par Guido Socher (homepage)

L´auteur:

Il y a quelques temps, nous avions dit que LinuxFocus voulait rendre disponible les articles au format PDF. Nous avons reçu un certain nombre de suggestions en réponse, que nous résumons ici, dans ce truc. Merci beaucoup pour toutes les suggestions.

Traduit en Français par:
Jean–Etienne Poirrier (homepage)

Truc LF: Générer des PDF depuis des documents html

Résumé:

C'est une petite astuce. A partir de maintenant, LinuxFocus aura au moins une nouvelle astuce tous les mois. Si vous avez des idées pour une nouvelle astuce, envoyez–les à guido(le signe « à »)linuxfocus.org

Introduction

Vous avez probablement remarqué que nous avons maintenant des fichiers PDF pour tous les articles dont la langue utilise l'ensemble des caractères iso8859–1. Cela n'a pas été facile à implémenter car nous voulions qu'ils soient générés automatiquement afin d'éviter que les documents texte/html et PDF diffèrent.

Voici notre expérience avec une liste d'options sur la manière de générer un PDF en général.

L'idée

Tous les systèmes Linux possèdent l'utilitaire Ghostscript ps2pdf. ps2pdf fonctionne très bien et la qualité des PDF générés est bonne. En d'autres mots, nous pouvons toujours générer les fichier PDF si nous gérons le document comme un fichier postscript.

Le système d'impression de Linux est basé sur postscript ; ainsi, cela devrait être simple !? Le problème est réellement de trouver une manière de le réaliser avec un script en ligne de commande. Vous ne souhaitez pas cliquer avec la souris lorsque vous avez besoin d'imprimer quelques centaines d'articles.

Si vous n'êtes pas concernés par les tables, les couleurs et les images, alors une combinaison de « lynx –dump

.... | nenscript » et ps2pdf fonctionnera. Si, par contre, vous avez besoin des tables et des images, continuez votre lecture.

Les candidats

html2ps

C'est un script Perl et la version testée ici était html2ps 1.0 beta3. La page d'accueil est http://user.it.uu.se/~jan/html2ps.html
Le programme fonctionne assez bien. Il requiert cependant beaucoup de modules Perl comme dépendances et il a des problèmes avec les tables des pages pour les structurer. C'est une bonne solution si vous avez une disposition graphique très simple.

LaTeX

Il y a un convertisseur de LaTeX vers PDF. En utilisant XSLT, vous pouvez transformer du HTML en LaTeX. Un pré−requis pour cela est d'avoir un fichier HTML syntaxiquement correct. Cela peut être réalisé avec l'utilitaire Tidy :

```
HTML --(tidy)--> XHTML --(XSLT)--> Latex --(pdflatex)--> PDF
```

Je n'ai pas investigué plus loin dans cette voie parce que je trouve XSLT et LaTeX trop lourds et complexes.

Télécommande de navigateur web

Si, d'une manière ou d'une autre, il était possible de commander à distance un navigateur web, alors nous aurions l'avantage d'un fichier PDF généré, identique à ce que vous voyez normalement dans votre navigateur web. Le problème est qu'il faut un affichage X11. Il n'est donc pas possible de le faire exécuter par un job cron.

Le projet Mozilla a amélioré l'impression et le rendu. Cela a cependant enlevé quelques possibilités de contrôle à distance que Netscape Communicator avait. La solution suivante ne fonctionnera donc qu'avec Communicator 4.x :

```
netscape −noraise −remote "openurl(http://unepage)"
sleep(10) # il n'y a pas moyen de savoir si une page est complètement chargée
          # donc, nous attendons un peu
netscape −noraise −remote saveas(unepage.ps,PostScript)
sleep(10)
ps2pdf unepage.ps
```

Quelques lecteurs m'ont dit qu'ils pensaient que la télécommande d'impression devrait être également possible avec Konqueror mais personne n'a pu me fournir de solution fonctionnelle.

htmldoc

Htmldoc est un utilitaire très bien écrit de http://www.htmldoc.org/. La commande suivante va faire exactement ce que nous voulons :

```
htmldoc −t pdf −webpage −f fichier.pdf fichier.html
```

Nous avons utilisé la version 1.8.24 et elle fonctionnait parfaitement. Le seul problème est que les fichiers PDF résultants sont, en moyenne, 10 fois plus gros que n'importe quel autre fichier PDF généré par les autres

solutions, peu importe l'option de compression que vous utilisez dans htmldoc. C'est un gros problème si vous avez des centaines de documents.

Conclusion

Nous utilisons maintenant une combinaison de télécommande Netscape et Htmldoc. Nous ne pouvons pas compter uniquement sur Htmldoc, vu la taille des fichiers générés. Si vous avez d'autres suggestions ou des idées à ce sujet, écrivez–nous !

2005–04–28, generated by lfparser_pdf version 2.51