

本周工作总结组会汇报

董若扬

School of Computer Science and Technology, USTC

2024-09-20



学习向量数据库



学习向量数据库

• 通过开发文档系统学习完 ArkTS 基本语法 感觉其声明式和响应式的特性和谷歌新推出的 Compose 框架很类似



学习向量数据库



观看学习了上个月底的会议回放,会议主要提出了端侧向量数据库的 CAP 定理,即要在性能,精度,存储空间三者中取得平衡。云端的向量数据库看重性能和召回,而功耗和内存更为有限的终端设备需要考虑有限的资源如何取得可接受的效果。

会议介绍了什么是向量数据库,向量数据库中相较于结构化数据库不同的搜索方式,不同的建立索引方式,以及构建端侧向量数据库的 benchmark。

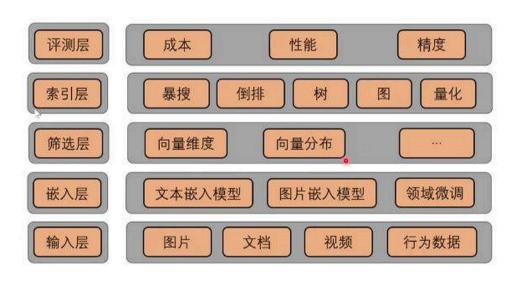


图 1 构建端侧向量数据库的 benchmark



阅读了两篇向量数据库相关的论文:

- Survey of vector database management systems
- A comprehensive survey on vector database: Storage and retrieval technique, challenge

里面介绍的相关技术还有很多前置知识没有学习,暂时看不懂,但内容和会议探讨的方向非常相关,就是如何做近似搜索,如何减少索引体积,如何在大模型中发挥效用。



学习向量数据库

重点了解会议和论文中提到的理论知识并作汇报。