

Control de Acceso y Estadísticas a las Páginas de un Servicio Web en el Ámbito Docente Universitario

Jesús M. Álvarez Llorente

Departamento de Ingeniería Electrónica, Sistemas Informáticos y Automática
Escuela Politécnica Superior. Universidad de Huelva
21071 Palos de la Frontera (Huelva)
e-mail: llorente@uhu.es

Resumen

En este trabajo se presenta un sistema desarrollado para el control de acceso y realización de estadísticas de visitas a las páginas de un sitio web relacionado con la docencia universitaria. Mediante un sencillo sistema de registro, que se puede incluir de forma trivial en las páginas del servidor web, podremos obtener diversos informes sobre la procedencia, costumbres y preferencias de nuestros visitantes, así como de la actividad general del sitio web, el interés que suscita cada sección, etc., lo que nos resultará muy valioso a la hora de organizar las páginas y su contenido para prestar el servicio al que se disponen, de la manera más eficaz.

Analizaremos la forma de obtener y almacenar la información necesaria, así como el procedimiento para obtener los informes más relevantes, y cómo interpretarlos en el ámbito de aplicación en que se está utilizando el sistema: el apoyo en la docencia de dos asignaturas de la carrera de Ingeniería Técnica en Informática.

1. Introducción

Cada día somos más los profesores universitarios que construimos sitios web como suplemento a nuestra actividad docente. Inicialmente se limitaban a una página en la que se mostraba el nombre de la asignatura, el temario, el horario y, con suerte, se podrían descargar algunos materiales. Poco a poco hemos descubierto las posibilidades de Internet para permitir ofrecer, a través de un sitio web, un conjunto de servicios de reconocida utilidad: información sobre la asignatura, descarga de apuntes y materiales,

publicación de noticias, foros de debate e intercambio, consulta de notas, tutorías virtuales, etc. [3] [4] [5] [6]

Una de las primeras inquietudes que tenemos como constructores de este tipo de sitios web es saber si resultan o no útiles al alumnado, y el primer impulso es incluir en las mismas un contador de visitas. Es fácil encontrar en la red servicios públicos y gratuitos de contadores que proporcionan información bastante completa acerca de las visitas a nuestras páginas: fecha, hora y dirección de las últimas conexiones, resúmenes organizados por horas, días, meses, dominios de Internet, etc. Como ejemplo, citamos los servicios *NetStat* (<http://es.nedstat.net>) y *CiberStats* (<http://www.infase.es/ciberstats/>). El problema planteado por estos servicios es doble:

- Por un lado, la información completa de accesos a las páginas es propiedad de las empresas que prestan el servicio, no quedando a nuestra disposición para la posible obtención de informes proporcionados por el servicio.
- La velocidad de conexión con los servidores que prestan los servicios suele ser notablemente más lenta, lo que ocasiona la pérdida de la información de acceso en visitas cortas.

Pero el control de accesos no sólo sirve para saciar nuestra curiosidad de quién, cuándo y desde dónde acceden a los servicios de nuestra web. Ya no nos sorprende enterarnos de que a determinado profesor le llegan, a través de correo electrónico o medios similares, y refugiándose en el supuesto anonimato de la red, insultos, amenazas o material ofensivo, a menudo enviado desde los mismos laboratorios de informática de las escuelas,

incluso utilizando los servicios web que el profesor pone a disposición del alumnado como apoyo a su docencia.

Por estas razones, se ha decidido diseñar un sistema propio que registre cada acceso a cada página de un sitio web y organice la información de registro de tal manera que permita obtener distintos informes sobre la actividad del sitio, sin que esto altere los servicios prestados.

El sistema construido forma parte de los servicios del sitio web correspondiente a las asignaturas de *Ampliación de Sistemas Operativos* y *Prácticas de Sistemas Operativos* de la carrera de Ingeniería Técnica Informática, integrados dentro de la página personal del autor, en el servidor web de la Universidad de Huelva (disponible en <http://www.uhu.es/jesus.alvarez>).

2. Objetivos

Por las razones anteriormente descritas, el sistema de control de acceso y estadísticas construido, se ha diseñado con los siguientes objetivos:

1. Registrar todos los accesos a las páginas, de manera que sea posible establecer quién y desde dónde se visita cada página.
2. Analizar la actividad de un visitante a lo largo de las distintas páginas del sitio web: qué páginas visita y en qué orden.
3. Analizar la repercusión que tienen sobre el acceso a la web determinadas circunstancias como la cercanía a fechas de exámenes, la incorporación de nueva información, etc.

Estos objetivos deben lograrse en todo caso sin que el funcionamiento normal del sitio web se vea afectado. Por supuesto, el sistema debe ser fácilmente ampliable de manera permanente, y en la medida de lo posible, portable. Además, resulta interesante que la información generada se pueda consultar a través del propio sitio web.

Con estas premisas se pretenden obtener, al menos, los siguientes informes:

- Recuento de accesos a cada página, desglosando las visitas totales, las realizadas en los últimos días, semanas, etc., lo que nos permitirá conocer la actividad reciente del sitio web.

- Conjunto de visitantes que han accedido a cada página, lo que nos permitirá conocer el perfil del usuario que accede a cada sección.
- Visitas realizadas a las páginas por un usuario a lo largo del tiempo, lo que nos permitirá conocer el conjunto de intereses del usuario.
- Estudio evolutivo del número de visitas a cada página, lo que nos permitirá conocer la incidencia de determinados eventos en la actividad del sitio.

También podemos considerar otros informes menos relevantes (debido a que proporcionan una información menos útil), como los ofrecidos por los mencionados sistemas *NetStat* y *CiberStat*: costumbres horarias, accesos por meses, días o semanas, enlaces de procedencia, etc.

3. Metodología y funcionamiento

La fórmula para conseguir los objetivos es muy simple: registrar el momento y procedencia de cada acceso a cada página. Esta información basta para obtener cada uno de los informes de acceso anteriormente planteados. El mecanismo de registro es el mismo empleado en los sistemas públicos de contadores: incluir en cada página una referencia a un elemento dinámico (un *cgi*) en forma de imagen (visible o no) que cada vez que se carga, registra un acceso. [7] [8] [10]

El mecanismo se puede extender a la descarga de archivos de cualquier tipo, de manera que la descarga de un archivo se puede contabilizar de la misma forma que la visita a una página ficticia.

Cada acceso se almacena en un fichero que contiene pares (*hora de acceso, dirección de procedencia*), existiendo un fichero individual para cada página del sitio. Tanto los ficheros de registro como el *cgi* de control se alojan en el mismo servidor web que alberga el conjunto de páginas. Esto soluciona los dos problemas que planteaba el uso de los contadores públicos: tenemos acceso completo a los datos, y el servicio se presta en las mismas condiciones que el resto servicios del sitio web.

El análisis de estos ficheros de la forma adecuada nos proporciona toda la información que necesitamos.

Simplemente contando el número de entradas registradas en cada fichero ya obtenemos el

número de visitas a la página. Un análisis detallado de cada entrada nos permitirá averiguar el número de visitas en distintos periodos de tiempo. Con esta información podemos construir el primero de los informes enumerados en los objetivos, que podemos ver en la figura 1, y que podemos actualmente consultar en la dirección <http://www.uhu.es/cgi-bin/cgi00.dat?command=getlogs>. Contiene la cuenta de accesos total, los accesos del día, del día anterior y el total de las últimas 48 horas. También está disponible esta tabla en una versión más completa que recoge los accesos durante la última semana y durante la anterior, lo que permite establecer si existe una tendencia al aumento o disminución en el número de visitas de cada página.



Figura 1. Informe de contadores de accesos

Este informe nos permite tener una idea de la cantidad de visitas que se realizan a las páginas de nuestro sitio web. Un histograma muestra visualmente la cantidad de visitas, y un franja de color rojo en cada barra indica las realizadas durante las últimas 48 horas. Esto nos da una idea rápida de la actividad reciente.

Desde este informe podemos acceder a un análisis detallado de accesos a cada página, simplemente seleccionado uno de los contadores. En este informe (figura 2) se muestra una tabla con una lista de accesos, incluyendo la fecha y hora de conexión, la dirección de Internet (dirección IP) del visitante y el nombre asociado a esa IP (nombre de dominio, si está disponible).

El análisis de este informe nos permite averiguar qué tipo de usuarios acceden a la información. La presentación del nombre de dominio asociado a cada IP nos permite averiguar

algo más sobre el sitio desde el que se accedió. Es fácil saber, a partir del nombre de dominio o dirección IP, si el acceso se realizó desde algún laboratorio de la propia Universidad, o desde otras universidades, países, etc. También podemos saciar curiosidades menos científicas, como averiguar si nuestros compañeros han visitado, por ejemplo, nuestra sección personal.



Figura 2. Informe de accesos a una página

Para mejorar el rendimiento del sistema, se mantiene una tabla que asocia cada IP con el nombre correspondiente. Esta tabla es accesible como un informe más, y nos proporciona información sobre el conjunto de visitantes que han accedido alguna vez a nuestra web. Dicha tabla (figura 3) nos sirve como paso intermedio para acceder al conjunto de visitas realizadas desde una determinada dirección IP.



Figura 3. Base de datos de direcciones IP

Este nuevo informe recoge todas las entradas de los ficheros de registro asociadas a una

dirección IP determinada, y las resume en una tabla, ordenadas por páginas o cronológicamente (ver figura 4).



Figura 4. Informe de accesos desde una dirección IP

Dado que el número de visitas desde determinados puntos de red puede ser muy alto (por ejemplo accesos desde laboratorios de prácticas) este informe se puede refinar especificando el rango de fechas entre las cuales se quieren estudiar las visitas.

Examinando el contenido de la tabla de visitas desde una dirección, podemos tener una primera idea de cuál ha sido el interés del visitante por las páginas del sitio, si ha realizado una única visita, o se conecta con una cierta frecuencia, el tipo de información que busca, etc.



Figura 5. Gráfico de una visita

Si ordenamos el informe cronológicamente y nos quedamos sólo con las visitas realizadas un día, podemos conocer exactamente cuál fue el orden en que se visitaron las páginas durante ese

día. Ese es el objetivo del siguiente informe, al que podemos acceder seleccionando una cualquiera de las visitas desde una dirección.

Se trata de un grafo que analiza la visita realizada desde una dirección IP durante un día, indicando el orden en que se han ido visitando las páginas y el tiempo que se ha permanecido en cada una, como podemos ver en la figura 5.

Este es uno de los informes más interesantes que podemos estudiar. Nos permite averiguar información muy variada, como puede ser:

- Cuál es la página de entrada al sitio web. En nuestro caso particular, se pueden considerar varios puntos de entrada teóricos: la página principal, la página principal de cada asignatura, etc. Muchas veces se accede directamente a una página de datos (por ejemplo unos apuntes o un artículo) y no a un índice. Esto nos indica que el acceso se hizo desde algún buscador o mediante un enlace guardado en la lista de preferidos.
- El tiempo que ha invertido en la lectura de cada página, lo cual nos puede dar una idea de si le interesó o no lo que allí había.
- Si al visitante le pareció interesante o no el contenido general de las páginas, en función de si visitó muchas páginas y secciones o terminó su visita rápidamente.
- El tipo de información que buscaba: sobre una u otra asignatura, investigación, proyectos, localización, información personal, etc.



Figura 6. Selección de páginas para gráfico de evolución de visitas

Otro de los informes más interesante que podemos obtener a partir de los datos registrados

es la evolución de visitas a las distintas páginas del sitio web. Se trata de obtener un gráfico en el que, para un conjunto de páginas del sitio, se muestre la evolución en el tiempo del número de visitas a esas páginas. En el sistema construido comenzamos por seleccionar el conjunto de páginas cuya evolución queremos examinar, en <http://www.uhu.es/cgi-bin/cgi00.dat?command=selegs> y el rango de fechas a considerar (ver figura 6). El sistema entonces genera los datos necesarios para trazar el conjunto de histogramas de la evolución de las visitas.

En la figura 7 podemos ver un gráfico de ejemplo en el que se han seleccionado los registros correspondientes a las páginas de las prácticas de la asignatura anual Sistemas Operativos, durante el periodo de clases del primer cuatrimestre (que coincide con el periodo de vida del sitio web). Podemos observar el desigual interés que despiertan estas páginas, por otro lado evidente, debido a dos causas: la diversidad de servicios que proporcionan, y que algunas no han llegado a tener contenidos operativos.

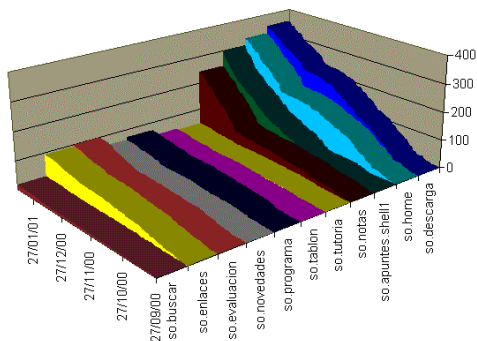


Figura 7. Gráfico de avance de visitas de la web de Prácticas de Sistemas Operativos

En la figura 8 tenemos otra vez el mismo gráfico, pero esta vez conteniendo sólo las cuatro secciones más visitadas del sitio:

- *so.descarga*: la página desde la que los alumnos pueden descargar apuntes, enunciados y otros materiales.
- *so.home*: la página inicial del sitio.
- *so.apuntes.shell1*: los apuntes utilizados en el primer cuatrimestre, en formato HTML.

- *so.notas*: la página en la que se pueden consultar las calificaciones de prácticas.

Podemos hacer las siguientes observaciones analizando el gráfico:

- La sección de notas ha tenido una actividad muy baja hasta el momento en que se comenzaron a publicar notas, pero aun sin ninguna información útil, suscitaba más interés que otras páginas (ver figura 7).
- Los apuntes en HTML tuvieron poca aceptación al principio (también se ofrecen en formato imprimible en la sección de descarga), pero más adelante empezaron a ser cada vez más visitados, hasta convertirse en la página más popular (si nos fijamos en la pendiente de la curva de visitas, vemos que actualmente es la que más rápidamente incrementa su contador).
- El acceso a la página principal sufrió un notable descenso en la frecuencia de visitas durante el periodo de vacaciones de Navidad.
- La página de descarga de apuntes y materiales también sufrió un descenso de visitas durante las vacaciones. También observamos que los visitantes acceden con mayor frecuencia a la página de descarga que a la principal, lo cual nos revela que probablemente han guardado el enlace entre sus favoritos y no tienen mayor interés en el resto de secciones.

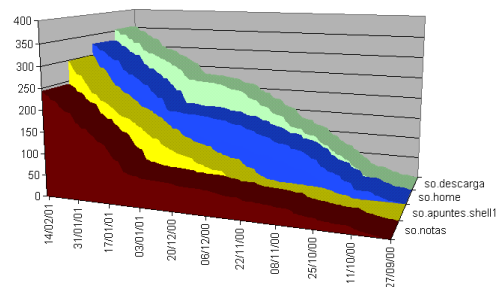


Figura 8. Avance de visitas en las secciones más activas de las páginas de prácticas de Sistemas Operativos

4. Conclusiones

El sistema construido demuestra que con sólo registrar una pequeña información sobre las visitas que se realizan a un sitio web, podemos

obtener gran cantidad de información sobre la actividad del mismo.

Por medio de la información almacenada podemos conocer la cantidad de visitas realizadas a cada página, saber quién, cuándo y desde dónde se realizaron esas visitas, rastrear la actividad de los visitantes y conocer la incidencia de diversos factores sobre la actividad del sitio. Con toda esta información podemos saber cuáles son las necesidades y costumbres de los usuarios de nuestros sitios web, más allá de lo que podemos averiguar mediante encuestas entre el alumnado.

Se trata, además, de un sistema fácilmente ampliable. Cada página que añadamos al conjunto web pasará a formar parte de las estadísticas tan pronto como incluyamos en ella una referencia invisible al sistema de registro, algo tan fácil como copiar y pegar cambiando un simple identificador. Por tanto, el hecho de realizar el registro de la actividad no afecta para nada al funcionamiento normal del sitio. En ningún caso es necesario utilizar navegadores concretos ni configuraciones especiales.

Todos los informes diseñados son accesibles actualmente desde cualquier ordenador a través de Internet, por lo que podemos considerar el sistema de estadísticas como una sección más del sitio.

La información registrada se almacena en el propio servidor web, quedando a disposición del administrador del sitio, lo cual nos permite diseñar nuevos informes en cualquier momento. También permite tener un historial detallado de las personas que han utilizado los servicios ofrecidos desde el sitio web.

5. Líneas futuras

Los aspectos futuros de este proyecto se centran en dos aspectos claves.

En primer lugar debemos continuar con la construcción y perfeccionamiento del sistema, ya que se pueden considerar gran variedad de informes que, aunque a priori parece que no van a ser especialmente interesantes, quizás revelen algunos datos curiosos o destacables que puedan ayudar a conocer las necesidades y gustos reales de los visitantes. Además, el proceso de generación del gráfico de avance de visitas está aún incompleto. Por el momento, el sistema sólo realiza la generación de la tabla de datos

correspondiente al gráfico, que se construye actualmente utilizando un programa comercial a partir de esos datos.

Por otra parte, el sistema lleva poco tiempo funcionando y todavía está en periodo de evaluación. Algunas estructuras de datos y algoritmos deberán ser optimizados para mejorar el rendimiento del sistema, que puede verse afectado por la acumulación de datos en los archivos de registro.

Aparte de estas dos líneas, queda aún pendiente mejorar la portabilidad del sistema. Actualmente sólo se utiliza dentro del ámbito de una página personal, y aunque se ha diseñado pensando en la portabilidad, por el momento ésta no puede realizarse sin algunos conocimientos de programación en C (lenguaje en el que se ha desarrollado íntegramente el sistema) [1][2][9].

Referencias bibliográficas

- [1] Brian W. Kernighan, Dennis M. Ritchie. *El Lenguaje de Programación C*. Prentice-Hall Hispanoamericana, 1991.
- [2] Byron S. Gottfried. *Programación en C (2ª edición)*. McGraw-Hill, 1999.
- [3] Daniel Philippe Chalat. *HTML y la Programación de Servidores Web*. Gestión 2000, 1996.
- [4] E. Stephen Mack, Janan Platt. *HTML 4.0*. Anaya Multimedia, 1998.
- [5] Ed. Tittel (et al.). *Fundamentos de Programación con HTML & CGI*. Anaya Multimedia, 1996.
- [6] Germán Galeano, Pablo Díaz, José Carlos Sánchez. *Manual Imprescindible de HTML 4*. Anaya Multimedia, 2000.
- [7] Isaac Cohen. *CGI/Pearl y JavaScript: Creación de Páginas HTML interactivas*. Gestión 2000, 1996.
- [8] Jerry Bradenbaugh. *Aplicaciones JavaScript*. Anaya Multimedia, 2000.
- [9] Kaare Christian. *Diccionario de C y UNIX*. Anaya Multimedia, 1990.
- [10] William Robert Stanek. *HTML, CGI, SGML, VRML, JAVA : Web Publishing: Unleashed*. Sams.Net, 1996.