

## LA REPRÉSENTATION DES TEXTES ENRICHIS

Les textes en ASCII ou en Unicode sont simplement des suites de caractères. Les éditeurs de texte sont les logiciels qui manipulent ces suites de caractères. Toutefois, quand on écrit un texte, on peut souhaiter lui donner une forme spéciale, plus jolie, plus lisible, comme le fait un imprimeur. On peut jouer sur la police de caractères — Times, Courier, etc. —, sur la taille des caractères — 11 points, 12 points, etc. —, sur leur forme — romain, italique, etc. —, leur graisse — maigre, gras, etc. On peut aussi souhaiter découper un texte en chapitres et mettre en valeur les titres des chapitres, etc. Or, les seules caractéristiques que l'on puisse exprimer avec un code comme l'ASCII, par exemple, sont la casse d'une lettre — minuscule ou majuscule — et le découpage en paragraphes, grâce au symbole retour chariot. Les traitements de texte sont les logiciels qui permettent ces mises en pages plus élaborées.

Ceci a amené à enrichir ces formats, de manière à :

1. qualifier certaines parties du texte, par exemple en mettant certaines parties en gras ou en italique,
2. structurer le texte en divisions : un texte n'est pas uniquement une suite de paragraphes, mais est hiérarchisé en parties, chapitres, sections, sous-sections. . . ,
3. présenter certaines informations sous forme de listes et de tables,
4. permettre de faire référence à d'autres textes,
5. donner des informations sur le texte : son titre, son ou ses auteur(s), sa date de création, sa langue, des mots-clés utilisés pour le rechercher parmi plusieurs textes, etc. Ces informations sur le texte, et non du texte, sont appelées des méta-données.

Toutes ces considérations sont, bien entendu, valables aussi bien pour les textes manuscrits ou imprimés que pour les textes traités par les ordinateurs.

L'un de ces formats enrichis, qui est utilisé en particulier pour écrire des pages web est appelé le format HTML. En HTML, pour mettre un passage en gras, on le délimite par les balises `<b>` et `</b>` et pour le mettre en italique, on le délimite par les balises `<i>` et `</i>`. Ainsi le texte :

```
1 Ma première page web
```

s'affiche dans un navigateur :

Ma *première* page **web**

Comme les parenthèses, les balises vont par deux : on ouvre le passage à mettre en gras avec la balise `<b>` et on le ferme avec la balise `</b>`. Une division du texte est délimitée par les balises `<div>` et `</div>`, ainsi le texte :

```
1 <div>Ma première page web
2 <div> comporte une première sous-division pour dire « Bonjour tout le monde !
   »</div>
3 <div> et une seconde qui finit par « À bientôt ! »</div>
4 </div>
```

s'affichera dans le navigateur en rendant ces divisions explicites. Comme les parenthèses, les balises peuvent s'emboîter les unes dans les autres, mais pas se chevaucher. On indique qu'un passage est un titre en le délimitant par les balises `<h1>` et `</h1>` et que c'est un sous-titre en le délimitant par les balises `<h2>` et `</h2>`. De même, les autres structurations du texte comme les énumérations ou les tableaux sont exprimées par d'autres balises. Quand on écrit un texte, il est fréquent de mentionner d'autres textes : par exemple, de parler dans une lettre d'un livre que l'on a lu. Dans le cas d'un texte manuscrit ou imprimé, on donne en général une référence du texte cité, par exemple le titre du livre et son auteur, afin que le lecteur puisse s'y référer s'il le souhaite. Quand on veut exprimer, dans une page web, une référence à une autre page, on peut faire mieux que simplement indiquer l'adresse de la page web en question

(par exemple l'adresse `http://fr.wikipedia.org/wiki/Hypertext_Markup_Language`); on peut changer l'apparence du passage où l'on fait la référence, pour indiquer au lecteur que s'il clique sur ce passage, le navigateur affichera la page demandée. On utilise pour cela les balises `<a>` et `</a>`: on encadre la partie du texte à qualifier par ces deux balises et on indique à l'intérieur de la balise `<a>` l'adresse de la page référencée. Par exemple le texte :

```
1 Pour les détails sur le langage HTML, on pourra consulter <a href="http://fr.
  wikipedia.org/wiki/Hypertext_Markup_Language">la page <i>Hypertext Markup
  Language de Wikipédia</i></a>.
```

qui affiche dans un navigateur :

Pour les détails sur le langage HTML, on pourra consulter la page [Hypertext Markup Language de Wikipédia](http://fr.wikipedia.org/wiki/Hypertext_Markup_Language).

Si l'on clique sur le passage en bleu et souligné, le navigateur affiche la page dont l'adresse est `http://fr.wikipedia.org/wiki/Hypertext_Markup_Language`. Un tel passage sur lequel on peut cliquer pour accéder à une autre page s'appelle un lien, et un texte qui contient au moins un lien est un hypertexte. Les balises `<body>` et `</body>` délimitent le texte à afficher dans le navigateur. On indique avant ces informations les méta-données relatives à la page : son titre, le format utilisé pour les lettres accentuées, etc. Voici, au bout du compte, un exemple de texte au format HTML :

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

L'en-tête situé entre les deux balises `<head>` et `</head>` indique d'une part que le texte est exprimé en UTF-8, c'est l'objet de la ligne :

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

et d'autre part que le titre de la page est Un exemple.

Le contenu est situé entre les balises `<body>` et `</body>`. On y retrouve les balises `<b>`, `</b>`, `<i>`, `</i>`, `<h1>`, `</h1>`, `<h2>`, `</h2>`, `<div>`, `</div>`, `<a>` et `</a>` que l'on a décrites. Dans un navigateur, le texte s'affiche de la manière ci-dessous :

**Un titre**

**Un sous-titre**

[Un lien](#)

Un passage important

### Exercice 1 (avec corrigé)

Écrire une page HTML qui présente les différents projets informatiques des élèves d'une classe.

```

1 <html>
2 <head>
3   <meta http-equiv="content-type" content="text/html; charset=UTF-8"></meta>
4   <title>Projets EEB1</title>
5 </head>
6 <body>
7   <h1>Les projets de la classe de S3 de l'<a href="https://www.eeb1.com/fr/">
   école européenne de Bruxelles 1</a></h1>
8   <div>
9     <a href="https://fr.wikipedia.org/wiki/Backgammon">Un programme qui <b>
   joue aux Backgammon</b></a>
10  </div>
11  <div>
12    <a href="https://fr.wikipedia.org/wiki/JPEG">Un programme qui <b>
   compresse des images</b></a>
13  </div>
14 </body>
15 </html>

```

Listing 1 – Page web de projets : [http://www.barsamian.am/2023-2024/S3ICT/TP1\\_Exemple\\_FR.html](http://www.barsamian.am/2023-2024/S3ICT/TP1_Exemple_FR.html)

## Exercice 2

Écrire une page HTML qui présente la liste des concerts et des spectacles présentés dans un théâtre.

## Exercice 3

Changer le texte HTML suivant pour que le mot « première » apparaisse non en italique, mais en gras.

```
1 Ma <i>première</i> page web
```

## Exercice 4

Ce texte HTML est incorrect. Comment le corriger ?

```
1 Il faut <i>comprendre/i> le codage des objets numériques pour les maîtriser.
```

## Exercice 5

Dans ce texte, vers quel site web pointe le lien ? Comment ce texte s'affiche-t-il dans un navigateur ? Quel est l'intérêt de regarder le source HTML de cette page avant de cliquer ?

```
1 Votre compte bancaire présente une anomalie. Cliquer <a href="http://grosse-
   arnaque.com">ici</a> pour avoir de l'aide.
```

## Exercice 6

Donner le source HTML du texte suivant sachant que le texte en bleu et souligné est un lien vers la page <http://www.monlivre.fr/page2> :

On pourra consulter la [page](#) suivante.