

Analyse Syntaxique - Année 2010-2011

Feuille 1

Automates et Expressions Rationnelles

Exercice 1.1

On considère le programme en C suivant :

```
main(){
    int a, b, c;
    a = 7.2;
    char *l'heure = "jjmmaa";
    float x = a, 3
    for (int i = 0; i < a; ++i){
        b = b + i*i
        c = c * (i+1);;
    'd'=5;
    printf("resultat: %d", c+d);
}
```

Indiquer les erreurs contenues dans ce programme et pour chacune d'elles, préciser la phase de compilation au cours de laquelle elle serait détectée : analyse lexicale, analyse syntaxique, ou phase ultérieure (vérification de types,...) .

Exercice 1.2

On considère le programme en C suivant :

```
main(){
    int a, b, c;
    a = 7;
    for (int i = 0; i < a; ++i){
        b = b + i*i;
        c = c*b;
    }
    affiche("resultat",b,c);
}
```

Indiquer sur ce programme quels sont tous les lexèmes contenus : mots réservés, identificateurs, opérateurs, constantes, séparateurs.

Exercice 1.3 *Constantes entières en C.*

L'ouvrage de référence du langage C indique les points suivants pour définir les constantes entières :

Une constante entière est composée d'une séquence de chiffres, si elle commence par 0 (le chiffre 0) elle est écrite en octal sinon en décimal. Les constantes en octale ne contiennent pas

les chiffres 8 ou 9. Une séquence de chiffres précédée de $0x$ ou $0X$ est considérée comme un entier hexadécimal, les chiffres hexadécimaux vont de a ou A à f ou F .

Donner un automate qui reconnaît les entiers du langage C, on distinguera dans les états terminaux, ceux qui reconnaissent les entiers décimaux, octaux ou hexadécimaux.

Exercice 1.4 Constantes flottantes en C.

L'ouvrage de référence du langage C indique les points suivants pour définir les constantes flottantes :

Une constante flottante se compose d'une partie entière, d'un point décimal, d'une partie fractionnaire, d'un e ou d'un E , d'un exposant entier éventuellement signé, et d'un suffixe de type éventuel, à savoir f , F , l ou L . Chacune des parties entières et fractionnaires se compose d'une séquence de chiffres. On peut omettre la partie entière ou la partie fractionnaire (mais pas les deux), on peut aussi omettre le point décimal ou le e suivi de l'exposant mais pas les deux.

Donner une expression rationnelle des constantes flottantes en C. Proposer un automate qui les reconnaît.

Exercice 1.5 Commentaires.

1. Donner un automate déterministe qui reconnaît le langage A^*abA^* .
2. En utilisant l'automate précédent, écrire des expressions rationnelles pour les langages formés des mots u tels que :
 - u ne contient pas le facteur ab ;
 - u admet ab pour suffixe, et c'est la première occurrence de ab dans u (autrement dit ab n'est facteur d'aucun préfixe strict de u).
3. Dédire des résultats précédents une expression rationnelle pour les commentaires en C (syntaxe `/*...*/`)

Exercice 1.6 URL.

Donner une expression rationnelle correspondant à un format simplifié d'URL. Une URL sera structurée de la façon suivante :

elle commence par le nom d'un protocole (`http`, `https`, `ftp`, `news`,...), l'adresse d'un serveur, un chemin absolu ou relatif, et des suffixes optionnels (requêtes précédées de '?', étiquettes précédées de '#').

Exemples :

`http://dept-info.labri.fr/~carrere/ERASMUS/erasmus.html#prog`

`http://www.commentcamarche.net/forum/index.php3?cat=1&page=2`

Exercice 1.7 dates.

1. En utilisant la syntaxe de `lex`, donnez une expression régulière pour les *noms* des nombres en Français (de 1 : "un" à 9999 : "neuf mille neuf cent quatre vingt dix neuf"). Les conventions utilisées sont les suivantes :
 - on utilise l'alphabet en minuscules
 - le caractère de séparation est toujours " " (espace),
 - "vingt", "cent" et "mille" sont invariables ("quatre vingt", "deux cent", "trois mille"),
 - on distingue les spécificités linguistique du Français, dont "treize", "vingt et un", etc.

2. En se basant sur l'expression précédente, donner une expression régulière pour les *noms de dates* : "dix janvier mille neuf cent quatre vingt douze", par exemple. Indications :
- le nom du jour dans le mois varie de "premier" à "vingt-neuf", "trente" ou "trente et un" selon le mois,
 - le caractère de séparation est toujours " " (espace),
 - on considérera que toutes les années peuvent être bissextiles...