

# 一般公開版「都道府県議会会議録検索システム」の概要

## Public Release of a Web-based System for Visualizing Prefectural Assembly Minutes in Japan

乙武北斗<sup>\*1</sup>  
Hokuto Ototake

高丸圭一<sup>\*2</sup>  
Keiichi Takamaru

内田ゆず<sup>\*3</sup>  
Yuzu Uchida

木村泰知<sup>\*4</sup>  
Yasutomo Kimura

<sup>\*1</sup> 福岡大学  
Fukuoka University

<sup>\*2</sup> 宇都宮共和大学  
Utsunomiya Kyowa University

<sup>\*3</sup> 北海学園大学  
Hokkai-gakuen University

<sup>\*4</sup> 小樽商科大学/AIP  
Otaru University of Commerce / AIP

This paper illustrates a web-based system for visualizing prefectural assembly minutes in Japan. It was public-released in February 2018. Although many local governments in Japan provide public access to political documents, researchers and scientists cannot take full advantage of them because they consist only of raw text. Therefore, there is a need for these to be converted into panel data and to visualize differences by region or over time. For ordinary citizens, visualizing differences between the topics discussed by assembly members can be useful in assessing their characteristics. The proposed system aims to provide a complete solution that meets all of these requirements. The system has five main functions: i) full-text search for political documents, ii) showing keywords in context, iii) map and time-series visualization, iv) cross-tabulation, and v) political keyword extraction. Although three existing systems offer some of these functions, the proposed system integrates all the five necessary functions and provides a complete solution that meets all of the requirements.

### 1. はじめに

筆者らは科研費研究課題「地方議会会議録コーパスの構築とその学際的応用研究」(2010～2013年度)において、全国の自治体の地方議会会議録の収集・整理を進め、423自治体の議会会議録を収録した「地方議会会議録コーパス」(約1億3400万文、約80GB)を構築した(第Ⅰ期コーパス)。これを活用し、情報工学の分野における情報抽出手法の開発[木村 2013], [葦原 2015], [木村 2011]や、言語学分野における言語使用や言語変異の実証的研究[高丸 2015], [井上 2013], [二階堂 2015]が進められてきた。第Ⅰ期コーパスは情報工学分野、言語学分野においては、有用な大規模言語資源としての利用可能であることが明らかとなった。しかし、収集対象の自治体の範囲や期間が統制されていないコーパスであるため、地域間の比較を統計的に行うような研究には活用しづらいという問題があった。

現在、科研費研究課題「議論の背景・過程・結果を関連づける地方政治コーパスの構築とその学際的応用」(2016～2019年度)において、新たに地方政治における議論の過程と結果、およびその背景を結びつけた「地方政治コーパス」の構築を進めている(第Ⅱ期コーパス)。このコーパスでは、比較分析を定量的に行うことができるように、地方議会会議録の収集条件(対象期間・対象自治体)を整えたコーパスの整備を行っている。新たに構築する会議録コーパスは、データ規模は小さくなるものの、時系列(時間に伴う変化)、および、横断面(ある時点の地域による変異)の分析が可能なものとなる。現在、第Ⅱ期コーパスの収集が進み、これを利用した統計的分析等が進められつつある。

収集したデータを学術研究だけでなく、一般の市民が関心を持ったことばを検索できるようにすることで、地方政治への関心をさらに高めることができ、筆者らとしても、研究成果の社会還元を行うことができる。このため、収集した会議録の全文検索および検索結果の可視化が可能なWebシステムの実装を進め、第Ⅱ期コーパスの一部を「都道府県議会会議録検索システム」(<http://local-politics.jp/47pref/>)として一般公開することとした。本稿では、2018年2月に一般公開したシステム

連絡先: 乙武北斗, 福岡大学工学部, 福岡県福岡市城南区七  
限 8-19-1, ototake@fukuoka-u.ac.jp

の概要について述べる。収集データについては2章で述べる。システムの概要について3章で述べる。検索結果の表示機能の特徴について4章に述べる。

### 2. 収録データ

#### 2.1 データの収集と整理

「都道府県議会会議録検索システム」には、平成23年4月の統一地方選挙から4年分(平成27年3月まで)の全国47都道府県議会の本会議会議録が収録されている。

会議録は句点を基準に1発言ごとに分割し、それぞれの発言の発言者名を抽出した上で、会議の開催情報と併せて、表1に示す17項目のフィールドをもつテーブルに登録した。

表1 データベースのフィールド

1. 識別子(主キー、発言文に固有に振られたID)
2. 自治体名(「北海道」「青森県」「宮城県」…)
3. 会議名(「平成24年2月定例会」「第335回」など)
4. 号(各会議の会期における号数(日数))
5. 年(開催年(和暦))
6. 月(開催月)
7. 日(開催日)
8. 会議種別(定例/臨時)
9. 会議名フルテキスト (例:「平成24年 2月 定例会(第335回)-02月27日-04号」)
10. 発言者フルテキスト(「62番岩本剛人君」)
11. 発言者ID(発言者テーブルを参照する外部キー)
12. 発言者名(例:「岩本剛人」)
13. 発言者の役職(例:「62番」)
14. 発言文(例:「次に、冬の節電対策について伺います。」)
15. 発言以外の記録文(「(拍手)」[「…君登壇」][「-----」など])
16. 原本URL
17. HTMLファイルのパス

収集した会議録から、会議名や発言者名の抽出を行った。さらに、発言者を同定するために人手によって、名寄せを行った。特に、議員については発言者を識別する「発言者ID」を付し、

議員の属性情報を格納したテーブルを外部参照できるようにした。(詳しくは[高丸 2017]参照)

なお、現在のところ、発言文と発言以外の記録文の分割は行っていない。このため、発言以外の記録文はすべて直前の発言者の発言の一部として、発言文に格納されている。この点は今後の課題である。

表 2 都道府県別の収録データ数

都道府県	発言レコード数	発言文字数	発言者数
北海道	115,091	6,690,886	142
青森	96,095	6,075,997	123
岩手	83,436	5,304,262	120
宮城	123,815	6,591,736	113
秋田	57,635	3,564,034	65
山形	33,075	2,219,422	98
福島	69,741	4,106,475	313
茨城	70,492	4,387,984	210
栃木	81,452	4,086,005	101
群馬	104,851	5,664,259	134
埼玉	139,901	6,216,179	189
千葉	133,345	6,966,264	150
東京	115,750	5,746,805	240
神奈川	101,895	5,991,067	148
新潟	90,476	5,008,553	104
富山	74,132	4,688,790	86
石川	72,677	4,413,817	110
福井	87,817	4,568,260	96
山梨	79,247	4,274,357	119
長野	136,055	7,597,746	133
岐阜	67,610	4,283,206	100
静岡	91,526	5,456,750	136
愛知	88,416	5,881,919	179
三重	75,893	5,047,765	123
滋賀	142,596	8,646,095	88
京都	72,824	5,031,745	166
大阪	95,929	4,645,928	181
兵庫	74,480	4,223,394	138
奈良	71,490	4,193,615	114
和歌山	62,970	3,473,289	111
鳥取	172,523	10,844,069	101
島根	106,708	6,010,462	95
岡山	111,268	6,392,762	111
広島	54,298	3,357,243	113
山口	82,116	4,995,372	95
徳島	69,015	3,797,876	94
香川	68,475	4,029,591	105
愛媛	69,609	3,999,989	108
高知	83,058	5,859,125	102
福岡	84,918	4,933,062	106
佐賀	97,474	5,760,078	92
長崎	135,966	6,585,947	112
熊本	82,164	4,771,578	123
大分	94,297	4,666,031	112
宮崎	103,344	6,546,640	88
鹿児島	113,984	7,266,893	100
沖縄	155,947	7,501,424	119
合計	1,651,676	92,290,795	2,542

## 2.2 収録データ数

表 1 に都道府県別の発言レコード数, 発言文字数(発言フィールドの文字数), 発言者数を示す. 総収録データは 1,651,676 レコードであり, 総文字数は 92,290,795 文字である. また, 名寄せの結果得られた発言者数は合計 2,514 名であった.

## 3. システム

### 3.1 概要

本システムは, 2 章で述べた 47 都道府県議会の本会議会議録を検索・視覚化するウェブアプリケーションである. ユーザは一般的なウェブブラウザを用いて, 会議録を検索し, 結果を視覚化することができる. システムのアウトラインを図 1 に示す. ①システムが検索クエリを受け取ると, ②クエリは全文検索エンジンに送られる. ③検索エンジンから検索結果を受け取った後, ④システムはユーザのリクエストに応じて検索結果を可視化して出力する. 全文検索エンジンには Elasticsearch<sup>1</sup>を用いた.

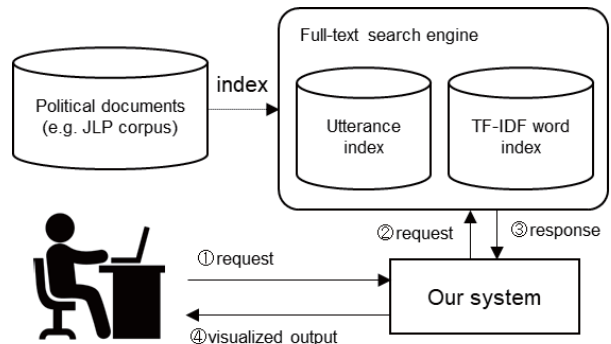


図 1 システムのアウトライン



図 2 システムのメニュー画面

一般公開システム<sup>2</sup>のトップページの様子を図 2 に示す. はじめて訪れたユーザが試用しやすいように, 検索文字列に「オリンピック」をデフォルト値として入力している. 右側の「検索」ボタン, または, 下部のアイコンをクリックすることにより, 検索を開始することができる. また, 上部には, 「発言検索」「マップ検索」「クロス

<sup>1</sup> <https://www.elastic.co/>

<sup>2</sup> <http://local-politics.jp/47pref/>

表検索「時系列グラフ」統計」の5つのメニューがあり、ここから本システムが有するすべての機能へアクセスすることができる。

「発言検索」メニューには「発言検索」「KWIC」「TF・IDF キーワード検索」のサブ項目がある。「マップ検索」メニューには「発言マップ検索」「TF・IDF マップ検索」のサブ項目がある。各機能の概要は次節以降に述べる。

「統計」メニューをクリックすると、本システムが提供するデータに関する以下の統計情報を表示することができる。

- 都道府県発言・文字数
- 都道府県発言・文字数(年度別)
- TF・IDF 上位 100(自治体・発言者単位)
- TF・IDF 上位 100(自治体の発言者平均)

## 4. 機能

### 4.1 全文検索(発言検索)

「全文検索」機能は、検索文字列を含む発言文(レコード)を抽出する。図3に全文検索の例を示す。「検索文字列の扱い」で「N-gram 分割」を選択すると、検索文字列を2-gram および3-gram に分割し、これらを含む検索結果を表示する。「フレーズ完全一致」を選択すると、検索文字列全体をそのまま含む検索結果を表示する。「複数検索語の扱い」では、空白によって区切られた複数の検索文字列を AND 検索するか、OR 検索するかを選択することができる。「出力結果のソート」では、結果の表示順を「スコア降順」、「日付降順(昇順)」、「自治体コード降順(昇順)」、「発言文字数降順(昇順)」から選択することができる。さらに「詳細な条件を設定」をチェックすることで、検索対象の自治体および発言時期を限定した検索を行うことができる。

The screenshot shows a search interface for 'オリンピック' (Olympics). It includes filters for search method (N-gram division, phrase match), search logic (AND/OR), and sorting options. Three results are displayed, each with a score and word count, and buttons to view the top 5 or bottom 5 sentences.

順位	都道府県	年度	議題	ID	スコア	文字数
1	長崎県	24	平成24年8月定期月議会-09月12日-02号	420000_240912_540	97.11	22
2	新潟県	25	DiscussNetPremium 平成25年12月定期会 本会議-12月09日-一般質問-03号	150002_H8013164_10	89.96	29
3	東京都	24	平成25年第1回定例会(第4号)	130001_250228_1512	87.43	61

図3 全文検索の例

### (1) 文脈の展開機能

検索結果として表示される情報は、議会における発言の一部分(1文)である。その発言の前後にどのような発言があるかを知りたいという要望に答えるために、全文検索の検索結果には「前5発言を取得する」「後5発言を取得する」ボタンを用意した。これをクリックすることで、5文単位で発言を遡ったり、発言を辿ったりすることができる。

### (2) KWIC 表示

また、上部のメニューから「発言検索→KWIC」を選択すると、検索結果を KWIC 形式[Luhn 1960]で表示し、前文脈と後文脈を一覧することができる。この例を図4に示す。

#	属性	前文脈	キーワード	後文脈
1.1	長崎県 24		オリンピック	ならぬ長崎オリンピック。
1.2	長崎県 24	ロンドンオリンピックならぬ長崎	オリンピック	。
2.1	新潟県 25		東京オリンピック	、東京都オリンピックという議論もあります。
2.2	新潟県 25	東京オリンピック、東京都	オリンピック	という議論もあります。
3.1	東京都 24	ロンドン・	オリンピック	のことを何か失敗したかのようにおっしゃっていたが、ロンドン・
3.2	東京都 24	のことを何か失敗したかのようにおっしゃっていたが、ロンドン・	オリンピック	は成功したオリンピックです。
3.3	東京都 24	のようにおっしゃっていたが、ロンドン・	オリンピック	は成功したオリンピックです。
4.1	静岡県 25	まず要領ですが、	オリンピック	、パラリンピックに関連したものでありますが、東京オリンピック
4.2	静岡県 25	オリンピック・パラリンピックに関連したものでありますが、東京	オリンピック	は夏季オリンピックでありますね。
4.3	静岡県 25	リンピックに関連したものでありますが、東京	オリンピック	でありますね。

図4 KWIC 表示の例

### 4.2 マップ検索

「マップ検索」機能は、検索結果の件数を都道府県別に集計し、地図上で塗りつぶしの濃度によって可視化する機能である。地図をクリックすることで、当該地域での発言を抽出して、全文検索表示することができる。マップ検索の例を図5に示す。検索文字列は「離合」である。

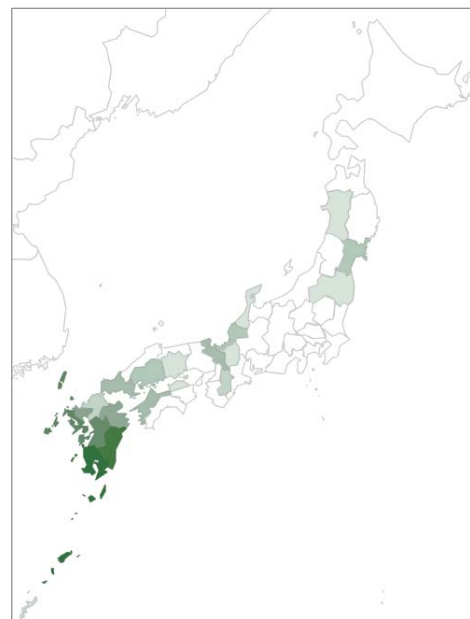


図5 マップ検索の例

### 4.3 クロス表検索

「クロス表検索」機能は、「行要素」および「列要素」にそれぞれ項目名を指定することで、検索結果の件数のクロス集計結果を表形式で表示する機能である。行要素・列要素として指定できる項目は、発言地域「都道府県」、発言時期「年度」または「四半期」、発言者「発言者名」または「議員生年年代(10年ごと)」の5種類である。クロス表内の数値を右クリックし「この条件で発言検索」をクリックすることで、行要素・列要素に指定した条件に合致する発言文を全文検索形式で表示できる。

クロス表検索の例を図6に示す。検索文字列に「オリンピック」、行要素に「都道府県」、列要素に「議員生年年代」を設定した検索結果である。

表を最大化する		すべてのセルを選択する		処理時間: 67 ms			
1	都道府県 \ 議員生年年代	1930	1940	1950	1960	1970	1980
2	北海道	0	12	34	27	27	0
3	青森	0	38	38	43	31	10
4	岩手	0	12	4	20	9	0
5	宮城	0	80	70	12	17	0
6	秋田	0	15	33	0	3	0
7	山形	0	5	32	51	1	0
8	福島	0	27	84	23	30	0
9	茨城	0	30	49	23	22	0
10	栃木	0	27	27	57	22	6
11	群馬	0	11	22	35	54	3
12	埼玉	6	63	49	51	75	16
13	千葉	0	104	69	105	47	8
14	東京	6	115	490	347	191	9
15	神奈川	9	16	88	70	83	11
16	新潟	1	25	50	53	34	0
17	富山	0	9	38	20	4	9
18	石川	0	26	30	29	6	0
19	福井	0	50	31	13	0	0
20	山梨	15	81	37	81	9	0
21	長野	1	16	88	49	35	0
22	岐阜	0	33	16	49	10	3
23	静岡	0	45	24	97	41	2
24	愛知	0	9	41	47	22	0
25	三重	0	16	26	18	21	0
26	滋賀	0	8	45	29	12	0
27	京都	0	7	15	35	14	0

図6 クロス表検索の例

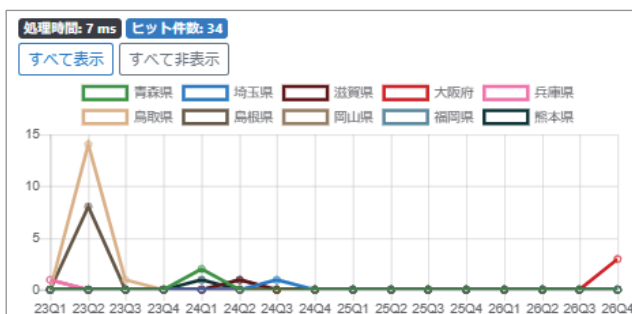


図7 時系列グラフの例

### 4.4 時系列検索

「時系列検索」機能は、都道府県別に四半期単位で検索結果の件数を集計し、折れ線グラフで表示する機能である。都道府県名または「すべて表示」「すべて非表示」をクリックすることで、グラフ上に表示する都道府県を切り替えることができる。時系列検索の例を図7に示す。検索文字列は「生レバー」である。

### 4.5 TF-IDF 検索

全都道府県の議員の発言を対象に TF-IDF によって特徴語を求めた。議員特徴語を検索するために「発言検索」メニュー、「マップ検索」メニューに「TF-IDF キーワード検索」「TF-IDF マップ検索」を用意した。

「TF-IDF キーワード検索」では、検索文字列の TF-IDF 値が高い順に、都道府県名、議員名、TF-IDF 値を一覧表示する。また、「TF-IDF マップ検索」では、検索文字列の都道府県内の議員の TF-IDF 値の合計を塗りつぶしの濃度によって地図上に可視化する。

## 5. まとめ

地方議会会議録の横断検索を提供する Web システムは既に複数存在する(例えば「chiholog」(<http://chiholog.net/>)や「議事ロックス」(<https://giji.rocks/>))。これらのサービスでは、現在進行形で最新の地方議会会議録が収集・提供されている。しかしながら、名寄せ処理が行われていないため、検索結果が「誰の」発言であるかを条件に入れた検索や集計を行うことはできない。

筆者らが一般公開した「都道府県議会会議録検索システム」は、4年間の限定された期間を対象とした横断検索システムであるが、「いつ」「どこで」「誰が」発言したかを検索し、検索結果を自動的に集計・可視化する機能を有する。「誰の」発言であるかが明らかであるため、TF-IDF に基づく議員別の特徴語を検索する機能を有する。本システムが、地方議会における議員の活動を明らかにし、市民の地方政治への参画意識が増進する手助けになるように、今後もデータの増強、システムの周知を進めていく予定である。

## 謝辞

本研究の一部は JSPS 科研費 JP16H02912, JP17K02739 による。

## 参考文献

- [木村 2013] 木村泰知, 関根聡:主辞に基づく政治問題抽出手法, 人工知能学会論文誌, 28(4), 370-378, 2013.
- [葦原 2015] 葦原史敏, 木村泰知, 荒木健治:地方議会会議録における節単位による議員の要望抽出, 電子情報通信学会論文誌, J98-D(11), 1390-1401, 2015.
- [木村 2011] 木村泰知, 洪木英潔, 高丸圭一, 乙武北斗, 小林哲郎, 森辰則:地方議員マッチングシステムにおける能動的質問のための質問生成手法, 人工知能学会論文誌, 26(5), 580-593, 2011.
- [高丸 2015] 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知:地方議会会議録コーパスにおけるオノマトペ出現傾向と語義の分析一, 人工知能学会論文誌, 30(1), SP2-K, 306-318, 2015.
- [井上 2013] 井上史雄:去った〇日, ことばの散歩道, 明治書院, 154-155, 2013.
- [二階堂 2015] 二階堂整, 川瀬卓, 高丸圭一, 田附敏尚, 松田謙次郎:地方議会会議録による方言研究—セミフォーマルと気づかない方言一, 方言の研究, 1, 299-324, 2015.
- [高丸 2017] 高丸圭一, 内田ゆず, 木村泰知:地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析, 宇都宮共和大学シティアライヴ学論叢, 18, 136-155, 2017.
- [Luhn 1960] Luhn, H. P.:Key word-in-context index for technical literature (kwic index), Journal of the Association for Information Science and Technology, 11.4, 288-295, 1960.