

Wednesday, 11 July 2007

## Convertir un fichier HTML en document Word

Récemment j'ai eu la tâche de convertir des fichiers HTML en des fichiers Microsoft Word (.doc) - et idéalement par ligne de commande: je voulais l'intégrer dans une application PHP que j'utilise localement pour générer du contenu HTML, et proposer automatiquement à mes clients ces mêmes fichiers en .doc.

J'ai fait quelques recherches et testé quelques logiciels spécialisés, mais même si certains étaient bien confortables aucun ne générait un résultat acceptable. En fait, le seul qui sache "bien" le faire est Word lui-même... La solution était donc simple: ouvrir un fichier HTML avec Word, et le sauvegarder sous un autre nom. Sauf que il fallait pouvoir le faire de manière automatisée. Quelques recherches plus tard, j'avais trouvé une liste des paramètres de ligne de commande de Word - dont un permet d'exécuter une macro au démarrage.

J'ai enregistré une macro très simple qui sauvegarde le fichier en cours et quitte Word. Ainsi, sur la ligne de commande je pouvais déjà ouvrir un fichier HTML, et le sauvegarder en doc - sauf que il l'enregistrait toujours au même endroit et le même dossier.

En visual basic on a accès au nom et chemin du document ouvert, donc j'ai utilisé ces propriétés pour sélectionner automatiquement le dossier cible et le nom de fichier. Voici ce que ça donne:

```
Sub SaveAsDoc()  
,  
' SaveAsDoc Macro  
' Saves the active file and exits the application.  
' Used to convert HTML to doc files.  
,  
    ChangeFileOpenDirectory ActiveDocument.Path  
    ActiveDocument.SaveAs FileName:=ActiveDocument.Name + ".doc", FileFormat _  
        :=wdFormatDocument, LockComments:=False, Password:="", AddToRecentFiles:= _  
        True, WritePassword:="", ReadOnlyRecommended:=False, EmbedTrueTypeFonts:= _  
        False, SaveNativePictureFormat:=False, SaveFormsData:=False, _  
        SaveAsAOCELetter:=False  
    Application.Quit  
End Sub
```

Après avoir ajouté la macro, on peut la lancer sur la ligne de commande avec la commande suivante:

```
winword.exe /mSaveAsDoc document.html
```

Cette commande ouvre Word, charge le fichier spécifié et lance la macro "SaveAsDoc". Le fichier est directement sauvegardé dans le même dossier que le fichier ouvert, avec une extension .doc ajoutée au nom original. On retrouve donc après un fichier nommé

```
document.html.doc
```

Assez confortable, si on peut supporter l'ouverture/fermeture de Word pour chaque fichier. Je n'ai pas testé en laissant Word ouvert par contre. À partir de PHP, le seul hic est que je ne peux pas lancer d'applications avec un exec() - mais là encore il y a une solution assez simple: je crée automatiquement un fichier de ligne de commande contenant toutes les commandes de conversion, et sur lequel il ne reste plus qu'à cliquer.

Gros avantages de cette solution: compatibilité parfaite + gratuité. Si quelqu'un a une meilleure solution ou un logiciel à proposer, dites-le moi.

Posté par Argh dans Web à 13:56