

法律分野の統合引用グラフを活用した質問応答の実現

丸山拓海 稲垣有二

弁護士ドットコム株式会社 リーガルブレイン開発室

{ta.maruyama, inagaki}@bengo4.com

概要

法律実務においては、法令や判例に加え、専門書籍やガイドラインなど、多様な情報源の関係性を総合的に把握することが不可欠である。しかし、これまでの研究は主に法令や判例間関係性に焦点を当てており、書籍やガイドラインとの連携については十分に検討されていなかった。本研究では、これらすべての情報源を包括的に結びつける大規模な引用グラフを構築し、その実務的有用性を検証した。その結果、質問応答タスクにおいて、本引用グラフを活用することで、適切な法令・判例の選出が可能となっただけでなく、法的な正確性を保ちながら、人手で作成した回答よりも好ましい回答を生成できることが明らかとなった。

1 はじめに

法律実務において適切な判断を行うためには、関連法令や類似判例、さらには専門書籍やガイドラインによる解説など、多様な情報源を総合的に検討することが求められる。

海外法令においては、法令や判例における関係性を構造化したグラフが構築され、分析や応用事例が数多く報告されている [1, 2, 3, 4]。日本においても、法令データの整備 [5] や民事判決のオープンデータ化 [6, 7] が進展しており、法律実務における効率的な情報活用の基盤となることが期待されている。

しかし、既存研究では法令・判例間関係性に主眼が置かれており、実務で重要な役割を果たす書籍やガイドラインまでを包括的に含めた引用グラフについては、十分に組み込まれていなかった。このような背景から、本研究では法令・判例・書籍・ガイドラインを包括的に結びつける引用グラフを構築し、実務における有用性を検証した。

本研究の主な貢献は以下の3点である：

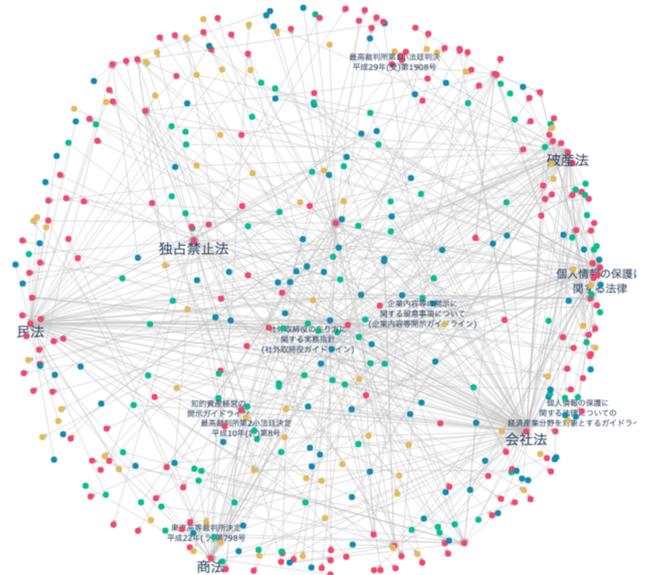


図1 コンテンツ同士の引用関係。赤・黄・緑・青色の頂点がそれぞれ、法令・判例・ガイドライン・書籍を意味する。直接引用関係にある頂点同士が線で結ばれている。

1. 法令や判例に加え、書籍やガイドラインといった多様な情報源を含む引用グラフを初めて構築した (2章)。
2. 法律実務を想定した、回答根拠を提示する質問応答タスクを定義した (3章)。
3. 大規模言語モデル (Large Language Model: LLM) と構築した引用グラフを組み合わせることで、質問に対して適切な根拠を提示することが可能となっただけでなく (4章)、専門家にとってより好ましい回答を生成できることを明らかにした (5章)。

2 引用グラフ

引用グラフでは、以下を頂点として扱う：

- **コンテンツ**: 書籍・ガイドライン・判例・法令。
- **チャンク**: コンテンツに含まれるテキストを一定の単位で分割したもの。

表 1 引用グラフの頂点数と引用関係を表す辺の数. 判例から判例への辺や法令から判例への辺は存在しないため, 未記載とした.

コンテンツ	頂点数	辺の数	
		引用法令	引用判例
書籍・ガイドライン	837,882	975,999	97,383
判例	460,450	400,259	-
法令	4,824,711	2,563,047	-
合計	6,123,043	3,939,305	97,383

コンテンツとチャンクは, 親子関係を表す辺で接続した. チャンク同士及びチャンクからコンテンツへの引用関係を表す辺は, チャンク頂点を持つテキストに出現する法令名・判例名をもとに構築した(2.2節). コンテンツ同士の引用関係(図 1)は, チャンクに紐づく引用法令・判例を, チャンクの親となっているコンテンツに集約することで構築した.

2.1 引用グラフの頂点

書籍・ガイドライン 弁護士ドットコム株式会社が提供する, BUSINESS LAWYERS LIBRARY¹⁾・弁護士ドットコム LIBRARY²⁾で閲覧可能な法律書籍と, 各省庁の web ページで公表しているガイドラインを合わせた, 約 2,000 文書を対象とした. PDF や web ページからテキスト抽出したのち, 本文テキストを固定長で区切り, チャンクとした.

判例 株式会社 LIC が提供する, 判例秘書³⁾の約 20 万件を対象とした. 判例データには, 判決本文に加え, 判決要旨や事件番号・掲載雑誌・判決年月日, 参照法令などが記述されている. 判例チャンクとして, 判決要旨を採用した. また, 判例データに含まれる参照法令をもとに, 判例から法令への引用関係を表す辺を構築した.

法令 e-Gov 法令検索⁴⁾から, 2024 年 9 月時点で取得可能な法令 XML ファイル約 11,000 件対象とした. 法令は条・項・号それぞれをチャンク頂点として扱った. また, 法令は同一名称であっても改正日や施行日が異なる複数のバージョンを持つ. そのため, 引用グラフ内ではこれらを区別して扱った.

2.2 引用グラフの辺

書籍チャンクやガイドラインチャンクの本文テキストに出現する, 法令名・判例名を抽出することで引用関係を表す辺を構築した. 表 1 に, コンテンツ

1) <https://www.businesslawyers.jp/lib>
 2) <https://library.bengo4.com/about>
 3) <https://www.hanreihisho.com/index.html>
 4) <https://laws.e-gov.go.jp/>

ごとの頂点数と引用関係の辺の数を示す. 文字列抽出は, 先行研究にならい, ルールベースの手法を採用した[8].

引用法令を一意に特定する際には, 引用元となる頂点を持つ日付情報(発行日・判決日・改正日等)に基づき, その時点における最新の法令へと対応付けた. 条以下の項や号の単位への引用も, 適宜後処理を加え, 条の単位への引用として扱った.

引用判例は, 抽出した文字列に含まれる, 裁判所・判決結果・判決年月日・出典の情報に基づいて, 判例を一意に特定できるもののみ辺を構築した.

3 引用グラフを用いた質問応答

構築した引用グラフの実務的有用性を評価するため, 法律実務を想定した質問応答タスクを定義した. 実務では法的判断の透明性が求められることから, 本タスクでは, 法的質問に対する回答に加え, その根拠となる法令・判例を必ず提示することを要件とする. 質問応答の過程では, 入力質問に対して関連文書検索により取得した書籍やガイドラインのチャンクを「関連文書」として参照し, 法令条文や判例要旨を「根拠情報」として提示する.

これらの要件を満たすため, 質問応答パイプライン(3.1節)を構築し, 以下の 2 点を検証した:

- (Q1) 回答の根拠として必要な法令・判例を適切に選出できるか? (4章)
- (Q2) 選出した関連文書・根拠情報を用いることで, LLM は法的観点を適切に踏まえた回答を生成できるか? (5章)

3.1 質問応答のパイプライン

検索拡張生成(Retrieval Augmented Generation: RAG)[9]をベースに, 図 2 のパイプラインを構築した. 処理の具体的な流れを以下に示す.

- (i) **クエリ埋め込み**: 質問文をベクトルへ変換する. 具体的には次の 2 種類の手法を用いた: (1) 質問文を BM25[10]・multilingual-e5-large[11]を利用して質問文をベクトルへ変換する. (2) LLM により仮回答を 3 件生成したのち(1)と同様の手法で, 仮回答のベクトルを 3 件取得する. 以下, この手法を HyDE とよぶ[12].
- (ii) **関連文書の検索**: クエリベクトルを用いてベクトル DB から関連文書を M 件取得する. なお, HyDE により複数の検索結果が得られる場

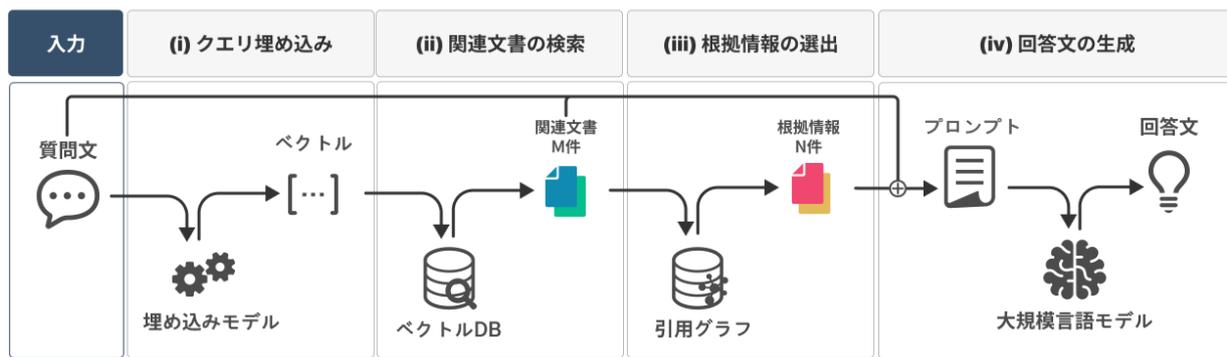


図2 質問応答パイプライン. (i) 質問文を元にベクトルへの検索クエリを生成する. (ii) 検索クエリを元に関連文書を M 件取得する. (iii) 引用グラフから関連文書に紐づく根拠情報を N 件取得する. (iv) 質問文と関連文書 M 件, 根拠情報 N 件を LLM に与え, 回答を生成する.

合は Reciprocal Rank Fusion (RRF)[13] によりスコアリングし, 上位 M 件のみを利用する.

(iii) **根拠情報の選出:** 2章で構築した引用グラフを用いて, 関連文書 M 件が引用している根拠情報を N 件取得する. 関連文書検索の順位を利用し, チャンクが引用している根拠情報を RRF によって並び替え, 上位 N 件を根拠情報として利用する.

(iv) **回答文の生成:** 関連文書 M 件と, その文書が引用している根拠情報 N 件の, 計 $M + N$ 件の参考文献と質問文を LLM に与え, 回答文を生成する. LLM には, Claude3.5 Sonnet[14] を利用する.

3.2 データセット

評価データセットとして, BUSINESS LAWYERS の実務 Q&A⁵⁾ の記事を用いた. 記事には, 企業法務のさまざまな場面を想定した質問と, 専門家の回答及びその解説が含まれている.

記事中の質問文を質問応答パイプラインへの入力とし, 記事中の回答及び解説文を正答例とした. また, 記事の解説本文に現れる引用法令・引用判例を, 回答を行う上で必要となる正解の根拠情報とした. 実験では, 根拠情報を1つ以上含む539記事を利用した.⁶⁾

4 根拠情報選出の検証

関連文書数・根拠情報数をそれぞれ $M = 30$, $N = 20$ としたときの根拠情報選出の性能を評価す

5) <https://www.businesslawyers.jp/practices>

6) 1つの記事に対して, 平均4.6件の根拠情報が付与されていた.

表2 根拠情報選出の性能評価 ($M = 30, N = 20$). 表中の $R@5 \cdot R@10, R@20$ は, それぞれ Recall@5 · Recall@10 · Recall@20 を意味する.

関連文書の検索手法	根拠情報の選出性能		
	R@5	R@10	R@20
Zero-Shot	18.84	24.32	26.87
BM25	17.06	23.99	28.98
multilingual-e5-large	20.71	28.33	35.87
HyDE (BM25)	24.11	32.79	38.18
HyDE (multilingual-e5-large)	23.14	31.29	36.96

る. ここで, LLM に対して質問と関連する法令 (法令番号+条番号) · 判例 (事件番号) を関連度順に20件ずつ列挙するように指示する, Zero-Shot の手法をベースラインとする.

表2より, BM25を除く全ての手法で, Zero-Shot に比べて高い Recall を達成している. これは, 引用グラフを用いたアプローチが, LLM への直接的な問い合わせよりも優れていることを示している.

また, HyDE を組み合わせることで, 関連法令の選出性能が大幅に改善されている. 特に, 埋め込み手法として BM25 を用いることで顕著に性能が向上している. これは, HyDE によって, 複数の法的キーワードを含む仮回答を生成でき, キーワードベースの類似度 (BM25) で, より多様な関連文書を取得できたことが要因だと考えられる. より詳細な検証は付録 A を参照されたい.

5 回答文生成の検証

質問文と JSON 形式で記述された参考文献 (関連文書・根拠情報) を LLM に入力し, 回答文を生成する. 最も基本的な手法として, ID · タイトル · 本文を持つ参考文献 JSON をリスト形式で入力する方法

を **Text-only** とする。入力する順序は質問文、書籍、ガイドライン、判決要旨、法令条文の順とする。これに加えて、引用グラフから得られた文書間の関係性を含める手法として、以下の3種類のプロンプトを検証した：

1. **Focus-literature**: 関連文書に引用法令・判例を集約する。具体的には、各関連文書 JSON に `cites` フィールドを設け、その文書が引用している法令条文・判例要旨のテキスト情報を記載する。
2. **Focus-law**: 法令条文に関連文書・判決要旨を集約する。具体的には、各法令条文 JSON に `cited` フィールドを設け、その条文を引用している関連文書および判例要旨のテキスト情報を記載する。
3. **RDF Triples**: **Text-only** と同様にテキスト情報を与える。さらに、引用関係を RDF 形式で記述したテキストを追加。

各プロンプトの詳細は付録 B を参照されたい。クエリ埋め込みには `multilingual-e5-large` を使用し、 $M = 30$, $N = 20$ とした。評価指標として、`ROUGE-L`[15], `BERTScore`[16] に加え、法的論点と結論に基づく関連度を `G-EVAL-4`[17] で評価した。また、参考文献を用いず質問文のみを LLM に入力する **Zero-Shot** をベースラインとした。

表 3 より、全てのプロンプト手法において、**Zero-Shot** を上回る性能を達成し、関連文書・根拠情報を追加することで、より適切な回答が生成可能となることが実証された。特に、`G-EVAL-4` による法的な観点での関連度スコアではその差が顕著に現れた。

6 回答文の質的評価

質的評価として、ランダムに選出した質問 60 件に対する回答文を、専門家により評価した。評価対象は、**Zero-Shot** を除く 4 種類のプロンプトで生成した回答文と、記事に含まれる回答文 (Reference) の計 5 種類とし、以下の 2 つの観点で評価を行った：

- **回答の適切さ**：質問に対して回答の適切性を二値 (適切：1, 不適切：0) を付与
- **回答の好ましさ**：5 種類の回答を好ましさ順に並び替え、最下位から順に 1~5 のスコアを付与

表 4 より、「回答の適切さ」において、Reference (98%) と提案手法による生成文 (97%) の間にほと

表 3 回答文生成の実験結果.

プロンプト	ROUGE-L	BERTScore	G-EVAL-4
Zero-Shot	21.85	83.36	77.90
Text-only	24.30	83.87	95.95
Focus-literature	24.40	83.91	95.61
Focus-law	24.61	84.07	95.89
RDF Triples	24.28	83.78	96.52

表 4 回答文生成の質的評価.

プロンプト	回答の適切さ	回答の好ましさ
Reference	98	1.4
Text-only	97	3.5
Focus-literature	97	3.3
Focus-law	97	2.8
RDF Triples	97	3.5

んど差は見られなかった。一方「回答の好ましさ」では、むしろ提案手法の方が高いスコアを獲得している。これらの結果から、提案手法では適切さを損なうことなく、より好ましい表現で回答を生成できることが明らかとなった。また、プロンプト手法間の「回答の好ましさ」に注目すると、**Text-only** と比べて、引用関係を付与する手法が劣っていることから、引用関係を与えるプロンプト手法には検討の余地があることが明らかとなった。

7 今後の展望

質問応答タスクでは、グラフ構造を活用した高度な手法 [18, 19] も数多く提案されている。今後は、これらの最新の研究動向を踏まえつつ、本研究で構築した引用グラフの特性を最大限に活かせる手法の開発に取り組んでいきたいと考えている。また、法的判断予測 [20, 21], 文書分類 [22, 23], 類似判例検索 [24, 25], 自動要約 [26, 27] などをはじめとする法律分野の応用タスクにおいても、本研究で構築した引用グラフの応用可能性について今後検討を進める予定である。

8 まとめ

本研究では、法令、判例、専門書籍、ガイドラインなどの多様な法律情報を統合した大規模な引用グラフを構築した。実務を想定した質問応答タスクにおいて、この引用グラフを活用することで、大規模言語モデルの単体と比較して、より適切な法令・判例の選出が可能となることを実証した。さらに、選出された関連文献を活用した回答生成により、人手で作成した回答よりも高い評価を得る回答の生成が可能であることを示した。

参考文献

- [1] Emile de Maat, Radboud Winkels, and Tom van Engers. Automated detection of reference structures in law. In **Proceedings of the 2006 conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference**, pp. 41–50, NLD, June 2006. IOS Press.
- [2] Tobias Milz, Michael Granitzer, and Jelena Mitrović. Analysis of a german legal citation network. In **Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management**. SCITEPRESS - Science and Technology Publications, 2021.
- [3] Ali Sadeghian, Lakshman Sundaram, Daisy Zhe Wang, William F Hamilton, Karl Branting, and Craig Pfeifer. Automatic semantic edge labeling over legal citation graphs. **Artif. Intell. Law**, Vol. 26, No. 2, pp. 127–144, June 2018.
- [4] Yanran Tang, Ruihong Qiu, Hongzhi Yin, Xue Li, and Zi Huang. CaseLink: Inductive graph learning for legal case retrieval. In **The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval**, March 2024.
- [5] 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, March 2024.
- [6] 久本空海, 城戸祐亮, 津金澤佳亨, 八木田樹. 民事判決のオープンデータ化へ向けた機械処理による判例仮名化の検証. 言語処理学会 第 28 回年次大会 発表論文集, pp. 1406–1410, March 2022.
- [7] 前田郁勝, 外山勝彦, 小川泰弘. 民事第一審判決書の XML データ化. 言語処理学会 第 30 回年次大会 発表論文集, pp. 1324–1328, March 2024.
- [8] 丸山拓海, 菅原祐太, 西野裕貴, 稲垣有二. 法律分野における大規模引用グラフの構築. 第 17 回データ工学と情報マネジメントに関するフォーラム, March 2025.
- [9] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. **arXiv [cs.CL]**, December 2023.
- [10] Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. **Found. Trends® Inf. Retr.**, Vol. 3, No. 4, pp. 333–389, 2009.
- [11] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual E5 text embeddings: A technical report. **arXiv [cs.CL]**, February 2024.
- [12] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In Jordan Boyd-Graber Okazaki, Anna Rogers, editor, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, p. 1762–1777. Association for Computational Linguistics, July 2023.
- [13] 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**, New York, NY, USA, July 2009. ACM.
- [14] Anthropic. The claude 3 model family: Opus, sonnet, haiku. https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude.3.pdf, March 2023.
- [15] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. **Annu Meet Assoc Comput Linguistics**, pp. 74–81, July 2004.
- [16] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with BERT. **arXiv [cs.CL]**, April 2019.
- [17] Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. G-eval: NLG evaluation using gpt-4 with better human alignment. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 2511–2522, Stroudsburg, PA, USA, 2023. Association for Computational Linguistics.
- [18] Ruosong Ye, Caiqi Zhang, Runhui Wang, Shuyuan Xu, and Yongfeng Zhang. Language is all a graph needs. In Yvette Graham and Matthew Purver, editors, **Findings of the Association for Computational Linguistics: EAACL 2024**, p. 1955–1973. Association for Computational Linguistics, March 2024.
- [19] Bowen Jin, Chulin Xie, Jiawei Zhang, Kashob Kumar Roy, Yu Zhang, Zheng Li, Ruirui Li, Xianfeng Tang, Suhang Wang, Yu Meng, and Jiawei Han. Graph chain-of-thought: Augmenting large language models by reasoning on graphs. In Martins Sriku-mar Lun-Wei Ku, editor, **In Findings of the Association for Computational Linguistics: ACL 2024**, p. 163–184. Association for Computational Linguistics, August 2024.
- [20] Junyun Cui, Xiaoyu Shen, and Shaochun Wen. A survey on legal judgment prediction: Datasets, metrics, models and challenges. **IEEE Access**, Vol. 11, pp. 102050–102071, 2023.
- [21] Hiroaki Yamada, Takenobu Tokunaga, Ryutarō Ohara, Akira Tokutsu, Keisuke Takeshita, and Mihoko Sumida. Japanese tort-case dataset for rationale-supported legal judgment prediction. **Artif. Intell. Law**, May 2024.
- [22] Tuggener Don, Pius von Däniken, Peetz Thomas, and Cieliebak Mark. LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Twelfth Language Resources and Evaluation Conference**, p. 1235–1241. European Language Resources Association, May 2020.
- [23] Limsopatham Nut. Effectively leveraging BERT for legal document classification. In Aletas Nikolaos, Androusoopoulos Ion, Barrett Leslie, Goanta Catalina, and Daniel Preotiuc-Pietro, editors, **Proceedings of the Natural Legal Language Processing Workshop 2021**, p. 210–216. Association for Computational Linguistics, 2021.
- [24] Yi Feng, Chuanyi Li, and Vincent Ng. Legal case retrieval: A survey of the state of the art. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 6472–6485, Stroudsburg, PA, USA, August 2024. Association for Computational Linguistics.
- [25] Jaspreet Singh Dhani, Ruchika Bhatt, Balaji Ganesan, Parikshet Sirohi, and Vasudha Bhatnagar. Similar cases recommendation using legal knowledge graphs. SAIL’ 23: 3rd Symposium on Artificial Intelligence and Law, February 2023.
- [26] Abhay Shukla, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal, and Saptarshi Ghosh. Legal case document summarization: Extractive and abstractive methods and their evaluation. In Yulan He, Heng Ji, Sujian Li, Yang Liu, and Chua-Hui Chang, editors, **Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, p. 1048–1064. Association for Computational Linguistics, 2022.
- [27] 菅原祐太, 宮崎桂輔, 山田寛章, 徳永健伸. 日本語法律 bert を用いた判決書からの重要箇所抽出. 言語処理学会 第 28 回年次大会 発表論文集, pp. 838–841.

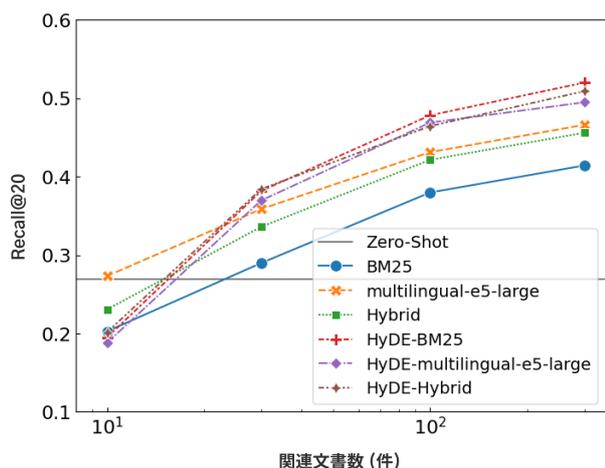


図3 根拠情報選出における Recall@20.

A 関連文書数と根拠情報選出

図2(ii)における関連文書の取得数 $M = 10, 30, 100, 300$ を変化させた際の、根拠情報選出の性能を評価した。評価には3.2節と同一のデータセットを使用した。関連文書の検索手法として、3.1節で述べた手法に加え、BM25とmultilingual-e5-largeを組み合わせたHybridを含めた計7種類の手法を比較した。なお、Hybridでは、BM25とmultilingual-e5-largeそれぞれで関連文書を100件ずつ取得し、各手法の正規化スコアの平均値に基づいて文書をランキングした後、上位 M 件の関連文書を採用する。

図3より、全ての手法において関連文書数の増加に伴い Recall@20 が向上することがわかる。特に、関連文書数を10件から30件に増加させた際に顕著な性能向上が確認された。また、HyDEを活用した手法群は、わずか30件程度の少ない文書数でも高い Recall を実現しており、効率的な検索が可能であることを示唆している。一方で、BM25やmultilingual-e5-largeを単体で利用する場合は、同等の性能を得るためにより多くの関連文書数を必要とすることが明らかとなった。

B 回答文生成のプロンプト

プロンプトの例を図4に示す。プロンプトは大きく分けて次の内容で構成されている：

- 回答生成の指示:** <instruction> タグで囲まれた範囲。「質問に対して短く簡潔に回答すること」「文献間の関係性を考慮すること」を指示する。
- 質問文:** <question> タグで囲まれたテキスト。こ

```
<instruction>
1. 与えられた参考文献に基づいて、質問に短く簡潔に回答してください。
2. 参考文献のコンテンツの間には引用関係があります。
   各文献間の関係性を考慮して回答してください。
</instruction>
<question>
{{質問文}}
</question>
<references>
<books>
[
  {
    index: 'BOOK_1',
    header: '{{書籍タイトル1}}',
    content: '{{書籍本文1}}'
  },
  {
    index: 'BOOK_2',
    header: '{{書籍タイトル2}}',
    content: '{{書籍本文2}}'
  },
  ...
]
</books>
<guidelines>
[{{<books>と同じ形式のガイドラインチャンクリスト}}]
</guidelines>
<precedents>
[{{<books>と同じ形式の判決要旨リスト}}]
</precedents>
<laws>
[{{<books>と同じ形式の法令条文リスト}}]
</laws>
</references>
</instruction>
```

図4 Text-only のプロンプト例。{{}} はプレースホルダーを意味しており、質問応答パイプライン(図2)の実行時に具体的な値が挿入される。

```
<references>
[
  <laws>
  [
    {
      index: 'LAW_1',
      header: '{{法令の条タイトル1}}',
      content: '{{法令条文1}}',
      cited: [
        {
          index: 'BOOK_1',
          header: '{{書籍タイトル1}}',
          content: '{{書籍本文1}}'
        },
        [{{ガイドラインチャンクのリスト}}]
        [{{判決要旨のリスト}}]
      ]
    },
    ...
  ]
  </laws>
]
</references>
```

図5 Focus-law のプロンプト例。references タグ内部以外は図4のプロンプトと同一であるため省略した。

こに入力された質問文を直接挿入する。

- 参考文献:** <references> タグで囲まれたテキスト。書籍チャンク・ガイドラインチャンク・判決要旨・法令条文をそれぞれ、<books>・<guidelines>・<precedents>・<laws> のタグで囲む。参考文献は、ID (例: BOOK_1, LAW_3, ...)・タイトル・本文テキストを属性とした JSON 形式として与える。

Focus-literature のプロンプトの場合は、属性として引用文献 (cites) という項目を追加し、その属性の値として引用法令や引用判例を追加する。Focus-law のプロンプト(図5)の場合は、属性として被引用文献 (cited) という項目を追加し、その属性値として、書籍チャンクやガイドラインチャンク・判決要旨を追加する。