NTCIR

Temporal Information Access (Temporalia-2)

NTCIR-12 Core Task
http://ntcirtemporalia.github.io

Hideo Joho (University of Tsukuba)

Adam Jatowt (Kyoto University)

Roi Blanco (Yahoo! Research)

Haitao Yu (University of Tsukuba)

Shuhei Yamamoto (University of Tsukuba)

Motivation: Temporal IR

- 1.5% of queries are explicit temporal queries
  - *Germany 1920s, Olympics 1956*

- Many queries are implicitly temporal, e.g.:
  - *einstein life*, *fashion trend*, *tokyo weather*, *olympics*

- Temporal queries require special treatment
  - By matching query temporality with document temporality

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)

Motivation: Importance of Time in IR

- Example results that do not match query temporal intent:
  - *tokyo weather* → doc with yesterday's results (?)
  - *economic forecast japan* → doc with past forecasts (?)
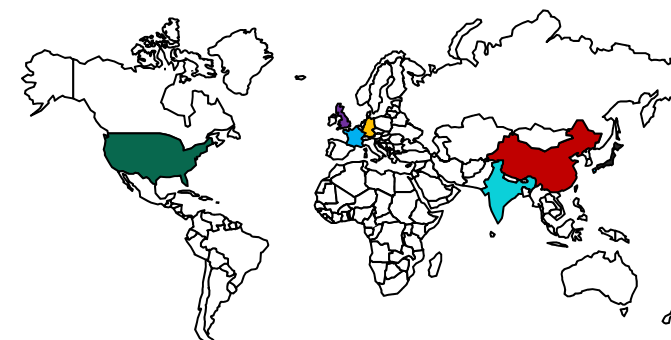 - *previous tokyo olympics* → doc about Olympics 2020 (?)

(PAST) Temporalia-1: English Lang.

1. Temporal Query Intent <u>Classification</u> (**TQIC**)

   • Classify query to temporal classes

2. Temporal Information Retrieval (**TIR**)

   • Rank documents for temporal queries

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		⌛

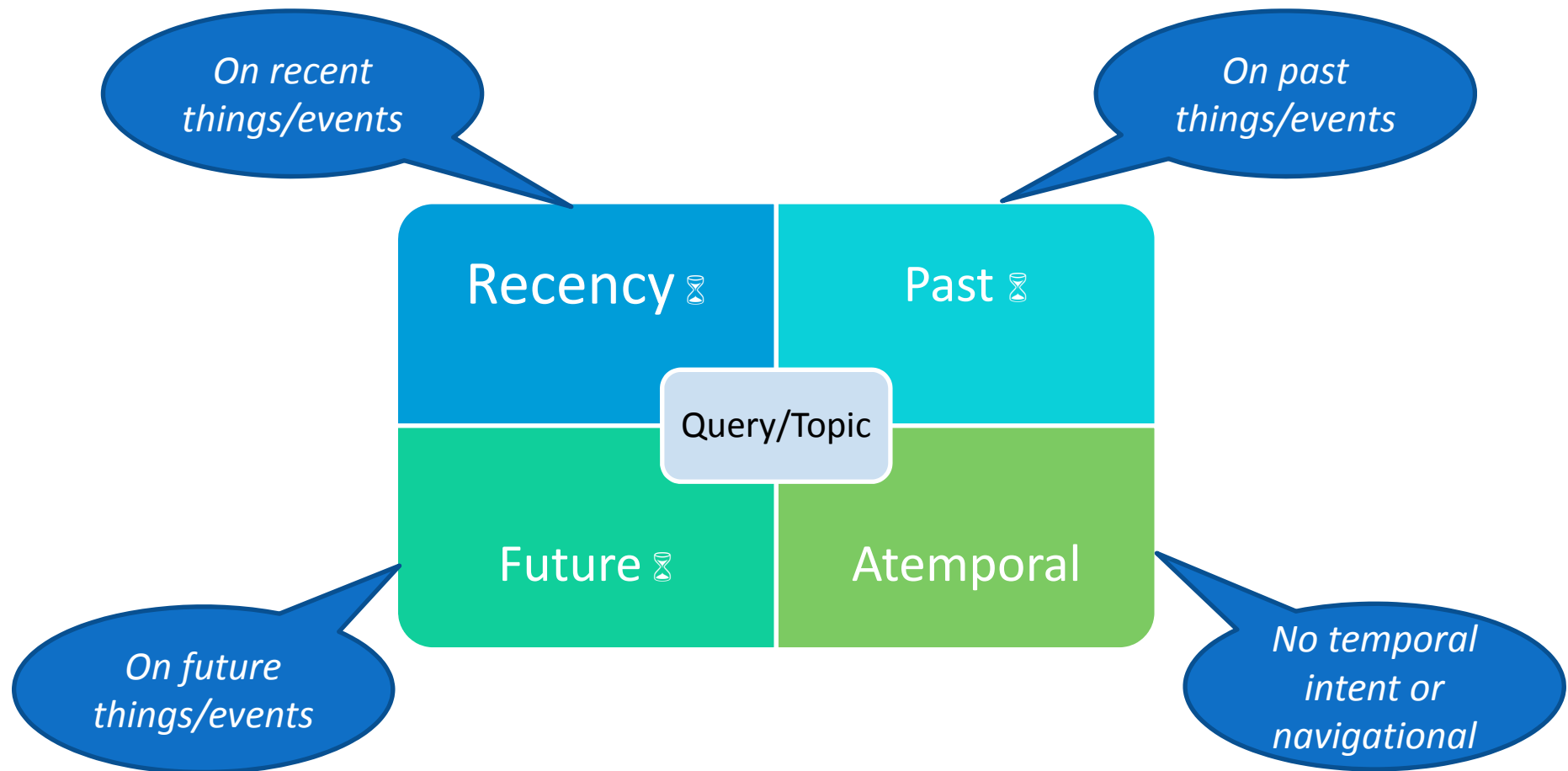# (Current) Temporalia-2: English & Chinese Lang.

1. **Temporal Intent <u>Disambiguation</u> (TID)**

   - For a given query, determine probability distribution over temporal intent classes

     - **Input:** query, query timestamp

     - **Output:** temporal intent class distribution

2. **Temporally <u>Diversified</u> Retrieval (TIR)**

   - For a given topic, retrieve documents relevant to each of temporal intent classes

     - **Input:** set of topic descriptions, timestamp and fixed temporal intent class distribution

     - **Output:** diversified ranking of documents

TEMPORALIA⏱

Temporal Classes of Queries/Topics

English Document Collection

- **LivingKnowledge Corpus** [http://livingknowledge.europarchive.org/]

  - Crawl time: April 2011 and March 2013

  - Source: **news** and **blog** feeds

  - Size: 20GB uncompressed (>5GB zipped)

    - 3.8 million documents from 1500 different blogs and news sources

    - Text only

- Provided annotations: Temporal Expressions, Named Entities and Sentence Boundaries

  - Participants can choose to use annotations or not

  - Participants can use external collections such as Wikipedia

Chinese Document Collection

- Dry Run: SogouCA, a collection of **news web pages**
  - Time span: June and July in 2012
  - Source: several popular news websites in China
  - Size: the total uncompressed size is 711MB

- Formal Run: Distilled subset of the **web page corpus** SogouT 2012
  - Crawl time: mid 2012
  - Size: subset of 135.4 million Web pages from 5.3 million Chinese Web sites

- Provided annotations:
  - Similar as in the English collection

Draft Schedule

Important Dates (Tentative)

Feb 28, 2015	NTCIR-12 Kick-off Event
Jun 2015	Final task guideline release, Task Registration due, Document Collection release
Jun 2015	Dry Run: Test topics release
Aug 2015	Dry Run: Run submission due
Sep 2015	Dry Run: Evaluation results returned
Oct 2015	Formal Run: Test topics release
Dec 2015	Formal Run: Run submission due
Feb 2016	Formal Run: Evaluation results returned
Mar 2016	Paper submission due
May 2016	Camera-ready submission due
Jun 2016	NTCIR-12 Conference

TEMPORALIA

Thank you!

Website: *http://ntcirtemporalia.github.io*
Facebook: *https://www.facebook.com/ntcirtemporalia*
Twitter: *https://twitter.com/ntcirtemporalia*
Email: *tc4fia@googlegroups.com*