



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

## Curiosidades de Baleares: Apache 78 - Microsoft IIS: 15 (10126 lectures)

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

Creado el 22/04/2004 02:03 modificado el 22/04/2004 02:03

*Un grupo de investigadores de la UIB (informáticos y economistas) estamos realizando un estudio de la web de Baleares y su correlación -si es que existe- con las actividades económicas de las islas. Una parte importante del trabajo es crear lo más automáticamente posible una base de datos con los sitios de Baleares y sus enlaces entre ellos. Llegamos a un punto que tenemos más o menos datos fiables. Por ello adelanto unas curiosidades, entre ellas el porcentaje de uso de los servidores que muestra el título.*

Ahora estamos trabajando en la primera parte del proyecto: recorrer con un robot los webs a partir de un conjunto inicial de webs y luego ir agregando nuevos sitios a la lista de dominios locales. La automatización de la decisión es una parte también muy importante, de hecho hay una matemática haciendo su tesis doctoral sólo en este tema.

El caso es que después de un par de meses ya tenemos una base de datos importante de webs locales (básicamente de personas y empresas radicadas en Baleares, o que parte importante de su actividad sea en las islas). Por ahora llevamos más de 1000 servidores y menos de 10.000, y más de 6.000.000 de enlaces -ya diremos el número exacto en cuanto validemos más datos. Más detalles “técnicos” al final.

### Tipos de servidores

Lo más llamativo es que la proporción de uso del Apache es bastante superior, más de 10 puntos, al [reportado por Netcraft](#)<sup>(1)</sup> para toda la red web. En la siguiente tabla tenéis los valores más significativos.

Tipo de servidor reportado	Porcentaje
Apache	77.65%
Microsoft IIS	14.85%
Oracle	1.42%
PHP	44.00%
Perl	19.83%

En esta tabla sólo se han tomado en cuenta los servidores que han desvelan esos datos en la “firma” de las cabeceras. Dichos servidores son aproximadamente el 95% del total de servidores en la base de datos.

### Distribución por países

Otro dato curioso es la lista de países donde están alojados los servidores de los sitios anteriores. En total hemos encontrado servidores alojados en 14 países distintos, casi el 60% están en España, luego le sigue USA y después



Alemania. Los datos completos en la siguiente tabla.

País	Porcentaje
AR	0.24%
AU	0.16%
CA	1.18%
CH	0.24%
DE	14.16%
DK	0.16%
ES	59.40%
FR	3.70%
GB	3.38%
IT	0.31%
NL	0.16%
NO	0.24%
SE	0.87%
US	16.60%

## Más datos técnicos

Para los fanáticos de datos y que sepan un poco del tema, aquí van los otros datos que calculamos, sobre todo para validar la muestra.

Número medio de enlaces a sitios distintos	4.184
Nro. de sitios aislados (sin enlaces desde otros sitios)	26.000
Longitud media del camino entre dos webs conectados	4.156
Diámetro del grafo (la mayor de las distancias más cortas)	13.000
Densidad de enlaces con bucles	.003
Densidad de enlaces con bucles	.003
<b>Coefficiente de Clustering = CC</b>	<b>.414</b>
Grado de Prestigio (centralidad interior)	.056
Grado de Intermediación	.154

Lo más curioso es que el coeficiente de *clustering* (CC) es elevadísimo, un 41%. Este coeficiente indica en las redes sociales si las personas que yo conozco se conocen también entre ellas. En el caso de web de Baleares dice que si un sitio tiene enlaces a dos distintos, la probabilidad de que haya enlace entre esos dos es de un 41%.



Por otro lado, todos los demás datos se corresponden con lo que dice la “teoría” en base a datos genéricos de Internet. Por ejemplo, con la evidencia hasta ahora, los webs de Baleares no conforman un grafo aleatorio, sino un *small world*, con una distribución *power-law* y además *scale-free*, tal como es la red web general. El hecho que sea *power-law* de alguna manera valida como correctos a nuestros datos, ya que en teoría si se cogen muestras más pequeñas de la red web general (que es una *power-law*), esos subgrafos también deberían seguir una distribución *power-law*.

O sea, la evidencia dice que nuestra red es coherente con Internet, a pesar del altísimo nivel de endogamia indicado por el CC :-)

NOTA: si estáis interesados en colaborar, sobre todo en acabar de reprogramar el robot en lenguaje Perl (ahora está en C++, e l nuevo debe estar en menos de un mes, ya tengo el esqueleto basado en WWW::Mechanize), seréis muy bienvenidos :-)

---

**Lista de enlaces de este artículo:**

1. [http://news.netcraft.com/archives/web\\_server\\_survey.html](http://news.netcraft.com/archives/web_server_survey.html)

---

E-mail del autor: [gallir\\_ARROBA\\_uib.es](mailto:gallir_ARROBA_uib.es)

Podrás encontrar este artículo e información adicional en: <http://bulma.net/body.phtml?nIdNoticia=2018>