

一個原來應該很簡單的事,卻變的很複雜.....

ERP不支援utf-8,因此有一些廠商的帳戶名稱無法輸入,因此要使用一個程式來修改資料
因為使用pspad來儲存utf-8h的資料,並配合autoit3來寫轉換程式
但是讀出來的資料就是亂碼????????????????,找了很多資料,還是?????

最後終於知道了,是pspad設定的問題,不是autoit3的問題

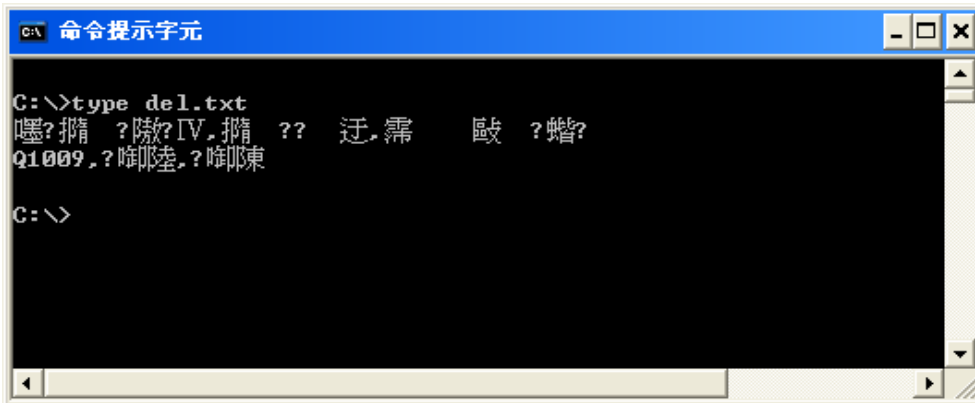
使用pspad來編輯Del.txt檔

#廠商代碼,廠商名稱,變更後名稱

Q1009,新峰,新峯

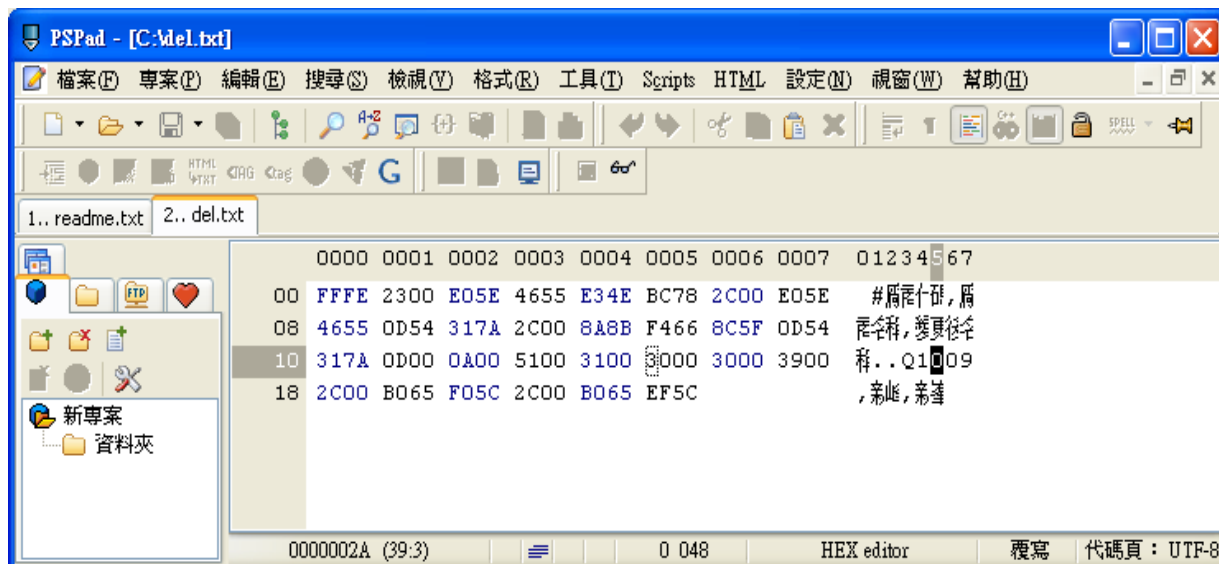


正確應該如下



有嘿為第一個字才對,使用別的程式來儲存,記事本寫好後儲存為 utf-8 的檔案,就可以了

使用 pspad 檢查兩個檔案 hex 編碼



竟然一模一樣,不相信.....

將兩個檔案拉到瀏覽器,IE 或 firefox,使用 notepad 儲存的檔案可以正確的被判別,但 pspad utf-8 的檔案還是有問題

使用 winmerge,比對,也一模一樣但是 notepad 儲存的是 utf-8(b) ,多了(b)

又以前知道嚟,是 utf-8 的第一個字,但它是如何造成的呢?

http://unicode.org/faq/utf_bom.html ,找到答案

Q: What is a BOM?

A: A *byte order mark* (BOM) consists of the character code U+FEFF at the beginning of a data stream, where it can be used as a signature defining the byte order and encoding form, primarily of unmarked plaintext files. Under some higher level protocols, use of a BOM may be mandatory (or prohibited) in the Unicode data stream defined in that protocol. [\[AF\]](#)

BOM(位元順序標示)

就是在檔案的啓始位置加入 U+FEFF,前面的 pspad 的 hex 檢視圖可以看到 FEFF,但 U+不見了

使用 linux 來檢視

```
[root@fc8 ~]# file *.txt
del_note.txt: Unicode text, UTF-8
del.txt: UTF-8 Unicode text, with CRLF line terminators
[root@fc8 ~]# hexdump del_note.txt
00000000 bbef 23bf bbe5 e5a0 8695 bbe4 e7a3 bca2
00000010 e52c a0bb 95e5 e586 8d90 a8e7 2cb1 aee8
00000020 e68a b49b bee5 e58c 8d90 a8e7 0db1 510a
00000030 3031 3930 e62c b096 b3e5 2cb0 96e6 e5b0
00000040 afb3 0a0d
00000044
[root@fc8 ~]# hexdump del.txt
00000000 e523 a0bb 95e5 e486 a3bb a2e7 2cbc bbe5
00000010 e5a0 8695 90e5 e78d b1a8 e82c 8aae 9be6
00000020 e5b4 8cbe 90e5 e78d b1a8 0a0d 3151 3030
00000030 2c39 96e6 e5b0 b0b3 e62c b096 b3e5 0daf
00000040 000a
00000041
```

很明顯不一樣

pspad 很好的程式,不可能就這樣,如何解決呢?

找了一下,如下圖,將 bom 打勾就好了

