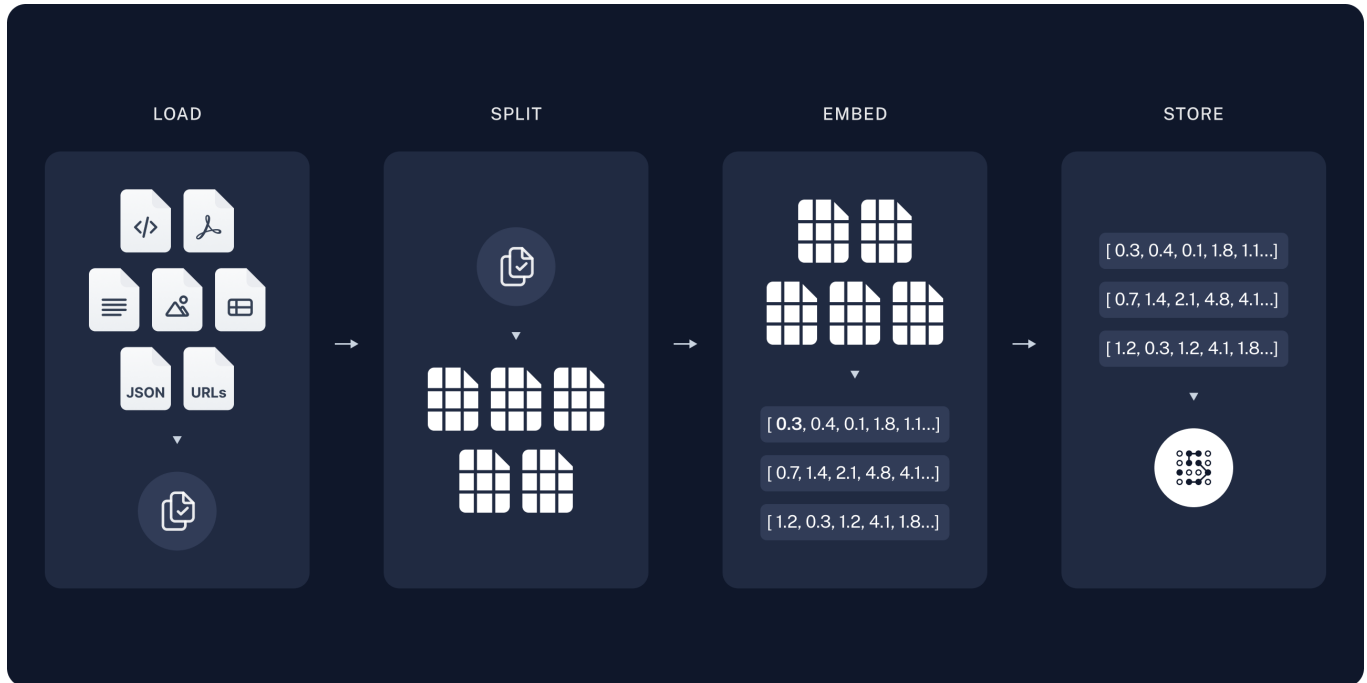


벡터스토어(Vector Store) 저장



벡터스토어 저장 단계는 Retrieval-Augmented Generation(RAG) 시스템의 네 번째 단계로, 이전 단계에서 생성된 임베딩 벡터들을 효율적으로 저장하고 관리하는 과정입니다. 이 단계는 향후 검색 과정에서 벡터들을 빠르게 조회하고, 관련 문서를 신속하게 찾아내는 데 필수적입니다.

벡터스토어 저장의 필요성

1. 빠른 검색 속도: 임베딩 벡터들을 효과적으로 저장하고 색인화함으로써, 대량의 데이터 중에서도 관련된 정보를 빠르게 검색할 수 있습니다.
2. 스케일러빌리티: 데이터가 지속적으로 증가함에 따라, 벡터스토어는 이를 수용할 수 있는 충분한 스케일러빌리티를 제공해야 합니다. 효율적인 저장 구조는 데이터베이스의 확장성을 보장하며, 시스템의 성능 저하 없이 대규모 데이터를 관리할 수 있도록 합니다.

3. **의미 검색(Semantic Search)** 지원: 키워드 기반 검색이 아닌 사용자의 질문과 의미상으로 유사한 단락을 조회해야 하는데, 벡터스토어는 이러한 기능을 지원합니다. 텍스트 자체가 저장되는 DB의 경우 키워드 기반 검색에 의존해야 하는 한계성이 있지만, 벡터스토어는 의미적으로 유사한 단락 검색을 가능케합니다.

(예시)

질문: "모바일 디바이스 상에서 동작하는 인공지능 기술을 소개한 기업명은?"

- 삼성전자는 삼성 가우스를 활용한 온디바이스 AI 기술도 소개했으며, 생성 AI 모델을 다양한 제품에 단계적으로 탑재할 계획

벡터스토어 중요성

벡터스토어 저장 단계는 **RAG 시스템의 검색 기능과 직접적으로 연결되어 있으며**, 전체 시스템의 **응답 시간과 정확성에 큰 영향**을 미칩니다. 이 단계를 통해 데이터가 잘 관리되고, 필요할 때 즉시 접근할 수 있도록 함으로써, 사용자에게 신속하고 정확한 정보를 제공할 수 있습니다.

코드

```
from langchain_community.vectorstores import FAISS

# 단계 4: DB 생성(Create DB) 및 저장
# 벡터스토어를 생성합니다.
vectorstore = FAISS.from_documents(documents=documents,
embedding=embeddings)
```

참고

- [벡터스토어 사용방법](#)

- LangChain VectorStores