



哈爾濱工業大學  
HARBIN INSTITUTE OF TECHNOLOGY

**2021 年秋季学期**  
**计算学部 《软件安全》**

**Lab 4 实验报告**

姓名	余涛
学号	1180300829
专业	信息安全
班号	1803202
手机号码	15586430583

## 1、实验项目描述

### 1、基于特征匹配的误用检测技术原理和方法

- (1) 掌握多模式匹配方法
- (2) 掌握基于双数组-自动机原理的多模式匹配原理

### 2、基于双数组-AC 算法的多模式特征匹配算法实现

- (1) 双数组-自动机的预处理：Next 表、Base 表、Check 表、失效函数、输出函数构建
- (2) 双数组-自动机的特征扫描流程
- (3) 合理的数据结构

### 3、利用构建的自动机扫描目标文件

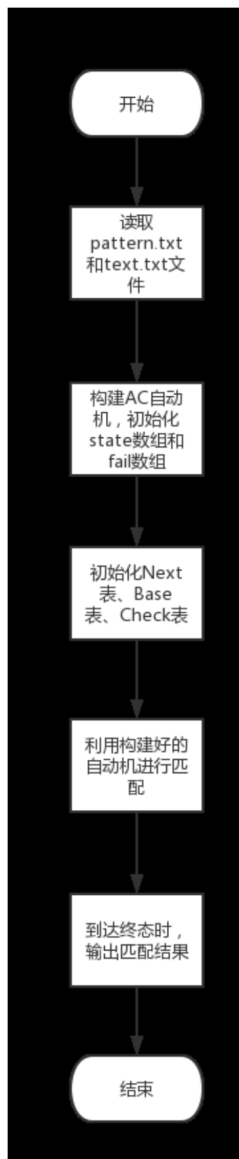
- (1) 扫描文件发现命中某个模式，需明确给出命中的模式和该模式在文件中的位置
- (2) 如命中多个模式，需全部列出

## 2、实验要求

- 1、实验数据准备。实验数据应简单实用：利用 ASCII 字符集做为输入集，不考虑多字节编码的中文、英文字符集。模式和待匹配文本可以只包含英文。
- 2、程序的输入部分（模式、待匹配文本）需以文件方式提供或者其它非固定的输入途径。
- 3、实验结果和实验数据一起给出：包括设定的模式有哪些？Next 表、Base 表、Check 表是什么？待匹配文本是什么？输出结果是什么。注意：仅给出匹配结果是不够的，必须在输入多模式后，给出 Next 表、Base 表、Check 表内容，Next 表中空间占用的百分比( $\text{Next 表中状态占用的空间} / \text{Next 表大小}$ )，然后输入待检测文本，再输出检测结果。
- 4、程序本身需要提交。

### 3、实验结果

#### 1、程序的流程图



#### 2、列出构建自动机所定义的数据结构，简单说明其功能

这几个数据结构的功能如下（ASCII 码包含 256 个字符，所以各表大小均设置为 256）：

（1）Next 为转向函数表（数组、链表），下标是位置偏移量，输出是状态值。

（2）Base 表（数组），下标是状态值，输出是 Base 值。Next 表中当前状态为  $s$ ，输入为  $c$  时，假设应跳转为状态  $t$ ，状态  $t$  在 Next 表中的位置 = 状态  $s$  的位置 + 状态  $s$  的 Base 值 + 输入  $c$  的 ASCII 码值。

(3) Check 表 (数组), 下标是状态值, 输出是下标状态的父状态的值

### 3、根据这些数据结构说明 Next 表、Base 表、Check 表、失效函数、输出函数的构建过程

只需要对初始 AC 算法进行修改即可。在先广遍历 AC 树的过程中逐层初始化 Next 表、Base 表、Check 表, 失效函数与转向函数一起构造, 输出函数和以前一样。Check 表下标与 Base 表一致, 值为下标状态的父状态的值。

在初始化 Next 表时, 首先选择第一层中按 ASCII 序最小的一个字符, 将其赋值给 Next[1], 并且令其状态为 1, 根节点中其余字符在 Next 中的位置按在 ACSCII 序中与前述最小字符的相对位置进行计算。Base 表和 Next 表其余值根据下述表达式进行计算:

$t \text{ 在 Next 中位置} = s \text{ 的位置} + s \text{ 的 base 值} + \text{输入 } c \text{ 的 ASCII 码值}$

。

## 4、数据结果:

```
All patterns are as follows:
he
his
she
hers
next table:
0 1 3 5 6 7 4 8 9 0 0 0 2
base table:
-103 -99 -101 -110 -110 -94 -107 9 9 9
check table:
0 0 0 1 1 2 3 4 5 6 0 0 0
fail:
0 0 0 0 0 1 0 2 3 2
The text to be matched is:he she is a he

All detected patterns are as follows:
0   he
3   she
4   he
```