

Temporal Information Access (Temporalialia)

NTCIR-11 Pilot Task

<https://sites.google.com/site/ntcirtemporalialia>



Hideo Joho

(Univ. of Tsukuba)



Adam Jatowt

(Kyoto University)



Roi Blanco

(Yahoo! Research)

Motivation



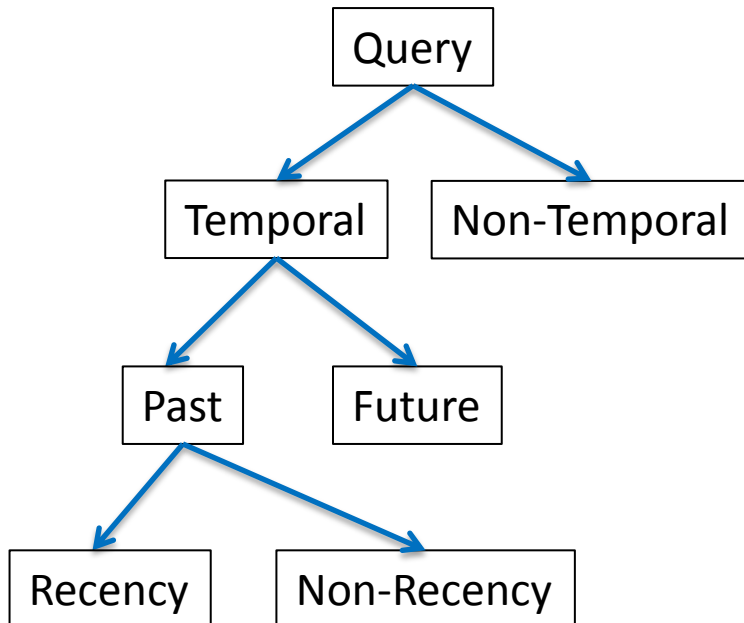
Proposed Subtasks in NTCIR-11

- **Temporal Query Intent Classification (TQIC)**
 - For a given query, participants classify it to different temporal classes such as atemporal, recency, past, future, etc.
- **Temporal Information Retrieval (TIR)**
 - For a given topic, participants retrieve relevant documents for different kinds of temporal queries

Extension to temporal summarization, visualization, and other applications will be considered in NTCIR-12.

Subtasks

Temporal Query Intent Classification (TQIC)

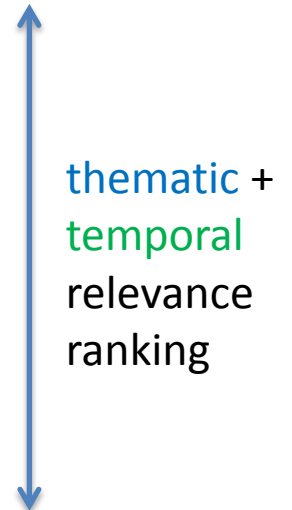


Temporal Information Retrieval (TIR)

Query

1. DocA
2. DocB
3. DocC
4. ...
5. ..

N. DocN



Example Queries

Non-Temporal

how to tie a necktie
pitgoras theorem
facebook

Temporal

watergate story
dollar yen rate
olympics 2016

Past

dollar yen rate
top comedies 2000s
calgary flood news

Future

google glass release date
olympics 2016
opening of new train line

Recency

dollar yen rate
san francisco weather
calgary flood news

Non-Recency

watergate story
top comedies 2000s
why hitler did not attack switzerland

**Queries will possibly be collected
from real search query logs**

SubTasks Data

- **Temporal Query Intent Classification (TQIC)**
 - query strings (100 for training/dry run, 100 for formal run)
 - query issuing times
- **Temporal Information Retrieval (TIR)**
 - 50 topic descriptions (title, description, narrative, query issuing time)
 - qrels (relevant document list)

Document Collection

- **LivingKnowledge Corpus** (150GB compressed)
 - Crawl time: April 2011 and March 2013
 - Source: Annotated News Feeds
 - Size: 20G uncompressed (>5G zipped)
 - 3.8M documents from 1500 different blogs and news sources
- Provided annotations: **Time Annotations**, **Named Entities** and **Sentence Splitting**

ClueWeb Corpora: ClueWeb09 (5TB, Jan-Feb 2009), ClueWeb12 (5.5 TB, Feb- May 2012) will be considered in NTCIR-12.

Example Document

```
<?xml version="1.0" encoding="UTF-8"?>
<doc id=20111004040101_5171>
<meta-info>
  <tag name="host">latimesblogs.latimes.com</tag>
  <tag name="date">2011-10-04</tag>
  <tag name="url">
    http://latimesblogs.latimes.com/the_big_picture/2011/09/the-new-oscar-rule-book-can-the-academy-really-curtail-awards-season-excess.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+PatrickGoldstein+%28L.A.+Times+-+Patrick+Goldstein%29</tag>
  <tag name="sourcers">http://feeds.latimes.com/PatrickGoldstein/</tag>
  <tag name="title">New Oscar rules: Can the Academy curtail awards season excess?</tag>
  <tag name="source-encoding">UTF-8</tag>
  <tag name="rsscategory">Patrick Goldstein</tag>
</meta-info>
<text><SE><E type="E:ORGANIZATION:CORPORATION">New Oscar</E> rules: Can the <E type="E:ORGANIZATION:GOVERNMENT">Academy</E> curtail awards <E type="T:DATE:DATE">season</E> excess?</SE>
<SE>The <E type="T:DATE:DATE">Oscar silly season</E> has officially begun.</SE>
<SE>That's the only way to look at the new <E type="E:FAC:BUILDING">Motion Picture Academy</E> rules governing how <E type="E:ORG_DESC:CORPORATION">studios</E> and <E type="E:PER_DESC">filmmakers</E> can promote their movies during <E type="T:DATE:DATE">Oscar season</E>, a period that <E type="T:DATE:DATE">these days lasts</E> longer than <E type="T:DATE:DATE">winter</E> in <E type="E:GPE:CITY">Siberia</E>.</SE>
<SE>Being a sports <E type="E:PER_DESC">fan</E>, <E type="E:ORGANIZATION:CORPORATION">I've</E> always thought that it was impossible for any <E type="E:ORG_DESC:OTHER">organization</E> to have more arcane rules than the <E type="E:ORGANIZATION:OTHER">NCAA</E>, but the <E type="E:ORG_DESC:EDUCATIONAL">academy</E> has easily topped that <E type="E:PER_DESC">body</E>.</SE>
<SE>Its new regulations are intended to stop <E type="E:ORGANIZATION:CORPORATION">Oscar-season</E> <E type="E:ORG_DESC:CORPORATION">excess</E>, but many believe they could easily lead to more over-the-top campaigning than ever.</SE>
<SE>When it comes to excess, nothing can really top an <E type="E:PERSON">Oscar</E> <E type="E:ORG_DESC:CORPORATION">shindig</E> like the <E type="N:CARDINAL">one</E> <E type="E:ORGANIZATION:CORPORATION">Arianna Huffington</E> threw <T val="201102">last February</T> at her <E type="E:FAC_DESC:BUILDING">house</E> for <E type="E:ORGANIZATION:CORPORATION">Harvey Weinstein's "The King's Speech</E>," which featured not just the A-list <E type="E:PER_DESC">cast</E> and <E type="E:PER_DESC">filmmakers</E> from the movie, but real <E type="E:NORP:NATIONALITY">British</E> <E type="E:PER_DESC">royalty</E>, notably <E type="E:PERSON">Earl Charles Spencer</E>, <E type="E:PER_DESC">brother</E> of the late <E type="E:PER_DESC">Princess</E> <E type="E:PERSON">Diana</E>.</SE>
<SE>The <E type="E:ORG_DESC:POLITICAL">party</E> generated <E type="N:QUANTITY:WEIGHT">tons</E> of <E type="E:PER_DESC">press</E> and publicity, and was clearly designed to create buzz for the film, which ended up winning the <E type="E:PERSON">Oscar</E> for best picture.</SE>
<SE>According to the new rules, a similar <E type="E:ORG_DESC:POLITICAL">party</E> <T val="2011">this year</T> could offer <E type="N:MONEY">just as much</E> pomp and circumstance, <E type="T:TIME">just as long</E> as it happened <E type="T:DATE:DATE">two weeks earlier</E>, before the nominations were announced.</SE>
<SE>Because <E type="E:ORGANIZATION:CORPORATION">"The King's Speech"</E> was already the <E type="E:PER_DESC">favorite</E> to win best picture even before the nominations, it seems clear that the <E type="E:ORG_DESC:POLITICAL">party</E> would have had <E type="N:MONEY">just as</E> much impact if it had been held in <T val="201101">mid-January</T> instead of <T val="201102">early February</T>.</SE>
```

More details at the Task's website: <https://sites.google.com/site/ntcirtemporalia>

Performance Metrics

- Thematic relevance
 - Graded relevance (highly relevant, relevant, partially relevant and not relevant)
- Temporal relevance
- Some function to integrate them, similarly as in:
 - GeoCLEF (Spatial relevance + Thematic relevance)
 - INEX (Structural relevance + Thematic relevance)

Draft Schedule

Sep 02, 2013	NTCIR-11 Kick-off Event
Dec 20, 2013	Task Registration Due
Jan 05, 2014	Document collection release
Jan 05, 2014	Training data for dry run release
Mar 30, 2014	Testing data for formal run release
Jun 30, 2014	Formal run submission due
Aug 1, 2014	Evaluation results release
Aug 1, 2014	Early overview draft release
Sep 1, 2014	Participant papers due
Nov 1, 2014	All camera-ready copy due
Dec 09-12, 2014	NTCIR-11 Conference

Thank you!

tc4fia@googlegroups.com

References

- H. Joho, A. Jatowt, and R. Blanco. *A Survey of Temporal Web Search Experience*. TempWeb 2013 Workshop @ WWW 2013
- A. Jatowt, C. M. Au Yeung and K. Tanaka. *Estimating Document Focus Time*. CIKM 2013
- R. Blanco, H. Halpin, D. M. Herzig, P. Mika, J. Pound, H. S. Thompson, D. T. Tran: *Repeatable and reliable search system evaluation using crowdsourcing*. SIGIR 2011
- M. Matthews, P. Tolchinsky, R. Blanco, J. Atserias, P. Mika, and H. Zaragoza. *Searching Through Time in the New York Times*, HCIR 2010