

議会議事録の自動タグ付けに基づく探索的閲覧システムの試作とそのユーザビリティ評価

成瀬 雅人[†], 白松 俊[†], 松島格也[‡]
名古屋工業大学[†] 京都大学[‡]

1. はじめに

地方議会の議事録は市民に向けて公開がなされているが、その議事録自体が非常に長く、また構造が分かりにくいいため、市民に読んでもらうために公開されているが、実際は市民が気軽に閲覧できる形にはなっていない。

そこで、本研究では地方議会議事録の発言と、探索の手掛かりとなるような外部データセットを紐づけ、その紐づけ結果をもとに探索的閲覧が可能な閲覧システムを試作する。閲覧の手掛かりとなるような外部データセットには、自治体の Web ページで公開されているような分野階層構造と、地図情報を用いる事とする。

そして、試作した探索的閲覧システムの評価実験を行うことで、本システムが議事録の閲覧および理解に対してどれほど支援が出来ているかを検証する方法を提案する。

なお、本研究では地方議会の議事録の例として名古屋市会の会議録¹を利用する。

2. 地図情報を用いたインタフェース

探索的閲覧 (Exploratory browsing) とは、ユーザー自身が必要な情報を十分に把握していない場合にリンクを辿りながら見つかった情報をヒントにして必要な情報を探す行動である [1]。

議論中の発言には議会のある地方の地名が多く含まれており、その地名を抽出して発言と地名を紐づける。紐づけた情報を元に閲覧者が住んでいる地域に関連するような内容が見たい場合に、地名から議事録の内容を検索するという事が可能になると考えられる。

発言に紐づける地図情報として本研究では施設名、駅名、地域名を用いる。

施設名と駅名はそれぞれ OSM 地図データから Overpass API を用いて "amenity" キーと "railway" キーを持つデータとして取得し、GeoJSON 形式で保存する。

また、地域名は Linked Open Addresses Japan² から住所データを Linked Open Data として取り出す。

これらの地域名や、施設名はそのままと議事録の検索にあまり適さないものが多く存在する。例として、正式名称が長くて全ては呼ばれないものや、固有名詞以外にも意味があり、関係のない所で使われる可能性が高いものがある。これらは名称を加工あるいは、使用しないようにする事によってノイズを削減する。

得られた施設名、駅名、地域名は発言の中に同じ単語が存在しているかどうかを確認し、単語が存在している発言との組み合わせを集めた GeoJSON ファイルを作成する事で地図情報の自動タグ付けを行う。

地図情報から議事録の発言を検索するインタフェースには、地図表示のための JavaScript ライブラリである leaflet.js³ を使用し、土台となる地図には OSM を用いる。

実際に試作したインタフェースを図 1 に示す。



図 1: 地図から検索するインタフェース

図 1 では画面の左側に名古屋市の地図が表示され、地図の右上にあるボックスから表示したい名古屋市の区名を選択する事ができるようになっている。また、地図上に表示されているピンをクリックする事によってその施設名、駅名、地域名が現れる発言が画面の右側に表示される。

Implementing a System for Exploratory Browsing of Council Minutes based on Automatic Tagging and Its Usability Evaluation
Masato Naruse[†], Shun Shiramatsu[†], Matsushima Kakuya[‡]
[†] Nagoya Institute of Technology, [‡] Kyoto University

¹ <http://www.kaigiroku.net/kensaku/nagoya/nagoya.html>

² <http://uedayou.net/loa/>

³ <http://leafletjs.com/>

3. 分野階層を用いたインタフェース

議事録の発言はその地域に関連する発言が多くなるため、本研究では紐づける外部データセットとして同地方の自治体の Web ページを利用する。Web ページに存在する階層構造からタグとして使うため各階層を分野名として取り出し、発言と紐づけを行う。本研究では議事録が名古屋市のものなので、名古屋市 Web ページの分野構造を外部データセットとして使用する。

議事録の発言と分野名を紐づける為、議事録の発言と分野名の内容を用いて文章の類似度計算を行う。

類似度計算には sentence2vec⁵ を使用する。コーパスには Wikipedia から、行政タグと名古屋市タグからそれぞれ 8 階層分を取り出してまとめた物を使用する。

Sentence2vec では、比較したい分野名以下の階層に存在する文章を一つの巨大な文として扱う。求めたコーパスから各文章の Paragraph vector [2] を求める。分野と発言の Paragraph vector が求められたら、分野と発言を一つずつ選んで比較を行う。比較には Cos 類似度を用いる。

Cos 類似度が 1 に近いほど内容が似ているとみなされる。本研究では、文書が類似しているとする閾値を経験的に 0.45 と設定し、この値を越えた組み合わせの場合に発言に対して類似している分野名をタグとして追加する事で分野と発言を紐づける。求めた類似度から、分野階層構造を用いて議事録を検索するためのインタフェースを図 2 に示す。

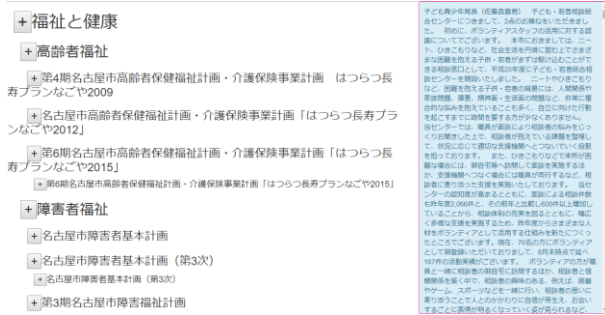


図 2: 分野名から検索するインタフェース

図 2 では、画面の左側に分野を提示し、検索したい分野を選ぶ事でその分野に関連した発言を画面の右側に表示している。また閲覧者は分野の中でも深く掘り下げて見たい分野を階層的に辿ることができ、より細かい範囲で閲覧したい情報あるいは興味のある分野を探索することができる。

4. 各インタフェースの評価実験計画

試作した地図情報を用いたインタフェースと分野階層を用いたインタフェースのユーザビリティを評価するため評価実験を行う予定である。

実験では試作した 2 つインタフェースと公開されている名古屋市議会の議事録ページを用いてそれぞれのユーザビリティを比較する。実験では 2 つのタスクを設定する。1 つ目は明確なタスク、例えば名古屋駅といった固有名詞に関連した発言の検索をタスクとして設定する。2 つ目は多少曖昧なタスク、例えば「あなたは昭和区に住む子育て中の主婦として気になる事を検索してください」といったタスクを設定する。これら 3 つのインタフェースと 2 つのタスクを組み合わせた 6 通りの条件に対し、以下の 3 つの評価項目で 7 段階評価を行う。

- 発言の検索しやすさ
- 素早く発言が探せたか
- 興味のある発言を探すことができたか

また曖昧なタスクを設定した場合に、与えられた曖昧なタスクから、どれだけ具体的な問題に落とし込めたかを評価する事も必要だと考えられる。評価方法として、関連しているあるいは重要だと考えた発言をいくつ見つける事ができたかを記録し、指標とするという方法が考えられる。

5. おわりに

本稿では探索的閲覧のための議事録の外部データセットとの紐づけ及び探索インタフェースの試作をした。今後は本稿ではできなかった評価実験を行い、よりユーザビリティの高いインタフェースへの改善を行う予定である。

また、本稿では名古屋市議会の議事録を扱ったが、他の自治体のデータでもこの研究が利用できるような拡張が必要だと考えられる。そして、この構造化手法から公的な討論の場で話のタネとして地方議会の議事録を活用することを目指す。

謝辞 本研究の一部は、JICE 研究開発助成、JSPS 科研費 (No. 17K00461)、および JST CREST (No. JPMJCR15E1) の支援を受けた。

参考文献

[1] White, R. W., et al. "Exploratory search: beyond the query-response paradigm." Morgan and Claypool, 2009.
 [2] Le, Q. V., et al. "Distributed Representations of Sentences and Documents." In ICML, Vol. 14, pp. 1188-1196, 2014.

⁵ <https://github.com/klb3713/sentence2vec>