

Python Code Text Splitter

PythonCodeTextSplitter splits text along python class and method definitions. It's implemented as a simple subclass of RecursiveCharacterSplitter with Python-specific separators. See the source code to see the Python syntax expected by default.

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

```
from langchain.text_splitter import PythonCodeTextSplitter
```

```
python_text = """
class Foo:

    def bar():

def foo():

def testing_func():

def bar():
"""
python_splitter = PythonCodeTextSplitter(chunk_size=30, chunk_overlap=0)
```

```
docs = python_splitter.create_documents([python_text])
```

```
docs
```

```
[Document(page_content='Foo:\n\n    def bar():', lookup_str='', metadata={},
lookup_index=0),
 Document(page_content='foo():\n\ndef testing_func():', lookup_str='', metadata=
{}, lookup_index=0),
 Document(page_content='bar():', lookup_str='', metadata={}, lookup_index=0)]
```