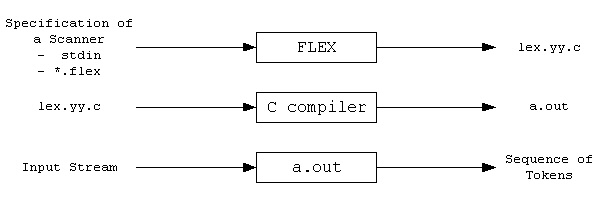
1. **Định nghĩa**

Flex là một công cụ để tạo scanner: các chương trình phân tích từ vựng trong văn bản. flex đọc các tệp đầu vào đã cho (hoặc đầu vào tiêu chuẩn của nó nếu không có tên tệp nào được đưa ra) để mô tả về máy quét để tạo. Mô tả ở dạng cặp biểu thức chính quy và mã C, được gọi là quy tắc. flex tạo ra output C, `lex.yy.c ', định nghĩa một routine yylex. Biên dịch và liên kết tệp này với thư viện `-lfl 'để tạo tệp executable. Khi executable chạy, nó sẽ phân tích đầu vào của nó cho các lần xuất hiện của các biểu thức chính quy. Bất cứ khi nào nó tìm thấy một biểu thức chính quy , nó thực thi mã C tương ứng.



First, FLEX reads a specification of a scanner either from an input file \*.lex, or from standard input, and it generates as output a C source file *lex.yy.c*. Then, *lex.yy.c* is compiled and linked with the "-lfl" library to produce an executable *a.out*. Finally, *a.out* analyzes its input stream and transforms it into a sequence of tokens.

Đầu tiên FLEX sẽ đọc các thông số của bộ scanner trong file .lex và chuyển thành mã nguồn C : lex.yy.c . Sau đó file lex.yy.c đươc biên dịch và liên kết với thư viện “-lfl” để tạo thành file thực thi a.out , file này sẽ thực hiện phân tích đầu vào và chuyển đổi ra thành chuỗi token.

1. **Cách matching input**

Khi scanner được chạy, nó sẽ phân tích đầu vào của nó để tìm các chuỗi khớp với bất kỳ mẫu nào của nó. Nếu nó tìm thấy nhiều hơn một kết quả khớp, nó sẽ lấy một văn bản khớp với hầu hết các văn bản. Nếu nó tìm thấy hai hoặc nhiều kết quả khớp có cùng độ dài, quy tắc được liệt kê đầu tiên trong tệp đầu vào flex được chọn.Sau khi so khớp được xác định, văn bản tương ứng (được gọi là token) sẽ được cung cấp trong con trỏ global yytext và độ dài của nó trong global yyleng. Hành động tương ứng với mẫu phù hợp sau đó được thực hiện (mô tả chi tiết hơn về các hành động tiếp theo) và sau đó đầu vào còn lại được quét cho một kết quả khớp khác.Nếu không tìm thấy kết quả khớp, thì quy tắc mặc định được thực thi: ký tự tiếp theo trong đầu vào được coi là khớp và được sao chép vào đầu ra tiêu chuẩn. Do đó, đầu vào flex đơn giản nhất là:%%mà tạo ra một máy quét chỉ đơn giản là sao chép đầu vào của nó (một ký tự một lần) vào đầu ra của nó.

1. **Thành phần**

Đầu vào file .lex bao gồm 3 phần với %% ngăn cách :

***definitions***

***%%***

***rules***

***%%***

***user code***

Ví dụ với bài toán xác định các kí tự là hoa hay thường và đếm số kí tự hoa, thường

* Phần ***definitions*** để định nghĩa các biến sẽ dùng trong yylex() và hàm main :

*%{*

*int count = 0;*

*%}*

* Phần ***rule*** sẽ liệt kê các quy tắc xử lí khi gặp một kí tự nào đó , gồm 2 yếu tố là pattern và action :

*[A-Z] {printf("%s capital letter\n", yytext);*

*count++;}*

*. {printf("%s not a capital letter\n", yytext);}*

*\n {return 0;}*

* Ở đây quy tắc đầu tiên khớp với chữ in hoa, quy tắc thứ hai phù hợp với bất kỳ kí tự nào ngoại trừ dòng mới và quy tắc thứ ba không nhận đầu vào sau enter

**3.1 Pattern:**

`.' tất cả kí tự trừ newline

`[xyz]' là x, y hoặc z

`[abj-oZ]' là a,b , từ j → o và Z

`[^A-Z]' bất kì kí tự nào trừ các chữ cái viết hoa

`[^A-Z\n]' bất kì kí tự nào trừ các chữ cái viết hoa hoặc a newline

`r\*'

zero or more r's, where r is any regular expression

`r+'

one or more r's

`r?'

zero or one r's (that is, "an optional r")

`r{2,5}'

anywhere from two to five r's

`r{2,}'

two or more r's

`r{4}'

exactly 4 r's

`{name}'

the expansion of the "name" definition (see above)

`"[xyz]\"foo"'

the literal string: `[xyz]"foo'

`\x'

if x is an `a', `b', `f', `n', `r', `t', or `v', then the ANSI-C interpretation of \x. Otherwise, a literal `x' (used to escape operators such as `\*')

`\0'

a NUL character (ASCII code 0)

`\123'

the character with octal value 123

….

**3.2 Action**

- Mỗi mẫu trong một rule có một hành động tương ứng, có thể là bất kỳ câu lệnh C nào. Pattern kết thúc ở ký tự khoảng trắng không thoát đầu tiên; Phần còn lại của dòng là action của nó.

- Nếu hành động trống, thì khi mẫu được khớp, input token sẽ bị loại bỏ.

- 1 action chỉ có ‘|’ tức là nó giống với action dưới

- Action nhiều dòng sẽ được cho vào block ‘{}’