# NTCIR-11 **Temporalia**

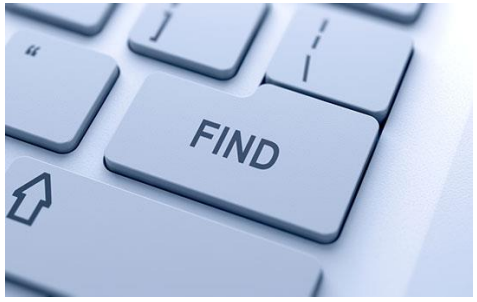
## Temporal Information Access


Hideo Joho *(Univ. of Tsukuba)*
Adam Jatowt *(Kyoto Univ.)*
Roi Blanco *(Yahoo! Research)*
Hajime Naka *(Univ. of Tsukuba)*
Shuhei Yamamoto *(Univ. of Tsukuba)*


https://sites.google.com/site/ntcirtemporalia

# Importance of Time in Search

- When search results do not match temporal intent behind query:
  - "tokyo weather" → doc with yesterday's results
  - "economy forecast japan" → doc with past forecasts
  - "previous tokyo olympics" → doc about achieved milestones in preparations for forthcoming Olympics
  - etc.



# Temporal IR

- 1.5% of queries are explicit temporal queries [Nunes 2008]
  - Germany 1920s, Olympics 1956
- Many queries are implicitly temporal, e.g.:
  - einstein childhood, past depressions, fashion trend, kyoto weather, olympics, population increase expectation
- Queries with temporal component require special treatment [Arikan 2009; Berberich 2010; Kanhabua 2010]
  - By matching query temporality with document temporality

Recent survey on Temporal IR contains over 150 citations [Campos 2014]

Arikan, I., Bedathur, S.J. and Berberich, K. Time Will Tell: Leveraging Temporal Expressions in IR. WSDM2009
Berberich et al. Language Modeling Approach for Temporal Information Needs. ECIR 2010
Kanhabua, N. and Norvag,K., Determining Time of Queries for Reranking Search Results. ECDL2010
Nunes, S., Ribeiro, C., and David, G. Use of Temporal Expressions in Web Search. ECIR2008
Campos, R., Dias, G. Jorge, A.M., and Jatowt, A. Survey of Temporal Information Retrieval and Related Applications. ACM Comput. Surv. 47(2): 15 (2014)

# Our Objective

- Offer standardized test bench for evaluating different aspects of **T**emporal **I**nformation **R**etrieval
  - Foster research to understand temporal aspects of search needs and to accommodate them in the best way

**Search aspects**:

- Intent
- Query
- Context
- Document processing
- Result ranking
- Result presentation
- User feedback
- ….

The problem is not trivial…

# NTCIR-11 Temporalia: Subtasks

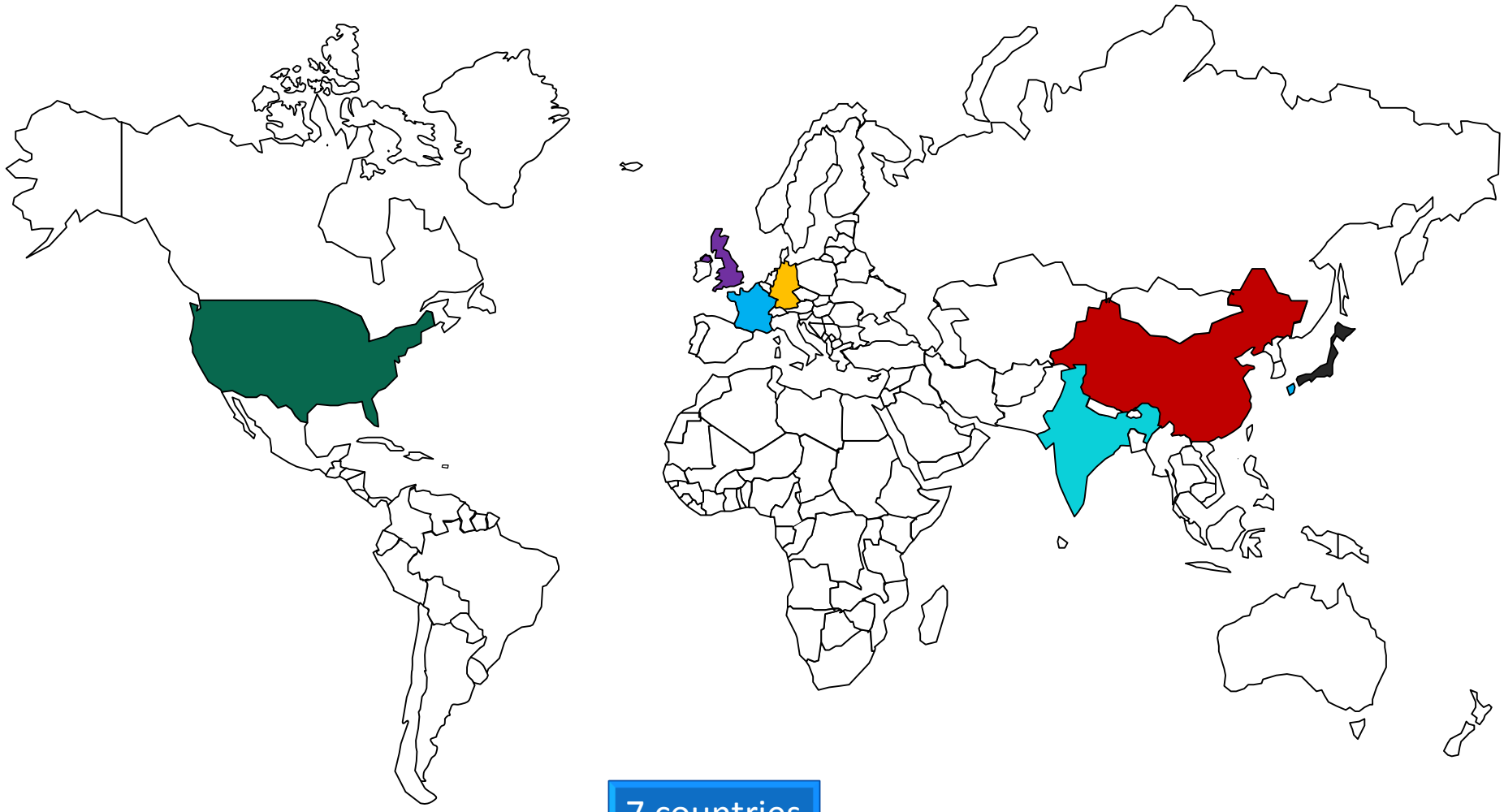
- Temporal Query Intent Classification (**TQIC**)
  - Classify query to temporal classes
- Temporal Information Retrieval (**TIR**)
  - Rank documents for temporal queries

**Search aspects**:

- Intent
- Query
- Context
- Document processing
- Result ranking
- Result presentation
- User feedback
- ….

Just a pilot task...

# Participation



7 countries

TEMPORALIA

# Participating Teams

- Total number of participating teams **9**
  - **6** for TQIC (17 runs)
  - **5** for TIR + 1 organizer (18 runs)

| Team ID | Institution | Country | TQIC | TIR |
|---|---|---|---|---|
| **HITSZ** | Graduate School of Harbin Institute of Technology at Shenzhen | P.R.C. | ⏳ | ⏳ |
| **HULTECH** | University of Caen | France | ⏳ | |
| **TUTA1** | University of Tokushima | Japan | ⏳ | ⏳ |
| **Andd7** | Dhirubhai Ambani Institute of Information and Communication Technology | India | ⏳ | ⏳ |
| **MPII** | Max Planck Institute for Informatics | Germany | ⏳ | |
| **UniMAN** | University of Manchester | UK | ⏳ | |
| **BRKLY** | U.C. Berkeley | USA | | ⏳ |
| **OSKAT** | Osaka Kyoiku University | Japan | | ⏳ |
| **ORG** | Temporalia Organizers | Japan, Spain | | ⏳ |

# Temporal Query Intent Classification

# Background: Search for Present, Past and Future Related Information on the Web

- Survey of 110 users about their recent search queries

| Answers | Frequency | Percentage |
|---|---|---|
| Information older than 1 year | 9 | 8.2% |
| One year ago | 0 | 0.0% |
| Several months ago | 5 | 4.5% |
| Several weeks ago | 2 | 1.8% |
| Last week | 2 | 1.8% |
| Several days ago | 9 | 8.2% |
| Yesterday | 9 | 8.2% |
| Today | 53 | 48.2% |
| Tomorrow | 1 | 0.9% |
| Several days later | 5 | 4.5% |
| Next week | 0 | 0.0% |
| Several weeks later | 1 | 0.9% |
| Several months later | 1 | 0.9% |
| One year later | 0 | 0.0% |
| More than one year later | 1 | 0.9% |
| Others (Please specify) | 12 | 10.9% |
| Total | 110 | 100.0% |

**Past: 24.5%**

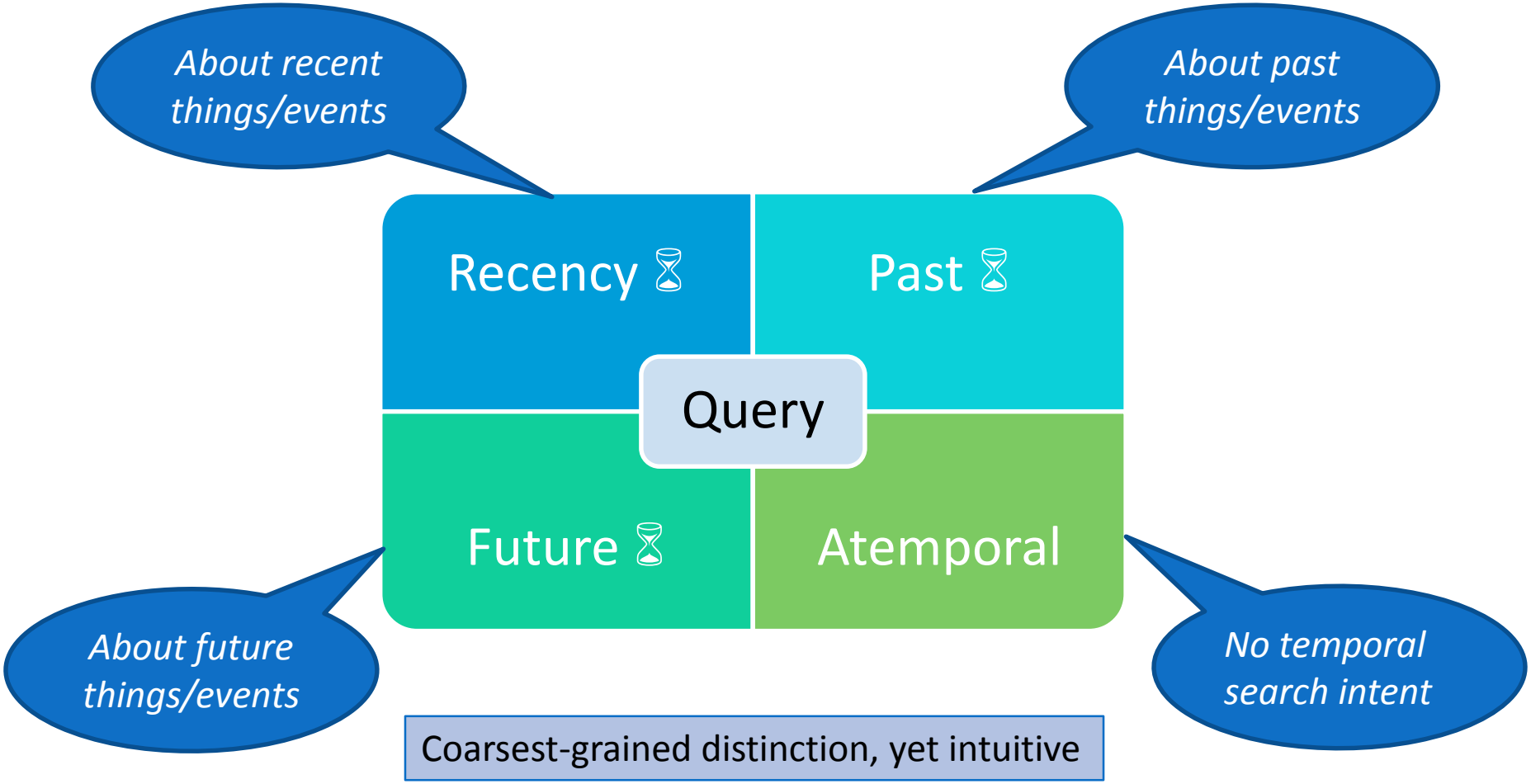
**Present: 57.3%**

**Future: 7.2%**

**Target time of sought information**

Joho, H., Jatowt, A., Blanco, R. A survey of temporal web search experience. TempWeb2013

# Query Temporal Categories



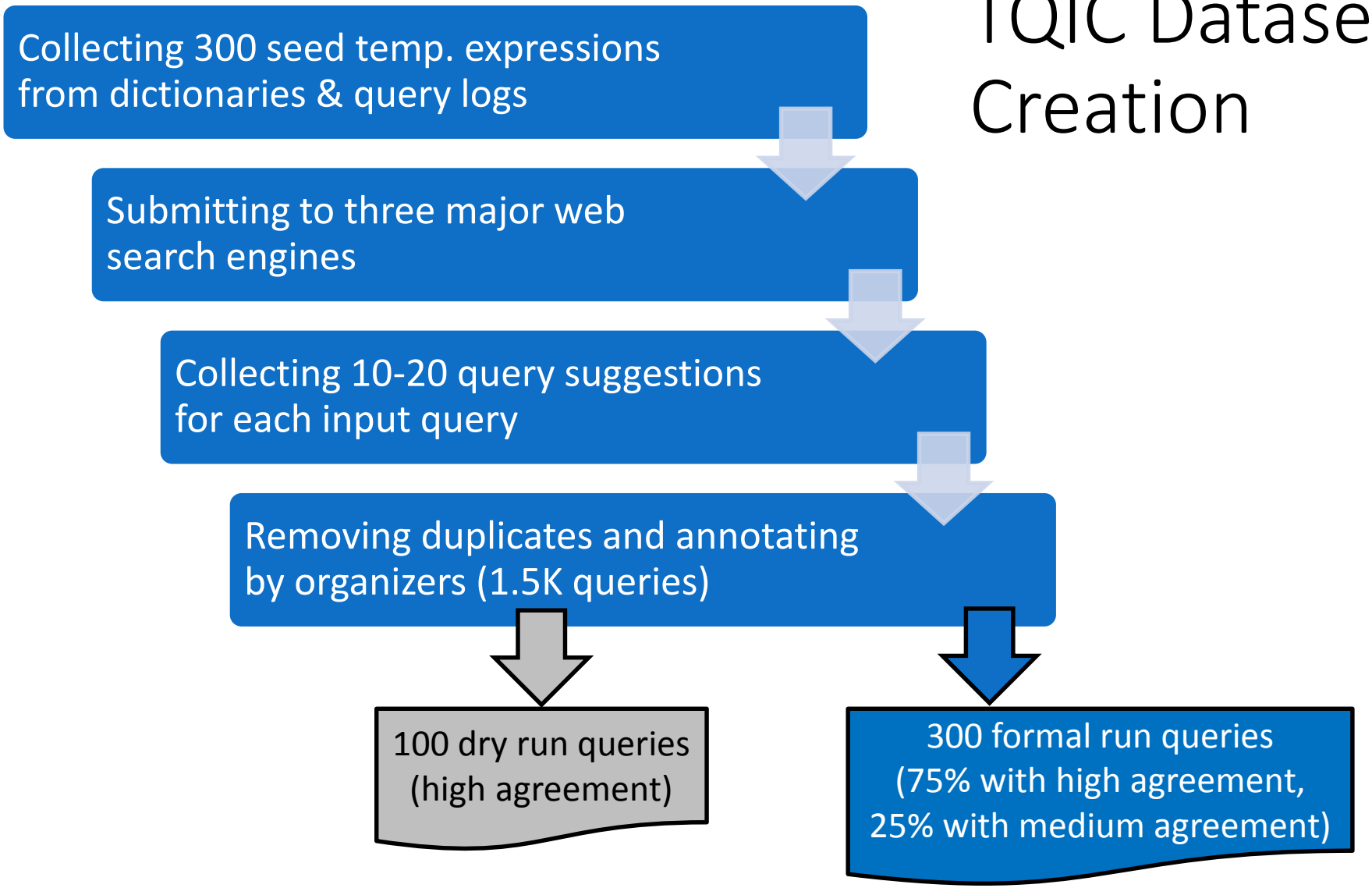

Coarsest-grained distinction, yet intuitive

# TQIC: Query Examples

| Class | Query |
|---|---|
| Past | Who Was Martin Luther |
| Past | when did the titanic sink |
| Past | Yuri Gagarin Cause of Death |
| Past | History of Coca-Cola |
| Past | price hike in bangladesh 2008 |
| Recency | apple stock price |
| Recency | Number of Millionaires in USA |
| Recency | time in london |
| Recency | Trendy Plus Size Clothing |
| Recency | Did the Pirates Win Today |
| Future | 2013 MLB Playoff Schedule |
| Future | College Baseball Regional Projections |
| Future | disney prices 2014 |
| Future | long term weather forecast |
| Future | release date for ios7 |
| Atemporal | blood pressure monitor |
| Atemporal | distance from earth to sun |
| Atemporal | how to start a conversation |
| Atemporal | New York Times |
| Atemporal | lose weight quickly |

Query issuing time provided

# TQIC Dataset Creation

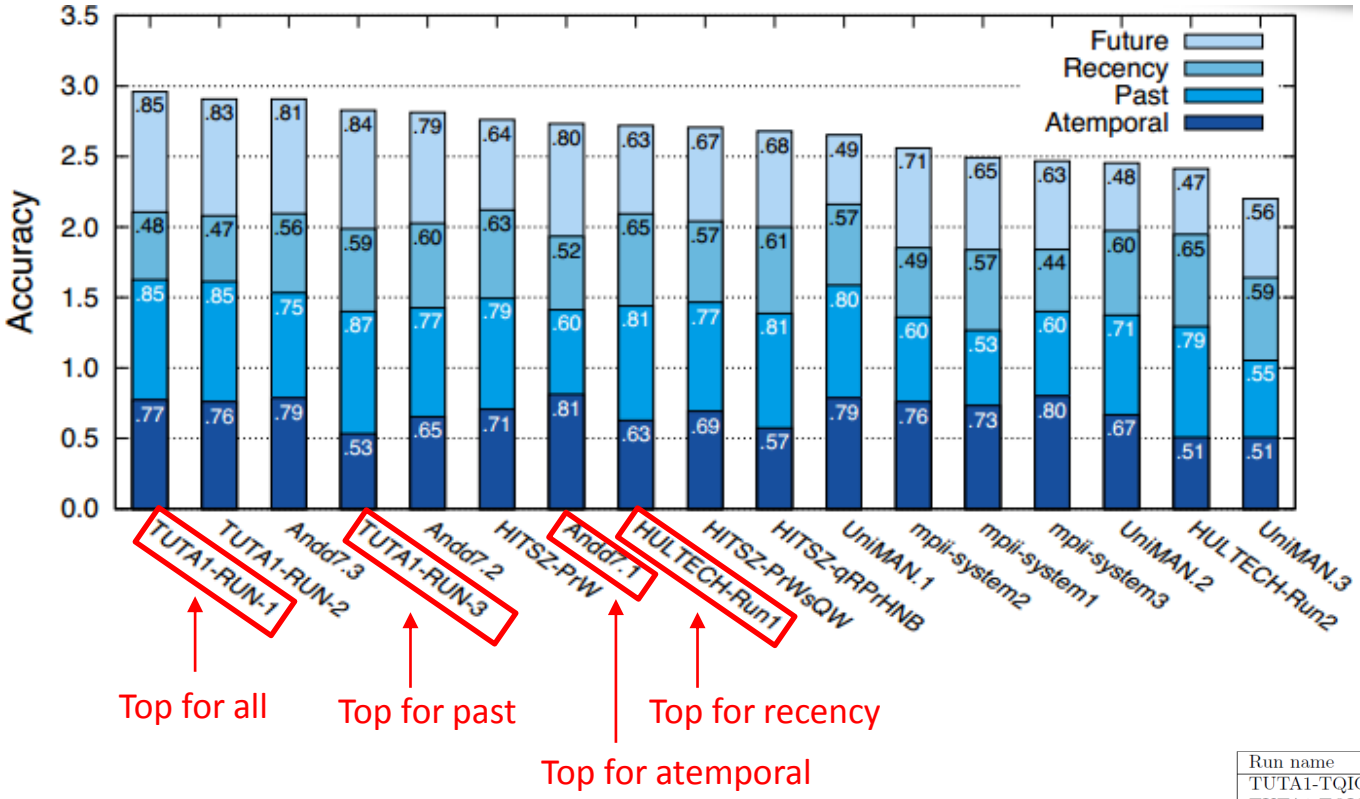


# TQIC Approaches

| Team | Runs | Num. of features | External data sources | External tools | Classifiers |
|---|---|---|---|---|---|
| HITSZ | 3 | 3 features + ngrams and POS ngrams | Web search results (titles, snippets) | POS tagger, NER | Classifier voting & rule-based method |
| HULTECH | 2 | 11 features | Web search results (snippets) | TempoWordnet [1], GTE [2] | Ensemble learning (8 classifiers) |
| TUTA1 (top) | 3 | 3 feature groups | AOL 500K query session | POS tagger, SUTime [3], NER | Semi-supervised & supervised classifiers (Log. Reg., SVMLin [4]) |
| Andd7 | 3 | 3 features + bag of words | - | Porter Stemmer | SVM, NB, DT & their agreement |
| MPII | 3 | 6 feature groups + ngrams | DMOZ directory | 2 thesauri, POS tagger, NER | NB, DT & simulated annealing |
| UniMan | 3 | 19 features | Wikipedia | POS tagger, TempoWordnet [1], ManTIME [5] | SVM & Random Forrest |

[1] https://tempowordnet.greyc.fr/
[2] http://www.ccc.ipt.pt/~ricardo/software
[3] http://nlp.stanford.edu/software/sutime.shtml
[4] http://vikas.sindhwani.org/svmlin.html
[5] http://www.cs.man.ac.uk/~filannim/projects/tempeval-3/

# TQIC Results




| | |
|---|---|
| Mean Acc. Future: | **0.68** |
| Mean Acc. Recent: | **0.56** |
| Mean Acc. Past: | **0.73** |
| Mean Acc. Atemp.: | **0.69** |
| *Mean Acc. All:* | ***0.66*** |

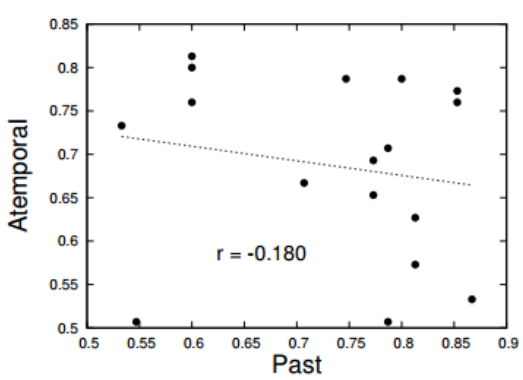
| Run name | Atemporal | Past | Recent | Future | All |
|---|---|---|---|---|---|
| TUTA1-TQIC-RUN-1 | 0.773 | 0.853 | 0.480 | **0.853** | **0.740** |
| TUTA1-TQIC-RUN-2 | 0.760 | 0.853 | 0.467 | 0.827 | 0.727 |
| tqic-Andd7.3 | 0.787 | 0.747 | 0.560 | 0.813 | 0.727 |
| TUTA1-TQIC-RUN-3 | 0.533 | **0.867** | 0.587 | 0.840 | 0.707 |
| tqic-Andd7.2 | 0.653 | 0.773 | 0.600 | 0.787 | 0.703 |
| tiqc-HITSZ-PrW | 0.707 | 0.787 | 0.627 | 0.640 | 0.690 |
| tqic-Andd7.1 | **0.813** | 0.600 | 0.520 | 0.800 | 0.683 |
| TQIC-HULTECH-Run1 | 0.627 | 0.813 | **0.653** | 0.627 | 0.680 |
| tiqc-HITSZ-PrWsQW | 0.693 | 0.773 | 0.573 | 0.667 | 0.677 |
| tiqc-HITSZ-qRPrHNB | 0.573 | 0.813 | 0.613 | 0.680 | 0.670 |
| tqic-UniMAN.1 | 0.787 | 0.800 | 0.573 | 0.493 | 0.663 |
| tqic-mpii-system2 | 0.760 | 0.600 | 0.493 | 0.707 | 0.640 |
| tqic-mpii-system1 | 0.733 | 0.533 | 0.573 | 0.653 | 0.623 |
| tqic-mpii-system3 | 0.800 | 0.600 | 0.440 | 0.627 | 0.617 |
| tqic-UniMAN.2 | 0.667 | 0.707 | 0.600 | 0.480 | 0.613 |
| TQIC-HULTECH-Run2 | 0.507 | 0.787 | **0.653** | 0.467 | 0.603 |
| tqic-UniMAN.3 | 0.507 | 0.547 | 0.587 | 0.560 | 0.550 |
| Mean | 0.687 | 0.733 | 0.565 | 0.678 | 0.665 |
| SD | 0.105 | 0.112 | 0.064 | 0.128 | 0.051 |

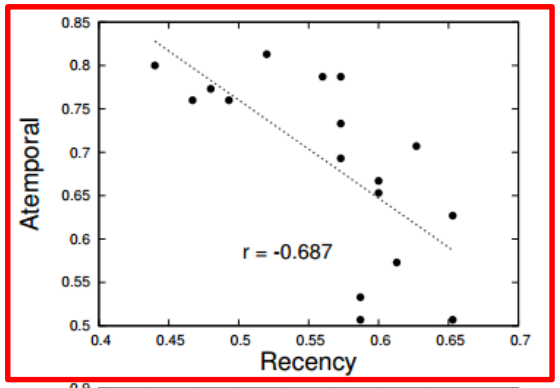
TEMPORALIA

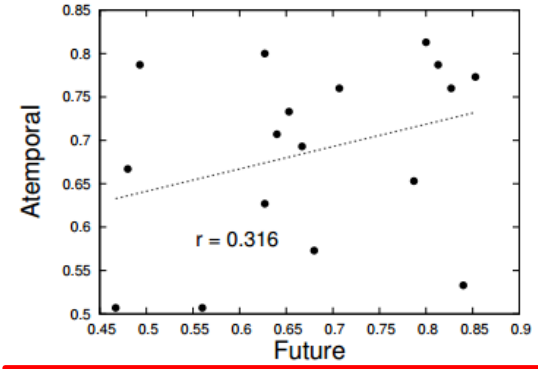
# TQIC: Confusion Matrix

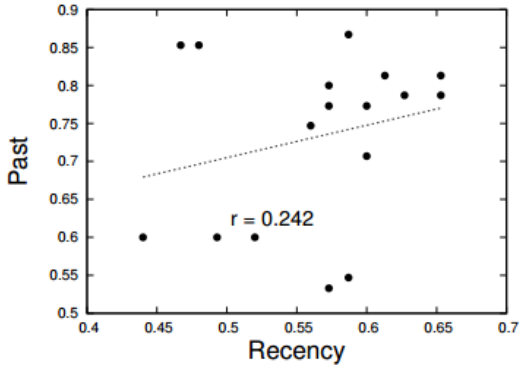
| | | Estimated class | | | | |
|---|---|---|---|---|---|---|
| | | Atemporal | Past | Recency | Future | total |
| Answer class | Atemporal | 783 (67.7%) | 111( 9.6%) | 193(16.7%) | 69( 6.0%) | 1,156 |
| | Past | 147(13.1%) | 836(74.2%) | 61( 5.4%) | 82( 7.3%) | 1,126 |
| | Recency | 154(13.5%) | 28( 2.5%) | 638(55.9%) | 322(28.2%) | 1,142 |
| | Future | 51( 4.4%) | 18( 1.6%) | 299(25.9%) | 786(68.1%) | 1,154 |
| | total | 1,135(24.8%) | 993(21.7%) | 1,191(26.0%) | 1,259(27.5%) | 4,578 |

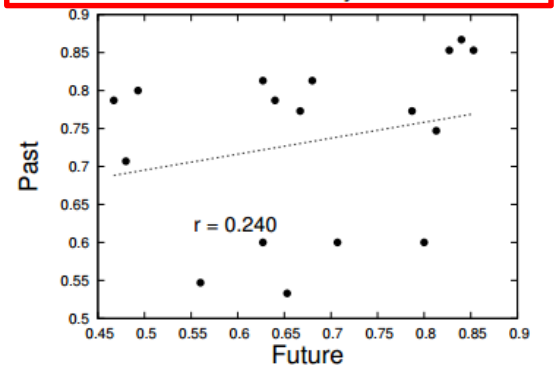
# Pearson's Correlation of Temporal Classes

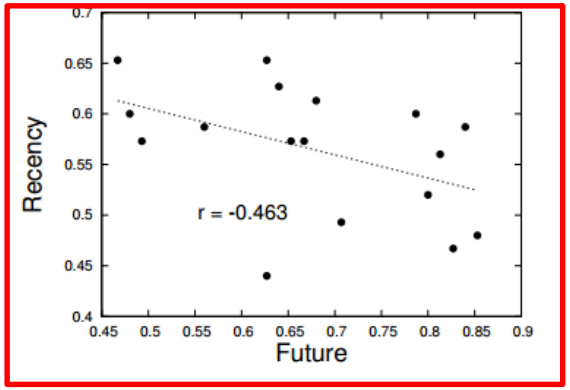


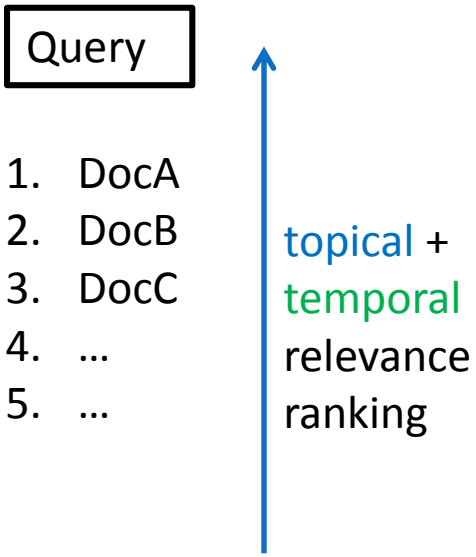



# Temporal Information Retrieval

# NTCiR-11: Temporal Information Retrieval

- Temporal Information Retrieval (TIR)
- Returning top 100 documents for each subtopic of same topic:
  - past, recency, future and atemporal




https://sites.google.com/site/ntcirtemporalia/

# Document Collection

- **LivingKnowledge Corpus** [http://livingknowledge.europarchive.org/]
  - Crawl time: April 2011 and March 2013
  - Source: annotated **news** and **blog** feeds
  - Size: 20GB uncompressed (>5GB zipped)
    - 3.8M documents from 1500 different blogs and news sources
    - Text only
- Provided annotations: Time Annotations, Named Entities and Sentence Splitting
  - Participants can choose to use annotations or not
  - Participants can use external collections such as Wikipedia

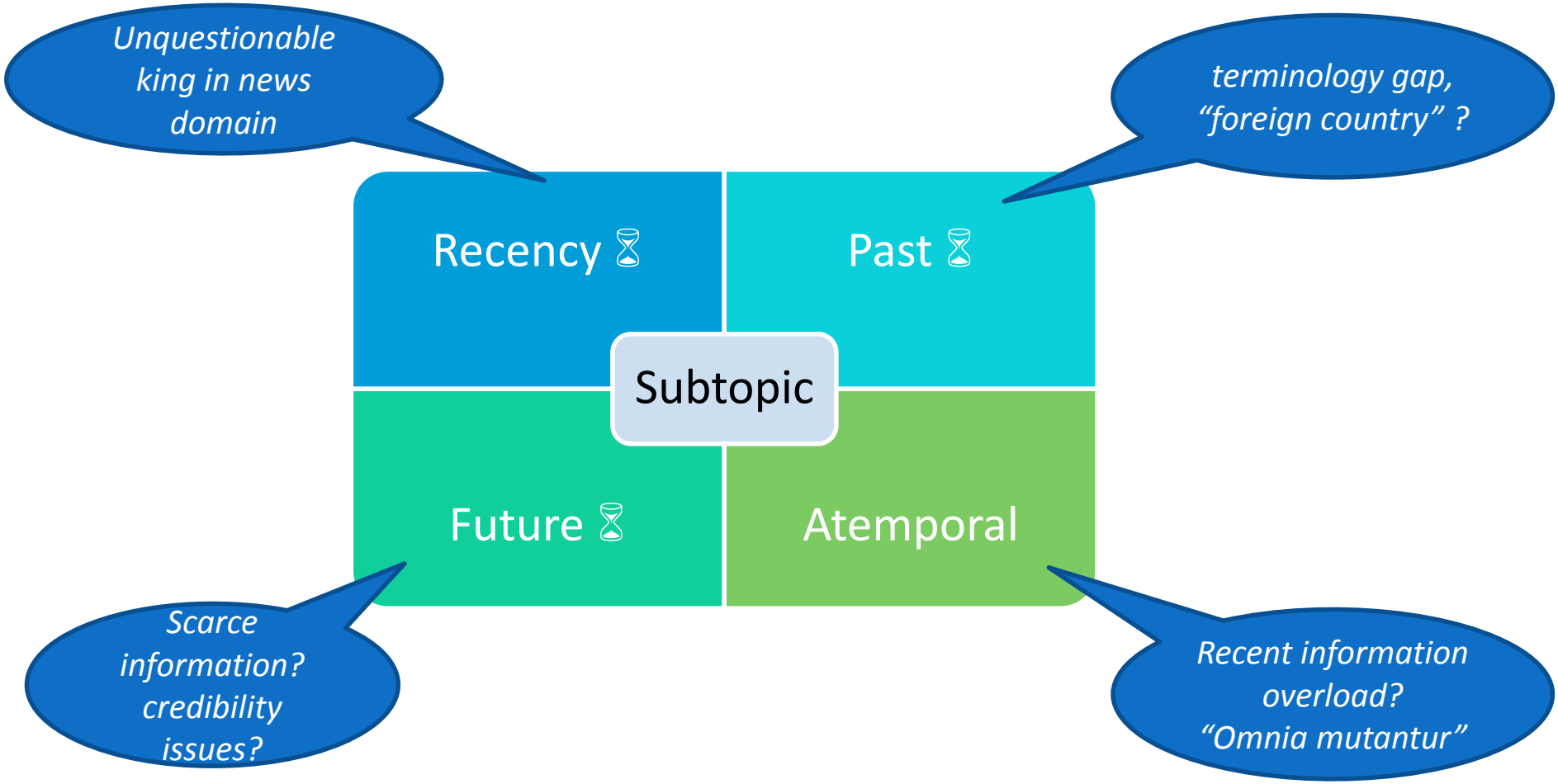
# Example Document

```
<?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&amp;
    utm_medium=feed&amp;utm_campaign=Feed%3A+PatrickGoldstein+%28L.A.+Times+-+Patrick+Goldstein%29</tag>
    <tag name="sourcerss">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: https://sites.google.com/site/ntcirtemporalia
M. Matthews et al. *Searching Through Time in the New York Times*, HCIR, 2010

TEMPORALIA

# Sub-Topic Temporal Categories and their Characteristics

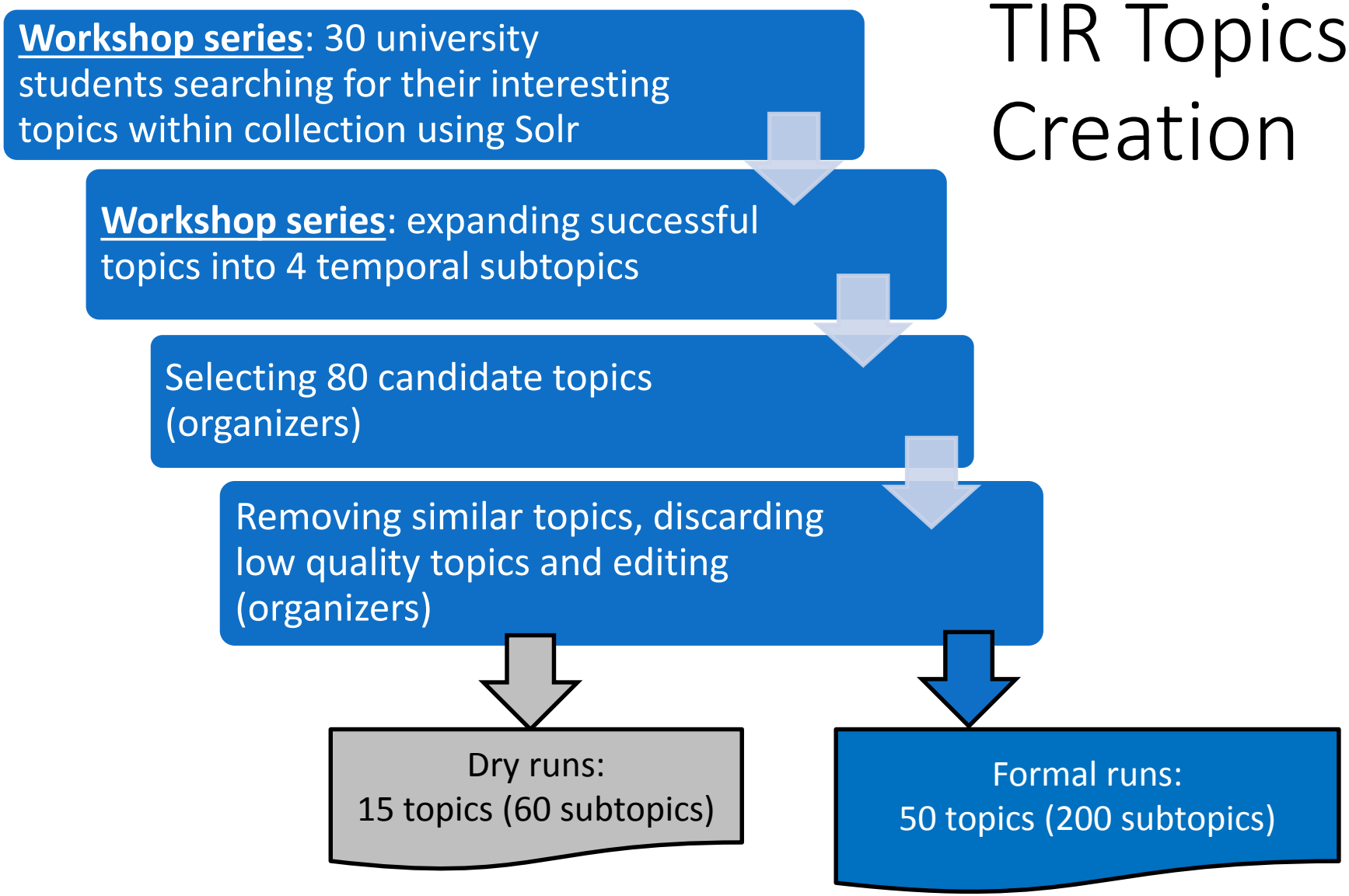


# Temporalia: Topic Examples

| | |
|---|---|
| **Title** | Girl with the Dragon Tattoo |
| **Description** | I've recently watched a film called Girl with the Dragon Tattoo, and really liked it. Therefore, I would like to gather information about the movie. |
| **Past question** | How did the casting of the film develop? |
| **Recency question** | What did the recent reviews say about the film? |
| **Future question** | Is there any plan about its sequel? |
| **Atemporal question** | What are the names of main actors and actresses of the film? |
| **Search date** | 28 Feb 2013 GMT+0:00 |

| | |
|---|---|
| **Title** | Father's Day |
| **Description** | I am from a country where Father's Day is not a common custom, and I would like to learn more about the Father's day and its relation to the role of father in society. |
| **Past question** | What is the history and origin of the Father's day? |
| **Recency question** | What famous persons has lately done during Father's day? |
| **Future question** | What is the future outlook for the problem of fatherlessness and how the father's role is supposed to change?. |
| **Atemporal question** | What are common gifts for Father's day? |
| **Search date** | 1 Mar 2013 GMT+0:00 |

# TIR Topics Creation



# TIR: Relevance Assessment (Formal Run)

1. About 36K documents pooled for evaluation
   - Depth of 20
2. Workshop participants found one highly relevant and one non-relevant document for each subtopic
   - Used as test questions for crowdsourcing platform after manual verification
3. Crowdsourcing with common settings [Kazai 2013]:
   - min. 3 judgments per document
   - 5 docs/task
   - min. 120sec work time/task
   - $0.1/task

CrowdFlower

G. Kazai, J. Kamps, and N. Milic-Frayling. An analysis of human factors and label accuracy in crowdsourcing relevance judgments. Information Retrieval, 16(2):138–178, 2013.

# TIR: Relevance Assessment

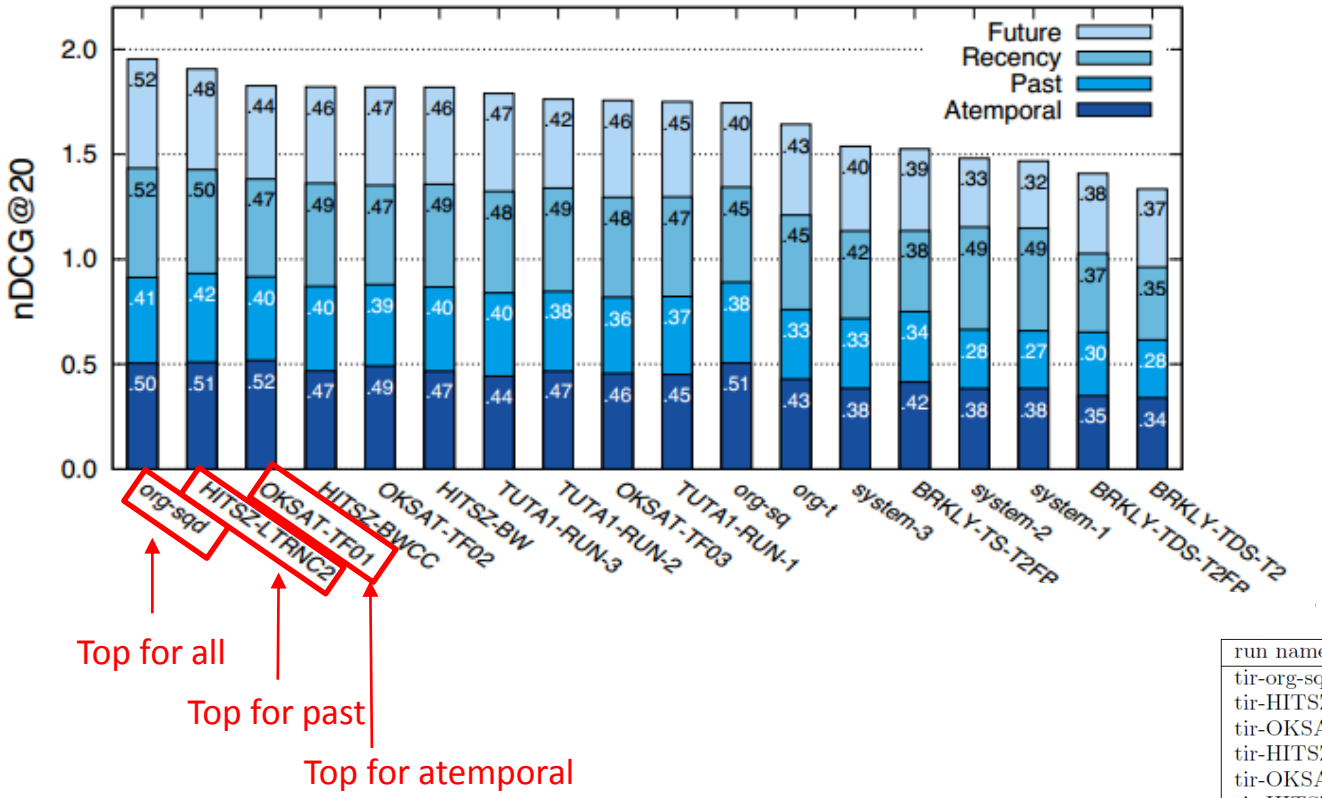
- **L2**: **Highly relevant** – exhaustive information related to search question
- **L1**: **Relevant** – some information related to search question contained (non exhaustive)
- **L0**: **Not relevant** – no information related to search question

| | Atemporal | Past | Recency | Future | All |
|---|---|---|---|---|---|
| *L2* | 1,985 (23.3%) | 1,719 (18.5%) | 2,271 (24.6%) | 2,102 (22.6%) | 8,077 (22.2%) |
| *L1* | 2,574 (29.9%) | 2,361 (25.5%) | 2,555 (27.6%) | 2,747 (29.5%) | 10,210 (28.1%) |
| *L0* | 3,976 (46.7%) | 5,196 (56.0%) | 4,417 (47.8%) | 4,449 (47.8%) | 18,038 (49.7%) |
| Total | 8,508 | 9,276 | 9,243 | 9,298 | 36,325 |

# TIR Approaches

| Team | Runs | Used tags | External data sources | External tools | Approach |
|---|---|---|---|---|---|
| HITSZ | 3 | time annotations of collection | - | Own TQIC classification method | Learning to rank, BM25 |
| TUTA1 | 3 | time annotations of collection | - | POS tagger | Learning to rank |
| Andd7 | 3 | time annotations of collection | - | POS tagger | Lucene |
| BRKLY | 3 | - | - | - | Logistic regression model, blind relevance feedback |
| OKSAT | 3 | time annotations of collection | Wikipedia, Web search results | - | Plural set of terms |
| ORG (top) | 3 | - | - | - | BM25 (Solr) |

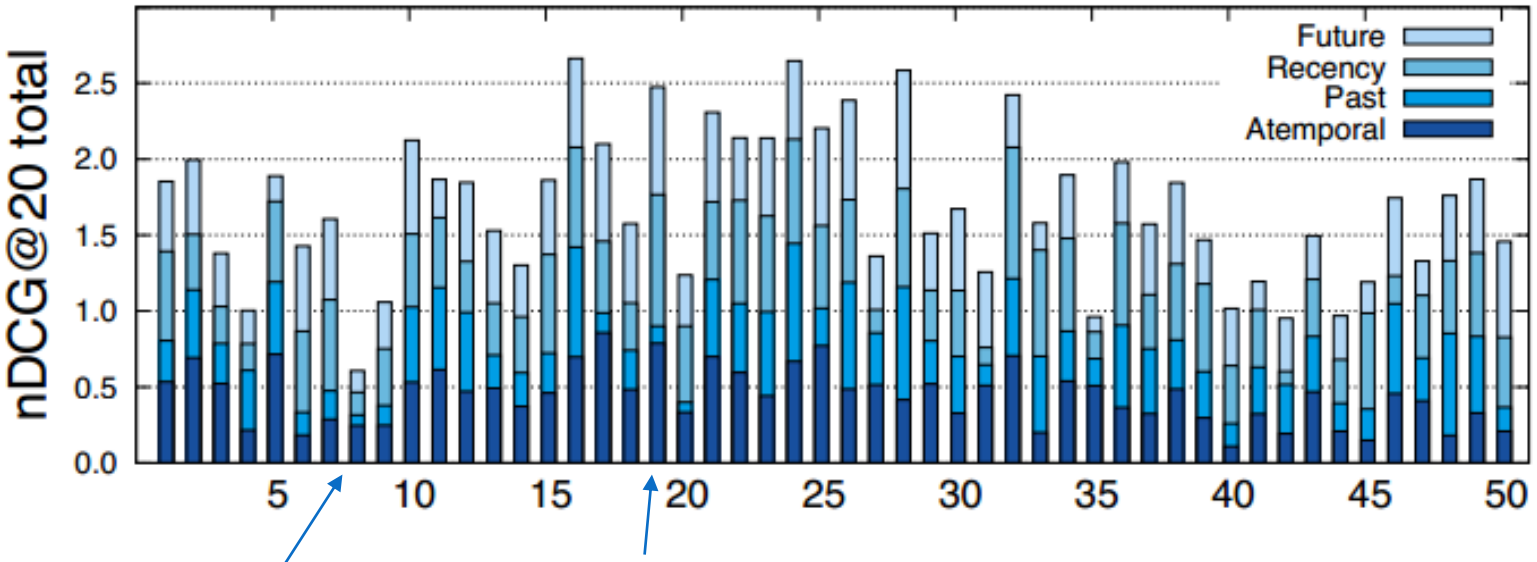
# TIR Results




| run name | Q@20 | P@20 | nDCG@20 |
|---|---|---|---|
| tir-org-sqd | **0.416** | **0.618** | **0.488** |
| tir-HITSZ-LTRNC2 | 0.410 | 0.602 | 0.477 |
| tir-OKSAT-TF01 | 0.383 | 0.583 | 0.457 |
| tir-HITSZ-BWCC | 0.385 | 0.590 | 0.455 |
| tir-OKSAT-TF02 | 0.383 | 0.584 | 0.455 |
| tir-HITSZ-BW | 0.384 | 0.590 | 0.454 |
| TUTA1-TIR-RUN-3 | 0.370 | 0.583 | 0.447 |
| TUTA1-TIR-RUN-2 | 0.369 | 0.579 | 0.441 |
| tir-OKSAT-TF03 | 0.372 | 0.572 | 0.439 |
| TUTA1-TIR-RUN-1 | 0.362 | 0.568 | 0.438 |
| tir-org-sq | 0.364 | 0.559 | 0.436 |
| tir-org-t | 0.348 | 0.531 | 0.411 |
| system-3 | 0.298 | 0.500 | 0.385 |
| TIR-BRKLY-TS-T2FB | 0.304 | 0.501 | 0.382 |
| system-2 | 0.290 | 0.478 | 0.370 |
| system-1 | 0.286 | 0.475 | 0.367 |
| TIR-BRKLY-TDS-T2FB | 0.274 | 0.468 | 0.352 |
| TIR-BRKLY-TDS-T2 | 0.251 | 0.445 | 0.334 |

| | |
|---|---|
| Mean nDCG@20 Future: | **0.43** |
| Mean nDCG@20 Recent: | **0.46** |
| Mean nDCG@20 Past: | **0.39** |
| Mean nDCG@20 Atemp.: | **0.44** |
| *Mean nDCG@20 All:* | ***0.42*** |

# TIR Results per Topic




**Topic**: *"**comet elenin**"*
- Past: *other comets spotted around the time when Elenin was discovered*

**Topic**: "***teaching English as a second language***"
- Future: *future learning methods*
- Past: *past learning methods before 2000*
- Recency: *latest technologies available*
- Atemp.: *definition of second languages*

Topics tend to have varying difficulty over different temporal classes

# Conclusions

- Pilot task
  - First task entirely devoted to temporal information retrieval
  - Subtasks:
    - Temporal Query Intent Classification (**TQIC**)
    - Temporal Information Retrieval (**TIR**)
  - 8 different teams participating
- Provided task-specific document collection
  - Available for anyone
- Much room for future improvement

# Round Table Discussion

- Day: Thu 11th December
- Time: 16:05-18:00
- Location: NII 20F, Room: 2001A

# Proposal of Temporalia-2

| | | |
|---|---|---|
| **Temporal Intent Disambiguation (TID)** Subtask | Challenge | Assign probability to each temporal intent class for a query |
| | Class (4) | Atemporal, past, recency, future |

| | | |
|---|---|---|
| **Temporally Diversified Retrieval (TDR)** Subtask | Challenge | **Diversified retrieval** |
| | Data: Living Knowledge news and blogs | |
| | Evaluation metric | To be decided |
| | | • Relevance<br>• Diversity<br>• Proportionality |

| | |
|---|---|
| **Temporal Search Result Presentation (TSRP)** Subtask | Challenge<br>• Open-ended & exploratory task<br>• Design of interactive interface with good visualization and flexibility for information access |

# Thank you for Participation!

## Any suggestions for improving Temporalia ver.2 are welcome!

tc4fia@googlegroups.com