

# 词法分析核心算法 预习

正则表达式  $\square$  NFA

NFA  $\square$  DFA

DFA化简

**提示：PPT中设置了一些动画，  
不要直接翻页显示后续内容，  
先思考一下应该怎么做，再翻  
页看结果**

## 正则表达式□NFA

- 等价性是首要的——

接受的符号串集合是相等的

- 可视为一个翻译任务

□语法制导翻译方法

□为每个语法规则设计翻译

方法，即，为每个语法规则

设计NFA的转换方法

# 正则表达式 $\square$ NFA

- 正则表达式的语法规则定义?

基本正则表达式?

$\epsilon$ ,  $a$

正则表达式运算?

$|$ 、 $\cdot$ 、 $*$ 、 $()$

- 问题变为

$\epsilon$ ,  $a$ 如何转换为NFA?

$|$ 、 $\cdot$ 、 $*$ 、 $()$ 如何转换为NFA?

## 正则表达式 $\square$ NFA

- 接受  $\{\epsilon\}$ ,  $\{a\}$  的NFA?

几个状态几条边就够了?

- 接受  $L(r \mid s)$  的NFA?

既接受  $L(r)$ , 也接受  $L(s)$

注意,  $r$  和  $s$  的NFA已转换好

$\square$  从初态经过  $r$  和  $s$  都能到达

终态

$\square$   $r$  和  $s$  的NFA上下摆放, 增

加必要的状态和边

## 正则表达式□NFA

- 类似的

$r \bullet s$ : 接受 $L(r)$ 连接 $L(s)$

□从初态相继经过 $r$ 和 $s$ 方能

到达终态

□ $r$ 和 $s$ 的NFA前后放置, 增

加必要的状态和边

$r^*$ : 接受 $r^0$ 、 $r^1$ 、 $r^2$ 、 $r^3$ 、...

□要能从初态直达终态、经

过 $r$ 到达、经过多次

## NFA $\square$ DFA

- NFA  $N \square$  DFA  $D$
- 等价性是首要的

等价性另一个角度——

任意符号串 $x$ ,  $N$ 和 $D$ 识别它

的结果相同

- $x$ 扔到 $N$ 中最终得到什么?

一个状态子集 $T$

- $x$ 扔到 $N$ 中最终得到什么?

单一状态 $s$

## NFA $\square$ DFA

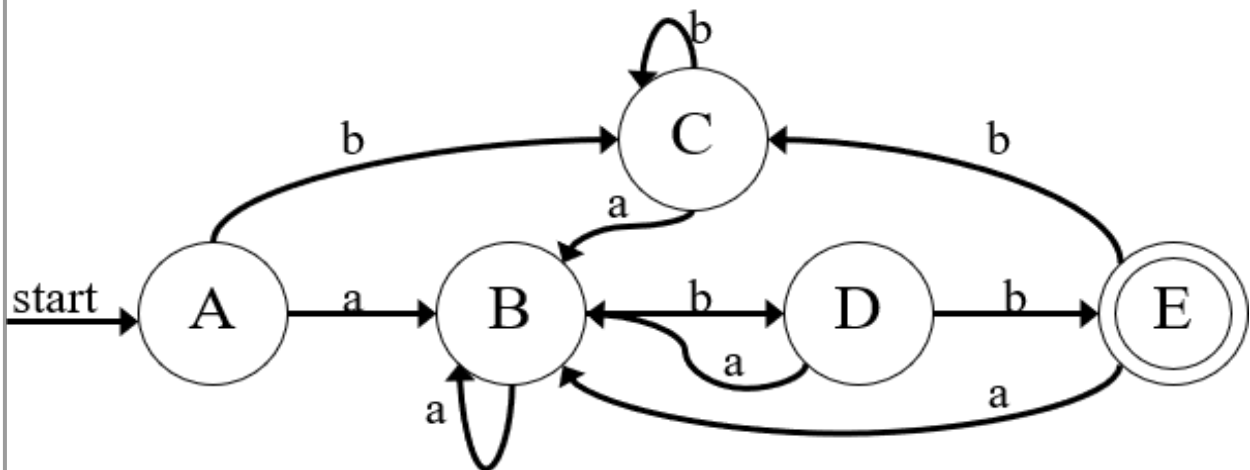
- D中状态s（未知）与  
N中状态集T（已知）对应  
 $\square$  我们可以用T表示s，我们  
已经构造出D的一个状态！
- 循此思路：有一个符号串x，  
我们扔给N，就得到一个状态集T，构造D的一个状态s
- 困难：符号串是无穷的！  
穷举x构造s不是一个算法！

## NFA $\square$ DFA

- 有序地枚举符号串，首先是最最简单的  
 $\epsilon$ ， $\epsilon$ 扔到N中得到？  
不只是初态，还有从初态开始只经过 $\epsilon$ 边就能到达的所有状态  
 $\square$ 这个状态集就是D的初态！
- 接下来反复计算已有状态  
(对应状态集) 的状态迁移



## DFA化简



- 对A和B两个状态，输入bb，  
一个到达非终态C，一个到达终态E，表明识别结果不同□bb可“区分”A和B
- A和C：任何符号串都无法区分—对词法分析无差别，  
可合并—实际算法是“分裂”

