LLM Apps
LangChain: Document Loaders

Versatility for Data Loading

- Variety of sources:
    - Web pages.
    - Databases.
    - YouTube, Twitter.
    - Excel, Pandas, Notion, Figma, HuggingFace, Github, etc.

- Variety of formats:
    - PDF.
    - HTML.
    - JSON.
    - Word, PowerPoint, etc.

Examples

- Loading a PDF document.

- Loading audio from a YouTube video.
  - Data extraction is slow.
  - Necessary to install ffmpeg (on Mac, with Brew).

- Loading contents of a web page.
  - Three options.
  - In this case, we see the need to clean and prepare the loaded data before using it.

Other

- There are other document loaders that you can see in the LangChain documentation.