

## **Mejorando Decisiones de Gestión en la Medición de Audiencias de Proyectos Web**

Ruby C. Morales M.<sup>1</sup> Sergio Troncoso U.<sup>2</sup>, Darío Barra S.<sup>2</sup>

<sup>1</sup> Universidad de La Frontera, Avda. Francisco Salazar 01145 Casilla 54-D, Temuco, Chile

<sup>2</sup> Datacare Ltda., Pablo Neruda #01765 Temuco, Chile  
rmorales@ufro.cl, sergio.troncoso@datacare.cl,  
dario.barra@datacare.cl

**Resumen.** El artículo presenta un acercamiento a la metodología de análisis de audiencias de un proyecto Web que aporta suficientes elementos para realizar decisiones de primer nivel de complejidad en relación a la gestión de un proyecto Web. Asociado a esta idea se presenta una experiencia de selección entre herramienta de análisis de archivos Log realizado entre cuatro herramientas, dos propietarias y dos de código abierto. La experiencia exigió la adaptación y creación de parámetros de medición para facilitar el análisis comparativo entre herramientas. La metodología generada para esta comparación, así como las variables utilizadas y sus pesos en el estudio, pueden orientar para su uso, gestión e inversión a otras empresas, clientes y gestores de proyectos Web interesados en la medición audiencias.

### **Introducción**

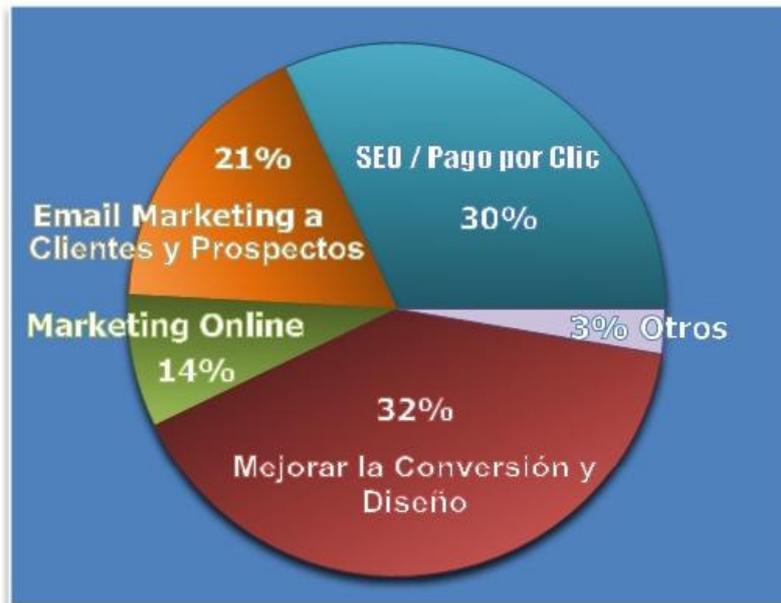
Desde la perspectiva de una empresa prestadora de servicios Web y desde la perspectiva de un propietario / cliente de un proyecto Web, es necesario contar con elementos de juicio para tomar decisiones de mejoramiento del o los servicios o prestaciones del o a los proyecto Web y otras inversiones. Este es un ejercicio periódico asociado con el ciclo de vida de un proyecto y en general, estas decisiones de gestión e inversión están orientadas por el objetivo del proyecto Web, sean estos con o sin fines de lucro, en el entendimiento que si se está publicado en Internet hay un compromiso

implícito con el cliente empresa / navegante de dar el mejor de los servicios de calidad.

Se entiende por proyecto Web aquel conjunto de código ejecutable que esta alojado en un servidor Web y que sirve de interfase entre el visitante y el servidor. Este conjunto de código ejecutable puede representar desde una simple página de información en la Web hasta un sistema complejo como un portal o un sistema de e-Commerce. En el contexto del artículo los componentes y niveles de complejidad que representan el contenido de un proyecto Web no inciden en el análisis de la medición de audiencias, por lo que se consideran estas componentes como una caja negra, solo se trabaja con la primera capa, aquella que está asociada directamente a la interfase con la que actúa el visitante de un proyecto Web. Desde el punto de vista del propietario de un proyecto Web, esta primera capa asociada a la interfase con la que actúa el visitante del proyecto Web es de vital importancia para ayudarlo a decidir sobre inversiones y estrategias de gestión del proyecto Web, gracias a las mediciones de audiencias.

Estos elementos de juicio en los que basar las decisiones de gestión / inversión de un proyecto Web están asociados a factores que pueden ser críticos, dependiendo del tipo de proyecto, tales como número de visitas, elección de productos, procedencia de los visitantes, contenidos populares, satisfacción de los usuarios, niveles de usabilidad, diseño de la página, retorno de la inversión de las campañas publicitarias [1].

Existen diferentes enfoques para realizar la definición de estos factores que serán la base de los elementos de juicio que sustentarán la toma de decisión sobre la gestión e inversión de proyectos Web. En la Figura 1, se muestran las principales áreas en que los responsables de un proyecto Web evalúan después de un análisis estadístico de audiencias para hacer sus inversiones o mejoramiento del proyecto Web en cuestión.



**Fig 1.** Principales áreas de evaluación para la inversión (adaptado de [2])

La conversión es la principal área de inversión para el tipo de clientes que atiende la empresa Datacare Ltda., en base a estos requerimientos se seleccionaron los software de prueba que pudieran proveer análisis de audiencia en esta área.

“La conversión es la habilidad que tiene un proyecto Web de hacer que el visitante ejecute la acción deseada.”[3]. Los pasos para el análisis de la conversión están basados en cinco aspectos: Entrada, Salida, Abandono, Desvío y Saltos, donde se identifican desde donde llegan los visitantes, en que punto dejan el proceso, cuando salen del proceso y si se van a su vez del sitio, si se distraen y caen en otras paginas (link) del sitio Web, se identifican las páginas a las que acceden, y se registra los saltos en las etapas, dado que no todo visitante sigue una visita lineal.

En general los factores que determinan el comportamiento de usuario visitante en un proyecto Web, y que el gestor necesita conocer son tres, los contenidos disponibles, los links relacionados y el diseño del sitio Web [3], [11].

Desde el punto de vista de los elementos asociados a un sistema Web, se tienen al menos tres enfoques de medición de audiencias en la Web.

El enfoque orientado al usuario (user-centric) basado en un panel de navegadores de una población de objetivo observada, monitorea a través de programas instalados en el computador del navegante, horarios y la frecuencia de usos. Este enfoque facilita establecer ranking de los sitios más visitados por una determinada población con cierto perfil demográfico de los usuarios de la Red. Por ejemplo se puede saber cuáles son los sitios que prefieren las mujeres v/s los hombres, o los adolescentes v/s los adultos. [4], [5], [12].

El enfoque orientado al sitio Web (site-centric) permite datos sobre la tendencia del tráfico, tecnología que usa el visitante y perfil del visitante pudiendo identificar hora y día con mayor número de visitas, que sistema operativo y que navegador usa el visitante, entre otros.

El enfoque orientado a la publicidad (Ad-Centric) estudia el tráfico de los banners y publicidad en línea. Recoge información útil para los avisadores, las agencias de publicidad y de medios y las centrales de publicidad que gestionan el avisaje de todo un conjunto de sitios Web.

Varios de estos sistemas usan cookies para saber el perfil demográfico de los usuarios y, de esa manera, dirigir su publicidad en función de ese objetivo. Recientemente, la Comisión Europea propuso una ley para prohibir el uso de cookies justamente porque cree que estos sistemas abusan de la privacidad de los usuarios. [4]

Los nombrados agrupan los enfoques generales que permiten dar una perspectiva de los factores que pueden ser de interés para un gestor de un proyecto Web o para una empresa de servicios Web pueda proveer, para otros y para si misma.

Saber quien y que, cuando y donde utiliza un sitio, “mi” sitio, permite al gestor tomar decisiones informadas sobre que hacer, donde y, o, como pedir, que enfatizar, cuando vender, cuando invertir, cuando hacer mantenimiento, etc. Le permite conocer a su mercado objetivo, la dinámica del como y el perfil de su mercado objetivo. En otras palabras retroalimentarse en forma dinámica y ajustar la marcha de su negocio / servicio / proyecto Web de acuerdo al comportamiento de la audiencia o influenciar en ella y medir los cambios.

Existen otros enfoques para realizar las mediciones de audiencias, la mayoría proviene de los estudios de marketing y de la medición de audiencias de los medios de comunicación como los utilizados en los proyectos internacionales World Internet Project (WIP) y Business and Information Technologies Project (BIT), [5], o como los utilizados en la minería Web (Web Mining) [6], pero todos comparten la base común de recolectar datos sobre el comportamiento del visitante, que medios usa y de donde viene, cuanto tiempo se queda y tratar de colegir su motivación y si los propósitos que guían los contenidos del proyecto Web se cumplen y en base a los datos recolectados y una vez analizados tomar las decisiones pertinentes.

La experiencia de seleccionar una herramienta de análisis de la medición de audiencias en base a archivos Log para mejorar la gestión / inversión de un proyecto Web, compartida a continuación refleja aproximadamente un cuarto del potencial que estas herramientas de medición de audiencia, ponen a disposición de un gestor de proyectos Web. Se compartirá el enfoque elegido y las razones que llevan a esta elección, la metodología diseñada para elegir una herramienta que aporte las mejores prestaciones de acuerdo a los requerimientos básicos establecidos y se concluirá con las recomendaciones decantadas de esta experiencia.

### **Análisis de Archivos Log**

Otros enfoques disponibles para la medición de audiencia son los sistemas de análisis de Logs y los contadores y trackers. [4]

Para el caso que se comparte en este artículo se eligió el análisis de archivos Logs. Técnicamente un archivo Log de un proyecto Web es un archivo creado por el servidor donde está instalado el proyecto y donde se registran las acciones que los navegadores realizan en la Web. Un ejemplo de archivo Log se ve en la Figura 2.

```
200.66.196.132 - - [03/Mar/2002:23:59:20 +0100] "GET
/directorio/16.html HTTP/1.1"
304"http://www.dominio.com/directorio/index.html"
"Mozilla/4.0 (compatible; MSIE 5.5; Windows 98; Win 9x
4.90)"
```

**Fig 2.** Ejemplo de un registro de un archivo Log de un servidor apache

A pesar de que no permite disponer de estadísticas en tiempo real por sí mismo, el mayor problema con esta tecnología es la identificación de las visitas únicas. Para identificar un usuario, el sistema se basa en el número de IP de donde vienen las visitas. En empresas grandes con más 100 empleados por ejemplo, todos ellos tienen la misma salida de Internet compartiendo un mismo número IP. Según el sistema de análisis de Logs, todos esos empleados son como un solo usuario. Al revés, varios proveedores de acceso Internet funcionan distribuyendo números de IP dinámicos a sus usuarios, es decir, un mismo usuario puede entrar en un mismo sitio dos veces usando dos números de IP distintos. [3]. Estos dos extremos generan las dificultades mencionadas que hacen del análisis de Log una técnica imprecisa al momento de calcular la cantidad exacta de visitas realizadas al proyecto Web. Pero sí se puede evaluar una tendencia dentro del proyecto Web respecto de las visitas y su procedencia IP.

Este tipo de archivo Log en su forma original presenta un contenido no procesado por lo que dificulta su manejo, y es necesario disponer de herramientas que permitan interpretar su contenido adecuadamente de acuerdo a algún propósito predefinido.

Estos registros Log guardan entre otros datos, la dirección IP del usuario, la fecha y hora de acceso, el URL de la página accedida, el protocolo utilizado para la transmisión de los datos, el código de error, el número de bytes transmitidos [6].

Cada vez que se produce una llamada a un archivo dentro de un proyecto Web, el servidor grabará un registro como el mostrado en la figura

anterior. Cada visita, cada HTML, cada imagen dentro de ese HTML generará un registro. Esto es lo que hace que el archivo Log sea difícil de manejar en su formato original.

Aunque estos archivos no están definidos para ser leídos directamente, tienen la doble ventaja que están presentes en todos los sistemas operativos y que son generados sin esfuerzo. Una vez definido su funcionamiento generará en forma automática, por lo tanto sin costo adicional, la bitácora del sistema, la que podrá ser accedida en el momento oportuno de acuerdo a los requerimientos del gestor.

El análisis de archivos Log se facilita utilizando software especializado para leerlos y generar los informes estadísticos definidos a través del perfil que estos software permitan conformar y que también podrán ser presentados en forma gráfica para facilitar su entendimiento.

En el caso que se comparte en este artículo fueron elegidos para participar en la prueba de selección dos software propietarios y dos software de código abierto, WebTrends 4.5, WebTrends 8, AWStats 6.8 y Webalizer 2.20 respectivamente.

### **Metodología de Selección**

Se eligió un proyecto Web que contiene los elementos a ser medidos y del cual se disponía de un respaldo de archivos Log que cubrían en forma extensa 10 días. Este proyecto por su complejidad representa a la mayoría de los proyectos Web atendidos por Datacare Ltda.

Se crearon los ambientes de prueba en los respectivos servidores de acuerdo al sistema operativo que cada software requería.

Se instalaron los software elegidos para las pruebas. Se definieron cada uno de los software con los factores elegidos que en conjunto son similares en cada de las alternativas que participan en la selección.

Se realizaron las pruebas en este ambiente controlado, cuidando de mantener las equivalencias para cada una de las alternativas. El ambiente de prueba construido se puede observar en la Figura 2.



**Fig 2.** Ambiente controlado de pruebas de software de análisis de archivos  
Log

Una vez preparados los ambientes de prueba, se realizaron las simulaciones aplicando el software de análisis de archivos Log en forma secuencial, generando cada uno de ellos el informe de análisis de audiencia correspondiente al mismo proyecto durante el mismo lapso de tiempo, pues se usó el mismo archivo Log sobre el cual actuaron los cuatro software en la prueba. El archivo Log del proyecto Web de prueba pesaba 4.4 GB.

Los atributos de configuración de cada uno de los softwares de análisis de archivos Log fueron elegidos en base a los estándares de calidad de software ISO/IEC 9126 [7], con objeto de tener un respaldo de metodologías probadas y maduras en la etapa de prueba. Los seis atributos utilizados son funcionalidad, fiabilidad, usabilidad, eficiencia, mantenibilidad, y portabilidad [8], definiendo los valores de acuerdo a la experiencia y las expectativas de mejoramiento de los software en uso respecto de los probados. Para la identificación de los requerimientos de calidad y los pesos que representan la valoración comparada entre los distintos atributos y sub-atributos se utilizó la calificación relativa basada en valores entre 1 y 10. [9].

El enfoque general que primó en las decisiones en las etapas del proyecto es el aportado por Garvin y sus cinco perspectivas, visión trascendental, la del usuario, la del productor, la del producto y la del valor [9].

## Resultados de la Simulación y Prueba

Una vez realizadas todas las pruebas se ordenaron los datos obtenidos en la Tabla 1. Por motivos de confidencialidad no se puede identificar explícitamente el nombre del proyecto Web elegido para realizar las pruebas de simulación, por lo que es llamado [www.10x.cl](http://www.10x.cl).

**Tabla 1.** Resultados por herramientas de análisis Log del proyecto Web del tipo complejo “www.10x.cl”

<b>Informe Resumen</b>	<b>WebTrends v 4.5</b>	<b>WebTrends v 8</b>	<b>AWStats v 6.8</b>	<b>Webalizer v 2.2</b>
<b>Resumen Visitas</b>				
Visitas	118.931	127.641	120.130	140.663
Promedio x día	9.910	10.636	3.828	11.721
Duración media de la visita	0:15:45	0:13:24	-	-
<b>Visitantes</b>	-			
que realizaron una sola visita	-	67.622	71.069	79.794
que realizaron más de una visita	-	18.998	-	-
<b>Cantidad</b>				
aciertos por página de bienvenida	182.548	-	130.422	130.424
Cantidad total de aciertos exitosos	18.620.002	16.784.670		158.30.968
<b>Accesos</b>				
total de accesos	-	18.483.171	18.481.137	20.421.790
Cantidad total de accesos fallidos		1.698.501	1.723.024	1.756.686
<b>Trafico</b>				
Trafico	-	75 GB - Pic.	133,18 GB	165,5 GB

La Tabla 1, muestra datos comparativos de las herramientas de análisis de archivos Log obtenidos de los informes resumen o general del proyecto Web analizado. Se puede observar que no todas las herramientas analizan los mismos puntos, y los puntos en común los valores obtenidos no están tan apartados unos de los otros. Las diferencias de los valores observados en los puntos comunes se deben básicamente a como reconoce e interpreta cada herramienta los datos del archivo Log y a los algoritmos de calculo que utilizan. Respecto de este tema, las cuatro herramientas probadas usaron algoritmos distintos.

Lo anterior es importante de tener en cuenta cuando las diferencias son significativas por que puede llevar a errores de apreciación si el tomador de decisiones no está informado, en particular en el caso de migración de herramientas, ya que es costumbre aunque sea una versión nueva de la misma herramienta que se esta utilizando, en el proceso de mejoramiento continuo, también afectan al mejoramiento de los algoritmos, como resultó en este caso con las dos versiones de WebTrends v 4.5 y v 8. La Tabla 2 muestra el resumen obtenido desde el punto de vista de la calidad del software.

En la tabla en la primera columna se utilizan números para codificar por problema de espacio y legibilidad los atributos medidos. 1:Funcionalidad; 2:Fiabilidad ; 3:Usabilidad; 4:Eficiencia; 5:Mantenibilidad ; 6:Portabilidad. La evaluación de la calidad de las herramientas fue realizada según los criterios de implementación y comportamiento utilizando una escala de 1 a 10 para su apreciación de acuerdo a la encuesta de experiencias realizada a un panel de especialistas del sector. Como resultado de esta consulta la escala de 1 a 10 quedo conformada como lo muestra la Tabla 3.

Criterios basados en la experiencia de un panel de ingenieros especialistas en la administración de las herramientas probadas concordaron en los valores para las sub atributos de los atributos de calidad de acuerdo a la norma ISO/IEC 9126 como se muestra en la Tabla 2.

**Tabla 2.** Calidad de las herramientas probadas  
Calidad Externa e Interna

<b>Calidad Externa e Interna</b>							
	<b>Peso %</b>	<b>Subcaracterística</b>	<b>Peso %</b>	<b>W.T. v 4.5</b>	<b>W.T. v 8.</b>	<b>AW v 6.8</b>	<b>W v 2.2</b>
<b>1</b>	21%	Aplicabilidad	39%	7	10	6	6
		Precisión	36%	6	8	6	7
		Interoperabilidad	7%	7	10	6	5
		Seguridad	8%	6	9	5	6
		Conformidad de funcionalidad	10%	7	9	8	8
<b>2</b>	22%	Madurez	40%	8	10	8	8
		Tolerancia a fallos	36%	6	8	7	6
		Recuperabilidad	14%	10	10	9	8
		Conformidad de fiabilidad	10%	8	9	9	7
<b>3</b>	15%	Entendibilidad	29%	6	9	9	8
		Facilidad de aprendizaje	23%	5	7	9	7
		Operabilidad	27%	7	9	8	7
		Atractividad	11%	7	10	9	6
		Conformidad de usabilidad	10%	7	9	9	7
<b>4</b>	18%	Comportamiento en el tiempo	41%	7	10	6	8
		Utilización de recursos	35%	6	6	8	9
		Conformidad de eficiencia	24%	9	9	7	8
<b>5</b>	16%	Analizabilidad	21%	8	10	9	9
		Cambiabilidad	23%	5	5	9	7
		Estabilidad	21%	9	9	8	8
		Testeabilidad	27%	7	9	9	8
		Conformidad de facilidad de mantenimiento	8%	9	9	9	8
<b>6</b>	8%	Adaptabilidad	25%	7	7	8	7
		Inestabilidad	25%	3	5	8	9
		Co existencia	25%	7	9	9	9
		Reemplazabilidad	13%	5	5	9	9
		Conformidad de portabilidad	13%	6	6	9	8
<b>Total de Valores Según Calidad</b>				<b>185</b>	<b>226</b>	<b>216</b>	<b>203</b>

**Tabla 3.** Valoración basada en la experiencia

Deficiente	Suficiente	Adecuado
1 -4	5 -7	8 - 10

Al no existir criterios preestablecidos para las sub atributos relacionadas con los software de análisis de archivos Log, se creó la escala de valoración de la calidad interna del 1 al 10 como se muestra en la Tabla 3. Por asunto de espacio se explica los rangos mostrados en la Tabla 3 a través del siguiente ejemplo. Para el sub atributo de seguridad el rango del 1 - 4 quiere decir que: es muy probable que los informes puedan ser violados y alterado su contenido. El rango Suficiente, que va de 5 – 7, quiere decir: puede que la seguridad del software tenga que ser complementada con la seguridad del servidor tras la ayuda de firewall. En el rango Adecuado que va de 8 – 10, quiere decir: el software es lo suficientemente capaz de rechazar un ataque sin la ayuda del sistema operativo u otro software de apoyo. Los valores de los rangos Suficiente y Adecuado representan las posiciones de escasamente suficiente (suficiente -) y escasamente adecuado (adecuado -) o el valor 5 y 8 respectivamente. Siendo los valores intermedio de cada rango el que corresponde a (+- suficiente y +- adecuado) o 6 y 9 respectivamente, completándose la valoración con suficiente y adecuado o 8 y 10 respectivamente. Las apreciaciones fueron consensuadas por el panel de especialistas de acuerdo a su experiencia con las herramientas. Los valores asignados para cada sub atributo de la calidad interna, están resumidos en la Tabla 2.

Los valores obtenidos según los niveles de calidad informados en la Tabla 2, muestran que el mejor comportamiento lo tienen las herramientas de análisis de archivos Log, WebTrends v.8 y AWStat v.6.8 con un puntaje de 226 y 216 respectivamente.

La elección de la herramienta a utilizar está en principio entre un software propietario, Webtrends v.8 y un software de código abierto, AWStat v.6.8.

El siguiente paso fue utilizar una tabla de comparaciones técnicas adaptada a las necesidades actuales y futuras de la empresa, donde se seleccionaron los aspectos técnicos que ya existen y se les agrego un

subconjunto de aspectos técnicos mínimos esperables para un cambio de tecnología. Además, se modificó la tradicional marca cumple o no cumple utilizada, una equis y casilla en blanco, por valores numéricos 1 y 0, para cumple o no cumple. De esta manera se puede utilizar en dos formas la tabla de comparaciones técnicas. Una forma visual para saber cuales son los aspectos técnicos disponibles o no en las herramientas comparadas y la forma numérica, donde se suman las ocurrencias de las características técnicas esperadas de las herramientas en comparación.

Para continuar con el análisis comparativo se intentó crear una tabla que midiera las características que cumplían al mismo tiempo todas las herramientas en el proceso de prueba de selección, obteniéndose la Tabla 4.

**Tabla 4.** Comparación y evaluación de aspectos técnicos

<b>Características / Capacidades Herramientas</b>	<b>Propietarias</b>	<b>Libres</b>
Archivo de caché de DNS	x	x
Análisis Archivos de registro Web/ FTP/ Mail	x	x
Informe número de "humanos" visitas	x	x
Estadísticas se basan en visitas	x	x
Informe de hosts	x	x
Informe de horas punta	x	x
Informe día de la semana	x	x
Informe navegadores (NB detectado)	x	x
Informe de Frases clave motores de búsqueda	x	x
Estadísticas diarias	x	x
Estadísticas mensuales	x	x
Estadísticas anuales	x	x
Gráficas estadísticas en una página	x	x

En esta etapa se aplicó esta tabla de comparación técnica modificada a las dos herramientas que quedaron en la etapa final del proceso de selección, WebTrends v.8 y AWStat v.6.8 como se observa en la Tabla 4. Los

valores de cumplimiento técnico para las dos herramientas en comparación son similares.

Ante prestaciones similares de ambas herramientas, entra a primar las consideraciones económicas. Para el caso WebTrends v.8 el costo de instalación es de 4 UF (2 UF x 2 H/H) más el valor de la licencia que se paga anualmente con un precio relativo al número de páginas vistas estimadas a priori [10]. En el caso de AWStat v.6.8 el costo de instalación es de 1UF (1 UF x 1 H/H) más cero valor de la licencia porque es software de código abierto. Los otros aspectos son los de mantenimiento, y en ambos casos se requiere de personal especializado complementándose éstos por las actualizaciones provistas por la propietaria y la necesidad de mantener el software de código abierto por la empresa usuaria. Un resumen de los elementos descritos se muestran en la Tabla 6 y 7.

Utilizando como guía las tablas de comparación técnicas provistas por el software propietario se creó la Tabla 5 que muestra las características y capacidades que son comunes a las herramientas de análisis Log comparadas. Estas caracterizas y capacidades las proveen tanto los software propietarios como los software de código abierto.

En la Tabla 5 se observa la comparación técnica de realizada entre las dos herramientas que pasaron todas las pruebas y están en la etapa de selección final, una de ellas es un software propietario y la otra es un una herramienta código libre. Se observa que las dos cumplen los criterios técnicos seleccionados, que en conjunto abarcan generosamente las necesidades de los clientes de las empresas y sus proyectos Web.

Como apoyo a la toma de decisión de la selección del software de análisis de archivos Log, se recopilaron datos sobre los aspectos de asociados al costo de utilización del software como se observa en la Tabla 6 y Tabla 7.

Estas tablas fueron conformadas recabando información desde las casas de software propietario y desde el entorno profesional consultados en cuanto a los valores hora hombre para profesionales que trabajan con software propietario y para profesionales que trabajan con código abierto.

Los tiempos estimados corresponden a información de referencia del software propietario y a las estimaciones producto de la experiencia. Estos valores pueden variar, si los tipos de proyectos Web son de mayor complejidad que los tomados como referencia en este estudio, así como la capacidad de carga y trabajo y número de servidores disponibles para hacer el trabajo.

Baste decir que existen proyectos Web cuyos Log pueden colapsar los servidores en horas de máximo tráfico.

Considerando los costos y las necesidades cubiertas, la decisión de la elección de la herramienta de análisis de archivos Log en esta experiencia, se inclinó por la herramienta de código abierto AWStat v.6.8.

**Tabla 5.** Características e informes compartidos por un software propietario y uno de código abierto.

<b>Características/Capacidades/ Informes</b>	<b>WebTrends v8 AWStats v6.8</b>
Informes de remitentes y dominios	<b>X</b>
Informes de motores de búsqueda	<b>X</b>
Informes de visitas	<b>X</b>
Informes geográficos	<b>X</b>
informes de búsqueda en el sitio	<b>X</b>
Informes de páginas y archivos	<b>X</b>
Informes de campaña de correo electrónico	<b>X</b>
Idioma	<b>X</b>
Informe disponible en todas las plataformas	<b>X</b>
Funciona con Apache común (CLF) Formato de registro	<b>X</b>
Funciona con IIS (W3C) Formato de registro	<b>X</b>
Funciona con formato de registro personalizado	<b>X</b>
Analizar Web / FTP / archivos de registro de correo	<b>X</b>
Actualización de las estadísticas	<b>X</b>
Interior búsqueda DNS inversa	<b>X</b>
Archivo de caché de DNS	<b>X</b>

<b>Características/Capacidades/ Informes</b>	<b>WebTrends v8 AWStats v6.8</b>
Informe número de "humanos" visitas	<b>X</b>
Informe único "humano" de visitantes	<b>X</b>
Informe duración período de sesiones	<b>X</b>
Informe de países	<b>X</b>
Nombre de dominio	<b>X</b>
Informe regiones (EE.UU. y Canadá estados)	<b>X</b>
Informe de las principales ciudades y regiones de los países	<b>X</b>
Informe ISP	<b>X</b>
Informe de hosts	<b>X</b>
Informe a los usuarios autenticados	<b>X</b>
Informe / filtro de robots (NB detectado)	<b>X</b>
Informe de horas punta	<b>X</b>
Informe día de la semana	<b>X</b>
Informe de la mayoría de los casos vistos páginas	<b>X</b>
Informe páginas de entrada	<b>X</b>
Informe páginas de salida	<b>X</b>
Informe de directorio de páginas	<b>X</b>
Informe páginas con el último tiempo de acceso / tamaño medio	<b>X</b>
Informe de compresión de estadísticas web (mod_gzip, mod_deflate)	<b>X</b>
Informe de tipos de archivo	<b>X</b>
Informe navegadores (NB detectado)	<b>X</b>
Informe sobre las versiones de los navegadores	<b>X</b>
Informe tecnología apoyada por el navegador para Java / Flash / PDF	<b>X</b>
Informe formato de audio apoyados por el navegador para Real / QuickTime / Mediaplayer	<b>X</b>
Informe de los motores de búsqueda utilizados (NB detectado)	<b>X</b>
Informe palabras / frases clave utilizadas en los motores de búsqueda (n ° detectado)	<b>X</b>
Informe externo la página web de referencia con o sin consulta	<b>X</b>
Informe sobre errores HTTP	<b>X</b>
Informe 404 errores	<b>X</b>
Estadísticas diarias	<b>X</b>
Estadísticas mensuales	<b>X</b>
Estadísticas anuales	<b>X</b>
Las estadísticas de exportación a PDF	<b>X</b>

**Tabla 6.** Elementos asociados al costo de utilización de software propietario

<b>H/H Certificado Webtrends</b>	<b>Valor Hora / UF</b>	<b>Horas Estimadas</b>
Instalación	2 UF	2 Hrs.
Configuración	2 UF	2 Hrs.
<b>Seguimiento</b>		
Monitoreo	1,5 UF	1.5 Hrs.
Mantenimiento	1,5 UF	-
<b>Interpretación</b>		
Interpretación de la Información	2 UF	40 Hrs.

**Tabla 7.** Elementos asociados al costo de utilización de software código abierto

<b>Horas Hombre</b>	<b>Valor Hora / UF</b>	<b>Horas Estimadas</b>
Instalación	1 UF	1 Hrs.
Configuración	1 UF	1 Hrs.
<b>Seguimiento</b>		
Monitoreo	0,5 UF	1.5 Hrs.
Mantenimiento	1,5 UF	1 Hrs.
<b>Interpretación</b>		
Interpretación de la Información	1 UF	45 hrs.

En la elección final además del aspecto económico primó las capacidades de independencia que provee el software de código abierto, su peso y facilidad de manejo además que cubre sobradamente las necesidades de la empresa.

Un ejemplo de lo dicho es que un software propietario de esta naturaleza tiene un peso aproximado de 300Mb y por su propia definición cubre todas las posibilidades, lo que se puede transformar en una debilidad cuando se trata de personalizarlo, especialmente al compararlo con un software de similares prestaciones (ver tabla 5) que pesa 20Mb. En lo que se refiere a la capacidad de análisis de archivos Log, los software

propietarios exceden las necesidades comparados con la capacidad de software código abierto elegido. Si la situación fuera más compleja como análisis de la medición de audiencias en línea, podría revertirse la situación.

## **Conclusiones**

La experiencia compartida en este artículo destaca el trabajo de selección donde se utilizaron estándares y experiencias disponibles, adaptando a la realidad de la empresa las comparaciones técnicas y el peso de las valoraciones en las subcategorías de los atributos en la tabla de calidad del Software.

El trabajo realizado permite aportar a la comunidad y a la empresa una metodología de evaluación de software de análisis de archivos Logs, que puede ser utilizada en cualquier momento de los procesos continuos de mejoramientos.

El caso presentado en este artículo puede ser replicado, y pueden ser utilizados con varios proyectos Web para generar los datos que permitan el análisis comparativo, pero la experiencia dice que basta elegir bien un solo proyecto Web, que sea complejo, es decir que cubra la medición de todos los atributos que se acostumbran a utilizar en otros proyectos, es decir, que en sí tenga la concatenación de todos los atributos que necesiten ser medidos por proyectos de menor o mediana complejidad, de tal manera que cubra en forma natural varios proyectos de distinta complejidad, y del cual se tengan archivos Log de un período razonable, por ejemplo 3 días como mínimo. De esta manera se podrá simular los análisis de los archivos Log en tiempo real o en tiempo diferido de acuerdo a las características de las herramientas de software a ser probadas.

La experiencia mostró que no se necesitan disponer de archivos Log en un rango mayor de 3 días, incluso si se cuenta con un proyecto Web con medición completa, bastará con disponer de un archivo Log de 1 hora, porque el propósito de la comparación no es medir la audiencia del proyecto Web, si no medir el comportamiento de análisis que presta las distintas herramientas de análisis de archivos Log que están siendo probadas.

Si se tratara de una empresa nueva, o se tratara de un primer proceso de selección de software de esta naturaleza, es factible conseguir un archivo Log de un proyecto Web equivalentes a los que se darán servicio con la herramienta que será seleccionada, para realizar las pruebas de análisis comparativo.

En el proceso de toma de decisión de la selección del software adecuado para la empresa, se dejó como último elemento de juicio al factor costo, para disponer de otros elementos de juicio antes de aplicar el factor costo para la selección, porque es sabido que al considerar las herramientas de código abierto, se tiende a olvidar los costos de mantenimiento, que por la especialización del o los profesionales a cargo y por la rápida migración de éstos entre empresas, el costo de un software de código abierto puede llegar a ser superior al costo inicial de un software propietario en el mediano largo plazo.

Desde la perspectiva del tomador de decisiones, el trabajo pone a su disposición elementos de juicio, y antecedentes para solicitarlos, ya sea éste gestor de una empresa prestadora de servicios a proyectos Web, o sea un gestor de un proyecto Web en particular.

Finalmente el trabajo también deja entrever los problemas éticos del uso de estas herramientas de medición de audiencias, cuando si no se tiene el cuidado necesario en la selección de los factores a medir, se puede fácilmente trasgredir los límites de la privacidad de los visitantes, comprometiendo derechos, leyes y responsabilidad social empresarial.

Si se entiende bien el problema, a los cibernautas, a los mercados objetivos, no es necesario traspasar los límites privados en la medición de audiencia, porque se puede llegar a conclusiones similares, mejorando los análisis estadísticos de los datos capturados en los procesos de la medición de audiencia.

Este un tema que queda abierto, porque dependerá del objetivo del estudio, la selección de elementos a medir teniendo en cuenta el elemento confianza que cada proyecto Web y sus gestores ofrecen a sus visitantes.

## Agradecimientos

Este trabajo a sido desarrollado gracias a la colaboración de la empresa Datacare Ltda. y al Departamento de Ingeniería de Sistemas de la Universidad de La Frontera, quienes han provisto de las herramientas técnicas, recursos y financiamiento.

## Referencias

- [1] e-interactive Webanalytics. Introducción Analítica Web.  
<http://www.web-analytics.es/analiticaweb/introduccion-analitica-web.pdf>. Agosto,2008
- [2] [http://www.isoftland.com/software/webtrends/analytics/3/@http://www.isoftland.com/docs/webtrends/analytics/WebTrends\\_ComoPotenciarTraficoWeb\\_es.pps](http://www.isoftland.com/software/webtrends/analytics/3/@http://www.isoftland.com/docs/webtrends/analytics/WebTrends_ComoPotenciarTraficoWeb_es.pps)
- [3] The iProspect Search Engine User Attitudes Survey, Developed by iProspect, WebSurveyor, Strategem Research, and Survey Sampling International. March 2004.
- [4] Webnova, La medición de audiencias, Las diferentes metodologías.  
<http://www.webnova.com.ar/articulo.php?recurso=708>. Septiembre, 2008
- [5] Proyectos de Medición de Audiencias, WIP-Chile y BIT-Chile,  
[http://fcom.altavoz.net/prontus\\_fcom/site/artic/20080418/pags/20080418230431.html](http://fcom.altavoz.net/prontus_fcom/site/artic/20080418/pags/20080418230431.html), Septiembre, 2008.
- [6] Sady C. Fuentes Reyes, Marina Ruiz Lovaina, Minería Web: Un Recurso Insoslayable para el Profesional de la Información, ACIMED, v.16, n.4, Ciudad de La Habana, Oct., 2007.
- [7] ISO/IEC, ISO/IEC 9126-1:2001 Software Engineering – Product quality. Part 1: Quality Model, Secretaría General de ISO, Ginebra, 2001.
- [8] Pablo Straub, Cómo Comprar Software de Calidad , IV Feria de Proveedores ChileCompra, Marzo de 2005.
- [9] David Garvin, “What Does 'Product Quality' Really Mean”, Sloan Management Review 25(18), Fall 1984.
- [10] Departamento de ventas y servicios de ABITS Chile  
<http://www.abits.cl/default.htm>
- [11] Philip Kotler, Los 10 Principios del Nuevo Marketing,  
<http://www.soportecorp.com/documentos/10%20principios%20del%20Nuevo%20Marketing.pdf>, Septiembre, 2008.

- [12] Javan Erfanian, Mobile Technology Evolution, IEEE Communications Society, March 2007,  
[http://www.ingenieros.cl/index.php?option=com\\_docman&task=doc\\_download&gid=428&Itemid=](http://www.ingenieros.cl/index.php?option=com_docman&task=doc_download&gid=428&Itemid=), Septiembre, 2008