

# Sistemas Estratégicos de Análisis de Audiencia en Internet

Por: Fernando Cayuela Fernández.

Empresa: REGISTRO, S.A.

- La audiencia en Internet depende de :
  - Posicionamiento
  - Usabilidad
  - Contenidos
  - Diseño
  - Rapidez.
- Su conocimiento y valoración dependen de una óptima planificación, implantación y ejecución de los "Sistemas Estratégicos de Análisis de Audiencia en Internet".

- **Sistemas de Análisis y volcado de datos.**
  - Escuchadores, marcadores y logs.
  - Sistemas de logs.
  - Volcado de datos.
- **Herramientas de extracción de los datos.**
  - Herramientas específicas.
  - Herramientas personalizadas.
- **Interpretación estadística y Filtros.**
  - Filtros de perfiles de usuario y áreas.
  - Filtros de mediciones de Impacto.

## Sistemas de análisis idóneo

Todos los sistemas de análisis tienen sus pros y sus contras.

### **Consideraciones:**

Entorno en el que tendremos el sistema.

Nivel de personalización del sistema.

Herramientas estándar:

Internas en el sistema.

Externas.

Herramientas personalizadas.

Mantenimiento.

Costes.

## Lectores de datagramas.(Packet sniffing)



Son aplicaciones que escuchan por el puerto por el que nos llegarán los accesos que queremos analizar (http: puerto 80), capturando la información relevante desde los datagramas que viajan por la red.

### Ventajas:

- .- Descargamos los servidores de procesos.
- .- Si el servicio es contratado externamente, no sobrecargamos nuestra red ni nuestros servidores con procesos de carga y análisis.

### Inconvenientes:

- .- Llegada de Datos críticos encriptados.
- .- Precio de la contratación con empresas externas.
- .- Todo cambio en la web o mejora de filtro de búsqueda debe ser notificado a la empresa contratada.



## Marcadores. (Clickstreams)

Son pequeñas instrucciones que se implantan en las páginas web, las cuales mandan la petición de acceso a esta página, a la empresa con la que hemos contratado este servicio o bien al servidor donde tenemos alojada la aplicación que nos tratará esta información (BB.DD).

En la mayoría de los casos, este código consiste en alojar una “cookie” en el PC cliente la cual se activará cada vez que entremos en la página. Su labor puede ser

- .- Simple: Un contador.
- .- Complejo: Envía los datos más completos posibles sobre este acceso (La 1ª vez por formulario).

### Ventajas:

- .- No cargamos los servidores con procesos.
- .- Si el servicio es contratado externamente, no sobrecargamos nuestra red ni nuestros servidores.

### Inconvenientes:

- .- PC cliente con “cookies” desahilitadas.
- .- Exige una actualización exhaustiva de todo el árbol web.
- .- El precio de coste.



## Sistemas de logs.

Fichero de texto con todos los eventos relativos a las funciones que cumple ese servidor en un puerto específico.

Sistema estándar más extendido:

Sistema Operativo = Microsoft Windows NT, 2000, 2003

Servidor = Internet Information Server (IIS)

Ficheros de log = extendido W3C (World Wide Web Consortium).

Ventajas:

- La información escalable y accesible.
- El coste de las herramientas estándar menor.
- Más facilidad para analizar y evaluar los datos.

Inconvenientes:

- Tamaño de los ficheros de log.
- Depuración exhaustiva de la información.
- Control horario (GMT-0).
- No contempla las sesiones de usuario.
- La herramienta debe darnos la información que el log no contempla.

Es el más extendido en pequeñas y medianas empresas o entidades.

## Sistemas de logs. (Cont.) Cabecera

#Software: Microsoft Internet Information Services 5.1

#Version: 1.0

#Date: 2004-05-03 12:31:11

#Fields: time c-ip cs-username cs-method cs-uri-stem cs-uri-query sc-status sc-bytes cs-bytes time-taken cs(User-Agent)  
cs(Cookie) cs(Referer)



Los campos principales son:

Día_____	date	Estado Win32_	sc-win32-status
* Hora_____	time	* Bytes enviados	sc-bytes
* Dirección IP del cliente	c-ip	* Bytes recibidos	cs-bytes
* Nombre de usuario	cs-username	* Tiempo consumido	time-taken
Nombre de servicio	s-sitename	Versión del protocolo	cs-version
Nombre de servidor	s-computername	Host_____	cs-host
Dirección ip del serv.	s-ip	* Agente de usuario	cs-(User-Agent)
Puerto del servidor	s-port	* Cookie_____	cs(cookie)
* Método_____	cs-method	* Sitio de referencia	cs(Referer)
* Recurso URI visitado	cs-uri-stem		
* Consulta URI solicitada	cs-uri-query		
* Estado del protocolo	sc-status		



# Sistemas de logs. (Cont.) Parametrización

Los archivos de logs del IIS estan en:

\\WINNT\system32\LogFiles\W3SVC1 (uno por cada “site”).

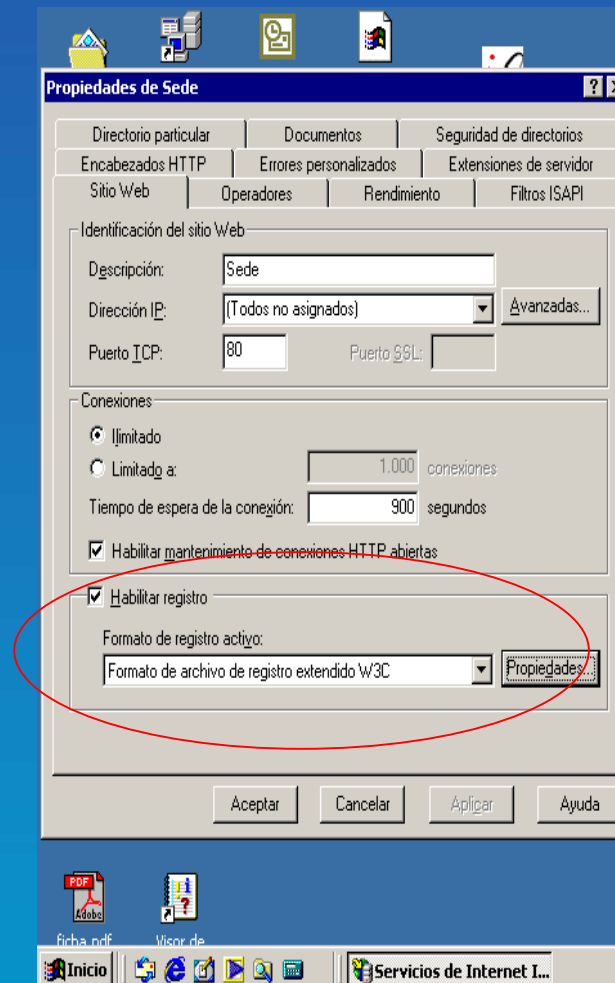
Para configurarlos accedemos a :

“Herramientas Administrativas”

“Administrador de servicios Internet”  
sobre el sitio web

pulsamos boton derecho  
opción “Propiedades”.

Habilitamos registro (“ ½”) y formato  
“Archivo de registro extendido W3C”



# Sistemas de logs. (Cont.) Parametrización

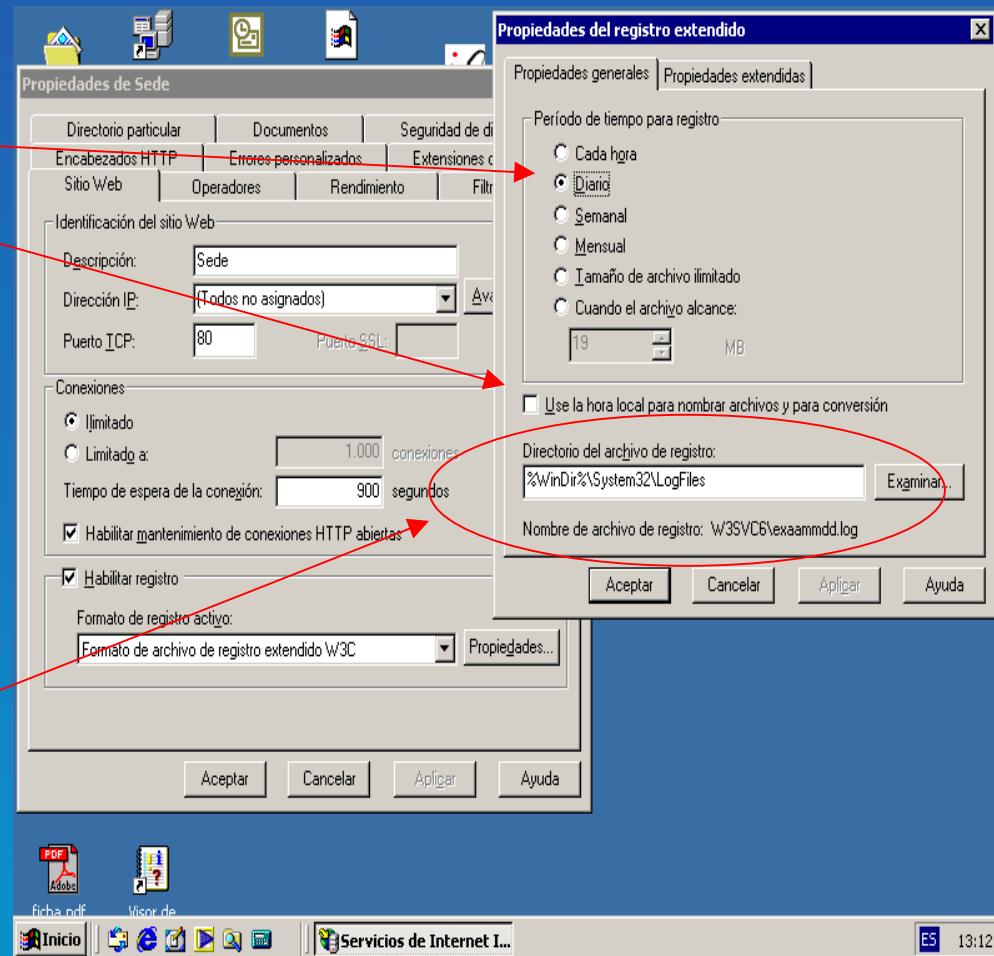


.- propiedades generales :

- periodo: Diario
- deshabilitar la hora local

El log W3C utiliza la hora "GMT:0"

Directorio de los ficheros de log

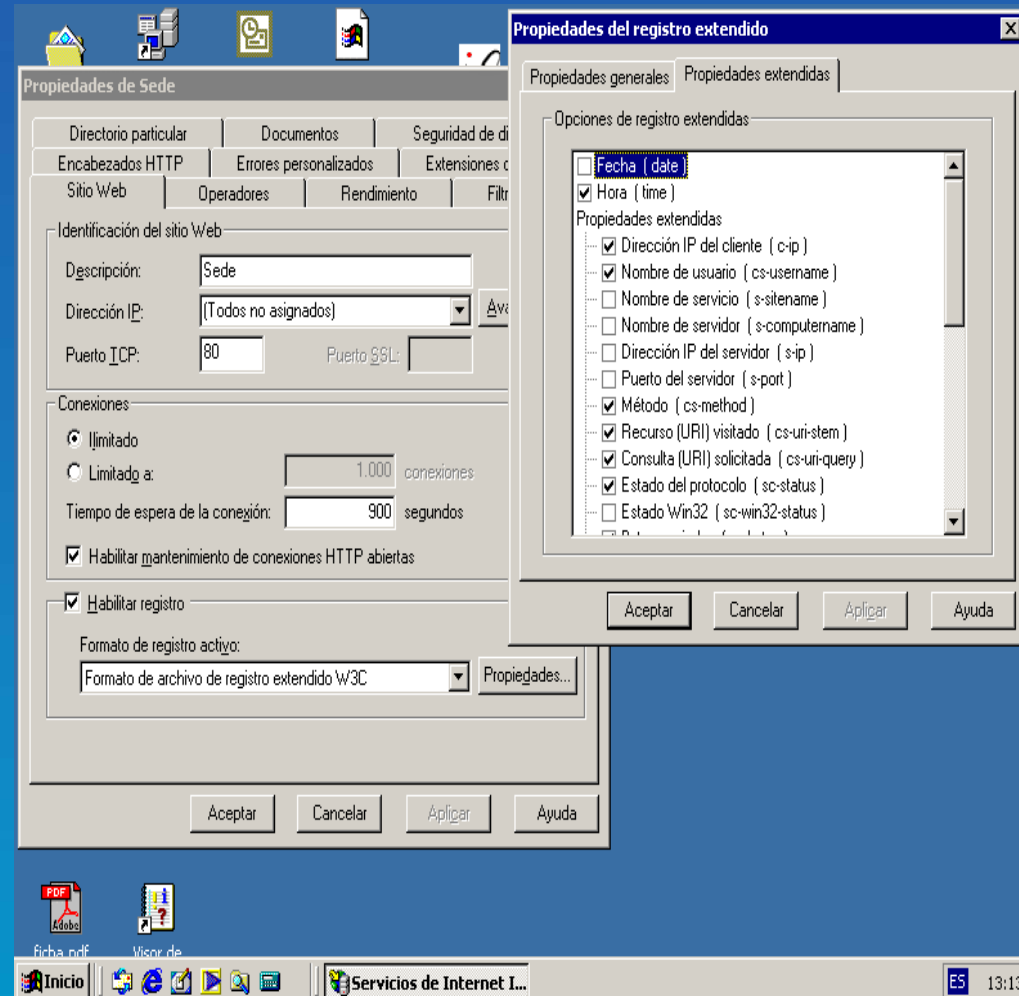


# Sistemas de logs. (Cont.) Parametrización



## Propiedades extendidas de este formato de log

.- Sólo habilitamos las opciones de registro marcadas con un asterisco en la tabla anterior en la tabla anterior



# Sistemas de logs. (Cont.) Volcado de datos



.- Por aplicaciones específicas.- (Ej:WebTrends)  
Perfiles de carga por cada sitio web. (Fastrend)

.- Por programas personalizados a BB.DD.

Tareas:

- .- Depurar la información
- .- Marcar las páginas principales
- .- Resolver las direcciones IP
- .- Separamos los accesos no validos.
- .- Minimizar el tamaño de los logs ya cargados.
- .- Calcular las visitas de los accesos válidos. Niveles:
  - .- Ip (Orden: ip, usuario, date, hora, autonumérico)
  - .- Sesión (Orden: ip, usuario, cookie, date, hora, autonumérico)

Nueva visita:

- .- Cualquier cambio secuencial según la ordenación.
- .- Siendo los tres primeros campos iguales según la ordenación pero con un margen de tiempo superior a 10 minutos.

# Herramientas de extracción de los datos.



## - **Herramientas específicas.** (Ej:WebTrends)

Por informes lanzados contra los perfiles de carga.

Ventajas:

- Con filtros de búsqueda.

- Pueden ser programados o tiempo real

- Rapidez de carga.

- Programable en el tiempo.

- Datos básicos y concretos

- Bajo coste.

Inconvenientes:

- Filtros solo para un solo "site".

- Sólo un "site" por perfil de carga.

# Herramientas de extracción de los datos. (Cont.)



## - Herramientas personalizadas.

Desarrolladas especialmente para la entidad.

Ej.: Entorno = Windows 2000 Server, IIS, Oracle, SAS o ASP

### Consultas por Cubos Olap:

Creamos los tres ejes del cubo como:

- .- Paginas
- .- Visitas
- .- Measures: Periodos temporales

En tiempo real: paginas ASP

En procesos internos (lanzados en horario nocturno)

### Consulta por SAS

- .- Consultas masivas de datos de Data Warehouse (Requerimientos de un sistema de almacenamiento para soporte a la decisión).
- .- Patrones de decisión (filtros entre tablas)
- .- Gráficos con applets de java y html para las tablas.
- .- Consultas no en tiempo real.

# Interpretación estadística y Filtros.



## **.- Filtros de perfiles de usuario y áreas.**

El perfil de usuario identifica el movimiento de un usuario en la web. Por donde ha entrado, páginas visitadas y por donde ha salido.

Filtros:

.- 1 Con cookies o acceso es autenticado.

Con una misma ip, nombre de usuario y cookie, en un margen de tiempo entre peticiones de páginas inferior a 10 minutos conseguimos conocer la navegación precisa de un usuario o de una visita.

.- 2 Sin cookies en nuestras páginas o ni acceso autenticado.

Sólo podremos contemplar un usuario o visita por la dirección ip del cliente y por su navegador, en un margen de tiempo entre peticiones de páginas inferior a 10 minutos, siendo estos datos algo imprecisos.

## Interpretación estadística y Filtros.(Cont.)



### **.- Filtros de perfiles de usuario y áreas. (Cont)**

¿Cómo ha entrado?

Por “Sitio de referencia” o “referer” y por “Recurso URI visitado”.

¿Por donde ha navegado?

Por “Recurso URI visitado” y por los parámetros que va utilizando en “Consulta URI solicitada”.

¿Por donde ha salido?

Por el último “Recurso URI visitado”.

¿Es una visita de calidad?

Evaluar:

El tiempo que ha durado la visita entre la primera página y la última.  
Si ha consultado ficheros clave como “PDF” o informes como “RPT”  
Si proviene de un buscador. (¿En que lugar estamos posicionados?)  
Si proviene del envío masivo de un boletín por correo electrónico.  
Si se ha tomado la molestia de rellenar el formulario de entrada.  
(muy positivo para la creación de una cookie).



## Interpretación estadística y Filtros.(Cont.)



### **.- Filtros de mediciones de Impacto.**

#### **.- Envío de boletines electrónicos masivo:**

.- Con página de redireccionamiento..

Ej.: href control.asp?enlace=www.cervantes.es&nombre\_usuario@dominio.es

.- Las veces que se ha abierto. Imagen pral. con otro nombre.

#### **.- Aceptación de un producto de nuestra web.**

.- Por “Recurso URI visitado” y por “Consulta URI solicitada”.

#### **.- Medición del impacto por el posicionamiento en un buscador.**

.- Filtramos el campo “referer” por el nombre del buscador.

Ej.: Google= <http://www.google.es/search?q=Instituto>

palabra buscada = Instituto

<http://www.google.es/search?q=%22Los+lunes+al+sol%22+%2B+ficha>

frase buscada = “Los lunes al sol”

palabra buscada = ficha

- Conclusiones

Una buena valoración depende de:

- Estudio en profundidad del entorno.
- Sistema de recopilación de datos.
- Análisis de los datos y filtros.
- Herramientas satisfactorias.
- Filtros concretos.
- Comparativas con otras entidades.