0.0.1 Le Monoïde Libre

Définitions

Déf. 1 (Alphabet)

Un *alphabet* X est un ensemble fini de symboles (lettres). La *taille* de l'alphabet est le nombre de symboles.

Déf. 2 (Mot)

Un *mot* sur l'alphabet X est une suite finie de lettres de X.
Formellement, on définit $[p] = (1, 2, 3, 4, ..., p)$ (suite entière ordonnée).
Alors un mot est une fonction

$$u : [p] \longrightarrow X$$

p, la longueur du mot u, est notée $|u|$.

Example 0.0.1.1 Mots-lettres
u = toto

$$\begin{array}{ccc} 1 & \mapsto & \text{t} \\ 2 & \mapsto & \text{o} \\ 3 & \mapsto & \text{t} \\ 4 & \mapsto & \text{o} \end{array}$$

Déf. 3 (Sous-mot)

w est un *sous-mot* de u si w est une sous-suite de lettres de u.
NB : L'ordre est conservé.

Example 0.0.1.2
u =voiture, w =vtre.

Déf. 4 (Facteur)

Un *facteur* w de u est un sous-mot de u dont les lettres sont adjacentes dans u.

- w est un facteur de u $\Leftrightarrow$ $\exists u_1, u_2$ t.q. $u = u_1 w u_2$
- w est un facteur gauche (*préfixe*) de u $\Leftrightarrow$ $\exists u_2$ t.q. $u = w u_2$
- w est un facteur droit (*suffixe*) de u $\Leftrightarrow$ $\exists u_1$ t.q. $u = u_1 w$

Déf. 5 (Factorisation)

On appelle *factorisation* la décomposition d'un mot en facteurs.

On peut munir l'ensemble des mots sur un alphabet X (X*) d'une opération : la *concaténation*.

Déf. 6 (Concaténation)

Soient $[p] \xrightarrow{u} X$, $[q] \xrightarrow{w} X$. On définit la concaténation de u et w, notée uw (quelquefois $u.w$ ou $u\hat{}w$) :

$$uw : \quad [p+q] \longrightarrow X$$
$$uw_i = \begin{cases} u_i & \text{pour} \quad i \in [1,p] \\ w_{i-p} & \text{pour} \quad i \in [p+1, p+q] \end{cases}$$

On démontre facilement que l'opération de concaténation ainsi définie a les propriétés suivantes :

- La concaténation est non commutative (en général)
- La concaténation est associative : $u(vw) = (uv)w = uvw$
- La concaténation admet un élément neutre, le mot vide, noté $\emptyset$ ou ε ou 1_X ou $\mathbb{1}_X$.

Cette opération munit donc l'ensemble des mots sur un alphabet X (noté X*) d'une structure de *monoïde*.

0.1 Langages et Expressions rationnelles

0.1.1 Définition

Déf. 7 (Langage)

Un langage sur un alphabet X est un ensemble de mots.
Si on note X^* l'ensemble de tous les mots qu'on peut former sur l'alphabet X (c'est donc un langage), on peut définir un langage sur X comme un sous-ensemble de X^*.

0.1.2 Opérations sur les langages

Déf. 8 (Opérations sur les langages)

On peut définir deux opérations binaires et une opération unaire sur les langages :
- L'*union* des langages est définie comme d'habitude (union ensembliste)
- Le *produit* des langages est défini de la manière suivante (on suppose l'opération de concaténation définie) : $L_1.L_2 \quad = \quad \{uv \,/\, u \in L_1 \text{ et } v \in L_2\}$
- L'*étoile* (ou fermeture) d'un langage est définie de la manière suivante :
  En généralisant, on peut proposer la notation

$$\begin{array}{lcl} A^0 & = & \{\mathbb{1}_X\} \\ A^1 & = & A \\ A^{i+1} & = & A.A^i \end{array}$$

  d'où :
  Avec $A^n = \{a_1 ... a_n \,/\, a_i \in A\}$ on définit l'*étoile* de A : $A^* = \bigcup_{n \geq 0} A^n$

0.1.3 Expressions rationnelles

Déf. 9 (Expression rationnelle)

Soit X un alphabet. On définit les expressions rationnelles récursivement de la façon suivante :
- Pour tout $x \in X$, x est une expression rationnelle
- ε est une expression rationnelle
- Si φ et ψ sont des expressions rationnelles, alors
  - $(\varphi|\psi)$,
  - $(\varphi.\psi)$,
  - et φ^* sont des expressions rationnelles.

Remarque : la définition précédente décrit un **langage** sur l'alphabet formé des symbole de X plus $\{(,), |, ., *\}$. Plus précisément, on définit ainsi une **syntaxe**. Il faut donner une **sémantique** à ces expressions :
- Pour tout $x \in X$, l'e.r. x dénote le langage $\{x\}$,
- L'e.r. ε dénote le langage $\{\varepsilon\}$,
- Si φ et ψ sont des expressions rationnelles, alors
  - l'e.r. $(\varphi|\psi)$ dénote l'union des langages dénotés par φ et ψ ;
  - l'e.r. $(\varphi.\psi)$ dénote le produit des langages dénotés par φ et ψ ;
  - et l'e.r. φ^* dénote l'étoile du langage dénoté par φ.

.1 Bases Mathématiques

1. Soit l'ensemble de personnes $\mathcal{E} = \{Jean, Marie, Paul, Kim, Sandy, Bob\}$.
   – Quelle est la cardinalité de $\mathcal{E}$?
   – Donner le contenu de $\mathcal{E}^2$
   – On sait que Jean aime Marie, Sandy, Kim et Paul, que Marie aime Paul, Sandy, Kim et Jean, que Bob aime Sandy et Kim, que Sandy aime Kim, que Kim et Paul n'aiment personne à part eux-mêmes, et que tout le monde s'aime soi-même. Exprimer cela de manière formelle par une relation $A = x$ aime y.
   – Quelles sont les propriétés de la relation A ?
2. L'équation $y^2 = 1 - x^2$ décrit-elle une fonction ? Même question pour l'équation $y = \sqrt{1 - x^2}$.
3. Donnez les propriétés formelles de la multiplication dans l'ensemble $\mathbb{N}$.

.2 Théorie des Langages Formels

1. 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.
2. Soit $v = abacbc$. Donner la liste des préfixes de v, la liste de ses suffixes, la liste de ses facteurs.
3. Soit le mot $u = aabcab$. Comment faire la liste des sous-mots de u ?
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 $X = \{a, b\}$. Calculer le produit $A.B$ pour les langages A et B suivants :

$A = \{a, ab, bb\}$	$B = \{\varepsilon, b, aa\}$
$A = \emptyset$	$B = \{a, ba, bb\}$
$A = \{\varepsilon\}$	$B = \{b, aba\}$
$A = \{aa, ab, ba\}$	$B = X^*$

6. 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 :
   $$L_1.L_2 \quad L_2.L_1 \quad L_1.\{\varepsilon\} \quad \emptyset.L_2 \quad L_1^3$$
   (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 ?
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. (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 $u \in \{a\}^*$.
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 $v = w^k$

.3 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. (a) Soient $e_2 = a(ba)^*$ et $e_2' = (ab)^*a$. A-t-on $L_2 \subset L_2'$? $L_2' \subset L_2$?
   (b) 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 ?
3. Soit $L_3 = \{ab, bba, acba, babb, ccacc\}$. Donner e_3.
4. Soit $L_4 = \{u \in X^*/|u|_a = 2k, k \in \mathbb{N}\}$. Donner e_4.
5. Soit $e_5 = a(b|c)^*(a^*|aa^*)ba(a^*|b^*|c^*)^*$. Peut-on simplifier e_5 ?
6. 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.
7. Soit $e_7 = (a|b|c)(a|b|c)^*$. Y a-t-il une différence entre L_7 et $\{a, b, c\}^*$?
8. Soit L_8 = Langage des identificateurs du langage Pascal (ex. `f33`, `A_voir`, `ok`, `pMin`...). Donner e_8[1]
9. Soit $L_9 = \{ab, abc, abcd, bc, bcd\}$. Donner e_9.
10. L_{10} est le langage sur l'alphabet {a, b} de tous les mots qui comprennent la séquence 'abb'. Donner e_{10}.
11. Proposer une expression rationnelle pour le langage L_{11} de tous les mots de $\{a, b, c\}^*$ dont *cac* est un sous-mot[2].
12. 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 ?
13. 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*.