A.1 Théorie des langages formels

1. Soit le mot $u = aabcab$. Comment faire la liste des sous-mots de u ?
2. Soit $X = \{a, b\}$. Calculer le produit $A.B$ pour les langages A et B suivants :
   $A = \{a, ab, bb\}$ $B = \{\varepsilon, b, aa\}$
   $A = \emptyset$ $B = \{a, ba, bb\}$
   $A = \{\varepsilon\}$ $B = \{b, aba\}$
   $A = \{aa, ab, ba\}$ $B = X^*$
3. Soit $v = abacbc$. Donner la liste des préfixes de v, la liste de ses suffixes, la liste de ses facteurs.
4. Donner l'algorithme qui, étant donné un mot (tableau de caractères), fournit (a) la liste de ses préfixes, et (b) la liste de ses facteurs.
5. Soit $x = abbcc$ un mot sur l'alphabet $X = \{a, b, c\}$.
   (a) Quelle est la valeur de $|x|$? et de $|x|_a$?
   (b) Donner un mot de X^3 qui ne soit pas un facteur de x.
   (c) Donner un sous-mot de x qui n'est pas un facteur de x.
   (d) Donner tous les facteurs de x qui appartiennent à X^3.
   (e) Donner les ensembles $Pre(x)$ et $Suf(x)$ des préfixes et suffixes de x.
6. (a) Soient t, u, v, w quatre mots de X^* tels que que $tu = vw$. Montrer qu'il existe un mot unique $z \in X^*$ tel que :
   – soit $u = zw$ et $v = tz$
   – soit $t = vz$ et $w = zu$
   (Lemme de Levi)
   (b) En utilisant ce lemme montrer que si u_1, u_2 et v sont trois mots de X^*, si $u_1 \in Pre(v)$ et si $u_2 \in Pre(v)$ alors soit $u_1 \in Pre(u_2)$ soit $u_2 \in Pre(u_1)$.
   (c) En utilisant ce théorème, et en appliquant un raisonnement par récurrence sur $|u|$, montrer que si $a \in X, b \in X, u \in X^*$, alors $ua = bu \Rightarrow a = b$ et $u \in \{a\}^*$.
7. Vérifier les propriétés suivantes du produit de langages :
   – Associativité
   – Distributivité par rapport à $\cup$
   – Non distributivité par rapport à $\cap$
   – Admet un élément neutre
   – Admet un élément absorbant
8. Combien y a-t-il de mots de 10 lettres sur l'alphabet $\{a, b, c, d\}$ comportant 3 occurrences de la lettre a et 4 occurrences de la lettre b ?
9. Admettons la définition suivante : les mots u et v de X^* sont dits **conjugués** si et seulement si $\exists u_1, u_2$ t.q. $u = u_1u_2$ et $v = u_2u_1$. Montrer :
   (a) que la "conjugaison" est une relation d'équivalence
   (b) que si u et v sont conjugués, alors $\exists w \in X^*, k, l \in \mathbb{N}$, t.q. $u = w^l$ et $v = w^k$
10. Soit l'alphabet $X = \{a, b\}$ et les langages $L_1 = \{a, ab, ba\}$ et $L_2 = \{\varepsilon, b, ab\}$.
   (a) Donner le résultat des opérations suivantes :
   $$L_1.L_2 \; L2.L_1 \; L_1.\{\varepsilon\} \; \emptyset.L_2 \; L_1^3$$
   (b) Si $L_3.L_4 = \{\varepsilon\}$, que peut-on dire des langages L_3 et L_4 ?
   (c) Si $L_3.L_4 = \emptyset$, que peut-on dire des langages L_3 et L_4 ?

A.2 Expressions rationnelles

Notation : e_i est une expression rationnelle, et L_i est le langage correspondant à e_i.

1. Soit $e_1 = a^*b^*$. Décrire L_1.
2. Soient $e_2 = a(ba)^*$ et $e'_2 = (ab)^*a$. A-t-on $L_2 \subset L'_2$? $L'_2 \subset L_2$?
3. Soient $e_2 = a(a|b)^*$ et $e'_2 = (a|b)^*a$. A-t-on $L_2 \subset L'_2$? $L'_2 \subset L_2$? Peut-on caractériser facilement la différence entre les deux langages ?
4. Soit $L_3 = \{ab, bba, acba, babb, ccacc\}$. Donner e_3.
5. Soit $L_4 = \{u \in X^*/|u|_a = 2k, k \in \mathbb{N}\}$. Donner e_4.
6. Soit $e_5 = a(b|c)^*(a^*|aa^*)ba(a^*|b^*|c^*)^*$. Peut-on simplifier e_5 ?
7. Soit $e_6 = (((a|\varepsilon)|b(ac)^*c)|(b^*c|\varepsilon))$. Trouver une expression e'_6 qui décrit le même langage et ne comprend pas le symbole ε.
8. Soit $e_7 = (a|b|c)(a|b|c)^*$. Y a-t-il une différence entre L_7 et $\{a, b, c\}^*$?
9. Soit L_8 = Langage des identificateurs du langage Pascal (ex. `f33`, `A_voir`, `ok`, `pMin`...). Donner e_8[1]
10. Soit $L_9 = \{ab, abc, abcd, bc, bcd\}$. Donner e_9.
11. L_{10} est le langage sur l'alphabet {a, b} de tous les mots qui comprennent la séquence 'abb'. Donner e_{10}.
12. Proposer une expression rationnelle pour le langage L_{11} de tous les mots de $\{a, b, c\}^*$ dont *cac* est un sous-mot[2].
13. Soit l'expression rationnelle $a^*b^*c^*zc^*b^*a^*$. Soit $L_\mathcal{E}$ le langage correspondant à cette expression. Lequel des langages $L_{\mathcal{G}_1}$ et $L_\mathcal{E}$ est-il inclus dans l'autre ?
14. Soient les expressions suivantes, compatibles avec le langage des expression régulières définies dans `emacs`. Pour chacune d'entre elle, donner une expression équivalente qui n'utilise que les 3 opérations union, produit et étoile.
    a^+b*, $[0-9]?$, $[\hat{}a-z]^*$

[1] On utilisera les abbréviations suivantes : $L = \{$A, B, ..., Z, a, ... z$\}$ et $C = \{$0, 1, ..., 9$\}$.

[2] Un *sous-mot* de u est une sous-suite de lettres — non nécessairement contiguë — de u. À distinguer d'un *facteur*. Exemple : *pis* est un sous-mot de *produits*.