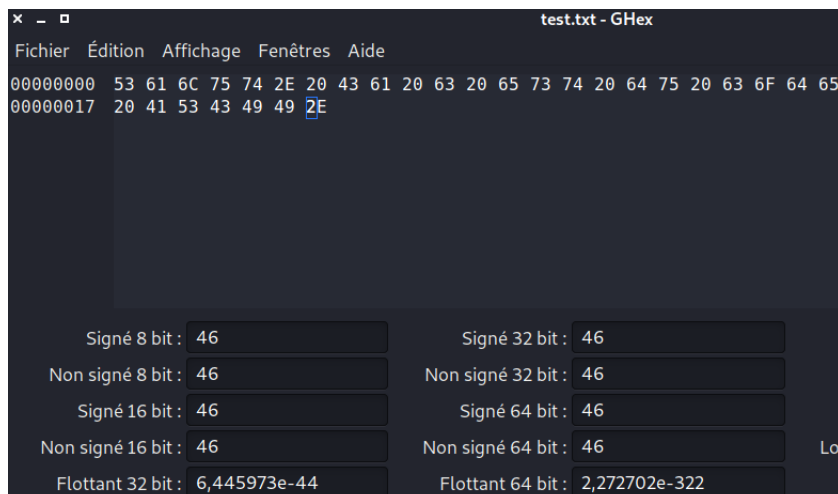


Représentation d'un texte en machine (TD):

Exercice 1: un premier code: Un fichier test.txt est corrompu. Néanmoins, lorsqu'on l'ouvre avec GHex (logiciel permettant d'examiner et de modifier des fichiers binaires), on peut lire le code suivant: Qui a t-il d'écrit dans ce fichier?



Exercice 2: un second code: Un hacker récupère les octets suivants: 01110000 01100001 01110011 01110011 00111101 01110011 01100101 01100011 01110010 01100101 01110100

Aider le à savoir ce qui se cache derrière cette série d'octets.

Exercice 3: évolution: Quand on est passé de la table ASCII à d'autres tables, les mémoires étant plus fiables qu'avant, de nouvelles techniques plus sûres de contrôle de parité ayant été inventées, le huitième bit a pu servir pour coder plus de caractères. Combien a t-on pu coder de caractères en plus.

Exercice 4: encodage: Lorsque je fais un commentaire dans un fichier python, j'ai l'affichage suivant:

```
#Il y a un problÃ"me d'encodage
```

prgm.py

Expliquer d'où vient le problème.

Exercice 5: UTF8: L'encodage UTF8 utilise entre 1 et 4 octets en respectant certaines règles:

- un texte en ASCII est codé de manière identique en UTF8. On utilise un octet commençant par un bit 0 à gauche (bit de poids fort)
- Si les bits de poids forts forment une suite de 1, ceci indique le nombre d'octets utilisés pour coder le caractère (les octets qui suivent commencent alors tous par 10).

UTF8	signification
0*** ****	1 octet codant 1 à 7 bits
110* **** 10** ****	2 octets codant 8 à 11 bits
1110 **** 10** **** 10** ****	3 octets codant 12 à 16 bits
1111 0*** 10** **** 10** **** 10** ****	4 octets codant 17 à 21 bits

1. Le symbole € correspond à la valeur décimale 8364. Convertir cette valeur en binaire.
2. Donner le codage UTF8 de ce caractère.
3. Étant donné un caractère codé en binaire sur n bits, combien d'octets sont nécessaires pour le coder en UTF?
4. Écrire un algorithme qui permet, à partir de la valeur décimale d'un caractère, de trouver le nombre d'octets nécessaires pour coder ce caractère en UTF8.