

- 扫描处理的过程
- 词法描述：正则语言与正则表达式
- 词法识别和分析算法：有穷自动机
- 扫描程序的实现

- 问题理解

- 输入: 符号词法要求的字符序列
- 输入描述: 正则表达式的含义
- 输出: 记号序列
- 输出描述: 正则表达式的名称 (记号)

- 算法设计

- 扩展的有穷自动机

- 编程实现

- 用代码实现有穷自动机: 双层开关语句法

- 记号, 属性
- 正则表达式
 - 四则运算表达式, 外延, 内涵
 - 字母表, 字符串集合及其运算 (并, 连接, 闭包), 正则语言, 等价性
 - 正则语言编写手法: 分解与综合
 - 正则表达式的扩展及程序设计语言记号的正则表达式; 二义性, 最长子串原理, 分隔符, 先行 (向前看)

- 字符串识别与有穷自动机的状态转移图：状态, 状态转移, 初始状态, 接受状态
- 确定的有穷自动机及其化简, 陷阱状态
- DFA 的三种形态: 状态转移图, 数学定义, 状态转换表
- DFA 接受的句子和语言
- 记号对应的 DFA 及 DFA 的构建手法: 分解与综合, 并联, 串联
- DFA 作为分析算法的问题及解决方案: 最长子串原理, 非消耗读取与分隔符, 接受记号的标注, DFA 的合并

编程实现

- 朴素方法
- 双层嵌套开关语句法
- 表驱动法
- 从正则表达式自动转换



本章要求

- 扫描处理的过程及相关概念
- 正则语言与正则表达式的概念、关联与区别
- 有穷自动机的相关概念
- 有穷自动机与正则表达式的关联
- 有穷自动机的识别过程
- 使用于扫描程序的 DFA
- 从有穷自动机构造扫描程序的主要方法
- 正则表达式到 DFA 的等价转换算法

★ 本章作业

2.1, 2.4, 2.5, 2.7, 2.8, 2.9, 2.20, 2.27(a), 2.30

编 译 方 法

第二章 词法分析

第二部分

冯速

fengsu@bnu.edu.cn

北京师范大学

信息科学与技术学院