

编译原理 Flex 作业

2015.10.19

作业目的

- ✂ 掌握 Flex 的基本使用
- ✂ 体验词法分析的过程
- ✂ 练习正则表达式

作业概览

✂ IPv4 地址识别与分类

✂ DOT 语法文件 token 划分

作业一：IPv4 地址识别与分类

IPv4 地址简介

IPv4 地址是 IP 协议提供的一种统一的地址格式，它为互联网上的每一个网络和每一台主机分配一个逻辑地址，以此来屏蔽物理地址的差异。

IPv4 地址的一般表示方式：

×××.×××.×××.×××

通过 "." 分隔，由 4 个区段组成，每个区段为十进制数字 0-255（省略数字前面多余的 0）。

IPv4 地址分类

每个 IP 地址包括两个标识码（ID），即网络 ID 和主机 ID。

Internet 委员会根据网络 ID 定义了 5 种 IP 地址类型以适合不同容量的网络，即 A 类 ~ E 类。

类别	IPv4 地址范围
A	0.0.0.0-127.255.255.255
B	128.0.0.0-191.255.255.255
C	192.0.0.0-223.255.255.255
D	224.0.0.0-239.255.255.255
E	240.0.0.0-255.255.255.255

作业说明

- ✂ 从标准输入中逐行读取待检验的字符串。
- ✂ 使用正则表达式匹配 IPv4 地址，若匹配失败则输出 Invalid，若匹配成功则输出 IPv4 地址的类别。
- ✂ Flex 中必须使用 C 代码，无需自定义 main、yywarp 函数，gcc 编译时添加 -lfl 或者 -ll 选项即可自动链接这些函数。

格式说明

- ✂输入格式：从 input.txt 中读取 IPv4 地址，每行一个 IPv4 地址。
- ✂输出格式：处理结果输出到 output.txt，每行一个。
- ✂匹配成功的必要条件：该行只包含一个合法的 IPv4 地址，没有任何其它字符。
(输入输出使用重定向就可以很方便的做到，无需在 flex 加入文件操作。)

输入	输出
10.0.0.1	A
20.0.00.1	Invalid
192.168.1.1	C
172.168.18.88	B

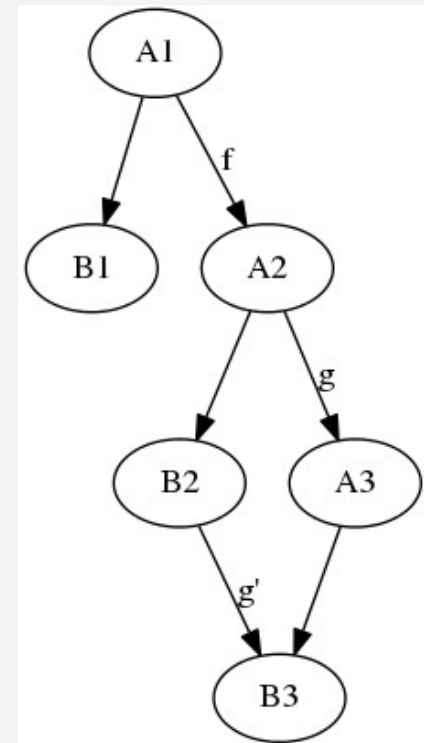
作业二：DOT 语法文件 token 划分

DOT 介绍

DOT 是一种图形描述语言，通过简单的语法即可定义图形，尤其适合流程图、状态机、网络图。

Graphviz 是 DOT 的绘图工具，核心是布局引擎，布局算法有：dot, neato, twopi, circo, fdp, sfdp，图形输出常见格式有：jpg, png, gif, ps, svg, pdf。

```
1 digraph g {  
2     A1 -> B1;  
3     A2 -> B2;  
4     A3 -> B3;  
5  
6     A1 -> A2 [label=f];  
7     A2 -> A3 [label=g];  
8     B2 -> B3 [label="g'"];  
9 }
```



DOT 介绍

详细语法说明参考

- <http://www.graphviz.org/content/dot-language>

图形类型

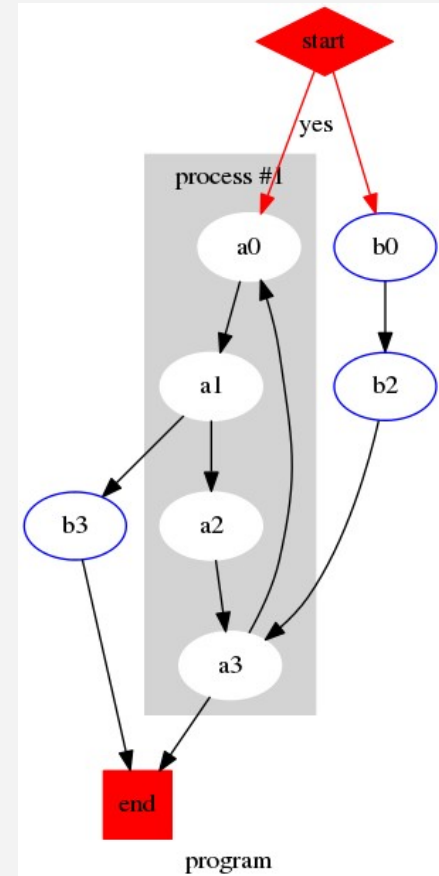
- graph(无向图) , digraph(有向图) , subgraph(子图) , strict(节点间无重复边)

组成元素

- node
 - 单个单词可省略双引号
- edge
 - "--" : 无向边
 - "->" : 有向边
- attribute
 - 全局属性
 - 图属性, 节点属性, 边属性
 - 单个节点属性
 - 单条边属性
- 注释
 - // , /* */, #

DOT 介绍

```
1 digraph G {
2   label = "program"
3   node [color=blue];
4   edge [size=0.75,not_exist=any];
5
6   subgraph cluster_0 {
7     style=filled;
8     color=lightgrey;
9     node [style=filled;color=white];
10    a0 -> a1 -> a2 -> a3;
11    label = "process #1";
12  }
13
14  start -> a0 [color=red,label="yes"];
15  start -> b0 [color=red];
16  b0 -> b2;
17  a1 -> b3;
18  b2 -> a3;
19  a3 -> a0;
20  a3 -> end;
21  b3 -> end;
22
23  start [shape=Mdiamond,color=red,style=filled,fillcolor=red];
24  end [shape=Msquare,color=red,style=filled,fillcolor=red];
25 }
```



Graphviz 安装使用

✂ Windows

下载安装即为图形界面：

http://www.graphviz.org/Download_windows.php

✂ Linux

一般系统内置，否则可通过 `sudo apt-get install graphviz` 安装。

每种布局算法其名字对应一个命令行工具，如最常用的 `dot`，绘图命令为：`dot graph.dot -T pdf -o graph.pdf`，其中 `-T` 指定输出格式。

作业说明

- ✂ 从标准输入中读取 DOT 文件，文件无词法、语法错误。
- ✂ 根据 token.txt 使用正则表达式匹配 token 并且输出。
- ✂ Flex 中必须使用 C 代码，无需自定义 main、yywarp 函数，gcc 编译时添加 -lfl 或者 -ll 选项即可自动链接这些函数。
- ✂ DOT 关键词大小写不敏感。
- ✂ DOT 只包含英文字母和符号，即在 ASCII 范围内。
- ✂ DOT 的 ID 与 C 语言相同，STRING 里面可包含转义的双引号，即\"。
- ✂ DOT 无需支持的特性
 - ✂ HTML STRING，冒号域（把这些特性加上，DOT 的整个词法就完整了）

格式说明

✂输入格式：把 input.dot 重定向到标准输入。

✂输出格式：处理结果输出到 output.txt，每行一个 token。

（输入输出使用重定向就可以很方便的做到，无需在 flex 加入文件操作。）

部分输入

```
digraph G {  
    label = "program"  
    node [color=blue;  
    edge [size=0.75,not_exist=any;  
  
    subgraph cluster_0 {  
        style=filled;
```

部分输出

```
digraph  
G  
{  
label  
=  
"program"  
node  
[  
color  
=  
blue  
]  
;
```

作业提交

提交文件：**ipv4.flex dot.flex**

提交时间：**10月31号晚12.00前提交**

提交方式：**由学委统一收齐发给TA**

命名规范：**学号 - 名字拼音 -hw1.zip**

如：**12345678-xiaoming-hw1.zip**

Reference

IP 地址 - 百度百科

http://baike.baidu.com/link?url=Prm0C63BuZNduZaKpT-2SgvUZG4usoonA7VnGct21wllds4pCcQGQawXaPnZSBCyOyPd4qUUgkkPoVewdBWXV_

Graphviz 官网

<http://www.graphviz.org/>

DOT 语法参考

<http://www.graphviz.org/content/dot-language>