

数据库系统及应用作业 4

傅申 PB20000051

题 1.

如下:

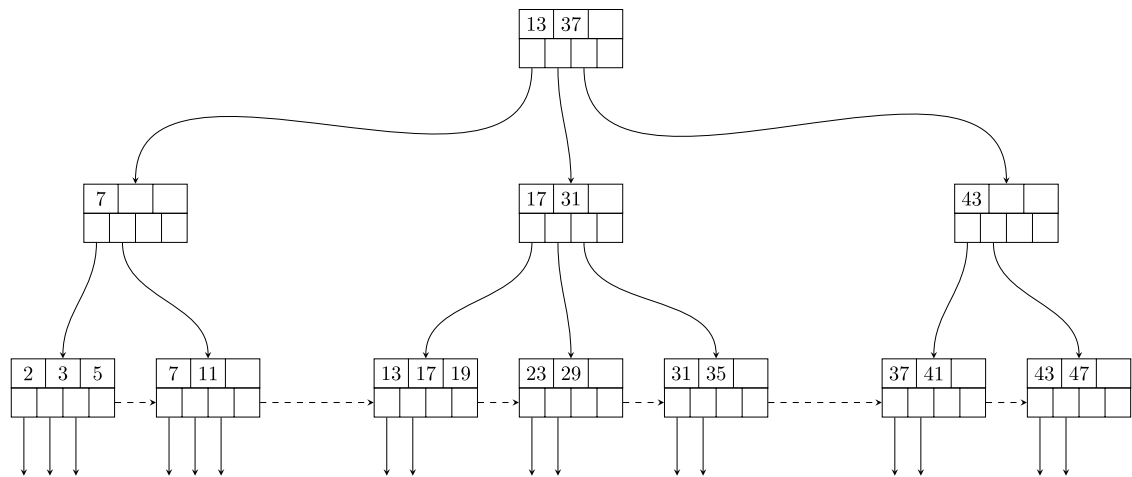


图 1: 插入了 35 后的 B+ 树

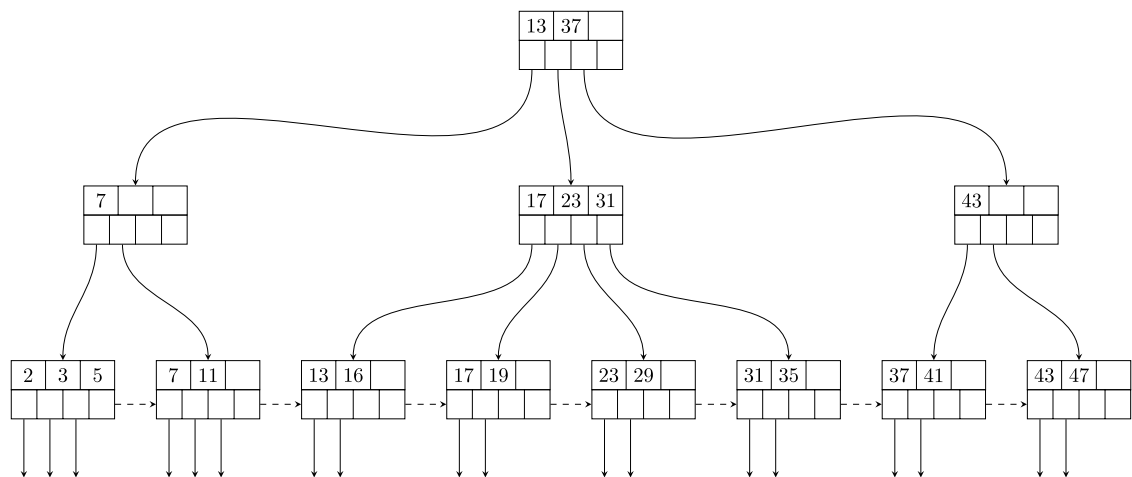


图 2: 插入了 16 后的 B+ 树

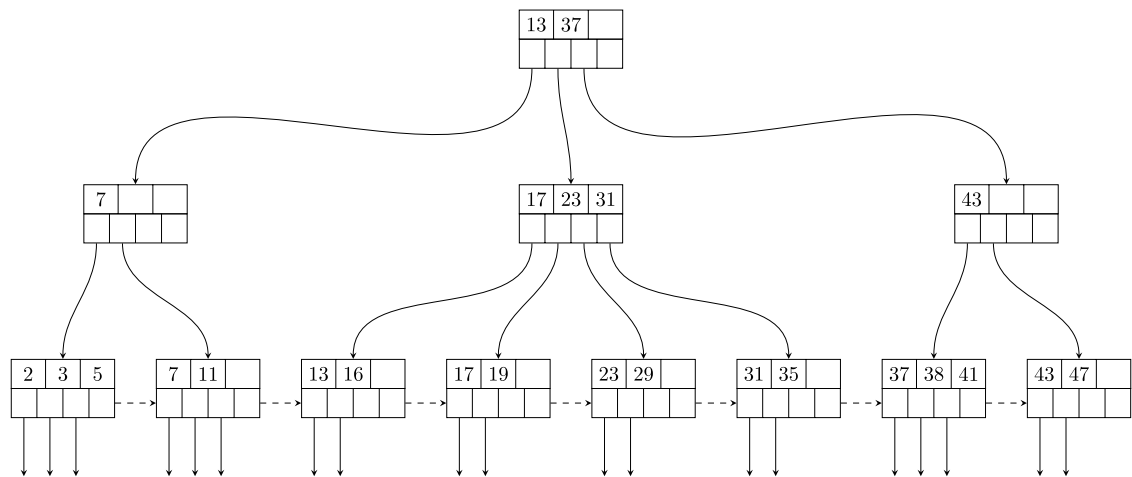


图 3: 插入了 38 后的 B+ 树

题 2.

- 1) 不能, 因为查询的 title 只需要包含关键字即可, 关键字可以出现在 title 的任何位置, 而 B+ 树是根据 title 的字典序排序的, 仅仅可以搜索前缀, 仍然需要扫描所有项.
- 2) 使用倒排索引, 如下图

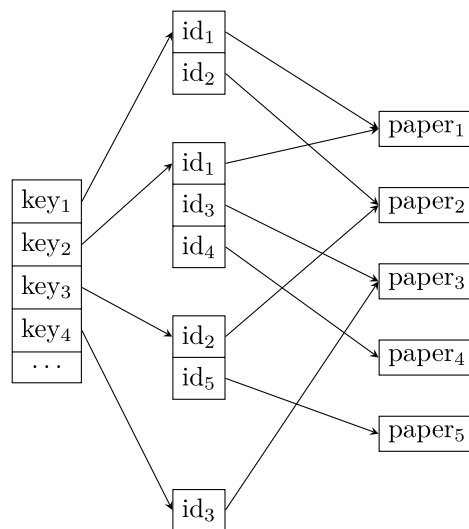


图 4: 倒排索引

插入 对文章的 title 分词, 然后在索引中查询关键字是否存在,

- 若关键词存在, 则将指针插入对应的桶中;
- 否则, 在索引中插入关键字, 增加对应的桶并在桶中插入指向文档的指针.

查询 在索引中查找所有关键字, 返回得到的文档集合的交集.

题 3.

- 1) 共有 6 个桶, 键值 E 所在的桶中的全部键值有: E, I, K.
- 2) 共有 6 个桶, 键值 B 所在的桶中的全部键值有: B, G, N.