

La evolución de los contenidos en Internet

Sebastián Cáceres, Analista

Existen básicamente dos tipos de contenidos *web*. El primero es al que se accede tradicionalmente a través de buscadores que ordenan los documentos obedeciendo a diversos criterios, como el número de vínculos al documento, el número de visitas recibidas, etc. Este grupo de contenidos, si bien es el más utilizado, constituye una parte relativamente pequeña del conjunto de información disponible. A principios del año 2001 incluía unos 2.500 millones de documentos, equivalentes a unos 25-50 terabytes (2^{40} bytes) de información, creciendo a un ritmo de unas 7,3 millones de páginas por día. Actualmente, el mayor motor de búsqueda tiene indexadas 3.300 millones de páginas *web*.

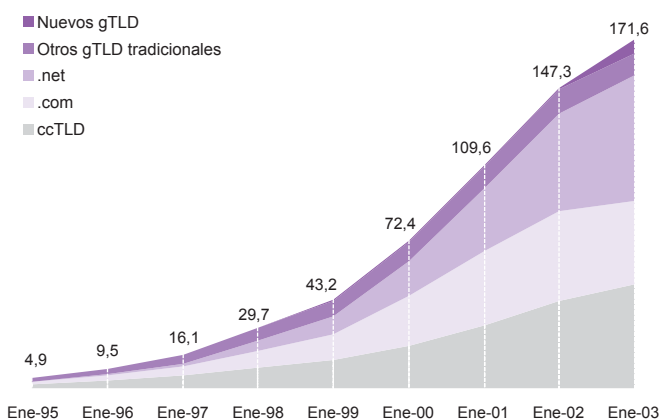
El otro grupo de contenidos, normalmente no identificados por los motores de búsqueda tradicionales, se conoce como “*web invisible*” o *deep web*, y consiste, principalmente, en bases de datos especializadas y en sitios *web* dinámicos, accesibles a través de Internet. La información contenida en este grupo podría alcanzar, según ciertas estimaciones,

un volumen cientos de veces mayor que el del otro grupo, unos 7.500 terabytes. Al tener en cuenta toda la información accesible vía *web*, es decir, las bases de datos conectadas a Internet, páginas dinámicas, sitios intranet, etc., se podrían contar unos 550.000 millones de documentos, accesibles públicamente en un 95% de los casos. Más de la mitad son bases de datos, un 13% son sitios internos, el 11% son publicaciones, un 5% son sitios de compras o subastas, otro 5% clasificados, un 3% portales y un 9% sitios de otro tipo.

Una forma de medir la evolución en el volumen de la información disponible es a través del número de ordenadores centrales

(*hosts*) conectados permanentemente a Internet, que proveen contenidos *web*. El número total de *hosts* es el indicador más común del

Evolución del número de *hosts* en el mundo Por dominio, en millones



gTLD: *general Top Level Domain*: Dominios de primer nivel generales. Los originales son “.com”, “.net”, “.gov”, “.org”, “.edu”, “.mil”, “.int”; y los nuevos “.biz”, “.aero”, “.coop”, “.name”, “.info”, “.museum”, “.pro”. **ccTLD:** *country code Top Level Domain*: Dominios de primer nivel con código nacional; por ejemplo, “.es”.

Fundación AUNA a partir de Internet Software Consortium (<http://www.isc.org/>)

tamaño mínimo de Internet, y se incluyen en este concepto todos los ordenadores centrales que tienen una dirección IP asociada, provean o no contenidos *web*.

Entre 1981 y 1991 el número total de *hosts* pasó de 213 a poco más de medio millón. Este número, sin embargo, representaba a nivel mundial menos de un *host* cada 10.000 habitantes. El cambio de concepto que se produce ese año en la arquitectura de Internet generó un crecimiento espectacular de la Red. Como consecuencia, el número de ordenadores centrales se multiplicó por 300 en la década 1991-2001. Más del 90%, sin embargo, se encontraba concentrado en los 23 países de mayor nivel de renta.

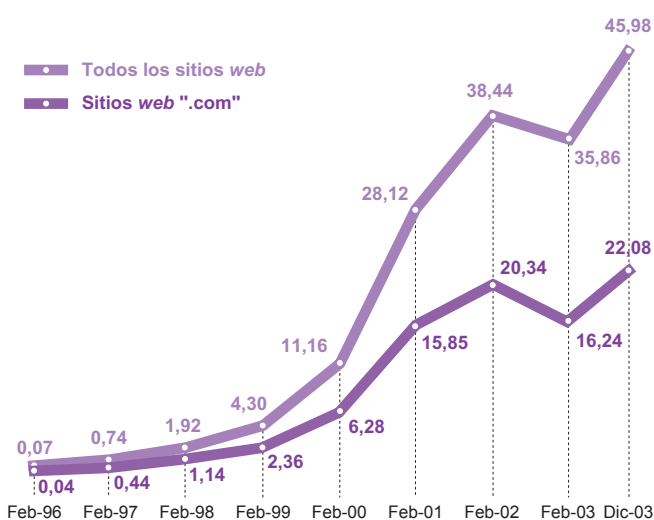
Durante el año 2002, el número de *hosts* ha crecido en unos 25 millones, superando los 170 millones. Dentro de los dominios de primer nivel generales (gTLD), los que más han aumentado son los “.net”, que lo han hecho en más de 14 millones. Los dominios nacionales de primer nivel (ccTLD) también han crecido fuertemente en este año, en unos 8 millones. Los *hosts* con otras extensiones tradicionales de dominios generales, especialmente los “.com”, han sufrido una importante caída en el período, al tiempo que los nuevos dominios generales apenas han comenzado a emerger (“.biz”, “.aero”, “.info”, etc.) Fuera de EE UU, la extensión “.jp” (Japón) sigue siendo la que mayor número de *hosts* con extensión de dominio nacional posee. Italia ha pasado en este año del sexto puesto

al segundo, superando a Canadá, Alemania, el Reino Unido y Australia.

La caída en el número de *hosts* con extensión “.com” ha tenido su paralelo en el número de sitios *web*. Con el fin de detectar el software utilizado por los servidores conectados a Internet, la empresa *Netcraft* obtiene periódicamente el número de *hosts* que proveen contenidos *web*, infiriendo el número de sitios *web* activos en cada momento. Entre febrero de 1995 y febrero de 2002, éstos pasaron de unos 20.000 a poco más de 38 millones, suponiendo una tasa media anual de crecimiento superior al 200%. En 2002, sin embargo, por primera vez en la historia su número ha sufrido

una caída. Su explicación se encuentra fácilmente, observando el comportamiento de los sitios *web* “.com”, que han caído aún más que el total,

Evolución del número de sitios *web* en el mundo. En millones



Fundación AUNA a partir de Netcraft

e incluso en valor absoluto. Durante el año 2003, la tendencia ha vuelto a revertirse, ya que el número de sitios *web* ha crecido en este año en un 30%, y el número de sitios *web* “.com” lo ha hecho en un 35%.

Un estudio paralelo al de Netcraft es el realizado por el *Online Computer Library Center*, OCLC, de EE UU. Aunque no tiene en cuenta el hosting virtual de sitios –sitios que comparten direcciones IP–, facilita información importante relativa a los contenidos y tipos de sitios *web*. Según sus investigaciones, se ha producido un importante crecimiento en la proporción de sitios privados –con algún tipo de restricción para el acceso a sus contenidos–, que han pasado de representar el 12% en 1998 al 29% del total de sitios en 2002. También hay un crecimiento en la proporción de sitios provisionales –con contenidos triviales o en construcción– y una importante pérdida de participación de los sitios públicos, que han pasado a representar un 35% del total, frente a un 55% en 1998.

Por regiones, es muy importante el crecimiento de los contenidos en idiomas asiáticos, que en sólo dos años han multiplicado por más de tres su

cuota en Internet. Esto ha sido en buena parte impulsado por la demanda de esos países, predominantemente de navegación. En China y Japón, los dos países de mayor demanda, las razones principales para conectarse son, junto al correo electrónico, la búsqueda de información y el entretenimiento. En el caso de Japón, donde gran parte de los usuarios de Internet acceden a través de un móvil, son muy populares los contenidos sencillos, como el correo electrónico, y los servicios que son posibles a través de éstos. La mayor penetración de la banda ancha, por otra parte, comienza a generar en esta región nuevas demandas, en especial de aplicaciones más complejas, como vídeo, juegos en red, etc.

La búsqueda de información y noticias es también la actividad más importante de los usuarios de la UE. En esta región, la industria

de los contenidos digitales emplea a unos cuatro millones de personas y ha sido valorada en más de 400.000 millones de euros, equivalentes

Actividades de los usuarios de Internet. Principales mercados, en % de los usuarios

China (1) Servicios más frecuentes	Japón (2) Propósito de Uso	UE (3) Usos	EEUU (4) Actividades más populares
e-mail 93	e-mail 96	e-mail 79	e-mail/Instant Messaging 88
Motores de búsqueda 68	Revistas por e-mail 76	Noticias 74	Navegación 76
Chats 45	Búsqueda de información 70	Viajes 63	Noticias 52
Carga/Descarga de software 45	eComercio 52	Formación 46	Entretenimiento 46
Adquisición de información 42	Noticias 49	Salud 38	eComercio 45

Fundación AUNA a partir de (1) CNNIC, (2) MPMHAPT, (3) Gallup Europe 2002, y (4) UCLA Internet Report 2002

al 5% del PIB comunitario. De ahí que, desde la Comisión Europea, un grupo de trabajo de la Dirección de Sociedad de la Información haya concluido en la necesidad de crear un “Observatorio de los contenidos digitales”, con el fin de realizar un seguimiento del mercado, y asistir a gobiernos y participantes de otros sectores en el proceso de creación, producción y diseminación de los contenidos.

En EE UU, las herramientas de información son también muy importantes. En este país, sin embargo, los usuarios no perciben Internet como una fuente de entretenimiento –lo que a corto plazo no está favoreciendo el desarrollo de nuevas aplicaciones de banda ancha– aunque sí ven la Red como una importante fuente de información.

La producción de contenidos en inglés, por otra parte, no es patrimonio exclusivo de los países angloparlantes sino que en muchos otros países las empresas prefieren utilizarlo en sus sitios *web* –en lugar de los idiomas nacionales– con el fin de alcanzar mercados más amplios. El efecto negativo que esto tiene en la producción de contenidos en lenguas locales es una de las razones por las cuales no se ha generalizado el acceso en muchos países, en especial, en los menos desarrollados de Asia y África.

En este último continente, prácticamente todos los países tienen algún tipo de sitio *web*, alojado en servidores locales o internacionales, que los representa, oficialmente o no. La mayoría de estos sitios se encuentran en inglés, aunque cada vez más comienzan a verse sitios bilingües o en idiomas locales. A finales de 2001, menos de 100.000 URL habían sido registradas en África, aunque se calcula que existen más contenidos africanos fuera del continente que dentro de él. La razón estriba en el alto coste de alojamiento por la falta de capacidad de los ISP locales. Los ISP externos son más baratos y proveen de mayor valor añadido a los contenidos. Los más activos productores –y usuarios– de contenidos africanos son, de este modo, los que viven fuera del continente, los descendientes de africanos y los exiliados, normalmente por conflictos bélicos. Gran parte de los sitios *web* son informales, a menudo generados por los propios usuarios, mediante comentarios y compilaciones: chats políticos, páginas *web* personales e información recolectada y cargada.

Finalmente, en el caso de los países árabes, la producción es deficiente, tanto con respecto a su patrimonio cultural como a creatividad. La mayoría de los recursos no se encuentran digitalizados, y en cuanto a nuevos contenidos, se aprecia una significativa escasez □