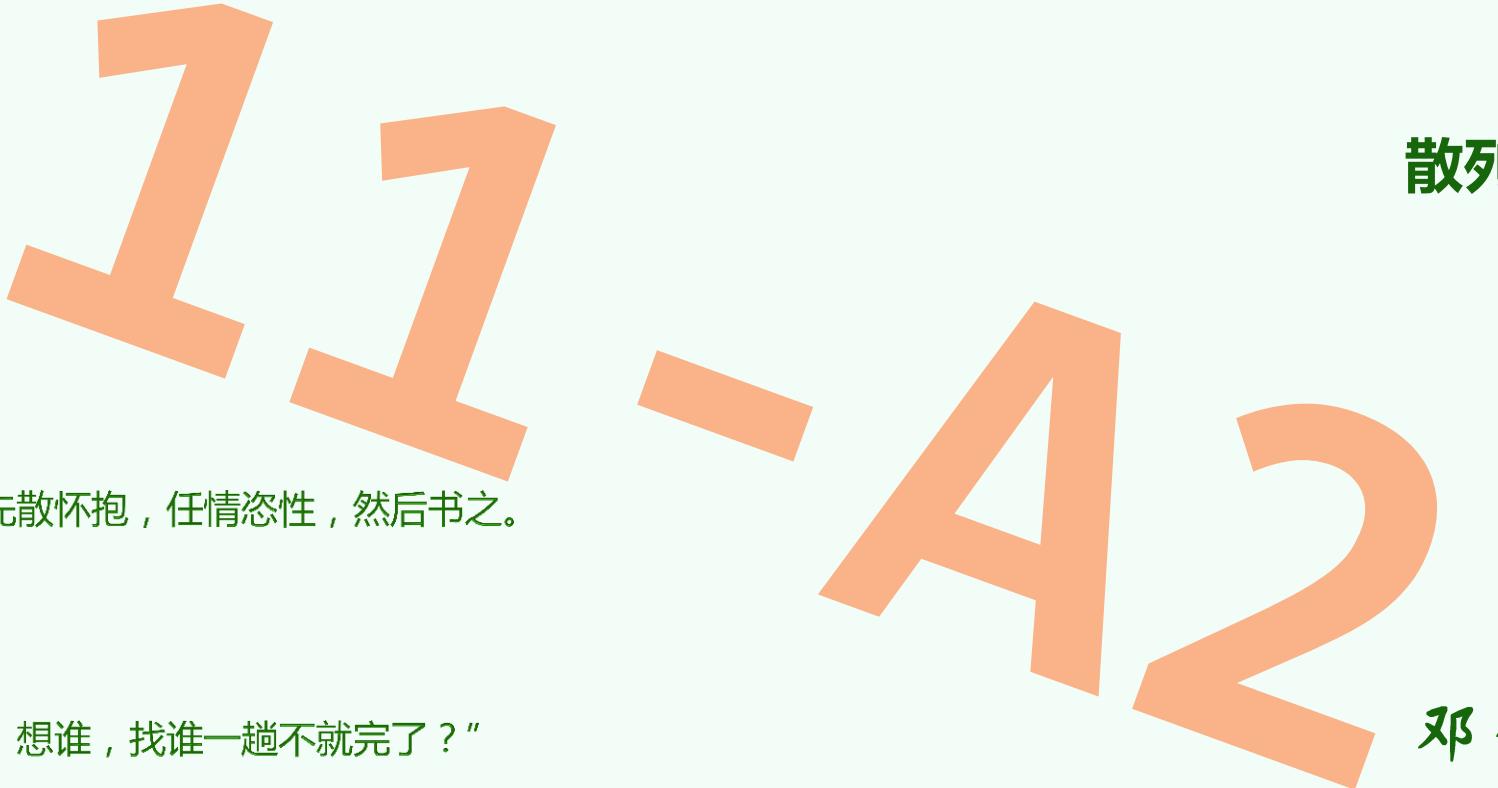


词典

散列：原理



服务 ~ 电话

General inquiries

Tel: Toll Free: 1-800-IBM-4YOU

E-mail: askibm@vnet.ibm.com
www.ibm.com/us/en/

Shopping

Tel: Toll Free: 1-888-SHOP-IBM

Sales Center

1-855-2-LENOVO (1-855-253-6686)

Mon - Fri: 9am-9pm (EST)

Sat - Sun: 9am-6pm (EST)

Customer Service

1-855-2-LENOVO (1-855-253-6686)

Mon - Fri: 9am-9pm (EST)

Sat - Sun: 9am-6pm (EST)



电话簿

需求：为一所学校制作电话簿

号码 \rightarrow 个人（教员、学生、员工）或办公室

蛮力：使用数组，按电话号码索引

时间 = $\Theta(1)$

以清华为例（2003）

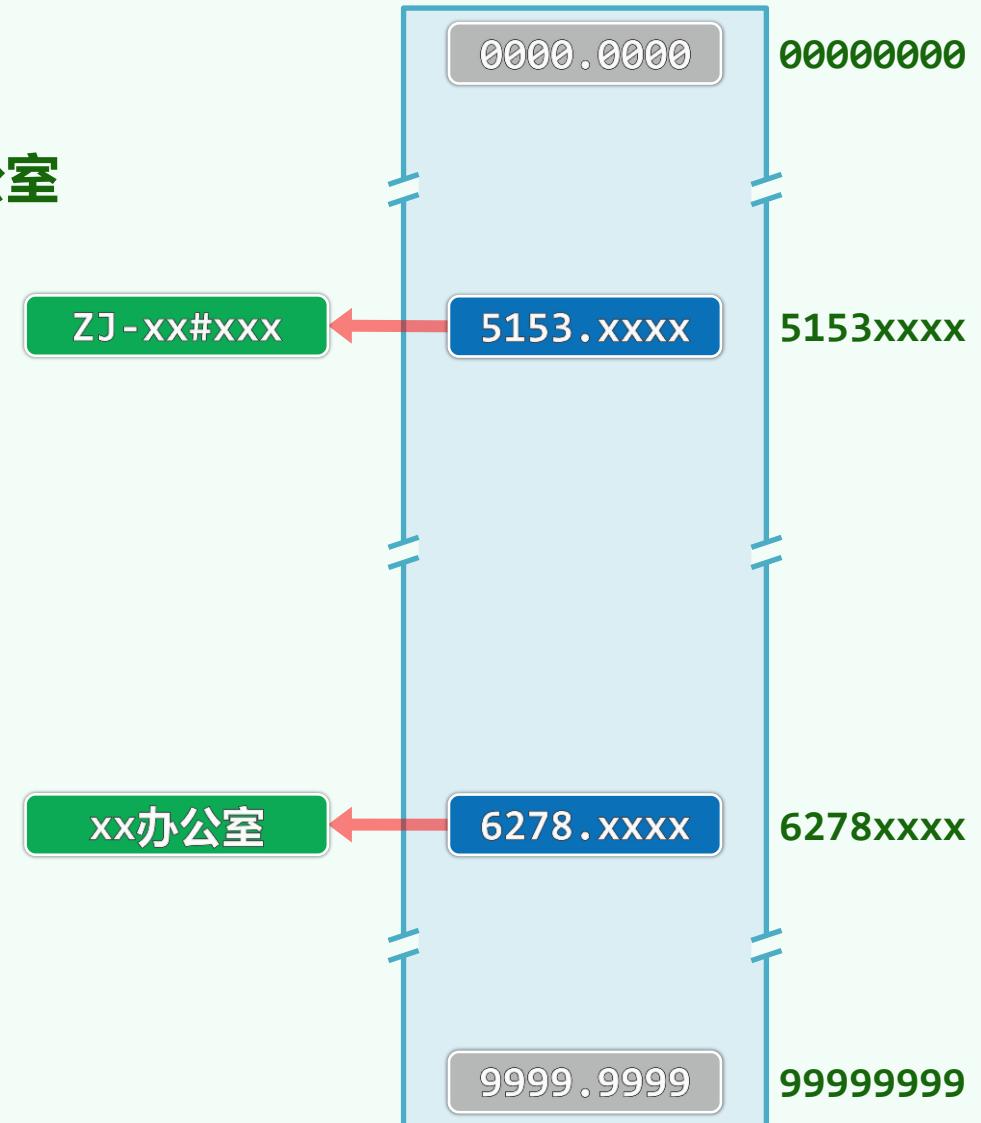
#可能的电话 = $R = 10^8 = 100M$

#实有的电话 = $N = 25,000 = 25K$

问题：空间 = $\Theta(R + N) = \Theta(100M + 25K)$

效率 = $25K / 100M = 0.025\%$

如何在保持查找速度的同时，降低存储消耗？



散列表 / 散列函数

❖ 桶 (bucket) : 直接存放或间接指向一个词条

❖ 桶数组 (bucket array) / 散列表 (hash-table)

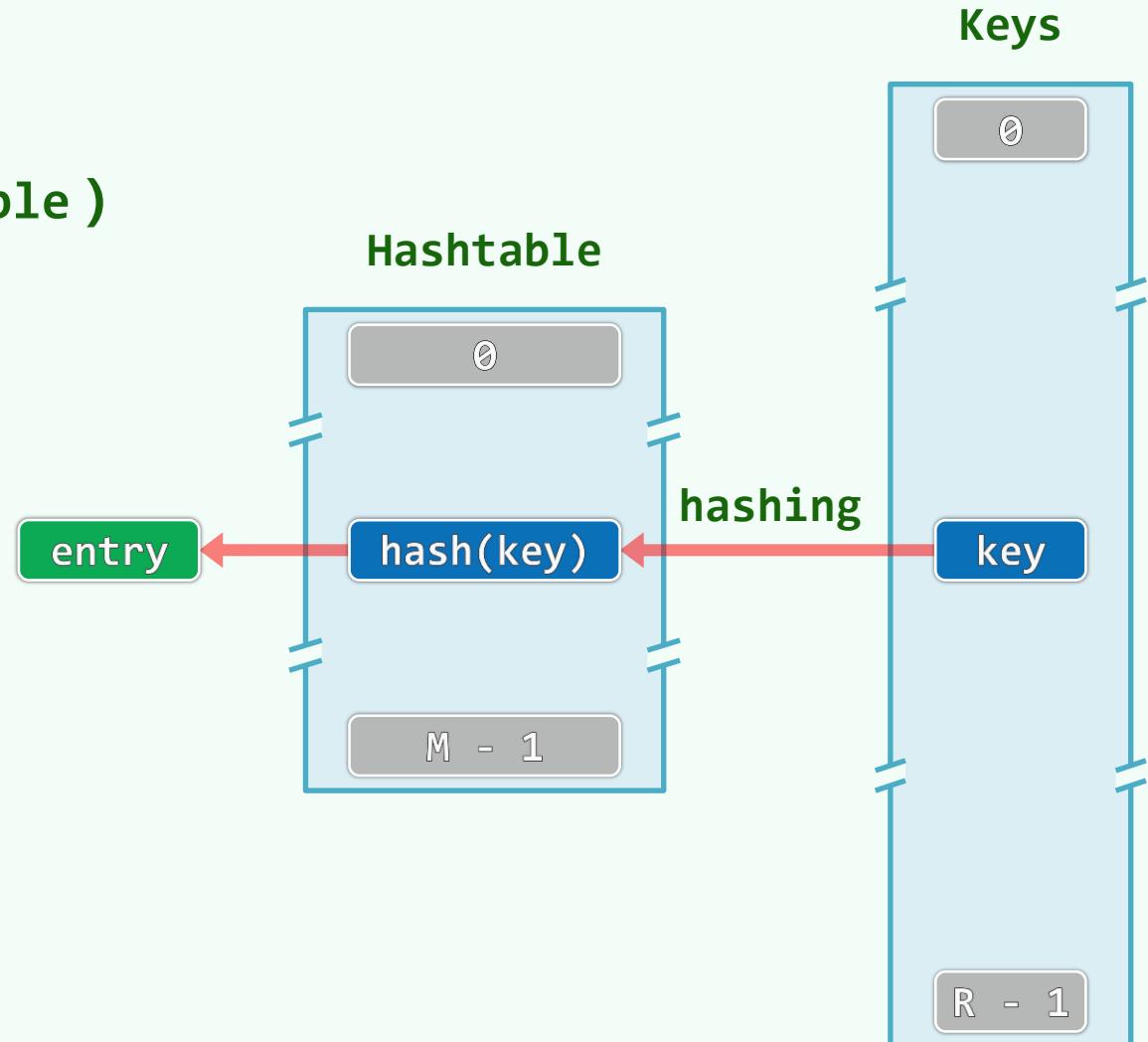
- 容量 : M
- 满足 : $N < M \ll R$
- 空间 : $\mathcal{O}(N + M) = \mathcal{O}(N)$

❖ 定址/杂凑/散列

- 根据词条的key (未必可比较)
- “直接” 确定散列表入口 (无论表有多长)

❖ 散列函数 : $hash() : key \mapsto \&entry$

❖ “直接” : $expected - \mathcal{O}(1) \neq \mathcal{O}(1)$



实例

