

TP 5 : Data Representation

Warming up :

For the 16-bit codes : 0000000000101010 and 1000000000101010 Give their values, if they are representing :

- a 16-bit unsigned integer ;
- a 16-bit signed integer ;
- two 8-bit unsigned integers ;
- two 8-bit signed integers ;
- a 16-bit Unicode characters ;
- two 8-bit ISO-8859-1 characters

How would you represent "Hello, how are you ?" in ASCII ? (look for the comma, question mark, and space characters in the ASCII table)

1 Character encoding in HTML

HTML pages have several ways to encode characters :

- Entity references which is an alternative name for a series of characters. You can use an entity in the `&name;` format, where name is the name of the entity
- Numeric references that give the code of a Unicode character, in the form `&#nnn;` (décimal) or `&#xnnn;`.
- Direct binary coding in one of the formats discussed in the course (UTF-8, ISO-8859-1, ...). This requires adding a tag of the form

```
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
```

in the header of the HTML document.

We will consider the names of the following European municipalities :

Crèvecœur (France)
L'Haÿ-les-Roses (France)
Kroměříž (Tchéquie)
Gödöllő (Hongrie)
Süßen (Allemagne)
Præstø (Danemark)

1. Create an HTML document that displays these names with their entity references.
2. Look for Unicode codes for non-ASCII characters and use them to recreate the list with numeric references.
3. Create an HTML document in UTF-8 directly with your text editor.

2 Analyzing an error

2.1 In the html pages

Visit the <http://www.lsv.fr/~fhh/tp05-1>. Describe what is happening using the tables of characters available on the internet ([https://fr.wikipedia.org/wiki/Table_des_caractères_Unicode_\(0000-0FFF\)](https://fr.wikipedia.org/wiki/Table_des_caractères_Unicode_(0000-0FFF))). And Provide a solution to correct this problem.

3 File encoding problem

Download the text file <http://www.lsv.fr/~fhh/tp05-2.txt> and view it in your browser.

Search "é" with Firefox and Chrome.

Search "é" with the command "grep" on the downloaded file the.

What do you see ?

Suggest a solution to correct this type of problem.

4 Floating point

For this question, you can complete the float-skel.c and float-full.c files (look at <https://sites.google.com/site/farzadjafarrahmani/home/architecture-systems-course>).

Reminder : In c, the type `float` is represented according to the IEEE 754 standard in the 32 bit. 1 bit for the sign, 8 for the exponent and 23 for the mantisse.

```
typedef struct { int signe; int exposant; int mantisse; } fc;
```

1. Write a C function that decomposes a float into its three components. For example, the representation of 2.5 in IEEE 754 is :

0 . 1000 0000 . 010 0000 0000 0000 0000 0000

In this case, the returned structure would contain `signe = 0`, `exposant = 0x80 = 128` and `mantisse = 0x200000 = 2097152`.

Reminder : To do this, it is convenient to use `typecast`.

2. Create a function that does the opposite, that is to say that returns the `float` corresponding to a given `fc` structure.
3. Realize the actual addition based on the addition of integers by going through the structures of `fc`. To simplify, we will make the following restrictions : (i) both operands are positive (ii) no special cases NaN / Inf etc.

Addition in the `fc` type is done in three steps :

- (a) Standardize the two values, i.e. if the two exponents are different, we adjust the mantissa of one of the two according to the difference.
- (b) Add the sum of the two mantissas, taking into account the "hidden" bit representing the 1.
- (c) Normalize the mantissa for whatever is in $[1, 2)$, while adjusting the exponent of the result.