Document Parser

`Document`s can represent files in various formats, such as PDF, DOC, TXT, etc. To parse each of these formats, there's a `DocumentParser` interface with several implementations included in the library:

- `TextDocumentParser` from the `langchain4j` module, which can parse files in plain text format (e.g. TXT, HTML, MD, etc.)
- `ApachePdfBoxDocumentParser` from the `langchain4j-document-parser-apache-pdfbox` module, which can parse PDF files
- `ApachePoiDocumentParser` from the `langchain4j-document-parser-apache-poi` module, which can parse MS Office file formats (e.g. DOC, DOCX, PPT, PPTX, XLS, XLSX, etc.)
- `ApacheTikaDocumentParser` from the `langchain4j-document-parser-apache-tika` module, which can automatically detect and parse almost all existing file formats

Here is an example of how to load one or multiple `Document`s from the file system: