Misos : ISO 8859−1, MS−DOS et Emacs

L'Echo de Linux (*Juillet 1996*)
Georges KO (kott@ccms.ntu.edu.tw)

Résumé

Cet article décrit l'installation, la configuration ainsi que l'utilisation du package Misos pour Emacs, permettant de gérer « au mieux » les caractères ISO dans un environnement hyper−hostile : MS−DOS et le clavier QWERTY. La configuration minimale de Misos ne requiert pas de connaissances en Elisp.

Sommaire

1. Introduction
    1. Problèmes
    2. Emacs ou la philosophie Unix en action
    3. Misos
2. Installation
3. Configuration
4. Utilisation
    1. Les modes
    2. Changement de mode
    3. Saisie
    4. Utilisation avec d'autres packages ou des programmes externes
5. À faire / À améliorer
6. Conclusion
7. *Credits*

Introduction

Problèmes

ISO 8859−1 (*ISO*) est le jeu de caractères latins standard sur Internet, aussi bien pour le courrier, les news, IRC (?), ainsi que le *World Wide Web*. Malheureusement, *MS−DOS*, utilise un jeu de caractères qui lui est propre, le plus souvent *PC Code Page 437* (*CP 437*), ou encore *PC Code Page 850* (*CP 850*), plus adapté au français. Il en résulte que l'utilisation de programmes tels que *NCSA Telnet*, *Telix* et autres programmes de communication pour MS−DOS avec Internet se révèle pénible, car les « à » se transforment en « alpha », « ç » en « tau », les majuscules accentuées en éléments graphiques, etc., sans parler de leur saisie avec un clavier *QWERTY*.

Pour gérer ces problèmes, plusieurs solutions possibles :

- Utiliser des filtres : pas très pratique car nécessitant des sauvegardes, manipulations à la main, pour−IRC−comment−je−fais, etc. ;
- Utiliser des tables de traduction et des macros pour les caractères accentués dans le programme de communication : pourquoi pas, mais si la machine est utilisée depuis plusieurs endroits distincts, pas très pratique non plus... ;
- Prendre les sources de telnet ou d'un programme de communication et bidouiller avec : j'ai que ça à faire... ;
- Prendre les sources de Lynx, d'Elm, de Pine, etc. et bidouiller avec : je commence demain... ;
- etc.

Bref, ce n'est pas simple à gérer... Il existe bien sûr d'autres solutions plus radicales : installer Windows (sur des 286 avec 2 Mo de RAM), installer Linux (dans le département de psycho **−NDLR: beuh, qu'est ce qu'il dit lui ? −**), etc.

Emacs ou la philosophie Unix en action

Emacs est un éditeur permettant, entre autre, de lire le courrier, lire les news et naviguer sur le Web. Si on arrivait à faire afficher avec Emacs les caractères ISO correctement sur un terminal MS−DOS, si on arrivait à saisir ces mêmes caractères avec le clavier le plus primaire qui soit, le clavier QWERTY américain, non seulement il sera possible d'éditer du texte normal à partir de ce terminal, mais en plus, ces fonctionnalités pourront être très facilement intégrées avec le courrier, les news et le Web (avec un minumum de connaissances en Elisp et des *packages* concernés). Tiens, *Misos* fait ça...

Misos

Misos (« ISO dans MS ») est un package Emacs permettant de (liste non−exhaustive) :

- Afficher et saisir des caractères ISO sur un terminal MS−DOS (exemples : telnet depuis un terminal MS−DOS sur une machine UNIX ; utilisation d'Emacs pour DOS pour éditer des fichiers avec caractères ISO) ;
- Afficher et saisir des caractères MS−DOS sur un terminal MS−DOS (exemples : éditer un fichier MS−DOS depuis un terminal MS−DOS avec telnet sur une machine UNIX ; utilisation d'Emacs pour DOS pour éditer des fichiers avec caractères MS−DOS) ;
- Afficher et saisir des caractères MS−DOS sur un terminal ISO (exemples : éditer un fichier avec caractères MS−DOS sous Linux (avec un jeu ISO)) ; les éléments graphiques (pour dessiner des cadres, tableaux, etc.) sont représentés par des « | », « / », « \ », « + », etc. ;
- Saisir les caractères de manière consistante ;
- Avec GNUS (lecteur de news et de courrier), W3 (pour le Web), etc., sélection automatique du *mode* à utiliser, selon le nom du newsgroup, l'expéditeur du courrier, l'URL, etc. ;
- Mise en majuscule/minuscule correcte selon le mode ;
- etc.

Installation

Étant donné que Misos a été écrit avec Emacs 19.31, il n'est **pour l'instant** utilisable qu'avec Emacs 19.30 et plus, des modifications de structures de données ayant eu lieu lors du passage à la version 19.30.

L'installation du package est simple :

1. Détarer/décompresser misos.tar.gz dans un répertoire connu de `load−path`. Si tu n'es pas *root*, ajouter dans ton fichier `$HOME/.emacs` : `(setq load−path (append load−path '("~/`*misos_rep*`")))`, où *misos_rep* est le répertoire où le package a été détaré/décompressé ;
2. Ajouter dans `$HOME/.emacs` : `(load "misos")` ;
3. Si t'utilises un terminal MS−DOS, placer la variable d'environnement

`MISOSTERM` à `ms` avant de lancer Emacs ;
4. C'est tout en principe... S'il y a un problème, m' écrire.

Configuration

La configuration minimale se fait grâce aux variables suivantes :

- `ms850iso` : si elle vaut `t`, les caractères MS−DOS sont interprétés avec CP 850 ; si elle vaut `nil`, avec CP 437. Défaut : `nil` ;
- `isoms850-one2one` : si `t`, un caractère ne peut−être représenté que par un seul caractère (« È » sera représenté par « E » , « © » par « c », etc.) ; si `nil`, un caractère pourra être représenté par plusieurs caractères (« È » par « ‘E » ou « E‘ » , « © » par « (c) », etc.) : c'est plus précis, mais ça allonge les lignes : dommage pour la présentation... Ceci est pour utiliser l'ISO sur terminal MS−DOS. Défaut : `nil` ;
- `ms850iso-one2one` : pareil que le précédent, sauf que c'est pour les caractères MS−DOS sur terminal ISO.
- `prefix-accent` : avec **−one2one* à `nil`, si `prefix-accent` vaut `t`, alors les accents sont préfixés (« È » donne « ‘E ») ; si à `nil`, ils sont suffixés (« È » donne « E‘ »). Défault : `nil`.

Une variable se fixe avec la fonction setq. Pour modifier une de ces variables, ajouter après le chargement de Misos : `(setq` *variable valeur*`)`, où *valeur* est soit `t`, soit `nil`.

Si aucune de ces variables n'est fixée dans le fichier `$HOME/.emacs`, la phrase (ISO) « LES ¾ DES INTERNÉS ACHÈTERONT DU MAÏS » sera affiché « LES 3/4 DES INTERNÉS ACHE‘TERONT DU MAI"S » sur un terminal MS−DOS.

Utilisation

Les modes

Une fois Emacs lancé, si tout se passe bien :−), Emacs a 5 nouveaux *modes mineurs* utilisables :

isoms−mode
: C'est le mode « utilisation d'ISO / terminal MS−DOS » ;

msms−mode
C’est le mode « utilisation de MS−DOS / terminal MS−DOS » ;

msiso−mode
C’est le mode « utilisation de MS−DOS / terminal ISO » ;

big5−mode
C’est le mode *Big5* pour le chinois. Undocumented :−) ;

raw−mode
C’est le mode par défaut, qui est contrôlé par la variable d’environnement `MISOSTERM`. Si elle est à `ms`, ce mode équivaut à *isoms−mode*, sinon il équivaut au mode « standard ». Cette dénomination *raw−mode* est trompeuse, car elle incite à croire qu’il n’y a pas de modification, alors qu’il y en a, vraiment. À changer (sans doute en *current−terminal−mode*).

La touche `F2` permet de passer d’un mode à l’autre. Si ça ne marche pas, `Esc x misos-switch`. Pour ne pas avoir tous les modes, copier `misos-switch` (dans *misos.el*) dans `$HOME/.emacs` après le `load` pour faire sa propre version. Pareil pour changer la touche associée (`global-set-key`...). Le mode utilisé devrait apparaître dans la *modeline* et dans l’*echo area* (ce dernier est utile lorsque la modeline n’affiche pas les modes mineurs, comme avec W3. Enfin, n’affiche pas *par défaut*).

Changement de mode

Ce qui se passe lorsqu’on change de mode :

- La table d’affichage (*display−table*) du *buffer* en question est modifiée, donc l’affichage des caractères accentués, etc. devrait se modifier aussi ;
- La table de syntaxe (*syntax−table*) du buffer est modifiée (Jérémie est un mot, mais le codage en ISO et en MS−DOS est différent) ;
- La table des majuscules/minuscules (*case−table*) du buffer est modifiée (la mise en majuscule de « fête » donnera « FÊTE » en ISO et CP 850, mais « FêTE » en CP 437 : si c’est transformé en « E », comment je fais, pour savoir que c’est un « ê » si je le minusculise plus tard ?) ;
- La table de saisie des caractères accentués, etc., est modifiée.

Saisie

Ce qui est important de saisir, c’est que lorsque le mode *XY* est utilisé, *X* réfère au codage utilisé et *Y* le terminal utilisé. Exemple : isoms : codage ISO sur terminal MS. Si *X* et *Y* ne collent pas au codage et terminal utilisés, l'affichage sera incorrect.

À partir de là, c'est simple, utiliser « `Control x 8` *touche(s)* » pour saisir un caractère, quel que soit le mode. Exemple : « `C-x 8 ' e` » pour « é », `C-x 8 C` pour un « © », etc. Je dis « `C-x 8` », mais ça pourrait être n'importe quoi. Idem pour « `' e` ». Pour avoir toute la liste des touches, regarder les fichiers *m−iso−map.el* et *m−ms−map.el* ou alors, dans un des modes, taper « `F1 m` » et regarder le buffer `*Help*` résultant. C'est quelquepart dedans.

Les caractères représentés par plusieurs caractères (exemple : © représenté par (c) en *isoms−mode* en CP 437), sont considérés par l'éditeur comme un seul caractère (tu vois trois caractères, un « (», un « c » et un «) », mais tu ne peux mettre le curseur qu'à deux endroits : sur le « (» et apres le «) » en appuyant sur flèche gauche. En fait, t'es déjà sur le caractère suivant le « © »). Ça va :−) ?

Utilisation avec d'autres packages ou des programmes externes

Exemples dans *m−hooks.el* : là, faut connaître Elisp... Disponible dans tout bon site miroir FTP GNU.

Le principe, c'est d'utiliser les *hooks*. Qu'est−ce qu'un hook ? Un hook, c'est une variable contenant une liste de fonctions à exécuter à des moments bien précis. Tous les packages bien faits d'Emacs (ainsi qu'Emacs lui−même) sont bourrés de hooks.

Par exemple, lorsque GNUS, le lecteur de news, a fini de préparer un article (il est prêt à l'afficher), il exécute toutes les fonctions du hook `gnus-article-prepare-hook`. Vu que je lis des newsgroups en chinois (Big5) et en français (ISO), il suffit donc de créer une fonction qui vérifie le nom du newsgroup (en *tw.**, c'est en Big5, le reste en *isoms−mode* ou équivalent (*raw−mode*)), et de l'ajouter à ce hook. Pareil lors d'un *folluwup* ou d'un *Reply*, avec les hooks `gnus-article-prepare-hook` et `gnus-mail-hook`. Pareil pour le courrier, pareil pour le Web (enfin, en principe, car les hooks en question ne semblent pas être exécutés dans W3), etc.

Pour utiliser Misos avec des programmes externes (Elm, Tin, Pine, Slrn, etc.), il suffit de les faire lancer Emacs avec les bons arguments (`-f isoms-mode` par exemple). Cependant, je ne vois pas l'intérêt d'utiliser ces programmes dans ce contexte, vu qu'Elm, Tin, Pine, Slrn etc. ne gèrent pas ces problèmes d'ISO et MS−DOS (si ?). `(emacs-advocacy (concat "Cela dit, même sans ces problèmes, je ne voyais déjà pas l'intérêt d'utiliser ces programmes" (smile) "."))`

A faire / à améliorer

- Améliorer la doc, qui est cet article (ben ouais, avec Elisp, la meilleure doc, c'est les sources...) ;
- Meilleure configurabilité (?) (utiliser une liste pour construire `misos-switch`, ne changer que le clavier de saisie, ou juste l'affichage, des choses comme ça) ;
- Gestion des claviers AZERTY et autres (c'est bête de devoir taper « `C-x 8 ' e` » lorsqu'on a déjà la touche... Pas compliqué à faire, mais je n'ai pas de clavier AZERTY :−)...) ;
- Ne pas être forcé à tout modifier lors d'un changement de mode (syntax−table, case−table, etc.) ;
- Utilisation avec Emacs < 19.30 : les modifications portent sur les tables de syntaxe, de majuscules/minuscules, etc. ;
- Faciliter une éventuelle extension à ISO 8859−2, 3, etc. en utilisant des listes au lieu de tout coder en dur ;
- Compléter la saisie des caractères MS−DOS (lettres grecques, éléments graphiques, etc.) : pas très important..;
- Euh...

Conclusion

Misos est le package qui me permet de nager dans un environnement totalement « Engish−only », « QWERTY−only », et « (presque) MS−DOS−only ». Parcequ'il est écrit en Elisp, je peux l'utiliser sur n'importe quel système d'exploitation, ainsi que sous MS−DOS. Parceque c'est pour Emacs, ça s'intègre donc bien (aux effets de bord près !) avec les autres packages et résoud donc en même temps les problèmes avec le courrier, les news, le Web, etc. Bon, c'est sûr que ce n'est pas spectaculaire à voir : pas de menus, pas d'utilitaires pour configurer, etc., c'est sobre et discret, mais fonctionnel et satisfaisant. Si t'en as pas besoin, sois quand même conscient(e) que ça existe...

Suggestions, commentaires, améliorations, critiques, questions, etc. bienvenus, en

ISO ou en MS−DOS si t'arrives pas à installer / configurer / utiliser :−) !

Credits

- Stephen G. Simpson pour *cemacs.el* (la gestion du mouvement du curseur en mode Big5 modifiée) ;
- K. Shane Hartman pour *picture.el* (picture−mode modifié pour insérer des caractères Big5) ;
- Howard Gayle pour *case−table.el* (fonctions modifiées pour utiliser des tables de syntaxe locales) et *iso−insert.el* (mécanismes d'insertion et de saisie).

Georges KO