

MPI: Langages

Cours

- Définir un alphabet, un mot, le mot vide ε , un langage
- Longueur $|u|$ d'un mot, $|u|_a$ nombre d'occurrence d'une lettre
- Concaténation, union, intersection de langages, complémentaire, étoile de Kleene
- Préfixe, suffixe, facteur, sous-mot
- Langage clot par opération, classes de langages
- Mirroir, palindrome (HP)
- Le monoïde libre (Σ^*, \cdot) et morphismes de mots (HP)
- Mots infinis, concaténation de mots infinis, morphismes (HP)

Questions de Cours

- Donner les préfixes, suffixes, facteurs et sous-mots de $abbab$ et $abcba$.
- **Lemme de Levy:** Soient $x, y, x', y' \in \Sigma^*$ tel que $xy = x'y'$. Montrer qu'il existe $z \in \Sigma^*$ tel que soit $x = x'z$ et $y' = zy$, soit $x' = xz$ et $y = zy'$.
- Soient $u, v \in \Sigma^*$ tel que $uv = vu$. Montrer qu'il existe $w \in \Sigma^*$ et $p, q \in \mathbb{N}$ t.q. $u = w^p$ et $v = w^q$

Petites questions

1. Donner le plus petit langage clot par préfixe contenant tous les mots qui peuvent s'écrire de la forme $\{a^n b^n : n \in \mathbb{N}\}$
2. Soient $u, v \in \Sigma^*$ et $a, b \in \mathbb{N}$ tel que $u^a = v^b$. Montrer qu'il existe $w \in \Sigma^*$ et $p, q \in \mathbb{N}$ tel que $u = w^p$ et $v = w^q$
3. Soient $x, y, z \in \Sigma^*$ tel que $xy = yz$ avec $x \neq \varepsilon$. Montrer qu'il existe $k \in \mathbb{N}$ et $u, v \in \Sigma^*$ tels que $x = uv, z = vu$ et $y = (uv)^k u$.
4. Un mot est dit *primitif* s'il n'est puissance d'aucun autre mot que lui-même. Montrer que pour tout mot non vide, il existe un unique mot primitif dont il est une puissance.

Exercices à rajouter

- Equations de mots https://diplome.di.ens.fr/informatique-ens/annales/2022_InfoLCR-rapport.pdf

Opérations de langages

Pour les paires de $L_i, L_{i'}$ suivantes sur $\Sigma = \{a, b, c\}$, expliciter les valeurs de $L_i \cap L_{i'}$, $L_i \cup L_{i'}$, $L_i \cdot L_{i'}$ et L_i^* :

1. $L_1 = \{abb, ba, c\}$, $L'_1 = \{abb, cc, \varepsilon\}$
2. $L_2 = \{a^n : n \in \mathbb{N}^*\}$, $L'_2 = L_2$
3. $L_3 = \{a^n b^n : n \in \mathbb{N}^*\}$, $L'_3 = \{c^n : n \in \mathbb{N}\}$
4. $L_3 = \{a^n b^n : n \in \mathbb{N}^*\}$, $L'_3 = \{a^n : n \in \mathbb{N}\}$
5. $L_4 = \{a^n b^m : n < m\}$, $L'_4 = \{b^n : n \in \mathbb{N}\}$
6. $L_5 = \{a^n b^m : n < m\}$, $L'_5 = \{b^n c^m : n < m\}$

Language contenant ε

Soit L un langage, montrer que $\varepsilon \in L \Leftrightarrow L \subseteq L^2$

Cloture de langages

Pour chacun des langages suivant, indiquer le plus petit langage L clot par étoile de kleene, préfixe, suffixe, sous-mot

1. $L = \{a^n : n \in \mathbb{N}\}$

1. $L = \{a^n b^n : n \in \mathbb{N}\}$

1. $L = \{a^n b^n : n \in \mathbb{N}\}$

Suite de mots par récurrence

On définit sur l'alphabet $\Sigma = \{a, b\}$ la suite de mots $(u_n)_{n \geq 0}$ définie par récurrence par $u_0 = a$, $u_1 = b$ et $u_{n+2} = u_{n+1}u_n$

1. Montrer que si $n \geq 2$, alors u_n se termine par ba si n est pair et par ab sinon.
2. Proposer un algorithme qui pour un $n \in \mathbb{N}$ calcule $|u_n|$ la longueur de u_n
3. On définit v_n comme étant u_n où l'on a retiré les deux dernières lettres. Montrer que pour tout $n \in \mathbb{N}_{\geq 2}$, v_n est un palindrome.

Mot circulaire

Étant donnés deux mots $u, v \in \Sigma^n$ et $p < n$, proposer un algorithme qui trouve des indices i, j tels que $u[i..i + p - 1]$ et $v[j..j + p - 1]$ ont le plus grand nombre de caractères identiques.

Les indices sont à prendre au sens circulaire (i.e. $w[|w| + l] = w[l]$, pour tout $l < |w|$).

Hypercubes et mots

On définit l'hypercube de dimension n comme étant le graphe non orienté $G = (V, E)$ avec $V = \{0, 1\}^n$ et E l'ensemble des couples (u, v) de mots qui ne diffèrent que d'un bit.

1. Montrer que l'hypercube de dimension n possède un cycle hamiltonien.
2. En déduire qu'on peut énumérer les mots de $\{0, 1\}^n$ en ne modifiant à chaque fois qu'un seul mot (une telle énumération est un *code de Gray*).
3. Proposer un algorithme prenant un n et renvoyant un code de Gray des mots de longueur n .
4. Proposer un algorithme (de bonne complexité) qui à un mot associe son suivant dans cette liste.

Mots univers

On dit qu'un mot $w \in \Sigma^*$ est n -univers si tous les mots de Σ^n sont des facteurs de w . On s'intéresse à créer les plus petits mots n -univers.

1. Montrer qu'un mot n -univers sur un alphabet à k lettres à au moins une longueur de $k^n + n - 1$

Soit $G = (V, E)$ un graphe **orienté**, on définit $L(G)$ le *graphe ligne* de G par le graphe orienté (E, E') avec E' l'ensemble des arêtes de la forme $((x, y), (y, z))$ pour $(x, y), (y, z) \in E$.

2. Donner le graphe ligne du cycle à 4 éléments et d'un arbre binaire parfait de hauteur 2.

On construit alors la famille des graphes de Bruijn $(DB(n))_{n \in \mathbb{N}^*}$ par $DB(1) = (\{0, 1\}, \{0, 1\}^2)$ et $DB(n+1) = L(DB(n))$.

3. Construire $DB(2)$
4. Montrer que pour tout $n \in \mathbb{N}^*$, chaque sommet de $DB(n)$ à autant d'arêtes sortantes que entrantes. Combien de sommets et d'arêtes $DB(n)$ possède t'il ?
5. Montrer que pour tout graphe orienté fortement connexe tel que pour tout sommet le degré entrant est le même que le degré sortant, il existe un cycle eulérien (un cycle passant par toutes les arêtes du graphe).

En déduire que pour tout $n \in \mathbb{N}^*$, $DB(n)$ possède un cycle eulérien.

6. En voyant les sommets de $DB(n)$ comme des mots dans $\{0, 1\}^{n-1}$, et en étiquetant les arêtes par 0 ou 1, montrer qu'il existe un mot n -univers sur l'alphabet $\{0, 1\}$ de taille $2^n + n - 1$
7. Généraliser la question précédente pour des alphabets plus grand.

Ensemble inévitables¹

On fixe un alphabet fini $\Sigma = \{a, b\}$. Pour $w \in \Sigma^*$, on écrit $w = w_1 \dots w_n$ où $n := |w|$ est la longueur de w . On dit qu'un mot $w \in \Sigma^*$ évite un mot $s \in \Sigma^*$ si s n'apparaît pas comme facteur de w . On dit que w évite un ensemble de mots $S \subseteq \Sigma^*$ s'il évite chaque mot de S .

1. Donner un mot de longueur au moins 12 qui évite $S = \{aaaa, aaab, aba, baaa, bab, bbbb\}$.
2. Donner un algorithme qui étant donné un ensemble fini de mot S et un w , teste si w évite S .

On dit qu'un ensemble $S \subseteq \Sigma^*$ est inévitable s'il n'existe qu'un nombre fini de mots $w \in \Sigma^*$ qui évitent S . Sinon, on dit que S est évitable.

3. Est-ce que S est évitable ? Et l'ensemble $\{aaa, abb, baa, abab\}$?
4. Montrer que pour tout Σ un alphabet, et pour tout $k \in \mathbb{N}$, il existe un $n \in \mathbb{N}$ tel que pour tout mot $w \in \Sigma^*$ avec $|w| > n$, il existe $p < q$ tel que pour tout $l \in \llbracket 0; k \rrbracket$, on a $w_{p+l} = w_{q+l}$
5. Montrer que, pour tout ensemble inévitable $S \subseteq \Sigma^*$, il existe un sous-ensemble $S' \subseteq S$ fini tel que S' soit inévitable.

¹oral info Ulm ENS 2019

Mots sans cube²

On dit qu'un mot w est sans carré s'il n'a pas de facteur de la forme u^2 pour $u \neq \varepsilon$. On dit qu'il est sans cube s'il ne contient pas de u^3 en

1. Donner tous les mots sans carré sur $\Sigma = \{a, b\}$

On définit $\bar{h} : \Sigma \rightarrow \Sigma^*$ par $\bar{h}(a) = ab$ et $\bar{h}(b) = ba$ que l'on étend en $h : \Sigma^* \cup \Sigma^{\mathbb{N}} \rightarrow \Sigma^*$ par $h(w_1 \dots w_{|w|}) = \bar{h}(w_1) \dots \bar{h}(w_{|w|})$ et $h(w_1 \dots w_n \dots) = \bar{h}(w_1) \dots \bar{h}(w_n) \dots$

2. Montrer que $h^i(a)$ est un préfixe de $h^{i+1}(a)$.
3. Montrer que pour tout i , $h^i(a)$ est sans cube
4. En déduire l'existence d'un mot arbitrairement grand sans cube.

²Oral de l'ENS Info 2019 et sujet de CCINP d'Informatique 2026

Facteurs de points fixe de morphismes³

On dit que $h : \Sigma^* \rightarrow \Sigma^*$ est un *morphisme* si pour tout $u, v \in \Sigma^*$, $h(uv) = h(u)h(v)$. On dit qu'un morphisme h est une *substitution* si pour tout $\alpha \in \Sigma$, $|h(\alpha)| \geq 1$. Une telle substitution est dite α -*prolongeable* pour $\alpha \in \Sigma$ si α est un préfixe strict de $h(\alpha)$.

1. Soit h une substitution α -prolongeable. Montrer que pour tout $n \in \mathbb{N}$, $h^n(\alpha)$ est un préfixe strict de $h^{n+1}(\alpha)$
2. En déduire l'existence d'un unique mot infini u commençant par α dont tous les préfixes $h^n(\alpha)$ sont préfixes, et tel que $h(u) = u$.

Dans ce cas, on définit $\lim_{n \rightarrow \infty} h^n(\alpha)$ l'unique mot infini qui soit point fixe de h commençant par α . Une substitution (pas forcément α -prolongeable) est dite *primitive* s'il existe $k \in \mathbb{N}$ tel que $\forall \alpha, \beta \in \Sigma$, $|f^k(\alpha)|_\beta > 0$.

Soit f une substitution primitive sur un alphabet Σ à au moins 2 lettres.

3. Montrer qu'il existe $k \in \mathbb{N}$ et $\alpha \in \Sigma$ tel que f^k soit α -prolongeable.
3. On définit $F(w)$ l'ensemble des facteurs de w pour $w \in \Sigma^*$. Soient $L = \bigcup_{n \in \mathbb{N}} F(f^{nk}(\alpha))$ et $w \in L$, montrer qu'il existe $B \in \mathbb{N}^*$ tel que tout $u \in L$ de longueur $\geq B$ contient w en facteur.

³Tiré de l'Info D 2026 de décembre

Langage sans étoile

Soit Σ un alphabet, on définit par induction la classe des langages sans étoile \mathcal{M} par:

- Σ^* est sans étoile,
 - Pour tout $\alpha \in \Sigma$, $\{\alpha\}$ est sans étoile,
 - Pour tout L_1, L_2 deux langages sans étoile, $L_1 \cup L_2$, $L_1 \setminus L_2$ et $L_1 \cdot L_2$ sont sans-étoile.
1. Montrer $\{ab^n : n \in \mathbb{N}\}$ est un langage sans étoile sur $\Sigma = \{a, b\}$.
 2. Montrer que pour tout $u \in \Sigma^*$, le langage des mots qui n'ont pas u comme facteur est sans étoile.
 3. En déduire que $\{(ab)^n : n \in \mathbb{N}\}$ est un langage sans étoile.

(*) Égalité pour les résiduels

Soit Σ un alphabet. Donner L un langage sur Σ tel que $\forall t \in \Sigma^*, ut \in L \Leftrightarrow vt \in L$ ssi $u = v$

(*) Lemme d'Higman

Soit Σ un alphabet fini à n lettres. On note le fait que u soit un sous-mot de v par $v \succ u$.

Montrer que pour toute suite de mot infinie $(v_i)_{i \in \mathbb{N}}$, on a l'existence de $i < j$ tel que $v_i \succ v_j$.