

将PDF文档转换为Markdown的智能工具Vision Parse

GitHubStore GitHubStore 2025年01月09日 09:16 河北

项目简介

使用最先进的视觉语言模型将 PDF 文档解析为格式精美的 Markdown 内容 - 只需几行代码!

Vision Parse 利用视觉语言模型的力量彻底改变文档处理:

- 📄 **扫描文档处理**: 智能识别扫描文档中的文本、表格和 LaTeX 方程, 并将其高精度提取为 Markdown 格式的内容
- 🎨 **高级内容格式**: 保留 Markdown 格式内容的 LaTeX 方程、超链接、图像和文档层次结构
- 🗣️ **多LLM支持**: 与OpenAI、Gemini和Llama等多个Vision LLM提供商无缝集成, 以实现最佳的准确性和速度
- 📁 **本地模型托管**: 支持 Ollama 本地模型托管, 以实现安全、免费、私密和离线文档处理

🚀 开始使用

先决条件

- 🐍 Python ≥ 3.9
- 🖥️ Ollama (如果你想使用本地模型)
- 🗣️ OpenAI 或 Google Gemini 的 API 密钥 (如果您想使用 OpenAI 或 Google Gemini)

安装

使用 pip 安装核心包 (推荐):

```
1 pip install vision-parse
```

安装 OpenAI 或 Gemini 的附加依赖项:

```
1 # For OpenAI support
2 pip install 'vision-parse[openai]'
```

```
1 # For Gemini support
2 pip install 'vision-parse[gemini]'
```

```
1 # To install all the additional dependencies
2 pip install 'vision-parse[all]'
```



从源安装包：

```
1 pip install 'git+https://github.com/iamarunbrahma/vision-parse.git#egg=vision-
```

设置 Ollama（可选）

有关如何在本地设置 Ollama 的信息，请参阅[Examples/ollama_setup.md](#)。

🕒 用法

基本用法示例

```
1 from vision_parse import VisionParser
2
3 # Initialize parser
4 parser = VisionParser(
5     model_name="llama3.2-vision:11b", # For local models, you don't need to p
6     temperature=0.4,
7     top_p=0.5,
8     image_mode="url", # Image mode can be "url", "base64" or None
9     detailed_extraction=False, # Set to True for more detailed extraction
10    enable_concurrency=False, # Set to True for parallel processing
11 )
12
13 # Convert PDF to markdown
14 pdf_path = "path/to/your/document.pdf" # Local path to your pdf file
15 markdown_pages = parser.convert_pdf(pdf_path)
16
17 # Process results
18 for i, page_content in enumerate(markdown_pages):
19     print(f"\n--- Page {i+1} ---\n{page_content}")
```



OpenAI 或 Gemini 模型使用

```
1 from vision_parse import VisionParser
2
3 # Initialize parser with OpenAI model
4 parser = VisionParser(
5     model_name="gpt-4o",
6     api_key="your-openai-api-key", # Get the OpenAI API key from https://plat
7     temperature=0.7,
8     top_p=0.4,
9     image_mode="url",
10    detailed_extraction=True, # Set to True for more detailed extraction
```

```
11     enable_concurrency=True,
12 )
13
14 # Initialize parser with Google Gemini model
15 parser = VisionParser(
16     model_name="gemini-1.5-flash",
17     api_key="your-gemini-api-key", # Get the Gemini API key from https://aist
18     temperature=0.7,
19     top_p=0.4,
20     image_mode="url",
21     detailed_extraction=True, # Set to True for more detailed extraction
22     enable_concurrency=True,
23 )
```

✅ 支持的模型

该软件包支持以下 Vision LLM模型：

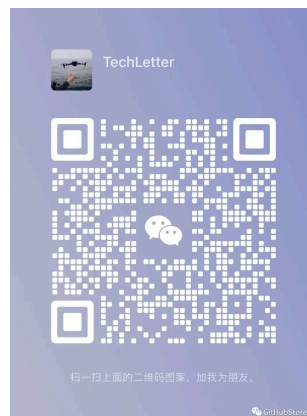
- OpenAI: gpt-4o 、 gpt-4o-mini
- Google Gemini: gemini-1.5-flash, gemini-2.0-flash-exp, gemini-1.5-pro
- 来自 Ollama 的 Meta Llama 和 Llava: llava:13b 、 llava:34b 、 llama3.2-vision:11b 、 llama3.2-vision:70b

项目链接

<https://github.com/iamarunbrahma/vision-parse>

扫码加入技术交流群，备注「**开发语言-城市-昵称**」

合作请注明



关注「**GitHubStore**」公众号

GitHub **GitHubStore**
分享有意思的开源项目