是的，LlamaIndex 提供了对 Microsoft Word 文档（.docx 格式）的解析支持。以下是相关的解析器和使用方法：

### 📄 1. DocxReader：专用的 Word 解析器

LlamaIndex 提供了 DocxReader，这是一个专门用于解析 Word 文档的读取器它会将 .docx 文件的内容提取为纯文本，并将文件类型信息存储在元数据中这种方式适用于结构简单的文档

**安装和使用：** 

pip install llama-index-readers-file docx2txt

```



```python

from llama\_index.readers.file.docs import DocxReader

reader = DocxReader()

documents = reader.load\_data("path/to/your/document.docx")

```

这种方法适用于结构简单的 Word 文档

---

### 📁 2. `SimpleDirectoryReader`：自动识别多种文件格式

`SimpleDirectoryReader` 是 LlamaIndex 的通用读取器，能够自动识别并处理多种文件格式，包括 Word、PDF、Markdown 它会根据文件扩展名选择合适的解析对于 Word 文件，它会使用 `DocxReader` 进行解。

\*\*安装和使用：\*\*



```bash

pip install llama-index-readers-file docx2txt

``





```python

from llama\_index.core import SimpleDirectoryReader

reader = SimpleDirectoryReader("path/to/your/directory")

documents = reader.load\_data()

``



这种方法适用于需要处理混合格式文件夹的场。

---

### 🧠 3. `DoclingReader`：结构化解析器（推荐

`DoclingReader` 是 LlamaIndex 提供的一个高级解析器，能够将 Word、PDF、HTML 等文档转换为结构化的 Markdown 或 JSON 格式，保留布局、标题、表格等。这种方式适用于需要保留文档结构的复杂档。

\*\*安装和使用：\*



```bash

pip install llama-index-readers-docling

``





```python

from llama\_index.readers.docling import DoclingReader

reader = DoclingReader()

documents = reader.load\_data("path/to/your/document.docx")

``



这种方法适用于需要保留文档结构的复杂档。

---

### ☁️ 4. LlamaParse：GenAI 原生解析平

LlamaParse 是 LlamaIndex 提供的一个 GenAI 原生解析平台，支持包括 Word 在内的多种文档。它能够将复杂文档转换为 LLM（大型语言模型）优化的格式，支持自然语言指令进行析。

\*\*安装和使用：\*



```bash

pip install llama-cloud-services

``





```python

from llama\_cloud\_services import LlamaParse

parser = LlamaParse(api\_key="your\_api\_key")

result = parser.parse("path/to/your/document.docx")

``



这种方法适用于需要处理复杂文档并进行结构化解析的景。

---

### ✅ 总结推荐

- \*\*结构简单的 Word 文档\*：使用 `DocxReader` 或 `SimpleDirectoryReade`。

- \*\*需要保留文档结构的复杂文档\*：使用 `DoclingReade`。

- \*\*需要 GenAI 优化解析的复杂文档\*：使用 LlamaPare。

如果您需要进一步的示例代码或有其他问题，欢迎继续提问！