Codage de l'information

Rakotoarimalala Tsinjo Tony

Cours 5

IT University - 2023

Codes à longueur fixe

On dit d'un code qu'il est de *longueur fixe* lorsque tous les mots qu'il contient ont la même longueur. Autrement dit $C \subset \mathcal{A}^*$ est un code de longueur fixe s'il existe un entier non nul n tel que

$$C \subset \mathcal{A}^n$$

Codes à longueur fixe

On dit d'un code qu'il est de *longueur fixe* lorsque tous les mots qu'il contient ont la même longueur. Autrement dit $C \subset \mathcal{A}^*$ est un code de longueur fixe s'il existe un entier non nul n tel que

$$C \subset \mathcal{A}^n$$

Exemple

- Les 128 mots (octets) utilisé pour le codage ASCII forment un code de longueur fixe.
- Le code associé à l'un quelconque des codages **ISO-8859** est un code de longueur fixe.

Codes à virgule

Les codes à virgule sont des codes pour lesquels un symbole (ou un groupe de symboles) de $\mathcal{A}$ est réservé pour marquer le début ou la fin des mots du code.
Ce symbole particulier est appelé ***une virgule***. Il permet de factoriser un message codé avec ce code

Exemple

- *Le Morse est un codage à virgule*

Codes préfixes

Un langage préfixe est un langage dans lequel aucun mot n'est le préfixe d'un autre mot du langage. Dit autrement, $L \subset \mathcal{A}^*$ est préfixe si pour tous mots $u, v \in L$, on a

$$u \notin Pref(v) \text{ et } v \notin Pref(u).$$

CODES PRÉFIXES

Un langage préfixe est un langage dans lequel aucun mot n'est le préfixe d'un autre mot du langage. Dit autrement, $L \subset \mathcal{A}^*$ est préfixe si pour tous mots $u, v \in L$, on a

$$u \notin Pref(v) \text{ et } v \notin Pref(u).$$

Exemple

- *Le langage* $L = \{00, 010, 10, 110, 111\}$ *est préfixe*
- *Les langages de longueur fixe sont des langages préfixes.*

Codes préfixes

Un langage préfixe est un langage dans lequel aucun mot n'est le préfixe d'un autre mot du langage. Dit autrement, $L \subset \mathcal{A}^*$ est préfixe si pour tous mots $u, v \in L$, on a

$$u \notin Pref(v) \text{ et } v \notin Pref(u).$$

Exemple

- *Le langage* $L = \{00, 010, 10, 110, 111\}$ *est préfixe*
- *Les langages de longueur fixe sont des langages préfixes.*

Théorème

Tout langage préfixe autre que $\{\epsilon\}$ *est un code.*

Démos: À faire comme exercice

Résiduel d'un langage

Le résiduel à gauche d'un langage $L \subset \mathcal{A}^*$ par un mot $u \in \mathcal{A}^*$ est le langage que l'on obtient en ne gardant des mots de L que ceux admettant u comme préfixe, et en supprimant ce préfixe de ces mots. En notant $u^{-1}.L$ ce résiduel, on a

$$u^{-1}.L = \{v \in \mathcal{A}^* | u.v \in L\}$$

(on peut définir le résiduel à droite de manière équivalente mais on n'utilisera pas cette notion dans le cours).

Résiduel d'un langage

Le résiduel à gauche d'un langage $L \subset \mathcal{A}^*$ par un mot $u \in \mathcal{A}^*$ est le langage que l'on obtient en ne gardant des mots de L que ceux admettant u comme préfixe, et en supprimant ce préfixe de ces mots. En notant $u^{-1}.L$ ce résiduel, on a

$$u^{-1}.L = \{v \in \mathcal{A}^* | u.v \in L\}$$

(on peut définir le résiduel à droite de manière équivalente mais on n'utilisera pas cette notion dans le cours).

Exemple

Avec le langage $L = \{00, 01, 110, 001\}$

- *si* $u_1 = 10$ *alors* $u_1^{-1}.L = \emptyset$
- *si* $u_2 = 1$ *alors* $u_2^{-1}.L = \{10\}$
- *si* $u_3 = 00$ *alors* $u_3^{-1}.L = \{\epsilon, 1\}$

Quotient d'un langage

Le QUOTIENT À GAUCHE D'UN LANGAGE $L \subset \mathcal{A}^*$ par un langage $M \subset \mathcal{A}^*$ est le langage obtenu en ne gardant que les mots de L ayant un préfixe dans M et en supprimant ce préfixe des mots gardés. En notant $M^{-1}.L$ ce quotient, on a

$$M^{-1}.L = \{v \in \mathcal{A}^* | \exists u \in M \text{ tel que } u.v \in L\}$$

Autrement dit

$$M^{-1}.L = \bigcup_{u \in M} u^{-1}.L$$

Quotient d'un langage

Le QUOTIENT À GAUCHE D'UN LANGAGE $L \subset \mathcal{A}^*$ par un langage $M \subset \mathcal{A}^*$ est le langage obtenu en ne gardant que les mots de L ayant un préfixe dans M et en supprimant ce préfixe des mots gardés. En notant $M^{-1}.L$ ce quotient, on a

$$M^{-1}.L = \{v \in \mathcal{A}^* | \exists u \in M \text{ tel que } u.v \in L\}$$

Autrement dit

$$M^{-1}.L = \bigcup_{u \in M} u^{-1}.L$$

Exemple

En reprenant l'exemple ci-dessus, et en posant $M = \{u_1, u_2, u_3\}$ alors

$$M^{-1}.L = \{\epsilon, 1, 10\}$$

Algorithme de Sardinas et Patterson

Soit $L \subset \mathcal{A}^*$ un langage. On définit à partir de L une suite (L_n) de langage en posant

$$L_0 = L \tag{1}$$

$$L_1 = L^{-1}.L - \{\epsilon\} \tag{2}$$

$$L_{n+1} = L^{-1}.L_n \cup L_n^{-1}.L \text{ pour tout } n \geqslant 1 \tag{3}$$

Algorithme de Sardinas et Patterson

Soit $L \subset \mathcal{A}^*$ un langage. On définit à partir de L une suite (L_n) de langage en posant

$$L_0 = L \tag{1}$$

$$L_1 = L^{-1}.L - \{\epsilon\} \tag{2}$$

$$L_{n+1} = L^{-1}.L_n \cup L_n^{-1}.L \text{ pour tout } n \geqslant 1 \tag{3}$$

Théorème

Un langage $L \subset \mathcal{A}^*$ est un code si et seulement si le mot vide n'appartient à aucun des langages L_n de la suite définie ci-dessus.

Demos: à admettre

Applications

1. Lorsque L est un langage de longueur fixe (non nulle), alors $L_1 = L^{-1}.L - \{\epsilon\} = \emptyset$, et par conséquent, pour tout entier $n \geqslant 2$, $L_n = \emptyset$. Le mot vide n'est dans aucun langage L_n , et L est bien évidemment un code.

Applications

1. Lorsque L est un langage de longueur fixe (non nulle), alors $L_1 = L^{-1}.L - \{\epsilon\} = \emptyset$, et par conséquent, pour tout entier $n \geqslant 2$, $L_n = \emptyset$. Le mot vide n'est dans aucun langage L_n , et L est bien évidemment un code.
2. Lorsque L est un langage préfixe (autre que le langage réduit au seul mot vide), $L_1 = \emptyset$ et il en est de même pour tous les langages L_n avec $n \geqslant 2$. Et tout langage préfixe est un code.

Pour le langage $L = \{1, 00, 01, 10\}$,

Pour le langage $L = \{1, 00, 01, 10\}$,
on a

$$L_1 = \{0\}$$
$$L_2 = \{0, 1\}$$
$$L_3 = \{0, 1, \epsilon\}$$

Puisque $\epsilon \in L_3$ donc L n'est pas un code.

En effet on a bien une double factorisation du mot 1001 avec des mots de L :

$$1001 = \underbrace{10} \ . \ \underbrace{01} = \underbrace{1} \ . \ \underbrace{00} \ . \ \underbrace{1}$$

Enfin pour le langage $L = \{000, 010, 011, 01001\}$, Enfin pour le langage $L = \{000, 010, 011, 01001\}$, on a

$$
\begin{aligned}
L_1 &= \{01\} \\
L_2 &= \{0, 1, 001\} \\
L_3 &= \{00, 10, 11, 1001\} \\
L_4 &= \{0\} \\
L_5 &= \{00, 10, 11, 1001\}
\end{aligned}
$$

Comme $L_3 = L_5$, on en déduit que tout indice pair de la suite L_n vaut L_4 et les impairs vaut L_3. Donc ϵ n'apparaît dans aucun de tous les L_i. Donc L est un code.

Écrire un programme en python (ou n'importe quel langage) qui décide si un langage donné en argument est un code ou pas.