Temporal Information Retrieval

Avisheek Anand
anand@kbs.uni-hannover.de
Web science @ L3S

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

- Selected projects
  - Retrieval, Exploration and Analytics for Web Archives (ERC Advanced Grant)
    - Real-time data processing for finance predictions
    - Cross-media analysis and interpretation
  - CUbRIK: Searching by computers and humans
  - ForgetIT: Concise Preservation via Managed Forgetting

Avishek Anand
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

Graphs

Knowledge Units

Angela Merkel
Chancellor of Germany

Profiles
Facebook Instagram

Quote:
The willingness to learn new skills is very high. Whoever decides to define their life politics knows that earning money isn’t the priority.

The text is our common goal, and Europe is our common future.
Where is Temporal Data?

**News Collections**

- The New York Times
- GDELT 2.0
- event repositories, > 20 years, 200k articles a day

**Encyclopaedic Collections**

- Wikipedia
- Web Archives
- > 15 PB, 150 Bi Captures
Web Archives as a Dataset

1998
Web Archives as a Dataset

2008
Web Archives as a Dataset
Web Archives as a Dataset
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.

Avishek 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]
Problem 1: 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...

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. 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 been

classified with the noncommunist world.
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
  
  \( nato \text{ intervention} @ 2001 \text{ - } 2004 \)

• We want to retrieve documents containing terms “nato”, “intervention” and valid between \( 2001 \text{ - } 2004 \)

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**
Index Sharding

Trade-off between costs of sequential and random reads

more random accesses  more wasted reads

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

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

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

Queries: AOL
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.
How do we automatically rank a daily batch of news?
News Ranking - Classical

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

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

Belongs to long running "Iraqi civil war"

Search

Historical Importance!

Not popular today!
Characterising Event History

Short Term Events highly popular

One Off Events

Long Running Events
Exploiting Historical Cues

Clusters after false-positive removal

Short-term Context

Cluster chaining

Long-term Context

Temporal profile from canonical topic assignment using Wikipedia

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

Avishek Anand
Ranking Events

**Current-day features**
- effective cluster size
- diversity features
- authority features

**Historical features**
- Temporal profile from Canonical topic assignment
- Cluster chaining

Event ranking using Pairwise Learning to rank (SVM Rank)
Experiments

<table>
<thead>
<tr>
<th>STICS Data</th>
<th>P@10</th>
<th>MAP@10</th>
<th>F1@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM-w/o Hist</td>
<td>0.421</td>
<td>0.404</td>
<td>0.233</td>
</tr>
<tr>
<td>LM-Hist</td>
<td>0.507</td>
<td>0.499</td>
<td>0.281</td>
</tr>
<tr>
<td>Event-Rank-w/o</td>
<td>0.582</td>
<td>0.608</td>
<td>0.437</td>
</tr>
<tr>
<td>Event-Rank- Hist</td>
<td>0.882</td>
<td>0.885</td>
<td>0.660</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gdelt Data</th>
<th>P@10</th>
<th>MAP@10</th>
<th>F1@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM-w/o Hist</td>
<td>0.393</td>
<td>0.465</td>
<td>0.275</td>
</tr>
<tr>
<td>LM-Hist</td>
<td>0.405</td>
<td>0.480</td>
<td>0.283</td>
</tr>
<tr>
<td>Event-Rank-w/o</td>
<td>0.578</td>
<td>0.690</td>
<td>0.517</td>
</tr>
<tr>
<td>Event-Rank- Hist</td>
<td>0.675</td>
<td>0.750</td>
<td>0.588</td>
</tr>
</tbody>
</table>

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

Search Intent

Ranking
Textual relevance
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

Rudy Giuliani (sort of) endorses Donald Trump

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

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.

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."

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

Canonicalisation of entities
Coverage Problem

Aspects are diverse across time

Time periods are aspect-diverse
• 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.
Experiments

- Main Metric: Subtopic Recall
- Other Metrics: $\alpha$-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

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>0.428</td>
<td>0.471</td>
<td>0.509</td>
</tr>
<tr>
<td>OnlyTime</td>
<td>0.426</td>
<td>0.497</td>
<td>0.534</td>
</tr>
<tr>
<td>IA-Select</td>
<td>0.442</td>
<td>0.491</td>
<td>0.516</td>
</tr>
<tr>
<td>MDIV</td>
<td>0.460</td>
<td>0.497</td>
<td>0.534</td>
</tr>
<tr>
<td>HistDiv</td>
<td>0.497</td>
<td>0.542</td>
<td>0.583</td>
</tr>
<tr>
<td>HistDiv-Burst</td>
<td><strong>0.509</strong></td>
<td><strong>0.560</strong></td>
<td><strong>0.601</strong></td>
</tr>
</tbody>
</table>
Highlights & Impact


• Hannover Impuls Ideas competition — runner up

Demo: bit.ly/archive-search

Avishek Anand
Searching Archives without a Full Text Index

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.

Temporal Archive Search Based on Tags.
http://tempas.l3s.de
Willkommen!

Lieber 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
Problem 3: Knowledge Creation

From Search to Enrichment
What's more impacting?

Super cyclonic storm (IMD scale)
Category 5 (Saffir–Simpson scale)

- 10k casualties
- $4.5 Billion

Category 5 major hurricane (SSHWS/NWS)

- 1.8k casualties
- $108 Billion
Wikipedia Coverage
Hurricane Katrina

New Orleans
From Wikipedia, the free encyclopedia

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 Chernoby,” when the Federal levee system failed during Hurricane Katrina in 2005. 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. 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. 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

No coverage for “Super Cyclone” in Wikipedia pages for Odisha, Berhampur, Ganjam
Reasonable coverage in important news media but no Wikipedia coverage
Why is the coverage low?

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

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

news

Pub. date: $t_k$

news article

• news title
• headline
• paragraphs
• named entities

suggest news $n$ to entity $e$?

specify the section in $e$ for $n$

to entity $e$

Rev. date: $t_{k-1}$

entity page

entity pages

• section template
• categories
• entities (anchors)
• …..
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.

---

**Task 1**

Automated News Suggestions for Populating Wikipedia Entity Pages.

in CIKM ’15: Besnik Fetahu, Katja Markert, Avishek Anand.

---

**Task 2**

article section placement

[state]:geography
[city]:climate

---

entities

news article

one classifier per entity type

THE WALL STREET JOURNAL.

As It Happened: Cyclone Reaches Orissa

Odisha

entity page

Bay of Bengal

Phailin

sections

wikipedia

Task 2

Task 1
News Suggestion Attributes: Task#1

Novelty & Redundancy

Novelty and Redundancy Measure

• novelty is measured w.r.t previously added news articles
• major events have wide coverage in news media
• place the news article into the correct section
News Suggestion Attributes: Task#1

Entity Salience

- reward entity appearing throughout the text
- reward entity appearing in the top paragraphs
- 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

Entity Salience: Relative Entity Frequency

$$\Phi(e, n) = \frac{|p(e, n)|}{|p(n)|} \sum_{p \in p(n)} \left( \frac{tf(e, p)}{\sum_{e' \neq e} tf(e', p)} \right)^{1/p}$$
News Suggestion Attributes: Task#1
Relative Entity Authority

- entities with `low authority’ have lower entry barrier
- a news article in which an entity co-occurs with `high authority’ entities conveys news the importance

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

Relative Entity Authority

Avishek Anand
Task#2: Section—template generation

Germanwings

- History
  - 1.1 1960s: Early years
  - 1.2 1970s: Jetliner introduction
  - 1.3 1960s: Modernisation
  - 1.4 1990s: Resuming a scheduled airline
  - 1.5 2000s: Getting on the new markets
  - 1.6 2010s: Solving the crisis
- Corporate affairs and identity
  - 2.1 Headquarters
  - 2.2 Brand history
  - 2.3 Financial and operational results
  - 2.4 Operations
    - 2.4.1 Adria Airways Technika
    - 2.4.2 Adria Flight Center
- Destinations
- Codeshare agreements
- Fleet
- Incidents and accidents
- Reference
- External links

Adria

- History
  - 1.1 1960s: Post-war reformation
  - 1.2 1970s: Jetliner introduction
  - 1.3 1980s: The wide-body era
  - 1.4 1990s: 2000s: Further expansion
  - 1.5 2010s: Fleet Lightning
- 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
- Miles & More
- Lounges
  - 7.1 Overview and access
  - 7.2 First Class Terminal
- Accidents and incidents
  - 8.1 Fatal
  - 8.2 Non-fatal
  - 8.3 Hijackings
- Criticism
  - 9.1 Employment relations
  - 9.2 Germanwings' accident crisis management
- See also
- Citations
  - 11.1 Notes
  - 11.2 References
- External links

Lufthansa

- History
- Corporate affairs and identity
- Destinations
- Codeshare agreements
- Fleet
- Incidents and accidents
- Reference
- External links

- Section templates per entity type
- Cluster based on the X-means algorithm

---

Interpretability in Search Systems

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

By Veena Mani, Business Standard, New Delhi | In Digital | February 09, 2018
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
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
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,
Why is a Result Relevant?

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

**Ranker 1**

Space hazards

**Ranker 2**

Interpretability
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