

Compression Arithmétique

I- Principe

Le codeur arithmétique traite le fichier dans son ensemble, en lui associant un unique nombre décimal rationnel. Ce nombre compris entre 0 et 1 dépend non seulement des symboles du fichier dans l'ordre où ils apparaissent mais aussi de leur distribution statistique.

II- Etapes

Considérons l'alphabet {A,E,I,O,U, !} avec les probabilités suivantes :

<i>Caractère</i>	<i>Probabilité</i>
A	0,2
E	0,3
I	0,1
O	0,2
U	0,1
!	0,1

But : coder EAI!

Etape 1 : Dans l'intervalle de probabilité $[0 ; 1]$, chaque symbole de l'alphabet se voit affecter un intervalle de probabilité

<i>Caractère</i>	<i>Probabilité</i>	<i>Intervalle</i>
A	0.2	$[0 ; 0,2[$
E	0.3	$[0,2 ; 0,5[$
I	0.1	$[0,5 ; 0,6[$
O	0.2	$[0,6 ; 0,8[$
U	0.1	$[0,8 ; 0,9[$
!	0.1	$[0,9 ; 1[$

Etape 2 : Codage du message EAI!

1. Initialisation de l'intervalle de travail $[0 ; 1[$
2. Le premier symbole est représenté par son intervalle de la première étape
3. Calcul des nouveaux intervalles des symboles suivants par la méthode suivante :

- nouvelle borne inférieure = ancienne borne inférieure + largeur ancien intervalle * borne inférieure symbole lu
--

- nouvelle borne supérieure = ancienne borne inférieure + largeur ancien intervalle * borne supérieure symbole lu

Sachant que ancienne « borne inférieure » et « largeur ancien intervalle » correspondent à la ligne précédente dans le tableau :

<i>Caractère</i>	<i>Intervalle</i>
Initialisation	[0 ; 1[
Lecture de E	[0,2 ; 0,5[
Lecture de A	[0,2 ; 0,26[
Lecture de I	[0,23 ; 0,236[
Lecture de I	[0,233 ; 0,2236[
Lecture de !	[0,23354 ; 0,2336[

Un nombre dans l'intervalle final, comme 0,23355, code de façon unique le message EAIL!

Etape 3 : Décodage EAIL!

$$V1=0,23355$$

Comme V1 est dans l'intervalle de E, la première lettre est E

$$V2=(V1-\text{Inf}(E))/\Delta(E)=(0,23355-0,2)/(0,5-0,2)=0,11133$$

Comme V2 est dans l'intervalle de A, la deuxième lettre est A

$$V3=(V2-\text{Inf}(A))/\Delta(E)=(0,11133-0)/(0,2-0)=0,5556$$

Comme V3 est dans l'intervalle de I, la troisième lettre est I

... et ainsi de suite jusqu'à la dernière lettre du mot, sachant que l'on peut soit définir sa taille pour pouvoir le décoder, soit définir un caractère d'arrêt.