Flowhog

WHAT IS RAG?

GOAL: Have AI that knows YOUR business in detail

DANI FUYÀ

Flowhog

RAG PROCESS

⚙️ DOCUMENT INGESTION

1. Parsing
2. Chunking
3. Embeddings
4. Storage

🔍 CONTEXT RETRIEVAL

1. Question Processing (Expansion + Embeddings)
2. Search (Embeddings + Keywords)
3. Reranking

🔍 RESPONSE GENERATION

LLM generates a response considering your question and sources relevant to that question

DANI FUYÀ

Flowhog

STEP 1 : DOCUMENT INGESTION

GOAL: Create a searchable knowledge base

1 - PARSE

Extract text from **PDFs**, **docs**, **Excel**, and **emails**

2 - CHUNK

Split text into pieces

3 - EMBED

Extract semantic meaning from each chunk

4 - STORE

Save each text piece and its meaning into a database

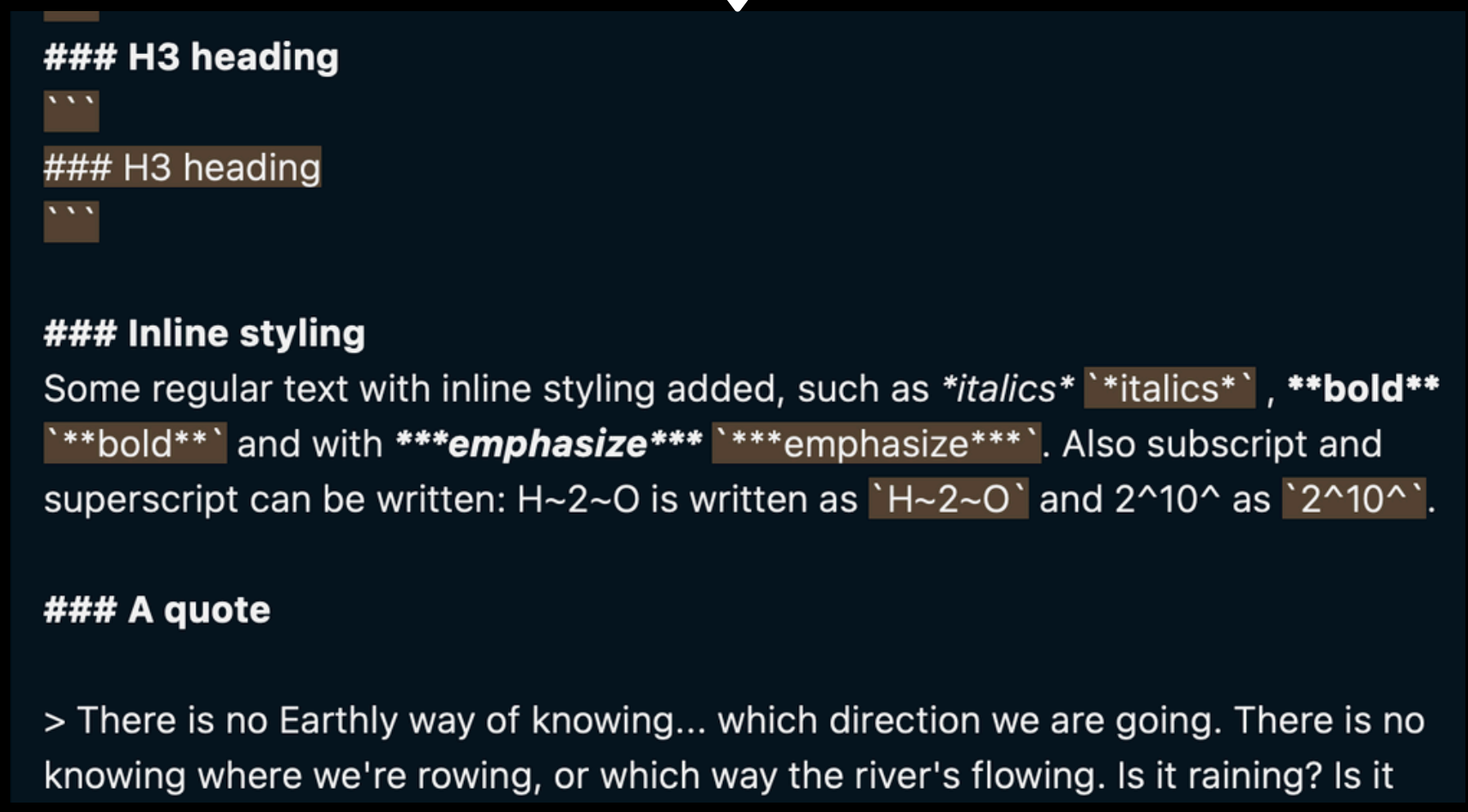
DANI FUYÀ

Flowhog

STEP 1.1: INGESTION > PARSING

GOAL: Extract text from documents

Name	Date modified	Type
Product Details.zip	24/09/2017 21:46	Compressed (zipp...
Product Enquiry.msg	24/09/2017 21:04	Outlook Item
Proposal Data.xlsx	24/09/2017 20:52	Microsoft Excel W...
Proposal Description.doc	24/09/2017 20:56	Microsoft Word 9...
Proposal Presentation.ppt	24/09/2017 20:54	Microsoft PowerP...
Scanned Docs.tif	24/09/2017 20:59	TIFF image



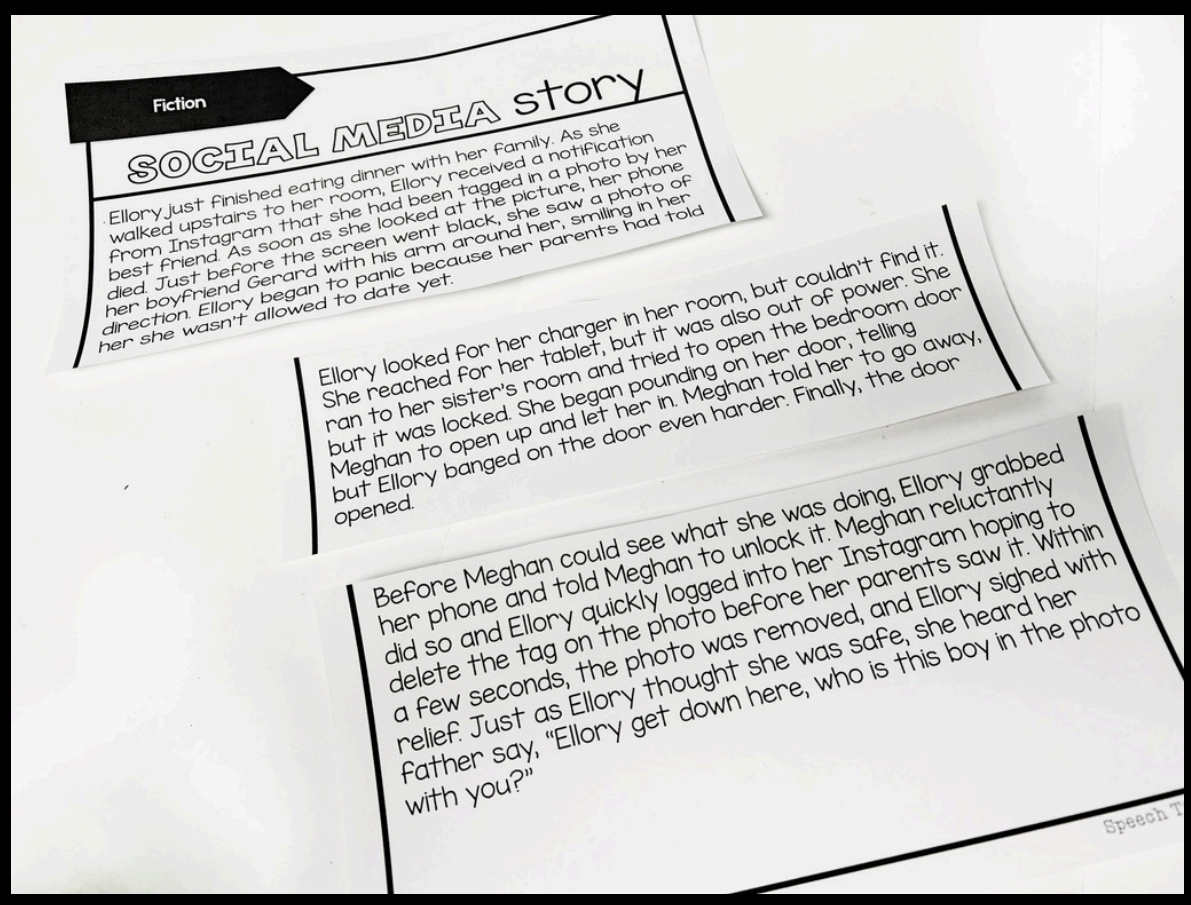
````
### H3 heading
```
### H3 heading
```

### Inline styling
Some regular text with inline styling added, such as *italics* `*italics*` , **bold** `**bold**` and with ***emphasize*** `***emphasize***`. Also subscript and superscript can be written: H~2~O is written as `H~2~O` and 2^10^ as `2^10^`.

### A quote

> There is no Earthly way of knowing... which direction we are going. There is no knowing where we're rowing, or which way the river's flowing. Is it raining? Is it
````

DANI FUYÀ

Flowhog

STEP 1.2: INGESTION > CHUNKS

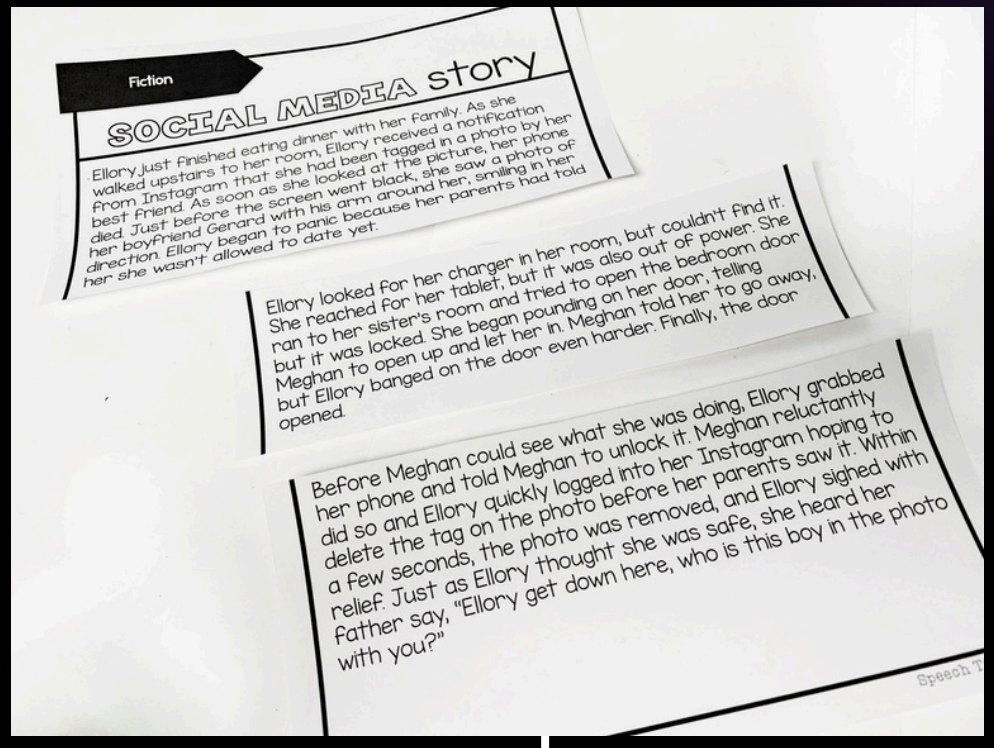
GOAL: Split text into pieces

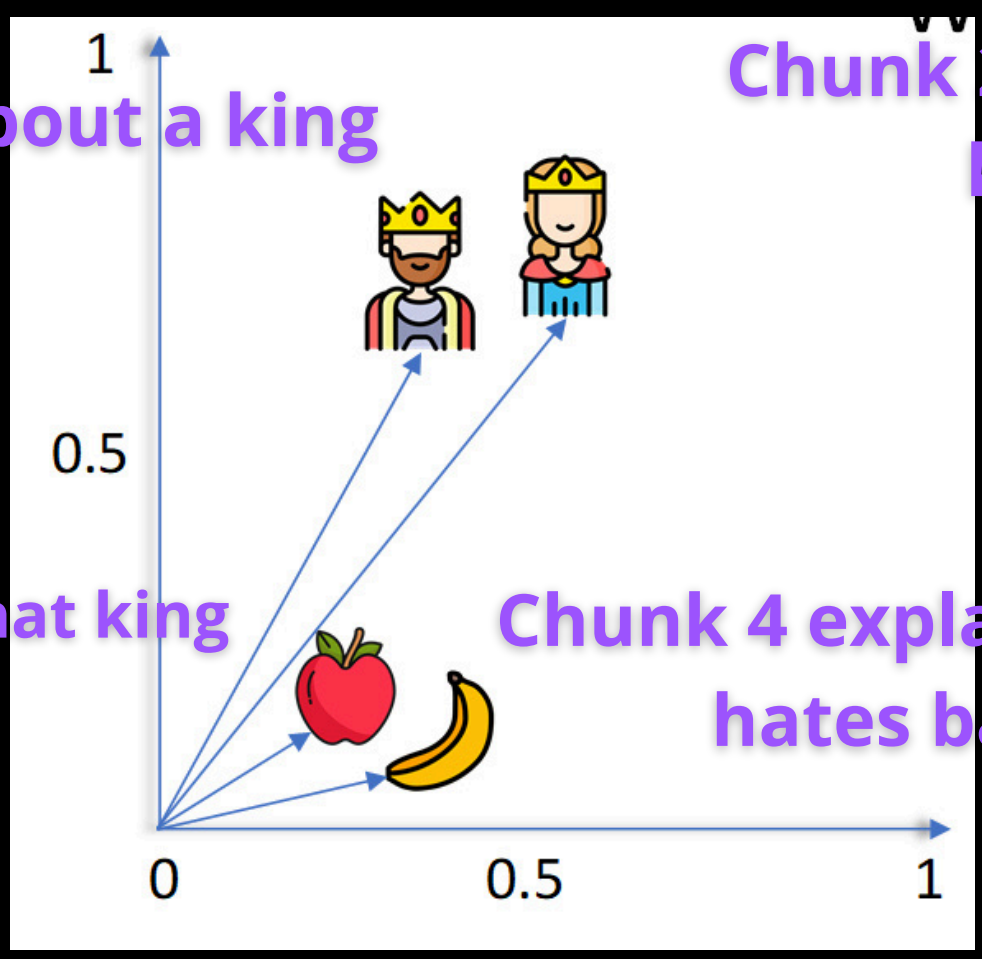
```
### H3 heading
```
H3 heading
```

### Inline styling
Some regular text with inline styling added, such as *italics* `*italics*` , **bold** `**bold**` and with ***emphasize*** `***emphasize***`. Also subscript and superscript can be written: H~2~O is written as `H~2~O` and 2^10^ as `2^10^`.

### A quote

> There is no Earthly way of knowing... which direction we are going. There is no knowing where we're rowing, or which way the river's flowing. Is it raining? Is it




DANI FUYÀ
```

Flowhog

STEP 1.3: INGESTION > EMBEDDINGS

GOAL: Represent what each text chunk is talking about as a vector



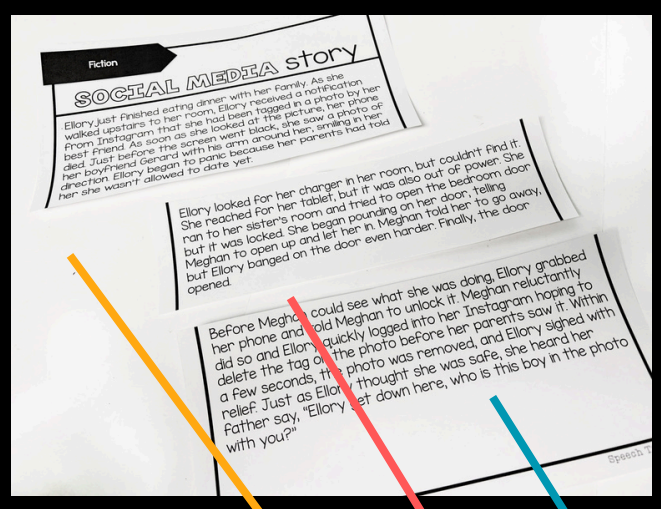


DANI FUYÀ

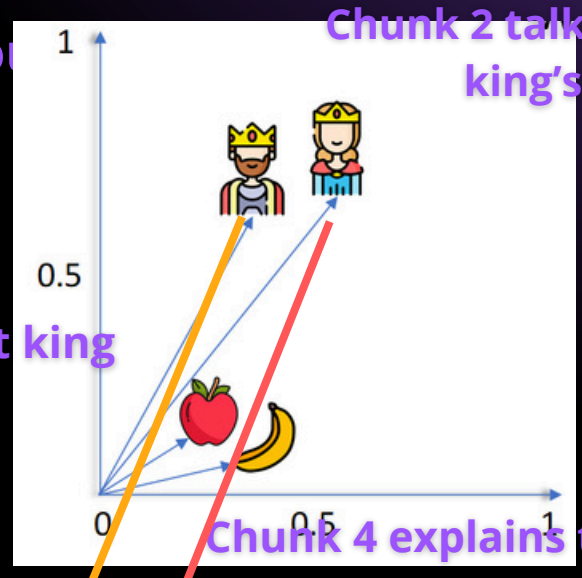
Flowhog

STEP 1.4: INGESTION > STORE

GOAL: Store in a DB each chunk and its embedding vector



Chunk 1 talks abou



Chunk 2 talks about the king's wife

Chunk 3 explains that king likes apples

Chunk 4 explains that king hates bananas

chunk	embedding
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
One of the operating systems shown…	[-0.05131773,0.022594599,0.09677809…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
National character string literals…	[0.052299663,0.013026265,0.0477994…
National character string literals…	[0.052299663,0.013026265,0.0477994…
NCHAR type is provided as the data…	[-0.044851042,0.014970196,0.068836…
NCHAR type is provided as the data…	[-0.044865385,0.014974983,0.068858…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…

DANI FUYÀ

Flowhog

STEP 2: CONTEXT RETRIEVAL

GOAL: Given a question, find the chunks that contain the answer to that question

1 - Question Embedding

Represent the question meaning as a vector

2 - Search in DB

Find chunks in db that are similar to the question

3 - Reranking

Pick the top 5 results

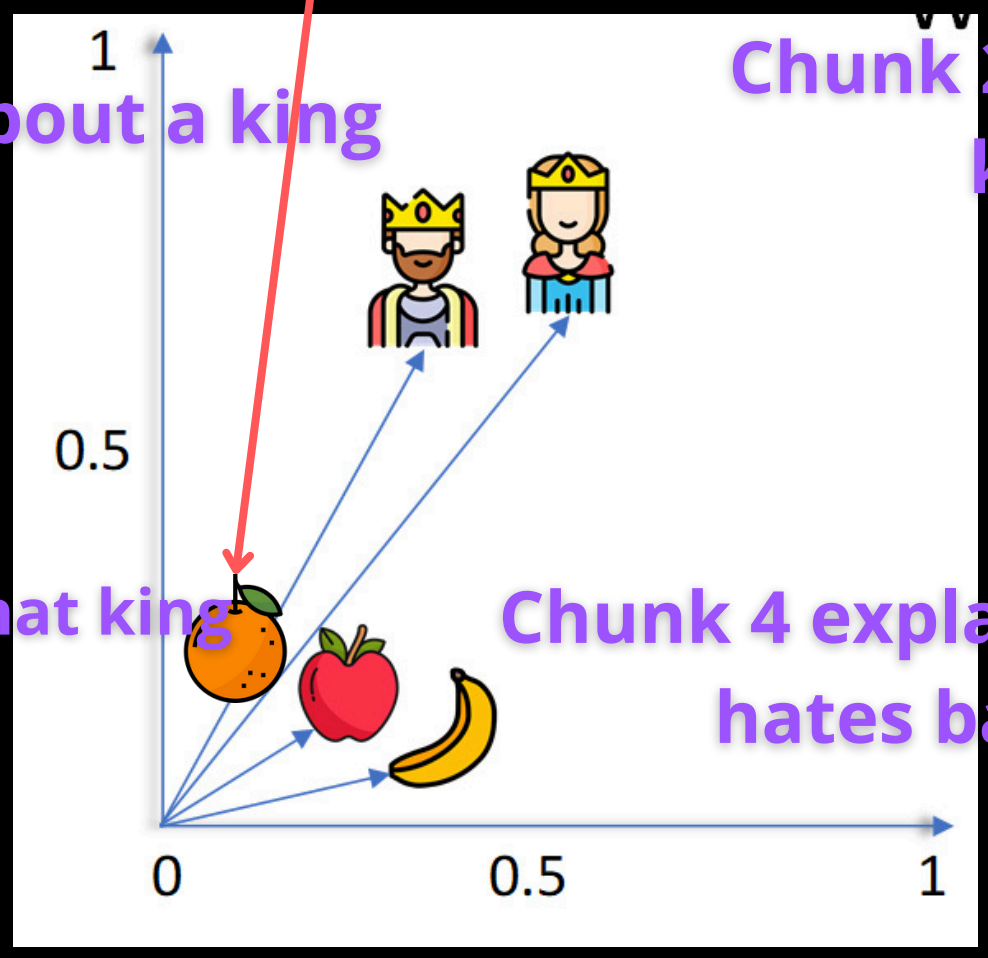
DANI FUYÀ

Flowhog

STEP 2.1: RETRIEVAL > EMBEDDINGS

GOAL: Represent user question as a vector so that you can compare it with text chunks in DB

Did he like oranges?



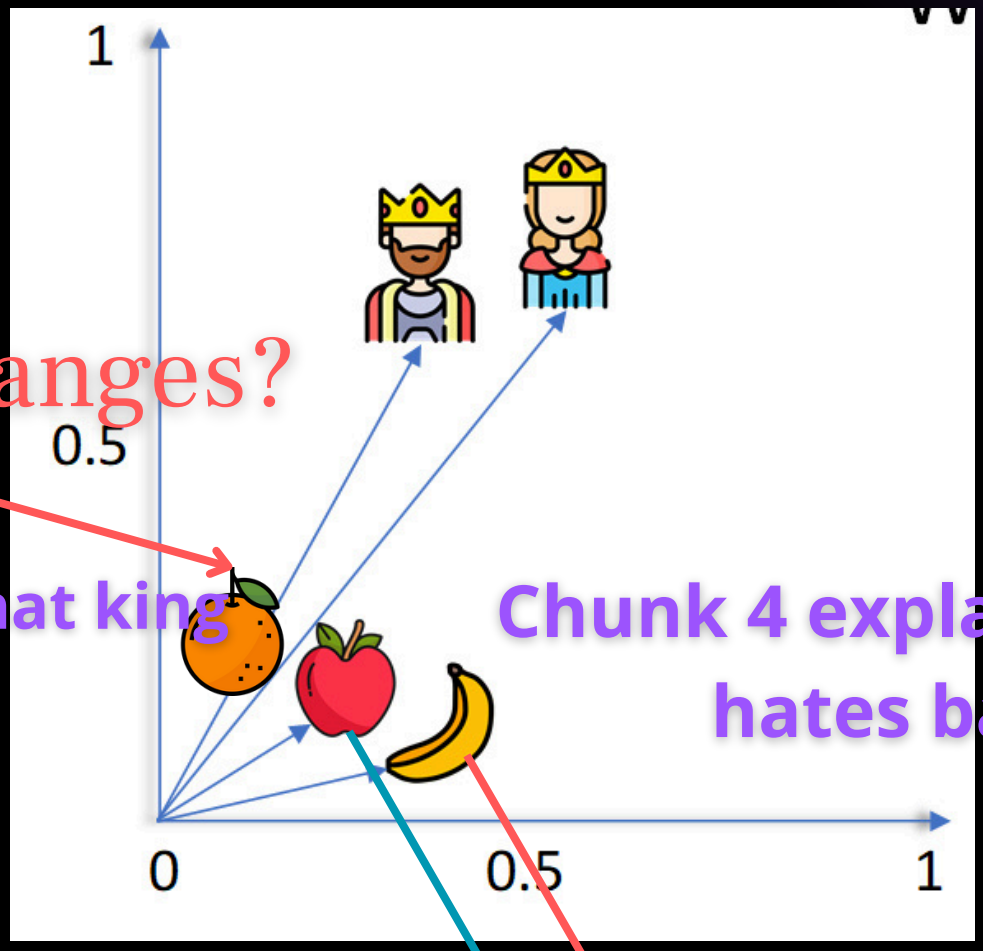
DANI FUYÀ

Flowhog

STEP 2.2: RETRIEVAL > SEARCH

GOAL: Extract chunks from DB that can potentially answer the user question

Did he like oranges?



Chunk 3 explains that king likes apples

Chunk 4 explains that king hates bananas

chunk	embedding
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
One of the operating systems shown…	[-0.0531773,0.022594599,0.09677809…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
National character string literals…	[0.052299663,0.013026265,0.0477994…
National character string literals…	[0.052299663,0.013026265,0.0477994…
NCHAR type is provided as the data…	[-0.044851042,0.014970196,0.068836…
NCHAR type is provided as the data…	[-0.044865385,0.014974983,0.068858…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…

DANI FUYÀ

Flowhog

STEP 2.3: RETRIEVAL > RERANKING

GOAL: Refine chunk candidates to increase probability that chunks selected are relevant to the question

Did he like oranges?

chunk	embedding
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
One of the operating systems shown…	[-0.0531773,0.022594599,0.09677809…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
This section explains how to delet…	[0.027757,0.059048597,0.063414864,…
The data type of the NCHAR type co…	[0.002209758,0.04273992,0.05689596…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
The interface for application deve…	[-0.04833659,-0.0066552167,0.10973…
National character string literals…	[0.052299663,0.013026265,0.0477994…
National character string literals…	[0.052299663,0.013026265,0.0477994…
NCHAR type is provided as the data…	[-0.044851042,0.014970196,0.068836…
NCHAR type is provided as the data…	[-0.044865385,0.014974983,0.068858…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…
This is the basic configuration wh…	[-0.0129189035,0.014315235,0.10044…

Reranker

More relevant

1. **Chunk 3**
2. **Chunk 4**

Less Relevant

DANI FUYÀ

Flowhog

STEP 3 : GENERATION

GOAL: LLM generates response using your company sources

Prompt

Context:

[1] **Chunk 3**

[2] **Chunk 4**

Question: Did he like oranges?

Answer: There's no evidence that he liked or disliked oranges. According to sections 1 and 2, he likes apples and hates bananas.

DANI FUYÀ