

EPubs

Contents

- Retain Elements

This covers how to load `.epub` documents into a document format that we can use downstream. You'll need to install the `pandocs` package for this loader to work.

```
from langchain.document_loaders import UnstructuredEpubLoader
```

```
loader = UnstructuredEpubLoader("winter-sports.epub")
```

```
data = loader.load()
```

Retain Elements

Under the hood, Unstructured creates different "elements" for different chunks of text. By default we combine those together, but you can easily keep that separation by specifying `mode="elements"`.

```
loader = UnstructuredEpubLoader("winter-sports.epub", mode="elements")
```

```
data = loader.load()
```

```
data[0]
```

```
Document(page_content='The Project Gutenberg eBook of Winter Sports
```

[Skip to main content](#)

```
sports.epub', 'page_number': 1, 'category': 'Title'}, lookup_index=0)
```