

蓄積された行政情報を簡単に発見できる
庁内システムを構築したい！

実証実験報告書

実証事業者：名古屋工業大学 白松研究室・ハイラブル株式会社
×
都市・交通局 航空空港課

目次

● 1. 実証実験の概要

- 背景・目的
- 対象となる業務の全体像
- 目標・KPI

● 2. 実証実験の結果

- KPIに関連する結果
- 工夫したポイント
- 実証実験で得られた発見

● 3. 課題と今後の展開

- サービス/システムにおける課題
- 導入に向けての課題と解決策

● 4. 課題担当課の評価等

- 実証実験で得られた発見
- 実証実験に対する評価

● 5. 参考資料

- サービス/システムの紹介

1. 実証実験の概要

背景・目的

データフォルダに蓄積された膨大な過去の情報から必要なものを職員が容易かつ的確に取り出せる仕組みを構築する。

- 背景：過去の膨大な行政文書から、必要な情報を発見する作業に苦労
 - 例：「5年前のあのときの経緯を説明する資料を作って」
 - 特に異動して日の浅い職員の場合、非常に時間がかかる
- 目的：膨大に蓄積された行政文書の中から必要な情報を職員が「容易かつ的確に」取り出せるようにしたい
 - 経験や知識量に関わらず、全ての職員が蓄積された行政情報を容易かつ正確に把握することで、新たな課題に対してより迅速で的確な意思決定が可能に
- アプローチ：大規模言語モデル (LLM; Large Language Model) を用いた検索拡張生成 (RAG; Retrieval-Augmented Generation) システムの構築

対象となる業務の全体像

大規模言語モデルを用いてデータフォルダに蓄積された行政文書から適した情報を検索し、適切に提示できるか検証する。

- 蓄積された膨大な行政文書から必要な情報を探し、資料作成や業務に応じた意思決定に活用
 - 過去の決定やプロジェクトの記録、法規や規程などが含まれる
- しかし、航空空港課だけでも2TBを超えるデータ量があり、目当ての文書を探し出すのは大変な労力が必要
 - 文書の保存年限やフォーマットの違いや、グループごとに異なるフォルダ構成等によって検索が困難に
- 特に経験の少ない職員は、必要な情報を見つける作業に苦労
 - 異動して日が浅いと、文書の保存場所や内容について知識が少ない場合が多い

目標・KPI

- 必要な情報を見つけるまでの時間短縮効果
 - 特に前提知識のない作業で大幅な時間短縮を目指す
- 見つけた必要な情報の正解率
 - 特に事前知識のない作業で大幅な正解率向上を目指す
- ファイル検索 (N-best解) の精度
 - 再現率、適合率、F値

2. 実証実験の結果

実証実験の方法

- 実験用の行政文書：航空空港課の実業務に関するファイルから選定
 - ファイル形式はdoc, docx, pptx, xls, xlsx, rtf
 - ファイル数は164
- 実験1: ファイル検索精度の検証
 - 航空空港課職員が、想定質問文と発見すべき正解ファイル(複数)を設定
 - システムのファイル検索機能を用いて適合率, 再現率, F値を測定
- 実験2: 正解率向上・作業時間短縮効果の検証
 - 被験者：航空空港課職員（前提知識あり）, DX推進課職員（前提知識なし, 行政経験あり）, 学生（前提知識, 行政経験なし）→ 各3名計9人
 - 手法：手作業・2種類のシステム（100文字分割, 4000文字分割）
 - 3つの手法それぞれで用意した一問一答に回答
 - 回答精度：正答率を検証
 - 作業時間：作業時間を検証
 - 作業時間の低減かつ正答率の向上を目指す。

開発したシステムのUI



回答



あいち航空ミュージアム100万人達成セレモニーの内容

あいち航空ミュージアムは、2017年11月30日にオープンし、2021年7月24日に累計来場者数100万人を達成しました。これを記念して、「来場者100万人達成感謝セレモニー」が以下の内容で開催されました。

日時	2021年8月20日（金）午前10時から午前10時15分まで
場所	あいち航空ミュージアム1階フライングボックス前ステージ
出席者	大村秀章愛知県知事、鈴木邦尚豊山町長など
	・ 閉会

状況をを入力してください。

あいち航空ミュージアム100万人達成セレモニーの内容を教えてください

ユーザ入力

参照リスト

参照リスト

- 1: R3/ミュージアム関係/0816記者発表資料（ミュージアム100万人記念）最終.docx
- 2: H29/イベント/02 記者発表資料（10万人達成セレモニー）.docx
- 3: H30/知事レク/05 知事発言要旨（1周年）3.docx
- 4: R4/ミュージアム/【記者発表資料】あいち航空ミュージアム100万人達成感謝セレモニー

文献

- 1: data/フォルダ構成/R3/ミュージアム関係/0816記者発表資料（ミュージアム100万人記念）最終.docx

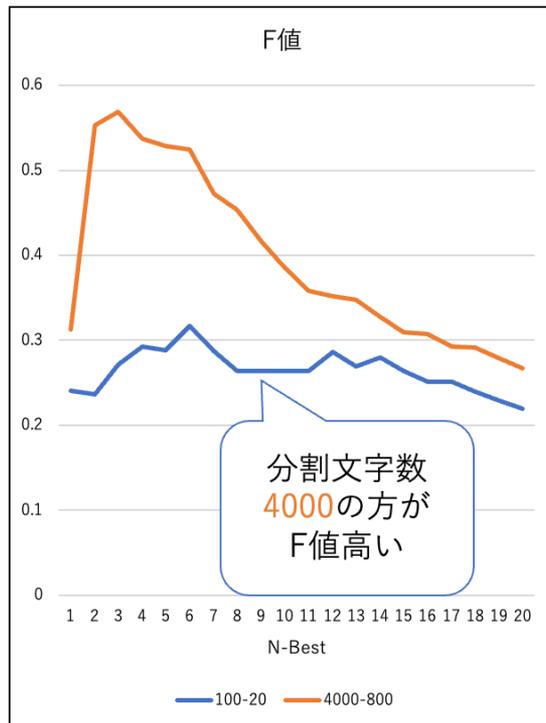
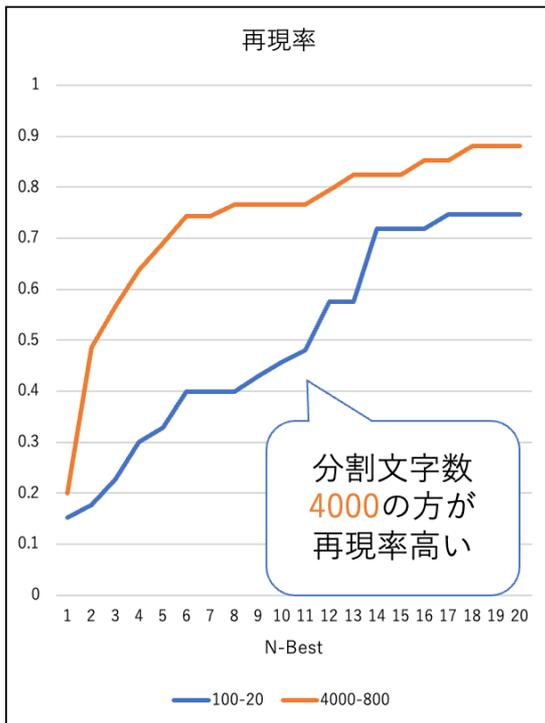
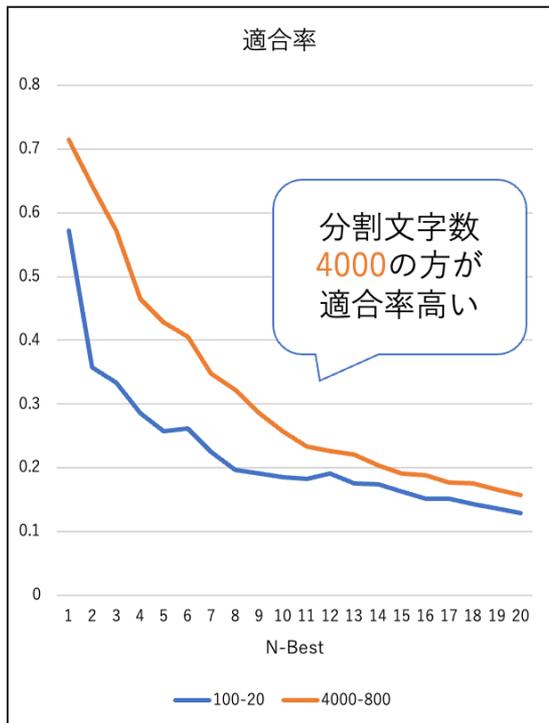
2021年 8月16日

文献内容

KPIに関する結果

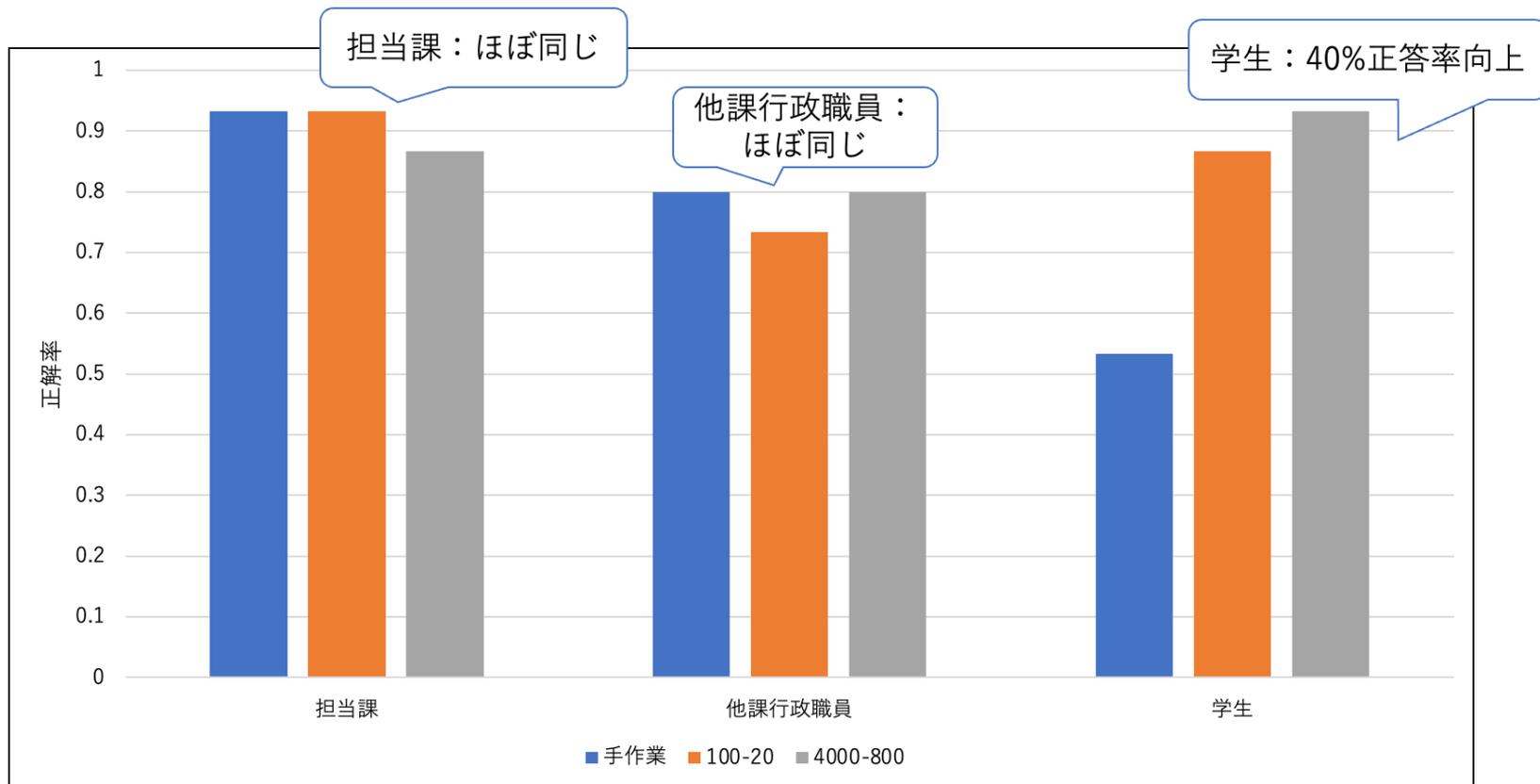
(100文字分割 / 4000文字分割) のF値等の比較

- N-Best (検索対象) による変動



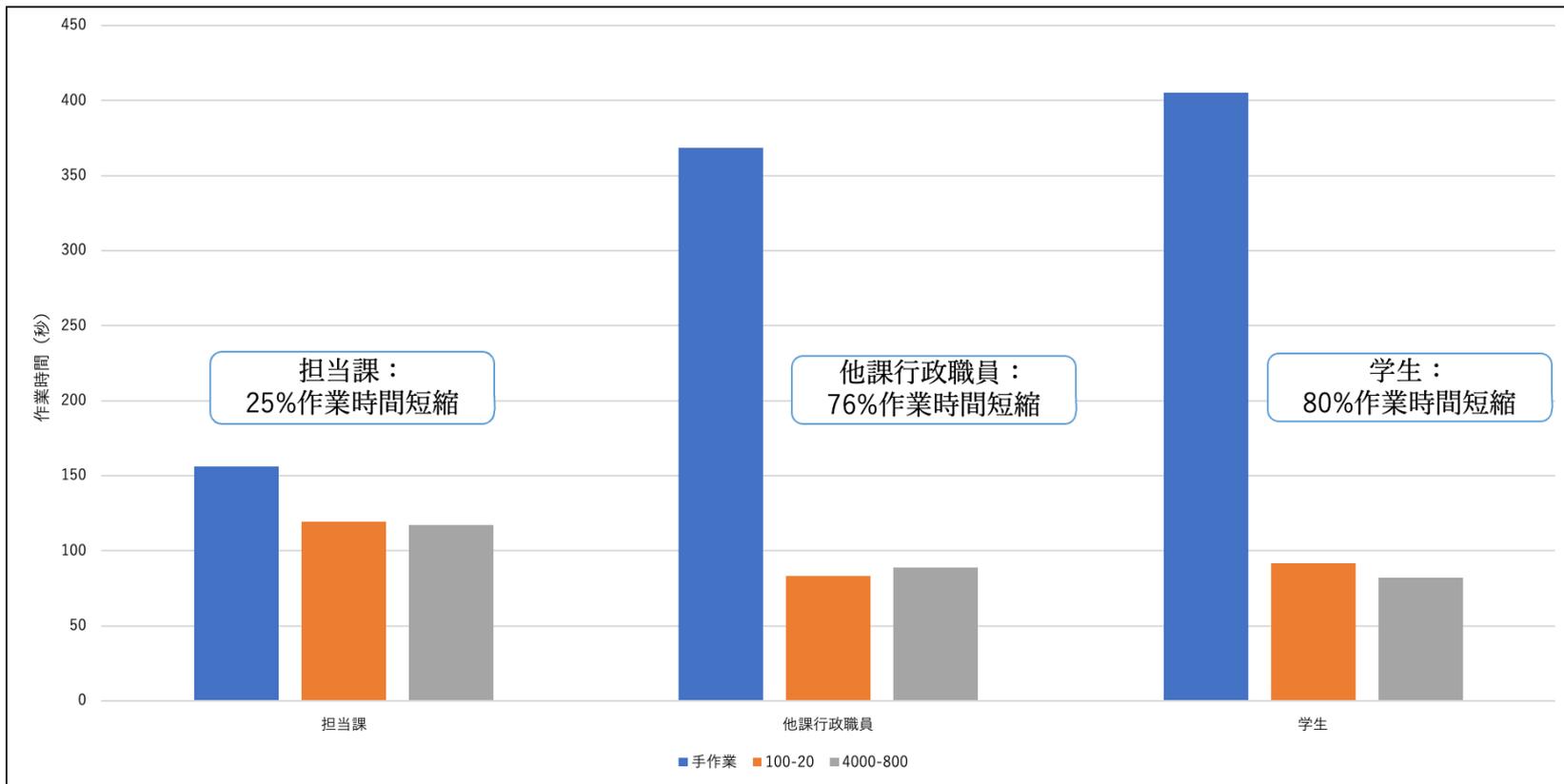
KPIに関する結果

(手作業 / 100-20 / 4000-800) ごとの短文問題正答率



KPIに関する結果

(手作業 / 100-20 / 4000-800) ごとの短文問題作業時間



KPI達成状況

- 必要な情報を見つけるまでの時間短縮効果：
削減幅24～80%、特に前提知識のない県職員や学生で大幅に時間短縮
 - 100文字分割：前提知識のある県職員: 24%削減、前提知識のない県職員: **78%**削減、学生: 77%削減
 - 4000文字分割：前提知識のある県職員: **25%**削減、前提知識のない県職員: 76%削減、学生: **80%**削減
- 見つけた必要な情報の正解率：
73～93%、特に学生でシステム不使用時53%→93%の40ポイント向上
 - 100文字分割：前提知識のある県職員：**93%**、前提知識のない県職員: 73%、学生: 87%
 - 4000文字分割：前提知識のある県職員：87%、前提知識のない県職員: **80%**、学生: **93% (40pt向上)**
- ファイル検索 (N-best解) の精度：
最大で0.57（ただしF値より再現率の重要性を確認）
 - 100文字分割：F値は最大で0.32 (N=6)、適合率は最大で0.57 (N=1)、再現率は最大1.0 (N=64)
 - 4000文字分割：F値は最大で**0.57** (N=3)、適合率は最大で**0.71** (N=1)、再現率は最大**1.0** (N=42)
 - 再現率を向上させることが重要であり、適合率が少し低くてもほとんど問題にならないことが明らかに
 - 適合率が最大0.71、F値が最大0.57でも、上記のように**93%**の正解率
 - なぜなら、RAGでは不正解文書が混じっていても正解文書のみを用いて説明を生成できるため

用語等の補足

- 再現率
 - 正解として見つけるべき項目のうち、システムが見つけられた項目の割合。正解が10個あり、8個見つけられた場合、再現率は0.8になる。
- 適合率
 - システムが正解と予測した項目のうち、実際に正解であった項目の割合。10個を正解と予測した中で、実際に正解であったのが7個なら、適合率は0.7になる。
- F値
 - 再現率と適合率はトレードオフの関係にあるため、それらの調和平均を取ったものがF値。
- N-Best解
 - 複数の候補の中から最も良いと思われる上位N個を選んだ解をN-Best解と呼ぶ。例えば、Googleで検索した上位10件の検索結果だけを見ると、N=10の「10-Best解」だけを見ていると言える。
- Retrieval Augmented Generation (RAG) と文書分割
 - 外部ソースの文書から取得した情報を検索し、それをLLMに入力して文章生成する技術をRAGと呼ぶ
 - LLMで入力できるテキストの長さには制限があるため、長い文書は分割して入力する必要がある
 - 本事業では、100文字分割と4000文字分割の2パターンを試行した

工夫したポイント

本システムの基本設計（図1）

1. ユーザのクエリと類似するDB内の関連文書を検索
2. 各関連文書から事前に抽出した時系列情報をプロンプトに追加
3. GPT-4 Turboにリクエスト
4. 生成結果と関連文書をユーザに提供

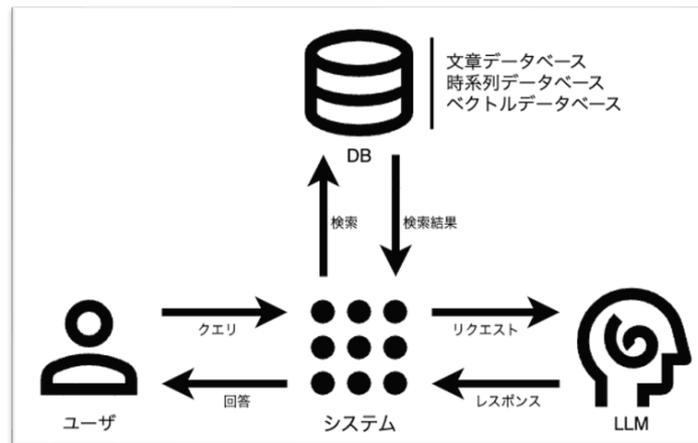


図1：本システムの基本設計

時系列情報を事前に文書から抽出し、プロンプトに追記

- 更新され続ける法規・規程の中では、同一テーマに関連する文書群に対して時間的に前後する情報が混在
- 事前に時系列的な整理と把握を行うことでシステム精度を向上

ユーザの置かれている状況を入力する欄を追加

- ユーザの状況を入力することで検索精度向上（図2）
- 状況を入力することでメタ認知に繋がる可能性

文書の分割文字数による検索精度と回答精度の変動を調査

- 今後の開発方針の決定に有用

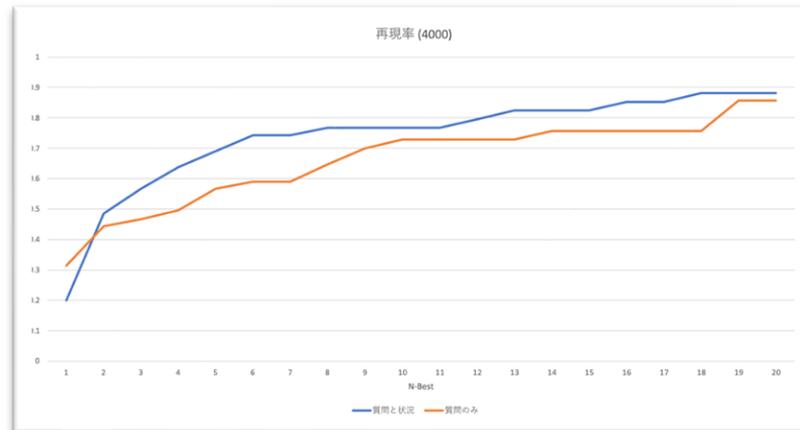


図2：状況の入力の有無による再現率の変動

実証実験で得られた発見

手作業では学生や他課行政職員は担当課職員の約2倍の作業時間を要している(図1:作業時間の比較)

手作業では、学生の正答率は担当課職員の約半分である(図2:正答率の比較)



システムを使用することでその差がほとんど無くなることがわかった(図1, 図2)

➤ システム使用は経験や知識の差を埋める効果がある

※担当課職員の時間短縮効果が低いのは、実験で検索対象としたファイル数を減らしたため経験者には手作業での発見が容易だったと考えられる。本来の膨大なファイル数を検索対象とした場合は、担当課職員についても時間短縮効果が確認できる可能性がある

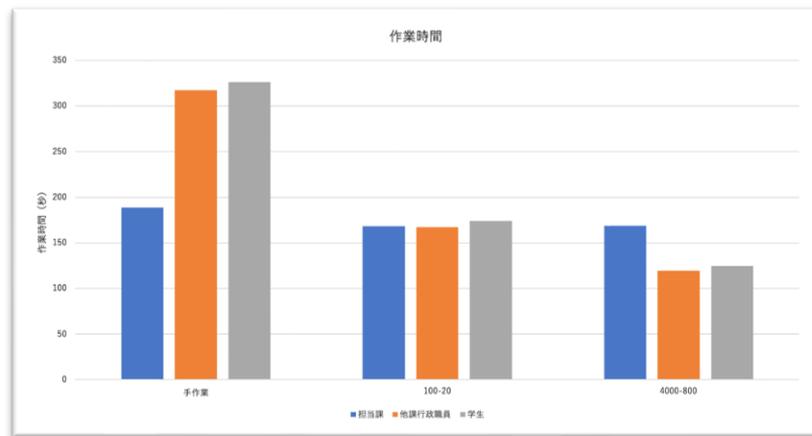


図1: 作業時間の比較

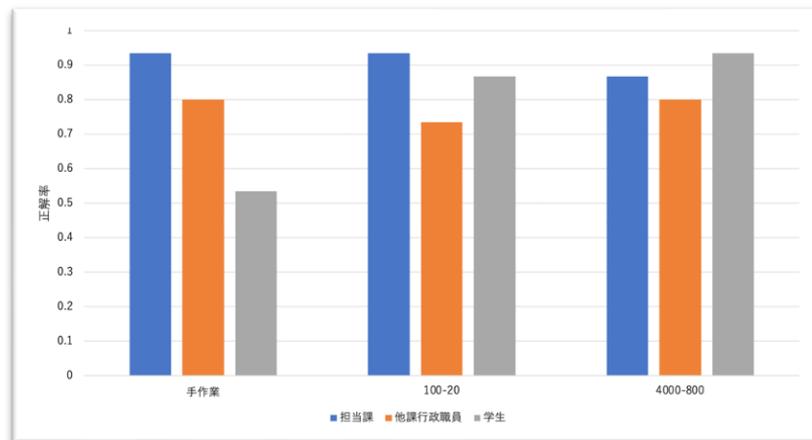


図1: 正答率の比較

3. 課題と今後の展開

サービス/ システムに関する課題

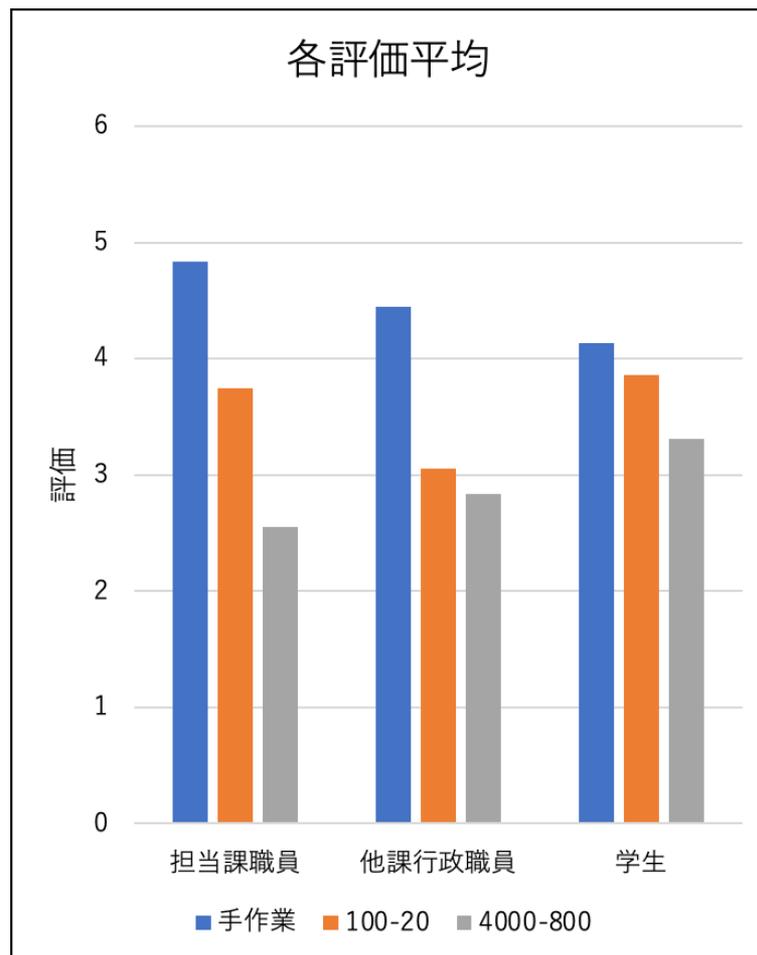
背景, 経緯, 理由に関する高度な記述問題について手作業とシステム使用による回答精度を比較した。
その結果, 手作業での回答の評価がより高くなった。

原因は

1. LLMの推論能力の限界
2. LLMに必要な情報が十分に供給されていないことが考えられる。

1については, 新たなモデルの開発に伴い, それらを試したり, オープンLLMを用いたファインチューニングによる試作を通して, 最適なモデルを選定する。

2について, 現在は文字数に基づいたテキスト分割を行っている。今後の研究では, LLMを活用して文脈を保持したテキスト分割にフォーカスを当て, 精度向上を目指す。



導入に向けての課題と解決策

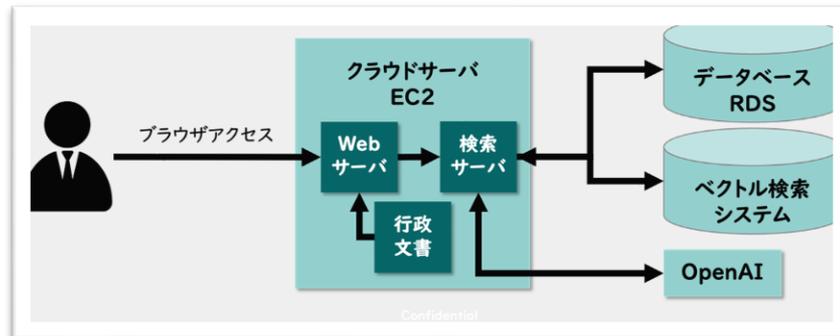
現状：

- ブラウザから検索ができる仕組みを Amazon Web Services (AWS) 上に構築済
- 行政文書の追加は別途メール等で共有することで可能
- ただし、システムはサーバ上に搭載されており利用者や文書がさらに増えたときのスケーラビリティが不足

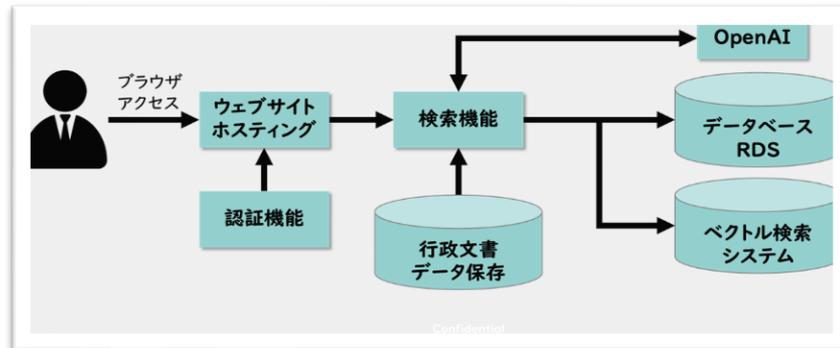
開発項目：

- 実証時より大規模な行政文書の管理
- 複数同時アクセスの処理
- アカウント管理
- 日々生み出される行政文書の受け渡し方法の確立
- 継続的な改善のための仕組み

現在のインフラ構成



改修後のインフラ構成案



4. 課題担当課の評価等

実証実験で得られた発見

大規模言語モデルを用いた応答システムにより、蓄積された行政文書から適した情報を検索し、適切に提示することが可能なことがわかった。一方で、実環境で業務利用するにはさらなる大量のファイルでの検証など、追加の検討が必要なことも判明した。

AI（LLM）は文の意味を「理解」するわけではなく、「語のつながり方の確からしさ」を見ているだけなので、できること（機能）には限界があることを改めて確認できた。

- a. 一つの文書から回答部分を取り出して自然な文章で回答する→ほぼできる
- b. 複数の文書から回答部分を取り出して自然な文章に組み立てる→できるが、たまに間違える
- c. 複数の文書の内容から導かれる（推察される）事柄を、文書に直接的に書かれていない言葉や表現も用いながらまとめる→できなさそう

今後の方向性としては、①上記bの精度向上、②実際の業務環境への適用に係る課題解決が想定される。

①については、状況の入力等、一見、回答の精度に影響を与えないように思われる情報が実際は影響を及ぼしているということは新たな発見であると同時に、使用者側のシステムへの聞き方も重要であると感じた。

②については、情報セキュリティに配慮すること、現在の業務環境からの変更（余計な作業）を最小限に止めることなどが課題となる。

実証実験に対する評価

「だれでも」「容易に」「的確に」必要な情報を取り出せるようにしたいという本実験の課題については、概ね期待通りの成果があり、現状の成果物であっても、実装されれば相当の業務負担軽減につながると評価できる。

一方、本実験は外部（AWS）に実験環境を構築する必要があり、情報セキュリティ上、事前に選別した少数のファイルのみを対象としたが、実環境では日々新たなファイルが作成されており、都度、外部にアップロードしなければならないのは実用性に欠ける。また、GPTのような外部のLLMとデータをやり取りすることは情報セキュリティ面でのリスクが残る。

従って、今後は、オープンLLM等の活用により（精度が多少犠牲になったとしても）県のネットワーク内で動く仕組みを構築していくことが望ましい。

5. 參考資料

