



科研成果系列发布（第三十九期）

时间：2022-01-18 作者： 文章来源：计算机科学与技术学院 浏览： 280



哈尔滨工程大学计算机科学与技术学院 科研成果系列发布（第三十九期）



基于倒排索引的数据组织模型研究

成果简介：

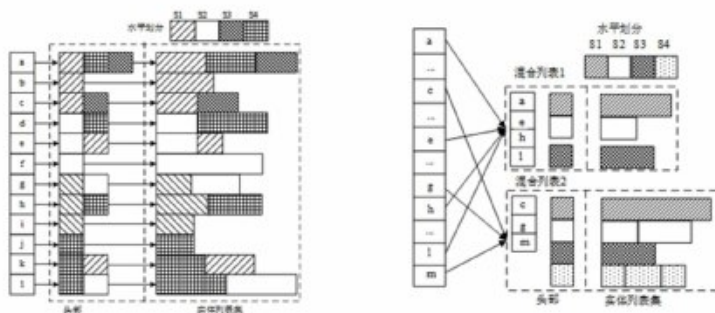
针对传统数据空间索引方法无法适用于高倾斜分布的大规模数据的问题，从负载均衡和划分角度对数据空间多维索引技术进行了研究，研究一种基于倒排索引的数据组织模型，旨在保持各个索引节点负载均衡、减少查询通信开销、提高数据空间查询处理性能。

技术路线：

- (1) 在垂直划分中，聚合在查询日志和实体中频繁出现的token词，以减少查询涉及倒排列表的聚合/合并开销；
- (2) 在水平划分中，结合超图理论和用户查询与实体间访问模式信息，把水平划分问题归约为超图划分问题，从而保持水平划分的负载均衡；
- (3) 结合垂直划分和水平划分策略，构建了二维混合索引。

应用领域：

数据管理、数据集成、多维索引。



联系人：王红滨

联系方式：18145148895

电子邮箱：wanghongbin@hrbeu.edu.cn