

Créer des documents professionnels avec *Markdown* et *Pandoc*

L'écriture numérique ouverte

Bernard Pochet

2021 (cc-by)

Résumé

Ce court document, qui peut être considéré comme un tutoriel, est destiné à montrer les possibilités de *Markdown* associé à *Pandoc* pour la rédaction, l'édition et la publication de documents pdf. Cette méthode de rédaction permet de se passer des traitements de texte qui mélangent la rédaction d'un texte et la mise en page de ce texte. Elle est à la fois simple et durable (les fichiers créés sont ouverts et peuvent être édités avec un simple éditeur).

1 Introduction

Markdown est un langage de balisage léger co-créé par Aaron Swartz et John Gruber en 2004 afin de générer facilement du code HTML pour le web. *Markdown* est prévu pour être facile à lire et facile à écrire¹.

Depuis sa création il a été étendu pour être utilisé pour produire d'autres types de documents dont le *pdf* que vous consultez actuellement. L'utilisation de *Pandoc*² qui fait appel à *LaTeX*³ permet de produire des documents avec une qualité identique à celle des documents produits directement avec *LaTeX* (Pierre, 2017).

L'objectif de ce court document n'est pas de présenter *Markdown* mais de démontrer les possibilités qu'il offre⁴.

La méthode décrite ci-dessous ne doit pas être considérée comme une complication supplémentaire mais fait au contraire partie d'un mouvement *Low Tech*. Ce document a été rédigé avec un simple éditeur de texte *Gedit*⁵ qui permet de se concentrer sur la rédaction du texte sans se préoccuper de sa forme finale (Dehut, 2018). Il rassemble les fonctions de base présentes dans tous les bons manuels (Mailund, 2019) ou des fonctions plus "pratiques" glanées au fil de lectures.

Markdown fait partie des solutions *Single Source publishing*, c'est à dire qu'avec un même fichier source on peut produire des documents dans différents formats⁶. Ce document sera spécifiquement utilisé pour illustrer la création d'un document *pdf*.

1. Comme les commandes *Markdown* sont assez simples, on peut les maîtriser en dix minutes.
2. <https://github.com/jgm/pandoc/wiki/Pandoc-User's-Manual-in-French>
3. <https://fr.wikibooks.org/wiki/LaTeX>
4. Pour une présentation de *Markdown*, lisez ce billet : <https://infolit.be/wordpress/quel-outil-pour-rediger>.
5. <https://fr.wikipedia.org/wiki/Gedit>
6. essentiellement HTML, ePub et pdf.

2 La description et la structure du document

Un des avantages de *Markdown* est de faciliter la gestion et la visualisation de :

- la description d'un document avec ses métadonnées ;
- la structure d'un document avec les niveaux de titre.

2.1 Les métadonnées

Pour faciliter le partage et la diffusion d'un document numérique, celui-ci doit contenir sa propre description. On parle de métadonnées. La structure de ces métadonnées fait l'objet de normes ⁷ et permet le partage d'informations entre outils ainsi que l'alimentation automatique des bases de données. Dans un document *Markdown*, les métadonnées se trouvent au début du document, dans une partie appelée *YAML*. Voici le contenu de l'entête *YAML* de ce document. Elle contient les métadonnées du fichier *Markdown* ainsi qu'une série d'instructions pour la création du document *pdf*.

```
---
documentclass: article
title: Créer des documents professionnels avec *Markdown* et *Pandoc*
subtitle: L'écriture numérique ouverte
author: Bernard Pochet
Affiliation: ULiège
date: 2021 (cc-by)
abstract: Ce court document est destiné à démontrer les possibilités de *Markdown* ...
keywords: Markdown, rédiger
subject: écriture numérique ouverte
Modified: 18 janvier 2021
right: CC-BY 0.4
lang: fr-CA
geometry:
- margin=2cm
- a4paper
fontsize: 11pt
toc: false
toc_depth:
numbersections: true
bibliography: library.bib
csl: apa.csl
links-as-notes: true
linkcolor: blue
---
```

Il y a donc des instructions de **mise en page** ⁸ :

- `documentclass`: avec *book* ou *report* ou *article* (et également *chapter* ou *part*). La mise en page sera différente pour ces différents types ⁹
- `lang`: le choix de `fr-CA` va franciser la mise en page
- `geometry`: pour définir les 4 marges et le format de la page
- `fontsize`: la taille des caractères

7. À l'heure actuelle, il n'est malheureusement pas encore possible d'intégrer l'ensemble des champs de description *Dubin Core* dans un document *Markdown*.

8. Cette liste n'est pas limitative. Il existe bien d'autres possibilités. Elle s'est construite progressivement au rythme des besoins. Voir le manuel Pandoc : <https://pandoc.org/MANUAL.html> pour d'autres instructions de formatage

9. Notez qu'utiliser le champ *abstract* avec le type *book* provoque une erreur.

- toc: true/false (crée une table des matières en début du document. Également lof: et lot: pour les figures et les tableaux)
- toc_depth: de 1 à x (précise la profondeur de cette table des matières)
- numbersections: true/false (pour numéroter les titres. Pour empêcher la numérotation d'un titre en particulier, il faut lui ajouter "{-}"¹⁰).
- bibliography: library.bib (le nom du fichier *bibtex* créé, par exemple, avec *Zotero*)
- csl: apa.csl (le style bibliographique)
- links-as-notes: true/false (créer des notes de bas de page pour tous les liens)
- linkcolor: blue (pour définir la couleur des liens)

Et des **métadonnées** :

- title:
- subtitle:
- author:
- date:
- abstract:
- Affiliation:
- keywords:
- subject:
- project:
- modified:
- right:

Les cinq premiers items servent aussi pour la mise en page puisque, comme pour ce document, avec le type *article*, ces informations apparaissent en tête du document.

Il est possible d'ajouter d'autres directives de mise en page spécifiques à *LaTeX*¹¹. Cependant, on arrive alors à un niveau de complexité plus élevé qui va à l'encontre de l'intérêt de cette méthode basée sur la simplicité.

2.2 Les niveaux de titre

Pour être correctement structuré, un texte doit être divisé en parties, chapitres, sous-chapitres, etc. Le niveau des titres va déterminer cette structure. Avec *Markdown*, ces niveaux sont simplement indiqués avec les balises "#" dont le nombre détermine la profondeur du titre :

```
# Titre de niveau 1
## Titre de niveau 2
### Titre de niveau 3
#### Titre de niveau 4
##### Titre de niveau 5
##### Titre de niveau 6
```

Dans le texte, la séparation en paragraphes est réalisée par un double saut de ligne.

3 Les fonctions de mise en page

3.1 La mise en forme

Les fonctions de base sont l'*italique* et le **gras**. Il est aussi possible de ~~barre~~ un texte.

Les indices et les exposants sont réalisés avec "H~2~O" pour H₂O et "X^2^" pour X².

10. Utilisé ici pour la section "Bibliographie".

11. y compris le choix de *Templates* de mise en page spécifique.

Ce passage :

Ceci est un texte italique, ****Ceci est un texte gras**** et ~~~Ceci est un texte barré~~~

Donnera :

Ceci est un texte italique, **Ceci est un texte gras** et ~~Ceci est un texte barré~~

(il est possible de les combiner).

3.1.1 Les listes

Markdown permet de créer des listes simples et des listes numérotées.

Pour les listes simples, il faut utiliser le tiret (ou le +) suivi d'un espace.

Ceci :

```
- item 1
- item 2
+ item 3
```

Donnera :

```
— item 1 12
— item 2
— item 3
```

Pour les listes numérotées, il faut mettre un chiffre (peu importe le chiffre mais le premier indique le début de la numérotation), suivi d'un point et d'une espace.

Ceci :

```
1. item 1
1. item 2
1. item 3
```

Donnera :

```
1. item 1
2. item 2
3. item 3
```

Les listes doivent être précédées d'un saut de ligne. Il est possible de combiner les deux types de listes comme par exemple :

```
1. item 1
2. item 2
   — item 2.1
   — item 2.2
3. item 3
```

Peut-être moins utile, il est aussi possible de créer des cases à cocher.

Ceci :

```
- [x] item 1, coché
- [ ] item 2, non-coché
  - [x] item 2.1, coché
```

12. Notez la qualité esthétique. Difficile d'obtenir un tel résultat avec *Mot* sous *Fenêtres* (*Word* sous *Windows*...;-).

- [] item 2.2, non-coché
- [] item 3, non-coché
- [X] item 4, coché

Donnera :

- item 1, coché
- item 2, non-coché
 - item 2.1, coché
 - item 2.2, non-coché
- item 3, non-coché
- item 4, coché

3.1.2 Les citations

Pour citer et mettre en évidence un passage de texte, il faut utiliser le caractère ">" en début de ligne.

Ce passage :

```
> "But here's the most important bit, saved for last: none of this matters if you want
to write a book. Quite a few people have told me that they want to write a book, but
they're not sure about which tools to use. My advice: all you need to write a book is
a program that allows you to write text into a file [Ball2018]."
```

Donnera :

“But here’s the most important bit, saved for last: none of this matters if you want to write a book. Quite a few people have told me that they want to write a book, but they’re not sure about which tools to use. My advice: all you need to write a book is a program that allows you to write text into a file (Ball, 2018).”

Le texte est centré dans la page avec une justification gauche et droite (comme le reste du texte). Notez au passage l’ajout de la citation à l’auteur du texte qui va reproduire le renvoi à la bibliographie suivant la norme sélectionnée dans l’entête *YAML* et ajouter la référence dans la bibliographie en fin de document (voir plus loin pour plus d’explications à propos des possibilités en matière de bibliographie).

3.1.3 Les codes et textes non formatés

Pour afficher du code sans formatage (comme fait à plusieurs reprises plus haut), il faut :

- soit encadrer le texte (ligne précédente et ligne suivante) avec ‘ ‘ ‘ (trois accents graves consécutifs) ;
- soit, pour une seule ligne par exemple, faire précéder le texte par une tabulation ou au moins trois espaces.

3.1.4 Les notes de bas de page

Dans ce document, les notes sont créées avec l’ajout d’un "^" et le texte de la note est encadré de "[" et "]" :

```
ceci est le texte^[Et ceci est la note.]
```

Les notes de bas de page sont numérotées automatiquement. La méthode utilisée ici est spécifique à la création de document *pdf*.

Il y a une méthode un peu plus complexe (et compatible *ePub* et HTML) qui se fait en deux temps, avec l’appel de note :

ceci est le texte^[numéro de la note].

et avec toutes les notes regroupées à la fin du document .md :

[^{numéro de la note}]: Et ceci est la note.

3.1.5 Les liens

Pour créer un lien vers une page Web, il faut utiliser la séquence `[]()`. Entre les crochets, on va placer le texte à afficher et, entre les parenthèses, l'URL cible :

`[texte à afficher](https://url_cible.org)`

Il est aussi possible de créer des liens à l'intérieur d'un document ¹³. Pour cela :

- il faut ajouter une ancre (ici, après un titre) : `# chapitre 10 {#chap10}`
- créer le lien avec la même séquence : `texte texte texte, voir [chapitre 10](@chap10)`

3.1.6 Les figures

La syntaxe pour insérer une figure est comparable à celle de la création d'un lien. Il faut seulement faire précéder l'instruction d'un "!". L'image "appelée" peut être un fichier présent sur l'ordinateur (attention à appeler le fichier au bon endroit sur le disque dur) ou une image quelque part sur Internet (il faut alors entrer une URL correcte).

S'il y a une légende (texte entre le `[et]`), les images sont centrées et automatiquement numérotées avec "Figure x :" suivi de la légende. *Pandoc* va éventuellement déplacer le texte avant et après les figures pour optimiser le remplissage des pages.

Ceci :

`![Le logo de *Markdown*](markdown-large.png)`

Donnera :



FIGURE 1 – Le logo de *Markdown*

S'il n'y a pas de légende (texte entre le `[et]`) ou si la commande se termine par un `\`, la figure sera alignée à gauche et la mise en page ne sera peut-être pas optimale.

Il est possible de modifier la taille de l'image en ajoutant `{ width=50% }` derrière l'appel d'image (ici, réduction de la taille de 50%).

3.1.7 Les équations mathématiques

Il est aussi possible d'intégrer des équations mathématiques avec la syntaxe *TeX Math* de *LaTeX*. Elles seront correctement reproduites dans le fichier pdf exporté. Ces équations sont encadrées par "\$\$".

13. Ne fonctionne pas si les liens sont transformés en note en bas de page comme pour ce document.

La séquence suivante :

```
$$x = \frac {-b \pm \sqrt{b^2 - 4ac}}{2a}$$
```

donnera :

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Encore une fois, l'objectif reste juste de montrer ce qui est possible et non de détailler cette syntaxe (vous trouverez un tutoriel très complet sur *WikiBooks*¹⁴).

3.1.8 Les tableaux

Markdown gère plutôt bien les tableaux. Il faut utiliser le “|” pour séparer les colonnes et le saut de ligne pour séparer les lignes.

La séquence suivante :

```
| **entête 1** (gauche) | **entête 2** (centre) | **entête 3** (droite) |
|:---|:---:|---:|
| cellule 1.1 | cellule 2.1 | cellule 3.1 |
| cellule 1.2 | cellule 2.2 | cellule 3.2 |
| cellule 1.3 | cellule 2.3 | cellule 3.3 |
: Titre du tableau
```

Donnera :

TABLE 1: Titre du tableau

entête 1 (gauche)	entête 2 (centre)	entête 3 (droite)
cellule 1.1	cellule 2.1	cellule 3.1
cellule 1.2	cellule 2.2	cellule 3.2
cellule 1.3	cellule 2.3	cellule 3.3

Pour aligner à gauche, centrer ou aligner à droite le contenu d'une colonne, il faut le déclarer dans la deuxième ligne en ajoutant “:” à gauche, “:” des deux côtés ou “:” à droite des trois “-” de la deuxième ligne. Les enrichissements (gras, italique...) sont autorisés pour toutes les cellules du tableau. Il existe sur le Web un générateur de tables¹⁵ *Markdown* qui facilite leur création.

Le titre du tableau apparaît au dessus, numéroté, comme pour les figures.

3.1.9 Autres codes utiles pour la mise en page

Deux petites commandes peuvent encore se révéler utiles :

- l'insertion d'un saut de page : “\newpage”
- l'insertion d'une ligne horizontale : “***”
- les espaces insécables (pour insérer une ligne vide ou empêcher le positionnement en début de ligne de “?”, “:” ou “;”) : “ ”
- la commande *LaTeX* pour créer un cadre de texte “\fbox{texte texte texte}”

14. <https://en.wikibooks.org/wiki/LaTeX/Mathematics>

15. https://www.tablesgenerator.com/markdown_tables

- l'échappement avec “\” pour afficher des caractères (comme [ou #) qui sont interprétés par *Pandoc* comme des commandes.

3.2 Les citations et la bibliographie

Un élément puissant de cette méthode est la facilité d'insertion de citations dans le texte et de génération d'une bibliographie à la fin du document.

Pour que cela fonctionne :

1. il faut que deux fichiers soient présents dans le même répertoire que le document .md :
 - un fichier *BibTex* créé avec un gestionnaire bibliographique. Avec *Zotero*, il faut utiliser la fonction “exporter les documents” et choisir le format *BibTex* pour créer un fichier .bib ;
 - copier le fichier .csl (*Citation Style Language*) dans le répertoire ;
2. que ces fichiers soient renseignés dans l'entête YAML (voir le chapitre sur les métadonnées)
3. insérer les citations :
 - soit avec l'appel [*@auteur0000*]¹⁶ pour une citation avec des parenthèses : “texte texte texte (Auteur, 0000)” ;
 - soit avec l'appel *@auteur0000* pour une citation directe : “texte texte texte Auteur (0000)”.
4. ajouter le titre “Bibliographie” en fin de document.

4 Les outils

Pour débiter, *Typora*¹⁷ est un logiciel (gratuit mais non libre) qui permet de se familiariser avec *Markdown*. Il guide l'utilisateur dans l'installation de *Pandoc* et de *LaTeX*. *MarkText*¹⁸ est une alternative libre à *Typora*.

Il est aussi possible d'utiliser *Markdown* avec *FocusFox*¹⁹, *Dillinger*²⁰, *Stack Edit*²¹, *editor.md*²² ou *Stylo*²³, via le navigateur, sans installation.

Le site *Alternative To*²⁴ liste l'ensemble des logiciels d'édition possibles avec des indications sur les plateformes disponibles et le caractère *libre* ou non.

5 La transformation du document

Avec *Typora* (voir ci-dessus) et quelques autres outils, il est possible d'exporter directement un fichier *Markdown* en fichier *pdf*. Ce fichier n'aura cependant pas la forme de ce document (mise en page stricte et professionnelle). Les fichiers *pdf* créés directement avec *Typora* seront bien des documents *pdf* mais ils ne bénéficieront pas de toute la puissance de mise en page de *LaTeX*²⁵, ils sont en fait transformés en fichiers *pdf* en passant par des fichiers *html*.

16. *@auteur0000* est la clé de citation enregistrée dans le gestionnaire bibliographique

17. <https://typora.io>

18. <https://github.com/marktext/marktext/releases>

19. <https://www.focusfox.co/>

20. <https://dillinger.io/>

21. <https://stackedit.io>

22. <https://pandao.github.io/editor.md/>

23. <https://stylo.huma-num.fr/>

24. <https://alternativeto.net/feature/markdown/>

25. Utilisation des données YAML, numérotation des pages, notes de base de page, équations mathématiques, citations et bibliographie. . .

Pour utiliser *Pandoc* et utiliser toute la puissance de *LaTeX*, la commande doit être entrée en *mode ligne* (“Terminal” sur une machine *Linux* ou *OSX*. Pour *Windows* il faut entrer “cmd” dans la zone de recherche).

Voici la commande :

```
pandoc -t votrefichier.md -o votrefichier.pdf
```

Il n’est pas nécessaire de maîtriser totalement le mode ligne pour effectuer cette manipulation. Pour que la commande fonctionne, il faut néanmoins préciser l’endroit où se trouve le fichier .md (en ajoutant l’arborescence : dir/dir/fichier.md) ou se placer dans le bon répertoire (avec la commande “cd” pour *change directory*).

Il y a d’autres solutions. Par exemple, pour ce document, édité dans *Gedit* (sous *Linux*), l’utilisation des “outils externes” permet d’appeler *Pandoc* sans sortir de l’éditeur.

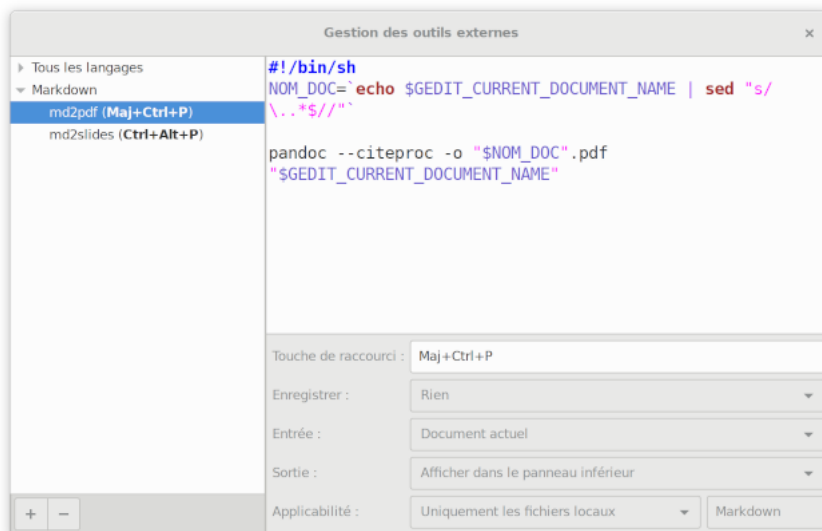


FIGURE 2 – Utilisation des outils externes de *Gedit*

Le site *typademic.ch*²⁶ propose également une interface Web pour transformer un document *markdown* en un document *pdf* (toujours avec l’utilisation de *LaTeX*) sans utilisation du *mode ligne*. Si vous utilisez *Stylo*²⁷, la transformation est également intégrée.

6 Un mot de conclusion

Avec cette méthode, vous allez probablement estimer qu’il y a de nombreuses fonctions proposées par les traitements de texte que vous ne retrouvez pas. C’est bien probable, mais l’objectif est ici d’illustrer une méthode simple, ouverte et durable pour rédiger et produire de beaux documents numériques (Audet et al., 2018).

Bibliographie

Audet, R., Blanc, J., Bourassa, R., Brassard, L.-O., Casenave, J., Hourez, J., Miranda, X., Mellet, M., Monjour, S., Paquin, M.-O., Pons, L., Soubret, J.-L., Sauret, N., & Youssef, E. (2018). *Version 0*.

26. <https://typademic.ch/>

27. <https://stylo.huma-num.fr/>

Notes sur le livre numérique. CRC en écritures numériques. <https://hal.archives-ouvertes.fr/hal-01922682>

Ball, T. (2018). *The Tools I Use To Write Books* [Personal blog]. <https://thorstenball.com/blog/2018/09/04/the-tools-i-use-to-write-books/>

Dehut, J. (2018). En finir avec Word ! Pour une analyse des enjeux relatifs aux traitements de texte et à leur utilisation. *L'Atelier des Savoirs*, 23 janvier 2018, 1-17. <https://eriac.hypotheses.org/80>

Mailund, T. (2019). *Introducing Markdown and Pandoc: using markup language and document converter*. Apress. <https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5852233>

Pierre, A. (2017). *De l'avantage de LaTeX sur les traitements de texte modernes*. <https://aurelienpierre.com/de-lavantage-de-latex-sur-les-traitements-de-texte-modernes/>