





# Mastering RAG

## Recursive Character Text Splitting

Splitter: Recursive Character Text Splitter   

Chunk Size:  

Chunk Overlap:  





Total Characters: 1366

Number of chunks: 4

Average chunk size: 341.5

The Marvel Universe is a vast and interconnected world filled with superheroes, villains, and epic storytelling that has captivated audiences for decades. Founded by visionaries such as Stan Lee, Jack Kirby, and Steve Ditko, Marvel Comics has introduced some of the most iconic characters in pop culture history. From its early beginnings in 1939 as Timely Publications to its transformation into Marvel Comics in the 1960s, the company has consistently pushed the boundaries of storytelling by creating relatable and dynamic characters. Heroes like Spider-Man, Iron Man, Captain America, and Thor have become household names, each with their own compelling backstories and struggles that resonate with fans across generations. Marvel's success extends beyond the pages of comic books. The launch of the Marvel Cinematic Universe (MCU) in 2008 with the release of Iron Man revolutionized the film industry, introducing interconnected storylines that culminated in epic crossover events such as The Avengers and Infinity War. The MCU's success is largely attributed to its ability to blend action, humor, and emotional depth while maintaining the essence of the beloved comic book characters. Audiences have followed the journeys of superheroes as they face powerful foes like Thanos and Loki, all while dealing with their own internal conflicts and responsibilities.



Splitter: Recursive Character Text Splitter   Chunk Size: 400 Chunk Overlap: 0 

Total Characters: 1366

Number of chunks: 4

Average chunk size: 341.5

The Marvel Universe is a vast and interconnected world filled with superheroes, villains, and epic storytelling that has captivated audiences for decades. Founded by visionaries such as Stan Lee, Jack Kirby, and Steve Ditko, Marvel Comics has introduced some of the most iconic characters in pop culture history. From its early beginnings in 1939 as Timely Publications to its transformation into Marvel Comics in the 1960s, the company has consistently pushed the boundaries of storytelling by creating relatable and dynamic characters. Heroes like Spider-Man, Iron Man, Captain America, and Thor have become household names, each with their own compelling backstories and struggles that resonate with fans across generations. Marvel's success extends beyond the pages of comic books. The launch of the Marvel Cinematic Universe (MCU) in 2008 with the release of Iron Man revolutionized the film industry, introducing interconnected storylines that culminated in epic crossover events such as The Avengers and Infinity War. The MCU's success is largely attributed to its ability to blend action, humor, and emotional depth while maintaining the essence of the beloved comic book characters. Audiences have followed the journeys of superheroes as they face powerful foes like Thanos and Loki, all while dealing with their own internal conflicts and responsibilities.



Unlike the first method which doesn't look for the document structure, this method recursively divides text using a predefined list of separators and intelligently merges the resulting smaller chunks into larger ones. The final chunks are optimized to contain no more than N characters, ensuring efficient text processing and context preservation.

It is parameterized by a list of characters.

The default list is:

- “\n\n” – Double new line, or most commonly paragraph breaks
- “\n” – New lines
- “ ” – Spaces
- “” – Characters



```
%pip install -qU langchain-text-splitters
```

```
text = """
```

```
The Marvel Universe is a vast and interconnected world filled with  
superheroes, villains, and epic storytelling that has captivated  
audiences for decades. Founded by visionaries such as Stan Lee, Jack  
Kirby, and Steve Ditko, Marvel Comics has introduced some of the most  
iconic characters in pop culture history. From its early beginnings in  
1939 as Timely Publications to its transformation into Marvel Comics in  
the 1960s, the company has consistently pushed the boundaries of  
storytelling by creating relatable and dynamic characters. Heroes like  
Spider-Man, Iron Man, Captain America, and Thor have become household  
names, each with their own compelling backstories and struggles that  
resonate with fans across generations. Marvel's success extends beyond  
the pages of comic books. The launch of the Marvel Cinematic Universe  
(MCU) in 2008 with the release of Iron Man revolutionized the film  
industry, introducing interconnected storylines that culminated in epic  
crossover events such as The Avengers and Infinity War. The MCU's  
success is largely attributed to its ability to blend action, humor, and  
emotional depth while maintaining the essence of the beloved comic book  
characters. Audiences have followed the journeys of superheroes as they  
face powerful foes like Thanos and Loki, all while dealing with their  
own internal conflicts and responsibilities."""
```

```
from langchain_text_splitters import RecursiveCharacterTextSplitter
```

The RecursiveCharacterTextSplitter is imported from the langchain-text-splitters package.





This class is used to split large text documents into smaller chunks efficiently while preserving context.

```
text_splitter = RecursiveCharacterTextSplitter(  
    # Set a really small chunk size, just to show.  
    chunk_size=400,  
    chunk_overlap=0,  
    length_function=len,  
)  
text_splitter.create_documents([text])
```

## Output

[Document(metadata={}, page\_content='The Marvel Universe is a vast and interconnected world filled with superheroes, villains, and epic storytelling that has captivated audiences for decades. Founded by visionaries such as Stan Lee, Jack Kirby, and Steve Ditko, Marvel Comics has introduced some of the most iconic characters in pop'),

...



```
Document(metadata={}, page_content='film industry,  
introducing  
interconnected storylines that culminated in epic  
crossover events such as  
The Avengers and Infinity War. The MCU's success is  
largely attributed to  
its ability to blend action, humor, and emotional depth  
while maintaining  
the essence of the beloved comic book characters.')
```

```
Document(metadata={}, page_content='Audiences have  
followed the journeys of  
superheroes as they face powerful foes like Thanos  
and Loki, all while  
dealing with their own internal conflicts and  
responsibilities.')] ]
```

The resulting list of Document objects contains multiple chunks of the text, each with overlapping portions to ensure smooth transitions. Here's a breakdown of the output:

### **First Chunk:**

“The Marvel Universe is a vast and interconnected world filled with superheroes, ... iconic characters in pop”

### **Second Chunk:**

“culture history. From its early beginnings in 1939 as Timely Publications to its transformation into Marvel Comics in the 1960s, ... Iron Man, Captain America, and”

### **Third Chunk:**

“Thor have become household names, each with their own compelling backstories and struggles that resonate ... Iron Man revolutionized the”

### **Fourth Chunk:**

“film industry, introducing interconnected storylines that culminated in epic crossover events such as The Avengers ... comic book characters.”

### **Fifth Chunk:**

“Audiences have followed the journeys of superheroes as they face powerful foes like Thanos and Loki, ... responsibilities.”