
Temporal Information Retrieval: Challenges and Opportunities

Omar Alonso¹, Jannik Strötgen²,
Ricardo Baeza-Yates³, Michael Gertz²

¹ Microsoft Corp, ² University of Heidelberg,
³ Yahoo! Research

Overview




- The Role of Temporal Information
- Time in Documents
- Research on Temporal Tagging
- Research Trends

Characteristics of Temporal Information

- Temporal Information is **well-defined**

Characteristics of Temporal Information

- Temporal Information is **well-defined**

- two points/intervals
→ temporal relationship
before 
after 
overlap 
...

Characteristics of Temporal Information

- Temporal Information can be **normalized** to some standard format

Characteristics of Temporal Information

- Temporal Information can be **normalized** to some standard format
 - expressions with same semantics
 - same value in standard format
 - March 28, 2011 → 2011-03-28
 - 2011/03/28 → 2011-03-28
 - today → 2011-03-28

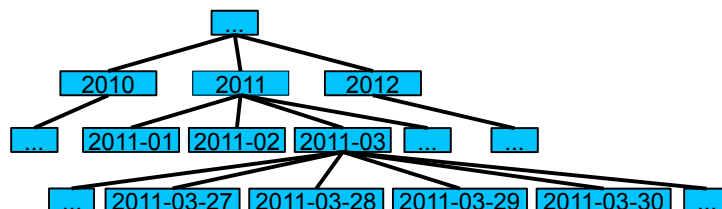
Characteristics of Temporal Information

- Temporal Information can be **organized hierarchically**

Characteristics of Temporal Information

- Temporal Information can be **organized hierarchically**

- different granularities (day, month, ...)
→ tree structure



Overview

- The Role of Temporal Information
- **Time in Documents**
- Research on Temporal Tagging
- Research Trends

Types of Temporal Information

- 4 types of temporal expressions
 - **Dates**
January 25, 2001
 - **Times**
11 a.m.
 - **Durations**
two weeks
 - **Sets**
twice a week

Occurrences in Documents

- explicit expressions
- implicit expressions
- relative expressions

Occurrences in Documents

- **explicit expressions**
 - refer to specific point in time
 - carry all information explicitly
 - January 25, 2010
 - November 2011
 - the 7th century
 - can be normalized without further knowledge

Occurrences in Documents

□ **implicit expressions**

- names of holidays
- names of events
Independency Day 2001
Labor Day 1999
- can be normalized if semantics is known:

Labor Day → first Monday in September

Occurrences in Documents

□ **relative expressions**

- may be underspecified or deictic
today
November 11
two weeks later
- context information is needed
→ reference time
→ relation to reference time

Types of Temporal Information

□ Example:

Document Creation Time: 1998-04-18
Hungarian astronaut Bertalan Farkas is leaving for the
United States to start a new career, he said today.
... On May 22, 1995, Farkas was made a brigadier general,
and the following year he was appointed military attache
... However, cited by District of Columbia traffic police in
December for driving under the influence of ...

The diagram illustrates the temporal relationships between the document's creation time and the events described in the text. A curved arrow points from the 'Document Creation Time: 1998-04-18' box to the word 'today' in the sentence 'he said today'. Another curved arrow points from the 'May 22, 1995' box to the phrase 'the following year' in the sentence 'and the following year he was appointed military attache'. A third curved arrow points from the 'December' box to the phrase 'for driving under the influence of ...' in the sentence 'However, cited by District of Columbia traffic police in December for driving under the influence of ...'.

- explicit and relative expressions with needed context information

Overview

- The Role of Temporal Information
- Time in Documents
- **Research on Temporal Tagging**
- Research Trends

Temporal Tagging

- Tasks
 - Extraction
 - Normalization

Temporal Tagging

- Extraction Task
 - identify temporal expressions
 - identify boundaries of expressions
- classification tasks

Temporal Tagging

- Normalization Task
 - normalize temporal expressions according to standard format
 - TIMEX2 or TIMEX3 contain attributes:
 - type
 - offset
 - value
 - ...

Temporal Tagging

- Normalization Task
 - normalize temporal expressions according to standard format
 - TIMEX2 or TIMEX3 contain attributes:
 - type → date, time, ...
 - offset → On March 28, 2011, the
 - value → 2011-03-28
 - ...

Temporal Taggers

- Extraction
 - rule-based systems
 - machine learning systems
- Normalization
 - (almost) only rule-based approaches

Evaluation Challenges

- Evaluate quality of temporal taggers
 - MUC 1995, 1997
 - only extraction
 - ACE TERN 2004, 2005, 2007
 - extraction and normalization
 - TempEval 2010
 - extraction and normalization

Evaluation Challenges

- Thanks to challenges:
 - annotation guidelines
 - TIDES TIMEX2
 - TimeML (with TIMEX3)
 - annotated corpora
 - TERN 2004, 2005, 2007
 - TimeBank
 - TempEval 2010

Example: HeidelTime

- rule-based system
- performs:
 - extraction
 - normalization

Example: HeidelTime

- Extraction
 - mainly regular expressions
 - pattern resources
 - additional features (POS tags, ...)

Example: HeidelTime

- Normalization
 - knowledge resources (mappings from patterns to normalized forms)
 - linguistic clues (tense of sentences...)

Example: HeidelTime

- Rules
 - type of expression
 - extraction pattern
 - normalization function

- Example

On March 27, 2011, I arrived in India.

Example: HeidelTime

- Expression rule

d1 = (reMonth) (reDay), (reYear4Digit)
 g1 g2 g3

Example: HeidelTime

- Expression rule

d1 = (reMonth) (reDay), (reYear4Digit)

- with

reMonth = (...|February|March|...)

reDay = (0?[1-9]|[1-2][0-9]|30|31)

reYear = (\d\d\d\d)

- (reMonth) (reDay), (reYear4Digit)
March 27 , 2011

Example: HeidelTime

- Normalization function

norm_r1(g1,g2,g3)=

g3-normMonth(g1)-normDay(g2)

- with

- normMonth("March") = "03"

- normDay("27") = "27"

- ...

- March 27, 2011 → 2011-03-27

Example: HeidelTime

- HeidelTime at TempEval 2010
 - best system for extraction
 - best system for normalization
- Extraction Results
 - $p=90\%$, $r=82\%$, $f=86\%$
- Normalization
 - type accuracy = 96%
 - value accuracy = 85%

Example: HeidelTime

- availability:
 - end of April 2011

Overview

- The Role of Temporal Information
- Time in Documents
- Research on Temporal Tagging
- **Research Trends**

Research Trends

- Exploratory Search
- Real-time Search
- Temporal Summaries
- Temporal Clustering
- Question Querying
- Temporal Question Answering
- Temporal Similarity
- Timelines and User Interfaces
- Searching in Time
- Web Archiving
- Spatio-Temporal Information Extraction

Exploratory Search

- How to expose temporal information in exploratory search systems?
- What's the best way of presenting temporal information as a retrieval *cue*?
- For which data sources exploratory search makes sense besides news?
- Is e-discovery a vertical application that can benefit from temporal information?

Real-time Search

- What is the best way to provide a timeline of events in micro-blogging?
- What is the lifespan of a main event?
- How fast and precise can we detect trending events?
- What is the ratio of new content on the topic stream?

Temporal Summaries

- ❑ When to show a timestamp or temporal expressions?
- ❑ Should the snippet present the matching lines in a timeline?
- ❑ Is a temporal summary a good surrogate for a document?
- ❑ For which kind of queries a temporal summary would be appropriate?
- ❑ Should temporal summaries be query independent?

Temporal Clustering

- ❑ Can we identify documents that are contemporary and therefore related?
- ❑ Which chronons can be more useful for clustering?
- ❑ How can we cluster micro-blogging data by time?
- ❑ Is a timeline the best way to cluster search results?

Question Querying

- in addition to text query:
 - temporal constraint
- advantage:
 - “interval querying” is possible
 - e.g., text query = “crisis”
temporal query= “1950-1970”

Question Querying

- How to combine textual and temporal scores?
- How close should textual and temporal matches occur in the document?
- What if a great textual match, “slightly” failed a temporal match? Vice versa?

Temporal Question Answering

- How can be dealt with inconsistent temporal information?
- How can temporal reasoning be executed if temporal relationships are missing?

Temporal Similarity

- Given one document, return temporal similar documents
- Should the temporal interval be the same?
- Or just the temporal focus?
- What if one document is subinterval of the another document?

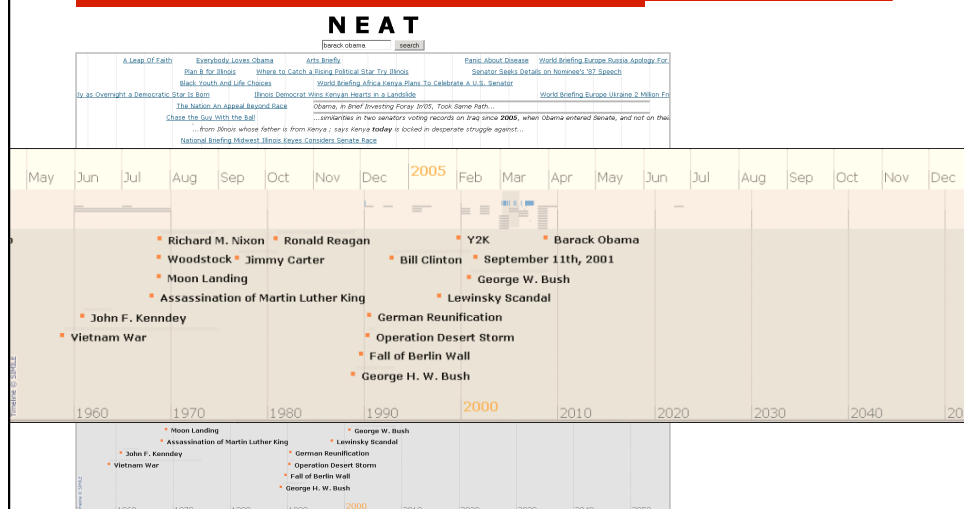
Timelines and User-Interfaces

- ❑ What is the appropriate way to present a timeline?
- ❑ Is a linear timeline the only way to present and anchor documents in time?
- ❑ How can one leverage document temporal measures to present a good display?
- ❑ Are there specific visualizations or user interfaces that can benefit from temporal information?

Alonso, Strötgen, Baeza-Yates, Gertz

TWAW 2011

Neat Example

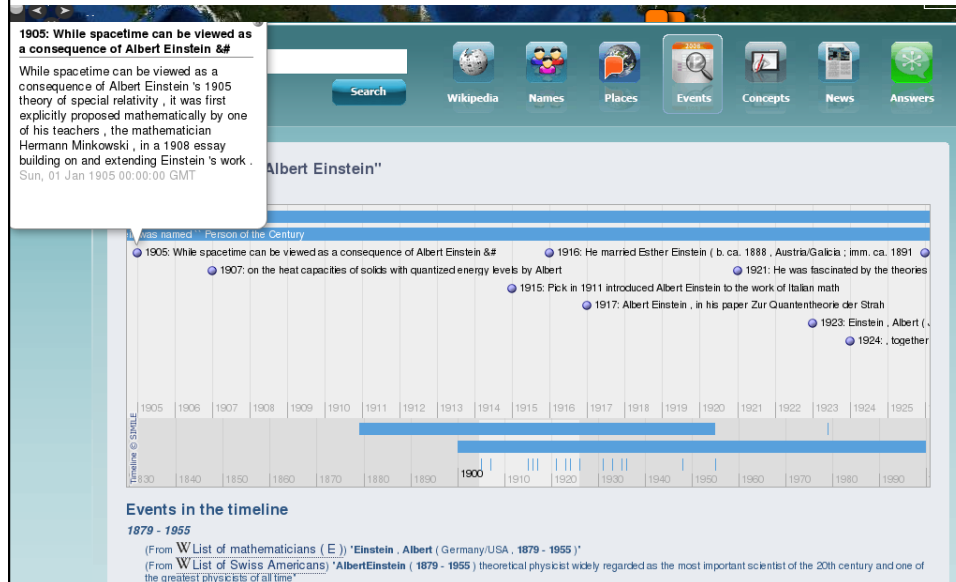


Annotation done via crowdsourcing

Alonso, Strötgen, Baeza-Yates, Gertz

TWAW 2011

Correlator Examples - sandbox.yahoo.com



Searching in Time

- ❑ Besides news, what other sources can one use to search in the past and/or the future?
- ❑ How far does one need to go in time to look at the past?
- ❑ Can we improve bibliographic search instead of just sorting by publication date?
- ❑ How can we evaluate the quality of such systems?

Time Explorer

□ Finding Relations among Entities in News

- Past, present or future!
 - Baeza-Yates, Searching the Future, 2005.
- The clue is the interface
- Part of the Living Knowledge EU project

□ Winner of the HCIR 2010 Challenge

- New York Times collection (1987-2007)
- Found many interesting examples
- Generates new NLP research problems

Time Explorer



(a) Basic Timeline



(b) Time period Selection



(c) Timeline with entity trends

Time Explorer



Figure 4: Searching the Future



Figure 5: Results

Web Archiving

- ❑ How can temporal information be used to predict which pages are likely to change over time?
- ❑ How can temporal coherence be achieved for any point in time or time interval?

Spatio-temporal Information Extraction

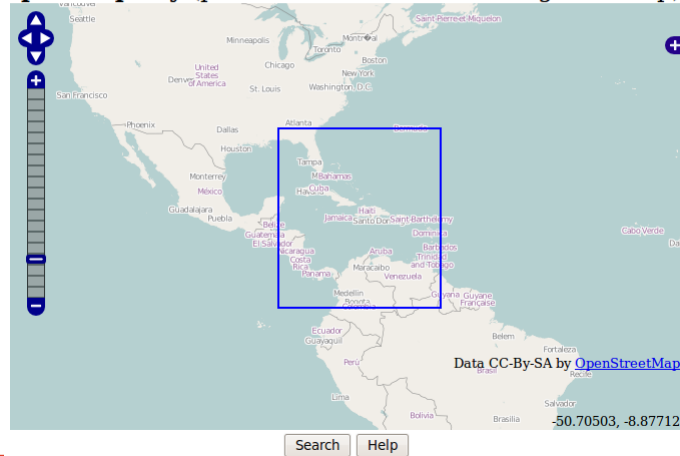
- ❑ “space” similar to “time”
- ❑ both can be normalized
- ❑ How to visualize time and map together?
- ❑ Which scenarios benefit from spatio-temporal exploration?

Spatio-temporal Information Extraction

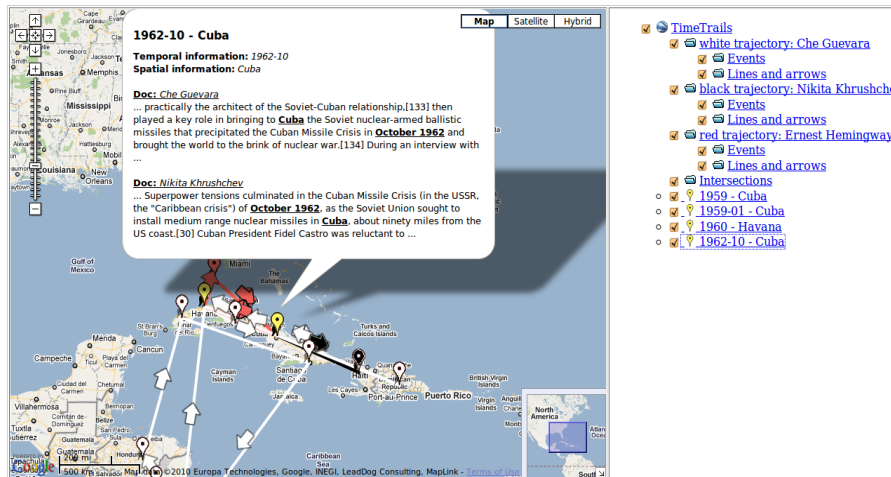
Textual search query:

Temporal query:

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



Spatio-temporal Information Extraction



Conclusions

- A way to enhance retrieval tasks
- Involves not only IR, but also NLP & HCI
- Multidisciplinary challenge
- Plenty of open problems