

Journal : Plouf, plouf, plouf, c'est toi qui fera mieux de l'epub

Posté par Ysabeau (site Web personnel) le 20/09/22 à 20:27. Licence CC BY-SA.
Étiquettes : epub, livre, électronique, writer, writer2html, calibre, lecture

Salut les gens et les autres,

suite à ce [commentaire](#) dans la rubrique sur la sortie de [LibreOffice 7.4](#), je me suis livrée à des tests (et à la rédaction d'un [tutoriel](#)) avec les trois logiciels dont je dispose pour générer des fichiers epub.

On a, à ma gauche, Writer de LibreOffice, avantage, c'est dedans (le logiciel je veux dire, rien à rajouter, mais, enfin voilà quoi), au milieu [Writer2html](#), une extension de LibreOffice (cvi version beta 1.7, la version sur le dépôt des extensions de LibreOffice non prise avec la 7.4), à ma droite Calibre. Le visualiseur est FBreader.



Oui, celui qui rend le mieux la structure du document, c'est Writer2html.

Un peu de code

Si vous voulez voir un bout de code (enfin trois).

EPUB généré par LibreOffice

```
<body class="body0">
<p class="para1"><span class="span1">Sucré</span></p>
<p class="para2"><span class="span2">Tâte aux pommes</span></p>
<p class="para3"><span class="span3">Pomes bien coupées</span></p>
<p class="para3"><span class="span3">Four brisée jetées</span></p>
<p class="para3"><span class="span3">Au four sont cuites</span></p>
<p class="para2"><span class="span2">Compte de fruits</span></p>
<p class="para3"><span class="span3">Laver, épilucher</span></p>
<p class="para3"><span class="span3">Couper, dans casserole</span></p>
<p class="para3"><span class="span3">Ajouter eau et cuire</span></p>
</body>
```

C'est cracra. Et en plus (c'est pas sur le bout de code), mais il n'a même pas pris le titre du document.

EPUB généré par Writer2html

```
<body dir="ltr">
<p class="heading">Recettes en haiku générées par Writer2html</p>
<h1 id="toc0">Salé</h1>
<h2 id="toc1">Chicons braisés</h2>
<p class="textintro">Braisés les chicons</p>
<p class="textintro">Huile, sel, poêle, lardons</p>
<p class="textintro">C'est tout et c'est bon</p>
<h2 id="toc2">Riz pilaf</h2>
<p class="textintro">Du riz bien rincé</p>
<p class="textintro">Doce, huile et oignon</p>
<p class="textintro">Eau, deux et demi</p>
<h2 id="toc3">Soupe de légumes</h2>
<p class="textintro">Légume, deux tiers</p>
<p class="textintro">Patates un tiers, eau, sel</p>
<p class="textintro">Cuire, crème, mixer</p>
<h2 id="toc4">Frites</h2>
```

Il reprend très fidèlement les styles du fichier.

Et enfin version Calibre :

```
<body class="calibre">
<h1 class="p-pl" id="calibre_p0_0">Salé<span class="calibre"></span></h1>
<h2 class="calibre">Chicons braisés<span class="calibre"></span></h2>
<p class="p-texte_20_intro">Braisés les chicons</p>
<p class="p-texte_20_intro">Huile, sel, poêle, lardons</p>
<p class="p-texte_20_intro">C'est tout et c'est bon</p>
</body>
```

Soit c'est moi (possible), soit c'est Calibre, mais j'ai autant de pages que de titre 2 (h2).

Le fichier martyre et autres considérations

Vous trouverez le [fichier test en téléchargement](#) ainsi que la version pdf. J'ai également copié-collé les recettes dans le [journal](#) qui m'a donné l'idée (pas du tutoriel mais celle de faire des recettes en haikus), il est basé sur ce [modèle de CV](#) aussi pour montrer qu'on peut utiliser ces modèles pour autre chose, et parce que j'avais la femme de faire une mise en forme particulière.

En me livrant à ce petit jeu, j'ai travaillé sur d'autres documents pour voir ce qui se passait. En effet, ces recettes en haiku, c'est un petit fichier de deux pages sans illustration, ni notes. J'ai essayé de passer mon guide [SPIP 4 - vu du côté de la rédaction](#). Calibre, à **chaque** fois n'a pas pu lire le boulot. J'ai un autre document similaire, même résultat. Similaire, vraiment. Ce sont deux documents longs, que j'ai commencé il y a quelques années, avec pas mal d'illustrations et et hors texte, et ça coince quoi. Je pense que le xml derrière doit être bien sale.

Une autre chose que j'ai constaté, c'est qu'il y a vraiment un problème avec les puces. Writer2html ne fait pas mieux que les autres. J'ai systématiquement la puce sur une ligne, le texte associé à la puce sur la ligne du dessous. Peut-être (à vérifier) faudrait-il que les puces ne soient pas suivies de tabulations.

Un genre d'appel à l'aide

Si vous avez sous la main des CSS pour EPUB qui fonctionnent au petit poil, je suis ultrapreneuse, merci d'avance. Si vous n'en avez pas j'essaierai de ne pas trop vous en vouloir (je devrais y arriver, je pense, avec beaucoup de persévérance).

Si vous m'avez lue jusque-là, bravo.

C'est un chanteur qui rentre dans l'EPUB

Posté par Faboulou (site Web personnel) le 20/09/22 à 22:24. Évalué à 10 (+9/-0).

Écoute un peu ton core, ?

Il apprendis les CSS !

Que désire ton WAN ?

Surtout pas de JS !

Dieu a voulu coder, ?

C'est pour ça qu'il t'a fait, ?

Avec des tags pratiques, ?

Ça c'est du sémantique !

Suis moi là !

Tout n'est qu'ordre et beauté, ?

Luxe, calme et W3C, ?

Redonne un sens à ta vie, ?

En y mettant de la **po-é-sie** !

—

Le post ci-dessus est une grosse connerie, ne le lit pas sérieusement.

Pandoc

Posté par Faboulou le 20/09/22 à 22:43. Évalué à 5 (+5/-0).

Un peu hors-sujet, mais par curiosité, j'ai regardé ce que ça donne avec pandoc, et il met des `<blockquote>` de partout. Il y ajoute aussi des `<section>` :

```
<h1 data-number="3">Pâtes</h1>
<section id="pâte-brisée" class="level2" data-number="3.1">
<h2 data-number="3.1.1">Pâte brisée</h2>
<blockquote>
<p>Farine, beurre</p>
</blockquote>
<blockquote>
<p>Mélanger et bien pétrir</p>
</blockquote>
<blockquote>
<p>Eau si nécessaire</p>
</blockquote>
</section>
```

La commande qui va bien:

```
pandoc recettes_en_haiku.odt -t epub -s -o recettes.epub
```

Re: Pandoc

Posté par Eticoumet le 21/09/22 à 06:38. Évalué à 4 (+4/-0).

Bonjour,

J'utilise Sigili pour rédiger mes epub. Alors, pas wysiwyg comme Writer mais super efficace.

C'est dans les dépôts Debian.

Ma chaîne de rédaction est la suivante : écriture du document avec Docbook (là, j'utilise le logiciel propriétaire dument payé Oxygen). Je génère le pdf "quivabien".

Ensuite, j'utilise xslRNG (la nouvelle version des processeurs xslt, dispo sur github) pour générer l'html.

J'importe le html dans Sigili pour faire mes epub.

Pour certains documents, j'utilise pandoc (génération de markdown par exemple)

Bonne journée.

Eric

exemple de document pdf généré (cours de chinois anglais/chinois (pinyin et hanzi) : [FSI Chinese](#); c'est un "work in progress" (il y a plus de 1400 pages à numériser ... et je fais le travail seul)

Re: Pandoc

Posté par Pandoc le 21/09/22 à 08:48. Évalué à 5 (+3/-0).

Il n'est pas évident de publier sous divers formats, web, pdf, etc... Le web et epub sont assez proches, mais le fait de vouloir avec une présentation "adaptée" pour l'impression, en ayant par le PDF, rend la chose compliquée, encore que cela dépend du type de document final. Pour des documents assez formatés / linéaires, comme tes cours, ça peut aller, mais pour un document avec le besoin d'avoir des illustrations ponctuelles, des tableaux, de diverses tailles, ça complexifie.

J'utilise un langage de balisage léger, que ça soit pandoc, docbook ou txt2tags le problème restera le même : comment prendre en compte les cas particuliers ? Pour ma part je m'en accomode, préférant utiliser un langage que je trouve pratique et rapide plutôt qu'un outil graphique comme Libreoffice.

Pour en revenir à la question initiale, en exportant le fichier de recettes en haiku, les titres sont en gris clairs sur blanc, ce n'est pas idéal, mais j'imagine qu'on peut adapter cela dans les options. Effectivement les puces ont une tabulation après leur inclusion, ce qui perturbe l'affichage...

Pour ma part, si j'avais un tel document à produire, je convertirais donc l'odt en txt2tags, avec pandoc<12 : <https://github.com/farvardin/pandoc>

ensuite je transformerais tous les titres niveau 2 en niveau 3 (pas d'outil dédié pour ça, mais ça devrait pouvoir se coder facilement, là j'ai fait ça à la main), pour éviter les sauts de page.

Mon css pour l'epub est ici, mais il n'y a rien d'extraordinaire...

<https://github.com/farvardin/textallion/blob/master/includes/epub.css>

le rendu donne ça en epub :

recettes haiku

Sauces

Mayonnaise

Jaunes d'œufs frais

Vinaigre et puis huile

Bien émulsionner

Vinaigrette

Vinaigre, huile

Moutarde, échalotes

Tout bien mélanger

- test listes
- test 2
- test 3
- test 4
- test 5

94%

et ça en pdf :

recettes haiku

Sauces

Mayonnaise

Jaunes d'œufs frais

Vinaigre et puis huile

Bien émulsionner

Vinaigrette

Vinaigre, huile

Moutarde, échalotes

Tout bien mélanger

- test listes
- test 2
- test 3
- test 4
- test 5

viii

là aussi, rien de spécial, on peut adapter les styles facilement

Re: Pandoc

Posté par Ysabeau (site Web personnel) le 21/09/22 à 10:14. Évalué à 3 (+0/-0).

Pour en revenir à la question initiale, en exportant avec writer2html le fichier de recettes en haiku, les titres sont en gris clairs sur blanc, ce n'est pas l'idéal

En fait ça dépend du visualiseur. C'est impeccable dans FBReader, voir l'illustration du journal et dans ma liseuse, c'est le mieux. Mais, effectivement, selon les visualiseurs, l'apparence peut changer. Il me semble que FBReader est le visualiseur qui a le meilleur rendu des EPUB.

Idéalement j'aurais dû, en effet, revoir la mise en forme du document pour en faire une version pour l'EPUB.

Concernant les titres, alors non le niveau 1 c'est le titre de chapitre, 2 le sous-chapitre, etc. Ce n'est pas une page d'un site web.

Merci pour les css. Pour le pdf, je fais ce que je veux puisque ça n'est « qu'une » photocopie numérique (c'est-à-dire bien plus qu'une photocopie papier). Autant pour la version EPUB, il faut rester dans la simplicité parce que tu ne maîtrises absolument pas le rendu à l'autre bout (police, taille de police, styles, etc.), et c'est très bien comme ça, autant pour la version PDF tu peux te laisser à toutes les fantaisies que tu veux.

—

« Tak ne veut pas qu'on pense à lui, il veut qu'on pense », Terry Pratchett, *Défaillé*.

Re: Pandoc

Posté par Ysabeau (site Web personnel) le 21/09/22 à 10:03. Évalué à 4 (+1/-0).

On a tous et toutes nos outils d'écritures favoris. L'un des énormes intérêts d'un outil comme un traitement de texte, outre que j'aime bien ne voir que ce que j'écris et pas le code qui l'emballe, est qu'il est beaucoup plus facile d'arranger et ré-arranger le document en déplaçant un chapitre, changeant un titre de niveau, etc. Ce que j'ai d'ailleurs fait dans le fichier de test. Et, c'est un outil super efficace pour les gens qui ne sont pas familiers avec le code.

Et je ne suis pas très convaincu (voire, pas du tout) que passer par le pdf soit une bonne chose. Le pdf est tout de même un drôle de format. En tout cas, pour mes documents.

—

« Tak ne veut pas qu'on pense à lui, il veut qu'on pense », Terry Pratchett, *Défaillé*.

Re: Pandoc

Posté par Eticoumet le 21/09/22 à 10:45. Évalué à 1 (+1/-0). Dernière modification le 21/09/22 à 10:49.

re-

Je voulais ajouter que je fais aussi des epub "audio". J'en même fait avec de la vidéo, mais les fichiers deviennent un peu gros.

Voilà un exemple:

[Rudgers Chinese Mandarin copie d'écran sur lecteur koodo](#)

Eric

Re: Pandoc

Posté par PsychoFox le 21/09/22 à 09:58. Évalué à 3 (+0/-0).

Si j'avais eu Pandoc j'aurais aussi fait le test avec. Mais il n'est pas dans ma distribution et il a l'air compliqué à installer (plein de dépendances si j'ai bien compris).

Merci d'avoir fait le test. En définitive, c'est, très clairement, Writer2html qui s'en sort le mieux sur ce document.

—

« Tak ne veut pas qu'on pense à lui, il veut qu'on pense », Terry Pratchett, *Défaillé*.

Re: Pandoc

Posté par PsychoFox le 21/09/22 à 13:37. Évalué à 3 (+0/-0).

Je suis étonné d'apprendre que tu n'ales pas pandoc dans les repos de ta distro. Quelle distro exotique utilises-tu?

Si non tu peux utiliser podman, docker ou [distrobox](#) pour l'utiliser facilement depuis n'importe quelle distros. C'est un peu le chroot des temps modernes.

<https://hub.docker.com/r/pandoc/core>

Re: Pandoc

Posté par Gill_Cgt le 21/09/22 à 13:50. Évalué à 2 (+0/-0).

Mouais, je trouve pas ça convainquant de devoir sortir l'artillerie lourde pour un usage normal et non technique de son poste de travail. Non, rien n'égale la disponibilité dans les dépôt, et au pire que le paquet sur sa distro soit dispo <https://pkgs.org/download/pandoc-common>

De mémoire, sa distribution est la Mageia.

—

"It is seldom that liberty of any kind is lost all at once." — David Hume

Re: Pandoc

Posté par PsychoFox le 21/09/22 à 13:57. Évalué à 3 (+0/-0). Dernière modification le 21/09/22 à 14:00.

Je suis d'accord que rien n'égale la dispo dans les repos de la distros mais apparemment ce n'est pas le cas. C'est pas forcément de l'artillerie lourde. Au final ça se résume à un alias et une image pas forcément plus grosse qu'un flatpak type qui se télécharge automatiquement.

Dans ton .bashrc:

```
alias pandoc='podman run --rm -v "$pwd:/data" -u $(id -u) $(id -g) pandoc/core'
```

(EDIT: Quelqu'un peut m'expliquer pourquoi le ping dollar est interprété de façon hasardeuse et écrit ce cryptique (mathjax)?

Et après l'utilises la commande plus ou moins de la même manière. Il faut juste se rappeler de n'utiliser que des chemins relatifs ascendants.

Et ça va moins foutre le bronx qu'un potentiel rpm pas fait pour ta propre distrib.

Re: Pandoc

Posté par PsychoFox le 21/09/22 à 14:03. Évalué à 3 (+0/-0). Dernière modification le 21/09/22 à 14:03.

[Lien vers l'alias correct](#) à mettre dans son .bashrc à cause d'un bug de linuxfr:

[alias pour podman](#)

Re: Pandoc

Posté par Ysabeau (site Web personnel) le 21/09/22 à 14:26. Évalué à 3 (+0/-0).

Pour le S, c'est euh, normalement anormal. C'est un bogue connu. Je me demande si en faisant précéder ce signe de " dans le bout de code ne serait pas une solution que j'ai peut-être vue quelque part.

J'essaie :

```
alias pandoc='podman run --rm -v "$pwd:/data" -u $(id -u) $(id -g) pandoc/core'
```

Bon essai pas concluant, mais, on est prévenu qu'il faut virer le truc devant ça devrait être bon. J'ai essayé d'autres graphies. En fait avec n'importe quoi devant le \$ ça fonctionne. Il existe apparemment aussi une image "minimale" qui est sans doute un peu plus petite et peut-être suffisante juste pour faire de l'epub.

—

« Tak ne veut pas qu'on pense à lui, il veut qu'on pense », Terry Pratchett, *Défaillé*.

Re: Pandoc

Posté par Gill_Cgt le 21/09/22 à 14:48. Évalué à 3 (+0/-0).

Oui pandoc est un outil en ligne de commande de format en ligne de commande.

J'ai vérifié aussi l'image docker, c'est tout de même un peu plus gros qu'un flatpak 380MB. Pas vraiment *green computing* donc mais pour tester rapidement c'est pratique. Il existe apparemment aussi une image "minimale" qui est sans doute un peu plus petite et peut-être suffisante juste pour faire de l'epub.

Re: Pandoc

Posté par Gill_Cgt le 21/09/22 à 19:23. Évalué à 3 (+1/-0). Dernière modification le 21/09/22 à 19:23.

Autant pour nous autres (sys/dev/geek/nerd/etc.) dont c'est le métier ça fait sens, mais si pour un usage standard tu commences à exiger « podman, docker » je trouve que c'est sortir l'artillerie lourde. C'était cela que je pointais.

—

"It is seldom that liberty of any kind is lost all at once." — David Hume

Note : les commentaires appartiennent à celles et ceux qui les ont postés. Nous n'en sommes pas responsables.