ISEN CIR ³

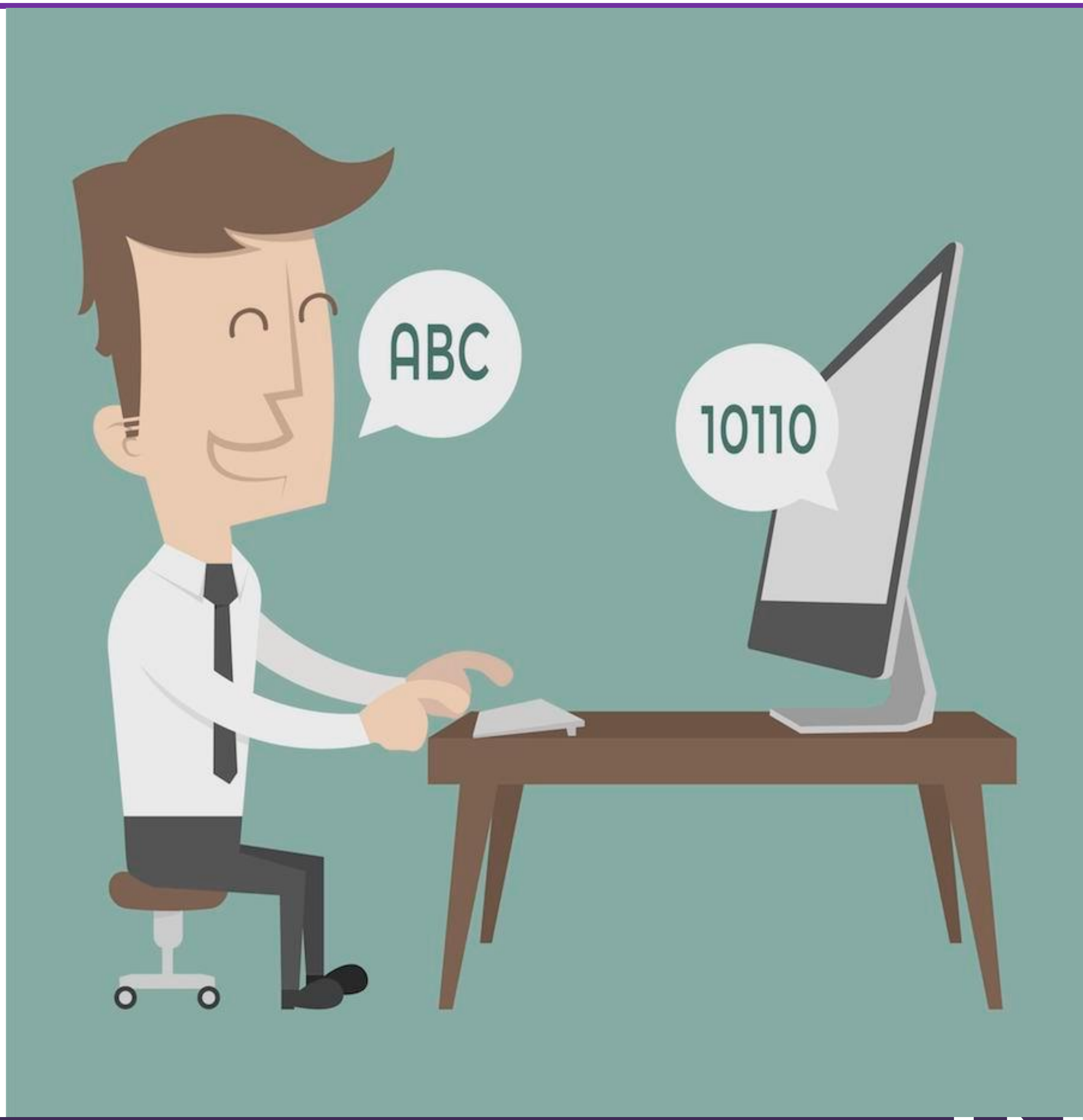
Théorie des langages

2. Alphabets, mots et langages

2021-2022

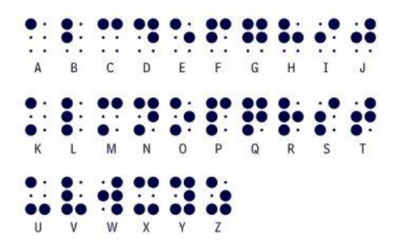
Plan

- Alphabets
- Mots
- Ensembles de mots
- Langages
- Opérations sur les mots et langages



- Définition :

Un **alphabet** est un ensemble fini de symboles



- Exemples :

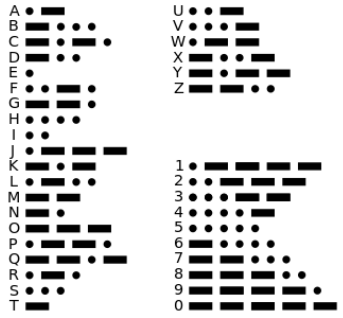
A1 = {0, 1}

A2 = {A, C, G, T}

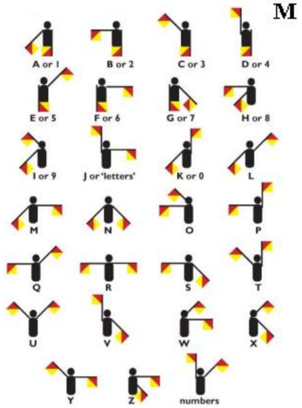
A3 = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F}

A4 = {if, then, else, X, Y, Z}

A5 = {😎, 😷, 🙄, …}



A	Alpha	N	November
B	Bravo	O	Oscar
C	Charlie	P	Papa
D	Delta	Q	Quebec
E	Echo	R	Romeo
F	Foxtrot	S	Sierra
G	Golf	T	Tango
H	Hotel	U	Uniform
I	India	V	Victor
J	Juliett	W	Whiskey
K	Kilo	X	X-ray
L	Lima	Y	Yankee
M	Mike	Z	Zulu



- Définition :

  Un **mot** est une suite finie et ordonnée d'éléments de l'alphabet.
  Elle peut être éventuellement vide.

  Le **mot vide** est noté ε (Epsilon)

- Exemples :
  - 0, 1, 11, 001, 101010, … sont des mots de l'alphabet {0, 1}
  - ACCGTTATGCCATTT est un mot de l'alphabet {A,C,G,T}
  - Bravo Oscar November Juliet Oscar Uniform Romeo est un mot en langage radio

- La **longueur** d’un mot **m** est le nombre de symboles constituant ce mot.
  Elle est notée **|m|**.
  La longueur du mot vide ε est 0.
  Le nombre d’occurrence d’un symbole **x** dans le mot **m** est noté $|m|_x$
- Exemples:

$\lvert\varepsilon\rvert = 0$	La longueur du mot vide est nulle
|0| = 1	Un caractère de l’alphabet
|11| = 2	Un mot de deux caractères
|ACCGTTC| = 7	Un mot de 7 caractères

| Bravo Oscar November Juliet Oscar Uniform Romeo| = 7

- L'ensemble de toutes les suites (les mots) de l'alphabet **X** est notée **X***.
- Aussi appelée « fermeture transitive »

- Pour un entier **n**, l'ensemble de tous les mots de **X*** de taille **n**, est noté X^n.

- Exemple : Soit l'alphabet B = {1, 0}

$\mathbf{B^0} = \{\varepsilon\}$

$\mathbf{B^1} = \{1, 0\}$

$\mathbf{B^2} = \{00, 11, 01, 10\}$

$\mathbf{B^3} = \{000, 001, 010, 011, 100, 101, 110, 111\}$

Tous les nombres binaires quelque soit leur longueur :

$\mathbf{B^*} = \{\varepsilon, 0, 1, 00, 11, 01, 10, 000, 001, 010, 011, 100, 101, 110, 111, 0000, 0001, ...\}$

Notations

Soit **X** un alphabet:

- **X*** est l'ensemble des mots
- **m** est un mot composé des symbole des l'alphabet **X**
- $m \in X^*$
- $|m|$ est la longueur de **m**
- **x** est une lettre de l'alphabet **X**
- $|m|_x$ est le nombre d'occurrences de **x** dans **m**

- Un **langage** est un ensemble de mots
  (C'est un sous-ensemble quelconque de **X***).
- Un langage **L** est un élément d'une partition de **X***.
- Il peut être **fini** ou **infini**.
- Un langage étant un ensemble, toutes les opérations ensemblistes s'appliquent aux langages:
  Intersection, union, complément, différence, …
- L'ensemble $\varnothing$ est le plus petit langage construit avec les éléments de **X**, on l'appelle le langage vide. **X*** est le plus grand.
- On dit que langage **M** est plus grand que le langage **L** ssi **L** est inclus dans **M**.

- Deux mots sont égaux s'ils ont des lettres identiques à des positions identiques.
- Ils ont ainsi la même longueur.

$x = x_1 \, x_2 \, x_3 \, \ldots \, x_m$

$y = y_1 \, y_2 \, y_3 \, \ldots \, y_n$

$x = y$ si et seulement si $m = n$ et $\forall \, i < n : x_i = y_i$

$x = x_1\, x_2\, x_3 \ldots x_m$

$y = y_1\, y_2\, y_3 \ldots y_n$

- La concaténation de deux mot **x** et **y**, notée **x.y** ou **xy** est le mot **z** tel que:

$$z = z_1\, z_2\, z_3 \ldots z_k$$

$$k = n + m$$

$$z_i = x_i \text{ si } i <= m \qquad \text{et} \qquad z_i = y_{i-m} \text{ si } i > m$$

- Exemple : la concaténation des mot **bon** et **jour** est le mot **bonjour**.
- Exercice : Quelles sont les propriétés de la concaténation?

- **a** et **b** sont des mots de **X***

- Préfixes :
  a est préfixe de **b** si et seulement si : ∃ **c** tel que **b = a.c**

  Exemple : « bon » est préfixe de « **bon**jour »

- Suffixes :
  a est suffixe de **b** si et seulement si : ∃ **c** tel que **b = c.a**

  Exemple : « jour » est suffixe de « bon**jour** »

Quelques propriétés:

- Le mot vide **ε** est un préfixe pour tous les mots.
- Le mot vide **ε** est un suffixe pour tous les mots.
- Aussi, chaque mot est préfixe de lui-même.
- Et chaque mots est suffixe de lui-même.
- Lorsque le préfixe (ou suffixe) d’un mot **w** est différent de **w**, on l’appelle « préfixe (ou suffixe) propre ».

- La puissance d'un mot **x**, notée $\mathbf{x^n}$, est définie par récurrence :

$$x^0 = \varepsilon$$

$$x^n = x \,.\, x^{n-1}$$

- Exemple : soit le mot $x = 011$

$$x^0 = \varepsilon$$

$$X^1 = 011$$

$$X^4 = 011011011011$$

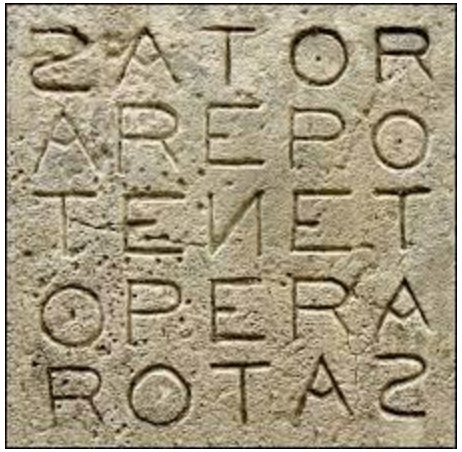
- Le mot miroir d'un mot $\mathbf{x = x_1\, x_2\, x_3 \ldots x_{n-1}\, x_n}$

  est défini par $\mathbf{x^m = x_n\, x_{n-1} \ldots x_3\, x_2\, x_1}$

- Théorème : pour deux mots **u** et **v**

  $\mathbf{(uv)^m = v^m\, u^m}$

Un palindrome est un mot égal à son mot miroir


SATOR
AREPO
TENET
OPERA
ROTAS


SATOR
AREPO
TENET
OPERA
ROTAS


SATOR
AREPO
TENET
OPERA
ROTAS


ROTAS
OPERA
TENET
AREPO
SATOR


ROTAS
OPERA
TENET
AREPO

- Egalité : **$a = b$**
- Puissance : **a^n** (n est un entier positif)
- Concaténation : **ab**
- Préfixes : a est préfixe de b si il existe c tel que **$b = ac$**
- Suffixes : a est suffixe de b si il existe c tel que **$b = ca$**
- Miroir d'un mot : **a^m** ou **a^-**

- Egalité: **A = B**
- Produit: **A.B** ou **AB** (ensemble des concaténations des mots de A et B)
- Fermeture de Kleene (étoile): **A^***
- Fermeture Positive: **A^+**
- Langage miroir: **A^m**

QU35710N5?