



Bisoños Usuarios de GNU/Linux de Mallorca y Alrededores | Bergantells Usuaris de GNU/Linux de Mallorca i Afegitons

Convertir ficheros entre distintos conjuntos de caracteres (UTF, latin1...)

(16712 lectures)

Per **Ricardo Galli Granada**, [gallir](http://mnm.uib.es/gallir/) (<http://mnm.uib.es/gallir/>)

Creado el 12/12/2001 20:55 modificado el 12/12/2001 20:55

Estaba ya cansado de no poder **usar el kword para escribir los artículos de Bulma** debido a que genera salida en UTF8, así que me puse a buscar como convertirlos a Latin1 o ISO-8859-{1,15} que es el estándar en Europa y Bulma.

Aunque el Quanta funciona muy bien, tiene sus problemas si lo que pretende es escribir un artículo de longitud media o larga. Las etiquetas HTML molestan bastante a la lectura y corrección y además tiene algunos bugs de formateo y salto de líneas.

Encontré dos soluciones razonables:

1. Grabar como SGML y luego convertir a HTML con el **sgmltools**.
2. Grabar como HTML y luego convertir los caracteres UTF8 generados por el kword a Latin1 con el **tcs**.

Según el código HTML generado, para escribir los artículos en Bulma me parece **más sencillo usar el tcs** y luego copiar el texto a los campos del formulario de Bulma.

tcs

El paquete Debian se llama igual que el comando, `tcs`. Básicamente lo que hace es convertir la codificación de caracteres de una serie de ficheros a otra codificación.

Para los impacientes, basta con grabar en HTML desde el Kword y luego ejecutar el `tcs`.

```
tcs -t latin1 art_bulma.html
```

Las opciones más importantes del `tcs` son

- `-f tipo_entrada`: especifica que codificación se usa en el fichero de original.
- `-t tipo_salida`: especifica la codificación que queremos para la salida.

Los tipos para la entrada puedes ser obtenidas con `tcs -l`:

```
$ tcs -l
tcs version = 'Sun Mar 26 02:42:25 EST 1995'
cs: utf utf1 ascii 8859-1 latin1 8859-2 8859-3 8859-4
8859-5 8859-6 8859-7 8859-8 8859-9 8859-10 8859-14 koi8
ucode cp866 av cp1251 ov sf1 sf2 jis(from) jis-kanji
ujis ms-kanji big5 gb euc-k tis viet1 viet2 viscii
msdos msdos2 ps2 macrom next atari unicode ebcdic utf-12
```

Como véis, también sirve para convertir desde diversas codificaciones, incluidas `msdos`, `ebcdic`, `atari`, etc.



sgmltools

Para los interesados en soluciones SGML/DocBook, podéis grabar desde el Kword en formato SGML (es un DTD DocBook-Oasis), que curiosamente sí graba en Latin1 (ISO-8859-1), y luego convertir a HTML con el sgmltools (paquete Debian sgml-tools o **linuxdoc-tools**).

Aunque este tema es más complejo y se merece un artículo más largo, para poner un artículo en Bulma basta con hacer:

```
sgmltools --backend=onehtml art_bulma.sgml
```

La opción `onehtml` indica que sólo hay que generar un fichero `.html`. Si se quiere en un subdirectorio hay que especificar `--backend=html`.

E-mail del autor: gallir_ARROBA_uib.es

Podrás encontrar este artículo e información adicional en: <http://bulma.net/body.phtml?nIdNoticia=1061>