**RAG流程**

RAG是一种结合信息检索（Retrieval）和生成式语言模型（Generation）的框架，旨在通过从外部知识库中检索相关信息来增强语言模型的回答质量。

1. **文档预处理（Document Preprocessing）**：

**输入**：原始文档（例如PDF、TXT等）。

**处理**： 提取文档中的文本内容（对于PDF，可能需要OCR处理图像型内容）将长文本分割成较小的块（chunks），以便于嵌入生成和检索。

**目的**：将原始文档转化为适合后续处理的结构化文本片段。

1. **嵌入生成（Embedding Generation）**：

**输入**：分割后的文本块。

**处理**： 使用嵌入模型（通常是基于Transformer的句嵌入模型）将每个文本块编码为高维向量。这些向量捕获文本的语义信息，用于后续的相似性检索。

**目的**：将文本转换为机器可理解的数值表示。

1. **向量存储（Vector Storage）：**

**输入**：文本块及其对应的嵌入向量。

**处理**： 将嵌入向量存储到向量数据库（如FAISS、Pinecone）中，构建索引以支持高效的相似性搜索。

**目的**：创建一个可快速查询的知识库。

1. **查询检索（Retrieval）：**

**输入**：用户的问题（query）。

**处理**： 将用户问题编码为嵌入向量。在向量数据库中执行相似性搜索（通常使用余弦相似度或欧氏距离），检索与问题最相关的k个文本块。

**目的**：从知识库中提取与问题相关的上下文。

1. **生成回答（Generation）：**

**输入**：检索到的文本块（上下文）和用户问题。

**处理**： 将检索到的上下文和问题组合成提示（prompt）。输入到不同的生成式语言模型（LLM）的API接口，生成多模态的回答。

**目的**：基于检索到的信息生成准确、流畅的回答。

**同义词筛选：**

超声相关：

“超声” 与 “超声波”。

“超声检查” 与 “超声诊断”“超声扫描”。

“超声探头” 与 “超声换能器”。

人体组织与器官：

“肝脏” 与 “肝部”。

“子宫” 与 “女子胞”“胞宫”。

“甲状腺” 与 “甲腺”。

疾病与异常：

“囊肿” 与 “囊性肿物”。

“脂肪肝” 与 “肝脂肪变性”。

超声技术与指标：

“B 超”与 “B 型超声” 与 “二维超声”。

“彩超” 与 “彩色多普勒超声”与“彩色超声”。

“M 型超声” 与 “M 超”。

“血流速度” 与 “血流速率”。

“回声”与 “反射波”。

        "ChaoShengWuLiJiChuBiDu(OCR).pdf",

        "ChaoShengXinDongTuXue(OCR).pdf",

        "ShiYongChaoShengZhenDuanXue(OCR).pdf",

        "ChaoShengZhenDuanLinChuang(OCR).pdf",

        "ChaoShengZhenDuanXue(OCR).pdf",