





Closed

[BUG] 知识库搜索，一模一样的内容搜不到 #1433  
H9990HH969 opened this issue on Sep 11 · 22 comments

用 txt 分得不是很好，用 csv 会好一点。  
主要大模型会乱回答，这个不太好



zRzRzRzRzRzR commented on Oct 19 · edited

Collaborator

针对以上问题，主要是以下几个方面的改进

1. 从读入数据的角度，自定义分词器和loader能更好的支持各种各样的文件
2. 从大模型回答的角度，更好的提示词和适合自己场景的提示词是很重要的，并且建议换一个支持文本更长，理解能力更强的模型，同时，换一个更好的embed模型例如bge-large-zh-v1.5，必要时可以微调嵌入模型。
3. 从框架的角度，目前在搜索部分 0.2.x和0.1.x确实算法不同，召回方面有一定的差距，在某些场景甚至是降低的，这个我们也意识到了，我们将会在这方面进行改进，提升召回率。
4. 从知识库的角度，尽量让知识库之间的混淆信息减少，以降低向量匹配时候的难度

