

项目编号：63

项目标题：利用 RAG（RetrievalAugmented Generation，检索增强生成）的 SAT 决策抓取和搜索应用程序

项目客户：Josh Tedjasaputra（客户体验洞察私人有限公司）

项目专长：软件开发；网络应用程序开发；人工智能（机器/深度学习、NLP）；人机交互（HCI）；物联网（IoTs）；云计算；大数据分析与可视化；计算机科学与算法；移动应用程序开发；

组数：4 组

主要联系人：Josh Tedjasaputra (Customer Experience Insight Pty Ltd)

背景介绍

本项目旨在开发一个代理 Python 应用程序，该程序可自动搜索、处理 SAT（国家行政法庭）裁决并编制索引。该应用程序应允许通过人工智能语言模型使用检索增强生成（RAG）进行自然语言查询，以提供与上下文相关的结果、摘要和见解。

此外，该系统必须允许授权用户手动上传新的 SAT 裁决文件，这些文件将被处理、嵌入并可立即搜索。主要目标是利用人工智能驱动的洞察力，实现对法庭裁决的高效搜索和分析（西澳大利亚州的法律民主化）

要求和范围：

- 为 SAT 决策开发基于网络的搜索和检索系统。
- 使用自动网络搜索器提取法庭裁决、元数据和摘要。
- 利用人工智能模型整合检索-增强生成（RAG）功能，生成符合上下文的决策摘要。
- 为自然语言查询和决策检索提供安全的用户界面。
- 通过实时索引和搜索功能实现手动添加文档。
- 实施错误处理、日志记录和健全的文档，以提高可维护性。

所需知识和技能

数据采集与处理：

- 抓取并存储带有元数据（日期、案件标题、涉及方、摘要）的 SAT 决定。

- 处理结构化和非结构化数据清理（删除重复数据、规范化）。
- 将经过处理的决策存储在可搜索的矢量数据库中（如 FAISS、Pinecone）。

搜索和检索界面：

- 基于网络的用户界面（Flask/FastAPI）允许自然语言查询（如 "显示有关环境法规的决定"）。
- 查询结果应包括决策链接、元数据和人工智能模型生成的摘要。

RAG 与人工智能集成：

- 使用人工智能模型嵌入进行基于向量的搜索和响应生成。
- 对案件细节和类似决定实施上下文感知检索。

手动添加文件：

- 允许通过身份验证的用户上传新的决定（PDF/DOCX）。
- 处理上传的文件，生成嵌入文件，并使其可即时搜索。

系统可扩展性和安全性

- 系统应在基于 Linux 的环境中运行，并配备安全的 API 端点。
- 用于手动上传文件和管理控制的用户认证。

预期成果/交付成果：

可进行 SAT 决策搜索和检索的功能性网络应用程序。

自动抓取器可收集 SAT 决定并以结构化格式存储。

使用 RAG 和 Llama 嵌入的人工智能搜索引擎。

具有实时索引功能的手动文件添加功能。

全面的文档（README、API 说明、用户指南）。

错误处理和日志记录框架，提高系统的可维护性。

