

正则表达式-自动机转换工具 开发文档

小组成员:

张许可 2009212639

耿卉卉 2009212626

2009/11/26

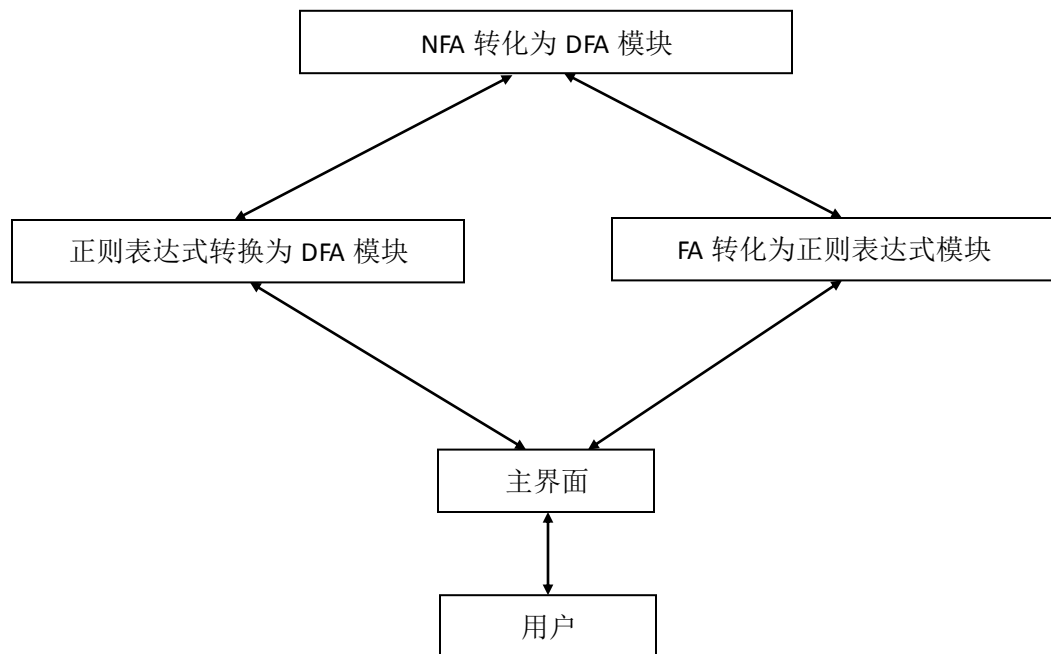
正则表达式-自动机的转换工具开发文档

- 正则表达式-自动机的转换工具开发文档-----1
- 1 开发环境 -----3
- 2 系统模块图 -----3
- 3 功能模块描述 -----3
 - 3.1 正则表达式转换为确定自动机（DFA） -----3
 - 3.2 有限自动机（FA）转换为正则表达式 -----4
- 4 研究路线的简要介绍（实现算法） -----5
- 5 项目进度预期安排 -----6
- 6 一些说明事项 -----6

1 开发环境

- Microsoft Windows XP Professional SP3
- Pentium(R) Dual-Core CPU E5300 @ 2.60GHz, 2G 内存, 406G 硬盘
- Eclipse Integrate Develop Environment
- Java 语言

2 系统模块图



3 功能模块描述

该转换工具实现了正则表达式和确定自动机的相互转换, 下面是相应的转换功能的一些简要描述。

3.1 正则表达式转换为确定自动机 (DFA)

基于最小正则表达式经过并运算、连接运算、括号运算和闭包运算产生的所有正则表达式的转换。先将正则表达式转换为非确定自动机 (NFA), 然后将非确定自动机转

换为确定自动机（DFA）。示例：

输入正则表达式 $r=(a|b)^*abb$ 可以生成以下的 NFA

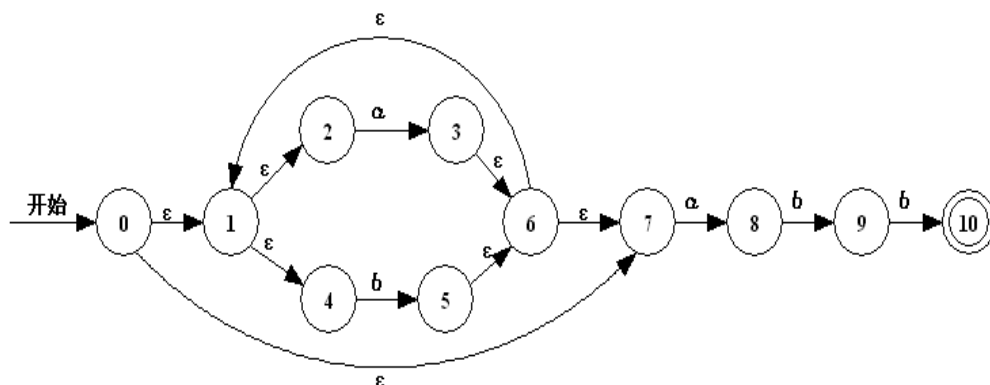


图 3-1 正则表达式 $r=(a|b)^*abb$ 生成的 NFA

由图 3-1 所示的 NFA 转换生成的 DFA

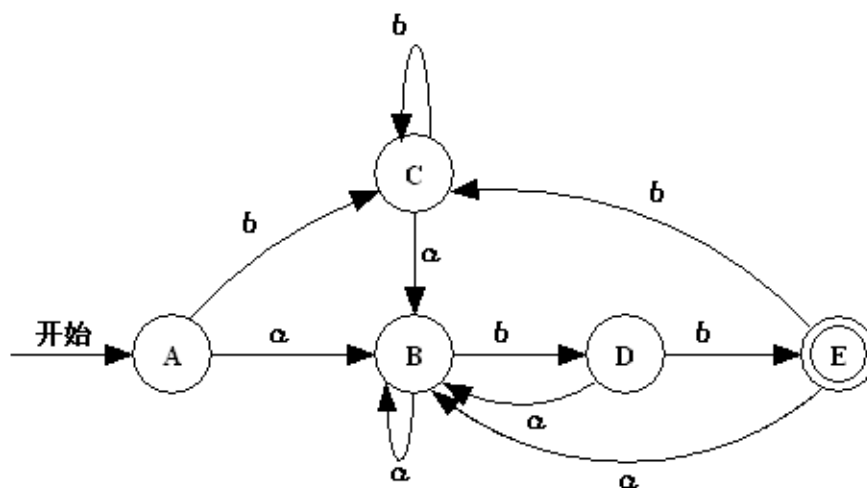


图 3-2 正则表达式 $r=(a|b)^*abb$ 生成的 DFA

注：由正则表达式到 NFA 和由 NFA 到 DFA 两步转换拟采用的算法见 [4] 中研究路线的简要介绍。

3.2 有限自动机（FA）转换为正则表达式

将给定的有限自动机转换为正则表达式，分为三步：

- 1、若给定一个非确定自动机（NFA）的描述（状态转移表），首先将其转换为确定自动机（DFA），转 3
- 2、若给定一个确定自动机的（DFA）的描述（状态转移表）

3、画出确定自动机（DFA）的状态转移图，然后将该 DFA 转换为相应的正则表达式进行输出。

示例：

表 3-1 DFA 状态转移表

	a	b
→A	B	C
B	B	D
C	B	C
D	B	E
*E	B	C

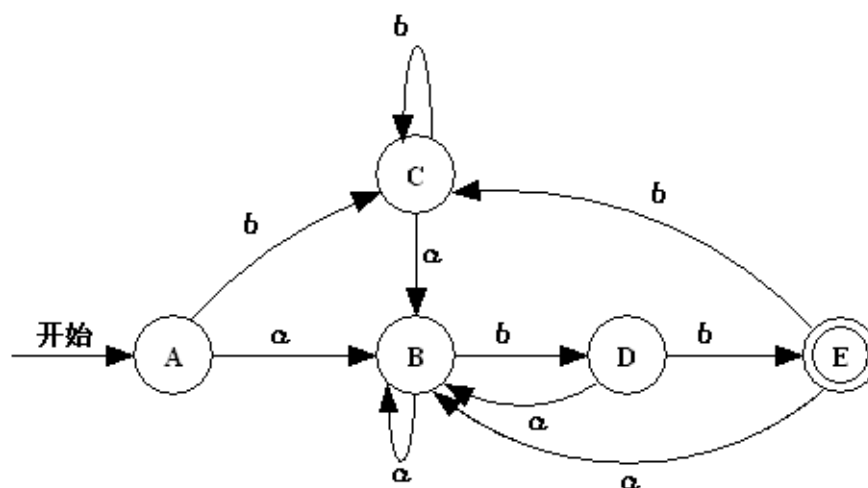


图 3-3 表 3-1 表示的 DFA

该 DFA 生成相应的正则表达式： $(a|b)^*abb$ ，然后输出。

注：由 NFA 到 DFA 和由 DFA 到正则表达式两步转换拟采用的算法见 [4] 中研究路线的简要介绍。

4 研究路线的简要介绍（实现算法）

简要介绍在实现正则表达式和确定自动机的相互转换中拟采用的算法，由于这些算法在课本中都有描述，在这里不再进行详述。

4.1 正则表达式转换为非确定自动机（NFA）拟采用算法

具体的实现将采用 Thompson 构造法。关于该方法的详细描述见课本 3.2.3。

4.2 非确定自动机（NFA）转化为确定自动机（DFA）拟采用算法

具体的实现将采用含惰性求值的子集构造法。关于该方法的详细描述见课本 2.3.5。

4.3 确定自动机（DFA）转化为正则表达式拟采用算法

具体的实现将采用状态消去法。关于该方法的详细描述见课本 3.2.2。

5 项目进度预期安排

下面是项目进度的预期安排时间表，可能会由于课程问题略有出入，开始时间为 2009 年 11 月 30 日：

表 5-1 项目进度预期安排表

功能模块	预计时间	完成标志
正则表达式->NFA	2~3 周	正确实现相应的转换功能
DFA->正则表达式	2~3 周	正确实现相应的转换功能
NFA-> DFA	2~3 周	正确实现相应的转换功能

6 一些说明事项

可能由于对项目的认识程度不够，或者由于自身能力有限，在开发过程中，如果遇到一些难以解决的问题，希望得到老师和助教的帮助。

如果在项目开始一段时间后，感觉不能很好的开发出项目所要求实现的全部功能，愿意参加课程考试。