Advanced Topics in Information Retrieval

Temporal Information Retrieval

Vinay Setty  Jannik Strötgen
vsetty@mpi-inf.mpg.de  jannik.stroetgen@mpi-inf.mpg.de

ATIR – June 23, 2016
Why **temporal information retrieval?**
Time in queries

Temporal information needs are frequent

query log analyses

- 1.5% queries with **explicit temporal intent** [Nunes et al. 2008]
- 7% queries with **implicit temporal intent** [Metzler et al. 2009]
- 13.8% **explicit**, 17.1% **implicit** [Zhang et al. 2010]

Different types of temporal information in IR

- time as **dimension of relevance**
- time as **query topic**
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Temporal information extraction

Temporal information is frequent

News articles.
Narrative documents.
Biographies.

temporal information can be normalized

same semantics → same value

“heute”, “aujourd’hui”, “today”, “June 8, 2016” → 2016-06-08
So far addressed: temporal tagging

**Addressed types of temporal expressions**

TimeML standard:  
[Pustejovsky et al.  2005]

- **dates** ("May 1, 2015", "today")
- **times** ("9 pm", "last night")
- **durations** ("three years")
- **set expressions** ("twice a day")

Dates and times may be:
- **explicit** ("May 1, 2015")
- **implicit** ("Christmas 2012")
- **relative** ("last night")
- **underspecified** ("Monday")

Normalization difficulty varies between types, but:

all are obvious temporal expressions
So far ignored: free-text temporal expressions

Idea
standard text phrases may be associated with temporal scopes
So far ignored: free-text temporal expressions

Idea

standard text phrases may be associated with temporal scopes

- **temponyms** [Kuzey et al. 2016a, 2016b]
  refer to arbitrary kinds of **named events or facts with temporal scopes** that are merely given by a text phrase but have unique interpretations **given the context and background knowledge**.

- **temponym tagging**
  is the **detection** and **normalization** of temponyms.

Goal

further temporal enrichment of documents
Examples of temponyms

John F. Kennedy’s death marked a watershed in the memories of almost every American.

President Obama awarded the nation’s highest military honor to a Union soldier who was killed more than 150 years ago during the Battle of Gettysburg.
Examples of temponyms

John F. Kennedy’s death marked a watershed in the memories of almost every American.

President Obama awarded the nation’s highest military honor to a Union soldier who was killed more than 150 years ago during the Battle of Gettysburg.

normalized temporal information
(temporal tagging)

publication date: 2014-11-06

1864
Examples of temponyms

John F. Kennedy’s death marked a watershed in the memories of almost every American.

President Obama awarded the nation’s highest military honor to a Union soldier who was killed more than 150 years ago during the Battle of Gettysburg.

normalized temporal information (temporal tagging, temponym tagging)

1963-11-22

1864
[1863-07-01, 1863-07-03]
Examples of temponyms

phrases with temponyms

The Cuban Revolutionary War
The second inauguration of Bill Clinton
2008 Mexico City plane crash
2016 WWW Conference
Examples of temonyms

**temporal tagging**

- The Cuban Revolutionary War
- The second inauguration of Bill Clinton
- 2008 Mexico City plane crash
- 2016 WWW Conference
Examples of temponyms

temporal tagging vs. temponym tagging

— vs. [1953-07-26,1959-01-01]
The Cuban Revolutionary War

— vs. 1997-01-20
The second inauguration of Bill Clinton

2008 vs. 2008-11-04
2008 Mexico City plane crash

2016 vs. [2016-04-11,2016-04-15]
2016 WWW Conference

temponyms add new or more precise temporal information
WWW’16 paper [Kuzey et al. 2016a]

- all temponyms, not only explicit ones, e.g., “during his presidency”
- often year-level temporal scopes

As Time Goes By: Comprehensive Tagging of Textual Phrases with Temporal Scopes

TempWeb’16 Approach [Kuzey et al. 2016b]

- explicit temponyms, day-level temporal scopes
- completely other approach than WWW’16 approach: temponym tagging with HeidelTime

Temponym Tagging: Temporal Scopes for Textual Phrases
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
How dynamic is the Web?


Data

- weekly crawls of **154 web sites** over one year
- **top-ranked web sites** from topical categories in Google Directory (extension of DMOZ) from different top-level domains
- at most **200K web pages per web site** per weekly crawl

<table>
<thead>
<tr>
<th>Domain</th>
<th>Fraction of pages in domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>41%</td>
</tr>
<tr>
<td>.gov</td>
<td>18.7%</td>
</tr>
<tr>
<td>.edu</td>
<td>16.5%</td>
</tr>
<tr>
<td>.org</td>
<td>15.7%</td>
</tr>
<tr>
<td>.net</td>
<td>4.1%</td>
</tr>
<tr>
<td>.mil</td>
<td>2.9%</td>
</tr>
<tr>
<td>misc</td>
<td>1.1%</td>
</tr>
</tbody>
</table>
How dynamic is the Web?

Web pages

- on average **8% new web pages** per week
- **peek** in creation of new pages at the **end of each month**
- after **9 months** about **50% of web pages** have been deleted

<table>
<thead>
<tr>
<th>Week</th>
<th>Fraction of Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
</tr>
<tr>
<td>15</td>
<td>0.4</td>
</tr>
<tr>
<td>20</td>
<td>0.2</td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

- week 1
  - 4.8 M pages
- week 45
  - one crawler crashed
- work from 2004!
How dynamic is the Web?

content

- based on w-shingles (contiguous sequence of w words)
- **after one year** more than **50% of shingles** are still available
- each week about **5% of new shingles** are created

![Graph showing fraction of shingles over weeks]

shingle size

- $w = 50$
- week 1
- 4.3 B unique shingles
How dynamic is the Web?

**links**

- after one year **only 24% of links** are still available
- on average **25% of new links** are created every week

![Fraction of Links](image)

**red:**
- first-week links

**blue**
- new links from 1st week pages

**white**
- new links from “new” pages
Dynamics and age

the Web is **highly dynamic**

- new content is continuously added
- old content is deleted and potentially lost forever

Web archives

- e.g., archive.org, internetmemory.org
- have been preserving **old snapshots** of web pages since 1996

improved digitalization

- e.g., using OCR (optical character recognition)
- have allowed (newspaper) archives to make old documents (e.g., from 1700s) searchable
Dynamics and age

several challenges

- How to index highly redundant document collections like web archives?
- How to make use of temporal information such as publication dates?
- How to search documents written in archaic language?
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Indexing Redundant Content

Zhang & Suel (2007): approach to index **highly-redundant document collections** (e.g., web archives)

main ideas:
- **break up documents** into shorter segments
- **segments should be shared** between overlapping documents
- use a **two-level index structure** to index associations between **words-and-segments** and **segments-and-documents**

\[
\begin{align*}
&d_1 \begin{array}{c} 
\text{aac} \\
\text{bab} \\
\text{ccb}
\end{array} \rightarrow s_1 \begin{array}{c} 
\text{aac} \\
\text{bab} \\
\text{ccb}
\end{array},
&s_2 \begin{array}{c} 
\text{bab} \\
\text{ccc}
\end{array},
&s_3 \begin{array}{c} 
\text{ccc}
\end{array}
\end{align*}
\]

\[
\begin{align*}
&a \rightarrow s_1, s_2, s_4, s_7, \ldots
&s_1 \rightarrow d_1, d_3, d_9, \ldots
\end{align*}
\]
Indexing Redundant Content

**Problem**

how to break up documents into smaller segments so that segments are **shared between overlapping documents**

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aac</td>
<td>acb</td>
</tr>
<tr>
<td>bab</td>
<td>abc</td>
</tr>
<tr>
<td>ccb</td>
<td>cba</td>
</tr>
</tbody>
</table>

**Naïve approach**

<table>
<thead>
<tr>
<th>aac</th>
<th>bab</th>
<th>ccb</th>
</tr>
</thead>
</table>

**Better approach**

<table>
<thead>
<tr>
<th>a</th>
<th>acb</th>
<th>abc</th>
<th>cb</th>
</tr>
</thead>
<tbody>
<tr>
<td>acb</td>
<td>abc</td>
<td>cb</td>
<td>a</td>
</tr>
</tbody>
</table>
Indexing Redundant Content

hash breaking (naïve approach)
- compute **hash code** $h[i]$ for each term $d[i]$ in document
- break document **at all indices** $i$ such that $h[i] \% w = 0$

Winnowing (as a better approach with guarantees)
- compute **hash code** $h[i]$ for all subsequences $d[i \ldots i+b-1]$ of length $b$
- **slide window** of size $w$ over the array of hash codes $h[0..|d|-b]$
  - if $h[i]$ is **strictly smaller** than all other values $h[j]$ in current window, **cut the document** between $i$ and $i-1$
  - if **multiple positions** $i$ in the current window with minimal value $h[i]$
    - if we have previously cut directly before one of them, don’t perform a cut
    - otherwise, cut before the rightmost position $i$ having minimal value $h[i]$
Indexing Redundant Content

query processing

- needs to be adapted to reflect that the same segment can occur in many documents
- when seeing a segment in a posting list of the first index, look up documents containing it in the second index
- effectiveness of skipping for conjunctive queries is reduced
  - terms could be spread over different segments in a document
  - segments can be contained in documents with arbitrary document identifiers
Time travel text search

text search on **version document collections**

**time-travel keyword query q@t**
- combines keywords q with a time of interest t to search “as of” the indicated time in the past

**time-travel inverted index**
- adds a **valid-time interval** $[t_b, t_e)$ to postings indicating when the information therein was current

![Diagram of a dictionary with postings list](https://example.com/diagram.png)
Time travel text search

**Dictionary**

```
a       g       z
```

**Posting list**

- `d_{123}, 2, [1, 4]`
- `d_{123}, 2, [4, 6]`
- `d_{125}, 2, [4, 8]`

**time-travel keyword query q@t**

- **read posting lists** for keywords in q
- **filter out postings** whose valid-time interval does not contain t, i.e.: \( t \notin [t_b, t_e] \)
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Time in queries

temporal information needs are frequent

query log analyses
- 1.5% queries with **explicit temporal intent** [Nunes et al. 2008]
- 7% queries with **implicit temporal intent** [Metzler et al. 2009]
- 13.8% **explicit**, 17.1% **implicit** [Zhang et al. 2010]

different types of temporal information in IR
- time as **dimension of relevance**
- time as **query topic**
Time in documents

Documents come with different kinds of **temporal information**
- **publication dates** (DCT): when document was published
- **temporal expressions**: time periods the document talks about

what is helpful depends on how time is used
Time as dimension of relevance

- temporal tagging is not needed
- **document creation time** and **query time** are utilized
- examples: news-related queries, freshness of search results
- besides improving search results:
  - query time for time-sensitive query auto-completion

[Sengstock & Gertz 2011]
Time as dimension of relevance

- temporal tagging is not needed
- **document creation time** and **query time** are utilized
- examples: news-related queries, freshness of search results
- besides improving search results:
  - query time for time-sensitive query auto-completion

[Sengstock & Gertz 2011]
Time as dimension of relevance

- temporal tagging is not needed
- **document creation time** and **query time** are utilized
- examples: news-related queries, freshness of search results
- besides improving search results:
  - query time for time-sensitive query auto-completion

[Sengstock & Gertz 2011]
Time as dimension of relevance

- temporal tagging is not needed
- **document creation time** and **query time** are utilized
- examples: news-related queries, freshness of search results
- besides improving search results:
  - query time for time-sensitive query auto-completion
    - [Sengstock & Gertz 2011]

**Suggestions** for query “work...” at **different times**:

- 6am: workwear
- 3pm: workforce
- 9pm: workout
Time as query topic

- temporal tagging is required
- temporal information in the content
- document creation time is not meaningful
- example: queries with explicit time expressions

\[ q_{text} = \langle \text{Germanwings} \rangle, \quad q_{time} = [2015-03-01, 2015-04-30] \]

<table>
<thead>
<tr>
<th>March 25, 2015</th>
<th>November 10, 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germanwings plane crash: Leaders visit Alps site</td>
<td>Lufthansa tries to force striking staff back to work</td>
</tr>
<tr>
<td>The German, French and Spanish leaders have arrived together in the French Alps to visit the scene where a Germanwings plane crashed on Tuesday, killing all 150 on board.</td>
<td>The carrier confirmed Tuesday it had applied for a German court order... Lufthansa is still recovering from the blow that it suffered when disaster struck its subsidiary Germanwings in March. ...</td>
</tr>
</tbody>
</table>


Source: Strötgen & Gertz 2016
Time as query topic

$q_{text} = \langle \text{Germanwings} \rangle$, $q_{time} = [2015-03-01, 2015-04-30]$
Time in queries

queries can be **temporally classified** along several dimensions

- query can refer to a single or multiple time periods
  - temporally unambiguous
    (e.g., fifa world cup 2014, battle of waterloo)
  - temporally ambiguous
    (e.g., summer olympics, world war)

- time period is explicitly mentioned or implicitly assumed
  - explicitly temporal
    (e.g., fifa world cup 2014, presidential election 2016)
  - implicitly temporal
    (e.g., superbowl, london bombing)
Time in queries

queries can be **temporally classified** along several dimensions

- query aims for information about the past, present, or future
  - **past**
    (e.g., historic map of rome, news reports about moon landing)
  - **recent**
    (e.g., orlando shooting, tesla stock price)
  - **future**
    (e.g., uefa euro final, academy awards 2016)

- query can refer to any time period at all
  - **atemporal**
    (e.g., muffin recipe, side effects of paracetamol, muscle cramps)
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Temporal document priors

freshness of documents

Li and Croft (2003): approach based on language models for queries favoring more recent documents
- analysis of publication dates of relevant documents

recency query
query 301: international organized crime
Temporal document priors

- freshness of documents

Li and Croft (2003): approach based on language models for queries favoring more recent documents
- analysis of publication dates of relevant documents

rather uniform

query 165: Tobacco Company Advertising and the Young
Temporal document priors

- Query likelihood approach with **temporal document prior** $P[d]$ depending on DCT $t$ of a document and **current time** $c$

  $$
P[d|q] \propto P[d] \cdot \prod_v P[v|d]
  \quad P[d] = \lambda e^{-\lambda(c-t)}
$$

- typically: uniform prior probability $P[d]$, i.e., $P[d]$ can be ignored
- now: exponential distribution for prior probabilities, i.e., recent documents have higher prior probability $P[d]$

**experiments show**

- ranking improvements – if applied on recency queries
Query classification

not all queries are equal

- treating every query as recency query decreases ranking quality
- it is important to distinguish queries

query logs can be analyzed to detect

implicitly temporal and atemporal queries;
temporally ambiguous and unambiguous queries

how?

question in assignment 5
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Time as query topic

Berberich et al. (2010): language modeling approach for temporal information needs

approach addresses main **shortcoming** of standard IR
- **temporal expressions** are treated as **terms**
- their semantics is lost

approach handles
- **explicitly temporal queries**
  i.e., queries with temporal expression
  e.g., “Michael Jordan 1990s”
Problems of standard IR approaches

- **temporal and geographic expressions**
  - (seem to be) treated as regular terms
  - semantics is lost
  → **should be extracted and normalized**

- **query functionality**
  - how to search for time intervals?
  - how to search for geographic regions?
  → **should be defined and provided**

- **results**
  - same ranking as for standard text search
  - no time-/geo-centric exploration features
  → **special ranking is required**
  → **time-/geo-centric exploration should be possible**
Time as query topic

temporal expressions are (often) vague
- precise time interval they refer to is uncertain
- this uncertainty needs to be reflected
- e.g., *in the* 1990s can refer to

approach models temporal expressions as sets of time intervals
- temporal expressions as four-tuples \((tb_l, tb_u, te_l, te_u)\)
- temporal expression \(T = (tb_l, tb_u, te_l, te_u)\) can refer to any time interval \([tb, te]\) such that the following holds

\[
\begin{align*}
  tb_l & \leq tb & \leq tb_u \\
  & & \land \\
  tb & \leq te & \land \\
  te_l & \leq te & \leq te_u
\end{align*}
\]
Time as query topic

documents
- modeled as set of textual terms \( d_{text} \) and set of temporal expressions \( d_{time} \)

queries
- modeled as set of textual terms \( q_{text} \) and set of temporal expressions \( q_{time} \)

query-likelihood approach
- assumes independence between textual terms and temporal expressions

\[
P[q|d] = P[q_{text}|d_{text}] \cdot P[q_{time}|d_{time}]
\]
Time as query topic

query likelihood of **textual part** $P[q_{text}|d_{text}]$ estimated with
- **unigram language model** with Jelinek-Mercer smoothing

query likelihood of **temporal part** $P[q_{time}|d_{time}]$ estimated
- assuming independence between temporal expressions
- assuming uniform probability for temporal expressions from document $d$
- assuming uniform probability for time intervals $Q$ can refer to
- assuming uniform probability for time intervals $T$ can refer to

Berberich et al. (2010)'s evaluation shows importance of treating time in a special way
Time as query topic

- independence between temporal expressions
  \[
P[q_{time} | d_{time}] = \prod_{Q \in q_{time}} P[Q | d_{time}]
  \]

- uniform probability for temporal expressions from \(d\)
  \[
P[Q | d_{time}] = \frac{1}{|d_{time}|} \sum_{T \in D_{time}} P[Q | T]
  \]

- uniform probability for time intervals \(Q\) can refer to
  \[
P[Q | T] = \frac{1}{|Q|} \sum_{[q_b, q_e] \in Q} P[[q_b, q_e] | T]
  \]

- uniform probability for time intervals \(T\) can refer to
  \[
P[[q_b, q_e] | T] = 1([q_b, q_e] \in T)
  \]
Time as query topic

Stroetgen & Gertz (2013): proximity-aware ranking

- no independence between terms and temporal expressions
- three dimensions: text, time, geo
  i.e., no independence between all three dimensions

Multi-word textual query:

- query: search engine

Document 1

search engine

Document 2

search engine
What did Alexander von Humboldt do between *late 18th century* and *early 19th century* in Latein America?
Time as query topic

multidimensional query model with **query dimensions**

- **textual** query \( (q_{text}) \)
- **temporal** query \( (q_{temp}) \): time intervals of interest
- **geographic** query \( (q_{geo}) \): regions of interest
Time as query topic

Textual search query: Alexander von Humboldt

Temporal query: [1794, 1802]

Spatial query (press SHIFT to draw rectangle on map):

Data CC-By-SA by OpenStreetMap

© Jannik Strötgen – ATIR-08
Time as query topic

multidimensional query

- $q_{text}$: Alexander
- $q_{temp}$: late 18th – early 19th century
- $q_{geo}$: box(Latin America)

Document 1

............ **Alexander** visited **Cuba** in **1800**............

.................................

...Until **2001**.................

.................................

...brother of **Paul**............

Document 2

................. **Paul** visited **Cuba** in **2001**.............

...............................

...Until **1800**..................

.................................

...brother of **Alexander**.......

© Jannik Strötgen – ATIR-08
Time as query topic

**term proximity score**
- proximity of terms satisfying all query dimensions

**final score**
- textual, temporal, geographic scores
- term proximity score
Time as query topic

evaluation – data set

- NTCIR GeoTime [Gey et al. (2010)]
  
  e.g., When and where did a volcano erupt in Africa during 2002?
  
  \[ q_{\text{text}} = \text{volcano erupt}; \ q_{\text{temp}} = 2002; \ q_{\text{geo}} = \text{box(Africa)} \]

comparison

- proximity-aware ranking approach
- text baseline: text score \((q_{\text{text}} = \text{volcano erupt 2002 Africa})\)
- boolean baseline: text score & boolean filtering
**Time as query topic**

- boolean baseline outperforms text baseline
- proximity-aware model outperforms both baselines
Outline

1. Left-over from last week: Temponym tagging
2. Dynamics of the Web
3. Indexing Redundant Content
4. Aspects of Time
5. Time as Dimension of Relevance
6. Time as Query Topic
7. Historic Document Collections
Historical Document Collections

- improved **digitization methods** (e.g., OCR)
- (very) **old documents** now being **digitally available**

**examples**
- The New York Times Archive (1851 – today)
- The Times Archive (1785 – now)
- Google Books (1500 – now)
- HathiTrust (1500 – now)
Historical Document Collections

challenges and opportunities

- **unknown publication dates** of documents can be estimated based on similar documents with known publication dates

- **vocabulary gap** between today’s queries and old documents needs to be bridged for effective information retrieval

- **longitudinal document collections** allow analyses that give insights into, e.g., the evolution of language and historic events
Historical Document Collections

IR on historical document collection suffers from **vocabulary gap** between today’s queries and old documents

- **language evolution** (e.g., “and if he hear thee, thou wilt anger him”)
- **terminology evolution** (e.g., Leningrad/Saint Petersburg)

Koolen et al. (2006) treat the problem as

- **cross-language information retrieval problem**
- translate documents using **rewriting rules** mined from the document collection
Historical Document Collections

**phonetic sequence similarity**

- **transcribe** historical and modern words **into phonemes**
  - veeghen (historical) $\rightarrow$ v e g @ n
  - vegen (modern) $\rightarrow$ v e g @ n
- find pairs of historical and modern word with same pronunciation
- split words into sequences of consonants and vowels
  - historical: v e e g h e n
  - modern: v e g e n
- **align sequences** and spot rewritings (e.g., ee $\rightarrow$ e, gh $\rightarrow$ g)
- rewritings that are **often observed** become **rewriting rules**
Summary

- **Temponyms**: phrases with temporal scopes
- **different aspects of time** can be distinguished in IR
- Web is **highly dynamic**
- **Temporal information** (e.g., publication dates and temporal expressions) can be leveraged for more effective IR
- **Web archives** keep often highly-similar old snapshots of web pages, allowing for efficient indexing and time-travel text search
- **Historical document collections** contain documents published long time ago, are challenging to search, but insightful to analyze

Thank you for your attention!
References

- Nunes et al. 2008: Use of Temporal Expressions in Web Search, ECIR.
- Metzler et al. 2009: Improving Search Relevance for Implicitly Temporal Queries, SIGIR.
- Zhang et al. 2010: Learning Recurrent Event Queries for Web Search, EMNLP.
- Pustejovsky et al. 2005: Temporal and Event Information in Natural Language Text, LRE journal.
- Kuzey et al. 2016a: As Time Goes By: Comprehensive Tagging of Textual Phrases with Temporal Scopes, WWW.
References

- Zhang & Suel 2007: Efficient Search in Large Textual Collections with Redundancy, WWW.
- Sengstock & Gertz 2011: CONQUER: A System for Efficient Context-aware Query Suggestions, WWW.
- Li & Croft 2003: Time-Based Language Models, CIKM.
- Berberich et al. 2010: A Language Modeling Approach for Temporal Information Needs, ECIR.
- Strötgen & Gertz 2013: Proximity²-aware Ranking for Textual, Temporal, and Geographic Queries, CIKM.
- Gey et al. 2010: NTCIR-GeoTime Overview: Evaluating Geographic and Temporal Search, NTCIR.
Thanks

some slides / examples are taken from / similar to those of:

- Klaus Berberich, Saarland University, previous ATIR lecture