README.md 2024-03-22

DockerによるWebアプリのビルドとデプロイ

このドキュメントでは、Dockerを使ってWebアプリケーションのRAGデータベース(SQLデータベースとベクトルデータベース)を構築し、 そのアプリケーションをDockerホスト上にデプロイする方法を説明します。

RAGデータベースの構築には、

- 1. 既に構築済みのデータベースファイルを利用する方法と、
- 2. Excelファイルから新たにデータベースを構築する方法があります。

既存のExcelファイル過去プロジェクト収集データ(サンプル)。xlsxをそのまま用いる場合、方法1で十分です。 Excelファイルを更新したい場合には方法2を選択します。

目次

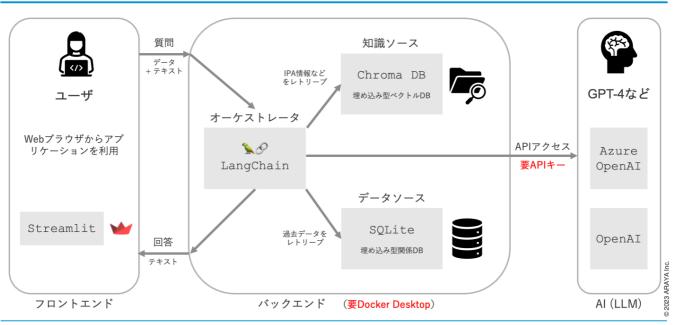
- 全体図
- 前提条件
- RAGデータベースの構築
- Webアプリケーションのデプロイ

全体図

このアプリケーションの全体図は次の通りです:

RAGアプリケーションの全体図





また, 上の全体図における各要素とソースコードの対応関係は概ね次の通りです:

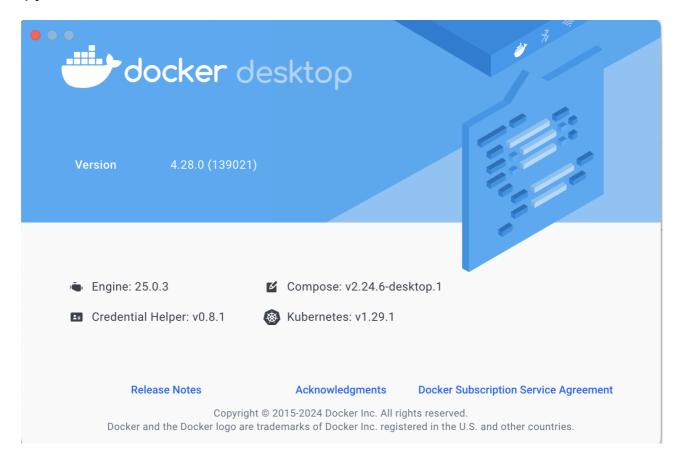
アプリの要素	ソースコード
フロントエンド	app/,src/app_util/

README.md 2024-03-22

アプリの要素	ソースコード
オーケストレータ	src/orchestrator/
知識ソース	src/rag_tabular/
データソース	src/rag_textual/
APIアクセス	util/api.py

前提条件

1. Docker Desktop (https://www.docker.com/products/docker-desktop/) がインストールされているとします。



- 2. Linuxコンテナを使用するので、DockerホストのOSがWindowsの場合は、WSL 2 バックエンドが有効化されているとします。参考: Dockerマニュアル (https://docs.docker.com/desktop/wsl/)
- 3. Azure OpenAI関係のセットアップが完了しているとします。具体的には次の条件が満たされているとします:
 - o Azure OpenAIのリソースが作成されており、APIキーとエンドポイントが取得できること。
 - そのリソースにgpt-4モデルとtext-embedding-3-smallモデルがデプロイされていること。更に、それらのモデル名とデプロイ名が同一であること。

なお、Microsoft Entra IDによるAzure OpenAl APIの認証方法には現状対応していません。

RAGデータベースの構築

1. Docker Volumeの作成

README.md 2024-03-22

```
docker volume create rag-tab-data docker volume create rag-txt-data
```

2. RAGデータベースの構築方法1: 既存のデータベースファイルを利用する

dockerディレクトリに移動し、次のコマンドを実行すると、 既存のデータベースファイルを上で作成した Dockerボリュームにコピーします。

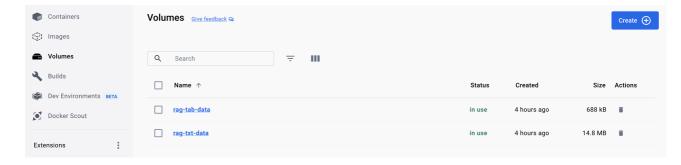
```
docker compose -f db-build-1.compose.yaml up
docker compose -f db-build-1.compose.yaml down
```

3. RAGデータベースの構築方法2: データベースを新たに構築する

まず、新たに使用するExcelファイルを用意し、プロジェクトのルートディレクトリ内に配置します。 次に、そのExcelファイルのプロジェクトルートから見た相対パスを環境変数RAG_TAB_PATHに設定します。 そのような準備をした上で次のコマンドを実行すると、データベースが新たに構築されます。

```
docker compose -f db-build-2.compose.yaml up docker compose -f db-build-2.compose.yaml down
```

4. RAGデータベースが無事に構築できたことをDocker Desktopで確認



Webアプリケーションのデプロイ

dockerディレクトリに移動し、次のコマンドを実行すると、WebアプリがDockerホスト上に立ち上がります。

```
docker compose up
```

Dockerホストからは、Webブラウザでhttp://localhost:8501にアクセスすることで、アプリケーションを利用できます。 Dockerホストと同じネットワークにある他の端末からは、DockerホストのIPアドレスを指定してアクセスすることができるはずです。