

法令改正・政策立案プロセスの把握と政策動向の理解支援 に向けた RAG 要約システムの構築

水島陵太¹ 林田歴¹ 水越尚子² 乙武北斗¹

¹ 福岡大学 ² 株式会社 polisee

{td252016,td252010}@cis.fukuoka-u.ac.jp nmizukoshi@polisee.com
otokake@fukuoka-u.ac.jp

概要

政府会議資料から法令改正・政策立案過程や政策動向を把握したい需要が高まるが、資料は回ごとに追加・更新され論点が分散し継続理解が難しい。そこで本研究では政府会議資料を OCR し、出典付き要約を生成する RAG システムを構築した。本システムは複数回にまたがる論点を根拠箇所とともに参照でき、理解支援と資料探索負担の軽減が期待できる。検索傾向の分析で、OCR では情報量の少ない短いチャンクが上位に現れやすいという課題が確認されたため、本稿では OCR での前処理や今後の改良方針について議論する。

1 はじめに

政府が公式に公開する一次情報をもとに、法令改正や政策立案のプロセスをモニタリングすることで、政策決定の背景や論点を理解し、企業で活用したいというニーズが高まっている。一方で、法令・政策立案に至るプロセスは、審議会、パブコメ、国会提出法案といった資料が複数にまたがり、資料や議論の痕跡が分散しやすい。加えて政府会議は、会議回ごとに資料が追加・更新され参照すべき資料が増え続けるとともに、規制強化・緩和といった議論の方向性が回次によって変化しやすいため、特に非専門家が政策決定の流れや論点を継続的に追うことは難しい。こうした状況に対して、法令改正・政策立案プロセスを可視化し、関連する会議体や資料を体系的に可視化するプラットフォームとして polisee¹⁾が提供されている。

また、国会議事録を対象とした要約システムの研究 [1] があるものの、政府会議資料を対象とした研究は少ない。本研究は、polisee により収集・整理さ

れた政府会議資料を一次情報として、RAG を用いて会議回をまたぐ根拠を横断的に検索・統合し、政策プロセスの進捗と論点推移を要約として把握するための支援基盤の第一歩を示す。

本システムのタスクは、(1) 根拠箇所の検索、(2) 根拠に基づく要約生成（必要に応じた複数資料の統合）、(3) 資料名・該当箇所の提示である。本稿では (1) と (2) に焦点を当て、理解支援の前提となる根拠検索の網羅性・ノイズの検証、および専門家による要約生成結果の適切さの評価について述べる。

2 データセット

2.1 使用した会議資料

政府会議に関連して公開されている資料には、議事概要に加え、図 1 のように議論の方向性を示すために事務局が作成した資料（事務局資料）、法令資料、およびパブリックコメント関連資料などが含まれる。法令データは API 化に関する検証が行われている [2] 一方で、これらの資料は、各省庁のホームページで会議体ごとに異なるフォーマットで公開され、図 2 に示すような複雑な図表を含むものも少なくない。また、法令改正や政策立案のプロセスを把握するためには、単一の会議体の資料のみでは不十分であり、関連委員会の資料を横断的に参照する必要が生じる場合が多い。

本研究の目的は、政府会議資料を対象とした要約システムを構築することである。今後は、本システムの出力した要約文書を活用し、事務局資料、議事録、パブリックコメントを統合的に参照することで、政策動向の推測へと応用することを検討する。ただし、検証段階で複数の委員会に加えて複数資料を同時に扱うと、資料の収集・整形や対応付けに関する処理が増え、システムが複雑化する。そこで本

1) <https://polisee.com/>

アンケート調査の背景と目的

- 当委員会は、個人情報安全・円滑に越境移転できる国際環境の構築において、事業者側のニーズを把握した上で、ビジネス様態や規模に応じて、複数の選択肢から利用しやすい越境移転のスキームを遊ぶことができる国際環境の構築を目指している。
- この観点より、様々な業種・規模の民間企業から、個人情報の越境移転の状況、越境移転規制を含む海外個人情報保護法への対応における課題等の情報を収集し、今後の当委員会での政策検討の材料とすることを目的にアンケート調査を実施した。

■アンケート調査の背景

- 当委員会では個人情報安全・円滑に越境移転できる国際環境の構築を目指す観点から、事業者側のニーズを把握した上で、ビジネスの様態や規模に応じて、複数の選択肢から利用しやすい越境移転のスキームを遊ぶことができるような環境構築を図ることとしている。
- 「個人情報保護委員会の国際戦略」（令和5年3月29日個人情報保護委員会決定）においても、国際動向の把握と情報発信に取り組みに当たり、企業ニーズを把握することの重要性について言及している。
- 企業活動のグローバル化に伴い、各企業が海外法規制への対応やデータの越境移転規制への対応が求められる機会が増えていると想定されることから、企業ニーズの一端として、企業が抱える課題や当委員会への意見についても把握することが求められている。

■アンケート調査の目的

- 上記の背景を踏まえ、様々な業種・規模の民間企業から、どのような個人情報の越境移転が実施されているか、越境移転規制を含む海外個人情報保護法への対応にあたりどのような課題があるのか等の情報を収集し、今後の当委員会での政策検討の材料とすることを目的にアンケート調査を実施した。

図1 事務局資料の例（個人情報保護委員会）

（参考）各設問分類に関する図解

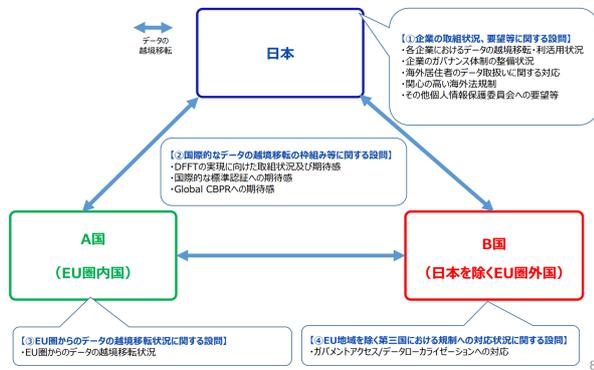


図2 図や表を含む政府会議資料の例（個人情報保護委員会）

稿では、分析を容易にするため、政府会議資料の対象と種類を限定する。

具体的には、会議体については(1)会議が一定期間にわたり継続的に開催され、資料が継続的に蓄積されていること、(2)他会議体の資料を広範に参照せずとも議論の流れを追跡できること、を満たすものを選定する。また資料種別については、まず事務局資料を対象とし、その後、議事録およびパブリックコメントへと対象を拡張する。以上の条件を踏まえ、本研究では政府会議資料の対象を polisee に事務局資料として登録されている個人情報保護委員会に限定した。対象期間は2016年1月16日から2025年6月25日までで2,448件の資料を使用した。

2.2 評価クエリ

提案システムは検証段階のため、事例分析に用いた質問は企業の法務・公共政策・渉外の部署が政策動向を把握するために、開催された会議全体における議論の論点・方向性を整理する利用場面を想定し

て表1のように用意した。なお、情報検索の評価では十分な数の質問が要求される一方で、質問数を増やすと、LLMの出力結果の妥当性を人手で評価することが困難であるため、質問数を限定した。

3 提案システムの概要

3.1 全体構成

本研究で提案するシステムは、入力されたクエリから gpt-5-mini を用いて日付を抽出し、抽出した日付に基づいて対象期間を設定する。次に、検索を行い、関連チャンクを取得し、検索上位20件をコンテキストとして gpt-5-mini に入力し、回答を生成する。なお、表1の q8 は LLM により対象期間が抽出できなかったため全期間で検索を行った。

3.2 ベクトル DB の構成

対象資料は PDF 形式で提供されているため、RAG で扱うために日本語に特化した文書画像解析エンジン YomiToku²⁾ を用いて PDF 文書の OCR およびレイアウト解析を行い、テキスト化した。YomiToku は、レイアウト解析に基づいて段落や表、図などを認識し、高精度な OCR を実現することができるため、本研究では YomiToku を採用した。YomiToku で解析したレイアウト情報のうち、見出しを抽出し、チャンク化の際にメタデータとして利用した。なお、レイアウト解析で抽出された page_header や page_footer は、ノイズとなる可能性が高いため、本研究では利用しなかった。

本研究ではベクトル DB に Qdrant³⁾ を使用し、レイアウト解析で抽出した段落をもとに文書をチャンク化し、表2の設定でメタデータを付与した。

3.3 検索手法

本研究では、政府会議資料のように専門用語・固有名詞・表現ゆれが混在し、さらに OCR やチャンク設計の影響を強く受ける文書集合において、検索手法の選択が検索結果の偏り・重複といった性質にどのような差が生じるかを明らかにするため、密ベクトル検索と、BM25 を用いたハイブリッド検索、SPLADE^[3] を用いたハイブリッド検索の3つの検索手法を比較した。

2) <https://github.com/kotaro-kinoshita/yomitoku>

3) <https://qdrant.tech/>

表1 評価クエリ

qid	クエリ
q1	この半年にどんな法改正の議論がありましたか。
q2	この半年にどんな基準の策定の議論がありましたか。
q3	この半年にどんなトピックの議論がありましたか。
q4	この半年で規制強化がありましたか。
q5	規制強化を主張した人は誰ですか。
q6	会議に一番出席した有識者は誰ですか。
q7	この半年でどんな海外の規制が議論されましたか。
q8	今後法改正はされそうでしょうか。

表2 チャンク化およびメタデータ付与の設定

項目	設定
分割単位	レイアウト解析で抽出した段落
チャンクサイズ	256
オーバーラップ	64
付与メタデータ	会議名, 資料名, 会議開催日, ページ番号, 資料 ID, セクション
総チャンク数	111,732

3.3.1 Dense

クエリとチャンクを埋め込み表現に変換し、内積に基づいて類似度を計算した。埋め込みモデルには多言語・多機能な埋め込みモデルとして報告されている bge-m3[4] を用い、関連度の上位 20 件をコンテキストとして使用した。

3.3.2 BM25 Hybrid

語彙一致に基づく BM25 と、bge-m3 を用いた密ベクトル検索を組み合わせたハイブリッド検索を行った。BM25 では MeCab を用いた形態素解析と Stopwords ISO⁴⁾ を用いたストップワード除去を行った。各 Retriever で関連度の高い上位 100 件ずつ取得し、RRF[5] によりランキングを統合した後、上位 20 件をコンテキストとして使用した。

3.3.3 SPLADE Hybrid

SPLADE による疎ベクトル表現と、bge-m3 を用いた密ベクトル検索を組み合わせたハイブリッド検索を行った。本研究では SPLADE の事前学習済みモデルとして、hotchpotch/japanese-splade-v2⁵⁾ を用いた。BM25 を用いたハイブリッド検索と同様に、各 Retriever で関連度の高い上位 100 件ずつ取得し、RRF によりランキングを統合した後、上位 20 件をコンテキストとして使用した。

4) <https://github.com/stopwords-iso/stopwords-iso>

5) <https://huggingface.co/hotchpotch/japanese-splade-v2>

4 評価方法

政府会議資料は会議回ごとに追加・更新され、論点や根拠が複数回・複数資料に分散し得る。このため、表 1 の q1, q2 のように複数回にまたがる論点整理を要する問いでは、検索結果が少数資料に偏ると参照範囲が狭まり、論点整理に必要な根拠を取りこぼすリスクがある。そこで本研究では、Retriever が返す検索上位について、(1) 参照先の偏り、(2) 短いチャンク混入、(3) 見出しチャンクの混入の 3 点を指標化し、検索手法間で比較する。また、要約内容の評価は、ドメイン特有の論点の妥当性や網羅性の判定が必要であり RAGAs[6] や QAGS[7] などの自動指標では評価が難しいため、自動評価は行わず、専門家 1 名による手動の評価を行った。

本システムは日付フィルタを用いるため、評価では 2024 年 1 月 1 日を基準として期間を固定した。LLM への入力には上位 20 件に制限されるため主な分析は $k = 20$ で行い、将来的なリランク導入時の改善余地を把握する目的で $k = 100$ でも同様に算出する。

検索結果全体の傾向を把握するために、各クエリ $q \in Q$ について検索上位 k 件を対象に指標値 $M_q@k$ を算出し、検索手法間の比較にはクエリ平均

$$\overline{M@k} = \frac{1}{|Q|} \sum_{q \in Q} M_q@k \quad (1)$$

を用いる。以下では $M \in \{\text{DC, LI, HM}\}$ と定義する。

(1) 参照先の偏り (DC@k) 検索上位 k 件に含まれ

表 3 検索結果指標のクエリ平均, DC@k:資料集中度, LI@k:短いチャンクの混入率(閾値 20 文字) HM@k:見出しチャンクの混入率.

手法	k	DC@k	LI@k	HM@k
Dense	20	0.519	0.338	0.441
Dense	100	0.538	0.348	0.439
BM25 Hybrid	20	0.469	0.163	0.294
BM25 Hybrid	100	0.511	0.245	0.338
SPLADE Hybrid	20	0.444	0.244	0.352
SPLADE Hybrid	100	0.480	0.233	0.299

る資料 ID のうち, 同じの資料 ID の出現数を $c_{id}(q)$ とすると,

$$DC_q@k = \frac{c_{id}(q)}{k} \quad (2)$$

と定義する. 値が大きいくほど, 検索結果が少数の資料に集中していることを示す.

(2) 短いチャンク混入率 (LI@k) 検索上位 k 件の各チャンクについて空白を除いた文字数を数え, 20 文字未満のチャンク数を $c_{short}(q)$ とすると,

$$LI_q@k = \frac{c_{short}(q)}{k} \quad (3)$$

と定義する. 値が大きいくほど, 情報量の少ない短いチャンクが上位に混入していることを示す. なお判定の閾値は, 検索上位チャンクの文字数を小規模に確認し, 20 文字を採用した.

(3) 見出しチャンク混入率 (HM@k) 検索上位 k 件においてチャンクが見出しと同一である割合を見出しチャンク混入率 (HM@k) として算出する. チャンク内容が見出しと一致した件数を $c_{match}(q)$ とすると,

$$HM_q@k = \frac{c_{match}(q)}{k} \quad (4)$$

と定義する. 値が大きいくほど, 見出しと同一のチャンクが検索上位を占める割合が高いことを示す.

5 評価結果

表 3 に各手法の検索結果指標の平均を示す. 表 3 より, DC@k は手法間で大きな差が見られず, 本設定では参照先の資料への集中は検索手法差としては顕著ではない. これは, レイアウト解析により段落単位で分割していても, OCR テキストでは見出し等の短いチャンクがクエリ語と一致しやすく, 上位検索を占め得ることを示唆する. また, 専門家確認でも要約の偏りが 2023 年 12 月に見られることが指摘された. さらに, HM@k および LI@k が一定値を示すことから, 検索上位が短文・見出しになること

で, 有効な根拠が減少し網羅性が損なわれた可能性がある.

6 検索手法の傾向と課題

本研究は RAG 要約システムを構築し, 検索結果の偏り・ノイズを指標で分析したが, 理解支援としての有用性を示すには評価設計が不十分である. 近年は, LLM を用いて要約を評価させる研究がある [8]. 今後は, 根拠の整合性・重要論点の網羅・出典提示の明確さ等の評価項目と判定基準を明確化し, 専門家評価に加えて LLM での評価を行う.

また, q8 では回答作成日より後の資料が引用される事例が見られた. 未来の資料が存在しないことを前提に, 対象期間内の審議内容から改正の言及や論点の推移を根拠として推測する評価・推論設定へ修正する必要がある. 近年, 文書間の関係をグラフとして統合し, 網羅的な検索を可能にする GraphRAG に関する研究があり [9], 会議回をまたぐ論点整理や根拠統合に有効な可能性がある. 最終的には, 本要約システムで対象期間の網羅的な要約を出力し, その要約に基づいて q8 の推測を行う枠組みを目指す.

そのためには, 見出しの上位化や参照先の偏りを低減し, 有効情報量を確保することが重要である. 具体的には, HyDE[10] や RAG-Fusion[11] を用いたクエリ展開の検討やリランク [12], ColBERT[13] を使った際の参照資料の偏りの傾向を検証する. さらに, OCR の限界に対して, ページ画像を入力とするマルチモーダル RAG システム [14] の検討や PDF 文書から画像を抽出し, 要約させることで根拠抽出を補完する必要がある.

7 まとめ

本研究は政府会議資料を OCR し, 出典付き要約を生成する RAG 要約システムを構築したうえで, 検索結果の偏り・ノイズを分析した. その結果, 検索上位には短文および見出しが一定割合混入し得ることが確認され, 生成に利用可能な根拠の有効情報量が低下し, 網羅性や議題の偏りを助長する可能性が示唆された. さらに OCR 文書では図表やレイアウト依存箇所の扱いに限界があり, 根拠抽出の補完が課題である. 今後は, 短文・見出しチャンクの上位化を抑制しつつ, 網羅的な要約を出力し, その要約と会議の進捗に基づいて今後の方向性を推測する枠組みを構築し, 法令改正・政策立案プロセスの把握と政策動向の理解支援を目指す.

参考文献

- [1] 名知浩一郎, 西村昭治. 国会会議録の言語モデルを利用した要約の評価. 情報知識学会誌, Vol. 35, No. 3, pp. 341–349, 2025.
- [2] 山内匠. 法令データの現状と法令分野へのデジタル技術適用の展望. 言語処理学会第 30 回年次大会発表論文集, pp. 1051–1055, 2024.
- [3] Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. **SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking**, p. 2288–2292. Association for Computing Machinery, New York, NY, USA, 2021.
- [4] Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. Bge m3-embedding: Multilingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation, 2024.
- [5] Gordon V. Cormack, Charles L A Clarke, and Stefan Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In **Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval**, SIGIR '09, p. 758–759, New York, NY, USA, 2009. Association for Computing Machinery.
- [6] Shahul Es, Jithin James, Luis Espinosa Anke, and Steven Schockaert. RAGAs: Automated evaluation of retrieval augmented generation. In Nikolaos Aletras and Orphee De Clercq, editors, **Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations**, pp. 150–158, St. Julians, Malta, March 2024. Association for Computational Linguistics.
- [7] Alex Wang, Kyunghyun Cho, and Mike Lewis. Asking and answering questions to evaluate the factual consistency of summaries. **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, 2020.
- [8] 山縣将貴, 加藤拓, 藤本拓, 吉村健. 質疑応答を用いた文書要約評価指標への llm 適用の検討と分析. 人工知能学会全国大会論文集 第 39 回 (2025), pp. 1Win437–1Win437. 一般社団法人人工知能学会, 2025.
- [9] 穴口史将, チャクラポルティシュデシナ, 森田武史, 江上周作, 鶴飼孝典, 福田賢一郎. 文書のチャンクに基づく graphrag と知識グラフのコミュニティ分類に基づく graphrag の比較分析. 人工知能学会全国大会論文集, Vol. JSAI2025, pp. 3P1OS46a02–3P1OS46a02, 2025.
- [10] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1762–1777, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [11] Zackary Rackauckas. Rag-fusion: A new take on retrieval augmented generation. **International Journal on Natural Language Computing**, Vol. 13, No. 1, p. 37–47, February 2024.
- [12] Rodrigo Nogueira and Kyunghyun Cho. Passage re-ranking with bert. **ArXiv**, Vol. abs/1901.04085, , 2019.
- [13] Omar Khattab and Matei Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In **Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval**, SIGIR '20, p. 39–48, New York, NY, USA, 2020. Association for Computing Machinery.
- [14] 福井琢, 宗像聡. 画像を含む文書から検索用洞察を生成することによるマルチモーダル rag システムの検索精度の改善. 人工知能学会全国大会論文集, Vol. JSAI2025, pp. 4Q1GS1001–4Q1GS1001, 2025.