

- 출처: LangChain 공식 문서 또는 해당 교재명
- 원본 URL: <https://smith.langchain.com/hub/teddynote/summary-stuff-documents>

Microsoft PowerPoint

- **Microsoft PowerPoint**
 - **Microsoft** 에서 개발한 프레젠테이션 프로그램
 - **Microsoft PowerPoint** 문서를 다운스트림에서 사용할 수 있는 문서 형식으로 로드하는 방법 다루기
 - **Unstructured 공식 문서** - 자세한 Unstructured 설정 방법 참조 가능

```
# API KEY를 환경변수로 관리하기 위한 설정 파일
import os
from dotenv import load_dotenv

# API KEY 정보로드
load_dotenv()                # true
```

- 사전에 **VS Code** 터미널에 설치할 것

```
pip install -qU unstructured python-pptx
```

```
from langchain_community.document_loaders import UnstructuredPowerPointLoader

# UnstructuredPowerPointLoader 생성
loader = UnstructuredPowerPointLoader("../06_Document_Loader/data/sample-ppt.pptx")

# 데이터 로드
docs = loader.load()

# 로드한 문서의 개수 출력
print(len(docs))                # 1 (셀 출력: 4.0s)
```

Unstructured

- **Unstructured** = 텍스트의 다양한 **chunks** 에 대해 다양한 **elements** 를 생성
- 기본적으로 결합되어 하나의 **document** 로 반환
 - **but mode="elements"** 를 지정 → 쉽게 요소들을 분리 가능

```
# UnstructuredPowerPointLoader 생성
loader = UnstructuredPowerPointLoader(
    "../06_Document_Loader/data/sample-ppt.pptx",
    mode="elements"
)

# 데이터 로드
docs = loader.load()

print(len(docs))                # 17
```

```
print(docs[0].page_content)    # 랭체인 한국어 튜토리얼
```

```
docs[0].metadata
```

- 셀 출력

```
{'source': '../06 Document Loader/data/sample-ppt.pptx',  
'category_depth': 0,  
'file_directory': '../06 Document Loader/data',  
'filename': 'sample-ppt.pptx',  
'last_modified': '2025-09-12T13:42:39',  
'page_number': 1,  
'languages': ['kor'],  
'filetype': 'application/vnd.openxmlformats-officedocument.presentationml.presentation',  
'category': 'Title',  
'element_id': '784918cd25f6cbe0084401b85fe2f3d8'}
```

-
- next: 웹 문서(`WebBaseLoader`)
-