

Leibniz
Universität
Hannover

Temporal Information Retrieval

Avishek Anand
anand@kbs.uni-hannover.de

Web science @ L3S

- Computer Science and interdisciplinary research on all aspects of the Web



Real-time data processing
for finance predictions



Cross-media analysis
and interpretation

- Internet: Communication and Networks
- Information: Accessing information and knowledge on and through the Web
- Community: Supporting communities and groups on the Web, for research, education, production and entertainment
- Society: Requirements (technological, social, legal) for the Web



Retrieval, Exploration and Analytics for Web Archives (ERC Advanced Grant)



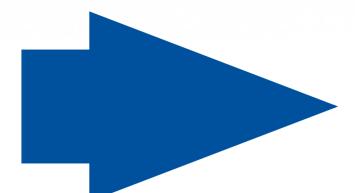
ForgetIT

ForgetIT: Concise
Preservation via
Managed Forgetting



CUBRIK: Searching by
computers and humans

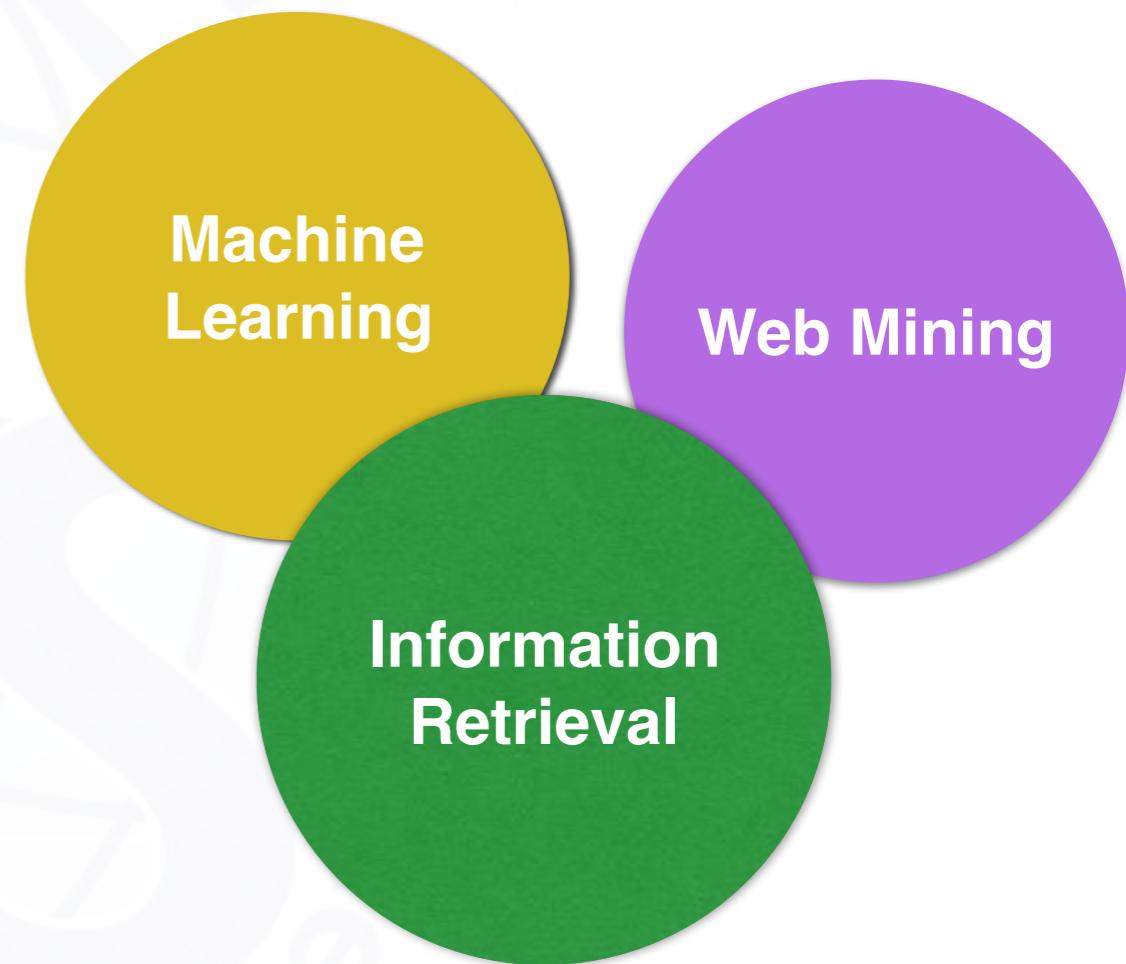
- Selected projects



My Research - Long Data

Temporal (time-varying) aspects of data

- Temporal Retrieval Models
- Temporal Indexing Methods
- Mining Temporal Collections
- Interpretability of Search

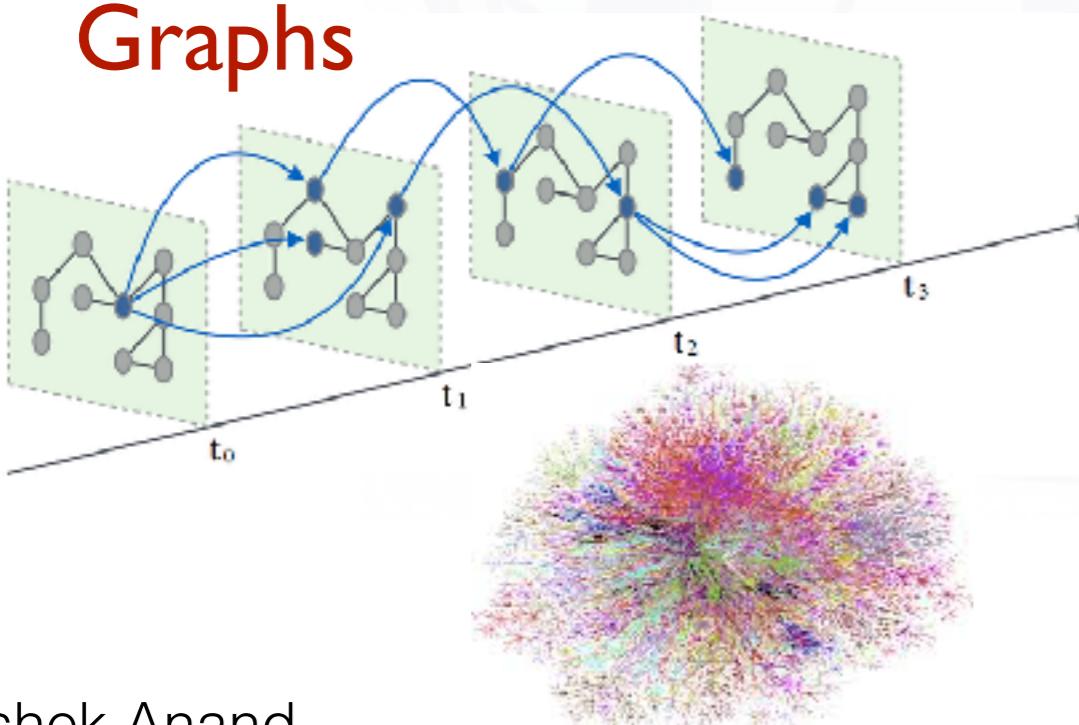


What kind of Data ?

Text Data

1. [Version 2](#) | [PloS One. 2009 Aug 21. \[published: Sep 14\] PMID:20001025.](#)
Adjuvant is necessary for a robust immune response to a single dose of H5N1 pandemic flu vaccine in mice.
1. [Version 2](#) | [PloS One. 2009 Aug 21. \[published: Sep 14\] PMID:20001025.](#)
Committee PB, Donnelly JJ, J...
Novartis Vaccines and Diagnostics, Cambridge, MA, USA; NOVARTIS VACCINES & Diagnostics; Novartis Vaccines; University of Siena; Novartis and Division of Immunobiologics, National Institute for Medical Research, The Ridgeway, London NW7 1AA.
Pandemic H5N1 MF59 adjuvant increases response to H5N1 influenza vaccine antigens are currently being manufactured. The MF59 adjuvant has an established safety profile in humans and a proven ability to increase responses to some influenza vaccines in humans. To inform initial decisions on the use of these vaccine components to protect human populations, we have immunized mice with MF59-adjuvanted or non-adjuvanted pandemic influenza vaccines. Immunizing unprimed mice with a single dose of MF59-adjuvanted vaccine elicits functional antibody titers equivalent to those associated with protection of humans from seasonal influenza. Without adjuvant, two doses are required for a robust antibody response. Unadjuvanted vaccines with 0.5 and 1 microgram of antigen elicit equivalent titers. These data support including MF59 in pandemic flu vaccines to rapidly protect young adults and children, who may have little or no previous exposure to influenza infection or immunization.
PMID: 20029611 [PubMed - as supplied by publisher]

Graphs



Knowledge Units



Angela Merkel

Chancellor of Germany

<http://time.com/time-person-of-the-year-2015-angela-merkel-choice/>

[angela-merkель](#)

Angela Dorothea Merkel is a German statewoman and former research scientist. Merkel has been the Chancellor of Germany since 2005, and the leader of the Christian Democratic Union since 2000. [Wikipedia](#)

Born: July 17, 1954 (age 62), Hamburg

Height: 1.65 m

Party: Christian Democratic Union of Germany

Spouse: Joachim Sauer (m. 1986); Ulrich Merkl (m. 1977–1982)

Education: German Academy of Sciences at Berlin (1986); Leipzig University (1973–1978)

Profiles



Quote

The willingness to learn new skills is very high.

Whoever decides to dedicate their life to politics knows that earning money isn't their priority.

The euro is our common fate, and Europe is our common future.



'The negro Pymbus like the Hyrcanian Khyad' –
It is not so, it begins with Pymbus:
The negro Pymbus, he whose aborigines,
Black as his purpose, add the night resemble
Whom me may couch'd in the ominous horse,
Hath now his dread am' black complexion smear'd
With heraldry more dismal; head to foot
Now is he trimly gaine, comely tickld
With blee of fathers, mothers, daughters, sons,
Behold am' impasted with the purifying streets,
That lend a tyramous and damned light
To their lord's murder, masted in wrath and ire,
And thus o'er-sized with coagulate gore
With eyes like carbuncles, the hellish Pymbus
Old grandam' from make.'
So proceed you.

Where is Temporal Data ?

News Collections



event repositories, > 20 years,
200k articles a day

Encyclopaedic Collections



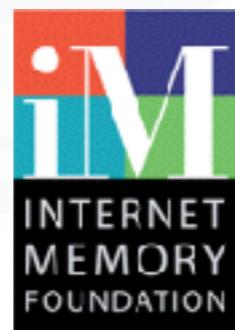
Rudolph the Red-Nosed Reindeer: Difference between revisions
From Wikipedia, the free encyclopedia

Revisions as of 09:58, 12 December 2015 (edit | history)
2003:12:14 00:01 (EST) (edit | history) (diff)
← Previous edit

Revision as of 10:13, 26 December 2015 (edit | history)
2015:12:27 12:46 (EST) (edit)
→ Next edit

Older	Newer
[Rudolph the Red-Nosed Reindeer] → [Rudolph the Red-Nosed Reindeer: The Movie]	[Rudolph the Red-Nosed Reindeer] → [Rudolph the Red-Nosed Reindeer: The Movie]
[Rudolph the Red-Nosed Reindeer] → [Rudolph the Red-Nosed Reindeer]	[Rudolph the Red-Nosed Reindeer] → [Rudolph the Red-Nosed Reindeer]

Web Archives



> 15 PB, 150 Bi Captures

Web Archives as a Dataset

Bauhaus-Universität Weimar



Aktuelles

Studium & Lehre

Wissenschaft & Kunst

Index & Suche

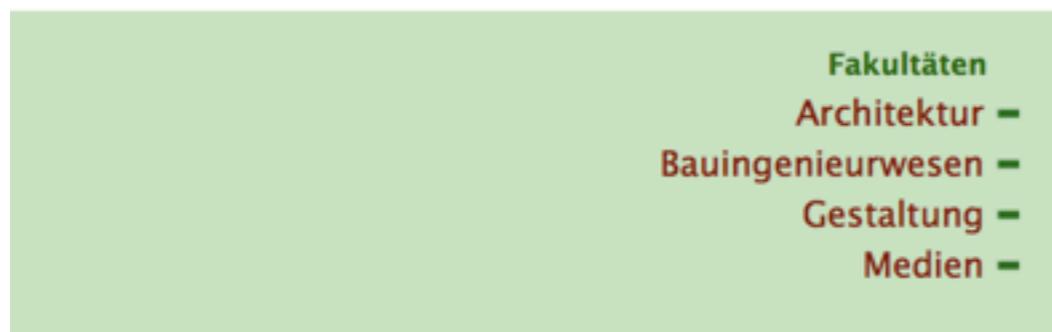
Struktur

Aktuelles	studium & Lehre	Wissenschaft & Kunst	Index & Suche	Struktur
news Aktuelle Informationen und Pinnwände	academics Über Weimar als Studienort, die Studiengänge und -angebote	science & arts Selbstdarstellung von Lehrstühlen, Instituten, Projekten und Personen	index & search Schnellübersicht, Suche nach Personen, e-mail und Telefonnummern	structure Telefon- und Personalverzeichnis der Einrichtungen und Mitarbeiter

[© Webwagen AG](#)



Web Archives as a Dataset



Pressemitteilung vom 29.04.2008

⇒ [bau.werke](#)

Die wissenschaftlich-künstlerische Vortragsreihe an der Bauhaus-Universität Weimar startet mit „Planstadt Weltstadt Monument“.

Pressemitteilung vom 29.04.2008

⇒ [Mediengestaltung studieren in Shanghai und Weimar](#)

Bewerbungsschluss für das Masterprogramm „Integrated International Media Art and Design Studies“ am 15. Mai 2008



Web Archives as a Dataset

The screenshot shows the homepage of the Bauhaus-Universität Weimar. At the top, there is a red header bar with the university's name. Below it is a navigation menu with several tabs: UNIVERSITÄT, ARCHITEKTUR, BAUINGENIEURWESEN, GESTALTUNG, and MEDIEN. Under the UNIVERSITÄT tab, there are links for Aktuell, Profil, Struktur, Studium, International, Forschung und Kunst, Partner und Alumni, and Service. The main content area features a large photograph of a modern, multi-story building with glass roofs and a crowd of people in front. Below the photo, there is a section titled "Die Bauhaus-Universität Weimar" with a brief description of the university's profile and study programs. To the right, there is a green box with the text "GEBEN SIE UNS IHR FEEDBACK ZUR NEUEN WEBSITE!" and a speech bubble icon. A large yellow star graphic with the year "2012" is overlaid on the bottom right of the screenshot.

Bauhaus-Universität Weimar

UNIVERSITÄT ARCHITEKTUR BAUINGENIEURWESEN GESTALTUNG MEDIEN

Aktuell Profil Struktur Studium International Forschung und Kunst Partner und Alumni Service

Die Bauhaus-Universität Weimar

Architektur, Bauingenieurwesen, Gestaltung und Medien – mit Ihren Fakultäten und Arbeitsgebieten verfügt die Bauhaus-Universität Weimar über ein einzigartiges Profil. Das Spektrum der Universität umfasst gegenwärtig über 40 Studiengänge und reicht von der Freien Kunst über Design, Web-Design, Visuelle Kommunikation, Mediengestaltung und -kultur bis zu Architektur, Bauingenieurwesen, Baustoffkunde und Umwelt sowie Management. Der Begriff »Bauhaus« im Namen unserer Universität steht heute für Experimentierfreudigkeit, Offenheit, Kreativität, Nähe zur industriellen Praxis und Internationalität. > mehr

GEBEN SIE UNS
IHR FEEDBACK
ZUR NEUEN WEBSITE!

2012

Web Archives as a Dataset

The screenshot shows the homepage of the Bauhaus-Universität Weimar. At the top, there's a navigation bar with tabs for UNIVERSITY, ARCHITECTURE AND URBANISM, CIVIL ENGINEERING, ART AND DESIGN, and MEDIA. Below the navigation is a banner featuring a collage of various objects related to Bauhaus, including a lamp, books, a camera, and architectural models. A red circular badge on the left side of the page contains the word "NEWS". To the right of the badge, there's a news item about a project called "Wissen erobern". Further down the page, there are two yellow call-to-action boxes: one for "KANDIDIERE FÜR DEINEN FACHSCHAFTSRAT" and another for "NEUE BAUHAUS VORTRÄGE". A large yellow star graphic with the word "today" in red is overlaid on the bottom right.

Bauhaus-Universität Weimar

UNIVERSITY ARCHITECTURE AND URBANISM CIVIL ENGINEERING ART AND DESIGN MEDIA

News Profile Structure Studies International Research and Art Partners and Alumni Bauhaus100 Students Representatives Services

NEWS

Bildungsprojekt »Wissen erobern« vermittelt Schülerinnen und Schülern wissenschaftliches Arbeiten

Am Freitag, 9. März, und am Montag, 12. März 2018, findet das Gemeinschaftsangebot »Wissen erobern. Schülerinnen und Schüler entdecken Welmarer Bibliotheken« für die Gymnasien des Welmarer Landes und der Stadt Weimar statt. Mit dem Projekt soll den Schülerinnen und Schülern Hilfestellung bei der Bearbeitung von Seminarfacharbeiten durch öffentlich zugängliche Bibliotheken gegeben werden. Thematisch dreht sich in diesem Jahr dabei alles um das Bauhaus-Jubiläum. [» more](#)

[» To all news in the BAUHAUS.JOURNAL ONLINE](#)

KANDIDIERE FÜR DEINEN FACHSCHAFTSRAT
#BUWählt
10.-16. Mai 2018

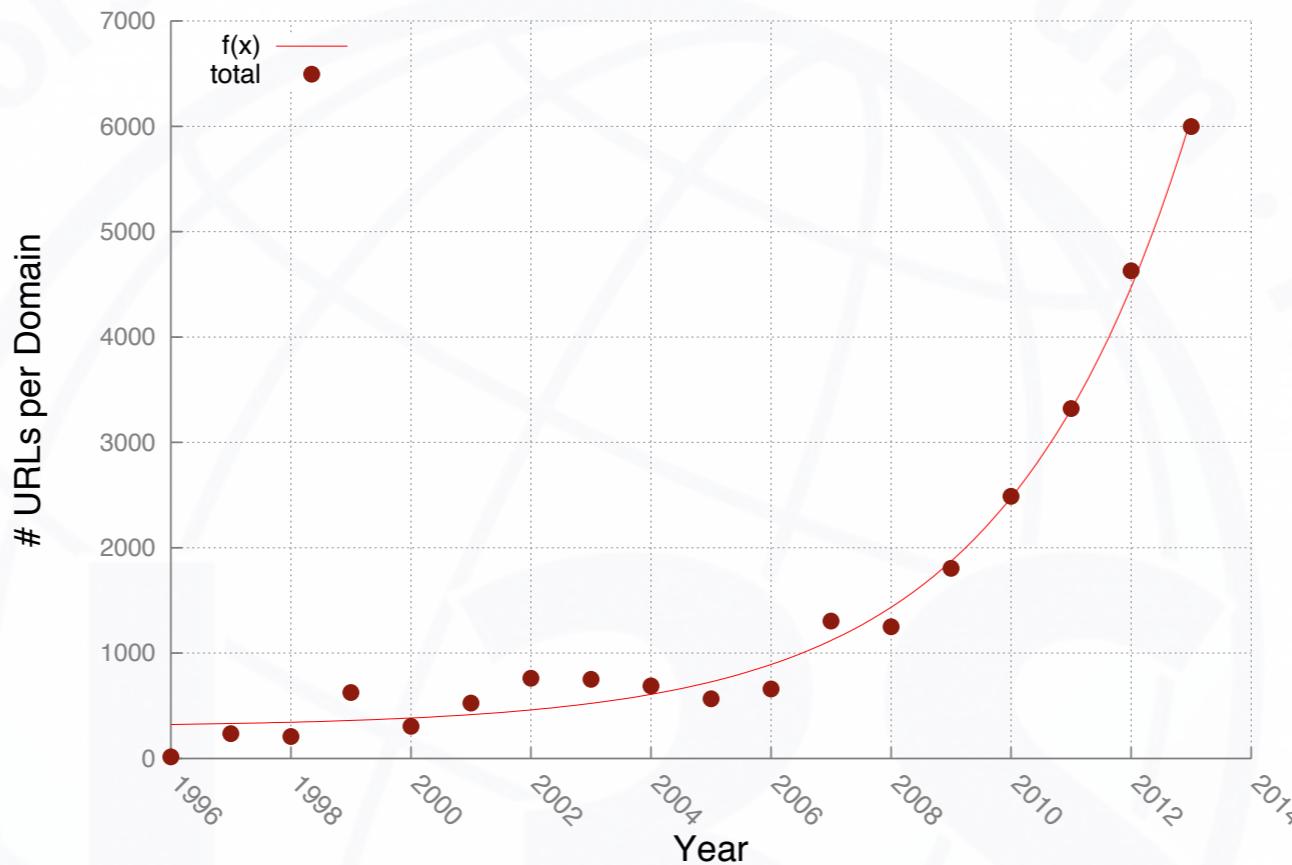
NEUE BAUHAUS VORTRÄGE

today

Why are they valuable ?

They encode **history** and the study of **long-term change** and **evolution** possible

Really How big is the Web? Predictions



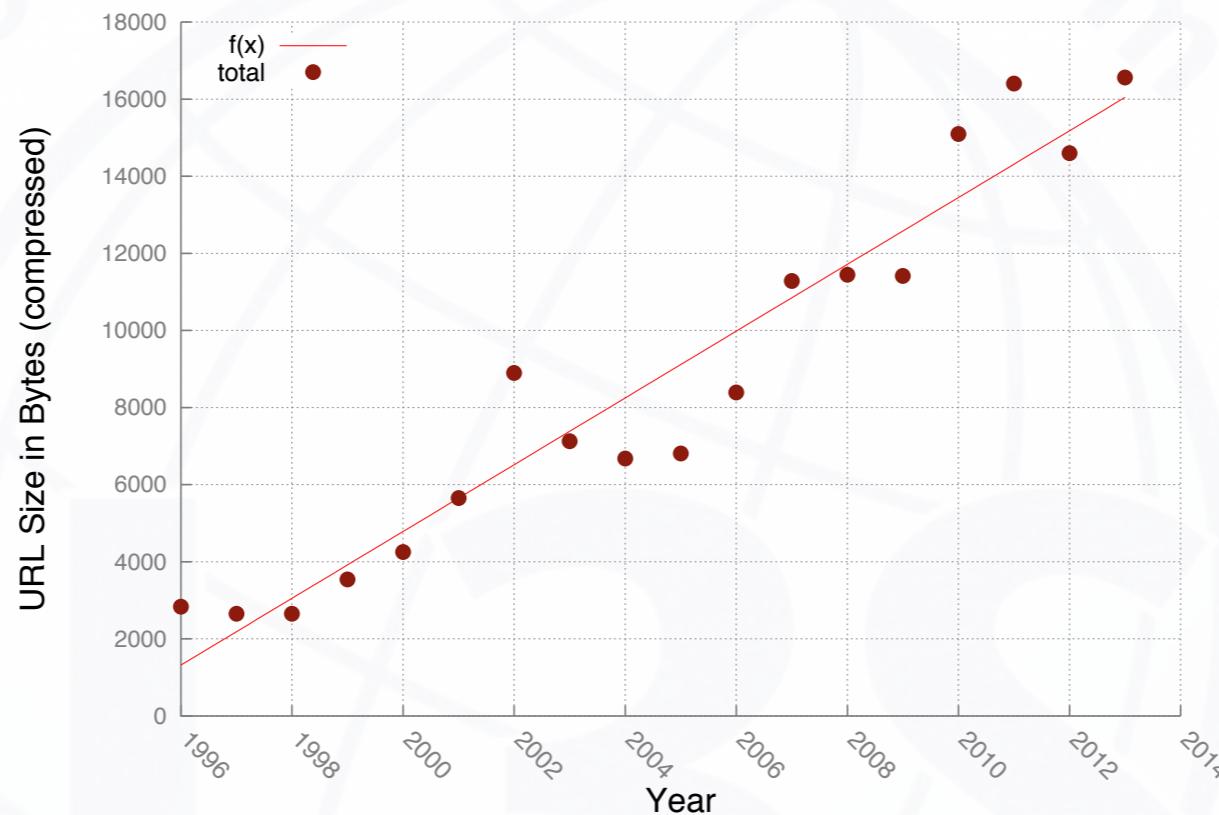
By 2017 the domain volume will **double that of 2013**

By 2020 the number of URLs will be **6.7 times** that of today

By 2030 almost **166 times** the number of URLs today

*The Dawn of Today's Popular Domains: A Study of the Archived German Web over 18 Years..
in JCDL'16: H. Holzmann, W.Nejdl, A.Anand.*

Really How big is the Web? Predictions



A new URL in 2017 will be born with **double the file size as of 2013**

long data is also big data

Research Issues

Problems of Access

- How do we access temporal collections ?
- How do we efficiently support temporal queries ?

[*SIGIR '11; SIGIR '12; CIKM '10; JCDL '16*]

Searching Temporal Collections

- How can we support temporal information needs ?
- How do we devise retrieval models taking history into account ?

[*CHIIR '16; SIGIR '16; WWW '16; WSDM '17*]

Enrichment

- How do we extract knowledge from temporal collections ?

[*CIKM '15; CIKM'16; EMNLP'17*]

Access

Search

Enrichment

Problem I: Efficient Access

From Analysis to Access

From Analysis to Access

Archives have great value towards

- historical information needs
- Searching the past

The history NATO interventions in the last 100 years...

Access

Tunis Asks **NATO** Intervention to End French-Algerian Conflict

TUNIS, Feb. 27. (AP)—Tunisian President Habib Bourguiba today called on the NATO powers to intervene to end the conflict between France and the Algerian rebels.

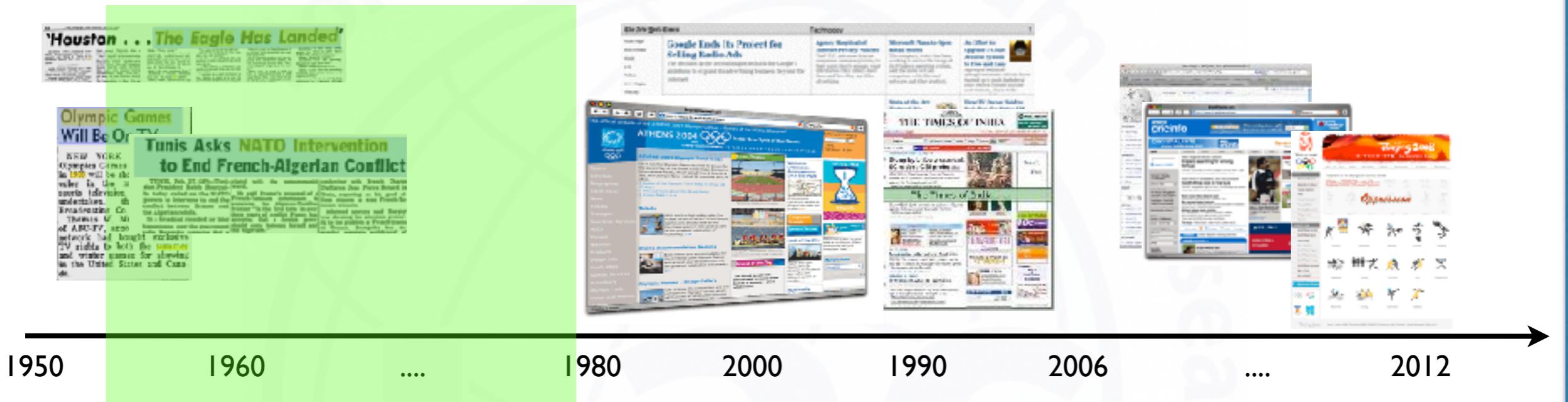
ciated with the noncommunist world.

He said France's proposal of a French-Tunisian commission to supervise the Algerian-Tunisian frontier "is the first time in over three years of conflict France has

conferring with French Charge D'affaires Jean Pierre Benard in Tunis, reporting on his good offices mission to ease French-Tunisian tensions.

Informed sources said Murphy

Time-Travel Text Search



nato intervention @ 1950-1980

[Tunis Asks Nato Intervention To End French-algerian...](#)

Spokane Daily Chronicle - Feb 27, 1958

Habib Bourguiba today called on the **NATO** powers to **intervene** to end the conflict between **France** and the Algerian rebels. In a broadcast recorded for later ...

[Nato .Intervention Ashed In French,...](#) Reading Eagle

Search as a primitive operation

Time-Travel Text Search

- Given a versioned collection of documents with valid time intervals
- Time-travel text queries

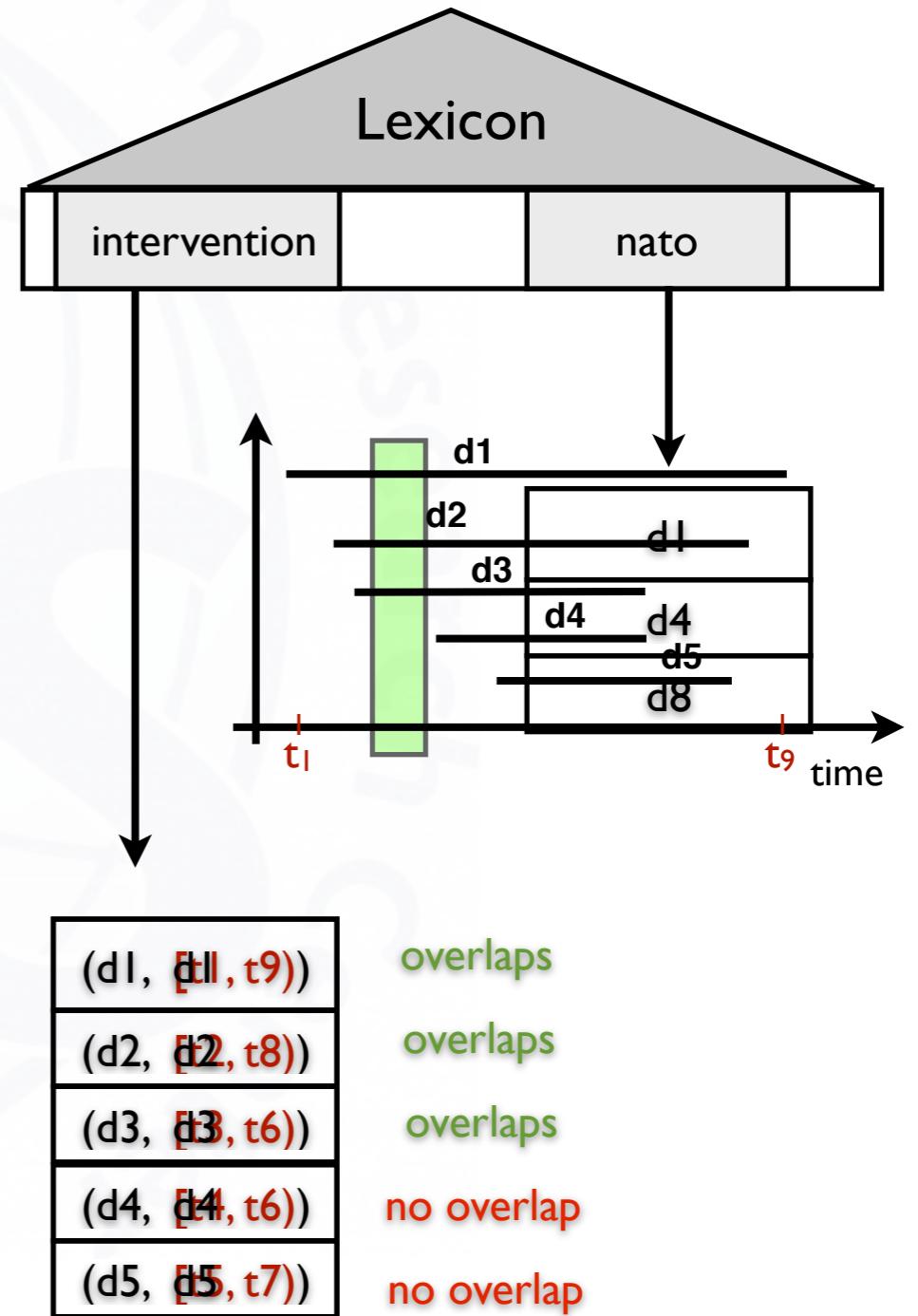
nato intervention @ 2001 - 2004

- We want to retrieve documents containing terms “nato”, “intervention” and valid between **2001 - 2004**

How do we efficiently retrieve these documents ?

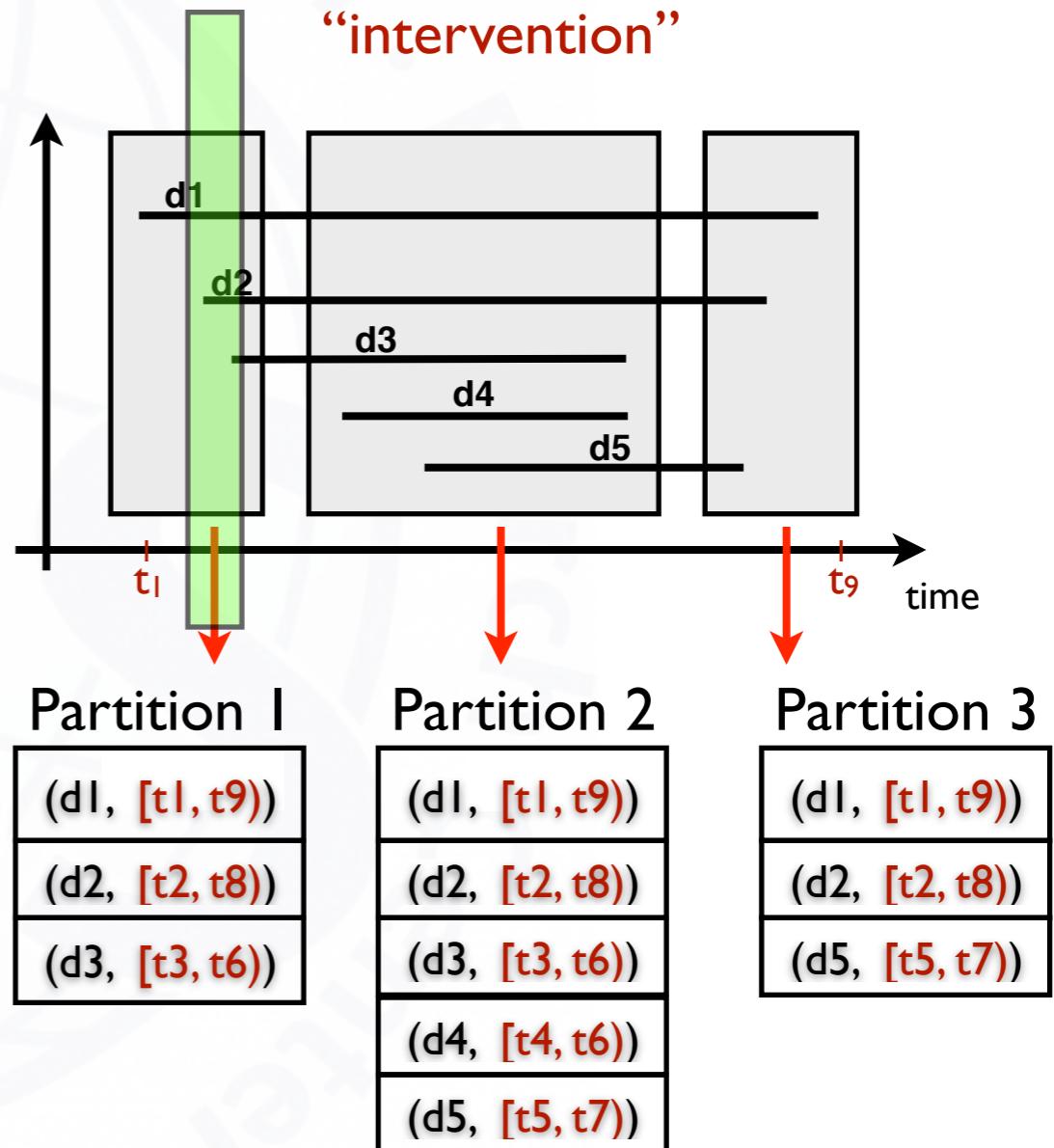
Time-Travel Index

- Inverted index processes keyword queries
- Intersection of posting lists for processing queries
- Versions have valid time intervals
- TTIX proposed by [Berberich et.al SIGIR '07] extends the standard inverted index



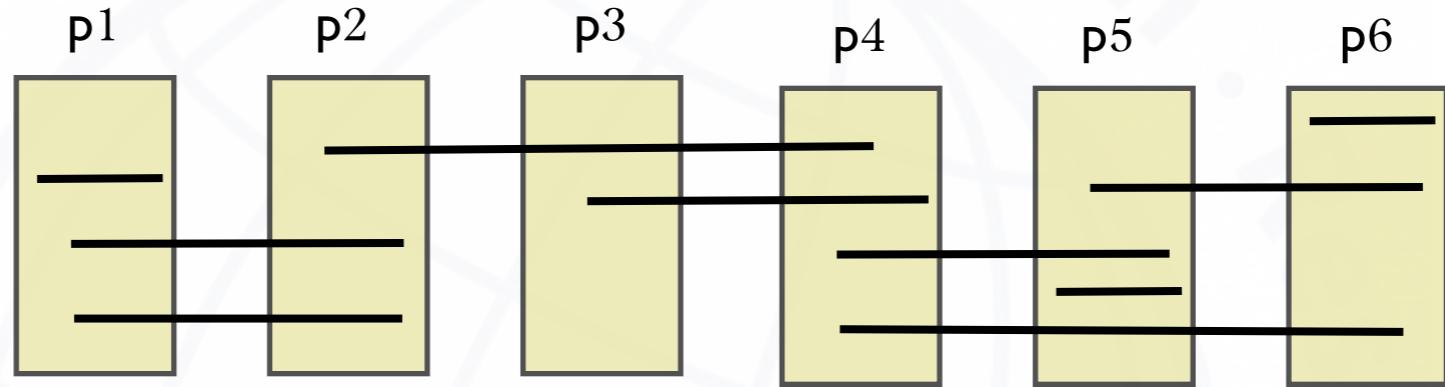
Vertical Partitioning

- Posting lists partitioned along the time-axis
- Each partition has a time span
- Query processing selects subset of partitions
- Postings are replicated



Trade-off between index size and performance

Effect of Replication

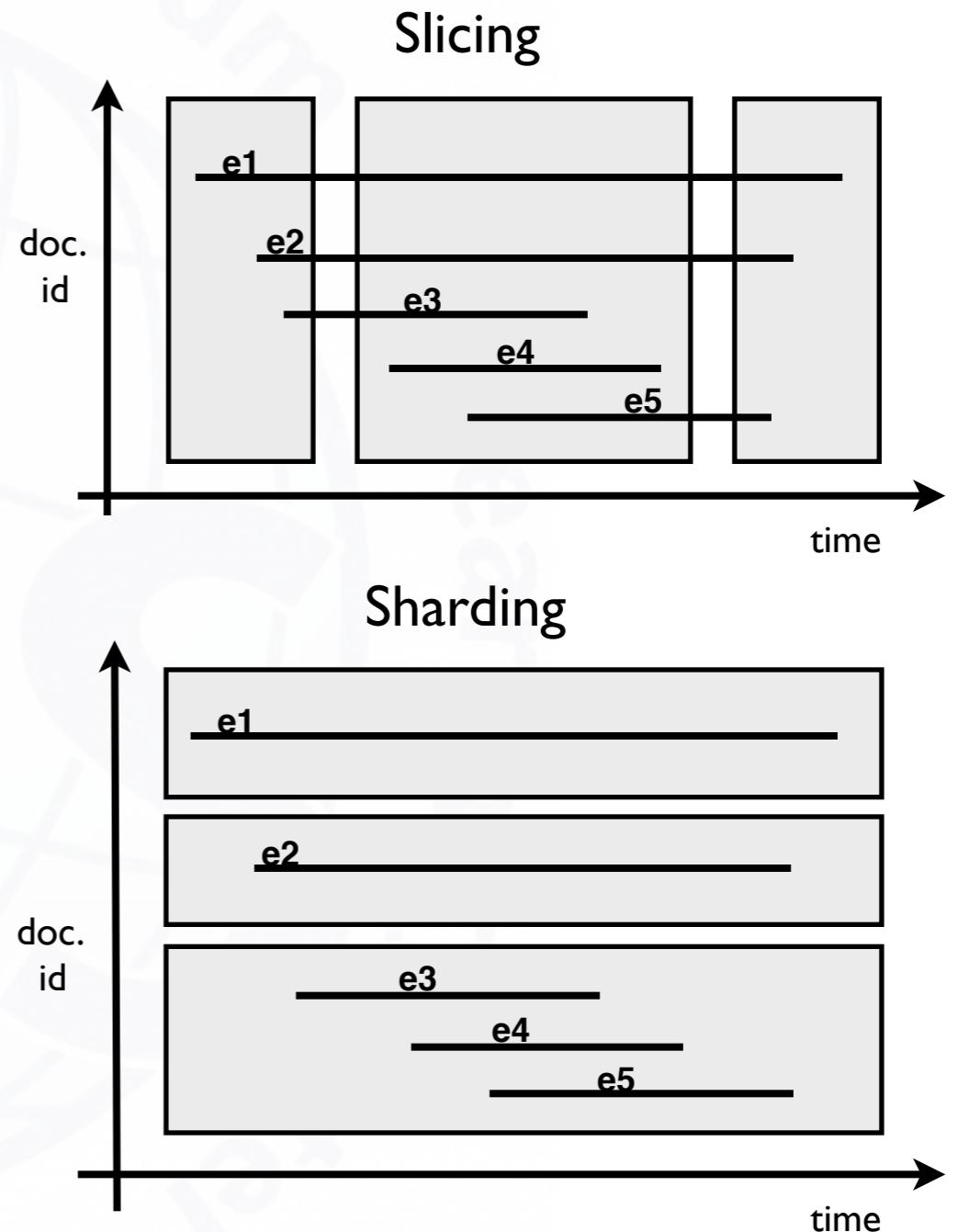


- Index size blowup due to replication of postings across partitions
- Index maintenance is hard

Can we partition a posting list without replicating postings ?

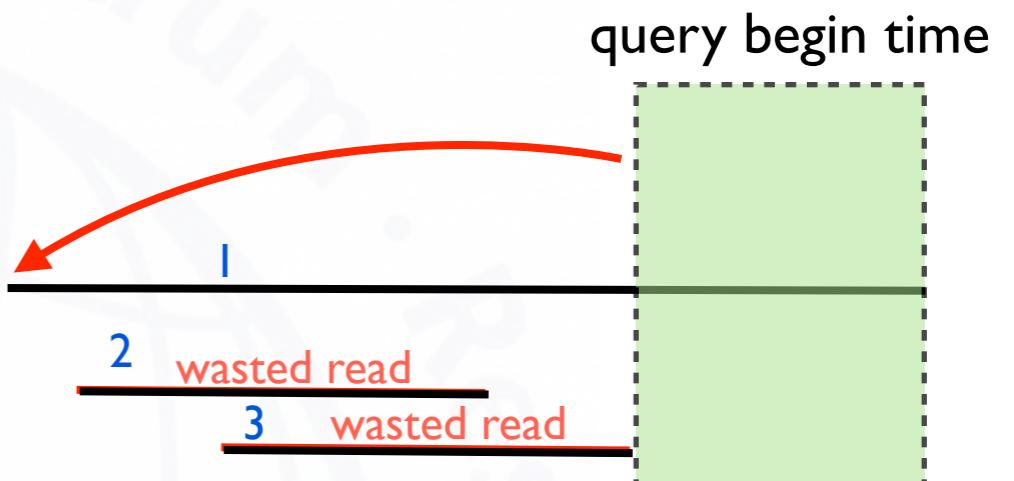
Index Sharding

- Partition documents in each posting list into sublists called **shards**
- Contents of each shard disjoint - no replication, no index blowup
- Postings stored in begin time order
- Access structure over each shard for efficient query processing

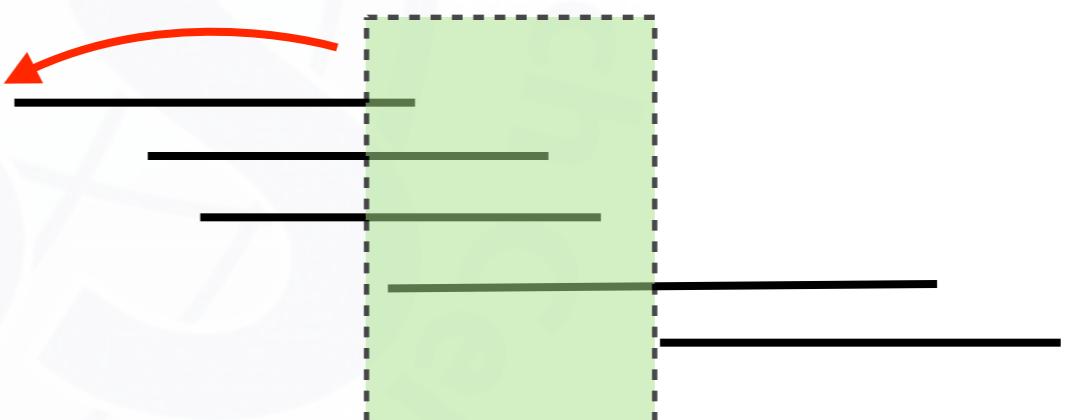


Staircase Property in a Shard

- Wasted reads are processed but do not overlap with the query time interval
 - Staircase property in a shard
 - Intervals arranged in begin time order
 - No interval completely subsumes another interval
 - Eliminates wasted reads

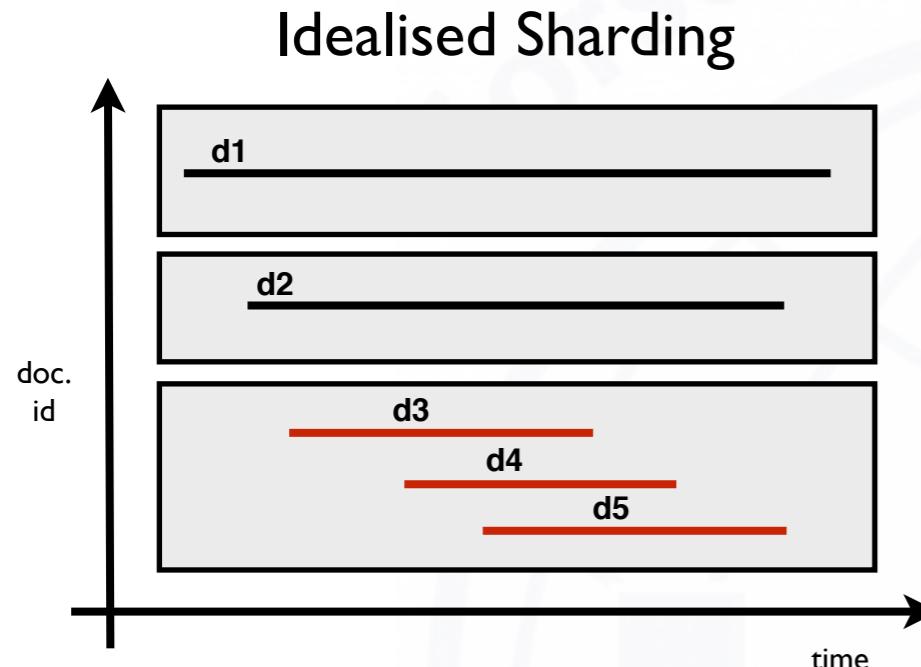


wasted reads due to subsumption

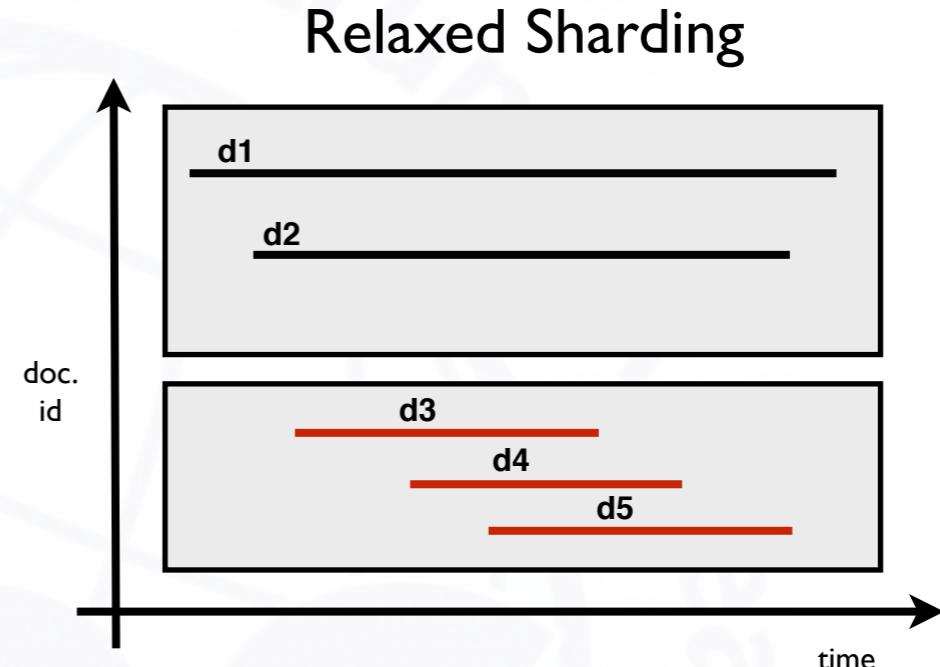


staircase property

Index Sharding



more random accesses



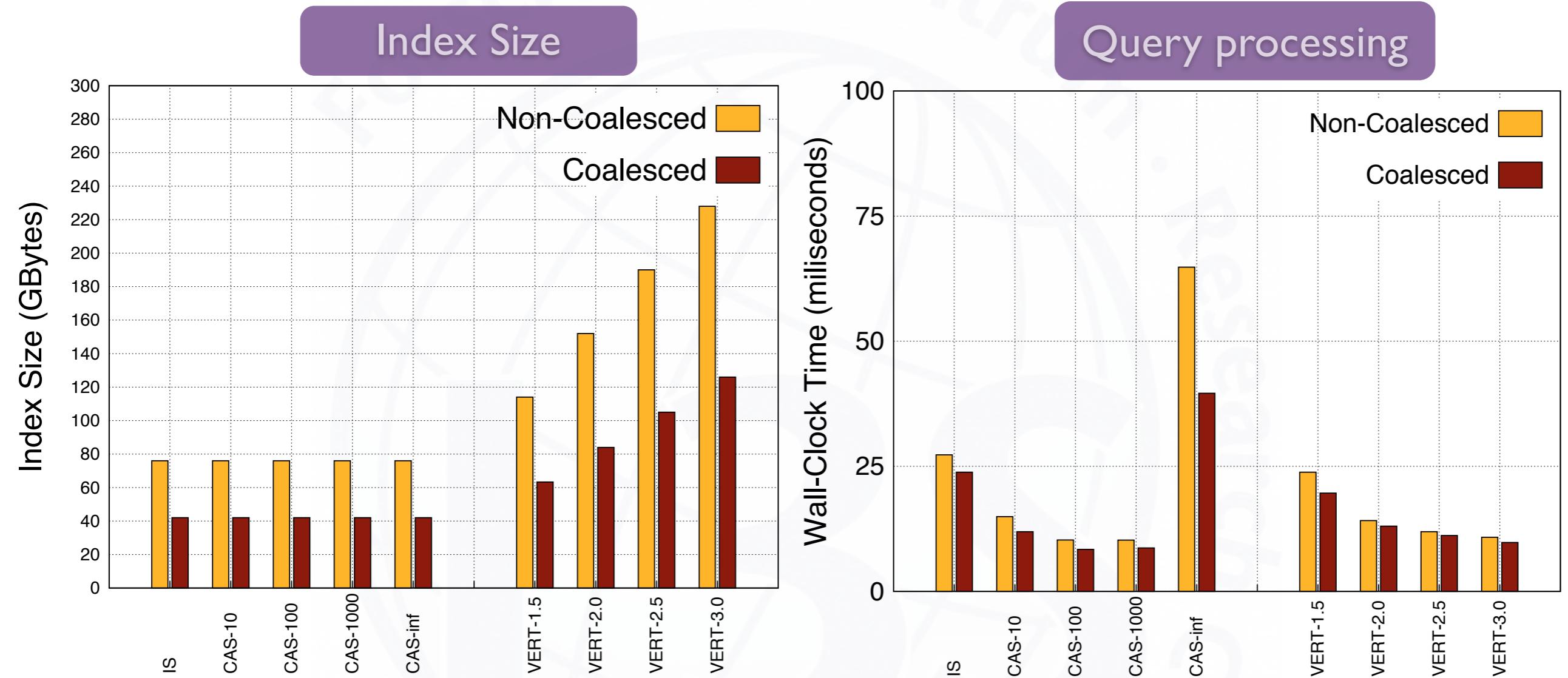
more wasted reads

- **Query processing performance bad if —**
 - more no. of shards
 - more violations

Trade-off between costs of sequential and random reads

Indexing Performance

Access



- No index size blowup in sharded index
- 22.2% improvement over best VERT for year granularity queries

Dataset: [Wikipedia](#)

Queries: [AOL](#)

*Temporal Index Sharding for Space-Time Efficiency in Archive Search,
in SIGIR'11. A.Anand, S. Bedathur, K. Berberich, R. Schenkel*

Problem 2: Result Ranking

From Access to Search

Issues in Temporal Ranking

- Current ranking methods focus on **freshness** of results
- **News ranking** – historical importance is also relevant
- Search intents could be temporal - Historical Information Needs
- Useful for social scientists, historians etc.

News Ranking

- Most consumption and production of news online
- More than 100k articles published daily
- Information Overload

The New York Times

REUTERS

the guardian

BBC
NEWS

CNN

ABC
News



USA
TODAY

H
THE
HUFFINGTON
POST



THE HINDU



THE TIMES OF INDIA

How do we automatically rank a daily batch of news ?

News Ranking - Classical



Belongs to long
running "Iraqi
civil war"

1

Orlando nightclub shooting: How the attack unfolded
How the Orlando Shooting Unfolded
What Happened Inside the Orlando Nightclub

2

ISIL attack on army barracks near Fallujah kills dozens

2

Microsoft to buy LinkedIn for \$26bn
LinkedIn bought by Microsoft for \$26.2bn in cash
Microsoft to buy LinkedIn for \$26 billion

3

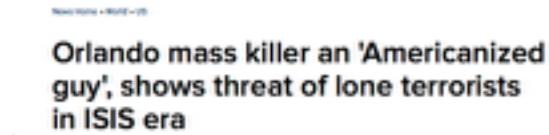
Car bomb rocks eastern Turkey
Report: 9 injured in car bomb explosion in southeast Turkey

Not popular
today!

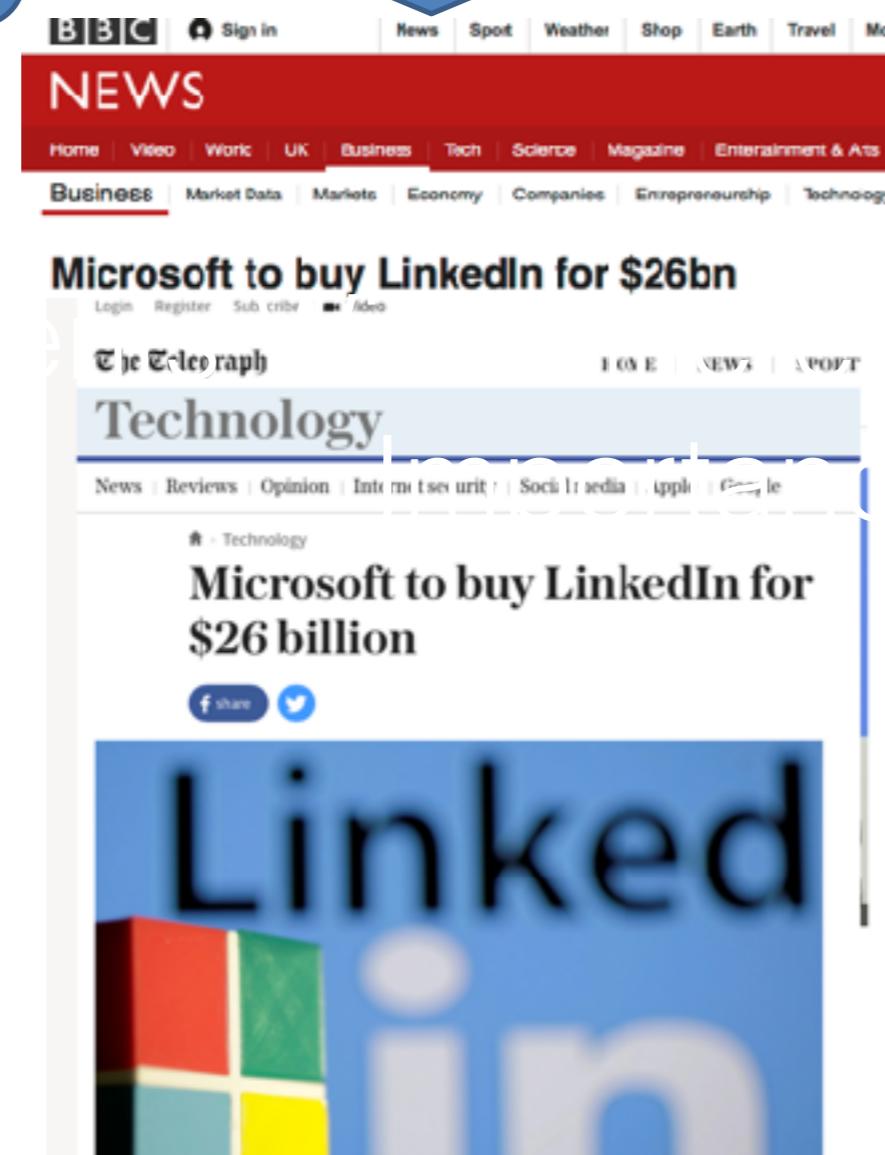
Historical Importance!

Characterising Event History

Short Term Events
highly popular



One Off Events

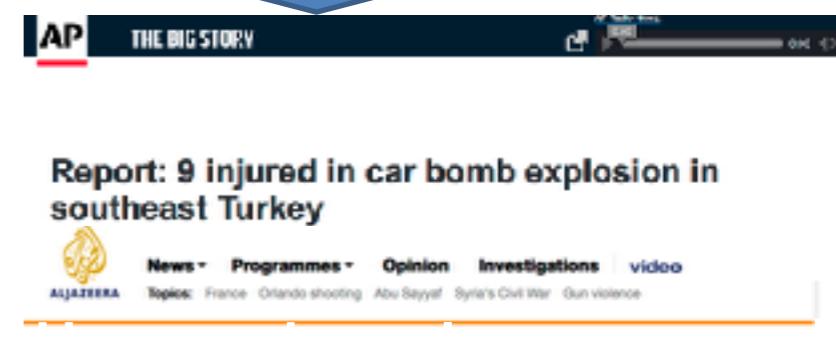


The Microsoft and LinkedIn logos are pictured together in Taipei, Taiwan. Microsoft will buy LinkedIn for \$26.2bn in cash. CREDIT: EPA

By James Titcomb

13 JUNE 2016 • 7:33PM

Long Running Events

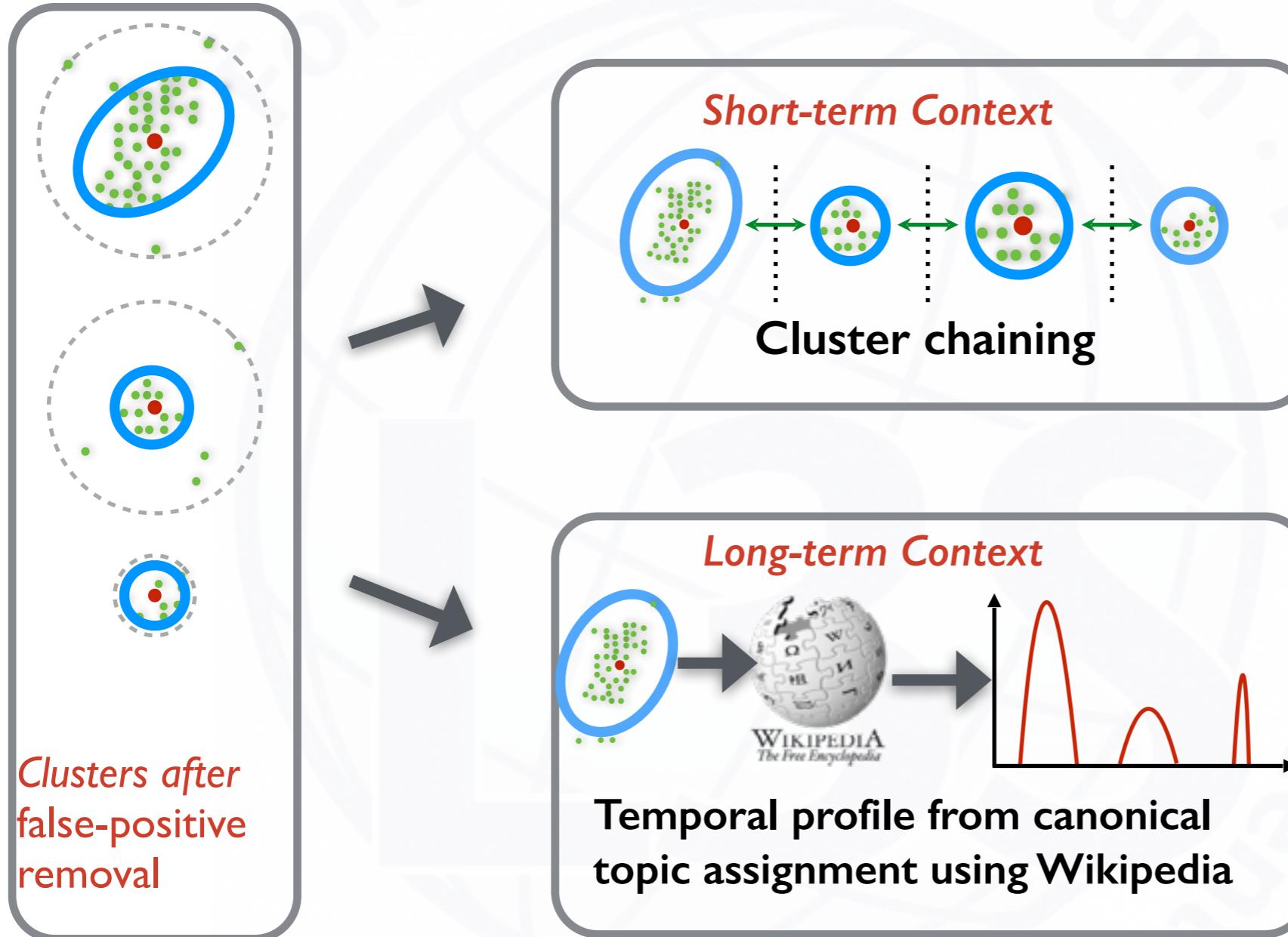


Dozens of Iraqi government forces and militia members have been killed in an attack on military barracks east of Fallujah by the Islamic State of Iraq and the Levant (ISL).

Military sources said 50 members of the Iraqi security forces and allied Shia militias were killed on Saturday.

Search

Exploiting Historical Cues

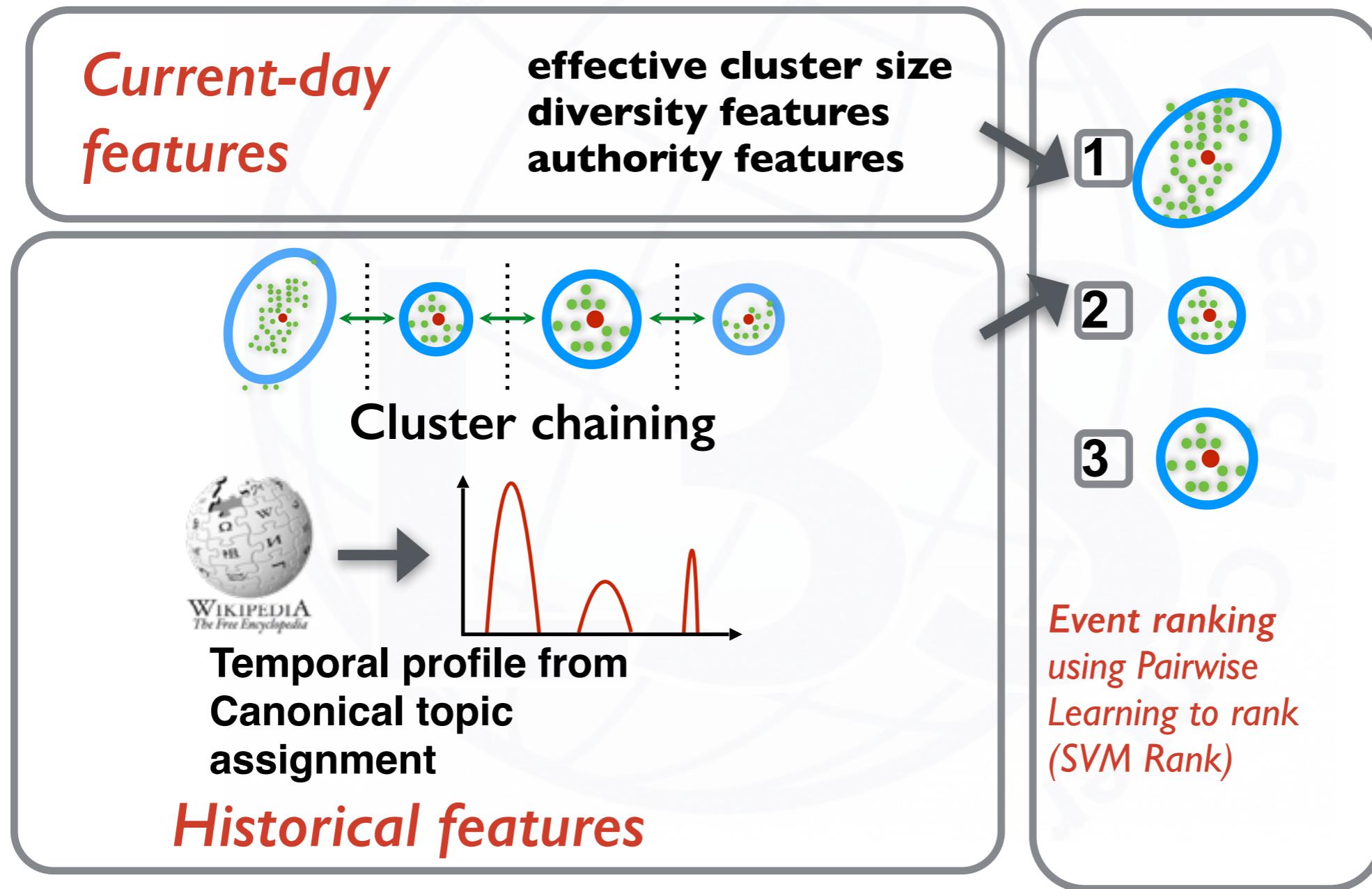


Exploit page views, edits whenever possible by canonicalizing articles to EP

Modeling Event Importance for Ranking Daily News Events.

in WSDM '17: A.Anand, V. Setty, A. Mishra, A.Anand

Ranking Events



Experiments

3k articles
a day

STICS Data	P@10	MAP@10	F1@10
LM-w/o Hist	0.421	0.404	0.233
LM-Hist	0.507	0.499	0.281
Event-Rank-w/o	0.582	0.608	0.437
Event-Rank- Hist	0.882	0.885	0.660

20k - 100k
articles a day

Gdelt Data	P@10	MAP@10	F1@10
LM-w/o Hist	0.393	0.465	0.275
LM-Hist	0.405	0.480	0.283
Event-Rank-w/o	0.578	0.690	0.517
Event-Rank- Hist	0.675	0.750	0.588

- First large scale longitudinal evaluation for News Ranking (365 days)
- Automatic evaluation using Wikipedia Current Events Portal

From Access to Search

The British NEWSPAPER Archive

Sign-in | Subscribe | Register

In partnership with BRITISH LIBRARY

HOME | BLOG | HELP & ADVICE | ABOUT | REGISTER | SEARCH

Home > Search > Results

Filter your search...

Date Range:
1st Jan 1950 - 31st Dec 1991
Remove all

Filter by Date
(Number of Articles)
1950 - 1999 [Remove]
1950 - 1959 (418,424)

Filter by Newspaper
(Number of Articles)
Yorkshire Post and Leeds Intelligencer (75,508)
Portsmouth Evening News (46,274)
Dundee Courier (26,695)
Sunderland Daily Echo and Shipping Gazette (25,238)
Aberdeen Evening Express (23,503)
+ more newspapers...

Filter by Region
(Number of Articles)
Yorkshire and the Humber, England (109,202)
South East, England (83,064)

Your results for: london

london

Keep filters Exact search

Search + Advanced

Page 1 2 3 4 5 ... 34869 | » Order by Relevance ▲ Date ▲

London
... London ...

Friday 08 January 1954, Yorkshire Evening Post, West Yorkshire, England
Type: Article Words: 1 Page: 8 Tags: none
Bookmark View

London,
... London, ...

Friday 24 February 1950, Derry Journal, Londonderry, Northern Ireland
Type: Article Words: 1 Page: 2 Tags: none
Bookmark View

Search Intent

Ranking Textual relevance

33

How Historians Search

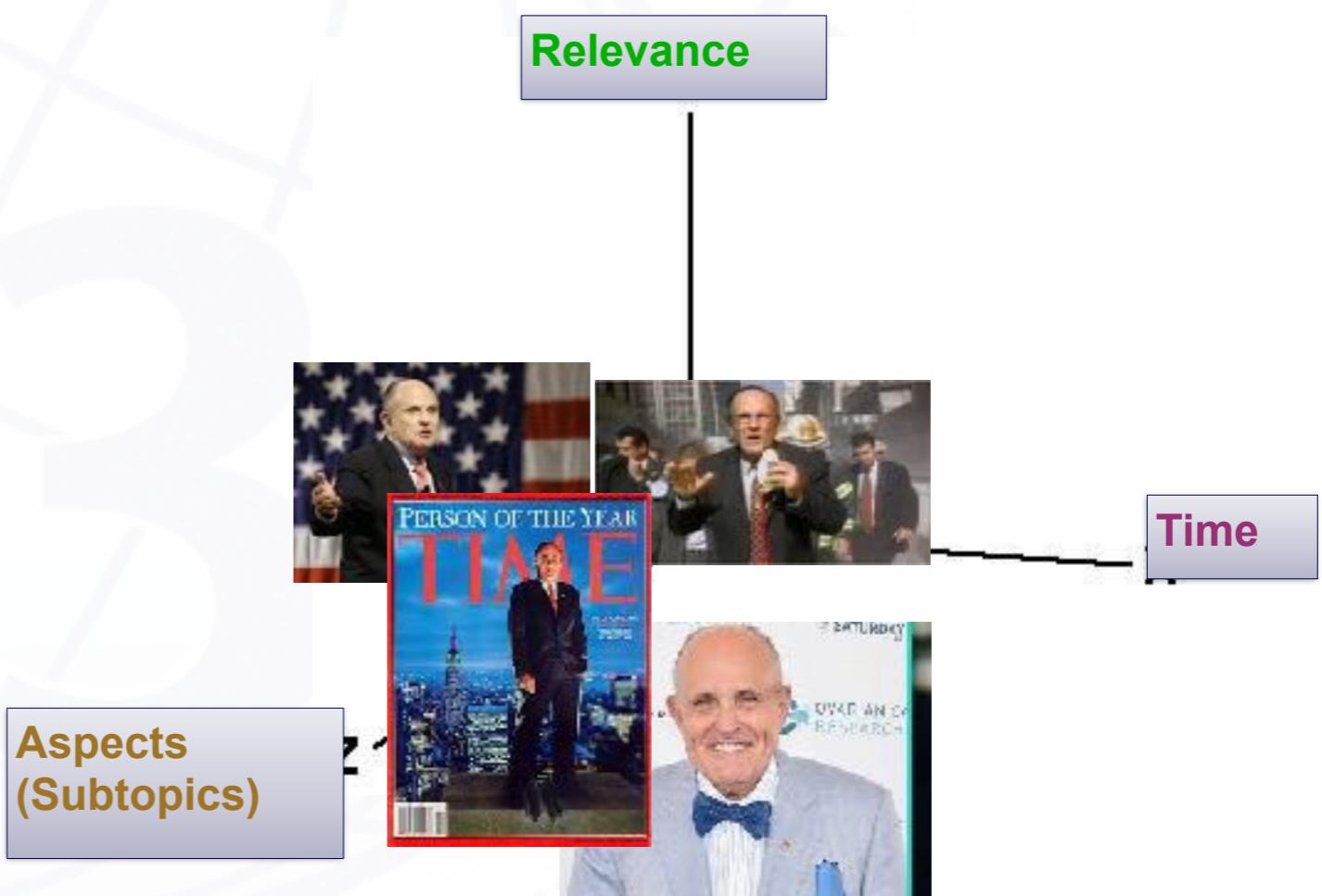


Primary Sources

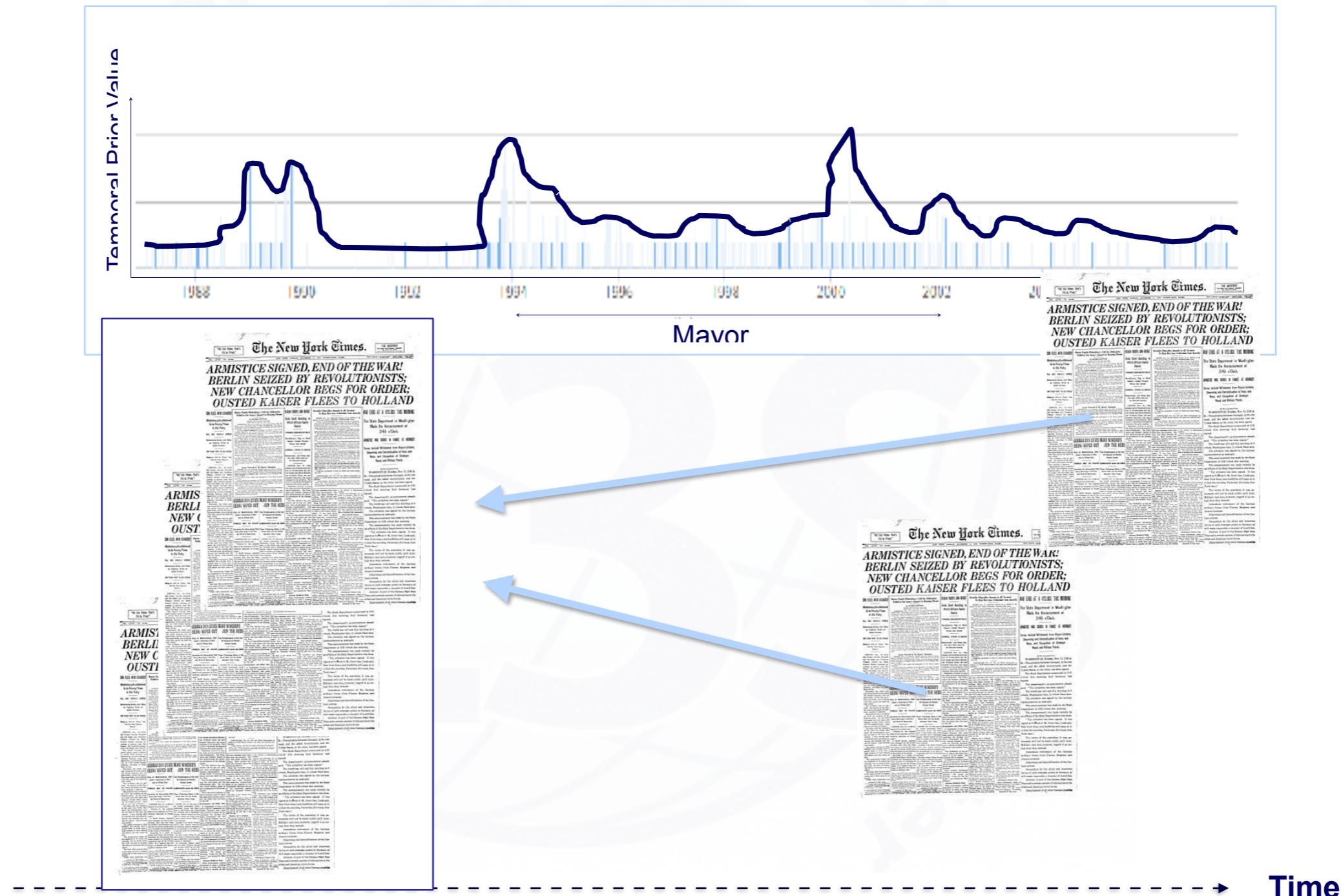
Browsing the Shelves

Problem Statement

*“I am looking for **relevant** documents regarding the most important **aspects** from the **time** period when these aspects were relevant.”*



Modelling Time



Modelling Aspects

cnn politics    

Rudy Giuliani (sort of) endorses Donald Trump

By Julia Manchester, CNN
① Updated 2114 GMT (0514 HKT) April 19, 2016



Entities as Aspects

Story highlights

Giuliani said he would endorse Trump, but not have a role on the campaign

He thinks Hillary Clinton would easily dispatch with Ted Cruz in a general election but have no idea how to handle Trump

When pressed by Cuomo to clarify what he meant, Giuliani repeated that he would endorse Trump, but not have a role on the campaign.

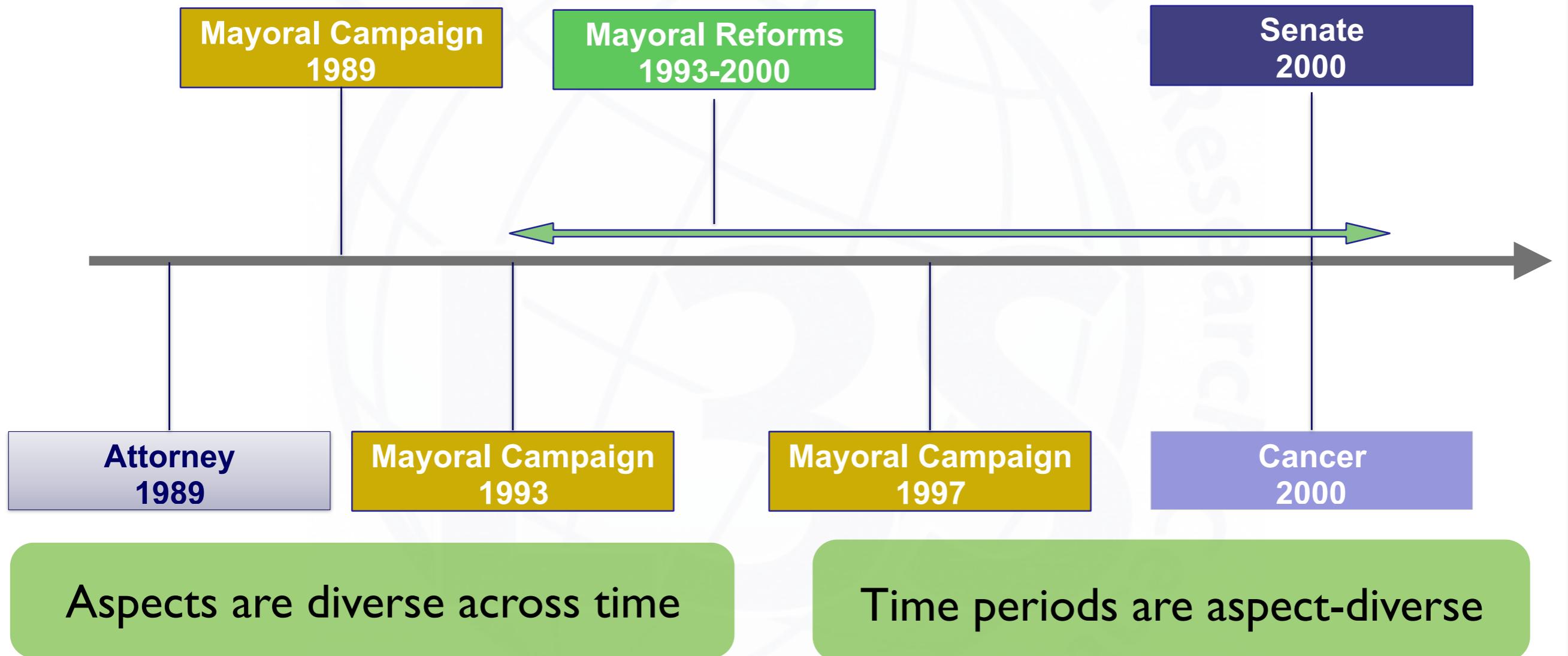
Washington (CNN) — Former New York City Mayor Rudy Giuliani endorsed Republican front-runner Donald Trump Tuesday, the day of the New York primary.

"I'll endorse, but I'm not a part of the campaign," Giuliani told CNN's Chris Cuomo on "New Day."

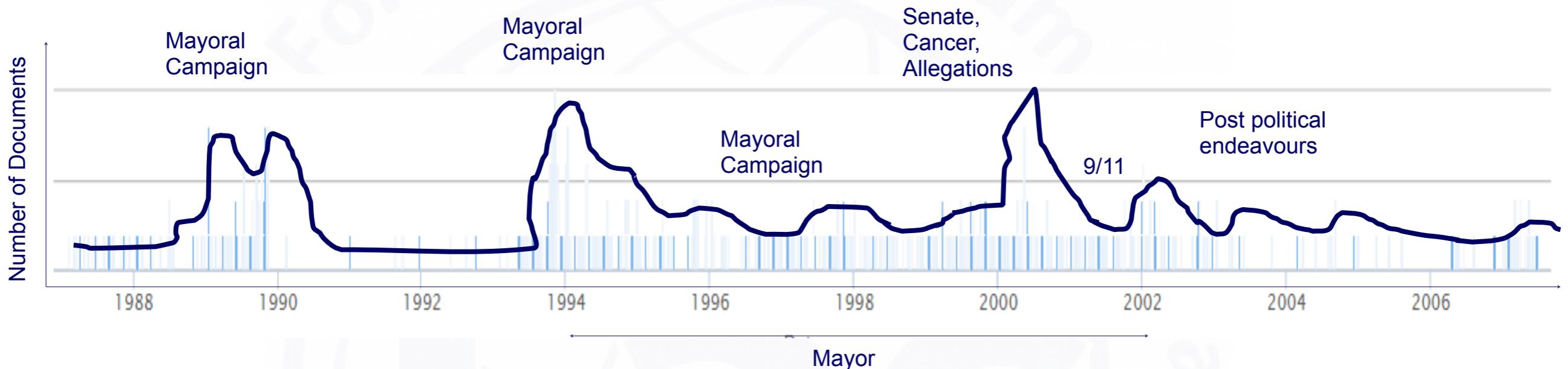


Canonicalisation of entities

Coverage Problem



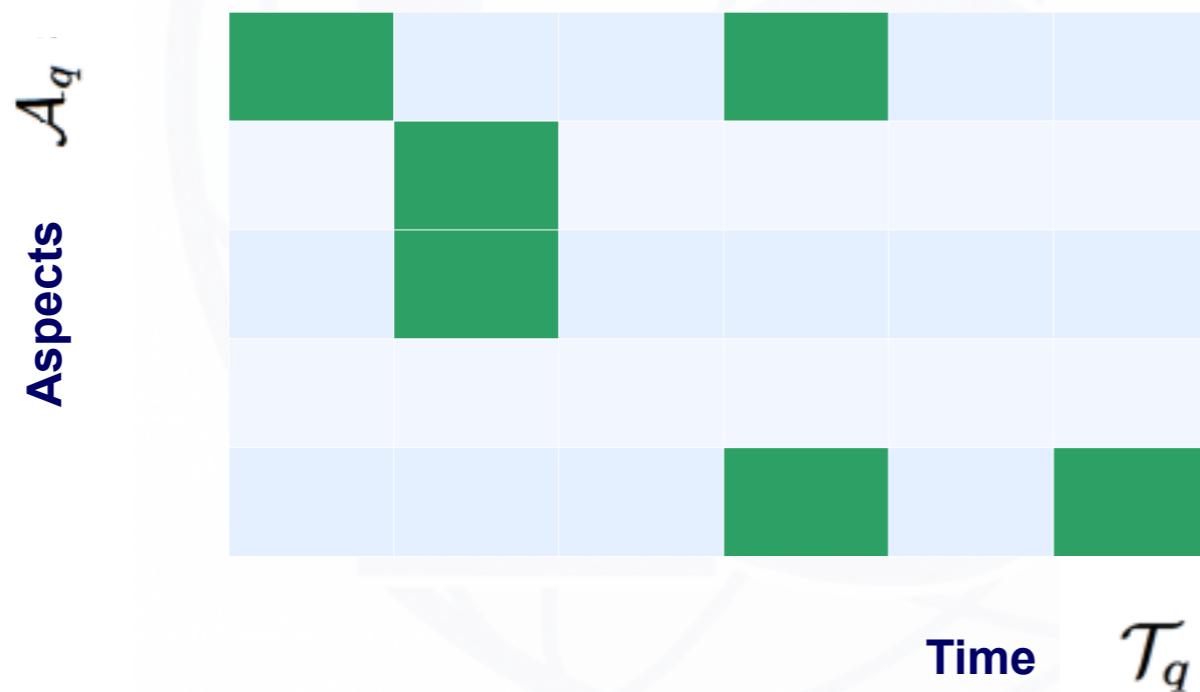
The HistDiv Approach



- Aspects are temporal in nature -
 - Aspect utility is updated using an **exponential decay function**.
- Time windows are aspect diverse -
 - Time window utility is updated based on **aspect coverage**
- We represent aspects as **entities** - **David Dinkins, Republican Party, Manhattan.**

Historical Search

- Standard diversity: coverage of aspect space (entity space)
- Temporal diversity: coverage of time space (temporal profile)
- **Historical diversity: coverage of aspect-time space**



- Greedy algorithm for set cover — specialised discounting

Test Collection for Historical Search

- TREC datasets are **short time spans**
 - topics and subtopics are not suited for historical search.
- Manually created topics and subtopics using relevant **wikipedia history sections**.
- NYT news archive 1987 - 2007
- Expert binary relevance judgements for 30 topics.

The screenshot shows a Wikipedia page for the Library of Alexandria. On the left, there is a sidebar with language links for various European languages. The main content area displays the "History" section of the article. The text discusses the library's purpose, its organization by Demetrius of Phaleron, and its role in translating works onto papyrus paper. It also mentions the fees charged for copying books and the library's influence over other cities like Rhodes and Athens. The text is heavily annotated with blue hyperlinks to other articles.

Nederlands
नेपाली
日本語
Norsk bokmål
Norsk nynorsk
Occitan
Polski
Português
Română
★ Русский
Slovenščina
Slovenčina
Српски / srpski
Srpskohrvatski / српскохрватски
Soomaalی
Svenska
Тоҷикӣ
Türkçe
Українська
ଓଡ଼ିଆ
Tiếng Việt
Wîlîmrîy
中文
Edit links

commerce bringing income to the library.^[11]

History [edit]

The library was arguably one of the largest and most significant in the world at the time. Its main purpose was to show off the vast knowledge and power of the Ptolemaic Empire. It was used to aid the ruler of Egypt.^[12]

According to the earliest source of information, the pseudographical *Periplus of the Erythraean Sea*, the library was initially organized by Demetrius of Phaleron,^[13] around 280 BCE—c.283 BCE). Other sources claim it was instead created by the ruler of Egypt, Ptolemy I Soter, around 290 BCE. The library was built in the Bruchelion (Royal Quarter) in the style of a Greek Temple or "House of Muses", whence the term "museum". The Library at Alexandria was in charge of collecting all the world's knowledge and translating works onto papyrus paper.^[14] It did so through an international network of scholars, scribes, and translators from all over the Mediterranean world. The Library at Alexandria was in charge of collecting all the world's knowledge and translating works onto papyrus paper.^[15] It did so through an international network of scholars, scribes, and translators from all over the Mediterranean world. The Library at Alexandria was in charge of collecting all the world's knowledge and translating works onto papyrus paper.^[16] It did so through an international network of scholars, scribes, and translators from all over the Mediterranean world. The Library at Alexandria was in charge of collecting all the world's knowledge and translating works onto papyrus paper.^[17] According to Galen, any book that was listed as "books of the library" were delivered to the owners.^[18] Official scribes then copied the work and delivered it to the owners. Other than collecting works, the library was well-patronized by the Ptolemaic dynasty with travel grants and stipends. According to Galen, Ptolemy III requested permission from the Athenians to copy their works, and Euripides, for which the Athenians demanded the enormous sum of 300 talents as guarantee. Ptolemy III happily paid the fee but kept the original copies to show the power of Alexandria over Athens during the Ptolemaic dynasty. The library was located between the mainland and the Pharos island, welcoming traders from all over the Mediterranean world. The editors of the Library of Alexandria are especially well known for their work on the *Encyclopedias* of the 1st century BC.

Experiments

- Main Metric: Subtopic Recall
- Other Metrics: α -NDCG, IA-ERR, IA-Precision, MAP
- Window size: year & month
- Aspects mined from AIDA & wikiminer (NE linking)
- Tuned for best performance in subtopic recall.
- IA-SELECT & PM2, MDIV, OnlyTime

Experiments

- Manually created topics and subtopics using relevant **wikipedia history sections**.
- NYT news archive 1987 - 2007
- Recall at the cost of precision

Subtopic Recall @ k

	10	15	20
LM	0.428	0.471	0.509
OnlyTime	0.426	0.497	0.534
IA-Select	0.442	0.491	0.516
MDIV	0.460	0.497	0.534
HistDiv	0.497	0.542	0.583
HistDiv-Burst	<u>0.509</u>	<u>0.560</u>	<u>0.601</u>

Highlights & Impact



*Expedition: A Time-Aware Exploratory Search System Designed for Scholars
in SIGIR '16: J. Singh, W. Nejdl, A. Anand.*

*History by Diversity: Helping Historians search News Archives.
in CHIIR '16: J. Singh, W. Nejdl, A. Anand.*

- Hannover Impuls Ideas competition — runner up

Demo: bit.ly/archive-search

Searching Archives without a Full Text Index

The screenshot shows the Tempas search interface. At the top, there's a navigation bar with "Tempas search" and a logo. Below it is a search bar with "bolzano" and a "Go!" button. To the right are three horizontal timelines: one for the year (2003-2011), one for the month (Jan-Dec), and one for the day (Feb-Mar-Apr-May-Jun-Jul-Aug-Sep-Oct-Nov-Dec). A list of search results follows:

- italy [blue button]
- hotel [blue button]
- travel [blue button]
- bozen [blue button]
- ari [blue button]
- pomodoro [blue button]
- scuola [blue button]
- opensource [blue button]
- ping.fm [blue button]
- trento [blue button]
- software [blue button]
- video [blue button]
- southtyrol [blue button]
- north [blue button]
- wittkowsky [blue button]
- tis [blue button]
- web [blue button]
- ÄÄyouth [blue button]
- zelig [blue button]

Below the results, there's a link to "www.sii.bz.it" and a date range "2009-12-08 - 2008-09-20". The main content area displays a large graphic for "SII" (Servizio Informativo aziende trasporto pubblico locale) featuring a sunburst design with orange arrows pointing outwards from a central image of a landscape. Text below the graphic reads: "Servizio Informativo aziende trasporto pubblico locale" and "Informations- und Serviceprovider der öffentlichen Nahverkehrsbetriebe". On the right side of the SII page, there are links for "ITALIANO", "DEUTSCH", and "ENGLISH", each with a small flag icon.

*Tempas: Temporal Archive Search Based on Tags.
in WWW '16: H. Holzmann, A. Anand.*

*On the Applicability of Delicious for Temporal Search on Web Archives.
in SIGIR '16: H. Holzmann, W. Nejdl, A. Anand.*

Searching Archives without a Full Text Index

Temporal
TimePortal

Search

2004/08/28 history

2002/11/22 history (closest)

2002/08/06 history

2001/09/23 history

2000/06/20 history

CDU

Homepage

Biographie

Wahlkreis

Wahl-ergebnisse

Ansichts-sache

Willkommen !

CDU

Liebe Internetfreunde,

als Ihre Bundestagsabgeordnete begrüße ich Sie ganz herzlich auf meiner Homepage.

Das Internet bietet eine ausgezeichnete Möglichkeit, sich über Themen meiner politischen Arbeit in Berlin und insbesondere für meinen Wahlkreis - Stralsund, Rügen, Grimmen - zu informieren.

Es ist vielen herkömmlichen Kommunikationswegen überlegen. So bietet es vor allem die einmalige Chance, nicht nur zu informieren, sondern auch unabhängig von Ort und Zeit miteinander zu kommunizieren.

Demokratie braucht Offenheit, denn nur sie schafft Vertrauen und nötige Kontrolle. Deshalb hat das Internet gerade in den letzten Monaten für die politische Arbeit deutlich an Bedeutung gewonnen.

Ich möchte Sie herzlich dazu einladen, sich über die Menüpunkte zum Beispiel über meine politischen Vorstellungen zu informieren. Auf Ihre Meinung bin ich sehr gespannt.

Weiterhin viel Freude beim Surfen und vielen Dank für Ihr Interesse!

Ihre Dr. Angela Merkel

Angela Merkel ...

... auf facebook

... auf MeinVZ

... auf CDU.TV

trum L3S Research Center

Problem 3: Knowledge Creation

From Search to Enrichment

Whats more impacting ?

10k casualties

\$4.5 Billions

Super cyclonic storm (IMD scale) Category 5 (Saffir–Simpson scale)	
	
05B making landfall	
Formed	October 25, 1999
Dissipated	November 3, 1999
Highest winds	<i>3-minute sustained:</i> 260 km/h (160 mph) <i>1-minute sustained:</i> 260 km/h (160 mph)
Lowest pressure	912 mbar (hPa); 26.93 inHg
Fatalities	~10,000 direct
Damage	\$4.5 billion (1999 USD)
Areas affected	India, Myanmar
Part of the 1999 North Indian Ocean cyclone season	

Category 5 major hurricane (SSHWS/NWS)	
	
on August 26, 2005	
Formed	August 23, 2005
Dissipated	August 31, 2005 ^[1] (Extratropical after August 30, 2005)
Highest winds	<i>1-minute sustained:</i> 175 mph (280 km/h)
Lowest pressure	902 mbar (hPa); 26.64 inHg
Fatalities	1,833 in U.S. confirmed ^[1]
Damage	\$108 billion (2005 USD)

1.8k casualties

\$108 Billions

Wikipedia Coverage

Hurricane Katrina

New Orleans

From Wikipedia, the free encyclopedia

Coordinates:  28°57'N 90°

Hurricane Katrina

See also: [Hurricane Katrina](#), [Effect of Hurricane Katrina on New Orleans](#) and [Drainage in New Orleans](#)

New Orleans was catastrophically affected by what the University of California Berkeley's Dr. Raymond B. Seed called "the worst engineering disaster in the world since Chernobyl," when the Federal levee system failed during Hurricane Katrina in 2005.^[44] By the time the hurricane approached the city at the end of August 2005, most residents had evacuated. As the hurricane passed through the Gulf Coast region, the city's federal flood protection system failed, resulting in the worst civil engineering disaster in American history.^[45] Floodwalls and levees constructed by the United States Army Corps of Engineers failed below design specifications and 80% of the city flooded. Tens of thousands of residents who had remained in the city were rescued or otherwise made their way to shelters of last resort at the Louisiana Superdome or the New Orleans Morial Convention Center. More than 1,500 people were recorded as having died in Louisiana, most in New Orleans, and others are still unaccounted for.^{[46][47]} Before Hurricane Katrina, the city called for the first mandatory evacuation in its history, to be followed by another mandatory evacuation three years later with Hurricane Gustav.

Hurricane Rita

Main article: [Hurricane Rita](#)



Hurricane Katrina at its New Orleans landfall.

No coverage for “Super Cyclone” in Wikipedia pages for Odisha, Berhampur, Ganjam

News Coverage

Odisha Super Cyclone

ON THIS DAY 1999
2005

29 October BBC NEWS

Search ON THIS DAY by date 29 October GO < >

Front Page | Years | Themes | Witness

1999: Super-cyclone wreaks havoc in India

A massive cyclone has swept through the state of Orissa in eastern India, killing an unknown number of people and leaving thousands more homeless.

The extent of the damage is difficult to determine. The area is almost impossible to reach, as the cyclone has torn down bridges and made roads and railways impassable.

All communications have been cut and the rescue effort is being hampered by the continuing bad weather.

Officials in the state capital, Bhubaneshwar, say nine deaths have been confirmed, but that number is expected to rise rapidly.

Many towns and villages have not been able to report casualty figures or damage assessments because telephone lines have been brought down.

CNN.com.asianow MAIN PAGE WORLD ASIA SOUTH ASIA INDIA INDONESIA TAIWAN CHINA ASIA PACIFIC ASIA BUSINESS SPORTS TECHNOLOGY SPACE HEALTH TRAVEL ASIA ASIA WEATHER TRAVEL ASIA S.E. ASIA POLITICS WEATHER BUSINESS SPORTS TECHNOLOGY SPACE HEALTH ENTERTAINMENT WORLD FORD TELEVISION NATURE WEATHER ANALYSIS ECONOMY Headline News Home

Thousands feared dead in India cyclone

October 29, 1999 Web posted at: 11:55 a.m. HKT (1855 GMT) CLOSE

In this story:
Much of Bhubaneshwar flooded
Hundreds of fishermen missing
No cyclone deaths
RELATED STORIES: INDEX

From staff and wire reports

BHUBANESHWAR, India (CNN) -- The death toll in a giant cyclone that tore across eastern India could run into thousands, the chief minister of the coastal state of Orissa said on Saturday.

"It could be thousands," Gindhar Gainang said in the state capital Bhubaneshwar. Sea water up to 1.5 meters (5 feet) deep has inundated the coast as far as 15 kilometers (3 miles) inland from the port of Paradip.

The New York Times

World

WORLD U.S. N.Y./REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

AFRICA AMERICAS ASIA PACIFIC EUROPE MIDDLE EAST

Where Death Got a Head Start, Indians Struggle

By BARRY BEARAK
Published: November 4, 1999

PARADIP, India, Nov. 3—Dead bodies, some washed up from the bay like driftwood, others bobbing like corks in the river, have become so plentiful in this port city that the authorities simply collect what they can and lay them out in tight bunches on the beach.

There are vague plans for a mass cremation, but today, with most of the corpses already five days old, only the dogs and crows and buzzards were busy with disposal.

Last Friday one of the mightiest cyclones in India's history chose Paradip as its first port of call. The storm would have been demon enough if it had merely flaunted its 160 miles-an-hour winds and then gone on its way. But the cyclone lingered here, punishing this small city of 80,000 people for 36 hours. The tidal surge -- three to five feet high -- lasted nearly a day.

To be in Paradip is to see the world turned upside down, with the bizarre and awful reality of it now flooding over its people the way the water did earlier.

FACEBOOK TWITTER GOOGLE+ EMAIL SHARE PRINT REPRINTS

Reasonable coverage in important news media but no Wikipedia coverage

Why is the coverage low ?

- Long tail entities
- Wikipedia editor bias
- Wikipedia recency bias
-

Lady Gaga

From Wikipedia, the free encyclopedia



Gaga briefly met with US president Barack Obama at a Human Rights Campaign fundraiser, he described the interaction as "intimidating" as she was dressed in 16-inch heels making her undoubtedly the tallest woman in the room.^[230]

Can we enrich and expand Wikipedia with high quality external resources like **News** ?

Citations in Wikipedia

X was born on August 4, 1961, [4]

- *Verifiability* one of the core principles in Wikipedia.
- Citations to external references as *evidence* for statements in Wikipedia articles
- *Reliability* and *authority* of sources for citations of Wikipedia statements
- Citations that point to *outdated* or *dead* URLs
- *Citation needed* for long-tail entities and newly added entities

Can we enrich Wikipedia by suggesting citations ?

Finding Citations for Wikipedia

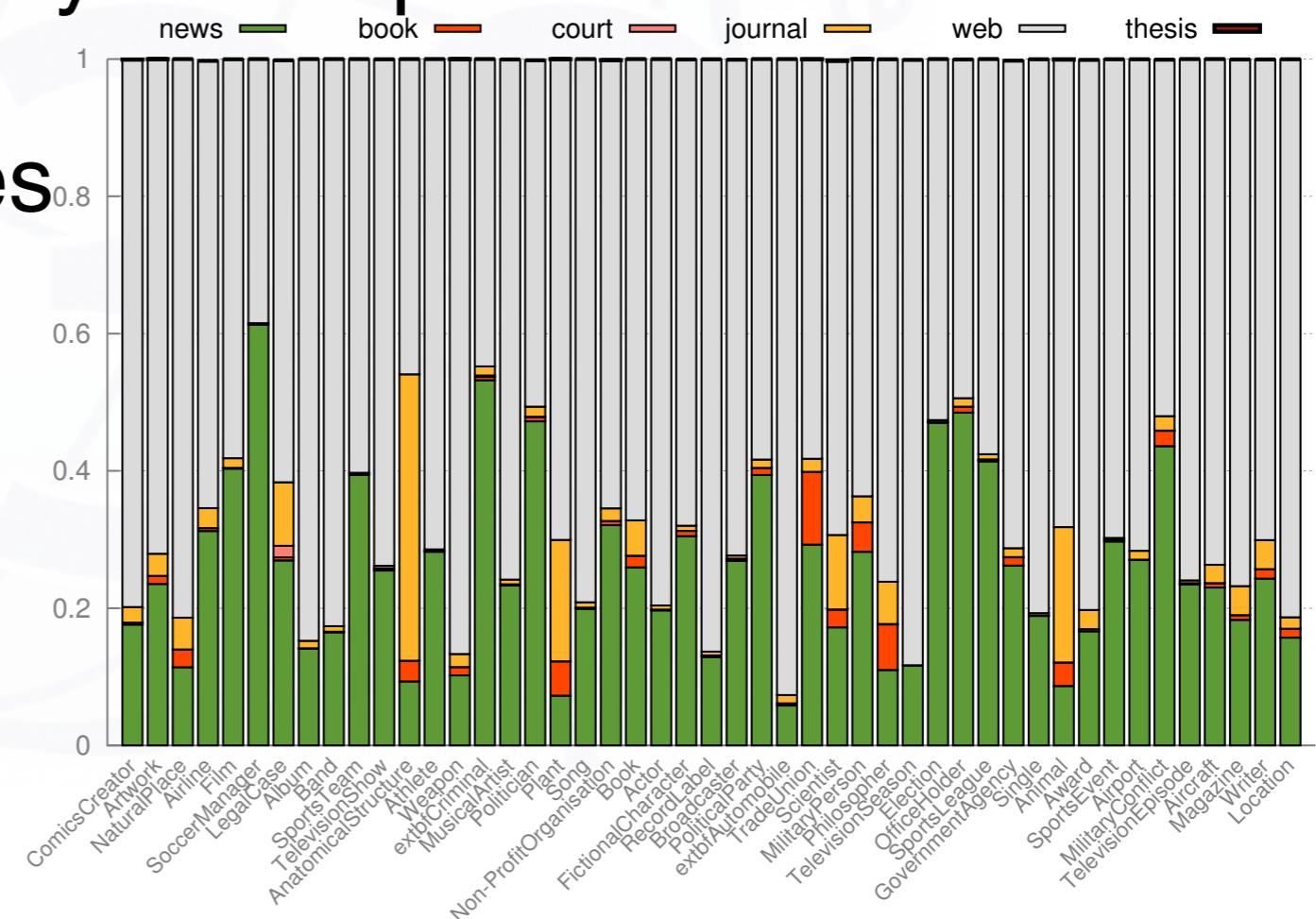
in CIKM '16: *Besnik Fetahu, Katja Markert, Wolfgang Nejdl, Avishek Anand.*

Fine-grained Citation Spans in Wikipedia

in EMNLP '17: *Besnik Fetahu, Katja Markert,, Avishek Anand.*

Why News ?

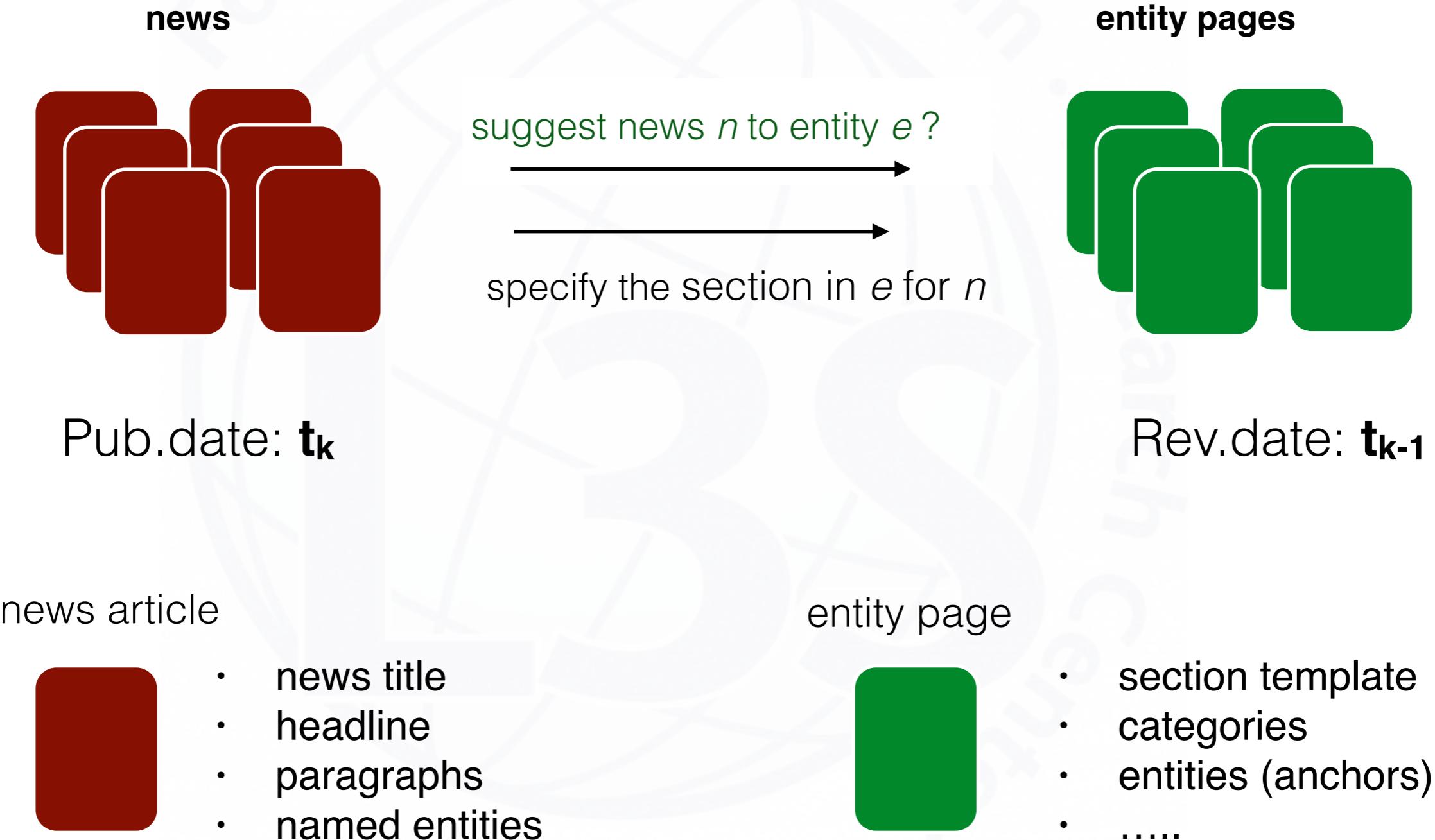
- Capture potentially important and impacting facts and events
- High news ref. density in Wikipedia
- Authoritative sources
- Editor control



How much is Wikipedia lagging behind news ?

in WebSci '15: Besnik Fetahu, Abhijit Anand, Avishek Anand.

Problem Statement



Approach: Two Tasks

feature extraction

THE WALL STREET JOURNAL. ≡ INDIA

4:49 pm IST
Oct 12, 2013

POLITICS

As It Happened: Cyclone Reaches Orissa

Some half a million people were evacuated from the southeastern Indian coast as Cyclone Phailin, a tropical storm from the Bay of Bengal, bore down on India. The states of Orissa and Andhra Pradesh, both of which have large coastal populations, were on high alert ahead of the storm's expected arrival.

news article

one classifier per entity type

Task 2

article section placement

[state]:geography

[city]:climate

...

article entity placement



Odisha

✗ Bay of Bengal ✗ Phailin

Odisha

Odisha (formerly known as Orissa)^[10] /oʊrɪsə/ [ɔɪrɪsə] is the west and north-west and Andhra Pradesh to the east. It was Mauryan Emperor Ashoka in 261 BCE who first recorded speakers.^[11] 1 April is therefore celebrated as Utkala of the state for over eight centuries until 13 April 1941.

Odisha is the 8th largest state by area in India, and the

Contents	[hide]
1	Eymology
2	History
3	Geography
4	Districts
5	Government and politics
6	Economy
6.1	Macro-economics
6.2	Industrial growth
6.3	Infrastructure development
7	Transportation
7.1	Air
7.2	Seaports
7.3	Railways
8	Demographics
9	Culture
10	Notable people
11	Literature
12	Dance
13	Music
4.4	Wikisource links

sections

wikipedia entity page

Task 1

Automated News Suggestions for Populating Wikipedia Entity Pages .
in CIKM '15: Besnik Fetahu, Katja Markert, Avishek Anand.

News Suggestion Attributes: Task#1

Novelty & Redundancy

Germanwings



Accidents and incidents recent

Germanwings Flight 9525 co-pilot deliberately crashed pl...

CNN - Mar 26, 2015

What questions do you have about the **crash**? Post on Twitter with the hashtag **#GermanwingsQs**. We'll take them to aviation experts and try to get them ...

Co-pilot 'wanted to destroy plane'

BBC News - Mar 26, 2015

Germanwings A320 crash: pilot locked out of cockpit

International - RFI - Mar 26, 2015

Germanwings crash: No algorithm can stop a pilot bent on killing ...

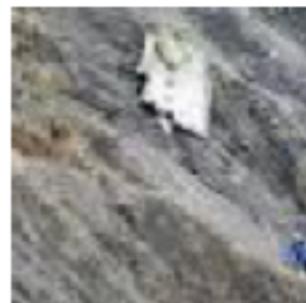
Opinion - Los Angeles Times - Mar 27, 2015

Germanwings Pilot Was Locked Out of Cockpit Before Crash in France

In-Depth - New York Times - Mar 25, 2015

Live updates: Germanwings plane crash

Blog - Washington Post (blog) - Mar 26, 2015



CNBC

BBC News

Los Angeles ...

News On 6

The Guardian

TIME

germ

IATA
4U ICAC
GWI

Founded

Operating base

Frequent-flyer program

Fleet size

Destinations

Parent company

Headquarters

Key people

Website

Explore in depth (10,979 more articles)

re

News Suggestion Attributes: Task#1

Entity Salience

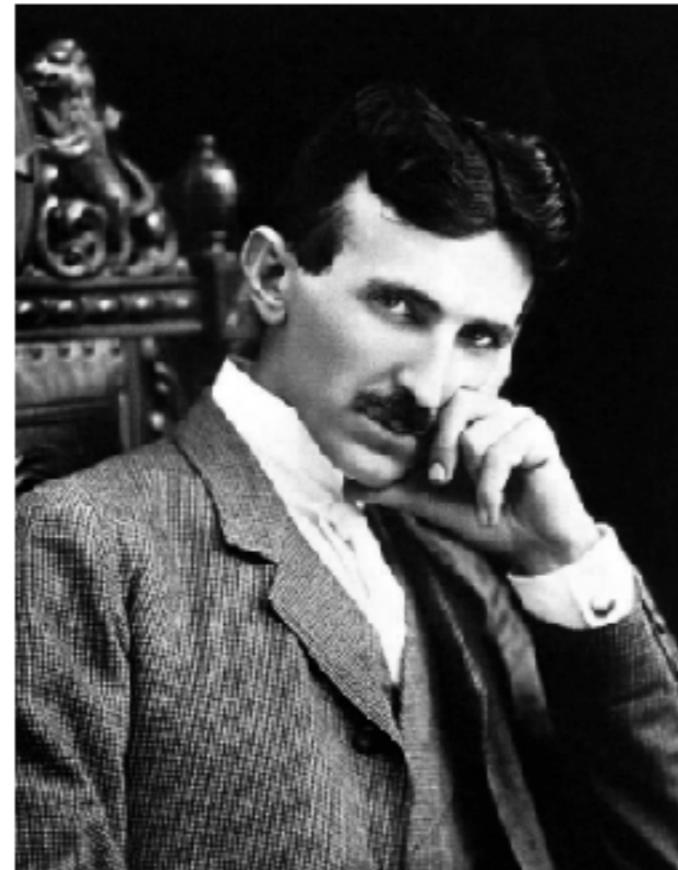
The inventor that inspired Elon Musk and Larry Page predicted smartphones nearly 100 years ago

CELENA CHONG | 1,591
Jul. 6, 2015, 5:43 PM

 FACEBOOK  LINKEDIN  TWITTER  PRINT

Nikola Tesla was a famous 20th century Serbian-American scientist who is most famous for designing the alternating current (AC) electricity supply system.

The world-famed engineer could speak in eight languages, had an eidetic memory and held 300 patents by the time he died. His name was borrowed for the red-hot car and energy storage company created by entrepreneur Elon Musk and he has been cited as an inspiration by Google co-founder Larry Page.



A It seems Tesla predicted the creation of the smartphone in a 1926 interview with John B. Kennedy.

- reward entity appearing throughout the text
- reward entity appearing in the top paragraph
- weigh an entity w.r.t its co-occurring entities

 **Nikola Tesla**
Elon Musk
Larry Page
John B. Kennedy

 **Tesla** is a central concept in the given news article

 **Elon Musk**

 **Larry Page**

 **John B. Kennedy**

$$\Phi(e, n) = \frac{|p(e, n)|}{|p(n)|} \sum_{p \in p(n)} \left(\frac{tf(e, p)}{\sum_{e' \neq e} tf(e', p)} \right)^{\frac{1}{p}}$$

Entity Salience: Relative Entity Frequency

News Suggestion Attributes: Task#1

Relative Entity Authority



Hillary Clinton
Broker for peace [edit]



Elias Taban



In 2012, United States Secretary of State Hillary Rodham Clinton met with Bishop Taban^[7] as part of a trip to urge Sudan and South Sudan to resolve an oil dispute that threatened to undermine existing peace agreements. Bishop Taban was a prominent public voice urging South Sudan to resume oil exports to the north and brokering peace between the two governments. The Bishop was also a strong advocate for the International Arms Trade Treaty,^[8] and traveled to New York for the treaty's final negotiations at the UN before its successful adoption. There, he was invited to address the full UN General Assembly on the need for an International Arms Trade Treaty.

In Sept. 2013, Bishop Taban was selected to receive the prestigious Global Citizen Award by the Clinton Foundation to be presented during the Clinton Global Initiative Annual Meeting in New York City.^[9] In honor of receiving the award, Bishop Taban sat down for an interview with David Brody of the Christian Broadcasting Network.^[10] Then, after the ceremony, the Bishop sat down with Secretary Clinton for an interview.^[11] He was honored for his vision and leadership in addressing global problems,^[12] which range in scope from the Arms Trade Treaty to access to clean water in Sudan to work with orphanages. He is praised as a respected broker for peace in both North and South Sudan, and the Clinton Foundation Global Citizens Award reflects that commitment.^[13]

opleseed

$$\hat{\Gamma}(e|\varphi(n)) = \frac{1}{|\varphi(n)|} \sum_{e' \in \varphi(n)} \mathbb{1}_{\Gamma(e') > \Gamma(e)}$$

Relative Entity Authority

Task#2: Section—template generation

Germanwings

Contents	[hide]
1 History	
1.1 Foundation	
1.2 Change of ownership	
1.3 Rebranding	
2 Corporate affairs	
3 Destinations	
4 Fleet	
4.1 Special liveries	
5 Services	
6 References	
7 External links	

Adria

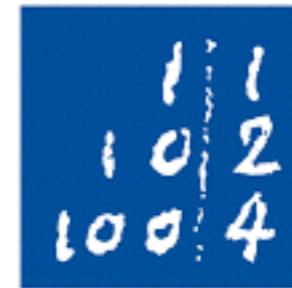
Contents	[hide]
1 History	
1.1 1960s: Early years	
1.2 1970s: Jetliner introduction	
1.3 1980s: Modernisation	
1.4 1990s: Becoming a scheduled airline	
1.5 2000s: Getting on the new markets	
1.6 2010s: Solving the crisis	
2 Corporate affairs and identity	
2.1 Headquarters	
2.2 Brand history	
2.3 Financial and operational results	
2.4 Subsidiaries	
2.4.1 Adria Airways Tehnika	
2.4.2 Adria Flight Center Center	
3 Destinations	
4 Codeshare agreements	
5 Fleet	
6 Incidents and accidents	
7 References	
8 External links	

Lufthansa

Contents	[hide]
1 History	
1.1 1950s: Post-war (re-)formation	
1.2 1960s: Jetliner introduction	
1.3 1970s–1980s: The wide-body era	
1.4 1990s–2000s: Further expansion	
1.5 2010s: Bell-lightening	
2 Corporate affairs and identity	
2.1 Headquarters	
2.2 Subsidiaries	
2.2.1 Airline subsidiaries	
2.2.2 Other operations	
2.3 Brand history	
2.4 Alliances and partnerships	
2.4.1 Commercial	
2.4.2 Technology	
2.5 Partner airlines	
2.6 Sponsorships	
6 Miles & More	
7 Lounges	
7.1 Overview and access	
7.2 First Class Terminal	
8 Accidents and incidents	
8.1 Fatal	
8.2 Non-fatal	
8.3 Hijackings	
9 Criticism	
9.1 Employment relations	
9.2 Germanwings' accident crisis management	
10 See also	
11 Citations	
11.1 Notes	
11.2 References	
12 External links	

- Section templates per entity type
- Cluster based on the X—means^[3] algorithm

[3] D. Pelleg, A. W. Moore, et al. X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML*, pages 727–734, 2000.



Leibniz
Universität
Hannover

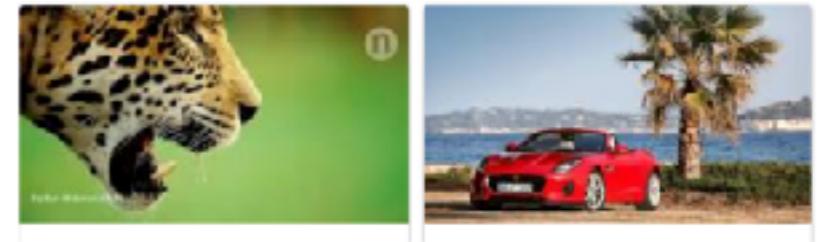
Interpretability in Search Systems

Search

Avishek Anand
anand@kbs.uni-hannover.de

The Challenge of Search

- Retrieval Models are functions that map **queries** to a sequence of documents **relevant for the query intent**
 - **Query Intent:** “*I want a politically correct way to catch a mouse.*”
 - **Queries :** 2.5 words long “mouse trap no poison”
 - **Documents:** Order of billions
 - Users typically do not go to **next page**
- **Ambiguity** in queries : “jaguar”, “clinton”,...
- **Modelling** of documents or sources:
 - Authenticity of documents, Centrality of topic..



A Brief History of Search

- Based on text features:
 - 80's-90's : TF-IDF, Probabilistic Models (BM25), Length Normalization
 - Closed form interpretable models
 - **No semantic information considered, Very heuristic**
 - State-of-the-art: **BM25**



Stolen Bicycles



Elmhurst police blotter: Bicycles stolen from Elmhurst College bike-sharing program



Theft, vandalism force Hong Kong bike sharing start-up Gobee to retreat from Europe

Asian bike sharing operators have embarked on overseas expansion over the past year as they burned through cash at home amid fierce competition

A brief History of Search

- Based on text features:
 - 00's - '10 : Language Models, Topic Models, Term Proximity Models
 - Have a closed form formula, Principled model
 - **Do not take into account user-based features**
 - **State-of-the-art:** Language Models w/ Topical Smoothing



Stolen Bicycles



Theft, vandalism force Hong Kong bike sharing start-up Gobee to retreat from Europe

Asian bike sharing operators have embarked on overseas expansion over the past year as they burned through cash at home amid fierce competition



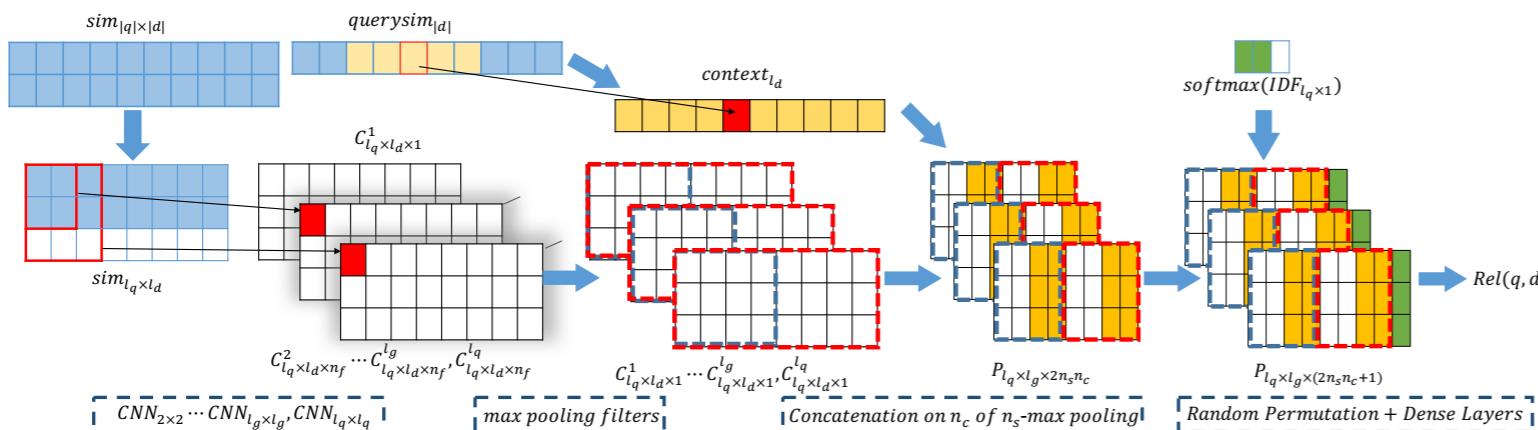
First Steps: What to Do When Your Bike Gets Stolen - WeLoveCycling ...

<https://www.welovecycling.com/.../first-steps-what-to-do-when-your-bike-gets-stolen/> ▾

Nov 20, 2015 - There are certain things you should do to higher your chances to see your bike again if it gets stolen. Let's hope you won't ever need this guide.

A brief History of Search

- Since 20 years there had not been much (< 10%) improvements over BM25
 - '07 - '13 : Learning to Rank — consider user behavior, query logs, clickthroughs
 - '13 - now : Neural Models — Deep Neural Models
 - **Higher Accuracy, but no closed form interpretation**



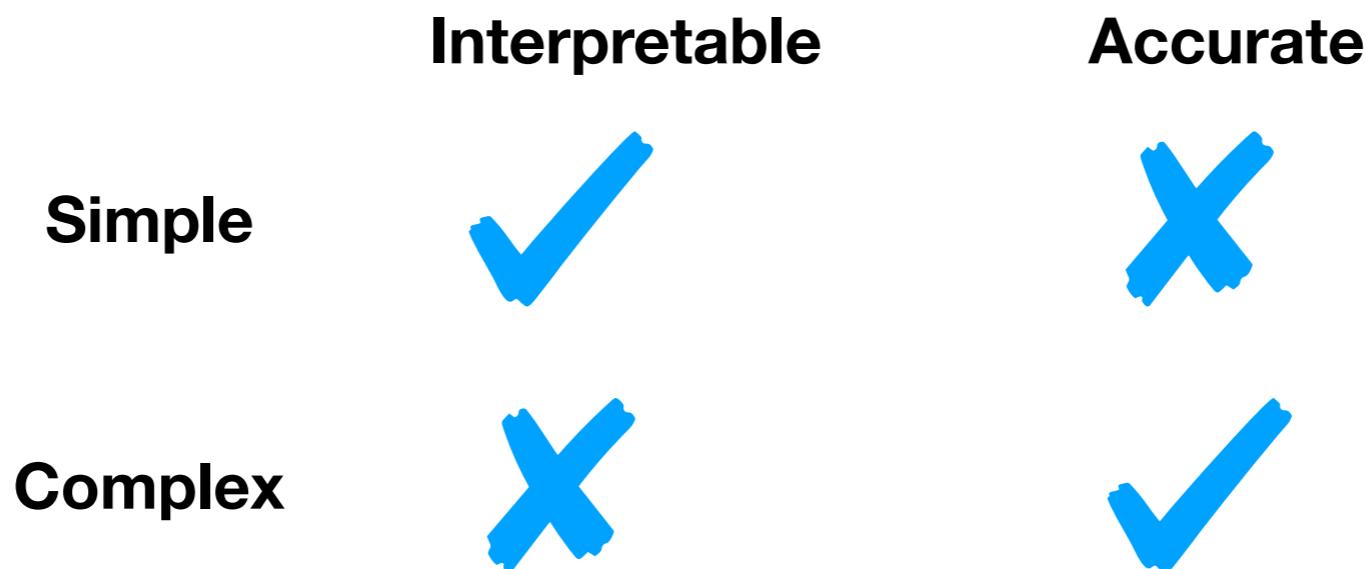
**Search bias,
Right for the wrong reasons**

Competition Commission imposes Rs 1.35 bn penalty on Google for search bias

The curse of Learning

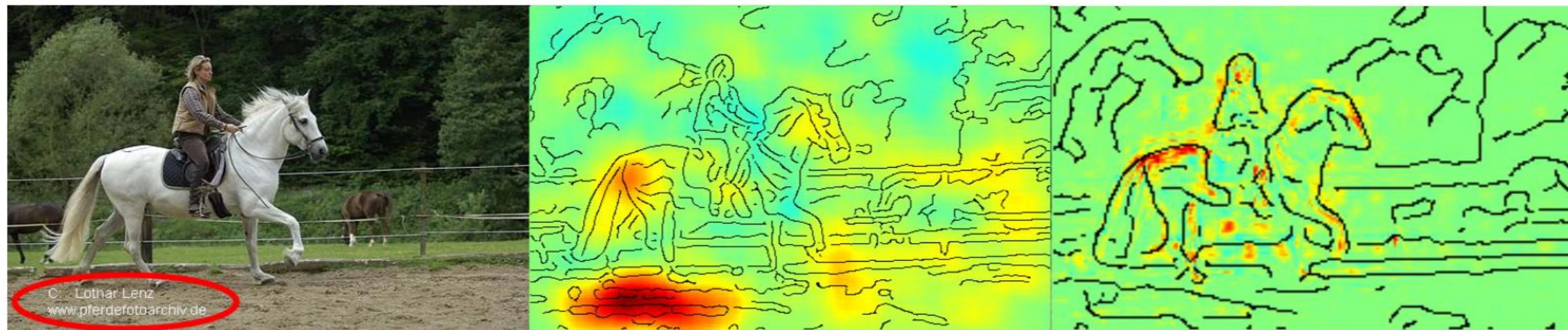
- Accurate models with more data typically entails
 - More features – trend towards automatically learned features
 - More complex features – non-linear spaces, complex aggregations

“Interpretability is the ability of a system to explain its reasoning”



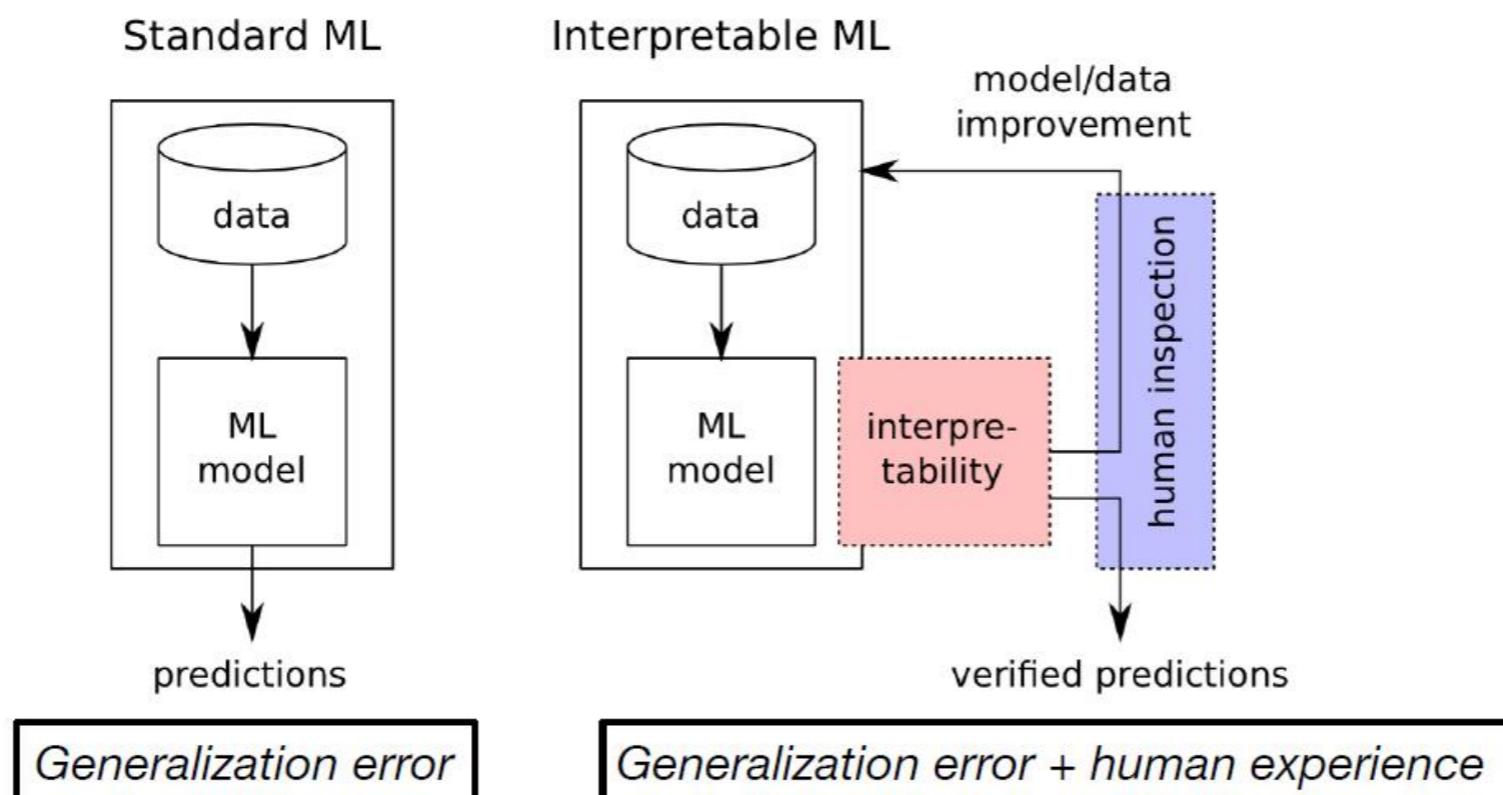
Why Interpretability

- **Right for the Right reasons:** a machine learning model is accurate and interpretable if
 - most of its predictions are correct and generalises well
 - and conforms to domain experts' knowledge about the problem



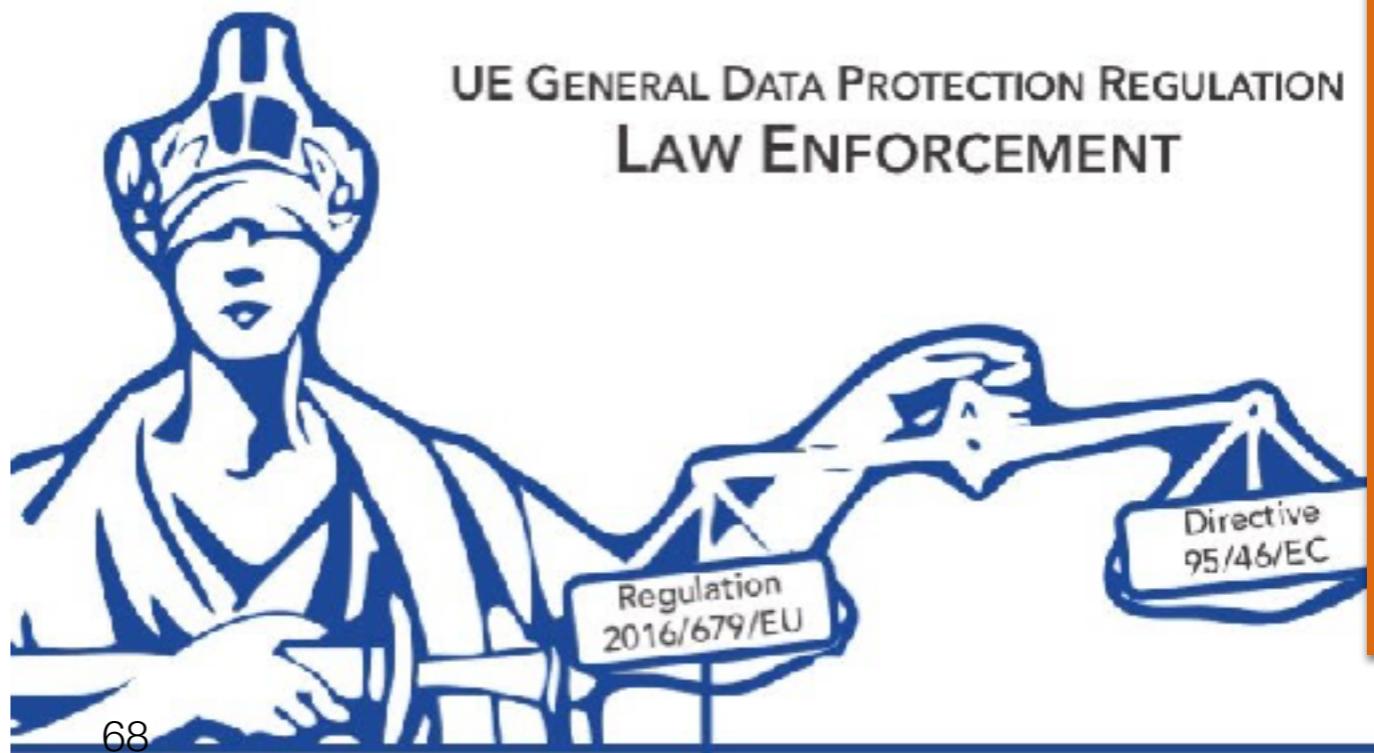
Why Interpretability

- **Insights into understanding and improving the model:**
 - Generalization error + human experience often results in better models



Why Interpretability

- **Compliance for legislation:** a machine learning model should explain itself
 - GDPR intends for “Right to Explanation”
 - Retain human decision to assign responsibility



Model Introspective Interpretability

- Given a trained model find the input/s that are most responsible for the output decision
- Approach: Sensitivity Analysis, Impact redistribution

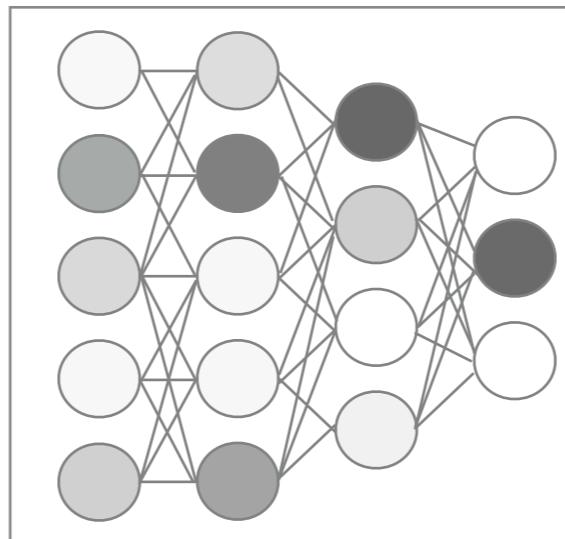
Input gradients +soc.religion.christian +alt.atheism

From: UST3012@ubulu.dpu.edu.kw
 Subject: Should teenagers ~~pick~~ church parents ~~don't~~ attend?
 Organization: U Texas Mail-to-News Gateway
 Lines: 10

Q. Should teenagers have the ~~freedom~~ to choose what ~~church~~ they go to?

My ~~friends~~ teenage kids do not like to go to church.
 If left up to them they ~~would~~ sleep, but that's not ~~an~~ option.
 They complain that they have no ~~friends~~ that go there, yet ~~don't~~
~~attempt~~ to make ~~friends~~. They ~~mention~~ not respecting their Sunday
 school teacher, and usually find a way to miss Sunday school but
 do make it to the ~~church~~ service, after their ~~parents~~ are thoroughly
 disgusted I might add. A never ending battle? It can just ruin your
 whole day if you let it.

Has ~~anyone~~ had this problem and how did it get resolved?
 f,



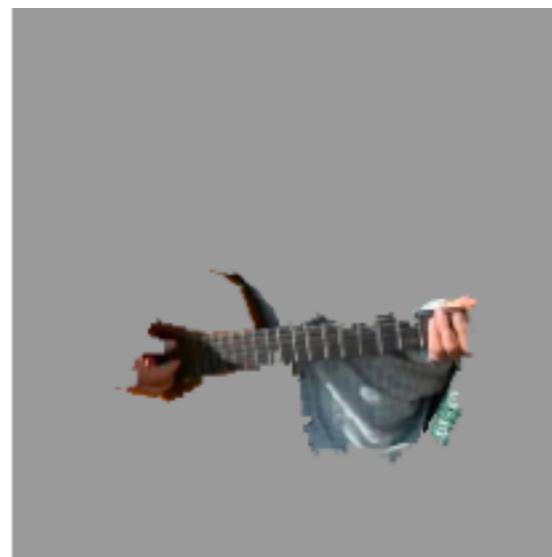
Sports
 Religion
 Politics

Model Agnostic Interpretability

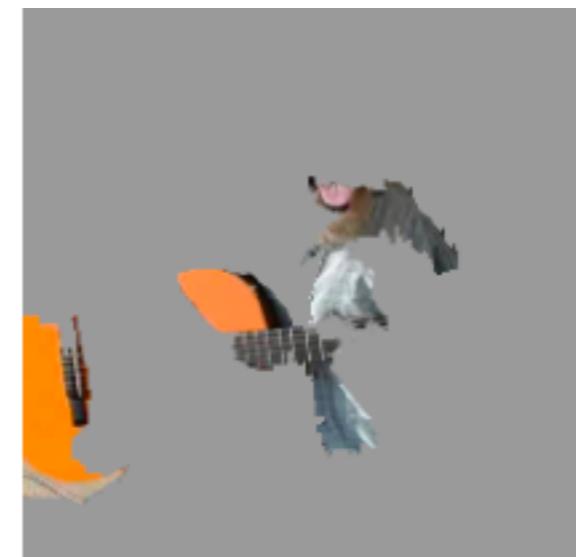
- Cannot peek into the model
- But can query the model infinite times
- **Approach:** Perturb the input to learn a locally interpretable model



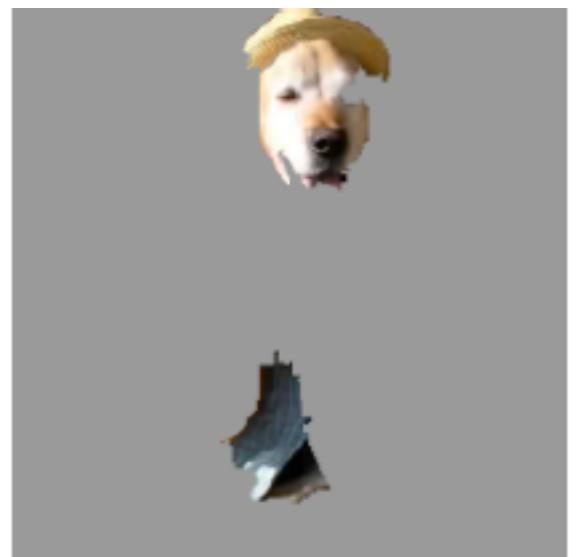
(a) Original Image



(b) Explaining *Electric guitar*



(c) Explaining *Acoustic guitar*



(d) Explaining *Labrador*

Interpretability aspects in search

- Why is a result **relevant** ?
 - Comparing rankers
- Which features contribute most ?
 - Summarizing important contributors to relevance
- Why is doc a **ranked above** doc b ?
- What **bias** does the ranker have ?
 - Semantic, syntactic, proximity-aware,..

Data protection | European Commission

<https://ec.europa.eu> › European Commission › Law › Law by topic ▾

Jan 9, 2018 - Data protection in the EU. The General Data Protection Regulation (GDPR), the Police Directive and other rules concerning the protection of personal data, as well as authorities tasked with enforcement. Data transfers outside the EU ...

What is the GDPR, its requirements and deadlines? | CSO Online

<https://www.csionline.com/>.../general-data-protection-regulation-gdpr-requirements-d... ▾

Feb 16, 2018 - GDPR is a regulation that requires businesses to protect the personal data and privacy of EU citizens for transactions that occur within EU member states. And non-compliance could cost ... or are a public authority. Some public entities such as law enforcement may be exempt from the DPO requirement.

[PDF] The EU General Data Protection Regulation - Allen & Overy

www.allenandover.com/.../Radical%20changes%20to%20European%20data%20protecti... ▾

Adoption of the GDPR marks a milestone in data protection laws in the EU. The Article 29 Working Party (WP29) which is composed of representatives of the EU national data protection authorities. (DPAs) has been working on guidance on various GDPR provisions to help organisations prepare. The first guidelines on data ...

General Data Protection Regulation (GDPR) – Final text neatly arranged

<https://gdpr-info.eu/> ▾

General Data Protection Regulation – Final legal text of the EU GDPR. The official PDF and its recitals as a neatly arranged website.

EU General Data Protection Regulation | DLA Piper Global Law Firm

<https://www.dlapiper.com/en/germany/focus/eu-data-protection-regulation/home/> ▾

EU data protection legislation is facing huge changes. Data protection laws are built on fundamental rights enshrined in the Charter of Fundamental Rights of the European Union which are the core building blocks of the EU's legal regime. Privacy issues arising from an exponential growth in consumer and mobile ...

GDPR: How Is It Different from U.S. Law & Why This Matters? - Lexology

<https://www.lexology.com/library/detail.aspx?g=4b2843f7-f67a-4015-bca9...> ▾

Sep 14, 2017 - This is part two of a multi-part look into the European Union's General Data Protection Regulation (GDPR) and why U.S. companies need to be aware of the law and how it may impact their business. We will conclude the series with a webinar on Jan. 17, 2018 that will review the series and provide further ...

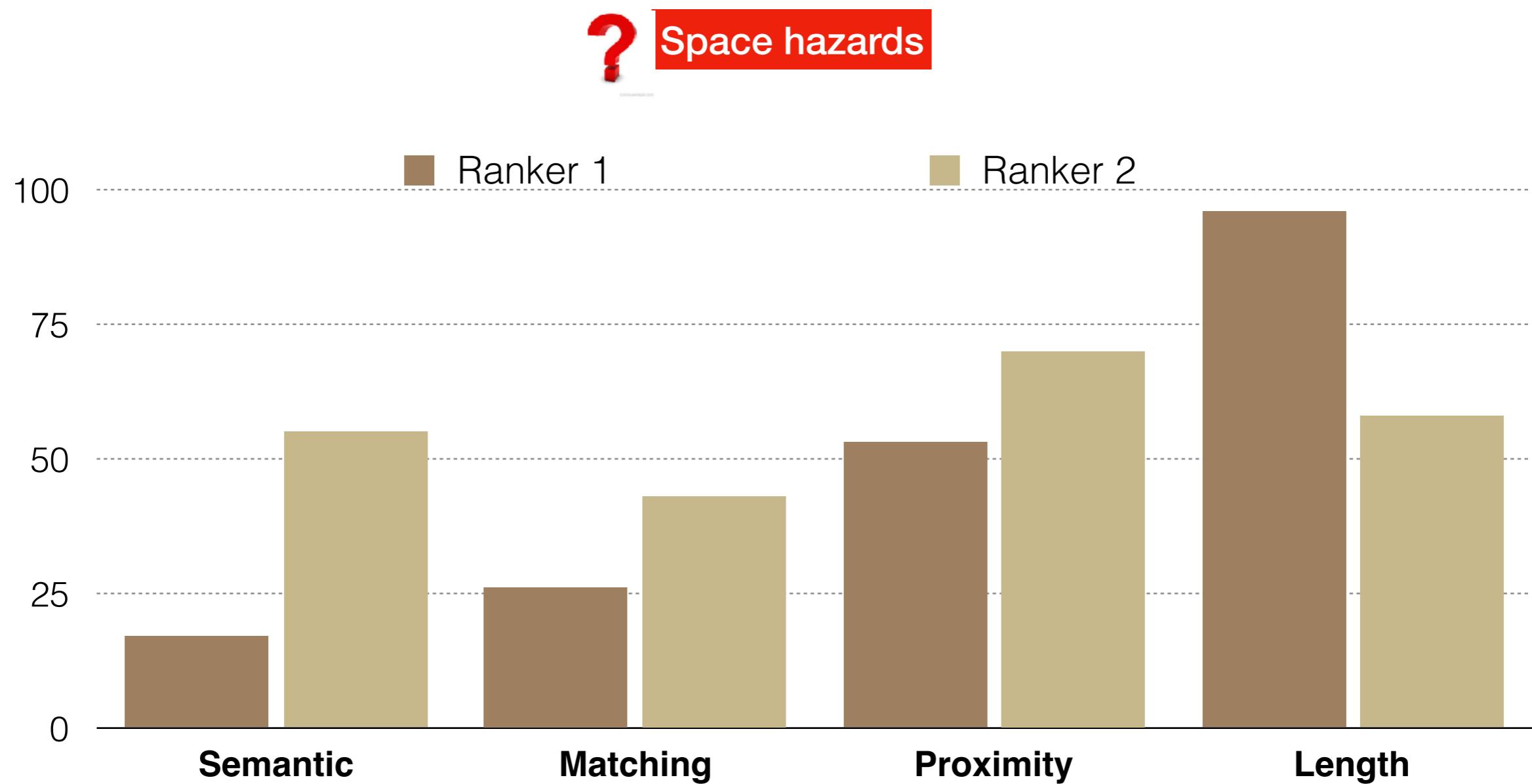
Why is a Result Relevant ?

- What words were considered important by the ranker for the result document ?

Space hazards	
	Ranker 1
sci.space (0.3)	<p>Yes, weightlessness does feel like falling. It may feel strange at first, but the body does adjust. The feeling is not too different from that of sky diving.</p> <p>>And what is the motion sickness >that some astronauts occasionally experience?</p> <p>It is the body's reaction to a strange environment. It appears to be induced partly to physical discomfort and part to mental distress. Some people are more prone to it than others, like some people are more prone to get sick on a roller coaster ride than others. The mental part is usually induced by a lack of clear indication of which way is up or down, ie: the Shuttle is normally oriented with its cargo bay pointed towards Earth, so the Earth (or ground) is "above" the head of the astronauts. About 50% of the astronauts experience some form of motion sickness, and NASA has done numerous tests in space to try to see how to keep the number of occurrences down.</p>
sci.med (-0.6)	<p>Yes, weightlessness does feel like falling. It may feel strange at first, but the body does adjust. The feeling is not too different from that of sky diving.</p> <p>>And what is the motion sickness >that some astronauts occasionally experience?</p> <p>It is the body's reaction to a strange environment. It appears to be induced partly to physical discomfort and part to mental distress. Some people are more prone to it than others, like some people are more prone to get sick on a roller coaster ride than others. The mental part is usually induced by a lack of clear indication of which way is up or down, ie: the Shuttle is normally oriented with its cargo bay pointed towards Earth, so the Earth (or ground) is "above" the head of the astronauts. About 50% of the astronauts experience some form of motion sickness, and NASA has done numerous tests in space to try to see how to keep the number of occurrences down.</p>

Why is a result Relevant ?

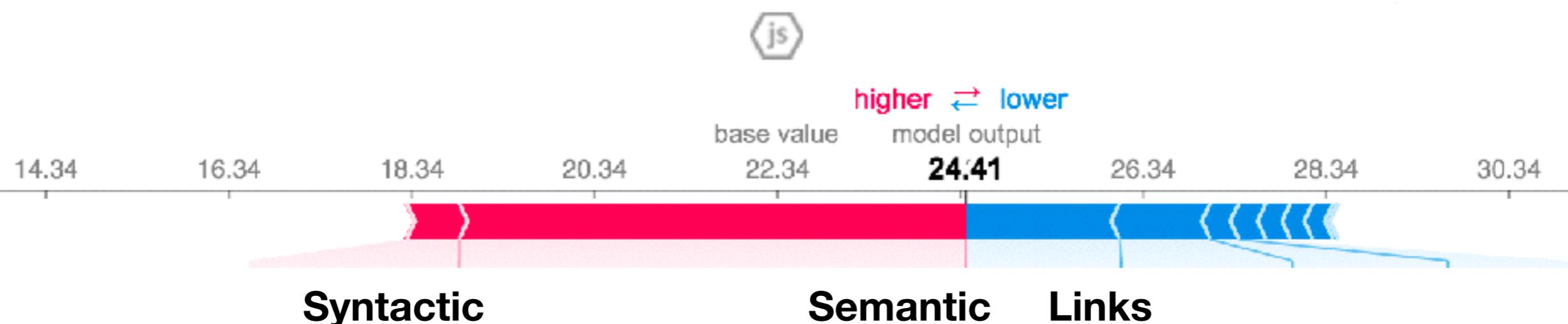
- Which IR features do the rankers adhere to ?



Which Feature contributes the most ?

Instance wise feature summarisation

- Learning to rank typically consists of 100s of features
- LinkedIn 2014 reports > 230 features
- Google rumoured to have > 800 features
- Can we summarise features most effective for a given query ?



Conclusion

- Long data is an interesting dataset along with novel research challenges
- Temporal indexing approaches exploit interval geometry for substantial query processing gains
- Retrieval models applicable for both recent and for historical intents
- News collections, Web archives contain uncovered facts in Wikipedia
- Need for interpretable primitives as diagnostic tools for IR researchers



Thanks Everyone...

Efficient Temporal Keyword Queries over Versioned Text, in CIKM '10
with S. Bedathur, K. Berberich, R. Schenkel

Temporal Index Sharding for Space-Time Efficiency in Archive Search, in SIGIR'11
with S. Bedathur, K. Berberich, R. Schenkel

Index Maintenance for Time-Travel Text Search, in SIGIR '12
with S. Bedathur, K. Berberich, R. Schenkel

History by Diversity: Helping Historians search News Archives. in CHIIR '16:
with J. Singh, W. Nejdl.

Modeling Event Importance for Ranking Daily News Events. in WSDM '17:
with A. Anand, V. Setty, A. Mishra

Automated News Suggestions for Populating Wikipedia Entity Pages in CIKM '15:
with B. Fetahu, K. Markert.

Finding Citations for Wikipedia in CIKM '16:
with B. Fetahu, K. Markert, W. Nejdl

Access

Search

Enrichment