

Overview of the NTCIR-16 QA Lab-PoliInfo-3 Task

Organizers:

1. Yasutomo Kimura, Otaru University of Commerce, Japan
2. Hideyuki Shibuki, National Institute of Informatics, Japan
3. Hokuto Ototake, Fukuoka University, Japan
4. Yuzu Uchida, Hokkai-Gakuen University, Japan
5. Keiichi Takamaru, Utsunomiya Kyowa University, Japan
6. Madoka Ishioroshi, National Institute of Informatics, Japan
7. Kazuma Kadowaki, The Japan Research Institute, Limited
8. Masaharu Yoshioka, Hokkaido University, Japan
9. Tomoyoshi Akiba, Toyohashi University of Technology, Japan
10. Yasuhiro Ogawa, Nagoya University, Japan
11. Minoru Sasaki, Ibaraki University, Japan
12. Kenichi Yokote, Hitachi, Japan

Advisers:

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



Table of contents

- Local assembly minutes
- QA Lab PoliInfo-3
 - Sub tasks
 - Question Answering
 - QA Alignment
 - Fact Verification
 - Budget Argument Mining
- Datasets
- Leaderboard
- Results



[Information](#) [Schedule](#) [Team](#) [Language](#)

NTCIR-16 QA Lab-PoliInfo-3

We designed four tasks
(Question Answering, QA Alignment, Fact Verification, Budget Argument Mining)
with the aim of providing appropriate information for solving political issues.

Tasks

Question Answering

Question Answering aims to generate a summary of the corresponding answers given a summary of the questions in the Minutes.

Dataset specification,
Evaluation methods

QA Alignment

QA Alignment aims to associate each question with an answer when given a question and answer in the form of a batch question and answer.

Dataset specification,
Evaluation methods

Fact Verification

Fact Verification aims to determine when summaries and minutes are given, whether the contents of the summaries really exist in the minutes.

Dataset specification,
Evaluation methods

Budget Argument Mining

Budget Argument Mining aims to extract the part of the meeting that is being discussed when a budget item or amount is given, and determines the argument label (decision, example, fact, etc.) of the content.

Dataset specification,
Evaluation methods

<https://poliinfo3.net/>

Background

- The information that exists on the web is a mixed bag, with various reports, claims, and opinions.
- Fact checking based on primary information is necessary.

What is primary information?

- Japanese local assembly minutes can be used as primary information.
- The minutes are texts that record who said what, when, and where.

What is Japanese local assembly ?

- Regular sessions four times a year
- Plenary sessions and committee meetings are held to discuss bills such as ordinances and budget.

City Assembly Hall



Examples of Tokyo Metropolitan Assembly Minutes

[トップ](#) > [会議の結果と記録](#) > [本会議の会議録](#) > [平成23年第2回定例会](#) > [第九号](#)

平成二十三年東京都議会会議録第九号

平成二十三年六月二十四日（金曜日）

出席議員 百二十六名

- 一番 小林 健二君
二番 加藤 雅之君
三番 小宮あんり君
四番 吉住 健一君
五番 桜井 浩之君
六番 山崎 一輝君
七番 福士 敬子君
八番 土屋たかゆき君
九番 山内れい子君
十一番 小山くにひこ君
十二番 くりした善行君
十三番 西沢けいた君
十四番 田中 健君
十五番 畔上三和子君
十六番 斉藤やすひろ君
十七番 栗林のり子君
十八番 遠藤 守君
十九番 大松あきら君

出席説明員

知事	石原慎太郎君
副知事	佐藤 広君
副知事	猪瀬 直樹君
副知事	吉川 和夫君
副知事	村山 寛司君
教育長	大原 正行君
東京都技監都市整備局長兼務	河島 均君
知事本局長	秋山 俊行君
総務局長	比留間英人君
財務局長	安藤 立美君
警視總監	池田 克彦君
主税局長	荒川 満君
生活文化局長	並木 一夫君
スポーツ振興局長	笠井 謙一君
環境局長	大野 輝之君
福祉保健局長	杉村 栄一君
産業労働局長	前田 信弘君
建設局長	村尾 公一君

午後一時開議

○議長（和田宗春君） これより本日の会議を開きます。

○議長（和田宗春君） この際、あらかじめ会議時間の延長をいたしておきます。

○議長（和田宗春君） 次に、日程の追加について申し上げます。

議員より、議員提出議案第四号、東京都省エネルギーの推進及びエネルギーの安定的な供給の確保に関する条例、知事より、東京都副知事の選任の同意について外人事案件六件がそれぞれ提出されました。これらを本日の日程に追加いたします。

○議長（和田宗春君） 昨日に引き続き質問を行います。

三十一番中村ひろし君。

（三十一番中村ひろし君登壇）

○三十一番（中村ひろし君） 質問に先立ち、東日本大震災で亡くなられた方々に謹んでお悔やみを申し上げますとともに、被災された方々に心からお見舞いを申し上げます。

初めに、都政運営の方針について、都の財政運営について質問します。

今回の議会は震災対策が中心となり、そのための補正予算一千三百七十四億円が議論されます。その財源は、基金からの繰り入れが約半分の七百五十億円となっています。今後、震災による厳しい景気状況のもと、都税収入も不透明です。十一月には東京都防災対応指針が示されるとのことですが、建物の耐震化や津波、高潮対策等にさらに多くの予算が必要になると考えられます。かねてから景気の低迷の影響もあり

Examples of Tokyo Metropolitan Assembly Minutes

平成二十三年東京都議会会議録第九号

平成二十三年六月二十四日（金曜日）

出席議員 百二十六名

When

二番 小宮のひろし君
四番 吉住 健一君
五番 桜井 浩之君
六番 山崎 一輝君
七番 福士 敬子君
八番 土屋たかゆき君
九番 山内れい子君
十一番 小山くにひこ君
十二番 くりした善行君
十三番 西沢けいた君
十四番 田中 健君
十五番 畔上三和子君
十六番 斉藤やすひろ君
十七番 栗林のり子君
十八番 遠藤 守君
十九番 大松あきら君

説明員

知事 石原慎太郎君
副知事 広君
副知事 直樹君
副知事 吉川 和夫君
副知事 村山 寛司君
教育長 大原 正行君
東京都技監都市整備局長兼務 河島 均君
知事本局長 秋山 俊行君
総務局長 比留間英人君
財務局長 安藤 立美君
警視總監 池田 克彦君
主税局長 荒川 満君
生活文化局長 並木 一夫君
スポーツ振興局長 笠井 謙一君
環境局長 大野 輝之君
福祉保健局長 杉村 栄一君
産業労働局長 前田 信弘君
建設局長 村尾 公一君

Where

Who

What

午後一時開議

○議長（和田宗春君）

これより本日の会議を開きます。

○議長（和田宗春君）

この際、あらかじめ会議時間の延長をいたしておきます。

○議長（和田宗春君）

次に、日程の追加について申し上げます。

議員より、議員提出議案第四号、東京都省エネルギーの推進及びエネルギーの安定的な供給の確保に関する条例、知事より、東京都副知事の選任の同意について外人事案件六件がそれぞれ提出されました。これらを本日の日程に追加いたします。

○議長（和田宗春君）

昨日に引き続き質問を行います。

三十一番中村ひろし君。

（三十一番中村ひろし君登壇）

○三十一番（中村ひろし君） 質問に先立ち、東日本大震災で亡くなられた方々に謹んでお悔やみを申し上げますとともに、被災された方々に心からお見舞いを申し上げます。

初めに、都政運営の方針について、都の財政運営について質問します。

今回の議会は震災対策が中心となり、そのための補正予算一千三百七十四億円が議論されます。その財源は、基金からの繰り入れが約半分の七百五億円となっています。今後、震災による厳しい景気状況のもと、都税収入も不透明です。十一月には東京都防災対応指針が示されるとのことですが、建物の耐震化や津波、高潮対策等にさらに多くの予算が必要になると考えられます。かねてから曇りの低気圧の影響もあり

Examples of Tokyo Metropolitan Assembly Minutes

Members

出席議員 百二十六名

一番	小林 健二君
二番	加藤 雅之君
三番	小宮あんり君
四番	吉住 健一君
五番	桜井 浩之君
六番	山崎 一輝君
七番	福士 敬子君
八番	土屋たかゆき君
九番	山内れい子君
十一番	小山くにはこ君
十二番	くりした善行君
十三番	西沢けいた君
十四番	田中 健君
十五番	畔上三和子君
十六番	斉藤やすひろ君
十七番	栗林のり子君
十八番	遠藤 守君
十九番	大松あきら君

Attendance staff

副知事	佐藤 広君
副知事	猪瀬 直樹君
副知事	吉川 和夫君
副知事	村山 寛司君
教育長	大原 正行君
東京都技監都市整備局長兼務	河島 均君
知事本局長	秋山 俊行君
総務局長	比留間英人君
財務局長	安藤 立美君
警視總監	池田 克彦君
主税局長	荒川 満君
生活文化局長	並木 一夫君
スポーツ振興局長	笠井 謙一君
環境局長	大野 輝之君
福祉保健局長	杉村 栄一君
産業労働局長	前田 信弘君
建設局長	村尾 公一君

Speakers include not only members but also governors, vice-governor, and local government staff.

午後一時開議

○議長（和田宗春君） これより本日の会議を開きます。

○議長（和田宗春君） この際、あらかじめ会議時間の延長をいたしておきます。

○議長（和田宗春君） 次に、日程の追加について申し上げます。

議員より、議員提出議案第四号、東京都省エネルギーの推進及びエネルギーの安定的な供給の確保に関する条例、知事より、東京都副知事の選任の同意について外人事案件六件がそれぞれ提出されました。これらを本日の日程に追加いたします。

○議長（和田宗春君） 昨日に引き続き質問を行います。

三十一番中村ひろし君。

（三十一番中村ひろし君登壇）

○三十一番（中村ひろし君） 質問に先立ち、東日本大震災で亡くなられた方々に謹んでお悔やみを申し上げますとともに、被災された方々に心からお見舞いを申し上げます。

初めに、都政運営の方針について、都の財政運営について質問します。

今回の議会は震災対策が中心となり、そのための補正予算一千三百七十四億円が議論されます。その財源は、基金からの繰り入れが約半分の七百五十億円となっています。今後、震災による厳しい景気状況のもと、都税収入も不透明です。十一月には東京都防災対応指針が示されるとのことですが、建物の耐震化や津波、高潮対策等にさらに多くの予算が必要になると考えられます。かねてから景気の低迷の影響もあり

Characteristics of Local Assembly Minutes

A transcript of a speech

○議長（和田宗春君） この際、あらかじめ会議時間の延長を承知していただきます。

○議長（和田宗春君） 次に、日程の追加について申し上げます。

議員より、議員提出議案第四号、東京都省エネルギーの推進及びエネルギーの安定的な供給の確保に関する条例、知事より、東京都副知事の選任の同意について外人事案件六件がそれぞれ提出されました。これらを本日の日程に追加いたします。

Primary information

A speech is very long

Questions and answers are separated

○三十一番（中村ひろし君） 質問に先立ち、東日本大震災で亡くなられた方々に謹んでお悔やみを申し上げますとともに、被災された方々に心からお見舞いを申し上げます。

初めに、都政運営の方針について、都の財政運営について質問します。

今回の議会は震災対策が中心となり、そのための補正予算一千三百七十四億円が議論されます。その財源は、基金からの繰り入れが約半分の七百五億円となっています。今後、震災による厳しい景気状況のもと、都税収入も不透明です。十一月には東京都防災対応指針が示されるとのことですが、建物の耐震化や津波、高潮対策等にさらに多くの予算が必要になると思われます。かねてから景気の低迷の影響もあり、長期的に見るとともに、少子化や高齢化の影響もいたわってきた中、今後は震災への対応も加わることで、

What is QA Lab-PoliInfo-3?

How have we proceeded QA Lab-PoliInfo?

2018.01 → 2019.06 → 2019.07 → 2020.12 → 2021.01 → 2022.06



NTCIR14
QA Lab PoliInfo



NTCIR15
QA Lab PoliInfo-2



NTCIR16
QA Lab PoliInfo-3

Summarization

Dialog Summarization

Question Answering

Segmentation

Topic Detection

QA Alignment

Classification

Entity Linking

Fact Verification

Stance Classification

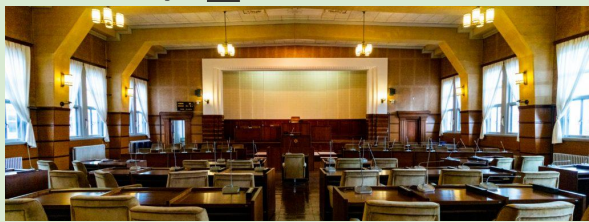
Budget Argument Mining



QA Lab-PoliInfo3

Information published by governments

Minutes (Transcript)



Report
(QA pairs by hand)

Newsletter
(summary by hand)

Budget Information

QA Alignment

Associating each question with its answer in the minutes

Question Answering

Answering a question based on the contents of the minutes

Fact Verification

Verifying the credibility of political claims using predefined primary sources.

Budget Argument Mining

Linking budget items to related argument

Sub tasks in QA Lab-PoliInfo3

- **QA Alignment**

- ▶ Input : Assembly minutes
- ▶ Output : Filled in the value of the QAID field in the minutes
- ▶ Evaluation: F_measure based on number of associated question and answer pairs

- **Question Answering**

- ▶ Input : Summary of questions and utterances of members in assembly minutes
- ▶ Output : A summary of the answers for each question
- ▶ Evaluation: ROUGE-1-F_measure, Mutual evaluation among participants

- **Fact Verification**

- ▶ Input : Summary of utterances and local assembly minutes
- ▶ Output : StartingLine, EndingLine and DocumentEntailment
- ▶ Evaluation : Accuracy

- **Budget Argument Mining**

- ▶ Input : Budget information and assembly minutes
- ▶ Output : Argument Class and related ID
- ▶ Evaluation: F_measure based on both argument class labeling and linking relatedID

Datasets

The dataset can be downloaded from GitHub

The screenshot shows the GitHub repository page for "NTCIR-16 QA Lab-PoliInfo-3". The repository is private and contains several datasets. The repository name is "PoliInfo3-DryRun-QuestionAnswering", which is highlighted with a red box. The repository description is "政治課題を解決する上で適切な情報を提示することを目的とした、4つのタスクを行います" and the URL is "https://poliinfo3.net". The repository is updated 8 days ago and has 0 stars, 0 forks, and 0 pull requests. Other datasets listed include "PoliInfo3-DryRun-BudgetArgumentMining" (updated 8 days ago), "PoliInfo3-DryRun-QAAlignment" (updated 9 days ago), and "PoliInfo3-DryRun-FactVerification" (updated 9 days ago).

The screenshot shows the GitHub repository page for "poliinfo3 / PoliInfo3-DryRun-QuestionAnswering". The repository is private and contains several files. The repository name is "PoliInfo3-DryRun-QuestionAnswering", which is highlighted with a red box. The repository description is "政治課題を解決する上で適切な情報を提示することを目的とした、4つのタスクを行います" and the URL is "https://poliinfo3.net". The repository is updated 8 days ago and has 0 stars, 0 forks, and 0 pull requests. Other datasets listed include "PoliInfo3-DryRun-BudgetArgumentMining" (updated 8 days ago), "PoliInfo3-DryRun-QAAlignment" (updated 9 days ago), and "PoliInfo3-DryRun-FactVerification" (updated 9 days ago).

Schedule

Meeting and dataset release

- **March 24, 2021: QA Lab-PoliInfo-3 first round table meeting**
- March 29, 2021: NTCIR-16 kickoff meeting
- June 15, 2021: QA Lab-PoliInfo-3 second round table meeting
- June 15, 2021: Dataset release

Dry Run

- **August 10--November 12, 2021: Dry run**
- November 01--12, 2021: Evaluation by participants (Question Answering)
- November 15, 2021: Evaluation result release

Formal Run

- November 22, 2021: Update of dataset for formal run
- **November 22--30, 2021: Formal run**
- November 30, 2021: Task registration due for formal run (not required for dry run participants)

NTCIR-16 CONFERENCE

- December 06--17, 2021: Evaluation by participants
- December 18--19, 2021: Evaluation by organizers
- December 20, 2021: Evaluation Result Release
- February 01, 2021: Task overview paper release (draft)
- March 01, 2022: Submission due for participant papers
- May 01, 2022: Camera-ready participant paper due
- June 14--17, 2022: NTCIR-16 Conference

Dataset

- **QA Alignment**
 - ▶ Training data : **143,798** utterances 2011- 2015
 - ▶ Test data : **24,302** utterances 2019
 - ▶ References : Tokyo Metropolitan Assembly Minutes
- **Question Answering**
 - ▶ Training data : **2,765** questions
 - ▶ Test data : **391** questions
 - ▶ Reference : Tokyo Metropolitan Assembly Minutes
- **Fact Verification**
 - ▶ Training data : **1,024** summary sentences
 - ▶ Test data : **298** summary sentences
 - ▶ References : Tokyo Metropolitan Assembly Minutes
- **Budget Argument Mining**
 - ▶ Training data : **1,248** money expressions
 - ▶ Test data : **520** money expressions
 - ▶ References : Budget and Assembly minutes (Diet, Otaru city, Ibaraki prefecture and Fukuoka city)

Participants

Participants

1. 10807010 Tokyo Institute of Technology
2. AKBL* Toyohashi University of Technology
3. ditlab Denso IT Laboratory
4. Forst* Yokohama National University
5. fuys* Fukuoka University
6. Ibrk* Ibaraki University
7. JRIRD* The Japan Research Institute, Limited
8. nukl* Nagoya University
9. OUC* Otaru University of Commerce
10. rVRAIN Universitat Politècnica de València
11. SMLAB National Agriculture and Food Research Organization & The University of Tokyo
12. takelab Osaka Electro-Communication University
13. TO*

* Task organizers are in the team

Leaderboard

Dashboard for NTCIR-16 QA Lab-PoliInfo-3



ログイン・参加申し込み

Login

Registration

Linking your github account

QUESTION ANSWERING

QA ALIGNMENT

FACT VERIFICATION

BUDGET ARGUMENT MINING

Login

Access token key

ログイン

Leaderboard

Leaderboard for NTCIR-16 QA Lab-PoliInfo-3

Dashboard for NTCIR-16 QA Lab-PoliInfo-3

Formal run

Dry run

Late submissions

ログイン・参加申し込み

QUESTION ANSWERING

QA ALIGNMENT

FACT VERIFICATION

BUDGET ARGUMENT MINING






概要・提出

FORMAL RUN

DRYRUN V20211027

DRYRUN V20210728

LATE SUBMISSIONS

ID	Team Name	Download	Submitted at	Score ↓	AC	RID	More...
299	JRIRD		2021/11/30 17:30	0.5106	0.5827	0.6170	▼
302	JRIRD		2021/11/30 18:05	0.4894	0.5654	0.6170	▼
300	OUC		2021/11/30 17:55	0.4468	0.5712	0.6596	▼
309	OUC		2021/11/30 22:01	0.4255	0.5385	0.6596	▼
263	OUC		2021/11/28 20:05	0.4255	0.4827	0.6596	▼

Leaderboard

Leaderboard for NTCIR-16 QA Lab-PoliInfo-3

Dashboard for NTCIR-16 QA Lab-PoliInfo-3

ログイン・参加申し込み

QUESTION ANSWERING

QA ALIGNMENT

FACT VERIFICATION

BUDGET ARGUMENT MINING






概要・提出

FORMAL RUN

DRYRUN V20211027

DRYRUN V20210728

LATE SUBMISSIONS

ID	Team Name	Download	Submitted at	Score ↓	AC	RID	More...
299	JRIRD		2021/11/30 17:30	0.5106	0.5827	0.6170	▼
302	JRIRD		2021/11/30 18:05	0.4894	0.5154	0.6170	▼
300	OUC		2021/11/30 17:55	0.4453	0.5112	0.6596	▼
309	OUC		2021/11/30 22:01	0.4255	0.5385	0.6596	▼
263	OUC		2021/11/28 20:05	0.4255	0.4827	0.6596	▼

Leaderboard

Number of Submissions in formal run

Table 10: Number of submissions in formal run

Team	QAA	QA	FV	BAM	Total
ditlab	32 (+1)	22	-	-	54 (+1)
OUC	-	-	-	21 (+4)	21 (+4)
AKBL	11	2	8	-	21
Forst	11	-	2 (+4)	-	13 (+4)
nukl	-	8 (+1)	-	-	8 (+1)
rVRain	-	-	-	7	7
SMLAB	-	-	-	6	6
takelab	-	-	-	5	5
10807010	-	-	4	-	4
JRIRD	-	-	-	3	3
fuys	-	-	-	1 (+1)	1 (+1)
Ibrk	-	-	-	- (+11)	- (+11)
Subtotal	54 (+1)	32 (+1)	14 (+4)	43 (+16)	143 (+22)
TO	1	1	1	1	4
Total	55 (+1)	33 (+1)	15 (+4)	44 (+16)	147 (+22)

QA Alignment

QA Alignment in QA Lab-PoliInfo3

- Purpose
 - The aim of the QA Alignment task is to associate each question with its answer in the minutes.
 - In QA Alignment, the goal is to align member's question with its corresponding answer from a governor or superintendent.
- Input
 - The minutes of a Japanese local assembly resemble a transcript.
 - In a question and answer session, an assembly member asks several questions at a time, and a prefectural governor or a superintendent answers the questions.
- Output
 - Filled in the value of the QAID field in the minutes
- Evaluation
 - F_{measure} based on number of associated question and answer pairs

Scores of QA Alignment subtask in formal run

ID	Team	F-Measure	Precision	Recall					
314	ditlab	0.8329	0.8703	0.8037	284	Forst	0.7753	0.7670	0.7883
286	ditlab	<u>0.8324</u>	<u>0.8691</u>	0.8038	294	AKBL	0.7750	0.8159	0.7450
281	ditlab	0.8298	0.8664	0.8005	198	Forst	0.7746	0.7854	0.7716
245	ditlab	0.8289	0.8651	0.8000	197	Forst	0.7746	0.7854	0.7716
259	ditlab	0.8289	0.8651	0.8000	210	ditlab	0.7718	0.7614	0.7875
282	ditlab	0.8275	0.8502	0.8096	283	Forst	0.7703	0.7615	0.7837
285	ditlab	0.8264	0.8650	0.7956	261	Forst	0.7703	0.7615	0.7837
237	ditlab	<u>0.8224</u>	0.8559	0.7959	262	Forst	0.7699	0.7594	0.7852
227	ditlab	0.8224	0.8556	0.7961	290	Forst	0.7662	0.7803	0.7608
267	ditlab	0.8221	0.8562	0.7955	218	ditlab	0.7649	0.7316	0.8085
243	ditlab	0.8219	0.8555	0.7954	199	ditlab	0.7638	0.7661	0.7659
242	ditlab	0.8190	0.8542	0.7913	216	AKBL	0.7591	0.7587	0.7694
226	ditlab	0.8161	0.8476	0.7912	205	Forst	0.7584	0.7774	0.7484
260	ditlab	0.8156	0.8204	0.8134	188	ditlab	0.7574	0.7567	0.7626
221	ditlab	0.8107	0.8376	0.7903	297	AKBL	0.7548	0.8263	0.7016
235	AKBL	0.8098	0.8000	0.8311	220	Forst	0.7522	0.7680	0.7446
241	ditlab	0.8074	0.8053	0.8134	186	ditlab	0.7343	0.7312	0.7427
236	AKBL	0.8050	0.8307	0.7858	184	ditlab	0.7336	0.7258	0.7473
215	AKBL	0.8002	0.7780	<u>0.8354</u>	185	ditlab	0.7322	0.7244	0.7454
182	AKBL	0.7931	0.7615	0.8435	196	Forst	0.7317	0.7366	0.7340
180	AKBL	0.7917	0.7990	0.7903	194	ditlab	0.6897	0.6646	0.7255
213	ditlab	0.7870	0.7970	0.7816	168	ditlab	0.6859	0.6721	0.7097
206	ditlab	0.7857	0.7974	0.7791	169	ditlab	0.6772	0.6651	0.6985
204	ditlab	0.7855	0.7938	0.7820	170	ditlab	0.6648	0.6191	0.7411
179	AKBL	0.7855	0.7927	0.7873	171	ditlab	0.6370	0.5956	0.7035
181	AKBL	0.7826	0.8001	0.7728	167	TO	0.6166	0.5991	0.6437
178	AKBL	0.7823	0.7598	0.8175	173	ditlab	0.5392	0.4887	0.6250
					195	Forst	0.0000	0.0000	0.0000

Question Answering

Question Answering

- Purpose
 - The purpose of the Question Answering task is to answer a question based on the contents of the minutes.
 - The goal is to identify question utterances similar to the input question and return a summarization of its answer utterances.
- Input : Summary of questions and utterances of members in assembly minutes
- Output : A summary of the answers for each question
- Evaluation : **Automatic evaluation** (ROUGE-1-F_measure), **Mutual evaluation** among participants

Table 5: Statistics of data for the Question Answering subtask

Dataset		Speeches		Sentences		Summaries	Date
		Question	Answer	Question	Answer		
Dry run	Train	611	1,758	60,979	28,283	2,765	June 2011 – December 2018
	Test	85	260	9,068	4,514	391	February 2019 – December 2019
Formal run	Train	1,465	4,842	150,194	72,128	7,627	September 2001 – December 2019
	Test	93	272	9,205	4,697	416	February 2020 – December 2020

Scores of Question Answering subtask in formal run (ROUGE scores)

Automatic evaluation

We consider the answer summary in **Togikaidayori(newsletter)** as the gold standard and calculated **ROUGE scores** (ROUGE-1 F-measure of content words on the leaderboard)

ID	Team	ROUGE (Recall)			ROUGE (F-measure)		
		N1	N2	R	N1	N2	R
Surface form							
310	nukl	0.4811	<u>0.2601</u>	<u>0.4277</u>	<u>0.4629</u>	<u>0.2501</u>	0.4112
313	nukl	0.4826	0.2616	0.4288	0.4631	0.2505	<u>0.4110</u>
311	nukl	0.4778	0.2499	0.4216	0.4571	0.2388	<u>0.4030</u>
288	ditlab	<u>0.4826</u>	0.2556	0.4256	0.4518	0.2364	0.3983
269	ditlab	0.4666	0.2443	0.4118	0.4485	0.2341	0.3962
271	ditlab	0.4651	0.2439	0.4106	0.4476	0.2340	0.3955
270	ditlab	0.4641	0.2424	0.4091	0.4461	0.2322	0.3936
273	ditlab	0.4834	0.2542	0.4252	0.4500	0.2335	0.3954
266	nukl	0.4508	0.2325	0.3988	0.4387	0.2240	0.3877
304	nukl	0.4554	0.2328	0.4012	0.4348	0.2209	0.3826
268	ditlab	0.4462	0.2235	0.3909	0.4263	0.2119	0.3740
258	nukl	0.4190	0.2068	0.3671	0.4225	0.2075	0.3708
291	ditlab	0.4560	0.2289	0.3994	0.4282	0.2112	0.3744
253	nukl	0.4411	0.2192	0.3885	0.4137	0.2016	0.3629
289	ditlab	0.4516	0.2200	0.3935	0.4187	0.2004	0.3648
225	ditlab	0.4339	0.2077	0.3793	0.4117	0.1978	0.3606
249	ditlab	0.4339	0.2077	0.3793	0.4117	0.1978	0.3606

ID	Team	ROUGE (Recall)			ROUGE (F-measure)		
		N1	N2	R	N1	N2	R
Surface form							
238	ditlab	0.4325	0.2093	0.3801	0.4083	0.1972	0.3588
222	ditlab	0.4415	0.2072	0.3832	0.4087	0.1925	0.3548
240	ditlab	0.4243	0.1999	0.3695	0.3979	0.1894	0.3472
246	ditlab	0.4259	0.1997	0.3722	0.3970	0.1853	0.3475
255	ditlab	0.4169	0.1907	0.3630	0.3956	0.1815	0.3453
244	ditlab	0.4229	0.1910	0.3629	0.4073	0.1862	0.3502
190	AKBL	0.4529	0.2096	0.3841	0.3850	0.1771	0.3267
228	ditlab	0.4448	0.1990	0.3819	0.4042	0.1813	0.3465
231	ditlab	0.4448	0.1990	0.3819	0.4042	0.1813	0.3465
256	ditlab	0.4207	0.1958	0.3679	0.3958	0.1832	0.3460
189	AKBL	0.3954	0.1843	0.3455	0.3808	0.1756	0.3325
223	ditlab	0.4224	0.1891	0.3616	0.4006	0.1798	0.3429
295	nukl	0.3824	0.1561	0.3301	0.3636	0.1472	0.3126
296	ditlab	0.3861	0.1519	0.3296	0.3466	0.1339	0.2948
293	ditlab	0.3623	0.1255	0.3036	0.3234	0.1106	0.2705
166	TO	0.2628	0.0482	0.2187	0.2422	0.0439	0.2001

Scores of Question Answering subtask in formal run (human evaluation results)

Manual evaluation

Each participant evaluated the results, including the other participants' results as well as summaries from **Togikaidayori**, in the following four aspects and gave a grade of A, B, or C, with A being the highest and C being the lowest.

ID	Team	Correspondence				Content				Well-formed				Overall			
		A	B	C	Score	A	B	C	Score	A	B	C	Score	A	B	C	Score
Gold		377	20	3	774	208	170	22	586	391	8	1	790	217	164	19	598
310	nukl	363	25	12	751	138	211	51	487	381	19	0	<u>781</u>	148	203	49	499
288	ditlab	348	33	19	<u>729</u>	138	200	62	<u>476</u>	379	17	4	775	142	200	58	<u>484</u>
269	ditlab	346	31	23	723	129	209	62	467	384	16	0	784	136	207	57	479
190	AKBL	320	42	38	682	104	196	100	404	381	6	13	768	103	203	94	409
166	TO	83	77	240	243	4	58	338	66	99	33	268	231	4	36	360	44

Fact Verification

Statistics of data for the Fact Verification subtask

The Fact Verification dataset consists of two types of files.

ID	Identification code
Line	Line number
Prefecture	Prefecture name
Volume	Volume
Number	Number
Year	Year
Month	Month
Day	Date
Title	Title
Speaker	Speaker
Utterance	Utterance

Table 6: Statistics of data for the Fact Verification subtask

Dataset		Truth	Misinformation
Dry run	Train	596	428
	Test	166	132
Formal run	Train	596	427
	Test	226	184

ID	Number that uniquely identifies the claim
Prefecture	Meeting location
Date	Date of the meeting
Meeting	Title of the meeting
Speaker	Speaker name of the utterance
UtteranceSummary	Summary of the utterance
UtteranceType	Type of utterance (question or answer)
ContextSummary	Summary of entire dialog before and after the utterance
ContextWord	Topic word related to the utterance
RelatedUtteranceSummary	Another utterance related to the utterance. Example: When "UtteranceType" is "answer", "RelatedUtteranceSummary" is a summary of the question from which the answer was based.
StartingLine	Target value of the task. "Line" of the predefined primary source corresponding to "UtteranceSummary". This field is -1 when "DocumentEntailment" is false.
EndingLine	Target value of the task. "Line" of the predefined primary source corresponding to "UtteranceSummary". This field is -1 when "DocumentEntailment" is false.
DocumentEntailment	Target value of the task (whether or not the claim is credible)

Results

Table 14: Scores of Fact Verification subtask in formal run

ID	Team	F-Measure	Precision	Recall
232	AKBL	0.8892	0.8951	0.9030
307	AKBL	<u>0.8874</u>	<u>0.8930</u>	0.9030
306	AKBL	0.8866	0.8917	0.9030
203	AKBL	0.8608	0.8668	0.8718
202	AKBL	0.8506	0.8559	0.8610
292	Forst	0.8389	0.8466	0.8451
201	AKBL	0.8098	0.8139	0.8238
200	AKBL	0.8098	0.8139	0.8238
257	Forst	0.8040	0.8113	0.8110
272	10807010	0.7963	0.8014	0.8329
191	10807010	0.7876	0.7984	0.8199
192	10807010	0.7822	0.7899	0.8146
274	10807010	0.7734	0.7808	0.8098
229	AKBL	0.4853	0.4917	0.4866
165	TO	0.4488	0.4488	0.4488

$$StartGS = \{startgs_1, \dots, startgs_N\}$$

$$EndGS = \{endgs_1, \dots, endgs_N\}$$

$$LabelGS = \{lgs_1, \dots, lgs_N\}$$

$$StartPRED = \{startpred_1, \dots, startpred_N\}$$

$$EndPRED = \{endpred_1, \dots, endpred_N\}$$

$$LabelPRED = \{lpred_1, \dots, lpred_N\}.$$

We calculated the scores using the following equation:

$$Recall = \frac{1}{N} \sum_i \frac{lineoverlap(i)}{endgs_i - startgs_i + 1}$$

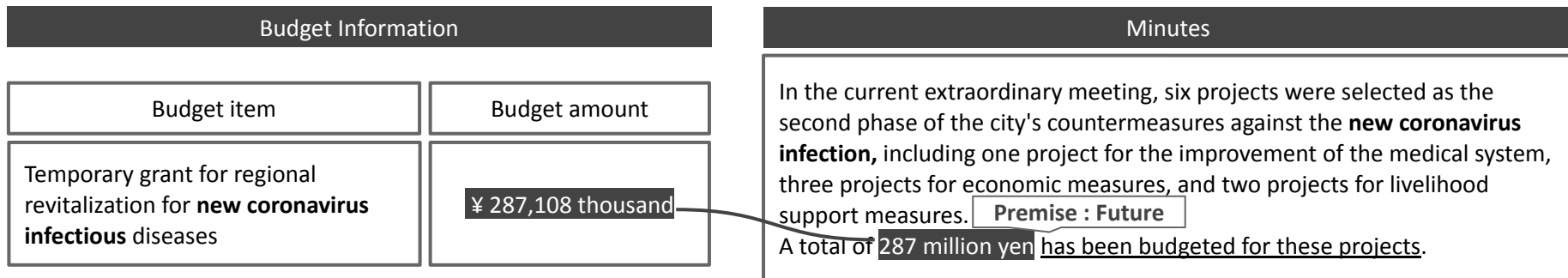
$$Precision = \frac{1}{N} \sum_i \frac{lineoverlap(i)}{endpred_i - startpred_i + 1}$$

$$F1 = \frac{1}{N} \sum_i Hm \left(\frac{lineoverlap(i)}{endgs_i - startgs_i + 1}, \frac{lineoverlap(i)}{endpred_i - startpred_i + 1} \right)$$

Budget Argument Mining

Budget Argument Mining

The goal of Budget Argument Mining is to identify argumentative components related to a budget item and then classify these argumentative components on the basis of their argumentative roles when budget information and minutes are given.



Input

1. Budget Information
 2. Minutes
- (1) National Diet
(2) Local Assembly

Output

1. Argument Classification
2. Related IDs Linking

Scores of Budget Argument Mining subtask in formal run

ID	Team	Score	AC	RID					
299	JRIRD	0.51064	0.58269	0.61702	234	OUC	0.08511	0.47500	0.17021
302	JRIRD	<u>0.48936</u>	0.56538	0.61702	233	OUC	0.08511	0.42308	0.17021
300	OUC	0.44681	<u>0.57115</u>	0.65957	219	OUC	0.08511	0.37308	0.17021
309	OUC	0.42553	0.53846	0.65957	211	OUC	0.08511	0.37308	0.17021
263	OUC	0.42553	0.48269	0.65957	212	OUC	0.08511	0.37308	0.14894
308	OUC	0.40426	0.56154	0.65957	176	rVRAIN	0.06383	0.48462	0.21277
305	OUC	0.40426	0.55769	0.65957	187	rVRAIN	0.06383	0.42692	0.21277
303	JRIRD	0.40426	0.54423	0.61702	298	rVRAIN	0.04255	0.48462	0.25532
251	OUC	0.40426	0.49038	0.65957	312	takelab	0.04255	0.39423	0.06383
252	OUC	0.40426	0.48846	0.65957	280	takelab	0.04255	0.39423	0.04255
250	OUC	0.40426	0.49038	0.57447	174	rVRAIN	0.00000	0.48462	0.17021
301	OUC	0.38298	0.47500	0.65957	183	OUC	0.00000	0.37308	0.12766
277	OUC	0.38298	0.42308	0.65957	209	takelab	0.00000	0.39423	0.00000
248	OUC	0.38298	0.49038	0.55319	208	takelab	0.00000	0.39423	0.00000
278	OUC	0.29787	0.46154	0.65957	279	takelab	0.00000	0.39423	0.00000
224	fuys	0.23404	0.56923	0.34043	276	SMLAB	0.00000	0.38269	0.00000
239	rVRAIN	0.17021	0.47885	0.21277	265	SMLAB	0.00000	0.38269	0.00000
287	rVRAIN	0.14894	0.47885	0.25532	275	SMLAB	0.00000	0.38269	0.00000
230	OUC	0.14894	0.49038	0.17021	175	SMLAB	0.00000	0.38269	0.00000
177	OUC	0.12766	0.37308	0.21277	217	OUC	0.00000	0.37308	0.00000
254	rVRAIN	0.10638	0.47885	0.17021	172	SMLAB	0.00000	0.35962	0.00000
					193	SMLAB	0.00000	0.32885	0.00000
					164	TO	0.00000	0.13462	0.00000

$$\text{Score} = \frac{1}{|S_{RID}|} \sum_{x,y \in S_{RID}} \{ACC(x,y) \times RIDC(x,y)\}.$$

x and y are the labels given to the same monetary expression of the system output and the gold standard data, respectively. S_{RID} is a set of monetary expressions in the gold standard data whose RIDs are not null, as shown in the following equation:

$$S_{RID} = \{y | y.RIDs \neq null\}.$$

ACC indicates whether the AC of a monetary expression is correct or not, as shown in the following equation:

$$ACC(x,y) = \begin{cases} 0 & (x.AC \neq y.AC) \\ 1 & (x.AC = y.AC). \end{cases}$$

$RIDC$ indicates whether an RID output by the system is included in the RIDs of the gold standard data or not:

$$RIDC(x,y) = \begin{cases} 0 & (x.RID \notin y.RIDs) \\ 1 & (x.RID \in y.RIDs). \end{cases}$$

Summary

- Local assembly minutes
- QA Lab PoliInfo-3
 - Sub tasks
 - Question Answering
 - QA Alignment
 - Fact Verification
 - Budget Argument Mining
- Datasets
- Leaderboard
- Results

Participants

1. 10807010 Tokyo Institute of Technology
2. AKBL* Toyohashi University of Technology
3. ditlab Denso IT Laboratory
4. Forst* Yokohama National University
5. fuys* Fukuoka University
6. Ibrk* Ibaraki University
7. JRIRD* The Japan Research Institute, Limited
8. nukl* Nagoya University
9. OUC* Otaru University of Commerce
10. rVRAIN Universitat Politècnica de València
11. SMLAB National Agriculture and Food Research Organization & The University of Tokyo
12. takelab Osaka Electro-Communication University
13. TO*

2018.01 → 2019.06



NTCIR14
QA Lab PoliInfo

2019.07 → 2020.12



NTCIR15
QA Lab PoliInfo-2

2021.01 → 2022.06



NTCIR16
QA Lab PoliInfo-3