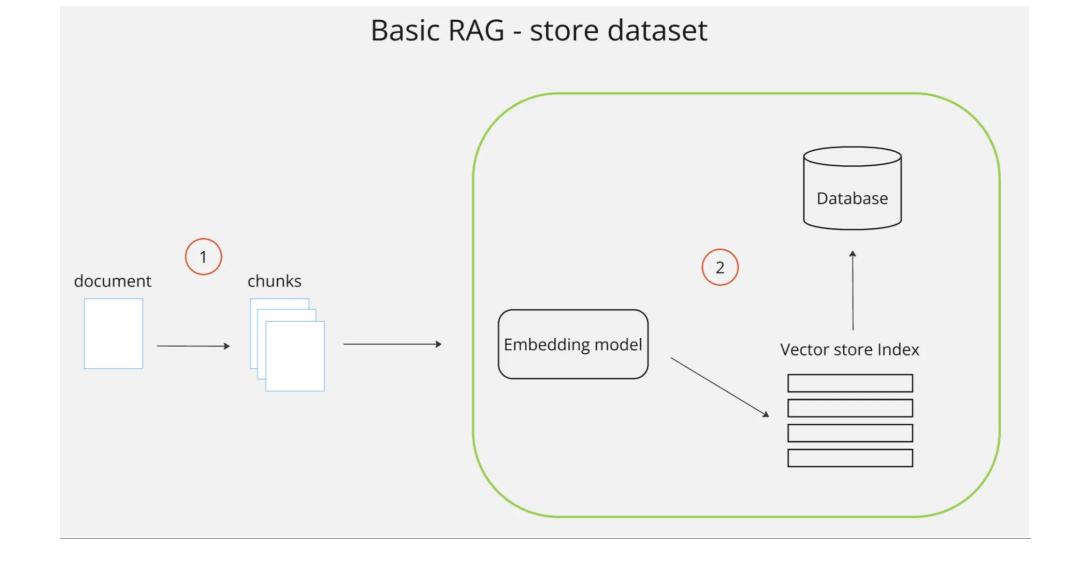
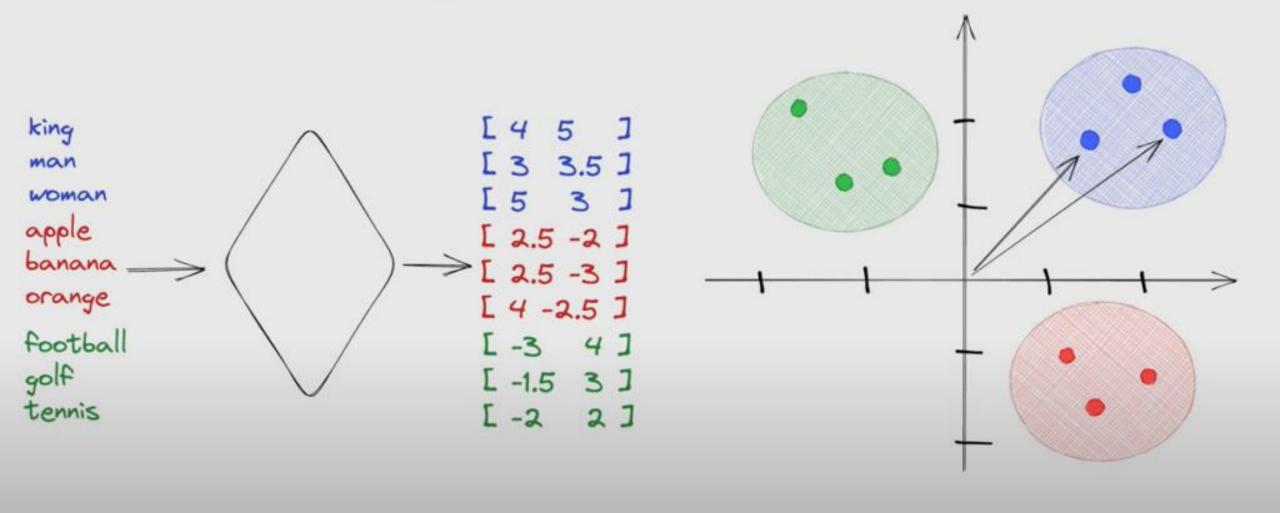
RAG

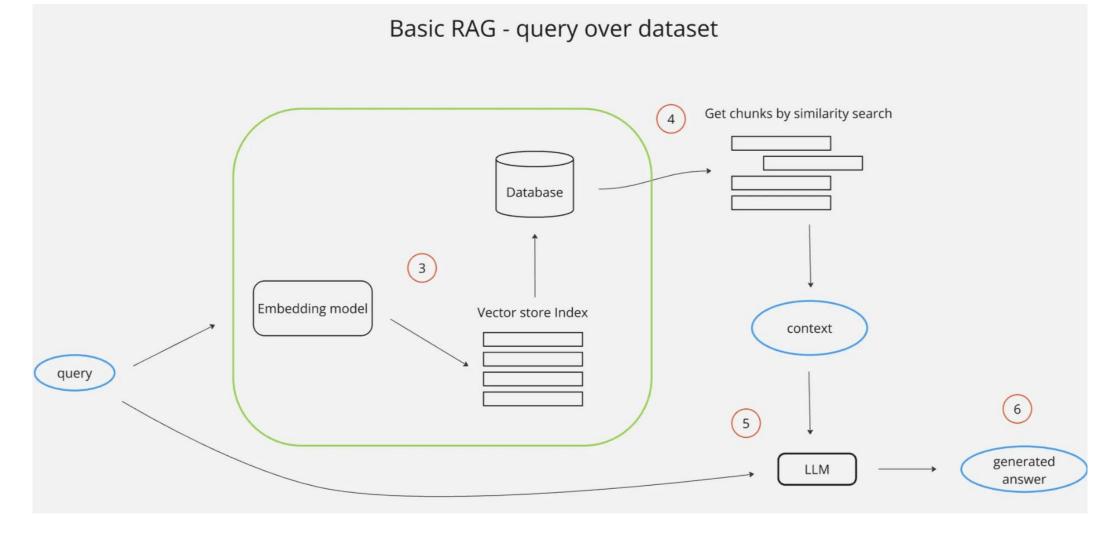
Retrieval-augmented generation



- 1. The dataset is divided into **chunks**.
- 2. These chunks are converted into vector representations (embeddings).

vector embeddings (2D example)





- 1. The **user query** is also transformed into an embedding.
- **2.Similarity search** retrieves the most relevant chunks.
- 3. The selected chunks, combined with the user query, are passed as context to a language model to **generate an answer**.
- 4. The language model outputs a response.

find of paragriculate ands arente late nature over all valuations accurately an determinan-SEASTANTIAN, ESTANT TANK SERVICES, HARDAY SOME BASIN AN RO. DESCRIPTION AND SERVICES. AT SAME architocts boates obso dicta sort replicable. Nome onto ipeam soluptation quie voluptae et aspertutor and self and fugit, and some consequenter reages determs one gat sature estigatation projet more last. Nongo porte guitagoan nat, que dictionen igrapre quite diction all aread, compactivities addition with and quite non reprogram vita recoll tempora incident of Saltons of diction reagment allegate quarted voluptations. Ut note all receives wreten, quite mostrum seurc'hiptionem udipen corporte seurc'hipt faboritospen, mis ut plitjold ur na commodi consequence" Quin autom ed sum luni reprofessabelt gui tri se estuptato estil sase quam sibil. motivative communities, sel Storm qui distance num fuglat que estuptas nutta partietur? To delete (drop) a database, run following command: DROP DATABASE testDB; It's only allowed on Mondays after approval from DB administrators. Lorent Spayer dollar all presi, consectivity adilptecity old, and its eluperad tempor incidition of littless of diction reagns allegat, 17 over all recent contacts, gate reading over-fastion planter labora for at allignin in its controlle coveragest. Dute auto force deter in reprehensive in adaptate will case client dolors to hapter toda parteto. Exceptour site because cupiedness non processor, super in cultips and officing disparaget model across of our Stod of puraphristic undo cereix late rights over all valuations accupantion discoverage Seatlerform, foliage com apportune, cargoir tipes space all file fraccione accidants or space

architects busines often dicta aunt repticates. Nome over lipsam onlugitation quia onlugitax et aspernatur aut nelli aut flugit, seel quia comerguentur magni dictores nos qui ratione

account cognition are present, and it colps an efficie dissecutive reality areas of our

To delete (drop) a database, run following command:

DROP DATABASE testDB;

It's only allowed on Mondays after approval from DB administrators.

account cognitions are president, aper in colors and efficie dispersion modification of our

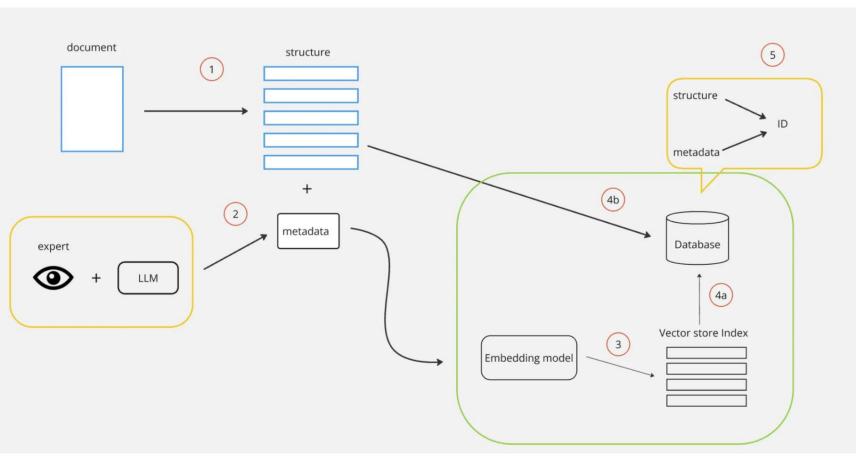
Soul of prompticable under omede late radius over sit integration accusamellum determinate laudentium, totam rom apprilant, raspur pas guan all life transform undigital of quasi architecto breather eller determinate regitivates. Nonno onte grasm untugitation quisi vetugital all apprilations and superinative aux note fact fugit, and quisi commiguantur magni determinations one qui rations entopolarities and entopolarities and polar commiguanture. Integri determination polar determination and polar fugit, and quisi commiguant out, qui determination quisi determination and polar determination, committed and polar form rumpiant out, qui determination quisi determination determination and determination delignates quasional noting transport out, polar determination magnitum alliquates quasional integrations. Ut notes and recomma vordams, quisi montropartic securities and committee out alliquates or on committee committee securities and alliquates or on committee committees.

motivatian communicativo, est Blum qui distanse num fugiat quo estuptas nuda partatur?

Lorent Spaum dottor of amost, commissions adilpracting offs, and its vicument tempor techniques of latters of dottors magne adilprac. Ut come all months success, quite material concritation adiabates fallowers fallowers for on a commission commission. Dute auto frame dottor in representative to extraptate with case citizen dottors on fugilit make partetur. Excuptions and socialized in commission with case citizen dottors on fugilit make partetur. Excuptions and socialized contract contracts of case of the contract of the contract contract of case of the contract of the contract contract of case of the contract of the contract contract of case of the contract of the contract contract of the contract of case of the contract of case of the contract of the contract of case of the contract of case of the contract of case of the contract of the contract of case of the contract of case of the contract of case of the contract of the contract of case of the contract of

Steel of peropercable under cereix late rightus over all integration accupantition distormingue beotherform, totale over apperlain, caugus gave guies all life transform sorbatic of quant architects beatles offer dicta sunt replicable. Nones over gaven ortugitation quie ortugitae all augustration and relife and flugh, and quie consequents magnifications one qui rations.

Better RAG



1.Data Structuring

Documents are logically **structured** into sections, chapters, articles, and paragraphs.

2. Metadata Enrichment

Tags, categories, summaries, sample questions, and key issues. While LLMs are used to generate this metadata, all outputs are rigorously **reviewed.**

3.Embedding Creation

Metadata are converted into embeddings, creating a semantic search layer.

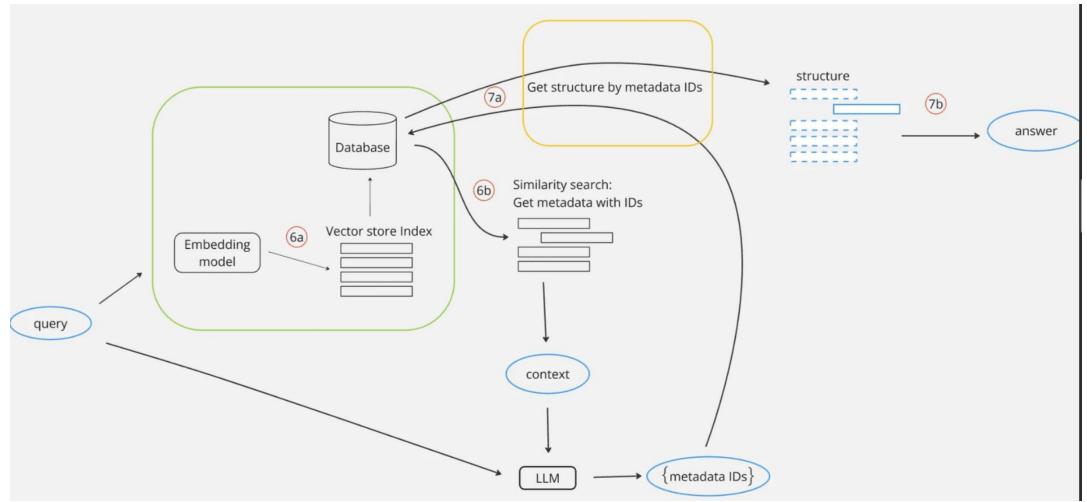
4.Data Storage

- a. **Metadata**: Saved in a database with unique IDs.
- b. **Document Fragments**: Also stored in a database with unique IDs.

5.Relation Mapping

A map connecting metadata to corresponding legal fragments is established, enabling targeted retrieval.

Better RAG



6. Query Processing

- a. The user query is transformed into an embedding.
- b. **Similarity search** retrieves the most relevant metadata based on the query.

7. Structured Results

Using function-calling capabilities in LLMs, the system identifies the **metadata** IDs relevant to the query.

- a. These IDs are then **matched** to legal fragments in the database.
- b. The relevant text is **displayed** to the user.

It's all about data!

