多格式文档处理器测试文档

# 概述

这是一个用于测试多格式文档处理器的示例Word文档。

该文档包含了各种文本格式和结构，用于验证Word文档处理功能。

# 支持的功能

* 本系统支持以下文档格式：
* PDF文档处理
* Word文档处理
* PowerPoint演示文稿
* Excel电子表格
* 图片OCR识别

# 技术规格

|  |  |
| --- | --- |
| 技术 | 描述 |
| 文本提取 | 使用python-docx库 |
| 向量化 | 使用OpenAI嵌入模型 |
| 存储 | FAISS向量数据库 |
| 搜索 | 语义相似度搜索 |