测试集的构建方法及系统、模型的评估方法及系统

# 说 明 书 摘 要

本说明书提供一种测试集的构建方法及系统、模型的评估方法及系统，其可以基于源数据库的多个文档块，将具有关联关系的文档块作为测试文档组合，基于大语言模型对目标文档进行问题抽取，以确定对应的测试问题，并将测试问题与其对应的测试文档组合作为一组标签数据，基于多组标签数据构成测试集，后续基于测试集对检索增强生成系统的检索能力进行评估，这样的设置方式使得本说明书提供的方法及系统不仅可以评估检索增强生成系统根据问题检索单一文档的能力，还能评估检索增强生成系统根据总结性的问题检索关联文档的能力，从而提升对检索增强生成系统的检索能力评估的全面性。

图 3