



# Data Structures and Algorithms ( 12 )

Instructor: Ming Zhang

Textbook Authors: Ming Zhang, Tengjiao Wang and Haiyan Zhao

Higher Education Press, 2008.6 (the "Eleventh Five-Year" national planning textbook)

<https://courses.edx.org/courses/PekingX/04830050x/2T2014/>

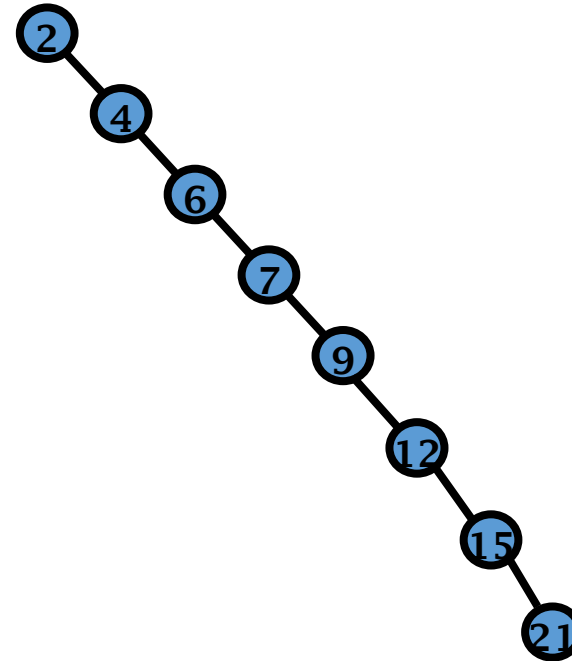
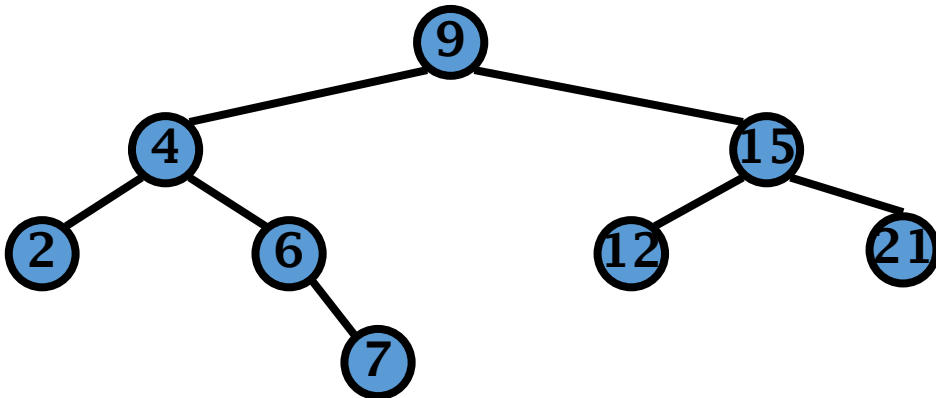


# Chapter 12 Advanced data structure

- 12.1 Multidimensional Array
- 12.2 Generalized Lists
- 12.3 Storage management
- **12.4 Trie**
- 12.5 Improved binary search tree

## 12.3 Trie

- Ideal situation: The average time of insertion, deletion, and search is  $O(\log N)$
- Input 9, 4, 2, 6, 7, 15, 12, 21
- Output 2, 4, 6, 7, 9, 12, 15, 21





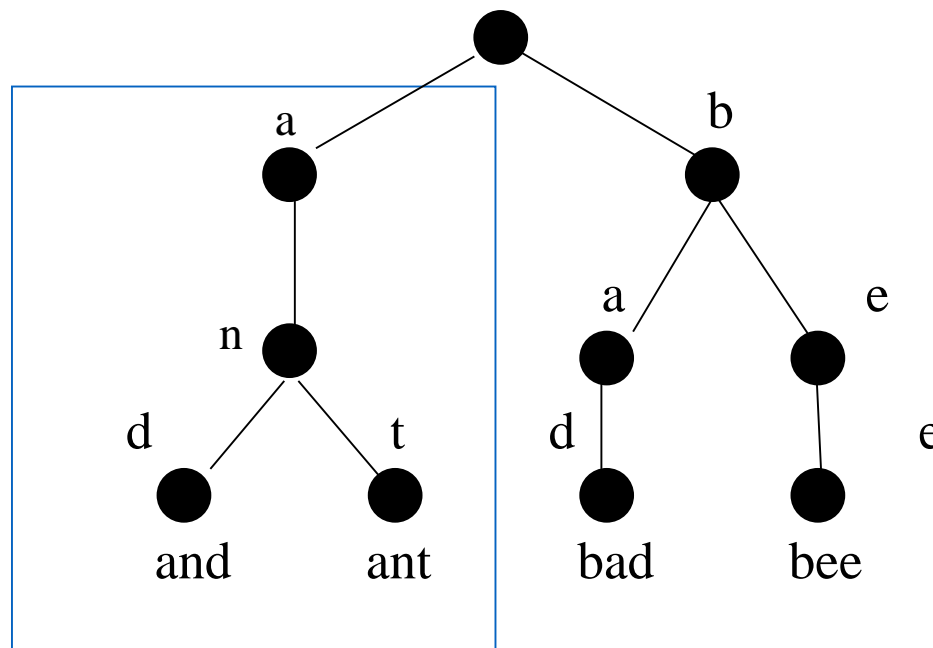
# Structure of Trie

- Space division of key
- “trie” comes from “retrieval”
- Application
  - Information retrieval
  - Large scale of English dictionary
- 26-branch Trie
- Binary Trie
  - Letters (numbers) represented as binary coding
  - Coding includes just 0 and 1

# Tree of English words: 26-branch Trie

Store words 'and, ant, bad, bee'

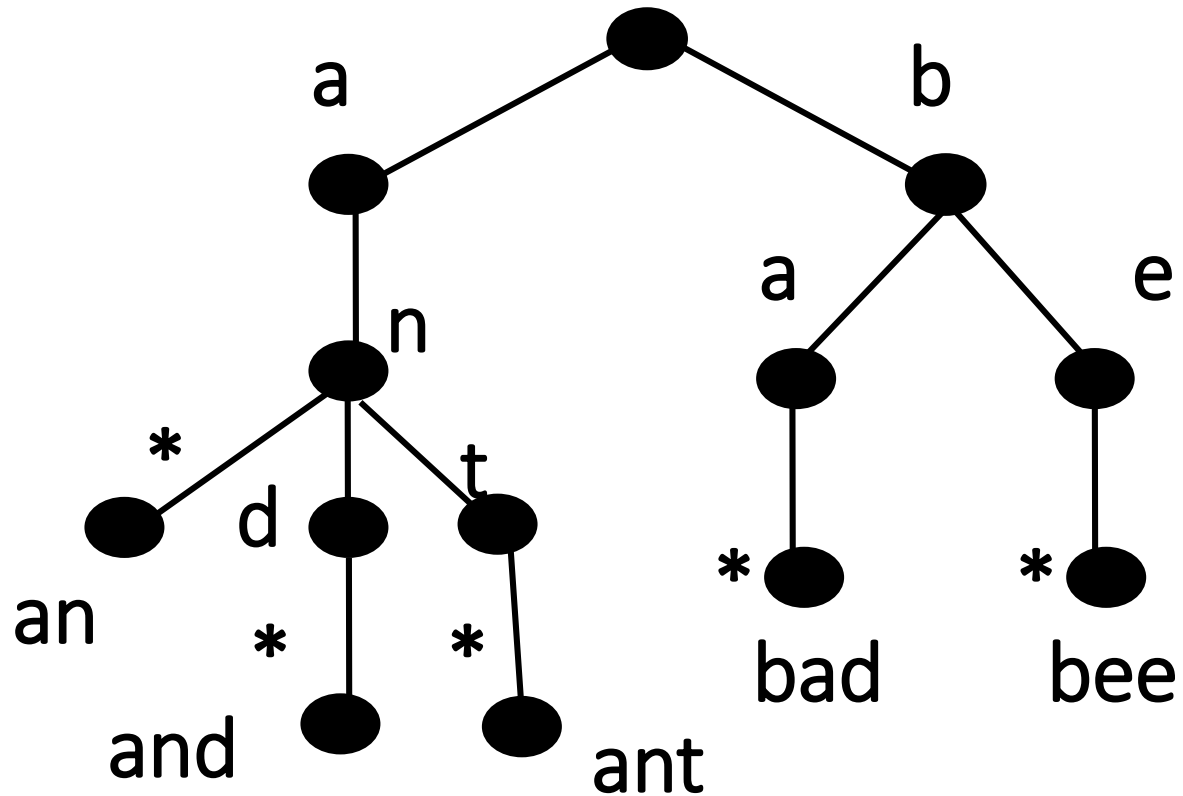
Subtree 'an' contains set {and, ant} that every word from the set has the same prefix 'an'.



A subtree contains the words with the same prefix

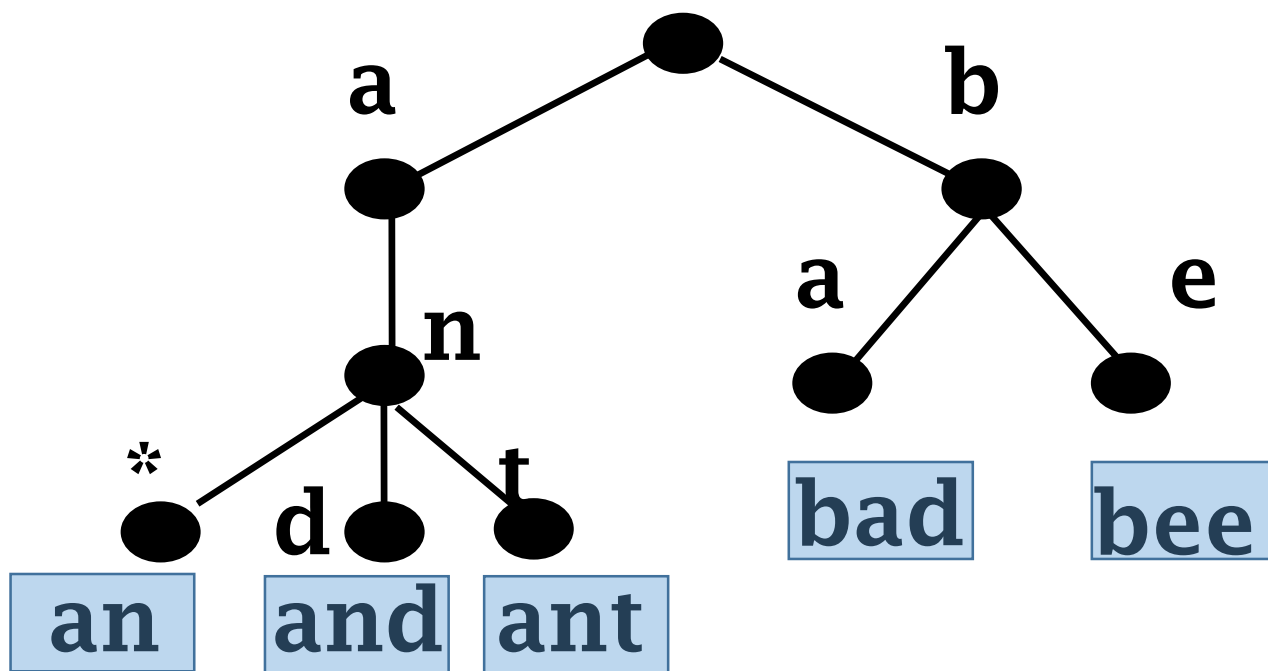
## 12.4 Trie

Store words **an, and ant, bad, bee**



## 12.4 Trie

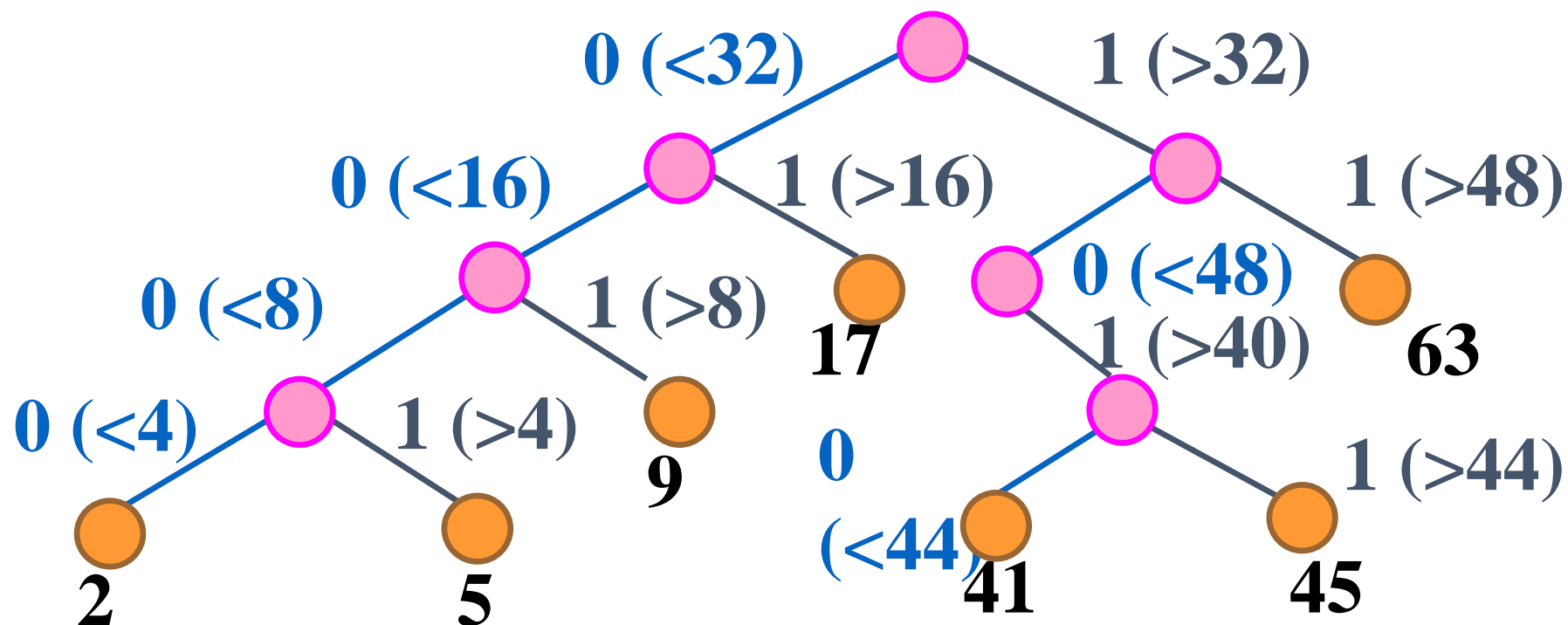
## Compact the Single Paths close to the leaf

Store words **an, and, ant, bad, bee**

## 12.4 Trie

## Binary Trie

Elements are 2, 5, 9, 17, 41, 45, 63

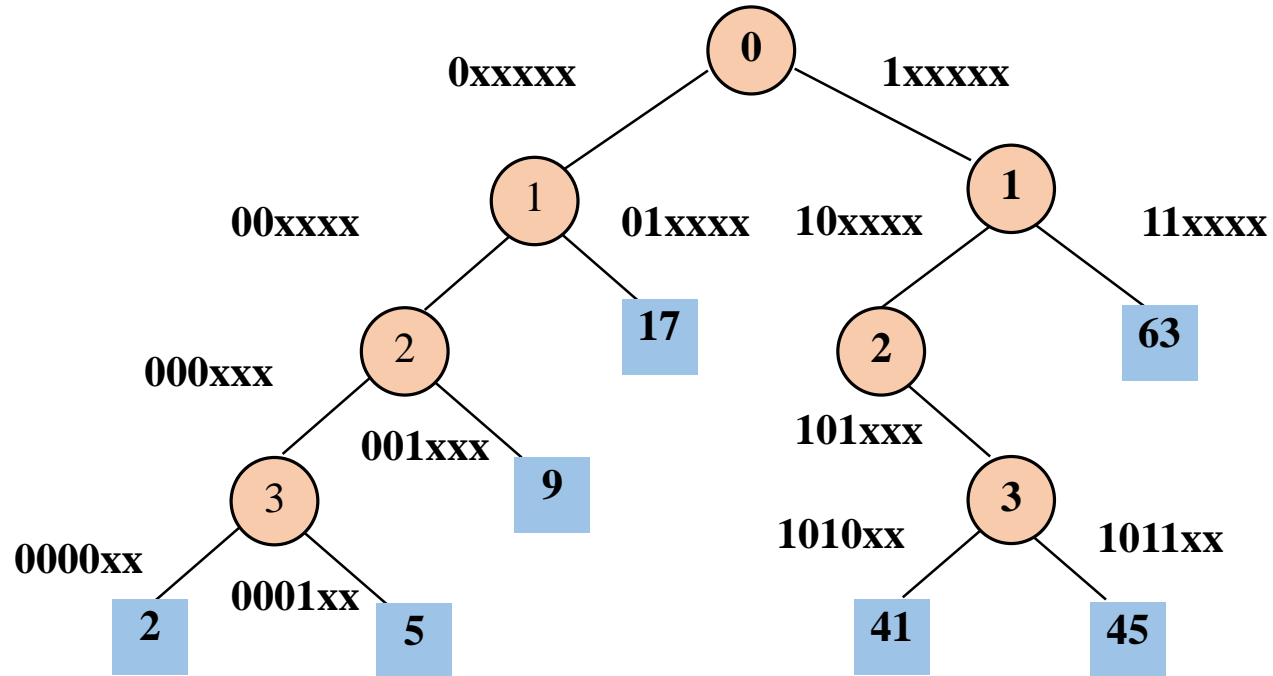






# PATRICIA Structure

## Compression



Code: 2: 000010 5: 000101 9: 001001

17: 010001 41: 101001 45: 101101 63: 111111



# Characteristics of PATRICIA Tree

- The compressed PATRICIA tree is a full binary tree
  - Every internal node represents a 1-bit comparison
  - Always at least two children are generated
- The number of comparisons will not exceed the length of the key



# Suffix Trees

Ascending Order

ababc

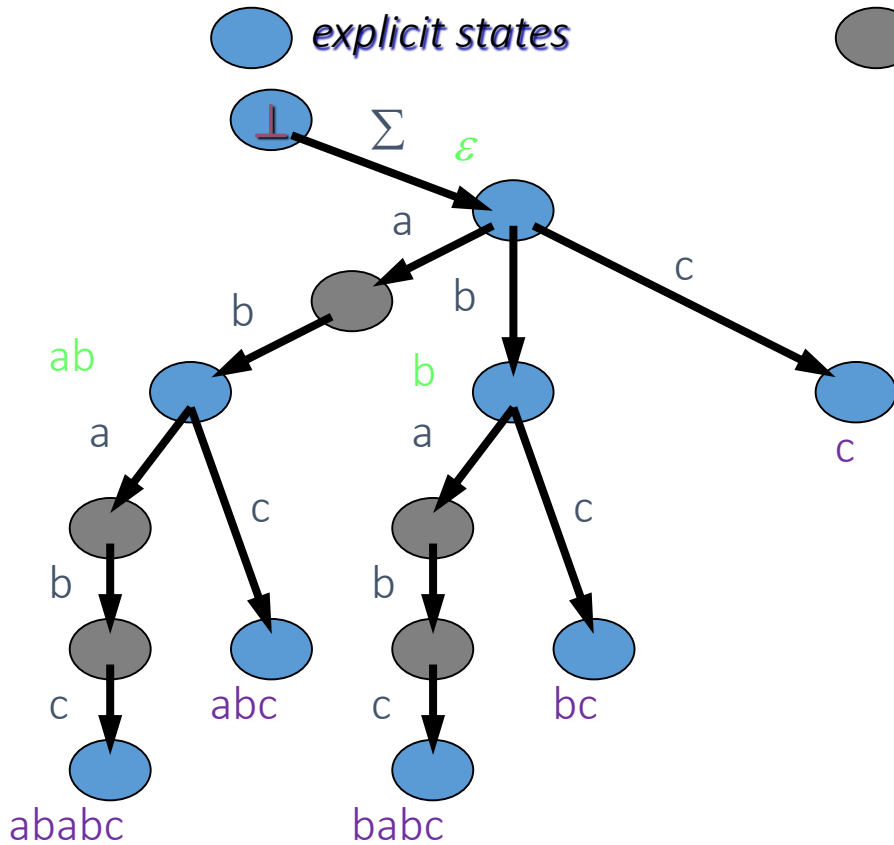
abc

abc

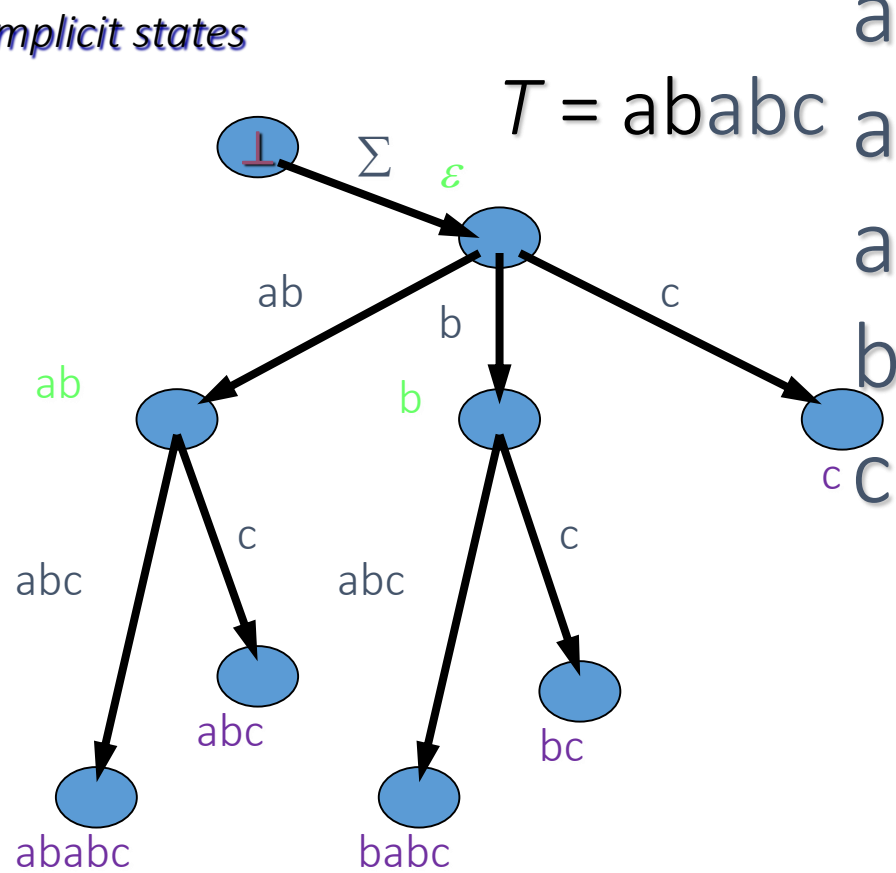
bc

c

$T = ababc$



Suffix Trie



Suffix Tree

## 12.4 Trie

# 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
---	---	---	---	---	---	---	---	---	---

Suffix Array

<b>3</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>-</b>
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

↑ **The longest common prefix array**

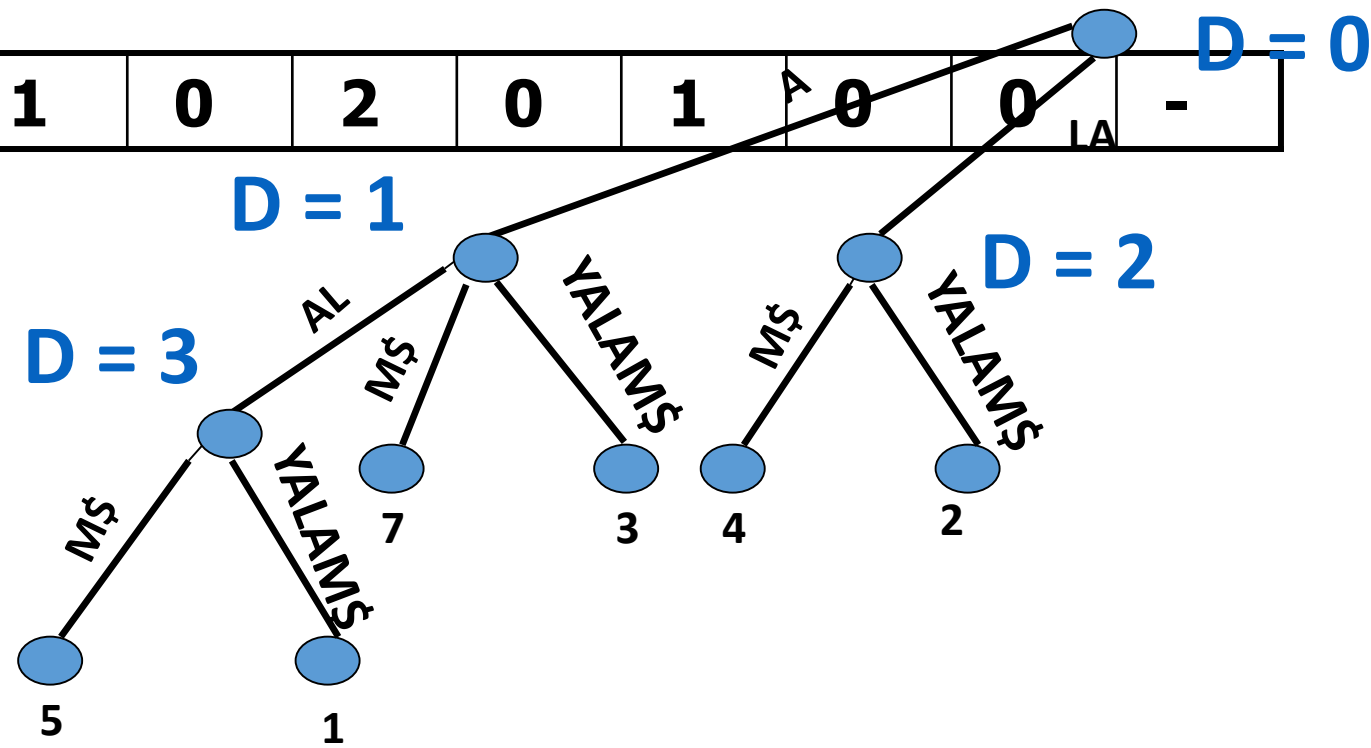
Suffix 5 and Suffix 1 share "ALA"

Suffix 1 and Suffix 7 share "A"      LCP always adjacent



# 12.4 Trie

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



SA

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

*lcp*



## 12.4 Trie

# Discussions

- Can Trie handle Chinese characters? What about PATRICIA Trie structure?
- Learn related document about Suffix Array and Suffix Tree. And think about their applications.



---

# Data Structures and Algorithms

## Thanks

the National Elaborate Course (Only available for IPs in China)

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

**Ming Zhang, Tengjiao Wang and Haiyan Zhao**

**Higher Education Press, 2008.6 (awarded as the "Eleventh Five-Year" national planning textbook)**