



【今さら聞けないRAG入門】

現場で役立つRAGの基礎と設計

～社内情報資産をビジネス価値に変えるAI活用の考え方～

株式会社オープン
観光ビッグデータ事業部

📅 2025年8月29日



自己紹介・事業部紹介

講師プロフィール



- ✓ 氏名:板坂 剛志
- ✓ 所属:観光ビッグデータ事業部
- ✓ 担当:機械学習やAI技術を応用したサービスの設計・開発
- ✓ 略歴:メーカー系SE、半導体メーカー、CADベンダー、国内SIer、外資系コンサル、個人事業主、海外MBA取得など、精力旺盛に活躍した後、2019年、(株)オープトーン入社

よろしく
お願いします！

事業部概要



観光予報プラットフォームを見る

<https://kankouyohou.com/>

とっとり宿泊予報プラットフォームを見る

<https://lp.tottori-dx.info/>

観光ビッグデータ事業部では、観光・地域創生分野に特化したIT基盤やデータ分析サービスを提供しています。

観光ビッグデータ事業部

- 📈 宿泊データ分析・需要予測
- 🏠 地域基盤構築・スマートシティ対応観
- 🤖 観光DX・AI活用支援
- 👥 地域課題改善に向けた取り組み

今日のゴール

今日のゴール

- ✓ RAG（検索拡張生成）の仕組みを
ざっくり理解する
- ✓ 社内データを"価値あるナレッジ"に変える
ヒントを持ち帰る
- ✓ 自社での活用イメージ・最初の一步をつかむ

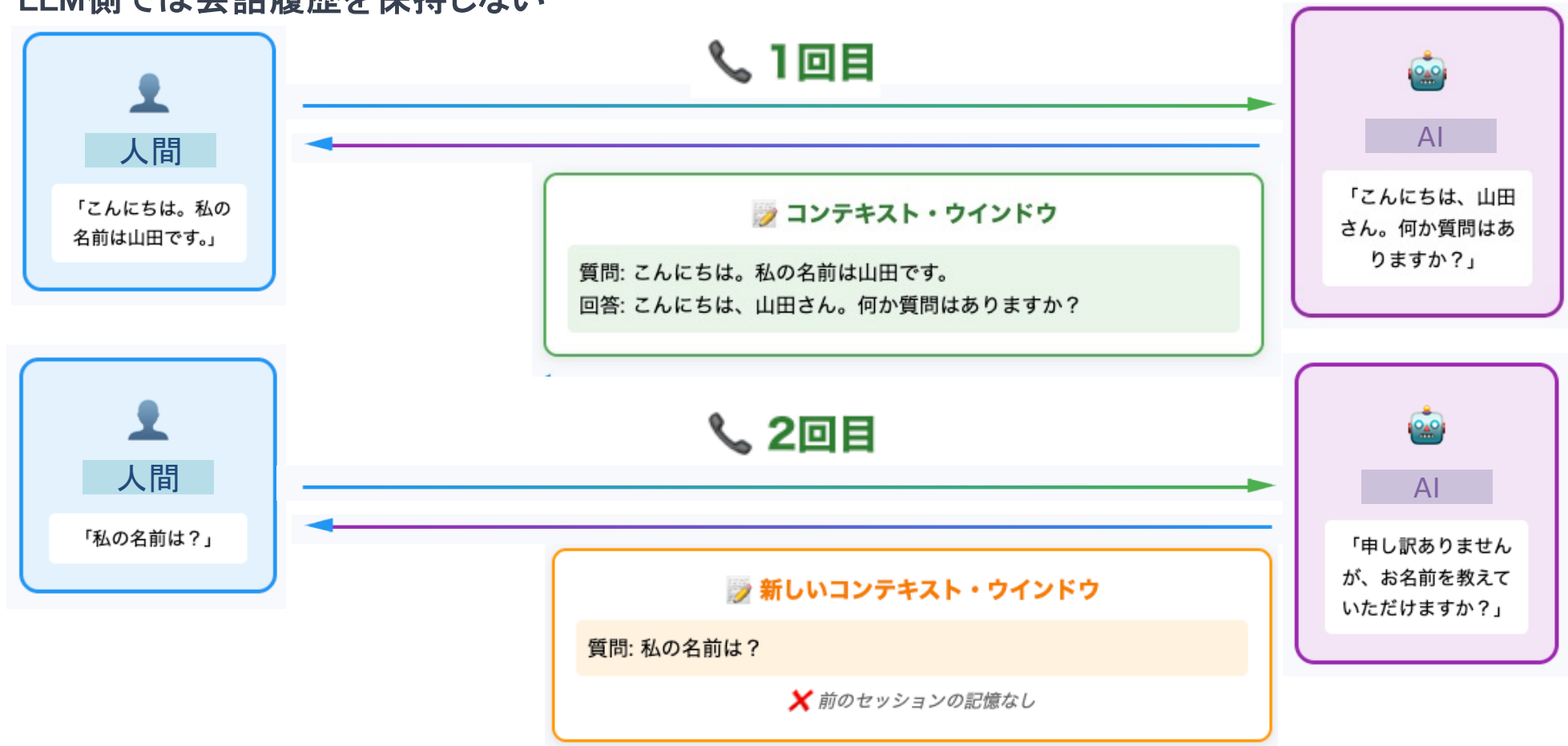


別途ご案内のダウンロードサービスを利用された方には、資料後半にサンプルコード特典が付いてきます。



コンテキスト・ウィンドウは有限

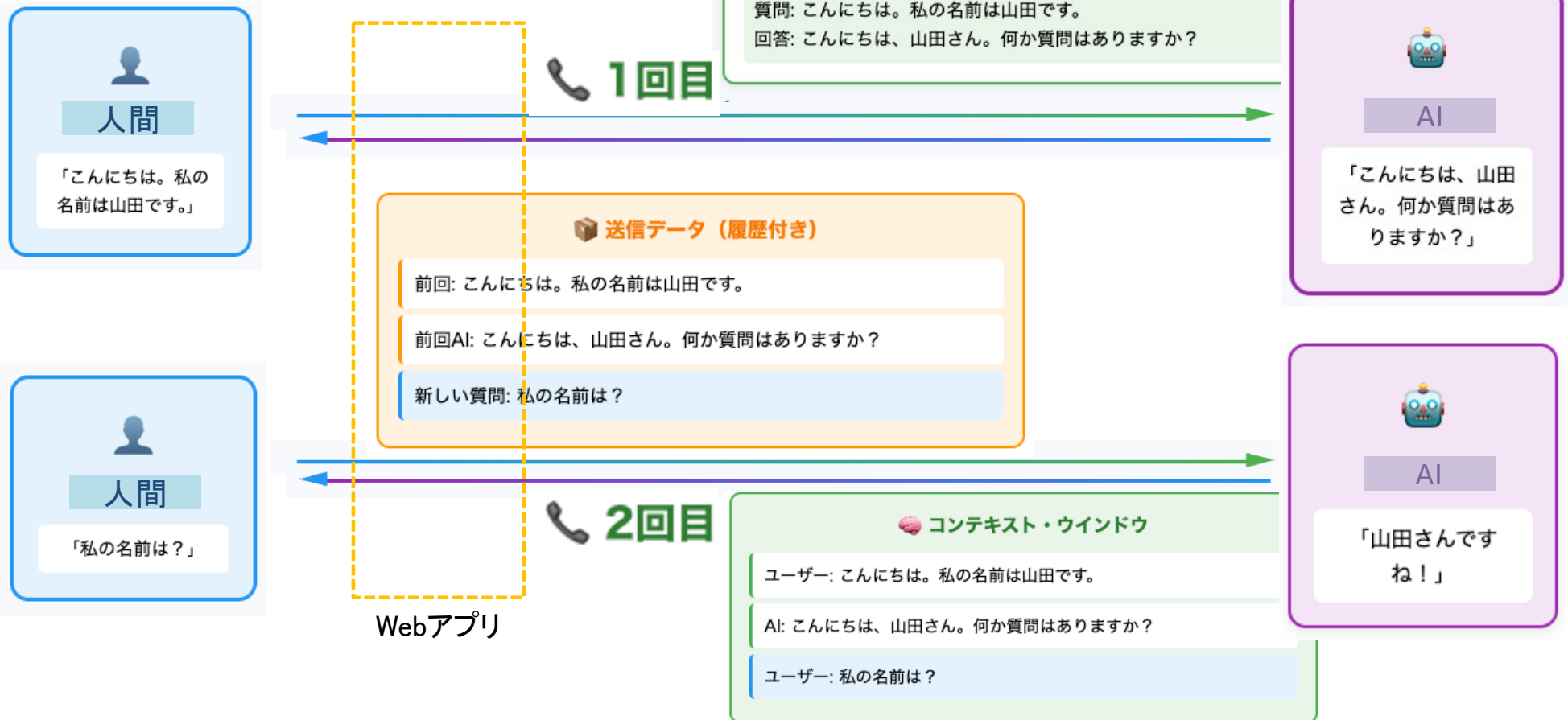
LLM側では会話履歴を保持しない





AIとの会話はステートレス

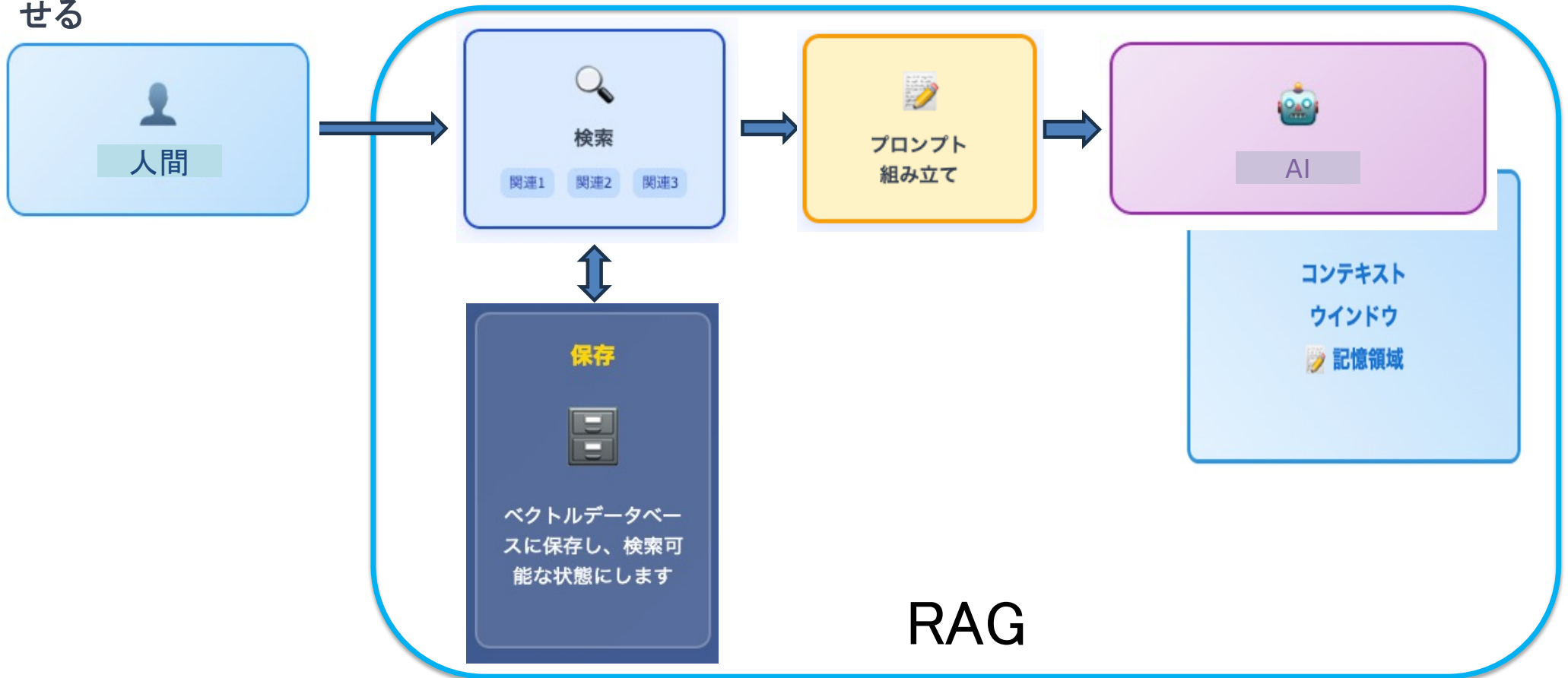
アプリ側で会話履歴を保持する





RAG (Retrieval Augmented Generation) : 検索拡張生成

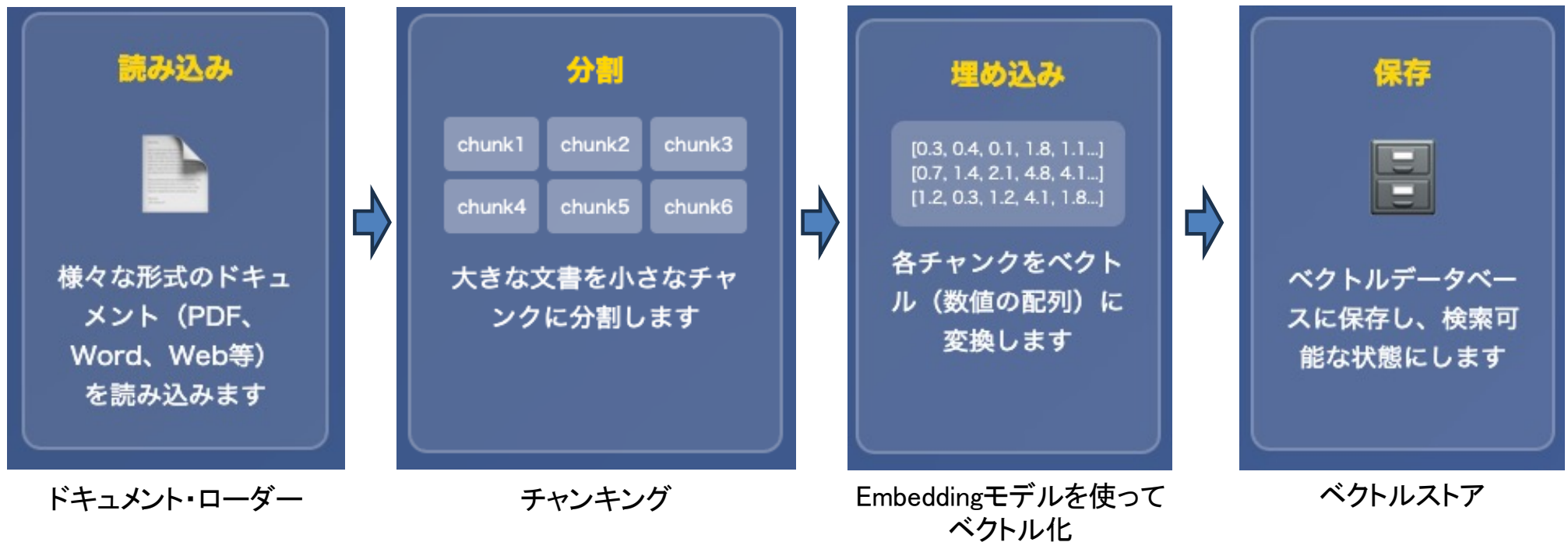
ベクトルストアから関連情報を取得して、ユーザ入力と合わせてプロンプトを作成してLLMへ問い合わせる





ベクトルストアの構築

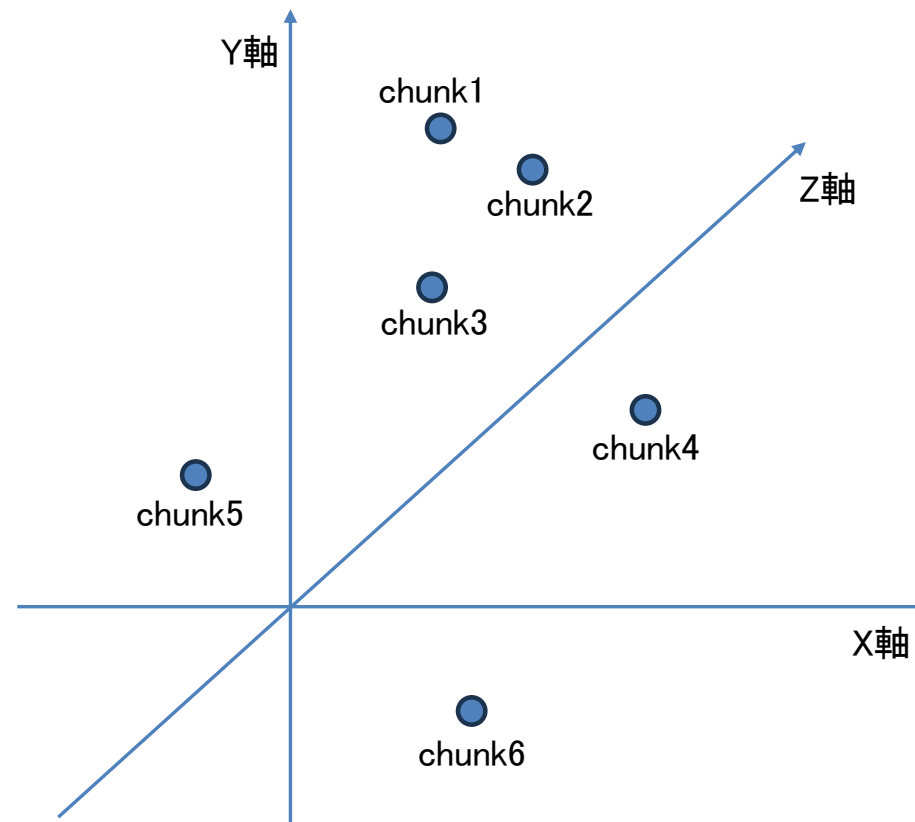
Embeddingモデルを使ってチャンクをベクトル(座標)に変換してベクトルストアに格納する





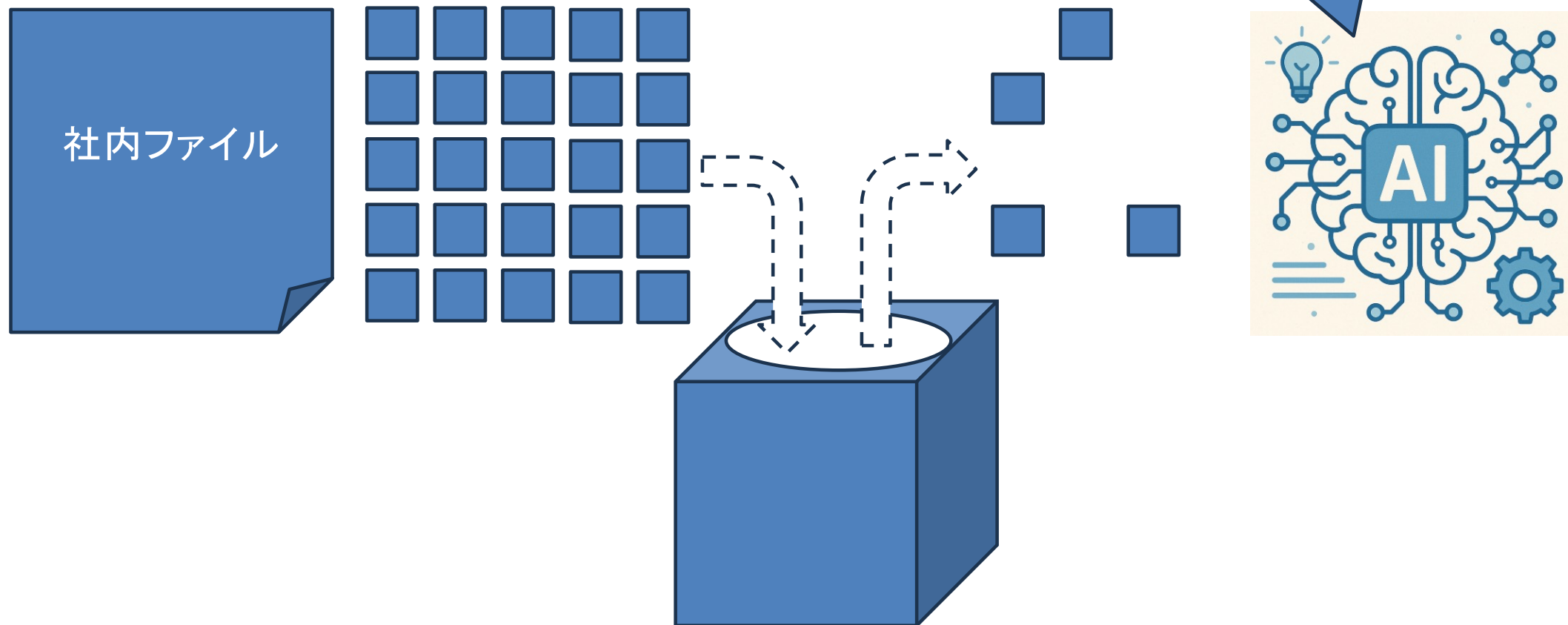
ベクトル空間

実際のエンベディングモデルでは数千次元ベクトルに変換される





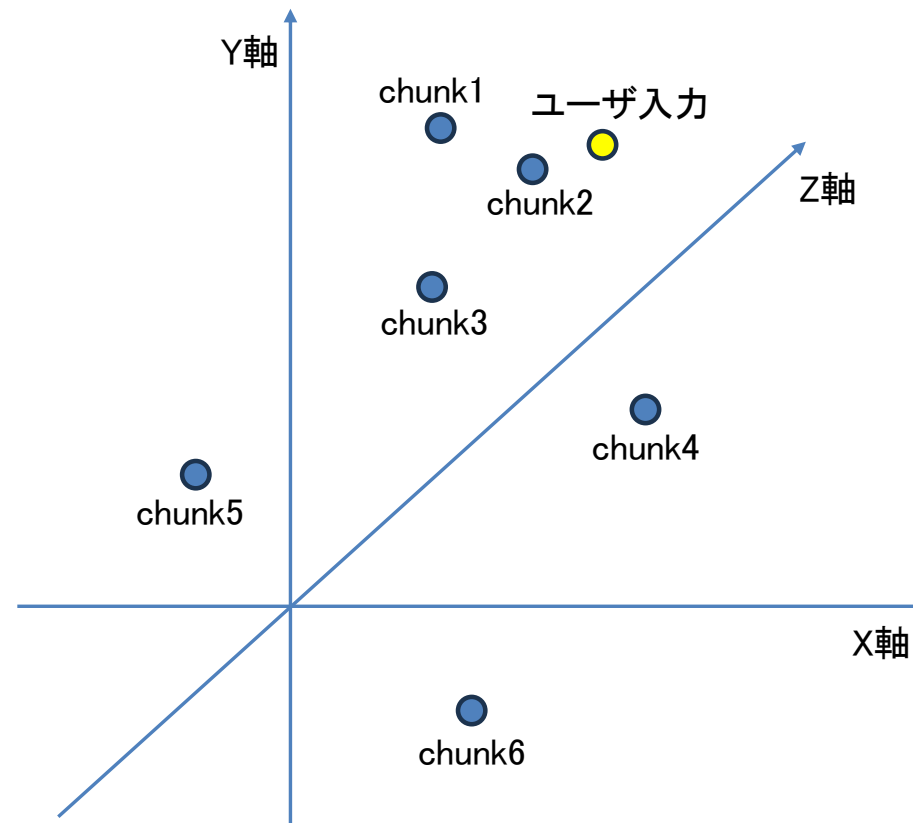
RAGのイメージ





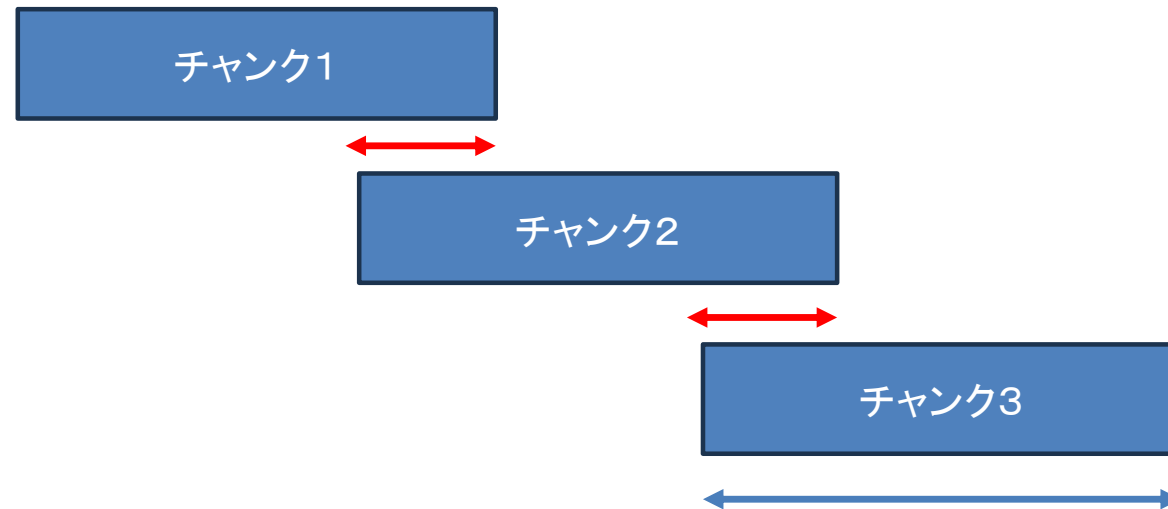
意味類似検索

- ・ユーザ入力をベクトル化して、各ベクトルとの距離から関連性の高さを判定する
- ・距離が近い＝関連性が高い＝ユーザ入力にAIが回答する際に有益な情報である



チャンキング

- ・チャンクサイズ(青線部分)
- ・オーバーラップサイズ(赤線部分)



チャンキング



利用用途／文書構造に合わせたチャンキング

商品1の説明
商品2の説明
商品3の説明

商品4
商品5
商品6
商品7
商品8の説明
商品9の説明
商品10の説明
商品11の説明
商品12の説明
商品13の説明
商品14の説明

商品1の説明

商品2の説明

商品3の説明





RAG事例紹介：日観振デジタルプラットフォーム



資料をダウンロード

利用申し込み

使ってる？デジプラ！

日本観光振興
デジタルプラットフォーム

地域のマーケティングデータ戦略支援ツール

**全国観光DMP・
高度化地域DMP**

「効果につながるデータ分析・活用術」をご紹介

お問い合わせはこちら

<https://lp.nihon-kankou-dx.info/>



RAG事例紹介：日観振デジタルプラットフォーム

新機能「AIレポート」を開発中

レポート管理

レポートの一覧です。内容を確認したり、ダウンロード・削除ができます。

データ登録

+ 新規レポート作成

観光動向レポート

観光動向分析

作成済み

📅 作成日時: 2025-08-04 17:45:34

📄 PDFダウンロード

表示 >

AIレポートテスト

観光動向分析

作成済み

📅 作成日時: 2025-07-18 16:35:51

📄 PDFダウンロード

表示 >

AIレポート 20250718

観光動向分析

作成済み

📅 作成日時: 2025-07-18 10:53:24

📄 PDFダウンロード

表示 >



事例紹介：日観振デジタルプラットフォーム

データ管理

[📁 レポート管理](#)[+ 新規レポート作成](#)

AIレポート作成に必要なデータファイルを管理します。カテゴリ別にテキスト、PDFなどファイルをアップロード・整理できます。

日観振デジタルPFのダッシュボードレポート



ファイルをここにドラッグ&ドロップするか、クリックして選択してください。

対応ファイル形式: .doc, .docx, .md, .pdf, .txt

アップロードされたファイル: 0 個

[📁 アップロード](#)

ステータス	ファイル名	アップロード日時	アクション
アップロード済み	4.5_姫路市レポート_2025-03-14T02_29_01.pdf	2025-07-17 10:51:21	削除
アップロード済み	4.4_観光施設別入込状況_2025-03-14T02_27_40.pdf	2025-07-17 10:51:18	削除
アップロード済み	4.3_姫路市観光入込概況_2025-03-14T02_25_47.pdf	2025-07-17 10:51:12	削除



事例紹介：日観振デジタルプラットフォーム

レポート作成

[← レポート一覧に戻る](#)[データ登録](#)

AIを活用した観光データ分析・アクションプランの管理

レポートタイトル *

観光動向レポート

レポートタイプ

観光動向分析

✓ 観光動向分析

SWOT分析

観光施策検討

インバウンド戦略

オーバーツーリズム対策

使用するデータ

日観振デジタルPFのダッシュボードレポート

すべて解除

使用するデータを選択してください

☒ 4.5_姫路市レポート_2025-03-14T02_29_01.pdf

☒ 4.4 観光施設別入込状況 2025-03-14T02_27_40.pdf



事例紹介：日観振デジタルプラットフォーム

レポート作成

AIを活用した観光データ分析・アクションプランの管理

← レポート一覧に戻る

データ登録

レポートタイトル *

観光動向レポート

レポートタイプ

観光動向分析

対象期間

2024年

対象地域

兵庫県姫路市

使用するデータ

日観振デジタルPFのダッシュボードレポート

使用するデータを選択してください

- ☒ 4.5_姫路市レポート_2025-03-14T02_29_01.pdf
- ☒ 4.4_観光施設別入込状況_2025-03-14T02_27_40.pdf
- ☒ 4.3_姫路市観光入込状況_2025-03-14T02_25_17.pdf

すべて解除



事例紹介：日観振デジタルプラットフォーム

AIレポート20250718

2025-07-18 10:53:24 に生成されたAIレポートです。

[← レポート一覧に戻る](#)[+ 新規レポート作成](#)[削除](#)

レポート内容

AIレポート20250718：兵庫県姫路市 観光動向分析レポート

対象地域：兵庫県姫路市

対象期間：2024年度（カレンダーイヤー2024として分析）

1. はじめに

姫路市は「国宝姫路城」をはじめとする豊富な歴史文化資源と、手柄山・書写山などの自然景観、姫路セントラルパークや好古園といった体験型レジャー資源を併せ持つ全国有数の観光地です。本レポートでは、提供資料をもとに2024年度の観光動向を分析し、施策立案やプロモーション戦略のための示唆を得ることを目的とします。

2. 分析の視点とデータソース

- 分析対象：姫路市内の観光資源、宿泊動向、入込客数、Web動向
- 期間設定：2024年1月～12月（資料の多くがカレンダーイヤー表示のため）

レポートについて質問

[会話を削除](#)[インバウンド動向についてももう少し詳しく](#)

インバウンド動向の詳細分析

2024年度の姫路市におけるインバウンド（訪日外国人）動向を、以下の切り口で詳しく解説します。前回レポートの概況を踏まえつつ、国籍・月次推移・属性・訪問行動の各データを整理しました。

1. 総量・月次推移

・延べ宿泊者数（来訪地：姫路市）

2024年のインバウンド延べ宿泊者数は、月別に見ると1月に約10千人泊からスタートし、春先にかけて徐々に増加。夏季ピークの7-8月にはそれぞれ約100-120千人泊に達し、9月以降は秋冬にかけてやや減少する典型的な季節変動を示しました。年間を通じて見ると、コロナ前の2019年度比でおおむね70～80%程度まで回復していると推定されます。



事例紹介：日観振デジタルプラットフォーム

AIレポート20250718

← レポート一覧に戻る

+ 新規レポート作成

削除

2025-07-18 10:53:24 に生成されたAIレポートです。

レポート内容

AIレポート20250718：兵庫県姫路市
観光動向分析レポート

対象地域：兵庫県姫路市
対象期間：2024年度（カレンダーイヤー2024として分析）

レポートについて質問

会話を削除

1. はじめに

姫路市は「国宝姫路城」をはじめとする豊富な歴史文化資源と、手柄山・書写山などの自然景観、姫路セントラルパークや好古園といった体験型レジャー資源を併せ持つ全国有数の観光地です。本レポートでは、提供資料をもとに2024年度の観光動向を分析し、施策立案やプロモーション戦略のための示唆を得ることを目的とします。

2. 分析の視点とデータソース

- 分析対象：姫路市内の観光資源、宿泊動向、入込客数、Web動向
- 期間設定：2024年1月～12月（資料の多くがカレンダーイヤー表示のため）
- 主なデータソース：
 - 宿泊旅行統計調査（観光庁）を基にした国内・インバウンド宿泊動向

4. 高価格帯マーケットの深耕

欧米豪市場では高価格帯利用が多いため、高級旅館・ホテルとの連携プランやMICEツアーの開発で付加価値を提供すると、客単価向上が期待できます。

以上、2024年度の姫路市におけるインバウンド動向の詳細分析でした。これらの示唆をもとに、国籍別・属性別にきめ細やかなプロモーション戦略を策定してください。

この2024年のインバウンド動向から2025年のアクションプランを考えて

2025年インバウンド向けアクションプラン（姫路市）

2024年のインバウンド動向を踏まえ、2025年に向けて以下のアクションプランを提案します。前回レポートで挙げた「欧米豪・東南アジア市場」「女性・カップル層」「季節変動」「城周辺集中」といったポイントを具体策に落とし込みました。

このレポートについて質問してください...

送信



AIレポート20250718：兵庫県姫路市観光動向分析レポート

対象地域：兵庫県姫路市

対象期間：2024年度（カレンダーイヤー2024として分析）

1. はじめに

姫路市は「国宝姫路城」をはじめとする豊富な歴史文化資源と、手柄山・書写山などの自然景観、姫路セントラルパークや好古園といった体験型レジャー資源を併せ持つ全国有数の観光地です。本レポートでは、提供資料をもとに2024年度の観光動向を分析し、施策立案やプロモーション戦略のための示唆を得ることを目的とします。

2. 分析の視点とデータソース

- 分析対象：姫路市内の観光資源、宿泊動向、入込客数、Web動向
- 期間設定：2024年1月～12月（資料の多くがカレンダーイヤー表示のため）
- 主なデータソース：
 - 宿泊旅行統計調査（観光庁）を基にした国内・インバウンド宿泊動向
 - QuickSightダッシュボードによるゾーン・施設別入込客数
 - 「JAPAN 47 GO」サイトの表示回数データ
 - 観光資源データベースによるカテゴリ別資源数

お疲れ様でした

今日のゴール

- ✓ RAG（検索拡張生成）の仕組みを
ざっくり理解する
- ✓ 社内データを"価値あるナレッジ"に変える
ヒントを持ち帰る
- ✓ 自社での活用イメージ・最初の一步をつかむ

ありがとう
ございました！



別途ご案内のダウンロードサービスを利用された方には、資料後半にサンプルコード特典が付いてきます。



Q&A

質疑応答

ご質問をお待ちしております

気軽に
質問して下さい



ダウンロード特典

ダウンロード特典





サンプル動作に必要なもの

プログラミング言語はPythonを使っているのでPython開発環境が必要です。

下記をpip installして下さい。

```
jupyter>=1.1.1
```

```
langchain>=0.3.27
```

```
langchain_community>=0.3.27
```

```
langgraph>=0.6.0
```

```
langchain_openai>=0.3.28
```

```
faiss-cpu>=1.11.0
```

```
pypdf>=5.9.0
```

```
unstructured>=0.18.11
```

OpenAI APIのアカウント(有料)を作成してAPIキーを環境変数OPENAI_API_KEYにセットして下さい。

サンプルはCursor/VSCodeのJupyter拡張またはGoogle Colabで実行させる事を前提としています。

第1章

RAG (Retrieval Augmented Generation)

検索拡張生成の仕組みを理解し、
サンプルコードを使って詳しく解説します





サンプルコード: Embeddingモデルの準備

```
[1] from langchain_openai import OpenAIEmbeddings

# 埋め込みモデルの準備
embeddings = OpenAIEmbeddings(
    model="text-embedding-3-large"
)
✓ 0.7s Python
```

```
[2] # Embeddingする
vector = embeddings.embed_query(
    text="埋め込みベクトルに変換するテキスト"
)
✓ 4.5s Python
```

```
[3] print(vector)
✓ 0.0s Python
```

```
... [0.01910506933927536, -0.02345600351691246, -0.012313704937696457, 0.008422964252531528, 0.028127681463956833, 0.039744120091199875, -0.0372618548572063
```

```
[4] print(len(vector))
✓ 0.0s Python
```

```
... 3072
```



サンプルコード: ベクトルストアの準備

FAISSの準備

- インデックス検索アルゴリズム
 - IndexFlatL2 ... L2ノルム（ユークリッド距離）を使用して検索
 - IndexFlatIP ... 内積を使用して検索
 - IndexIVFFlat ... InvertedFileデータ構造を使用して検索
 - 文書から単語を引くのではなく、単語から文書を逆引きするデータ構造
 - 単語「浅草寺」の転置リスト
 - 文書ID: 5, 位置: [3, 15, 27], 出現回数: 3
 - 文書ID: 12, 位置: [8], 出現回数: 1
 - 文書ID: 23, 位置: [2, 19], 出現回数: 2
 - 「車」と「自動車」は別の単語として処理されてしまう（基本的に完全一致検索）

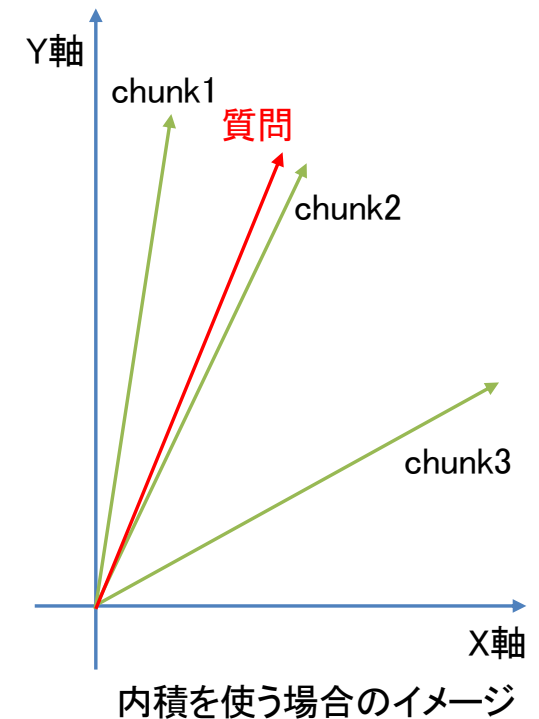
```
import faiss
from langchain_community.docstore.in_memory import InMemoryDocstore
from langchain_community.vectorstores import FAISS

# L2ノルム（ユークリッド距離）検索アルゴリズムのインデックスを準備
index = faiss.IndexFlatL2(
    len(
        embeddings.embed_query(
            text="埋め込みベクトルに変換するテキスト"
        )
    ) # 次元数を指定

# ベクトルストアを準備
vector_store = FAISS(
    embedding_function=embeddings,
    index=index,
    docstore=InMemoryDocstore(),
    index_to_docstore_id={},
)
```

[5] ✓ 1.0s

Python





サンプルコード:ドキュメントの準備(1／3)

```
from langchain_core.documents import Document

# ドキュメントの準備 (日本の観光地情報)
documents = [
    Document(
        page_content="浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",
        metadata={
            "source": "japan-tourism-doc",
            "prefecture": "東京都",
            "category": "歴史・文化",
            "annual_visitors": "30000000"
        },
    ),
    Document(
        page_content="富士山は日本最高峰の山で、世界文化遺産に登録されています。登山シーズンは7-9月で多くの登山者が訪れます。",
        metadata={
            "source": "japan-tourism-doc",
            "prefecture": "静岡県・山梨県",
            "category": "自然・山岳",
            "annual_visitors": "5000000"
        },
    ),
    Document(
        page_content="清水寺は京都を代表する古寺で、桜と紅葉の名所として知られています。清水の舞台からの眺望は絶景です。",
        metadata={
            "source": "japan-tourism-doc",
            "prefecture": "京都府",
            "category": "歴史・文化",
            "annual_visitors": "6000000"
        },
    ),
]
```



サンプルコード:ドキュメントの準備(2／3)

```
,
Document(
    page_content="沖縄美ら海水族館は世界最大級の水槽を持つ水族館で、ジンベエザメやマンタが人気です。",
    metadata={
        "source": "japan-tourism-doc",
        "prefecture": "沖縄県",
        "category": "レジャー・体験",
        "annual_visitors": "4800000"
    },
),
Document(
    page_content="ユニバーサル・スタジオ・ジャパンは大阪にあるテーマパークで、ハリーポッターエリアが特に人気です。",
    metadata={
        "source": "japan-tourism-doc",
        "prefecture": "大阪府",
        "category": "レジャー・体験",
        "annual_visitors": "14500000"
    },
),
]
```

[6] ✓ 0.0s

Python



サンプルコード:ドキュメントの準備(3／3)

```
for document in documents:
    print(document.model_dump_json(indent=2))
```

[7] ✓ 0.0s Python

```
... {
  "id": null,
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "東京都",
    "category": "歴史・文化",
    "annual_visitors": "30000000"
  },
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",
  "type": "Document"
}
{
  "id": null,
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "静岡県・山梨県",
    "category": "自然・山岳",
    "annual_visitors": "5000000"
  },
  "page_content": "富士山は日本最高峰の山で、世界文化遺産に登録されています。登山シーズンは7-9月で多くの登山者が訪れます。",
  "type": "Document"
}
{
  "id": null,
  "metadata": {
    ...
  },
  "page_content": "ユニバーサル・スタジオ・ジャパンは大阪にあるテーマパークで、ハリーポッターエリアが特に人気です。",
  "type": "Document"
}
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...



サンプルコード:ドキュメントをベクトルストアに追加(1／3)

```
from uuid import uuid4

# オプションのidを生成
uuids = [str(uuid4()) for _ in range(len(documents))]
print(uuids)
```

[24] ✓ 0.0s

Python

... ['5116e509-e0ff-47e7-b677-510786a95d13', 'e117283e-9825-44e2-969d-2e9577ade99f', 'b4e459e1-61c8-4927-9725-b5686382e7cb', 'b20ace53-eb52-427d-9559-548259

```
# ベクトルストアにドキュメントを追加 (id付き)
vector_store.add_documents(
    documents=documents,
    ids=uuids
)
```

[25] ✓ 1.4s

Python

... ['5116e509-e0ff-47e7-b677-510786a95d13',
'e117283e-9825-44e2-969d-2e9577ade99f',
'b4e459e1-61c8-4927-9725-b5686382e7cb',
'b20ace53-eb52-427d-9559-548259a5df44',
'059e25c2-f12b-46d6-8616-1b7ce5047cba']



サンプルコード:ドキュメントをベクトルストアに追加(2/3)

```
_documents = vector_store.get_by_ids([uuids[0], uuids[1]])  
for document in _documents:  
    print(document.model_dump_json(indent=2))
```

[27] ✓ 0.0s

Python

```
... {  
  "id": "5116e509-e0ff-47e7-b677-510786a95d13",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "東京都",  
    "category": "歴史・文化",  
    "annual_visitors": "30000000"  
  },  
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",  
  "type": "Document"  
}  
{  
  "id": "e117283e-9825-44e2-969d-2e9577ade99f",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "静岡県・山梨県",  
    "category": "自然・山岳",  
    "annual_visitors": "5000000"  
  },  
  "page_content": "富士山は日本最高峰の山で、世界文化遺産に登録されています。登山シーズンは7-9月で多くの登山者が訪れます。",  
  "type": "Document"  
}
```

サンプルコード:ドキュメントをベクトルストアに追加(3/3)

```
def get_vectors_by_ids(vector_store, ids):
    """IDのリストからベクトルのリストを取得"""
    vectors = []
    idx_to_id = {v: k for k, v in vector_store.index_to_docstore_id.items()}

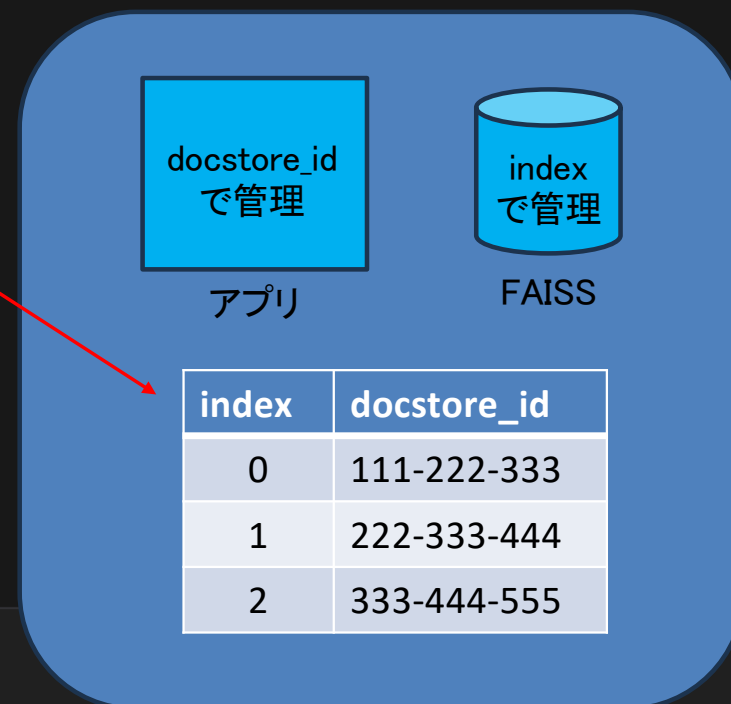
    for target_id in ids:
        if target_id in idx_to_id:
            faiss_idx = idx_to_id[target_id]
            vector = vector_store.index.reconstruct(faiss_idx)
            vectors.append(vector)
        else:
            vectors.append(None)

    return vectors

vectors = get_vectors_by_ids(vector_store, [uuids[0], uuids[1]])
for vector in vectors:
    print(f"取得したベクトル: {vector}")
```

[37] ✓ 0.0s

... 取得したベクトル: [0.03528976 0.00540935 -0.00297535 ... 0.00642017 0.00369646
0.00605644]
取得したベクトル: [0.00855165 0.00792902 -0.00979268 ... 0.00504882 -0.0010176
-0.00350071]



Python

第2章

ベクトルストアで直接検索

文字列クエリとの類似度で検索（同期・非同期・スコア付き）や
埋め込みクエリとの類似度検索を実践します





文字列クエリとの類似度で検索（同期）

```
results = vector_store.similarity_search(  
    query="歴史的な建物がある観光地",  
)
```

[12] ✓ 0.5s

Python

```
for result in results:  
    print(result.model_dump_json(indent=2))
```

[13] ✓ 0.0s

Python

```
... {  
  "id": "da04c853-ecb0-48d9-b38a-719fc32d02c9",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "東京都",  
    "category": "歴史・文化",  
    "annual_visitors": "30000000"  
  },  
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",  
  "type": "Document"  
}  
{  
  "id": "3248a3fb-42af-49a6-8429-4477e107e04d",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "京都府",  
    "category": "歴史・文化",  
    "annual_visitors": "60000000"  
  },  
  "page_content": "清水寺は京都を代表する古寺で、桜と紅葉の名所として知られています。清水の舞台からの眺望は絶景です。",  
  "type": "Document"  
}  
{  
  "id": "6a875242-e6d6-4815-90ee-faf53b850991",  
  "metadata": {  
    ...  
  },  
  "page_content": "富士山は日本最高峰の山で、世界文化遺産に登録されています。登山シーズンは7-9月で多くの登山者が訪れます。",  
  "type": "Document"  
}
```



文字列クエリとの類似度で検索（非同期）

```
[14] import nest_asyncio
      nest_asyncio.apply() # イベントループのネストを許可

      results = await vector_store.asimilarity_search(
      |   query="自然を楽しめる観光地",
      )
      ✓ 1.3s Python

[15] for result in results:
      |   print(result.model_dump_json(indent=2))
      ✓ 0.0s Python

... {
  "id": "3248a3fb-42af-49a6-8429-4477e107e04d",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "京都府",
    "category": "歴史・文化",
    "annual_visitors": "6000000"
  },
  "page_content": "清水寺は京都を代表する古寺で、桜と紅葉の名所として知られています。清水の舞台からの眺望は絶景です。",
  "type": "Document"
}
{
  "id": "c77a6e72-97be-4bbc-b5cf-bc94ac6bb601",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "沖縄県",
    "category": "レジャー・体験",
    "annual_visitors": "4800000"
  },
  "page_content": "沖縄美ら海水族館は世界最大級の水槽を持つ水族館で、ジンベエザメやマンタが人気です。",
  "type": "Document"
}
{
  "id": "6a875242-e6d6-4815-90ee-faf53b850991",
  "metadata": {
    ...
  },
  "page_content": "富士山は日本最高峰の山で、世界文化遺産に登録されています。登山シーズンは7-9月で多くの登山者が訪れます。",
  "type": "Document"
}
```



文字列クエリとの類似度で検索（類似度スコア付き）

```
[16] ✓ 0.8s Python

results = vector_store.similarity_search_with_score(
    query="水族館",
)

# スコアが小さいほど類似度が高い
for doc, score in results:
    print(doc.model_dump_json(indent=2))
    print(f"スコア: {score}")
    print("-" * 10)

[17] ✓ 0.0s Python

... {
  "id": "c77a6e72-97be-4bbc-b5cf-bc94ac6bb601",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "沖縄県",
    "category": "レジャー・体験",
    "annual_visitors": "4800000"
  },
  "page_content": "沖縄美ら海水族館は世界最大級の水槽を持つ水族館で、ジンベエザメやマンタが人気です。",
  "type": "Document"
}
スコア: 0.9978859424591064
-----
{
  "id": "6a875242-e6d6-4815-90ee-faf53b850991",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "大阪府",
    "category": "レジャー・体験",
    "annual_visitors": "14500000"
  },
  "page_content": "ユニバーサル・スタジオ・ジャパンは大阪にあるテーマパークで、ハリーポッターエリアが特に人気です。",
  "type": "Document"
}
スコア: 1.5677359104156494
...
  "type": "Document"
}
スコア: 1.6367518901824951
-----
```



埋め込みクエリとの類似度で検索(1 / 2)

```
# 埋め込みを作成
vector = embeddings.embed_query(
    text="寺院"
)
print(vector)
```

[19] ✓ 0.8s

Python

... [0.010530274361371994, 0.008984951302409172, -0.013709225691854954, 0.007675105705857277, 0.0024798759259283543, -0.013341291807591915, -0.0140698011964

```
results = vector_store.similarity_search_by_vector(
    embedding=vector,
)
```

[20] ✓ 0.0s

Python



埋め込みクエリとの類似度で検索(2/2)

```
[21] ✓ 0.0s Python
... {
  "id": "3248a3fb-42af-49a6-8429-4477e107e04d",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "京都府",
    "category": "歴史・文化",
    "annual_visitors": "6000000"
  },
  "page_content": "清水寺は京都を代表する古寺で、桜と紅葉の名所として知られています。清水の舞台からの眺望は絶景です。",
  "type": "Document"
}
{
  "id": "da04c853-ecb0-48d9-b38a-719fc32d02c9",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "東京都",
    "category": "歴史・文化",
    "annual_visitors": "3000000"
  },
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",
  "type": "Document"
}
{
  "id": "c77a6e72-97be-4bbc-b5cf-bc94ac6bb601",
  "metadata": {
    ...
  },
  "page_content": "ユニバーサル・スタジオ・ジャパンは大阪にあるテーマパークで、ハリーポッターエリアが特に人気です。",
  "type": "Document"
}
```

第3章

Retrieverを使って検索

ベクトルストアからRetrieverを作成し、
フィルタリング付きの検索機能を実装します





ベクトルストアをRetrieverに変換してクエリーする場合

```
# retrieverを準備
retriever = vector_store.as_retriever(
    search_type="similarity",
    search_kwargs={
        "k": 3
    },
)
```

[22] ✓ 0.0s

Python



Retrieverを使った検索(基本)

```
[23] results = retriever.invoke(  
      | input="テーマパーク"  
      )  
✓ 1.9s Python
```

```
[24] for result in results:  
      | print(result.model_dump_json(indent=2))  
✓ 0.0s Python
```

```
... {  
  "id": "6a875242-e6d6-4815-90ee-faf53b850991",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "大阪府",  
    "category": "レジャー・体験",  
    "annual_visitors": "14500000"  
  },  
  "page_content": "ユニバーサル・スタジオ・ジャパンは大阪にあるテーマパークで、ハリーポッターエリアが特に人気です。",  
  "type": "Document"  
}  
{  
  "id": "c77a6e72-97be-4bbc-b5cf-bc94ac6bb601",  
  "metadata": {  
    "source": "japan-tourism-doc",  
    "prefecture": "沖縄県",  
    "category": "レジャー・体験",  
    "annual_visitors": "4800000"  
  },  
  "page_content": "沖縄美ら海水族館は世界最大級の水槽を持つ水族館で、ジンベエザメやマンタが人気です。",  
  "type": "Document"  
}  
{  
  "id": "da04c853-ecb0-48d9-b38a-719fc32d02c9",  
  "metadata": {  
    ...  
  },  
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",  
  "type": "Document"  
}
```



Retrieverを使った検索(フィルタリング付き)

```
# フィルタリング付きで検索
results = retriever.invoke(
    input="人気の観光地",
    filter = {
        "category": "歴史・文化"
    },
)
```

[25] ✓ 0.9s

Python

```
for result in results:
    print(result.model_dump_json(indent=2))
```

[26] ✓ 0.0s

Python

```
... {
  "id": "da04c853-ecb0-48d9-b38a-719fc32d02c9",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "東京都",
    "category": "歴史・文化",
    "annual_visitors": "30000000"
  },
  "page_content": "浅草寺は東京最古の寺院で、雷門と仲見世通りで有名な観光名所です。年間約3000万人の観光客が訪れます。",
  "type": "Document"
}
{
  "id": "3248a3fb-42af-49a6-8429-4477e107e04d",
  "metadata": {
    "source": "japan-tourism-doc",
    "prefecture": "京都府",
    "category": "歴史・文化",
    "annual_visitors": "6000000"
  },
  "page_content": "清水寺は京都を代表する古寺で、桜と紅葉の名所として知られています。清水の舞台からの眺望は絶景です。",
  "type": "Document"
}
```

第4章

RAGを使ったチャットボット

簡単なサンプルコードを使って
RAGチャットボットの実装方法を解説します





RAGを使ったチャットボット(1／3)

```
from langchain.chat_models import init_chat_model
```

```
# ChatModelの準備
```

```
chat_model = init_chat_model(  
    model_provider="openai",  
    model="gpt-4.1",  
    temperature=0.9,  
)
```

[27] ✓ 0.0s

Python

```
from langchain_core.prompts import ChatPromptTemplate
```

```
# PromptTemplateの準備
```

```
message = """提供されたコンテキストのみを使用して、この質問に答えて下さい。  
{question}
```

```
Context:
```

```
{context}
```

```
"""
```

```
prompt_template = ChatPromptTemplate.from_messages(  
    messages=[
```

```
        ("human", message)
```

```
    ]
```

```
)
```

[28] ✓ 0.0s

Python

RAGを使ったチャットボット(2/3)

```
from langchain_core.runnables import RunnablePassthrough

# chainの準備
chain = (
    {
        "context": retriever,
        "question": RunnablePassthrough()
    }
    | prompt_template
    | chat_model
)
```

[29] ✓ 0.0s

Python



```
message = """提供されたコンテキストのみを使用して、この質問に答えて下さい。
{question}

Context:
{context}
"""
```



RAGを使ったチャットボット(3／3)

```
output = chain.invoke(  
    input="歴史・文化カテゴリの観光地について教えてください。"  
)
```

[30] ✓ 2.1s

Python

```
print(output.content)
```

[31] ✓ 0.0s

Python

... 歴史・文化カテゴリの観光地として、以下の2つが挙げられます。

1. 浅草寺（東京都）

東京最古の寺院で、雷門や仲見世通りで有名な観光名所です。年間約3,000万人の観光客が訪れます。

2. 清水寺（京都府）

京都を代表する古寺で、桜や紅葉の名所として知られています。特に「清水の舞台」からの眺望が絶景です。年間約600万人が訪れます。