

Markdown Text Splitter

MarkdownTextSplitter splits text along Markdown headings, code blocks, or horizontal rules. It's implemented as a simple subclass of RecursiveCharacterSplitter with Markdown-specific separators. See the source code to see the Markdown syntax expected by default.

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

```
from langchain.text_splitter import MarkdownTextSplitter
```

```
markdown_text = """
# 🦜🌀 LangChain

⚡ Building applications with LLMs through composability ⚡

## Quick Install

```bash
Hopefully this code block isn't split
pip install langchain
```

As an open source project in a rapidly developing field, we are extremely open to
contributions.
"""
markdown_splitter = MarkdownTextSplitter(chunk_size=100, chunk_overlap=0)
```

```
docs = markdown_splitter.create_documents([markdown_text])
```

```
docs
```

```
[Document(page_content='# 🦜🌀 LangChain\n\n⚡ Building applications with LLMs\nthrough composability ⚡', lookup_str='', metadata={}, lookup_index=0),\n Document(page_content='Quick Install\n\n```bash\n# Hopefully this code block\nisn't split\n\n\ninstall langchain". lookup_str='', metadata={}, lookup_index=0).
```

[Skip to main content](#)

```
we are extremely open to contributions.', lookup_str='', metadata={},  
lookup_index=0)]
```