NTCIR-14 QA Lab-PoliInfo の Formal Run Dataset の構築

木村泰知（小樽商科大学, 理化学研究所AIP）
渋木英潔（横浜国立大学）
高丸圭一（宇都宮共和大学）
内田ゆず（北海学園大学）
乙武北斗（福岡大学）
石下円香（国立情報学研究所）
阪本浩太郎（横浜国立大学, 国立情報学研究所）
三田村照子（カーネギーメロン大学）
神門典子（国立情報学研究所, 総合研究大学院大学）

NTCIR-14 QA Lab PoliInfo

https://poliinfo.github.io/

Githubでデータセットを公開しました

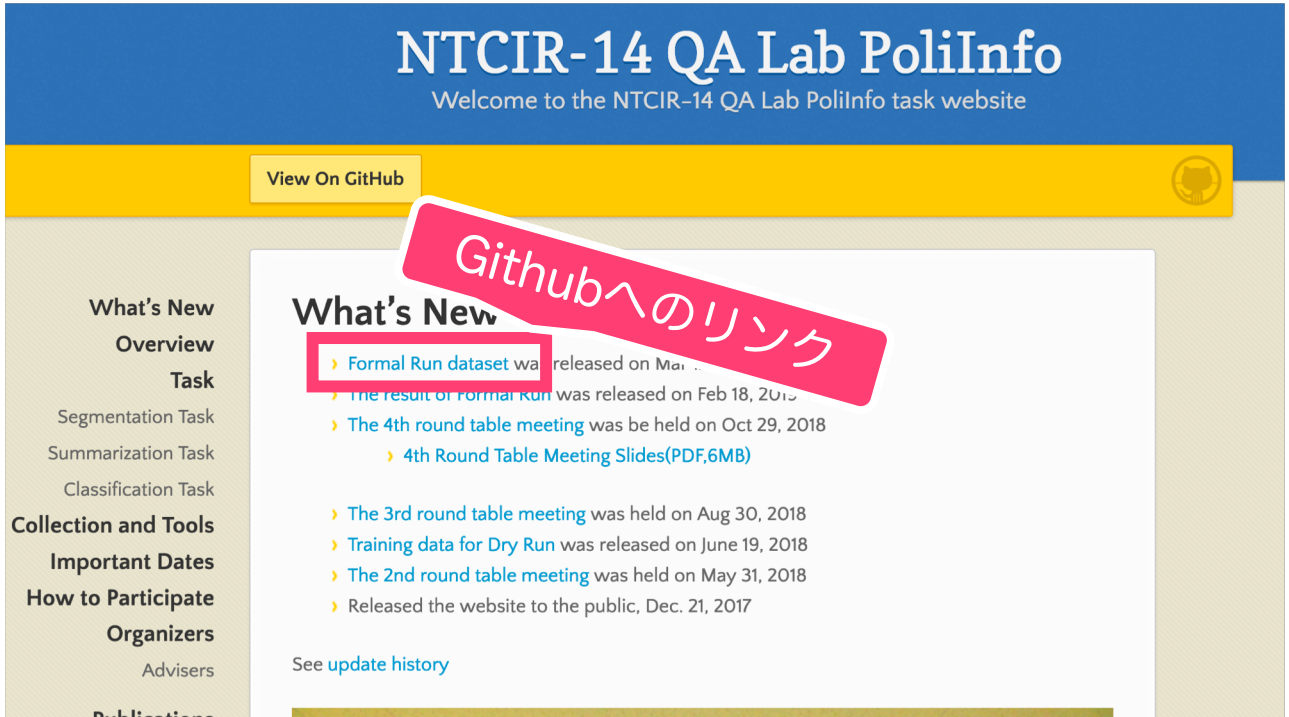
https://github.com/kmr-y/NTCIR14-QALab-PoliInfo-FormalRunDataset

もくじ

- NTCIR14 QALab-PoliInfo とは
- 政治情報
  - 地方議会会議録
  - 議会だより
- ３つのタスクについて
  - Segmentation task
  - Summarization task
  - Classification task
- データセットの構築方法
  - クラウドソーシングなど
- Github での公開

「QALab PoliInfo」で検索

https://poliinfo.github.io/



NTCIR14 QALab-PoliInfoとは

- NTCIRとは
  - ▶ 共通のデータセットを利用した評価型ワークショップ
- QALab
  - ▶ 現実世界における質問応答システムの実現を目指している
    - ✓ 1回目から3回目まで「世界史の大学入試問題」
    - ✓ 4回目から「政治情報」を対象としている
- QA Lab-PoliInfo
  - ▶ 地方議会会議録を用いたタスクの設定
    - ✓ 引用元の一次情報を見つける
    - ✓ 根拠のある意見を見つける
    - ✓ 議論の構造を考慮して要約する
- NTCIR14 QALab-PoliInfo Formal Run
  - ▶ 開催期間：2018年11月下旬から12月上旬まで
    - ✓ Dry Runは2018年8月に開催された
    - ✓ Dry Runから，データセットを修正している

議会会議録

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

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
	広君
	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
警視総監	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

いつ　どこで　誰が　何を　発言したのか正確にわかる

議会会議録

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

出席議員

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

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
警視総監	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

発言者は議員以外にも，知事，副知事，担当局長などがいる

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思われます。かねてから景気の低迷の影響もあり

地方議会会議録の特徴

議員発言をそのまま記録

1次情報

○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
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。

自動要約

話し言葉で読みづらい
一括質問一括答弁のため質問と答弁が離れている

○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
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。
る。その財
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
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

意見抽出

根拠のある意見が含まれている

議会だより

自動要約の正解

トップ　>　都議会だより　>　293号

都議会だより293号

人手で要約した文書

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。

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。

３つの課題

① Segmentation

1次情報をみつける

新聞記事

xxxxxxxxxxxxxxx

知事が「xxxxxxx」

と述べており… xxx

③ Classification

根拠のある意見を

みつける

地方議会会議録

議員

質問1 xxxxxxxxxx

質問2 xxxxxxxxxx

質問3 xxxxxxxxxx

知事

答弁1 xxxxxxxxxx

答弁2 xxxxxxxxxx

答弁3 xxxxxxxxxx

② Summarization

質問と答弁の構造を

考慮して要約する

議会だより

(要約の正解)

議員「xxx」

知事「xxx」

政治情報を対象にした３つの課題

1.Segmentation Task

- 目標　一次情報をみつける

2.Summarization Task

- 目標　質問と答弁の構造を考慮して要約する

3.Classification Task

- 目標　根拠のある意見をみつける

政治情報を対象にした３つのタスク

1.Segmentation Task

- ▶ 入力：質問と答弁の「発言の引用 の対(つい)」とその発言が含まれる「議会会議録」
- ▶ 出力：引用を理解するために読むべき発言の範囲（「開始行」「終了行」）

2.Summarization Task

- ▶ 入力：議会会議録に含まれる「発言」と要約の「制限字数」と「議会会議録」
- ▶ 出力：発話者の意図が伝わる「要約」

3.Classification Task

- ▶ 入力：政策が記述された「課題文」と議会会議録中の「発言文（一文のみ）」
- ▶ 出力：３値に分類「事実検証可能な根拠を伴う賛成」「事実検証可能な根拠を伴う反対」「その他」
  - ✓ ３つのサブクラス
    - 「関連（有り，無し）」
    - 「事実検証(可能，不可能)」
    - 「立場(支持，不支持，どちらでもない)」

Segmentation

Segmentation task　開始行について

- 開始行の決定方法
  - ▶ まず，開始行を説明する前に「都議会会議録」と「都議会だより」の関係について説明します

Segmentation task　開始行について

● 開始行の決定方法

▶ 「東京都議会だより」と「都議会会議録」の関係について

都議会だより

議会だよりは，自治体が，議会の内容を人手でまとめて，
住民に提供している情報誌であり，会議録の要約といえる．

トップ　>　都議会だより　>　293号

都議会だより293号

山下太郎（民主党）

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。	質問1(山下議員)
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。	答弁1(石原知事)
東京の総合防災力 更に高める取組が必要。	質問2(山下議員)
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。	答弁2(石原知事)

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。

「議会だより」は各トピックごとに
「質問」と「答弁」を対応させながら要約している．

東京都議会会議録

10 行ごと：発言者：ラベル

001 - 010 行	議長	議事進行
011 - 020 行	議長	議事進行
021 - 030 行	議長	議事進行
031 - 040 行	議長	議事進行
041 - 050 行	議長	議事進行
051 - 060 行	議長	議事進行
061 - 070 行	議長	議事進行
071 - 080 行	議長	議事進行
081 - 090 行	石原知事	所信表明
091 - 100 行	石原知事	所信表明
101 - 110 行	石原知事	所信表明
111 - 120 行	石原知事	所信表明
121 - 130 行	石原知事	所信表明
131 - 140 行	石原知事	所信表明
141 - 150 行	石原知事	所信表明
151 - 160 行	石原知事	所信表明
161 - 170 行	石原知事	所信表明
171 - 180 行	石原知事	所信表明
181 - 190 行	石原知事	所信表明
191 - 200 行	石原知事	所信表明
201 - 210 行	石原知事	所信表明
211 - 220 行	石原知事	所信表明
221 - 230 行	石原知事	所信表明
231 - 240 行	議長	議事進行
241 - 250 行	議長	議事進行
251 - 260 行	山下議員	質問概要
261 - 270 行	山下議員	質問1
271 - 280 行	山下議員	質問1
281 - 290 行	山下議員	質問2
291 - 300 行	山下議員	質問3
301 - 310 行	山下議員	質問4
311 - 320 行	山下議員	質問5・6
321 - 330 行	山下議員	質問7
331 - 340 行	山下議員	質問8
341 - 350 行	山下議員	質問9
351 - 360 行	山下議員	質問10
361 - 370 行	山下議員	質問11
371 - 380 行	山下議員	質問12
381 - 390 行	山下議員	質問13
391 - 400 行	山下議員	質問14
401 - 410 行	山下議員	質問15
411 - 420 行	山下議員	質問16
421 - 430 行	山下議員	質問16
431 - 440 行	山下議員	質問16
441 - 450 行	山下議員	質問16
451 - 460 行	山下議員	質問16
461 - 470 行	山下議員	質問17
471 - 480 行	山下議員	
481 - 490 行	山下議員	
491 - 500 行	山下議員	
501 - 510 行	山下議員	質問18
511 - 520 行	山下議員	質問19
521 - 530 行	石原知事	答弁1
531 - 540 行	石原知事	答弁2
541 - 550 行	石原知事	答弁3
551 - 560 行	石原知事	答弁3

東京都議会会議録(本会議)は1年間で約3万行の記述がある．
ここでは，平成23年第二回(6月)定例会の560行を例に説明する．

	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。
	257行目	東日本大震災より3カ月余りが過ぎました。
	（中略）	…（中略）…
	265行目	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。
質問1 山下議員	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。
	267行目	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。
	268行目	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。
	269行目	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。
	270行目	私たちは、この未曾有の複合災害に対していち早く被災地支援と都内の震災対策を充実させること、そして補正予算の編成を知事に申し入れいたしました。
	271行目	また、各議員は、党の被災地支援活動やＮＰＯと連携した取り組みを行うなど、被災地支援に取り組んでまいりました。
	272行目	そこで伺います。
	273行目	都は、児童生徒への心のケアや、災害時要援護者の救護など、医療人材の継続的な派遣や、地元雇用を推進する自治体事業、キャッシュ・フォー・ワークといった取り組みへの支援をするなど、被災者の皆さんが希望を見出し、一歩踏み出すことのできるよう、生活再建をともにサポートしていくことが重要です。
	274行目	また、各県が創造的復興、もしくは再生を目指し、独自復興計画を策定、実現させていくことを都が後押しし、安全な地域社会の再建に寄与していく必要があります。
	275行目	このように被災地が取り組むべき課題は山積し、日々刻々地域ごとに状況が変化しております。
	276行目	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。
	277行目	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いらっしゃいます。
	278行目	故郷から遠く離れ、いつ帰れるのかという思いを持って生活している皆さんに、都は寄り添う形でその生活を支えていくべきと考えます。
	279行目	避難者は、見知らぬ東京での生活が不安であり、特に高齢者の方々については、引きこもりがちになるなど、孤立化も懸念されます。
	280行目	先日、特別区の都営住宅で、自治会の皆さんが避難者と懇談会を開き、福島での共通の話題で盛り上がりました。
	281行目	こうしたかかわり合いをみやす場でもあるミニ懇談会を開催し、避難者同士や地域との交流機会を創造すること を求められております。
	282行目	また、福祉も含めた総合的な相談を区市町村や災害復興まちづくり支援機構、ＮＰＯなどと連携して開催するなど、広い協働の形で避難者の暮らしを支えることも重要と考えます。
	283行目	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。
		の被害想定の検討や防災対策の総点検、そして東京の総合的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。
質問2 山下議員	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一層取り組み、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。
	285行目	現在、各連邦県や市町村で地域防災計画などを見直す動きが出ています。
	286行目	今回の震災による大津波は、近年研究が進みつつあった平安期の貞観地震に類似したものといわれています。
	287行目	高知県や茨城県では、既に江戸期の地震の実例を盛り込み、地域防災計画の策定や浸水想定を行っています。
	288行目	東京においても、江戸期に三連動地震による大津波、これに続く暴風雨や富士山噴火による複合災害が起きており、過去の災害分析からも改めて被害想定を研究すべきと考えます。
	289行目	実践的訓練やライフラインの耐震化、減災化のさらなる推進も必要です。
	290行目	福島原発事故を踏まえるのであれば、近い将来必ず起きるといわれている東海地震による静岡県浜岡原発事故リスクをも想定した放射能対策も行わなければなりません。
	291行目	地震、津波の被害想定の検討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。
答弁1 石原知事	527行目	山下太郎議員の代表質問にお答えいたします。
	528行目	まず、被災地支援についてでありますが、震災発生から三カ月たちましたが、依然として被災地には大地震がもたらした惨状が広がっており、被災者の生活再建も遅々として進んでおりません。
	529行目	電力、農林水産物等の供給の多くを被災地に依存してきた首都東京が、全国の先頭に立って刻苦する被災地を支援していくことは当然であります。
	530行目	眼前に立ちはだかる瓦れきの撤去や仮設住宅の建設、被災地域の経済再生など、山積する現地の課題を解決へと導くには、日々変化する現地の実情に即した手だてを講じることが、東都に課じることこそが何よりも重要だと思います。
	531行目	都は、震災後早期に被災地に現地事務所を開設して、被災地が真に必要とするニーズを的確にとらえ、現場を持つ都の強みを最大限に生かして、必要な物資の搬送や医療、福祉のスタッフの派遣など、機を逃すことなく支援をしてまいりました。
	532行目	国の対応は、厳しい被災地の現状に対して後手後手に回っておりますが、東京は、今後とも復興に向けみずから踏み出す被災地、被災者を強力に後押ししていきたいと思っております。
答弁2 石原知事	533行目	次いで、東京の総合防災力の向上についてでありますが、東日本大震災は、東北地方に甚大な被害をもたらしただけではなく、東京にも帰宅困難者の発生といった直接的な被害を生じさせました。
	534行目	さらには、遠隔地の災害にもかかわらず、計画停電による都民生活の混乱などの連鎖的な被害をもたらしました。
	535行目	こうした被害の発生によりまして、交通ネットワーク、電力の供給網など、ふだんは当たり前と思っている都市機能の脆弱性が明らかになったことから、防災対策を抜本的に見直す必要があります。
	536行目	地震の専門家によりますと、世界で発生するマグニチュード六以上の地震のうち、約二割はこの日本で発生しております。
	537行目	いずれにしろ、この日本はご承知のように世界最大のファイアリング、火山噴の上に存在する国土でありまして、また、東京では、体に感じないものも含めて、
	538行目	十分に一度は地震が起きているということであります。
	539行目	地震大国日本の中で、高度に集中、集積が進んだ東京において、防災力の向上は待ったなしの課題であります。
		このため、今回の震災以上の被害が懸念される首都直下型地震に備えた防災対策の一層の強化と、東海、東南海、南海連動地震も視野に入れた新たな視点からの対応を加えた、東京都防災対応指針を本年十一月に策定いたします。
	540行目	指針の内容は、地域防災計画の修正に反映させて、東京の防災力向上への道筋を早期に示していきたいと思っております。
	541行目	

東京都議会会議録のJsonファイル(Pref13_tokyo.json)

	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
	257行目	東日本大震災より3カ月余りが過ぎました。
	（中略）	…（中略）…
	265行目	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
		組むことを表明させていただきます。
質問１ 山下議員	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
	267行目	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
		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
	268行目	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
	269行目	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
	270行目	私たちは、この未曾有の複合災害に対していち早く被災地支援
	271行目	また、各議員は、党の被災地支援活動やＮＰＯと連携した取り
	272行目	そこで伺います。
	273行目	都は、児童生徒への心のケアや、災害時要援護者の救護など、
		り組みへの支援をするなど、被災者の皆さんが希望を見出し、
	274行目	また、各県が創造的復興、もしくは再生を目指し、独自復興計
		このように被災地が取り組むべき課題は山積し、日々刻々地域
	275行目	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
	276行目	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
	277行目	故郷から遠く離れ、いつ帰れるのかという思いを持って生活し
	278行目	避難者は、見知らぬ東京での生活が不安であり、特に高齢者の
	279行目	先日、特別区の都営住宅で、自治会の皆さんが避難者と懇談会
	280行目	こうしたかかわり合いをふやす場でもあるミニ懇談会を開催し
	281行目	また、福祉も含めた総合的な相談を区市町村や災害復興まちづ
	282行目	要と考えます。
		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
	283行目	の被害想定の検討や防災対策の総点検、そして東京の総合防災
質問２ 山下議員	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
		市としていかなければなりません。
	285行目	現在、各道府県や市町村で地域防災計画などを見直す動きが出
	286行目	今回の震災による大津波は、近年研究が進みつつあった平安期
	287行目	高知県や茨城県では、既に江戸期の地震の実例を盛り込み、地
	288行目	東京においても、江戸期に三連動地震による大津波、これに続
		きと考えます。
	289行目	実践的訓練やライフラインの耐震化、減災化のさらなる推進も
	290行目	福島原発事故を踏まえるのであれば、近い将来必ず起きるとい
		ん。
	291行目	地震、津波の被害想定の検討や防災対策の総点検、そして東京


```
265:
  ID:          "130001_230623_19"
  Line:        266
  Prefecture:  "東京都"
  Volume:      "平成23年_第２回"
  Number:      "2"
  Year:        23
  Month:       6
  Day:         23
  Title:       "平成23年_第２回定例会(第８号)"
  Speaker:     "山下太郎"
  Utterance:   "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。"
266:
  ID:          "130001_230623_20"
  Line:        267
  Prefecture:  "東京都"
  Volume:      "平成23年_第２回"
  Number:      "2"
  Year:        23
  Month:       6
  Day:         23
  Title:       "平成23年_第２回定例会(第８号)"
  Speaker:     "山下太郎"
  Utterance:   "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。"
```

Segmentation のJsonフォーマット(Formal run バージョン)

- Segmentation Task
  - ▶ 質問と答弁の要約と会議録を与えます
    - ✓ 右のデータ構造のJsonファイル
    - ✓ 東京都議会会議録のJsonファイル（43MB，約11.5万行）
  - ▶ 質問の範囲と答弁の範囲を答えてもらいます
    - ✓ 質問の開始行
    - ✓ 質問の終了行
    - ✓ 答弁の開始行
    - ✓ 答弁の終了行

表記	説明
ID	識別子 タスク名ー開催年ー言語ータスク種別ー連番
Prefecture	都道府県名
Date	日付
Meeting	会議名 （議会だよりの記載）
MainTopic	メイントピック
SubTopic	サブトピック
QuestionSpeaker	質問者
QuestionSummary	質問の要約
AnswerSpeaker	答弁者
AnswerSummary	答弁の要約
QuestionStartingLine	質問の開始行
QuestionEndingLine	質問の終了行
AnswerStartingLine	答弁の開始行
AnswerEndingLine	答弁の終了行

Segmentation のJsonフォーマット(Formal run バージョン)

{
"ID":"Segmentation-2018-JA-FormalSample-00001",
"Prefecture":"東京都",
"Date":"23-6-23",
"Meeting":"平成23年_第２回定例会",
"MainTopic":"東京の防災力向上は待ったなし
被災地復興には総合的取組を",
"SubTopic":"都政運営",
"QuestionSpeaker":"村上英子（自民党）",
"QuestionSummary":"〔1〕国難のもとにおける東京の役割について所見を。〔2〕「10年後の東京」計画改定の視点は。〔3〕都民への責任果たし得る財政運営を。〔4〕被災地復興支援と首都東京の防災力向上に今後どう取り組むのか、所見を。","AnswerSpeaker":"知事本局長",
"AnswerSummary":"　〔2〕高度防災機能備えた都市へ取組強化等で安全・安心社会をつくり、節電意識徹底等で環境と経済が両立した都市実現等。",
"QuestionStartingLine":763,
"QuestionEndingLine":793,
"AnswerStartingLine":1240,
"AnswerEndingLine":1243
},

埋める

表記	説明
ID	識別子 タスク名ー開催年ー言語ータスク種別ー連番
Prefecture	都道府県名
Date	日付
Meeting	会議名 （議会だよりの記載）
MainTopic	メイントピック
SubTopic	サブトピック
QuestionSpeaker	質問者
QuestionSummary	質問の要約
AnswerSpeaker	答弁者
AnswerSummary	答弁の要約
QuestionStartingLine	質問の開始行
QuestionEndingLine	質問の終了行
AnswerStartingLine	答弁の開始行
AnswerEndingLine	答弁の終了行

Segmentation task のデータセット

Training

質問	298
答弁	298

298セット

質問と答弁を対(つい)としているので
298セットとなる

Test

質問	83
答弁	83

83セット

Summarization

SummarizationのJsonフォーマット(Formal run バージョン)

- Summarization Task
  - ▶ 入力：議会会議録に含まれる「発言(原文)」と要約の「制限字数」と「議会会議録」
    - ✓ 議会会議録はPref13_tokyo.json
      - 議会会議録の文脈を利用することができる
  - ▶ 出力：質問と答弁の構造を考慮した「要約」

表記	説明
ID	識別子 タスク名ー開催年ー言語ータスク種別ー連番
Prefecture	都道府県名
Date	日付
Meeting	会議名 （議会だよりの記載）
Speaker	発言者
StartingLine	開始行
EndingLine	終了行
Main topic	メイントピック
Sub topic	サブトピック
Summary	要約文
Length	制限字数
Source	原文

SummarizationのJsonフォーマット(Formal run バージョン)

{

"ID": "Summarization-2018-JA-Dry-00002",

"Prefecture": "東京都",

"Date": "23-6-23",

"Meeting": "平成23年第2回定例会",

"Speaker": "石原慎太郎",

”StartingLine": 541,

”EndingLine": 550,

”Main topic": ” 東京の総合防災力を更に高めよ
環境に配慮した都市づくりを",

"Sub topic": ”原子力発電所",

"Summary": " 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。 ",

"Length": ”100字以内",

"Source": "次いで、原子力発電の安全性についてでありますが、今回の福島原発の事故は、ただいまの質問にあった四年前、柏崎原発が被災した際、東京都が国に安全性確保を要求した云々の話とは全く次元の違う問題であります。¥¥n先日、海江田経済産業大臣は、原発の短期的な安全については確認されたと表明しております。¥¥nみずからの民主党政権が原発を管轄し、検査もしているのですから、安全性については国家に聞いたらいかがですか。¥¥n改めて申し上げることもなく、原子力の利用に当たっても、安全の確保は当然であります。¥¥nなぜフランスでできている安全管理が日本にできないのか。¥¥nインシデントすら隠して信頼性を失ってきたこれまでの姿勢を反省し、再考すべきであると思います。¥¥n同時に、エネルギー確保は、国家、社会の存立に直結するがゆえに、エネルギー戦略も速やかに構え直す必要があります。¥¥n豊富な電力が安定的に供給されるからこそ、我が国を支えるものづくりがここまで発展できたわけでありまして、新しいＩＴの技術が急速に進歩してまいりました。¥¥nかつて、完全な管理技術を前提とすれば、東京湾に立派な原子力発電をつくってもいいと発言したことがございますが、当然のことであります。¥¥n日本の頭脳、心臓部の将来を、一段の発展を考えたとき、その裏づけとなる電力をほかに依存し続けることは不合理だと考えます。¥¥n"

},

埋める

表記	説明
ID	識別子 タスク名ー開催年ー言語ータスク種別ー連番
Prefecture	都道府県名
Date	日付
Meeting	会議名 （議会だよりの記載）
Speaker	発言者
StartingLine	開始行
EndingLine	終了行
Main topic	メイントピック
Sub topic	サブトピック
Summary	要約文
Length	文字数
Source	原文

Summarization task のデータセット

Training

質問	298
答弁	298

596セット

質問と答弁を分けているので
Segmentationの2倍になる

Test

Single	79
Multi	67

146セット

Multiの例

〔1〕研修会等で支援センター活用を働きかける。
〔2〕支援策充実を働きかける

Classification

ラベル付けについて

- 目的
  - ▶ 政治家の発言から「事実検証可能な根拠を伴う意見」を見つけることです。
- 作業
  - ▶ 「課題文」と「政治家の発言(一文のみ)」を読み、次の３つのラベルを付けてもらいます。

1. 「関連」... 有り or 無し
2. 「事実検証可能」 ... 可能 or 不可能
3. 「立場」 ... Support , Against or Other

ラベル付けについて

▶ 課題文とは

✓ 政治課題と意見が含まれており、「政治家の発言(一文のみ)」と比較する文のことです。

- 例「築地市場を豊洲に移転するべきである」

▶ 関連(Relevance)とは

✓ 「政治家の発言(一文のみ)」と「課題文」との間にかかわりがあることです。

- 関連のある例　「新銀行東京や築地市場の移転問題は非の立場です。」
- 関連のない例　「お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります」

▶ 事実検証可能(Fact-checkability)とは

✓ 「政治家の発言(一文のみ)」に施設名や日付や金額などの具体的な表現が含まれ、Fact Check(事実の検証)ができることです。

- 事実検証できる例　「このような中、東京都はこの八月三十日に、豊洲の土壌汚染対策工事として、ゼネコン系の三つのＪＶと合計約五百四十二億円の契約を交わしています。」
- 事実検証できない例　「私は、築地の場外市場のようなあの温かみのある、あの何ともいえないにぎわいをつくるのが、都の責任の一つであると考えます」

▶ 立場(Stance)とは

✓ 「課題文」に対する発言者の立ち位置です。Support(支持), Against(不支持), Neutral(中立,どちらでもない) から一つ選びます。

- Support は、課題文の内容を支持していること
  - Support の例 「豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。」
- Against は、課題文の内容を支持していないこと
  - Against の例「豊洲の新市場予定地では108箇所で液状化が発生し、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。」
- Other　は、支持、不支持がはっきりしない場合や中立の場合
  - （発言者の言葉ではない）引用や報告は意見ではありません.

関連、事実検証可能、立場のラベルをつける例　(1/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、３つのラベルをつける

文ID	発言文（文脈のない一文）　※ 連続した発言ではありません。	関連	事実検証可能	立場
1	豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。			
2	豊洲の新市場予定地では108箇所で液状化が発生し、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	新銀行東京や築地市場の移転問題は非の立場です。			
4	このような中、東京都はこの八月三十日に、豊洲の土壌汚染対策工事として、ゼネコン系の三つのＪＶと合計約五百四十二億円の契約を交わしています。			
5	豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。			
6	お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります			
7	京築地域ではどのように三次医療が確保されているのかお聞かせ願います			

関連、事実検証可能、立場のラベルをつける例　(2/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、３つのラベルをつける

文ID	発言文（文脈のない一文）　※ 連続した発言ではありません。	関連	事実検証可能	立場
1	豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。	有		
2	豊洲の新市場予定地では108箇所で液状化が発生し、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。	有		
3	新銀行東京や築地市場の移転問題は非の立場です。	有		
4	このような中、東京都はこの八月三十日に、豊洲の土壌汚染対策工事として、ゼネコン系の三つのＪＶと合計約五百四十二億円の契約を交わしています。	有		
5	豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。	有		
6	お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります	無		
7	京築地域ではどのように三次医療が確保されているのかお聞かせ願います	無		

関連、事実検証可能、立場のラベルをつける例　(3/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、３つのラベルをつける

文ID	発言文（文脈のない一文）　※ 連続した発言ではありません。	関連	事実検証可能	立場
1	豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。	有	可能	
2	豊洲の新市場予定地では108箇所で液状化が発生し、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。	有	可能	
3	新銀行東京や築地市場の移転問題は非の立場です。	有	不可能	
4	このような中、東京都はこの八月三十日に、豊洲の土壌汚染対策工事として、ゼネコン系の三つのＪＶと合計約五百四十二億円の契約を交わしています。	有	可能	
5	豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。	有	不可能	
6	お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります	無	可能	
7	京築地域ではどのように三次医療が確保されているのかお聞かせ願います	無	不可能	

関連、事実検証可能、立場のラベルをつける例　(4/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、３つのラベルをつける

文ID	発言文（文脈のない一文）　※ 連続した発言ではありません。	関連	事実検証可能	立場
1	豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。	有	可能	支持
2	豊洲の新市場予定地では108箇所で液状化が発生し、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。	有	可能	不支持
3	新銀行東京や築地市場の移転問題は非の立場です。	有	不可能	不支持
4	このような中、東京都はこの八月三十日に、豊洲の土壌汚染対策工事として、ゼネコン系の三つのＪＶと合計約五百四十二億円の契約を交わしています。	有	可能	どちらでもない
5	豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。	有	不可能	どちらでもない
6	お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります	無	可能	どちらでもない
7	京築地域ではどのように三次医療が確保されているのかお聞かせ願います	無	不可能	どちらでもない

関連、事実検証可能、立場　と　Class の関係

関連	事実検証	立場	Class
有	可能	支持	Fact-checkable Support
有	可能	不支持	Fact-checkable Against
有	可能	どちらでもない	Other
有	不可能	支持	Other
有	不可能	不支持	Other
有	不可能	どちらでもない	Other
無	可能	支持	Other
無	可能	不支持	Other
無	可能	どちらでもない	Other
無	不可能	支持	Other
無	不可能	不支持	Other
無	不可能	どちらでもない	Other

課題文と文数と注釈者

14の課題文
Training data の異なり文数　約1万文
Training data の総文数　約3万文

注釈者総数　46人
異なり注釈者数　20人

番号	番号	課題文	注釈者
1	02	カジノを含む統合型リゾートを推進するべきである	3
2	04	集団的自衛を認めるべきである	3
3	05	八ッ場ダムを建設を進めるべきである	3
4	07	高齢者への医療助成を増やすべきである	3
5	08	私学助成を推進するべきである	3
6	09	中京都構想を推進するべきである	5
7	10	オスプレイを配備する	3
8	12	特定秘密保護法案を進めるべきである	3
9	13	道州制を導入するべきである	3
10	16	子ども医療費を無料化にするべきである	5
11	19	正規の教員を増やすべきである	3
12	22	生活保護の基準額を引き下げるべきである	3
13	25	東京にオリンピックを招致するべきである	3
14	26	行政の判断で空き家を処理できるようにするべきである	3

課題文	注釈者	関連有り	関連無し	事実検証可能	事実検証不可能	支持	不支持	どちらでもない	FactCheckable-Support	FactCheckable-Against	Other	文数
04-集団的自衛を認めるべきである-1515-A.json	A	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1515
04-集団的自衛権の行使を容認する-1515-B.json	B											1515
04-集団的自衛権の行使を容認する-1515-C.json	C											1515
16-子ども医療費を無料化にするべきである-241-D.json	D											241
16-子ども医療費を無料化にするべきである-241-E.json	E											241
16-子ども医療費を無料化にするべきである-241-F.json	F											241
16-子ども医療費を無料化にするべきである-241-G.json	G											241
16-子ども医療費を無料化にするべきである-241-H.json	H											241
09-中京都構想の推進について-226-B.json	B											226
09-中京都構想の推進について-226-I.json	I											226
09-中京都構想の推進について-226-J.json	J											226
09-中京都構想を推進するべきである-226-G.json	G											226
09-中京都構想の推進について-226-C.json	C											226
02-カジノを含む統合型リゾートの推進-1165-J.json	J											1165
02-カジノを含む統合型リゾートの推進-1165-K.json	K											1165
02-カジノを含む統合型リゾートの推進-1165-F.json	F											1165
08-私学助成を推進するべきである-502-L.json	L											502
08-私学助成を推進するべきである-502-G.json	G											502
08-私学助成を推進するべきである-502-M.json	M											502
10-オスプレイを配備するべきである-1174-J.json	J											1174
10-オスプレイを配備するべきである-1174-L.json	L											1174
10-オスプレイを配備するべきである-1174-A.json	A											1174
22-生活保護の基準額を引き下げるべきである-827-N.json	N											827
22-生活保護の基準額を引き下げるべきである-827-I.json	I											827
22-生活保護の基準額を引き下げるべきである-827-A.json	A											827
05-八ッ場ダムの建設について-886-I.json	I											886
05-八ッ場ダムを建設を進めるべきである-886-F.json	F											886
05-八ッ場ダムの建設について-886-H.json	H											886
26-行政の判断で空き家を処理できるようにするべきである-298-N.json	N											298
26-行政の判断で空き家を処理できるようにするべきである-298-O.json	O											298
26-行政の判断で空き家を処理できるようにするべきである-298-P.json	P											298
25-東京にオリンピックを招致するべきである-834-H.json	H											834
25-東京にオリンピックを招致するべきである-834-M.json	M											834
25-東京にオリンピックを招致するべきである-834-I.json	I											834
19-正規の教員を増やすべきである-345-M.json	M											345
19-正規の教員を増やすべきである-345-C.json	C											345
19-正規の教員を増やすべきである-345-G.json	G											345
12-特定秘密保護法案を進めるべきである-462-Q.json	Q											462
12-特定秘密保護法案を進めるべきである-462-R.json	R											462
12-特定秘密保護法案を進めるべきである-462-S.json	S											462
13-道州制を導入するべきである-1152-G.json	G											1152
13-道州制を導入するべきである-1152-B.json	B											1152
13-道州制を導入するべきである-1152-H.json	H											1152
07-高齢者への医療助成を増やすべきである-664-F.json	F											664
07-高齢者への医療助成を増やすべきである-664-T.json	T											664
07-高齢者への医療助成を増やすべきである-664-C.json	C											664
												31807

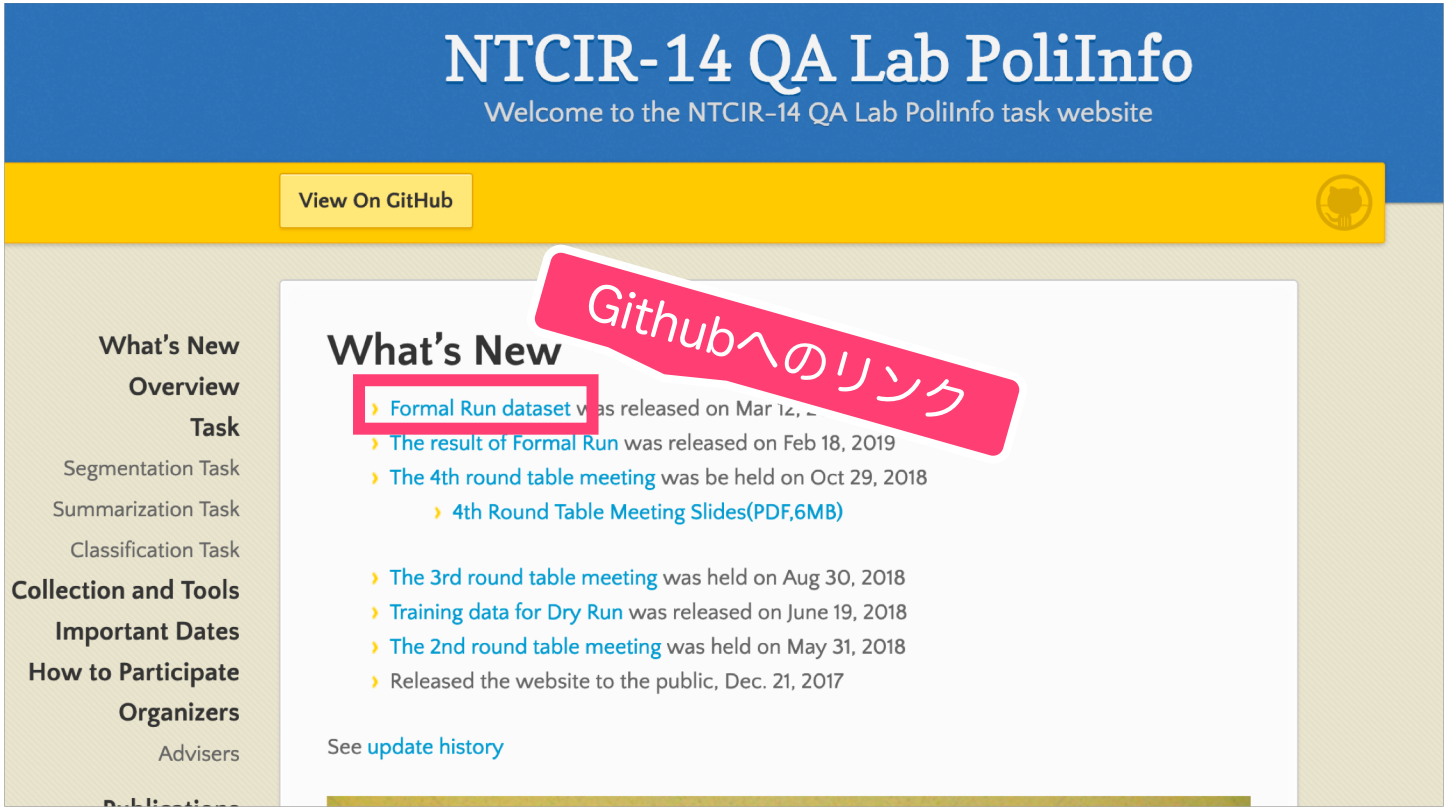
Classification Training dataについて

14の課題文

異なり文数　約1万文
3〜5人の注釈文の総数　約3万文

注意：注釈者への説明の際に、３つのサブクラスの用語を異なる表現（例えば、「根拠あり」と「事実検証可能」）で説明していますが、注釈付けするラベルの意図は、同じです。

Github での公開(1/2)



Github での公開(1/2)

kmr-y / NTCIR14-QALab-PoliInfo-FormalRunDataset Private

Unwatch 2 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

No description, website, or topics provided.

Edit

Manage topics

20 commits | 1 branch | 0 releases | 1 contributor

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

kmr-y Update README.md — Latest commit e868160 34 minutes ago

Classification	READMEと評価スクリプト追加	6 hours ago
Segmentation	READMEと評価スクリプト追加	6 hours ago
Summarization	READMEと評価スクリプト追加	6 hours ago
README.md	Update README.md	34 minutes ago

README.md

Overview

Github での公開(2/2)

kmr-y / NTCIR14-QALab-PoliInfo-FormalRunDataset

Unwatch 2 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

Branch: master | NTCIR14-QALab-PoliInfo-FormalRunDataset / Segmentation / | Create new file | Upload files | Find file | History

kmr-y 東京都議会会議録のJsonファイル — Latest commit 3defdef a day ago

..		
test	Formal Run Dataset	2 months ago
testGS	Formal Run Dataset	2 months ago
training	Formal Run Dataset	2 months ago
Pref13_tokyo.json	東京都議会会議録のJsonファイル	a day ago
README.md	READMEと評価スクリプト追加	7 days ago
poliinfo_check_format2.py	READMEと評価スクリプト追加	7 days ago
poliinfo_eval_formal_segmentation.py	READMEと評価スクリプト追加	7 days ago

まとめ

- NTCIR14 QALab-PoliInfo とは
- 政治情報
  - 地方議会会議録，議会だより
- ３つのタスクについて
  - Segmentation task, Summarization task, Classification task
- データセットの構築方法
  - クラウドソーシングなど
- Github での公開
  - ✓ https://github.com/kmr-y/NTCIR14-QALab-PoliInfo-FormalRunDataset

関連発表

P8:ポスター(4)　3月15日(金) 13:10-14:40　ES会議室

[P8-11]　NTCIR-14 QA Lab-PoliInfo Formal RunにおけるSummarization Taskの結果と評価

渋木, 木村, 乙武, 内田, 高丸, 阪本, 石下, 三田村, 森, 神門