

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.
4. A snail can sleep for three years.
5. The longest word in the English language is 'pneumonoultramicroscopicsilicovolcanoconiosis.'
6. The elephant is the only mammal that can't jump.
7. The letter 'Q' is the only letter not appearing in any U.S. state name.
8. The heart of a shrimp is located in its head

....Many more.....



Terminal

>> What is an interesting fact about the English language?

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

fact

"Dreamt" is the only English word that ends with the letters "mt"

Prompt

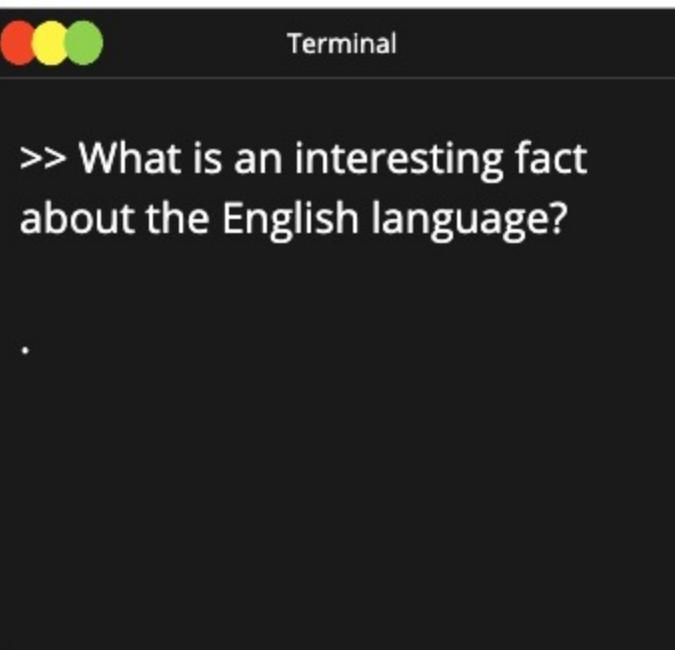
Use the following interesting fact to answer the user's question: "Dreamt" is the only English word that ends with the letters "mt"

Here is the user's question:
What is an interesting fact about the English language?

facts.txt

- 1. "Dreamt" is the only English word that ends with the letters "mt."
- 2. An ostrich's eye is bigger than its brain.
- 3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....



fact

Prompt

"Dreamt" is the only English word that ends with the letters "mt"

Use the following interesting fact to answer the users question:
{FACT}

Here is the user's question:
{QUESTION}

We need to load up the content in facts.txt

We need to be able to search its contents

LangChain provides classes to help load data from different types of files

These are called 'loaders'

Files we want to load

facts.txt

report.pdf

users.json

blog.md

Loader classes provided by LangChain

TextLoader

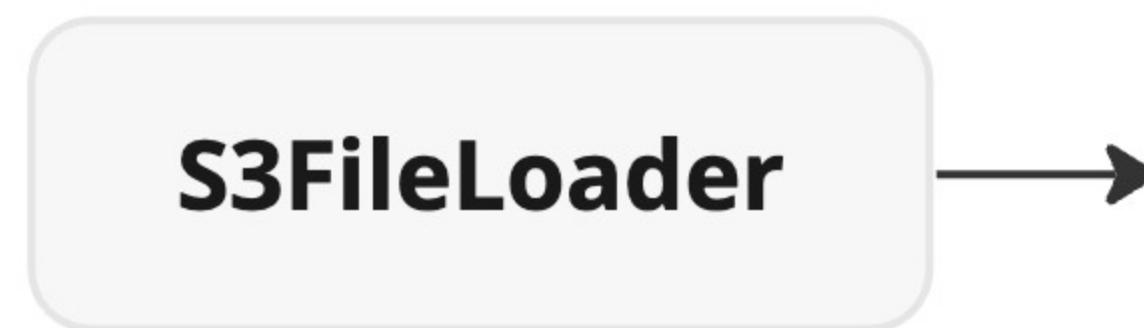
PyPDFLoader

JSONLoader

Unstructured MarkdownLoader

LangChain also provides classes to load up any kind of file from different locations

Somewhat confusingly, these are *also* called loaders



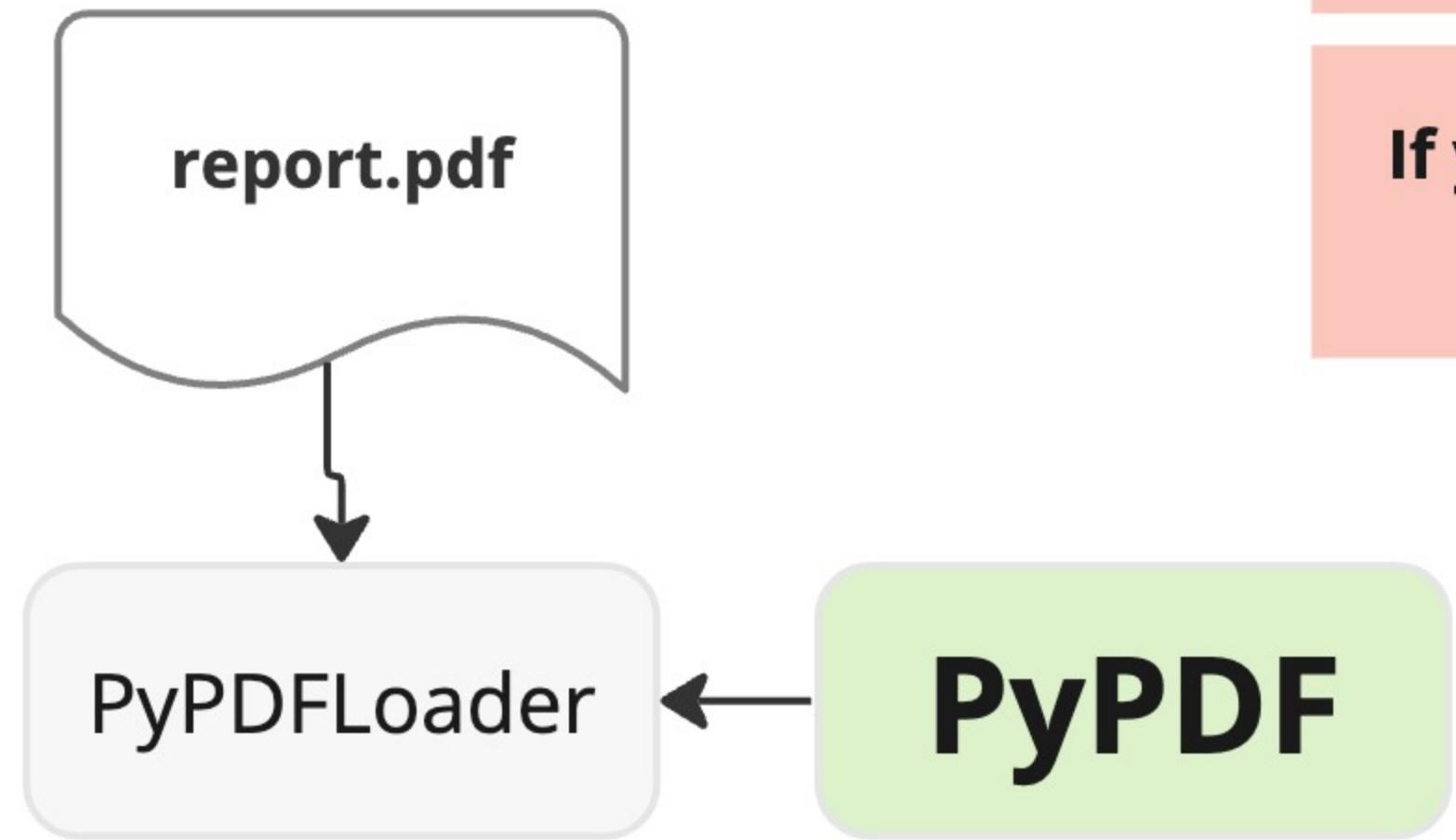
Amazon S3

users.json

blog.md

report.pdf

facts.txt



Many loaders require you to install extra packages to use

If you don't, you'll get an error message telling you to install it

Package that must be installed for LangChain to be able to load a pdf

facts.txt

- 1. "Dreamt" is the only English word that ends with the letters "mt."
- 2. An ostrich's eye is bigger than its brain.
- 3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....



TextLoader



Document

page_content	<ul style="list-style-type: none">1. "Dreamt" is the only English word that ends with the letters "mt."2. An ostrich's eye is bigger than its brain.3. Honey is the only natural food that is made without destroying any kind of life. <p><i>....Many more.....</i></p>
metadata	{ "source": "facts.txt" }

Result of loading a file is a "Document"

Contains the contents of the file ("page_content")

Contains info about where this document came from ("metadata")

User's Question

What is an interesting fact about the English language?

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

Prompt

Use the following interesting fact to answer the users question: {FACT}

Here is the user's question:
{QUESTION}

What should be the "fact" input?

User's Question

What is an interesting fact about the English language?

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

Prompt

Use the following interesting fact to answer the users question:
1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

Option #1

Put the entire contents of the "facts.txt" file into the prompt

Downside: longer prompt = **costs more to run**

Downside: longer prompt = **takes longer to run**

Downside: longer prompt = **LLM has a harder time finding relevant facts**

User's Question

What is an interesting fact about the English language?

facts.txt

- 1. "Dreamt" is the only English word that ends with the letters "mt."
- 2. An ostrich's eye is bigger than its brain.
- 3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

Prompt

Use the following interesting fact to answer the user's question:

- 1. "Dreamt" is the only English word that ends with the letters "mt."

Here is the user's question:
{QUESTION}

Option #2

Put only the top 1-3 relevant facts into the prompt

Upside: shorter prompt! Less expensive, faster, more focused.

Downside: need to somehow find the most relevant facts

User's Question

What is an interesting fact about the English language?

facts.txt

- 1. "Dreamt" is the only English word that ends with the letters "mt."
- 2. An ostrich's eye is bigger than its brain.
- 3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

1. "Dreamt" is the only English word that ends with the letters "mt."

2. An ostrich's eye is bigger than its brain.

3. Honey is the only natural food that is made without destroying any kind of life.

4. *The Python programming language was named after "Monty Python's Flying Circus"*

5. Additional fact

6. Additional fact

Prompt

Use the following interesting fact to answer the user's question: 1. "Dreamt" is the only English word that ends with the letters "mt."

Here is the user's question:

{QUESTION}

Possible way to implement the search

Split up the text by each fact

Find facts that contain words from the user's question

User's Question

What is an interesting fact about a language?

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

1. "Dreamt" is the only English word that ends with the letters "mt."

2. An ostrich's eye is bigger than its brain.

3. Honey is the only natural food that is made without destroying any kind of life.

4. *The Python programming language was named after "Monty Python's Flying Circus"*

5. Additional fact

6. Additional fact

Prompt

Use the following interesting fact to answer the users question: 4. *The Python programming language was named after "Monty Python's Flying Circus"*

Here is the user's question:
{QUESTION}

What if the user's question doesn't contain the word "English"?

Alternative - Semantic Search

Understand the goal of the users
search

We are going to implement this using
embeddings

We are going to try to score how much a given sentence talks about two things.

One score for how much the sentence talks about
bravery

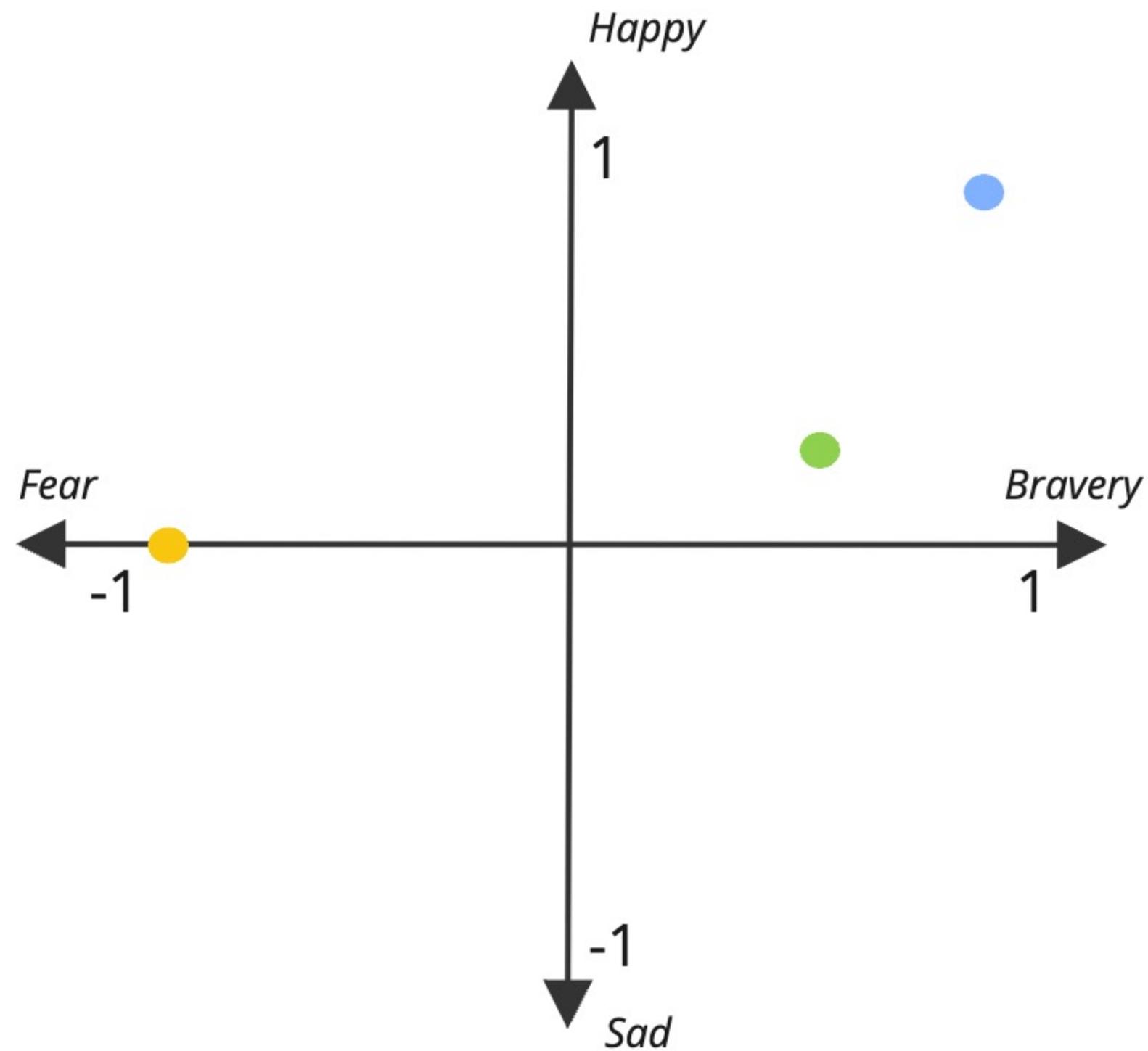
One score for how much the sentence talks about
happiness

Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)
The happy child jumped bravely from rock to rock	1	1
The child was not timid and had a good time jumping from rock to rock	.5	.3
Although filled with great fear, the child jumped from rock to rock	-1	0

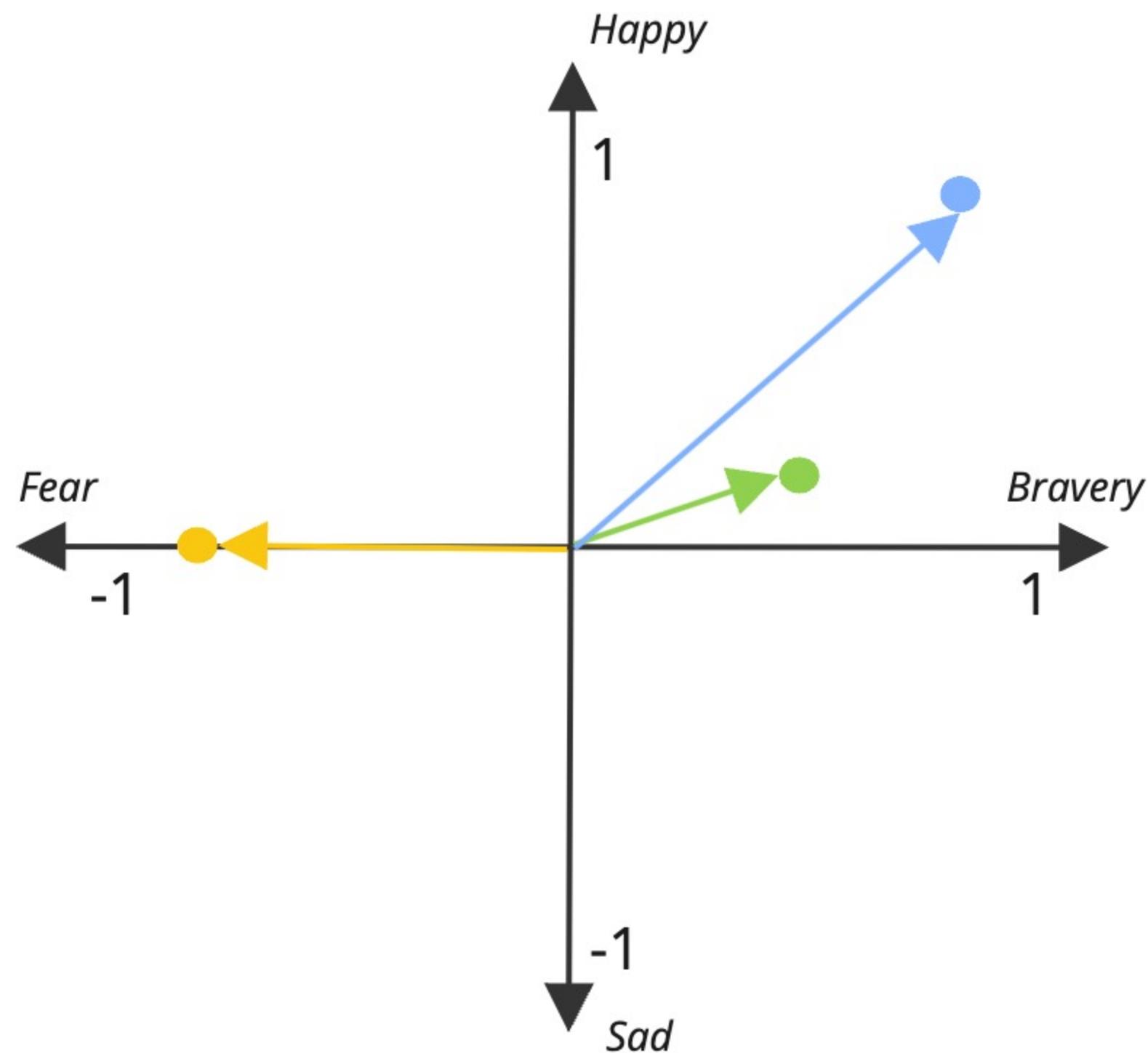
Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)	Embedding
The happy child jumped bravely from rock to rock	1	1	[1, 1]
The child was not timid and had a good time jumping from rock to rock	.5	.3	[.5, .3]
Although filled with great fear, the child jumped from rock to rock	-1	0	[-1, 0]

An embedding is a list of numbers between -1 and 1 that score how much a piece of text is talking about some particular quality

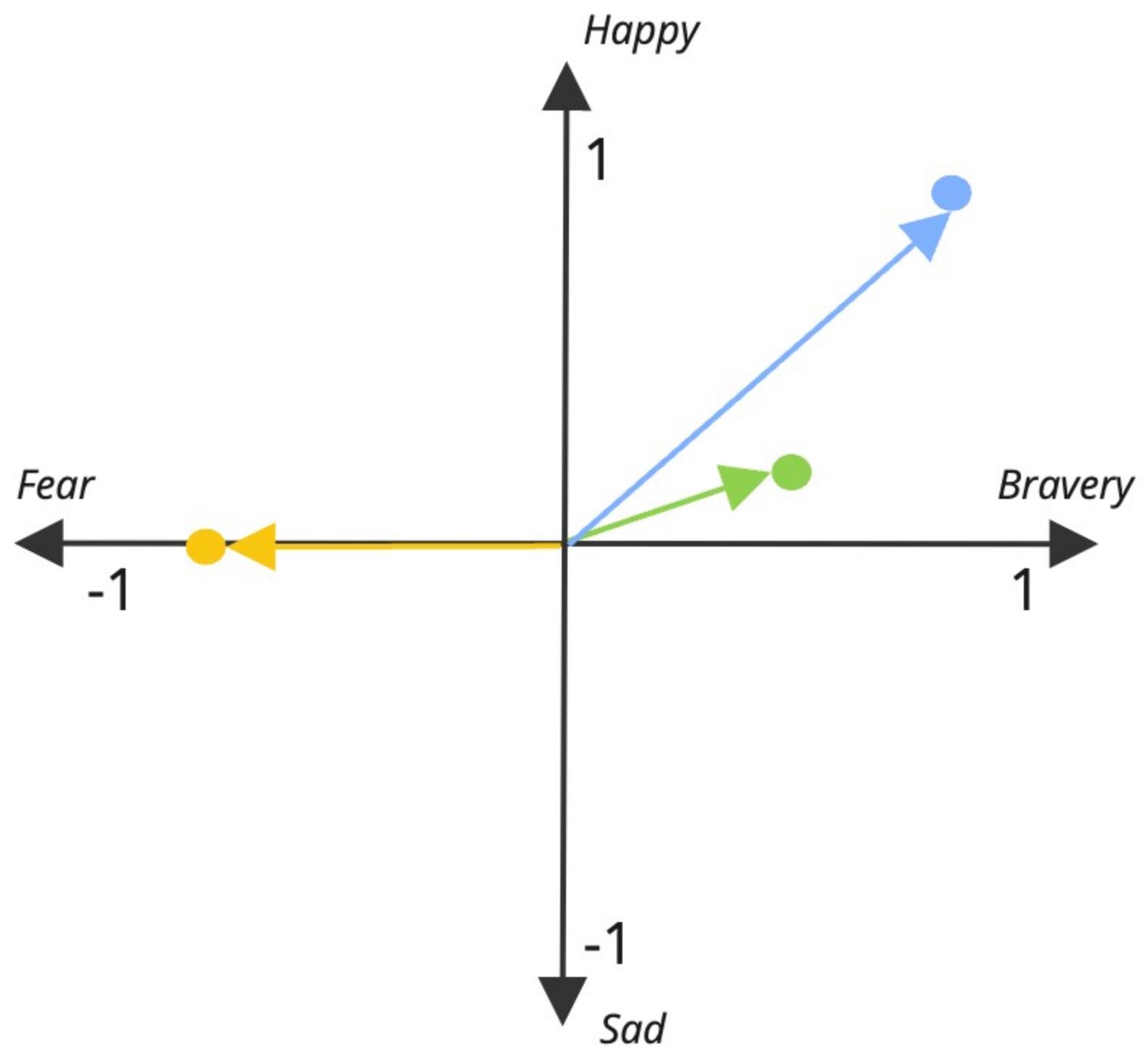
These embeddings only rate **2 qualities**, but real embeddings frequently have **700 to 1500 "dimensions"** or qualities that they score



Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)
The happy child jumped bravely from rock to rock	1	1
The child was not timid and had a good time jumping from rock to rock	.5	.3
Although filled with great fear, the child jumped from rock to rock	-1	0

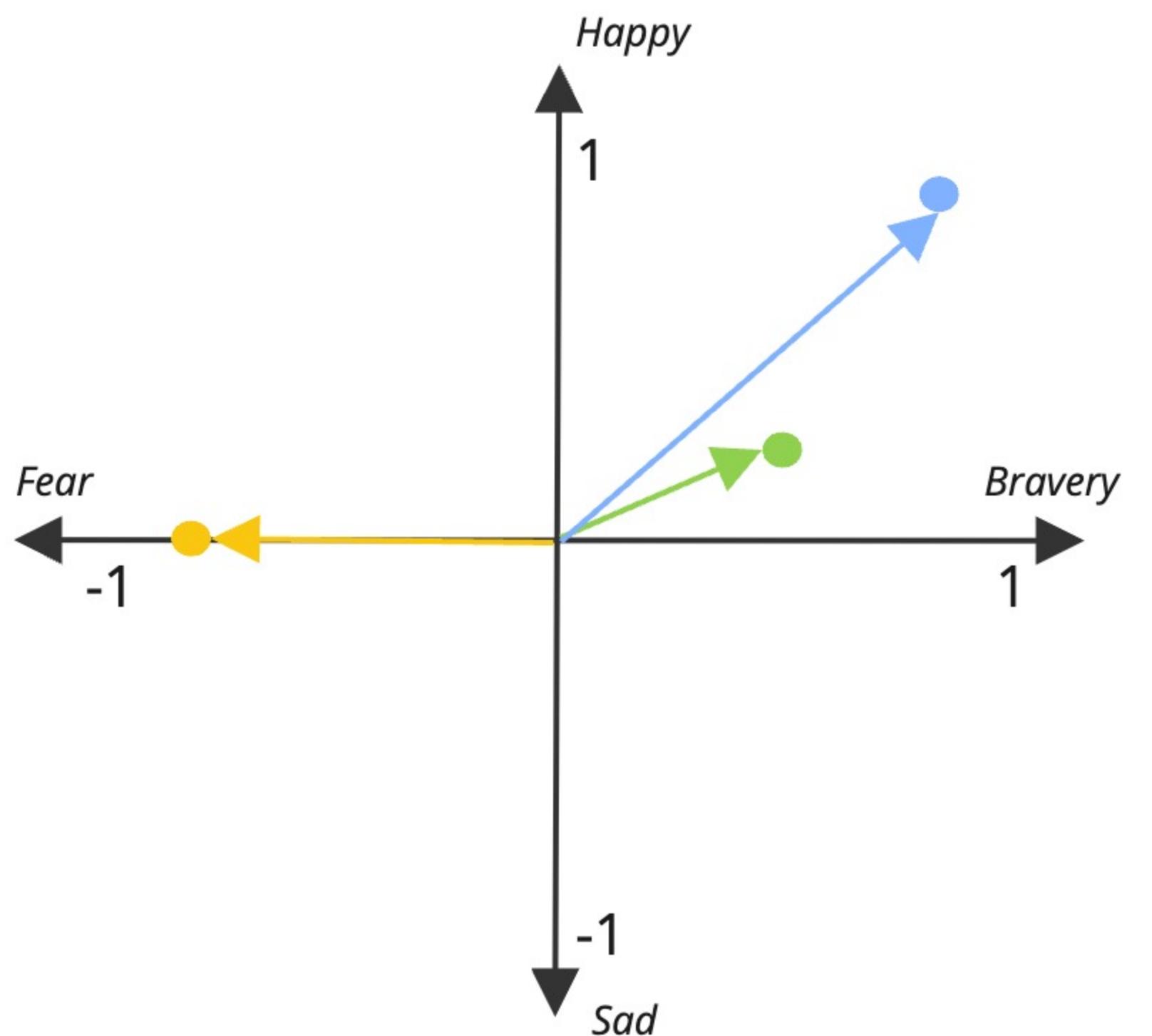


Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)
The happy child jumped bravely from rock to rock	1	1
The child was not timid and had a good time jumping from rock to rock	.5	.3
Although filled with great fear, the child jumped from rock to rock	-1	0



Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)
The happy child jumped bravely from rock to rock	1	1
The child was not timid and had a good time jumping from rock to rock	.5	.3
Although filled with great fear, the child jumped from rock to rock	-1	0

How can we tell how similar these sentences are?



Squared L2

Using the **distance between two points** to figure out how similar they are

Cosine similarity

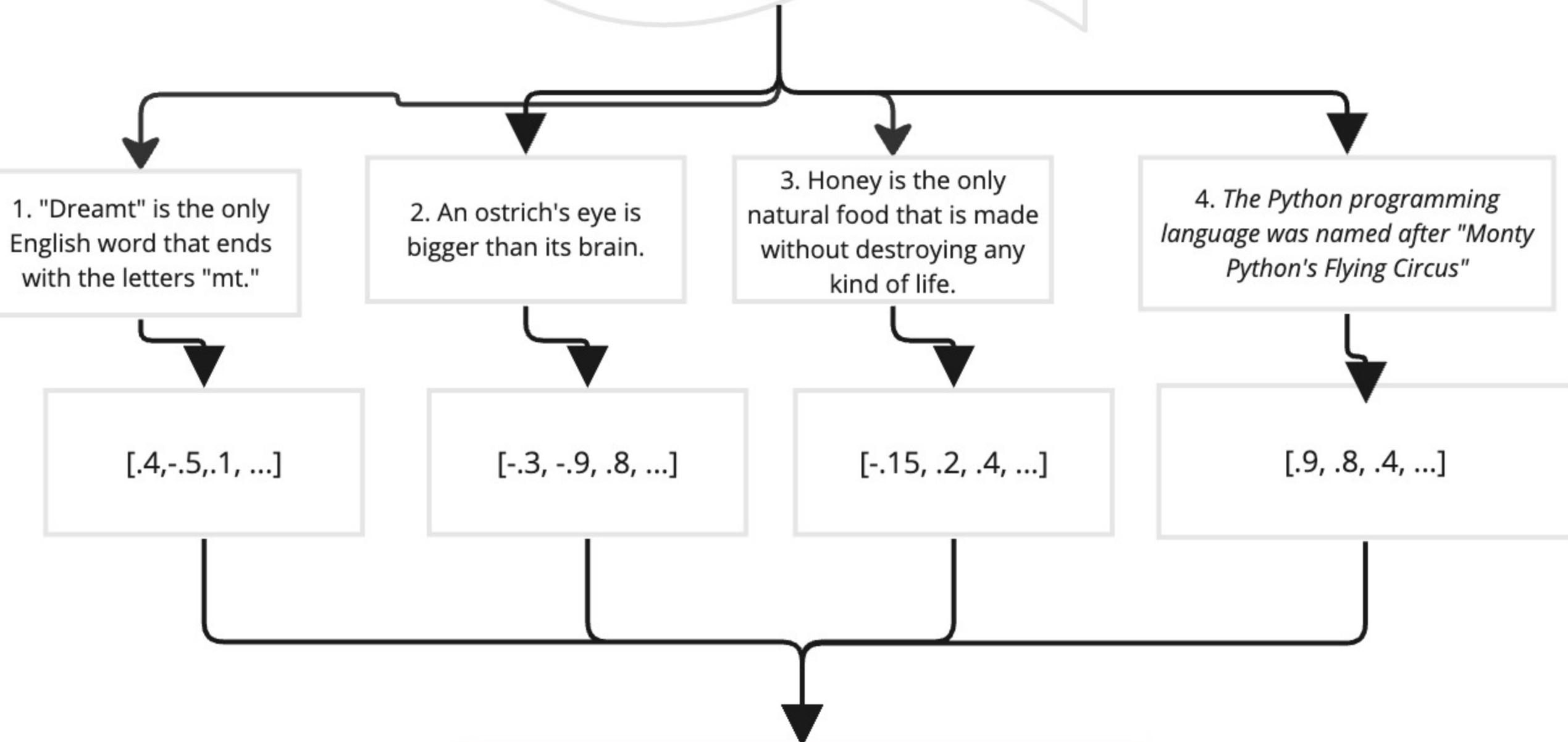
Using the **angle between two vectors** to figure out how similar they are

Sentence	Bravery Score (1 = bravery, -1 = fear)	Happiness Score (1 = happy, -1 = sad)
The happy child jumped bravely from rock to rock	1	1
The child was not timid and had a good time jumping from rock to rock	.5	.3
Although filled with great fear, the child jumped from rock to rock	-1	0

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....



Split the text into separate chunks

Calculate embeddings for each chunk

Store embeddings in a database specialized in storing embeddings

Vector Store

User's Question

What is an interesting fact about the English language?

[.3, .2, -.4, ...]

Vector Store

1. "Dreamt" is the only English word that ends with the letters "mt."

Prompt

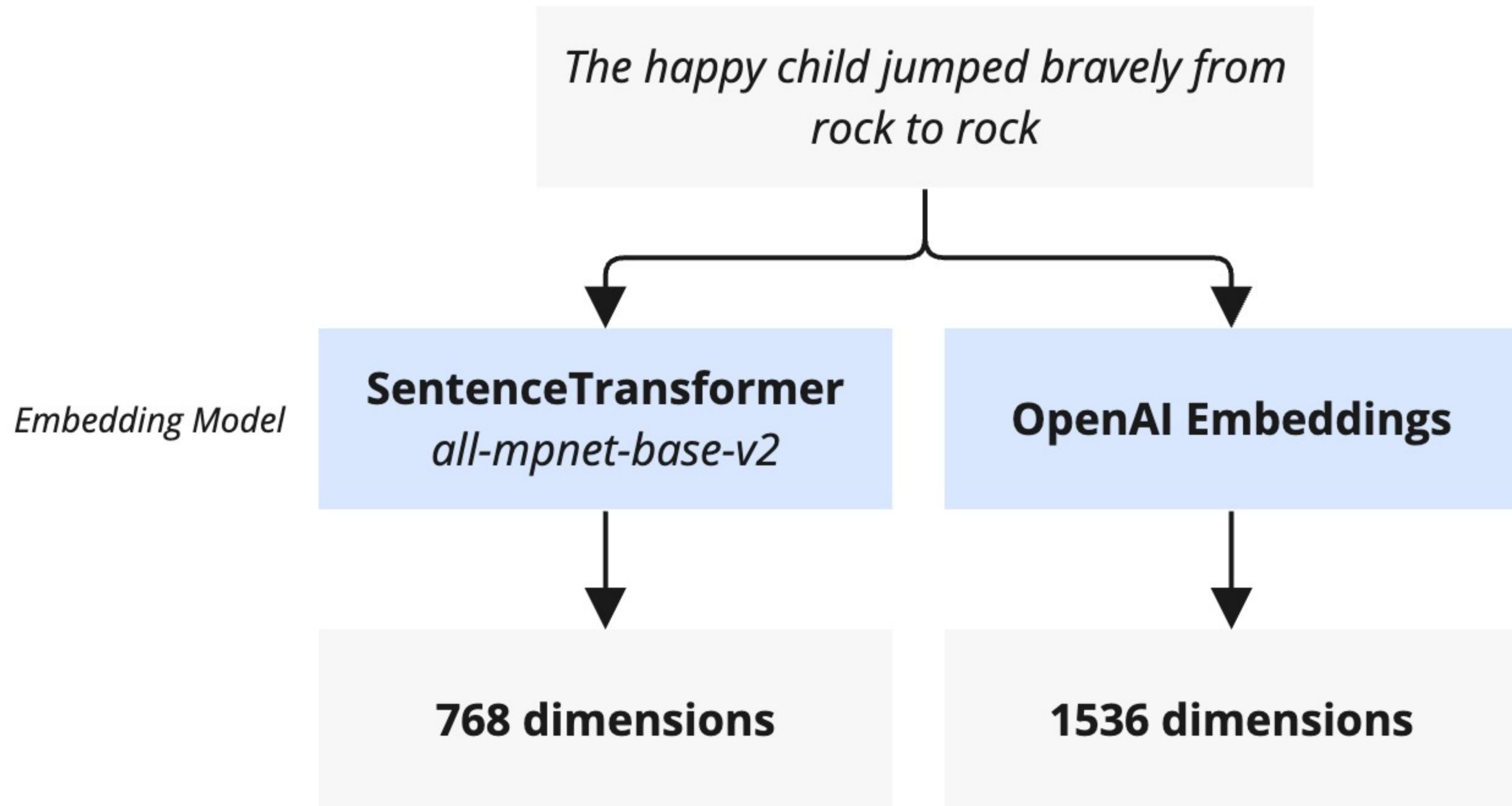
Use the following interesting fact to answer the user's question: {fact}

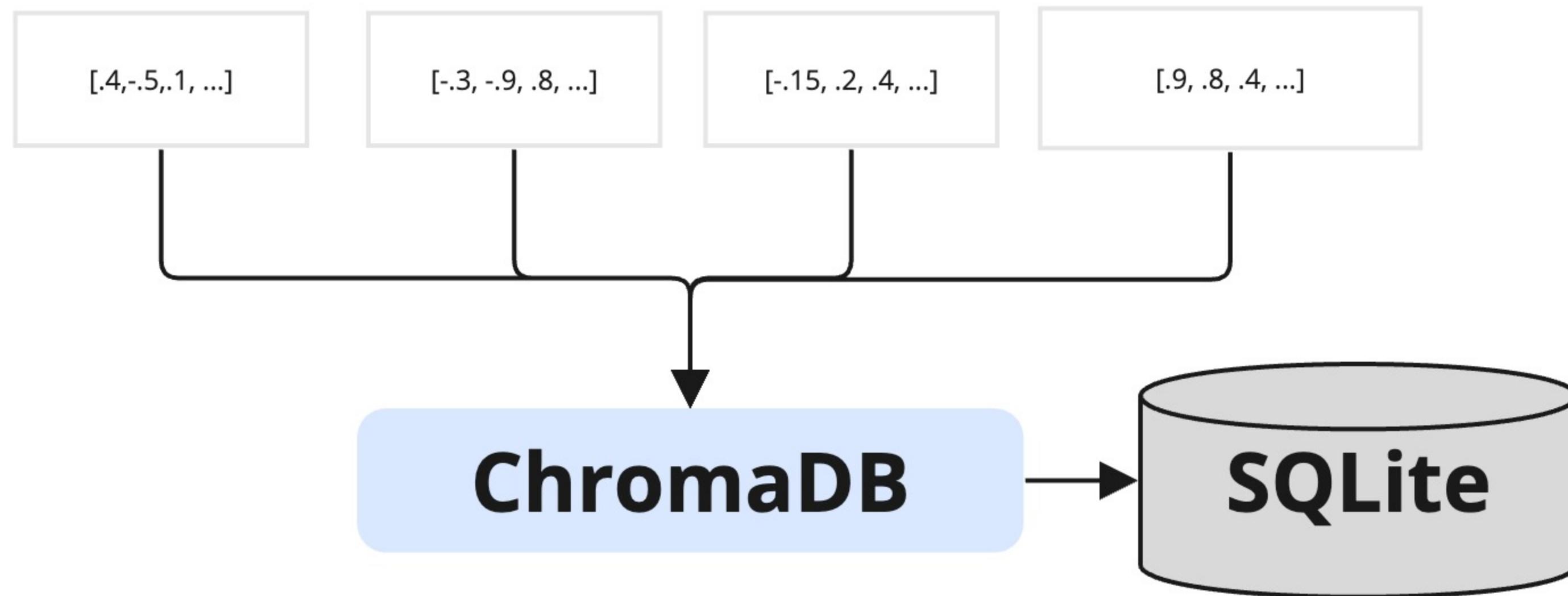
Here is the user's question:
{question}

Create embedding out of the user's question

Do a similarity search with our stored embeddings to find the ones most similar to the user's question

Put the most relevant 1-3 facts into the prompt along with the user's question





facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....

1. "Dreamt" is the only English word that ends with the letters "mt."

[.4,-.5,.1, ...]

2. An ostrich's eye is bigger than its brain.

[-.3, -.9, .8, ...]

3. Honey is the only natural food that is made without destroying any kind of life.

[-.15, .2, .4, ...]

4. *The Python programming language was named after "Monty Python's Flying Circus"*

[.9, .8, .4, ...]

[.4,[.4,-.5,..1] ...]

[.4,-.5,.1, ...]

[.4,-.5,.1, ...]

[-.3[-.3,-.8, .8] ...]

[-.3, -.9, .8, ...]

[-.3, -.9, .8, ...]

[-.1[5,1.2, .4, .4] ...]

[-.15, .2, .4, ...]

[-.15, .2, .4, ...]

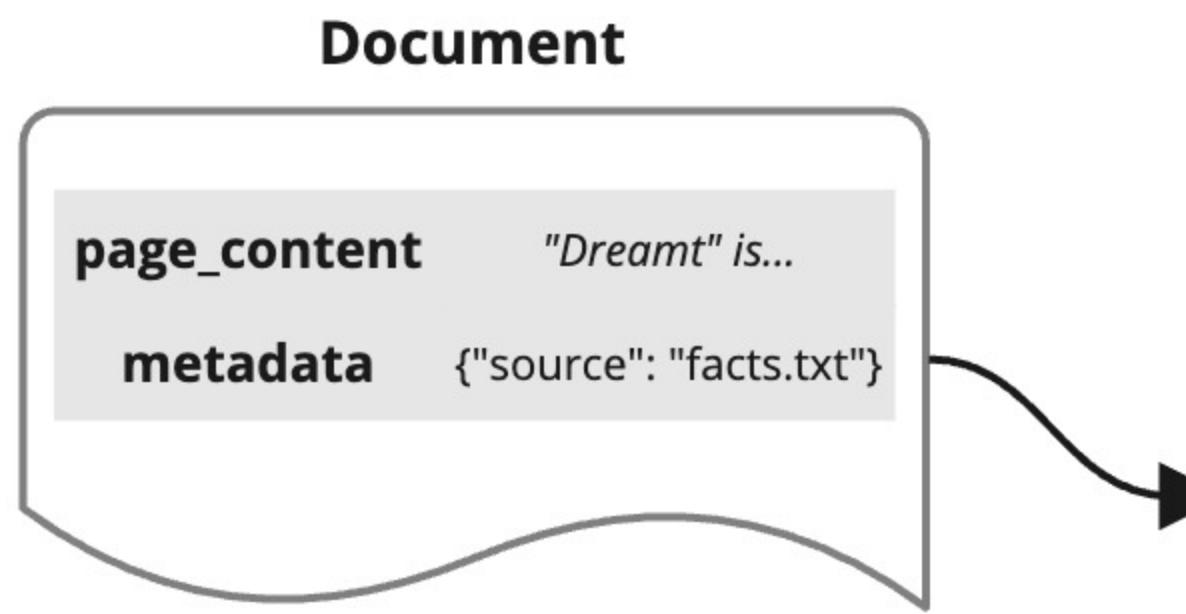
[.9,[.8, .4] ...]

[.9, .8, .4, ...]

[.9, .8, .4, ...]

Vector Store

Vector Store

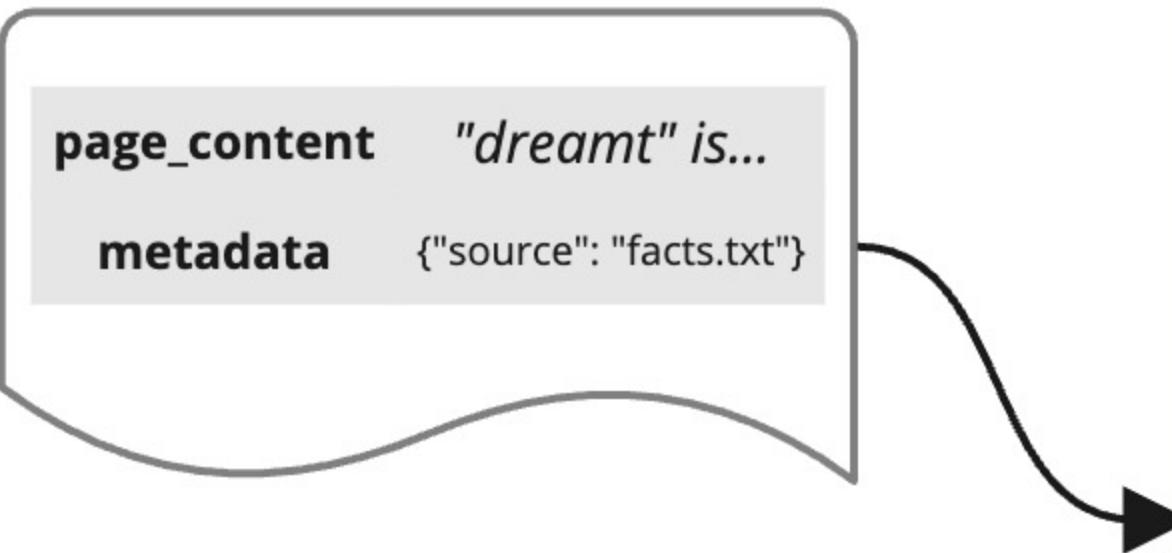


id	metadata	document	embedding
lk2j4	{"source": "facts.txt"}	<i>"Dreamt"</i> is...	[.24, -.1, .4,...]
lk2j			

Option #1 for avoiding duplicates

Check to see if there is another record with the same content

Document

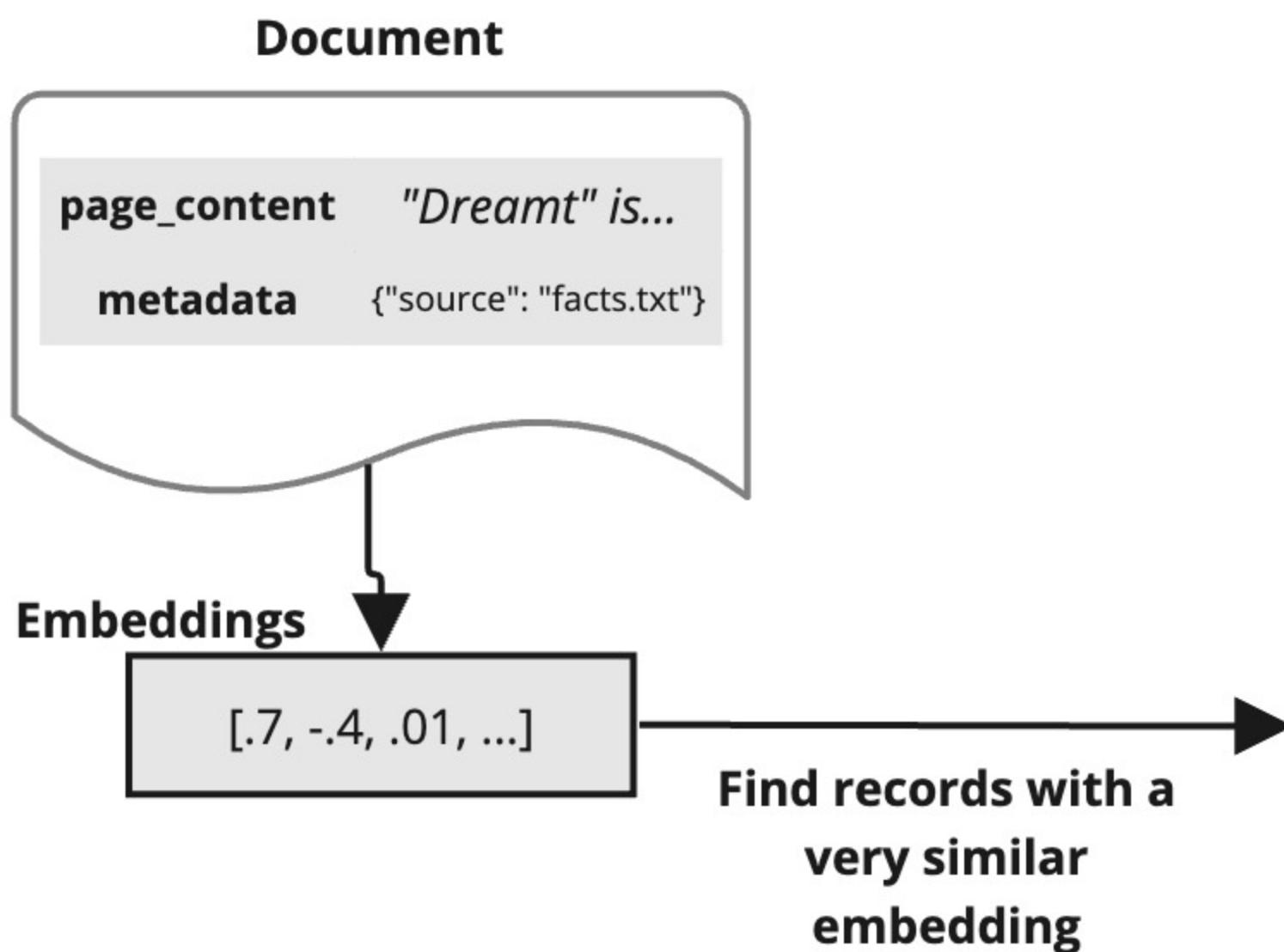


Vector Store

id	metadata	document	embedding
2503	{"source": "facts.txt"}	Isaac Newton invented the cat flap	[.3, -.2, .001]
5902	{"source": "facts.txt"}	"Dreamt" is the only word in the English...	[-.9, .8, .1]

Option #2 for avoiding duplicates

Do a similarity lookup to see if there are any *very similar* records already in the vector store



Vector Store

id	metadata	document	embedding
2503	{"source": "facts.txt"}	Isaac Newton invented the cat flap	[.3, -.2, .001]
5902	{"source": "facts.txt"}	"Dreamt" is the only word in the English...	[.7, -.4, .01, ...]

User's Question

What is an interesting fact about the English language?



[.3, .2, -.4, ...]



Vector Store



1. "Dreamt" is the only English word that ends with the letters "mt."

Prompt

Use the following interesting fact to answer the user's question: {fact}

Here is the user's question:
{question}

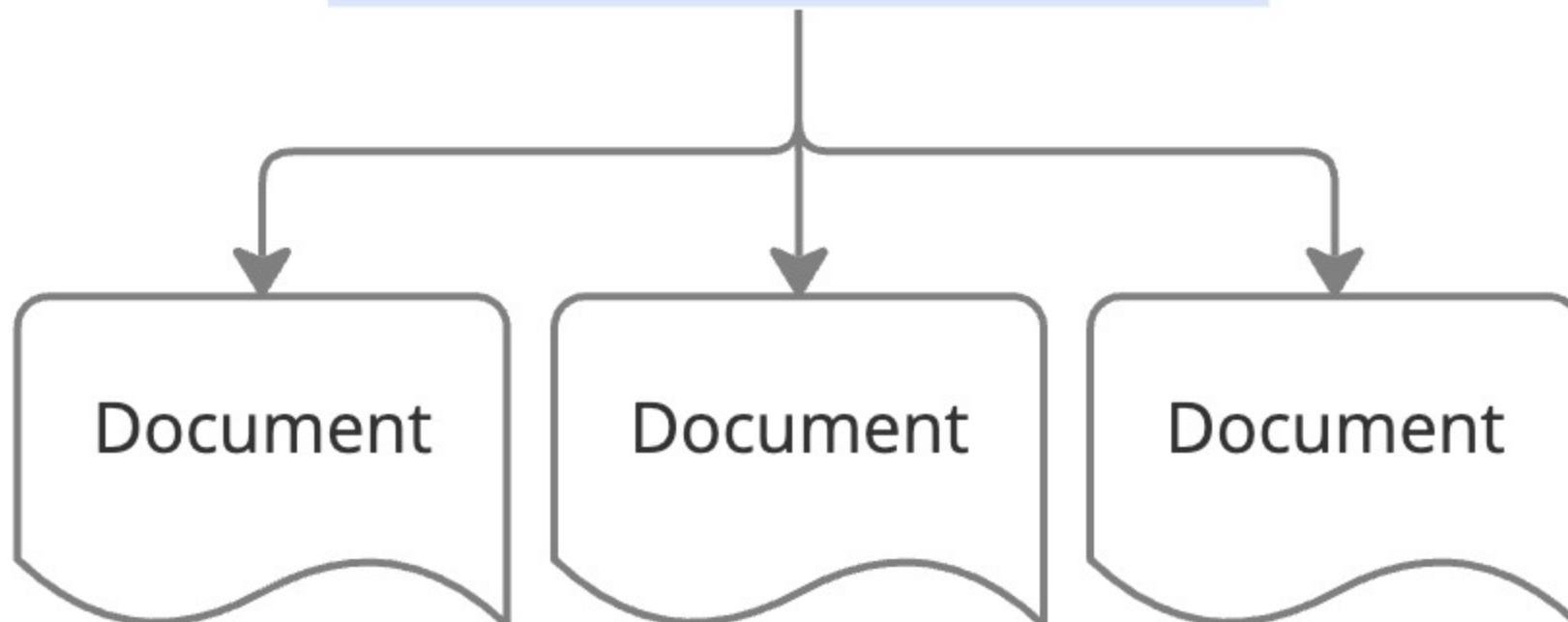
Create embedding out of the user's question

Do a similarity search with our stored embeddings to find the ones most similar to the user's question

Put the most relevant 1-3 facts into the prompt along with the user's question

*What is an interesting
fact about the English
language?*

Retriever



A retriever is an object that can take in a string and return some relevant documents

To be a "Retriever", the object must have a method called "get_relevant_documents" that takes a string and returns a list of documents

chroma

similarity_search(string)

as_retriever()

retriever

get_relevant_documents(string)



What is an interesting fact about the English language?

Vector Store

ChatPromptTemplate

SystemMessagePromptTemplate

Use the following interesting fact to answer the users question: {fact}

HumanMessagePromptTemplate

Here is the user's question:
{question}

LLMChain

We **could** set this all up manually

But LangChain already does it!

RetrievalQA

What is an interesting fact about the English language?

Vector Store

- 1. "Dreamt" is the only English word that...
- 2. An ostrich's eye is bigger than....
- 3. Honey is the only natural food...

- 4. A snail can sleep for three years.
- 5. The longest word in the English language...
- 6. The elephant is the only mammal...

- 86. Broccoli and cauliflower ...
- 87. The dot over an 'i' or 'j' is ...
- 88. A group of owls is called a...

- 118. The original Star-Spangled Banner...
- 119. The average adult spends more time...

SystemMessagePromptTemplate

Use the following interesting fact to answer the users question: {fact}

HumanMessagePromptTemplate

Here is the user's question:
{question}

LLMChain

chain_type="stuff"

Take some context from the vector store and "stuff" it into the prompt

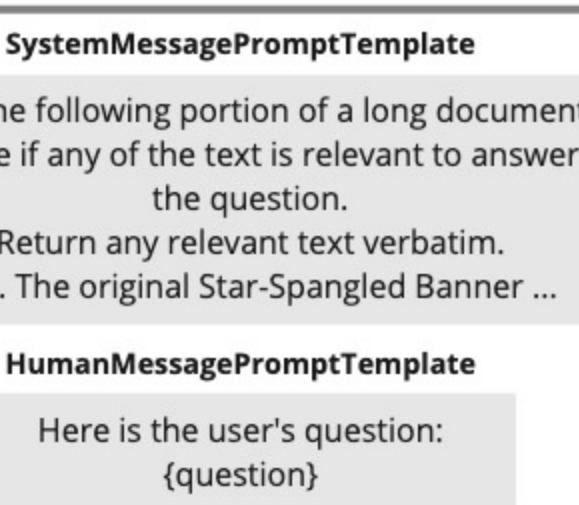
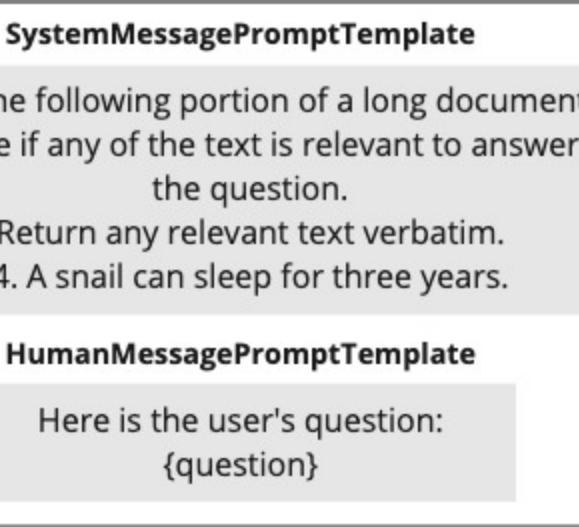
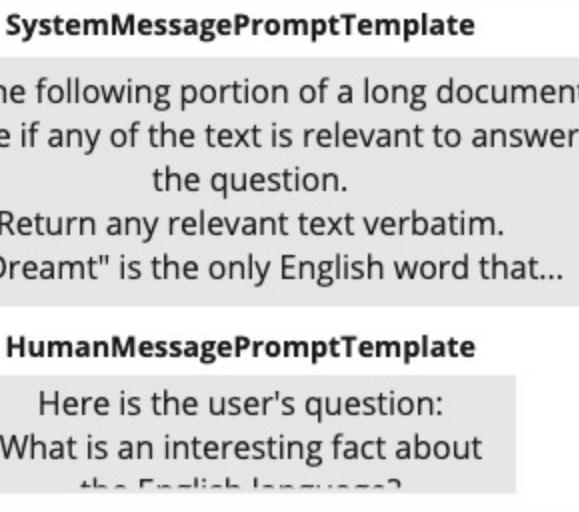
What is an interesting fact about the English language?

Vector Store

- 1. "Dreamt" is the only English word that...
- 2. An ostrich's eye is bigger than....
- 3. Honey is the only natural food...

- 4. A snail can sleep for three years.
- 5. The longest word in the English language...
- 6. The elephant is the only mammal...

- 118. The original Star-Spangled Banner ...
- 119. The average adult spends more time...

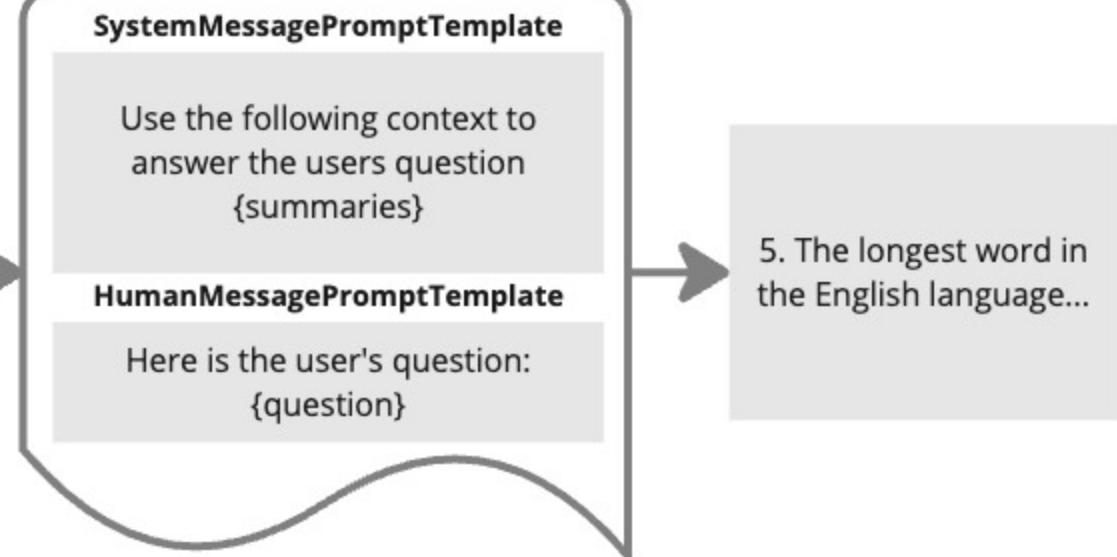


1. "Dreamt" is the only English word that...

5. The longest word in the English language...

English is the third most spoken language in the world

!!!!



5. The longest word in the English language...

chain_type=
"map_reduce"

Build a summary of each document, then feed each summary into a final question

What is an interesting fact about the English language?

Vector Store

- 1. "Dreamt" is the only English word that...
- 2. An ostrich's eye is bigger than....
- 3. Honey is the only natural food...

- 4. A snail can sleep for three years.
- 5. The longest word in the English language...
- 6. The elephant is the only mammal...

- 86. Broccoli and cauliflower ...
- 87. The dot over an 'i' or 'j' is ...
- 88. A group of owls is called a...

HumanMessagePromptTemplate
Use the following pieces of context to answer the question at the end. In addition to giving an answer, also return a score of how fully it answered the user's question.
{content}

HumanMessagePromptTemplate
Here is the user's question:
{question}

SystemMessagePromptTemplate
Use the following pieces of context to answer the question at the end. In addition to giving an answer, also return a score of how fully it answered the user's question.
{content}

HumanMessagePromptTemplate
Here is the user's question:
{question}

SystemMessagePromptTemplate
Use the following pieces of context to answer the question at the end. In addition to giving an answer, also return a score of how fully it answered the user's question.
{content}

HumanMessagePromptTemplate
Here is the user's question:
{question}

Score: 90
1. "Dreamt" is the only English word that...

Score: 50
5. The longest word in the English language...

Score: 100
English is the third most spoken language in the world

!!!!

Find highest score

1. "Dreamt" is the only English word that...

chain_type=
"map_rerank"

Find relevant part of each document and give it a score of how relevant it is.

What is an interesting fact about the English language?

Vector Store

1. "Dreamt" is the only English word that...
2. An ostrich's eye is bigger than....
3. Honey is the only natural food...

4. A snail can sleep for three years.
5. The longest word in the English language...
6. The elephant is the only mammal...

86. Broccoli and cauliflower ...
87. The dot over an 'i' or 'j' is ...
88. A group of owls is called a...

SystemMessagePromptTemplate
Use the following context to answer the question:
{context}

HumanMessagePromptTemplate
Here is the user's question:
{question}

1. "Dreamt" is the only English word that...

HumanMessagePromptTemplate
Here is the user's question:
{question}

AIMessagePromptTemplate
1. "Dreamt" is the only English word that...

HumanMessagePromptTemplate
We have a chance to refine the answer using this additional context:
4. A snail can sleep for three years.
5. The longest word in the English language...
6. The elephant is the only mammal...

1. "Dreamt" is the only English word that...

HumanMessagePromptTemplate
Here is the user's question:
{question}

AIMessagePromptTemplate
1. "Dreamt" is the only English word that...

HumanMessagePromptTemplate
We have a chance to refine the answer using this additional context:
86. Broccoli and cauliflower ...

The added context was not relevant

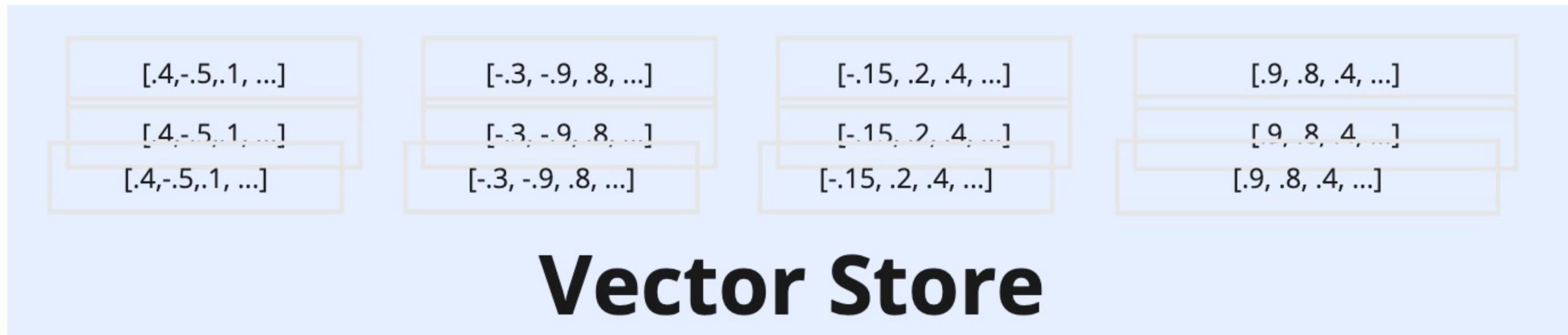
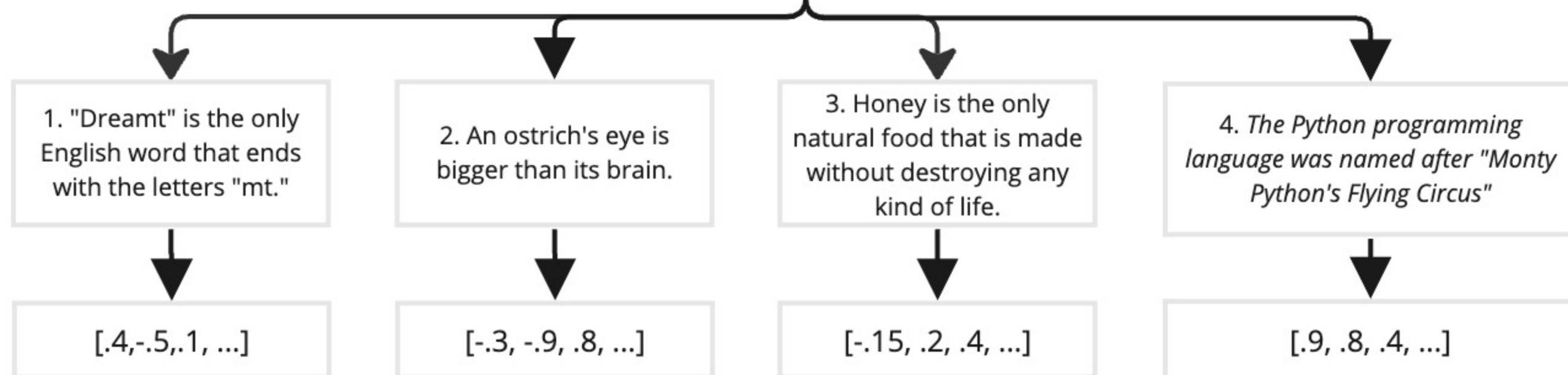
chain_type=
"refine"

Build an initial response, then give the LLM an opportunity to update it with further context

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

....Many more.....



Vector Store

User's Query

What is the Krebs cycle?

Biology Textbook

Relevant Doc #1

The Krebs cycle is a series of chemical reactions used by all aerobic organisms to **release stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins into ATP and carbon dioxide.**

Relevant Doc #2

The Krebs cycle is a crucial metabolic pathway that **releases stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins into ATP and carbon dioxide.**

Docs are repetitive - more diversity would be better

facts.txt

1. "Dreamt" is the only English word that ends with the letters "mt."
2. An ostrich's eye is bigger than its brain.
3. Honey is the only natural food that is made without destroying any kind of life.

1. "Dreamt" is the only English word that ends with the letters "mt."

1. "Dreamt" is the only English word that ends with the letters "mt."

3. Honey is the only natural food that is made without destroying any kind of life.

3. Honey is the only natural food that is made without destroying any kind of life.

[.4,-.5,.1, ...]

[-.15, .2, .4, ...]

EmbeddingsRedundantFilter

1. "Dreamt" is the only English word that ends with the letters "mt."

3. Honey is the only natural food that is made without destroying any kind of life.

**What is an interesting
fact about the English
language?**



RetrievalQA Chain



chroma retriever

`get_relevant_documents(str): [Document]`

**What is an interesting
fact about the English
language?**



RetrievalQA Chain



Won't work

EmbeddingsRedundantFilter



chroma retriever

`get_relevant_documents(str): [Document]`

We can't easily insert or
inject the
EmbeddingsRedundant
Filter into this process

**What is an interesting
fact about the English
language?**



RetrievalQA Chain



custom retriever

```
1 class CustomFilterRetriever:  
2     def get_relevant_documents(self, query):  
3         # Code to use Chroma to find relevant docs  
4         # and remove any duplicate records
```

We need to build a custom retriever

A retriever is a class that has a "get_relevant_documents" function that receives a string and returns a list of documents

Our function will take in the user's query, calculate the strings embeddings, find relevant documents in Chroma, remove duplicates, and return the result

**Reminder: to calculate embeddings manually, we can use
the "embed_query" function**

```
1 from langchain.embeddings import OpenAIEmbeddings
2
3 embeddings = OpenAIEmbeddings()
4
5 result = embeddings.embed_query("hi there")
6
7 print(result) # [-.3, .4, .002, ...]
```

Reminder: to find similar documents, we can use Chroma's "similarity_search()" function

```
1 from langchain.vectorstores import Chroma
2
3 embeddings = OpenAIEmbeddings()
4 db = Chroma(persist_directory="emb", embedding_function=embeddings)
5
6 result = db.similarity_search("what is an interesting fact about...")
7
8 print(result) # List of relevant documents
```

New thing! If you want to find similarities to an embedding you already calculated, you can use the "similarity_search_by_vector" function instead

```
1 from langchain.vectorstores import Chroma
2
3 embeddings = OpenAIEmbeddings()
4 db = Chroma(persist_directory="emb", embedding_function=embeddings)
5
6 # Calculate embeddings for the user's query
7 emb = embeddings.embed_query("what is an interesting fact about...")
8
9 # Find similar documents by using the embedding we just calculated
10 results = db.similarity_search_by_vector(emb)
11
12 print(results) # List of relevant documents
```

New thing! Turns out Chroma can remove duplicates for us automatically using the "max_marginal_relevance_search_by_vector" function!

```
1 from langchain.vectorstores import Chroma
2
3 embeddings = OpenAIEmbeddings()
4 db = Chroma(persist_directory="emb", embedding_function=embeddings)
5
6 # Calculate embeddings for the user's query
7 emb = embeddings.embed_query("what is an interesting fact about...")
8
9 # Find similar documents by using the embedding we just calculated
10 results = db.max_marginal_relevance_search_by_vector(
11     embedding=emb,
12     lambda_mult=0.8 # 0 to 1. Higher values allow for similar docs
13 )
14
15 print(results) # List of relevant documents
```

To make a custom retriever, we can define a class that extends "BaseRetriever". Our retriever must define "get_relevant_documents" and "aget_relevant_documents"

```
1 from langchain.schema import BaseRetriever  
2  
3 class RedundantFilterRetriever(BaseRetriever):  
4     def get_relevant_documents(self, query):  
5         # must return a list of documents  
6         return []  
7  
8     async def aget_relevant_documents(self):  
9         # must return a list of documents  
10        return []
```