

Busquedas de páginas en tu servidor web: htdig

Por Paco Aldarias Raya

Impreso: 13 de abril de 2004

Email: [pacolinux arroba inicia punto es](mailto:pacolinux@pacodebian.es)

Web: <http://pagina.de/pacodebian>

Con Linux Debian. En Valencia (España)

Este documento es de libre reproducción siempre que se cite su fuente.

Realizado con: L^AT_EX

Índice

Índice	1
1. Introducción	1
2. Requisitos	1
3. Instalación	2
4. Configuración	2
5. Crear un fichero de páginas	2
6. Inicializar la base de datos	2
7. Uso	2
8. Bibliografía	2
Índice alfabético	3

1. Introducción

Htdig es "WWW Search Engine Software" es decir un motor de búsqueda de páginas web.

Vamos a ver la forma de poder tener un buscador de ficheros o páginas web, dentro de nuestro propio servidor web. Para ello usaremos htdig

El objetivo es poder poner un texto y busque en nuestra web (o otra web) aquellas páginas o ficheros que contengan dicho texto.

La página <http://bulma.net/htdig/> usa este tipo de software para las busquedas.

2. Requisitos

Tener instalado el servidor apache.

3. Instalación

Desde consola como usuario root:

```
apt-get install htdig
```

4. Configuración

Desde consola como usuario root, editaremos el fichero: nano /etc/htdig/htdig.conf

Deberemos poner la dirección de nuestra web aquí:

```
start_url: http://aldarias.dyndns.org/paginas/
```

5. Crear un fichero de páginas

Para el fichero htdig.txt con las páginas q me interesan

```
find /var/www/paginas/ -type f -name \*.html -print | sed -e 's| /var/www/pa
```

6. Inicializar la base de datos

Desde consola como usuario root:

```
htdig -v -s htdig.txt
```

7. Uso

Abriremos el buscador así:
`http://127.0.0.1/search.html`

8. Bibliografía

1. Web Oficial <http://www.htdig.org>

Índice alfabético

htdig, 2