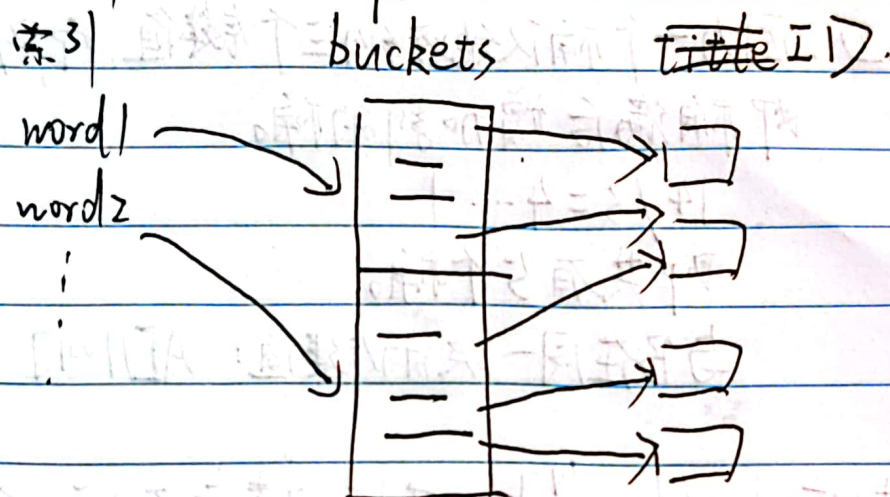


1. 解: 1) 能提高效率,

因为使用B+树进行查找时可以迅速定位到相关记录, 而不需要对整个表进行顺序查找, 使查找更高效。

2) 可以使用倒排索引。



插入: 对于 title 中的关键词, 先进行遍历, 并将文档的 ID 添, 加到对应关键词的桶中。

查找: 当查询到对应关键词时, 只需检索倒排索引中对应关键词的桶, 而不需要扫描整个文档表, 提高了查询效率。



2. 解: (1) 取前3位构成桶数组。

则有 110, 001, 101, 010, 000, 111, 100, 011 这8个桶。

其中与E一个桶的全部键值有: 10/00, 10/01
E K

(2) 因为一个桶仅能容纳三个键值, 而 $2/3 < 80\%$, 所以即桶满后增加新的桶。

$$14/3 = 4 \dots 2$$

即共有5个桶。

与B在同一块的键值: A[1100] B[0011], C[0001]

3. 解: 树高 $h = \lceil \log_n \frac{N}{2} \rceil$, 其中 $\lceil x \rceil$ 表示不小于 x 的最小整数。

读块时间: $t = S + T + R$

解压时间: c

$$\text{总时间} = h \cdot (t + c) + m$$

~~$$= \lceil \log_n \frac{N}{2} \rceil (S + T + R) + m$$~~

$$= \lceil \log_n \frac{N}{2} \rceil (S + T + R + c) + m$$

