

Remplacement automatisé de caractères en Emacs

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 18 septembre 2008. Dernière mise à jour le 19 septembre 2008

<https://www.bortzmeyer.org/emacs-replacement-automatise.html>

Un petit exercice de programmation en Emacs Lisp. Comment remplacer facilement un ensemble de caractères par un autre ?

Beaucoup de pages Web aujourd'hui utilisent tout le jeu de caractères Unicode et elles ont bien raison, le Web étant certainement un des services d'Internet le mieux « unicodisé ». Ainsi, on trouvera sur une page Web comme <<http://morris.blogs.nytimes.com/2008/08/11/photography-as-a-weapon/>>, pourtant en anglais, des caractères Unicode comme les jolis guillemets ("[Caractère Unicode non montré]¹ *this information that you heard? It's wrong.*[Caractère Unicode non montré]", utilisant les caractères U+0201c et U+0201d. Mais ces caractères ne sont pas présents dans ASCII ni même dans Latin-1. Comme je ne suis pas encore pas encore passé à UTF-8 <<https://www.bortzmeyer.org/pas-encore-utf8.html>>, ces caractères me gênent lorsque je veux copier-coller une partie d'une page Web dans, par exemple, un courrier. Emacs accepte le texte mais, à l'enregistrement, proteste que :

```
These default coding systems were tried to encode text
in the buffer 'toto':
(iso-latin-1 (15 . 342396) (51 . 342393) (60 . 342397) (185
. 342393))
However, each of them encountered characters it couldn't encode:
iso-latin-1 cannot encode these: \ "
```

Comment les convertir sans procéder à un rechercher/remplacer par caractère (car il n'y a pas que les guillemets) ? Une des forces d'Emacs est d'être programmable, dans son langage dédié, un dialecte de Lisp. On peut donc écrire :

1. Car trop difficile à faire afficher par L^AT_EX

```
(defun to-latin1 ()
  (interactive "*")
  (goto-char (point-min))
  (replace-string "something" "other thing")
  (goto-char (point-min))
  (replace-string "some other thing" "yet another thing")
  ...
)
```

On peut alors appeler cette fonction (avec `M-x to-latin1` ou bien en l'attachant à une touche) et la conversion est faite.

Bon, il reste à mettre les caractères à changer. Le moyen le plus simple est de convertir le code en chaîne. La syntaxe Emacs Lisp pour cela est :

```
"\x53979"
```

Comment ai-je trouvé le chiffre (hexadécimal) 53979? Emacs n'utilise pas les points de code de la norme Unicode, malheureusement. Ce code 53979 est donc spécifique à Emacs. Pour le trouver, le plus simple est donc de mettre le curseur sur le caractère problématique et de faire `C-x =`. Le code est alors affiché en bas dans la mini-fenêtre.

Voici donc le résultat final :

```
(defun to-latin1 ()
  (interactive "*")
  (goto-char (point-min))
  (replace-string "\x53979" "'")
  (goto-char (point-min))
  (replace-string "\x5397c" "\"")
  (goto-char (point-min))
  (replace-string "\x5397d" "\"")
)
```

Merci à Christopher J. Madsen (cjm <<http://www.cjmweb.net/>>) pour son aide sur la syntaxe.