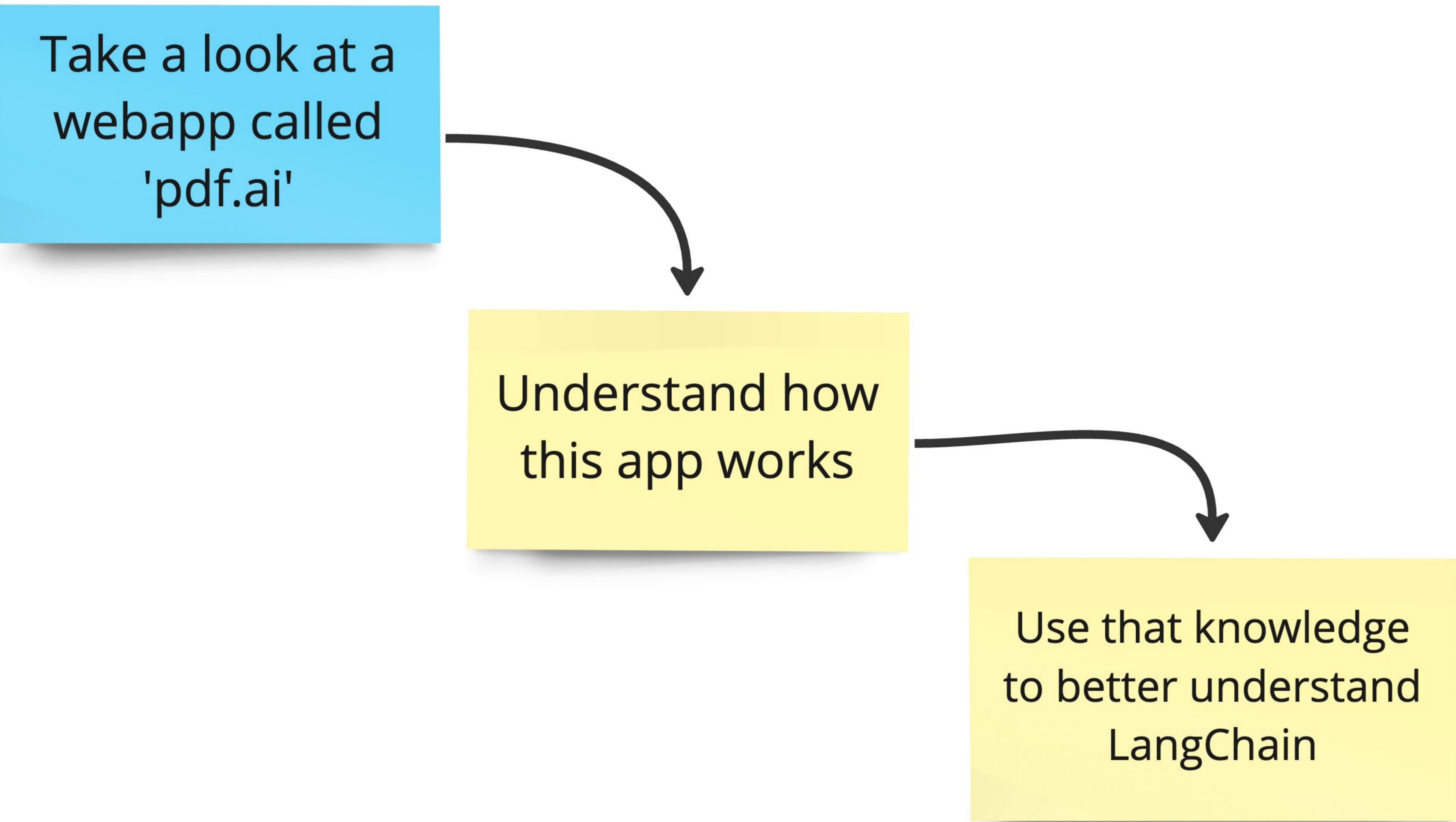


So....What is  
LangChain?

Take a look at a  
webapp called  
'pdf.ai'



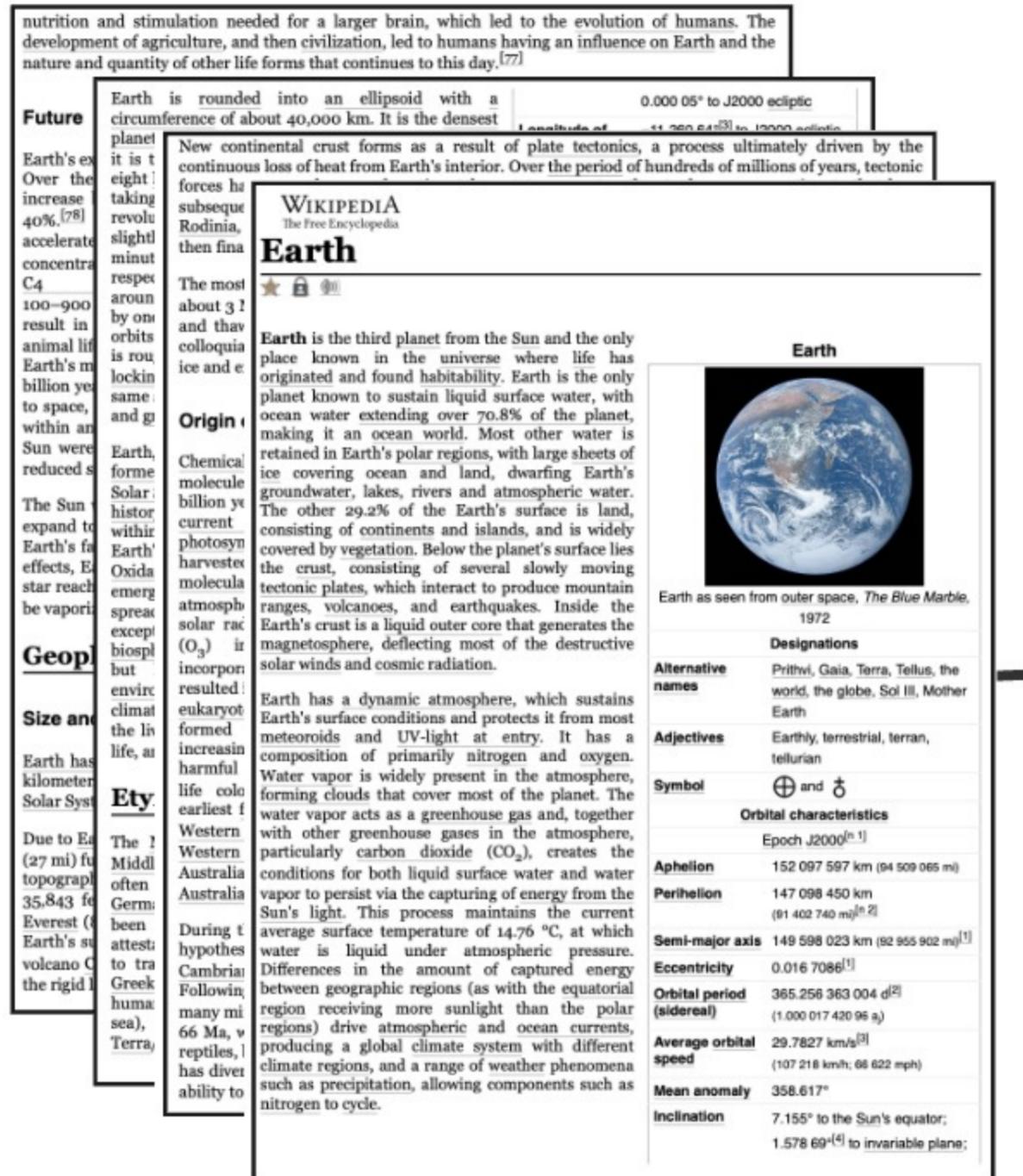
Understand how  
this app works

Use that knowledge  
to better understand  
LangChain

### User's Question

Where does the word  
"Earth" come from?

All the text from the PDF



**Option #1 for answering a user's question**  
Take ALL the text from the PDF + the users  
question and send it off to ChatGPT

ChatGPT

## **Option #1 for answering a user's question**

Take ALL the text from the PDF + the users question and send it off to ChatGPT

Probably a bad idea

Can only send a *limited* amount of text to ChatGPT

Even if we could send an unlimited amount, text models usually don't do well with a ton of text

more text = more \$ to run the request

## Option #2 for answering a user's question

1

When a user uploads a pdf, extract all the text, divide it into chunks

2

Store a summary of what each chunk is talking about

3

Later, when a user asks a question, find the chunk of text *most relevant* to the user's question

4

Send the most relevant chunk + the users question to ChatGPT

# Earth



**Earth** is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

Earth has a dynamic atmosphere, which sustains Earth's surface conditions and protects it from most meteoroids and UV-light at entry. It has a composition of primarily nitrogen and oxygen. Water vapor is widely present in the atmosphere, forming clouds that cover most of the planet. The water vapor acts as a greenhouse gas and, together with other greenhouse gases in the atmosphere, particularly carbon dioxide ( $\text{CO}_2$ ), creates the conditions for both liquid surface water and water vapor to persist via the capturing of energy from the Sun's light. This process maintains the current average surface temperature of 14.76 °C, at which water is liquid under atmospheric pressure. Differences in the amount of captured energy between geographic regions (as with the equatorial region receiving more sunlight than the polar regions) drive atmospheric and ocean currents, producing a global climate system with different climate regions, and a range of weather phenomena such as precipitation, allowing components such as nitrogen to cycle.



Earth as seen from outer space, *The Blue Marble*, 1972

## Designations

**Alternative names** Prithvi, Gaia, Terra, Tellus, the world, the globe, Sol III, Mother Earth

**Adjectives** Earthly, terrestrial, terran, tellurian

**Symbol**  $\oplus$  and  $\odot$

## Orbital characteristics

Epoch J2000<sup>[n 1]</sup>

**Aphelion** 152 097 597 km (94 509 065 mi)

**Perihelion** 147 098 450 km  
(91 402 740 mi)<sup>[n 2]</sup>

**Semi-major axis** 149 598 023 km (92 955 902 mi)<sup>[1]</sup>

**Eccentricity** 0.016 7086<sup>[1]</sup>

**Orbital period (sidereal)** 365.256 363 004 d<sup>[2]</sup>  
(1.000 017 420 96 a.)

**Average orbital speed** 29.7827 km/s<sup>[3]</sup>  
(107 218 km/h; 66 622 mph)

**Mean anomaly** 358.617°

**Inclination** 7.155° to the Sun's equator;

# All Text from the PDF is extracted

## Earth

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

Earth has a dynamic atmosphere, which sustains Earth's surface conditions and protects it from most meteoroids and UV-light at entry. It has a composition of primarily nitrogen and oxygen. Water vapor is widely present in the atmosphere, forming clouds that cover most of the planet. The water vapor acts as a greenhouse gas and, together with other greenhouse gases in the atmosphere, particularly carbon dioxide ( $\text{CO}_2$ ), creates the conditions for both liquid surface water and water vapor to persist via the capturing of energy from the Sun's light. This process maintains the current average surface temperature of 14.76 °C, at which water is liquid under atmospheric pressure. Differences in the amount of captured energy between geographic regions (as with the equatorial region receiving more sunlight than the polar regions) drive atmospheric and ocean currents, producing a global climate system with different climate regions, and a range of weather phenomena such as precipitation, allowing components such as nitrogen to cycle.

## Earth

Earth as seen from outer space, *The Blue Marble*, 1972

## Designations

Alternative Prithvi, Gaia, Terra, Tellus, the world, the globe, Sol III, Mother Earth

Adjectives Earthly, terrestrial, terran, tellurian

## Symbol

## Orbital characteristics

Epoch J2000<sup>[n 1]</sup>

Aphelion 152 097 597 km (94 509 065 mi)

Perihelion 147 098 450 km  
(91 402 740 mi)<sup>[n 2]</sup>

Semi-major axis 149 598 023 km (92 955 902 mi)<sup>[1]</sup>

Eccentricity 0.016 7086<sup>[1]</sup>

Orbital period (sidereal) 365.256 363 004 d<sup>[2]</sup>  
(1.000 017 420 96 a.)

Average orbital speed 29.7827 km/s<sup>[3]</sup>  
(107 218 km/h; 66 622 mph)

Mean anomaly 358.617°

Inclination 7.155° to the Sun's equator;

## Text Chunk #1

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

Earth has a dynamic atmosphere, which sustains Earth's surface conditions and protects it from most meteoroids and UV-light at entry. It has a composition of primarily nitrogen and oxygen. Water vapor is widely present in the atmosphere, forming clouds that cover most of the planet. The water vapor acts as a greenhouse gas and, together with other greenhouse gases in the atmosphere, particularly carbon dioxide (CO<sub>2</sub>), creates the

conditions for both liquid surface water and water vapor to persist via the capturing of energy from the Sun's light. This process maintains the current average surface temperature of 14.76 °C, at which water is liquid under atmospheric pressure. Differences in the amount of captured energy between geographic regions (as with the equatorial region receiving more sunlight than the polar regions) drive atmospheric and ocean currents, producing a global climate system with different climate regions, and a range of weather phenomena such as precipitation, allowing components such as nitrogen to cycle.

### Earth

Earth as seen from outer space, The Blue Marble, 1972

#### Designations

Alternative Prithvi, Gaia, Terra, Tellus, the names

world, the globe, Sol III, Mother Earth

Adjectives Earthly, terrestrial, terran, tellurian

#### Symbol and

#### Orbital characteristics

Epoch J2000[n 1]

AJ 2000 = 152 007 507 Julian (21 500 000.000)

## Text Chunk #2

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

*This is repeated for all the  
text in the entire pdf*

## Text Chunk #1

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

## Text Chunk #2

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

## Embedding Creation Algorithm

## Embedding For Chunk #1

*Data that describes the raw essence of what chunk #1 is about*

"I feel  
great!"

## Embedding Creation Algorithm

[0.9, -.84, .71, 0.9, .....]

This is the "embedding", an array  
of numbers, all between -1 and 1

# Embeddings

Array of numbers that describes the absolute fundamental essence of what some text is about

*Exactly 1536 numbers long for the encoding scheme we will be using*

"I feel great!"

[0.9, -0.84, 0.71, 0.9, .....]

Might be a score of how 'happy' the text is

Might be a score of how much the text is talking about potatoes

Might be a score of whether the text is talking about hiking a mountain

"I feel  
great!"

→ [0.9, -.84, .71, 0.9, .....]

"I am sad"

→ [0.9, 0.1, -0.94, 0.4, .....]

"The sun tastes  
like jellybeans"

→ [0.23, -0.26, .14, -0.9, .....]

## Text Chunk #1

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

## Text Chunk #2

Earth is the third planet from the Sun and the only place known in the universe where life has originated and found habitability. Earth is the only planet known to sustain liquid surface water, with ocean water extending over 70.8% of the planet, making it an ocean world. Most other water is retained in Earth's polar regions, with large sheets of ice covering ocean and land, dwarfing Earth's groundwater, lakes, rivers and atmospheric water. The other 29.2% of the Earth's surface is land, consisting of continents and islands, and is widely covered by vegetation. Below the planet's surface lies the crust, consisting of several slowly moving tectonic plates, which interact to produce mountain ranges, volcanoes, and earthquakes. Inside the Earth's crust is a liquid outer core that generates the magnetosphere, deflecting most of the destructive solar winds and cosmic radiation.

## Text Analysis Algorithm

## Embedding For Chunk #1

*Data that describes the raw essence of what chunk #1 is about*

## Embedding For Chunk #2

*Data that describes the raw essence of what chunk #2 is about*



# Vector Store

Embedding For Chunk #1

Embedding For Chunk #2

Embedding For Chunk #3

Embedding For Chunk #4

Embedding For Chunk #5

Special database  
made for storing  
embeddings



Remember, an 'embedding' is  
a special array of numbers  
that encodes what a chunk of  
text is talking about

*Question I submitted*

Where does the  
word 'earth' come  
from?

PDF.AI Server

Embedding  
Algorithm

[0.34, -0.15, 0.4, 0.9, ....]



Where does the word 'earth' come from?

**[0.34, -0.15, 0.4, 0.9, .....]**

**How similar is this text  
to all of the text  
chunks?**

Embedding For Chunk #3

## Vector Store

Embedding For Chunk #1

Embedding For Chunk #2

Embedding For Chunk #4

Embedding For Chunk #5

# Super Relevant Text From the PDF

## Question from User

Where does the word 'earth' come from?

### Chunk #3

#### Etymology

The Modern English word Earth developed, via Middle English, from an Old English noun most often spelled eorðe. It has cognates in every Germanic language, and their ancestral root has been reconstructed as \*erbō. In its earliest attestation, the word eorðe was already being used to translate the many senses of Latin *terra* and Greek γῆ gē: the ground, its soil, dry land, the human world, the surface of the world (including the sea), and the globe itself.

## *Question from User*

Where does the word 'earth' come from?

## *Super Relevant Text From the PDF*

### **Chunk #3**

#### **Etymology**

The Modern English word Earth developed, via Middle English, from an Old English noun most often spelled eorðe. It has cognates in every Germanic language, and their ancestral root has been reconstructed as \*erþō.

In its earliest attestation, the word eorðe was already being used to translate the many senses of Latin terra and Greek γῆ: the ground, its soil, dry

## *Prompt to Send to ChatGPT*

Dear ChatGPT,

Did you know that the Modern English word Earth developed, via Middle English, from an Old English noun most often spelled eorðe.?

Also, Where does the word 'earth' come from?

*Prompt to Send to ChatGPT*

Dear ChatGPT,

Did you know that the Modern English word Earth developed, via Middle English, from an Old English noun most often spelled eorðe.?

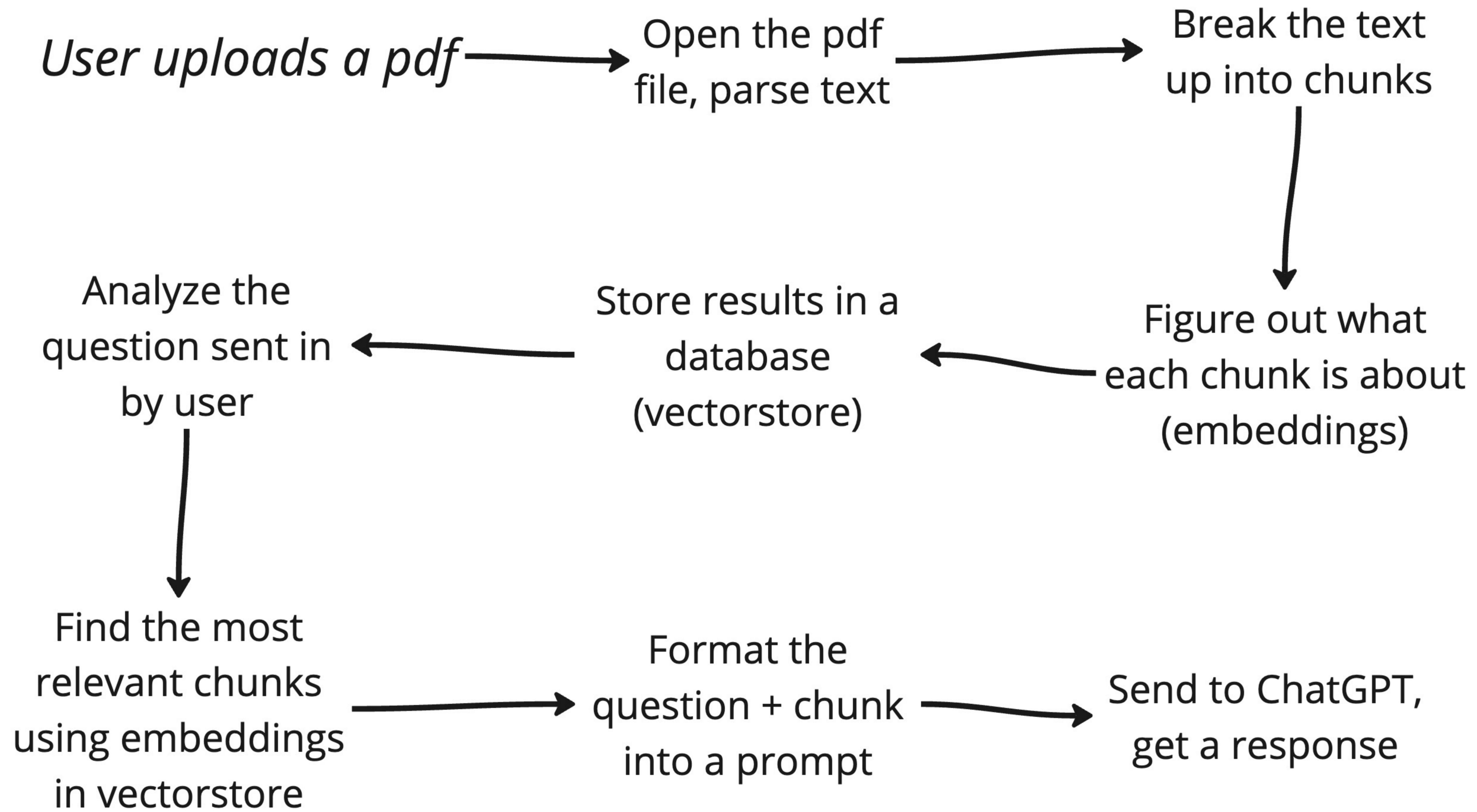
Also, Where does the word 'earth' come from?

*Answer From ChatGPT*

**Old English noun most often spelled eorðe.**

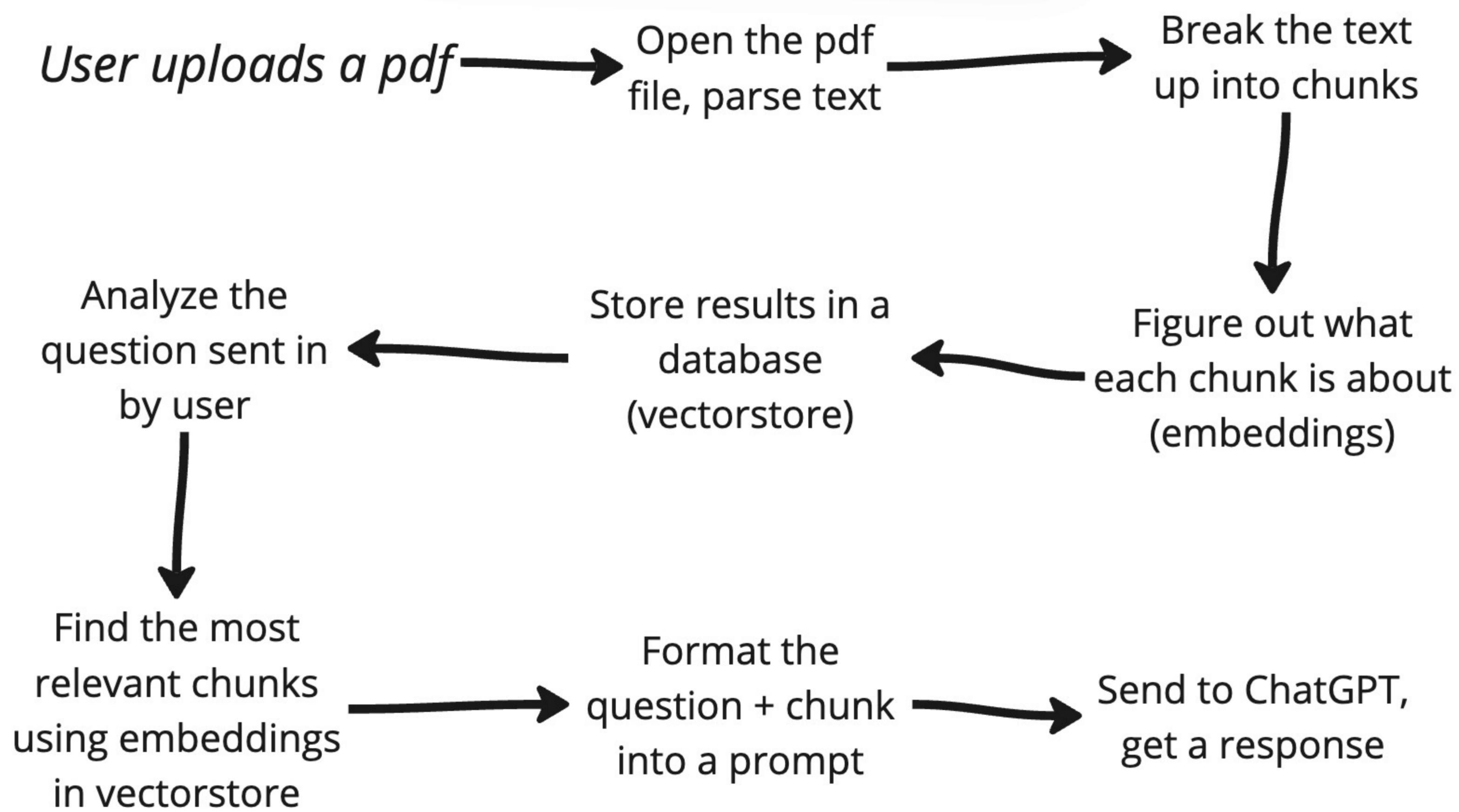
Wow chatgpt  
is amazing

*us*



# Langchain's Goal:

Provide tools to  
automate each of  
these steps



*User uploads a pdf* →

**Open the pdf file,  
parse text**

```
1 from langchain.document_loaders import UnstructuredPDFLoader
2 loader = UnstructuredPDFLoader("earth.pdf")
3 data = loader.load()
4
5 print(data) # Document(page_content="Earth is a planet...")
```

Class that makes loading +  
parsing a PDF really easy

*User uploads a pdf* →

Open the pdf file,  
parse text

UnstructuredPDFLoader

PyPDFLoader

JSONLoader

UnstructuredHTMLLoader

S3FileLoader

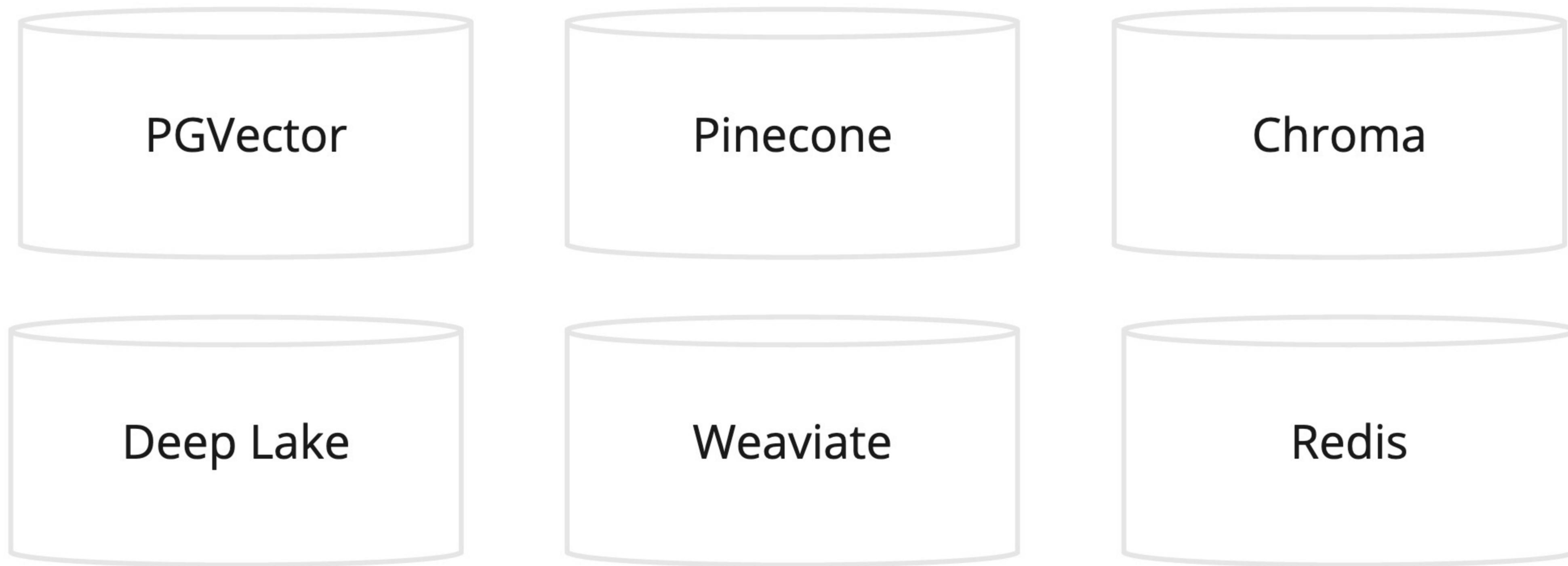
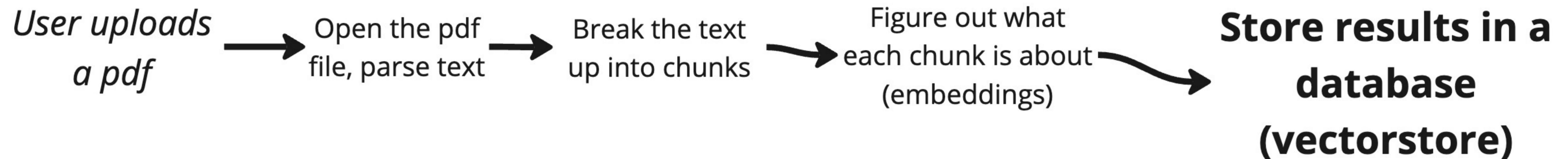
DiscordChatLoader

UnstructuredExcelLoader

GitHubIssuesLoader

GoogleDriveLoader

**Classes that make it  
easy to load data  
from *anywhere***



**There are a *ton* of popular vector databases. We could spend a ton of time figuring out how each one works...**

....Or we could use some classes provided by Langchain that completely wrap up the implementation details *and are interchangeable*

## PGVector

```
1  from langchain.vectorstores.pgvector import PGVector
2
3  db = PGVector.from_documents(
4      embedding=embeddings,
5      documents=docs,
6      collection_name="earth",
7      connection_string=CONNECTION_STRING,
8  )
9
10 query = "Where does the word earth come from?"
11 docs = db.similarity_search(query)
```

## Redis

```
1  from langchain.vectorstores.redis import Redis
2
3  rds = Redis.from_documents(
4      embeddings=embeddings
5      documents=docs,
6      redis_url="redis://localhost:6379",
7      index_name="earth"
8  )
9
10 query = "Where does the word earth come from?"
11 docs = db.similarity_search(query)
```

## Pinecone

```
1  from langchain.vectorstores.pinecone import Pinecone
2
3  db = Pinecone.from_documents(
4      embedding=embeddings,
5      documents=docs,
6      index_name="earth"
7  )
8
9  query = "Where does the word earth come from?"
10 docs = db.similarity_search(query)
```

## Weaviate

```
1  from langchain.vectorstores.weaviate import Weaviate
2
3  db = Pinecone.from_documents(
4      embedding=embeddings,
5      documents=docs,
6      weaviate_url=weaviate_url
7  )
8
9  query = "Where does the word earth come from?"
10 docs = db.similarity_search(query)
```

# Goal of LangChain

Provide *interchangeable* tools to automate each step of a text generation pipeline.

