Langages Formels
Série 5 - Lemme de l'étoile

20 Octobre 2025

Pensez à justifier vos réponses.

Lemme de l'étoile : Application

Le lemme de l'étoile (appelé aussi lemme de pompage) dit que si un langage L est régulier et infini, alors il doit respecter certaines conditions :

Lemme de l'étoile : Si L est un langage infini et régulier, alors
$\exists N \in \mathbb{N}$,
$\forall w \in L$ tel que $|w| \geq N$,
$\exists x, y, z$ tels que $w = xyz$, $|xy| \leq N$ et $|y| \geq 1$,
$\forall i \in \mathbb{N}$ on a $w' = xy^iz \in L$.

Mais pour montrer qu'un langage est régulier, non seulement c'est difficilement appliquable, mais en plus on a déjà plein de moyens plus efficaces, comme créer un automate fini (déterministe ou non déterministe) ou une expression régulière (ou, comme on le verra plus tard, une grammaire régulière).

En revanche, **le lemme de l'étoile va s'avérer très utile pour montrer qu'un langage n'est pas régulier**. On applique alors la contraposée :

Lemme de l'étoile (contraposée) : soit un langage L infini. Si :
$\forall N \in \mathbb{N}$,
$\exists w \in L$ tel que $|w| \geq N$,
$\forall x, y, z$ tels que $w = xyz$, $|xy| \leq N$ et $|y| \geq 1$,
$\exists i \in \mathbb{N}$ tel que $w' = xy^iz \notin L$,
alors L n'est pas régulier.

Exemple

Soit le langage $L_1 = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$ (i.e. w contient autant de a que de b). Montrons qu'il n'est pas régulier :

On applique donc le lemme (ou plus précisément sa contraposée). Il faut trouver pour tout N, un mot de longueur supérieure ou égale à N, pour lequel pour tout découpage respectant les conditions spécifiées, il existe un i tel que le mot où on pompe i fois la partie "y" n'est plus dans le langage.

- $\forall N$,
- $\exists w = a^N b^N \in L$, $|w| = 2N \geq N$ (on exprime le mot en fonction de N, ainsi on a bien un mot qui existe pour tout N, et qui appartient bien au langage L),
- $\forall x, y, z$ tels que $w = xyz$, $|xy| \leq N$ et $|y| \geq 1$, on peut exprimer tous ces découpages ainsi :
  $x = a^j$ $\quad j \geq 0, k \geq 1, j + k \leq N$
  $y = a^k$
  $z = a^{N-j-k} b^N$

  Cette forme résume bien tous les découpages possibles : par les conditions du lemme, les deux premiers facteurs x et y doivent être de longueur totale inférieure ou égale à N. Ils ne peuvent donc pas contenir plus de N caractères à eux deux, et donc ne ils ne peuvent contenir que des "a" dans ce mot (puisque dans ce mot les N premiers caractères sont tous des "a").
- $\exists i = 2$ tel que $w' = xy^iz = xyyz = a^j a^k a^k a^{N-j-k} b^N = a^{N+k} b^N \notin L$: pour i=2, quelque soit le découpage, on obtient un mot de la forme $a^{N+k} b^N$ (avec $k \geq 1$), qui n'appartient donc pas à L.

On a donc bien montré que $\forall N, \exists w \in L_1$ (avec $|w| \geq N$), $\forall x, y, z$ (tels que $w = xyz$, $|xy| \leq N$ et $|y| \geq 1$), $\exists i$ tel que $w' = xy^iz \notin L_1$.
Donc ce langage L_1 n'est pas régulier.

Exercices

1. Quand on montre qu'un langage n'est pas régulier avec le lemme, le choix du mot est important :
   - Montrez que le langage $L_2 = \{a^i b^j c^k \text{avec i=j ou i=k } , i, j, k \geq 0\}$ n'est pas régulier, en utilisant $\forall N$ le mot $a^N c^N$.
   - Et si à la place de choisir le mot $w = a^N c^N$, on avait choisi le mot $w_2 = a^{\lceil \frac{N}{2} \rceil} c^{\lceil \frac{N}{2} \rceil}$ (où $\lceil x \rceil$ est la partie entière supérieure de x) ? Faites de nouveau la preuve que L_2 n'est pas régulier, mais en utilisant le mot w_2 cette fois-ci. Est-il aussi simple d'exprimer tous les découpages respectant les conditions du lemme ?

2. Montrez que les langages suivants ne sont pas réguliers, sur l'alphabet $\Sigma = \{a, b\}$:
   - $L_3 = \{w \mid w = a^i b^j,\ i < j\}$
   - $L_4 = \{w \mid w = a^i b^j,\ i > j\}$
   - $L_5 = \{w \mid w \in \Sigma^*,\ |w|_a \neq |w|_b\}$
   - $L_6 = \{w \mid w \text{ est un palindrome}\}$ (rappel : un palindrome est un mot qui se lit à l'identique de gauche a droite et de droite à gauche, comme par exemple "kayak")
   - $L_7 = \{w \mid w \text{ n'est pas un palindrome}\}$
   - $L_8 = \{a^i \mid i \text{ est un nombre premier}\}$