

# NTCIR-17 QA Lab-PoliInfo-4

## Organizers:

1. Yasuhiro Ogawa, Nagoya University, Japan
2. Yasutomo Kimura, Otaru University of Commerce, Japan
3. Hideyuki Shibuki, Besna Institute Inc.
4. Hokuto Ototake, Fukuoka University, Japan
5. Yuzu Uchida, Hokkai-Gakuen University, Japan
6. Keiichi Takamaru, Utsunomiya Kyowa University, Japan
7. Madoka Ishioroshi, National Institute of Informatics, Japan
8. Kazuma Kadowaki, The Japan Research Institute, Limited
9. Tomoyoshi Akiba, Toyohashi University of Technology, Japan
10. Minoru Sasaki, Ibaraki University, Japan
11. Akio Kobayashi, National Agriculture and Food Research Organization, Japan

## Advisers:

1. Masaharu Yoshioka, Hokkaido University, Japan
2. Tatsunori Mori, Yokohama National University, Japan
3. Kenji Araki, Hokkaido University, Japan
4. Satoshi Sekine, RIKEN, Japan
5. Teruko Mitamura, Carnegie Mellon University, USA



# NTCIR-17 QA Lab-PoliInfo-4

- Background
- Related work
- What is QA Lab PoliInfo-4 ?
- Sub tasks
  1. Question Answering-2
  2. Answer Verification
  3. Stance Classification-2
  4. ~~Budget Argument Mining-2~~ Minutes-to-Budget Linking
- Summary

# Background

- Fake news articles are easily propagated through various online media platforms
- In political news, the one reason is the complexity of political resources
- To deal with such fake news, we have addressed these tasks from NTCIR-14 to NTCIR-16.

# Related Work

	<b>CLEF-2020 Profiling Fake News Spreaders on Twitter</b>	<b>CLEF-2022 CheckThat! Lab</b>	<b>NTCIR 16 QA Lab-PoliInfo-3</b>
Dataset	Twitter	Twitter News articles	Assembly minutes, newsletter
Task	Profiling Fake News Spreaders on Twitter	Task 1 - Check-Worthiness Estimation Task 2 - Verified Claim Retrieval Task 3 - Fake News Detection	Task 1: Question Answering Task 2: QA Alignment Task 3: Fact Verification Task 4: Budget Argument Mining
Language	English and Spanish	Arabic, Bulgarian, Dutch, English, German, Spanish, and Turkish	Japanese

# What is QA Lab-PoliInfo-4 ?

- We aim to resolve issues related to fake news detection and fact checking on political information (minutes and newsletters)
- In particular, we are conscious of accurately dealing with primary information
  1. Question Answering-2 (QA2)
  2. Answer Verification (AV)
  3. Stance Classification-2 (SC2)
  4. ~~Budget Argument Mining-2 (BAM2)~~ Minutes-to-Budget Linking (MBLink)

NTCIR14  
QA Lab PolInfo

NTCIR15  
QA Lab PolInfo-2

NTCIR16  
QA Lab PolInfo-3

NTCIR17  
QA Lab PolInfo-4

Summarization

Dialog  
Summarization

Question Answering

Question  
Answering-2

Segmentation

Topic Detection

QA Alignment

Answer Verification

Classification

Entity Linking

Fact Verification

Stance  
Classification-2

Stance  
Classification

Budget Argument  
Mining

Minutes-to-Budget  
Linking



# Question Answering-2 (QA2)

## Purpose:

- ▶ The second question answering (QA2) subtask is **the same** as the question answering (QA) subtask at NTCIR-16 QA Lab-PoliInfo-3.
- ▶ Collaborate with the answer verification (AV) subtask at NTCIR-17 QA Lab-PoliInfo-4.

## Input:

- ▶ Questions of Q&A sessions in the newsletters
- ▶ The minutes of the Tokyo Metropolitan Assembly

## Output:

- ▶ Short answer of the question in a few lines

## Evaluation:

- ▶ ROUGE (for automatic evaluation)
- ▶ Quality questions (for human evaluation)



# Answer Verification (AV)

Purpose:

- ▶ Answer Verification aims to check whether the results of QA is correct.

Two Stages:

1. Expanding training data including fake answers
  - Participants tackle **to make fake answers** or **seemingly wrong answers** that **baseline classifiers will misjudge**.
2. Building binary classifiers that can judge the test data set correctly
  - These classifiers are used as new baseline classifiers in the 1<sup>st</sup> stage

# AV 1<sup>st</sup> Stage: Fake-answer Generation

## Purpose:

- ▶ The 1<sup>st</sup> stage of AV aims to generate answers difficult for classifiers to judge
- ▶ This idea is from GAN (Generative Adversarial Networks)

## Input:

- ▶ Sets of **a question and its answer** in Q&A sessions in Newsletters of the Tokyo Metropolitan Assembly
- ▶ **The minutes** of the Tokyo Metropolitan Assembly

## Output:

- ▶ **Fake answers** of the question in a few lines
- ▶ Correct but **seemingly wrong answer** the question in a few lines

## Evaluation:

- ▶ Decrease rate of the accuracy of estimated results by fact checking baseline system

# AV 2<sup>nd</sup> Stage: Fact-checking Classification

## Purpose:

- ▶ The 2<sup>nd</sup> stage of AV aims to judge whether the answer is fake

## Input:

- ▶ **Questions** of Q&A sessions in the newsletters
- ▶ List of **correct or fake answers** of the question in a few lines
- ▶ **The minutes** of the Tokyo Metropolitan Assembly

## Output:

- ▶ List of **“True”** or **“False”** (binary) (“True” indicates that the answer is not fake.)

## Evaluation:

- ▶ Accuracy

# Stance Classification-2 (SC2)

## Purpose:

- ▶ The stance classification-2 aims to classify the stance of a political party
- ▶ It expands target local governments (Tokyo, Sapporo... ) and is more difficult than NTCIR15 stance classification-1.

## Input:

- ▶ A **topic**
- ▶ An assembly member's **opinion sentences** on the given topic without explicit statement of neither approval nor disapproval.

## Output:

- ▶ Relevance (Binary) : Either approval or disapproval on a given topic

## Evaluation:

- ▶ Accuracy (for automatic evaluation)

# ~~Budget Argument Mining - 2~~

## Minutes-to-Budget Linking (MBLink)

### Purpose:

- ▶ **Minutes-to-Budget Linking** aims to link minutes to budget tables when a sentence of assembly minutes is given, and extracts the evidence of the discuss.

### Input:

- ▶ Budget tables (HTML); each cell has an ID assigned by the organizer
- ▶ Assembly minutes; each sentence has an ID assigned by the organizer

### Output:

- ▶ For each sentence in the minutes,
  - ✓ “linkedCellIDs”: a list of cell IDs that are related to the sentence.
  - ✓ “containsReason”: whether the sentence contains a reason for the claim.

### Evaluation:

- ▶ “linkedCellIDs” and “containsReason” are evaluated using F-measure.

# NTCIR-17 PolInfo-4 Important Dates

## IMPORTANT DATES

- July 8, 2022 Task Proposal Submission Due
- August 26, 2022 Acceptance Notification of Task Proposals
- September 28, 2022 NTCIR-17 Kickoff Event
- **November 8, 2022** **First table meeting for PolInfo-4**
- January 2023: Dataset Release\*
- March 2023: Second round table meeting
- January-June 2023: Dry Run\*
- June 2023: Release of Formal Run dataset
- June-July 2023: Formal Run\*
- August 1, 2023: Evaluation Result Release
- August 1, 2023: Draft Task Overview Paper Release
- September 1, 2023: Draft Participant Paper Submission Due
- November 1, 2023: All Camera-ready Paper Submission Due
- December 2023: NTCIR-17 Conference in NII, Tokyo, Japan

(\* indicates that the schedule can be different for different tasks)

# What is QA Lab-PoliInfo-4 ?

## Question Answering-2

Input : Question  
Output : Answer

We consider Q&A summary in the newsletter to be the correct answers.

## Stance Classification-2

Input : Topic and Opinion sentences  
Output : Approval or disapproval

Diet or local assembly



Person asking a question



Person answering the question

## Assembly minutes

○議長（石川良一君） これより質問に入ります。  
百十五番小山くにひこ君。  
（百十五番小山くにひこ君登壇）

○百十五番（小山くにひこ君） 令和二年第四回定例会に当たり、都民ファーストの会東京都議団を代表し、小池知事及び教育長、関係局長に質問いたします。

初めに、過日、名簿整理がある議員先生、有馬朗人さんがご逝去されました。ここに謹んで哀悼の意を表し、心よりご冥福をお祈りいたします。

質問に先立ち、新型コロナウイルス感染症によりお亡くなりになられた方々に対しまして、心よりご冥福をお祈り申し上げます。また、今なお療養中の方々におかれましては、一日も早いご回復を祈ります。また、国内外において、新型コロナウイルスの第三波というべき状況が到来しています。東京でも多くの新規陽性者が発生し、重症者数の推移も予断を許さない状況です。医療崩壊を起こさないため、新型コロナウイルス対応に当たってくださっている医療従事者、医療機関への支援を一層強化しながら、これまでの知見を踏まえた、めり張りのついた対策を進め、感染拡大の防止をまずはしっかりと行いながら、社会経済活動との両立を図っていく必要があります。

## Answer Verification

1<sup>st</sup> stage:  
Input : Q&A in newsletters and assembly minutes  
Output : Fake or seemingly wrong answers

2<sup>nd</sup> stage:  
Input : Assembly minutes and questions and questionable answers  
Output : “True” or “False”

## Minutes-to-Budget Linking

Input : Budget tables and assembly minutes  
Output :  
1. Linked cell ID  
2. Presence or absence of reasons