

Reescritura de URL - URL rewriting

Del URL al servidor Web

Cuando escribe la dirección <http://www.carpediem.fr/corpo/home.html> en su navegador, su Proveedor de Acceso Internet convertirá esta dirección en un IP, consultando un nombre de servidor. De esta forma sabrá a cual ordenador dirigirse. Por lo tanto será “redirigido” hacia el ordenador que corresponde a su inicial url. Si su dominio es configurado correctamente hacia el servidor WEB de ese ordenador, el, corresponderá a una ubicación sobre el disco duro o más precisamente a un directorio y a un archivo (carpeta corpo, fichero home.html).

Esto es lo que pasa cuando introduce un url en su navegante.

Definición de la reescritura de URL

A la base, la reescritura de url es una técnica que consiste en reescribir (por el servidor WEB) en una forma más simple, URLs complejos. Por url complejos entendemos direcciones de tipo:

Ejemplo 1:

http://ressources.carpediem.fr/flux_xml/index.php?g=hetero&c=51&n=3&lg=it

URL con muchos parámetros que a primera vista nos parecen totalmente extranjeros. Efectivamente, un url que nos indique el contenido será más pertinente para nosotros pero sobre todo para los motores de búsqueda.

Ejemplo 2:

http://ressources.carpediem.fr/flux_xml/it-hetero-51-videos-3.html

Esta técnica sirve a hacer que el url del ejemplo 2 corresponde al url del ejemplo 1 sin que el fichero del ejemplo 2 existiera físicamente en el ordenador. No es una redirección, sino una reinterpretación, una reescritura.

Esta técnica es más y más difundida porque ella fomenta una buena referencia. En efecto, los motores de búsqueda toman en cuenta los elementos de un url en función de la pertinencia de la página respectiva. Es ideal para tratar por encima unas palabras esenciales suplementarias. Para más informaciones sobre la referencia consulte nuestra documentación en formato PDF.

Los archivos .htaccess

Los archivos .htaccess son ficheros de configuración de servidores WEB que permiten definir las reglas en la carpeta y en todos sus directorios (que no tienen ficheros .htaccess en el interior). Pueden ser utilizados a gestionar el acceso a ciertos archivos, proteger el acceso a una carpeta a través de una contraseña, definir las páginas de errores personalizadas...

Entre las funciones más esenciales permitidas por el fichero .htaccess se encuentra la función de la reescritura “por vuelo”, es esta la operatividad que nos va a concernir. Es útil puntualizar que ciertos anfitriones de los

sitios WEB no han activado el modulo de reescritura sobre sus servidores. En este caso usted tiene dos posibilidades: o renuncie al url rewriting o cambie simplemente el anfitrión del sitio WEB. Recuérdese que si obtiene un cierto número de negocios, Carpe Diem os puede recibir gratis en su sitio.

Ensayemos juntos con un ejemplo sencillo

Abre su cliente FTP, conectase a su servidor FTP y crea una carpeta nombrada “rewriting”. En esta ponga una página html simple conteniendo solamente una línea con texto.

Ejemplo: **Sin embargo funciona**

Lo van a nombrar ok.html. Ponga en esta carpeta un fichero que tenga el nombre .htaccess conteniendo las siguientes líneas:

```
Options +FollowSymlinks
RewriteEngine on
RewriteRule ^blablabla.html$ ok.html [L]
```

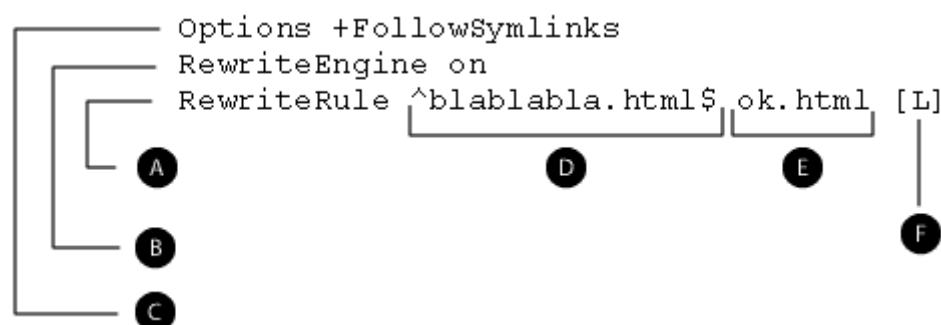
Después trata de entrar en el blablabla.html sabiendo que usted no dispone de este fichero en su sitio WEB. El resultado tendría que ser algo parecido a lo siguiente:

<http://www.mywebsite.com/rewriting/blablabla.html>

¡Si han seguido el ejemplo, usted tendría que ver una página WEB con una línea de texto en grueso que os diga claramente que funciona y sin embargo el fichero blablabla.html no existe!

Puede ser que su alojador de sitio WEB no os permita añadir el “FollowSymLinks” en las opciones del servidor WEB. Esta primera línea autoriza al servidor WEB a seguir las líneas simbólicas en la carpeta respectiva. Su presencia permite corregir un contingente defecto de configuración en el fichero httpd.conf del servidor. A este fichero evidentemente que usted no podrá acceder en el caso de los alojamientos mutuales.

Esquema explicativo del contenido del fichero .htaccess



Puede suprimir esta línea (C) sin problemas.

La segunda línea (B) contiene instrucciones de activación de la reescritura del url. Ella es obligatoria. Ella siempre debe ser insertada en el fichero .htaccess dando sencillamente al servido WEB la instrucción de lanzar el motor de reescritura.

Pasemos ahora a la tercera línea. La palabra Rewrite Rule introduce cualquier regla de reescritura. La parte D es la primera parte de esta norma, es la que determina la cadena de caracteres que el modulo tendrá que buscar para reescribir. Ella contiene dos caracteres especiales marcadas al principio de la cadena con (^) y al final de la cadena con (\$). La parte E corresponde a la cadena de reemplazo. Generalmente, esta parte corresponde al nombre del fichero que existe realmente en su servidor FTP. En lo que concierne la parte F, esta corresponde al “flag” que significa que esta regla es la última que se aplica en este caso.

Traducción:

Si en la carpeta “rewriting”, se entra sobre el fichero que tiene el nombre blablabla.html, entonces se trata del archivo ok.html.

Ejemplo más avanzado de utilización

Ejemplo: tiene un sitio que propone una serie de vídeos clasificados por orientación sexual (hetero o homosexual) repartidos en distintas categorías en tres idiomas: Francés, Inglés, Italiano. Aquí esta el esquema de su URL: <http://www.mywebsite.com/index.php?g=hetero&c=51&lg=it&n=3>

En este ejemplo, “g” muestra si somos en la sección hetero o homosexual, “c” siendo el número de la categoría, “lg” significa la idioma, lengua y “n” es el número de la pagina porque en función de las categorías existen varias páginas con vídeos. Aquí el valor “c” corresponde a la categoría “negritas”. Es obvio que no traerán tráfico con sus url.

Sin duda preferirían que sus visitantes accedan a una página de url de este tipo:

<http://www.mywebsite.com/hetero-it-51-videos-3.html>

Si apliquemos la lógica del ejemplo anterior llegamos fácil a:

```
Options +FollowSymlinks
RewriteEngine on
RewriteRule ^hetero-it-51-videos-3.html$ index.php?g=hetero&c=51&lg=it&n=3 [L]
```

¿Por qué hacer una línea por página? ¿No podemos elaborar una regla que engloba cada caso? Analizamos punto a punto las posibles variaciones de este URL. Por consiguiente, tenemos hetero u homosexual en valores a través del parámetro “g”, un número a través del parámetro “c”, 3 idiomas a través del parámetro “lg”, respectivamente Francés, inglés, italiano y un número a través del “n”.

```
Options +FollowSymlinks
RewriteEngine on
RewriteRule ^(hetero|gay)-(fr|en|it)-([0-9]+)-([0-9]+)\.html$ index.php?g=$1&c=$3&lg=$2&n=$4 [L]
```

Explicaciones

Hemos explicado ya en el ejemplo anterior la ventaja de las dos primeras líneas, no vamos a revenir acerca este asunto. En esta regla encontramos 4 expresiones regulares relacionadas por guiones.

Empecemos con las primeras dos: **(hetero|gay)** y **(fr|en|it)** que designan varias opciones. En el primer caso, o hetero o homosexual; en el segundo caso, o fr, o en, o todavía it. Lo que da por ejemplo: hetero-it.

Después encontramos dos veces la misma cadena de caracteres ([0-9]+). Las paréntesis cuadrados [] determinan un intervalo, por conseguido [0-9] determina el intervalo de números “0” hasta “9”.

El signo “+” que sigue inmediatamente el intervalo significa “uno o más casos de la expresión que preceda”, en este ejemplo nuestro intervalo [0-9]. Para resumir esta expresión regulada significa cualquier número entero superior o equivalente con 0. En el primer caso, esta será “c” (respectivamente el id de la categoría), en el segundo caso “n” (respectivamente el número de la página)

Por fin las paréntesis que rodean todo de « ([0-9]+) » solicitan al motor de reescritura a clasificar la cadena encontrada y a guardarla en una variable interna con el fin de utilizarla más tarde.

De este modo \$1 representa la primera expresión regular, \$2 la segunda, \$3 la tercera y así sucesivamente. Teniendo en cuenta que el punto decimal tiene una significación especial en las expresiones, es natural que en nuestro caso sea precedido por el carácter de escape ‘\’.

Después de comprender la parte izquierda de la expresión, os pondrá fácil a entender la parte derecha de la expresión.

```
RewriteRule ^(hetero|gay)-(fr|en|it)-([0-9]+)-([0-9]+)\.html$ index.php?g=$1&c=$3&lg=$2&n=$4 [L]
```

Tomamos el ejemplo del principio en donde “g” tenía la valor “hetero”, “c” tenía la valor “51”, “lg” tenía por la valor “it” y en final “n” tenía la valor “3” el url siendo esto:

<http://www.mywebsite.com/hetero-it-51-videos-3.html>

Tenemos ‘hetero’ o ‘homosexual’ seguido por un guión, después una de las tres lenguas seguidas por un guión, después el número superior o equivalente con 0 seguido por un guión y al final otro número seguido por ‘.html’.

¡Nuestro url encuadra esta regla, la reescritura va a efectuarse!

Algunas expresiones regularías a conocer:

.	Cualquier carácter
[abad]	Cualquier carácter de esta lista
[^abad]	Cualquier carácter que no está en esta lista (otro del a, b, c o d)
Blanco negro	Alternativa, o ‘blanco’ o ‘negro’
+	Una o N caso(s) de la expresión que preceda (N>1)
*	0 o N caso(s) de la expresión que preceda (N>0)
(texto)	El agrupación permite el uso de las referencias invertidas (\$1,... \$n) Esta también utilizada para demarcar una alternativa como en (blanco negro)
^	Marca el inicio de la línea
\$	Marca el fin de la línea
\	Permite la inserción de cualquier carácter que sigue y la descarta de la especial significación, por ejemplo \.

Varios flags útiles

Aquí están varios flags útiles para facilitar el mantenimiento de un sitio:

- [L]** Si una reescritura es efectuada, c'èst la forma reescrita que sera utilizada en entrada para las reglas siguientes. El flag [L] permite de salir prematuramente de la bucle.
 Dans ces deux formes une redirection est effectuée.
 Si l'argument code n'est pas précisé, une redirection 302 (déplacé temporairement) est effectuée. Si
- [R]** vous souhaitez faire savoir au navigateur/robot qu'une page a été remplacée définitivement, utiliser **[R=código]** le code 301 como dans :
- ```
RewriteRule ^ancien\.html$ http://domaine.tld/nouveau.html [R=301,L]
```
- Dans ce cas précis, une reescritura "externa" s'impose (utilización de http://...)
- [F]** Forbidden - interdicto. Retorne un código 403, por ejemplo : RewriteRule ^secret.html\$ - [F]  
 NoCase, ou « insensible à la casse ». La règle suivante :
- [NC]** RewriteRule ^script\.php\$ programme.php [NC,L]  
 S'appliquera aussi bien à script .php, SCRIPT.PHP ou ScRiPt .PhP
- [G]** Gone. Esta página no existe más o entra un membrete. http 410
- [C]** Chain, hilera con regla o con reglas siguientes hasta la primera regla que no empiece con [C]  
 Apache interprete este flag como sigue: Si hay una reescritura (la regla fue comprobada), la regla que sigue esta ejecutada junto a la hilera reescrita en introducción. Si la regla no se verifica, todas las reglas que siguen hasta la primera que no tiene el flag [C] se apliquen.
- [QSA]** Query String Append Anidí al final de la expresión QUERY STRING, después la reescritura. Se guarda hasta la última regla de la reescritura. A menudo esta utilizada con el flag [L], igual el [QSA,L]

|                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [L]               | Si se efectuó una reescritura, esa es la forma reescrita que será utilizada como introducción para las reglas que siguen. El flag L permite la salida prematura del bucle.                                                                                                                                                                                                                                                                                        |
| [R]<br>[R=código] | De estos dos modos se realiza una redirección. Si el argumento código no está especificado, una redirección 302 (desplazada temporal) se efectúa. Si quieren indicar a su navegante/al robote que una página fue remplazada por definitivo, utiliza el código 301 como en el siguiente:<br><br><pre>RewriteRule ^ancien\.html\$ http://domaine.tld/nouveau.html [R=301,L]</pre> En este caso, una reescritura externa es necesaria (utilización del http:// ... ) |
| [F]               | Forbidden- prohibido, interdicción Muestra un código 403 por ejemplo : RewriteRule ^secret.html\$ - [F]                                                                                                                                                                                                                                                                                                                                                           |
| [NC]              | Nocase- o no puede ser arrompido La siguiente regla:<br><br><pre>RewriteRule ^script\.php\$ programme.php [NC,L]</pre> Se aplica igual de bien al script php, SCRIPT.PHP o ScRiPt .PhP                                                                                                                                                                                                                                                                            |
| [G]               | Gone. Esta página no existe más o entra un membrete. http 410                                                                                                                                                                                                                                                                                                                                                                                                     |
| [C]               | Chain, hilera con regla o con reglas siguientes hasta la primera regla que no empiece con [C]<br>Apache interpreta este flag como sigue: Si hay una reescritura (la regla fue comprobada), la regla que sigue esta ejecutada junto a la hilera reescrita en introducción. Si la regla no se verifica, todas las reglas que siguen hasta la primera que no tiene el flag [C] se apliquen.                                                                          |
| [QSA]             | Query String Append Anidí al final de la expresión QUERY STRING, después la reescritura. Se guarda hasta la última regla de la reescritura. A menudo esta utilizada con el flag [L], igual el [QSA,L]                                                                                                                                                                                                                                                             |

Esta lista no esta completa, porque existen otros flag admitidos. La lista completa esta descrita en la documentación del modulo mod\_rewrite en el sitio de Apache.

[http://httpd.apache.org/docs-2.1/mod/mod\\_rewrite.html](http://httpd.apache.org/docs-2.1/mod/mod_rewrite.html)