

THE REPRESENTATION OF ENRICHED TEXTS

Texts in ASCII or Unicode are simply sequences of characters. Text editors are the software that manipulates these sequences of characters. However, when we write a text, we can wish it gives a special, prettier, more readable form, as does a printer. We can play with the font — Times, Courier, etc. —, on character size — 11 points, 12 points, etc. —, on their form — roman, italic, etc. —, their weight — thin, fat, etc. We can also wish to divide a text into chapters and highlight the chapter titles, etc. However, the only characteristics that we can express with a code like ASCII, for example, are the case of a letter — lowercase or uppercase — and division into paragraphs, thanks to the carriage return symbol. Word processors are software that allow these more elaborate layouts.

This led to the enrichment of these formats, so as to:

1. qualify certain parts of the text, for example by putting certain parts in bold or italics,
2. structure the text in divisions: a text is not only a sequence of paragraphs, but is organized into parts, chapters, sections, subsections. . . ,
3. present certain information in the form of lists and tables,
4. allow to refer to other texts,
5. give information about the text: its title, its author(s), its creation date, language, keywords used to search for it among several texts, etc. This information on the text, and not text, are called metadata.

All these considerations are, of course, valid both for handwritten or printed texts than for texts processed by computers.

One of these enriched formats, which is used in particular for writing web pages is called HTML format. In HTML, to put a passage in bold, it is delimited by the tags `` and `` and to put it in italics, it is delimited by the tags `<i>` and `</i>`. Thus the text:

```
1 My first webpage
```

displays, in a browser, as:

My *first* **web**page

Like parentheses, tags come in pairs: we open the passage to make it bold with the `` tag and close it with the `` tag. A division of the text is delimited by the `<div>` and `</div>` tags, so the text:

```
1 <div>My first web page
2 <div> has a first subdivision to say "Hello everyone!"</div>
3 <div> and a second that ends with "See you soon!"</div>
4 </div>
```

will be displayed in the browser making these divisions explicit. Like parentheses, tags can nest within each other, but not overlap. We indicate that a passage is a title by delimiting it with the tags `<h1>` and `</h1>` and that it is a subtitle by delimiting it with the tags `<h2>` and `</h2>`. Likewise, other structuring of the text such as enumerations or tables are expressed by other tags. When writing a text, it is common to mention other texts: for example, to speak in a letter about a book that one has read. In the case of a handwritten or printed text, we generally give a reference to the quoted text, for example the title of the book and its author, so that the reader can refer to it if he wishes. When we want to express, on a page web, a reference to another page, we can do better than simply indicate the address of the web page in question (for example the address http://fr.wikipedia.org/wiki/Hypertext_Markup_Language); we can change the appearance of the passage where the reference is made, to indicate to the reader that

if he clicks on this passage, the browser will display the requested page. We use for this the tags `<a>` and ``: we frame the part of the text to be qualified by these two tags and we indicate inside the `<a>` tag the address of the referenced page. For example the text:

```
1 For details on the HTML language, you can consult the Wikipedia Hypertext
  3  Markup Language page.
```

displays, in a browser:

For details on the HTML language, you can consult [the *Wikipedia Hypertext Markup Language*](http://fr.wiki- pedia.org/wiki/Hypertext_Markup_Language) page.

If you click on the blue and underlined passage, the browser displays the page whose address is `http://fr.wikipedia.org/wiki/Hypertext_Markup_Language`. Such passage on which we can click to access another page is called a link, and text that contains at least one link is hypertext. The `<body>` and `</body>` tags delimit the text to be displayed in the browser. We indicate before this information the meta-data relating to the page: its title, the format used for accented letters, etc. In the end, here is an example of text in HTML format:

```
1 <html>
2 <head>
3   <meta http-equiv="content-type" content="text/html; charset=UTF-8">
4   </meta>
5   <title>An example</title>
6 </head>
7 <body>
8   <h1>A title</h1>
9   <h2>A subtitle</h2>
10  <div><a href="http://www.wikipedia.org/">A link</a></div>
11  <div><b>An important passage</b></div>
12 </body>
13 </html>
```

The header located between the two tags `<head>` and `</head>` indicates on the one hand that the text is expressed in UTF-8, this is the purpose of the line:

```
1 <meta http-equiv="content-type" content="text/html; charset=UTF-8">
2 </meta>
```

and on the other hand that the title of the page is An example.

The content is located between the `<body>` and `</body>` tags. There we find the tags ``, ``, `<i>`, `</i>`, `<h1>`, `</h1>`, `<h2>`, `</h2>`, `<div>`, `</div>`, `<a>` and `` that we have described. In a browser, the text is displayed as below (French translation):

A title

A subtitle

[A link](#)

An important passage

Exercise 1 (with answer)

Write an HTML page that presents the different IT projects of the students in a class.

```
1 <html>
2 <head>
3   <meta http-equiv="content-type" content="text/html; charset=UTF-8"></meta>
4   <title>EEB1 Projects</title>
5 </head>
6 <body>
7   <h1>The projects of the S3 class of the <a href="https://www.eeb1.com/fr/">
8     European School of Brussels 1</a></h1>
9   <div>
10    <a href="https://fr.wikipedia.org/wiki/Backgammon">A program that <b>
11      plays Backgammon</b></a>
12  </div>
13  <div>
14    <a href="https://fr.wikipedia.org/wiki/JPEG">A program that <b>compresses
15    images</b></a>
```

Listing 1: Project webpage: http://www.barsamian.am/2023-2024/S3ICT/TP1_Example_EN.html

Exercise 2

Write an HTML page that presents the list of concerts and shows presented in a theater.

Exercise 3

Change the following HTML text so that the word “first” appears not in italics, but in bold.

```
1 My <i>first</i> web page
```

Exercise 4

This HTML text is incorrect. How to correct it?

```
1 You have to <i>understand</i> the coding of digital objects to master them.
```

Exercise 5

In this text, to which website is the link pointing? How is this text displayed in a browser? What is the point of looking at the HTML source of this page before clicking?

```
1 Your bank account has an anomaly. Click <a href="http://big-scam.com">here</a>
  > for help.
```

Exercise 6

Give the HTML source of the following text knowing that the text in blue and underlined is a link to the page <http://www.mybook.com/page2>:

We can consult the following [page](#).