



Pontificia Universidad Católica de Valparaíso

Facultad de Ingeniería

Escuela de Ingeniería Informática

Ingeniería Civil en Informática

**WEB MINING COMO APOYO A LA EVALUACIÓN
DE USABILIDAD DE SITIOS WEB**

Informe final del Proyecto para optar al Título profesional de
Ingeniero Civil en Informática

Autor:

Eduardo Andrés Zamora Rodríguez

Profesor Guía:

Cristian Alexandru Rusu

JULIO 2008

Este trabajo está dedicado a mi familia, a Claudia, a mis amigos y compañeros de carrera, ellos son mi motivación para querer ser un buen profesional. Gracias a quienes me apoyaron en la realización de este trabajo participando o guiándome en éste

Glosario de términos

Data mining: Consiste en la extracción no trivial de información que reside de manera implícita en los datos. Dicha información era previamente desconocida y podrá resultar útil para algún proceso. En otras palabras, la minería de datos prepara, sondea y explora los datos para sacar la información oculta en ellos

Web mining: Es una metodología de recuperación de la información que usa herramientas de la minería de datos para extraer información tanto del contenido de las páginas, de su estructura de relaciones (enlaces) y de los registros de navegación de los usuarios.

Log: Un registro oficial de eventos durante un periodo de tiempo en particular. En este caso es un archivo que informa sobre las conexiones a un servidor.

Usabilidad: Es la medida en que un producto puede ser usado por usuarios específicos para alcanzar metas específicas con efectividad, eficiencia y satisfacción, en un contexto específico de uso

Link: Es un elemento de un documento electrónico que hace referencia a otro recurso, por ejemplo, otro documento o un punto específico del mismo o de otro documento.

Combinado con una red de datos y un protocolo de acceso, un link (hiperenlace) permite acceder al recurso referenciado en diferentes formas, como visitarlo con un agente de navegación, mostrarlo como parte del documento referenciador o guardarlo localmente.

Lista de abreviaturas o siglas

URL: Uniform Resource Locator

WUM: Web Usage Mining

HCI: Human – Computer Interaction

PUCV: Pontificia Universidad Católica de Valparaíso

Resumen

Debido a la necesidad de mejorar constantemente la usabilidad de los sitios web para mejorar la experiencia del usuario y brindar mayor accesibilidad, *Web Usage Mining* (WUM) ha surgido como una nueva herramienta útil para obtener información acerca de la interacción de los usuarios con la web. En el contexto de ingeniería de usabilidad y desarrollo de sitios web es importante conocer a los usuarios para aplicar mejor las guías, principios y métricas de usabilidad. La idea principal del presente proyecto es investigar acerca del uso de los datos entregados por WUM con el fin hacer mejoras en la usabilidad de sitios web, complementando las pruebas y evaluaciones tradicionales de usabilidad y viendo la relación entre los métodos.

En el presente trabajo se han realizado dos casos de estudio con evaluaciones de usabilidad sobre un sitio web, una de ellas con la ayuda de *Web Usage Mining*. Este último caso es sobre el sitio de la Pontificia Universidad Católica de Valparaíso (PUCV), y sobre éste se muestran las actividades realizadas y los resultados conseguidos.

El método propuesto en este trabajo y seguido en la realización de las actividades es un método que integra la minería web y las actividades más clásicas y adecuadas para este tipo de evaluación, obteniendo así un proceso más completo. Las actividades realizadas sobre el sitio fueron: pruebas de usabilidad, ordenamiento de tarjetas, cuestionarios, evaluaciones heurísticas y WUM. Los resultados obtenidos se presentan de forma resumida y ordenada.

El proyecto obtuvo los resultados deseados, validando así la propuesta de integración de *Web Usage Mining* a un proceso de evaluación de usabilidad. Se presentan conclusiones, ventajas y desventajas obtenidas de la realización de las evaluaciones con sus fundamentos y observaciones.

Palabras clave: usabilidad, web usage mining, data mining, evaluación., web.

1. Introducción

La Web es sin lugar a dudas el lugar más visitado de hoy y es el más indicado para el intercambio de información. Muchos negocios se han formado en su entorno y la época moderna se ha integrado en ella haciendo de la Web parte de nuestra cotidianidad. Es claro entonces que cuando se tiene un sitio web se debe tratar de que este sirva bien a nuestros objetivos, siendo accesible y fácil de usar, para que la información ofrecida esté disponible de la mejor manera.

La mayoría de los sitios web evolucionan constantemente en busca de optimizar su funcionalidad a medida que su entorno cambia o se descubren nuevas oportunidades de mejorar. Tomando esto en cuenta, varias metodologías y herramientas han aparecido para asistir la tarea de mejorar los sitios web. Dentro de estas metodologías se encuentra el *Web Usage Mining*.

Web Usage Mining (WUM) es una extensión de *Data Mining* pero enfocado más hacia datos textuales, ya que son estos los que se encuentran mayormente en la Web. WUM usa como fuente la interacción del usuario con la Web, adquiriendo y analizando la información relevante respecto a ésta para así obtener resultados como: la cantidad de visitas por intervalo de tiempo, horas de alta y baja audiencia, páginas más visitadas, páginas de entrada más frecuentes del sitio, ranking de palabras clave usadas en los buscadores para llegar al sitio y otra información que nos puede ayudar a descubrir nuevas tendencias, determinar pautas de navegación, patrones de interacción, errores de diseño y otros. Para obtener información relevante para el análisis se usan principalmente los logs de servidor.

Varias herramientas se han creado para WUM pero la mayoría se limita a la entrega de estadísticas y datos como los anteriormente nombrados, siendo necesaria la interacción de un experto que tome estos datos, entienda su significado y vea como puede usarlos para mejorar el sitio web en cuestión. La mayoría de la literatura se basa en técnicas, algoritmos

y beneficios del WUM pero poco se habla de cómo interpretar sus resultados y como usarlos dentro del contexto de evaluación de usabilidad web.

La creación y mantenimiento de un sitio Web es una tarea que debe ser vista como un esfuerzo igual al de la ingeniería de software, por lo que se debe seguir un enfoque similar que involucre requerimientos, análisis, desarrollo y pruebas. Estas pruebas deben encontrarse durante toda la vida del proyecto y pueden comenzar sobre bosquejos en papel y extenderse hasta que el sitio lleve años de uso. Es en este último caso donde los logs de servidor adquieren vital importancia, dado que son una huella que refleja uso real de los visitantes del sitio en condiciones naturales de trabajo y pueden reflejar el uso de grandes cantidades de usuarios.

Los logs de servidor aportan mucha información útil pero también carecen de una total aceptación respecto a la veracidad de la información contenida y se dice que pueden guiar a falsas conclusiones si no se tiene cuidado al analizarlos, es por esto que muchas de las personas que han trabajado con ellos sugieren complementar el análisis de los logs con pruebas de usabilidad o más bien, incluir el análisis de los logs dentro de un proceso de pruebas más completo. En el presente trabajo se expondrán las limitaciones y el potencial que ofrecen los logs de servidor en cuanto a aportar información útil para el proceso de pruebas de usabilidad.

Varias metodologías de pruebas e inspecciones de usabilidad han sido desarrolladas a lo largo del tiempo debido a la necesidad de comprobar y medir la satisfacción del usuario, la eficiencia de las interfaces, el tiempo de realización de las tareas, entre otros. La mayoría de estos métodos se aplica cómodamente a interfaces computacionales del estilo software tradicional, pero necesitan algunas modificaciones para aplicarse sobre interfaces Web, debido a que los usuarios interactúan de forma distinta con cada interfaz. Las técnicas que se describirán en el presente informe y que utilizarán para el análisis práctico, han sido elegidas por complementarse bien con las interfaces Web, por su simplicidad y tomando en cuenta los recursos disponibles para realizarlas. Estas técnicas son ordenación de tarjetas (Card Sorting), evaluación heurística, pruebas de usabilidad basadas en tareas (en el laboratorio de usabilidad) y cuestionarios.

La idea principal del presente proyecto es investigar acerca del uso de los datos entregados por WUM con el fin hacer mejoras en la usabilidad de sitios web, complementando estos datos con pruebas tradicionales de usabilidad y viendo la relación entre los métodos. Para comprobar esto se tomarán dos sitios web, uno de ellos analizado con dos herramientas de WUM (nombradas en el presente informe) y se usarán estos datos para apoyar pruebas tradicionales de usabilidad con las que se analizará el sitio en busca de problemas. Se muestra como crear una integración entre las salidas del WUM y las entradas de las pruebas de usabilidad y además buscar una forma en que estos métodos se complementen y apoyen. Se finaliza entregando una propuesta de integración del WUM con las pruebas y evaluaciones de usabilidad, con su respectiva validación y comprobación.

En el comienzo del documento se establece una base teórica para el trabajo práctico a realizar, describiendo los conceptos básicos involucrados en el proyecto, revisando el estado del arte y trazando el camino seguido durante el transcurso del proyecto. Más adelante se desarrolla la descripción de la metodología de pruebas de usabilidad que se utiliza, explicando los pasos a seguir y las técnicas elegidas. Se analiza el alcance del experimento y de sus resultados y se describe la propuesta de integración. Posteriormente se describe la realización de las actividades de evaluación y sus resultados obtenidos, terminando con la revisión y validación de la propuesta basándose en los datos encontrados.

Discusión bibliográfica

Para el siguiente trabajo se ha sacado información principalmente de Internet. La mayoría de las referencias encontradas son documentos PDF o contenido en sitios web, debido a que es donde se encuentra la mayoría información y la más actualizada. Cabe comentar que de todas formas ha sido difícil encontrar información objetiva libre de propósitos comerciales o publicitarios, al igual que documentos gratuitos de buena calidad, especialmente en lo que respecta a casos prácticos y ejemplos de aplicaciones reales de Web Usage Mining, por lo que se ha decidido basar la propuesta final mayoritariamente en la información obtenida en la realización del caso de estudio desarrollado en el presente proyecto.

La bibliografía encontrada hace referencia a trabajos que utilizan Web Usage Mining principalmente de dos formas:

Como una actividad desarrollada sola, sin complementos, con fines de usabilidad o comerciales.

Como una actividad desarrollada en un estudio de mercadotecnia más complejo

Por estas razones el proyecto aquí presente desarrolla un acercamiento centrado en la usabilidad, donde Web Usage Mining se realiza como parte de un estudio de usabilidad más grande, junto a otras actividades de evaluación de usabilidad.

Muchas referencias de las que se presentan en este trabajo están en inglés, ya que es el idioma ocupado en la mayoría de los trabajos realizados.

2. Análisis de objetivos y metodología

2.1 Objetivo General

Analizar el potencial de *Web Usage Mining* como apoyo a la evaluación de usabilidad de sitios web.

2.2 Objetivos específicos

Comprender el estado del arte en usabilidad web y *Web Usage Mining*.

Encontrar una manera de complementar los resultados obtenidos mediante *Web Usage Mining* y pruebas de usabilidad tradicionales.

Proponer una manera de integrar *Web Usage Mining* al proceso de evaluación de usabilidad de sitios web.

Obtener datos empíricos prácticos que apoyen los análisis y validen la propuesta final.

Estos objetivos han sido cumplidos en el presente trabajo y sus respectivos resultados serán descritos posteriormente. No hubo mayores inconvenientes en cumplir los objetivos, pero sí requirieron de una buena cantidad de trabajo. El desarrollo de los objetivos y del trabajo en sí mismo cumplió con lo esperado en un comienzo y superó las expectativas de resultados.

2.3 Metodología de trabajo

El desarrollo del presente trabajo comienza reuniendo y analizando información acerca de los conceptos que involucra el tema, el estado del arte, el contexto, las metodologías de solución y las diferentes aproximaciones. Una vez analizada esta información se pasa a la etapa experimental, donde se realiza el proceso de evaluación sobre dos sitios web, apoyando sólo uno de los procesos con *Web Usage Mining*, dejando el otro proceso de la

manera clásica para posteriores comparaciones, con el objetivo de reunir datos empíricos y de analizar casos prácticos reales. La fase final es realizar las conclusiones basadas en la información obtenida mediante investigación y experimentación. El trabajo termina con la propuesta de una manera de integrar WUM al proceso de pruebas y evaluaciones de usabilidad, presentando un enfoque dirigido a los evaluadores y no al software. Esta propuesta es apoyada por los casos prácticos que se realizan en la etapa experimental.

La división de trabajo por semestre ha sido la siguiente:

Proyecto 1: Investigación y elección de herramientas y metodologías.

Proyecto 2: Experimentación y conclusiones

2.4 Plan de Trabajo

El desarrollo del proyecto fue dividido en las siguientes etapas:

Primer semestre

1. Investigar respecto a los conceptos básicos que involucran Web Usage Mining y usabilidad.
2. Investigar del estado del arte de Web Usage Mining y usabilidad web.
3. Investigar documentos de casos de estudio reales efectuados por terceros y estudiar sus resultados, viendo la relación entre los datos obtenidos y las propuestas hechas para mejorar los sitios web que se analizaron.
4. Realizar una propuesta preliminar de cómo integrar Web Usage Mining a un proceso de evaluación de usabilidad de sitios web basado en la información teórica estudiada.

Segundo semestre

5. Analizar el sitio web seleccionado mediante las herramientas de WUM y analizar los resultados.

6. Analizar los sitios mediante pruebas de usabilidad convencionales con apoyo de los datos obtenidos con WUM dependiendo del caso.
7. Estudiar los resultados de la evaluación de usabilidad y analizar la relevancia de WUM dentro del proceso de evaluación realizado.
8. Proponer basado en los resultados obtenidos una manera de integrar WUM al proceso de evaluación de usabilidad de sitios web.

El desarrollo del trabajo tuvo ciertos atrasos causados por la falta de acceso al servidor del sitio analizado. Una vez que se garantizó el acceso el proceso fue fluido y sin complicaciones. Se recomienda a futuros evaluadores asegurar el acceso al servidor de forma anticipada.

3. Diseño y resolución del trabajo realizado

3.1 Web Mining

3.1.1 Definición de minería web

Web mining (minería Web) puede definirse como el descubrimiento y análisis de información útil en la World Wide Web a través de técnicas de *data mining* [1]. Existen tres tipos de información que puede analizarse, lo que da cabida a tres áreas de *web mining* o tres dominios de aplicación: *Web Content Mining* se ocupa de analizar el contenido de la web, es decir la información textual, *Web Structure Mining* analiza la estructura de la web, con sus links y la organización de los mismos, y por último, *Web Usage Mining* (WUM) que es el más relevante para el presente proyecto y se define a continuación.

WUM es el proceso de descubrir e interpretar patrones de acceso de los usuarios a los sistemas de información web, minando los datos recogidos de la interacción de los usuarios con el sistema [14]. Esta es una definición comúnmente usada, pero queda un poco corta tomando en cuenta el estado del arte actual en WUM, ya que pocas herramientas son capaces de entregar un análisis completo de caminos comunes o patrones de uso, como pudo verse en el desarrollo del presente trabajo. Hoy en día la información entregada por la mayoría de los softwares e WUM corresponde a estadísticas y datos cuantitativos respecto al uso del sitio. La mayoría de esta información es generada automáticamente por los servidores Web y se almacena en archivos llamados *Log files* o logs de acceso al servidor. Otras fuentes de información del usuario incluyen la referencia a otros sitios o páginas de la Web y registros de usuario en bases de datos vía formularios en línea.

Como en cualquier otra aplicación de minería de datos, el éxito del proceso depende del conocimiento que se descubre y su disponibilidad en un momento determinado, su accesibilidad por los usuarios que realmente lo van a emplear y, sobre todo, por su validez

y fiabilidad. Por ello es muy importante medir de forma exacta y precisa el tráfico sobre el servidor web.

3.1.2 Actividades del proceso de *Web Usage Mining* (WUM)

El proceso de web mining consta de varias actividades, las cuales se describen a continuación.

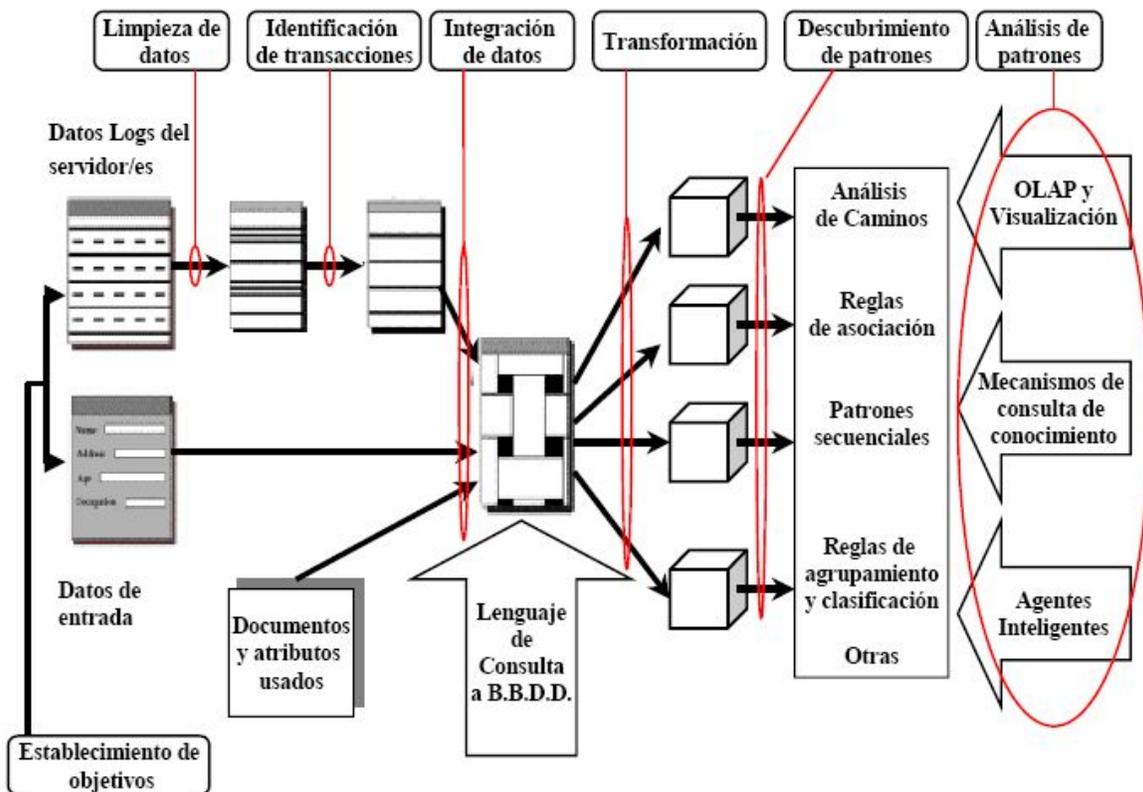


Ilustración 3.1: Actividades del proceso de Web Usage Mining por Bamshad Mobasher [3]

a) *Establecimiento de objetivos*

Se comienza como en todo proyecto con el establecimiento de los objetivos, estas serán las metas que guiarán la ejecución del proyecto, definiendo la estrategia y estableciendo marcos, restricciones y validaciones. Distintos objetivos pueden ser “mejorar la usabilidad del sitio”, “conocer o descubrir clientes”, “encontrar fallas de diseño”, “encontrar los

sectores mas populares del sitio”, etc. Estos objetivos son fijados conociendo las posibilidades que entrega la herramienta utilizada y tomando siempre en cuenta el modelo del negocio y las necesidades de los usuarios.

b) Adquisición de los datos

Una vez definidos los objetivos se pasa a reunir los datos que serán parte del análisis. Estos son por lo general Logs de servidor, aunque también se puede agregar datos de los clientes/usuarios, datos demográficos, datos de facturación y marketing, etc. [4]. El log de servidor se almacena en archivos de texto plano (ASCII) en diferentes formatos (NCSA, W3C y Microsoft IIS), aunque algunos servidores pueden almacenarlo directamente en una base de datos. La información que se almacena depende también de cada servidor y su configuración particular, pero en general los campos incluidos habituales son: la fecha y hora de la petición, la máquina (y/o dirección IP) del cliente, el método de acceso (GET, PUT...), el archivo accedido (URL), el resultado de la petición, el tamaño en bytes de los datos recibidos y devueltos, la identificación del cliente (User-Agent), cookies y la página de procedencia (Referrer). [5].

c) Limpieza de los datos

Cuando los datos a analizar han sido recolectados se procede a la limpieza de los datos. El objetivo de la limpieza es excluir todo tipo de accesos que entorpezcan el análisis [6], de esta manera se puede obtener mejores y más fiables resultados. Datos comunes que son eliminados en la limpieza son por ejemplo las peticiones por archivos de multimedia, como imágenes, videos y sonidos, datos generados por robots de búsqueda, spiders y otros, peticiones a páginas dinámicas o scripts que no influyen en el contenido, etc. Una técnica común para filtrar los datos es fijarse en el sufijo de la URL, así sufijos como gif, jpeg, jpg, js, cgi, pl sugieren que esa entrada puede ser eliminada [4]. Además de la limpieza como tal, un poco de pre-procesamiento puede ser necesario en los datos, integrando datos de distintos servidores, combinando logs múltiples o incorporando datos de referrer logs (logs de sitios de referencia).

d) Identificación de transacciones

Una vez que los datos están limpios viene el proceso de identificación de las transacciones de usuario y las sesiones. Una sesión de usuario está formada por todas las páginas consultadas por un usuario durante una sola visita al sitio. Una transacción es un conjunto de páginas homogéneas que han sido visitadas en una sesión. El tamaño de una transacción puede variar desde una sola página consultada hasta todas las consultadas en una sesión de usuario [3]. Para identificar a un usuario se puede asociar aquellos accesos que pertenezcan a la misma IP, además para reducir errores y ambigüedades como las que se comentarán mas adelante, se puede identificar también el navegador, su versión y su sistema operativo. Para identificar una sesión el método más usado es fijar un tiempo límite entre el acceso a cada página, un tiempo típico es de entre 20 y 30 minutos entre un acceso y otro.

Existen variados métodos para la construcción de las sesiones, que buscan cerrar los vacíos que puedan producirse al juntar los accesos. Estos vacíos pueden producirse por el uso de caché, proxys o el uso de los botones de atrás y adelante del navegador, se explicará con más detalle estos problemas en un apartado especial. Un método simple para llenar estos vacíos o incongruencias es comprobar si existe un medio de comunicación (link) entre una página y otra.

e) Integración y transformación de la información

En esta fase se usa (si se requiere) los datos extraordinarios que fueron nombrados en un principio, tales como datos del usuario y datos de las páginas de referencia. El objetivo es complementar los datos para obtener mayor información. Los datos obtenidos a través del registro de los usuarios están cobrando importancia para identificar a los usuarios de forma específica, pero también hay muchas aproximaciones que prefieren mantener al usuario anónimo. Cuando los datos han sido integrados se procede a transformarlos al formato que más se adecue a los procesos de data mining que seguirán.

f) Descubrimiento de patrones

Cuando los datos están listos viene el análisis o procesamiento de éstos para descubrir la información deseada. Dependiendo de la información que se busque utilizaremos distintas técnicas, así pues, las técnicas que más se emplean para realizar minería de uso de la Web

son: agrupamiento y clasificación, detección de reglas de asociación, análisis de caminos, y detección de patrones secuenciales, Bamshad Mobasher [4] define estas técnicas como sigue.

- Agrupamiento y clasificación: Las técnicas de agrupamiento o *clustering* distribuyen comportamientos de individuos similares en grupos homogéneos. Esto es, dos elementos con características parecidas pertenecerán al mismo grupo, y las características de un grupo (definidas por un elemento prototipo o ideal) serán diferentes a las de otro grupo. Dependiendo de la información almacenada en los archivos log, es posible detectar grupos de usuarios como:
 - Aquellos que visitan gran cantidad de páginas con un tiempo de estancia muy similar en todas ellas.
 - Los que visitan un número pequeño de páginas en sesiones cortas.
 - Los que visitan un número pequeño-mediano de páginas con tiempo variable en cada una de ellas.
 - Etc.

Una vez descubiertos los prototipos o perfiles de cada grupo, se pueden usar las características de cada uno de ellos para realizar *clasificación*.

- Reglas de asociación: Las técnicas de descubrimiento de reglas de asociación se aplican a los conjuntos de sesiones y/o transacciones. En este caso, el problema consiste en descubrir todas las asociaciones y correlaciones entre los accesos y usos de la Web por parte de los usuarios. Con estas técnicas, por ejemplo, es posible encontrar correlaciones tales como las siguientes:
 - El 40% de clientes/usuarios que accedieron la página web con URL /entidad/productos/producto1.html, también accedieron a /entidad/productos/producto2.html
 - El 30% de clientes/usuarios que accedieron a /entidad/anuncio/oferta-especial.html, efectuaron un pedido interactivo en /entidad/productos/producto1.

- Análisis de Caminos: Esta técnica supone la generación de algunas formas de grafos orientados que representan relaciones entre páginas web. Este grafo puede ser un esquema físico en el que las páginas web son los nodos del grafo y los hiper-enlaces entre las páginas son las flechas dirigidas entre nodos. Pueden formarse otros grafos a partir de los tipos de páginas web, con arcos que representen la similitud entre páginas, o creando arcos que muestren el número de usuarios que van desde una página a otra. El análisis de caminos podría utilizarse para determinar los caminos más frecuentemente seguidos en un sitio web. Otros ejemplos de información que puede descubrirse a partir de un análisis de camino son:

 - El 70% de clientes/usuarios que accedieron a /entidad/productos/archivo2.html lo hicieron a partir de /entidad siguiendo por entidad/novedades, /entidad/productos, y /entidad/productos/archivos1.html.
 - El 80% de clientes/usuarios que accedieron al lugar empezaron por /entidad/productos.
 - El 65% de clientes/usuarios abandonaron el lugar después de consultar cuatro o menos páginas.

- Patrones Secuenciales: El problema de descubrir patrones secuenciales se centra en localizar la presencia de un conjunto de elementos seguida por otro elemento en un conjunto de transacciones o visitas ordenadas en el tiempo. En un histórico de transacciones de un servidor web, la visita de un cliente se guarda por un periodo de tiempo asociado. El descubrimiento de patrones secuenciales en los históricos de acceso al servidor web permite a las organizaciones, predecir los patrones de visita de usuarios. Analizando esta información, el sistema de minería de uso de la Web puede determinar relaciones temporales entre elementos de datos tales como las siguientes:

 - El 30% de clientes/usuarios que visitaron /entidad/productos/, habían hecho una búsqueda en Yahoo dentro de la semana anterior con la palabra clave w. Esto es posible con configuraciones especiales del servidor y con algunas herramientas de minería de uso de la Web.

- El 60% de clientes/usuarios que efectuaron una orden de compra online en la página A, también efectuaron una orden online en la página B con un lapso máximo de 15 días.

Otro tipo de dependencia de datos son las secuencias ocurridas en un intervalo dado. Por ejemplo, puede ser interesante encontrar las características comunes de todos los clientes/usuarios que visitaron un archivo en particular dentro del periodo de tiempo $[t_1, t_2]$. O al contrario, puede ser de interés un intervalo de tiempo (un día, una semana, etc...) en el cual un archivo en concreto es el más accedido.

g) Análisis de los patrones encontrados

Una vez que ya se han minado los datos se deben analizar para entender su significado. Hoy en día existen herramientas que permiten facilitar este análisis, sin embargo no hay mucha experiencia en el campo debido al poco tiempo que este lleva de desarrollo y a la poca experiencia y trabajo en ésta área. Todavía no se ha llegado a un nivel de automatización del análisis de la información y patrones, aunque hay algunas propuestas que se presentarán en el estado del arte. Algunas de las técnicas más comunes son las visualizaciones gráficas tipo árbol o grafo, donde las páginas del sitio son los nodos y las aristas son los links entre ellas, mediante colores y tamaños se puede expresar prioridad, cantidad de visitas, cantidad de contenido, etc.

Muchos de los análisis se basan en la búsqueda de oportunidades comerciales y estudios del estilo de CRM (Customer Management Relationship) pero este proyecto se enfocará en la utilización de la información para mejorar la usabilidad del sitio analizado.

3.1.3 Ventajas y problemas de WUM

Las ventajas de WUM son principalmente el conocimiento no obvio que se puede descubrir de los datos obtenidos de la interacción de los usuarios, y la posibilidad de realizar este proceso en bases regulares, buscando una continua optimización del sitio web. WUM provee una manera de automatizar la recolección de datos respecto a los usuarios, disminuyendo la necesidad de realizar otras tareas más trabajosas como encuestas y cuestionarios. Las desventajas principales de WUM están relacionadas con los errores e

información inexistente que puede haber en los logs de servidor debido a los siguientes factores:

- *El uso de cachés, proxys y botones de adelante/atrás del navegador:* la existencia de cachés en la red hace que aunque un cliente visite una página determinada, la petición pueda no llegar al servidor por estar almacenada en un servidor intermedio, y por tanto la petición no queda registrada en el log. Lo mismo sucede con los botones del navegador, ya que la petición no llega al servidor.
- *Dificultad en la identificación del usuario y de las sesiones:* Debido a la navegación anónima, servidores Proxy y ISPs de asignación dinámica de Ips. También causan dificultad a la hora de limpiar los datos los bots, spiders y crawlers de los motores de búsqueda e indexación.

Este asunto se desarrollará con detalle más adelante en el informe.

3.2 Usabilidad

3.2.1 Definición

Usabilidad es la medida en que un producto puede ser usado por usuarios específicos para alcanzar metas específicas con efectividad, eficiencia y satisfacción, en un contexto específico de uso [7]. Este concepto no es único para las interfaces web o computacionales sino que es aplicable para todo producto u objeto que sea usado por una persona, tal como puede ser una tetera, un refrigerador o un celular.

En el ámbito de la informática y la computación existe una ciencia que se preocupa de lo concerniente a la usabilidad entre otras cosas. HCI o Human Computer Interaction es el estudio de la interacción entre los usuarios (humanos) y las computadoras. Es una materia interdisciplinaria que involucra ciencias de la computación, psicología, sociología, antropología, diseño, entre otras. Este estudio ha evolucionado a medida que las formas de

interactuar evolucionan también, pero el objetivo siempre se mantiene constante: hacer de la interacción una experiencia lo más satisfactoria y eficiente posible.

En la usabilidad existen dos problemas o asuntos fundamentales:

¿Cómo se puede desarrollar un sistema para asegurar su usabilidad?
¿Cómo se puede medir o demostrar la usabilidad de un sistema interactivo?

Para resolver estas preguntas se han desarrollado principios, guías y reglas de usabilidad, las cuales sirven como apoyo al desarrollo y evaluación de sistemas. Así mismo se han desarrollado métricas que permiten medir y comparar resultados. Para obtener estos resultados y para medir usabilidad y otros atributos importantes se han creado distintas maneras de evaluar una interfaz o un sistema a través de pruebas de usabilidad, evaluaciones, inspecciones, etc.

3.2.2 Principios, guías y reglas de usabilidad

Cuando se quiere medir usabilidad se puede tomar en cuenta algunos atributos que son muy importantes en un sistema, estos son la facilidad de aprendizaje, la eficiencia, la facilidad de recordar el uso, la propensión a errores y la satisfacción subjetiva del usuario con el sistema. Midiendo cada uno de estos atributos se puede tener una idea del nivel de usabilidad del sistema, ¿pero como diseñar y desarrollar un sistema para asegurar que cumpla bien estos atributos? Si bien no se puede asegurar la buena usabilidad de un sistema (o un sitio web) completamente, se puede esperar obtener mejores resultado al seguir los principios guías y reglas de usabilidad.

IBM Easy of Use [8] define principios y guías como sigue:

Principios: Ideas fundamentales y creencias usadas para guiar la toma de decisiones y obtener un resultado general.

Guías: Cursos de acción recomendados que están en soporte de un conjunto de principios y son específicos a un dominio en particular como puede ser la web.

Ejemplos de principios de usabilidad son:

Ordenar los elementos de una interfaz de manera coherente y significativa para el usuario, agrupando los relacionados.

Hacer que las tareas que parecen simples para el usuario sean simples de hacer.

Mantener a los usuarios informados del estado del sistema. Dar retroalimentación.

En el diseño actual de sistemas y sitios web en particular, se recomienda seguir un enfoque centrado en el usuario (user centered), tomando en cuenta sus necesidades y requerimientos. Siguiendo este enfoque y guiándose por los principios, guías y reglas de usabilidad se puede obtener un mejor resultado. En [9] se proponen 10 guías para un diseño web basado en el usuario, las que se resumen a continuación:

Visibilidad: Hacer que los elementos importantes de una interfaz resalten claramente para que el usuario pueda saber que puede y que no puede hacer sólo con una simple mirada. A esto también se le conoce como *affordances* o aquello que es posible hacer.

Reducir la carga de la memoria: Hacer que los elementos del sitio sean consistentes y significativos para que el usuario no tenga que recordar cada cosa que ve. A esto se le conoce también como *recognition rather than recall* que quiere decir que es mejor que el usuario reconozca un ítem a que tenga que recordarlo.

Retroalimentación: mantener al usuario informado del estado del sistema y de sus acciones.

Accesibilidad: Organizar la información de forma que sea fácilmente accesible y asimilable. Este término también hace referencia a garantizar el acceso a la información a distintos usuarios con distintas capacidades, perfiles y tecnologías.

Orientación/navegación: dotar al sitio de una coherencia en su estructura y elementos que permitan una fácil navegación.

Errores: saber prevenirlos y manejarlos cuando ocurren. El usuario no debe perder su trabajo ni sentirse desconcertado por un error.

Satisfacción: Es uno de los factores más influyentes en el éxito de un sitio web y uno de los más difíciles de garantizar. La satisfacción de un usuario es subjetiva y afectará su visión del sitio / sistema en varios aspectos.

Legibilidad: hacer el texto y la información en general fácil de leer.

Lenguaje: el lenguaje debe ser conciso y acorde con los usuarios, de naturaleza neutra y sin demasiados modismos. Es común hacer un sitio con soporte a más de un idioma.

Diseño visual: es la cara del sitio y la primera impresión que recibe el usuario, debe ser simple, acorde a los objetivos del sitio, claro, preciso y agradable.

3.2.3 Pruebas, evaluaciones e inspecciones

Como se menciona anteriormente, al tener en cuenta los principios a la hora de desarrollar un sistema es posible aumentar la posibilidad de obtener mejores resultados, pero no se puede asegurar completamente que el sistema tendrá un buen nivel de usabilidad. Para esto se debe entonces realizar pruebas de usabilidad, inspecciones y evaluaciones. Realizar este tipo de actividades es recomendado desde las primeras etapas del diseño y de forma iterativa durante el desarrollo del proyecto. Desde el bosquejo de una interfaz en papel hasta un prototipo funcional o el sistema ya con años de uso, existen varias instancias donde se pueden realizar estas actividades.

a) Inspecciones de usabilidad

Una inspección de usabilidad es cuando uno o varios expertos revisan y evalúan un sistema (en este caso un sitio web) para identificar potenciales problemas de usabilidad [10]. Generalmente las inspecciones de usabilidad tienden a encontrar fallas de alto nivel en las reglas de diseño y consistencia. Las inspecciones son generalmente más baratas y fáciles de realizar que las pruebas de usabilidad.

Las inspecciones de usabilidad más comunes son:

- Evaluación Heurística: la evaluación heurística es un método para la detección de problemas de usabilidad en el diseño de una interfaz e implica tener un grupo de evaluadores que examinan la interfaz y juzgan su adecuación con principios reconocidos de usabilidad [6].

- Recorrido cognitivo: El recorrido cognitivo es un método de inspección que se centra en la evaluación de un diseño a través de exploración. En él, un grupo de pares evalúan el diseño a través de la consecución de una o más tareas.

b) Pruebas de usabilidad

Las pruebas de usabilidad involucran tener gente de los usuarios objetivo evaluando el sitio (o sistema) y realizando ciertas tareas. Las pruebas involucran por lo general un experimento controlado, donde los usuarios son observados mientras realizan las tareas. Esta observación puede ser directa o indirecta, mediante cámaras, espejos o software de captura de interfaces. Estas pruebas sirven generalmente para encontrar problemas más específicos del dominio y en el flujo de tareas. Las pruebas de usabilidad están bien valoradas ya que dan un acercamiento a la experiencia del usuario de forma individual, siendo posible capturar incluso expresiones faciales y verbales mientras usa el sistema, ya no hay suposiciones acerca del comportamiento del usuario sino que hay una vista directa a él. Además se puede conocer la opinión directa del usuario mediante cuestionarios o comunicación verbal.

Los pasos a seguir para realizar una prueba de usabilidad son generalmente:

Plan: Las tareas, usuarios y el entorno deben estar dentro del contexto, ser representativos y adecuados. Los escenarios deben ser bien pensados y las sesiones deben contar con la infraestructura adecuada.

Realización de las sesiones: Los usuarios realizarán las tareas designadas mientras se observan los puntos que se busca conocer.

Análisis de las salidas: A través de métricas o evaluaciones subjetivas se analiza la información obtenida.

Estas pruebas se realizan comúnmente para medir lo siguiente:

El tiempo de realización de las tareas

La cantidad de errores cometidos

Los puntos críticos donde se cometen los errores

La capacidad de los usuarios de recordar el uso del sistema

La respuesta emocional del usuario ante el sistema

La satisfacción subjetiva

3.2.4 Usabilidad web

Usabilidad web es hacer un sitio de tal manera que los usuarios puedan encontrar lo que buscan de forma rápida y efectiva. Es aplicar los conceptos anteriormente mencionados al desarrollo de sitios web para lograr un sitio que ayude a los usuarios a alcanzar sus objetivos de forma simple, efectiva y con un alto grado de satisfacción.

Si bien los conceptos de usabilidad que se mencionaron anteriormente se aplican a todas las interfaces en general, los sitios web tienen también sus propias reglas, principios y guías específicas debido a sus características únicas que los diferencian, por ejemplo, de las interfaces estilo Windows.

Algunos conceptos o principios básicos muy generales son por ejemplo:

El sitio debe ser fácil de navegar

Usar convenciones a las que el usuario ya está acostumbrado

Las páginas deben cargar rápido

La información debe ser fácil de adquirir

Las restricciones no deben ser puestas en los usuarios

El diseño de páginas web ha tenido siempre un conflicto entre arte y ciencia. Si bien es una mezcla entre ambos, muchas veces el deseo de que el sitio sea visualmente impactante o complejo compromete seriamente la usabilidad, por otra parte, muchas veces la excesiva simplificación en el diseño nos deja un sitio con muy buena usabilidad, pero con poco atractivo para el usuario que termina por preferir otro sitio. En esta constante búsqueda de un diseño que llegue mejor a los usuarios es muy útil contar con herramientas y métodos como las que se mencionan en este trabajo, que nos ayudan a tener acceso a información acerca de lo que quieren, hacen y piensan los usuarios.

Un concepto importante que ha nacido en el ámbito de la usabilidad web es la accesibilidad, la cual significa básicamente asegurar que el sitio web sea accesible para todos los usuarios, sin importar el hardware o software que usen e incluso sus capacidades cognitivas o sensoriales. Asegurar accesibilidad implica el uso correcto de las tecnologías y seguir una serie de guías de accesibilidad cuando se desarrolla un sitio. Algunos ejemplos son:

Usar texto alternativo para las imágenes (ALT text)

Tener una versión solo texto del sitio.

Usar CSS para el formato en vez de marcos

Usar tecnologías compatibles con la gran mayoría de los navegadores

Se debe recordar que el objetivo de un sitio es dar a conocer cierta información o prestar cierto servicio a todos los usuarios que estén interesados en estos. Construir un sitio con mala accesibilidad es como tener un hospital sin acceso a minusválidos.

3.2.5 Métricas en usabilidad web

Como se menciona anteriormente, si se siguen los principios de usabilidad y accesibilidad se puede obtener mejores resultados pero no se puede asegurar completamente que nuestro sitio será 100% usable y del agrado de los usuarios. Se debe realizar pruebas e inspecciones y usar herramientas que nos ayuden en el proceso. Pero cuando se hacen estas pruebas ¿Qué es lo que se quiere medir?, hay que definir primero un conjunto de métricas que sean de interés y que permitan medir y comparar nuestros resultados.

En [11] se define métrica como un valor numérico o nominal asignado a características o atributos de un ente computado a partir de un conjunto de datos observables y consistentes con la intuición. Para que una métrica sea válida se debe confirmar la validez del atributo, de la unidad de medida, del instrumento (o método) que mide y del protocolo. Es posible usar estas métricas mediante la medición de ciertos eventos, o comprobando el cumplimiento de ciertos requisitos dados por heurísticas conocidas. De esta manera algunas heurísticas y métricas pueden ser [11]:

- Comprensión global del sitio: se puede comprobar la existencia de un esquema de organización global, un mapa del sitio, índices o una visita guiada.

- Métricas de éxito: sirven para analizar la realización de tareas
 - $((n^{\circ} \text{ tareas terminadas} + (\text{a medio terminar})/2) * 100) / n^{\circ} \text{ total de tareas}$
 - Tiempo que le toma al usuario completar la tarea
 - Índice de errores
 - Satisfacción subjetiva

- Sobre la usabilidad del contenido del sitio (lectura): se basan en preguntas tipo comprensión de lectura al usuario (puede depender de las capacidades del usuario)
 - Tiempo que le toma responder las preguntas
 - Cantidad de errores
 - Memoria: $(\text{correctas} - \text{incorrectas}) * 100 / n^{\circ} \text{ total}$
 - Tiempo que demora el usuario en dibujar un mapa del sitio
 - Satisfacción subjetiva

- Métricas de confiabilidad: para verificar la correctitud de los enlaces
 - N° de enlaces rotos
 - N° de páginas muertas (error 404)
 - N° de enlaces no implementados
 - N° enlaces inválidos

- Métricas de contenido: basados en estudios de la cantidad de información que es recomendable presentar según las capacidades de aprendizaje del usuario
 - Cantidad de palabras por página
 - Longitud de párrafos

- Legibilidad del texto
- Cantidad de imágenes

- Métricas de navegación: nos dicen sobre la estructura del sitio, se puede comprobar la existencia de indicador de camino, etiqueta de posición actual, la permanencia de los controles de navegación, el promedio de enlaces por página entre otros.

- Métricas de eficiencia
 - Tiempo de descarga
 - Versión solo texto
 - Cantidad de imágenes con ALT
 - Cantidad de frames (marcos)

3.3 Estado del arte

3.3.1 Estado del arte en Usabilidad

La usabilidad está cobrando gran importancia en la web en estos últimos años, debido al gran crecimiento que esta presenta, aumentando significativamente la cantidad de información y el número de usuarios con acceso. Debido a esto existe una gran demanda y una alta competitividad, donde los usuarios pueden elegir fácilmente entre una página y otra. Se hace necesario entonces, tener un sitio web usable, accesible y agradable para que los usuarios lo prefieran por sobre la competencia o simplemente para que cumpla su objetivo de la mejor forma posible.

En el sitio de IBM [8] se puede encontrar un completo documento que nos guía desde el principio del proyecto de desarrollar un sitio web, proporcionando guías de usabilidad web y consejos prácticos obtenidos a través de numerosas investigaciones y experiencia. Nos muestra que las consideraciones sobre usabilidad existen a lo largo de toda la vida del proyecto y se extienden hasta su mantenimiento, de forma iterativa y siempre con el enfoque centrado en el usuario. IBM no es único sitio que se puede encontrar que ofrece

material sobre usabilidad web, existen cada vez más sitios donde se pueden encontrar *papers*, cursos, consultorías y otros, demostrando la relevancia que está tomando el tema.

3.3.2 Difusión de la usabilidad

Existen numerosos sitios y estudiosos que se preocupan de la difusión de la usabilidad, haciendo ver su importancia y ofreciendo guías e información útil al respecto, con el objetivo de crear conciencia y mejorar el desarrollo de los sistemas, disminuyendo la cantidad de diseños basados en el sistema y aumentando los centrados en el usuario.

Useit.com: el sitio web de Jakob Nielsen [2], quien es una de las autoridades más respetadas en el ámbito mundial sobre usabilidad en la web, su carrera profesional se extiende por empresas como Bellcore, IBM y Sun Microsystems. En ese sitio es posible encontrar sus opiniones personales y estudios sobre el tema, además de noticias, columnas y enlaces a sitios relacionados. Jakob Nielsen actualmente trabaja en el Nielsen Norman Group, un grupo formado por expertos que ofrece servicios como entrenamiento, capacitación, pruebas de usabilidad y publicaciones entre otros.

IBM Easy of Use: es un área del sitio web de IBM que está enfocado en entregar información útil respecto a la usabilidad de sistemas, Windows y Web. Esta sección está más orientado a las empresas, para que estas logren construir sistemas fáciles de usar mediante principios y guías ofrecidos en el sitio. También se ofrecen seminarios, servicios y productos relacionados.

Usability.gov: ofrece guías desde el plan de proyecto hasta la refinación y pruebas

Usablenet.com: ofrecen un enfoque hacia la usabilidad y la accesibilidad.

Así como los sitios anteriormente nombrados existen muchos otros en diferentes idiomas y con diferentes enfoques (comercial, no comercial, publicaciones, servicios, etc.) pero todos reconocen y difunden la importancia de la usabilidad en los productos y servicios que se han de construir.

3.3.3 Ingeniería de la usabilidad

La ingeniería de la usabilidad describe un enfoque pragmático hacia el diseño de interfaces de usuario que enfatiza los métodos empíricos y definiciones operacionales de requerimientos del usuario. Se propone que especificaciones de usabilidad sean hechas explícitas como requerimientos [7]. Quienes siguen esta aproximación se dedican durante el proceso de desarrollo a tareas de análisis de tareas (task analysis) para luego pasar a prototipar y realizar pruebas de usabilidad, en un ciclo iterativo.

El ciclo de vida en ingeniería de usabilidad se puede resumir como sigue:

1. Conocer a los usuarios
2. Análisis competitivo
3. Puesta de metas
4. Diseño de la interacción orientado a metas
5. Diseño iterativo
6. Estudios posteriores de seguimiento

Este proceso es altamente adaptable y a la vez bien estructurado y se aplica tanto al desarrollo de software tradicional como al desarrollo de sitios y aplicaciones web. En ingeniería de la usabilidad se puede destacar un conjunto de tareas que lo diferencian de los enfoques tradicionales y que se preocupan de seguir un enfoque centrado al usuario.

Algunos ejemplos de estas tareas son:

Aplicar técnicas de análisis de requerimientos de usabilidad
Extraer metas de usabilidad desde el análisis de requerimientos para guiar el diseño
Aplicar un enfoque estructurado al diseño de la interfaz de usuario
Aplicar técnicas de evaluación iterativas para validar los diseños antes que se implementen

Integrar la ingeniería de la usabilidad tanto en el desarrollo de software como en desarrollo de proyectos web
--

3.3.4 Herramientas de apoyo en usabilidad

Existen variadas herramientas hoy en día que ayudan al proceso pruebas de usabilidad. Estas herramientas ayudan optimizando o automatizando cualquiera de las tres áreas o etapas del proceso: recolección y captura de datos, análisis y recomendaciones, y críticas. La automatización de ciertas tareas permite disponer mejor del tiempo y las personas del equipo de evaluación, siendo posible realizar más y mejores pruebas. Otro punto a favor que las herramientas de automatización ofrecen es la menor necesidad de expertos evaluadores. Sin embargo, estas herramientas ofrecen un complemento a las técnicas estándares de evaluación y prueba y no son un reemplazo de ellas ya que distintas pruebas obtienen diferentes resultados y medidas como la satisfacción subjetiva del usuario difícilmente pueden ser captadas por herramientas de automatización.

Es posible distinguir tres tipos automatización; de captura, de análisis y de recomendaciones [12].

a) Automatización de captura

En este nivel el enfoque se dirige a capturar automáticamente las acciones que el usuario realiza en la interfaz del sistema (estilo Windows o web). Métodos comunes son grabar las acciones del usuario en un video y guardar las acciones en logs. Las acciones pueden estar a nivel de sistema o a nivel de teclado. Estos softwares ofrecen generalmente funcionalidades como análisis de logs, traducción de los eventos simples a eventos significativos, acceso remoto a las acciones del usuario y otros. Las herramientas de web usage mining se encuentran en este nivel, ya que analizan los logs de servidor. Algunas herramientas para interfaces no web son KALDI, Morae Recorder y Morae Observer. Herramientas para WUM están descritas en el apartado de estado del arte de WUM.

b) Automatización de análisis

Se distinguen aquí cuatro aproximaciones al análisis: basado en métricas, basado en tareas, comparación de patrones e inferencial. El enfoque basado en métricas implica obtener resultados cuantitativos que miden la performance tales como tiempo de compleción de tareas, eficiencia del usuario, número de operaciones para completar la tarea, etc. El enfoque basado en tareas analiza las diferencias entre el modelo previo del diseñador respecto a las tareas del usuario y lo que el usuario hace en la práctica. El enfoque de comparación de patrones analiza los logs en búsqueda de descubrir el comportamiento de los usuarios. El enfoque inferencial del análisis de los logs utiliza técnicas estadísticas y de visualización tales como promedios, porcentajes, grafos y mapas conceptuales.

Muchas herramientas de WUM también caen en esta categoría y serán nombradas posteriormente. Algunas herramientas no web son AMME, QUIP y Morae manager.

c) Automatización de las recomendaciones

Estas herramientas se basan principalmente en la comparación de la información obtenida en la interfaz con guías y principios de diseño y sirven como apoyo a las inspecciones más que a las pruebas de usabilidad. Las herramientas de WUM escapan a esta categoría, así como la mayoría de las herramientas antes mencionadas. En el estado de arte actual existen pocas herramientas capaces de automatizar las recomendaciones de actividades más complejas como la pruebas de usabilidad, pero existen algunas que pueden complementar bien las inspecciones. Ejemplos de estas herramientas son UsableNet LIFT y Ergoval.

3.3.5 Estado del arte de Web Usage Mining

La aplicación de técnicas de data mining en la web está cobrando cada vez más fuerza. Existen ya diversas compañías que ofrecen herramientas de web mining e incluso asesorías completas de análisis y recomendaciones. También han nacido diferentes aproximaciones al problema de capturar data de uso de la interacción usuario/sitio web, tratando de compensar los problemas que presenta el analizar los logs de servidor. Cabe destacar que el proceso de minería web descrito anteriormente en la definición es un ideal que las herramientas existentes parecen no alcanzar aun. A continuación se presenta el estado del arte de los conceptos más importantes que involucran WUM.

3.3.6 Descubrimiento de patrones de acceso

La idea principal es ganar conocimientos acerca de quienes usan el sitio web. Pensando en que este es diseñado con un enfoque centrado al usuario, mientras mejor se conozca a los usuarios mejor se puede diseñar el sitio y se obtendrán mejores resultados. Las compañías de E-commerce están descubriendo que necesitan innovar y responder a las cambiantes necesidades y comportamiento de los clientes. Enfocándose en encontrar nuevas maneras de mejorar la experiencia del cliente para simplificar el negocio y aumentar la tasa de captación de clientes [13]. Se necesita estar al tanto de las nuevas tendencias y necesidades que puedan surgir, para no quedar atrás y tener siempre una ventaja competitiva. El descubrimiento de patrones de acceso ayuda actualmente a tareas como descubrir perfiles de usuario, mejorar la estructura de un sitio conociendo la manera en que los usuarios recorren la web, optimizar la colocación de anuncios, y la personalización de los servicios o el sitio web en general acorde a un perfil de usuario. Como se comentaba al inicio del documento, pocos softwares actuales proveen de patrones de acceso y tan sólo los software de más alto nivel, y costo por supuesto, son capaces de entregar este tipo de información.

3.3.7 Personalización web

La personalización es el siguiente paso en mejorar la experiencia del usuario al visitar un sitio. Aunque la web está creciendo de forma exponencial, la capacidad del individuo para leer y digerir el contenido está esencialmente fija. El potencial económico completo de la web no será descubierto a menos que se provea de tecnologías faciliten el acceso a los recursos web. Actualmente la personalización web es la aproximación más prometedora para solucionar este problema, y web mining, particularmente WUM, se considera como un componente crucial de cualquier sistema eficaz de personalización web [14].

Hoy existen varios trabajos acerca de aproximaciones a la personalización web, usando distintos lenguajes y herramientas. La aproximación más común y aceptada entre ellas es el uso de agentes inteligentes para realizar las tareas de personalización. En [15] se propone un sistema basado en capas, donde en cada capa existe un agente encargado de la respectiva tarea.

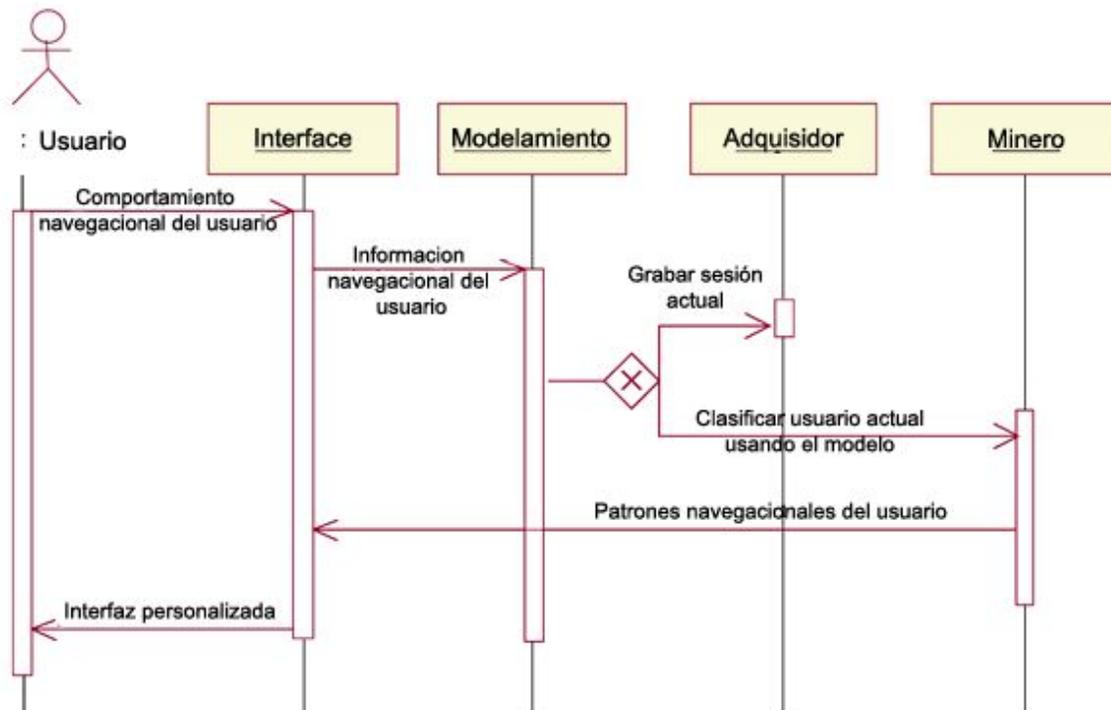


Ilustración 3.2: Diagrama de secuencia que muestra las interacciones entre las capas y agentes [15]

En [16] se propone una aproximación usando Autoclass C, dividiendo el trabajo en módulos e integrando la información obtenida en los logs de servidor con la información de registro de los usuarios.

3.3.8 WUM mediante huellas y sesiones

Una solución alternativa al registro en un log es el empleo de huellas. Una huella es un rastro o marca que queda registrada por parte del usuario al acceder a un determinado objeto. Las huellas se basan en la inclusión dentro de las páginas que se quieren controlar de una referencia (enlace) a un elemento adicional, que va a provocar una nueva petición por parte del cliente para acceder a ese elemento, con el efecto colateral de registrar el acceso a la página que la contiene. Este elemento adicional es un enlace a una página dinámica que recibirá como parámetros aquella información que se quiera registrar sobre la página que se está visitando y la almacenará, preferiblemente en una base de datos. Esta

página puede estar en un servidor distinto al que se quiere controlar. En este punto se introduce el concepto de *webmart* como almacén de la información de uso de un sitio web.

El proceso es el siguiente: un usuario hace una petición de una página a un servidor y éste se la entrega. Cuando el navegador presenta la página en pantalla, accede al enlace de la huella, enviando información del visitante que se incorpora al *webmart*, en la tabla operacional. Esta tabla contiene datos interesantes como el objeto (URL) accedido, la fecha y hora, la máquina (o su dirección IP) origen, el método de acceso, la página de procedencia, las cookies enviadas, la identificación del navegador cliente, la resolución de pantalla del cliente, etc. Otra ventaja es que ya no es necesario el log del sitio a controlar. Además, el análisis puede hacerse en tiempo real e incluso el servicio puede ser prestado por una organización independiente. [5]

3.3.9 Herramientas de WUM actuales

Existen varias empresas y herramientas en el mercado, que permiten analizar logs de servidor para encontrar información. Algunas de estas son:

- **Webmining.cl:** una empresa chilena que ofrece varios softwares y servicios orientados al web mining, entre ellos destacan WebMining Log Sessionizator XPert 7.10 y Web Usage Stats. La primera es una herramienta que permite encontrar reglas de asociación que describen el comportamiento de los visitantes del sitio. La segunda es un servicio para monitoreo remoto de sitios web. Con este servicio se puede monitorear la actividad que está teniendo el sitio y saber por ejemplo, cuáles páginas son las más vistas, desde cuáles sitios externos están llegando visitantes, desde que regiones o puntos de la red se está accediendo al sitio web, etc.
- **Webtrends:** es una empresa que brinda servicios de análisis web (web analytics) enfocados a la administración y al sector comercial, llevando a cabo estudios de mercadeo que prometen mejorar la relación con el cliente, la atracción de nuevos clientes y una mejora en el retorno de la inversión. Empresas como esta por lo general complementan el análisis de los logs con varias otras técnicas y herramientas para

lograr un análisis completo del estado en que se encuentra el sitio respecto a sus clientes. Tienen un enfoque de negocios más que de mejora de usabilidad.

- **AWStats:** Es un software gratuito de análisis de logs de servidores de varios tipos. El software está registrado con licencia GNU GPL y fue desarrollado en Perl. Ofrece un análisis bien completo respecto a las visitas que ha recibido el sitio web, entregando datos demográficos, de plataformas, de tráfico y lugares de acceso.
- **AlterWind Log Analyzer:** ALterWind ofrece un analizador de logs en su versión profesional pagada y lite gratis. Ambos proporcionan información respecto al tráfico, búsquedas en los buscadores y estadísticas de horas y días pico.
- **Analog:** Una de las herramientas gratuitas más utilizadas hasta el momento, provee información de tráfico mensual, diario y por horas. Soporta varios lenguajes y plataformas. Entrega información respecto a las búsquedas, páginas más visitadas, peticiones más comunes, etc.
- **Visitorator:** Software gratuito que ofrece muchas de las funcionalidades antes mencionada, pero con la diferencia de ofrecer agrupamiento o clustering visual mostrando grupos de usuarios con similitudes. La agrupación se hace desde un alto nivel hasta niveles más bajos, siguiendo un modelo de árbol desplegable.

3.3.10 WUM y la usabilidad en la práctica

Como ya se ha definido, usabilidad web es hacer un sitio de tal manera que los usuarios puedan encontrar lo que buscan de forma rápida y efectiva. Un sitio web con una buena usabilidad puede traer grandes beneficios en un negocio. Estudios de IBM muestran que por cada 1 libra (app 480 pesos) invertida en mejorar la usabilidad de un sitio hay un retorno de entre 10 y 100 libras [17]. Todo gira en torno a mejorar la experiencia del usuario al visitar el sitio. Con este fin ya se han realizado algunas evaluaciones utilizando

Web Usage Mining de diversas formas y con distintos resultados. A continuación se presentan casos de estudio donde se muestra el uso de WUM para mejorar sitios web y conocer información respecto a los usuarios.

a) Análisis de sesiones de la web del Cindoc[6]

En el Centro de Información y Documentación Científica (Cindoc) que es una unidad de servicios de información y documentación para la actividad científica, se realizó un estudio con WUM con el objetivo es estudiar la viabilidad de la minería de uso web y el análisis de sesiones como metodologías válidas para el estudio de la usabilidad y del diseño de la arquitectura web del Cindoc. Para ello, se analizan 2.748 sesiones obtenidas del archivo de transacciones del servidor central del centro durante el mes de octubre de 2003. La minería de datos se llevó acabo preparando los logs en tres pasos: limpieza de datos, identificación del usuario, identificación de la sesión y construcción de la sesión. Posteriormente se utilizó el software 3dv8 Enterprise (2004) para minar los datos.

Como resultado del análisis se pudo obtener datos concernientes a los accesos y a las sesiones. El análisis de accesos mostró que la gran mayoría provenía de los buscadores, seguido por accesos de enlaces en otros sitios y finalmente por la URL directa en el navegador (o en favoritos). Esto les llevó a la conclusión de la URL era difícil de recordar o que no era fácil de intuir. Se descubrió que los accesos más populares se efectuaban a los documentos con mayor información debido a los buscadores, y que estos documentos se encuentran en los niveles más profundos de navegación por lo que se sugirió maneras más fáciles de acceder a estos documentos desde los niveles más altos. Se descubrieron también secciones que se prestaban para confusión al comprobar que gran porcentaje de los usuarios que ingresaban a ellas decidieron regresar inmediatamente. Otro punto importante es que muy pocos accesos se realizaban hacia el área de investigación, el cual es uno de los objetivos prioritarios del sitio, por lo que fue sugerido difundir más el tipo de actividades que realiza el centro.

En este caso se realizaron las recomendaciones directamente desde los resultados del WUM sin posterior complemento por pruebas, o evaluaciones de usabilidad.

<i>Formas de acceso</i>	<i>Sesiones %</i>
Buscadores	1.060 42,78
Enlaces	757 30,55
Url	661 26,67
Total	2.478 100,00

Ilustración 3.3: Resumen de las formas de acceso en el Cindoc

b) WUM para la evaluación de sitios de e-learning

En este caso se optó por utilizar WUM debido a la dificultad de observar y evaluar el uso de sitios de aprendizaje basados en web, debido a su modelo de interacción distribuido y virtual. Además, las herramientas que brindan estos sitios no proveen la suficiente información para analizar las tendencias de uso de los estudiantes. El objetivo en este estudio es que los resultados brinden a los académicos información que ayude a mejorar estructura de sus cursos y el manejo de los recursos, que sea posible entregar una interfaz personalizada a los alumnos y que los administradores del sitio reciban información que les permita optimizar la eficiencia del sitio.

Un primer caso de estudio [18] para esta investigación se centro en WebCT (Web Course Tools) y se analizó un curso por 11 días, resultando en 15.953 accesos. Al análisis se le dio énfasis en la topología y navegabilidad. Un análisis de secuencias navegacionales comunes mostró que muchos estudiantes tenían problemas presentando tareas y terminaban recurriendo al Chat por ayuda. En este caso de estudio el análisis no fue terminado pero se concluyó que WUM presentaba gran funcionalidad y se presentarán proyectos posteriores que continúen este trabajo.

Un segundo caso de estudio [19], también centrado en WebCT realizado por la Universidad de Cantabria (España) analizó las actividades de los alumnos entre el 1º de Octubre y el 31 de diciembre del 2004 contando con 10.000.000 entradas, las que fueron reducidas a 2.206.024, identificando 3.235 alumnos y 322 cursos en un total de 48.691 sesiones. La información fue analizada mediante un sistema de minería no comercial hecho en C# y un sistema MOLAP de base de datos. Se descubrieron varias estadísticas acerca de los horarios

pico, tráfico y cantidad de uso en el tiempo pero el estudio de patrones no fue terminado. Se realizó un segundo estudio [20] más tarde en el WebCT con los alumnos del segundo semestre de un curso entre el 2005 y el 2006 donde se buscaba mejorar la estructura de los cursos, las categorías de las páginas, los tiempos de estudio, la secuencia del estudio, también obtener información respecto a datos demográficos de los estudiantes y revisar los objetivos fijados por los académicos.

c) WUM en un sitio de ventas multicanal online al por menor (Multi channel e-tailer)

En [21] se muestra como caso de estudio el análisis vía WUM de un sitio de ventas anónimo que cuenta con ventas online y 5000 tiendas de ventas al por menor en Europa. Los objetivos de este análisis son atraer usuarios para convertirlos en clientes y evaluar el sitio web comparado con otros canales de distribución. Para el estudio se utilizaron métricas web, análisis de clusters, reglas de asociación, minería de secuencias de navegación, cuestionarios y análisis de mercado. La fuente de datos fueron 92.467 sesiones de los de un periodo de 21 días, con 13.653 usuarios que compraron en línea en un periodo de 8 meses. Se analizaron 760.535 peticiones que fueron contrastadas con un mapa conceptual del negocio.

Para entender mejor los datos se trabajó en construir ciertas estrategias sobre el análisis de los datos definiendo restricciones como “un patrón es de interés si...”, “un cluster es de interés si...” y se definió claramente el dominio y el entorno de la compañía mediante los estudios de mercado.

Algunos de los resultados obtenidos fueron:

El sitio es usado preferentemente para adquirir información, pero las compras se hacen físicamente en la mayoría de los casos.

Los clientes pueden necesitar asistencia personal

90% de los clientes que compra más de una vez no cambia sus preferencias

Además se pudo obtener información demográfica respecto a los consumidores

En este estudio se puede ver que hubo una preocupación por complementar WUM con cuestionarios para averiguar las opiniones directas de los usuarios y entender su nivel de satisfacción subjetiva respecto al servicio.

d) *Otros ejemplos*

Para terminar con los casos de estudio se presentan tres ejemplos en los que el WUM ha sido utilizado para averiguar información respecto a los usuarios

El primer caso corresponde a la Arizona Health Sciences Library [22] donde se estudió data de logs de un periodo de 6 meses usando el método de regresión logística multinomial. La metodología de este estudio se basó en construir modelos teóricos del comportamiento de los usuarios y mediante WUM comprobar las probabilidades de cumplimiento de éstos modelos. Los resultados indican que los usuarios se comportan como se esperaba reafirmando los modelos navegacionales creados para el sitio.

El segundo caso es el análisis de un sitio web de una radio online [23]. En este estudio también se utilizaron estudios de mercado para conocer las categorías relevantes para el estudio. Los resultados mostraron que los usuarios se comportan online de manera similar a la manera tradicional, y la importancia de fijar objetivos de negocio claros.

El último caso [24] si bien no tiene resultado abiertamente (gratuitamente) publicados posee una idea interesante. Se trata de analizar mediante herramientas de WUM sitios de oferta de trabajos como Yahoo! Careers (HotJobs) y FlipDog para encontrar los requerimientos de calificación para el empleo de conductor de entregas. Este concepto puede ser ampliado para la búsqueda de la última información concerniente a un empleo, entregando estadísticas respecto a los requerimientos, número de ofertas promedio, aumento o disminución del salario medio ofrecido, etc. Este último caso nos muestra como el WUM también puede tomar distintos usos o caminos para brindar un mejor servicio al usuario.

Si bien en los casos anteriormente expuestos WUM se aplicó con algunas preocupaciones por la usabilidad, estudios de e-commerce y usabilidad web han propuesto que la observación del usuario juega un papel clave y que los análisis web como WUM no pueden

obtener la imagen completa del comportamiento del usuario. Se propone que para complementar el estudio son necesarias pruebas de usabilidad, que dan una mirada más al detalle de los usuarios [13].

3.4 Descripción de las pruebas a realizar

Muchos tipos de evaluaciones, pruebas e inspecciones de usabilidad han sido desarrollados a través de los años con el objetivo de crear mejores productos software y lograr mayor satisfacción del usuario. Con el crecimiento del uso de las interfaces web también se ha visto el crecimiento de interés en métodos de evaluación que se adecuen mejor a este tipo de interfaces, en [25] se puede encontrar un amplio espectro de actividades enfocadas a la evaluación de usabilidad web y su respectiva descripción. Para el presente proyecto se han elegido cuatro actividades debido a su relativa facilidad de realización, bajo costo y adecuación con las interfaces web.

3.4.1 Ordenamiento de tarjetas (Card Sorting)

Card Sorting es una técnica enfocada en obtener información respecto a la estructura y organización contenida en el sitio desde el punto de vista de los usuarios, el objetivo principal es aprender acerca de cómo ellos piensan respecto al contenido del sitio y como organizarían la información. Esto ayuda a desarrollar una estructura del sitio que tenga sentido para los usuarios.

Para realizar esta técnica los participantes deben revisar ítems o tópicos del sitio web escritos en tarjetas y organizarlos en grupos según su relación, para crear categorías que ellos mismos pueden nombrar. Existen dos tipos de Card Sorting, el abierto y el cerrado. En el card sorting abierto los participantes pueden agrupar las tarjetas en categorías que ellos encuentren coherentes y lógicas, pudiendo crearlas ellos mismos, en el card sorting cerrado los participantes deben agrupar las tarjetas en categorías predefinidas. La realización de cada uno depende si se está tratando de descubrir las categorías o comprobar la efectividad de las que ya se han descubierto. Además cabe comentar que card sorting es útil tanto antes de la puesta en marcha del sitio como después de un tiempo de uso. Es posible aplicar está

técnica en la etapa de diseño para ayudar en la creación de la arquitectura y se puede usar posteriormente para probar la arquitectura existente.

Los pasos a seguir para realizar esta técnica se describen como sigue:

1. Se reparten las tarjetas en una mesa o escritorio con espacio suficiente para que el participante (o los participantes) pueda trabajar con comodidad, cada tarjeta debería incluir sólo un tópico del sitio web.
2. Los participante deben ordenar las tarjetas en pequeñas pilas según su similitud o de alguna manera que tenga sentido para ellos. Debe ponerse a su disposición material que les permita agrupar las tarjetas cómodamente como elásticos, sobres u otros. Si los participantes desean (o si es requerido) pueden agrupar los grupos ya formados en categorías más altas.
3. Una vez que los grupos están listos se le pide a los participantes que nombren cada categoría, para esto deben tener disponible el material adecuado.

El método anteriormente descrito es utilizado para realizar card sorting en persona, pero éste método puede ser aplicado también a distancia, mediante e-mails o software como EZsort, WebCAT o WebSort.

3.4.2 Evaluación Heurística

Como se mencionó con anterioridad este tipo de inspección de usabilidad es probablemente el más conocido y consiste en uno o más expertos en HCI explorando un sistema, identificando problemas de usabilidad y clasificando cada problema encontrado como una violación de uno o varios principios de usabilidad o heurísticas. La lista de heurísticas a utilizar debe ser detallada previamente como un documento previo necesario para esta actividad junto con los objetivos del proyecto, la audiencia objetivo y las partes del sitio que se van a analizar. Los evaluadores deben tener conocimiento de esta información antes de poder realizar el proceso.

Una sesión de evaluación dura en promedio entre 60 a 90 minutos. El evaluador puede recorrer el sitio según su criterio, patrones de navegación de ejemplo o realizando tarea específicas [32]. Durante este recorrido el evaluador identifica los problemas y los relaciona con una o más heurísticas. Una vez completada la revisión individual de cada evaluador, estos se juntan y presentan los resultados de su evaluación, presentando los problemas, recomendaciones y aspectos positivos encontrados. Un encargado o coordinador compone una lista de los problemas encontrados por todo el conjunto de evaluadores y se la entrega a cada evaluador para que este asigne a cada problema una puntuación de severidad, frecuencia y criticalidad. Las puntuaciones de severidad van desde 0 hasta 4 donde los niveles se pueden expresar como:

0. Esto no es un problema
1. Sólo un problema cosmético
2. Problema menor
3. Problema mayor
4. Problema catastrófico.

La frecuencia es el número de ocurrencias encontradas de cada problema y la criticalidad puede ser calculada como $\text{severidad} + \text{frecuencia}$. Al terminar de calcular estos datos para cada problema se ordena la lista general de problemas en orden descendente para obtener un ranking ordenado de los problemas y quizás sus probables soluciones propuestas o recomendaciones.

3.4.3 Pruebas de usabilidad

Si bien el término pruebas de usabilidad puede usarse de forma más general y abstracta, es a la actividad de tener usuarios “reales” probando el sitio a lo que se le llama pruebas de usabilidad en la mayoría de los casos. En una prueba de usabilidad se tiene a usuarios realizando tareas típicas o escogidas en el sitio mientras que se recolectan datos de ésta interacción mediante la observación directa, anotaciones, grabaciones y otros métodos.

Al realizar pruebas de usabilidad se debe tener presente que los usuarios influyen directamente en el resultado de las pruebas, por lo que es conveniente tener usuarios que sean parte de la audiencia objetivo del sitio, que sean una muestra representativa y que los participantes sepan que se está probando al sitio y no sus capacidades como usuario. Se debe tener en cuenta objetivos claros para cada prueba, ya que estos fijarán las tareas y escenarios que los usuarios deberán realizar. En una sesión de pruebas de usabilidad se busca encontrar generalmente problemas de usabilidad del sitio, información respecto a la performance de los usuarios y el nivel de satisfacción de los usuarios con el sitio.

El proceso de pruebas de usabilidad una vez que los usuarios han sido reunidos puede ser descrito como sigue:

1. En una reunión inicial (personal o grupal) se explica el proceso, se describe el sitio y se responden preguntas
2. El usuario pasa a sentarse frente a un PC que tiene una página de comienzo para su navegación en el sitio. Durante 30 a 60 minutos aproximadamente el usuario interactúa con el sitio desarrollando las tareas que han sido programadas para él en un documento. Dependiendo de lo que se quiera estudiar pueden variar las tareas y la forma de completarlas, de esta manera se puede por ejemplo dar una lista de preguntas, las cuales para ser respondidas por el usuario necesitarán que realice ciertas tareas en el sitio. También puede asignársele una lista de tareas y observar el desarrollo de éstas en el video de la sesión del usuario. Generalmente métodos de captura automatizados son usados para documentar las sesiones de los usuarios tales como grabación de la pantalla en tiempo real, anotación de las acciones en logs, grabación de las expresiones faciales mediante web cam y grabación o anotación de los comentarios del usuario
3. Al finalizar la realización de las tareas se puede tener una discusión abierta (individual o grupal) con el usuario para obtener su opinión respecto a asuntos que se consideren importantes. También se pueden realizar encuestas o cuestionarios.

3.4.4 Cuestionarios

Los cuestionarios son el método más utilizado y más simple para conocer elementos subjetivos como la satisfacción del usuario en el uso del sitio y el nivel de agrado con la interfaz. También son útiles para recolectar información demográfica y de perfil de usuario. Dependiendo del objetivo del cuestionario se puede buscar respuestas a las siguientes incógnitas:

- a. Quienes usan el sitio
- b. Que buscan los usuarios en el sitio
- c. Que tan satisfechos están los usuarios con el sitio
- d. Que experiencia tiene los usuarios con el sitio o con sitios similares
- e. Que les gusta más/menos a los usuarios del sitio
- f. Que ideas, sugerencias o comentarios pueden aportar los usuarios

Si bien la realización de cuestionarios tiende a ser una actividad por si sola, se puede realizar cuestionarios dentro de otras actividades como complemento. Las oportunidades más comunes para realizar cuestionarios son antes y después de actividades como card sorting y pruebas de usabilidad, para recoger las opiniones del usuario como resultado de la realización de la actividad y para conocer al usuario (su perfil) que ha realizado la actividad. Al igual que las otras actividades mencionadas, los cuestionarios pueden realizarse antes o después de la puesta en marcha del sitio. Los pasos a seguir para realizar una encuesta pueden describirse de la siguiente manera:

1. Decidir el objetivo de la encuesta
2. Decidir el público objetivo de la encuesta
3. Desarrollar el cuestionario, este debe ser relativamente corto, de preguntas fáciles de responder y con una lógica que el usuario entienda.
4. Realizar la actividad

En [25] se pueden encontrar más recomendaciones con respecto a los cuestionarios y a la realización de éstos, la mayoría de las cuales hace referencia a hacer las preguntas adecuadas de la manera adecuada. Los cuestionarios pueden realizarse de forma presencial o remota, mediante e-mails o páginas web (online), siendo éste último método uno de los más masivos, pero que garantiza menos una respuesta a conciencia y verdadera del participante. Los cuestionarios son una actividad muy difundida y antigua por lo que existen varios ejemplos disponibles (pocos gratuitos), empresas dedicadas a realizarlos y cuestionarios que ya casi son un estándar en HCI, estos son WAMMI, QUIS, SUMI y otros. En [26] se puede encontrar ejemplos de varios cuestionarios incluyendo los anteriores.

3.4.5 Web Usage Mining

Anteriormente en el informe se ha descrito el proceso de WUM con sus ventajas y problemas, ahora se presentarán algunos asuntos concernientes al potencial del uso de los logs de servidor como fuente de datos para evaluaciones de usabilidad, analizando lo que puede esperarse y lo que no del procesamiento de estos logs. Posteriormente se describirán las herramientas que se usarán en la etapa de pruebas.

A diferencia de otras fuentes de información como cuestionarios y pruebas, los logs de servidor sólo prestan utilidad después que el sitio ha recibido una cierta cantidad de uso, esto es a la vez una ventaja inherente de los logs, ya que son una fuente de datos fácil de adquirir debido a que los servidores pueden generarlos por defecto y presentan datos acerca de uso real que ha recibido el sitio, aunque la validez de estos datos es cuestionada debido a tres factores que se presentarán más adelante. Los logs de servidor pueden ser analizados para obtener información cuantitativa pero escasamente cualitativa, ya que los datos obtenidos son en su mayoría estadísticos y numéricos. Si bien algunas herramientas publicitan que datos cualitativos y estratégicos pueden ser revelados, estos pueden llevar frecuentemente a asunciones erróneas e inferencias falsas si no se comprueban con otros métodos como pruebas de usabilidad.

En [31] se describen las siguientes opciones, considerando lo que se puede y no se puede obtener.

Se puede obtener estadísticas de uso útiles como:

Número de peticiones al servidor
Cantidad total de transacciones y peso de los datos transferidos
Visitas por página, número de peticiones por IP
Número de peticiones por dominio
Estado de las transacciones
Totales y promedios relativos a duraciones, horas y fechas
Sitios desde donde los usuarios accedieron al presente sitio
Navegadores y versiones de éstos

Estos datos son más seguros y básicos y presentan menos porcentaje de errores y por ende menos dudas respecto a su veracidad. Otra información si bien es ofrecida por herramientas y servicios, tiene un menor grado de credibilidad debido a los siguientes tres factores: datos que no capturan los logs, datos que se capturan de forma incompleta e inferencias “blandas” o débiles obtenidas de los datos del log.

a) Datos que no capturan los logs

Servidores Proxy, cachés y el uso de los botones adelante y atrás del navegador comúnmente dejan huecos en el registro de la navegación de los usuarios lo que hace incierta la identificación de:

<p>Individuos: a menos que la sesión de usuario sea iniciada mediante el acceso con contraseña y nombre de usuario (log in) las entradas en los logs no pueden identificar con seguridad a un usuario en particular.</p> <p>Número exacto de usuarios: ya que una IP puede no tener necesariamente una relación uno a uno con los usuarios. Es posible que una IP pueda pertenecer a varios computadores o que un usuario pueda tener varias IPs en el tiempo.</p> <p>Objetivos y metas de los usuarios: ya que las razones por las cuales se accedió a los archivos no queda registrada.</p>

Sitio al que el usuario fue después: esta transacción queda grabada sólo en el servidor del sitio siguiente.

b) Datos que se capturan de forma incompleta

Número exacto de peticiones reales: es difícil saber cuantas de estas peticiones fueron realmente hechas por usuarios que querían acceder a los documentos y no por robots o indexadores, también es difícil estimar cuantas de las peticiones de los usuarios fueron recibidas por algún servidor intermedio, que puso a disposición del usuario una copia del caché.

c) Inferencias “blandas” o débiles obtenidas de los datos del log

De los problemas anteriormente mencionados se puede concluir que las siguientes inferencias pueden estar erradas o deformadas:

Número de peticiones es igual a la cantidad de uso

Las sesiones de usuario pueden ser bien demarcadas y aisladas

En una sesión se puede calcular exactamente el número de páginas vistas por sesión, tiempo promedio de vista por página, páginas de entrada y salida más usadas

Información demográfica exacta puede ser recolectada

A pesar de los inconvenientes, los logs de servidor siguen siendo considerados una valiosa fuente de información respecto al uso del sitio y es parte del objetivo del presente proyecto analizar empíricamente el potencial de esta técnica como a poyo un proceso más amplio de evaluación de usabilidad.

3.4.6 herramientas seleccionadas

Con el fin de realizar WUM sobre los sitio a evaluar se han elegido 2 herramientas debido a sus características.

1. Alterwind Log AnalyzerPro

Es un software no gratuito que ofrece un demo gratuito totalmente funcional por 30 días, trabaja con una interfaz gráfica amigable y fácil de usar mostrando un análisis bastante variado con resúmenes estadísticos, datos cuantitativos y visualizaciones gráficas. Es compatible con los tipos de logs más importantes tales como NCSA, XFL, ELF, CLF, W3C y otros.

En [27] se puede encontrar un reporte de ejemplo entregado por la herramienta, el cual contiene en resumen:

Un histórico de cantidad de visitas por meses, días y horas

Una lista de locaciones desde donde se accede al sitio junto con el volumen de peticiones

Un análisis de las visitas al sitio por servidor, usuarios autenticados y robots

Análisis de duración de las visitas

Rankings de formatos de archivos más pedidos, páginas más solicitadas, buscadores más usados y los términos escritos en ellos para realizar la búsqueda, sistemas operativos más comunes y navegadores web.

Una integración de los resultados mediante hipervínculos en cada página de resultados

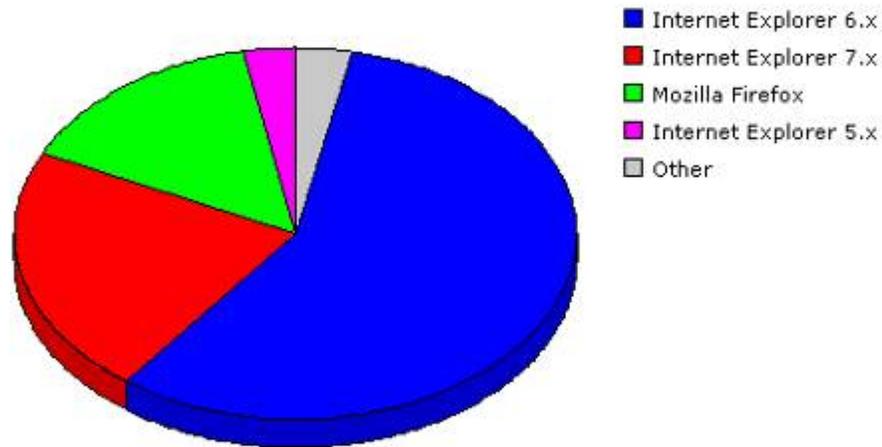
La posibilidad de expandir ciertos resultados para obtener más detalle

Gráficos de ciertos resúmenes

Most Used Browsers by Visitors

Time Period:

02-09-2007 4:12:49 - 04-10-2007 10:16:11



Rank	Browser	Hits	Visitors
1	Internet Explorer 6.x	9629043 62.50%	149939 56.95%
2	Internet Explorer 7.x	3654344 23.72%	57887 21.99%
3	Mozilla Firefox	1766947 11.47%	38602 14.66%
4	Internet Explorer 5.x	74308 00.48%	8141 03.09%

Ilustración 3.4: Ejemplo de reporte de Alterwind Log Analyzer

2. Analog

Es una de las herramientas gratuitas de web mining más conocidas y puede incluso competir con softwares comerciales más potentes y difundidos, en [28] se muestra que cuenta con un bajo pero presente grado de preferencia entre los utilizadores de herramientas de web mining. Esta herramienta ofrece al igual que la anterior un conjunto de estadísticas y datos cuantitativos, con reportes en varios idiomas y con gran compatibilidad de sistemas operativos y formatos de logs, permitiendo incluso personalizar el formato. En [29] se puede ver un reporte de ejemplo, el cual contiene básicamente ítems similares a la herramienta anterior, con la inclusión de análisis de búsqueda interna en el sitio. El

fundamento de utilizar dos herramientas distintas con reportes similares es verificar la confiabilidad de los resultados.

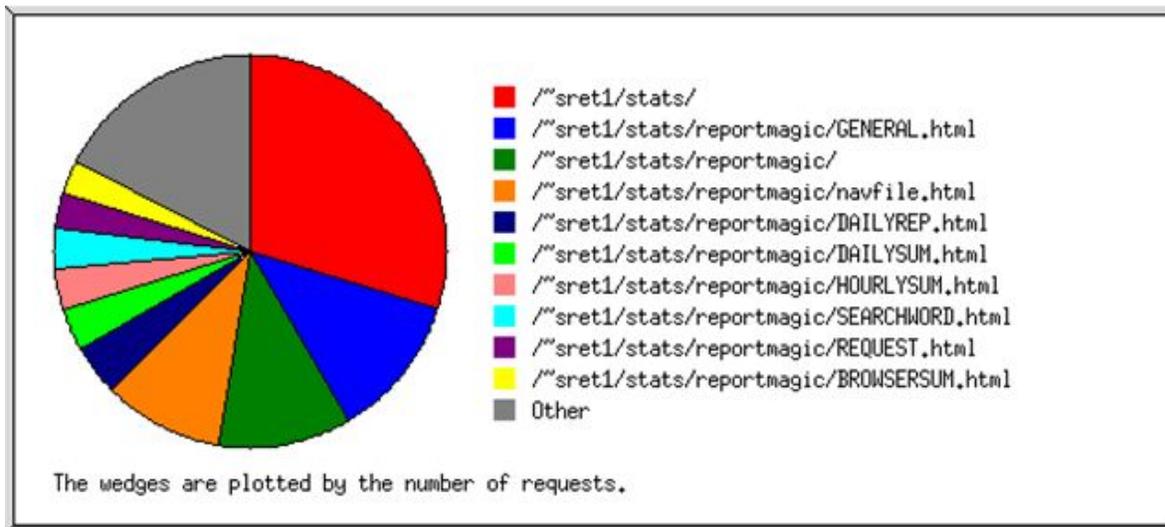


Ilustración 3.5: Ejemplo de reporte de Analog

3.5 Alcance del experimento

El presente proyecto toma sujetos de prueba en su mayoría de la región de Valparaíso con perfil de estudiantes universitarios, siendo la mayoría estudiantes de la PUCV. El proyecto busca obtener resultados a pequeña escala, sobre un sitio web en particular, obteniendo resultados concretos respecto a la utilidad de WUM en un proceso más complejo de evaluación de usabilidad. Se espera que los resultados de la investigación sean lo suficientemente contundentes como para obtener inferencias válidas y realistas.

3.6 Plan de la evaluación

Se presenta a continuación el plan de la evaluación de usabilidad que se ha efectuado durante el presente proyecto. Este plan puede haber sufrido ciertas modificaciones dependiendo de cómo acontecieron los eventos, pero su estructura general se ha mantenido.

3.6.1 Definición de objetivos

El objetivo principal de una evaluación de usabilidad (independiente del o los sitios sobre los que se realice) es descubrir problemas de usabilidad presentes en el sitio. Es relevante también adquirir información de los usuarios respecto a su interacción con el sitio (user feedback) y de sus preferencias y modelos mentales.

Cada una de las actividades a realizar, tal como se les describió, sirve para obtener distintos tipos de información y complementar la evaluación. En card sorting se investiga que estructura del sitio tiene más sentido para los usuarios, evaluación heurística analiza el sitio desde un alto nivel para verificar que cumpla con ciertos principios, pruebas de usabilidad obtienen información respecto a la interacción del usuario con el sitio, los cuestionarios se enfocan en conocer el perfil y opinión del usuario y web mining brinda información estadística y cuantitativa respecto al uso histórico del sitio.

En particular, se observarán los siguientes aspectos principales en los sitios analizados:

Facilidad y efectividad de la navegación: observando si los usuarios encuentran lo que buscan de forma fácil, si la estructura del sitio es lógica para ellos y si pueden entender las maneras de acceso (menú, links, etc.)

Utilidad del contenido: comprobando si el sitio tiene lo que los usuarios esperan de este y si el contenido está dispuesto de una forma útil para el usuario.

Efectividad de la presentación de la interfaz: observando si el diseño y gráficas ayudan o confunden al usuario, si son de su agrado o si le causan rechazo.

Éxito en la realización de las tareas: tomando el tiempo de realización, tasa de errores y satisfacción al realizar las tareas.

3.6.2 Marco temporal

Las etapas del proceso de evaluación se realizaron en dos ciclos, uno por cada sitio web. En el primer ciclo se realizó el proceso de evaluación de manera clásica, sin WUM. En el segundo ciclo WUM se realizó con anterioridad a las demás actividades viendo como este sirvió para guiar la realización de éstas y también como los resultados se complementan.

Las etapas a seguir fueron las siguientes

1. Determinar metas, marco de tiempo, audiencia, plan de reclutamiento, disponibilidad del sitio. (Agosto)
2. Reclutar sujetos de pruebas, probar las pruebas y hacer ajustes (principios de Septiembre)
3. Realizar las pruebas en dos ciclos (Septiembre y Octubre)
4. Compilar y analizar los datos, sacar conclusiones (Octubre, Noviembre).

3.6.3 Audiencia y plan de reclutamiento

La audiencia como se describe en el alcance del experimento son en su mayoría estudiantes universitarios de la zona de Valparaíso. El plan de reclutamiento cuenta con dos estrategias, reclutar gente que esté dispuesta a colaborar de forma desinteresada y reclutar gente mediante el ofrecimiento de un regalo o premio. Los sujetos de prueba se asignan a distintos horarios dependiendo de su disponibilidad y la disponibilidad del examinador y los recursos. Finalmente se construye una malla o carta con los horarios y las actividades.

3.6.4 Desarrollo de los documentos necesarios

Se utilizan básicamente los siguientes documentos en las actividades de evaluación:

Consentimiento: para dejar constancia de la voluntad del usuario para ser sujeto de pruebas y de estudiar los datos que se obtengan de su actividad.

Cuestionario de entrada: para obtener datos demográficos y de perfil de usuario.

Listado de tareas: para las pruebas de usabilidad, estas serán las tareas que el usuario deberá realizar.

Cuestionario de salida: para obtener la opinión del usuario después de realizar la actividad.

Cuestionarios específicos: pueden hacerse cuestionarios para usuarios que ya conozcan el sitio con anterioridad y que deseen expresar su opinión respecto a éste.

Listado de heurísticas

Documento de objetivos para evaluadores heurísticos

Para construir los cuestionarios se utilizan los ejemplos entregados en [26], personalizando las preguntas que sea necesario para que los cuestionarios y tareas sean adecuadas para cada sitio a analizar. Es importante conocer cuales son los puntos relevantes de cada sitio y tener en lo posible acceso a los planes de los desarrolladores del sitio, para conocer como ellos esperaban que el sitio fuera navegado.

3.6.5 Prueba y práctica de las actividades

Antes de realizar las actividades, estas deben ser probadas y aprobadas. Se debe realizar todas las actividades por el propio examinador y algunos otros usuarios para saber si las tareas son efectivamente realizables, para tener un primer estimado de los tiempos de realización de las actividades y probar la efectividad de las pruebas. Con respecto a web mining se debe asegurar que los servidores a analizar cumplan con los requisitos y que los logs estén en el formato adecuado, también se debe comprobar que las herramientas funcionen correctamente y que los resultados sean coherentes.

3.6.6 Realización de las actividades

Las cinco actividades a realizar pueden ser desarrolladas en distinto orden dentro de un marco de tiempo límite para todas y su realización dependerá de la disponibilidad del examinador y los sujetos de prueba. El proceso general para la realización de estas actividades ya ha sido descrito, así que a continuación se detallarán algunas particularidades para el desarrollo de las actividades en el presente proyecto.

a) Web Usage Mining

Del servidor del sitio web a analizar se extraen los logs generados en una cantidad de tiempo razonable, la que puede ir de dos semanas a un mes, dependiendo de la cantidad de

visitas que reciba el sitio. Los logs del sitio son analizados con las herramientas antes descritas para obtener los datos cuantitativos, resúmenes estadísticos y grafos de navegación correspondientes.

b) Ordenamiento de tarjetas

Esta actividad puede ser realizada en cualquier sala disponible que tenga un escritorio o mesa con espacio suficiente. Antes de comenzar la actividad se entrega al participante un consentimiento y un cuestionario de entrada. Para cada tarjeta se utilizará un tópico del sitio web correspondiente y se realizan actividades separadas para cada sitio web. Al finalizar la actividad se entrega un corto cuestionario de salida.

c) Evaluación heurística

Una cantidad de tres a cinco evaluadores es el número óptimo para realizar esta actividad, los evaluadores son estudiantes universitarios de ingeniería civil informática con conocimientos en HCI. A cada evaluador se le hace entrega de un consentimiento especial para esta actividad, una lista con las heurísticas que deberá observar en el sitio web y un documento que describe los objetivos de la actividad, la audiencia objetivo del sitio e información valiosa que debe ser tomada en cuenta al momento de analizar, tal como la que puede obtenerse de los desarrolladores del sitio y de la actividad de WUM.

Al finalizar la evaluación del sitio se integran los resultados de cada evaluación y se pide a los evaluadores que asignen sus indicadores de severidad, frecuencia y criticalidad (o criticidad) a la lista total de problemas.

d) Pruebas de usabilidad

Esta actividad es probablemente la más compleja. Se comienza dando la bienvenida al sujeto de prueba en el horario que se tenga acordado con él, se le explica brevemente el objetivo de la actividad y se responden preguntas. Se le hace entrega del consentimiento para que lo firme y se le hará pasar a su PC correspondiente en el laboratorio de usabilidad donde debe responder un cuestionario de entrada y posteriormente realizar tareas en el sitio de una lista de tareas específicas para cada sitio.

Un sujeto de prueba analiza solo un sitio web por sesión y se graba su interacción con el mismo mediante un software disponible en el laboratorio de usabilidad, no se grabará directamente al usuario mediante cámaras web u otro dispositivo. Al finalizar las tareas se le hace entrega de un corto cuestionario de salida. Dependiendo del acuerdo al que se haya llegado, se le hace entrega al participante de un premio o incentivo.

e) Cuestionarios

Esta actividad se realiza antes y después del ordenamiento de tarjetas y pruebas de usabilidad, y de forma separada sobre participantes que tengan experiencia previa con cualquiera de los sitios a analizar. El objetivo principal de los cuestionarios es recoger la opinión del usuario respecto al sitio y conocer un mínimo perfil del usuario.

3.7 Descripción de los sitios web a analizar

a) La Clínica X

La Clínica X (llamado sólo así por motivos de confidencialidad) es uno de los principales y más respetados centros de atención médica del país y su sitio web cuenta con varias y distintas áreas destinadas a ofrecer distintos tipos de servicios e información. Las funcionalidades más importantes del sitio son:

Informativos acerca de las distintas áreas de la clínica
Informativos útiles para usuarios que busquen información médica
Servicios de búsqueda interna
Noticias
Envío de preguntas médicas a especialistas
Suscripción a un newsletter

La clínica cuenta con varios cursos gratuitos que se realizan en sus instalaciones, cuenta además con un canal interno de televisión y una revista que circula gratuitamente. Todas estas actividades forman parte de una filosofía de mejorar la calidad de vida y salud de las

personas. El sitio web se suma como un canal informativo integrado en la filosofía mencionada difundiendo información médica útil para emergencias, “que hacer en caso de”, preguntas al doctor y artículos de interés. La clínica también cuenta con un área de docencia e investigación que realiza cursos, seminarios y otras actividades por lo que el sitio también cuenta con un sector dedicado a promover esta área.

Por otra parte el sitio web busca dar una ventaja especial a sus clientes proporcionándoles una cuenta en el sitio. Además se encuentra la funcionalidad de buscar a un doctor miembro de la clínica mediante su apellido y especialidad, lo que útil cuando uno conoce ciertos especialistas como referencia. También es posible enviar un saludo a una persona hospitalizada en la clínica mediante el sitio web. Por último el sitio posee la información necesaria para que los usuarios se puedan contactar con la clínica vía mail o fono y sepan donde se encuentra y que ofrece.

La información del sitio se encuentra dividida principalmente por áreas de interés, organizando los temas relacionados bajo un área específica, como por ejemplo “recién nacidos”, “urgencias”, “maternidad”, etc. Existe un sistema de búsqueda interna que permite a los usuarios encontrar información que buscan dentro del sitio sin tener que recorrerlo completamente, los resultados de la búsqueda se encuentran organizados por el área donde se encuentran las ocurrencias.

El público objetivo de La Clínica X es el sector A1 de la población, especialmente aquellos residentes en la región y ciudad donde esta se encuentra. Los principales usuarios del sitio web son:

Personas en busca de atención médica.

Personas en busca de información médica.

Clientes de la clínica (anteriores y actuales pacientes, familiares de los pacientes, etc.)

Docentes, estudiantes y profesionales de la salud.

Es el sitio de la Clínica X el cual se analizará de la manera “clásica”, es decir, sin WUM.

b) Sitio de la Pontificia Universidad Católica de Valparaíso

La PUCV es una universidad estatal ubicada en la quinta región, alberga un gran número de carreras y estudiantes, personal docente y administrativo. Es una de las universidades más antiguas del país y tiene gran prestigio a nivel nacional. El sitio web de la PUCV se caracteriza por su gran cantidad y variedad de contenido, tratando de dar acceso desde la página de inicio a la mayoría de las áreas que la universidad comprende. Para esto cuenta con varios menús y accesos directos. Las áreas del sitio pueden resumirse en:

Noticias e informaciones de interés respecto a seminarios, fechas importantes, actividades culturales y otros.

Páginas para las distintas áreas de la universidad (finanzas, asuntos estudiantiles, pastoral, etc.)

Páginas para las distintas carreras

Información geográfica y de ubicación de las distintas sedes

Funcionalidades para los alumnos y docentes (navegador académico, aula virtual, web mail, etc.)

Vínculos a otros sitios

La universidad cuenta con gran variedad de áreas y actividades las cuales pueden estar dirigidas a los alumnos, docentes, ex alumnos o a la comunidad en general. La diversidad inherente del mundo universitario hace que el sitio sea complejo y extenso, dando acceso a todas estas áreas.

El sitio web de la PUCV está construido de forma tal que sea la unión de varios sitios menores pertenecientes a las distintas áreas y actividades de la universidad. Así por ejemplo, desde la página inicial es posible acceder al sitio de la Dirección de Deportes y Recreación, el cual contiene su vez varias páginas relacionadas a su área. De esta forma la página inicial del sitio es un portal a los distintos sitios pertenecientes a la PUCV.

El público objetivo del sitio web de la PUCV es principalmente:

Alumnos, docentes y funcionarios pertenecientes a la universidad

Personas que buscan estudiar alguna carrera universitaria

Sobre este sitio web se realizará el proceso de evaluación incluyendo WUM.

3.8 Propuesta de cómo integrar WUM en un proceso de evaluación de usabilidad

Web Usage Mining es un tema recurrente y popular cuando se habla de usabilidad web y marketing en los sitios web, muchas compañías ya están realizando WUM en bases regulares y algunos estudios se encuentran publicados mostrando los resultados positivos obtenidos en su caso particular [6] [19] [22] [23]. Sin embargo también existen detractores de la técnica que dicen que la información en los logs es poco fiable y que no puede ser usada para evaluar la usabilidad de un sitio.

3.8.1 Discusión teórica respecto a la validez de WUM

A lo largo del presente informe se han descrito las ventajas y desventajas más reconocidas de Web Usage Mining, se ha explicado las causas de las desventajas lo más objetivamente posible. Estas desventajas o problemas que presentan los logs de servidor con respecto a los datos que capturan y al análisis que se puede hacer sobre ellos ha llevado a crear una visión pesimista en ciertos sectores del área de administración de sitios web. Hasta el momento, la mayoría de los autores consultados concuerdan en que los logs de servidor son una fuente de datos útil, en distintos grados. Analizando estos se obtiene un conjunto de estadísticas que son difíciles de conseguir de otra forma y que reflejan el uso del sitio independientemente de que este sea por robots, usuarios ficticios o usuarios reales.

La mayoría se mantiene escéptico a la hora de hacer inferencias de usabilidad directamente desde esta información y aconseja discreción a la hora de tratar de responder preguntas de usabilidad usando sólo la información entregada por los logs. Esto no debe parecer extraño, ya que como se ha mostrado en el presente informe, los datos entregados por las herramientas de análisis de WUM son datos cuantitativos, resúmenes e índices estadísticos que poco dicen acerca de las metas de los usuarios en el sitio.

Se pueden encontrar variadas opiniones respecto al uso de los logs que varían desde el pesimismo sobre su uso hasta el apoyo y promoción de este. En [33] se puede encontrar “... ¿No existe alguna otra manera de obtener buenos datos sobre que tan popular es el sitio?, en una palabra, no...” “los logs de acceso son buenos revelando el numero de hits versus el número de accesos... admitidamente, eso no es mucho como para construir una estrategia de mercado basada en eso, pero es probablemente lo mejor que se puede obtener”. En [31] dice “...Por ahora, es esencial recordar que la verdadera medida de uso y el verdadero número de usuarios individuales de un sitio, se mantienen desconocidos. Sin embargo, compilados de la manera correcta los logs de servidor web todavía pueden proveer algunos significativos indicadores estadísticos de uso del sitio”. Como último ejemplo, se puede encontrar en [30] “... muchos críticos dicen que estas y otras debilidades hacen que los archivos logs no sean de utilidad en estudios de usabilidad. Tec-Ed no concuerda, estas y otras debilidades simplemente previenen a los archivos logs de servir como base única para cualquier recomendación de usabilidad. De hecho, como los únicos datos fácilmente accesibles acerca del uso real de los sitios web, los archivos logs son extremadamente valuales”.

El enfoque que se ha tomado en el presente proyecto busca integrar WUM en un proceso de evaluación de usabilidad web más tradicional, con pruebas que involucren usuarios y observación directa. Mediante este enfoque se puede estudiar de forma práctica si los datos que se obtienen mediante WUM son de utilidad efectiva en el proceso de pruebas y si son lo suficientemente confiables como para incluir está técnica dentro del arsenal de técnicas de pruebas de usabilidad que se utilizan con frecuencia.

Se puede agregar que en el caso más optimista, un análisis ideal de los logs de servidor entregaría [30]:

Quienes visitan el sitio web

Los caminos que toman los usuarios

Cuanto tiempo los usuarios pasan en cada página del sitio

Por donde se van los usuarios del sitio y por donde entran

El nivel éxito que tienen los usuario al interactuar con el sitio

Estos aspectos son entregados en mayor o menor medida por algunas herramientas, pero por los motivos entregados anteriormente en este informe, donde se ha descrito los problemas que hace que los logs queden cortos a la hora de un análisis completo, estos ítems no pueden ser descubiertos de forma 100% confiable o correcta, por lo que se hace imperativo complementarlos o corroborarlos con otros análisis y evaluaciones. En el caso particular de este proyecto WUM es una de cinco actividades realizadas como parte de una evaluación de usabilidad de distintos sitios web.

3.8.2 Como usar WUM dentro del proceso de evaluación de usabilidad

Habiendo propuesto que WUM necesita ser integrado en un proceso más completo de evaluación de usabilidad, ya que por si solo no entrega la información necesaria para hacer inferencias y recomendaciones sólidas, queda ahora observar de que forma WUM puede ser integrado a este proceso de evaluación.

Se proponen en este proyecto dos hipótesis o enfoques no excluyentes, sino más bien complementarios, estos son:

WUM debe ser realizado con anterioridad a las demás actividades, ya que los resultados obtenidos pueden ser de utilidad para guiar la realización de las otras actividades.

Los resultados de todas las actividades encuentran un complemento en la etapa de análisis, es decir, los resultados de WUM se complementan con los resultados del resto de las actividades una vez que se pasa a analizar la información obtenida.

Sea cual sea el enfoque, parece conveniente de todas maneras realizar la minería con anterioridad a las demás actividades, ya que ambos enfoques pueden combinarse para permitir que WUM ayude en los siguientes aspectos:

1. Desarrollar las preguntas que se usarán en otras actividades
2. Desarrollar hipótesis que se pueden probar mediante otras actividades

3. Probar y apoyar hipótesis que se desarrollen en otras actividades
4. Obtener datos para mejorar el sitio web en aspectos técnicos

1. Desarrollar las preguntas que se usarán en otras actividades:

Mediante los datos obtenidos con WUM se puede enfocar ciertas preguntas para los cuestionarios de las distintas actividades. Se puede por ejemplo encontrar que cierta página X que es relevante para los dueños del sitio es poco accedida en comparación con el resto. Basados en esta información se puede buscar en la actividad de ordenamiento de tarjetas si los contenidos ofrecidos por la página X se encuentran dispuestos correctamente según las agrupaciones que hagan los usuarios y se les puede consultar si la agrupación que presenta X les parece coherente, comparado con la que realizaron ellos.

Se puede prestar especial atención a la página X en la evaluación heurística, expresándole a los evaluadores que dentro de los objetivos del sitio está difundir el contenido de la página X y que esta es en realidad poco accedida, y preguntándoles porque creen ellos que esto sucede, que heurísticas se relacionan con el problema y que recomendaciones ellos proponen.

En la actividad de pruebas de usabilidad se puede incluir la tarea de buscar en el sitio cierta información que se sabe está incluida en la página X, posteriormente se puede preguntar a los usuarios si tuvieron dificultades accediendo a esta información.

Por último también es posible incluir en los cuestionarios específicos para los usuarios regulares del sitio preguntas que busquen conocer la opinión del usuario sobre la página X, si conocían su contenido y si les es de importancia.

2. Desarrollar hipótesis que se pueden probar mediante otras actividades:

Tomando el caso anterior, puede ser posible desarrollar la hipótesis de que la página X es poco accedida porque es difícil encontrarla en el sitio, esto puede ser probado mediante pruebas de usabilidad, donde se le asigna a los usuarios la tarea de encontrar cierta información o realizar cierta actividad que puede ser encontrada en la página X, midiendo

la tasa de errores y el tiempo de realización de esta tarea es posible comprobar o refutar esta hipótesis.

3. Probar y apoyar hipótesis que se desarrollen en otras actividades:

Una vez realizadas las otras actividades se contará con una gran cantidad de datos respecto al uso del sitio y a las opiniones y modelos mentales de los usuarios, estos datos pueden llevar a desarrollar hipótesis e inferencias que requieran revisar nuevamente la información obtenida mediante WUM. Como ejemplo, se puede dar en la actividad de ordenamiento de tarjetas que los usuarios formen en su mayoría muchos grupos con pocos tópicos cada uno, expresando en el cuestionario de salida comentarios que indiquen que les fue difícil formar grupos más significativos ya que los contenidos del sitio parecen ser muy variados y con poca relación entre ellos. Estos datos nos pueden llevar a la hipótesis de que la mayoría de los usuarios tiende a encontrar lo que busca en el sitio buscando sólo en la página de inicio o mediante herramientas de búsqueda, pero que no hacen recorridos en profundidad ya que sienten que pueden perderse fácilmente en la gran cantidad de diverso contenido del sitio. Esto se puede comprobar en cierta medida revisando si los resultados de WUM reflejan gran cantidad de sesiones de caminos cortos, que comienzan en la página principal, también se puede verificar si los términos usados en los buscadores web para encontrar el sitio son muy variados, al igual que los términos usados en la búsqueda interna del sitio (si tiene habilitada). Esto podría comprobar que los usuarios prefieren buscar mediante las herramientas en vez de navegar el sitio.

4. Obtener datos para mejorar el sitio web en aspectos técnicos

Además de la posibilidad de asistir en las tareas mencionadas, WUM brinda datos cuantitativos únicos que no puede obtenerse mediante las otras pruebas con solidez, tales como los horarios y días con más visitas, los navegadores más usados y los sistemas operativos ocupados. Estos datos pueden ser útiles a la hora de trabajar en la compatibilidad del sitio con las tecnologías usadas por los usuarios y para asegurar disponibilidad del sitio en los horarios más críticos. Si bien estos aspectos no se incluyen directamente en lo que se conoce por usabilidad, son importantes para la accesibilidad y los tiempos de carga del sitio junto con la carga de trabajo del servidor. Conociendo estos datos se puede concentrar

esfuerzos en asegurar que el sitio sea compatible con la mayor cantidad de usuarios, que esté disponible de preferencia en las horas de mayor uso y que cargue rápido y sin mayor demora.

3.9 Caso de estudio

Las actividades realizadas en esta evaluación han sido Web Usage Mining, ordenamiento de tarjetas, pruebas de usabilidad, evaluación heurística y cuestionarios. Como estas actividades han sido descritas anteriormente en el informe, no se explicará en que consisten en esta parte sino que se describirá como fueron realizadas y los resultados obtenidos en cada una de ellas, comenzando con WUM, la cual fue realizada con anterioridad a las demás actividades para estudiar la primera hipótesis propuesta.

3.9.1 Realización de la evaluación de usabilidad sobre el sitio web de la PUCV

Los resultados que serán descritos en el presente informe no son muy extensos ni detallados debido a que el enfoque es evaluar el aporte de WUM en el proceso y no dar recomendaciones de usabilidad a la PUCV.

A continuación se describirán los pasos seguidos en la evaluación según lo establecido en el plan.

3.9.2 Audiencia y plan de reclutamiento

El plan de reclutamiento se ha realizado según lo propuesto, reclutando participantes ofreciéndoles un premio o incentivo por participar. Las personas reclutadas son en su mayoría alumnos de la PUCV, con edades entre los 18 y 26 años y de ambos sexos. El total de voluntarios reclutados ha sido de 28 personas, varias de las cuales participaron tanto en las pruebas de usabilidad como en el ordenamiento de tarjetas, obteniendo el siguiente número de participantes por cada actividad de evaluación:

Pruebas de usabilidad: 9 voluntarios

Ordenamiento de tarjetas: 10 voluntarios

Cuestionarios para usuarios frecuentes: 15 voluntarios

Las cantidades están dentro de los óptimos recomendados para las pruebas de usabilidad [34] y un poco bajo el óptimo para el ordenamiento de tarjetas [35].

3.9.3 Desarrollo de los documentos necesarios

Los documentos realizados para las actividades son los mencionados en el plan de la evaluación, esto son:

Consentimiento

Cuestionarios de entrada

Listado de tareas para las pruebas de usabilidad

Cuestionarios de salida para ambas actividades con voluntarios

Listado de heurísticas

Documento de objetivos para evaluadores heurísticos

Tarjetas para Card Sorting

Cuestionarios para usuarios frecuentes del sitio

Algunos de estos documentos son bastante genéricos así que no vale la pena describirlos aquí directamente, estos pueden ser encontrados en el anexo “documentos” del presente informe.

3.9.4 Prueba y práctica de las actividades

Las pruebas pilotos se realizaron exitosamente con tres sujetos de prueba una semana antes de la etapa de evaluación con voluntarios. Como resultado de estas pruebas piloto se hicieron modificaciones en los formularios de entrada y salida, así como también en la lista de tareas de las pruebas de usabilidad y en la lista de temas de ordenamiento de tarjetas. Los problemas encontrados en las pruebas piloto fueron los siguientes:

La actividad de pruebas de usabilidad era demasiado extensa para el usuario, ya que contaba con 22 tareas, demorando los sujetos de prueba entre 45 y 60 minutos en realizarlas, siendo 30 a 45 minutos el óptimo buscado.

El formulario de entrada (que es el mismo para pruebas de usabilidad y ordenamiento de tarjetas) contenía muchas preguntas enfocadas a usuarios comunes del sitio, lo cual difería del perfil real de los voluntarios seleccionados, los cuales no conocían el sitio.

El formulario de salida contenía preguntas muy generales y no enfocadas a la actividad realizada. La corrección llevó a que exista un formulario de salida para cada actividad.

La lista de temas fue reducida de 90 a 75 temas, eliminando aquellos que podían confundir demasiado a los participantes o que eran poco entendibles

3.9.5 Realización de las actividades

A continuación se describirá la realización de las 5 actividades realizadas en el proceso de evaluación, comenzando por WUM para mostrar como este se acopló al proceso. Ambas hipótesis serán tomadas en cuenta en el desarrollo del análisis, tomando en cuenta los resultados de WUM al inicio y fin de las actividades.

a) Web Usage Mining (WUM)

Para la realización de la minería fueron utilizados logs de acceso del servidor del sitio de la PUCV, específicamente el servidor que aloja las páginas bajo el dominio www.pucv.cl, es decir, aquellas páginas como dae.ucv.cl, inf.ucv.cl, etc. No fueron analizadas porque sus accesos no se encuentran en este servidor. Es posible realizar un análisis integrado utilizando algunas técnicas de minería pero no era parte del objetivo y alcance de este proyecto. Los accesos analizados fueron los realizados entre el 2 de septiembre del 2007 y el 4 de octubre del 2007, los siguientes datos resumen las características de volumen de los datos

Total de hits: 14.568.556

Total de visitantes: 289.644

Total de ips únicas: 108.997

Total de transferencia: 123,386 GB

Estas estadísticas muestran que el servidor es bien concurrido y recibe gran cantidad de tráfico durante un mes. A continuación se expondrá un resumen de los resultados más relevantes obtenidos mediante las herramientas, posteriormente se podrá ver como estos datos influyeron en la realización de las demás actividades de evaluación. El análisis temporal de los accesos muestra que existen tendencias definidas en los accesos. La mayoría de los accesos se realiza en el comienzo de la semana y su cantidad decae a medida que la semana avanza.

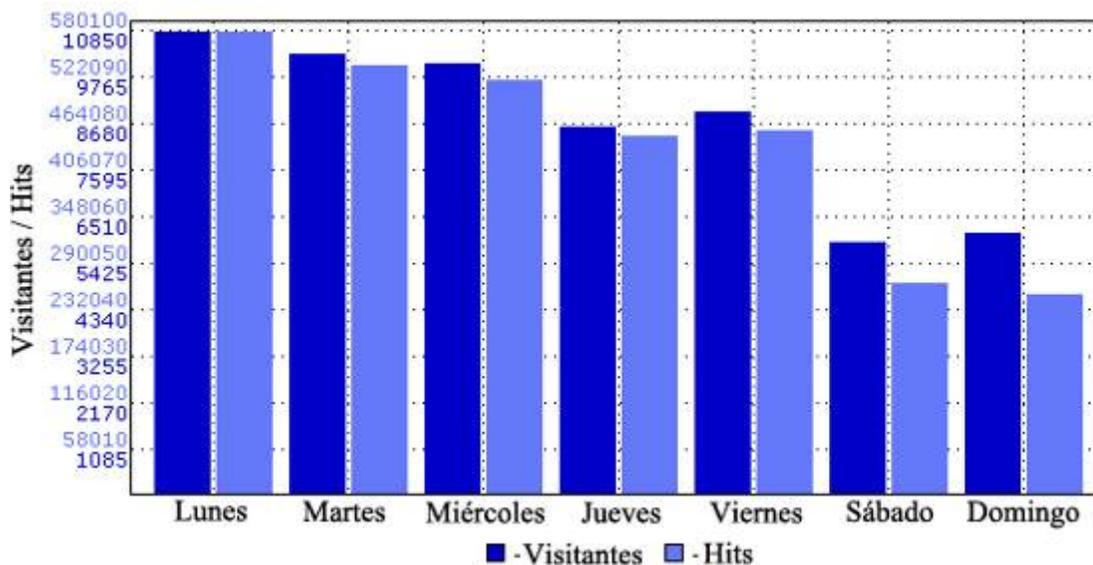


Ilustración 3.6: Gráfico de accesos según día de la semana

El análisis por hora del día muestra que los accesos se producen en su mayoría entre las 9 am y las 0 horas, siendo la hora peak las 16 horas.

Un listado ordenado de forma descendente de las áreas más visitadas es el siguiente:

1. La página principal del sitio: 80,19%
2. Recursos educativos sobre estadística descriptiva: 3%

3. Carreras: 1% siendo derecho la más visitada con 0.08%, psicología e ingeniería comercial con 0,07 y kinesiología e ingeniería civil industrial con 0,06
4. Admisión: 0.23%
5. Postgrado: 0.2%
6. Bienvenidos: 0.2%
7. Accesos a la biblioteca vía “acceso directo”: 0.2%
8. Escuela de ingeniería bioquímica: 0.18%
9. Escuela de agronomía: 0,15 %
10. Buscador: 0,12% (siendo “búsqueda avanzada por Internet” la opción más elegida)
11. Versión de bienvenida en inglés: 0,1%
12. Facultades: 0,07%
13. Vicerectoría de educación y estudios avanzados: 0,07%
14. Dirección de cooperación académica internacional: 0,05%
15. Directorio telefónico: 0,04%
16. Programa de movilidad estudiantil: 0,03%

Cabe destacar que además de la portada, la web de estadística es un punto de entrada importante con un 3% de las entradas, esto se debe búsquedas en Internet relacionadas con el tema como se expondrá más adelante en el análisis.

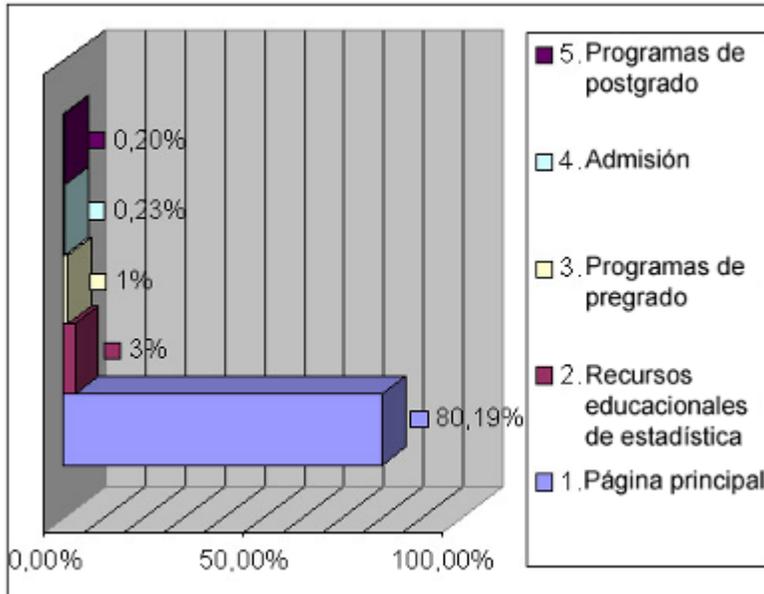


Ilustración 3. 7: 5 primeras áreas accedidas

Otros datos relevantes respecto al uso del sitio son los siguientes:

El tiempo promedio de vista de una página es de 2 minutos

24% de los visitantes sólo ve 1 página

Alrededor del 50% ve el sitio por 1 minuto o menos

la mayor cantidad de visitas es dentro de Chile (70%) seguido de EEUU (11%) y México (4%)

Google es el buscador más utilizado con un 98 %

Las búsquedas más comunes son

1. pucv

2. universidad católica de Valparaíso

3. percentil

4. ucv

5. estadística descriptiva

6. foto

7. exactitud

8. redondeo de datos

9. curtosis

10. ucv.cl

Los navegadores más utilizados son Explorer (87%) y Firefox (12%)

El 14% de los accesos corresponde a 158.251 (la red interna de la PUCV)

Un 60% del tráfico es archivos de video

Respecto al tráfico como se menciona en el último punto cabe destacar que el documento más descargado es el formulario de ingreso especial, seguido por el video de introducción a la universidad, le sigue un número de la revista correo universitario del 2004, la malla curricular de agronomía, documentos de las jornadas de historia de Chile, y el video introducción de facultad de ciencias jurídicas.

Respecto a la manutención de los enlaces del sitio su estado es media/alta con un 60% de 200 (OK) y un 5% de 404 (documento no encontrado). En cuanto a la limpieza y actualización del sitio se encontraron algunos errores que serán descritos más adelante.

Como conclusión respecto a la realización de esta actividad se puede decir que es relativamente sencilla dependiendo de la herramienta seleccionada, ya que se probó con varias herramientas y todas entregan resultados similares, pero difieren en gran medida en su usabilidad. Realizar el análisis de los datos obtenidos es más complicado ya que se debe realizar con cuidado para no hacer asunciones falsas o conjeturas equivocadas. Además es difícil interpretar los resultados cuando los sitios web no están diseñados para ser estudiados. Este tema se profundizará en las conclusiones finales.

b) Cuestionarios para usuarios frecuentes

Este cuestionario tiene como objetivo obtener algunos datos respecto a los usuarios del sitio, al las preferencias de éste y a su vez contrastar algunos datos obtenidos mediante WUM. A continuación se presenta el resumen de los datos obtenidos de un total de 15 cuestionarios. Fueron omitidas de los resultados aquellas opciones que no obtuvieron elecciones, el formulario original puede encontrarse en el apartado Documentos.

1. Género:

Femenino: 80%

Masculino: 20%

2. ¿A qué rango de edad perteneces (años)?

15-18: 27%

19-21: 40%

22-30: 33%

3. ¿Cuántas horas pasas a la semana usando el computador?

Entre 0 y 10: 27%

Entre 11 y 25: 27%

Más de 25: 46%

4. Selecciona tus softwares preferidos de las opciones (todos los que uses seguido)

Windows XP: 100%

Windows 98, NT o 2000: 20%

Linux: 20%

Firefox: 47%

Explorer: 47%

Opera: 13%

Netscape: 7%

Otros: 7 %

5. Escribe tu profesión, carrera o ocupación principal

En este punto destacan estudiantes de contador auditor e ingeniería

6. ¿Cuál es tu experiencia interactuando con sitios Web?

Media baja: 20%

Media: 33%

Media Alta: 40%

Experto: 7%

7. ¿Estudias en la PUCV?

Si: 100%

8. ¿Qué tan seguido usas el sitio Web de la universidad?

Algunas veces al mes: 27%

Una vez por semana: 13%

Varias veces a la semana: 60%

9. ¿Cuándo y en que horario prefieres usar el sitio web de la PUCV?

Entre Jueves y Domingo: 13%

Toda la semana: 80%

de 9 a 13 hrs.: 27%

de 13 a 18 hrs.: 27%

de 18 a 24 hrs: 13%

10. Cuando ingresas al sitio, ¿Qué áreas visitas o buscas?

Noticias: 13%

Información académica: 47%

Trámites: 40%

Información sobre la PUCV (carreras, facultades, etc): 13%

Biblioteca: 27%

Web mail: 53%

Área de admisión: 7%

Postgrados: 7%

Lo uso para acceder a otros sitios: 26% destacando el Navegador Académico

11. ¿Desde donde accedes comúnmente al sitio de la PUCV?

Desde la Universidad: 67%

Desde tu casa: 27%

Otro: 13% destacando los ciber cafés

12. ¿Qué es lo que más te gusta del sitio web de la PUCV?

En este punto destacan los colores y la gran cantidad de información que este posee

13. ¿Qué cosas sientes que faltan en el sitio o que están mal utilizadas?

En este punto destaca la falta de estructura del sitio

14. ¿Tienes algún comentario u opinión respecto al sitio que quieras expresar?

En este punto destaca que los encuestados parecen estar conformes con el sitio.

Es posible observar que si bien los datos no se corresponden perfectamente con los de WUM, no difieren demasiado. En la sección de análisis se profundizará en este tema.

c) Pruebas de usabilidad y sus respectivos cuestionarios

Una vez refinadas las pruebas y documentos se llevaron a cabo las pruebas de usabilidad. Los voluntarios acudieron al laboratorio de usabilidad de la PUCV en los horarios estipulados, aunque hubo faltas y voluntarios que no cumplieron con su asistencia. El tiempo que cada voluntario participó en el laboratorio fue en promedio una hora los que realizaron una sola actividad y 1 hora 45 minutos aquellos que realizaron ambas.

La sesión con cada usuario puede ser resumida como sigue:

1. Bienvenida y comprobación de asistencia
2. Explicación de la o las actividades a realizar
3. Firma de consentimiento y completación del formulario de entrada
4. Realización de las actividades y apoyo a consultas
5. Completación del formulario de salida correspondiente
6. Entrega del incentivo y despedida.

Las pruebas se llevaron a cabo con el hardware y software existente en el laboratorio de usabilidad, siendo el software utilizado: Windows XP, Brother Eye, CamStudio y Mozilla Firefox. La funcionalidad de asistencia remota de Windows Xp que permite ver lo que el usuario realiza en su PC desde otro PC fue utilizada sólo una vez, ya que provocó que Brother Eye terminara con errores, perdiendo el video grabado, el cual tuvo que ser recuperado de los archivos temporales del computador. No se capturó video sobre el usuario (expresiones faciales) mientras interactuaba con el sitio.

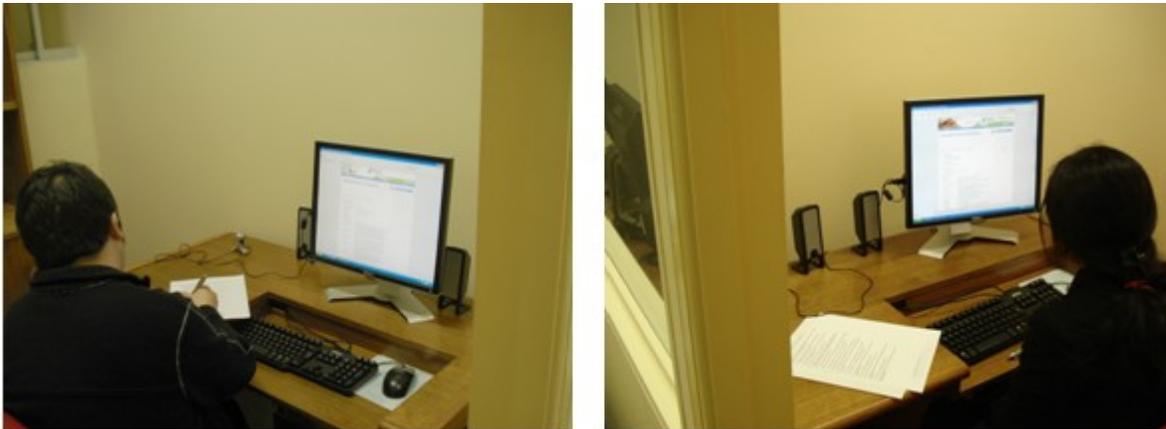


Ilustración 3.8 y 3.9: Participantes en las pruebas de usabilidad

Los voluntarios que realizaron esta actividad no reportaron haber tenido problemas mayores ni errores en la visualización del sistema operativo o el navegador. La lista de tareas les fue entregada explicándoles que la compleción de las tareas no es obligatoria, por lo que si alguna tarea les resultaba demasiado complicada podían continuar con la siguiente, marcando las tareas completadas con un visto y las no completadas con una x. Para efectos del estudio de las grabaciones de cada sesión de pruebas en el sitio, se les pidió a los participantes que presionaran el botón inicio o *home* del navegador cada vez que iniciaran una nueva tarea.

Los datos recolectados sobre los participantes en el formulario de entrada son los siguientes (las opciones sin elecciones fueron omitidas):

1. Género:

Femenino: 44%

Masculino: 56%

2. ¿A qué rango de edad perteneces?

15-18: 33%

19-21: 33%

22-30: 33%

3. ¿Qué actividades realizas generalmente en el computador?

Navegar por Internet: 78%

Trabajos (Excel, Word, etc): 33%

Usar programas avanzados: 11%

Recreación y ocio: 78%

4. ¿Cuántas horas pasas a la semana usando el computador?

Entre 0 y 10: 22%

Entre 11 y 25: 44%

Más de 25: 33%

5. Selecciona tus softwares preferidos de las opciones (todos los que uses seguido)

Windows XP: 100%

Linux: 11%

Firefox: 56%

Explorer: 33%

6. Escribe tu profesión, carrera o ocupación principal

En este punto destaca estudiante de ingeniería

7. ¿Cuál es tu experiencia interactuando con sitios Web?

Media: 67%

Media Alta: 33%

8. ¿Estudias en la PUCV?

Si: 67%

No: 33%

9. ¿Qué tan seguido usas el sitio Web de la universidad?

Nunca: 33%

Algunas veces al mes: 22%

Una vez por semana: 22%

Varias veces a la semana: 22%

10. Cuando ingresas al sitio, ¿Qué áreas visitas o buscas?

No lo uso: 33%

Información académica: 33%

Trámites: 11%

Biblioteca: 11%

Web mail: 33%

Otros: 11% destacando el navegador académico.

La lista de tareas que los usuarios tuvieron que trabajar fue la siguiente:

1. ¿Puedes encontrar el sitio de la escuela de ingeniería informática?
2. ¿Puedes hacer lo anterior usando el buscador del portal?
3. ¿Cual es el puntaje en que cerró Ingeniería Civil Bioquímica el 2007?

4. ¿Puedes encontrar el procedimiento de postulación al fondo solidario de crédito universitario?
5. Encuentra el código de pedido del libro “Usabilidad. Diseño de sitios Web” de Jakob Nielsen
6. ¿Puedes encontrar cuantas vacantes se abrieron para las carreras el 2007?
7. Encuentra el proceso de postulación para estudiar un semestre en el extranjero
8. ¿Cual es el horario de las actividades deportivas para el segundo semestre del 2007?
9. ¿Cuándo comienzan las vacaciones de verano?
10. ¿De que trata el taller de inserción laboral?
11. Encuentra el video de la Facultad de Filosofía y Educación
12. ¿Que promedio en la PSU requiere la Beca al Ingreso Destacado?
13. ¿Cuáles son las formas de pago que ofrece la PUCV para las mensualidades?
14. ¿Cuál es el valor del programa de Doctorado en Ciencias Físicas?
15. ¿Cuándo es el Simposio Internacional Biotecnología Marina 2007?
16. ¿Puedes encontrar información sobre la Acreditación?
17. ¿Puedes encontrar la ficha de postulación para ingreso especial a la PUCV?
18. ¿Puedes encontrar un formulario para retirar una asignatura que ya no quieras tomar?

Esta lista de tareas fue construida tomando en consideración los resultados de WUM acerca de cuales eran las áreas más accedidas del sitio y los objetivos de la PUCV anteriormente mencionados. Es posible observar que las tareas involucran varias áreas del sitio y están desarrolladas desde una perspectiva que simula las necesidades del usuario.

Los datos respecto a la realización de las tareas se presentan en la siguiente tabla, donde se encuentra las veces que cada tarea fue completada, el grado de dificultad que los participantes percibieron de la tarea (siendo 1 muy fácil y 5 imposible), el promedio del tiempo que les tomó a los participantes completar las tareas y del tiempo que les tomo antes de desistir y pasar a la tarea siguiente.

Tarea	Veces terminada	% de completación	Índice de dificultad (1-5)	Promedio de tiempo de completación	Promedio de tiempo de deserción
1	6	67%	3	44 seg.	2 min. 37seg.
2	4	44%	4	3 min. 30 seg.	1 min. 52 seg.
3	9	100%	1	55 seg.	---
4	1	11%	4	---	2 min. 40 seg.
5	6	67%	4	4 min. 40seg.	4 min. 20 seg.
6	6	67%	3	1 min. 26 seg.	4 min. 3 seg.
7	7	78%	3	1 min. 50 seg.	1 min. 55 seg.
8	5	56%	4	2 min.	2 min. 45 seg.
9	7	78%	3	59 seg.	2 min. 25 seg.
10	2	22%	4	2 min. 13 seg.	3 min. 26 seg.
11	6	67%	3	1 min. 29 seg.	2 min. 50 seg.
12	9	100%	1	34 seg.	----
13	5	56%	4	1 min. 11. seg.	1 min. 54 seg.

14	9	100%	2	1 min. 4 seg.	---
15	5	56%	4	2 min. 48 seg.	1 min. 33 seg.
16	2	22%	4	30 seg.	2 min. 23 seg.
17	7	78%	2	40 seg.	55 seg.
18	4	44%	4	2 min. 35 seg.	2 min. 13 seg.

Tabla 3.1: Tabla de completión de tareas

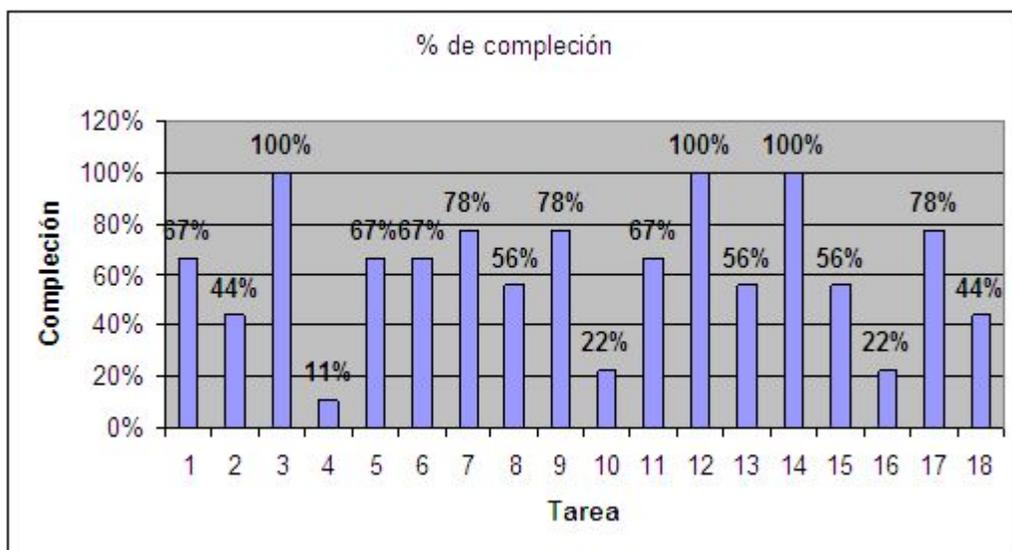


Ilustración 3.10: Gráfico de porcentaje de completión de las tareas

Más adelante se analizarán estos datos con mayor detalle y en relación con el resto de la información adquirida.

Los formularios de salida de esta actividad muestran la opinión del participante respecto al sitio de forma más completa, ya que se sabe el uso que este le ha dado. Cabe destacar que las opiniones de los usuarios frecuentes encuestados difieren con la de los participantes en las pruebas de usabilidad, puede ser porque los últimos tuvieron que realizar tareas más

precisas y durante más tiempo, por lo que se encontraron con más problemas que el resto. Un resumen de los formularios de salida es el siguiente:

1. ¿Cual es tu impresión general del sitio?

En esta pregunta los participantes se mostraron en general disconformes con el sitio, siendo el aspecto crítico la gran cantidad de información de la que dispone. El consenso es que es positivo que el sitio posea gran cantidad de información, pero que la forma en que está dispuesta lo hace confuso y complicado.

2. ¿Cual es tu impresión de la herramienta de acceso directo?

La herramienta obtuvo buenos comentarios, haciendo ver que es útil pero que está mal aprovechada, los participantes piensan que podría ser mejor.

3. ¿Cual es tu impresión de la herramienta de búsqueda?

Esta herramienta presentó serios problemas en las tareas lo que se ve reflejado en la opinión de los participantes. Estos muestran desagrado por la herramienta destacando su poca utilidad y confuso uso.

4. ¿Qué es lo que te gustó más del sitio?

En esta pregunta los participantes destacan que hay áreas como la de facultades y carreras y el área de becas que son fáciles de usar y fueron de su agrado. Se destaca positivamente la utilidad de la información en algunas áreas como las mencionadas.

5. ¿Qué es lo que menos te gustó del sitio?

En esta pregunta destaca el desagrado por la herramienta de búsqueda y por lo confuso que puede llegar a ser navegar el sitio debido al desorden que perciben los participantes en la estructura.

6. ¿Sientes que le falta algo al sitio? ¿Qué cosa?

Los participantes aquí destacan la necesidad de mejorar la herramienta de búsqueda y de simplificar y estructurar el sitio, para que su navegación sea más fácil.

7. Si le tuvieras que describir el sitio a alguien en una frase o dos, ¿Qué le dirías?

Las opiniones en esta área se podrían resumir en “complicado y desordenado”.

8. ¿Hubo alguna tarea que se hayas tenido que realizar de una forma distinta a lo que te imaginabas? ¿Cuál?

En esta pregunta la respuesta de la mayoría fue “no”. Aunque algunos participantes destacaron que en ciertas tareas no se les ocurría nada excepto usar el buscador.

9. ¿Cuál tarea fue la más complicada? ¿Por qué?

Aquí las opiniones fueron variadas, ya que cada usuario tuvo sus propios inconvenientes, pero se destacan la búsqueda en biblioteca y la búsqueda de formularios.

10. ¿Tienes algún comentario o acotación que desees expresar?

Si bien la mayoría de los participantes no deja comentarios, los que lo hicieron destacan que el sitio debe mejorar

En cada una de los siguientes ítems los participantes marcaron el recuadro que más se ajustó a tu opinión. En las tablas que siguen el recuadro marcado es aquel que corresponde al promedio medido del ítem.

1. Apreciación general del sitio

	1	2	3	4	5	6	7	
Terrible								Excelente
Confuso								Simple
Frustrante								Satisfactorio
Aburrido								Entretenido

2. Elementos Visuales

		1	2	3	4	5	6	7	
Leer la información es	Muy difícil								Muy fácil
La agrupación de elementos es	Muy confusa								Muy clara
El estilo visual es	Desagradable								Agradable
El tamaño de los elementos es	No adecuado								Adecuado

3. Aprendizaje

		1	2	3	4	5	6	7	
Recorrer el sitio es	Difícil								Fácil
Realizar tareas (actividades) es	Frustrante								Simple
El lenguaje utilizado es	Incoherente								Muy claro
Acostumbrarse al uso es	Muy lento								Muy rápido

4. Contenido

		1	2	3	4	5	6	7	
La información contenida es	Inútil								Muy útil
La información es actual	Muy poco								Bastante
La información es	Confusa								Entendible
El sitio contiene la información y funcionalidades que esperaba	Ninguna								Todas

5. Desempeño

		1	2	3	4	5	6	7	
El tiempo de carga es	Muy lento								Muy rápido
La ocurrencia de errores es	Muy alta								Muy baja
Siempre puedo saber donde me encuentro dentro del sitio	Difícilmente								Siempre
Acceder a las áreas del sitio es	Muy difícil								Muy simple

Es posible observar que los promedios no son muy bajos, si no que más bien medios/altos mostrando que si bien existen problemas, la apreciación general del sitio no es pésima.

d) Ordenamiento de tarjetas y sus respectivos formularios

El ordenamiento de tarjetas o Card Sorting fue realizado en el laboratorio de usabilidad y en otras dependencias de la universidad, contando con las 75 tarjetas con los temas, tarjetas de colores para los títulos de los grupos formados por los participantes y el material de apoyo correspondiente (lápices, elásticos, sobres, etc). Los participantes no reportaron

haber tenido dificultades con el espacio o los materiales disponibles y las actividades fueron realizadas cómodamente.

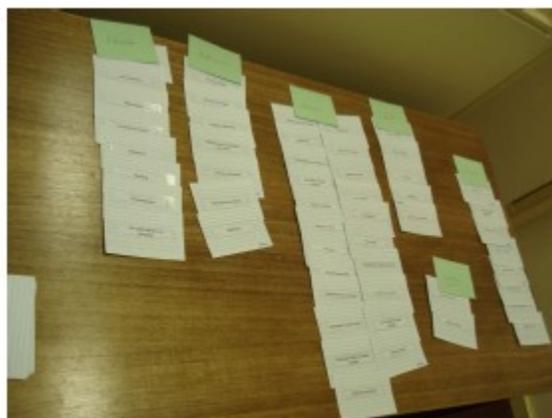


Ilustración 3.11 y 3.12: Ordenamiento de tarjetas en el laboratorio de la PUCV

El resumen de los cuestionarios de entrada no será expuesto ya que no difiere de los que se han expuesto anteriormente. El análisis de los datos recogidos mediante el ordenamiento de tarjetas se llevó a cabo en una planilla Excel de la forma descrita en [37] y a continuación se presenta la lista de temas y el resumen de los aspectos más relevantes.

Aranceles	Matrículas	Aula virtual
Formularios de inscripción y retiro	Web mail	Fondo solidario de C.U.
Ediciones universitarias	Contraloría	UCV TV
UCV radio	Fe- Pucv	Programa Vincular PUCV
Administración y finanzas	Deporte universitario	Dirección de cooperación internacional

Colegio Rubén castro	Navegador académico	Análisis y Ensayos de Laboratorio
Becas	Asuntos estudiantiles	Procesos docentes
Postgrados y postítulos	Investigación	Sesiones del tribunal de mérito
Facultades	Carreras	Unidades académicas
Cursos y seminarios	Bienvenidos	Palabras del rector
Sedes	Contacto	Buscador
Admisión	Agenda cultural	Noticias
Biblioteca	Ex alumnos	Dirección de informática
Directorio	Estructura organizacional	Pastoral PUCV
Semestres de estudios en el exterior	Calendario académico 2007	Sistemas de ingreso a las PUCV
Ingreso especial	Seguro estudiantil PUCV	Programas de becas de arancel
Consultoría Profesional e Innovación Tecnológica	Programación estudios generales 2007	Intercambio estudiantil
Beca Isabel Brown caces	Crédito de la ley 20.027	Beca excelencia académica
Bolsa de trabajo	Fondos concursables	Autogestión estudiantil
Talleres artísticos culturales	Prevención de drogas	Comunidad Rapa Nui
Desarrollo personal	Trabajo comunitario	Servicio dental

Seguro de accidente escolar	Seguro estudiantil PUCV	Asistentes sociales
Oficina técnica VRADE	Desarrollo curricular y formativo	Acreditación
Magíster académico	Dirección de investigación e innovación	Dirección de estudios avanzados
FONDEF	Voluntariados	Intercambio de idiomas

Tabla 3.2: Lista de temas usados para el ordenamiento de tarjetas

Un resumen de los resultados del ordenamiento sin refinar es el siguiente:

La cantidad promedio de grupos formados fue de 13, con un mínimo de 9 y un máximo de 17

La cantidad promedio de tarjetas por grupo fue de 7, con un mínimo de 3 y un máximo de 18

Una vez realizado el ordenamiento de tarjetas abierto (open card sorting) se debe formar grupos estandarizados con los grupos formados por los usuarios, así por ejemplo el grupo “Proceso de admisión” y el grupo “Área de admisión” pueden juntarse en un nuevo grupo llamado “Admisión”. De esta forma se pasó de un total de casi 60 grupos a un total de 36 grupos distintos, los cuales son:

1. Investigación
2. Programas de estudio
3. Medios de comunicación
4. Links y sitios relacionados
5. Asuntos estudiantiles

6. Admisión e ingreso
7. Actividades extraprogramáticas
8. Docencia
9. Ayudas estudiantiles
10. Becas
11. Intercambio internacional
12. Pastoral
13. Recursos
14. Gestión estudiantil
15. Estudios avanzados
16. Calendario, agenda y fechas importantes
17. Estructura organizacional
18. Directorio
19. Federación de estudiantes
20. Administración y finanzas
21. Fondos
22. Inscripción y retiro
23. Sedes, facultades, carreras y unidades académicas
24. Servicios
25. Matricula, créditos y aranceles

26. Formación curricular
27. Beneficios estudiantiles
28. Novedades
29. Desarrollo personal
30. Bienvenida
31. Informaciones varias
32. Oficinas dedicadas
33. Talleres y ayuda social
34. Inicio PUCV
35. Cultura y recreación
36. Sin clasificación

El paso de estandarizar los temas debe realizarse con cuidado para no parcializar los datos obtenidos de la participación de los voluntarios. Es posible observar que a pesar del proceso de filtrado y estandarización la cantidad final de temas sigue siendo lo que se corresponde con los datos que se presentan a continuación:

El promedio final de tarjetas por grupo es de 12, con un mínimo de 3 y un máximo de 34

El promedio de la cantidad de grupos a la que cada tarjeta fue asignada por los participantes es de 6 grupos, con un mínimo de 2 y un máximo de 9

El promedio del porcentaje de acuerdo de cada tarjeta es de 19%, con un mínimo de 11% y un máximo de 50%

El promedio del porcentaje de acuerdo de cada grupo es de 17%, con un mínimo de 10% y un máximo de 32%

Las tarjetas con una con la ocurrencia más elevada de acuerdo son:

- “Sistemas de ingreso a las PUCV” en el grupo “Admisión e ingreso” con un 90%
- “Ingreso especial” en el grupo “admisión e ingreso” con un 80%
- “Asuntos estudiantiles” en el grupo “Asuntos estudiantiles” con un 80%

Un 9% de las tarjetas está en 2 o 3 grupos, un 62% entre 4 y 6 grupos y un 28% entre 7 y 9 grupos

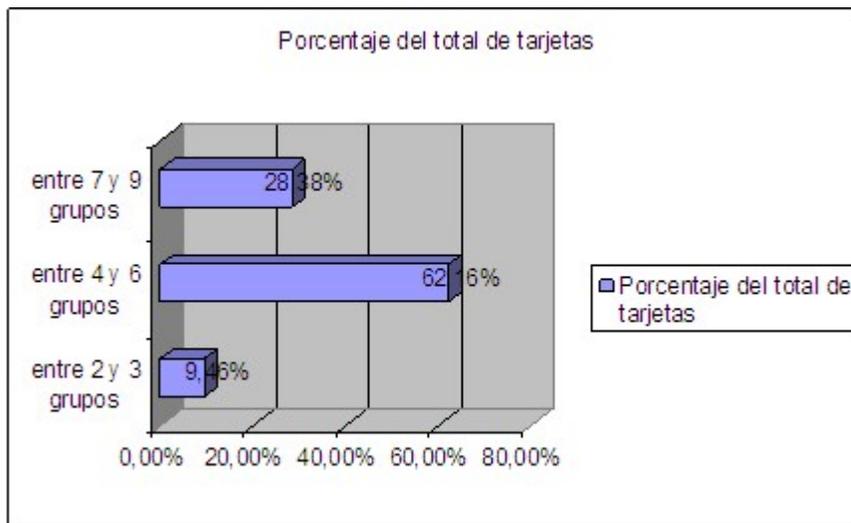


Ilustración 3.13: dispersión de las tarjetas

El nivel o porcentaje de acuerdo de una tarjeta muestra la exclusividad de una tarjeta respecto a un grupo, a modo de ejemplo, si una tarjeta aparece 10 veces en un solo grupo A tendrá un 100% de acuerdo en ese grupo, mientras que si aparece 9 veces en el grupo A y 1 vez en otro grupo B, tendrá un 90% de acuerdo en A y un 10% en B.

Si bien los promedios anteriormente entregados son de utilidad, se debe tener cuidado a la hora de analizarlos. Por ejemplo, una tarjeta que aparezca 9 veces en A y 1 en B obtiene un promedio de 50% de acuerdo ($(9+1)/2$), al igual que una tarjeta que obtenga 5 selecciones en A y 5 en B. Para remediar el problema anterior se sugiere prestar atención a los valores máximos de cada tarjeta y cada grupo.

El cuestionario de salida de esta actividad recogió los siguientes resultados:

1. ¿Cual es tu impresión general de la actividad, fue muy difícil?

La tarea no fue difícil para la mayoría, aunque algunos participantes destacaron que la gran variedad de temas hacía la actividad menos simple.

2. ¿Hubieron temas que fueron difíciles de agrupar?, ¿Cuáles?

En este punto varios temas fueron mencionados, destacando “Oficina técnica VRADE”, “Procesos Docentes” y “Sesiones del tribunal de mérito”

3. ¿Sientes que hay temas que faltan o sobran en el sitio?

En este punto los participantes sienten que sobra “Comunidad Rapa Nui” y que faltan las mallas de las carreras.

4. ¿Tienes algún comentario o acotación que desees expresar?

Se destaca en esta pregunta que algunos participantes opinan que si los temas en las tarjetas son links dentro del sitio, entonces varios links serían poco significativos o confusos para ellos.

e) Evaluación heurística

La evaluación heurística contó con 4 evaluadores que encontraron 33 problemas, los cuales fueron ordenados por orden de criticidad (frecuencia + severidad) de la siguiente forma:

1	No existe una sección o documento de ayuda de uso del sitio	Criticidad
2	El sitio al agregar tantas informaciones en su zona central, hace irrelevante	7,00

	las noticias. Además todas están resaltadas de la misma forma por lo que el sitio sugiere cual es más importante.(La información se hace carente de utilidad). Si se hace clic en “ver más” la lista que despliega es gigante y hace uso excesivo del scroll	
3	No existe ningún tipo de estructura en los distintos menús, todas las opciones se presentan al mismo nivel sin una organización clara	7,00
4	Existe inconsistencia en el uso del símbolo "flecha a la derecha" se usa acompañado de las opciones de Volver, Arriba, Cerrar	6,50
5	La opción de "Búsqueda por Personas" no está disponible	6,25
6	Al utilizar la opción "Búsqueda en el Sitio", cuando la búsqueda no produce resultados, el usuario no es informado de ello, simplemente la página queda en blanco.	6,25
7	No orienta al usuario a cerca del lugar del sitio en donde se encuentra, ni siquiera la URL indica en que lugar del sitio estamos ubicados.	6,25
8	Carece de mapa de navegación, de apoyo para algún usuario que no conozca el tipo de estructura organizacional	6,25
9	Mucha información en la página inicial.	6,00
10	En el buscador, al ingresar al buscador en Internet, existe un botón "Volver" que no vuelve a ninguna parte. Al igual que en Búsqueda Avanzada en Internet. Y en todos los "buscadores" dentro de búsqueda avanzada por Internet.	6,00

11	No se diferencian las páginas ya visitadas.	6,00
12	En la sección "Concursos Académicos" existe botón que dice "Cerrar" que no cierra nada.	5,75
13	El sitio principal está optimizado sólo para cierta resolución. A altas resoluciones hay un sector importante de la página que se pierde.	5,75
14	No utiliza metáforas conocidas para la realización de acciones comunes por ejemplo Home, mail.	5,75
15	En la página principal no existe una opción o link visible para el buscador, el usuario debe navegar y encontrar la herramienta.	5,50
16	En Mozilla, al presionar sobre el banner superior derecho, se abre la misma página inicial, pero sin los menús superior y lateral.	5,50
17	En la sección "Bienvenidos" el enlace a Estructura Organizacional" no lleva a nada... la pagina se queda igual.	5,50
18	Existe inconsistencia en el lenguaje utilizado en las páginas, la opción de "Búsqueda en el Sitio" se presenta en Inglés	5,50
19	No permite regresar a un estado o página anterior, pues no posee mecanismos de volver en todas las páginas	5,25
20	Se ocupa el color morado para los link, siendo que este color se ocupa generalmente para hacer notar que una página ha sido visitada.	5,25

21	Posee muchas animaciones que pueden distraer la atención del usuario.	5,25
22	Existe un banner al lado superior derecho que va cambiando cada vez que se actualiza la página, esto puede provocar que el usuario se confunda.	5,25
23	En la mayoría de los sitios que pertenecen a la Universidad (Deporte Universitario y otros), la interfaz del sitio, difiere de la interfaz de la pagina principal, si se supone que pertenece a la misma institución, debería tener el mismo formato.	5,25
24	Es común en los sitios web que un banner sea un link. En la página central sector sur oeste hay banners que llevan a sitios como Navegador Académico y Universia, pero en la parte de arriba hay banner promocionando al postgrado y a su lado un banner de cómo pagar la universidad. Estos últimos a pesar de ser informativos no llevan ningún lado, una persona que quiera informarse del posgrado o cómo pagar hará instintivamente clic y se desilusionará.	5,25
25	En el menú lateral izquierdo se muestra una "flecha" donde debería existir una opción, siendo que no existe y está en blanco.	5,25
26	El banner superior derecho a veces muestra solo información, y otras es un link hacia otra página.	5,00
27	En la sección "Concursos Académicos" existe un mensaje poco claro que dice "Bajar perfil postulante" y es siempre el mismo mensaje	5,00
28	Hay links que abren una nueva ventana y otros que navegan sobre la misma, sin ninguna distinción o criterio aparente	5,00

29	Existe inconsistencia en el lenguaje utilizado en el menú de la izquierda, hay vínculos en inglés y español. Por ejemplo "International Programs"	4,75
30	Cuando se accede a través del Link a "Búsqueda Avanzada" se espera encontrar el servicio que permita realizar una búsqueda de ese tipo, sin embargo lo que se encuentra es información teórica a cerca de los buscadores y listas de buscadores y/o directorios	4,75
31	No utiliza términos o frases estándar, por ejemplo usa la palabra "Portada" para hacer referencia la página inicial del sitio;	4,75
32	Usa términos o frases poco familiares al usuario. Por ejemplo "Formas Gráficas" en el menú Servicios de la derecha	4,75
33	El botón ir de la sección de accesos directos, es innecesario en algunas opciones que son ejecutadas al momento de ser seleccionadas.	4,75

Tabla 3.3: tabla de resultados de la evaluación heurística

Es posible observar que esta alta cantidad de errores hace ver que el sitio presenta obvios problemas, algunos de los cuales se analizarán enseguida.

3.9.6 Análisis de los datos y problemas de usabilidad encontrados

Los siguientes 3 problemas fueron encontrados usando sólo Web Usage Mining

1. Problema de navegabilidad y acceso vía buscadores

El siguiente problema fue encontrado tomando en cuenta que el 3% de los accesos se realizaban a un área del sitio que contiene recursos educativos para profesores respecto al tema de estadística descriptiva. Esta área reportó más accesos incluso que carreras o admisión, lo que puede parecer extraño. Esta interrogante se complementó con el resultado del análisis de las búsquedas en los motores como google, la que muestra que dentro de los

10 términos más buscados se puede encontrar “percentil”, estadística descriptiva”, “exactitud”, “redondeo de datos” y “curtosis” (5 de los 10 primeros términos), términos se encuentran presentes en el área anteriormente mencionada.

La situación es clara, el área mencionada es bien recurrida por los buscadores y una búsqueda de “percentil” entrega como primera ocurrencia el área mencionada. Pero esto no debería ser un problema, sino más bien una ventaja, una oportunidad de dar a conocer el sitio a más personas y tener un área que sea de frecuente uso y utilidad. El problema radica en que las páginas de esta área accedidas a través de los buscadores son páginas aisladas, las cuales no poseen hipervínculos hacia otro lado, no poseen menú ni identificación de ningún tipo lo cual deja al visitante en un callejón sin salida. Este problema no es exclusivo de esta área, otras búsquedas como “Neruda” o “chilean currency” nos llevas a otros callejones sin salida.

Tras investigar la naturaleza del problema se ha descubierto que este radica en la construcción de algunas secciones del sitio en base a marcos (frames). Cuando el motor de búsqueda encuentra la información devuelve sólo el marco principal que contiene la información, dejando a la página sin cabecera y menú, provocando los problemas mencionados.

2. Problema de actualización del sitio y acceso vía buscadores

Este problema fue también descubierto mediante los datos entregados por WUM, los cuales mostraron que una búsqueda comúnmente realizada era “usos y aplicaciones de la dilatación de la materia” y otras similares que tenían que ver con física. Estas búsquedas llevaban a un área del sitio con el mismo problema de marcos mencionado anteriormente y con otro problema nuevo, cuando se seguía el hipervínculo llamado “Portada” se llegaba a una versión del portal de la PUCV que data del 2003. Esta versión del portal está totalmente funcional y puede pasar bastante tiempo en la navegación del visitante antes de que se de cuenta que está leyendo información de hace 5 años, el visitante puede incluso no notar el problema y considerar la información como actual y correcta. El análisis de minería muestra que en el intervalo analizado (2 septiembre al 4 de octubre del 2007) este portal antiguo recibió 245 visitas. Este problema si bien fue encontrado por primera vez con

WUM, también se encontró en las pruebas de usabilidad, donde un participante utilizando la herramienta de búsqueda interna del sitio encontró este portal antiguo y trató de resolver la tarea usando la información que contenía.



Ilustración 3.14: fragmento del sitio del 2003

3. Problema de carga del servidor y transferencia de archivos

El análisis de WUM muestra que un 60% del tráfico en el servidor corresponde a archivos de video en formato .mpg (35%) y .avi (25%), además de un 4% en documentos .pdf y .doc. El tráfico de estos archivos constituye un total considerable y se debe principalmente al peso de éstos y a la manera en que se presentan. La mayoría de los archivos requiere de descarga para ser visualizados y tienen nombres poco significativos que hacen que muchas veces los usuarios deban descargar el archivo completo para saber si era lo que buscaban. En el caso específico de los videos estos pueden ser vistos en forma streaming solo si se tiene el plug-in necesario y de forma poco práctica, ocupando toda la ventana y sin mayor información sobre el video.

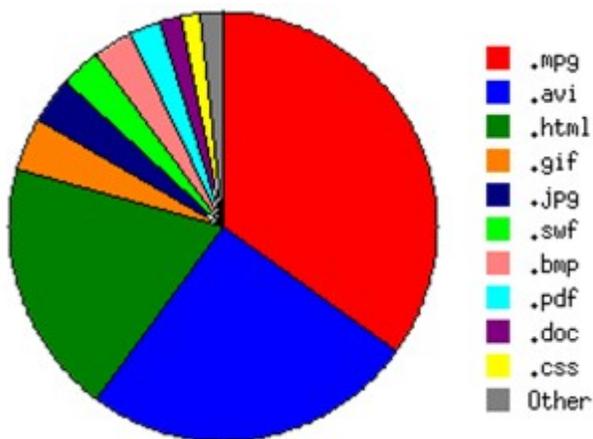


Ilustración 3.15: Tráfico del servidor por tipos de archivo

Los problemas anteriormente descritos fueron encontrados solamente analizando los datos entregados por WUM. Los siguientes problemas que se expondrán a continuación toman en cuenta los resultados de una o más actividades, algunas de las cuales involucran también la minería web realizada.

4. Problema de accesibilidad para usuarios del navegador Firefox

Analizando los cuestionarios de entrada para las actividades y cuestionarios para usuarios frecuentes del sitio, es posible encontrar que 50% de los encuestados usa o prefiere Firefox para navegar por Internet y muchos de ellos lo utilizan de forma exclusiva. Resultó curioso entonces que los datos de WUM sobre los navegadores mostraran que sólo un 12% de los visitantes al sitio utilizaran este navegador. Al profundizar en esta interrogante se encontró que el sitio tiene problemas con el navegador Firefox, sobre todo para las páginas dinámicas como impresión de cupones de pago, matrículas, certificados y otros. En la imagen se puede apreciar la diferencia que existe al navegar el sistema de impresión de cupones de pago con el navegador Explorer o con Firefox.

Estimado alumno, tu cuenta corriente registra cuotas impagas.
 Para regularizar tu deuda comunícate con nosotros al Call Center 227 30 30, una ejecutiva de crédito y cobranza te entregará soluciones concretas de pago.

ARANCEL		CUOTAS	
Valor de Arancel	\$ 796.500	Cantidad	5
BENEFICIOS		Valor	\$ 92.800
PAGO POR CREDITO PARA CUPONERA	\$332.500	ABONOS	
Total Beneficios	\$332.500	Total Abonos	\$ 0
Saldo (Arancel-Beneficios)	\$464.000		

ESTADO DE CUOTAS						
N°	Valor de Cuota	Abono	Total a Pagar	Vencimiento	Estado	Pagar
1	\$92.800	\$0	\$92.800	31/08/2007	Vencida	<input type="checkbox"/>
2	\$92.800	\$0	\$92.800	28/09/2007	Vencida	<input type="checkbox"/>
3	\$92.800	\$0	\$92.800	31/10/2007	Vencida	<input type="checkbox"/>
4	\$92.800	\$0	\$92.800	30/11/2007	Por Pagar	<input type="checkbox"/>

Ilustración 3.16: Página en Explorer

Estimado alumno, tu cuenta corriente registra cuotas impagas.
 Para regularizar tu deuda comunícate con nosotros al Call Center 227 30 30, una ejecutiva de crédito y cobranza te entregará soluciones concretas de pago.

ARANCEL		CUOTAS	
Valor de Arancel	\$ 796.500	Cantidad	5
BENEFICIOS		Valor	\$ 92.800
PAGO POR CREDITO PARA CUPONERA	\$332.500	ABONOS	
Total Beneficios	\$332.500	Total Abonos	\$ 0
Saldo	\$464.000		

Ilustración 3.17: Página en Firefox

Este problema no es menor, ya que es posible que hasta un 50% de los visitantes se sientan frustrados al encontrar problemas de accesibilidad cuando navegan el sitio.

5. Problemas con la herramienta de búsqueda del sitio

Estudios muestran que el 50% de los usuarios prefiere una navegación orientada a la búsqueda (search oriented) frente a una orientada a links jerárquicos (hierarchy links oriented). Esto nos dice que el 50% de nuestros hits en un sitio pueden ir directo a la herramienta de búsqueda, y es posible observar que los motores y tecnologías de búsqueda en la web son un tema recurrente y que avanza rápidamente. En el ámbito de WUM se ofrecen algunas herramientas nuevas que permiten analizar específicamente la búsqueda interna del sitio, ofreciendo estudios que muestran la relevancia de descubrir que es lo que los usuarios buscan dentro de un sitio web.

Frente a esta importancia que han adquirido las herramientas de búsqueda interna en el sitio, se presenta el problema en cuestión. El portal de la PUCV cuenta con un buscador el cual la mayoría de los participantes reconoció desconocer, ya que no se encuentra resaltado ni a la vista. Este buscador cuenta con varias opciones las cuales no parecen claras para el navegante, según mostraron las pruebas de usabilidad (específicamente la tarea n° 2). Cuando los participantes ingresaron al área de búsqueda se veían confundidos por las opciones, siendo la opción más visitada “Búsqueda avanzada”, dato que puede ser comprobado mediante WUM (se encuentra dentro de las 15 áreas más accedidas).

En el área de búsqueda, de las cinco opciones de búsqueda que se ofrecen, sólo una permite buscar en el sitio y ésta herramienta se encuentra completamente en inglés, haciendo que la búsqueda sea complicada ya que las opciones de búsqueda no son claras para los usuarios, problemas que quedaron demostrados en las pruebas de usabilidad, donde un comportamiento común de los participantes fue expandir cada una de las opciones ofrecidas, sólo para seleccionar ninguna. Otro punto crítico es la capacidad de búsqueda en sí, ya que hay búsquedas muy simples que no obtienen resultados como por ejemplo “matrícula” y además una búsqueda sin resultados entrega una página totalmente en blanco, sin ningún texto o imagen de carácter explicativo, lo que llevó a algunos participantes a pensar que la herramienta estaba descompuesta o no funcional. Estos y otros problemas

relativos a la búsqueda interna del sitio pueden encontrarse en la evaluación heurística, en los problemas nº 5, 6, 10, 15, 18 y 30. Finalmente cabe destacar que los participantes de las pruebas de usabilidad recurrieron a la búsqueda interna en varias ocasiones para poder resolver las tareas, y la mayoría de las veces la herramienta no fue de ayuda. Es posible observar en la tabla de compleción de tareas que la tarea nº2 tiene un bajo porcentaje de cumplimiento y un alto índice de dificultad, junto con un alto tiempo de compleción.

6. Problemas con la búsqueda en biblioteca

Según lo encontrado con Web Usage Mining, la búsqueda en biblioteca se encuentra dentro de las 10 primeras áreas más accedidas (nº 6). Esta área a pesar de su alto nivel de acceso presentó varios problemas en las pruebas de usabilidad, siendo la tarea nº 5 (buscar un libro en biblioteca) la tarea con el promedio de tiempo de compleción más largo. En primer lugar, el sistema de búsqueda no parece ser la opción que más llama la atención de los usuarios, ya que si bien está presente en la página principal del sistema de biblioteca, muchos participantes recorrieron exhaustivamente los menús buscando la opción que les pareciera correcta para buscar el libro. Una vez que los participantes se decidían a utilizar la búsqueda encontraban varias complicaciones. El sistema presenta al usuario 12 opciones de búsqueda ninguna de las cuales es autoexplicativa, llevando a los usuarios a elegir la que les parece más lógica la que fue en general “sistema de biblioteca”, la opción correcta es “Opac (Catalogo local)”. Si el usuario no selecciona una opción aparece un mensaje de error al respecto, pero en vez de dejarlo recuperar el control y realizar una búsqueda de la manera indicada el usuario es llevado a una página de error 404. Una vez que los participantes ya sabían como realizar la búsqueda se encontraban con un problema más en su camino, presionar “enter” el campo de búsqueda en vez de accionar la búsqueda hace que la página se refresque, perdiendo el texto de búsqueda escrito y la opción seleccionada.

Este sistema en general hace que el usuario se pierda en su utilización, llevándolo a distintas partes inesperadas, no respetando las metáforas y atajos a los que los usuarios ya están acostumbrados. El otro sistema de búsqueda de biblioteca, al cual se accede en el menú de búsqueda interna del sitio parece ser más directo y obvio para los usuarios, pero cuenta con varios errores como el de presionar “enter” el cual lleva a otra página sin sentido para los usuarios.

7. Problemas de íconos, metáforas y links

Resumiendo un poco muchos de los problemas reportados en la evaluación heurística, es posible observar que el sitio posee problemas respecto a la representatividad de los íconos, el uso de metáforas y la claridad y resalte de los links. En las pruebas de usabilidad esto se hizo presente en varias tareas y fue uno de los factores importantes en la compleción de estas.

En la tarea nº 11 se pide al participante que busque un video, el cual se encuentra ubicado en la página de la respectiva facultad que se muestra en el video. La mayoría de los participantes encontró 2 problemas relacionados en la ejecución de esta tarea. El primero lo encontraban cuando accedían a “Facultades y carreras” donde se puede apreciar una lista de las facultades y sus respectivas carreras, en esta página cada carrera tiene un link hacia su respectiva página, pero las facultades no poseen tal link, para acceder a la lista de facultades se debe elegir una opción en el menú que nos lleva a una página distinta. Prácticamente el 100% de los participantes intentó acceder a la facultad solicitada mediante el título presente en “Facultades y carreras” y todos probaron como segunda opción la imagen que se presenta al lado del título, la cual posee un símbolo “>” que hizo pensar a muchos usuarios que algo se desplegaría o alguna acción estaba asociada a éste. El segundo problema lo encontraron una vez que llegaron a la página adecuada de la facultad, donde si bien hay poca información muchos usuarios tuvieron problemas en distinguir el link que pertenecía al video del resto de los links que lleva a otras páginas. Hubiera bastado con una metáfora conocida y un icono para que los usuarios lo encontraran enseguida.

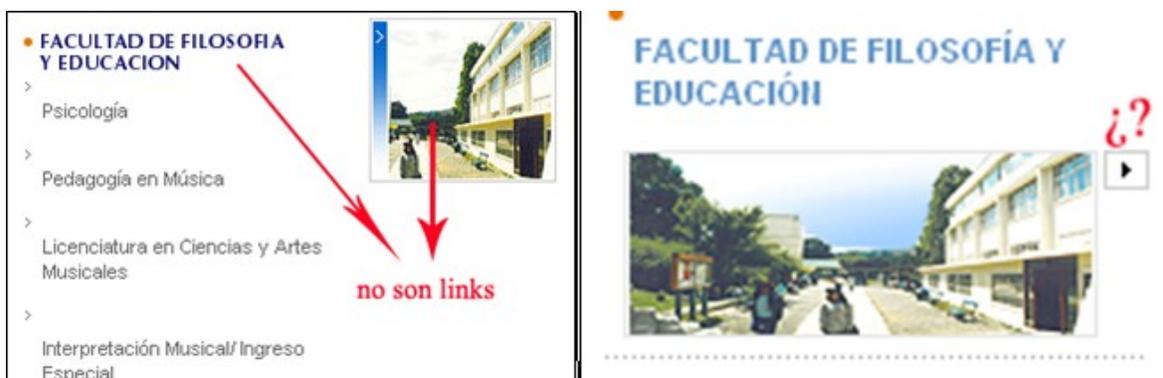


Ilustración 3.18 y 3.19: A la izquierda: el texto y la imagen eran obvios links para el usuario.

A la derecha: el recuadro con la punta de flecha no es un link, no hace nada.



Ilustración 3.20: 43% de los participantes no vio el link

Este problema no es aislado se repite en varias áreas del sitio y se puede encontrar problemas relacionados en la evaluación heurística, en los problemas 2, 4, 11, 14, 15, 20, 27, 30, 31 y 32. Todos estos problemas resultan en una navegación más compleja, donde el usuario debe escanear visualmente todo el sitio y debe pensar mucho más para poder encontrar lo que busca, y puede terminar no encontrándolo (como sucedió seguido en las pruebas de usabilidad).

8. Problema de inconsistencia de estilo

Otro problema altamente reportado tanto en la evaluación heurística (nº 18, 23, 24, 26, 28 y 29) como en los cuestionarios de salida es la inconsistencia que tiene el sitio a través de las distintas áreas que lo componen. Esta inconsistencia se extiende tanto al estilo gráfico y diseño como a las metáforas, uso de lenguaje e idioma. Este tipo de problema hace que un usuario tenga que reaprender toda una interfaz cada vez que accede a un área distinta del portal, aumentando su carga en memoria y contradiciendo varias heurísticas y principios de

usabilidad. Se puede destacar también que muchas de estas áreas poseen su propio dominio dificultando la minería web.



Ilustración 3.21: Muchas áreas del sitio presentan su propia interfaz distinta

9. Problema de estructura y navegación

Cuando se entregaron los resultados del ordenamiento de tarjetas ya se adelantó algo sobre este problema. En esta actividad es posible observar que el grado de acuerdo alcanzado por cada grupo es bajo y que sólo un 9% de las tarjetas estuvo entre 2 y 3 grupos y ninguna tarjeta fue exclusiva de un grupo. Esto hace ver que la gran cantidad y variedad de los temas se hace difícil de digerir para los participantes, los que no llegaron a un acuerdo total sobre como debería ser la estructura del sitio. Lamentablemente esto nos hace ver el problema pero no nos ofrece una solución concreta, ya que la mayoría de los participantes estructuró el sitio (muchos sin conocerlo) de manera similar a como este ya está construido. Y la manera en que este está construido ocasiona problemas que se pueden encontrar mencionados en la evaluación heurística en los ítems 2, 3, 7, 8, 9 y 15.

Estos problemas se observaron frecuentemente en las pruebas de usabilidad, donde se hizo obvio que a los participantes les costaba encontrar enlaces claros y significativos donde acceder para ejecutar la tarea. Muchos usuarios se pierden en el camino y desisten, otros recorren el portal perdidos hasta que desisten y esto puede verse reflejado en WUM que muestra que 80% de los accesos es sólo al sitio principal, además de decirnos que 24% de los visitantes sólo ve una página y que la mitad de estos ve el sitio por un minuto o menos. Cabe destacar además que sólo 6 tareas de las pruebas de usabilidad tienen promedios de compleción menor al minuto y que más de la mitad tiene altos índices de dificultad, siendo que son tareas que deberían ser rutinarias y accesibles.

En la etapa de definición de objetivos de la evaluación de usabilidad se establecieron 4 puntos que se pretendían observar, los cuales se mostrarán ahora a modo de resumen de los resultados generales de la evaluación.

- *Facilidad y efectividad de la navegación:*

Medio/bajo. El sitio cuenta con problemas de estructura, metáforas poco significativas, inconsistencia en los estilos e idiomas lo que hace que el sitio sea confuso y poco navegable. Muchas tareas que deberían ser simples tuvieron tiempos inapropiados de compleción debido a los diversos problemas existentes.

- *Utilidad del contenido: Medio/alto.*

Los contenidos del sitio web son del agrado de los usuarios, tal como ellos lo dieron a conocer en las encuestas y formularios. El sitio cuenta con una gran cantidad de información útil para alumnos y posibles alumnos, aunque el foco del sitio parece ser captar alumnos nuevos más que asistir a los existentes. La manera en que el contenido está dispuesto no es del agrado del promedio, pero la información está.

- *Efectividad de la presentación de la interfaz: Media/baja.*

Los íconos, vínculos, imágenes y contenido del sitio se presentan de forma inconsistente y en desorden. Ya que el contenido es útil la efectividad para presentarlo es baja, como se muestra en el mediocre porcentaje de compleción de las tareas y en sus largos tiempos de

compleción. Los formularios mostraron el desagrado del diseño del sitio por parte de los usuarios más experimentados y visitantes frecuentes, mientras que parecía complacer a los visitantes más remotos.

- *Éxito en la realización de las tareas: Media.*

El promedio general de compleción de las tareas fue de un 62%, con un promedio de tiempo de 1 minuto y 42 segundos. Esto muestra un resultado medio, no óptimo pero tampoco pésimo, que refleja la baja satisfacción del usuario ante el sitio frente a tareas críticas y frecuentes como buscar un libro en biblioteca, y la media/alta satisfacción frente a tareas cortas y específicas (por lo general orientadas a la captación de alumnos) como encontrar el puntaje de cierre de una carrera para el año anterior.

Recomendaciones y propuesta de soluciones

Aunque el objetivo del presente proyecto no es realizar un análisis de usabilidad a fondo que incluya soluciones, se presentan a continuación recomendaciones que puedan ayudar a la solución de los problemas descritos, con el fin de mostrar la validez y fidelidad del experimento realizado.

1. Para el problema de navegabilidad y acceso vía buscadores

Se recomienda prestar mayor atención a los buscadores, la indexación y la tecnología utilizada en el sitio. Las páginas con marcos (frames) ya no son utilizadas ya que provocan varios problemas [38] como los que se han dado acá así que es recomendable reconstruir aquellas páginas que los usen usando mejores tecnologías como CSS (Cascade Styling Sheets). Tener un sitio que aparezca en un buen lugar en los motores de búsqueda puede ser una gran ventaja si se sabe aprovechar.

2. Problema de actualización del sitio y acceso vía buscadores

Además de lo descrito en el punto anterior se recomienda hacer una revisión de aquellas páginas que están activas en el servidor y de los enlaces que componen el sitio, buscando que todos estén correctos y actualizados.

3. Problema de accesibilidad para usuario del navegador Firefox

Se recomienda prestar atención a la tecnología y métodos utilizados para construir el sitio, para que esta sea compatible con la mayoría de los navegadores. No se debe obligar al usuario a renunciar a sus preferencias.

4. Problemas con la herramienta de búsqueda del sitio

Habiendo remarcado la importancia de la búsqueda interna se recomienda que la herramienta de búsqueda sea mejorada por una más dinámica, en español y con mayor potencia. Además la herramienta de búsqueda debería aparecer claramente en el portal del sitio para su fácil acceso por lo visitantes. Se recomienda reemplazar la herramienta de acceso directo para darle ese lugar a una barra de búsqueda, o en su defecto complementar ambas.

5. Problemas con la búsqueda en biblioteca

Se recomienda en este punto hacer más claro e intuitivo el uso de la herramienta, enfocando su diseño no a como está construido el sistema de biblioteca, sino a que es lo que necesitan buscar los usuarios. A modo de ejemplo, poner menos opciones de búsqueda, donde la opción por defecto sea la búsqueda en el catálogo de libros que la biblioteca dispone para los estudiantes, y donde esta opción sea algo entendible por un estudiante como “libros para préstamo” o “libros disponibles” y no “Opac”. También se recomienda poner atención a las reglas y costumbres de navegación, donde por ejemplo presionar “enter” en un formulario es considerado como “submit” o ejecutar. Tómese en consideración que a algunos participantes les tomó casi 5 minutos encontrar un libro en el sistema.

6. Problemas de íconos, metáforas y links

Se recomienda al igual que en punto anterior prestar atención a las costumbres y convenciones en el diseño de sitios web, además de realizar una revisión de la navegación lógica sobre el sitio. Se recuerda que poner un hipervínculo es fácil y que los usuarios tienden a buscar imágenes o íconos que los guíen con mayor facilidad ya que el usuario prefiere reconocer antes que recordar o descubrir desde cero.

7. Problema de inconsistencia de estilo

Se recomienda pensar en el sitio como un todo integrado, una parte crucial de la organización que debe estar alineada con la estrategia y debe reflejar el carácter de la PUCV. El sitio como esta ahora refleja sólo una unión azarosa de páginas web sin consistencia ni sustancia. Mantener un estilo y consistencia de convenciones y metáforas a lo largo del sitio ayudará a los usuarios a reducir su carga en memoria, mejorar la realización de sus tareas y aumentará su satisfacción en el uso del mismo.

8. Problema de estructura y navegación

Este es el problema principal y más crítico del sitio, y probablemente el más difícil y costoso de corregir. Se recomienda para solucionar los problemas presentes en ésta área una reestructuración del portal además de la inclusión de guías para la navegación. Esta reestructuración debe tener por objetivo aminorar la carga de información que es presentada al usuario de una sola vez, permitiéndole absorber ésta más rápido. Recomendaciones puntuales podrían ser a modo de ejemplo:

reducir la cantidad de noticias presentadas en la página inicial
reducir la cantidad de links poco relevantes
integrar un menú que despliegue más opciones de forma jerárquica
añadir textos alternativos explicativos (tooltips) en links e imágenes importantes
añadir un guía de navegación
añadir índices y anclas en páginas cuyo texto se extiende demasiado verticalmente

También sería interesante poder dar a elegir vistas personalizadas a los usuarios del sitio, los cuales pueden identificarse vía usuario y contraseña si así lo desean, de esta forma un usuario regular del sitio podría sentirse más cómodo con éste. Sea como fuere que se implemente el cambio este debe guiar al navegante por ciertos caminos lógicos más que ofrecerle toda la diversidad de la universidad de golpe.

9. Problema de carga del servidor y transferencia de archivos

Para este problema se recomienda incorporar tecnología que permita la visualización en tiempo real de los archivos sin tener que descargarlos. En el caso de los videos la tecnología streaming de flash (formato .flv) utilizada ampliamente en la actualidad en sitios como Youtube.com es recomendada. Esto reduciría tanto la carga en el servidor como en el cliente. Otra recomendación nace de la tarea nº 11 de las pruebas de usabilidad, en la que muchos usuarios ingresaron al vínculo “Video PUCV” en la portada pensando que era una sección donde podían encontrar videos sobre la PUCV. La implementación de esta área sería una buena idea, mucho más accesible que los videos así como están puestos ahora.

3.9.7 Análisis y validación de la propuesta de integración de WUM al proceso de evaluación

A lo largo del presente se han presentado varios estudios teóricos, investigación de conceptos y casos prácticos, posteriormente los conocimientos adquiridos se pusieron en práctica al realizar una evaluación de usabilidad completa sobre un sitio web. Todo este proceso tiene como intención hacer un aporte a la evaluación de usabilidad de sitios web, en este caso, mediante Web Usage Mining (WUM).

Anteriormente a la descripción de la realización de la evaluación de usabilidad se hizo una propuesta sobre como integrar WUM a un proceso más completo de evaluación basados en el hecho de que los logs de acceso del servidor no son fuente suficiente de información para hacer análisis y recomendaciones de usabilidad. La evaluación se hizo basada en la propuesta, buscando contrastar la teoría con práctica, realizando las 5 actividades propuestas de forma integrada. ¿Qué fue lo que se descubrió?, ¿Es válida la propuesta?

La propuesta cuenta con 2 hipótesis no excluyentes y 4 puntos, los cuales parten de la base que WUM debe realizarse dentro de un proceso más completo de evaluación de usabilidad. Estas hipótesis y puntos son:

- WUM debe ser realizado con anterioridad a las demás actividades, ya que los resultados obtenidos pueden ser de utilidad para guiar la realización de las otras actividades.

- Los resultados de todas las actividades encuentran un complemento en la etapa de análisis, es decir, los resultados de WUM se complementan con los resultados del resto de las actividades una vez que se pasa a analizar la información obtenida.
 1. Desarrollar las preguntas que se usarán en otras actividades
 2. Desarrollar hipótesis que se pueden probar mediante otras actividades
 3. Probar y apoyar hipótesis que se desarrollen en otras actividades
 4. Obtener datos para mejorar el sitio web en aspectos técnicos

Los primeros dos puntos se corresponden con la primera hipótesis y el tercer punto se corresponde la segunda. El cuarto punto puede darse en cualquiera de la dos hipótesis.

El resultado de las actividades muestra que ambas hipótesis y cuatro puntos son posibles, dando los siguientes ejemplos para cada punto:

1. Desarrollar las preguntas que se usarán en otras actividades

En el presente caso, debido a la gran cantidad de accesos que presentaba la página principal frente al resto de las áreas (80% frente a un 1% la siguiente más alta) surgió la pregunta n° 10 de los formularios de entrada y usuario frecuente. Esta pregunta buscaba obtener la opinión de los usuarios respecto a que buscaban en el sitio cuando ingresaban a él y se obtuvo que la mayoría sólo buscaba el web mail y el navegador académico (ambos se encuentran juntos), cuyos accesos no quedan registrados en el servidor analizado.

Otra pregunta desarrollada influenciada por los resultados de WUM es la pregunta 11 del cuestionario para usuarios frecuentes. Esta pregunta busca saber desde donde prefieren acceder al sitio los usuarios, dado el alto nivel de accesos registrados desde dentro de la universidad. Los cuestionarios confirmaron la preferencia de los usuarios de acceder desde dentro de la universidad.

2. Desarrollar hipótesis que se pueden probar mediante otras actividades

Debido a los datos entregados por WUM que nos dicen que un importante porcentaje de usuarios visita en poca profundidad el sitio (app una página por sesión) y por poco tiempo (un minuto o menos) surgieron dos hipótesis interesantes:

Los usuarios encuentran lo que buscan rápidamente dentro del sitio

Los usuarios se frustran rápidamente en el sitio y lo dejan

Las pruebas de usabilidad muestran ambas son posibles, dependiendo del objetivo de la visita. Para visitas rutinarias, que varios usuarios confirmaron hacer, al web mail o navegador académico la primera hipótesis es cierta, mientras que para visitas menos usuales como buscar un formulario para un trámite (tarea nº 18) la segunda hipótesis es probable.

3. Probar y apoyar hipótesis que se desarrollen en otras actividades

En este punto WUM se mostró útil en esta evaluación, aportando con datos que corroboraron varios de los problemas encontrados y reportados en el presente informe. A modo de ejemplo se encuentra el problema de accesibilidad con Firefox que fue encontrado al contrastar los datos de los cuestionarios con los datos de WUM, también el problema de la búsqueda interna del sitio en que WUM comprueba lo obtenido mediante las pruebas de usabilidad respecto a la opción de búsqueda más lógica para los usuarios (y la vez errada), por último cabe mencionar el problema del sistema de búsqueda de biblioteca donde WUM ayuda a resaltar la importancia del problema siendo esta área la 6ta en la lista de área más accedidas, recibiendo 3.119 visitantes en un mes muchos de los cuales deben haber encontrado todos los problemas descritos.

4. Obtener datos para mejorar el sitio web en aspectos técnicos

Este punto puede ser probablemente donde hubo mejores resultados. Tres problemas fueron encontrados sólo usando WUM antes de realizar el resto de las actividades, además de encontrar marcadas y posiblemente valiosas tendencias en el aspecto temporal del ingreso al sitio. Se descubrieron problemas de accesibilidad con Firefox, de carga del servidor debido a la descarga de videos, problemas de actualización del sitio y problemas con los buscadores y la navegación del sitio.

Conclusión de la propuesta

Como conclusión los 4 puntos son posibles haciendo que ambas hipótesis sean también posibles. Es por lo tanto recomendable incluir WUM en un proceso de evaluación de usabilidad de sitios web, ya que aportará con información útil por si sólo como también en complemento de las demás actividades. Queda probado que es conveniente realizar WUM con anterioridad a las demás actividades ya que puede ser de utilidad.

Como parte del experimento general del proyecto se realizaron dos evaluaciones de usabilidad sobre dos sitios web distintos, uno de estos procesos se realizó sin WUM y el otro con este. El realizador de este proyecto considera que comparativamente, la evaluación con WUM es más completa que la otra ya que posee mayores fundamentos y consta de datos particulares obtenidos de los participantes y de datos generales obtenidos del total de usuarios del sitio.

Respecto a lo propuesto al comienzo del proyecto sobre encontrar una relación entre las salidas de WUM y las entradas de las pruebas de usabilidad, en realidad no existe una correlación fija e inalterable. No se puede asegurar con seguridad que ciertos datos específicos de WUM serán siempre útiles y necesarios para ciertas actividades de evaluación. En el caso particular de este proyecto los ejemplos dados anteriormente muestran como si existió cierta relación entre la minería y las actividades, pero esta relación puede variar dependiendo de cada sitio y sus datos encontrados. Por lo tanto la relación encontrada es que los datos de salida de WUM pueden servir como guía en la realización de las otras actividades, pero la forma en que esto sucede dependerá del caso particular analizado.

Por otro lado, se confirma la necesidad de complementar WUM con otras técnicas de evaluación de usabilidad, ya que en el desarrollo de la evaluación sobre el sitio de la PUCV se vislumbró que técnicas como pruebas con usuarios y encuestas son mejores a la hora de encontrar problemas. Además en la lista de problemas encontrados puede verse que hubo problemas encontrados con y sin WUM, lo que ratifica que se debe complementar y no usar por si sólo.

Para finalizar se presentarán las ventajas y desventajas de Web Usage Mining encontradas en la realización del presente proyecto

Ventajas y desventajas de WUM

Ventajas

- Información técnica fácilmente disponible: los logs no son difíciles de obtener en la mayoría de los casos, quizás se tengan mayores complicaciones si el evaluador no es el dueño del servidor, pero si existe buena comunicación y disponibilidad debería haber buenos resultados.
- Se pueden encontrar problemas directamente: como se ha visto en la realización del presente proyecto, 3 problemas fueron encontrados sólo usando la información entregada por las herramientas, lo cual fundamenta bien el uso de estas herramientas para mejorar un sitio web.
- Herramientas gratuitas: existen algunas herramientas gratuitas y accesibles y se espera que existan más en el futuro, cuando se haya realizado más trabajo en el área.

Desventajas

- El análisis puede ser confuso: las herramientas de WUM actualmente no permiten hacer una distinción clara del sitio, como lo sería por ejemplo, si generaran un mapa interactivo del sitio. En la Ilustración nº 22 se puede observar como un evaluador podría pensar que la primera ocurrencia en la lista es la página más visitada, cuando en realidad las primeras ocurrencias son todas partes de la página principal (menú, banner, footer, top, etc.). Este tipo de problema hace difícil examinar los resultados de forma rápida e intuitiva.
- El sitio no siempre está diseñado para que sea fácil de analizar: como se puede ver en la imagen las carpetas y archivos no tienen por lo general un nombre significativo que indique que es y donde está sólo con leerlo.
- Los accesos a otros servidores se deben integrar o la información queda incompleta: como se ha descrito antes, en servidor guarda sólo los accesos que éste a recibido, si se

quiere realizar un estudio de minería de uso web más completo se deberá integrar los datos encontrados por otros servidores relacionados al del sitio en cuestión.

Ilustración 3.22: Análisis de las páginas más visitadas

Rank	Page	Hits	%
1	/p4_extension/site/port/menu.html	274892	10.81%
2	/fset/head.html	262754	10.34%
3	/p4_extension/site/port/portada.html	260643	10.25%
4	/	249914	09.83%
5	/footer/foot.html	261760	10.30%
6	/fset/cont1.html	237877	09.36%
7	/prontus_marquee/site/edic/base/port/marquee.xml	490528	19.30%
8	/p3_carrera/site/port/carreras.html	22346	00.88%
9	/p3_admision/site/port/admision.html	5773	00.23%
10	/postgrado/pags/cont.html	4967	00.20%
11	/p3_postgrado/site/port/postgrado.html	4769	00.19%
12	/site/pags/top/top.html	4891	00.19%



Página inicial

4. Conclusiones

En el presente informe se ha logrado describir el marco teórico base para la realización del proyecto, presentando las definiciones, ejemplos, estado del arte, casos prácticos, propuesta de integración, realización de las actividades de evaluación y comprobación de la propuesta. Se ha logrado satisfactoriamente terminar con lo establecido al inicio del proyecto y en los objetivos.

El desarrollo del proyecto tuvo algunos percances y atrasos debido principalmente a la dependencia de las empresas dueñas de los servidores y los logs de acceso, los cuales son indispensables para el proyecto. La recopilación de información ha sido menos productiva de lo planeado debido a la escasa información gratuita respecto al tema de WUM y sus casos prácticos en la web. De todas formas el estudio teórico y práctico han sido completados satisfactoriamente, y se ha logrado realizar una propuesta de integración de WUM al proceso de evaluación de sitios web, logrando buenos resultados.

Como se explica más detalladamente en la sección anterior, los objetivos propuestos respecto a la propuesta de integración y su validación fueron cumplidos como se esperaba. WUM resulta de utilidad en un proceso de evaluación de usabilidad, apoyando y complementando los resultados obtenidos mediante otras actividades y aportando con datos propios y únicos. La minería ayudó a encontrar varios problemas por si sola y como complemento del resto de las actividades y sus datos sirven también para proponer soluciones y mejoras.

Si bien el proyecto ha terminado este da lugar a posibles trabajos futuros para otros tesis o estudiantes. Mucho trabajo queda por ser desarrollado en el área, especialmente mejorando las tecnologías y herramientas existentes para lograr un análisis más preciso y detallado. Herramientas que provean análisis de caminos más completos y que generen informes más “amigables” y fáciles de comprender. El área de la personalización web a

partir de web mining y herramientas de captura de datos de la interfaz promete trabajo y mejoras que pueden llegar a buenos resultados.

La documentación y casos prácticos encontrados en la literatura tienden a enfocarse en estudios de mercado, donde se busca conocer al usuario para poder convertirlo en cliente o afianzarlo como comprador, pero no se debe olvidar que uno de los objetivos principales de la web es entregar y difundir información, la cual crece cada día y hace necesario mejores formas de difundirla. La usabilidad juega un papel importante en la ayuda a la mejor difusión y comprensión de la información, al mejor uso de los sistemas de todo tipo (gubernamentales, médicos, de ayuda al ciudadano, etc.), a la inclusión de las personas con discapacidades al mundo de los sistemas de información y otras facetas que no necesariamente tienen que ver con el comercio electrónico. Por esto es que este proyecto ha dado resultados exitosos y novedosos al abarcar el tema de *Web Usabe Mining* desde el punto de vista de la usabilidad, en un proceso de evaluación clásico.

5. Bibliografía

- [1]. Mariano Silva V , *Web Mining: definiciones y aplicaciones*, Webmining.cl, 2001, <http://www.webmining.cl/doc.asp?docid=33>
- [2]. Jakob Nielsen, *Jakob Nielsen's Website*, <http://www.useit.com/>
- [3]. DAEDALUS , *Minería Web Documento Básico DAEDALUS*, DAEDALUS – Data, Decisions and Language, S.A 2002.
- [4]. Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, *Web Mining: Information and Pattern Discovery on the World Wide Web*, Department of Computer Science University of Minnesota. <http://maya.cs.depaul.edu/~mobasher/webminer/survey/survey.html>
- [5]. Julio Villena, José Carlos González, Emma Barceló, Juan Ramón Velasco, *Minería de uso de la web mediante huellas y sesiones*, Depto. Ingeniería de Sistemas Telemáticos–ETSIT–Universidad Politécnica de Madrid.
- [6]. José Luis Ortega Priego , *Análisis de sesiones de la web del Cindoc: una aproximación a la minería de uso web*, El profesional de la información, 2005, mayo-junio, v. 14, n. 3, pp. 190-198.
- [7]. Dr. Cristian Russu, *Curso: Human Computer Interaction*, PUCV 2006,
- [8]. IBM, *IBM Ease of Use – Web Design Guidelines*, <http://www-03.ibm.com/easy/page/561>
- [9]. Aimée Truchard , *Usability Techniques Ten Guidelines for User-Centered Web Design*, Usability Interface, Vol 5, No. 1, July 1998, <http://www.stcsig.org/usability/newsletter/9807-webguide.html>

- [10]. Lisa Halabi, *Expert usability review vs. usability testing*, WebCredible.com, Enero 2007, <http://www.webcredible.co.uk/user-friendly-resources/web-usability/website-review.shtml>
- [11]. J.M. Cueva Lovelle, *Métricas de usabilidad en la Web*, 2004.
- [12]. Melody Y. Ivory and Marti A. Hearst, *The State of the Art in Automating Usability Evaluation of User Interfaces*, University of California, Berkeley.
- [13]. Techsmith Corporation, *White paper: Web analytics and usability testing: understanding the total customer experience*, Techsmith Corporation, 2005
- [14]. Cyrus Shahabi, Farnoush Banaei-Kashani, *Efficient and Anonymous Web-Usage Mining for Web Personalization*, Department of Computer Science, Integrated Media Systems Center, University of Southern California, Los Angeles, California 90089-2561, USA,
- [15]. Girardi, Leandro Balby Marinho, Ismenia Ribeiro de Oliveira , *A system of agent-based software patterns for user modeling based on usage mining*, Rosario, Federal University of Maranhao Brazil
- [16]. Massimiliano Albanese, Antonio Picariello, Carlo y Lucio Sansone, *A Web Personalization System based on Web Usage Mining Techniques*, Departamento de Informática y Sistemas Universidad de Nápoles Federico II
- [17]. Trenton Moss, *Web usability: The basics* , Enero 2004, <http://www.webcredible.co.uk/user-friendly-resources/web-usability/basics.shtml>
- [18]. Leticia dos Santos Machado, Karin Becker, *Distance Education: a Web Usage Mining Case Study for the Evaluation of Learning Site*”, Facultad de Informática – Pontificia Universidad Católica de Rio Grande do Sul – Brazil
- [19]. M. E. Zorrilla, E. Menasalvas, D. Marín, E. Mora,J. Segovia, *Web Usage Mining Project for Improving Web-based Learning Sites*.

- [20]. Marta Zorilla, *Introducion al Web Mining*, Universidad de Cantabria
- [21]. Myra Spiliopoulou, Bettina Berendt, Ernestina Menasalvas, *Evaluation in Web Mining*
- [22]. Paul J. Bracke, *Web usage mining at an academic health sciences library: an exploratory study*, Arizona Health Sciences Library, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=521513>
- [23]. Mena J.; Pettit R., *Web-Mining Case Study: An Internet Radio Website*, Interactive Marketing, Volumen 3, Número 1, Julio/Septiembre 2001, <http://www.ingentaconnect.com/content/pal/im/2001/00000003/00000001/art00007>
- [24]. Zhizhong Li, Dirk Roesler, Katalin Meszlery, *Discovering development of personnel qualification requirements by web mining: a case study*, International Journal of Internet and Enterprise Management, <http://www.inderscience.com/offer.php?id=4939>
- [25]. Usability.gov, *Usability Methods*, <http://www.usability.gov/methods/>
- [26]. Web-Based User Interface Evaluation with Questionnaires <http://www.acm.org/~perlman/question.html>
- [27]. Reporte de ejemplo de Alterwind, <http://www.alterwind.com/loganalyzer/samplerreport/>
- [28]. KDnuggets, *Poll: web mining tools*, KDnuggets.com, 2001, <http://www.kdnuggets.com/polls/2001/web-mining.htm>
- [29]. Analog, *Web Server Statistics for analog documentation*, <http://www.chiark.greenend.org.uk/~sret1/stats/>
- [30]. Tec-Ed inc, *Assessing Web Site Usability from Server Log Files*, 1999.
- [31]. Susan Haigh, Janette Megarity, *Measuring Web Site Usage: Log File Analysis*, 1998.

- [32]. Michael D. Levi, Frederick G. Conrad, *Usability Testing of World Wide Web Sites*, 2002.
- [33]. MoreBussines.com, *Web Site Access Logs: Guess (Literally) Who's Coming to Visit*, 1999, http://www.morebusiness.com/getting_started/website/d935705350.brc
- [34]. Jakob Nielsen, *Why You Only Need to Test With 5 Users*, 2000, <http://www.useit.com/alertbox/20000319.html>
- [35]. Jakob Nielsen , *Card sorting: how many users to test*, 2004, <http://www.useit.com/alertbox/20040719.html>
- [36]. Donna Maurer , *Card Sorting: a Definitive Guide*, 2004, http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide
- [37]. Joe Lamantia , *Analyzing Card Sort Results with a Spreadsheet Template*, 2003.
- [38]. Jakob Nielsen, *Why Frames Suck (Most of the Time)*, 1996. <http://www.useit.com/alertbox/9612.html>

Anexos

Ver Anexos en: