

QA Lab-4:

QALab-PoliInfo

<https://poliinfo.github.io/>

YASUTOMO KIMURA *5 *6, HIDEYUKI SHIBUKI*1, KOTARO SAKAMOTO*1, *2,

MADOKA ISHIOROSHI*2, TERUKO MITAMURA*3, NORIKO KANDO*2, *4, TATSUNORI MORI*1,

*1: YOKOHAMA NATIONAL UNIVERSITY, *2: NATIONAL INSTITUTE OF INFORMATICS, *3: CARNEGIE MELLON UNIVERSITY,

*4: THE GRADUATE UNIVERSITY FOR ADVANCED STUDIES (SOKENDAI), *5: OTARU UNIVERSITY OF COMMERCE, *6: RIKEN AIP

キックオフミーティングの流れ

1. QALab-PoliInfo の説明 (15分程度)

- 目標
- 関連タスク
- Dataset
- Tasks
 - ✓ Extraction task(仮)
 - ✓ Summarization task
- Important Dates

2. 参加者との議論 (60分程度)

- 参加者の自己紹介
- 提案タスクについての疑問・修正案など
- 英語タスクについてのアドバイス

目標

QALabの目標

- 現実世界へ質問応答を適用した時に生じる問題に対して、解決することを目指している
(例えば、大学入試という現実世界の問題)

QALab-PoliInfo では

- 政治情報（地方議会会議録）を対象として，
 - ✓ 発言者の意図を変えずに根拠・条件・例示の抽出をすること（抽出タスク）
 - ✓ 発言者の意図を変えずに要約すること（要約タスク）

フェイクニュースなど政治・経済における情報の信憑性が問われる場面が多くなっている
これを受けて、FNC-1、CLEFのFact-checking Labなどが開催された

信憑性判断のためには、一次情報にあたることが肝要である

QA Lab-PoliInfoでは、地方議会会議録を一次情報として抽出・要約行うタスクを設定する

抽出（根拠・条件・例示）の必要性

問題：発言の引用によっては（文脈の欠落など）、発言者の意図と異なった印象を与えてしまう。

「A議員は〇〇と主張した」といった二次情報がある場合

例えば、

- 根拠 「Xのためには〇〇をすべきだ」
- 条件 「Yだとしたら〇〇をすべきだ」
- 例示 「Zなどの〇〇をすべきだ」

根拠・条件・例示といった文脈が欠落して、

「〇〇をすべきだ」と引用された場合、

読者に「（いかなる状況においても）〇〇すべきだ」といった誤解を与えかねない。

そのような誤解を防ぐために、

引用された発言の文脈（根拠、条件、例示など）に該当する範囲の

一次情報を提示することが必要である。

欠落すべきでない文脈の例

1. 根拠のない例

- 高度防災都市の実現のためには、住宅市街地などの道路や敷地、建物に対する対策を総合的に進めていく必要があります。

2. 条件のない例 (オリンピックの例)

- 日本全体での二〇二〇年の招致を考えていくのであれば、国とともに推進する招致としていかなければならない…

3. 例示（限定の）のない例

- 医療機関などのライフライン機能の維持に支障を来し、

- 一次情報における、文脈を含めた引用箇所の範囲（extraction taskの結果）は長い
- 不適切に文脈を欠落させると、発言者の意図と異なった印象を与えてしまう
- 発言者の意図と異なった印象を与えないように要約する必要がある

関連タスク (Shared task)

政治情報の信憑性に関するshared task が設定されている

1. Fake News Challenge Stage 1 (FNC-I)
2. CLEF-2018 「Fact checking Lab」
3. NTCIR-14 「QALab-PoliInfo Task」

EXAMPLE HEADLINE

"Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ..."

CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an \$800 million Led Zeppelin reunion deal. ..."

EXAMPLE HEADLINE

"Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ..."

CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an \$800 million Led Zeppelin reunion deal. ..."

CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED

見出しと記事内容の関係を4つに分類する

見出し 「ロバート・プラントは、8億ドルのレッド・ツェッペリンの再契約を破棄した」

1. AGREE :

「レッド・ツェッペリンのロバート・プラントは、5億ユーロ(=8億ドル)の再契約を取り下げた。」

2. DISAGREE :

「いいえ、ロバート・プラントは、レッド・ツェッペリンを再結成させるため8億ドルの契約を破棄しませんでした。」

3. DISCUSSED :

「ロバート・プラントは、報道によると、8億ドルのレッド・ツェッペリン・再結成の契約を破棄した。」

4. UNRELATED :

「リチャード・ブランソンのヴァージン・ギャラクティックは本日スペースシップツーを打ち上げる予定です。」



CLEF-2018 Fact checking Lab

Task 1: Check-worthiness :

- 68 TRUMP With the United States, as he said, not so much. 0
- 69 TRUMP So Ford is leaving. 1
- 70 TRUMP You see that, their small car division leaving. 1
- 71 TRUMP Thousands of jobs leaving Michigan, leaving Ohio. 1
- 72 TRUMP They're all leaving. 0

Task 2: Factuality :

- 68 TRUMP With the United States, as he said, not so much. N/A
- 69 TRUMP So Ford is leaving. 1 Ford Motor Company is moving their small car division out of the USA. TRUE
- 70 TRUMP You see that, their small car division leaving. 1 Ford Motor Company is moving their small car division out of the USA. TRUE
- 71 TRUMP Thousands of jobs leaving Michigan, leaving Ohio. 2 Thousands of jobs are being lost in Michigan and Ohio due to Ford Motor Company moving their small car division out of the USA. FALSE

関連タスクとの比較

	FNC-1	Fact checking	NTCIR QALab-PoliInfo
Dataset	新聞記事	大統領選のディベート	都議会会議録, および, 議会だより
Task	FNC-1 見出しと記事内容の分類 1. Agree 2. Disagree 3. Discussed 4. Unrelated	Check-worthiness 真偽判断すべきか否かの2値分類 Factuality 真偽判断と根拠の抽出	Extraction 発言者の意図を正確に伝えるための 一次情報の範囲を抽出 Summarization 発言者の意図を正確に 伝えることができる要約
Number of data	2,586 articles	1,400 sentences x 3 files	-
Language	English	English and Arabic	Japanese



東日本大震災

被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。

知事 全国の先頭に立ち刻苦する被災地を支援するのは当然。今後も強力に後押しする。

東京の総合防災力

更に高める取組が必要。

知事 新視点の対応加えた防災対応指針を11月に策定し防災力向上の道筋を早期に示す。

原子力発電所

安全性に対する基本認識は。

知事 原子力利用に当たって安全の確保は当然。信頼性失ってきた姿勢を反省し再考すべき。同時にエネルギー確保は国家・社会の存立に直結、戦略を速やかに構え直す必要あり。

議員 山下太郎	256行目	私は、都議会民主党を代表して、都政の主要課題について知事並びに関係局長に伺います。 東日本大震災より3カ月余りが過ぎました。 …（中略）… しかし、私たちは、この結果を尊重するとともに、もう1方の公選によって私たちに付託された都民の期待を踏まえ、今後も都民の生活を第1とする都政の実現に取り組むことを表明させていただきます。 まず、東日本大震災における被災地支援と東京の防災対策について伺います。 三月十一日、マグニチュード九・〇、最高震度七の強く長い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、甚大な被害を引き起こしました。 福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範に拡散しました。 …（中略）… 被災地のニーズを的確に把握し、被災地、被災者が真に必要なとする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。現在、都内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをしております。 現在、都内には福島県などから自主避難してきた約5,000名の避難者の皆さんが都営住宅などに仮住まいをしております。 …（中略）… 都は、コミュニティにも配慮した避難者に対する支援の取り組みを行っていくべきと考えますが、都の見解を伺います。 東日本大震災を教訓に、東京においても発災時における社会対応力の強化や防災リーダーなど、地域人材の育成などに1層取り組み、東京を災害に強い持続可能な都市としていかなければなりません。 …（中略）… 地震、津波の被害想定の見直しや防災対策の総点検、そして東京の総合防災力をさらに高める取り組みが必要だと考えますが、知事の見解を伺います。 次に、原子力発電所の安全性について伺います。 …（中略）… そこで改めて、原発の安全性に対する石原知事の基本認識について伺います。 次に、首都圏の防災対策について伺います。 …（中略）… ご清聴ありがとうございました。
	257行目	
	（中略）	
	265行目	
	266行目	
	267行目	
	268行目	
	（中略）	
	276行目	
	277行目	
	（中略）	
	283行目	
	284行目	
	（中略）	
	291行目	
	292行目	
	299行目	
	300行目	
	524行目	

2つの言語資源からデータセットの作成

1. 東京都議会
2. 都議会だより

```
"L266-283": {  
  "Prefecture": "東京都",  
  "Date": "23-6-23",  
  "Meeting": "平成23年第2回定例会",  
  "StartingLine": 266,  
  "EndingLine": 283,  
  "MainTopic": "東京の総合防災力を更に高めよ<br>環境に配慮した都市づくりを",  
  "SubTopic": "東日本大震災",  
  "Speaker": "山下太郎",  
  "Summary": "被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。",  
  "Source": "まず、東日本大震災における被災地支援と東京の防災対策について伺いま",  
  "GikaiDayoriURL": "http://www.gikai.metro.tokyo.jp/newsletter/293/",  
}
```




1. Extraction Task
2. Summarization Task



目標

- 発言者の意図を忠実に再現できるように抽出する.
- まず、要約や引用されている文から、第一情報が記載されている箇所を見つける.

評価方法（案）

- 人手による正解範囲とシステムの出力範囲を「文単位」で比較して、Recall, Precision を計算する



東日本大震災

被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。

知事 全国の先頭に立ち刻苦する被災地を支援するのは当然。今後も強力に後押しする。

東京の総合防災力

更に高める取組が必要。

知事 新視点の対応加えた防災対応指針を11月に策定し防災力向上の道筋を早期に示す。

原子力発電所

安全性に対する基本認識は。

知事 原子力利用に当たって安全の確保は当然。信頼性失ってきた姿勢を反省し再考すべき。同時にエネルギー確保は国家・社会の存立に直結、戦略を速やかに構え直す必要あり。

議員 山下太郎

269行
(19,204文字)

256行目
257行目
(中略)
265行目

266行目
267行目

268行目

(中略)
276行目

277行目

(中略)
283行目

284行目

(中略)

291行目
292行目

299行目
300行目

524行目

私は、都議会民主党を代表して、都政の主要課題について知事並びに関係局長に伺います。
東日本大震災より3カ月余りが過ぎました。
… (中略) …
しかし、私たちは、この結果を尊重するとともに、もう1方の公選によって私たちに付託された都民の期待を踏まえ、今後も都民の生活を第1とする都政の実現に取り組むことを表明させていただきます。
まず、東日本大震災における被災地支援と東京の防災対策について伺います。
三月十一日、マグニチュード九・〇、最高震度七の強く長い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、甚大な被害を引き起こしました。
福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範に拡散しました。
… (中略) …
被災地のニーズを的確に把握し、被災地、被災者が真に必要なとする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。現在、都内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをいらっしゃいます。
現在、都内には福島県などから自主避難してきた約5,000名の避難者の皆さんが都営住宅などに仮住まいをいらっしゃいます。
… (中略) …
都は、コミュニティにも配慮した避難者に対する支援の取り組みを行っていくべきと考えますが、都の見解を伺います。
東日本大震災を教訓に、東京においても発災時における社会対応力の強化や防災リーダーなど、地域人材の育成などに1層取り組み、東京を災害に強い持続可能な都市としていかなければなりません。
… (中略) …
地震、津波の被害想定の見直しや防災対策の総点検、そして東京の総合防災力をさらに高める取り組みが必要だと考えますが、知事の見解を伺います。
次に、原子力発電所の安全性について伺います。
… (中略) …
そこで改めて、原発の安全性に対する石原知事の基本認識について伺います。
次に、首都圏の防災対策について伺います。
… (中略) …
ご清聴ありがとうございました。

18行
(1,129文字)

8行
(478文字)

東日本大震災

被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。

知事 全国の先頭に立ち刻苦する被災地を支援するのは当然。今後も強力に後押しする。

東京の総合防災力

更に高める取組が必要。

知事 新視点の対応加えた防災対応指針を11月に策定し防災力向上の道筋を早期に示す。

東日本大震災

知事の見解

被災者が真に必要なとする支援に

継続して取り組むべき

東京の総合防災力

さらに高める取組が必要

問題例

平成23年6月の定例会第2回の山下太郎議員の「東日本大震災 被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。」という内容が記述されている範囲を開始行と終了行で示しなさい。

—— 対象データ ——

議員 山下太郎

265行目

…(前略)…

しかし、私たちは、この結果を尊重するとともに、もう1方の公選によって私たちに付託された都民の期待を踏まえ、今後も都民の生活を第1とする都政の実現に取り組むことを表明させていただきます。

266行目

まず、東日本大震災における被災地支援と東京の防災対策について伺います。

三月十一日、マグニチュード九・〇、最高震度七の強く長い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、甚大な被害を引き起こしました。

福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範に拡散しました。

原発周辺の住民の皆さんは、自宅があるのに帰れない深刻な状況が続いています。

…

被災地のニーズを的確に把握し、被災地、被災者が真に必要なとする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。現在、都内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをいらっしゃいます。

283行目

284行目

東日本大震災を教訓に、東京においても発災時における社会対応力の強化や防災リーダーなど、地域人材の育成などに1層取り組み、東京を災害に強い持続可能な都市としていかなければなりません。

…(後略)…

—— 出力 ——

開始行 266行目 終了行 283行目

要約の目標

- （情報を詰め込む要約ではなく）読み手が誤解しないように、発話者の意図を正しく伝える
- 長い文章をまとめながらも、発話者の伝えたいことを削除しない
- 発話者の利用した表現をできる限り利用する

評価方法（案）

- ROUGE
- Gold Standard Data（一行程度）を読み、○△×の三段階評価

問題例 次の山下太郎議員の発言をその意図が正確に伝わるように40文字以内に要約しなさい。

入力

まず、東日本大震災における被災地支援と東京の防災対策について伺います。
三月十一日、マグニチュード九・〇、最高震度七の強く長い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、甚大な被害を引き起こしました。福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範に拡散しました。
原発周辺の住民の皆さんは、自宅があるのに帰れない深刻な状況が続いています。
…（中略）…
被災地のニーズを的確に把握し、被災地、被災者が真に必要とする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。現在、都内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをしております。

出力

東日本大震災 被災地が真に必要とする支援に継続して取り組むべき。知事の見解は。



Feb 20, 2018: QALab-PoliInfo Kickoff meeting in NII (room 1901, 1902)

Mar 13, 2018: 19:10-19:40 1st round table meeting at Okayama Convention Center in NLP 2018

Mar 20, 2018: NTCIR-14 Kickoff event in NII

Jun 2018: Dataset Release

Jul 2018: Task Registration Due

Jul 2018: Dry Run

Nov 2018: Formal Run

Feb 1, 2018: Evaluation Result Release

Feb 1, 2018: Task overview paper release (draft)

Mar 15, 2019: Submission due of participant papers

Jun 2019: NTCIR-14 Conference & EVIA 2019 in NII, Tokyo

参加者との議論(60分程度)

1. QALab PolilInfo の説明 (15分程度)

- ねらい・目標
- 関連タスク
- Dataset
- Tasks
 - ✓ Extraction task(仮)
 - ✓ Summarization task
- Schedule

2. 参加者との議論 (60分程度)

- 参加者の自己紹介
- 提案タスクについての疑問・修正案など
- 英語タスクについてのアドバイス



qalab-admin

- qalab-admin@nii.ac.jp

Web site

- <https://poliinfo.github.io/>

Twitter

- <https://twitter.com/NTCIRQALab>
- <https://twitter.com/NTCIRQALabJA>

Facebook

- <https://www.facebook.com/groups/479281228887018/>