

Représentation de chaînes de caractères sous forme de nombres

Christophe Dronne

25 Octobre, 2023

Préliminaires

On définit une base quelconque, notée b , utilisant comme symboles les lettres de notre alphabet en plus d'autres symboles utilisées comme caractères typographiques. Cette base servira à représenter un nombre sous forme de texte, de la même manière que le nombre $(13303790)_{10}$ peut être représenté sous le nombre hexadécimal $(caffee)_{16}$ qui pourrait être interprété comme le mot "caffée".

Une implémentation possible et simple de cette base pourrait utiliser les lettres "a", "b", "c" ... "z" pour représenter les nombres 0, 1, 2, ... 25, avec $b = 26$. Cette base va être utilisée pour illustrer des exemples, mais de manière générale tout formule et propriété va être applicable à une toute base supérieure ou égale à 2.

Le but de cet exercice est de trouver une implémentation optimale des fonctions communes utilisées pour les chaînes de caractères en informatique avec ces nombres en base b .

Lors de cet exercice, on va souvent faire utilisation de la fonction de la partie entière, notée $E(x)$, et de la fonction de la partie fractionnaire, notée $frac(x)$, tel:

$\forall x \in \mathbb{R}$, on a:

$$E(x) \in \mathbb{N} \tag{1}$$

$$frac(x) = |x| - E(|x|) \tag{2}$$

Or, on prendra seulement en compte des valeurs positives, donc:

$$frac(x) = x - E(x) \tag{3}$$

On a les propriétés suivantes:

$$0 \leq \text{frac}(x) < 1$$

$$\Rightarrow E(\text{frac}(x)) = 0$$

$$0 \leq x < 1$$

$$\Rightarrow \text{frac}(x) = x$$

$$\Rightarrow E(x) = E(\text{frac}(x)) = 0$$

$$\text{frac}(\text{frac}(x)) = \text{frac}(x) - E(\text{frac}(x))$$

$$\Rightarrow \text{frac}(\text{frac}(x)) = \text{frac}(x) - 0$$

$$\Rightarrow \text{frac}(\text{frac}(x)) = \text{frac}(x)$$

$\forall (x_1; x_2) \in \mathbb{R}^2$, on a:

$$E(x_1 + x_2) = E(x_1) + E(x_2)$$

si, et seulement si:

$$0 \leq \text{frac}(x_1) + \text{frac}(x_2) < 1$$

sinon:

$$E(x_1 + x_2) = E(x_1) + E(x_2) + 1$$

Donc, $\forall x \in \mathbb{R}, \forall n \in \mathbb{N}$, on a:

$$E(x + n) = E(x) + n$$

car:

$$\text{frac}(x) + \text{frac}(n) = \text{frac}(x) + 0 = \text{frac}(x)$$

et

$$0 \leq \text{frac}(x) < 1$$

$$\text{Donc } E(x + n) = E(x) + E(n)$$

De plus, on a: $E(n) = n$, donc

$$E(x + n) \text{ est bien égal à } E(x) + n$$

Longueur d'un nombre

Les chaînes de caractères ont toujours une longueur associée. Pour donner un exemple, la chaîne "bonjour" a une longueur de 7, car elle est constituée de 7 caractères. Cependant, le nombre $(\text{bonjour})_b$ n'a pas de 'longueur', donc il va falloir trouver une définition mathématique de la longueur d'un nombre.

On pose:

$\forall b \in \mathbb{N} - \{0; 1\}$, une base quelconque

$\forall n \in \mathbb{N}$, la quantité de chiffres dans k

$\forall k \in \mathbb{R}$, tel que $1 \leq k < b$

Tout entier naturel l (exclu de 0) de base b peut s'écrire tel:

$$l = k \times b^{n-1}$$

Exemple:

$534 = 5.34 \times 10^{3-1}$, avec $b = 10$, $n = 3$, $l = 534$, $k = 5.34$,

suivant les conditions posées.

On pose la formule de la 'longueur' (c-à-d quantité de chiffres) d'un nombre:

$$n = E(\log_b(l)) + 1 \quad (4)$$

Par convenance, on va noter la fonction qui attribue à un entier naturel non nul sa quantité de chiffres $len(l)$, tel que:

$$len(l) = E(\log_b(l)) + 1 \quad (5)$$

Démonstration:

On pose:

$\forall b \in \mathbb{N} - \{0; 1\}$, une base quelconque

$\forall n \in \mathbb{N}$, la quantité de chiffres dans k

$\forall k \in \mathbb{R}$, tel que $1 \leq k < b$

On a:

$$1 \leq k < b$$

k et b sont positifs, donc:

$$\Rightarrow \log_b(1) \leq \log_b(k) < \log_b(b)$$

$$\Rightarrow 0 \leq \log_b(k) < 1$$

$$\Rightarrow E(\log_b(k)) = 0$$

On reprend la formule de l :

$$\begin{aligned}
 l &= k \times b^{n-1} \\
 \Rightarrow \log_b(l) &= \log_b(k \times b^{n-1}) \\
 \Rightarrow E(\log_b(l)) &= E(\log_b(k \times b^{n-1})) \\
 \Rightarrow E(\log_b(l)) &= E(\log_b(k) + \log_b(b^{n-1})) \\
 \Rightarrow E(\log_b(l)) &= E(\log_b(k) + n - 1)
 \end{aligned}$$

$(n - 1) \in \mathbb{N}$, et $\forall (n_1; n_2) \in \mathbb{N}^2$, $(n_1 + n_2) \in \mathbb{N}$,
donc $\forall x \in \mathbb{R}$, $E(x + n) = E(x) + n$, donc:

$$\begin{aligned}
 \Rightarrow E(\log_b(l)) &= E(\log_b(k)) + n - 1 \\
 \Rightarrow E(\log_b(l)) + 1 &= E(\log_b(k)) + n \\
 \text{or, on sait que } E(\log_b(k)) &= 0, \text{ donc:} \\
 \Rightarrow E(\log_b(l)) + 1 &= 0 + n
 \end{aligned}$$

On retrouve bien:

$$n = E(\log_b(l)) + 1$$

On a donc vérifié l'équation et justifié la validité de la fonction $len(l)$.

Exemple:

On reprend la base $b = 26$ définie dans les préliminaires:

$$\begin{aligned}
 l &= (\text{bonjour})_{26} = (481363913)_{10} \\
 len(l) &= E(\log_{26}(l)) + 1 \\
 \log_{26}(l) &\approx 6.1361 \\
 E(\log_{26}(l)) &= 6 \\
 len(l) &= E(\log_{26}(l)) + 1 = 6 + 1 \\
 len(l) &= 7
 \end{aligned}$$

En comptant la quantité de chiffres dans $(\text{bonjour})_{26}$, on retrouve bien 7.

Indice d'un nombre

On s'intéresse à trouver le chiffre d'indice $n \in \mathbb{N}$ d'un entier naturel quelconque l représenté en base quelconque b . On note la fonction qui permet de trouver ce chiffre $ind(l; n)$.

On pose:

$$ind(l; n) = E\left(\frac{l}{b^n}\right) - E\left(\frac{l}{b^{n+1}}\right) \times b \quad (6)$$

On note que il s'agit de l'indice **partant de la droite** à $n = 0$. Pour avoir le chiffre d'indice n **partant de la gauche**, il suffit de prendre $len(l) - n$ comme indice, soit $ind(l; len(l) - n)$, ce qui permet de suivre les conventions informatiques.

Démonstration:

On pose:

$$\forall x \in \mathbb{N}$$

$$\forall n \in \mathbb{N}, \text{ l'indice du nombre recherché, tel que } 0 \leq n < len(x)$$

On peut écrire x tel:

$$x = \varepsilon_1 \times b^{n+1} + c \times b^n + \varepsilon_2 \quad (7)$$

avec:

$$0 \leq c < b; c \in \mathbb{N}$$

$$0 \leq \varepsilon_1; \varepsilon_1 \in \mathbb{N}$$

$$0 \leq \varepsilon_2 < b^n; \varepsilon_2 \in \mathbb{N}$$

On cherche à trouver une formule de c en fonction de x , n , et b .

$$c = \frac{x - \varepsilon_1 \times b^{n+1} - \varepsilon_2}{b^n}$$

$$\Rightarrow c = \frac{x - \varepsilon_2}{b^n} - \varepsilon_1 \times b$$

$$\Rightarrow c + \varepsilon_1 \times b = \frac{x - \varepsilon_2}{b^n}$$

$$\Rightarrow c + \varepsilon_1 \times b = \frac{x}{b^n} - \frac{\varepsilon_2}{b^n}$$

$$\begin{aligned}
(c + \varepsilon_1 \times b^{n+1}) &\in \mathbb{N}, \text{ donc } c + \varepsilon_1 \times b^{n+1} = E(c + \varepsilon_1 \times b^{n+1}) \\
\Rightarrow c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n} - \frac{\varepsilon_2}{b^n}\right) \\
\Rightarrow \text{frac}\left(\frac{x}{b^n}\right) &= \text{frac}\left(\frac{\varepsilon_1 \times b^{n+1} + c \times b^n + \varepsilon_2}{b^n}\right) \\
\Rightarrow \text{frac}\left(\frac{x}{b^n}\right) &= \text{frac}\left(\varepsilon_1 \times b + c + \frac{\varepsilon_2}{b^n}\right)
\end{aligned}$$

$$\begin{aligned}
(\varepsilon_1 \times b + c) &\in \mathbb{N}, \text{ donc:} \\
\text{frac}(\varepsilon_1 \times b + c) &= 0 \\
\Rightarrow \text{frac}\left(\varepsilon_1 \times b + c + \frac{\varepsilon_2}{b^n}\right) &= \text{frac}\left(\frac{\varepsilon_2}{b^n}\right) \\
\Rightarrow \text{frac}\left(\frac{x}{b^n}\right) &= \text{frac}\left(\frac{\varepsilon_2}{b^n}\right) \\
\Rightarrow \text{frac}\left(\frac{x}{b^n}\right) - \text{frac}\left(\frac{\varepsilon_2}{b^n}\right) &< 1
\end{aligned}$$

$$\begin{aligned}
\text{On reprend la formule de } (c + \varepsilon_1 \times b^{n+1}): \\
c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n} - \frac{\varepsilon_2}{b^n}\right) \\
\Rightarrow c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n}\right) - E\left(\frac{\varepsilon_2}{b^n}\right)
\end{aligned}$$

$$\begin{aligned}
\text{Or, on a:} \\
0 &\leq \varepsilon_2 < b^n \\
\Rightarrow 0 &\leq \frac{\varepsilon_2}{b^n} < \frac{b^n}{b^n} \\
\Rightarrow 0 &\leq \frac{\varepsilon_2}{b^n} < 1 \\
\Rightarrow E\left(\frac{\varepsilon_2}{b^n}\right) &= 0
\end{aligned}$$

$$\begin{aligned}
c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n}\right) - E\left(\frac{\varepsilon_2}{b^n}\right) \\
\Rightarrow c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n}\right) - 0 \\
\Rightarrow c + \varepsilon_1 \times b^{n+1} &= E\left(\frac{x}{b^n}\right)
\end{aligned}$$

$$c + \varepsilon_1 \times b^{n+1} = E\left(\frac{x}{b^n}\right)$$

Il suffit plus qu'à exprimer $\varepsilon_1 \times b^{n+1}$ en fonction de x , n et b .

$$\varepsilon_1 = \frac{x - c \times b^n - \varepsilon_2}{b^{n+1}}$$

$$\Rightarrow \varepsilon_1 = \frac{x}{b^{n+1}} - \frac{c}{b} - \frac{\varepsilon_2}{b^{n+1}}$$

$$\begin{aligned}
\varepsilon_1 &\in \mathbb{N}, \text{ donc } \varepsilon_1 = E(\varepsilon_1) \\
\Rightarrow \varepsilon_1 &= E\left(\frac{x}{b^{n+1}} - \frac{c}{b} - \frac{\varepsilon_2}{b^{n+1}}\right)
\end{aligned}$$

Or, on a:

$$\begin{aligned} \text{frac}\left(\frac{x}{b^{n+1}}\right) &= \text{frac}\left(\frac{\varepsilon_1 \times b^{n+1} + c \times b^n + \varepsilon_2}{b^{n+1}}\right) \\ \Rightarrow \text{frac}\left(\frac{x}{b^{n+1}}\right) &= \text{frac}\left(\varepsilon_1 + \frac{c}{b} + \frac{\varepsilon_2}{b^{n+1}}\right) \\ \Rightarrow \text{frac}\left(\frac{x}{b^{n+1}}\right) &= \text{frac}\left(\frac{c}{b} + \frac{\varepsilon_2}{b^{n+1}}\right), \text{ car } \varepsilon_1 \in \mathbb{N} \end{aligned}$$

Donc:

$$\text{frac}\left(\frac{x}{b^{n+1}}\right) - \text{frac}\left(\frac{c}{b}\right) - \text{frac}\left(\frac{\varepsilon_2}{b^{n+1}}\right) = 0$$

Donc, on peut décomposer la formule:

$$\begin{aligned} \varepsilon_1 &= E\left(\frac{x}{b^{n+1}} - \frac{c}{b} - \frac{\varepsilon_2}{b^{n+1}}\right) \\ \Rightarrow \varepsilon_1 &= E\left(\frac{x}{b^{n+1}}\right) - E\left(\frac{c}{b}\right) - E\left(\frac{\varepsilon_2}{b^{n+1}}\right) \end{aligned}$$

De plus, on a:

$$\begin{aligned} b^{n+1} &> c \times b^n > \varepsilon_2 \geq 0 \text{ (d'après leur définition)} \\ \Rightarrow 1 &> \frac{c}{b} > \frac{\varepsilon_2}{b^{n+1}} \geq 0 \\ \Rightarrow E\left(\frac{c}{b}\right) &= 0 \\ \Rightarrow E\left(\frac{\varepsilon_2}{b^{n+1}}\right) &= 0 \end{aligned}$$

Donc:

$$\begin{aligned} \varepsilon_1 &= E\left(\frac{x}{b^{n+1}}\right) - E\left(\frac{c}{b}\right) - E\left(\frac{\varepsilon_2}{b^{n+1}}\right) \\ \Rightarrow \varepsilon_1 &= E\left(\frac{x}{b^{n+1}}\right) - 0 - 0 \end{aligned}$$

$$\varepsilon_1 = E\left(\frac{x}{b^{n+1}}\right)$$

On a trouvé une formule pour ε_1 en fonction de x , n et b . On note:

$$c + \varepsilon_1 \times b^{n+1} = E\left(\frac{x}{b^n}\right)$$

$$\Rightarrow c = E\left(\frac{x}{b^n}\right) - E\left(\frac{x}{b^{n+1}}\right) \times b^{n+1}$$

Concaténation de chaînes de caractères

Soient deux entiers naturels u et v représentés en une même base b .

On note $u \parallel v$ la concaténation de u et v , tel que le résultat soit égal à un nombre composé de tous les chiffres de u , puis tous les chiffres de v .

Définition:

On pose:

$$(u; v) \in \mathbb{N}$$

$b \in \mathbb{N} \setminus \{0; 1\}$ la base dans laquelle u et v sont représentés.

On définit l'opération de concaténation de deux nombres:

$$u \parallel v = u \times b^{\text{len}(v)} + v$$

Exemple:

On pose $(u; v) \in \mathbb{N}$, tel que:

$$u = 125$$

$$v = 784$$

On a:

u et v sont représentés en décimal/base 10 (uniformité de base)

$$u \parallel v$$

$$= u \times b^{\text{len}(v)} + v$$

$$= 125 \times 10^{\text{len}(784)} + 784$$

$$= 125 \times 10^3 + 784$$

$$= 125000 + 784$$

$$= 125784$$

Propriétés:

- Non-commutativité:

$\forall (u; v) \in \mathbb{N}^2$, on a:

$$u \parallel v \neq v \parallel u$$

Exemple:

$$53 \parallel 21 = 5321$$

$$21 \parallel 53 = 2153$$

$$\Rightarrow 53 \parallel 21 \neq 21 \parallel 53$$

- Associativité:

$$\forall (u; v; l) \in \mathbb{N}_*^3, \text{ on a:}$$

$$(u \parallel v) \parallel l = u \parallel (v \parallel l)$$

Exemple:

$$\begin{aligned} & (53 \parallel 21) \parallel 93 \\ &= 5321 \parallel 93 \\ &= 532193 \end{aligned}$$

$$\begin{aligned} & 53 \parallel (21 \parallel 93) \\ &= 53 \parallel 2193 \\ &= 532193 \end{aligned}$$

Cependant, pour si on inclut 0, la concaténation n'est pas forcément commutative.

Exemple:

$$\begin{aligned} & (50 \parallel 0) \parallel 67 \\ &= 500 \parallel 67 \\ &= 50067 \end{aligned}$$

$$\begin{aligned} & 50 \parallel (0 \parallel 67) \\ &= 50 \parallel 67 \\ &= 5067 \end{aligned}$$

L'erreur provient du fait que $0 \times b^a = 0$ et $0 + v = v$, mais dans le cadre d'une implémentation en informatique, ceci ne pose pas de problème lorsqu'on choisit un espace (" ") comme représentation de 0 (ainsi, "bonjour" équivaut à " bonjour").