



「研究大学におけるAIの可能性を考える」

Researcher2VecとLLMによる検索システム

IR

岡山大学 研究・イノベーション共創機構

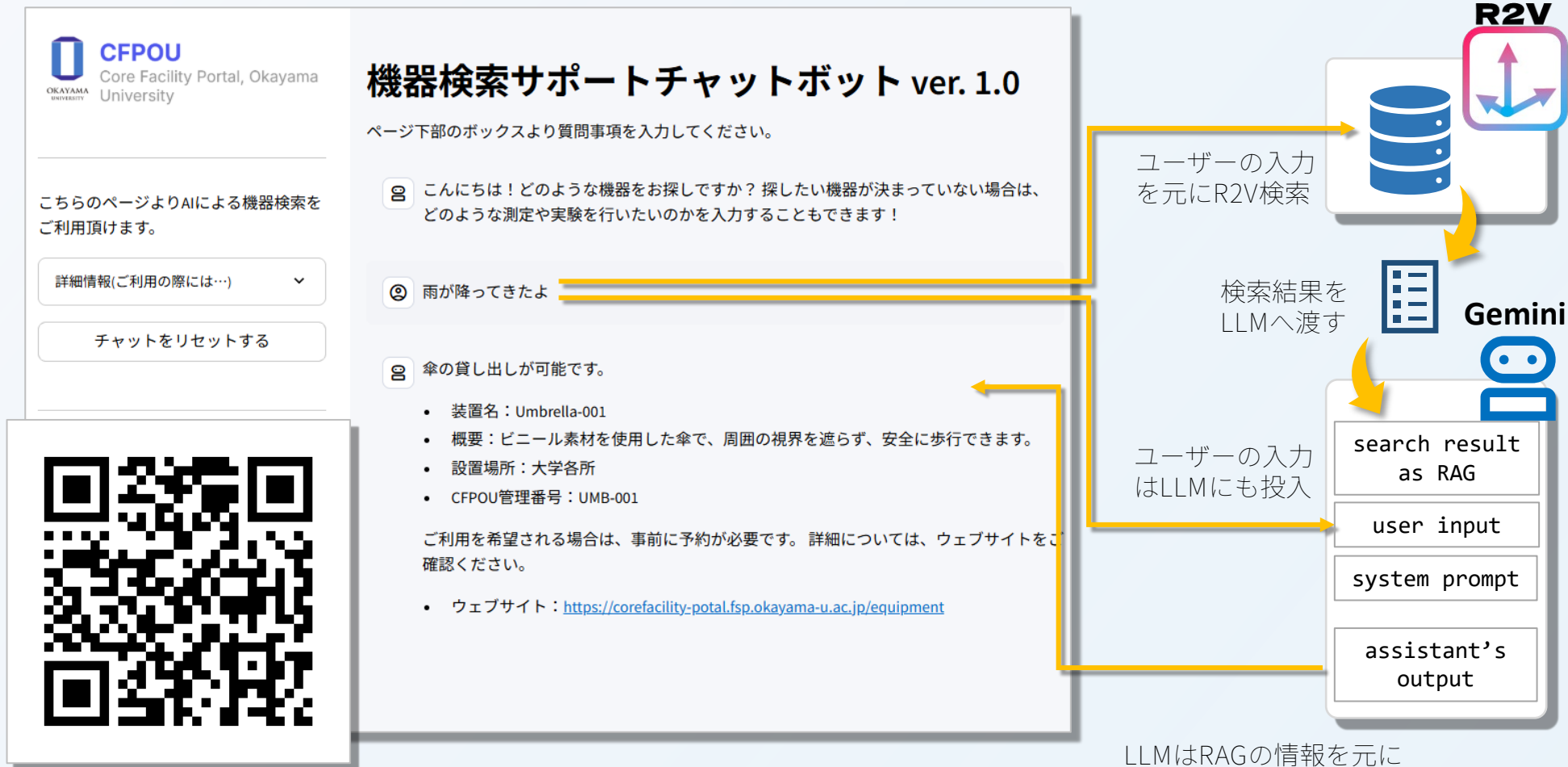
URA 小林 健輔

- 現状の機器共用ポータルサイトではキーワード検索に限界
 - キーワード一致かつタグ付けされていなければ検索できない.
 - ユーザーエクスペリエンス(UX)は非常に悪い. どうしたものか. . .
- MIRAI-DXでは統計数理研究所 持橋先生のResearcher2Vec(R2V)^[1]を活用
 - 論文情報から研究者をベクトル化し, 研究者の専門性を考慮した検索が可能に
 - 元データを論文ではなく機器情報に入れ替えれば, 機器検索ができる!
- ただし, R2Vを使えば専門性は考慮されるが, 出力はリスト形式
 - 依然としてUXを高めなければならない問題は残る
 - R2VとLLMを組み合わせればよいのでは?

R2VとLLMを組み合わせた, 機器検索チャットボットの作成・実装へ

[1] 持橋大地, “Researcher2Vec: ニューラル線形モデルによる自然言語処理研究者の可視化と推薦”

- R2Vにより専門性を考慮した検索を行い，出力をRAGとしてLLMへ投入



- なぜ2週間での短期開発が可能であったのか？
 - 機器情報は既存サイト用のデータが再利用可能な形式で利用できた

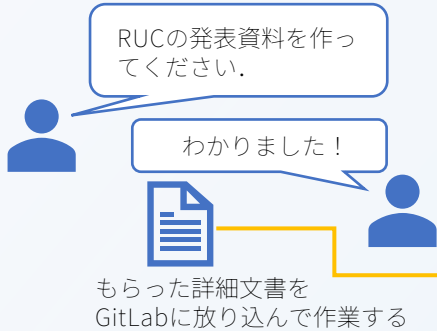
- 当システムを応用する際に問題になるのは、すぐに使えるデータの有無
 - WordやExcel, PDFのデータは再利用が難しい
(修飾は人間には有効だがAIには無効.むしろ**文書構造**を提供すべき)

 - 業務で扱うデータを再利用可能に, AIとの協働を見据えて準備すべき

- 「とりあえず動かせ」の仕組みも大事
 - トライアンドエラーを何度も実行するアジャイル型の取り組み
打ち合わせの間にワンループでもまわせ, ができるか
AIを使ったら5分で実装できるかもしれない

- 岡山大学 研究IRチームはGitLabを使い，業務データを管理

例えば，



名前	最新コミット	最終更新
20260213_発表資料.pptx	データを追加	22時間前
251208_RUCs_第9回_A4チ...	データを追加	22時間前
README.mnd	データを追加	22時間前

プロジェクト情報

- 2個のコミット
- 1個のブランチ
- 0個のタグ
- 5.4MB プロジェクトストレージ

README

プロジェクト管理・データ管理をGitLabで行う

20260213_RUC

1 # 20260213_RUC

2 ## To do

3

4

5 以下ppt中。

6 # 「研究大学におけるAIの可能性を考える」\n\nResearcher2VecとLLMによる検索システム

7 ## 岡山大学 研究・イノベーション共創機構\n\nnURA 小林 健輔

8

9 ## 背景

10 ## 現状の機器共用ポータルサイトではキーワード検索に限界

11 - キーワード一致かつタグ付けされていない場合は検索できない。

12 - コーディングスペースリソース(OX)は非常に悪い。どうしたものか...

13 ## MIRAI-OXでは統計検定研究所 特許発生のResearcher2Vec(R2V)[1]を活用

14 - 論文情報から研究者をベクトル化し、研究者の専門性を考慮した検索が可能に

15 - 元データを論文ではなく機器情報に入れ替えれば、機器検索ができる！

16 ## ただし、R2Vを使えば専門性は考慮されるが、出力はリスト形式

17 - 偶然としてOXを念の付けなければならない問題は残る

18 - R2VとLLMを組み合わせた場合は？

19 ## **R2VとLLMを組み合わせた、機器検索チャットボットの作成・実装へ**

20

21 ## システム構成

22 ## R2Vにより専門性を考慮した検索を行い、出力をRAGとしてLLMへ投入

23

24 ## 研究大学が向き合ふべき課題と活用術

25 ## なぜ2週間での期間限定が可能であったのか？

26 機器情報は既存サイト用のデータが再利用可能な形式で利用できた

27 ## 当システムを応用する際に問題になるのは、すぐに使えるデータの有無

28 - WordやExcel、PDFのデータは再利用が難しい\n\n(移動は人間には有効だがAIには無効。むしろ**文書整理**を推奨すべき。)

29 - 業務で扱うデータを再利用可能に、AIとの協働を見据えて準備すべき

30 ## 「とりあえず動かせ」の仕組みも大事

31 - トライアンドエラーを何度も実行するアジャイル型の取り組み

32 打ち合わせの間にツールでできてしまえば、安全に多歩行が可能にしたい

33 AIを使ったら5分で実装できるかもしれない

34

35 ## 参考資料 プロンプト詳細

36 ****

37 Use the following context as your learned knowledge, inside <context>/</context> XML tags.

38

39 <context>

40 # 装置一般名: 傘

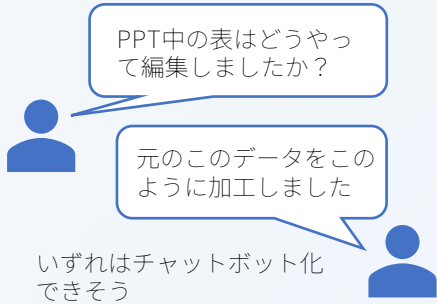
41 - 装置名: Umbrella-001

42 - 概要: ビニール素材を使用した傘で、周囲の視界を遮らす。安全に多歩行が可能にしたい

43 ****

目次 | 1 | 1 | スペース4 | UIF

こんなときは？



実作業はローカルPC，もしくはGitLabのweb IDEで行う

- AIフレンドリーなデータを蓄積する必要あり！
 - 手順や作業内容をAIがわかるテキストデータにして蓄積する
 - オープンサイエンスにおけるデータ管理と同じ

- AI-based CFPOUで使用しているプロンプトの例
(一般的なRAGの実装を単純化し、面倒なところは後段LLMに処理させている)

ここがRAG

```
Use the following context as your learned knowledge, inside
<context></context> XML tags.

<context>
# 装置一般名称：傘
- 装置名：Umbrella-001
- 概要：ビニール素材を使用した傘で、周囲の視界を遮らず、安全に歩行できます。
- 設置場所：大学各所
- CFPOU管理番号：UMB-001
</context>

When answer to user:
- If you don't know, just say that you don't know.
- If you don't know when you are not sure, ask for clarification.
Avoid mentioning that you obtained the information from the context.
And answer according to the language of the user's question.
```

ここがプロンプト
(実際にはもう少し長文)

```
あなたは誠実で優秀な日本人のアシスタントです。常に日本語で回答してください。
与えられた知識を元に応答を行い、不確かな場合にはわからないことを伝えます。

岡山大学で共用が可能な測定装置や機器に関して、ユーザーから質問されるため、利用可能な
装置を提案します。
必ず最初に機器の装置一般名称を答えて、その後に装置名や拠点、管理番号を適宜改行を入れて
答えます。
```

ユーザーの入力は
ここに追加される

“user”：“雨に対応できる装置は？”

実際に使用しているRAGの一部