

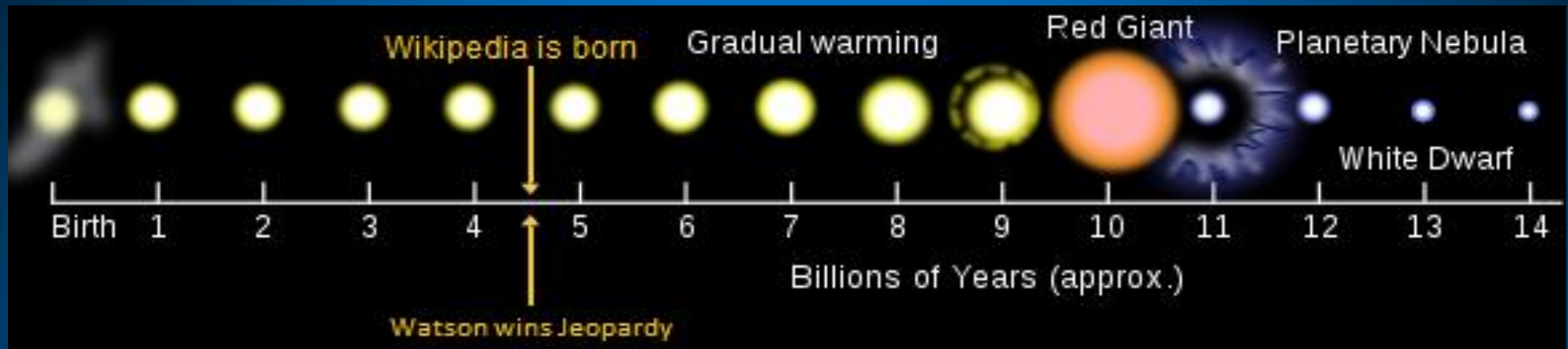
Entity Extraction and Linking with Applications to Web Search

Silviu Cucerzan

Machine Learning and Intelligence

MSR

Where We Are in the Semantic Space



- Game changer: Wikipedia + hardware advancements
small number of entity classes → large entity collection → Watson wins jeopardy
- Current state:
arguably, current entity extractions systems outperform non-expert humans on entity disambiguation for random documents/domains.
- Next big steps:
 - move from one large entity collection to many thousands of domain specific collections, business collections, and personal collections
 - learn to link to Web entities

NIST/LDC Entity Linking Evaluations

Knowledge base derived from Wikipedia Oct. 2008

818,741
entries

Text corpus:
(for the 2013 evaluation)

Newswire: 1 mil docs
Web Text: 1 mil docs
Discussion Forums: .5 mil docs
2.5 million documents

Annotated Data:

			Total mentions	PER	ORG	GPE
2009	Evaluation	Newswire	3904	627	2710	567
2010	Training	Web data	1500	500	500	500
	Evaluation	Newswire	1500	500	500	500
2011	Evaluation	Web data	750	250	250	250
		Newswire	1491	500	491	500
2012	Evaluation	Web data	759	250	259	250
		Newswire	1471	702	388	381
2013	Evaluation	Web data	755	216	318	221
		News and Web	1477	686	701	803
2014	Evaluation	Forums	713			
		News and Web	3318	3162	1007	1065
		Forums	1916			

NIST/LDC EL - Task Description

For a string at a given offset in a document, determine which entity from the provided knowledge base (if any) is being referred to by the string. Cluster all entities in the test set.

```
<query id="EL005833">
```

```
<name>IAF</name>
```

```
<docid>eng-WL-11-174596-12954631</docid>
```

```
<offset>...</offset>
```

```
</query>
```

```
<query id="EL005836">
```

```
<name>IAF</name>
```

```
<docid>eng-NG-31-142148-10021195</docid>
```

```
<offset>...</offset>
```

```
</query>
```

```
<query id="EL05838">
```

```
<name>IAF</name>
```

```
<docid>eng-WL-11-174596-12954257</docid>
```

```
<offset>...</offset>
```

```
</query>
```

```
<query id="EL05847">
```

```
<name>IAF</name>
```

```
<docid>eng-NG-31-147166-10475895</docid>
```

```
<offset>...</offset>
```

```
</query>
```

Israeli Air Force

Islamic Academy
of Florida

Israeli Air Force

Indian Air Force

LDC Knowledge Base



E0123252: Italian Air Force

E0247721: Iraqi Air Force

E0265128: Israeli Air Force

E0290069: Indonesian Air Force

E0384804: Italian Armed Forces

E0707328: Indian Armed Forces

...

NIL

Example: which “IAF”? *Is answering just a text lookup?*

```
<query id="EL005833">
  <name>IAF</name>
  <docid>eng-WL-11-174596-12954631</docid>
  <offset>...</offset>
</query>
<query id="EL005836">
  <name>IAF</name>
  <docid>eng-NG-31-142148-10021195</docid>
  <offset>...</offset>
</query>
<query id="EL05838">
  <name>IAF</name>
  <docid>eng-WL-11-174596-12954257</docid>
  <offset>...</offset>
</query>
<query id="EL05847">
  <name>IAF</name>
  <docid>eng-NG-31-147166-10475895</docid>
  <offset>...</offset>
</query>
```

```
<DOCID> eng-WL-11-174596-12954631 </DOCID>
<DOCTYPE SOURCE="blog"> BLOG TEXT </DOCTYPE>
<DATETIME> 2008-05-24T12:55:00 </DATETIME>
<HEADLINE> Syria stalls IAEA visit... </HEADLINE>
<TEXT> <POST>
<POSTER> GayandRight </POSTER>
<POSTDATE> 2008-05-24T12:55:00 </POSTDATE>
Gee, I wonder why....
```

Syria has not yet accepted a request by the International Atomic Energy Agency to visit the site bombed by the IAF on September 6, which Washington says was a nuclear reactor, Reuters reported Friday.

The news agency quoted diplomats in Vienna as saying that Damascus was stalling its approval of the UN delegation visit, demanding more details on the proposed inspection.

Syrian atomic energy chief Ibrahim Othman came to Vienna earlier this month to speak with IAEA head Mohamed ElBaradei on the matter, but the two did not agree on the timing or nature of a visit, diplomats said.

The agency received a letter from Syria several days ago asking for more details on the trip, one US diplomat said. The IAEA has replied and is now waiting for Damascus's response, he added.

```
</POST>
</TEXT>
```

Israeli Air Force

Italian Air Force, Italian Armed Forces, Indonesian Air Force, Iraqi Air Force, Israeli Air Force, Indian Armed Forces

Knowledge Graph



Which “Washington”?

The quantum state of entities.

386 Wikipedia entities can be referred to as **Washington**
(based on the August 5, 2013 Wikipedia dump).

- Washington, D.C.
- United States
- United States Department of State
- Federal government of the United States
- ...

... which **Syria**?

... which **Damascus**?

The answer depends on:

- The granularity of the knowledge base
- The disambiguation of the other entities in the document

```
<DOCID> eng-WL-11-174596-12954631 </DOCID>
<DOCTYPE SOURCE="blog"> BLOG TEXT </DOCTYPE>
<DATETIME> 2008-05-24T12:55:00 </DATETIME>
<HEADLINE> Syria stalls IAEA visit... </HEADLINE>
<TEXT> <POST>
<POSTER> GayandRight </POSTER>
<POSTDATE> 2008-05-24T12:55:00 </POSTDATE>
Gee, I wonder why....
```

Syria has not yet accepted a request by the International Atomic Energy Agency to visit the site bombed by the IAF on September 6, which Washington says was a nuclear reactor, Reuters reported Friday.

The news agency quoted diplomats in Vienna as saying that Damascus was stalling its approval of the UN delegation visit, demanding more details on the proposed inspection.

Syrian atomic energy chief Ibrahim Othman came to Vienna earlier this month to speak with IAEA head Mohamed ElBaradei on the matter, but the two did not agree on the timing or nature of a visit, diplomats said.

The agency received a letter from Syria several days ago asking for more details on the trip, one US diplomat said. The IAEA has replied and is now waiting for Damascus's response, he added.

```
</POST>
</TEXT>
```


MSR Entity Extraction and Linking System

- The best evidence for entity disambiguation is the set of co-occurring entities (rather than the plain text)



- Extract and disambiguate all entities in a target document



- Employ document-based features as well as local features

Knowledge Base

Surface Forms

[Washington](#) were defeated by the [Eagles](#).

distribution over entities

Wikipedia sources:

- anchor text of interlinks
- processed page titles
- redirects
- infobox fields

OTHER ENTITIES ASSOCIATED WITH THIS NAME (386)

[George Washington](#)
[George Washington Carver](#)
[University of Washington](#)
[Washington Redskins](#)
[Washington Irving](#)
[Booker T. Washington](#)
[Harold Washington](#)
[United States Naval Observatory](#)
[Washington County, Maryland](#)
[Washington County, Oregon](#)
[Washington County, New York](#)
[Washington University in St. Louis](#)
[Washington County, Utah](#)
[Washington Wizards](#)
[Washington Capitals](#)
[Washington County, Florida](#)
[Washington County, Alabama](#)
[Washington County, Wisconsin](#)
[Washington County, Virginia](#)
[Washington County, Vermont](#)
[Washington, Utah](#)
[Washington Terrace, Utah](#)
[Washington, Vermont](#)
[Washington, Virginia](#)
[Washington, West Virginia](#)
[Mount Washington \(New Hampshire\)](#)
[Martha Washington](#)
[Dinah Washington](#)
[USS Washington](#)
[Federal government of the United States](#)
[USS Washington \(BB-56\)](#)
[Washington State](#)
[Lake Washington](#)
[Washington State University](#)
[Washington Mystics](#)
[William Washington](#)
[Washington and Lee University](#)
[Washington, Tyne and Wear](#)
[George Washington University](#)
[Washington Bartlett](#)
[Union Station \(Washington, D.C.\)](#)

Washington Redskins	Washington	1	Washington	0	y	0	0.026	16715
Philadelphia Eagles	Eagles	1	Eagles	32	y	B	0.107	11556

Knowledge Base



Information associated with each entity:

- **Topics**

Wikipedia categories, list pages, interlinked entity mentions in enumerations, interlinked entity mentions in tables, ...

- **Contexts**

parentheticals in titles, infobox information, frequently co-occurring words

- **Triggers**

Wikipedia bidirectional linkage

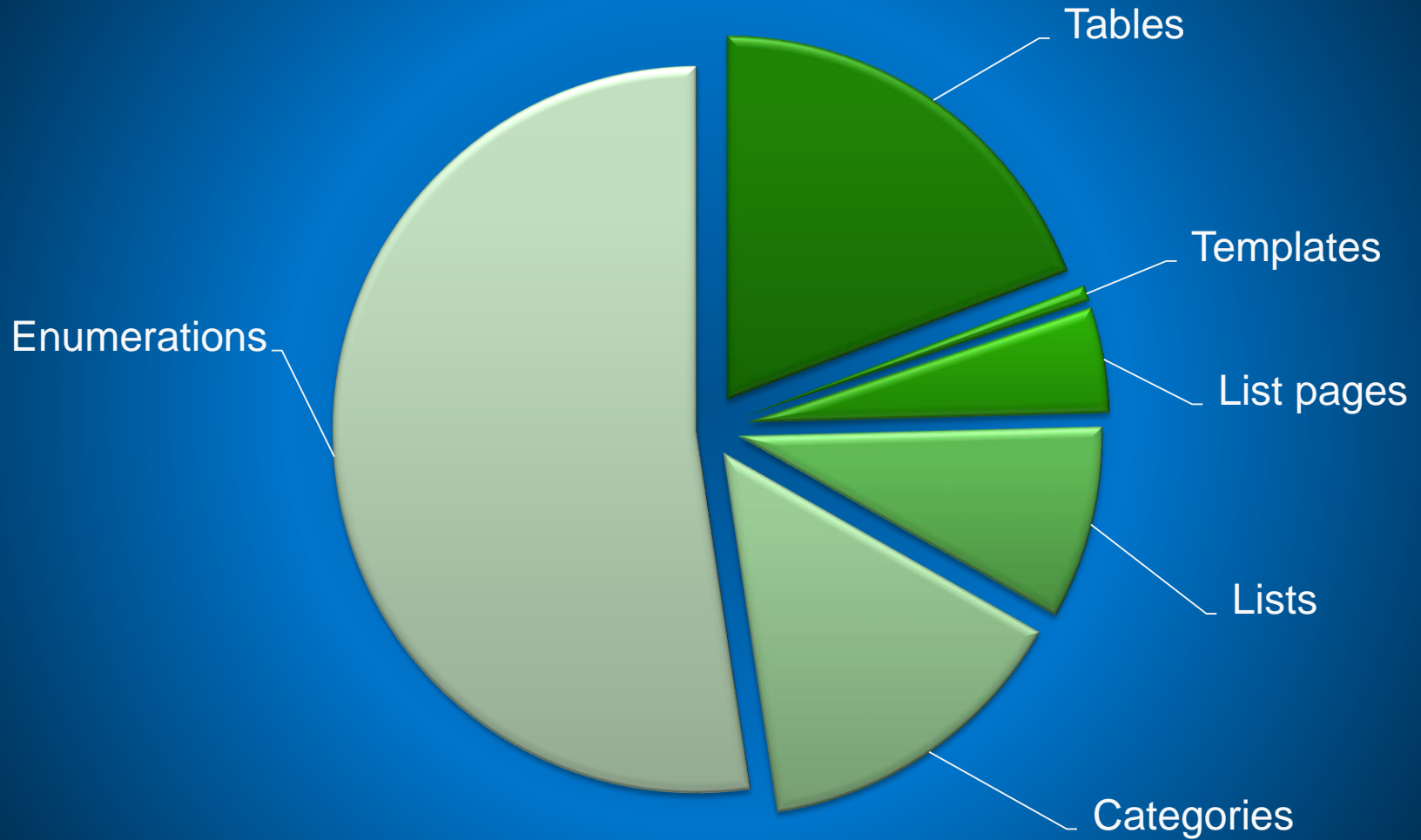
- **Entity Types**

14 types: Disambiguation, Common, Person, Geo-political entity, Location, Organization, Event, Vehicle, Work of art, ..., Other

- **Geo-coordinates**

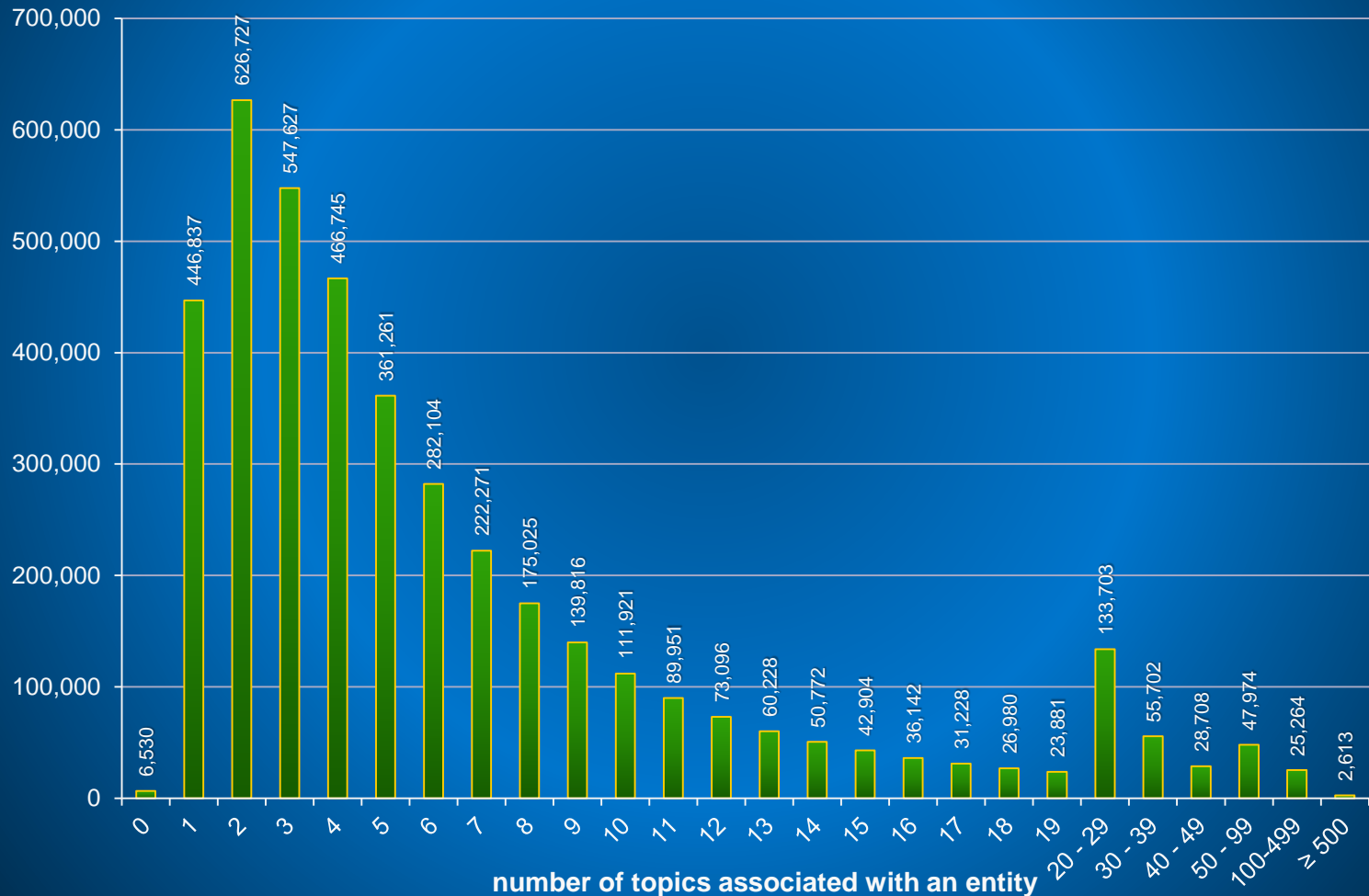
- **Time information**

Knowledge Base – Topics Stats (1)



Knowledge Base – Topics Stats (2)

number of entities



Knowledge Base



Linguistic Resources

Derived from the Wikipedia collection:

- name normalization

e.g.:

Ben	Benjamin, Benjy, Benedict
Bernard	Bernie, Barney, Bernardus, Bernhard
Betty	Elizabeth, Beth, Betsy
Bill	William, Billy, Bil

- entity-type contexts

e.g.:

result in →	B 216	C 456	D 4	L 8	M 4	O 2	P 14	T 6
← was founded.	B 18	G 2	L 42	M 2	O 198	P 8		

- word-capitalization statistics

...

System Architecture

The analysis of an input text is done in three stages, with the following main roles:

- Stage 1
 - text normalization
 - sentence breaking
- Stage 2:
 - surface form boundary detection
- Stage 3: disambiguation
 - latent document model construction
 - feature computation
 - entity candidate ranking

Surface Form Detection

- Capitalization, lexical resources, known surface forms
- Soft boundaries

TAC 2011 test: query: EL_00279

string: Bordeaux

doc: AFP_ENG_20081006.0534.LDC2009T13.sgm

e.g.: *Bordeaux-based wine merchant, Jeffrey Davies, said that while the crisis triggered by the terror attacks on New York and Washington had hit **US wine** sales, the economic meltdown had global implications. [...]*

"The big spenders that were ordering the top wines in top restaurants have been taken out," Davies said.

*After the attacks, sales of **Bordeaux wine** to the United States fell by 29 percent in volume during the final quarter of 2001 -- the key Thanksgiving, Christmas and New Year period, which accounts for half of annual sales.*

Composite(Bordeaux, Bordeaux wine)

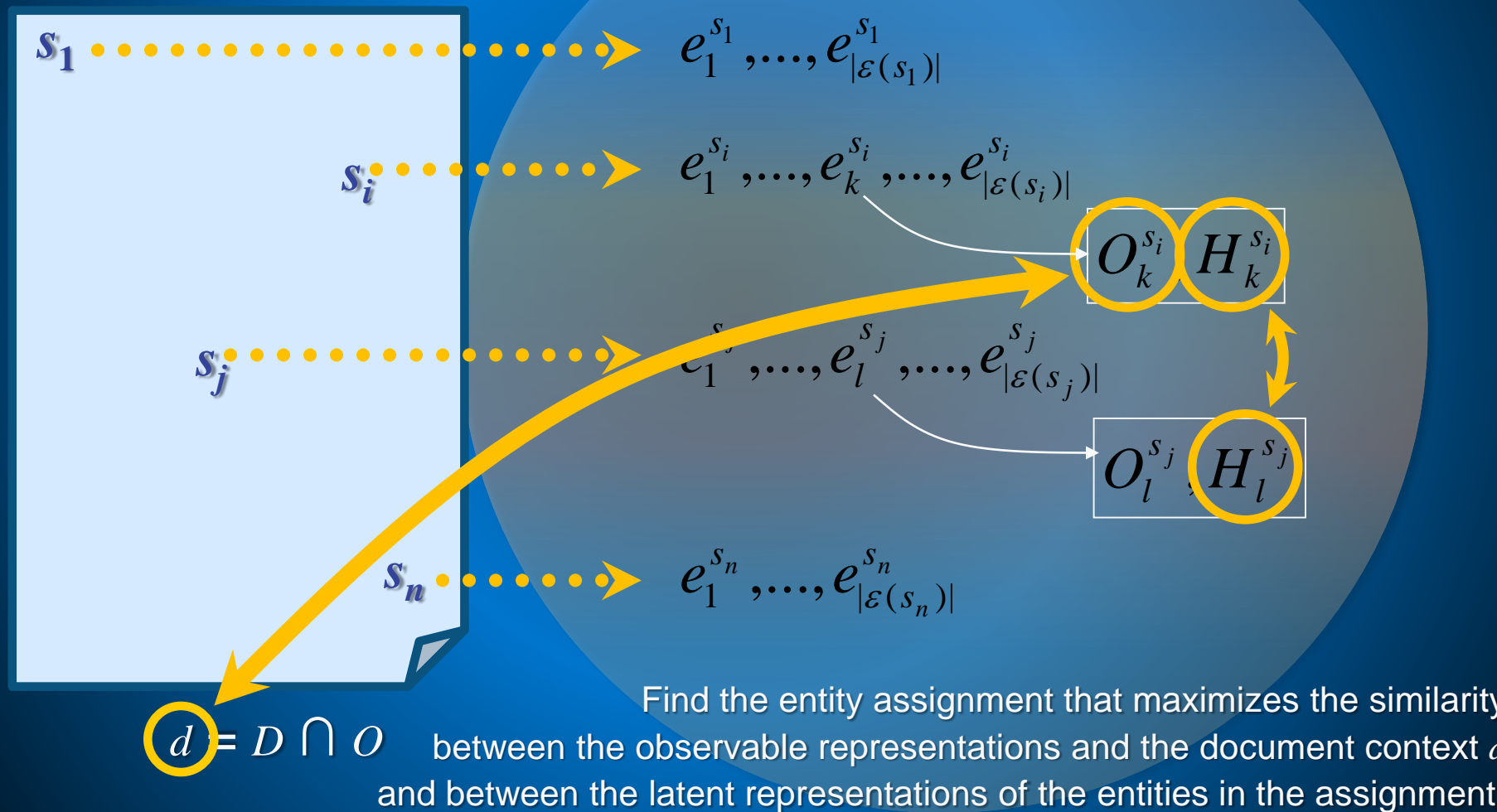
Composite(US, US wine)

Employ **composite surface forms** and let the disambiguation process determine the best entity in the context → resolve boundaries afterwards

Disambiguation - Intuition

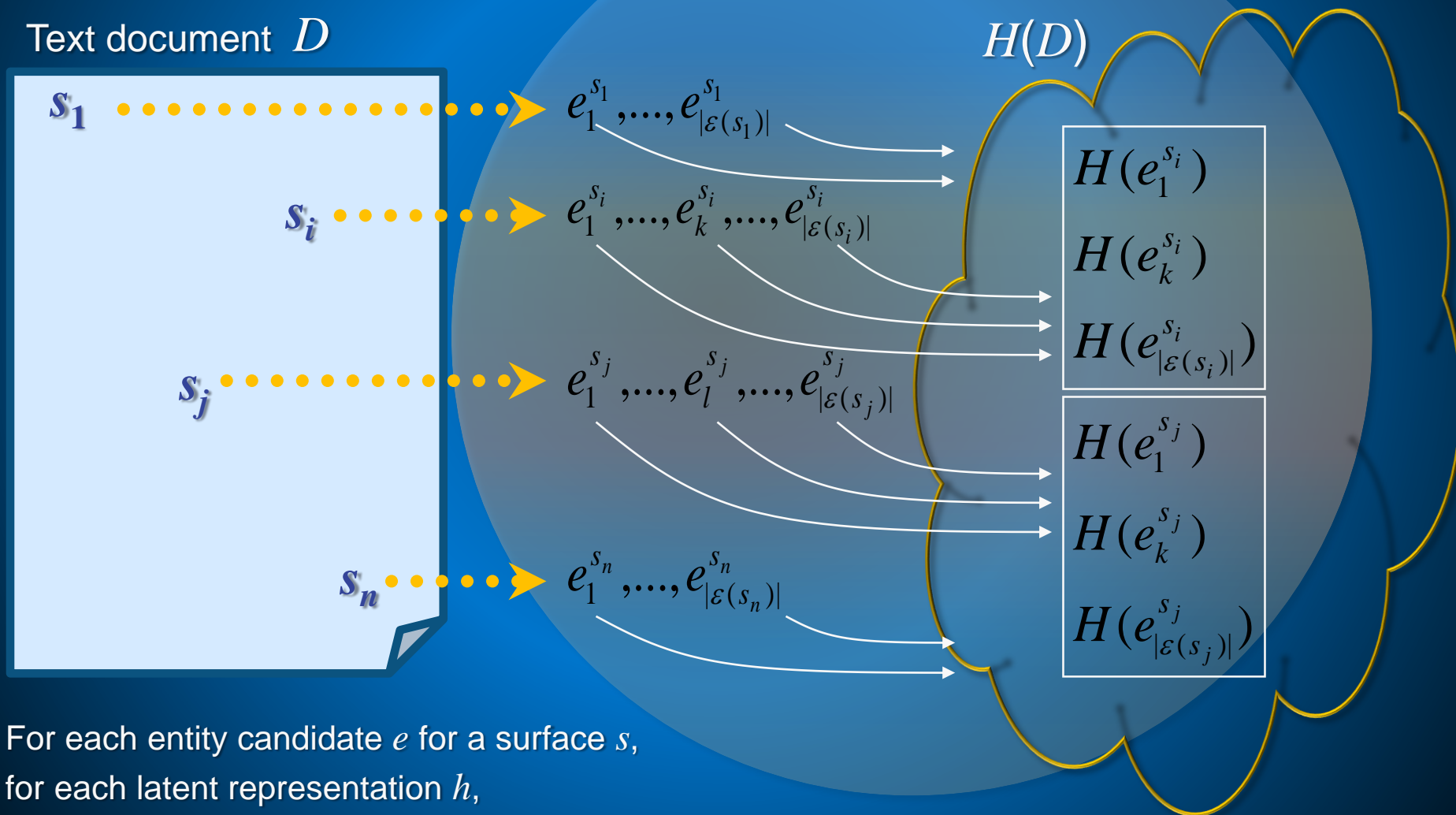
Each entity has multiple vectorial representations

Text document D



The Latent Document Representation

Build (noisy) latent vectorial representations of the document by aggregating the latent vectors from all entity candidates.



For each entity candidate e for a surface s ,
for each latent representation h ,
compute the similarity between $h(e)$ and $h(D) - h(s)$

Latent Features

And during that [Super Bowl](#) week, the hottest topic of conversation was [Peyton Manning](#), not his younger brother [Eli](#), who wound up leading the [New York Giants](#) to the title.

[Arizona](#), [Miami](#), [Tennessee](#), [Washington](#) and the [New York Jets](#) all have been rumored as possible destinations now; [Washington Redskins](#) [offense](#) [hater](#) in [Indianapolis](#), [Tom Moore](#), worked for the [Jets](#) as a [consultant](#) last season.

"There will be no other [Peyton Manning](#)," [Irsay](#) said, adding that he hoped Wednesday's joint appearance would serve to "honor incredible memories and incredible things that he's done for the franchise, for the city, for the state."

This marks the end of a strong marriage between a player and team.

After being a [No. 1](#) draft pick himself, [Manning](#) started every meaningful game for 13 seasons in [Indianapolis](#) - 227 in a row, including the playoffs - and took the [Colts](#) from perennial also-ran to one of the [NFL's](#) model franchises and the [2007 Super Bowl](#) title.

Local Features

- Depart from the one-sense-per-discourse paradigm
- Employ a latent paragraph model in addition to the latent document model
- Employ lexico-syntactic patterns to weight the latent contributions to the paragraph model

e.g.:

possessive constructions: “*Denmark’s Prime Minister visited Romania*”

conjunctive constructions: “*Denmark and Romania*”

Training (1)



Battle of Waterloo

From Wikipedia, the free encyclopedia

Coordinates:  50°41'N 4°24'E

The **Battle of Waterloo** was fought on Sunday, 18 June 1815, near [Waterloo](#) in present-day [Belgium](#), then part of the [United Kingdom of the Netherlands](#). An [Imperial French](#) army under the command of [Emperor Napoleon](#) was defeated by the armies of the [Seventh Coalition](#), comprising an Anglo-allied army under the command of the [Duke of Wellington](#) combined with a [Prussian](#) army under the command of [Gebhard von Blücher](#). It was the culminating battle of the [Waterloo Campaign](#) and Napoleon's last. The defeat at Waterloo ended his rule as Emperor of the French, marking the end of his [Hundred Days](#) return from exile.

Upon Napoleon's return to power in 1815, many states that had opposed him formed the Seventh Coalition and began to mobilise armies. Two large forces under Wellington and Blücher assembled close to the north-eastern border of France. Napoleon chose to attack in the hope of destroying them before they could join in a coordinated invasion of France with other members of the coalition. The decisive engagement of this three-day Waterloo Campaign (16–19 June 1815) occurred at the Battle of Waterloo. According to Wellington, the battle was "the nearest-run thing you ever saw in your life".^[a]

Battle of Waterloo

Part of the [War of the Seventh Coalition](#)



Battle of Waterloo by [William Sadler](#)

Date 18 June 1815

This text is about [\[\[Battle of Waterloo|Waterloo\]\]](#). Allegedly, Napoleon tried to escape to North America, but the [\[\[Royal Navy|Royal Navy\]\]](#) was blockading French ports to forestall such a move. He finally surrendered to [\[\[Captain \(Royal Navy\)|Captain\]\]](#) [\[\[Frederick Lewis Maitland \(Royal Navy officer\)|Frederick Maitland\]\]](#) of [\[\[Her Majesty's Ship|HMS\]\]](#) "[\[\[HMS Bellerophon \(1786\)|Bellerophon\]\]](#)" on 15 July. There was a campaign against French fortresses that still held out; [\[\[Longwy|Longwy\]\]](#) capitulated on 13 September 1815, the last to do so. The [\[\[Treaty of Paris \(1815\)|Treaty of Paris\]\]](#) was signed on 20 November 1815. [\[\[Louis XVIII of France|Louis XVIII\]\]](#) was restored to the throne of France, and Napoleon was exiled to [\[\[Saint Helena|Saint Helena\]\]](#), where he died in 1821.

Training (2)

ORIGINAL TRAINING TEXT:

This text is about [[Battle of Waterloo|**Waterloo**]]. Allegedly, Napoleon tried to escape to North America, but the [[Royal Navy|**Royal Navy**]] was blockading French ports to forestall such a move. He finally surrendered to [[Captain (Royal Navy)|**Captain**]] [[Frederick Lewis Maitland (Royal Navy officer)|**Frederick Maitland**]] of [[Her Majesty's Ship|**HMS**]] "[[HMS Bellerophon (1786)|**Bellerophon**]]" on 15 July. There was a campaign against French fortresses that still held out; [[Longwy|**Longwy**]] capitulated on 13 September 1815, the last to do so. The [[Treaty of Paris (1815)|**Treaty of Paris**]] was signed on 20 November 1815. [[Louis XVIII of France|**Louis XVIII**]] was restored to the throne of France, and Napoleon was exiled to [[Saint Helena|**Saint Helena**]], where he died in 1821.

THE ANALYSIS OF THE TRAINING TEXT:

This text is about **Waterloo**. Allegedly, **Napoleon** tried to escape to **North America**, but the **Royal Navy** was blockading French ports to forestall such a move. He finally surrendered to **Captain Frederick Maitland** of **HMS 'Bellerophon'** on 15 July. There was a campaign against French fortresses that still held out. **Longwy** capitulated on 13 September 1815, the last to do so. The **Treaty of Paris** was signed on 20 November 1815. **Louis XVIII** was restored to the throne of **France**, and **Napoleon** was exiled to **Saint Helena**, where he died in 1821.



Treaty of Paris (and 7 other training examples)

Training (3)

Treaty of Paris

← Features values →

1. Treaty of Paris
2. Treaty of Paris (1763)
3. Treaty of Paris (1783)
4. Treaty of Paris (1814)
5. **Treaty of Paris (1815)**
6. Treaty of Paris (1898)
7. Treaty of Paris (1920)
8. Treaty of Paris (1951)
9. Treaty of Paris (1856)
10. Paris Peace Treaties, 1947
11. Treaty of Paris (1810)
12. Paris Peace Conference, 1919
13. Treaty of Peace with Italy, 1947
14. Paris Peace Accords
15. Treaty of Paris (1259)
16. Pacte de Famille
17. Treaty of Paris (1229)
18. Treaty of Paris (1303)
19. Treaty of Paris (1323)
20. Treaty of Paris (1355)
21. Treaty of Paris (1623)
22. Treaty of Paris (1657)
23. Treaty of Paris (1900)
24. Bonn–Paris conventions
25. Treaty of Paris (band)
26. Treaty of Paris (1796)
27. Peace of Paris (1783)
28. Treaty of Paris (1626)
29. Treaty of Paris (1857)
30. Treaty of Paris (1802)

$\phi_1(x, y)$	$\phi_2(x, y)$	$\phi_3(x, y)$	$\phi_4(x, y)$...	$\phi_{n-1}(x, y)$	$\phi_n(x, y)$
$\phi_1(x, \hat{y})$	$\phi_2(x, \hat{y})$	$\phi_3(x, \hat{y})$	$\phi_4(x, \hat{y})$...	$\phi_{n-1}(x, \hat{y})$	$\phi_n(x, \hat{y})$



LINEAR CLASSIFIER (LOGISTIC REGRESSION)

- Trained on the pairs (truth, other_candidate):

$$\frac{P(1|\hat{y})}{P(1|y)} > 1 \quad \sim \quad \frac{P(\hat{y}|x)}{P(y|x)} > 1 \quad \sim \quad \sum_{k=1}^n \beta_k (\phi_k(x, \hat{y}) - \phi_k(x, y))$$

- Current system trained on ~90 million data points

Which Knowledge Base?

Knowing More Is Not Always an Advantage

- Mapping between entity collections is not trivial

Almost 100,000 out of 820,000 Wikipedia 2008 pages in the knowledge base changed their title as of 2014

- More comprehensive may lead to NIL answers

e.g.: Birmingham

TAC 2011: EL_00256

NILxxxx ← Birmingham campaign

The army moved to Albany, Ga., in 1961. Some observers say Albany was a failure for Dr. King, but others say it played an important part in preparing the movement for Birmingham.

TAC 2011: EL_00258

Birmingham, Alabama → E0609361

A map of hate groups from the Southern Poverty Law Center in Birmingham, Ala., shows there are 33 active white supremacist groups that have formed in Pennsylvania.

Gold standard for both queries is E0609361: Birmingham, Alabama

Results

NIST/LDC Evaluations

Accuracy:	MSR logistic regression	best result in the TAC Evaluation
2013 system		
TAC 2011 test set	89.3 %	86.8% (MSR)
TAC 2012 test set	80.4 %	76.2% (MSR)
TAC 2013 test set	83.2 %	83.2% (MSR)
2014 system		
TAC 2014 train set	87.2 %	
TAC 2014 test set	86.8 %	86.8% (MSR)

Google-Microsoft-Yahoo Challenge

	Precision	Recall	F-measure
ERD 2014 Train set	83.7%	72.6%	0.778
ERD 2014 Test set	83.3%	69.9%	0.760

Entity Extraction and Linking Applications

The image is a collage of screenshots from Microsoft Bing and MSN Now, illustrating entity extraction and linking applications. The top left shows the Bing homepage with a search bar. The middle left shows the MSN Now 'WHAT'S TRENDING' page, which includes a list of 'biggest movers' (now) such as John Travolta, Gavin Smith, Brooke Shields, Melissa Gilbert, Hillary Clinton, and Beastie Boys. The bottom left shows a 'MEET THE PRESS' interview with Rick Santorum, with a sidebar listing entities like 'religion', 'higher education', 'economy', 'Michigan', and 'Rick Santorum'. The right side shows a Bing Social page for Peyton Manning, displaying recent public updates and social media links.

Entity Extraction and Linking Applications

MSN NOW WHAT'S TRENDING
Wednesday, May 9, 11:10 am
Analyzing 144,186,249 Bing searches and 12,124,446 social updates.

biggest movers (now)

- john travolta
- gavin smith
- brooke shields
- melissa gilbert
- hillary clinton
- beastie boys
- christina aguilera

MEET THE PRESS
Feb. 26
Presidential candidate Rick Santorum discusses the crisis in Afghanistan, the upcoming Michigan primaries and the role more
Find term in this clip:
religion
higher education
economy
Michigan
Rick Santorum

PEYTON MANNING
See all social results for this topic
RECENT PUBLIC UPDATES
CloudFantasy: Peyton Manning will be officially cut by the Indianapolis Colts tomorrow. Now everyone is talking about where ...
WTH-TV: Peyton Manning may leave Colts | WTH-TV.com
2DaySports.com: Report: Peyton Manning is Leaving the Colts (read)
Yahoo Search: Say it ain't so! Peyton Manning won't be playing for the Colts this season. Where do you think he'll end up?
ESPN 1420: It's official, the Colts will release Peyton Manning.
wrapupp.com: Colts to Cut Peyton Manning | wrapupp
100.9 The Eagle: Peyton Manning, Colts To Split? Star QB Reportedly Done in Indianapolis
Fanatic Sports and Cards: Colts to release Peyton Manning | Fanatic Sports and Cards

Microsoft

Applications: Video Annotation

Video

Video

Latest videos

Full broadcasts

Most viewed

Topics

Headliners

Roundtables

Web only: Take Two

MTP Minute

More David Gregory

MTP History

More NBC shows

Nightly News

Morning Joe

Latest videos

Axelrod: 'We want to keep this economy growing'

Will goals in Afghanistan be met?

now Debating control and blame on Gulf oil spill

next Takeaways from Tuesday

MEET THE PRESS

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

3 instances

advertisement

BOEING

Applications: Trending Topics

The screenshot shows the Bing Social interface. At the top, there's a navigation bar with links to Web, Images, Videos, Shopping, News, Maps, More, MSN, and Hotmail. The Bing logo is on the left, and a search bar is in the center. Below the search bar, the word "Social" is displayed. On the left side, there's a "Hottest Social Topics" list with items like Peyton Manning, Chris Mortensen, Selena Gomez, Acrylic Paint, Gregg Wylde, Sales, Super Tuesday, Beast Boy, Miley Cyrus, and Stainless Steel. Each item has a right-pointing arrow. Below this list are filters for "SOURCE" (All, Twitter, Facebook) and "TIMEFRAME" (Past Hour, Past Day, Past Week). The main content area is titled "PEYTON MANNING" and includes a link to "See all social results for this topic". Below this is a section for "RECENT PUBLIC UPDATES" with several news snippets, each accompanied by a small logo and a brief description of the update, including links to the source and the time it was posted. A button on the right says "See recent tweets on Bing Twitter Maps". At the bottom, there's a section for "WITHIN THIS TOPIC" with keywords like Indianapolis Colts, ESPN, Chris Mortensen, Andrew Luck, and ESPNNews. A link at the very bottom says "See more social results for Peyton Manning".

File Edit View Favorites Tools Help

Web Images Videos Shopping News Maps More | MSN Hotmail

Sign in Rewards Redmond, Wa

bing MS Beta

Social

Hottest Social Topics

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

PEYTON MANNING

See all social results for this topic

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...> - Facebook - 34 minutes ago

Yahoo Search: Say it ain't so! Peyton Manning won't be playing for the Colts this season. Where do you think he'll end up? <http://search.yahoo.com/sea...> - Facebook - 53 minutes ago

ESPN 1420: It's official, the Colts will release Peyton Manning. <http://espn1420.com/pey...> - Facebook - 49 minutes ago

wrapupp.com: Colts to Cut Peyton Manning | wrapupp <http://www.wrapupp.com/nfl...> - Facebook - 51 minutes ago

100.9 The Eagle: Peyton Manning, Colts To Split? Star QB Reportedly Done In Indianapolis <http://www.huffingtonpost.com/201...> - Facebook - 50 minutes ago

Fanatic Sports and Cards: Colts to release Peyton Manning | Fanatic Sports and Cards <http://www.fanaticsportsandcards.com/col...> - Facebook - 11 minutes ago

WITHIN THIS TOPIC

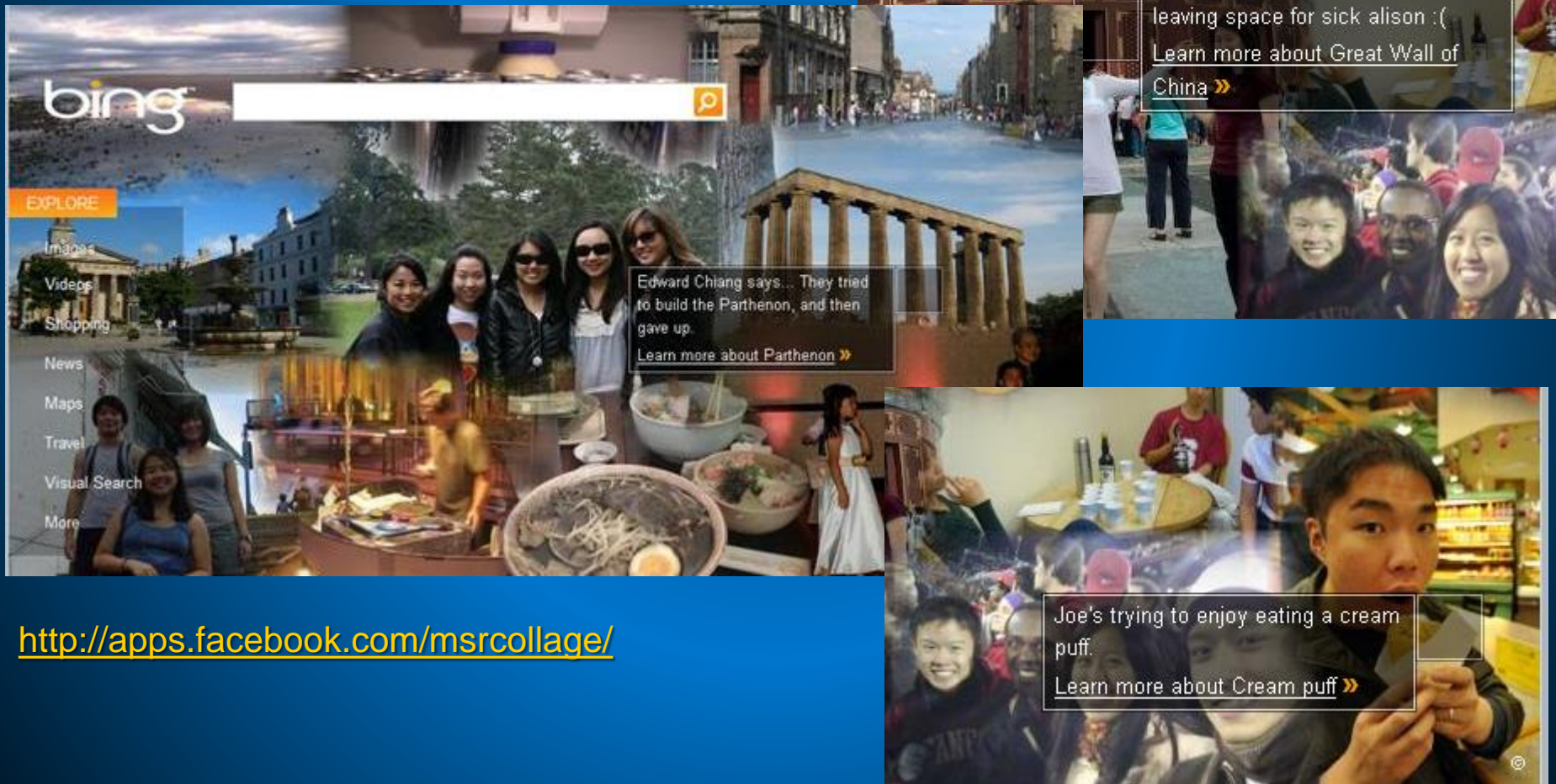
Keywords Indianapolis Colts, ESPN, Chris Mortensen, Andrew Luck, ESPNNews

See more social results for Peyton Manning

See recent tweets on Bing Twitter Maps

Applications: Facebook Collage

Emre Kiciman, Chun-Kai Wang



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

Applications: Search Ads Matching

- 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:

replace entities in a query with related entities

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

“weight loss acai berry best energy weight loss herbal pill”

- acai_berry best energy extract herbal_pill weight_loss
- acai_berry best energy extreme herbal_pill pure weight_loss
- acai_berry best energy extreme herbal_pill review weight_loss
- acai_berry best energy extreme herbal_pill weight_loss
- acai_berry best energy flush herbal_pill weight_loss
- acai_berry best energy health herbal_pill juice mn weight_loss
- acai_berry best energy herbal_pill information weight_loss
- acai_berry best energy herbal_pill ingredient juice weight_loss
- acai_berry best energy herbal_pill juice weight_loss
- acai_berry best energy herbal_pill oprah product weight_loss
- acai_berry best energy herbal_pill oprah weight_loss
- acai_berry best energy herbal_pill oprah_winfrey weight_loss
- acai_berry best energy herbal_pill organic weight_loss
- acai_berry best energy herbal_pill plant weight_loss
- acai_berry best energy herbal_pill product review weight_loss
- acai_berry best energy herbal_pill product weight_loss
- acai_berry best energy herbal_pill pure weight_loss
- acai_berry best energy herbal_pill research weight_loss
- acai_berry best energy herbal_pill review weight_loss
- acai_berry best energy herbal_pill sell store weight_loss
- acai_berry best energy herbal_pill side_effects weight_loss
- acai_berry best energy herbal_pill store weight_loss
- acai_berry best energy herbal_pill supplement weight_loss
- acai_berry best energy herbal_pill tablet weight_loss
- acai_berry best energy herbal_pill tea weight_loss
- acai_berry best energy herbal_pill testimonial weight_loss
- acai_berry best energy herbal_pill truth weight_loss
- acai_berry best energy herbal_pill weight_loss
- acai_berry best energy herbal_pill weight_loss wholesale
- acai berry extract best energy weight loss herbal pill 4
- pure acai berry extreme best energy weight loss herbal pill 5
- acai berry extreme reviews best energy weight loss herbal pill 3
- extreme acai berry best energy weight loss herbal pill 8
- acai berry flush best energy weight loss herbal pill 3
- acai berry health juice mn best energy weight loss herbal pill 7
- acai berry information best energy weight loss herbal pill 4
- acai berry juice ingredients best energy weight loss herbal pill 3
- acai berry juice used best energy weight loss herbal pill 4
- acai berry juice best energy weight loss herbal pill 4
- acai berry products oprah best energy weight loss herbal pill 5
- oprah acai berry best energy weight loss herbal pill 6
- oprah winfrey acai berry best energy weight loss herbal pill 7
- organic acai berry best energy weight loss herbal pill 4
- acai berry plants best energy weight loss herbal pill 11
- acai berry plant best energy weight loss herbal pill 3
- acai berry product review best energy weight loss herbal pill 4
- acai berry products best energy weight loss herbal pill 9
- pure acai berry best energy weight loss herbal pill 6
- acai berry research best energy weight loss herbal pill 7
- acai berry reviews best energy weight loss herbal pill 6
- acai berry pills reviews best energy weight loss herbal pill 9
- stores that sell acai berry best energy weight loss herbal pill 4
- side effects acai berry best energy weight loss herbal pill 4
- acai berry stores best energy weight loss herbal pill 4
- acai berry supplement best energy weight loss herbal pill 5
- acai berry supplements best energy weight loss herbal pill 4
- acai berry tablets best energy weight loss herbal pill 4
- acai berry tea best energy weight loss herbal pill 4
- acai berry testimonials best energy weight loss herbal pill 4
- acai berry truth best energy weight loss herbal pill 8
- acai berry best energy weight loss herbal pill 9
- acai berry it best energy weight loss herbal pill 5
- weight loss acai berry best energy weight loss herbal pill 10
- acai berry weight loss best energy weight loss herbal pill 4
- acai berry lose weight best energy weight loss herbal pill 10
- wholesale acai berry best energy weight loss herbal pill 6

Applications: Query Suggestion

e.g.

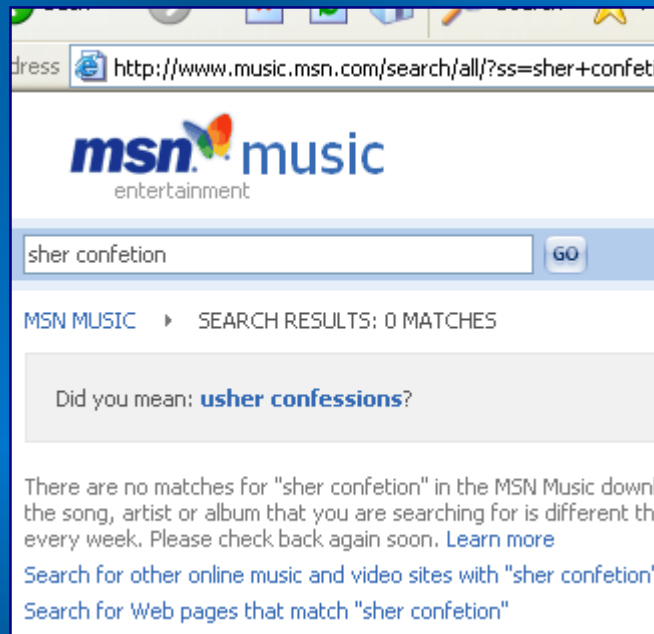
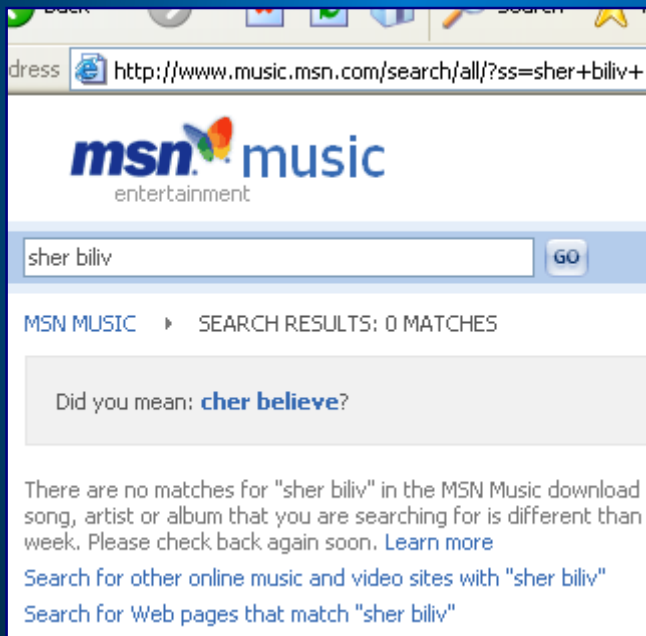
john travolta flics → john travolta movies
john travolta filmography
robert de niro flics
robert de niro movies
~~john wayne flics~~

bank of america credit card →
bank of america visa
wells fargo credit card
~~bank of brazil credit card~~
~~bank of america credit valentine~~

Applications: Spelling Correction

selin dion color of love → celine dion colour of my love

gensis selin → genesis selling england by the pound



sher biliv



cher believe

sher cofetion



usher confessions

Applications: Web Factoid Question Answering

bing 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](#)

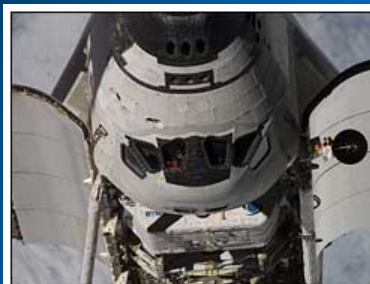
[discovery.yukozimo.com/who-discovered-america-before-columbus](#) ▼

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

Applications: Context-aware Search



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

QUOTE

"This is a major milestone for the program."

Lead flight director John Curry

WHAT DO YOU THINK?

Go To Comments >

folded map after it's been spread out on a car's dashboard. NASA managers are wary about putting too much pressure on the array while trying to get it back in its box.

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

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

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

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

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

(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 folded map after it's been spread out on a car's dashboard. NASA managers are wary about putting too much pressure on the array while trying to get it back in its box.

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

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

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

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

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

Live Search

heat shield

Web Images News Maps More Products Beta

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

Heat Shield - www.target.com SPONSORED SITES
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](#)

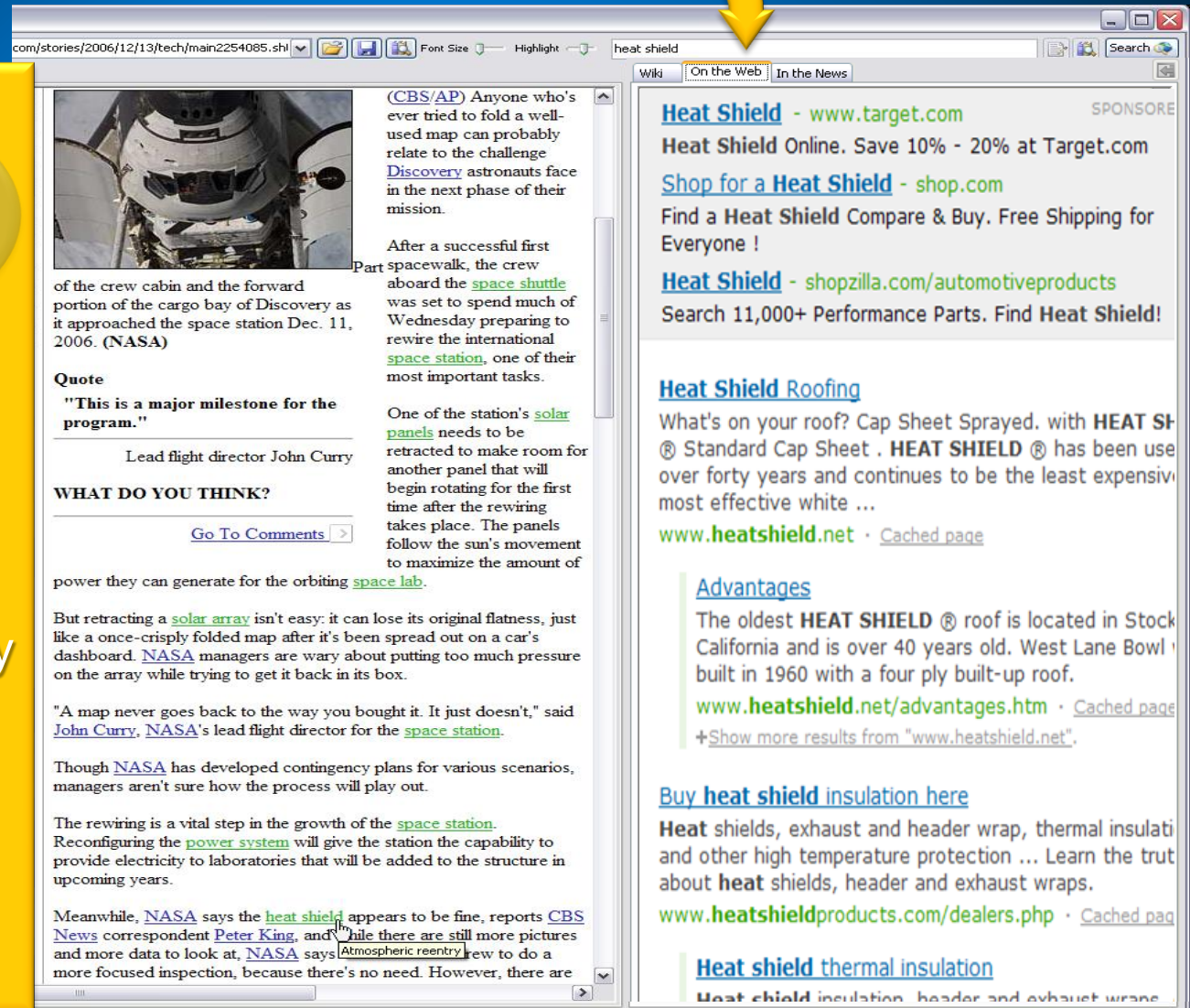
Advantages
The oldest **HEAT SHIELD**® roof is located in Stockton, California and is over 40 years old. West Lane Bowl was built in 1960 with a four ply built-up roof.
www.heatshield.net/advantages.htm • [Cached page](#)
+Show more results from "www.heatshield.net".

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.heatshieldproducts.com/dealers.php • [Cached page](#)

Heat shield thermal insulation
Heat shield insulation, header and exhaust wraps, other

Communicating Concepts as Context

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



The screenshot shows a web browser window with the address bar displaying "com/stories/2006/12/13/tech/main2254085.shl". The main content area features a photograph of the space shuttle Discovery's cargo bay, with the text: "Part of the crew cabin and the forward portion of the cargo bay of Discovery as it approached the space station Dec. 11, 2006. (NASA)". Below the photo is a quote from Lead flight director John Curry: "This is a major milestone for the program." A section titled "WHAT DO YOU THINK?" includes a "Go To Comments" link. The article continues with text about the shuttle's solar panels and the power system. A sidebar on the right contains several links: "Heat Shield - www.target.com", "Heat Shield Online. Save 10% - 20% at Target.com", "Shop for a Heat Shield - shop.com", "Heat Shield - shopzilla.com/automotiveproducts", "Heat Shield Roofing", "Advantages", "Buy heat shield insulation here", and "Heat shield thermal insulation".

com/stories/2006/12/13/tech/main2254085.shl

Font Size Highlight heat shield

Wiki On the Web In the News

[Heat Shield](#) - [www.target.com](#) SPONSORE

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 over forty years and continues to be the least expensive 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 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.heatshieldproducts.com/dealers.php](#) · [Cached page](#)

[Heat shield thermal insulation](#)

Heat shield insulation, header and exhaust wraps

How Important Is the Browsing Context?

Case study: news browsing

Toolbar



URL of news site
browsed page



Query submitted
to Search Engine
(Bing, Google, Yahoo)

FOXNEWS.COM HOME > WORLD > NATIONAL

Thousands Protest Immigration Policies in New York

Saturday, April 01, 2006

Associated Press

E-Mail

Share:

NEW YORK — Thousands of immigrant rights supporters formed a line stretching more than a mile long Saturday as they marched across the Brooklyn Bridge, waving flags from more than a dozen countries in a demonstration for immigration reform in Congress.

Heralded by a parade of floats, whistles and drums, a group of mostly Latin American immigrants gathered in downtown Brooklyn and trudged a path laden with symbols of the city's immigrant strength on their way to a plaza in lower Manhattan.

The marchers must have been tired, as they crossed a bridge designed by a German, and finished in a square at the edge of Chinatown in an area that once held the Irish slums depicted in the 1927 film "Gangs of New York."



53.9%

relevant

33.0%

irrelevant

5.8%

cannot say

7.3%

navigational for news



$P = 96\%$









$R = 90\%$

Applications: Topic Pages

TOPIC PAGES:

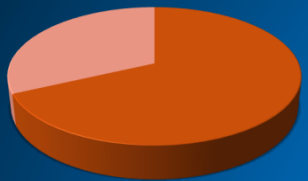
William Shatner

- William Alan Shatner (**born** on March 22, 1931) is a Canadian **actor** who gained fame for playing **Captain James Tiberius Kirk**, captain of the starship USS Enterprise in the television show **Star Trek** from 1966 to 1969 and in seven of the subsequent **movies**. 
- "Shatner is the epitome of the post-ironic, 21st-century **television** and pop-culture hero," says Robert Thompson, a Syracuse University **television** and pop-culture historian. 

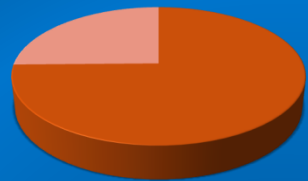
www.answers.com/topic/william-shatner
www.imdb.com/name/nm0000638/bio
www.rt.com/smart/William_Shatner.html
www.buddytv.com/star-trek.aspx
- Decades after his much-maligned album, "**The Transformed Man**," boldly took "**Lucy in the Sky With Diamonds**" into strange new worlds, Shatner performed **songs** from Has Been before 4,000 cheering fans. 
- Actor William Shatner has sold his **kidney stone** for \$25,000, with the money going to a housing **charity**, it was announced Tuesday. 
- "**Up Till Now**" by William Shatner allegedly "is riddled with discrepancies about the fateful night of august 9, 1999" when Shatner found Nerine in their pool at their studio city **home**. 
- At age 73, Shatner reinvented himself yet again with a recurring **role** as a nutty attorney on the last season of The Practice , which snagged him his first **Emmy** in 2004 and another in 2005 for playing the same part on the spin-off **Boston Legal**. 
- Shatner's **Star Trek** sidekick Leonard Nimoy will be the only original cast member on board for the prequel – he'll reprise his role as Mr. Spock in portions of the **film**. 
- In 1999, Shatner suffered public personal tragedy when his third **wife**, Nerine , accidentally **drowned** in their swimming pool. 
- In the latest **Priceline ads**, Shatner bursts forth as the **Priceline** Negotiator, a mashup of James Bond and Ron Popeil who will do anything to help people broker better deals. 
- William Shatner has signed up to host a celebrity-**interview** show on the Biography Channel titled "Shatner's Raw Nerve" which will premiere sometime next year, reports Daily Variety. 



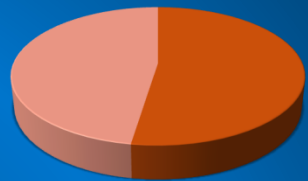
Aspect Models Derived from Query Logs



Topic : Michael Phelps
Aspects : sport, girlfriend, ...

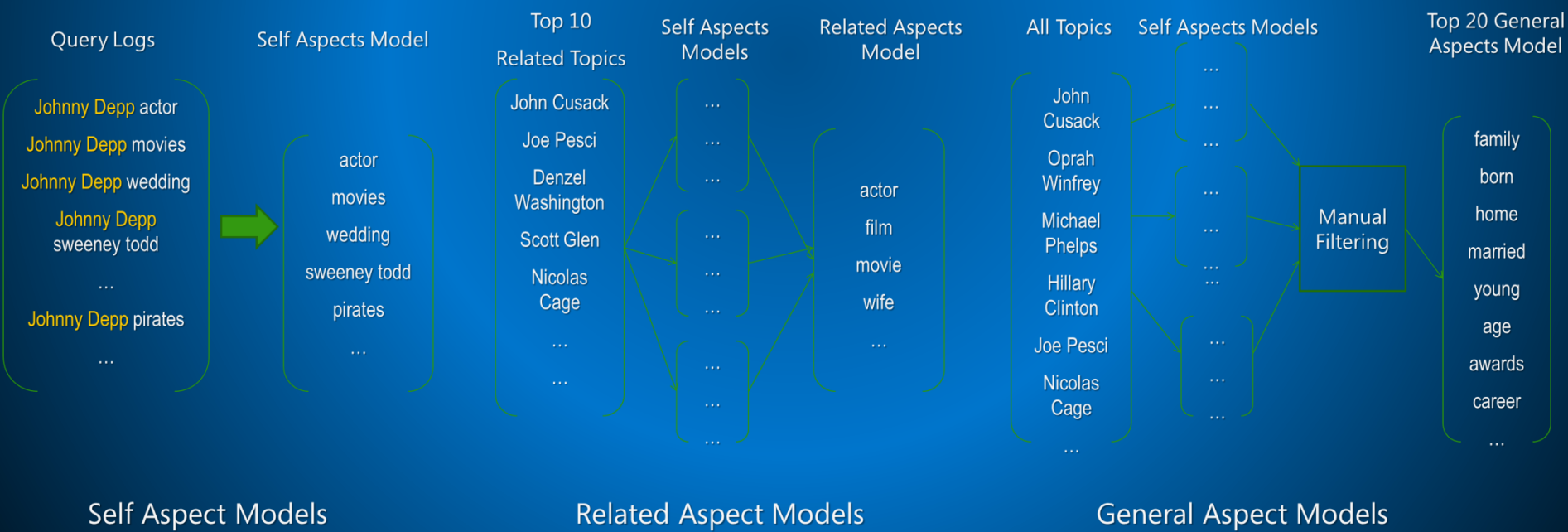


Topic : George Clooney
Aspects : movies, pet, pictures, ...



Topic : Oprah Winfrey
Aspects : show, tv, website, ...

Distribution of **topic only queries** and **topic + aspect queries** in a query log sample.



Sentence Selection by Exploiting Web Redundancy

Sentence Selection Goals

Grammatical	Classify sentences using lexical and language modeling features
Relevant	Aspect specific context vectors.
Diverse	Explicitly maintain covered aspects. Iteratively select sentences that cover new aspects.

What is a *typical* sentence for a given aspect ?

1. After **The Pianist** **Adrien Brody** has appeared in four very different movies.
2. **Adrien Brody** received widespread recognition when he was cast as the lead in Roman Polanski's **The Pianist** (2002).
3. **Adrien Brody** is a New York actor who is known to international audiences as the star of Roman Polanski's 2002 film, **The Pianist**.

Rachel Weisz (The Constant Gardener), **Adrien Brody** (**The Pianist**), Mark Ruffalo (Zodiac), and Rinko Kikuchi (Babel) star in THE BROTHERS BLOOM [...]

Sentence Selection Methods

A denotes aspect vector, S denotes the candidate sentence vector, and A_s denote the aspects in S .

Full Aspect (Full-context): $f(s) = \text{cosine}(A, S)$

Novelty (NS-Full Context):

until $|I| < n$ *do*

$$(1) n(s) = \lambda * f(s) + (1 - \lambda) * \min_{s_i \in I} (1 - \text{cosine}(s, s_i))$$

$$(2) I = I \cup \underset{s \in S}{\operatorname{argmax}} n(s)$$

Typical:

for each $a \in A$ *do*:

$$(1) V_a = \sum_{s \in T_a} s$$

$$(2) C_a = \lambda * A + (1 - \lambda) * V_a$$

$$(3) \forall s \in S, t(s) = \text{cosine}(s, C_a)$$

$$(4) I = I \cup \underset{s \in S}{\operatorname{argmax}} t(s)$$

Diversity: *until* $|I| < n$ *do*

$$(1) I = I \cup \underset{s \in S}{\operatorname{argmax}} f(s)$$

$$(2) A = A - \delta * A_s$$

Diversity + Typical (DS-Typical):

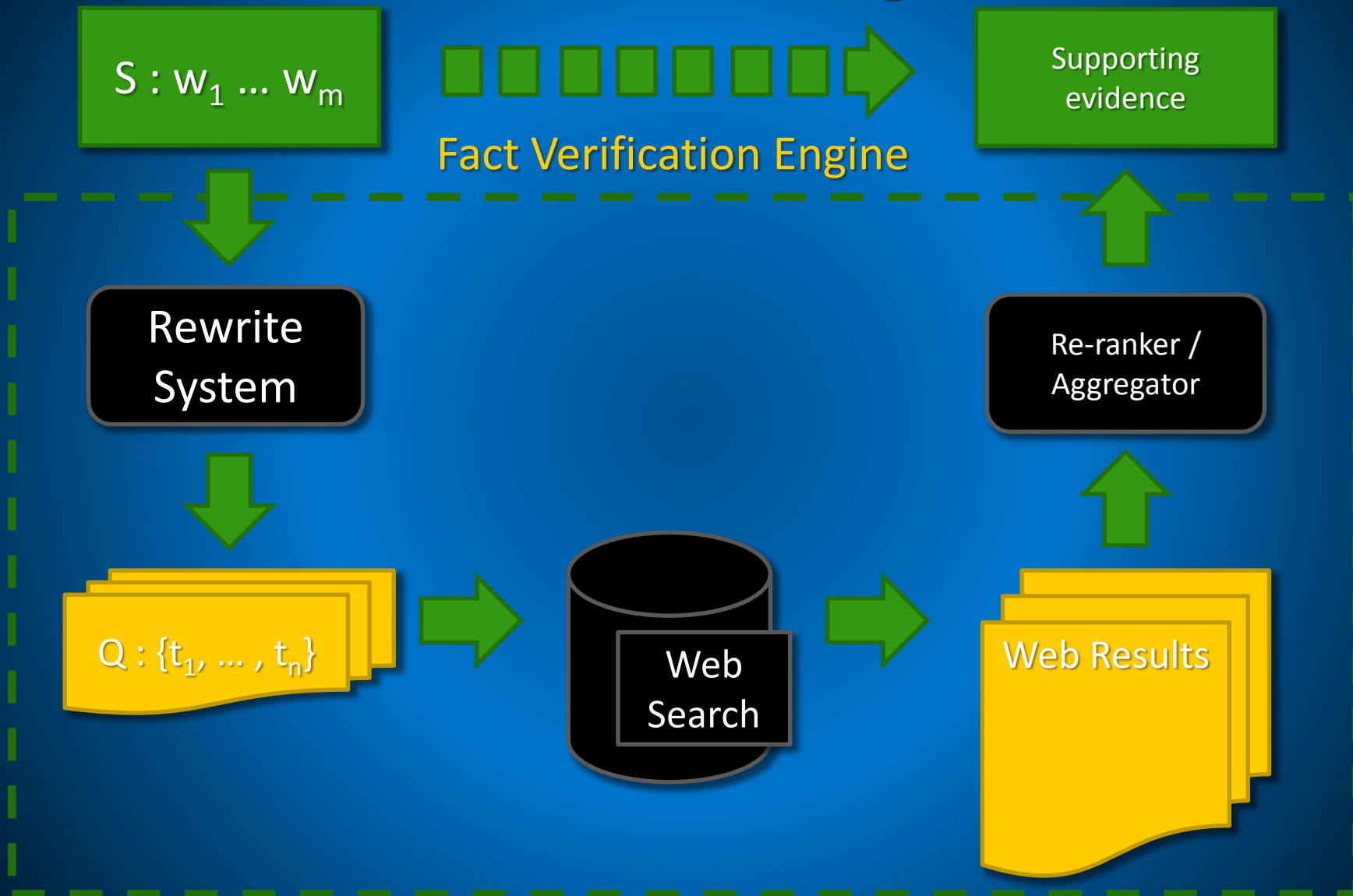
Apply **Diversity** to **Typical** ranking.

Results

Bette Midler, Harvey Keitel, Mario Cuomo, Newt Gingrich, Billy Bragg, Holly Hunter, Mario Lemieux, Reese Witherspoon, Bob Brady, Joe Theismann, Marion Jones, Roberto Benigni, Carmen Electra, Julie Walters, Matthew Santos, Saxby Chambliss, Elton Brand, Lindsey Graham, Monica Lewinsky, Sean Young

Method	Precision	Grammar	Non-Redund.	Novelty	Recall
Web Search	0.37 (0.12)	0.61 (0.22)	0.57 (0.23)	0.07 (0.09)	0.30 (0.16)
LexRank	0.22 (0.11)	0.61 (0.18)	0.71 (0.26)	0.19 (0.20)	0.34 (0.21)
DS-Typical	0.50 (0.17)	0.83 (0.13)	0.93 (0.15)	0.29 (0.17)	0.53 (0.19)

Applications: Fact Checking



Learning From Wikipedia Citations

- **Example paragraph from Wikipedia:**

In late 1988, Obama entered Harvard Law School. He was selected as an editor of the Harvard Law Review at the end of his first year,^[34] and president of the journal in his second year.^{[30][35]} During his summers, he returned to Chicago, where he worked as an associate at the law firms of Sidley Austin in 1989 and Hopkins & Sutter in 1990.^[36] After graduating with a J.D. magna cum laude from Harvard^[37] in 1991, he returned to Chicago.^[34] Obama's election as the first black president of the Harvard Law Review gained national media attention^{[30][35]} and led to a publishing contract and advance for a book about race relations,^[38] which evolved into a personal memoir. The manuscript was published in mid-1995 as *Dreams from My Father*^[38]

Wikipedia Citation Dataset

data points
600,000

<subject> Martin Wendt </subject>

<statement> As of 2009, his total live tournament winnings exceed \$900,000. </statement>

<ref> <http://pokerdb.thehendonsmob.com/player.php?n=29302> </ref>

<subject> Gamma-Aminobutyric acid </subject>

<statement> In general, GABA does not cross the blood-brain barrier, although certain areas of the brain which have no effective blood brain barrier, such as the periventricular nucleus, can be reached by drugs such as systematically injected GABA. </statement>

<ref> <http://physrev.physiology.org/content/79/2/511.full> </ref>

<subject> Jason Beghe </subject>

<statement> Kennedy and Beghe often hung out together outside the Metropolitan Museum of Art and in Central Park, and were monitored by Kennedy's Secret Service detail. </statement>

<ref> <http://www.cnn.com/US/9907/22/1fk.growing.up.in.nyc/> </ref>

<subject> Los Angeles </subject>

<statement> Los Angeles enjoys a Mediterranean climate, with an average of 35 days with measurable precipitation annually. </statement>

<ref> <http://www.weatherbase.com/weather/weather.php3?...> </ref>

<subject> Murder </subject>

<statement> More than 500,000 people have been killed by firearms in Brazil between 1979 and 2003. </statement>

<ref> <http://news.bbc.co.uk/2/hi/americas/4628813.stm> </ref>

<subject> Fox hunting </subject>

<statement> An amendment which allowed licensed hunting under stricter conditions, advocated by the then Prime Minister Tony Blair and some members of the government's independent inquiry on fox hunting (including its chairman Lord Burns), was voted down. </statement>

</statement>

<ref> <http://www.guardian.co.uk/uk/2000/jan/24/hunting.rura...> </ref>

<subject> Armenian Genocide </subject>

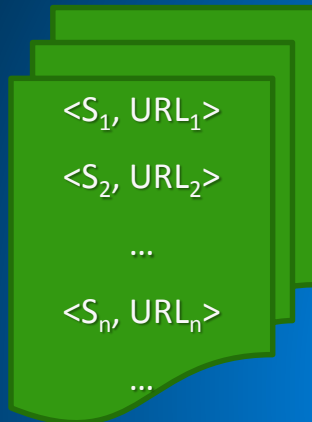
<statement> Despite his previous public recognition and support of Genocide bills, as well as the election campaign promises to formally recognize the Armenian Genocide, the U.S. President, Barack Obama, although repeating that his views on the issue have not changed, has thus far abstained from using the term 'genocide'. </statement>

<ref> <http://www.politifact.com/truth-o-meter/promises/obameter/...> </ref>

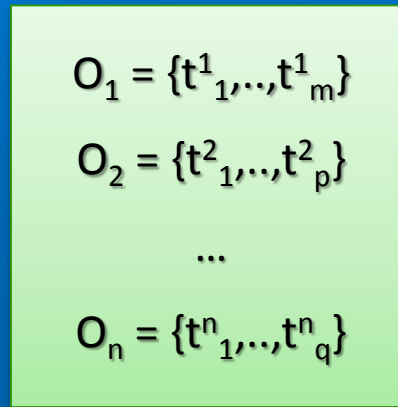
Rewriting System - Framework

Assume the documents cited in Wikipedia are the most relevant on the Web

Wikipedia Training



Overlapping Terms



$Search(O_i) = URL_i ?$

Features : {POS, sense #, tf*idf, wikiFreq, keyphraseness, offset, multi-word}

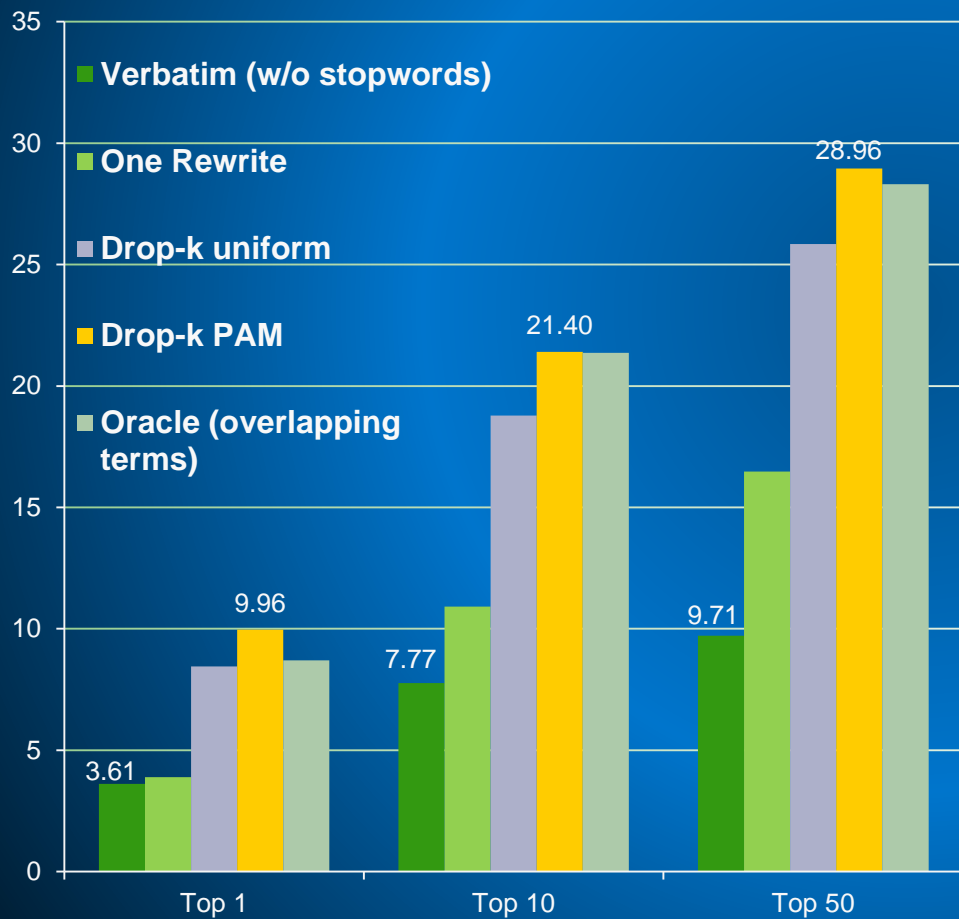
S: About 31% of the population is made up of [foreign nationals] living in [Saudi Arabia]

O:



Results

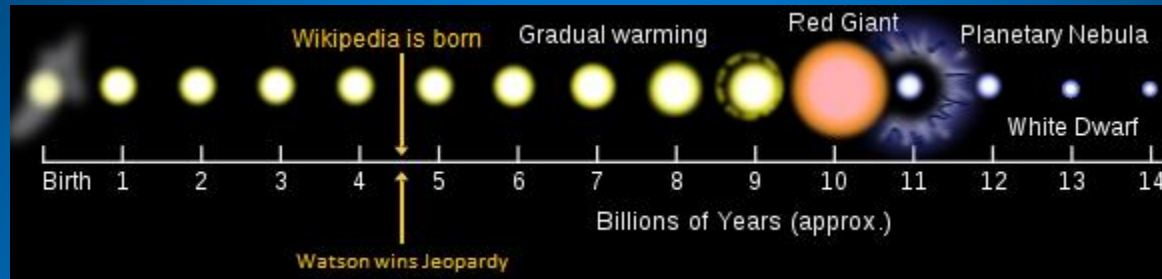
Percentage of Wikipedia citations
retrieved in top n results



Amazon Turk Experiments
(scale 0 – 2)

Reference Type	Quality	Ease of Effort	Trust-worthiness
Wikipedia	1.37	1.26	1.48
One Rewrite	1.15	1.18	1.36
Drop-k PAM	1.15	1.24	1.45

Where We Go From Here



- Web-scale knowledge collections
- Web-scale annotation and fact extraction
- Instant access to knowledge in any environment
- Context-awareness
- Personalized views of the knowledge
- Fact engines