

Programmation 1

Examen

En Java le code de chaque caractère est donné par le standard Unicode, et est représenté en utilisant 2 octets. Le codage UTF-8 est un codage qui permet de réduire la taille de la représentation d'un caractère, pour les caractères les plus utilisés, quitte à allonger cette taille pour les caractères les moins utilisés.

Le codage UTF-8 est défini par :

- Les caractères de code 0 à 127 sont codés sur un octet dont le bit de poids fort est toujours nul.
- Les caractères de code supérieur à 127 sont codés sur plusieurs octets. Dans ce cas, les bits de poids fort du premier octet forment une suite de 1 de longueur égale au nombre d'octets utilisés pour coder le caractère, les octets suivants ayant 10 comme bits de poids fort.

Définition du nombre d'octets utilisés	
Représentation binaire UTF-8	Signification
0xxxxxxx	1 octet codant 1 à 7 bits
110xxxxx 10xxxxxx	2 octets codant 8 à 11 bits
1110xxxx 10xxxxxx 10xxxxxx	3 octets codant 12 à 16 bits

Exemples de codage UTF-8 :

Caractère	Code du caractère	Codage binaire UTF-8
A	65	01000001
é	233	11000011 10101001
€	8364	11100010 10000010 10101100

Nous nous proposons de définir une classe permettant d'utiliser un caractère sous sa forme Unicode, ou sous sa forme code UTF-8.

Question 1 (4 points)

Définir la méthode `bits(char c, int d, int f)` de la classe `Util`, qui retourne un octet contenant les bits d'indices `d` à `f` du caractère `c`. Ces bits sont rangés à droite dans l'octet résultat. Les bits ont des indices qui vont de 0 à 15, le bit d'indice 0 est à droite.

c	<code>bits(c, 6, 12)</code>
<code>0000110101001100</code>	<code>110101</code>

Question 2 (4 points)

Définir la méthode `code(char c)` de la classe `Caractere` qui retourne un tableau d'octets représentant le codage UTF-8 du caractère `c`. On utilisera la méthode décrite dans les tableaux précédents.

Question 3 (4 points)

Définir la méthode `deCode(byte [] t)` de la classe `Caractere` qui retourne le caractère dont le code UTF-8 est contenu dans le tableau `t`.

Question 4 (2 points)

Définir les constructeurs de la classe `Caractere`.

Question 5 (2 points)

Définir les accesseurs `getChar()` et `getUTF8()` de la classe `Caractere` qui retournent respectivement le caractère ou son code UTF-8.

Question 6 (4 points)

Définir la méthode `toStringUTF8()` de la classe `Caractere` qui retourne une chaîne de caractères représentant le tableau des octets du code UTF-8. La valeur de chaque élément du tableau étant rangée dans la chaîne résultat sous forme binaire.

`new Caractere("€").toStringUTF8()` a pour valeur "11100010 10000010 10101100"

```
public class Util {
    static byte bits(char c, int d, int f){...}
}
```

```
public class Caractere {
    private char leCaractere;
    private byte[] leCaractereUTF8;

    public Caractere(char c){...}
    public Caractere(byte [] t){...}
    public char getChar(){...}
    public byte[] getUTF8(){...}

    private char deCode (byte[] t){...}
    private byte[] code(char c){...}

    String toStringUTF8(){ }
}
```