



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



QA Lab is aimed at **complex real-world question answering** (QA) technologies

- NTCIR-11 QA Lab
 - NTCIR-12 QA Lab-2
 - NTCIR-13 QA Lab-3
- } QA using **university entrance exams** data

Various technologies are required!

Previous tasks

1. Multiple-choice question type → Text Entailment
2. Term question type → Information Extraction
3. Essay question type → Automatic Summarization

However, the **data** we prepared were **depleted** in QA Lab-3.

Therefore, we will tackle a **new domain QA**



QA Lab-PolInfo is **QA for political information** using **Japanese regional assembly minutes**, to show summaries on the opinions of assembly members, and the reasons and conditions for such opinions.

The **importance of fact checking** owing to the **negative impact of fake news**

- International Fact-Checking Day, April 2 from 2017 <http://factcheckingday.com/>

However, fact-checking is difficult for general Web search engines

- because of the 'filter bubble' developed by Eli Pariser

For fact checking,

- we should **confirm the primary sources** such as the assembly minutes according to **critical thinking**

Japanese regional assembly minutes



This is a single speech given by an assembly member that is a request to the governor!

〇十番(やながせ裕文君) 福祉先進都市を実現する。外添知事の所信表明を頼もしく聞いておりました。このチャレンジは都民の誰もが歓迎するものであり、一刻も早い実現を望むものです。保育、子育て、介護、障害者支援など、手厚い予算が組まれ、都民福祉の向上は着実に図られているものと思います。一人一人が活躍できる社会へという知事の基本理念に賛同するものですが、それを実現し、持続していくためには、もっともっと自己改革と経済成長への取り組みが必要だという観点から質問をしていきます。

まず、財政についてであります。知事は、必要なところに必要な予算をつけるといいます。当然ながら、円税収減を体験した都政において、盤石な基盤とは到底いえない。これまで、長期の財政計画を策定すべきだ、都債発行額にキヤそこで、財政規律を保つため、目標となる財政運営上の指標が。知事には、今現在困難な状況にある人だけでなく、将来爆発的にまた、歳出を抜本的に削減する改革は進んでいません。何か新しかし、好調な税収のもと、事業をスクラップすることへのモチベ知事は、必要な改革を必要とときに言うといいますが、自主的な天下りし放題の監理団体には、さまざまな理由をつけて一十億円を。ちました。もう庁内を熟知されていることと思います。知事は、この内部改革をどのように取り組んでいくのか、見解を伺います。また、都民の理解をなかなか得られないのが都市外交であります。二十八年度十九億円が計上されており、知事の海外出張経費は。見解を伺います。

二〇二〇年東京大会の開催費用は大丈夫でしょうか。昨年末には、総額一兆八千億円という金額が報道されました。していることですが、状況の変化に応じてコストを算出すればいい。知事は、三兆円は必要だと発言していますが、これは何か根拠が。か的なニュアンスを感じますが、国立施設すら負担を求められてい。そこで、知事に質問ですが、現時点で大会の総費用をどれくらい。また、知事は第三者的な立場からではなく、大会経費の把握や分担。二〇二〇年大会の成功は、いかに費用面でもコンパクトに実施できるかにかかっています。運営費用を低減するとともに、これ以上混乱を招かないようにリーダーシップを発揮していただきたいと思。また、受動喫煙防止の条例化については、これはなぜ先送りをするのでしょうか。国の動向に関係なく、むしろ国に先駆けて、開催都市としての責務を果たすべきです。一刻も早い条例制定を望みますが、見解を伺います。安定した税収を得るためには、経済成長が欠かせません。政府は、GDP六百兆円を目指すとの意欲的な姿勢を打ち出しました。都はこれに応じ、的確な施策を全力で講じるべきです。有効な施策は、競争政策、規制緩和、民営化でしょう。一日九百万人以上という世界一の乗客数を誇る東京の地下鉄が一元化し、民営化されれば、莫大な富を生み出します。そして都民の利便性は飛躍的に向上します。しかし、懸念されるのはメトロ株の扱い。昨年、会計検査院が、国と都の売却に向けた協議がたざらしくなっていると指摘。東京メトロ株は復興財源確保法に基づき、売却収入を東日本震災の復興債の償還財源に充てることとなっています。この指摘を受けて、国はメトロ株をどうしていか、注視が必要ですが、地下鉄一元化を展望し、都は安易にメトロ株を手放してはいけません。地下鉄一元化及びメトロ株の売却について、知事の所見を伺います。経済を活性化していく上で、都はその巨大な資産を生かすべきです。国土交通省は、日本全体の不動産資産二千四百兆円のうち、公的不動産が五百七十兆円を占め、この効率的な運営が経済の活性化につながるとして、PRE戦略を進めています。都では、財務局が所管する土地のうち、百九十万平米が未利用地となっており、これは東京ドーム四十個分の広さに相当します。福祉利用など検討してきましたが、目的に合った土地は少なく、そのほとんどが資材置き場など暫定利用されています。将来の行政需要を考慮するなどという理由で莫大な土地を塩漬けにすることなく、民間活用を図るべきであります。これらの未利用地を早期に売却することを検討すべきと考えますが、見解を伺います。観光産業への投資を急ぐべきです。現在の急激な外国人観光客。法案審議が進んでいません。知事は早急に方針を示すべ。また、カジノ誘致とともに、。私は、カジノ施設は外国人。先日、パチンコに行くため、。しかし、ギャンブル依存症や。向き合う必要があり、莫大な。昨年の会派要望では、これ。ギャンブル依存症を引き起。パチンコ税の導入を検討すべきと考えますが、所見を伺います。最後に、今週末に開催される東京マラソンについて伺います。記念すべき第十回目を迎える中、抽せん倍率が十倍と高どまりし、なかなか当選できないという声を多数いただいています。ランナー団体の調査でも、九回エントリーしても一回も当選していない、五回以上エントリーしてもなかなか当選できない、非常に多くの方がこういう声を届けられています。同様の課題を持つ神戸マラソンでは、二万人のランナーのうち、二千人を初出場枠として設定しています。東京のよさを知ってもらうこの取り組みに向けて、これまで抽せん以外に外れて走れなかった方が優先的に参加できるような仕組みをつくるなど、一人でも多くの方に東京マラソンを体験してもらうべきと考えますが、見解を伺い、質問を終わります。(拍手)

Transcript of a speech. However, the speech is very long.

It is difficult to understand the contents, including the opinions of the assembly member at a glance.

New information access technologies to support user's understanding are expected.



For confirmation of primary information source

- When a citation is given, we need to identify the corresponding texts in primary sources

→ Segmentation task

- If the texts is too long, we need to summarize them

→ Summarization task

For critical thinking

- We need to get the whole view of opinions

→ Classification task



Segmentation Task

- Given Japanese regional assembly minutes and a brief citation
- Extract a text corresponding to the citation from the minutes

Summarization Task

- Given a text including an assembly member's opinion
- Make a summary which guarantees to keep the opinion

Classification Task

- Given a text including political keyword
- Classify the description whether merit or demerit

Indicate the range of the content.

Question

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示しなさい。

Speaker	Line	Sentence	Output
Taro Yamashita		…(略)…	
	265	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。	0
	266	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。	1
	267	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。	1
	268	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。	1
		… (略) …	
	269	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。	1
	276	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されています。	1
		… (略) …	
	283	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。	1
284	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。	0	
	… (略) …		

Segmentation Task



Indicate the range of the content.

Question

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示しなさい。

Speaker	Line	Sentence	Output
Taro Yamashita		…(略)…	
	265	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。	0
	266	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。	1
	267	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。	1
	268	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。	1
		… (略) …	
	269	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。	1
	276	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。	1
		… (略) …	
	283	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。	1
284	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。	0	
	… (略) …		

Summarization Task



Question 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

Summarize within 40 letters

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。

Output

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。



Input : An utterance which includes political keyword

Output : Classification concerning the description whether merit or demerit

Keywords “Tsukiji market”

1. Merit:

「築地市場は、施設配置が鉄道時代のもので、トラック輸送が主流の現在の物流に対応していません。」

2. Demerit:

「豊洲の新市場予定地では108箇所では液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。」

3. Both merit and demerit :

「築地市場が東京都の台所として主要な存在であることは周知のとおりですが、その一方で、鮮魚だけをとってみれば、築地以外に足立、大田といった市場があり、さらには野菜、果物や肉、花などを扱う大切な市場が存在し、都民の生活を地域に根差して支えています。」

4. Neither merit nor demerit:

「思い返せば、ここ数年、築地の豊洲移転は都議会の大きな課題でした。」

Related work (Shared task)

Shared tasks on the credibility of political information

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



	FNC-1	Fact checking	NTCIR QALab-PoliInfo
Dataset	News article	Political debate	Assembly minute and News letter
Task	Classification <ol style="list-style-type: none"> 1. Agree 2. Disagree 3. Discussed 4. Unrelated 	Check-worthiness <ul style="list-style-type: none"> • Binary Classification Factuality <ul style="list-style-type: none"> • Binary Classification • Extraction 	Classification Segmentation Summarization
Number of data	2,586 articles	1,400 sentences x 3 files	-
Language	English	English and Arabic	Japanese

Data and Resource

We provided the **Japanese Regional Assembly Minutes Corpus**.

- JSON format data of the Tokyo metropolitan assembly for 4 years
- Data fields

Identifier	Prefecture name
Volume	Number
Year	Month
Day	Period
Title	Speaker expression
Speaker ID	Speaker name
Speaker position	Speech
URL	HTML file

Participants can use any resources (and need to report)

Data characteristics

- (1) **Dialog** including questions and answers
- (2) **Beliefs and attitudes** of the assembly member
- (3) Mental spaces for other assemblymen
- (4) Contexts, including reasons
- (5) Several topics in the political documents
- (6) **Colloquial Japanese** including dialect and slang

Evaluation

We will discuss the appropriate representation, evaluation metrics and methodologies with the participants

The discussions will be held through round table meetings, mailing lists, and other means.

Scope

This task will contribute to the development as follows :

- QA technologies,
- information extraction,
- semantic representation,
- context understanding,
- information credibility,
- automated summarization,
- dialog system,
- and others



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

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

Apr 19, 2018: 1st round table meeting in NII (room 1901, 1902)

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



qalab-admin

- qalab-admin@nii.ac.jp

Web site

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

Twitter

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

