

Décrire et manipuler un document numérique

Cours 10 : Autres formats de documents

Loïc Grobol <lgrobol@parisnanterre.fr>

2022-03-28

Markdown

Markdown est un format de documents créé par Aaron Swartz et John Gruber en 2004 sur les principes suivants :

- Facile à lire et à éditer par des humains sans outil spécialisé.
- Facile à convertir dans d'autres formats numériques standard : par défaut en HTML.

Le nom est un jeu de mot sur *markup*, qui désigne les formats dits « balisés » comme HTML ou XML.

La spécification originale et un programme pour convertir de Markdown ont été publiés à <https://daringfireball.net/projects/markdown> en 2004.

Cette spécification est volontairement informelle, un des objectifs étant de permettre des évolutions futures vers des formats dérivés (ou **dialectes**) de Markdown adaptés à différents usages, ce qui a effectivement été le cas.

La légèreté du format a permis son inclusion dans énormément d'applications. Essentiellement, si vous voulez proposer du texte formaté en évitant de créer un éditeur WYSIWIG, Markdown est une bonne solution.

Pour visualiser, tester et expérimenter avec Markdown, vous pouvez utiliser un des nombreux éditeurs en ligne disponible, par exemple <https://dillinger.io> ou un éditeur local comme Marktext.

Évidemment — c'est tout le principe —, vous pouvez aussi écrire en Markdown dans n'importe quel éditeur de texte (bloc note, notepad++, ...).

Syntaxe

La spécification originale décrit Markdown en termes de balises HTML, et on va faire pareil.

Un fichier Markdown est un fichier texte, généralement avec l'extension `.md`, qui comprend du texte, ainsi que des annotations de format et structures, qui utilisent exclusivement des caractères non alphabétiques.

Ce qui suit est une brève introduction à la syntaxe. Voir par exemple les spécifications GFM ou Pandoc pour des dialectes modernes et complets.

Base

En Markdown, on écrit du texte en l'écrivant. Plus clairement :

```
Ceci est un texte.
```

Équivaut en HTML à

```
<p>Ceci est un texte.</p>
```

Les espaces multiples ne sont pas préservées et un retour à la ligne équivaut à une espace. En revanche une ligne blanche marque un changement de paragraphe.

```
Ceci est un texte.  
Le paragraphe    continue ici.  
  
Mais ici c'est en est un autre.
```

```
<p>Ceci est un texte. Le paragraphe continue ici.</p>  
<p>Mais ici c'en est un autre</p>
```

Emphase

Markdown a, comme HTML, deux niveaux d'emphase :

- `*hello*` équivaut à `hello` et est en général marqué en italiques.
- `**hello**` équivaut à `hello` et est en général marqué en gras.

```
Ceci est un texte avec emphase.
```

```
<p>Ceci est un texte avec emphase.
```

On peut utiliser `_` à la place de `*`

Titres

Les titres sont notés en débutant une ligne avec autant de `#` que le niveau voulu :

```
# Titre de niveau 1  
## Titre de niveau 2  
### Titre de niveau 3
```

```
<h1>Titre de niveau 1</h1>
<h2>Titre de niveau 2</h2>
<h3>Titre de niveau 3</h3>
```

Listes non-ordonnées

Les listes non-ordonnées se notent en démarrant une ligne par -, * ou +.

```
- un item
- un autre item

- un autre item dans une autre liste
- et un autre
```

```
<ul>
<li>un item</li>
<li>un autre item</li>
</ul>

<ul>
<li>un autre item dans une autre liste</li>
<li>et un autre</li>
</ul>
```

On peut également créer des sous-listes, signalées par l'indentation

```
- un item
  sur plusieurs lignes
- un autre item
  - avec
  - des sous-items
```

```
<ul>
<li>un item sur plusieurs lignes</li>
<li>un autre item
  <ul>
    <li>avec</li>
    <li>des sous-items</li>
  </ul>
</li>
</ul>
```

Listes ordonnées

On crée des listes ordonnées de la façon suivante

1. Une liste
2. Avec des numéros
3. Ordonnée, donc

```
<ol>
<li>Une liste</li>
<li>Avec des numéros</li>
<li>Ordonnée, donc</li>
</ol>
```

Lien

Les liens hypertextes peuvent être notés *inline*

Voici un `[lien](http://example.com)` inline.

Voici un `lien` inline.

Ils peuvent aussi être référencés ailleurs dans le document en leur donnant une étiquette arbitraire

Voici un `[lien][addresse]` étiqueté.

...

`[addresse]: http://example.com`

Voici un `lien` étiqueté.

HTML

Dans beaucoup d'applications, on autorise aussi l'insertion de balises HTML directement dans Markdown :

Du texte `*Markdown*` avec du `**HTML**` `<small>dedans</small>`.

Du texte `Markdown` avec du `HTML` `<small>dedans</small>`.

Éditeurs avancés

Plusieurs outils utilisant Markdown proposent des fonctionnalités en plus de Markdown qui pourraient vous être utiles comme

- L'édition collaborative de documents : démo.
- Générer des diapositives, comme pour ce cours ! Consultez le fichier source de celui-ci, par exemple.
- Des documents exécutables contenant des programmes, comme RMarkdown.

Exercice

Reprendre l'exercice lettre en formatant à l'aide de Markdown plutôt que HTML. Attention : il ne sera pas possible de tout représenter.

Pour vous entraîner, revoir les points précédents ou en apprendre plus, vous pouvez également suivre le tutoriel interactif de <https://www.markdowntutorial.com/fr>.

Exercice : rechercher-remplacer avec des RegEx

En plus de faire des recherches avec des RegEx, on peut s'en servir pour faire des remplacements.

Dans un éditeur local comme Notepad++, c'est déjà intégré. <https://regexr.com/> vous permet de le faire en ligne.

Le principe est simple : comme avec un rechercher-remplacer dont vous avez l'habitude, mais ce qu'on recherche, c'est une expression régulière.

Par exemple, essayez de remplacer `\bc\w*\b` par `chats` dans

```
Les chefs de section des commissariats sont invités à courir.
```

Plus intéressant : on peut aussi conserver une partie de l'expression régulière : remplacer `\bc(\w*)\b` par `C$1` (dans Regexr) ou par `C\1` (dans Notepad++) dans

```
Les chefs de section des commissariats sont invités à courir.
```

Principe : `$n` (ou `\n`) où `n` est un nombre sera remplacé par le contenu du `n`-ième groupe de capture (signalé par des parenthèses).

Maintenant, utilisez ce nouvel outil pour convertir du Markdown en HTML : exercice.

Ou si vous peinez, vous pouvez vous rafraîchir la mémoire sur les RegEx avec les premiers niveaux (jusqu'à *Intermediate* inclus) de <https://regexcrossword.com>