

# 这是一封简短的实验报告

为识别词法，我们需要做到识别ident，注释，数字和关键字。同时为了识别行号和列号，我定义了整型的lineno和colno。

## 对换行符：

- 当识别到时，lineno++，colno=1

## 对ident

- 构造正则表达式为[a-zA-Z][a-zA-Z0-9]\*
- 若识别到，colno加等长度数，下面都是如此。

## 对数字

- 构造正则表达式以匹配8进制，10进制和16进制的数
  - 0[0-7]\*
  - [1-9][0-9]\*
  - 0x[1-9a-fA-F][0-9a-fA-F]\*
  - 0X[1-9a-fA-F][0-9a-fA-F]\*
  - 0x0
  - 0X0

## 对注释

- 对单行注释：
  - 使用"//".\\*来匹配，由于这一行都被匹配了，所以没有统计colno的必要。同时，由于.不会匹配换行符，所以换行符仍可以被识别并进行相关lineno和colno操作
- 对多行注释
  - 使用%x COMMENT 定义环境COMMENT，通过下列代码，在识别到/\*时进入COMMENT。此时不需要识别任何数字或ident，故仅用.和\n来识别字符即可。在识别到\*/退出COMMENT环境，回到最初的INITIAL环境。

```
"/*"      { colno += strlen(yytext);BEGIN(COMMENT); }
<COMMENT>"*/" { colno += strlen(yytext);BEGIN(INITIAL); }
<COMMENT>.    {colno += strlen(yytext);}
<COMMENT>\n   {lineno++; colno = 1;}
```

## 对关键字

- 枚举了所有可能的关键词，并将其放至ident识别代码前，使其不被识别为ident

```
"int"|"void"|"const"|"if"|"while"|"break"|"continue"|"return"|"+"|"-"|"*"|"/"|"="|"%"|"=="|"!="  
printf("%-25s%-3d  (%d,%d)\n", yytext, lineno, colno,colno + strlen(yytext));  
colno += strlen(yytext);  
}
```

## 对于错误命名

- 由于示例1里有错误命名，故用 `[0-9][a-zA-Z0-9_]+` 来匹配处理了一下。也就是说ident不能以数字开头。

## 对其他字符（基本上也就是空格了）

- 直接使用  来匹配，并将其优先级放至最低。