



数据结构与算法 (十二)

张铭 主讲

采用教材：张铭，王腾蛟，赵海燕 编写
高等教育出版社，2008.6（“十一五”国家级规划教材）

<http://www.jpku.pku.edu.cn/pkujpk/course/sjjg>

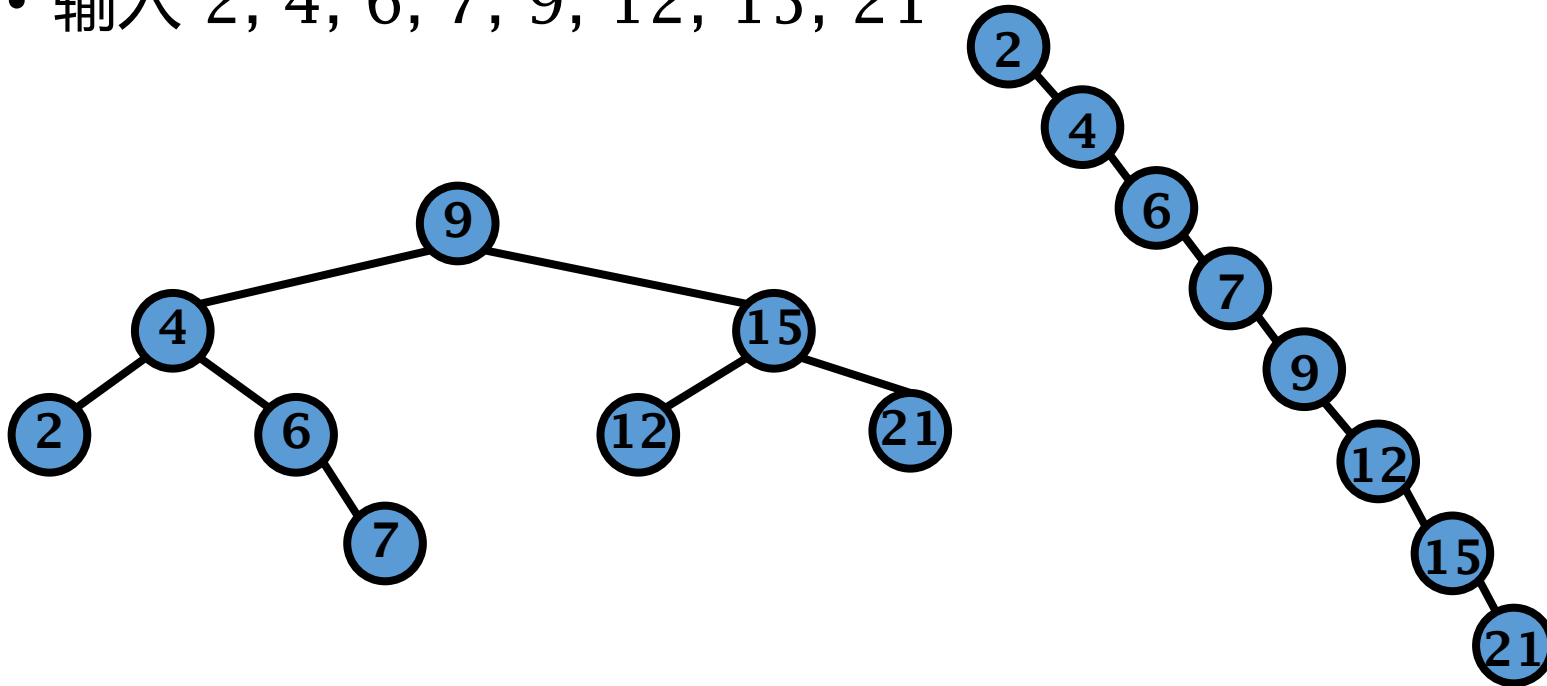


第十二章 高级数据结构

- 12.1 多维数组
- 12.2 广义表
- 12.3 存储管理
- 12.4 Trie 树
- 12.5 改进的二叉搜索树

12.4 Trie 树

- 理想状况：插入、删除、查找时间代价为 $O(\log n)$
- 输入 9, 4, 2, 6, 7, 15, 12, 21
- 输入 2, 4, 6, 7, 9, 12, 15, 21





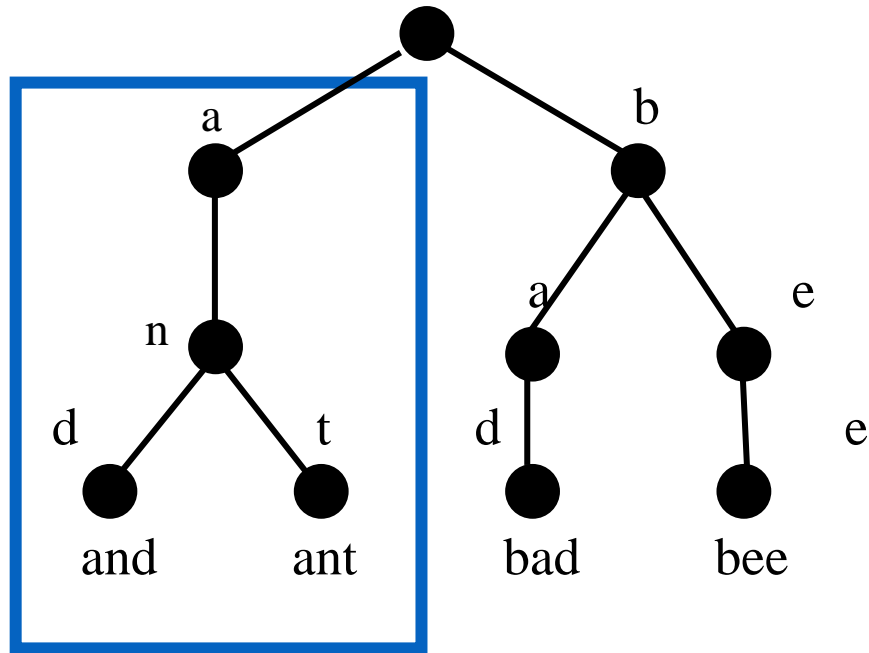
Trie 结构

- 关键码对象空间分解
- “trie”这个词来源于 “retrieval”
- 主要应用
 - 信息检索 (information retrieval)
 - 自然语言大规模的英文词典
- 字符树——26叉Trie
- 二叉Trie树
 - 用每个字母（或数值）的二进制编码来代表
 - 编码只有0和1

英文字符树——26叉Trie

存单词and、ant、bad、bee

“an”子树代表相同前缀an-具有的关键码集合{and, ant}

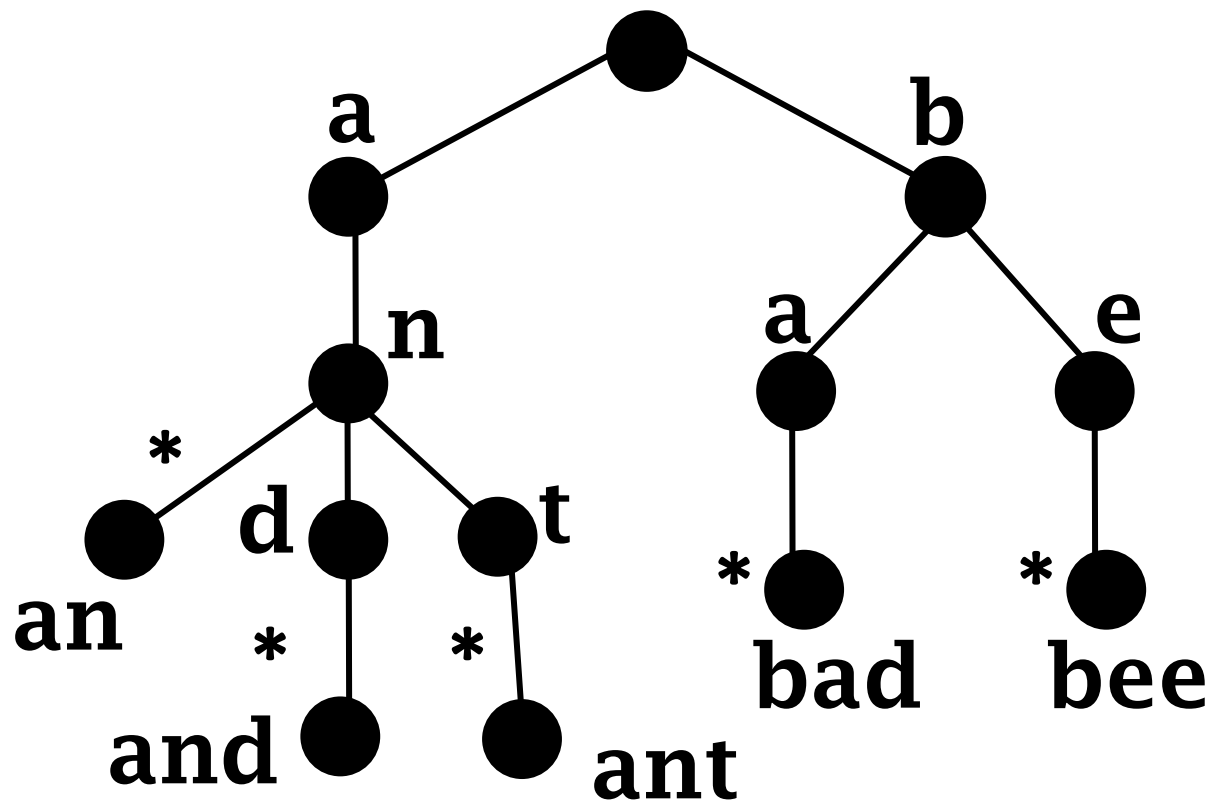


- 一棵子树代表具有相同前缀的关键码的集合

12.4 Trie 树

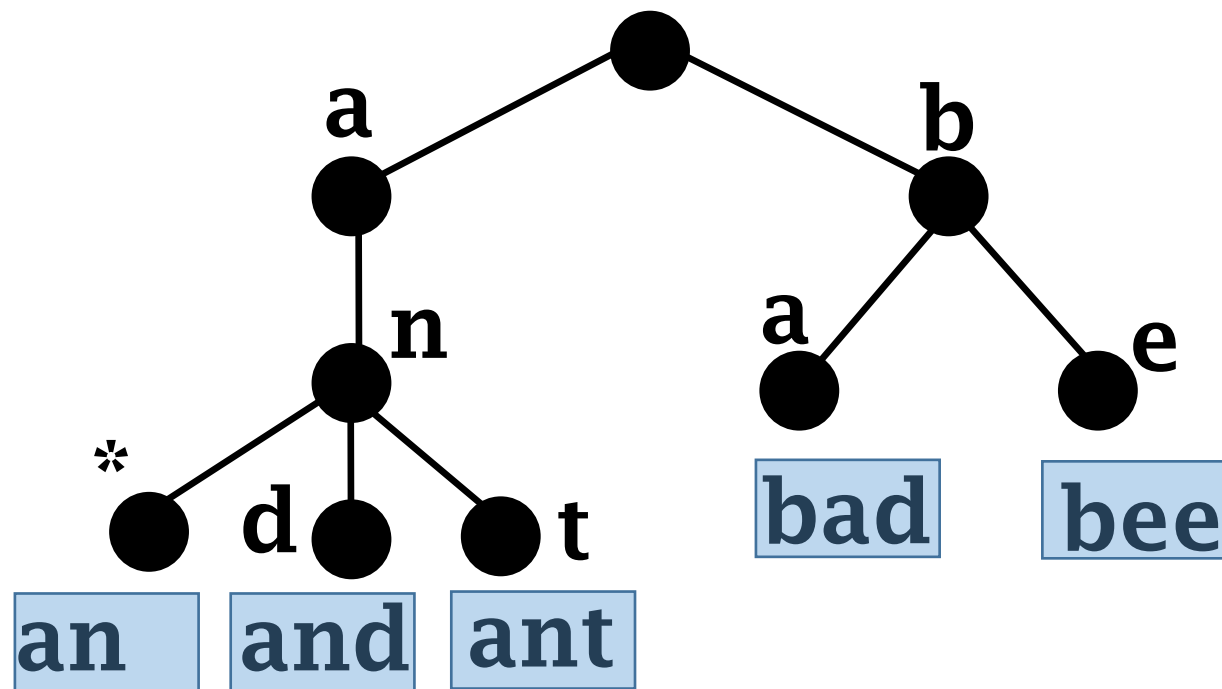
不等长的字符树，加 “*” 标记

存储单词 an, and, ant, bad, bee



压缩靠近叶结点的单路径

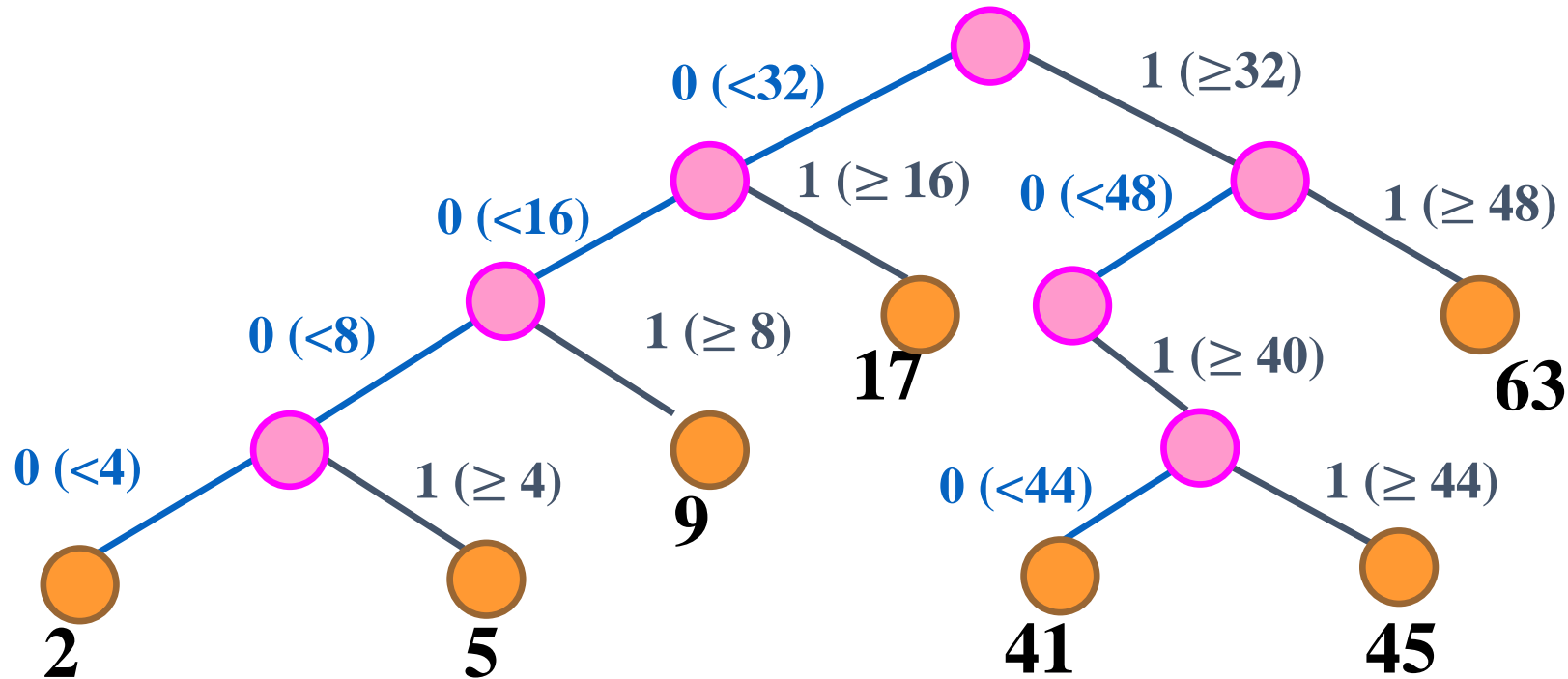
存储单词 an, and, ant, bad, bee



12.4 Trie 树

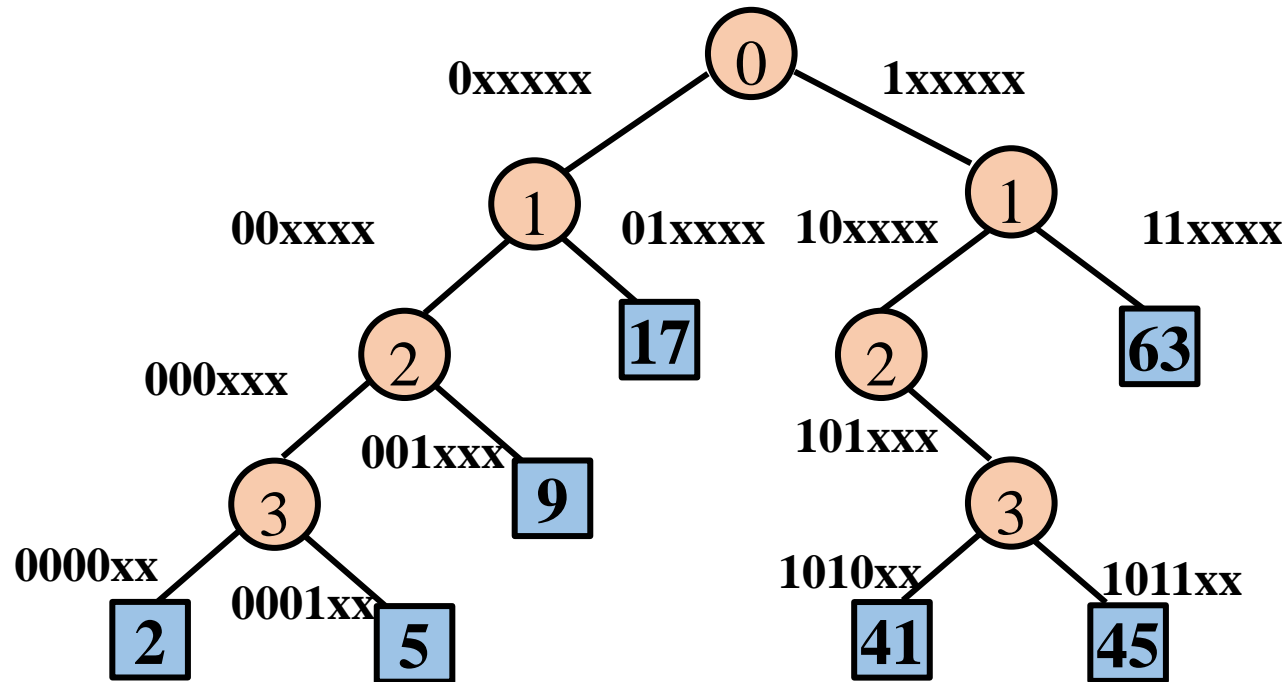
二叉 Trie 结构

元素为 2、5、9、17、41、45、63



二叉 Trie 结构

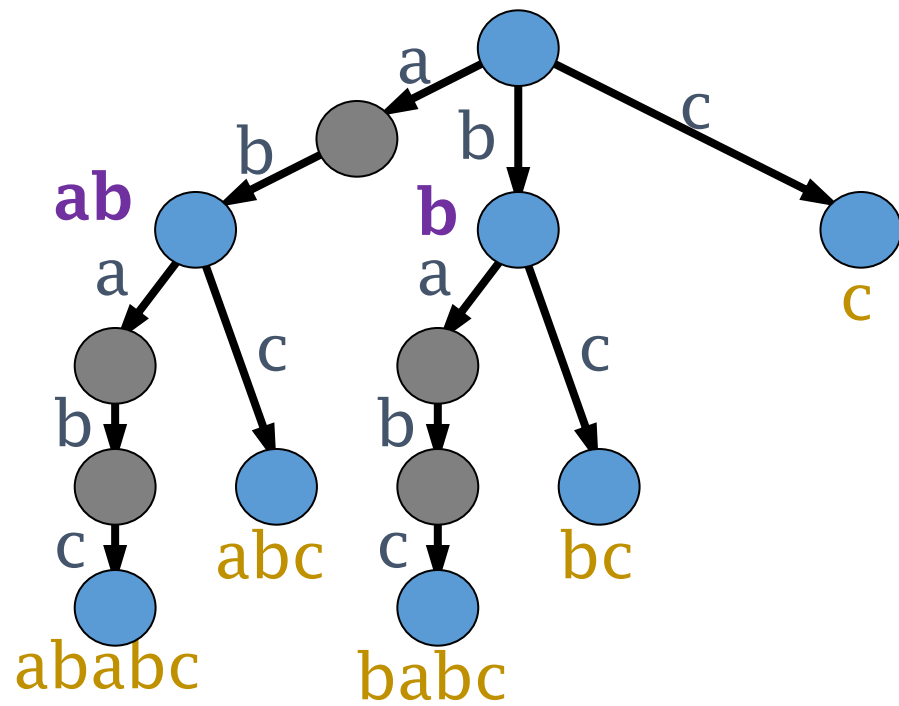
压缩



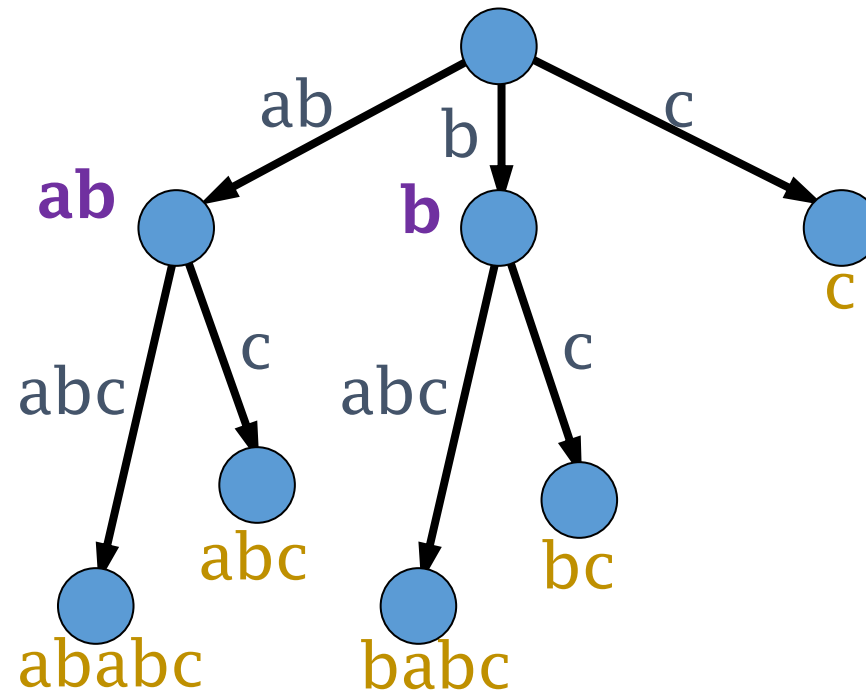
编码: 2: 000010 5: 000101 9: 001001
17: 010001 41: 101001 45: 101101 63: 111111

后缀树 (Suffix Trees)

ababc 后缀子串: ababc, babc, abc, bc, c



Suffix Trie



Suffix Tree



后缀数组 (Suffix Array)

5	ALAM\$
1	ALAYALAM\$
7	AM\$
3	AYALAM\$
6	LAM\$
2	LAYALAM\$
0	MALAYALAM\$
8	M\$
4	YALAM\$
9	\$

MALAYALAM\$
0 1 2 3 4 5 6 7 8 9

5	1	7	3	6	2	0	8	4	9
---	---	---	---	---	---	---	---	---	---

后缀数组

3	1	1	0	2	0	1	0	0	-
---	---	---	---	---	---	---	---	---	---

最长公共前缀数组

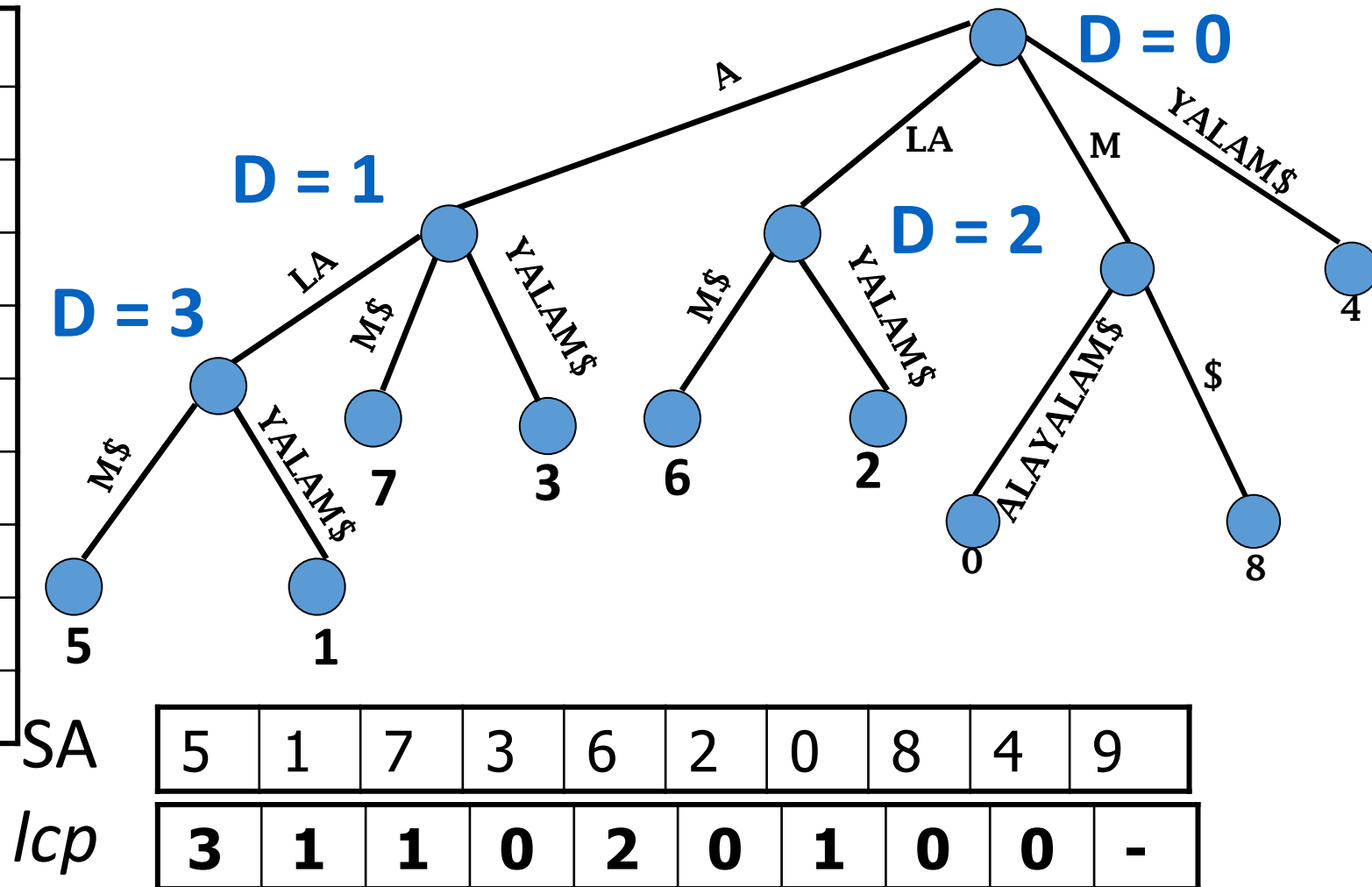
↑
后缀5和1共享 "ALA"
后缀1和7共享 "A"

LCP总是相邻的



12.4 Trie 树

5	ALAM\$
1	ALAYALAM\$
7	AM\$
3	AYALAM\$
6	LAM\$
2	LAYALAM\$
0	MALAYALAM\$
8	M\$
4	YALAM\$
9	\$





12.4 Trie 树

思考

- 中文是否适合组织字符树？是否适合二叉 Trie 结构？
- 查阅后缀树、后缀数组的文献，思考其应用场景。



数据结构与算法

谢谢聆听

国家精品课“数据结构与算法”

<http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/>

张铭，王腾蛟，赵海燕

高等教育出版社，2008. 6。“十一五”国家级规划教材