1. Core and Foundational Retrievers

This section covers the fundamental building blocks of RAG systems, which are the basic methods for retrieving documents based on a query. These retrievers are essential for establishing a baseline for information retrieval.

Retriever Type	How It Works	Best For
Vector-Store Retriever	Retrieves documents from a vector database by finding semantically similar embeddings. Can use **similarity search** or **MMR** to balance relevance and diversity.	General-purpose semantic search.
BM25 Retriever	A keyword-based method that ranks documents on exact keyword matches. It uses a statistical approach to improve upon basic keyword search by accounting for term frequency and document length.	Hybrid search scenarios where exact term matching is important.

2. Advanced and Specialized Retrievers

These retrievers are designed to handle more complex queries and document structures. They are used to enhance the quality of the retrieved context by going beyond a single search.

Retriever Type	How It Works	Key Benefit
Multi-Query	Uses a Large Language Model (LLM) to generate several different versions of a single user query. Retrieves results for each version to get a richer document set.	More comprehensive results by exploring different query perspectives.
Parent Document (Auto Merging)	Breaks documents into a hierarchy of small and large chunks. It searches on the	Preserves context in long documents for better summarization.

	small chunks but retrieves the larger parent chunks to provide more context.	
Recursive	Follows explicit relationships or links (like citations or metadata) between documents to retrieve a complete set of related information.	Navigating interconnected document structures or knowledge bases.
Document Summary	Generates summaries for each document. The search is performed on these summaries, and only the full documents of the most relevant summaries are retrieved.	Efficiently handling large, diverse document sets.

3. Query Fusion and Scoring Strategies

This final topic focuses on the advanced technique of combining search results from multiple queries to produce a single, highly-ranked list. The **Query Fusion Retriever** is the component that handles this process.

Fusion Strategy	How It Works	Key Advantage
Reciprocal Rank Fusion (RRF)	Combines results based only on their **rank**. It is completely independent of the raw scores.	Highly robust and stable; it is not sensitive to differences in scoring scales or outliers.
Relative Score Fusion (RSF)	Normalizes each query's scores by dividing them by the maximum score in that query's result set.	Preserves the relative **magnitude** of confidence from each retriever.
Distribution-Based Fusion (DSF)	The most advanced method. It analyzes the statistical distribution of scores for each query (e.g., mean and standard deviation) and normalizes them using a z-score.	The most statistically principled and robust method for handling noisy and varied score distributions.