

**Título del Proyecto
de Investigación a que corresponde el Reporte Técnico
PARCIAL:**

Retroalimentación implícita de relevancia de textos durante la
búsqueda de información mediante eyetracking

Tipo de financiamiento

Financiamiento externo no administrado por la UACJ

Autores del reporte técnico:

Dr. Francisco López Orozco
Dr. Rogelio Florencia Juárez
Ing. Jorge Alberto Medina Franco

Retroalimentación implícita de relevancia de textos durante la búsqueda de información mediante eyetracking (REPORTE PARCIAL)

Resumen del reporte técnico en español (máximo 250 palabras)

La búsqueda de información en los motores de búsqueda de Internet es simple y natural. Sin embargo, no toda la información es relevante. ¿Cómo un usuario puede determinar que un enlace no solo es relevante si no también efectivo para satisfacer su búsqueda de información? Una respuesta a esta pregunta es mediante la retroalimentación implícita, es decir, datos que se obtienen sin tener que manifestar de manera directa la recolección de estos. Llevando a cabo un análisis y estudio con usuarios a partir de datos oculométricos obtenidos de manera implícita se logró inferir elementos clave respecto a su comportamiento durante las tareas de búsqueda de información en los motores de búsqueda web. Se observaron y determinaron comportamientos que utilizan los usuarios para tomar una decisión en las tareas de búsqueda de información y en particular el comportamiento al momento de visualizar los *snippets* de la página de resultados. A partir de los datos obtenidos se revela lo que un usuario considera relevante al momento de investigar, observar y/o decidir respecto a la búsqueda de información. La conglomeración de los datos obtenidos de los distintos usuarios permitió realizar un prototipo de *snippet* que refleja la visión de estos en cuanto a relevancia se refiere.

Resumen del reporte técnico en inglés (máximo 250 palabras):

The search for information in the Internet search engines is simple and natural. However, not all information is relevant. How can a user determine that a link is not only relevant but also effective to satisfy their information search? One answer to this question is through implicit feedback, that is, data that are obtained without having to directly manifest the collection of these. By carrying out an analysis and study with users based on implicitly obtained oculometric data, it was possible to infer key elements regarding their behavior during the information search tasks in the web search engines. Behaviors that users use to make a decision in the information search tasks and in particular the behavior at the time of viewing the snippets of the results page were observed and determined. From the data obtained, what a user considers relevant when investigating, observing and / or deciding regarding the search for information is revealed. The conglomeration of the data obtained from the different users allowed to make a snippet prototype that reflects their vision regarding relevance.

Palabras clave: retroalimentación implícita, relevancia de textos, búsqueda de información, eye tracking

Usuarios potenciales

Desarrolladores de sitios web con énfasis en experiencia de usuarios (UX), usabilidad, testers, etcétera.

Reconocimientos

Al Laboratorio de Tecnologías Emergentes en Ciencias de la Computación (LabTEC2) de la División Multidisciplinaria de la UACJ en Ciudad Universitaria por facilitar el uso del equipo y sus instalaciones. A la Universidad de Grenoble-Alpes (UGA) por el apoyo financiero recibido para efectuar una estancia de investigación. A los estudiantes de pregrado de UACJ, principalmente de los programas de Ing. de Software e Ing. en Sistemas Computacionales que participaron durante la fase experimental.

1.-INTRODUCCIÓN

Para tener acceso a la información hoy en día tan solo es cuestión de unos segundos, vivimos en una época donde es más sencillo que nunca verificar hechos, responder dudas y comunicarnos con personas a kilómetros de distancia. De igual manera esta información a la que se tiene acceso no tiene limitantes, uno incluso podría argumentar que es infinita debido a que día tras día surgen nuevas ideas y nuevos conocimientos a los que se puede tener acceso. La herramienta que facilita la búsqueda y obtención de esta información se le conoce como motor de búsqueda, la cual su principal función es que a partir de palabras clave que le otorga el usuario el motor arroja un listado de sitios web que contienen información relacionada a las palabras clave que el usuario proporcionó. A partir de ahí, surgen varias preguntas, ¿cómo sabe el usuario seleccionar el enlace correcto del listado que se le muestra? Los motores de búsqueda cuentan con un elemento visual que tiene como objetivo reflejar una pequeña parte de información que está contenida dentro del enlace. A esto se le conoce como *snippet*. Es en este objeto donde se enfoca la mayor parte de esta investigación, a través de la obtención de datos oculométricos implícitos utilizando un *eyetracker*. Se realizaron pruebas a múltiples usuarios donde estos realizaron búsquedas de información en el motor de búsqueda y mediante la captura de video utilizando el *eyetracker* se obtuvieron datos sobre los puntos de enfoque de la vista de los usuarios, lo que permite analizar en qué se enfoca más un usuario dentro de los *snippets* para poder llegar a una conclusión, con el objetivo de determinar qué elementos contenidos dentro de los *snippets* son los que más le interesan a los usuarios y determinar si estos tienen algún factor de impacto que determine si el enlace que el usuario optó por seleccionar es relevante para llevar a cabo una búsqueda de información exitosa. El objetivo de esta investigación y las pruebas fue para presentar un prototipo de *snippet* con los elementos más relevantes y de mayor utilidad según los datos obtenidos de los usuarios. No se extendió esta investigación al desarrollo de un sistema completo pues se trata de un estudio inicial que busca ilustrar los resultados y dar paso a futuras investigaciones y mejoras para las distintas áreas que abarcan los motores de búsqueda, *snippets* y los navegadores web como tal.

2. PLANTEAMIENTO

En muchas situaciones la información que se obtiene de los motores de búsqueda está estrictamente ligada a las palabras clave que el usuario proporciona para su búsqueda. Si bien, esto es visto como lo correcto e ideal, muchas veces la información que es recuperada no es la indicada para el usuario pues se busca algo más técnico y la generalización que presentan los motores de búsqueda es que los resultados principales que arrojarán se puedan clasificar en dos secciones: aquellos que son sitios o artículos patrocinados y aquellos que se encuentran en su lugar por la cantidad de visitas que reciben día a día. Al día de hoy no existe una clara distinción entre la información clave y la información popular en los motores de búsqueda populares a excepción de los académicos como Google Académico (Scholar Google) y otros. La principal problemática en esta situación es el poder saber qué información es la relevante para los usuarios fuera de palabras clave o enlaces populares. Lo que se pretendió en este proyecto es que a través de los datos obtenidos mediante la investigación realizar un análisis estadístico de los elementos más relevantes contenidos dentro de un *snippet*, y con estos realizar las pruebas necesarias para poder generar un prototipo de *snippet*.

Antecedentes

El estudio realizado por Marcos y Gonzalez [1] buscaba identificar si la intención de las consultas afectaba a la forma en que el usuario explora la página de resultados de un buscador. Para ello utilizan parámetros a los cuales se les conoce como fijaciones. Éstas determinan el movimiento ocular a través de una secuencia que determina los puntos donde se realizó el enfoque de la vista y en cómo se fue formando un tramo que permite determinar la trayectoria de lo que visualizó el usuario. En el caso del estudio realizado se muestran las fijaciones que abarcan el título del sitio, el *snippet*, *url* e imágenes de los primeros resultados contando enlaces patrocinados. Este estudio se lleva a cabo utilizando un *eyetracker* para análisis del movimiento ocular de los usuarios y detección de patrones de lectura o atención de en dónde se concentraba más la vista. Buscaban mejorar los buscadores de Internet para el mejor funcionamiento de las páginas que se muestran al momento de que los usuarios realicen una consulta. El trabajo realizado por

Georg Buscher [2] describe un algoritmo que permite determinar de manera automática la relevancia del texto contenido dentro de los sitios a partir de los movimientos oculares. El algoritmo es capaz de detectar y distinguir entre una lectura continua y cuando existe un parpadeo. Basado en los resultados obtenidos por el algoritmo se determina la relevancia de un documento. Posteriormente, éstas son clasificadas dentro de un parámetro al que se le conoce como "Características". El siguiente paso es que aquellos puntajes que se le asignaron a estas características se acumulen para así dar lugar al último paso del algoritmo donde se determinan los umbrales para distinguir entre lectura y un simple paso del movimiento ocular sin retención de atención. De esta manera se determina el comportamiento adecuado para cada situación. El objetivo final del algoritmo es realizar un método para predecir la relevancia a partir del movimiento ocular de manera explícita a la retroalimentación de los usuarios. Otro trabajo es realizado por Salojärvi [3], donde se basa en inferir si es posible tomar la retroalimentación implícita y determinar lo relevante al usuario al momento de recuperar información en algún sitio web o artículo. Se le encomiendan al usuario tareas específicas de búsquedas de información para modelar a partir de movimientos oculares cuál de los títulos es relevante. El objetivo general es determinar la relevancia de los títulos de los documentos para poder realizar una predicción con una precisión razonable a partir de unas pocas características.

Marco teórico

Eyetracking

Se le llama *eyetracking* al proceso de medición del punto de la vista, es decir a donde está enfocándose uno o bien la medición del movimiento ocular relativo a la cabeza. Para poder llevar a cabo el proceso de *eyetracking* se utiliza un *eyetracker*. Su función principal es llevar a cabo mediciones de las posiciones de los ojos y movimientos oculares. Estos dispositivos se ven utilizados con frecuencia los campos de investigación del sistema visuales, psicología, psicolingüística, marketing, dispositivos de captura para la interacción hombre-máquina y diseño de producción [4]. En el *eyetracking* se le conoce como fijación al punto de enfoque de la vista, mientras que al camino que sigue la vista del usuario entre fijaciones se le conoce como sacada.

Hoy día, la tecnología eyetracking está basada en mediciones no invasivas ya que no requiere contacto directo con el sujeto de prueba. Se usa luz infrarroja que es reflejada en los ojos para posteriormente ser captada por una cámara o un sensor óptico. En la Fig. 1 se representa el comportamiento del movimiento ocular al realizar la lectura de texto y capturada por una cámara[5].

Este un texto sin relevancia, su objetivo es meramente ilustrativo y servir como ejemplo en alguna representación grafica.

Fig. 1: Ejemplos de fijaciones y el movimiento ocular.

Snippet

Un *snippet* es un extracto de una página web. Su objetivo es que al momento de realizar búsquedas de información en el navegador se concentren los puntos principales de los distintos sitios de manera que le sea más sencillo identificar al usuario los sitios web que contengan la información relevante con la búsqueda que realizó. Está compuesto por un título, una descripción donde se encuentra contenida parte del texto del sitio web o descripción del sitio. Por último la parte final, consiste en el URL del sitio web que normalmente viene de un color distinto al del título y la descripción para resaltarlo y el usuario pueda identificar sitios más fácilmente (Fig. 2) [6].

The diagram illustrates three examples of search snippets. Each example consists of a title, a URL, a date, and a description. Red arrows point from labels to the corresponding parts of the snippets. The labels are: 'Fecha del snippet.' (pointing to a date), 'Título del snippet.' (pointing to a title), 'URL / Enlace del snippet.' (pointing to a URL), and 'Descripción del snippet.' (pointing to a description). The snippets are:

- Snippet 1:**
Title: Un software hace que las webcams rastreen los ojos de los usuarios
URL: <https://www.tendencias21.net/Un-software-hace-que-las-webcams-rastreen-los-ojos-de...>
Date: 8 jun. 2016 - La técnica de seguimiento ocular se utiliza en neuromarketing para detectar los puntos de mayor interés de una imagen. Ahora, un nuevo ...
Description: (truncated)
- Snippet 2:**
Title: ¿Qué es el "Eye Tracking" y para qué nos sirve? | C2 Intelligent ...
URL: <https://www.solucionesc2.com/que-es-el-eye-tracking-y-para-que-nos-sirve/>
Date: 13 jul. 2012 - El concepto de Eye Tracking se basa en una tecnología de seguimiento ocular que nos permitirá saber comportamientos visuales del usuario ...
Description: (truncated)
- Snippet 3:**
Title: Cómo funciona el seguimiento ocular - Tobii Dynavox
URL: <https://www.tobiidynavox.com/es/sobre-tobii.../como-functiona-seguimiento-ocular/>
Date: (truncated)
Description: El seguimiento ocular, o interacción a través de la mirada, es una tecnología que se usa para saber adónde mira una persona que está viendo la pantalla de un ...

Labels and arrows:

- Red arrow pointing to the date: **Fecha del snippet.**
- Red arrow pointing to the title: **Título del snippet.**
- Red arrow pointing to the URL: **URL / Enlace del snippet.**
- Red arrow pointing to the description: **Descripción del snippet.**

Text on the right: **Snippets de búsqueda típicos.**

Fig. 2: Ejemplos de un snippet básico.

Motor de búsqueda web

Se le llama motor web al software que permite realizar búsquedas de información en la *World Wide Web*. Por lo general los motores de búsqueda presentan los resultados en forma de lista donde se muestra un resumen o extracto del sitio web donde se encuentra la información deseada; a este resumen se le conoce de manera formal como *snippet*.

Web Crawler

Un *web crawler* o araña es un *bot* que utilizan los motores de búsqueda para navegar a través del Internet con el objetivo de catalogar los sitios. Su funcionamiento general es: se tiene una lista de URL iniciales (sitios semilla). Una vez que se visitan estos sitios, la araña identifica los hipervínculos existentes en el sitio y los agrega a la lista de URL que debe visitar. El siguiente paso es archivar los sitios y a medida que entra a los sitios se guarda la información y de igual manera se crean copias de ésta. La información que se guarda puede ser vista como si fuera el sitio web verdadero pero esta almacenada en un formato de *snapshot* [7].

3. METODOLOGÍA

Se siguió una metodología experimental y se establecieron los siguientes cuestionamientos desde el punto de vista del usuario: ¿qué elementos que conforman a un *snippet* tradicional y son más relevantes al momento de realizar una tarea de búsqueda de información? y ¿qué efecto tiene sobre el usuario alterar la estructura y/o elementos ya previamente establecidos? ¿cómo puede saber qué enlace es el indicado con tan solo visualizar el *snippet*?, ¿de las consultas que realiza el usuario existen casos de falsos positivos y positivos falsos tanto para el *snippet* tradicional como el prototipo?

Hipótesis

Aquí se enunció la hipótesis: “Si se modifica la estructura actual del *snippet* entonces la manera en que el usuario realiza tareas de búsqueda de información cambiará”.

Definición de variables

Las variables independientes son los parámetros que se les entregaron a los usuarios para realizar sus tareas de búsqueda de información.

El primer paso fue visualizar de qué manera se les presenta a los usuarios la información para cada una de las áreas con el objetivo de analizar qué impacto tendrían estos resultados al momento de la colecta de datos durante las pruebas.

En la Fig. 3 se resumen las áreas evaluadas y la estructura de la experimentación.

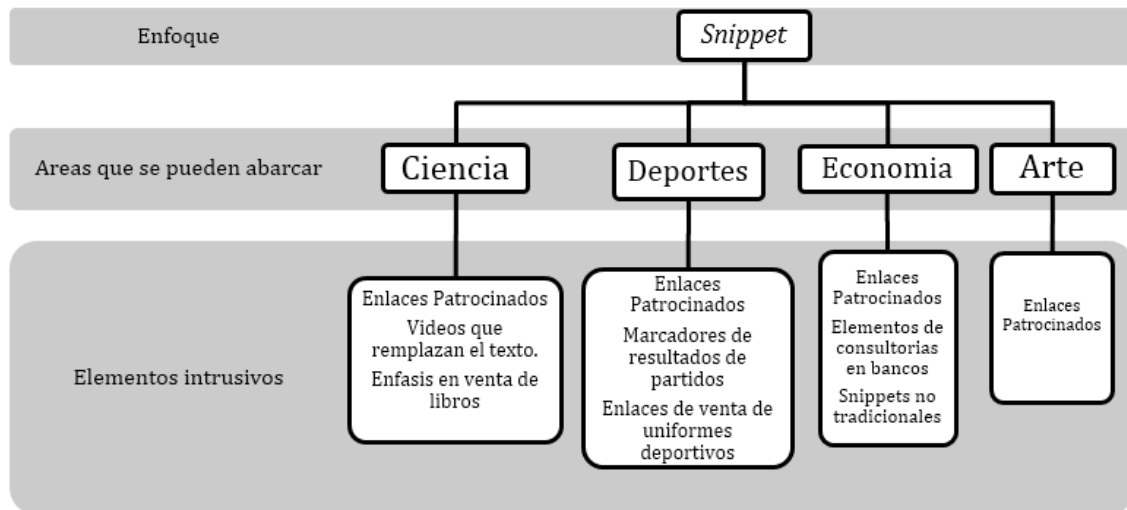


Fig. 3.- Esquema para marcar la delimitación del área de interés en tareas de búsqueda de información.

Las variables dependientes son las que genera el *eyetracker* al momento de hacer la captura de los datos oculares de los sujetos de prueba. La Tabla 1 es una representación de los datos que arroja el *eyetracker* en cada prueba.

Tabla 1.- Datos obtenidos de un usuario al pasar un experimento.

TIME (Segundos)	TIME_TICK (Entero)	FPOGX (Decimales)	FPOGY (Decimales)	FPOGS (Segundos)	FPOGD (Segundos)
150.78	3682998817	0.38	0.30	150.64	0.14
150.80	3683056430	0.38	0.30	150.64	0.16

Donde **TIME** es el segundo o fracción de segundo en el que sucede una fijación, **TIME TICK** es la etiqueta con la que se identifica fijaciones específicas, **FPOGX & FPOGY** representan el punto de enfoque del usuario en decimales en el eje “X” y “Y”

respectivamente, **FPOGS** representa el inicio de las fijaciones en segundos desde que el dispositivo fue calibrado, y **FPOGD** representa la duración de la fijación en segundos.

Elección de un diseño experimental

Éste debía permitir que la información se recolecte de manera implícita [8] y todas las pruebas debían poder llevarse a cabo en un lapso de tiempo de duración corto (5-15 min). La selección de participantes para las pruebas se optó por sujetos únicos sin capacidad de recursión, es decir, no podrían realizar múltiples pruebas fuera del día asignado a cada uno. Para determinar el tamaño de la muestra que se requería para llevar a cabo el experimento se utilizó la siguiente fórmula (1):

$$\text{Tamaño de la muestra} = \frac{N * Z^2 * pZ(1 - p)}{(N - 1) * e^2 + Z^2 * p * (1 - p)} \quad (1)$$

En donde N = tamaño de la población, e = margen de error, Z = nivel de confianza deseado y p = la proporción que se desea encontrar. Con una población de 60 personas, un nivel de confianza del 95% y un margen de error de 5 la muestra da un total de 53 sujetos para realizar las pruebas. Una vez que se definió de manera concreta el número de sujetos a utilizar lo siguiente fue la asignación de estos a un grupo para su análisis.

Selección y asignación de los sujetos a grupos

La selección de los sujetos de prueba se basó en que fueran pertenecientes a la UACJ en el campus de Ciudad Universitaria e IIT/IADA. Para cada sujeto que realizó la prueba se fue generando una bitácora que incluye el nombre del participante, la carrera perteneciente, y notas sobre la manera en que lleva a cabo la prueba y ciertos patrones que tiene en común con otros sujetos.

Procedimiento y recogida de datos

Para realizar las pruebas a cada participante se le daba una serie de instrucciones que aseguran comprendiera lo que debe hacer sin interferir directamente sobre las acciones que él tomará al momento de comenzar la prueba. Como se mencionó con anterioridad la recolección de datos sucede de manera automática y es implícita debido a que al momento de que el usuario está realizando la prueba el *eyetracker* se encarga de capturar todo movimiento ocular del usuario.

4. RESULTADOS

Con base en las pruebas que se realizaron se hizo un análisis estadístico para determinar las variables que afectan al momento de que un usuario realiza tareas de búsqueda de información. El *eyetracker* arroja una hoja de datos en formato XML que refleja los datos que se obtienen inmediatamente cuando se finaliza una prueba. Ésta consta de 57 columnas con distintas variables que abarcan desde el ancho y alto del monitor, variables para identificar cada elemento, y variables respecto a la posición de los ojos, pupilas y vista del usuario en los distintos ejes de la pantalla. Se hizo una simplificación conveniente de las 57 columnas para poder llevar a cabo el análisis de los datos. La Tabla 2 refleja los datos relevantes para la investigación.

Tabla 2.- Datos con las columnas relevantes para la investigación.

TIME (Seg)	FPOGX (Eje X)	FPGOY (Eje Y)	FPOGS (Seg.)	FPOGD (Seg)	FPOGV (ID)	BPOGX (Eje X)	BPOGY (Eje Y)	BPOGV (ID)
26.482	0.220	0.056	25.727	0.756	1	0.199	0.023	1
26.499	0.221	0.056	25.727	0.772	1	0.249	0.046	1
26.515	0.221	0.056	25.727	0.789	1	0.26	0.085	1

Donde **TIME** representa el tiempo en segundos en que ocurre una fijación, **FPOGX** y **FPOGY** son la fijación del usuario en el eje X y Y respectivamente en relación a la posición del monitor, **FPOGS** es la fijación desde que el sistema fue inicializado/calibrado. **FPOGD** es la duración de la fijación del usuario en segundos, **FPOGV** si se encuentra con el valor 1 significa que la fijación es válida, si se encuentra un valor 0 en la casilla quiere decir que esa fijación no fue capturada de manera correcta. **BPOGX** y **BPOGY** representan la mejor estimación del punto de enfoque del usuario en el eje X y eje Y y finalmente **BPOGV** es 1 cuando la fijación es válida, si se encuentra con un 0 el dato no es válido.

Además para poder llevar a cabo un mejor control y análisis de las fijaciones se creó la clasificación de la Tabla 3.

Tabla 3.- Categorización de elementos principales que componen a un *snippet*.

ID	ELEMENTO
1	Título Sitio
2	Enlace
3	Descripción
4	Fecha
5	Otros

En la Tabla 3 se observa los cuatro elementos principales que componen un *snippet*, el **Título del sitio**, el **Enlace**, la **Descripción**, la **Fecha**. El orden de **otros datos** se refiere a cualquier otro elemento poco común que normalmente no aparece en los *snippets* tradicionales, esto incluye imágenes, videos o calificaciones del sitio.

Adicionalmente se hizo un conteo general que se muestra en la Tabla 4 que permite ver el elemento que el usuario mayormente se enfocó y aquel que no fue de mucha atención. A esto se le llamó como análisis intra-*snippet* debido a que se enfoca en los elementos que conforman un *snippet* y su estructura interna.

Tabla 4.- Conteo de las fijaciones que realizó un usuario para cada uno de los elementos del *snippet* durante la búsqueda de información.

ID	SIGNIFICADO	TOTAL
1	Título Sitio	43
2	Enlace	69
3	Descripción	154
4	Fecha	0
5	Otros	6

El otro análisis fue el inter-*snippet* para determinar en qué momento un usuario pasa de un *snippet* a otro y que es lo primero que el usuario visualiza en un *snippet*.

Tabla 5.- Análisis inter-*snippet*.

Tiempo (Seg)	FPOGD (Seg)	ID	Visualización de <i>snippets</i>	Primera ocurrencia
752.54	0.476	5	Primer <i>snippet</i>	6/54
752.56	0.493	5		fijaciones

752.58	0.509	5		en primera ocurrencia respecto a total
752.59	0.526	5		
752.61	0.542	5		
752.63	0.559	5		

La Tabla 5 muestra el análisis inter-*snippet* con el tiempo de ocurrencia de una fijación, su duración y el elemento que el usuario está visualizando en el momento.

La Fig. 4 muestra que es lo que el usuario visualizó en mayor y menor proporción.



Fig. 4. – Veces que un usuario visualizó algún elemento determinado del *snippet*.

En la Fig. 5 se muestra cuánto dura un usuario visualizando un elemento determinado, es una escala de mínimo – máximo.

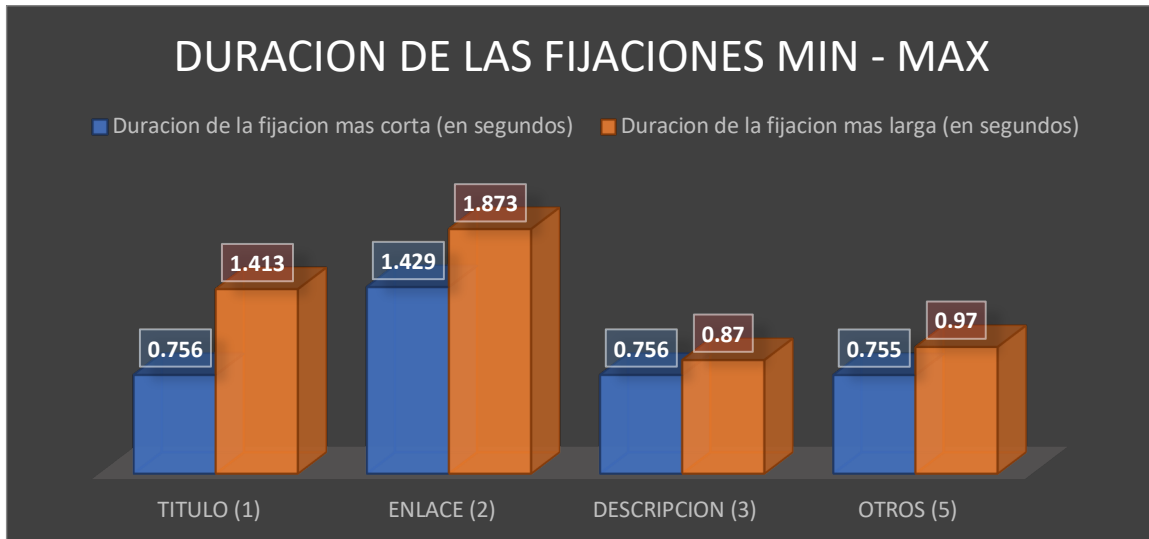


Fig. 5. – Gráfica comparativa del mínimo y máximo de duración para las fijaciones de cada elemento del *snippet*.

El siguiente paso fue crear un conglomerado de la información y datos de todos los usuarios que realizaron las pruebas para poder analizarla de manera detallada y notar patrones e indicaciones de las preferencias que tienen los usuarios. Se presenta la Fig. 6.

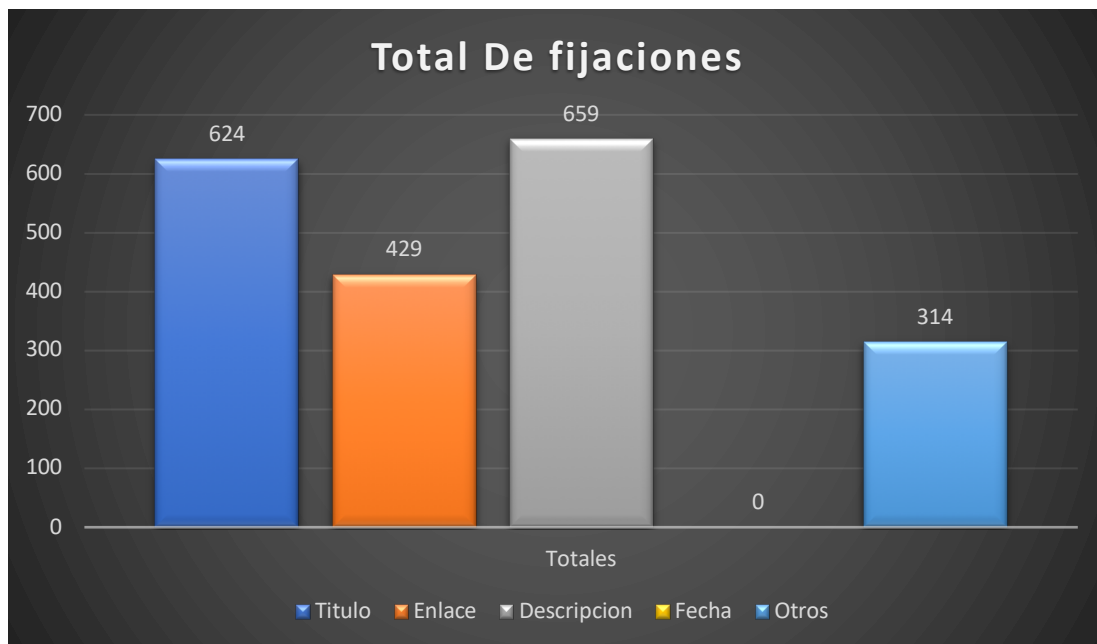


Fig. 6. – Gráfico del total de fijaciones de todos los usuarios para cada elemento del *snippet*.

La tendencia de lo que visualiza más un usuario es la descripción, seguido del título del sitio, en el siguiente lugar de la jerarquía se encuentra el enlace seguido, por último, todo otro elemento que no forma parte de un *snippet* tradicional (imágenes, videos, calificación del sitio, etc.). Es necesario recalcar la falta de fijaciones para el elemento de fecha, si bien este elemento no se encuentra presente en todos los *snippets* en la página de resultados y usualmente está relegado a un espacio pequeño la falta de atención a este elemento es un gran punto de interés ya que demuestra cómo afecta no solo la posición del elemento si no la frecuencia de este.

La Tabla 6 muestra otra recopilación de datos.

Tabla 6.- Duración de las fijaciones de los usuarios para cada elemento del *snippet*.

Usuarios	Fijaciones (Tit / Seg)	Fijaciones (URL / Seg)	Fijaciones (Descrip /seg)	Fijaciones (Fecha / Seg)	Fijaciones (Otros / Seg)
Sujeto 1	0.583	0.795	0.607	0	0
Sujeto 2	1.248	0.928	1.240	0	0
Sujeto 3	0.928	0	0.862	0	0
Sujeto 4	0.674	0.761	0	0	0
Sujeto 5	0.983	0.711	0.667	0	0.811
Sujeto 6	0.578	0.688	0.651	0	0.598
Sujeto 7	0.709	0	0.727	0	0.714
Sujeto 8	0.561	0.620	0.588	0	0.517
Sujeto 9	1.021	1.651	0.863	0	0.862
Sujeto 10	2.428	0	1.010	0	0
Sujeto 11	0.838	0.851	0.902	0	0.959
Sujeto 12	0.744	0.837	0.877	0	0
Sujeto 13	1.220	1.193	1.561	0	1.700
Sujeto 14	0.714	0.609	0.650	0	0
Sujeto 15	0.616	0.960	0.937	0	0.524
Sujeto 16	0.781	0.881	0.778	0	0.684
Sujeto 17	0.762	0.558	0.713	0	0.731

Sujeto 18	0.884	0.901	0.549	0	0.608
Sujeto 19	1.016	1.421	0.901	0	0.895
Promedio	0.9104	0.7563	0.7942	0	0.5056

En la Tabla 6 se obtuvo un promedio general agrupando la duración de las fijaciones de cada elemento de los distintos participantes de las pruebas, esto con el objetivo de tener una idea clara de cuánto es el tiempo de atención que se dedica en promedio a cada elemento que forma parte del snippet.

En contraste con la Fig. 6 donde se tiene la cantidad de fijaciones para cada elemento se puede notar una variación muy particular, claramente se puede recalcar que el elemento más visualizado por los usuarios no es aquel que acapara la mayor atención de los usuarios. Esto da una idea sobre cómo es que el usuario divide su atención en los *snippets*, es decir, a qué elemento le otorga más peso al momento de procesar la información y decidir seleccionar un enlace para realizar su tarea de búsqueda de información.

Por último, en base a la recopilación de datos también se pudo obtener la cantidad de snippets que visualizan los usuarios al momento de realizar sus tareas de búsqueda de información, esto es meramente una ayuda para comprender cómo es el patrón de búsqueda de los usuarios y saber qué tanto esfuerzo deben dedicar para encontrar el resultado esperando en cuanto a una tarea de búsqueda de información se refiere. Se presenta la Tabla 7.

Tabla 7. – Conteo de *snippets* visualizados por los usuarios.

Usuarios	<i>Snippets</i> visualizados
Sujeto 1	6
Sujeto 2	2
Sujeto 3	2
Sujeto 4	3
Sujeto 5	13
Sujeto 6	4
Sujeto 7	3
Sujeto 8	6
Sujeto 9	4
Sujeto 10	4
Sujeto 11	5

Sujeto 12	5
Sujeto 13	3
Sujeto 14	2
Sujeto 15	3
Sujeto 16	4
Sujeto 17	6
Sujeto 18	3
Sujeto 19	4
Promedio	4.31

La recopilación de estos datos no solo se realizó con el objetivo de obtener una tendencia de cómo se comportan los usuarios en las tareas de búsqueda de información respecto a los *snippets* como se planteó anteriormente. Este análisis es el primer paso para crear un prototipo de *snippet* donde se altere la estructura de éste de acuerdo a los datos que se puede observar en las Tablas 6 y 7 con el objetivo de poder capturar los distintos elementos que le son más relevantes al usuario, acomodarlos de manera que se dé prioridad a aquellos que tienen un mayor impacto en los usuarios y ver de qué manera afecta el proceso de búsqueda de información de los usuarios. Con estos datos establecidos se pudo llegar a lo siguiente: El prototipo de *snippet* que se desarrolló tiene como objetivo ilustrar lo que los usuarios comunicaron de forma implícita, es decir que el *snippet* reflejara las necesidades de estos y en teoría debe alterar el proceso de tareas de búsqueda de información de los usuarios. Para realizar el prototipo de *snippet* se realizó lo siguiente.

Debido a que se trabajó en las pruebas utilizando el motor de búsqueda de Google se optó por utilizar un API de búsqueda personalizada que permitiera recuperar y mostrar resultados de una búsqueda en la web. El orden en cómo se encuentran colocados dentro del código afecta en la manera que se le presenta al usuario el *snippet*, es decir, si se coloca el elemento que hace referencia al extracto de texto al inicio, seguido del título y por último el enlace se le presentará al usuario.

Teniendo listo el prototipo de *snippet* y los datos que indican la estructura que tomará este, se debe probar el prototipo para realizar una comparativa con los datos obtenidos en las pruebas con el *snippet* tradicional y el *snippet* prototipo. La prueba se llevó a cabo exactamente de la misma manera que fue planteada con anterioridad la única diferencia es la cantidad de sujetos de prueba que deberá ser reducido para poder concluir en un tiempo aceptable y analizar los datos. Las indicaciones para la prueba son las mismas que ya fueron establecidas previamente, teniendo eso en mente se realizó la prueba participante por participante.



Fig. 7. – Nueva versión de los *snippets* que visualizarán los usuarios.

Como se puede ilustrar en la Fig. 7 cada usuario que realizó las pruebas visualizaba los *snippets* de la siguiente manera. El enfoque principal se le dio a la descripción cambiando el tono de color, tamaño y posición donde se encuentra, seguido se encuentra el título del sitio al cual también se le modificó el color y tamaño para que no fuera opacado

completamente por la descripción. Por último, se encuentra el enlace del sitio cuya única modificación fue referente al tamaño de este para que fuera lo suficientemente notable con los participantes.

Una vez realizadas las pruebas con el nuevo prototipo de *snippet* se recopilaron los datos de manera similar a como se realizaron en las primeras pruebas, pero esta vez únicamente enfocándonos a la cantidad de fijaciones para cada elemento y la duración de dichas fijaciones. Con base a estos parámetros establecidos se obtiene la siguiente tabla de datos representada en la Fig. 8.

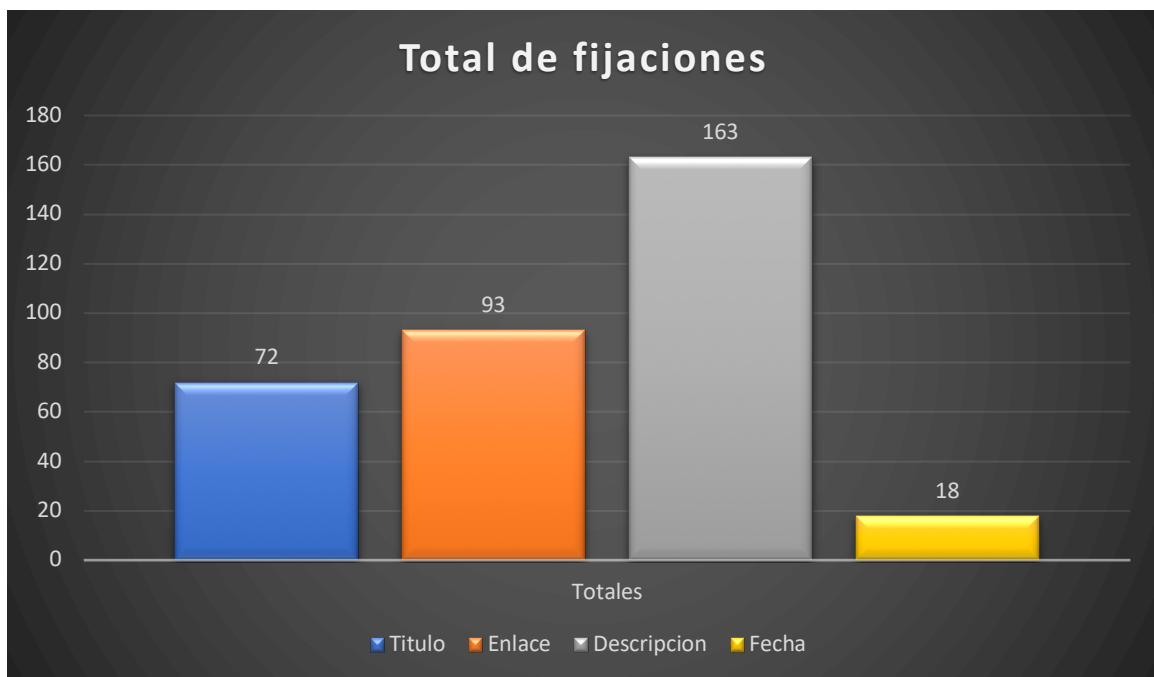


Fig. 8. – Instancias de fijaciones de los sujetos de prueba con el nuevo *snippet*.

En la Fig. 8 se ilustra la cantidad de fijaciones que cada sujeto de prueba realizó para cada uno de los elementos del nuevo *snippet*, en primera parte un aspecto que se hace recalcar bastante es el elemento de la fecha, haciendo referencia a la Fig. 6 (específicamente el elemento de fecha cuyo total resulto en 0) de manera que no existe ni una sola fijación en ese elemento ahora se puede notar este cambio de que los usuarios lo visualizaron al menos lo suficiente para notar su presencia. Otro aspecto que se debe recalcar es que la descripción, al igual que en las primeras pruebas realizadas, está posicionado como el elemento más visualizado. Sin embargo, la mayor diferencia aquí es

que el título ya no quedó como el segundo elemento más visualizado, en esta ocasión el enlace tomó esa posición a pesar de encontrarse en una posición distinta y alejada del elemento más visualizado.

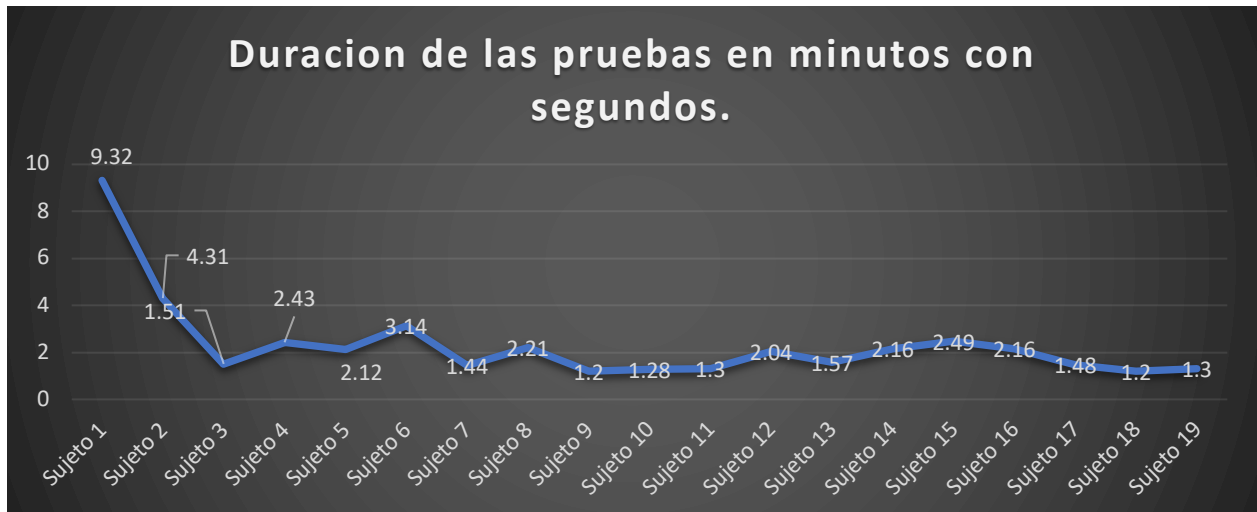


Fig. 9. – Gráfico de la duración de las pruebas en minutos con segundos con el *snippet* tradicional.

Este gráfico representa el tiempo que le tomó a cada usuario realizar su prueba de inicio a fin, los valores marcados se encuentran en función de minutos con segundos, es importante recordar que la prueba no concluía sino hasta el mismo usuario decidiera que había encontrado el enlace que le serviría para su tarea de búsqueda de información. Por esta misma razón existe una variación muy significativa. En algunos casos cada usuario está realizando la prueba de manera natural por lo tanto el tiempo que les tome debe reflejar la misma manera en que ellos realizan las tareas de búsqueda de información para su beneficio, ocio o curiosidad, el tiempo promedio de duración de estas pruebas fue de 2 minutos con 35 segundos. Ahora, como se observa en el siguiente gráfico representado en la Fig. 10, representa la duración de las pruebas de los usuarios utilizando el prototipo de *snippet* se obtuvieron los siguientes datos.

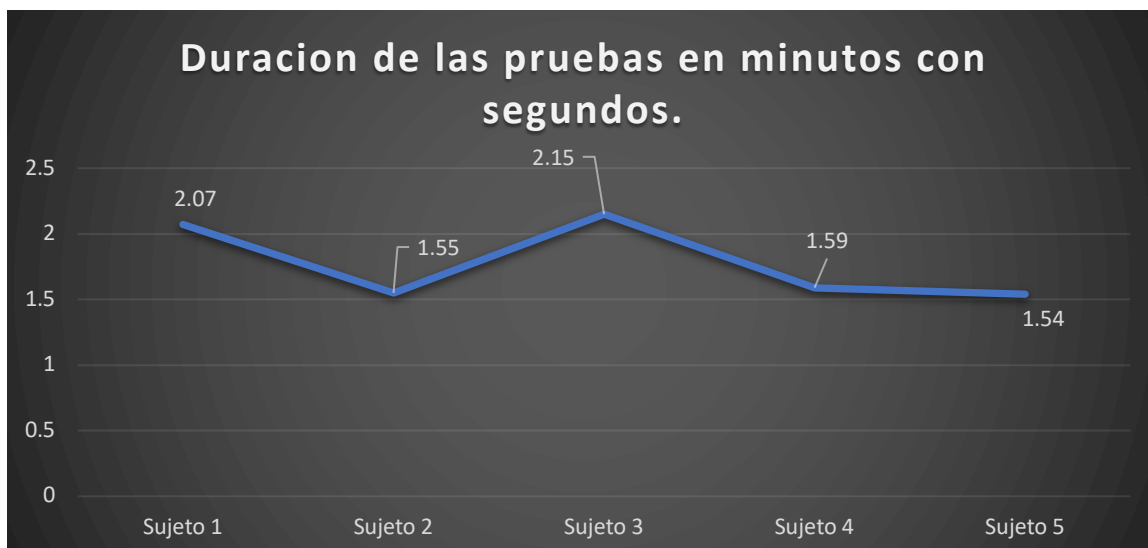


Fig. 10.- Gráfico de la duración de las pruebas con el nuevo *snippet*.

En el gráfico que se presenta en la Fig. 10 se puede notar cómo el margen de la duración de las pruebas de los usuarios se mantiene en un margen menor de tres minutos y encima de minuto y medio. Es decir que los usuarios en el transcurso de dos minutos observaron y procesaron la información necesaria para elegir un enlace en base al *snippet* y poder llegar a la conclusión de que el enlace que ellos seleccionaron les sería útil. El tiempo promedio de estas pruebas fue de 2 minutos con 18 segundos.

Discusiones

Efectivamente existe un claro cambio de comportamiento en los usuarios al modificar los *snippets* tradicionales incluso cuando estos cambios son únicamente de posición y color. La atención que le presta cada usuario a los distintos elementos hace que salgan a relucir características y propiedades individuales de estos que tal vez antes pasaban desapercibidas. Si se toma en cuenta un patrón que se visualiza en las pruebas iniciales donde la gran mayoría de los participantes optaban por ignorar el *snippet* que estuviera correlacionado directamente con Wikipedia. En la nueva propuesta de *snippet* al estar la descripción primero se crea esta necesidad de que el usuario lea un extracto del sitio antes de juzgarlo por el enlace o título del sitio. Se evita esa predisposición de saltar un enlace únicamente debido al enlace al que pertenece por lo tanto esto permite que más información sea analizada y procesada por el usuario. Otro aspecto de suma importancia es que como se estableció con anterioridad es que con el nuevo prototipo de *snippet* la

relación que se sigue ahora es que el elemento que es más visualizado también es el elemento que los usuarios más tiempo pasan observando. Aquí se puede definir que el iniciar con la descripción guía a los usuarios a leer detalladamente lo que esta contiene por lo tanto la mayor información que procesarán provendrá de la descripción y de ahí basarán su decisión para elegir un enlace.

5. CONCLUSIONES

La pregunta de investigación que se estableció al inicio de este proyecto fue si ¿alterar el contenido de los *snippets* impacta el proceso de toma de decisiones de los usuarios para llevar a cabo las tareas de búsqueda de información? De ser así ¿De qué manera? La respuesta a esta pregunta fue ratificada porque se demostró que alterar la estructura básica del *snippet* impacta de manera significativa la manera en que los usuarios llevan a cabo el proceso de toma de decisiones. El prototipo de *snippet* presentado permite que los participantes se adentren al contenido del sitio de manera más adecuada a la que tradicionalmente están acostumbrados, es decir que el participante deja de lado el estigma de ignorar sitios web que reconoce o sabe cómo está estructurada la información dentro de estos. Esto crea una situación donde el usuario no se deja guiar por el prestigio del sitio o el título del artículo, si no por el contenido que realmente se encuentra dentro de este. Al tener que procesar más información la expectativa sería que las tareas de búsqueda de información duraran más en tiempo a comparación de los *snippets* tradicionales los cuales les es común a los usuarios descartar ciertos enlaces con base a la estructura que contiene el *snippet* tradicional. Sin embargo, es fácil inferir que la razón por la cual el prototipo de *snippet* presenta resultados más satisfactorios en cuestiones de tiempo en contraste al tradicional es debido a que el prototipo presenta muchos menos casos de positivos falsos y falsos positivos. Con anterioridad planteamos la cuestión “¿De las consultas que realiza el usuario existen casos de falsos positivos y positivos falsos tanto para el *snippet* tradicional como el prototipo?” los casos de positivos falsos (seleccionar un *snippet* creyendo que es el indicado, pero no resulta ser el caso) y falsos positivos (el contenido dentro del sitio es el que el usuario requiere pero el *snippet* no lo refleja) no suceden dentro del prototipo del *snippet*. Esto se debe a que como ya se estableció el primer elemento que el usuario enfoca su atención es la descripción, donde

se encuentra el extracto de texto relevante a la consulta que realizó el usuario, debido a esto el usuario puede identificar de manera más veloz si un sitio le es o no relevante en base al contenido que se encuentra dentro del sitio. A diferencia del caso tradicional, donde el usuario se deja llevar más por el título o nombre del sitio y creando la posibilidad de descartar algún sitio que contenía la información que justo estaba buscando meramente por el prestigio o estigma que se tiene hacia éste y de igual manera creer que sitios con prestigio contarán con la información que requiere.

REFERENCIAS (bibliografía)

[1] Mari-Carmen Marcos and Cristina. González-Caro. Comportamiento de los usuarios en la página de resultados de los buscadores. Un estudio basado en eye tracking. *El profesional de la información*, 2010.

[2] Dengel A. Buscher, G. and L. van Elst. Eye movements as implicit relevance feedback. *Extended Abstracts on Human Factors in Computing Systems. ACM*, 2008.

[3] Kojo I. Simola J. & Kaski S. Salojärvi, J. ¿Can relevance be inferred from eye movements in information retrieval. *Proceedings of WSOM '03, Workshop on Self-Organising Maps*, 2003.

[4] K. S. Jacob, R. J. K. & Karn. Eye tracking in human-computer interaction and usability research: ready to deliver the promises. *The mind's eye: cognitive and applied aspects of eye movement research*, 2003.

[5] H. D. Crane and C. M. Steele. Generation-v dual-purkinje-image eyetracker. *Applied Optics*, 1985.

[6] Stephan Spencer. Anatomy of a google snippet. *Search engine land*, 2010.

[7] Review of Web Crawlers with Specification and Working. Review of web crawlers with specification and working. *International Journal of Advanced Research Computer and Communication Engineering*, 2016.

[8] E Ramirez. El método experimental. *Univerrsidad de Jaén*, 2010.

ANEXOS

Productos generados

- 1 capítulo de libro publicado
- 1 estudiante del programa de Ingeniería de Software graduado con su proyecto de titulación, base del presente reporte.
- 1 estudiante de licenciatura en proceso de desarrollo de proyecto de titulación intracurricular
- 2 estudiantes como prestadores de servicio social

****Nota: El reporte técnico tendrá una extensión mínima de 5 cuartillas y máxima de 30, a espacio y medio.**

CONSIDERACIONES:

- Los reportes deben estar escritos en español o en inglés.
- Se debe entregar en formato Word acorde a este formato.
- El texto debe ser escrito en hoja tamaño carta a espacio y medio, y los márgenes deberán encontrarse al menos a una pulgada (2.54 cm). La totalidad del texto debe escribirse en minúsculas, utilizando las mayúsculas sólo al principio de las oraciones y para los títulos de capítulos.
- Se recomienda usar el tipo de letra Arial tamaño 10, o Times new Roman tamaño 12.
- Todas las páginas deben estar numeradas en secuencia comenzando desde la portada.
- La extensión total del texto es de un mínimo de 5 cuartillas y un máximo de 30 cuartillas, con un interlineado de espacio y medio.
- Archivos de Excel de tablas y gráficas deben ser adjuntados al reporte enviado electrónicamente.
- Las figuras, fotografías y tablas, serán insertadas en el cuerpo del texto y numeradas en forma consecutiva comenzando con 1 y de manera independiente de las tablas. El número y descripción de la figura, tabla, etc., deberá colocarse antes de la misma.
- Se recomienda evitar el uso de sombras y líneas punteadas que no permitan una legibilidad clara de imágenes.
- Las fórmulas y ecuaciones deben hacerse con un editor de ecuaciones como el que viene en Word. Estarán centradas y separadas del texto. La numeración será consecutiva comenzando con 1. El número de la fórmula deberá encerrarse entre paréntesis y colocarse a la derecha de la fórmula lo más cercano posible al margen derecho.
- Las referencias bibliográficas en el texto deben ser en cualquier estilo reconocido como APA, MLA, ISO, etc.
- Los anexos se colocarán al final del documento después de la bibliografía, utilizando caracteres alfabéticos para distinguirlos: Anexo A, Anexo B, etc. La información

contenida en los anexos es importante pero no indispensable para la comprensión del trabajo. Se recomienda colocar en los anexos mapas, fotografías, tablas, desarrollos matemáticos, diagramas, etc.