



# Combining eye-tracking technologies with web usage mining for identifying Website Keyobjects



Juan D. Velásquez\*

Department of Industrial Engineering, Universidad de Chile, Av. República 701, P.O. Box 8370439, Santiago, Chile

## ARTICLE INFO

### Article history:

Received 3 August 2012  
Received in revised form  
4 January 2013  
Accepted 14 January 2013  
Available online 4 February 2013

### Keywords:

Web mining  
Eye tracking  
Website Keyobjects

## ABSTRACT

This paper introduces a novel approach for collecting and processing data originated by web user ocular movements on a web page, which are captured by using an eye-tracking tool. These data allow knowing the exact web user's eye position on a computer screen, and by combining them with the sequence of web page visits registered in the web log, significant insights about his/her behavior within a website can be extracted.

With this approach, we can improve the effectiveness of the current methodology for identifying the most important web objects from the web user's point of view, also called Website Keyobjects. It takes as input the website's logs, the pages that compose it and the interest of users in the web objects of each page, which is quantified by means of a survey. Subsequently, the data are transformed and preprocessed before finally applying web mining algorithms that allow the extraction of the Website Keyobjects.

With the utilization of the eye-tracking technology, we can eliminate the survey by using a more precise and objective tool to achieve an improvement in the classification of the Website Keyobjects. It was concluded that eye-tracking technology is useful and accurate when it comes to knowing what a user looks at and therefore, what attracts their attention the most. Finally, it was established that there is an improvement between 15% and 20% when using the information generated by the eye tracker.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

In recent years, there has been a new research area called Web Mining, which studies different ways of extracting information from data generated on the Web. With this knowledge, it is possible to develop techniques and algorithms to attract and retain users on a website. Specifically, this area applies data mining techniques to data originated on the Web with the aim of obtaining valuable information to continuously improve a website in terms of design and content, among other aspects.

An interesting result that has been achieved in this field are the Website Keywords (Velásquez et al., 2005), defined as a word or set of words used by users in their information search process, and which characterizes the content of a website or page. After finding them, the sites can be redesigned according to the needs and requirements of its users and, thus, be in the vanguard. Although identifying the Website Keywords of a site helps to know the preferences of users, the methodology that discovers them only focuses on the textual content, leaving out the analysis

of the multimedia content of websites. For this reason, in Dujovne and Velásquez (2009) the methodology was extended and integrated with both the textual content and the multimedia content in their analysis.

In Dujovne and Velásquez (2009) web objects are defined as “any structured group of words or a multimedia resource that is present on a web page that has metadata which describe its content”. Also, they characterized a Website Keyobject as a web object that captures the attention of the users and that characterizes the contents of a website (Velásquez et al., 2011). From the above definitions it is possible to deduce that every website consists of a set of web objects and that the set of Website Keyobjects it has is a subset of the former.

Also in Velásquez et al. (2011) the designed methodology allowed the identification of the Website Keyobjects of a site. These, like the Website Keywords Velásquez (2011), give guidelines for the websites to be redesigned according to user requirements. This methodology requires knowledge of the time spent by users on the web objects, i.e., how long a user spends looking at each web object. To determine the permanence time they propose two steps, sessionization and the application of a survey.

*A. Sessionization:* It is the process that reconstructs user sessions. A session is the sequence of web pages visited by a user while browsing a site. The necessary data for a session

\* Tel.: +562 2978 4834.  
E-mail address: [jvelasqu@dii.uchile.cl](mailto:jvelasqu@dii.uchile.cl)  
URL: <http://wi.dii.uchile.cl/>

reconstruction is registered in the web log files. Through this process it is possible to determine, for example, the permanence time of a user on each page, among other things.

B. *Application of a survey*: It refers to the implementation of a survey on a control group, wherein web objects were sorted according to their importance within each page of the website.

By mixing the two processes, the time spent by users on the web object is estimated. However, thanks to the eye-tracking technology it is possible to dispense with this survey. Eye-tracking technology allows knowing what a person looks at as a function of time (Ali-Hasan et al., 2008). By applying this technology to users browsing a site it is possible to measure the time spent on each web object. In this sense, if it was possible to quantify the permanence time of a control group of users on web objects, an improvement to the methodology developed in Velásquez et al. (2011) could be made, i.e., determining the Website Keyobjects more accurately.

Section 2 describes the related work with the paper proposal, explaining concepts like web mining, the nature of the data originated on the Web and how eye-tracking techniques have facilitated the application of its usability for analyzing computer-based systems. Section 3 describes the elemental issues for understanding how the eye-tracking techniques work. Section 4 shows a methodology for identifying Website Keyobjects, the main problem to be tackled in this paper. In Section 5, the whole set of experiments performed for demonstrating the practical utility of eye-tracking tools in the scope of web mining is shown. Finally, Section 6 presents the paper's conclusions and future work.

## 2. Related work

In this section a conceptual framework about the traditional web mining process, which uses data originated on the Web, and the possibility of using other kinds of data, for instance those generated by the application of eye-tracking tools, will be introduced.

### 2.1. Data originated on the web

The data that are originated on the Web are classified into three types (Velásquez and Jain, 2010): content, structure and usage.

1. *Content*: It refers to the objects on Web pages, including text, images, sounds and videos, in other words, all that can be seen on a page. Thus, the text can be presented in a semi-structured, highly structured or unstructured way. On the other hand, the multimedia content requires metadata that describe it; however, it is uncommon to find these descriptions.
2. *Structure*: They are the links between pages. Usually when there is a link between two pages, these are related by their content. If a set of pages are linked together, a common information community is created.
3. *Usage*: It is the data generated by users in their navigation process, as Web servers store each request made by users in a file called a web log.

### 2.2. Web mining

Web mining is the application of data mining to data originated on the Web (Chang et al., 2001; Velásquez and Jain, 2010). It is conceived as a product of the intersection of several research

areas such as databases, information retrieval, artificial intelligence, especially the sub-areas of machine learning and language processing (Etzioni, 1996). Research in this field is experiencing significant growth because of the large amount of data available for analysis (Kosala and Blockeel, 2000). This is no small task, considering that the Web is a large collection of heterogeneous, declassified, distributed, time-varying, semi-structured high-dimensional data (Pal et al., 2002).

Since the data originating from the Web can be classified into three categories, it is natural that Web mining also branches into these categories. Their nature differs in the way that they are set up as different problems. The sub-areas are (Velásquez and Palade, 2008; Kosala and Blockeel, 2000; Velásquez and Jain, 2010):

1. *Web content mining* (WCM): It is targeted to discover useful information from Web documents. WCM is not only limited to analyzing the text of Web pages, but also includes other types of documents such as images and videos. However, the analysis of this data type, called multimedia data mining, does not receive as much attention as text analysis.
2. *Web structure mining* (WSM): It is the sub-area which studies the links on Web pages. The pages and links are modeled as the nodes and arcs of a directed graph, respectively. The arc starts on the node that represents the page that has the link and ends at the node representing the page being targeted.
3. *Web usage mining* (WUM): It is focused on the application of data mining techniques to discover useful patterns that can predict user behavior while interacting on the Web. To discover these patterns, the browsing sessions, understood as the sequence of pages that a user visits while surfing the web, are analyzed.

### 2.3. Web user perception

One important question when a web page is designed is about what the web user will perceive and understand concerning the content and structure shown (Lee et al., 2004). Perception is a function of thinking, i.e., how we recognize and interpret an object by using our senses for encoding, storing and integrating information related to some previous knowledge.

Understanding web user perception is important for creating a better website design, which will finally attract and retain the users. However, it is not a simple task, because this perception is related to the web user's knowledge background, i.e., his/her personal experience (Novak et al., 2000).

According to the conceptual model of flow proposed by Hoffman and Novak (1997) during the web user's navigation, the website structure is perceived as a cognitive experienced state, which is determined by "high levels of skill and control; high levels of challenge and arousal; focused attention; and is enhanced by interactivity and telepresence". Perhaps the unique factor that can be influenced directly by construction of the website is the web user's focused attention. However, it could be a double-edged sword, i.e. attention can be captured because the website content is attractive or because the content is unpleasant, in which case the web user will never again visit the site.

In any case, to understand what the web user is looking for, or at least to know what elements in the website capture his/her attention it is useful to improve the site structure and content. In that regard, many approaches have been addressed, usually from the information and knowledge extracted from web data (Kosala and Blockeel, 2000; Pal et al., 2002) and also from the usability theory for creating web-based systems (Nielsen and Pernice, 2009).

## 2.4. Usability approaches

According with Nielsen (2004), usability is “a quality attribute that assesses how easy user interfaces are to use” and it has five quality components to be analyzed:

- Learnability: Is it easy for the users to perform basic tasks the first time they see the application?
- Efficiency: How quickly can the user perform tasks once they have learned the basics about the application’s design?
- Memorability: If the users visit the application again after a period, is it easy for them to remember how to use the basic operations?
- Errors: Are the users making several errors? How bad are these errors?
- Satisfaction: Are the users happy with the application design?

Of course to analyze the usability of any application by using the above components, it is necessary to implement qualitative and quantitative measures about the user experience. In that sense, a number of evaluation methods have been created (Jay et al., 2008):

- Questionnaires: A set of questions created to obtain the user’s opinions about an application.
- Observation: It is the direct observation of the user activities in an application, where an evaluator is taking notes about how to improve it.
- Interview: Usually, the interviews are a follow up after an observation process and these consist of a conversation with the users for obtaining their impressions about the application.
- Focus group: It is a kind of interview method where a selected group of persons are asked for a determinate theme.
- Think-aloud: The users are invited to verbalize their thoughts, feelings and opinions about the use of a particular application.
- Screen capture: It is implemented with a set of screenshots about the application, which are shown to the users and next a questionnaire is applied to get the users’ impressions and opinions.
- Eye-tracking: The user’s eye movements in an application are followed. These data are used to analyze the user behavior in an application.

Each method has advantages and disadvantages and usually for researching purposes, a combination of two or more of them are necessary, mainly because its application always has a subjective component. For instance, the first six methods need to apply some kind of survey for getting the user point of view. In the case of eye-tracking tools, more precise quantitative measure can be obtained because an independent computerized system is used for tracking the user’s eye position on the computer’s screen, and to use these data a certain indicator about what the user seeing is extracted, which can be used as a measure of the user interest in a web page. However, the eye position also can be influenced by external situations that could modify the user’s mood, altering the fixation of their eyes. In conclusion, the eye-tracking method is arising as one of the most objective in terms of the usability measure (Nielsen and Pernice, 2009).

## 2.5. Eye tracking

Eye tracking is a technique by which an individual’s eye movements are measured. Thus, a researcher can know what a person is looking at in each moment and the sequence in which their eyes

move from one place to another. Following the eye movements of people can help researchers of the human–computer interaction to understand the processing of visual information and the factors that may impact upon the usability of the interface. Thus, recordings of eye movements can provide a source of objective data for the evaluation of interfaces, which in turn may provide information to improve their design (Poole and Ball, 2005).

## 2.6. Eye movements

When looking at any scene, the eyes of a person move between points that capture their attention, which are able to recreate a brain image of the scene. While there are models of complex eye movements consisting of five steps (Duchowski, 2003), the typical model (which is enough for studies of eye tracking) consists of two concepts, fixation and saccades. Fixation is defined as the moment in which the eyes are fixed on an object and it is possible to appreciate it in detail, while the saccades correspond to rapid eye movements between two fixations (Nielsen and Pernice, 2009). It is important to note that while producing a saccade we are blind (we are not aware of what is between the two objects that caught our attention). However, our brain can interpret this “image sequence” as a continuum, and therefore, our perception seems more like a video than a sequence of images.

### 2.6.1. Visual attention

Visual attention is a phenomenon that has been studied for about a 100 years and still cannot be understood. Early studies were limited by technology, and corresponded only to observation and introspection. At present, this field is studied by different disciplines such as psychophysics, cognitive neuroscience and computer science, to name a few (Duchowski, 2003). Overall, human vision consists of two parts, a small central area with a very high resolution, called foveal vision, and the large majority of the visual field with a low resolution, called peripheral vision.

Usually, the fact of paying attention to regions of interest is related to eye movements (overt attention). However, we can also pay attention to peripheral objects without making such movements (covert attention) (Ali-Hasan et al., 2008).

On the other hand, there are two known ways in which attention is guided, bottom-up and top-down. The first, derived only from the visual scene, posits that regions of interest attract our attention strongly enough to not look at the rest of the scene (foveal vision). In turn, top-down is driven by cognitive factors such as knowledge, expectation and current goals. Under this model, people are more inclined to look around (peripheral vision); as an example, an individual who drives regularly is more likely to notice the gas stations while doing other activities than someone who does not drive (Frintrop et al., 2010).

Nowadays, it is not yet clear what it is that really captures our attention, nor how we respond to different stimuli. There is an evidence that we pay attention to spatial locations, features and objects. Most researchers believe that these theories are not mutually exclusive and moreover, that visual attention can be developed in each of these sub-areas. It is worth noting that humans can pay attention simultaneously to multiple regions of interest (a maximum of five) (Poole and Ball, 2005).

### 2.6.2. The Eye-mind hypothesis

Considering the theories described in the previous section, in Duchowski (2003) posited the following model:

1. Given a stimulus such as an image, the scene is seen largely in parallel through peripheral vision and therefore, at low

resolution. At this stage, the interesting features of the image can “appear”.

2. In this moment, attention is disconnected from the foveal vision (high resolution), but the eyes quickly move toward the first region of attraction.
3. Once the eyes are positioned, the fovea is aligned to the region of interest and attention is linked to perception, i.e., the user's attention has been captured and, therefore, it is possible to observe in high resolution.

In Nielsen and Pernice (2009) a simple model is proposed, which in its essence is the same. Their hypothesis states that “People are usually thinking about what they are looking at. Although they do not always understand what they see or are not fully focused on it; if they are observing something, then they are paying attention, especially when they are concentrated on a particular task”.

### 2.7. Eye tracking and perception on the web

How human beings allocate attention within different parts of a web page is an interesting issue to investigate. In fact, knowing which visual objects assist in or distract web user navigation from finding what is being looked for, could be the key to modifying website structure, content and web object characteristics, such as position, shape, etc. within a web page (Michailidou et al., 2008).

Eye-tracking systems have been used to investigate how sighted persons use and perceive the web object features on a website. This new knowledge is very important for web page designers, web masters, advertisers, etc., because many innovations can be improved for web page design, creation and visual representation (Buscher et al., 2009).

The correct free text distribution and other web object presentation on a web page can significantly improve the web user comprehension regarding what she/he reads and understands. In fact, web user behavior is not the same when faced by a text-only page versus an illustrated one. The correct content and distribution on a web page can facilitate both web user comprehension and the information searching task, by making the page more attractive (Buscher et al., 2010).

An open question is which regions within a web page the user recognizes from previously visited ones. Indeed, if that knowledge can be extracted, compact versions of the page could be created, assisting the web user during an exploration or search task.

To combine the knowledge extracted by analyzing web user perception by using an eye-tracking tool with the knowledge extracted from web user behavior by applying web mining algorithms in data originated on the website, is the objective of this paper. Both techniques are complementary and useful for the final goal: to understand what the web user is looking for on a website (Velásquez, 2012).

## 3. Eye-tracking techniques

To understand how the eye-tracking technology can be used in a web mining process, in this section we will review the basis concepts about its operation (Holmqvist et al., 2011).

### 3.1. Electrooculography

In the 1950s, *electrooculography* (Kaufman et al., 1993) was the most-used eye-tracking technique. It was based on measuring the difference of electrical potential of the skin through the use of electrodes located around the eyes. This is possible since the cornea is kept some 10ths of mV more positive than the retina,

which produces the difference of potential that is measured, and which varies according to the movement of the eyes. This technique measures the relative position of the eyes with respect to the head; therefore, it is not appropriate for calculating the point of attention (Bulling et al., 2011).

### 3.2. Scleral contact lenses

This technique consists of attaching a mechanical or optical reference to a contact lens to be used directly on the eyes. It is necessary that the contact lens be particularly large, so that it extends over the cornea and sclera, since this reduces the possibility of it moving across the eye (Duchowski, 2003). Different types of references on contact lenses have been used, the most common being a small coil, which can be located externally by applying an electromagnetic field (Roberts et al., 2008; Van der Geest and Frens, 2002). While it is one of the most precise techniques for measuring eye movements, it is also the most invasive, and causes discomfort when being used.

### 3.3. Reflection of the cornea and pupil center based on video

This is the eye-tracking technique which is most commonly used nowadays. It consists of a standard desktop computer with an infrared camera mounted under a monitor, with image-processing software to locate and identify the reflection of the cornea and the center of the pupil. With these characteristics, it is possible to dissociate eye movements from the position of the head, which makes it possible to calculate the attention focus of users (Poole and Ball, 2005).

In operation, an infrared light from an LED is directed toward the user to create noticeable reflections of the features of the eyes, thus making tracking easier (infrared light is used to avoid dazzling the user). Light enters the retina and a lot of it is reflected, so that the pupil appears as a bright, well-defined disc (the effect known as bright pupil). The corneal reflection is also generated by infrared light, appearing as a small but strong shine (Poole and Ball, 2005). Fig. 1 shows the bright pupil effect and the corneal reflection.

A digital camera captures and records the reflection light and the collected data are analyzed to extract the eye rotation through variations in the corneal reflection, also called “glint” or the first Purkinje reflection (Morimoto et al., 2000).

In Fig. 2  $I_1$  is the reflection from the cornea's external surface;  $I_2$  is the reflection from the cornea's internal surface;  $I_3$  is the reflection from the anterior surface of the crystalline lens and finally  $I_4$  is the reflection from the other surface of the crystalline lens.

Once the image processing software has analyzed the  $I_1, \dots, I_4$  corneal reflex, the vector that results from them is measured, and by using some trigonometric calculations, the point of attention can be found.

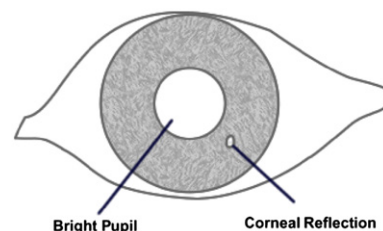


Fig. 1. Corneal reflection and bright pupil.

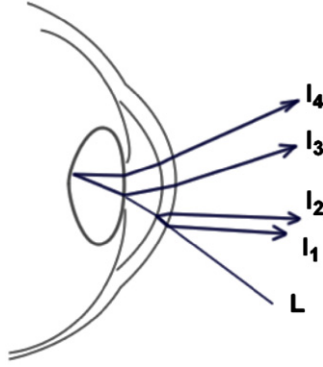


Fig. 2. Pukinje images with a direct light.

#### 4. Methodology to find Website Keyobjects

In Dujovne and Velásquez (2009) was defined a web object as “a structured group of words or multimedia content that is present on a Web page, and which has metadata describing its content.” In the previous definition, the metadata are fundamental as they are the basis of the information to construct the vector representing the content of the page. In addition, two multimedia files can be compared by their metadata, a problem which can be dealt with more easily than comparing files directly, because it only compares text. Along with this, they defined the Website Keyobject as “one or a group of web objects that attract user attention and which characterize the contents of a web page or site”. They provide knowledge about the content and format of most interest to users of a website, therefore finding them may be useful to improve the site both in presentation and content.

The sub-processes that allow for the identification of the Website Keyobject are (Velásquez et al., 2011):

1. *Sessionization*: the objective of this stage is to complete the sequence of pages visited by the different web users of the site. In addition, this sequence must have associated with it the residence times of each web user per page visited during his/her session.
2. *Addition of metadata*: the first step of this stage is to identify the objects that make up the pages of the site. Once identified, the concepts that describe each of the objects must be defined. Then, this information must be stored in a database. The gathering of these data should be done together with the web master to ensure that the concepts accurately reflect the content of objects.
3. *Web object*: By using the metadata as described above, web objects can be made as a simple ontology based on the work by the MPEG to include metadata to videos. In this sense an XML document will be associated to each web object present in a web page. Despite the complex semantic analysis of multimedia, metadata is used to define the web object within it. In this sense, a set of  $N$  objects is defined by  $\mathbf{o} = \{o_1, \dots, o_N\}$ , and each object is defined by a set of  $M$  concepts,  $\mathbf{c} = \{c_1, \dots, c_M\}$ . According to Velásquez et al. (2011), a Website Keyobjects' core ontology is represented by the tuple

$$\mathcal{O} := (\mathcal{C}, \leq_c, \sigma, \leq_{\mathcal{R}}, \mathcal{A})$$

with the following properties:

- A set of concepts defined by the set  $\mathcal{C}$  as,

$$\mathcal{C} = \{\text{Website } (s), \text{ Webpages } (\mathbf{w}), \text{ WebObjects } (\mathbf{o}), \text{ WebConcepts } (\mathbf{c})\}$$

- The concept hierarchy (or organizational part-of hierarchy)  $\leq_c = \{s, w_i, o_j, c_k\}, \forall w_i \in \mathbf{w}, o_j \in \mathbf{o}, c_k \in \mathbf{c}$ . This hierarchy can be interpreted that a given Website ( $s$ ) is represented by Webpages ( $\mathbf{w}$ ), each Webpage ( $w_i \in \mathbf{w}$ ) can be represented as the composition of WebObjects ( $\mathbf{o}$ ), and each WebObject ( $o_j$ ) is characterized by a composition of WebConcepts ( $\mathbf{c}$ ).
- Set of part-of relations  $\mathcal{R} = \{r_1, r_2, r_3\}$  where  $r_1 = \text{site-has-page}$ ,  $r_2 = \text{page-has-object}$ , and  $r_3 = \text{object-has-WebConcepts}$ .
- The signature  $\sigma$  is defined for relations  $r_i \in \mathcal{R}$  according to  $\sigma(r_1) = (s, w_i), \sigma(r_2) = (w_i, o_j), \sigma(r_3) = (o_j, c_k)$ .
- Relations' hierarchy ( $\leq_{\mathcal{R}}$ ) in this case will be represented by the identity, given the flat relation between concepts.
- Logical axioms are represented by the empty set  $\mathcal{A} = \emptyset$ .

4. *Web object similarity*: To compare two web objects a similarity measure  $do : |\mathbf{o}| \times |\mathbf{o}| \rightarrow [0, 1]$  was defined by using a given distance measure (e.g., an edit distance).

Given two objects  $o_i$  and  $o_j$  such that  $|o_i| = E$  and  $|o_j| = F$ , where  $E, F \geq 0 \wedge E \leq F$ , and given  $o_i \rightarrow c_k^i, k \in \{1, \dots, M\}$  as the  $k$ th concept of the object  $o_i$ . We need to find a representation for which all WebObjects can be compared. A compare function that could be used, but not limited to, is based on the edit distance between the pairwise alignment between WebConcepts of objects  $o_i$  and  $o_j, \forall o_i, o_j \in \mathbf{o}, i \neq j$ .

Once all the concepts are paired, the objects concepts are ordered in such a way that every concept is in the same relative position in relation to each object. Then a string that consists of a symbol representing each concept is created to represent each object. This string has the structure shown in the following equation:

$$o = \text{WebConcept}_1, \dots, \text{WebConcept}_M \Rightarrow o = c_1, \dots, c_M, \quad (1)$$

where  $c_k \in \{o_i \rightarrow c\}$  represents a set of all concepts in object  $o_i$ . Then, given two webobjects  $o_i, o_j$  the web object similarity can implemented by

$$do(o_i, o_j) = 1 - \frac{L(o_i, o_j)}{\max(|o_i|, |o_j|)}, \quad (2)$$

where  $L$  is the Leveshtian/Edit distance or any similarity that allow to compare how different/similar are two string sequences, for instance:

$$L(o_i, o_j) = \begin{cases} |o_i|, & F = 0, \\ |o_j|, & E = 0, \\ \min\{L(o_i \rightarrow c_{E-1}^i, o_j) + 1, \\ L(o_i, o_j \rightarrow c_{F-1}^j) + 1, \\ L(o_i \rightarrow c_{E-1}^i, o_j \rightarrow c_{F-1}^j) + z(o_i, o_j)\} & \text{else} \end{cases} \quad (3)$$

where

$$z(i, j) = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{otherwise.} \end{cases} \quad (4)$$

5. *Permanence time on objects*: after defining the objects, it is necessary to estimate how long a web user is seeing a particular object on a web page. In that sense, two possibilities are explored in this paper: to apply a survey about the interest in the web object and to apply an eye-tracking tool. Both approaches are applied on a control group of users. In the first one, a survey is performed on each user asking to distribute a total of 10 points of interest among all objects on a page. With these data, they estimated the percentage spent by each user on the objects on each page. In the second one,

an eye-tracking tool is used to follow what the user is seeing in a web page. Because each web object is bounded by a group of pixels, we can know how long time a web user was visualizing a part of the page.

6. *Vector of user behavior*: finally, for each session identified, the  $n$  objects that captured more attention from the user were selected, thus defining the Important Object Vector (IOV) according to the following equation:

$$v = [(o_1, t_1) \dots (o_n, t_n)]. \quad (5)$$

7. *Clustering algorithms*: once all the cleaning and transformation of data has been done, the following step is the processing of clustering algorithms on user sessions, represented by Important Object Vector. To run these algorithms, it is of crucial importance to define a measure of distance, or similarity, between these vectors.

8. *Similarity measures for sessions*: The similarity defined for comparing two IOV  $\alpha, \beta$  is calculated by using the following equation:

$$st(\alpha, \beta) = \frac{1}{i} * \left( \sum_{k=1}^i \min \left( \frac{\tau_k^\alpha}{\tau_k^\beta}, \frac{\tau_k^\beta}{\tau_k^\alpha} \right) * do(o_k^\alpha, o_k^\beta) \right), \quad (6)$$

where  $do: |\mathbf{o}| \times |\mathbf{o}| \rightarrow [0,1]$  is the similarity between two WebObjects (e.g., an edit distance) and  $\tau_k^\alpha, \tau_k^\beta$  the time spent for the user seeing a WebObject.

## 5. Experiment and results

On a site, the methodology designed by Velásquez et al. (2011) hereinafter the original methodology will be executed, and also the modified methodology will be executed, to which the permanence times calculated with an eye tracker will be incorporated.

### 5.1. Experiment

To capture the interest of users in the different objects of the pages, two procedures will be used, using an eye tracker and the application of a survey. In this sense, experiments were conducted on the site [www.mbauchile.cl](http://www.mbauchile.cl) belonging to the Master in Business Administration and Management program of the Department of Industrial Engineering of the University of Chile. Fig. 3 shows the site, which consists of 124 pages and 163 different objects that appear on the site 2047 times. Thus, the average number of objects per page is 12.55. With regard to visits, each month 4158 different people access the site, 6111 sessions are recorded and 26,589 pages are seen.

On the other hand, before selecting the controlled users who would participate in the experiment, the business expert was asked for information on the target of the site under study with the objective of choosing a representative sample of individuals who visit the site. Considering this information, 33 people were selected. Of these, 16 were male and 17 female, and their average age was 24.3 years. Regarding the knowledge and use of the Web, 15 of them defined themselves as expert web users, 12 considered that their knowledge is average and only six of them defined themselves as basic users.

Finally, for the development of the experiments eye tracking software and hardware tools were used. Regarding the hardware, the Tobii T120 Eye Tracker<sup>1</sup> was used, which consists of a 17-in. monitor to which two infrared emitters and a light sensor are

incorporated. This hardware has a time resolution of 120 Hz and has a margin of error of 0.5. With respect to software, the Tobii Studio Enterprise Edition was used, a solution which easily allows mapping what is displayed on the monitor with the same location that users observe.

#### 5.1.1. Capturing of data

To discover the composition of the site <http://www.mbauchile.cl>, a crawler that generated the list of pages that make up the site was implemented. This crawler was developed using the Beautiful Soup Python library. Once the list of pages to be analyzed was captured, the following procedure was used in identifying the objects of the site. To separate each page into the objects that compose it, two criteria were considered, differences in the content (concepts) and the spatial separation between objects of each page. In order to know the coordinates of the objects within the pages, the Python Imaging Library (PIL) was used, which allows working with images on the Python interpreter. A script was used, which from the images of the pages of the site generated the coordinates of the various objects belonging to each page. After defining the objects, by using the methodology explained in Section 4, the pixels that compound the web objects are identified as shown in Fig. 4 and the concept that described them were generated.

Regarding the visits to the site, it was possible to recover the web log from the server that hosts the site together with the system administrator. It was not possible to recover all of the requests as the historic records were not stored. However, the requests corresponding to the month of August 2011 were obtained. During this month, 3031 different people visited the site, in 5480 sessions. In total 28,832 pages were viewed, there were 156,259 requests and a traffic of 3.20 GB was reached. In order to measure the interest of users in the web objects, their permanence time on each was measured and estimated. This was done in two ways, by using an eye tracker and by applying the following survey used by Velásquez et al. (2011).

First a list of web objects is identified and three questions were asked to the controlled users. The first one was “Which object was the most appealing to you within the whole Website?”, the second question was to make a top 20 list with all the objects of the site being the first the most appealing for them. Finally for each page which had two or more objects, 10 points must be awarded between all of them having the most points the object which was the most appealing to the user within a certain page. An example about how the survey works is shown in Table 1 which contain the answers given by one user. The survey was applied on web pages containing two or more objects.

The data contained in Table 1 were interpreted as follows: one page can contain two or more objects (until four), the column *Favorite* shows the favorite object and the regarding four columns the points given by the user to each object. For instance, the page ID 9 has the objects IDs 11 and 12. From these two objects, the favorite one for the user is the ID 11 because 7 points were given, where only 3 points were given to object ID 12.

Next step was to propose to 15 of the controlled users the following situation: “You have an intention to apply for an MBA program, but have not yet made a final decision, therefore your first step will be to get informed. In the search for information you have arrived at the website <http://www.mbauchile.cl>, which provides relevant data about the MBA program offered by the University of Chile. Starting from the home site, navigate freely until you can make a decision or decide to take a new step”. In this way, it was attempted to emulate the typical navigation of users at home. The remaining 18 individuals did not navigate freely, but were instructed to look at the pages shown without following any links.

<sup>1</sup> <http://www.tobii.com/en/eye-tracking-research/global/>



Fig. 3. A web page extracted from the MBA program website.

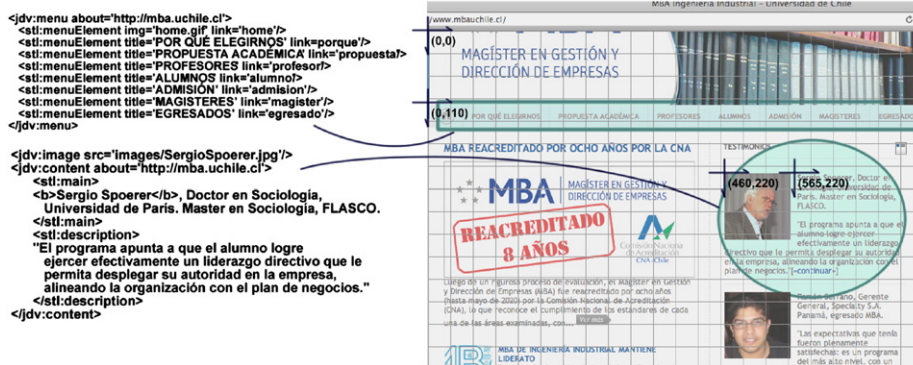


Fig. 4. Web objects identified in the MBA program, with position and metadata.

Table 1  
Calculating the importance of a Web object from the web user point of view.

Page ID	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	Favorite	Score(x <sub>1</sub> )	Score(x <sub>2</sub> )	Score(x <sub>3</sub> )	Score(x <sub>4</sub> )
2	33	35	34	–	33	7	2	1	–
9	11	12	–	–	15	7	3	–	–
20	4	5	–	–	4	5	5	–	–
21	23	24	–	–	23	7	3	–	–
24	29	30	31	–	31	2	2	6	–
28	25	26	–	–	25	8	2	–	–
31	28	27	–	–	28	6	4	–	–
32	19	22	21	20	19	5	3	1	1
39	38	38	–	–	38	4	5	1	–

Each of them was presented semi-random pages of the site. Users could move to the next page when deemed convenient, but if they spent more than a minute on a page, they were automatically redirected to the following page. The number of pages displayed to users was not higher than 30.

On the other hand, after facing the eye tracker, no matter how their eye movements were captured, subjects were asked to answer a survey in which they had to indicate the objects that most captured their attention. To measure the interest of the users, they had to distribute 10 points, as they chose, over the

objects on each page, considering that the more points an object had, the more interest he or she had.

### 5.1.2. Selection, data cleaning and transformation

Once the 163 objects on the site were identified, the business expert validated this separation, but also brought together, broke up and removed some pre-selected objects. Following this validation, as some objects changed, it was necessary to re-calculate their positions on the pages. To do this, the script that generated the location (in pixels) of objects on the page was modified and re-executed. Then the coordinates of the pixels of the objects were normalized according to the size of the stimulus, so that the coordinates of the objects came from being stored as integers between 0 and the length of the object, to a double-precision value between 0 and 1. Regarding the concepts (which describe objects), the way in which they were generated was validated by the business expert, so it was not necessary to make any changes, except for grouping or breaking up the concepts of objects that were modified by the expert. Then, the script that calculates the conceptual similarity between objects was implemented. To implement this script, the algorithm defined previously by Velásquez et al. (2011) was followed. Data transformation was

separated from user interest according to the way in which the data were captured.

The first step in transforming this data was grouping them according to the point observed (in a small neighborhood), increasing the duration of fixation according to the number of records grouped. A known statistical result is the range of duration of fixations, which are between 150 and 600 milliseconds (ms) (Duchowski, 2003), where the lowest threshold at which the brain can understand what is being observed is 150 ms. As noted above, the eye tracker used has a resolution of 120 Hz, therefore it captures information every 8 ms. With this information, the records (already grouped) with the fixation-duration field lower than 150 ms were considered invalid, as in this case, the user interviewed never noticed whether he or she was looking at a point at that moment. With the invalid records, whether because of a fixation-duration lower than 150 ms or because it was indicated by the validity-left and validity-right attributes, an indicator was constructed, called acceptance-percentage, which showed the percentage of time in which the eye movements of the controlled user upon each stimulus (website) were correctly measured. If this indicator was small, it would be a mistake to consider the records corresponding to the stimuli for further analysis, because results would simply be extrapolated which could be valid or not. For this reason, the records corresponding to stimuli with low acceptance-percentage were eliminated. Stimuli with acceptance-percentage higher than the average of all values plus two standard deviations were considered valid, resulting at a minimum 86%. From this result it is possible to consider the data capture with the eye tracker as successful. Finally, the permanence times were averaged.

On the other hand, the results of the survey corresponded to files where the page, the object, and points of interest assigned by the user were registered. These data were transformed so that the 10 points assigned by each user, corresponded to 100% of user interest in the stimulus. For example, if a person assigned 5 points to object *X* on page *Y*, it was assumed that the interest of the user in object *X* was 50% on page *Y*. Similar to the previous case, 0 was assigned to the interest of the user in objects that did not get points. Afterwards these results were averaged and stored. Starting from web requests recovered, the next step was to run the process of sessionization. The result was a set of sessions, in which each of these contained a list of pages associated with the amount of time spent on these. Then, for each register, the page was replaced by the objects that made it up, and the time spent on the page was weighed by the percentage of permanence on the object. Additionally the objects that appeared on more than one page were considered with special attention, because for a session the same object could have been seen twice. In this case, the times for both pages were added up, because they were paid attention twice. Then the most important *n* objects in each session were selected. The criterion for selecting the most important objects was the time spent on them.

### 5.1.3. Application of data mining algorithms

The methodology to find Website Keyobjects designed in Velásquez et al. (2011) grouped the Important Objects Vectors (IOV see Eq. (5)) by using three techniques: Self-Organizing Feature Maps, *k*-means and Association Rules. The results of these techniques were sets of vectors where the elements were similar to each other, but different when taking elements from different sets. The criteria used to determine if an object was a Website Keyobject was to select the objects that appeared more times in the clusters shown by the three algorithms.

### 5.1.4. Experiments

As a first experiment, the same strategy used in Velásquez et al. (2011) was developed, in which the results of the survey

about the importance of the web objects, from the user point of view, was used for estimating the time spent by a user seeing an object in a visited page during his/her session. In that sense, the average of points assigned in Table 1 per web object were calculated.

Because the sessionization process gives us the total time spent by the web user in a visited page, an estimation for the time spent seeing each web object in the page is necessary. This estimation is obtained by weighing the time spent in each page with the average of points from the above survey.

In the second experiment, the above described survey is eliminated and the time spent for the web user seeing the web objects in a visited page is calculated by using the data generated with the application of the eye tracking tool.

## 5.2. Results

The SOFM network was implemented in Python using  $\{12 \times 12, 14 \times 14, 18 \times 18, 24 \times 24\}$  neurons in the chart with a toroidal topology. However, after several tests regarding the number of neurons of the SOFM network, it was found that one constituted by  $12 \times 12$  neurons gave the best results.

This network gave 8 clusters for each experiment. The output of this algorithm was modified to deliver lists of session identifiers belonging to each cluster, in order to count the number of occurrences of objects in different clusters.

Regarding the *k*-means technique, it was implemented in Java 1.42, using *k* as the number obtained by the previous technique, i.e., 8. On the other hand, for the implementation of Association Rules, the Weka platform<sup>2</sup> was used. Only objects in the IOVs that were processed according to the platform were considered. The Apriori algorithm was used and it was asked to generate only 30 rules with a maximum confidence of 0.9. Both the algorithm and the platform to be used were the ones used by Velásquez et al. (2011) when he implemented this methodology.

### 5.2.1. Precision, recall and F-measure

To compare the results of both experiments, three ranges were selected, taking 10, 20 and 30 Website Keyobjects. Table 2 shows the precision achieved by experiments in the three ranges selected. Table 3 shows recall, and Table 4 shows the F-measure indicator.

An improvement of 15–20% can be noticed by comparing both experiments, which validates the fact that this technology is useful for measuring the interest of users. On the other hand, one can note that when choosing 30 Website Keyobjects the accuracy obtained decreases. This is because for this number a selection of objects that are not relevant begins. At this point, if we consider  $n=24$  as threshold it is concluded that the accuracy of the second experiment is 93%. Regarding recall, when taking 10 or 20 Website Keyobjects, this indicator is considerably smaller. This is expected since some Keyobjects are no longer being considered. However, in the second experiment, better results for all ranges observed are obtained. Finally, F-measure is a measure that combines both accuracy and recall indicators, therefore it was expected to achieve better results in the second experiment.

## 6. Conclusion and future work

In the present paper it was proved that using an eye tracker to measure the amount of time users spend looking at the different

<sup>2</sup> <http://www.cs.waikato.ac.nz/ml/weka/>



**Table 2**  
Precision.

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

**Table 3**  
Recall.

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

**Table 4**  
F-measure.

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

objects on a web page instead of conducting a survey to estimate these values improve the accuracy of finding the Website Key-objects of a site. To achieve this result, a comprehensive study on data originating from the Web, on the mathematical models used to describe the behavior of web users and on existing eye-tracking tools was carried out.

One of the limitations that eye tracking has is that it only determines what a person looks at. This is insufficient when what is sought is to qualify what a person looks at. In other words, these tools cannot determine if what is observed is liked or disliked. Therefore, the results provided by the eye tracker must be considered as the measure (always positive) of the interest of a person.

With the restriction described above, inherent in eye tracking, the following step to improve the methodology to find Website Keyobjects, the need to consider what users feel or think while looking at a web object naturally arises. This could be achieved by using elements of neurotechnology, such as electrodes that measure what brain area is more excited when a person is navigating a website. Emotiv.com is a company that manufactures this type of equipment (for this type of research) at a low cost.

On the other hand, a second improvement to the methodology consists of further exploring the data generated by the eye tracker. During the course of this research only these data were used to determine what objects the controlled users looked at. However, there is a wide range of studies that can use these data, such as determining from the trajectory of the eye movements if a person understands what he or she sees or not. Furthermore, no data on pupil dilation were considered, it being known that the more dilated they are, the greater the interest of a person in what they see. However, it is still not possible to classify this interest as good or bad.

Other ways to enrich this methodology are not explored herein, for example, to establish an ontology to standardize metadata. This could establish more expressive relationships between objects. In this same area, it could be possible to automate the process generating the metadata, because so far it is a slow and tedious process that, being manual, allows the introduction of unintentional errors. Finally, it could be possible to analyze the way in which this methodology operates, to investigate whether changing the structure of it can achieve

better results than using three different data mining algorithms. Thus, creating a new way of ranking the objects may help in this goal.

## Acknowledgment

This work was supported partially by the FONDEF project D10I- 1198 entitled, WHALE: Web Hypermedia Analysis Latent Environment and the Millennium Institute on Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16).

## References

- Ali-Hasan, N.F., Harrington, E.J., Richman, J.B., 2008. Best practices for eye tracking of television and video user experiences. In: *Proceeding of the First International Conference on Designing Interactive User Experiences for TV and Video. UXTV '08*. ACM, New York, NY, USA, pp. 5–8.
- Bulling, A., Ward, J., Gellersen, H., Tröster, G., 2011. Eye movement analysis for activity recognition using electrooculography. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (4), 741–753.
- Buscher, G., Cutrell, E., Morris, M.R., 2009. What do you see when you're surfing? using eye tracking to predict salient regions of web pages. In: *Proceeding of the CHI 2009*. ACM, New York, NY, USA, pp. 21–30.
- Buscher, G., Dunais, S. T Cutrell, E., Morris, M.R., 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, pp. 42–49.
- Chang, G., Healey, M., McHugh, J., Wang, J., 2001. *Mining the World Wide Web*. Kluwer.
- Duchowski, A., 2003. *Eye Tracking Methodology: Theory and Practice*. Springer Verlag.
- Dujovne, L.E., Velásquez, J.D., 2009. 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*. Springer-Verlag, Berlin, Heidelberg, pp. 301–308.
- Etzioni, O., 1996. The world-wide web: quagmire or gold mine?. *Commun. ACM* 39 (November), 65–68.
- Frintrop, S., Erich, R., Christensen, H.I., 2010. Computational visual attention systems and their cognitive foundations: a survey. *ACM Trans. Appl. Percept.* 7 (January), 6:1–6:39.
- Hoffman, D., Novak, T., 1997. A new marketing paradigm for electronic commerce. *Info. Soc.* 13 (1), 43–54.
- Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- Jay, C., Lunn, D., Michailidou, E., Harper, S., Yesilada, Y., 2008. End user evaluations. In: Harper, S., Yesilada, Y. (Eds.), *Web Accessibility*. Springer, London, UK, pp. 107–126.
- Kaufman, A.E., Bandopadhyay, A., Shaviv, B.D., 1993. An eye tracking computer user interface. In: *Proceedings of the IEEE 1993 Symposium on Research Frontiers in Virtual Reality*. IEEE, pp. 120–121.
- Kosala, R., Blockeel, H., 2000. Web mining research: a survey. *ACM SIGKDD Explor. Newsl.* 2 (1), 1–15.
- Lee, S., Lee, W., Stout, P., 2004. A comparison of objective characteristics and user perception of web sites. *J. Interact. Advert.* 4 (2), 61–75.
- Michailidou, E., Harper, S., Bechhofer, S., 2008. Visual complexity and aesthetic perception of web pages. In: *Proceedings of the 26th Annual ACM International Conference on Design of Communication. SIGDOC '08*. ACM, New York, NY, USA, pp. 215–224.
- Morimoto, C., Koons, D., Amir, A., Flickner, M., 2000. Pupil detection and tracking using multiple light sources. *Image Vision Comput.* 18 (4), 331–335.
- Nielsen, J., 2004. *Usability Engineering*. Academic Press, San Diego, CA, USA.
- Nielsen, J., Pernice, K., 2009. *Eyetracking Web Usability*. New Riders Pub.
- Novak, T., Hoffman, D., Yung, Y., 2000. Measuring the customer experience in online environments: a structural modeling approach. *Market. Sci.* 19 (1), 22–42.
- Pal, S.K., Talwar, V., Mitra, P., 2002. Web mining in soft computing framework: relevance, state of the art and future directions. *IEEE Trans. Neural Networks* 13 (5), 1163–1177.
- Poole, A., Ball, L., 2005. Eye tracking in human-computer interaction and usability research: current status and future prospects. In: *Encyclopedia of Human Computer Interaction*, pp. 211–219.
- Roberts, D., Shelhamer, M., Wong, A., 2008. A new wireless search-coil system. In: *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. ACM, pp. 197–204.
- Van der Geest, J.N., Frens, M.A., 2002. Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. *J. Neurosci. Methods* 114 (2), 185–195.
- Velásquez, J.D., 2011. Web site keywords: a methodology for improving gradually the web site text content. *Intell. Data Anal.* 15 (4), 327–348.

- Velásquez, J.D., 2012. Improvement of a methodology for website keyobject identification through the application of eye-tracking technologies. In: Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops. BI 2012. IEEE Computer Society, pp. 59–63.
- Velásquez, J.D., Dujovne, L.E., L'Huillier, G., 2011. Extracting significant website key objects: a semantic web mining approach. *Eng. Appl. Artif. Intell.* 24 (8), 1532–1541.
- Velásquez, J.D., Jain, L.C., 2010. *Advanced Techniques in Web Intelligence—Part 1*, first ed. Springer Publishing Company, Incorporated.
- Velásquez, J.D., Palade, V., 2008. *Adaptive Web Sites: A Knowledge Extraction from Web Data Approach*. IOS Press.
- Velásquez, J.D., Ríos, S., Bassi, A., Yasuda, H., Aoki, T., 2005. Towards the identification of keywords in the web site text content: a methodological approach. *Int. J. Web Inf. Syst.* 1 (1), 11–15.