



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

Nuevo buscador del Web Bulma basado en htDig (10100 lectures)

Per Ricardo Galli Granada, [gallir](http://mnm.uib.es/gallir/) (<http://mnm.uib.es/gallir/>)
Creado el 14/03/2001 02:53 modificado el 14/03/2001 02:53



El nuevo buscador de WebBulma es el htDig y permite buscar no sólo en el contenido de los artículos, sino también en los mensajes de las listas de correo de Bulma y los comentarios de los lectores. [ACTUALIZACIÓN] Además, como DaniRC no podía encontrar un artículo por problemas de acentos mal puestos ;-), modificamos el programa, los diccionarios y las reglas para que ignore en la generación de índices y búsquedas.

Desde el día 10 de marzo lo hemos puesto en marcha con Guillem Cantallops en forma experimental y ahora ya está a pleno [funcionamiento](#)⁽¹⁾. El buscador es el [htDig](#)⁽²⁾ y lo hemos configurado para que interprete las palabras en castellano, palabras típicas informáticas en inglés y las principales derivaciones de las palabras raíces, por ejemplo: *fichero-ficheros, limitar-limite-limites-limitamos-..* etc. Es muy potente, hasta inclusive está configurado para que trabaje con sinónimos de palabras técnicas, por ejemplo: *IDSN-RDSI, mail-email-emilio, fichero-ficheros*, etc.

El buscador actualiza su base de datos cada madrugada y genera unas bases de datos con índices por palabras y los punteros a cada página web. De esta forma la búsqueda es muy rápida y asigna pesos (los ) dependiendo de la aparición de las palabras. Cinco Tuxes indican el máximo.

Para poder asignar pesos correctos, hemos reorganizado las páginas web para que defina como título de la página al título de la noticia y el META information es el resumen de la misma. Por lo que desde ahora es muy importante poner buenos títulos y resúmenes para que los artículos aparezcan correctamente ordenados. Eso no sólo beneficiará a nuestro buscador, sino también a todos los buscadores.

Y ya que el buscador también funciona con los mensajes a las listas, estaría bien que los *subjects* fuesen del tema del mail, así facilitaremos la vida a todos.

Las configuraciones y base de datos que usamos están disponibles en:

- [Diccionarios](#)⁽³⁾
- [Ficheros de configuración](#)⁽⁴⁾
- [Ejecutables](#)⁽⁵⁾
- [cgi-bin](#)⁽⁶⁾
- [Base de datos](#)⁽⁷⁾ actualizadas diariamente

Modificación para el castellano/español

Los ejecutables que están son el resultado de unas pequeñas modificaciones al String.cc y htString.h para que ignore los acentos. Los diccionarios y reglas `espa~nol.*` también están modificados para ignorar acentos y no expandir en tantas palabras derivadas.

Los ejecutables son directamente usables con los diccionarios de los enlaces anteriores. El que quiera los *diffs* de los fuentes del htDig, que se ponga en contacto con nosotros en la lista Bulmailing (ver enlace en la página).



En cuanto lo termine de escribir, pondré una guía para escribir las noticias y hacer que aparezcan primeros en las herramientas de búsqueda.

--ricardo

Lista de enlaces de este artículo:

1. <http://bulma.net/htdig/>
2. <http://www.htdig.org/>
3. <http://bulma.net/dev/htdig/common/>
4. <http://bulma.net/dev/htdig/conf/>
5. <http://bulma.net/dev/htdig/bin/>
6. <http://bulma.net/dev/htdig/cgi-bin/>
7. <http://bulma.net/dev/htdig/db/>

E-mail del autor: gallir_ARROBA_uib.es

Podrás encontrar este artículo e información adicional en: <http://bulma.net/body.phtml?nIdNoticia=558>