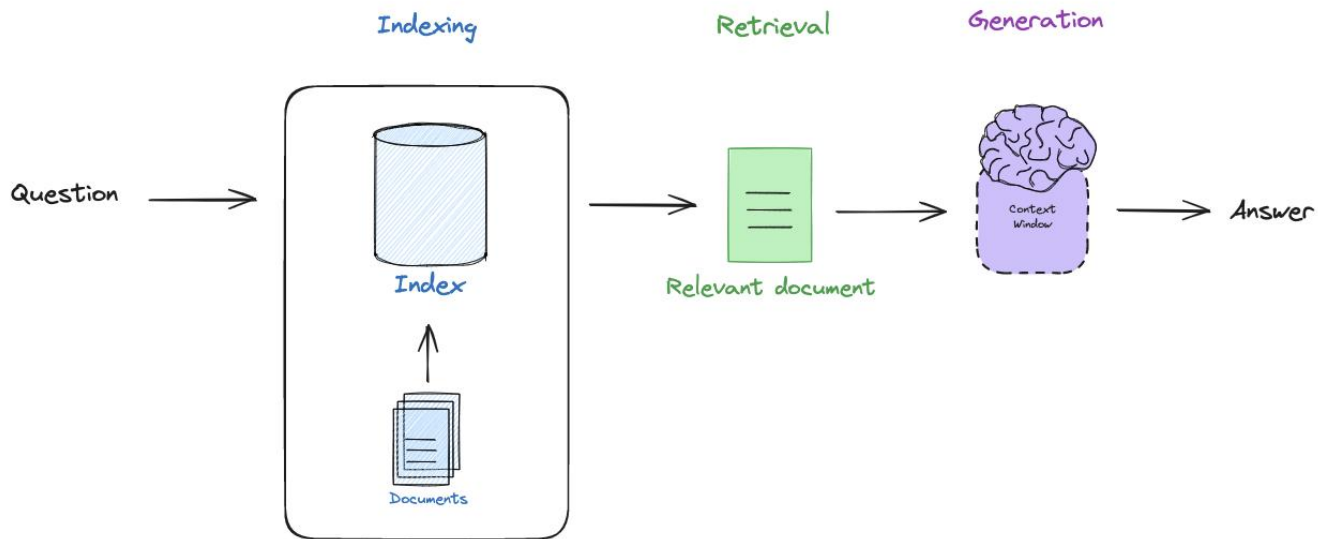# RAG from scratch: Indexing

Lance Martin
Software Engineer, LangChain
@RLanceMartin

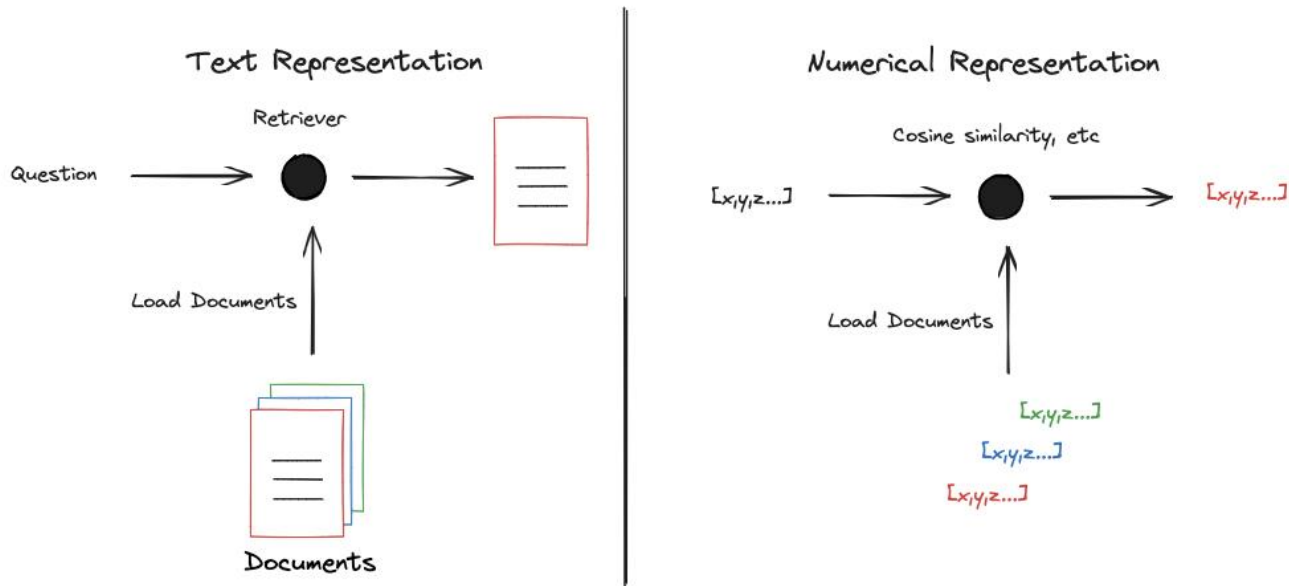**LangChain**

RAG motivation

Indexing   Retrieval   Generation

Question

Index

Documents

Relevant document

Context Window

Answer

Document loading
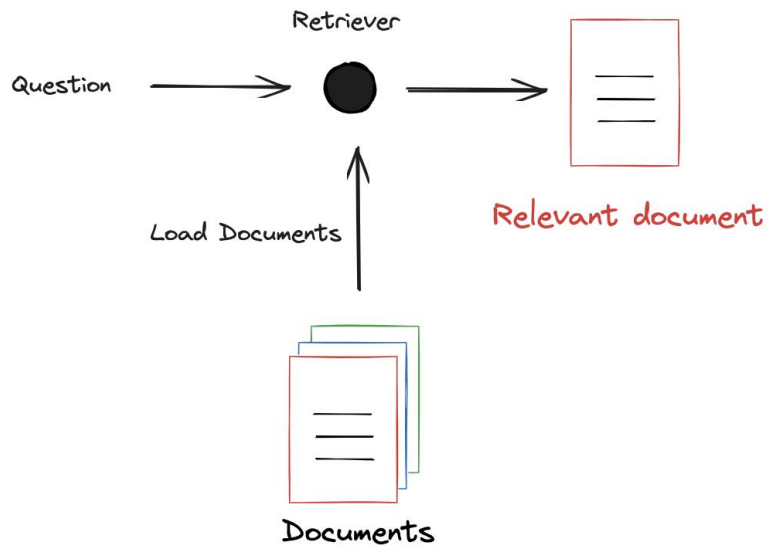
Numerical representation for search

Statistical and machine learned representations
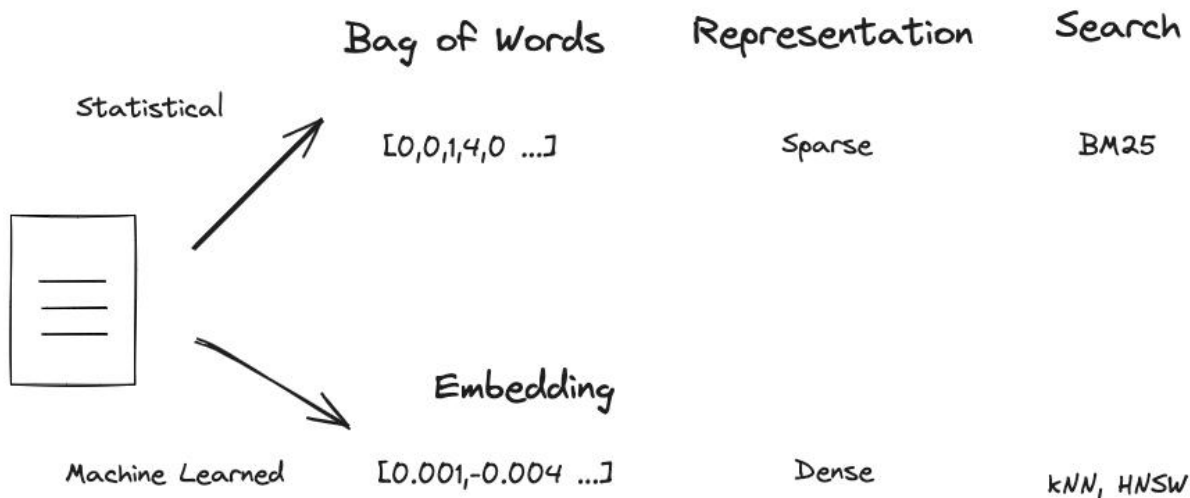
Bag of Words   Representation   Search

Statistical

[0,0,1,4,0 ...]   Sparse   BM25

Embedding

Machine Learned   [0.001,-0.004 ...]   Dense   kNN, HNSW

Loading, splitting, and embedding

🦜🔗 LangChain

Code walk-through