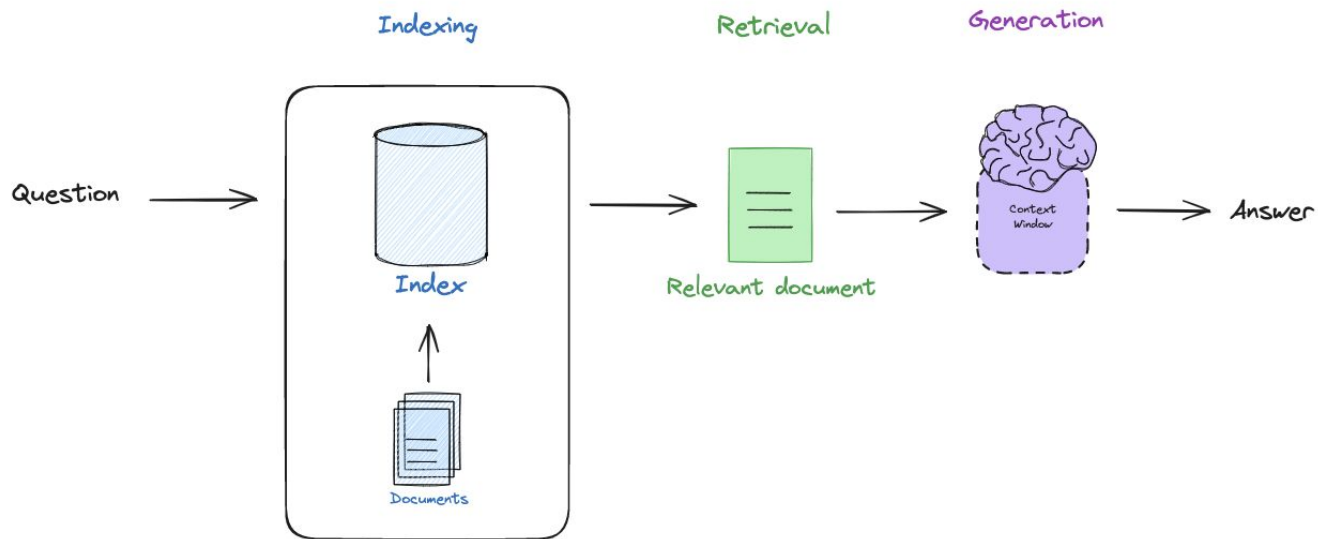


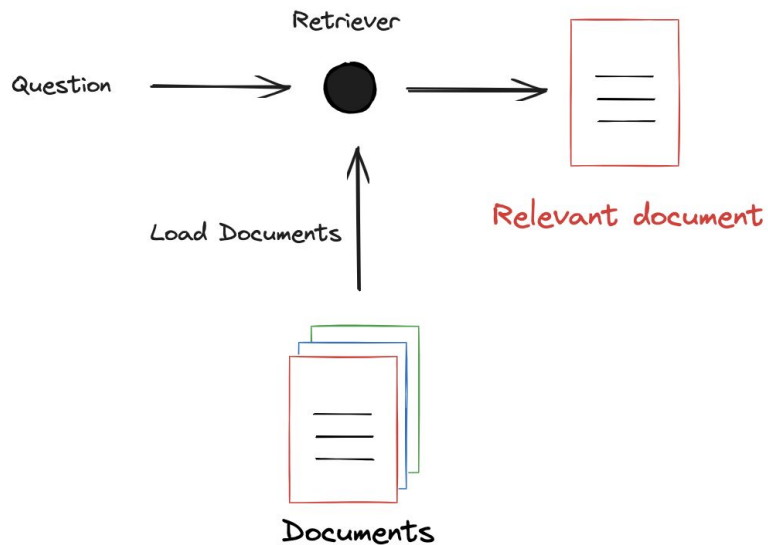
# RAG from scratch: Indexing

Lance Martin  
Software Engineer, LangChain  
[@RLanceMartin](https://twitter.com/RLanceMartin)

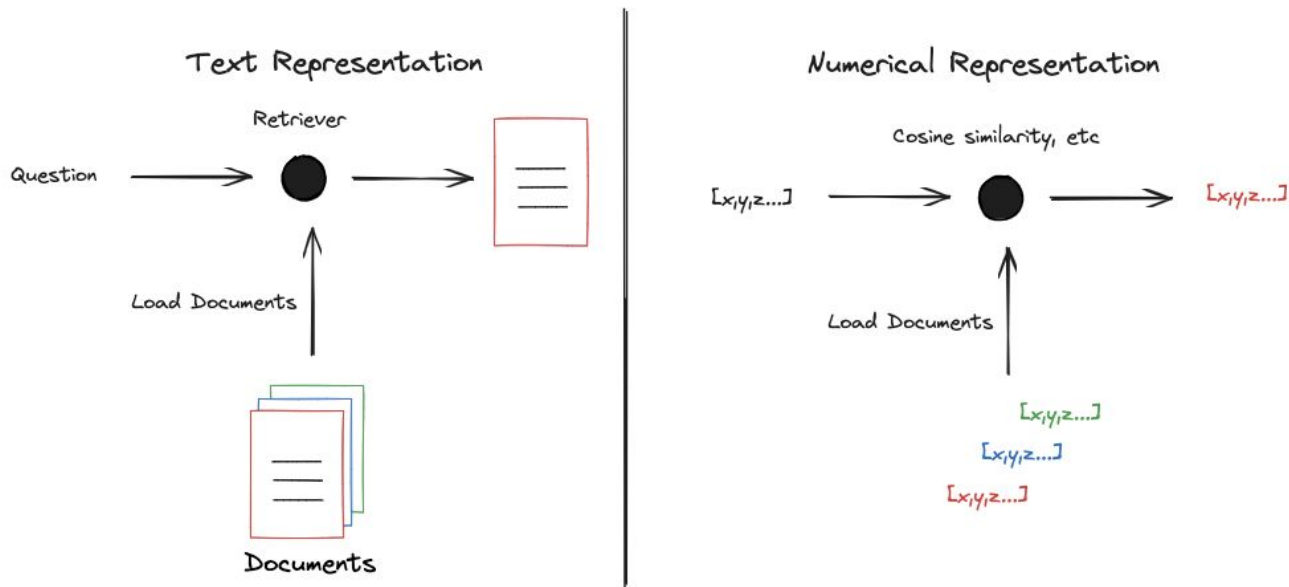
## RAG motivation



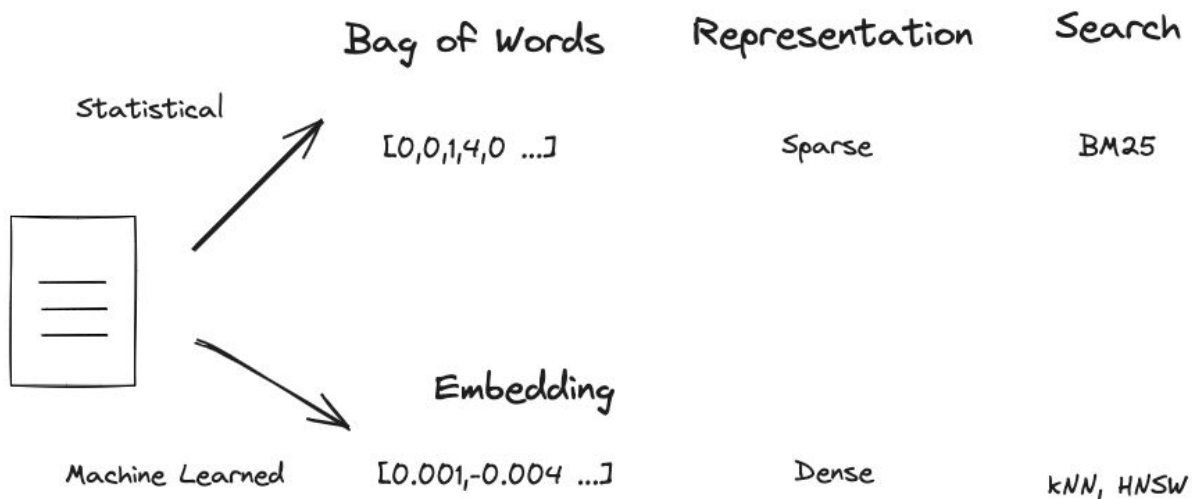
## Document loading



## Numerical representation for search



# Statistical and machine learned representations

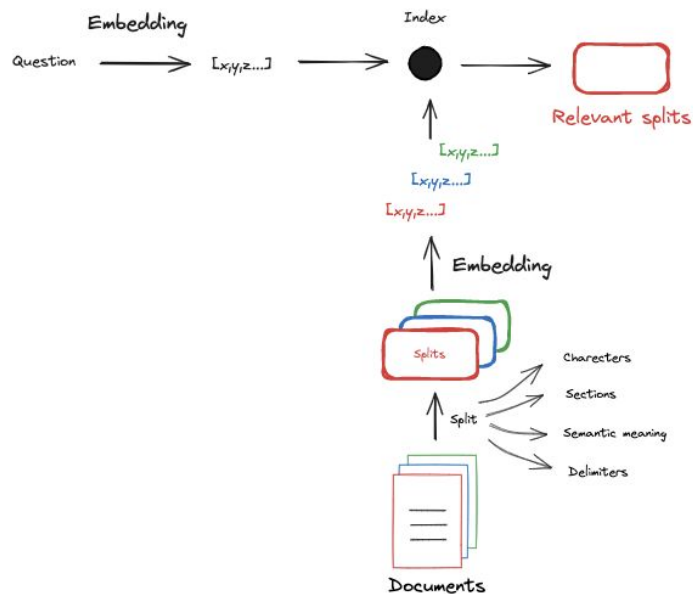


<https://simonwillison.net/2023/Oct/23/embeddings/>

<https://www.pinecone.io/learn/series/nlp/dense-vector-embeddings-nlp/>

[https://cameronrwolfe.substack.com/p/the-basics-of-ai-powered-vector-search?utm\\_source=profile&utm\\_medium=reader2](https://cameronrwolfe.substack.com/p/the-basics-of-ai-powered-vector-search?utm_source=profile&utm_medium=reader2)

## Loading, splitting, and embedding



## Code walk-through