

[Print to PDF](#)

Character Text Splitter

This is a more simple method. This splits based on characters (by default “\n\n”) and measure chunk length by number of characters.

1. How the text is split: by single character
2. How the chunk size is measured: by length function passed in (defaults to number of characters)

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

```
from langchain.text_splitter import CharacterTextSplitter
text_splitter = CharacterTextSplitter(
    separator = "\n\n",
    chunk_size = 1000,
    chunk_overlap = 200,
    length_function = len,
)
```

```
texts = text_splitter.create_documents([state_of_the_union])
print(texts[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. \n\nLast year COVID-19 kept us apart. This year we are finally together again. \n\nTonight, we meet as Democrats Republicans and Independents. But most importantly as Americans. \n\nWith a duty to one another to the American people to the Constitution. \n\nAnd with an unwavering resolve that freedom will always triumph over tyranny. \n\nSix 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. \n\nHe thought he could roll into Ukraine and the world would roll over. Instead he met a wall of strength he never imagined. \n\nHe met the Ukrainian people. \n\nFrom President Zelenskyy to every Ukrainian, their fearlessness, their courage, their determination, inspires the world.' lookup_str='' metadata={} lookup_index=0
```

[Skip to main content](#)

Here's an example of passing metadata along with the documents, notice that it is split along with the documents.

```
metadatas = [{"document": 1}, {"document": 2}]
documents = text_splitter.create_documents([state_of_the_union,
state_of_the_union], metadatas=metadatas)
print(documents[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. \n\nLast year COVID-19 kept us apart. This year we are finally
together again. \n\nTonight, we meet as Democrats Republicans and Independents.
But most importantly as Americans. \n\nWith a duty to one another to the American
people to the Constitution. \n\nAnd with an unwavering resolve that freedom will
always triumph over tyranny. \n\nSix 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. \n\nHe thought he could roll into
Ukraine and the world would roll over. Instead he met a wall of strength he never
imagined. \n\nHe met the Ukrainian people. \n\nFrom President Zelenskyy to every
Ukrainian, their fearlessness, their courage, their determination, inspires the
world.' lookup_str='' metadata={'document': 1} lookup_index=0
```