref and grounding
    - Reference to earlier posts is important
- Grounding as a Cultural Expert
  - Read fiction and learn about the world
    - EDL as a source of supervision
  - Ground into historical references
    - It was the age of wisdom, it was the age of foolishness…
- Temporal grounding
  - Multiple interpretations

The Future of EDL: Languages

- Ground from many languages
  - Also into other languages?
  - Important in some domains (e.g. Health)

- Size
  - Grounding into many languages (today) means carrying 10G of cross lingual embeddings
  - Installing on you own machines becomes difficult/infeasible

Outline

- Motivation, Task, Application 15 min
- Traditional “Ancient” Approaches 15 min
- Modern Approaches
  - Language Universal EDL 30min
  - Multi-lingual Common Space Construction 30min

Coffee Break

  - Cross-lingual Transfer Learning 20min
  - Cross-lingual Neural Entity Linking 20min
- Remaining Challenges and New Directions 30min
- Demos, Resources and QA 20min

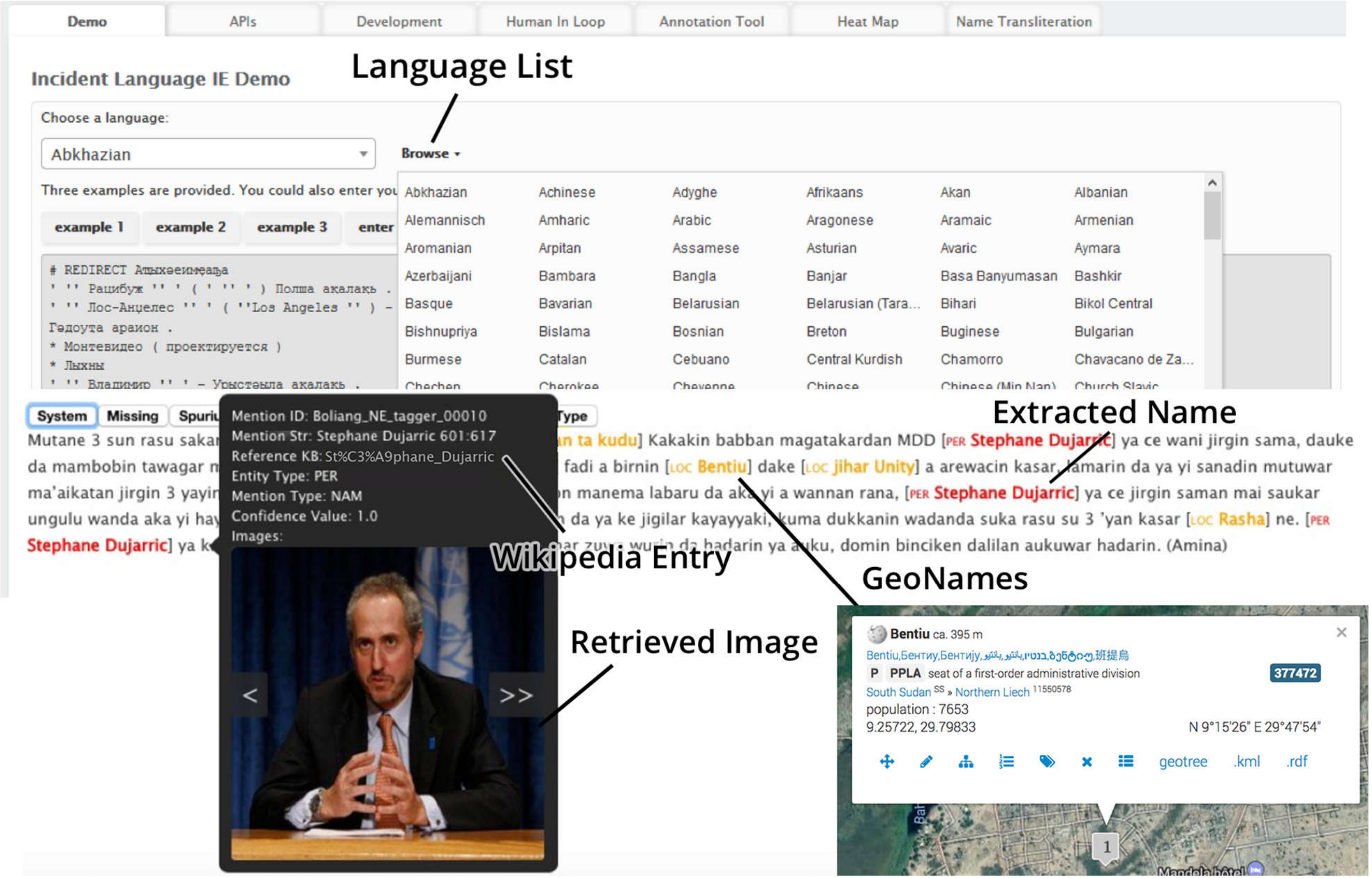
References

- **Tutorial:** **http://nlp.cs.rpi.edu/ccl.pptx**
- **Reading List:** **http://nlp.cs.rpi.edu/kbp/2018/elreading.html**

Systems, Data and Resources Publicly Available

- Cross-lingual IE
  - Re-trainable Systems:
  - http://blender02.cs.rpi.edu:3300/elisa_ie/api
  - Data and Resources:
  - http://nlp.cs.rpi.edu/wikiann/
  - Demos:
  - http://blender02.cs.rpi.edu:3300/elisa_ie
  - http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap
  - https://elisa-ie.github.io/heatmap/demo/
- Cross-media IE
  - http://nlp.cs.rpi.edu/multimedia/event2017/navigation_dark.html

Demo: Cross-lingual EDL for 282 Languages



Incident Language IE Demo

Choose a language:

Abkhazian

Browse ▾

Abkhazian, Achinese, Adyghe, Afrikaans, Akan, Albanian
Alemannisch, Amharic, Arabic, Aragonese, Aramaic, Armenian
Aromanian, Arpitan, Assamese, Asturian, Avaric, Aymara
Azerbaijani, Bambara, Bangla, Banjar, Basa Banyumasan, Bashkir
Basque, Bavarian, Belarusian, Belarusian (Tara..., Bihari, Bikol Central
Bishnupriya, Bislama, Bosnian, Breton, Buginese, Bulgarian
Burmese, Catalan, Cebuano, Central Kurdish, Chamorro, Chavacano de Za...
Chechen, Cherokee, Cheyenne, Chinese, Chinese (Min Nan), Church Slavic

Three examples are provided. You could also enter you

example 1 | example 2 | example 3 | enter

```
# REDIRECT Алыхәеимҿаҕа
' '' Рацибуж '' ' ( ' '' ' ) Полша акалакь .
' '' Лос-Анџелес '' ' ( ''Los Angeles '' ) –
Гәдоута араион .
* Монтевидео ( проектируется )
* Лыхны
' '' Владимир '' ' – Урыстәыла акалакь .
```

System | Missing | Spuri... | ...Type

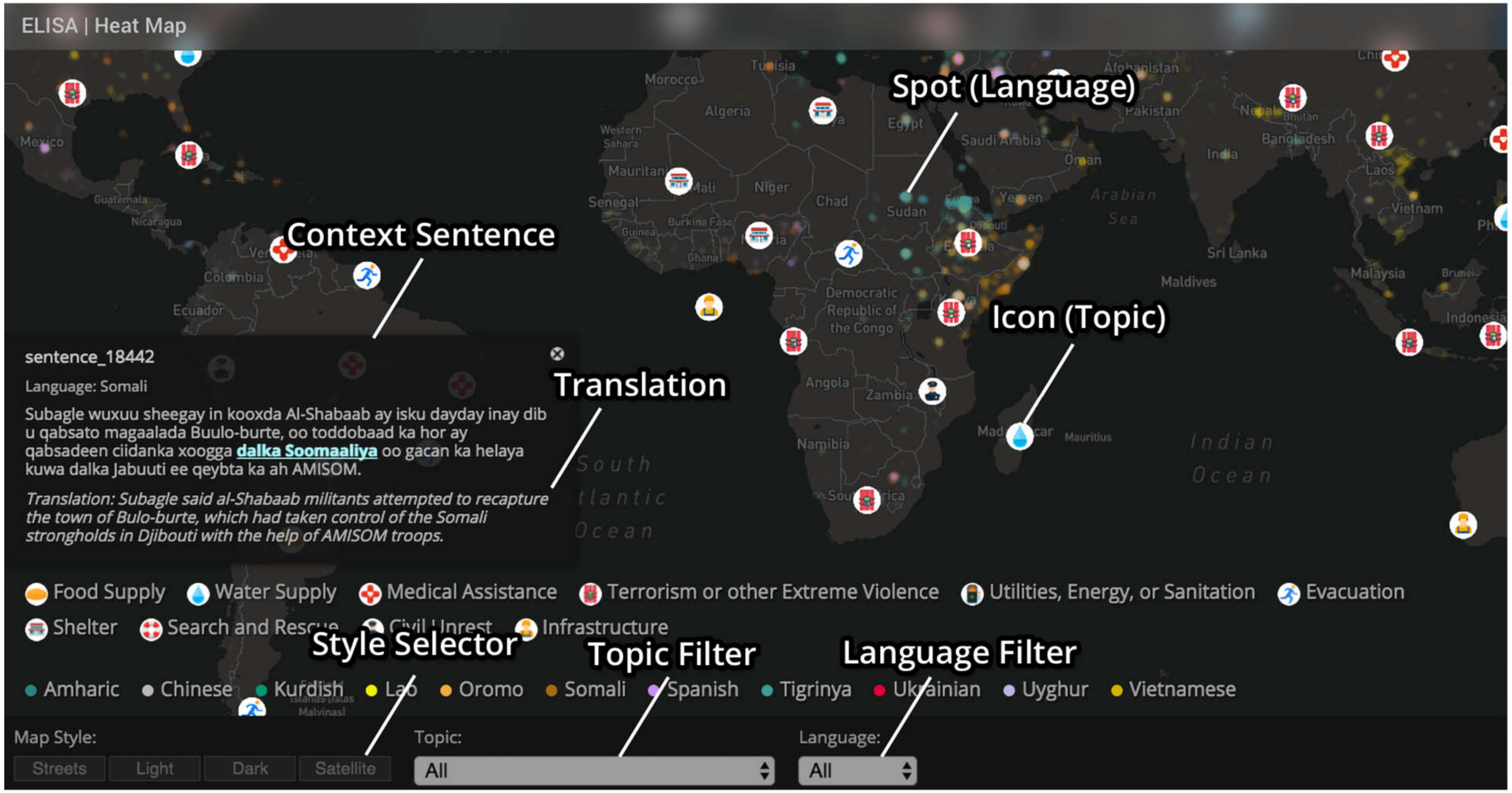
Mutane 3 sun rasu sakar ... an ta kudu] Kakakin babban magatakardan MDD [PER Stephane Dujarric] ya ce wani jirgin sama, dauke da mambobin tawagar m ... fadi a birnin [LOC Bentiu] dake [LOC jihar Unity] a arewacin kasar, lamarin da ya yi sanadin mutuwar ma'aikatan jirgin 3 yayin ... n manema labaru da aka yi a wannan rana, [PER Stephane Dujarric] ya ce jirgin saman mai saukar ungulu wanda aka yi hay ... da ya ke jigilar kayayyaki, kuma dukkanin wadanda suka rasu su 3 'yan kasar [LOC Rasha] ne. [PER Stephane Dujarric] ya k ... ar zuwa wurin da hadarin ya auku, domin binciken dalilan aukuwar hadarin. (Amina)

Mention ID: Boliang_NE_tagger_00010
Mention Str: Stephane Dujarric 601:617
Reference KB: St%C3%A9phane_Dujarric
Entity Type: PER
Mention Type: NAM
Confidence Value: 1.0
Images:

< >>

Bentiu ca. 395 m
Bentiu,Бентиу,Бентију,بانتیو,بانتيو,בנטיו,ბენთიუ,班提烏
P PPLA seat of a first-order administrative division 377472
South Sudan SS » Northern Liech 11550578
population : 7653
9.25722, 29.79833 N 9°15'26" E 29°47'54"
geotree .kml .rdf

Demo: Cross-lingual EDL for 282 Languages



sentence_18442

Language: Somali

Subagle wuxuu sheegay in kooxda Al-Shabaab ay isku dayday inay dib u qabsato magaalada Buulo-burte, oo toddobaad ka hor ay qabsadeen ciidanka xoogga **dalka Soomaaliya** oo gacan ka helaya kuwa dalka Jabuuti ee qeybta ka ah AMISOM.

Translation: Subagle said al-Shabaab militants attempted to recapture the town of Bulo-burte, which had taken control of the Somali strongholds in Djibouti with the help of AMISOM troops.

Spot (Language)

Context Sentence

Icon (Topic)

Translation

Style Selector

Topic Filter

Language Filter

- Re-trainable Systems: http://blender02.cs.rpi.edu:3300/elisa_ie/api
- Data and Resources: http://nlp.cs.rpi.edu/wikiann/
- Demos: http://blender02.cs.rpi.edu:3300/elisa_ie http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap

Thank You

NSF ARL DARPA Google IBM AIR FORCE RESEARCH LABORATORY

ALFRED P. SLOAN FOUNDATION 1934 BOSCH U.S. DEPARTMENT OF HOMELAND SECURITY Disney IARPA BE THE FUTURE

Home Publications

Overview

Recent years have wit
wide spectrum of hete
speech, and multimed
Much of this informatio
makes it accessible to
can read that particula
computer processing
research area of Blend
extraction (IE) on a ma
generation of informat
communicate with co
keyword search, and
concise, and trustable