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

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.
午後一時開議

○議長（和田宗春君） これより本日の会議を聞きます。

○議長（和田宗春君） これより、この際、あらかじめ会議時間の延長をいただきます。

○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、東京都電エネルギーの推進及びエネルギーの安定的供給の確保に関する条項、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） 今年の引き続き質問を行います。
三十分中村ひろし君
（三十分中村ひろし君登壇）

○三十分中村ひろし君 質問に先立ち、東日本大震災で亡くなった方々に黙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、都財政の状況について、都の財政状況について質問します。
今回の議題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。それ以外は、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、都財政も緊迫です。十一月には東京都防災対策指針が示されることで、建物の耐震化や、防災面での対策がさらに多くの留意が要されると思うおもいます。かねてより景気の低迷の影響もあり、
平成三十三年東京都議会議錄第九号

午後一時開議

○議長（和田宗春君） これより本日の議事を開きます。

○議長（和田宗春君） こと際、あらかじめ被災の場合の延長をいたしております。

○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、議案提出第四号、東京都防災業務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項、財務より、東京都防災の選任の同意について大審案件件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を悼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。

初めて、都防災の方針について、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一千三十七億四千数億円の審議されます。その財源は、基金からの繰り入れが約全体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、必要最小限で不透明です。十一月には東京都防災対策が示されることがありますが、建物の耐震化や消火、防犯対策等にさらに多くの苦労があることでしょうから、これから震災の影響の影響を認められることを願っております。
Speakers include not only members but also governors, vice-governor, and local government staff.
Characteristics of Local Assembly Minutes

A transcript of a speech

A speech is very long
Questions and answers are separated
What is QA Lab-PoliInfo-3?
How have we proceeded QA Lab-PoliInfo?

- **2018.01**
  - NTCIR14 QA Lab PoliInfo
  - Summarization
  - Segmentation
  - Classification

- **2019.06**
  - NTCIR15 QA Lab PoliInfo-2
  - Dialog Summarization
  - Topic Detection
  - Entity Linking
  - Stance Classification

- **2019.07**
  - NTCIR15 QA Lab PoliInfo-2

- **2020.12**
  - NTCIR16 QA Lab PoliInfo-3
  - Question Answering
  - QA Alignment
  - Fact Verification
  - Budget Argument Mining

- **2021.01**

- **2022.06**

---

**QA Lab PoliInfo**

- **NTCIR14**
- **NTCIR15**
- **NTCIR16**

---

**Questions**
- How have we proceeded QA Lab-PoliInfo?
Verifying the credibility of political claims using predefined primary sources.

Linking budget items to related argument

Information published by governments

Minutes (Transcript)

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

Newsletter (summary by hand)

Budget Information

Report (QA pairs by hand)
**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 related ID
Datasets

The dataset can be downloaded from GitHub
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

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
## Dashboard for NTCIR-16 QA Lab-PoliInfo-3

<table>
<thead>
<tr>
<th>Login</th>
<th>Registration</th>
<th>Linking your github account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Login</td>
<td>Access token key</td>
<td></td>
</tr>
</tbody>
</table>

- **Login**
- **Registration**
- **Linking your github account**
<table>
<thead>
<tr>
<th>ID</th>
<th>Team Name</th>
<th>Download</th>
<th>Submitted at</th>
<th>Score</th>
<th>AC</th>
<th>RID</th>
</tr>
</thead>
<tbody>
<tr>
<td>299</td>
<td>JRIRD</td>
<td></td>
<td>2021/11/30 17:30</td>
<td>0.5106</td>
<td>0.5827</td>
<td>0.6170</td>
</tr>
<tr>
<td>302</td>
<td>JRIRD</td>
<td></td>
<td>2021/11/30 18:05</td>
<td>0.4894</td>
<td>0.5654</td>
<td>0.6170</td>
</tr>
<tr>
<td>300</td>
<td>OUC</td>
<td></td>
<td>2021/11/30 17:55</td>
<td>0.4468</td>
<td>0.5712</td>
<td>0.6596</td>
</tr>
<tr>
<td>309</td>
<td>OUC</td>
<td></td>
<td>2021/11/30 22:01</td>
<td>0.4255</td>
<td>0.5385</td>
<td>0.6596</td>
</tr>
<tr>
<td>263</td>
<td>OUC</td>
<td></td>
<td>2021/11/28 20:05</td>
<td>0.4255</td>
<td>0.4827</td>
<td>0.6596</td>
</tr>
</tbody>
</table>
### Leaderboard for NTCIR-16 QA Lab-PoliInfo-3

<table>
<thead>
<tr>
<th>ID</th>
<th>Team Name</th>
<th>Download</th>
<th>Submitted at</th>
<th>Score</th>
<th>AC</th>
<th>RID</th>
<th>Late Submissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>299</td>
<td>JRIRD</td>
<td></td>
<td>2021/11/30 17:30</td>
<td>0.5106</td>
<td>0.5827</td>
<td>0.6170</td>
<td>v</td>
</tr>
<tr>
<td>302</td>
<td>JRIRD</td>
<td></td>
<td>2021/11/30 18:05</td>
<td>0.4894</td>
<td>0.5754</td>
<td>0.6170</td>
<td>v</td>
</tr>
<tr>
<td>300</td>
<td>OUC</td>
<td></td>
<td>2021/11/30 17:55</td>
<td>0.4440</td>
<td>0.5112</td>
<td>0.6596</td>
<td>v</td>
</tr>
<tr>
<td>309</td>
<td>OUC</td>
<td></td>
<td>2021/11/30 22:01</td>
<td>0.4255</td>
<td>0.5385</td>
<td>0.6596</td>
<td>v</td>
</tr>
<tr>
<td>263</td>
<td>OUC</td>
<td></td>
<td>2021/11/28 20:05</td>
<td>0.4255</td>
<td>0.4827</td>
<td>0.6596</td>
<td>v</td>
</tr>
</tbody>
</table>
Table 10: Number of submissions in formal run

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

<table>
<thead>
<tr>
<th>ID</th>
<th>Team</th>
<th>F-Measure</th>
<th>Precision</th>
<th>Recall</th>
<th>284</th>
<th>Forst</th>
<th>0.7753</th>
<th>0.7670</th>
<th>0.7883</th>
</tr>
</thead>
<tbody>
<tr>
<td>314</td>
<td>ditlab</td>
<td>0.8329</td>
<td>0.8703</td>
<td>0.8037</td>
<td>294</td>
<td>AKBL</td>
<td>0.7750</td>
<td>0.8159</td>
<td>0.7450</td>
</tr>
<tr>
<td>286</td>
<td>ditlab</td>
<td>0.8324</td>
<td>0.8691</td>
<td>0.8038</td>
<td>198</td>
<td>Forst</td>
<td>0.7746</td>
<td>0.7854</td>
<td>0.7716</td>
</tr>
<tr>
<td>281</td>
<td>ditlab</td>
<td>0.8298</td>
<td>0.8664</td>
<td>0.8005</td>
<td>197</td>
<td>Forst</td>
<td>0.7746</td>
<td>0.7854</td>
<td>0.7716</td>
</tr>
<tr>
<td>245</td>
<td>ditlab</td>
<td>0.8289</td>
<td>0.8651</td>
<td>0.8000</td>
<td>210</td>
<td>ditlab</td>
<td>0.7718</td>
<td>0.7614</td>
<td>0.7875</td>
</tr>
<tr>
<td>259</td>
<td>ditlab</td>
<td>0.8289</td>
<td>0.8651</td>
<td>0.8000</td>
<td>283</td>
<td>Forst</td>
<td>0.7703</td>
<td>0.7615</td>
<td>0.7837</td>
</tr>
<tr>
<td>282</td>
<td>ditlab</td>
<td>0.8275</td>
<td>0.8502</td>
<td>0.8096</td>
<td>261</td>
<td>Forst</td>
<td>0.7703</td>
<td>0.7615</td>
<td>0.7837</td>
</tr>
<tr>
<td>285</td>
<td>ditlab</td>
<td>0.8264</td>
<td>0.8650</td>
<td>0.7956</td>
<td>262</td>
<td>Forst</td>
<td>0.7699</td>
<td>0.7594</td>
<td>0.7852</td>
</tr>
<tr>
<td>237</td>
<td>ditlab</td>
<td>0.8224</td>
<td>0.8559</td>
<td>0.7959</td>
<td>290</td>
<td>Forst</td>
<td>0.7662</td>
<td>0.7803</td>
<td>0.7608</td>
</tr>
<tr>
<td>227</td>
<td>ditlab</td>
<td>0.8224</td>
<td>0.8556</td>
<td>0.7961</td>
<td>218</td>
<td>ditlab</td>
<td>0.7649</td>
<td>0.7316</td>
<td>0.8085</td>
</tr>
<tr>
<td>267</td>
<td>ditlab</td>
<td>0.8221</td>
<td>0.8562</td>
<td>0.7955</td>
<td>199</td>
<td>ditlab</td>
<td>0.7638</td>
<td>0.7661</td>
<td>0.7659</td>
</tr>
<tr>
<td>243</td>
<td>ditlab</td>
<td>0.8219</td>
<td>0.8555</td>
<td>0.7954</td>
<td>216</td>
<td>AKBL</td>
<td>0.7591</td>
<td>0.7587</td>
<td>0.7694</td>
</tr>
<tr>
<td>242</td>
<td>ditlab</td>
<td>0.8190</td>
<td>0.8542</td>
<td>0.7913</td>
<td>205</td>
<td>Forst</td>
<td>0.7584</td>
<td>0.7774</td>
<td>0.7484</td>
</tr>
<tr>
<td>226</td>
<td>ditlab</td>
<td>0.8161</td>
<td>0.8476</td>
<td>0.7912</td>
<td>188</td>
<td>ditlab</td>
<td>0.7574</td>
<td>0.7567</td>
<td>0.7626</td>
</tr>
<tr>
<td>260</td>
<td>ditlab</td>
<td>0.8156</td>
<td>0.8204</td>
<td>0.8134</td>
<td>297</td>
<td>AKBL</td>
<td>0.7548</td>
<td>0.8263</td>
<td>0.7016</td>
</tr>
<tr>
<td>221</td>
<td>ditlab</td>
<td>0.8107</td>
<td>0.8376</td>
<td>0.7903</td>
<td>220</td>
<td>Forst</td>
<td>0.7522</td>
<td>0.7680</td>
<td>0.7446</td>
</tr>
<tr>
<td>235</td>
<td>AKBL</td>
<td>0.8098</td>
<td>0.8000</td>
<td>0.8311</td>
<td>186</td>
<td>ditlab</td>
<td>0.7343</td>
<td>0.7312</td>
<td>0.7427</td>
</tr>
<tr>
<td>241</td>
<td>ditlab</td>
<td>0.8074</td>
<td>0.8053</td>
<td>0.8134</td>
<td>184</td>
<td>ditlab</td>
<td>0.7336</td>
<td>0.7258</td>
<td>0.7473</td>
</tr>
<tr>
<td>236</td>
<td>AKBL</td>
<td>0.8050</td>
<td>0.8307</td>
<td>0.7858</td>
<td>185</td>
<td>ditlab</td>
<td>0.7322</td>
<td>0.7244</td>
<td>0.7454</td>
</tr>
<tr>
<td>215</td>
<td>AKBL</td>
<td>0.8002</td>
<td>0.7780</td>
<td>0.8354</td>
<td>196</td>
<td>Forst</td>
<td>0.7317</td>
<td>0.7366</td>
<td>0.7340</td>
</tr>
<tr>
<td>182</td>
<td>AKBL</td>
<td>0.7931</td>
<td>0.7615</td>
<td>0.8435</td>
<td>194</td>
<td>ditlab</td>
<td>0.6897</td>
<td>0.6646</td>
<td>0.7255</td>
</tr>
<tr>
<td>180</td>
<td>AKBL</td>
<td>0.7917</td>
<td>0.7990</td>
<td>0.7903</td>
<td>168</td>
<td>ditlab</td>
<td>0.6859</td>
<td>0.6721</td>
<td>0.7097</td>
</tr>
<tr>
<td>213</td>
<td>ditlab</td>
<td>0.7870</td>
<td>0.7970</td>
<td>0.7816</td>
<td>169</td>
<td>ditlab</td>
<td>0.6772</td>
<td>0.6651</td>
<td>0.6985</td>
</tr>
<tr>
<td>206</td>
<td>ditlab</td>
<td>0.7857</td>
<td>0.7974</td>
<td>0.7791</td>
<td>170</td>
<td>ditlab</td>
<td>0.6648</td>
<td>0.6191</td>
<td>0.7411</td>
</tr>
<tr>
<td>204</td>
<td>ditlab</td>
<td>0.7855</td>
<td>0.7938</td>
<td>0.7820</td>
<td>171</td>
<td>ditlab</td>
<td>0.6370</td>
<td>0.5956</td>
<td>0.7035</td>
</tr>
<tr>
<td>179</td>
<td>AKBL</td>
<td>0.7855</td>
<td>0.7927</td>
<td>0.7873</td>
<td>167</td>
<td>TO</td>
<td>0.6166</td>
<td>0.5991</td>
<td>0.6437</td>
</tr>
<tr>
<td>181</td>
<td>AKBL</td>
<td>0.7826</td>
<td>0.8001</td>
<td>0.7728</td>
<td>173</td>
<td>ditlab</td>
<td>0.5392</td>
<td>0.4887</td>
<td>0.6250</td>
</tr>
<tr>
<td>178</td>
<td>AKBL</td>
<td>0.7823</td>
<td>0.7598</td>
<td>0.8175</td>
<td>195</td>
<td>Forst</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Dataset</th>
<th>Speeches</th>
<th>Sentences</th>
<th>Summaries</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Question</td>
<td>Question</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Answer</td>
<td>Answer</td>
<td></td>
</tr>
<tr>
<td>Dry run</td>
<td>Train</td>
<td>611</td>
<td>60,979</td>
<td>2,765</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>85</td>
<td>9,068</td>
<td>391</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,758</td>
<td>28,283</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>260</td>
<td>4,514</td>
<td></td>
</tr>
<tr>
<td>Form run</td>
<td>Train</td>
<td>1,465</td>
<td>150,194</td>
<td>7,627</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>93</td>
<td>9,205</td>
<td>416</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4,842</td>
<td>72,128</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>272</td>
<td>4,697</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>June 2011 – December 2018</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>February 2019 – December 2019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>September 2001 – December 2019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>February 2020 – December 2020</td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>ID</th>
<th>Team</th>
<th>ROUGE (Recall)</th>
<th>ROUGE (F-measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>310</td>
<td>nukl</td>
<td>0.4811</td>
<td>0.2601</td>
</tr>
<tr>
<td>313</td>
<td>nukl</td>
<td>0.4826</td>
<td><strong>0.2616</strong></td>
</tr>
<tr>
<td>311</td>
<td>nukl</td>
<td>0.4778</td>
<td>0.2499</td>
</tr>
<tr>
<td>288</td>
<td>ditlab</td>
<td>0.4826</td>
<td>0.2556</td>
</tr>
<tr>
<td>269</td>
<td>ditlab</td>
<td>0.4666</td>
<td>0.2443</td>
</tr>
<tr>
<td>271</td>
<td>ditlab</td>
<td>0.4651</td>
<td>0.2439</td>
</tr>
<tr>
<td>270</td>
<td>ditlab</td>
<td>0.4641</td>
<td>0.2424</td>
</tr>
<tr>
<td>273</td>
<td>ditlab</td>
<td><strong>0.4834</strong></td>
<td>0.2542</td>
</tr>
<tr>
<td>266</td>
<td>nukl</td>
<td>0.4508</td>
<td>0.2325</td>
</tr>
<tr>
<td>304</td>
<td>nukl</td>
<td>0.4554</td>
<td>0.2328</td>
</tr>
<tr>
<td>268</td>
<td>ditlab</td>
<td>0.4462</td>
<td>0.2235</td>
</tr>
<tr>
<td>258</td>
<td>nukl</td>
<td>0.4190</td>
<td>0.2068</td>
</tr>
<tr>
<td>291</td>
<td>ditlab</td>
<td>0.4560</td>
<td>0.2289</td>
</tr>
<tr>
<td>253</td>
<td>nukl</td>
<td>0.4411</td>
<td>0.2192</td>
</tr>
<tr>
<td>289</td>
<td>ditlab</td>
<td>0.4516</td>
<td>0.2200</td>
</tr>
<tr>
<td>225</td>
<td>ditlab</td>
<td>0.4339</td>
<td>0.2077</td>
</tr>
<tr>
<td>249</td>
<td>ditlab</td>
<td>0.4339</td>
<td>0.2077</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Team</th>
<th>ROUGE (Recall)</th>
<th>ROUGE (F-measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>238</td>
<td>ditlab</td>
<td>0.4325</td>
<td>0.2093</td>
</tr>
<tr>
<td>222</td>
<td>ditlab</td>
<td>0.4415</td>
<td>0.2072</td>
</tr>
<tr>
<td>240</td>
<td>ditlab</td>
<td>0.4243</td>
<td>0.1999</td>
</tr>
<tr>
<td>246</td>
<td>ditlab</td>
<td>0.4259</td>
<td>0.1997</td>
</tr>
<tr>
<td>255</td>
<td>ditlab</td>
<td>0.4169</td>
<td>0.1907</td>
</tr>
<tr>
<td>244</td>
<td>ditlab</td>
<td>0.4229</td>
<td>0.1910</td>
</tr>
<tr>
<td>190</td>
<td>AKBL</td>
<td>0.4529</td>
<td>0.2096</td>
</tr>
<tr>
<td>228</td>
<td>ditlab</td>
<td>0.4448</td>
<td>0.1990</td>
</tr>
<tr>
<td>231</td>
<td>ditlab</td>
<td>0.4448</td>
<td>0.1990</td>
</tr>
<tr>
<td>256</td>
<td>ditlab</td>
<td>0.4207</td>
<td>0.1958</td>
</tr>
<tr>
<td>189</td>
<td>AKBL</td>
<td>0.3954</td>
<td>0.1843</td>
</tr>
<tr>
<td>223</td>
<td>ditlab</td>
<td>0.4224</td>
<td>0.1891</td>
</tr>
<tr>
<td>295</td>
<td>nukl</td>
<td>0.3824</td>
<td>0.1561</td>
</tr>
<tr>
<td>296</td>
<td>ditlab</td>
<td>0.3861</td>
<td>0.1519</td>
</tr>
<tr>
<td>293</td>
<td>ditlab</td>
<td>0.3623</td>
<td>0.1255</td>
</tr>
<tr>
<td>166</td>
<td>TO</td>
<td>0.2628</td>
<td>0.0482</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>ID</th>
<th>Team</th>
<th>Correspondence</th>
<th>Content</th>
<th>Well-formed</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>Score</td>
</tr>
<tr>
<td>310</td>
<td>nukl 363</td>
<td>25</td>
<td>12</td>
<td>751</td>
<td>138</td>
</tr>
<tr>
<td>288</td>
<td>ditlab 348</td>
<td>33</td>
<td>19</td>
<td>729</td>
<td>138</td>
</tr>
<tr>
<td>269</td>
<td>ditlab 346</td>
<td>31</td>
<td>23</td>
<td>723</td>
<td>129</td>
</tr>
<tr>
<td>190</td>
<td>AKBL 320</td>
<td>42</td>
<td>38</td>
<td>682</td>
<td>104</td>
</tr>
<tr>
<td>166</td>
<td>TO 83</td>
<td>77</td>
<td>240</td>
<td>243</td>
<td>4</td>
</tr>
</tbody>
</table>
Fact Verification
The Fact Verification dataset consists of two types of files.

<table>
<thead>
<tr>
<th>ID</th>
<th>Number that uniquely identifies the claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>Line number</td>
</tr>
<tr>
<td>Prefecture</td>
<td>Meeting location</td>
</tr>
<tr>
<td>Volume</td>
<td>Volume</td>
</tr>
<tr>
<td>Number</td>
<td>Number</td>
</tr>
<tr>
<td>Year</td>
<td>Year</td>
</tr>
<tr>
<td>Month</td>
<td>Month</td>
</tr>
<tr>
<td>Day</td>
<td>Date</td>
</tr>
<tr>
<td>Title</td>
<td>Title</td>
</tr>
<tr>
<td>Speaker</td>
<td>Speaker name of the utterance</td>
</tr>
<tr>
<td>Utterance</td>
<td>Utterance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Identification code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>Line number</td>
</tr>
<tr>
<td>Prefecture</td>
<td>Prefecture name</td>
</tr>
<tr>
<td>Volume</td>
<td>Volume</td>
</tr>
<tr>
<td>Number</td>
<td>Number</td>
</tr>
<tr>
<td>Year</td>
<td>Year</td>
</tr>
<tr>
<td>Month</td>
<td>Month</td>
</tr>
<tr>
<td>Day</td>
<td>Date</td>
</tr>
<tr>
<td>Title</td>
<td>Title</td>
</tr>
<tr>
<td>Speaker</td>
<td>Speaker</td>
</tr>
<tr>
<td>Utterance</td>
<td>Utterance</td>
</tr>
</tbody>
</table>

### Table 6: Statistics of data for the Fact Verification subtask

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Truth</th>
<th>Misinformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry run</td>
<td>Train</td>
<td>596</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td></td>
<td>428</td>
</tr>
<tr>
<td></td>
<td>Train</td>
<td>596</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>184</td>
</tr>
<tr>
<td>Formal run</td>
<td>Train</td>
<td>596</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td></td>
<td>427</td>
</tr>
<tr>
<td></td>
<td>Train</td>
<td>596</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>184</td>
</tr>
</tbody>
</table>

- **ID**: Identification code
- **Line**: Line number
- **Prefecture**: Meeting location
- **Volume**: Volume
- **Number**: Number
- **Year**: Year
- **Month**: Month
- **Day**: Date
- **Title**: Title
- **Speaker**: Speaker name of the utterance
- **Utterance**: 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)
Table 14: Scores of Fact Verification subtask in formal run

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

\[
\begin{align*}
\text{StartGS} & = \{\text{startgs}_1, \ldots, \text{startgs}_N\} \\
\text{EndGS} & = \{\text{endgs}_1, \ldots, \text{endgs}_N\} \\
\text{LabelGS} & = \{\text{lg}_1, \ldots, \text{lg}_N\} \\
\text{StartPRED} & = \{\text{startpred}_1, \ldots, \text{startpred}_N\} \\
\text{EndPRED} & = \{\text{endpred}_1, \ldots, \text{endpred}_N\} \\
\text{LabelPRED} & = \{\text{lpred}_1, \ldots, \text{lpred}_N\}.
\end{align*}
\]

We calculated the scores using the following equation:

\[
\text{Recall} = \frac{1}{N} \sum_{i} \frac{\text{lineoverlap}(i)}{\text{endgs}_i - \text{startgs}_i + 1}
\]

\[
\text{Precision} = \frac{1}{N} \sum_{i} \frac{\text{lineoverlap}(i)}{\text{endpred}_i - \text{startpred}_i + 1}
\]

\[
F1 = \frac{1}{N} \sum_{i} Hm\left(\frac{\text{lineoverlap}(i)}{\text{endgs}_i - \text{startgs}_i + 1}, \frac{\text{lineoverlap}(i)}{\text{endpred}_i - \text{startpred}_i + 1}\right)
\]
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.

In the current extraordinary meeting, six projects were selected as the second phase of the city's countermeasures against the new coronavirus infection, including one project for the improvement of the medical system, three projects for economic measures, and two projects for livelihood support measures. A total of 287 million yen has been budgeted for these projects.

<table>
<thead>
<tr>
<th>Budget Information</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporary grant for regional revitalization for new coronavirus infectious diseases</td>
<td>Premise: Future</td>
</tr>
<tr>
<td>¥ 287,108 thousand</td>
<td>A total of 287 million yen has been budgeted for these projects.</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>ID</th>
<th>Team</th>
<th>Score</th>
<th>AC</th>
<th>RID</th>
<th>Score</th>
<th>AC</th>
<th>RID</th>
<th>Score</th>
<th>AC</th>
<th>RID</th>
</tr>
</thead>
<tbody>
<tr>
<td>299</td>
<td>JRIRD</td>
<td>0.51064</td>
<td>0.58269</td>
<td>0.61702</td>
<td>233</td>
<td>OUC</td>
<td>0.08511</td>
<td>0.47500</td>
<td>0.17021</td>
<td></td>
</tr>
<tr>
<td>302</td>
<td>JRIRD</td>
<td>0.48936</td>
<td>0.56538</td>
<td>0.61702</td>
<td>219</td>
<td>OUC</td>
<td>0.08511</td>
<td>0.42308</td>
<td>0.17021</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>OUC</td>
<td>0.44681</td>
<td>0.57115</td>
<td>0.65957</td>
<td>211</td>
<td>OUC</td>
<td>0.08511</td>
<td>0.37308</td>
<td>0.17021</td>
<td></td>
</tr>
<tr>
<td>309</td>
<td>OUC</td>
<td>0.42553</td>
<td>0.53846</td>
<td>0.65957</td>
<td>221</td>
<td>OUC</td>
<td>0.08511</td>
<td>0.37308</td>
<td>0.14894</td>
<td></td>
</tr>
<tr>
<td>263</td>
<td>OUC</td>
<td>0.42553</td>
<td>0.48269</td>
<td>0.65957</td>
<td>176</td>
<td>rVRAIN</td>
<td>0.06383</td>
<td>0.48462</td>
<td>0.21277</td>
<td></td>
</tr>
<tr>
<td>308</td>
<td>OUC</td>
<td>0.40426</td>
<td>0.56154</td>
<td>0.65957</td>
<td>187</td>
<td>rVRAIN</td>
<td>0.06383</td>
<td>0.42692</td>
<td>0.21277</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>OUC</td>
<td>0.40426</td>
<td>0.55769</td>
<td>0.65957</td>
<td>298</td>
<td>rVRAIN</td>
<td>0.04255</td>
<td>0.48462</td>
<td>0.25532</td>
<td></td>
</tr>
<tr>
<td>303</td>
<td>JRIRD</td>
<td>0.40426</td>
<td>0.54423</td>
<td>0.61702</td>
<td>312</td>
<td>takelab</td>
<td>0.04255</td>
<td>0.39423</td>
<td>0.06383</td>
<td></td>
</tr>
<tr>
<td>251</td>
<td>OUC</td>
<td>0.40426</td>
<td>0.49038</td>
<td>0.65957</td>
<td>280</td>
<td>takelab</td>
<td>0.04255</td>
<td>0.39423</td>
<td>0.04255</td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>OUC</td>
<td>0.40426</td>
<td>0.48846</td>
<td>0.65957</td>
<td>174</td>
<td>rVRAIN</td>
<td>0.00000</td>
<td>0.48462</td>
<td>0.17021</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>OUC</td>
<td>0.40426</td>
<td>0.49038</td>
<td>0.57447</td>
<td>183</td>
<td>OUC</td>
<td>0.00000</td>
<td>0.37308</td>
<td>0.12766</td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>OUC</td>
<td>0.38298</td>
<td>0.47500</td>
<td>0.65957</td>
<td>209</td>
<td>takelab</td>
<td>0.00000</td>
<td>0.39423</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>277</td>
<td>OUC</td>
<td>0.38298</td>
<td>0.42308</td>
<td>0.65957</td>
<td>208</td>
<td>takelab</td>
<td>0.00000</td>
<td>0.39423</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>248</td>
<td>OUC</td>
<td>0.38298</td>
<td>0.49038</td>
<td>0.55319</td>
<td>279</td>
<td>takelab</td>
<td>0.00000</td>
<td>0.39423</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>OUC</td>
<td>0.29787</td>
<td>0.46154</td>
<td>0.65957</td>
<td>265</td>
<td>SMLAB</td>
<td>0.00000</td>
<td>0.38269</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>fuys</td>
<td>0.23404</td>
<td>0.56923</td>
<td>0.34043</td>
<td>275</td>
<td>SMLAB</td>
<td>0.00000</td>
<td>0.38269</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>rVRAIN</td>
<td>0.17021</td>
<td>0.47885</td>
<td>0.21277</td>
<td>175</td>
<td>SMLAB</td>
<td>0.00000</td>
<td>0.38269</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>287</td>
<td>rVRAIN</td>
<td>0.14894</td>
<td>0.47885</td>
<td>0.25532</td>
<td>217</td>
<td>OUC</td>
<td>0.00000</td>
<td>0.37308</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>OUC</td>
<td>0.14894</td>
<td>0.49038</td>
<td>0.17021</td>
<td>172</td>
<td>SMLAB</td>
<td>0.00000</td>
<td>0.35962</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>OUC</td>
<td>0.12766</td>
<td>0.37308</td>
<td>0.21277</td>
<td>193</td>
<td>SMLAB</td>
<td>0.00000</td>
<td>0.32885</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>254</td>
<td>rVRAIN</td>
<td>0.10638</td>
<td>0.47885</td>
<td>0.17021</td>
<td>164</td>
<td>TO</td>
<td>0.00000</td>
<td>0.13462</td>
<td>0.00000</td>
<td></td>
</tr>
</tbody>
</table>

Score = \frac{1}{|\text{SID}|} \sum_{x,y \in \text{SID}} \{\text{ACC}(x, y) \times \text{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. \text{SID} is a set of monetary expressions in the gold standard data whose RIDs are not null, as shown in the following equation:

\text{SID} = \{y | \exists x : \text{RID}(x, y) \neq \text{null}\}.

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

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

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

\text{RIDC}(x, y) = \begin{cases} 0 & (x,\text{RID} \notin y,\text{RIDs}) \\ 1 & (x,\text{RID} \in y,\text{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. dltlab 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 Politecnica de Valencia
11. SMLAB National Agriculture and Food Research Organization & The University of Tokyo
12. takelab Osaka Electro-Communication University
13. TO*