

Manuel de FLEX

1. FLEX :

FLEX est un générateur d'Analyseur Lexical. Il est souvent utilisé avec un analyseur syntaxique. Un Analyseur Lexical est un programme permettant d'analyser un flux de chaînes de caractères et le segmenté en lexèmes (Tokens). Ces derniers représentent les entités lexicales d'un langage. A titre d'exemple : les identifiants de variables et de fonctions, les mots clés, les opérateurs, etc...

Un document FLEX comporte trois parties principales dont chacune d'elles est encadrée par le symbole « %% »:

1.1. Première partie: Les expressions régulières

Cette section permet de donner les expressions régulières des différentes entités lexicales (lexème) de notre langage de programmation sous la forme suivante :

<identificateur_de_l'entité> <expression régulière>

- *Un identificateur_de_l'entité* : Doit commencer par une lettre, et ne comportant que des caractères alphanumériques, des underscores (_) et des tirets (-).
- *Une expression régulière* : Doit être une *expression régulière* valide.

Exemple :

Entité lexicale Expression régulière

chiffre [0-9]

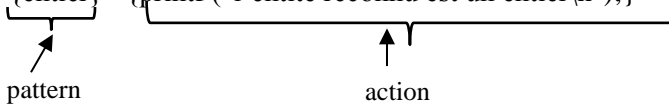
entier {chiffre} + /*utilisation d'une entité lexicale déjà définie*/

1.2. Deuxième partie : Les règles de traduction

Cette section est capitale. Elle comporte l'ensemble des actions associées à chaque entité lexicale. Une règle de traduction est de la forme suivante : **<pattern> <action>**

- **Un pattern**: c'est l'expression régulière décrivant un lexème.
- **Une action**: c'est le code C qui sera exécuté à chaque fois où le lexème correspondant apparaît.

Exemple : `{entier} {printf ("l'entité reconnu est un entier\n");}`



1.2. Troisième partie : Code additionnel

C'est la dernière section d'un document FLEX. Elle contient le **post-code C**. Ce dernier représente le code que nous voulons exécuter. Il y sera recopié tel quel à la fin du fichier lex.yy.c.

Exemple :

```
// Déclarations C (pré-code)
% {
    # include <stdio.h>
% }

// Définitions

chiffre [0-9]
entier {chiffre} +
reel {chiffre} + "." {chiffre} *

%%

// Règles de traduction

{entier} {ECHO; printf ("\n"); }
{reel} {printf ("%s", yytext); printf ("\n"); }

%%

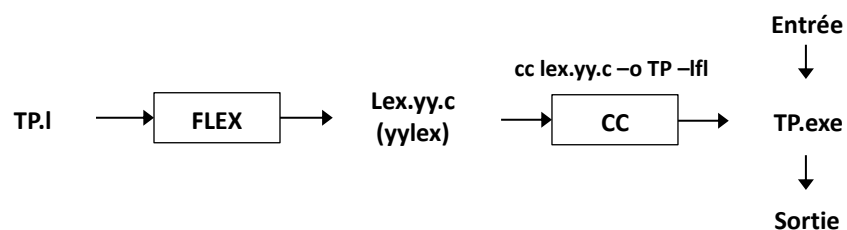
// Code Additionnel (post-code C)
int main () {
    yylex ();
    return 0 ;
}
```

2. Macro-action de FLEX :

<i>Fonctions</i>	yylex	Permet de lancer l'analyseur lexical.
	yywrap	Elle est appelée par le lexer quand il rencontre la fin du fichier. Elle doit, soit obtenir un nouveau flux d'entrée et retourner simplement la valeur 0, soit renvoyer 1, signifiant que la totalité des flux a été consommée et que le lexer a fini sa tâche.
	yyterminate	Permet de provoquer la fin d'exécution du lexer.
	ECHO	Affiche l'unité lexicale reconnue (équivalente à <code>printf("%s", yytext)</code>)
<i>Variables</i>	yytext	Récupère le texte formant le lexème reconnu.
	yylen	Détermine la longueur du texte contenue dans <i>yytext</i>
	yyval	Est une variable globale utilisée par FLEX pour stocker la valeur correspondante au Token reconnu.
	yylineno	Est le numéro de la ligne courante.
	yyin	Fichier d'entrée.
	yyout	Fichier de sortie.

3. Commandes de Compilation :

FLEX lit un fichier de l'extension (*.l), et génère un code C (lex.yy.c) pour la compilation.



Pour compiler le programme : \$ flex TP.l
 \$ cc lex.yy.c -o TP -lfl

Pour exécuter le programme : \$./TP

Pour arrêter le programme : \$ Ctrl+c

4. Caractères spéciaux :

" \ [] ^ - ? . * + | () \$ / { } % < >

Pour utiliser ces caractères comme "caractères ordinaires", il faut les protéger en plaçant dans une chaîne entourée de double-quotes (") ou en les plaçant après un \. Les caractères \n, \t correspondent respectivement au saut de ligne et à la tabulation.

5. Expressions régulières :

" : Une chaîne de caractères entourée par double-quotes représente la chaîne elle-même.

"**abc**" signifie la chaîne **abc**

[] : Une chaîne de caractères entre crochets représente un de ses éléments. Dans ce contexte, « | » et «->» indique un intervalle et ^ désigne l'exclusion.

[xyz] : x, y ou z

[a-zA-Z] : toutes les lettres minuscules et majuscules

[^0-9] : tous les caractères sauf les chiffres

. : Tout caractère sauf \n

| : Opérateur d'alternance

x|y|z : équivalent à [xyz]

[a-z][A-Z] : toutes les lettres minuscules et majuscules

* et + : Opérateur de répétition (* : zéro ou plusieurs fois, + : une ou plusieurs fois)

(x|y)* : chaîne de longueur positive ou nulle constituée des caractères x ou y

[a-z]+ : chaîne de longueur strictement positive constituée de lettres minuscules

? : Opérateur d'occurrence zéro ou une fois

ab?c : chaîne **abc** ou chaîne **ac**

/ : Condition de reconnaissance

ab/cd : chaîne **ab** seulement si elle est suivie de la chaîne **cd**

\$ et ^ : Début de ligne et fin de ligne

ab\$: chaîne **ab** en fin de ligne

^ab : chaîne **ab** seulement si elle est en début de ligne (après **\n** ou **\$**)

{} : Opérateur de répétition bornée - Définition

a{1, 5} : chaîne de longueur comprise entre 1 et 5 constituée du caractère **a**

a{2, } : chaîne de longueur supérieure ou égale à 2 constituée du caractère **a**

a{2, 2} : chaîne de longueur 2 constituée du caractère **a**

a{digit} : chaîne prédéfinie de nom **digit**

Expression	Exemple
abc	abc
abc*	ab, abc, abcc, abccc, ...
abc+	abc, abcc, abccc, ...
a(bc)+	abc, abcbcb, abcbcbcb, ...
a(bc) ?	a, abc
[abc]	a, b, c
[a-z]	a, b, c, d, ... z
[a\ -z]	a, -, z
[-az]	-, a, z
[a-zA-Z0-9]+	Un ou plusieurs caractères alphanumériques
[\t\n]+	Espaces
[^ab]	Tous les caractères sauf : a, b
[a^b]	a, ^, b
[a b]	a, , b
a b	a ou b