

文書構造グラフによる規定文書QAの根拠補完型RAG

連絡先：hirayama.h77@gmail.com

平山 大世¹, 松藤 彰宏², 小川祐輝², 坂地泰紀¹, 野田五十樹¹

1Yin-A-50

北海道大学¹, パナソニックエレクトリックワークス株式会社²



1. 生の規定文書PDFを見てみよう

第2章 試験及び検定業務

第1節 型式試験

(型式試験の申請)

第2条 型式試験を申請しようとする者（以下「型式試験申請者」という。）は、消防法施行規則（昭和36年自治省令第6号。以下「規則」という。）別記様式第2号に定める型式試験申請書正副各1通に、規則第35条第3項に定める見本（型式試験申請者の名称又はその略称が記載されているものに限る。）及び同条第4項各号に定める書類（以下「見本等」という。）並びに次条に規定する振込用紙の振込票（2以上の申請に係る手数料を一度に振り込む場合に貼付すること。以下同じ。条、第5条及び第26条において）を添付するものとする。

2 前項の見本等の提出は、原簿が持参することが困難であり、見本等に係る説明の必要がないと認める場合は、郵送等により提出することができる。

3 書類は、次表に掲げるところによるほか、次の各号によるものとする。（う）（き）

(1) 書類は、正副ごとに産業標準化法（昭和24年法律第185号）第20条第1項に定める日本産業規格（以下「JIS」という。）P0138のA4の大きさのファイルに一括して編冊すること。（い）（き）

(2) 消防法（昭和23年法律第186号。以下「法」という。）第21条の4第2項の型式承認を受けた型式のうち、法第21条の5又は法第21条の6の規定により当該型式承認の効力が失われていない型式（以下「既承認型式」という。）を有する者に、既承認型式による製品と同一の製品の製造を委託することを前提として型式試験の申請を行う場合の取得し、試験申請書の書類については、検定対象機械器具等の技術資料「社内検査体制等概要調書」を添付すること。（あ）（い）（き）

見出し階層

章・節・項が根拠の文脈を決める

参照表現

「前項」「次の各号」などで離れた条項も根拠になる

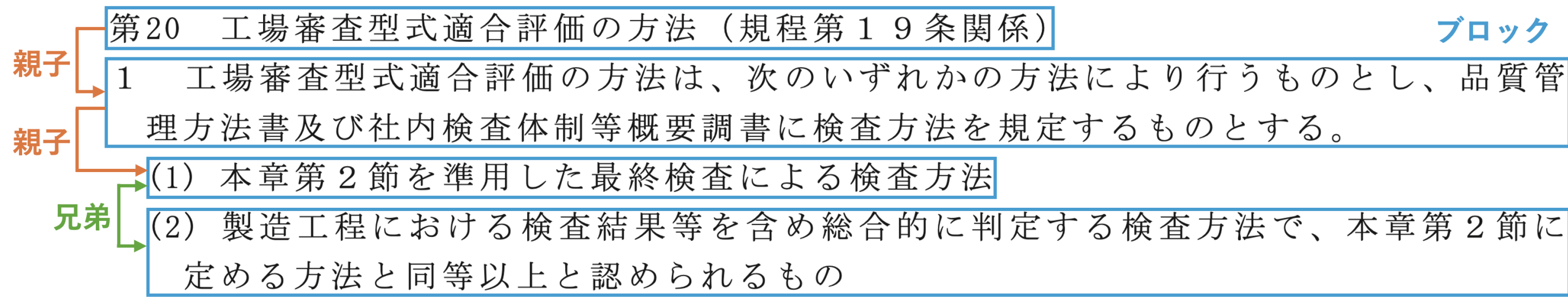
課題：根拠を漏れなく取得したい

列挙

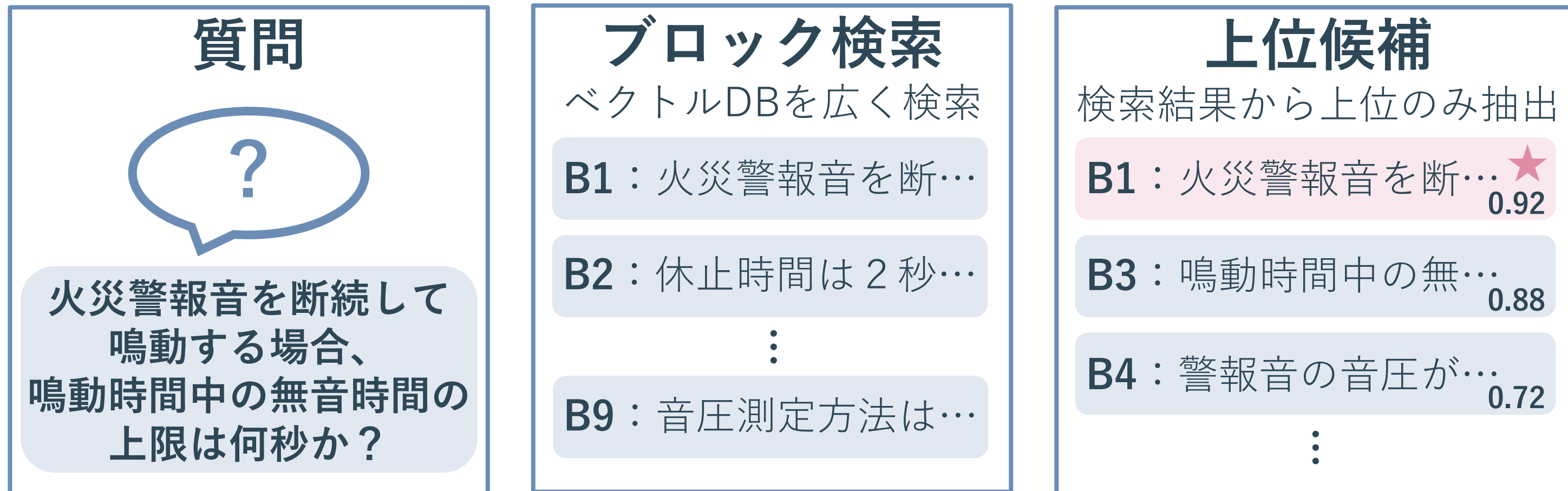
(1)~(5) や ア~オ の列挙はすべて揃えるべき根拠

2. 提案：文書構造グラフ×根拠補完

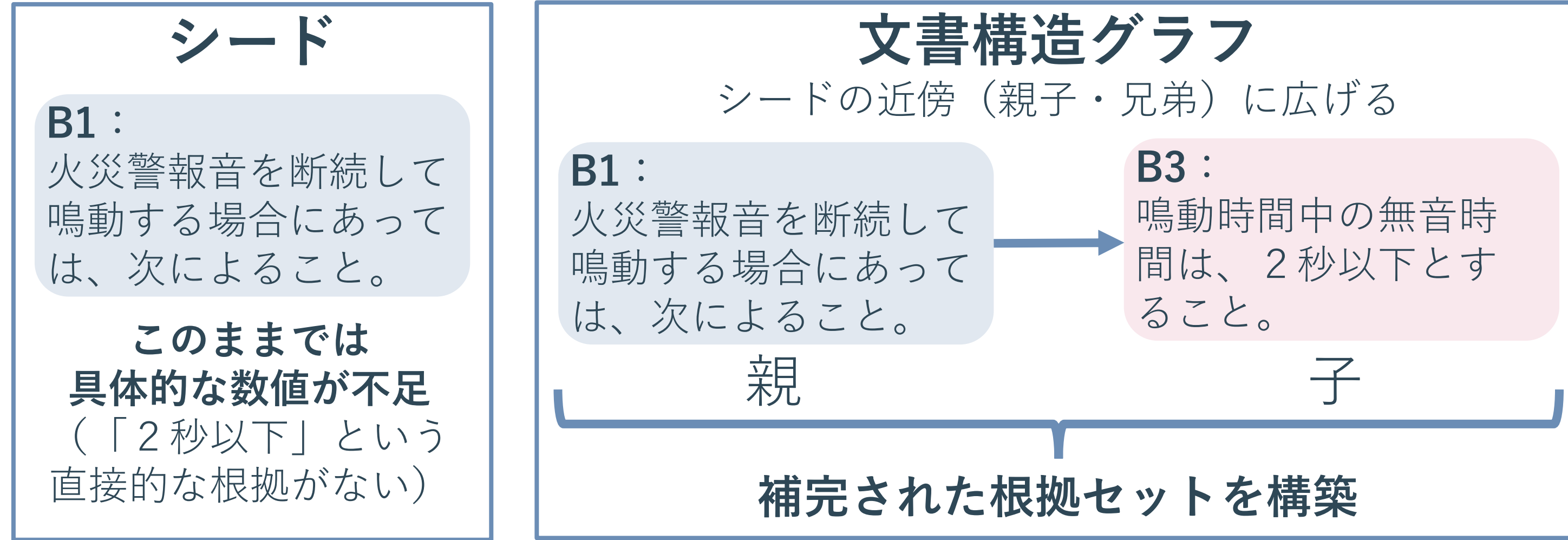
手順1：PDFを文書構造グラフに変換 NLP2026で発表（上のQR）



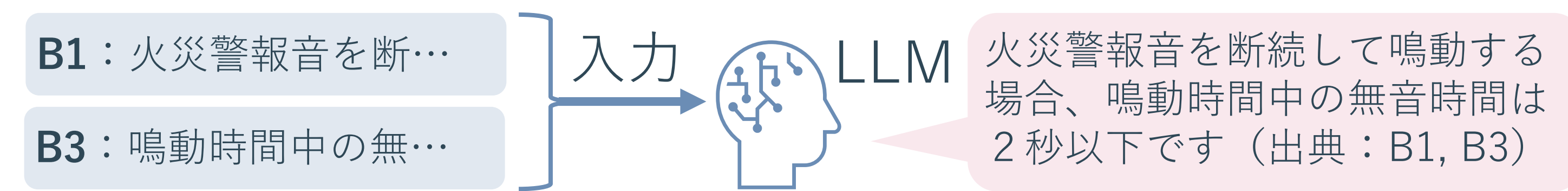
手順2：ブロック（箇条文）検索



手順3：近傍へ広げて根拠を補完



手順4：回答生成（根拠となる引用元を提示）



3. 実験設定と評価

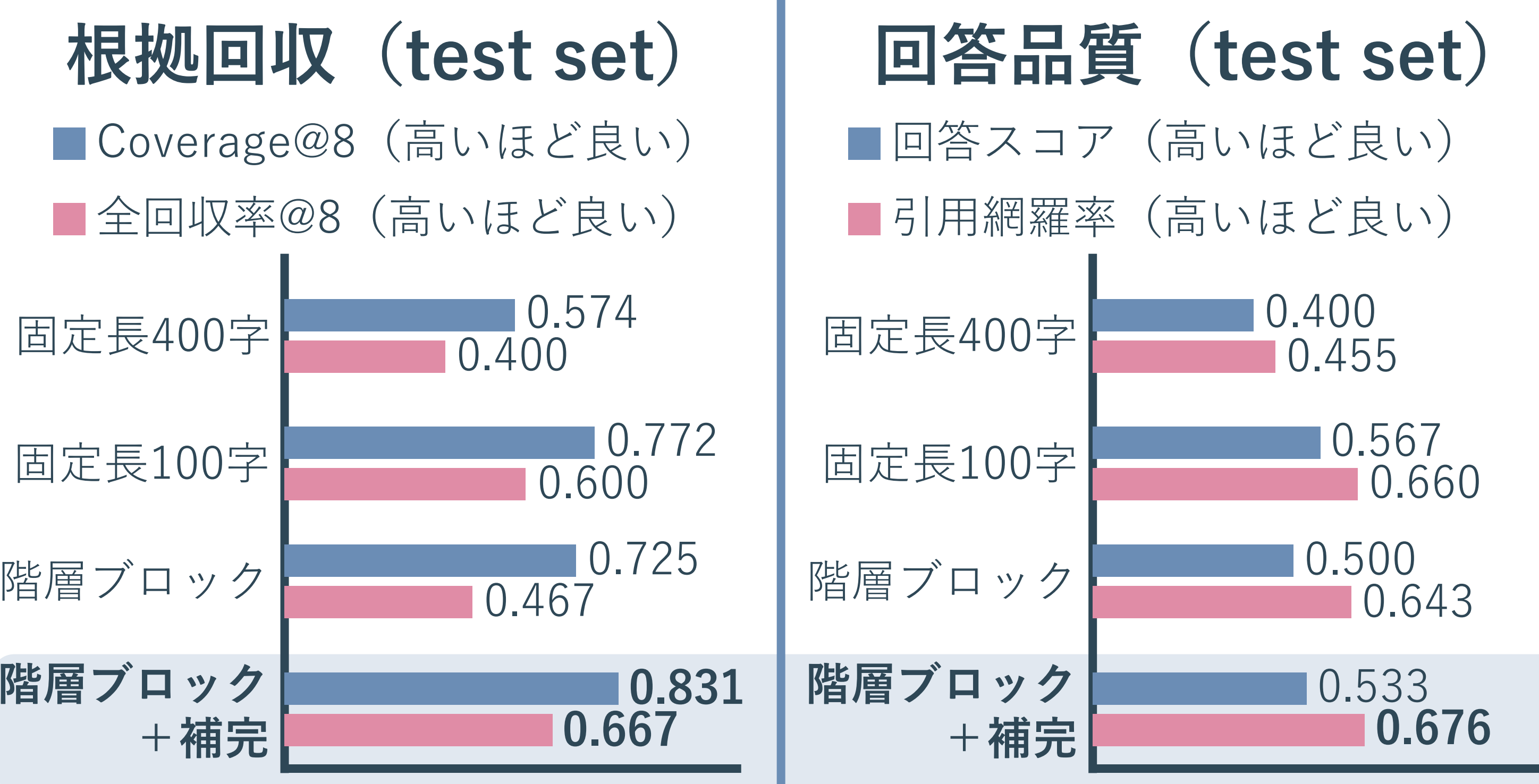
実験設定

- 対象文書：火災報知設備の感知器及び発信機の検定細則
- 階層ブロック数：189
 - 平均ブロック長：本文のみ56文字（+見出し：177文字）
- QA数：75問（dev60 / test 15）
- 比較手法：固定長チャンク（400字, 100字）、階層ブロック（補完なし）
- 回答生成：Qwen2.5-7B
- 採点：Qwen2.5-14B
- 回答入力根拠数：上位8件

評価

- Coverage@k：gold根拠の平均回収率
- 全回収率@k：gold根拠をすべて回収した割合
- 回答スコア：正解1 / 部分正解0.5 / 不正解0の平均
- 引用網羅率：生成回答の引用がgold根拠をどれだけ覆うか

4. 結果



- 提案法が Coverage@8 = 0.831, 全回収率@8 = 0.667 で最高となった
- 補完：Coverage@8 は0.725→0.831、全回収率@8は0.467→0.667に改善
- 回答品質でも引用網羅率が改善し dev で回答スコア・引用網羅率が最良

5. まとめ

結論：階層ブロック+文書構造グラフ補完は根拠網羅性を改善

考察

階層ブロックは見出し・箇条書き境界に沿うため固定長チャンクより文脈断片化を抑えやすい

限界

「第〇条」などの遠い参照は親子兄弟補完だけでは拾いきれない

今後

- 複数PDF・他分野の規定文書への拡張
- 遠い参照への対応