Représentation de chaînes de caractères sous forme de nombres

Christophe Dronne

25 Octobre, 2023

Préliminaires

On définit b le nombre d'éléments de la base $\mathbb{B}$, définie avec l'ensemble $\{"a", "b", ..., "z", "A", "B", ..., "Z"\}$ pour une quantité quelconque de caractères possibles (ici l'alphabet en minuscule et majuscule pour simplifier).

Ainsi, pour donner un exemple plus concret, le nombre $(2)_{10}$ pourrait être représenté sous forme $\mathbb{B}$ tel: $(2)_b$ ou, d'une meilleure façon, $"b"_{\mathbb{B}}$. Pour rendre le tout plus compréhensible, on adoptera cette dernière notation.

Conversement, la chaîne $"bonjour"_{\mathbb{B}}$ pourrait être représentée sous forme décimale, tel: $(50713653423)_{10}$

Le but du tout est de trouver les avantages/désavantages d'une représentation sous forme de nombre de chaînes de caractères en informatique. On prendra particulièrement soin de toujours indiquer la complexité des opérations, à usage exclusif du langage de programmation Python (3.11).

Indice d'un nombre

On essaie de simuler la fonctionalité d'indices présents dans les chaînes de caractères en Python.

$\forall b \in \mathbb{N}^+ - \{0, 1\}$ représentant le nombre d'éléments dans $\mathbb{B}$
$\forall x \in \mathbb{N}^*$ représenté sous forme $\mathbb{B}$
$\forall n \in \mathbb{N}$, tel que $n <$ la quantité de chiffres dans $x_{\mathbb{B}}$

On note $x_{\mathbb{B}}[n]$ le chiffre dans x sous forme $\mathbb{B}$ d'indice n partant de la droite.

Ainsi, pour $x = "bonjour"_{\mathbb{B}}$, on a: $x[0] = "r"_{\mathbb{B}}$ (0 étant la place des unités) et $x[2] = "o"_{\mathbb{B}}$.

De plus, on note $E(x)$ la fonction de la partie entière, tel que $E(3,14) = 3$.

On pose:

$$x_{\mathbb{B}}[n] = E\left(\frac{x - E\left(\frac{x}{b^{n+1}}\right) \times b^{n+1}}{b^n}\right) \tag{1}$$

Prenons $x = 4321$ et $b = 10$ (donc le système décimal).
On essaie d'extraire tous les chiffres dans $(x)_b$.

$$x[3] = 4 \Leftrightarrow x[3] = E\left(\frac{4321}{1000}\right) \Leftrightarrow x[3] = E\left(\frac{4321}{10^3}\right) \tag{2}$$

On remarque que $b = 10$ et $n = 3$ pour $x[3]$, donc $10^3 = b^n$. On note:

$$x[3] = E\left(\frac{4321}{b^n}\right) = E\left(\frac{x - 0}{b^n}\right) \tag{3}$$

On continue pour n = 2:

$$x[2] = 3 \tag{4}$$

Or, $43 - 40 = 3$, soit $\frac{4321-4000}{100} = 3,21$, donc:

$$x[2] = E\left(\frac{4321 - 4000}{100}\right) = 3 \tag{5}$$

À nouveau, on constate qu'on a: $n = 2$
$b = 10$
$100 = 10^2 = b^n$
et

$$4000 = E\left(\frac{4321}{1000}\right) \times 1000 = E\left(\frac{x}{b^{n+1}}\right) \times b^{n+1} \tag{6}$$

Finalement, on peut noter $x[2]$, tel:

$$x[2] = E\left(\frac{x - E\left(\frac{x}{b^{n+1}}\right) \times b^{n+1}}{b^n}\right) \tag{7}$$

On retrouve la formule posée, et on remarque que pour $n = 3$ (premier cas, voir haut de la page), on a:

$$E\left(\frac{x}{b^{n+1}}\right) \times b^{n+1} = E\left(\frac{4321}{10000}\right) \times 10000 = 0 \tag{8}$$

Donc la formule peut aussi être utilisée pour $x[3]$.

Indice d'un nombre en Python

Afin de uniformiser la notation d'indice de chaîne de caractères avec celle de l'informatique, on pose:
$\forall x \in \mathbb{N}^*$ en base $\mathbb{B}$
$i \in \mathbb{N}$, l'indice d'un chiffre dans $x_{\mathbb{B}}$ partant de la gauche (et non la droite comme pour n)
$l \in \mathbb{N}^*$ la quantité de chiffres dans $x_{\mathbb{B}}$.

Ainsi, on a:
$i = l - n$; ou, pour adopter une notation de Python: $i = \text{len}(x) - n$