

# Langages de balisage légers et logiciels de conversion de documents

Nicolas Poulain

28 mars 2012

## Table des matières

<b>1</b>	<b>Présentation</b>	<b>1</b>
<b>2</b>	<b>Les langages de balisage légers et les wikis</b>	<b>2</b>
<b>3</b>	<b>Les langages de balisage légers et la bureautique</b>	<b>3</b>

## 1 Présentation

Pour saisir et mettre en forme des textes ou des documents textuels comportant des insertions d'images, de figures ou de tableaux, on utilise généralement un traitement de texte WYSIWYG<sup>1</sup>, propriétaire comme Microsoft Word ou libre comme OpenOffice.

Les défauts majeurs de ces logiciels sont nombreux :

1. Le rédacteur d'un document se concentre presque autant sur le fond que sur la forme. Outre le temps passé, les conséquences sur le rendu sont nombreuses
  - Les mises en forme les plus hétéroclites sont autorisées au dépens de la lisibilité ;
  - Le résultat final est souvent discutable du point de vue de la typographie car les règles n'en sont pas respectées ni par l'utilisateur ni par le logiciel ;
  - L'utilisation des styles est souvent anarchique et les documents mal structurés, ce qui rend la production automatique de sommaire ou d'index impossible ;
  - L'insertion d'images ou de figures provoque des décalages mal maîtrisés.
2. En ce qui concerne les documents longs, l'inclusion de documents annexes au sein du document maître donne des résultats aléatoires ;
3. L'interopérabilité n'est pas assurée entre les logiciels, elle ne l'est pas même entre les différentes versions d'un même logiciel, ce qui nous amène au dernier point ;
4. La pérennité des documents n'est pas certaine puisque la compatibilité ascendante ne fonctionne pas toujours et qu'un document écrit il y a quelques années risque d'être perdu, faute du logiciel capable de le lire.

À l'opposé de la composition dans un logiciel de traitement de texte, on peut écrire des documents dans des langages de balisage. Il en existe de nombreux : LaTeX, HTML, DocBook, etc. Les fichiers sont enregistrés au format texte brut et doivent être interprétés par un logiciel afin d'être consultés.

En ce qui concerne HTML et LaTeX, où pratiquement toutes les mises en formes sont possibles, le problème vient de la difficulté à écrire les balises<sup>2</sup>. Pour écrire un titre suivi d'une phrase contenant un mot en gras puis une liste non numérotée, on saisira respectivement :

– en LaTeX

---

1. Un WYSIWYG pour *What you see is what you get* est une interface utilisateur qui permet de composer visuellement le résultat voulu. C'est une interface intuitive : l'utilisateur voit directement à l'écran à quoi ressemblera le résultat final.

2. Dans le cas du format DocBook, c'est même humainement presque impossible de l'écrire à la main tant l'enchevêtrement des balises est inextricable. On le génère avec un logiciel WYSIWYG...

```
\section{Le titre du paragraphe}
```

Voici un mot en `\textbf{gras}` puis une liste :

```
\begin{enumerate}
  \item c'est simple ;
  \item c'est efficace.
\end{enumerate}
```

– en HTML

```
<h1>Le titre du paragraphe</h1>
```

```
<p>Voici un mot en <strong>gras</strong> puis une liste :</p>
```

```
<ul>
  <li> c'est simple ;</li>
  <li> c'est efficace.</li>
</ul>
```

Comme on le voit, la syntaxe est accessible mais au goût de nombreux utilisateurs il y a trop de commandes de mise en forme qui nuisent à la lisibilité du texte lors de la saisie. C'est dommage car ces deux formats ouverts et universels ont chacun leur avantage :

- HTML peut être lu sur n'importe quelle plateforme ou terminal du monde entier car ses spécifications, gérées le W3C<sup>3</sup>, sont respectées par les navigateurs web.
- le logiciel LaTeX produit des documents de qualité unanimement reconnue. Il prend en charge la mise en page, l'utilisateur n'ayant qu'à se concentrer sur le fond et sa structure.

Il existe une alternative qui est à la fois simple, interopérable et efficace : les langages de balisage légers.

## 2 Les langages de balisage légers et les wikis

Un langage de balisage léger est un langage utilisant une syntaxe simple, conçue pour qu'un fichier en ce langage soit aisé à saisir avec un éditeur de texte simple, et facile à lire dans sa forme non formatée.

Les wikis ont grandement contribué à populariser ce type de langage. Le principe est de saisir des balises accessibles aux non initiés, un moteur se chargeant de la conversion en HTML avant la publication.

```
Le titre du paragraphe
=====
```

Voici un mot en **\*\*gras\*\*** puis une liste :

```
* c'est simple ;
* c'est efficace.
```

Avantages :

- les balises sont visuelles et le texte reste lisible ;
- le nombre de balises et de règles à mémoriser est peu important ;
- les balises étant constituées de caractères non alphabétiques, on peut utiliser un correcteur d'orthographe.

Il existe de nombreux langages de balisage légers : Creole, Markdown, AsciiDoc, etc. Chacun a ses avantages, mais tous sont simples.

Dans la section suivante, nous allons voir qu'il est possible de faire de la bureautique avec ces langages et nous verrons lequel choisir en fonction de l'usage qu'on souhaite en faire.

---

3. Un WYSIWYG pour *What you see is what you get* est une interface utilisateur qui permet de composer visuellement le résultat voulu. C'est une interface intuitive : l'utilisateur voit directement à l'écran à quoi ressemblera le résultat final.

### 3 Les langages de balisage légers et la bureautique

On vient de voir qu'au sein des wikis, les langages de balisage légers sont transformés en HTML.

C'est maintenant que les choses deviennent intéressantes : il existe des logiciels permettant d'exporter et de mettre en forme vers différents formats pour différents usages : la diffusion web, bien sûr mais aussi l'export pour un traitement de texte, l'impression, la lecture sur tablette ou liseuse d'e-book ou encore la vidéo-projection.

Ces logiciels de conversion sont nombreux, en voici quatre avec leurs principaux formats d'import et d'export.

Logiciel	Import	Export web	Export Bureautiq	Export ue TeX	Export LBL
Txt2tags	T2t	HTML, XHTML, SGML,	DocBook, Lout, MagicPoint, PageMaker	LaTeX	Creole, AsciiDoc, PmWiki, MoinMoin, AsciiDoc, DokuWiki
MMD	Markdown	HTML, XHTML	OpenDocument	LaTeX	
Pandoc	Markdown, LaTeX, HTML, Textile, RST	HTML, XHTML, HTML5, EPUB, Slidy,S5, DZSlides	OpenDocument, ODT, DOCX, DocBook	LaTeX, ConTeXt, Beamer	Markdown, RST, AsciiDoc, Textile, MediaWiki
AsciiDoc	asciidoc	HTML, XHTML	Docbook	LaTeX	

Fonctionnalités	Txt2tags	MMD	Pandoc	Asciidoc
en-tête (titre, auteur, date)	x	x	x	x
sections (numérotées ou non)	x	x	x	x
paragraphes	x	x	x	x
listes à puces, numérotées et de définition	x	x	x	x
texte en gras, italique, souligné, barré	x	x	x	x
couleurs et tailles de texte				x
police à espacement constant	x	x	x	x
coloration syntaxique de code source			x	x
gestion des liens (internet, courriel, etc.)	x	x	x	x
Références internes		x	x	x
images	x	x	x	x
tableaux (gestion de bordure et d'alignement)	x	x	x	x
tableaux (fusion de cellules)				x
Légendes (images et tableaux)		x	x	x
Citations		x	x	x
Notes de bas de page		x	x	x
formules mathématiques (LaTeX)	x	x	x	x