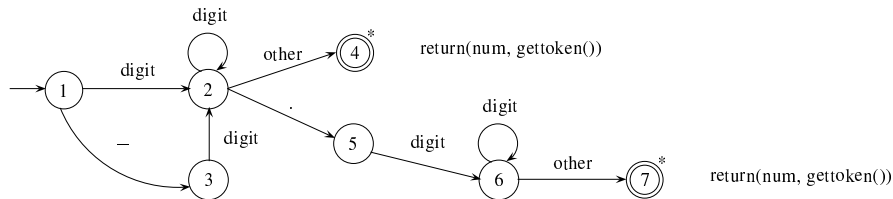


## Travail pratique #1 Analyses lexicale et syntaxique

### Questions

- (10 points.) Énumérez, en ordre alphabétique, toutes les chaînes faisant partie des ensembles suivants. **Note** : par définition, une chaîne ne peut apparaître plusieurs fois dans un ensemble ; soit elle y apparaît, soit elle n'y apparaît pas.
  - L'ensemble des préfixes de **carte**.
  - L'ensemble des suffixes propres de **joker**.
  - L'ensemble des sous-chaînes propres de **main**.
  - L'ensemble des sous-séquences de **aaaab**.
  - L'ensemble  $\{\mathbf{ab, bc, ca}\}^4 \cap \{\mathbf{a, b, c}\} \cdot \{\mathbf{ab, bc, ca}\}^3 \cdot \{\mathbf{a, b, c}\}$ .
- (10 points.) Les langages suivants sont décrits informellement. Pour chacun, donnez une expression régulière ou une définition régulière qui le génère. Donnez une explication dès que votre réponse est le moins complexe.
  - L'ensemble des chaînes sur l'alphabet  $\{\mathbf{a, b, c}\}$  dans lesquelles il apparaît au moins un **b** et au moins deux **c**. Par exemple, la chaîne **caacbaa** est dans le langage mais pas la chaîne **aaaccaca**.
  - L'ensemble des chaînes sur l'alphabet  $\{0, 1, \dots, 9\}$  dont le total des chiffres atteint au moins 5. Considérez que les naturels commencent par zéro. Par exemple, la chaîne **26** est dans le langage mais pas la chaîne **3010**.
  - L'ensemble des chaînes sur l'alphabet  $\{\mathbf{a, b, c, d}\}$  qui ont à la fois **abc** et **bcd** comme sous-chaînes. Par exemple, les chaînes **dbc dcabca** et **abcd** sont dans le langage mais pas la chaîne **ddabcaa**.
  - L'ensemble des chaînes sur l'alphabet  $\{\mathbf{a, b, c}\}$  qui ont à la fois **abc** comme sous-chaîne et **cab** comme sous-séquence. Par exemple, la chaîne **abcaabbcc** est dans le langage mais pas la chaîne **ccaabb**.
  - L'ensemble des chaînes sur l'alphabet  $\{\mathbf{a, b, c, d}\}$  qui n'ont pas la chaîne **acdc** comme sous-séquence. Par exemple, la chaîne **aabbccdd** est dans le langage mais pas la chaîne **dcbadcbadcba**.

3. (10 points.) Supposons qu'on s'intéresse aux chaînes sur l'alphabet  $\{a, b, c, d\}$ . Supposons que, pour une raison quelconque, on souhaite distinguer les chaînes qui contiennent au moins une lettre qui se répète en deux positions successives. Autrement dit, une chaîne avec répétition contient (au moins) l'une des sous-chaînes **aa**, **bb**, **cc** ou **dd**. Quand on identifie une chaîne qui contient une répétition quelque part, il faut produire le jeton **idRep**. Sinon, quand on identifie une chaîne non-vide sans répétition, il faut produire le jeton **id**. Donnez un ou des diagrammes de transition qui reconnaissent les motifs correspondant à ces jetons.
4. (10 points.) Dans les exercices reliés au chapitre 3 donnés sur le site web, on retrouve divers diagrammes de transition. Transformez le dernier diagramme de la dernière page de ce document en fragment de code. Par souci de simplicité, le diagramme est recopié ici. Pour produire le code, suivez la recette donnée vers la fin des notes de cours du chapitre 3. Vous devez produire du code pour les sept états du diagramme.



5. (10 points.) Soit  $L_5$  le langage tiré de l'alphabet  $\{a, b\}$  qui contient les chaînes qui ont exactement deux fois plus de 'a' que de 'b'. Je propose la grammaire hors-contexte  $G_5$  ci-bas en prétendant qu'elle génère  $L_5$ . Faites une démonstration rigoureuse (par exemple, par induction) de  $L_5 \subseteq L(G_5)$ ; l'autre inclusion est triviale.

$$\begin{array}{l}
 G_5 : \quad S \rightarrow \epsilon \\
 \quad \quad | \quad a S a S b \\
 \quad \quad | \quad a S b S a \\
 \quad \quad | \quad b S a S a \\
 \quad \quad | \quad S S
 \end{array}$$

**Indice :** Je vous suggère de considérer les bilans  $f(v) = |v|_a - 2|v|_b$  pour les préfixes  $v$  successifs d'un mot quelconque de  $L_5$ . La notation  $|v|_a$  désigne le nombre de 'a' dans  $v$ ;  $|v|_b$ , le nombre de 'b'. Par exemple, la chaîne **aaabab** est dans  $L_5$  et donne lieu à la suite de bilans 0, 1, 2, 3, 1, 2 et 0. Il faut noter que la suite de bilans doit commencer et finir par 0. Aussi, elle monte à coups de 1 et descend à coups de 2.

6. (10 points.) Considérons la grammaire du numéro précédent. Donnez la dérivation à gauche d'abord, la dérivation à droite d'abord et l'arbre de dérivation pour la chaîne **abbaaa**.

7. (10 points.) Considérons encore la même grammaire. Dites si la grammaire est ambiguë. Justifiez votre réponse à l'aide d'une démonstration ou d'un contre-exemple, selon que la grammaire est exempte d'ambiguïté ou pas.
8. (10 points.) Éliminez la récursion à gauche dans la grammaire suivante (inspirée de la syntaxe des expressions booléennes) en ordonnant les non-terminaux ainsi :  $E$ ,  $T$ ,  $F$ ,  $A$ . Décrivez en détail les étapes effectuées par l'algorithme (boucles sur  $i$  et  $j$ ).

$$\begin{aligned}
 E &\rightarrow E \text{ ou } T \mid T \\
 T &\rightarrow T \text{ et } F \mid F \\
 F &\rightarrow \text{non } F \mid A \\
 A &\rightarrow (E) \mid \text{faux} \mid \text{vrai} \mid \text{id}
 \end{aligned}$$

9. (5 points.) En imitant le pseudo-code présenté aux pages 41–46 des notes de cours du chapitre 4, donnez l'implantation d'une fonction ' $S()$ ' qui effectue l'analyse syntaxique descendante prédictive des chaînes générées par  $S$ . Supposez que les fonctions ' $peekToken$ ' et ' $readToken$ ' sont fournies.

$$S \rightarrow aSa \mid bSb \mid c$$

10. (5 points.) À la façon de la page 48 des notes de cours du chapitre 4, donnez la trace des opérations effectuées par l'analyseur prédictif sans récursion sur la chaîne d'entrée `bbbaabdccddd`.

NON- TERMINAL	SYMBOLE D'ENTRÉE				
	a	b	c	d	\$
$S$	$S \rightarrow aSc$	$S \rightarrow bSd$	$S \rightarrow \epsilon$	$S \rightarrow \epsilon$	$S \rightarrow \epsilon$

11. (10 points.) Considérons la grammaire de la question 5 à nouveau. Calculez les ensembles  $FIRST(S)$  et  $FOLLOW(S)$ .
12. (10 points.) Considérons la grammaire suivante. Les ensembles  $FIRST$  et  $FOLLOW$  sont précalculés. Construisez la table d'analyse pour cette grammaire.

$$\begin{aligned}
 A &\rightarrow B \mid dB \\
 B &\rightarrow cA \mid Cb \mid a \\
 C &\rightarrow aB \mid \epsilon
 \end{aligned}$$

Non-term.	FIRST	FOLLOW
$A$	{a, b, c, d}	{b, \$}
$B$	{a, b, c}	{b, \$}
$C$	{a, $\epsilon$ }	{b}

## Remise des travaux

Vous devez remettre le travail dans la boîte de dépôt de **monPortail** au plus tard le **25 février**. Veuillez ignorer la date de remise indiquée dans le plan de cours. Les autres modalités de remise demeurent telles que spécifiées dans le plan de cours.