
UNA APLICACIÓN DE HERRAMIENTAS DE EYE-TRACKING PARA ANALIZAR LAS PREFERENCIAS DE CONTENIDO DE LOS USUARIOS DE SITIOS WEB

LARRY GONZÁLEZ*
JUAN D. VELÁSQUEZ*

Resumen

En el presente artículo introducimos el seguimiento ocular de los usuarios que visitan un sitio Web, como una nueva fuente de datos para analizar su comportamiento de navegación y preferencias. La aplicación práctica de este nuevo enfoque, es la identificación de los objetos más importantes de un sitio, desde el punto de vista del usuario, también llamados *WebSite KeyObjects*. La captura de este tipo de datos se logra a través del uso de una herramienta de eye-tracking, para dar seguimiento al movimiento ocular de un usuario en una pantalla del computador. Utilizando estos datos, se propone mejorar la efectividad de una metodología de identificación de Website KeyObjects, la cual necesita de la aplicación de una encuesta de percepciones a los usuarios para categorizar la importancia de los contenidos de las páginas web. Con el uso de la tecnología eye-tracking, la encuesta se vuelve innecesaria, lográndose un análisis más objetivo y certero de las preferencias de los usuarios en un sitio web. Esta nueva forma de captura de datos fue probada en un sitio web nacional, correspondiente a un programa de MBA. Al contrastar los resultados de la metodología para la identificación de Website Keywords usando la encuesta versus el uso de los datos obtenidos por sistema de eye-tracking, se puede observar una mejora en la detección de dichos objetos de entre un 15 % al 20 %, lo cual demuestra la utilidad práctica de este nuevo enfoque en el contexto del análisis del comportamiento del usuario en la Web.

Palabras Clave: *Web Mining, Eye Tracking, Website Keyobjects.*

*Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

1. Introducción

En los últimos años, el *Web Mining* se ha transformado en una de las áreas de investigación más prometedoras para entender los cambios relacionados con el fenómeno de la Web. Gracias a las técnicas, métodos y algoritmos consignados en esta disciplina, es posible extraer información valiosa que permite una mayor comprensión sobre el comportamiento de navegación y preferencias de los usuarios de un sitio.

Dependiendo del tipo de dato a procesar, *Web Mining* puede ser dividido en tres grandes grupos [17]: Web Structure Mining (WSM), en el análisis de la evolución de la estructura de hipervínculos de un sitio, Web Content Mining (WCM), relacionado con el análisis de los contenidos y Web Usage Mining (WUM) para conocer un poco más de la navegación de los usuarios en el sitio.

Un resultado interesante, y que surge de la combinación del WCM y WUM, son las *Website Keywords* [19], que se definen como una palabra o un conjunto de palabras utilizadas por los usuarios en su proceso de búsqueda de información y que caracterizan el contenido de una página o sitio web. Al encontrarlas, los sitios pueden ser re-diseñados de acuerdo a las necesidades y requerimientos de sus usuarios y, de esta forma, estar a la vanguardia. Si bien identificar los *Website Keywords* de un sitio ayuda a conocer las preferencias de los usuarios, la metodología utilizada para su identificación se enfoca principalmente en el contenido textual, dejando prácticamente de lado el análisis del contenido multimedia de los sitios web [15]. Por esta razón en [5] extendieron esta metodología e integraron, tanto el contenido textual como el contenido multimedia en su análisis.

Se define un *Web Object* como cualquier “*grupo estructurado de palabras o un recurso multimedia que está presente en una página web que posee metadatos que describen su contenido*”. Asimismo, un *methodologyIdentifyingKeyobjects* es un *Web Object* que captura la atención de los usuarios y que caracteriza el contenido de un sitio web [16]. A partir de las definiciones anteriores se puede deducir que todo sitio Web está formado por un grupo de *Web Object* y que el conjunto de *Website Keyobject* que posee es sub-conjunto del primero.

En [16] se propuso una metodología que permitió identificar los *Website Keyobjects* de un sitio. Estos, al igual que *Website Keywords*, dan directrices para que los Web Sites sean re-diseñados en función de las necesidades de los usuarios. Esta metodología requiere conocer el tiempo de permanencia de los usuarios. Esta metodología requiere conocer el tiempo de permanencia de los usuarios en los *Web Objects*, es decir, cuánto tiempo gasta un usuario mirando

cada *Web Object*. Para determinar el tiempo de permanencia, plantean dos pasos: *Sesionización* y *Aplicación de una encuesta*.

1. **Sesionización:** es el proceso que reconstruye las sesiones de usuarios. Una sesión es la secuencia de páginas web que visita un usuario mientras navega en un sitio. Mediante este proceso es posible determinar, por ejemplo, el tiempo de permanencia de un usuario de cada página, entre otras cosas.
2. **Aplicación de una encuesta:** se refiere a la aplicación de una encuesta sobre un grupo de control, con la que los *Web Objects* fueron ordenados según su importancia, dentro de cada página del sitio Web.

Al mezclar ambos procesos [16], se logra estimar tiempo de permanencia de los usuarios en los *Web Object*. Lo anterior no está exento de ruidos, siendo el principal la aplicación de una encuesta que por sus características, los datos que entrega son altamente subjetivos.

La utilización de la tecnología *Eye Tracking* permitirá cuantificar del tiempo de permanencia de un grupo de usuarios de control en los *Web Objects*, con lo que se tendrá una medida más certera y objetiva de su interés en los objetos que componen el sitio.

El presente artículo muestra en la sección 2 un estado del arte respecto de las técnicas de extracción de información desde los datos originados en la Web, de las herramientas de seguimiento ocular, conocidas como *Eye Tracking* y como se combinan en el análisis del comportamiento de navegación y preferencia de los usuarios que visitan un sitio web. En la sección 3, se profundiza en la explicación del método propuesto para mejorar la detección de los *Web Objects*, lo que luego conducirá a la identificación de los *Website Keyobjects*, a través de los experimentos detallados en la sección 4. Finalmente, en la sección 5, se muestran las principales conclusiones y trabajos futuros a desarrollar.

2. Trabajo Relacionado

En esta sección, se explicarán los fundamentos teórico-prácticos del funcionamiento de la Web, cómo la navegación de un usuario en un sitio puede ser monitoreada, la aplicación de las herramientas de *Eye Tracking* en la detección de los objetos observados por los usuarios y como se pueden combinar todas estas técnicas para la detección de los *Website Keyobjects*.

2.1. La Web

En palabras de su creador, *Berners-Lee*, la Web “*es el universo de información accesible desde la red, una encarnación del conocimiento humano*”. Por su parte, en [18] se definió la Web como “*un canal masivo para la difusión e intercambio de información*”.

2.1.1. Datos Originados en la Web

Los datos que se originan en la Web son clasificados en tres tipos: contenido, estructura y usabilidad.

Contenido: se refiere a los objetos presentes en las páginas Web, como texto, imágenes, sonidos y videos, en otras palabras, todo lo que se puede ver en una página.

Estructura: Son los enlaces o *links* entre las páginas. Por lo general cuando existe un enlace entre dos páginas, éstas están relacionadas por su contenido.

Usabilidad: Son los datos generados por los usuarios en su proceso de navegación, en tanto los servidores Web almacenan cada petición realizada por los usuarios en un archivo llamado *web log*.

2.2. Web Mining

El *Web Mining* es la aplicación de Data Mining a los Datos originados en la Web [3]. Se concibe como un producto del cruce de varias áreas de investigación, como Bases de Datos, Recuperación de la Información, Inteligencia Artificial, especialmente, las subáreas de aprendizaje de máquina y procesamiento del lenguaje. La investigación en este campo está experimentando un importante crecimiento, a causa de la gran cantidad de datos disponibles para ser analizados [9].

Los datos originados en la Web pueden ser clasificados en tres categorías [17]:

- **Web Content Mining (WCM):** El objetivo es descubrir información útil desde los documentos Web [9]. WCM no está limitado sólo al análisis del texto de las páginas Web, sino que también incluye otros tipos de documentos, como imágenes y videos [18]. Sin embargo, el análisis sobre este tipo de datos, denominado *Multimedia Data Mining*, no recibe tanta atención como el análisis de texto.
- **Web Structure Mining (WSM):** Estudia los enlaces presentes en las páginas Web. Las páginas y enlaces se modelan como los nodos y los arcos

de un grafo dirigido respectivamente. El arco parte en el nodo que representa a la página que posee el enlace y termina en el nodo que representa a la página que es apuntada.

- Web Usage Mining (WUM): Está enfocado en la aplicación de técnicas de *Data Mining* para descubrir patrones útiles que puedan predecir la conducta del usuario mientras interactúa en la Web [9], a través del análisis de las sesiones de navegación entendidas como la secuencia de páginas que un usuario visita mientras navega en un sitio Web.

2.3. Percepción del Usuario Web

Un aspecto importante a considerar en el diseño de una página Web es la percepción que el usuario tendrá respecto de cuánto realmente entiende acerca del contenido y estructura que se le presenta [10]. La percepción es una función del pensamiento, i.e., cómo reconocemos e interpretamos un objeto a través de nuestros sentidos para luego codificarlo y memorizarlo, integrando información referente a nuestros conocimientos previos.

Entender la la percepción del usuario web, es sumamente importante para luego hacer mejores sitios, los que finamente atraerán y mantendrán a los usuarios. Sin embargo, no se trata de una tarea trivial a realizar, por cuanto la percepción está muy relacionada con el conocimiento previo que pueda tener un usuario, y más aun, con su experiencia personal [13].

De acuerdo al modelo conceptual de flujo propuesto por [6], durante la navegación, el usuario percibe la estructura de un sitio como un estado de cognición basado en su experiencia, el cual está determinado por: "*altos niveles de habilidad y control; altos niveles de desafío y excitación; atención focalizada, la que se ve reforzada por la interactividad y la telepresencia*". Quizás el único factor que podría ser influenciado a través del diseño del sitio web es la focalización de la atención del usuario. Sin embargo, esto podría ser una espada de doble filo, i.e., la atención puede ser capturada por el contenido del sitio es atractivo o por que es realmente desagradable, lo que motivará al usuario a nunca más volver a visitar sus páginas.

Independiente de cómo se realice el proceso, entender un poco más qué atrae al usuario cuando visita un sitio, o al menos más de claridad respecto de qué objetos le llaman la atención, serán factores claves en la mejora continua de la estructura y contenido del sitio. Al respecto, se han desarrollado varias teorías, técnicas y algoritmos para extraer información y conocimiento desde los datos originados en la Web [9]. También son destacables los trabajos en el ámbito de la teoría y practica de la usabilidad en el desarrollo de los sistemas basados en la Web [12].

2.4. Abordando la Usabilidad

En [11], se define la usabilidad como *“un atributo que permite evaluar cuán fácil para el usuario resulta ser la interfaz que se le presenta”*. Este atributo posee cinco componentes a ser analizados:

- **Facilidad de Aprendizaje:** ¿Es fácil para los usuarios realizar tareas básicas la primera vez que usan la aplicación?
- **Eficiencia:** ¿Cuán rápido el usuario puede realizar tareas una vez que ha aprendido lo básico respecto del diseño de la aplicación?
- **Memorabilidad:** Si los usuarios usan nuevamente la aplicación después de un tiempo, ¿Es fácil para ellos recordar cómo usar las operaciones básicas?
- **Errores:** ¿Están cometiendo muchos errores los usuarios? ¿Son muy malos los errores?
- **Satisfacción:** ¿Están los usuarios felices con el diseño de la aplicación?

Para analizar la usabilidad de una aplicación, utilizando los conceptos anteriormente descritos, es necesario definir una serie de indicadores que ya sea en forma cualitativa o cuantitativa permitan medir la experiencia que ha tenido un usuario frente a los estímulos que le presenta un sitio web. En este sentido, se han desarrollado varios métodos [7]:

- **Cuestionarios.** Un conjunto de preguntas creadas para conocer la opinión de un usuario frente a una aplicación.
- **Observación.** Consiste en observar directamente las acciones que ejecuta un usuario frente a una aplicación. Aquí un evaluador humano toma notas para luego mejorar la aplicación.
- **Entrevistas.** Usualmente, las entrevistas vienen después del proceso de observación y consisten en una conversación con el usuario para conocer sus impresiones acerca de una aplicación.
- **Focus group.** Es un tipo de método para hacer entrevistas donde un grupo escogido previamente de personas son consultadas por un determinado tema.
- **Pensamientos en voz alta.** Los usuarios son invitados a verbalizar sus pensamientos, sentimientos y opiniones respecto de una aplicación particular.

- Captura de pantalla. Se crea un conjunto de imágenes respecto del funcionamiento de la interfaz de una aplicación, mostrándole a los usuarios lo que se ve en la pantalla para que luego respondan un cuestionario con sus impresiones y opiniones.
- Eye-tracking. Este tipo de herramientas tecnológicas, permiten dar seguimiento al movimiento ocular del usuario frente a la pantalla del computador. Estos datos son utilizados para analizar el comportamiento de uso de la aplicación por parte del usuario.

Por supuesto que cada método expuesto tiene sus ventajas y desventajas al momento de proponer una investigación científica respecto del análisis del comportamiento del usuario en la Web. En algunos casos, para lograr una mayor objetividad en el análisis, se pueden combinar dos o más métodos. Sin embargo, el uso de los sistemas de eye-tracking están cobrando una tremenda relevancia, puesto que se trata de una forma objetiva de obtener datos referente a lo que el usuario mira en un determinado momento. Por esta razón, se les ha comenzado a usar intensivamente en el ámbito del análisis de la usabilidad de aplicaciones [12].

2.5. Eye Tracking

Es una técnica mediante la que los movimientos oculares de un individuo son medidos. De esta manera, un investigador puede conocer lo que una persona está mirando en cada momento y la secuencia en la que sus ojos se desplazan de un lugar a otro. Seguir los movimientos oculares de los usuarios puede ayudar a los investigadores de la Interacción Humano-Computador a entender el procesamiento de la información visual y los factores que pueden tener repercusiones en la usabilidad de la interfaz. Así, las grabaciones de los movimientos oculares pueden proporcionar una fuente de datos objetiva para la evaluación de interfaces, que a su vez pueden otorgar información para mejorar el diseño de las mismas [14].

2.5.1. Movimientos Oculares

Cuando observamos una escena cualquiera, los ojos de una persona se mueven entre puntos que capturan su atención, con la que logran recrear una imagen cerebral de la escena [12]. Si bien existen modelos de los movimientos oculares complejos que constan de cinco pasos [4], el modelo típico (y suficiente para los estudios de *Eye Tracking*) está constituido por dos conceptos: *fixation* (o fijación) y *saccades* (o movimientos sacádicos). La fijación se define como el momento en el que los ojos permanecen fijos sobre un objeto y es posible

apreciarlo en detalle; mientras que los movimientos sacádicos corresponden a los rápidos movimientos oculares entre dos *fixations* [12]. Es importante notar que mientras se produce un movimiento sacádico permanecemos ciegos (no somos conscientes de lo que está entre los dos objetos que capturaron nuestra atención). Sin embargo, nuestro cerebro es capaz de interpretar esta “secuencia de imágenes” como un continuo, y por ende, nuestra apreciación parece más un video que una secuencia de imágenes.

2.5.2. Atención Visual

Es un fenómeno que ha sido estudiado por cerca de cien años y que todavía no se logra comprender. Los primeros estudios estaban limitados por la tecnología, y correspondían sólo a observación e introspección. En la actualidad, este campo es estudiado por distintas disciplinas como psicofísica, neurociencia cognitiva y ciencias de la computación, por nombrar sólo algunas [4]. En términos generales, la visión humana tiene dos partes: una pequeña zona central con una resolución muy alta, llamada *visión foveal*, y la gran mayoría del campo visual con una baja resolución, llamada la *visión periférica*.

Usualmente el hecho de prestar atención a regiones de interés está relacionado con realizar movimientos oculares (*overt attention*). Sin embargo, también podemos colocar atención en objetos periféricos sin realizar este tipo de movimientos (*covert attention*).

Por otro lado, se conocen dos formas en las que la atención es guiada: *bottom-up* y *top-down*. La primera, derivada sólo de la escena visual, establece que las regiones de interés atraen nuestra atención lo suficientemente fuerte como para que no observemos el resto de la escena (visión foveal). A su vez, *top-down* es conducida por otros factores cognitivos, como el conocimiento, la expectativa y las metas actuales. Bajo este modelo, las personas son más propicias a ver a su alrededor (visión periférica); a modo de ejemplo, un individuo que conduce con regularidad, más propenso a notar las estaciones de combustible mientras realiza otra actividad que alguien que no lo conduce.

En la actualidad, aún no está claro qué es lo que realmente captura nuestra atención, ni cómo respondemos a diferentes estímulos. Existe evidencia de que prestamos atención a ubicaciones espaciales, características y objetos. La mayoría de los investigadores creen que estas teorías no son excluyentes entre sí y que, además, la atención visual puede ser desarrollada en cada una de estas subáreas. Vale la pena mencionar que los humanos podemos prestar atención simultáneamente a múltiples regiones de interés (máximo cinco).

2.5.3. La Hipótesis Mente-Ojo

Considerando las teorías descritas en la sección anterior, se planteó el siguiente modelo:

1. Dado un estímulo como una imagen, la escena es vista en su mayor parte en paralelo, a través de la visión periférica y, por lo tanto, en baja resolución. En esta etapa, las características interesantes de la imagen pueden “aparecer”.
2. En estos momentos, la atención está desconectada de la vista foveal (alta resolución), pero los ojos son rápidamente posicionados en la primera región que ha atraído.
3. Una vez que los ojos son posicionados, la fovea se alinea hacia la región de interés y la atención está ligada con la percepción, es decir, la atención del usuario ha sido capturada y, por ende, se logra observar en alta resolución.

In [12] se propuso que *“las personas están usualmente pensando en lo que están mirando. Aunque no siempre entienden lo que ven o no están totalmente enfocados en esto; si están observando algo, entonces están colocando atención, especialmente cuando están concentrados en una tarea en particular”*.

2.6. Técnicas Comúnmente Usadas en Eye Tracking

2.6.1. Electrooculografía

En la década de 1950, la electrooculografía [8] fue la técnica más usada de *Eye Tracking*. Se basaba en la medición de la diferencia de potencial eléctrico de la piel mediante el uso de electrodos ubicados al rededor de los ojos. Esto es posible, dado que la córnea se mantiene unas décimas de mV más positiva que la retina con lo que se produce la diferencia de potencial que es medida y que varía de acuerdo al movimiento de los ojos. Esta técnica mide la posición relativa de los ojos con respecto a cabeza, por lo que no es adecuada para calcular el punto de atención [1].

2.6.2. Lentes de Contacto Esclerales

Esta técnica consiste en adjuntar una referencia mecánica u óptica a un lente de contacto que será usado directamente sobre los ojos. Es necesario que el lente de contacto sea particularmente grande, de modo que se extienda sobre la córnea y la esclerótica, pues así se reduce la posibilidad de que se desplace sobre el ojo [4]. Se han usado distintos tipos de referencias sobre los lentes

de contacto siendo la más común una pequeña bobina, que puede ser ubicada desde el exterior al aplicar un campo electromagnético [1]. Si bien es una de las técnicas más precisas para medir los movimientos oculares, es también la más invasiva y causa malestar al usarlo.

2.6.3. Foto/Vídeo Oculografía

Estas técnicas son de las más atractivas debido a su versatilidad y simplicidad [2]. Normalmente, consisten en una serie de fotos y/o vídeos que guardan los movimientos oculares y que posteriormente son analizadas de forma manual o automática. Varios de estos métodos requieren que la cabeza esté fija, por ejemplo, mediante una mentonera.

2.6.4. Reflejo de la Córnea y Centro de la Pupila basado en Vídeo

Es la técnica *Eye Tracking* más usada en la actualidad. Consiste en un computador estándar de escritorio con una cámara infrarroja montada debajo de un monitor, con el software de procesamiento de imágenes para localizar e identificar el *Reflejo de la Córnea* y el *Centro de la Pupila*. Con estas características, es posible disociar los movimientos oculares de la cabeza, con lo que es posible calcular el punto de atención de los usuarios [14].

En funcionamiento, una luz infrarroja de un LED es dirigida hacia el usuario para crear reflejos notorios de las características de los ojos y, con esto, conseguir que sean más fáciles de rastrear (se ocupa luz infrarroja para evitar deslumbrar al usuario). La luz entra en la retina y una gran parte de ella se refleja, por lo que la pupila aparece como un disco brillante y bien definido (efecto conocido como *pupila brillante*). El reflejo de la córnea también es generado por la luz infrarroja, apareciendo como un pequeño, pero fuerte brillo [14]. En la Figura 1 se puede apreciar el efecto *pupila brillante* y el reflejo de la córnea. Una vez que el software de procesamiento de imágenes ha identificado el centro de la pupila y la ubicación del reflejo de la córnea, el vector que resulta de ellos se mide, y con algunos cálculos trigonométricos, el punto de atención puede ser encontrado.

2.7. Website Key Object

En [16] se definió un *Web Object* como “*un grupo estructurado de palabras o contenido multimedia, que está presente en una página Web y que posee meta datos que describen su contenido*”. En la definición anterior los meta datos son fundamentales ya que son la base de la información para construir el vector que representará el contenido de la página. Además, dos archivos

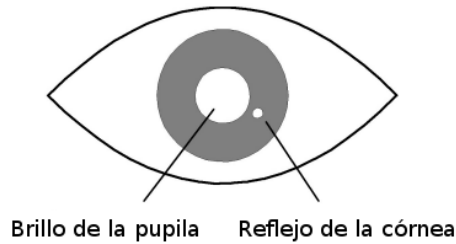


Figura 1: Reflejo de la córnea y brillo de la pupila.

multimedia pueden ser comparados mediante sus metadatos, problema que es considerablemente más abordable que el comparar directamente los archivos, debido a que sólo se compara texto. Junto con lo anterior, definieron [16] los *Web Site Key Object* como “uno o un grupo de *Web Object* que atraen la atención del usuario y que caracterizan el contenido de una página o sitio Web”. Estos proveen conocimiento acerca del contenido y formato que más interesan a los usuarios de un sitio Web, por lo que encontrarlos puede ser útil para mejorar el sitio tanto en presentación como en contenido.

2.7.1. Implementación

Considerando la primera definición, para que una página Web esté compuesta por *Web Object*, es necesario que sus objetos posean los meta datos, cosa que por lo general no ocurre. Existe una gran variedad de formas para agregarlos a las páginas Web, dependiendo de la ontología que se quiera ocupar. En [16] se asoció a cada objeto un documento XML que contiene los meta datos que describen su contenido y la página a la que este pertenece. Asimismo, en la página Web también se establece la relación entre el objeto y el documento XML. Para esto se usan *tags* de HTML. El siguiente es el formato con el que se guardaron los meta datos:

- identificador de la página
- objeto:
 - identificador.
 - formato.
 - concepto(s).

Cada objeto debe poseer al menos un concepto asociado pues en caso contrario, el objeto no tiene significado alguno. Cada concepto es un grupo de tres

sustantivos pues, en español, permiten una completa mas no suficiente definición de cualquier concepto. Cada concepto debe pertenecer a una categoría que los agrupe. Al usar estas categorías los conceptos podrán relacionarse con otros.

2.7.2. Comparación entre Objetos

Para comparar dos *Web Objects*, en ikom se guiaron por la noción de que los objetos son un cúmulo de conceptos. El siguiente procedimiento muestra cómo lograron establecer una medida de similitud entre objetos.

Considere los objetos O_1 y O_2 tal que $|O_1| = N$, $|O_2| = M$, $N \geq 0$, $M \geq 0$ y $N \leq M$. Además, sea $C_i(O)$ el i -ésimo concepto del objeto O .

a) Enlazar conceptos: El siguiente algoritmo permite enlazar los conceptos de O_1 y O_2 .

- $\forall C_i(O_1), i = 1..N$
 - $\forall C_j(O_2), j = 1..M$
 - Comparar $C_i(O_1)$ con $C_j(O_2)$ (Comparación de conceptos).
 - Guardar el resultado de la comparación en un contador.
 - Guardar un enlace entre $C_i(O_1)$ y $C_j(O_2)$ con el contador más grande (Conceptos más parecidos).

Para contrastar conceptos, se comparan las palabras que los componen. Si una palabra del primer concepto está en el segundo, se suma 1 al contador; si posee un sinónimos, se suma 0,5.

b) Ordenar conceptos: Una vez almacenados todos los enlaces entre los conceptos correspondientes a O_1 y O_2 , son ordenados de modo que queden en la misma posición relativa en función de los enlaces recientemente creados, es decir, que el primer concepto de O_1 termine enlazado con el primer concepto de O_2 .

c) Transformar conceptos en un string: Como se mencionó anteriormente, cada concepto debe pertenecer a una categoría. Cada categoría, la que es representada por un carácter. Luego, al reemplazar cada concepto por el carácter que representa a su categoría, el cúmulo de conceptos queda transformado en un string.

d) Aplicar una distancia de comparación de caracteres, e.g. Levenshtein. Entonces, como O_1 y O_2 fueron transformados en strings, son comparados mediante la distancia de Levenshtein, con lo que se tiene un indicador de qué tan parecidos son dos objetos. Mientras menor sea la distancia, más parecidos serán los objetos.

Finalmente la ecuación 1 define una similitud que permite comparar objetos, donde $L(O_1, O_2)$ es la distancia de Levenshtein de los strings que representan a los objetos y $|O|$ es el número de conceptos del objeto O .

$$do(O_1, O_2) = 1 - \frac{L(O_1, O_2)}{\max(|O_1|, |O_2|)} \quad (1)$$

2.8. Metodología para Encontrar Website Keyobject

A continuación se describen los subprocesos que permiten identificar los Website Keyobject.

Sesionización: Detallada con anterioridad (subsección 2.2), el objetivo de esta etapa es finalizar con la secuencia de páginas que visitaron los diferentes usuarios de un sitio. Además, esta secuencia debe tener asociados los tiempos de permanencia de los usuarios en cada página.

Incorporación de Metadatos: El primer paso de esta etapa es identificar los objetos que componen las páginas del sitio. Una vez identificados, se deben definir los conceptos que describen a cada uno de los objetos. Luego, esta información debe ser almacenada en una base de datos. El levantamiento de estos datos debe ser en conjunto con el *webmaster*, para asegurar que los conceptos reflejen fielmente el contenido de los objetos.

Tiempos de permanencia en objetos: Luego de definir los objetos, se realizó una encuesta a un grupo de usuarios de control de modo que cada persona entrevistada distribuyera un total de 10 puntos de interés entre todos los objetos de una página. Con estos datos, se estimó el porcentaje de permanencia de cada usuario en los objetos de cada página.

Posteriormente, en la lista de sesiones se reemplazó cada registro de página visitada por los objetos que componen dicha página, intercambiando el tiempo de permanencia de la página por el tiempo ponderado de permanencia en cada objeto.

Vector de comportamiento del usuario: Finalmente, para cada sesión identificada se seleccionaron los n objetos que capturaron más la atención del usuario, definiendo así el *Important Object Vector* (IOV) según la ecuación 2

$$v = [(o_1, t_1) \dots (o_n, t_n)] \quad (2)$$

Algoritmos de Clustering

Una vez realizada toda la limpieza y transformación de datos, se procede a procesar algoritmos de *clustering* sobre las sesiones de los usuarios, representadas mediante el *Important Object Vector*. Para poder ejecutar estos algoritmos, es de crucial importancia definir una medida de distancia, o similitud, entre estos vectores.

Medidas de similitud para sesiones: Se definió la similitud entre dos IOV mediante la ecuación 3.

$$st(\alpha, \beta) = \frac{1}{i} * \left(\sum_{k=1}^i \min\left(\frac{\tau_k^\alpha}{\tau_k^\beta}, \frac{\tau_k^\beta}{\tau_k^\alpha}\right) * do(o_k^\alpha, o_k^\beta) \right) \quad (3)$$

En la ecuación 3, α y β corresponden al identificador de las sesiones de usuarios a ser comparadas, τ_k^α corresponde al tiempo de permanencia del usuario α en el objeto o_k y $do(o_k^\alpha, o_k^\beta)$ es la similitud entre los respectivos objetos, definida en la ecuación 1. La ecuación 3 está definida entre los valores 0 y 1, siendo 0 cuando los IOV no se parecen en nada y 1 cuando son idénticos.

Esta medida de similitud fue ocupada como parámetro de entrada para los algoritmos de clustering. En [16] implementaron tres de estos algoritmos, principalmente para comparar los resultados que entregaron estas técnicas. Los algoritmos de clustering ocupados fueron: *Self Organizing Feature Maps*, *K-means* y *Association Rules*.

3. Mejorando la Detección de los Web Objects

Se plantea como hipótesis de investigación que si se puede cuantificar el tiempo de permanencia de un grupo de los usuarios en los *Web Objects*, entonces es posible mejorar la precisión al identificar los *Website Keyobject*.

3.1. Diseño del Experimento

Sobre un sitio, se ejecutará la metodología diseñada [16], en adelante la metodología original, y además se ejecutará la metodología modificada, a la cual se le incorporarán los tiempos de permanencia calculados con un *Eye Tracker*.

3.1.1. El Entorno

El sitio. En cuanto al sitio, se requiere satisfacer tres requerimientos: en primer lugar, debe tener un número de páginas adecuado. Además, la cantidad de objetos por páginas no puede ser excesivo y, por último, debe poseer una gran cantidad de visitas almacenadas (sesiones).

El grupo de control. No elegir una muestra suficientemente representativa de los usuarios que visiten el sitio, se podría llegar a resultados que no correspondan a la realidad. De contar con abundantes recursos (tiempo, dinero, personas), se recomienda altamente estudiar al menos a 39 personas distintas,

pues en este caso, el Teorema del Límite Central señala que la media y la varianza de la muestra serán similares a las de la población, donde aplicado a este caso, las variables aleatorias corresponden al tiempo de permanencia de los usuarios en los objetos Web.

El *Eye Tracker* En la actualidad, la mayoría de los *Eye Tracker* comerciales son precisos, pero su costo es elevado. Por lo que el elegido para realizar el experimento debe estar dentro del alcance económico del proyecto y su precisión debe permitir identificar los objetos que mira un usuario.

3.1.2. Captura y Transformación de Datos

Páginas. Una vez elegido el sitio, será examinado en todos sus componentes. Para realizarlo esto se empleará un *crawler* que recuperará desde la Web todas las páginas disponibles que conforman el sitio. De cada página, se almacenará su url, imagen y tamaño de la misma en píxeles.

Objetos. Como primer paso, la demarcación de los objetos debe ser, idealmente, realizada con el administrador del sitio. En caso de no ser posible trabajar en conjunto con él, será necesario que valide esta segregación. Para describir el contenido de cada objeto, se procederá según lo definido por [16], por lo que se guardará su formato y una lista de conceptos descriptores del objeto. Además, se almacenará su tamaño en píxeles y la ubicación de los mismos en la página. Luego, una vez cargados los datos de los objetos y los conceptos, se procederá a calcular la distancia conceptual entre objetos, según la ecuación 1. Estas distancias serán guardadas en una tabla relacional, de modo que al ejecutar los algoritmos de *Data Mining* no sea necesario volver a calcular la separación entre objetos.

Weblog. El archivo *weblog* (ver 2.1.1) será recuperado con el administrador de sistemas de la máquina donde está alojado el sitio. Posteriormente se llevará a cabo la sesionización, cuyos resultados serán almacenados en una tabla. Estos resultados darán a conocer la secuencia de páginas visitadas por los usuarios y la duración de su estancia en cada página.

Interés de los usuarios. Para capturar el interés de los usuarios en los diferentes objetos de las páginas, se procederá de dos maneras: uso de un *Eye Tracker* y aplicación de una encuesta.

1. *Eye Tracker*. La idea principal es dado los tamaños de páginas y objetos en píxeles y la ubicación de estos últimos, mapear las coordenadas que entrega el *Eye Tracker*, también en píxeles, sobre los objetos de una página que observó un usuario en cierto instante. Una vez mapeado al objeto correspondiente, se almacenará el tiempo que gastó cada usuario en cada objeto del sitio. Luego, este valor se transformará en un porcentaje de

permanencia del usuario en los objetos para finalmente promediar los tiempos de permanencia de todos los usuarios en los objetos.

2. Encuesta. Luego de realizar la medición con el *Eye Tracker*, se procederá a realizar la misma encuesta que aplicaron en [16] a los participantes del experimento para luego comparar ambas mediciones y contrastar lo que indican que vieron los usuarios, con lo que realmente vieron.

Usuarios de control Los campos sexo, edad, profesión, nivel académico, dominio de Internet y conocimiento del sitio se almacenarán para llevar control sobre la muestra seleccionada para el estudio, con el objetivo de que sea lo más representativa posible.

3.1.3. Comparación de Resultados

Las distintas versiones de la metodología para encontrar *website keyobjects* entregarán como resultados distintos conjuntos de objetos. Para poder comparar estos resultados, nuevamente se usará el conocimiento experto del encargado del sitio, quien establecerá si un objeto clasificado como *website keyobjects* realmente lo es. Una vez adquirido este conocimiento se compararán los conjuntos obtenidos por las metodologías mediante el indicador *precision*. Este indicador es usado para evaluar la predicción de un algoritmo de clasificación. Con estos resultados, se concluirá y se establecerá la validez de la hipótesis.

4. Aplicación del Experimento y Análisis de Resultados

4.1. Entorno

El sitio elegido: Los experimentos se desarrollaron sobre el sitio *www.mbauchile.cl*, perteneciente a la dirección del Magíster en gestión y dirección de empresas del Departamento de Ingeniería Industrial de la Universidad de Chile. Está compuesto por 124 páginas y 163 objetos distintos que aparecen en 2.047 veces en el sitio. Con esto, el número promedio de objetos por página es 12,55. Con respecto a las visitas, cada mes acceden 4.158 personas distintas al sitio, se contabilizan 6.111 sesiones y se ven 26.589 páginas.

Los usuarios entrevistados: Antes de seleccionar a los usuarios de control que participarían en el experimento, se solicitó al experto del negocio información sobre el mercado objetivo del sitio en estudio con el objetivo de

elegir una muestra representativa de los individuos que visitan el sitio. Teniendo en cuenta esta información, se seleccionaron 33 personas. De estas, 16 son de sexo masculino y 17 femenino, mientras que su promedio de edad es de 24,3 años. Con respecto al conocimiento y uso de la Web, 15 de ellos se declararon expertos al navegar por la Web, 12 consideran que su conocimiento es regular y sólo 6 de ellos se declararon usuarios básicos.

El Eye Tracker Utilizado: Para el desarrollo de los experimentos se contó con herramientas de Software y Hardware de *Eye Tracking*. Acerca del hardware, se usó el *Eye Tracker Tobii T120*, que consiste en un monitor de 17 pulgadas al que se le incorpora dos emisores infrarrojos y un sensor de luminosidad. Este hardware posee una resolución temporal de 120 Hz y tiene un margen de error de 0,5°. Con respecto al software, se usó el programa *Tobii Studio Enterprise Edition*, solución que permite mapear fácilmente lo que se muestra en el monitor, con el lugar del mismo que los usuarios observan.

4.2. Captura de Datos

Para conocer la composición de el sitio <http://www.mbauchile.cl> se implementó un *crawler* ocupando la librería de Python *Beautiful Soup* que generó la lista de páginas que componen el sitio. Luego se procedió a identificar los objetos del sitio considerando dos criterios: diferencias en el contenido (Conceptos) y la separación espacial entre objetos de cada página.

Para conocer cuáles eran las coordenadas de los objetos dentro de las páginas se empleó la librería *Python Imaging Library (PIL)*, que permite trabajar con imágenes sobre el interprete de Python. Se implementó un script, que a partir de las imágenes de las páginas del sitio, generaba las coordenadas de los distintos objetos pertenecientes a cada página. Una vez definidos los objetos fueron generados los conceptos que describen su contenido. Se procedió de forma análoga a [16], es decir, cada concepto fue creado de forma manual. Con respecto a las visitas del sitio, se logró recuperar el *weblog* desde el servidor donde se aloja el sitio en conjunto con el administrador de sistemas, para el mes de Agosto de 2011, en el que 3.031 usuarios visitaron el sitio, en 5.480 sesiones.

Para medir el interés de los usuarios en los objetos web, se midió y estimó su tiempo de permanencia en ellos. Esto se realizó de dos formas: mediante el uso de un *Eye Tracker* y mediante la aplicación de la encuesta empleada por [16].

Eye Tracker. A 15 de los usuarios de control, se les planteó la siguiente situación: “Usted tiene la inquietud de postular a un programa MBA, pero aún no toma una decisión definitiva, por lo que su primer paso será informarse. En

la búsqueda de información ha llegado al sitio <http://mbauchile.cl>, que provee datos relevantes sobre el MBA que dicta la Universidad de Chile. Partiendo desde el *home* del sitio, navegue libremente hasta que pueda tomar una decisión o decida realizar un nuevo paso.” De esta forma, se buscó emular la navegación típica de los usuarios en sus casas. Los restantes 18 individuos no navegaron libremente, sino que se les instruyó a mirar las páginas que se les presentaran, sin seguir ningún enlace. A cada uno de ellos se les presentaron páginas semi-aleatorias del sitio. Los usuarios podían pasar a la página siguiente cuando lo estimaran conveniente, pero si pasaban más de un minuto en una página, automáticamente se le redireccionó a la página posterior. El número de páginas presentadas a los usuarios no fue superior a 30.

Encuesta. Luego de ser enfrentados al *Eye Tracker*, independiente de la forma en que se capturaron sus movimientos oculares, se solicitó a los sujetos responder una encuesta en la que por cada página vista, indicaran los objetos que más capturaron su atención. Para medir el interés prestado por los usuarios, se les indicó repartir 10 puntos, como ellos quisieran, sobre los objetos de cada página, teniendo en cuenta que mientras más puntos tenía un objeto, más interés les prestó el usuario.

4.3. Selección, Limpieza y Transformación de Datos

Una vez que se identificaron los 163 objetos del sitio, el experto del negocio validó esta separación, pero también agrupó, disgregó y eliminó algunos objetos pre-seleccionados.

Eye Tracker El primer paso para transformar estos datos fue agruparlos según el punto observado (en una vecindad pequeña), aumentando el tiempo de duración de la fijación según la cantidad de registros agrupados. Un resultado estadístico conocido es el rango de duración de las fijaciones, que están entre los 150 y 600 milisegundos (*ms*) [4], por lo que la cota inferior para que el cerebro comprenda lo que se está observando es de 150 *ms*. Como se señaló anteriormente, el *eye tracker* empleado tiene una resolución de 120 Hz, por lo que captura información cada 8 *ms*. Se logró una identificación de objetos superior al 86 % lo cual es considerado como excelente.

La encuesta. Los resultados de la encuesta correspondieron a archivos donde se registró la página, el objeto, y los puntos de interés que le asignó el usuario. Estos datos fueron transformados, de modo que los 10 puntos que asignó cada usuario, correspondiera al 100 % del interés del usuario en dicho estímulo. A modo de ejemplo, si una persona asignó 5 puntos al objeto X en la página Y se asumió que el interés del usuario sobre el objeto X fue del 50 % en la página Y. Análogamente al caso anterior, se asignó 0 al interés del usuario en

los objetos que no obtuvieron puntos. Luego estos resultados se promediaron y almacenaron.

A partir de peticiones web recuperadas, se procedió a ejecutar el proceso de sesionización. Se obtuvo como resultado un conjunto de sesiones, en las que cada una de estas contenía una lista de páginas, asociadas a la cantidad de tiempo que permanecieron en estas. Luego, para cada registro se reemplazó la página por los objetos que la conformaban y se ponderó el tiempo de permanencia en la página por el porcentaje de permanencia en el objeto. Además se consideraron con especial cuidado los objetos que aparecieron en más de una página, pues para una sesión, el mismo objeto pudo haber sido visto dos veces. En este caso, se sumaron los tiempos correspondiente a ambas páginas, pues se le prestó atención dos veces. A continuación, se procedió a seleccionar los n objetos más importantes de cada sesión. El criterio de selección de los objetos más importantes fue el tiempo de permanencia en ellos. La determinación de n se realizó de forma análoga a cómo se procedió en [16].

4.4. Aplicación de las Técnicas de Web Mining

La metodología para encontrar *Website Keyobject* [16] agrupa los vectores de comportamiento de los usuarios mediante tres técnicas: *Self Organizing Feature Maps*, *K-means* y *Association Rules*. Los resultados de estas técnicas son conjuntos de vectores donde sus elementos eran parecidos entre sí, pero distintos al tomar elementos de conjuntos diferentes. El criterio que usaron para determinar si un objeto era un *Website Keyobject* fue seleccionar los objetos que aparecieron más veces en los clusters arrojados por los tres algoritmos.

Self Organizing Feature Maps (SOFM) es un modelo de red neuronal que procesa una base de datos resultando en un mapa (usualmente bidimensional) donde casos similares se mapean en regiones cercanas. Ocupa un modelo de aprendizaje no supervisado, por lo que difiere de las redes neuronales comunes. Al igual que en el trabajo desarrollado por [16] se empleará una red de forma toroidal.

En este algoritmo cada neurona será definida como un IOV (Ver ecuación 2). La forma de actualizar los IOV, mientras se produce el entrenamiento, será modificada de modo que compare los vectores de objetos importantes presentes en cada una de las sesiones. Para cada IOV, se debe encontrar la neurona más parecida a este y actualizar los pesos de la red en función de las distancias calculadas. Este proceso se repetirá hasta que los cambios en los pesos de la red sean menores a un ϵ .

K-means Es probablemente el método más conocido para realizar *clustering*. Consiste en partir con K IOVs aleatorios denominados centros. Para los

IOV restantes se calcula el centro más cercano a estos y se enlazan generando subconjuntos de IOVs. Luego se calcula el centro de masas de cada subconjunto y se vuelve a iterar hasta que la cantidad de cambios en los subconjuntos sea baja, es decir, cuando los subconjuntos converjan. La debilidad de este método es el valor inicial K que es desconocido. En este caso particular, para determinar el centro de masas de los subconjuntos mencionados, se calculará el objeto cuya distancia promedio a todos los demás objetos en el subconjunto es la menor.

Association Rules Es una técnica que permite encontrar relaciones entre distintos atributos de un conjunto de datos. De integrar los identificadores de las sesiones a los datos, no se encontraría ningún tipo de regla, ya que para todos los ejemplos el identificador es distinto. Como entrada para este algoritmo se implementó una pequeña transformación de datos para dejar los IOV como la lista de los objetos más vistos (Ecuación 4).

$$r = (o_1, \dots, o_n) \tag{4}$$

Los experimentos. Como primer experimento, se desarrolló la misma estrategia empleada en [16], donde fueron considerados como tiempos de permanencia en los objetos los resultados de la encuesta sobre las páginas del sitio. Por otro lado, el segundo experimento consistió en reemplazar en la metodología anterior los tiempos calculados a partir de los datos capturados por el *Eye Tracker*.

4.5. Resultados Obtenidos

Luego de realizar varias pruebas con respecto a la cantidad de neuronas de la red SOFM se encontró que la grilla de 12x12 entregó los mejores resultados, con ocho *clusters* para cada experimento. La salida de este algoritmo se modificó para que entregara listas de los identificadores de sesiones pertenecientes a cada cluster, con el objetivo de poder contar el número de apariciones de los objetos en los distintos *clusters*.

Con respecto a la técnica *K-means*, se ocupó como $K=8$, para poder comparar los resultados con los obtenidos en el SOFM. Por otro lado, para la ejecución de *Association Rules* se usó la plataforma *Weka*. Se consideró solo los objetos presentes en los IOVs que fueron transformados de acuerdo a la plataforma. Se utilizó el algoritmo *Apriori* y se solicitó que generara sólo 30 reglas con una confianza mínima de 0,9. Tanto el algoritmo, como la plataforma a ocupar fueron los mismos empleados por [16] cuando implementó esta metodología.

4.5.1. *Precision, Recall y F-measure*

Para comparar los resultados obtenidos por ambos experimentos se seleccionaron tres rangos, tomando 10, 20 y 30 *Website Keyobjects*.

La Tabla 1 muestra la precisión alcanzada en por experimentos para los tres rangos seleccionados. La Tabla 2 muestra el *Recall* y la Tabla 3 el indicador *F-measure*.

Experiment	10	20	30
First	70 %	72 %	68 %
Second	70 %	93 %	96 %

Tabla 1: *Precision*

Experiment	10	20	30
First	21 %	43 %	58 %
Second	21 %	61 %	77 %

Tabla 2: *Recall*

Experiment	10	20	30
First	32.3 %	53.8 %	61.6 %
Second	32.3 %	73.7 %	85.5 %

Tabla 3: *F-measure*

Se puede notar un aumento (15 % al 20 %) en la precisión al comparar ambos experimentos, con lo que se valida que esta tecnología resulta útil para medir el interés de los usuarios. Por otro lado, se puede notar que cuando se eligen 30 *Website Keyobjects* la precisión obtenida no aumenta significativamente. Esto ocurre pues a partir de 30 objetos se comienza a designar otros que no son relevantes. En este punto, si consideramos como *threshold* $n = 24$, se obtiene que la precisión del segundo experimento es 94 %.

Con respecto a *Recall*, cuando se toman 10 o 20 *Website Keyobjects* este indicador es considerablemente pequeño. Esto es esperable pues se están dejando de considerar *Keyobjects*. Sin embargo, en el segundo experimento se obtienen mejores resultados para todos los rangos observados.

Finalmente, *F-measure* es una medida que mezcla ambos indicadores *precision* y *recall*, por lo que en el segundo experimento se logren mejores resultados era esperable.

5. Conclusiones

En la presente investigación se probó que usar un *Eye Tracker* para medir la cantidad de tiempo que los usuarios gastan observando los distintos objetos de una página web en lugar de realizar una encuesta para estimar estos valores mejora la precisión a la hora de encontrar los *Website Keyobjects* de un sitio.

Para lograr este resultado se realizó un amplio estudio acerca de los datos originados en la Web, los modelos matemáticos que se usan para describir el comportamiento de los usuarios en la Web y las herramientas existentes de seguimiento ocular.

Una de las limitantes que tiene el seguimiento ocular es que sólo determina qué es lo que una persona observa. Esto es insuficiente cuando se busca calificar lo que una persona mira, en otras palabras, con estas herramientas no se puede determinar si lo que se observa me gusta o me desagrada. Por esto, los resultados entregados por el *Eye Tracker* deben ser considerados como el módulo (siempre positivo) del interés de una persona.

Como trabajo futuro, se plantea considerar que sienten o piensan los usuarios mientras miran un objeto Web. Esto se puede lograr al usar elementos de neurotecnología, como electrodos que midan que zona del cerebro se excita más cuando una persona navega por un sitio Web. *Emotiv.com* es una empresa que fabrica este tipo de aparatos (sólo para este tipo de investigaciones) a bajo costo.

Agradecimientos: El presente trabajo fue realizado gracias al aporte del proyecto Fondef D10I1198 y del Instituto Milenio de Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16).

Referencias

- [1] A. Bulling, J.A. Ward, H. Gellersen, and G. Tröster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2011.
- [2] J. J. Cerrolaza, A. Villanueva, and R. Cabeza. Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In *Proceedings of the 2008 symposium on Eye tracking research & #38; applications*, ETRA '08, pages 259–266, New York, NY, USA, 2008. ACM.

- [3] G. Chang, M. Healey, J. McHugh, and J. Wang. *Mining the World Wide Web*. Kluwer, 2001.
- [4] A. T. Duchowski. *Eye tracking methodology: Theory and practice*. Springer Verlag, 2003.
- [5] L. E. Dujovne and J. D. Velásquez. Design and implementation of a methodology for identifying website keyobjects. In *Proceedings of the 13th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems: Part I*, KES '09, pages 301–308, Berlin, Heidelberg, 2009. Springer-Verlag.
- [6] D. Hoffman and T. Novak. A new marketing paradigm for electronic commerce. *Information Society*, 13(1):43–54, 1997.
- [7] C. Jay, D. Lunn, E. Michailidou, S. Harper, and Y. Yesilada. End user evaluations. In S. Harper and Y. Yesilada, editors, *Web Accessibility*, pages 107–126. Springer, London, UK, 2008.
- [8] A. E. Kaufman, A. Bandopadhyay, and B. D. Shaviv. An eye tracking computer user interface. In *Proceedings., IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, pages 120–121. IEEE, 1993.
- [9] R. Kosala and H. Blockeel. Web mining research: A survey. *ACM SIGKDD Explorations Newsletter*, 2(1):1–15, 2000.
- [10] S.J. Lee, W.N. Lee, and P.A. Stout. A comparison of objective characteristics and user perception of web sites. *Journal of interactive advertising*, 4(2):61–75, 2004.
- [11] J. Nielsen. *Usability Engineering*. Academic Press, San Diego, CA, USA, 2004.
- [12] J. Nielsen and K. Pernice. *Eyetracking web usability*. New Riders Pub, 2009.
- [13] T.P. Novak, D.L. Hoffman, and Y.F. Yung. Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19(1):22–42, 2000.
- [14] A. Poole and L. J. Ball. Eye tracking in human-computer interaction and usability research: current status and future prospects. *Encyclopedia of human computer interaction*, pages 211–219, 2005.

- [15] J. D. Velásquez. Web site keywords: A methodology for improving gradually the web site text content. *Intelligent Data Analysis*, 16(2):327–348, 2011.
- [16] J. D. Velásquez, L. E. Dujovne, and G. L’Huillier. Extracting significant website key objects: A semantic web mining approach. *Engineering Applications of Artificial Intelligence*, 24(8):1532–1541, 2011.
- [17] J. D. Velásquez and L. C. Jain. *Advanced Techniques in Web Intelligence - Part 1*. Springer Publishing Company, Incorporated, 1st edition, 2010.
- [18] J. D. Velásquez and V. Palade. *Adaptive Web Sites*. IOS Press, 2008.
- [19] J. D. Velásquez, S. Ríos, A. Bassi, H. Yasuda, and T. Aoki. Towards the identification of keywords in the web site text content: A methodological approach. *International Journal on Web Information Systems*, 1(1):11–15, 2005.