



Modules

Retrieval

Text Splitters

Semantic Chunking

Semantic Chunking

Splits the text based on semantic similarity.

Taken from Greg Kamradt's wonderful notebook:

https://github.com/FullStackRetrieval-com/RetrievalTutorials/blob/main/5_Levels_Of_Text_Splitting.ipynb

All credit to him.

At a high level, this splits into sentences, then groups into groups of 3 sentences, and then merges one that are similar in the embedding space.

Install Dependencies

```
!pip install --quiet langchain_experimental  
langchain_openai
```

Load Example Data

```
# This is a long document we can split up.  
with open("../../state_of_the_union.txt") as  
f:  
    state_of_the_union = f.read()
```

Create Text Splitter

```
from langchain_experimental.text_splitter  
import SemanticChunker  
from langchain_openai.embeddings import  
OpenAIEmbeddings
```

```
text_splitter =  
SemanticChunker(OpenAIEmbeddings())
```

Split Text

```
docs =  
text_splitter.create_documents([state_of_the_un.  
print(docs[0].page_content)
```

Madam Speaker, Madam Vice President, our First Lady and Second Gentleman. Members of Congress and the Cabinet. Justices of the Supreme Court. My fellow Americans. Last year COVID-19 kept us apart. This year we are finally together again. Tonight, we meet as Democrats Republicans and Independents. But most importantly as Americans. With a duty to one another to the American people to the Constitution. And with an unwavering resolve that freedom will always triumph over tyranny. Six days ago, Russia's Vladimir Putin sought to shake the foundations of the free world thinking he could make it bend to his menacing ways. But he badly miscalculated. He thought he could roll into Ukraine and the world would roll over. Instead he met a wall of strength he never imagined. He met the Ukrainian people. From President Zelenskyy to every Ukrainian, their fearlessness, their courage, their determination, inspires the world. Groups of citizens blocking tanks with their bodies. Everyone from students to retirees teachers turned soldiers defending their homeland. In

this struggle as President Zelenskyy said in his speech to the European Parliament “Light will win over darkness.” The Ukrainian Ambassador to the United States is here tonight. Let each of us here tonight in this Chamber send an unmistakable signal to Ukraine and to the world. Please rise if you are able and show that, Yes, we the United States of America stand with the Ukrainian people. Throughout our history we’ve learned this lesson when dictators do not pay a price for their aggression they cause more chaos. They keep moving.