

探索的閲覧のための地方議会議事録の構造化手法の検討

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に含まれる発言を、探索の手掛かりとなるような外部データセットと紐づけを行うことで、探索的な閲覧が可能な形へ構造化する。閲覧の手掛かりになるような外部データセットの例として、自治体の Web ページで公開されているような分野の階層構造や、地図上の地理データを用いる。

この構造化によって、長く構造の分かりづらい議事録の閲覧および理解が支援されるかどうかを検証する。

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. 紐づける外部データセット

関連した内容の多い、同地方の自治体の Web ページを紐づけの対象に取ることにより、地方議会でのどの発言がどの分野に近い発言であるかを調べる事ができる。本研究では名古屋市 Web ページの分野階層²を用いる。

また、議論中の発言には地名が含まれている場合が多く、発言から地名を抜き出して発言と地名を紐づける事ができれば、閲覧者が自分の住んでいる地域に関連する内容の議事録を見たいときなどに、地域名から議事録を検索するという事が可能となる。本研究では地図情報として、Linked Open Addresses Japan³を利用する。

また、これらの外部データセット以外にも、自治体の担当部署ごとに関連する発言を紐づけるなど様々なデータセットとの紐づけが今後の展望として考えられる。

3. 探索的閲覧のためのインターフェイス

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]。

探索的閲覧を支援するために、使いやすく理解しやすいインターフェイスを作成することは重要な要素となってくる。そこで本研究では、自治体 Web ページの公開している分野の階層構造から分野名によって議事録を検索するためのインターフェイスと、地図を用いて閲覧者が自分の興味のある地図上の位置からその地域に関連のある議事録を検索するインターフェイスを検討する。図 1, 2 に、分野名から発言を探索する場合と地図から探索する場合のインターフェイスのモックアップを示す。

図 1 では、閲覧者に分野を提示し、検索したい分野を選ぶ事でその分野に関連した発言が表示されている、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。

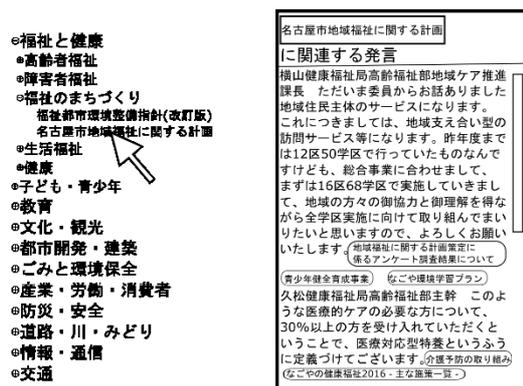


図 1: 分野名から検索するインターフェイス

Considering Structuring Method of Council Minutes for Exploratory Browsing

Masato Naruse[†], Shun Shiramatsu[†], Matsushima Kakuya[‡]

[†]Nagoya Institute of Technology, [‡]kyoto University

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

² <http://www.city.nagoya.jp/shisei/category/53-0-0-0-0-0-0-0-0-0.html>

³ <http://uedayou.net/loa/>



図 2: 地図から検索するインターフェイス

図 2 では、閲覧したい地域を選ぶ事でその地域に関連する発言を閲覧する事ができる。

4. 議事録との類似度推定

地図と議事録を紐づける為、議事録の発言に現れた地名をタグとして発言に追加する。タグとして付与する地名は Linked Open Addresses Japan を用いて統一された形態で付与する。

また、分野名との議事録の紐づけを行うため、議事録中の発言と分野との内容がどれほど類似しているかを検証する。本研究では自治体 Web ページの分野の階層構造を利用し、分野ごとに内容を取得する事で、議事録の発言と比較を行う。

文章の比較には sentence2vec¹ を使用し、Paragraph vector [2] を求める。まず、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。例として、上記図 1 より福祉のまちづくりであれば、福祉都市環境整備指針(改訂版)、名古屋市地域福祉に関する計画が含まれる事になる。議事録の全発言と全分野をコーパスとしてこの文の Paragraph vector を求める。なおベクトルの次元数は 100 とする。同様に議事録からも発言を全て抜き出して Paragraph vector を求め、これらと比較する。比較には Cos 類似度を用いる。文書 a の Paragraph vector の i 番目の要素を $s(a, i)$ とおくと、文書 a と文書 b の Cos 類似度は以下のように求められる。

$$\text{Cos}(a, b) = \frac{\sum_{i=1}^m s(a, i)S(b, i)}{\sqrt{\sum_{i=1}^m s(a, i)^2} \cdot \sqrt{\sum_{i=1}^m s(b, i)^2}}$$

この値は -1 から 1 の範囲をとり、1 に近いほど二つの文書は内容が類似しているとみなされる。この値が閾値を越えた場合、議事録の発言に関連する分野名をタグとして追加することで分野名と類似する議事録の発言を得る事ができる。

表 1 に、発言の要約と、その発言に自動付与されたタグの例を示す。表 1 では比較的妥当なタグ

表 1: 発言に自動付与されたタグの例

発言の要約	自動付与されたタグ
公園における暑さ対策につきまして、来年度の予算においてどのように反映しているか…(以下略)	「公園経営の取り組み」 「主な公園」 「公園経営基本方針」
耐震性防火水槽の設置につきましては、そういった地震の…(以下略)	「平成 28 年秋の火災予防運動」「地震防災対策強化地域とは」「名古屋市建築物耐震改修促進計画」
(略) 東山動植物園における生息域外保全の取り組みにつきまして…(以下略)	「水の環境活 2050 なごや戦略について」 「なごや生物多様性歳時記季節の生きものカレンダー」 「なごや環境学習プラン」

が付与された例を示したが、類似度の閾値の設定によっては、妥当でないタグが付与される事例もあった。閾値を高くすればタグの質が高くなるが付与されるタグの数が少なくなるため、適切な閾値を得る必要があり、どのように適切な閾値を取得するかが、今後の課題として挙げられる。

5. 関連研究

木村ら [3] は、地方議会の会議録のフォーマットが自治体ごとに異なっている事や、各研究者が重複するようなデータの電子化を行っている事を背景に、地方議会会議録コーパスの構築を行っている。本研究においても名古屋市以外の会議録を対象とした場合にフォーマットの違いをどうするかという問題点が挙げられる。

6. おわりに

本稿では探索的閲覧のための地方議会議事録の構造化手法の検討について述べた。今後は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。また、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 本研究は、JICE 研究開発助成、JSPS 科研費 (25870321)、および JST CREST 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.
- [3] 木村 他: 地方議会会議録コーパスの構築とその利用. 第 26 回人工知能学会全国大会, 3B3-NFC-4-3, 2012.

¹ <https://github.com/klb3713/sentence2vec>