



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ESTUDIO DEL COMPORTAMIENTO FISIOLÓGICO DE USUARIOS DE LA WEB  
COMO DETERMINANTE DEL GÉNERO, BASADO EN HERRAMIENTAS DE EYE  
TRACKING, ELECTROENCEFALOGRAMA Y TÉCNICAS DE MINERÍA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

EDUARDO DANIEL NEIRA REYES

PROFESOR GUÍA:  
SR. JUAN DOMINGO VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:  
SR. FRANCISCO MOLINA JARA  
SR. ALBERTO CABEZAS BULLEMORE

SANTIAGO DE CHILE  
2016

# Resumen Ejecutivo

RESUMEN DE LA MEMORIA  
PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL  
POR: EDUARDO NEIRA REYES  
FECHA: 23/12/2015  
PROF. GUÍA: SR. JUAN VELÁSQUEZ

Este Trabajo de Título, tiene como objetivo general, examinar diferencias en el comportamiento fisiológico entre hombres y mujeres cuando navegan por la Web, mediante la aplicación de herramientas de eye tracking, electroencefalograma y técnicas de minería de datos.

Este objetivo se quiere lograr en el marco del proyecto AKORI, un proyecto elaborado por WIC Chile Research Centre y el Laboratorio de Neurosistemas de la Universidad de Chile, que pretende desarrollar una plataforma informática para mejorar la estructura y contenido de sitios web mediante el uso de web intelligence y variables fisiológicas. En este proyecto ya se han elaborado 6 memorias de investigación de estudiantes de pregrado del DII de la U. de Chile, de las cuales se han podido validar o rechazar distintas hipótesis que relacionan la dilatación pupilar y la actividad bioeléctrica cerebral con el entorno web y que han abierto líneas de investigación.

Para esta nueva etapa del proyecto AKORI, se quiere que los experimentos representen de mejor manera el comportamiento web de las personas, utilizando sitios web reales y considerando el objetivo por el cual un usuario visitaría un sitio web. Este objetivo puede ser muy distinto dependiendo de las características de las personas, por lo cual estas características deben ser tomadas en cuenta al momento de realizar un estudio.

Dado lo anterior, se ha elegido el género como variable de segmentación ya que ha sido utilizada como base para la segmentación de mercado. Por esto, la hipótesis de investigación señala que *“es posible diferenciar por sexo a usuarios de la Web, utilizando su comportamiento pupilar y su actividad bioeléctrica cerebral”*.

Para cumplir con el objetivo, primero se analizó el estado del arte de las materias relacionadas. Luego, se diseñó e implementó un experimento en el laboratorio ubicado en el DII de la U. de Chile, en el que participaron 10 hombres y 10 mujeres. Finalmente se analizaron los datos obtenidos del experimento.

En la fase de análisis, primero se hizo un estudio exploratorio del comportamiento ocular, llegando a conclusiones consistentes con la literatura. Luego se realizó el proceso KDD, utilizando 5 modelos de predicción. De estos modelos, el con mejor desempeño fue el Árbol de Decisión, obteniendo un *Accuracy* superior al 70%, lo que permitió validar la hipótesis de investigación.

Finalmente, se concluye que se cumple con el objetivo de este trabajo, lo que puede traer implicancias para el desarrollo de la plataforma informática pretendida por el proyecto AKORI.

# Agradecimientos

En primer lugar, quisiera agradecer al profesor Juan Velásquez por depositar su confianza en sus alumnos y por darme esta oportunidad de elaborar el Trabajo de Título junto a él.

Quisiera agradecer también a cada miembro del grupo WIC, en especial a los de la “salita”, quienes me acompañaron en casi todo el proceso y dispusieron de su ayuda y “buena onda” en todo momento.

También agradezco al profesor Pedro Maldonado, por darse el tiempo de compartir su conocimiento conmigo y resolver mis dudas.

Agradezco a cada persona que se ofreció voluntariamente para realizar el experimento, lo cual fue fundamental para el desarrollo de mi trabajo.

Y finalmente, agradezco a mis amigos y a mi familia por su preocupación y apoyo en todo el proceso. Y en especial agradezco a mi hermano Álvaro por su importante colaboración en el finiquito de este trabajo.

Eduardo Daniel Neira Reyes

# Tabla de Contenido

Resumen Ejecutivo.....	i
Capítulo 1: Introducción .....	1
1.1. Antecedentes Generales .....	1
1.2. Descripción del Proyecto.....	3
1.3. Objetivos.....	4
1.3.1. Objetivo General .....	5
1.3.2. Objetivos Específicos .....	5
1.4. Metodología.....	5
1.5. Contribución de la Memoria y Alcances .....	6
1.6. Estructura del Informe.....	6
Capítulo 2: Marco Conceptual.....	8
2.1. Movimientos Oculares y su Medición.....	8
2.1.1. El Ojo Humano y la Atención Visual .....	8
2.1.2. Eye Tracking .....	12
2.2. Actividad Bioeléctrica Cerebral y su Medición .....	13
2.2.1. El Cerebro y las Neuronas .....	13
2.2.2. Potenciales Eléctricos [10].....	14
2.2.3. Electroencefalografía .....	14
2.3. Diferencias de Género .....	16
2.3.1. Diferencias en el Comportamiento Web .....	16
2.3.2. Electroencefalograma .....	19
2.3.3. Resumen de la Sección .....	19
2.4. Proceso KDD .....	19
2.4.1. Minería de Datos.....	20
Capítulo 3: Experimento .....	30
3.1. Diseño del Experimento.....	30
3.1.1. Instrumentación .....	30
3.1.2. Grupo Experimental .....	33

3.1.3.	Tarea.....	34
3.1.4.	Sitio Web.....	34
3.2.	Implementación del Experimento .....	36
3.2.1.	Grupo Experimental .....	36
3.2.2.	Sitio Web.....	37
3.2.3.	Procedimiento.....	40
3.2.4.	Datos .....	43
Capítulo 4:	Análisis de Resultados.....	44
4.1.	Comportamiento Ocular .....	44
4.1.1.	Detección de Fijaciones .....	45
4.1.2.	Resultados.....	48
4.1.3.	Conclusiones .....	53
4.2.	Proceso KDD .....	53
4.2.1.	Selección.....	54
4.2.2.	Pre – Procesamiento.....	55
4.2.3.	Transformación.....	58
4.2.4.	Minería de Datos.....	62
4.2.5.	Evaluación e Interpretación .....	69
Capítulo 5:	Conclusiones.....	72
5.1.	Conclusiones Generales .....	72
5.2.	Recomendaciones y Trabajo Futuro .....	74
Bibliografía	.....	75
Anexos	.....	82
A	Preguntas del Experimento.....	82
B	Consentimiento Informado.....	83
C	Microsoft Excel con Preguntas.....	86

# Índice de Tablas

Tabla 2.1: La Evolución y la Evidencia en el Comportamiento Web.....	18
Tabla 3.1: Especificaciones técnicas del Emotiv EPOC Neuroheadset .....	33
Tabla 3.2: Criterios para la Clasificación de Sitios Web.....	35
Tabla 3.3: Grupo Experimental .....	36
Tabla 3.4: Sitios Web.....	37
Tabla 4.1: Cantidad de Objetos Seleccionados .....	47
Tabla 4.2: Comportamiento Ocular del Hombre en Sitio Web Neutro.....	49
Tabla 4.3: Comportamiento Ocular de la Mujer en Sitio Web Neutro.....	49
Tabla 4.4: Comportamiento Ocular del Hombre en Sitio Web Masculino .....	50
Tabla 4.5: Comportamiento Ocular de la Mujer en Sitio Web Masculino .....	51
Tabla 4.6: Comportamiento Ocular del Hombre en Sitio Web Femenino.....	52
Tabla 4.7: Comportamiento Ocular de la Mujer en Sitio Web Femenino.....	52
Tabla 4.8: Duración mínima y cantidad de fijaciones.....	59
Tabla 4.9: Niveles de Descomposición con Transformada de Wavelet Discreta .....	61
Tabla 4.10: Cantidad de fijaciones a considerar por tipo de sitio .....	62
Tabla 4.11: Resultado de diferencias estadísticas, sitio neutro .....	64
Tabla 4.12: Resultado de diferencias estadísticas, sitio masculino .....	65
Tabla 4.13: Resultado de diferencias estadísticas, sitio femenino .....	67
Tabla 4.14: Resultados en sitio neutro.....	70
Tabla 4.15: Resultados en sitio masculino .....	70
Tabla 4.16: Resultados en sitio femenino .....	70
Tabla 1: Preguntas del Experimento .....	82

# Índice de Figuras

Figura 2.1: Globo Ocular .....	9
Figura 2.2: Morfología de una Neurona .....	14
Figura 2.3: Posición de Electrodo del EEG .....	15
Figura 2.4: Proceso KDD.....	20
Figura 2.5: Taxonomía de la Minería de Datos .....	21
Figura 2.6: Frontera de Decisión .....	24
Figura 2.7: Transformación con Función Kernel .....	25
Figura 2.8: Ejemplo de Árbol de Decisión.....	27
Figura 2.9: Matriz de Confusión .....	28
Figura 3.1: Pantalla del Laboratorio con Eye Tracker .....	31
Figura 3.2: Emotiv EPOC Neuroheadset .....	32
Figura 3.3: Posición de los Electrodo según el Sistema 10-20 .....	32
Figura 3.4: werd.com .....	38
Figura 3.5: wikipedia.org .....	39
Figura 3.6: designsponge.com .....	39
Figura 3.7: Humectación de Electrodo.....	41
Figura 3.8: Software de Emotiv .....	41
Figura 3.9: Pupila detectada vista en Sofey .....	42
Figura 4.1: Screenshot con Objetos Marcados.....	46
Figura 4.2: Screenshot con Objetos Seleccionados .....	47
Figura 4.3: Ejemplo de 5 segundos de registro de los 14 canales.....	55
Figura 4.4: Ejemplo de señal sin registro del Eye Tracker.....	56
Figura 4.5: Ejemplo de señal corregida con Interpolación .....	56
Figura 4.6: Señal con sacadas y señal sin sacadas (corregida).....	57
Figura 4.7: Corrección de Alta Frecuencia.....	57
Figura 1: Consentimiento Informado, Parte 1 .....	83
Figura 2: Consentimiento Informado, Parte 2 .....	84
Figura 3: Consentimiento Informado, Parte 3 .....	85
Figura 4: Sujeto en quinta pregunta del experimento.....	86
Figura 5: Sujeto con experimento completado.....	86

# Capítulo 1: Introducción

En este capítulo se introduce al lector en el tema de memoria. Se incluyen los antecedentes generales, lo que es el contexto en cual se encuentra este proyecto. Se describe el proyecto de memoria y su justificación. Luego se declaran los objetivos de este proyecto, su metodología, sus alcances y resultados esperados. Finalmente se introduce al lector en cómo se estructura y continúa este informe.

## 1.1. Antecedentes Generales

En los últimos años, se ha visto cómo la Web ha sido utilizada como canal de comunicación por las organizaciones o empresas para lograr diferentes tipos de objetivos, como dar a conocer una propuesta de valor, realizar publicidad o vender productos. Se ha visto que el uso de este medio de comunicación ha ido en aumento debido al constante crecimiento de usuarios de la Web [1].

Las características de los sitios en la Web serán determinantes en las acciones que los usuarios realicen, por lo que es entonces importante entender el comportamiento de los usuarios de la Web si se quiere utilizar este importante canal de manera exitosa.

En particular, la Web Intelligence Consortium Chile Research Centre quiere aportar al estudio del comportamiento del usuario en la Web a través de un proyecto FONDEF (AKORI: Advanced Kernel for Ocular Research and web Intelligence), en el que se pretende desarrollar una plataforma informática para mejorar la estructura y contenido de sitios web mediante el uso de web intelligence y herramientas de análisis de exploración visual (eye tracking) y actividad bioeléctrica cerebral (electroencefalograma). Este proyecto es además compartido con el Laboratorio de Neurosistemas de la Facultad de Medicina de la Universidad de Chile.

Previamente al proyecto AKORI, se realizaron estudios que dan origen a los conceptos que se aplican actualmente en el WIC en relación al análisis del comportamiento del usuario en la Web [2], [3]. L. E. Dujovne y J. D. Velásquez diseñaron una metodología para identificar los website keyobject. Un website object es definido como “cualquier grupo estructurado de palabras o un recurso visual que pertenece a una página web que tiene meta-data y que describe el contenido de ésta”. Luego, un website keyobject es definido como “cualquier objeto que capture la atención del usuario y caracterice el contenido del sitio”[4]. Esta metodología se basa en el tiempo en que una persona dedica al observar un website object y en el resultado de encuestas hechas al usuario, lo que es una fuente subjetiva de información.



Pero luego, esta metodología es mejorada cuando Gonzales y Velásquez introducen la tecnología de eye tracking para reemplazar a las encuestas [5]. Con ello, se da el puntapié inicial para seguir descubriendo conocimiento acerca de este tema utilizando estas herramientas.

En el proyecto AKORI se han elaborado 7 memorias a la fecha, siendo 6 de investigación, sólo de estudiantes de pregrado del Departamento de Ingeniería Industrial (DII) de la U. de Chile. Estas han consistido principalmente en investigar el aporte de las tecnologías de eye tracking y electroencefalograma a la identificación de website keyobject y a la identificación de zonas de dilatación pupilar.

Martínez fue el primero de estos memoristas. El, en [4] integra la dilatación pupilar como variable a la metodología propuesta anteriormente para la identificación de website keyobject.

Se basó en muestras que consistían en registrar la navegación de 23 usuarios con herramientas de eye tracking en el sitio web del programa de MBA del Departamento de Ingeniería Industrial de la U. de Chile.

Obtuvo una lista de website keyobject, la cual fue evaluada por un experto, llegando a la conclusión de que la dilatación pupilar es un aporte a la metodología previa.

En la misma línea de investigación, Dupré en [6] utiliza medidas de centralidad de teoría de grafos para la identificación de website keyobject. El utiliza los datos capturados por Martínez y obtiene conclusiones similares, es decir, las medidas de centralidad son un aporte a la identificación de website keyobject. En particular, la medida del vector propio es la que tuvo mejor evaluación.

Slanzi en [7], integra la actividad bioeléctrica cerebral medida por un electroencefalograma (EEG) a la identificación de website keyobject. El realiza el mismo método para tomar muestras de Martínez a 20 usuarios distintos, obteniendo resultados favorables.

En otra línea de investigación, Jadue en [8] estudia la presencia de patrones de comportamiento de la pupila en usuarios web al momento de tomar una decisión. El momento de toma de decisión la define como el momento en que un usuario hace click cuando navega en la Web.

En su experimento, muestra una grilla de 9 imágenes de la misma categoría al usuario y el usuario debe hacer click en una de las imágenes. Por ejemplo, se muestra una grilla de imágenes de distintos tipos de mesas y el usuario debe elegir una imagen bajo su propio criterio y hacer click en la imagen. Para esto utiliza 90 categorías distintas y el registro de 25 usuarios. Finalmente, detecta la presencia de patrones de dilatación pupilar al momento de hacer click en una de las imágenes.

Salas en [9] sigue la misma línea de investigación que Jadue, incorporando el potencial eléctrico medido por el electroencefalograma. El realiza el mismo experimento que Jadue a 18 usuarios distintos y no logra obtener diferencias significativas del potencial eléctrico entre la toma de decisión y la no toma de decisión, por lo que no logra

validar su hipótesis y demostrar el aporte del potencial eléctrico medido en el electroencefalograma.

Por otro lado, Aracena en [10] estudia la existencia de relación entre neurodatos (electroencefalograma), dilatación pupilar y emocionalidad. Su experimento consistió en exponer 180 imágenes con distinto contenido emocional a los usuarios, mientras era registrada su actividad bioeléctrica cerebral, dilatación pupilar y las calificaciones que los usuarios otorgaban a las imágenes.

Sólo se obtuvo el registro de 10 personas y no se logró obtener una relación significativa entre las variables estudiadas, por lo que no logró validar su hipótesis.

Finalmente, se pretende que con este proyecto de memoria se descubra nuevo conocimiento y que en conjunto a las otras memorias, se avance en la misión de lograr el objetivo del proyecto AKORI.

## **1.2. Descripción del Proyecto**

Como se ha mencionado, se han obtenido resultados favorables en las memorias anteriores en la utilización de ciertas herramientas. Se han validado las hipótesis de que la dilatación pupilar, la actividad bioeléctrica cerebral y las medidas de centralidad de teoría de grafos pueden ser una fuente importante de información para detectar website keyobject. Además, se encontró relación entre la dilatación pupilar y la toma de decisión (hacer click) en un sitio web.

Sin embargo, algunos experimentos pueden carecer de realismo. En el experimento de navegación por el sitio del programa de MBA de la Universidad de Chile, no se utiliza el sitio web real, sino que son imágenes que simulan el sitio web. Además el usuario visita todas las páginas del sitio con un orden definido. Esto es así por limitaciones de los recursos del laboratorio en el cual se realizó el experimento.

Por otro lado, no se contempla el objetivo o el interés por el cual un usuario visita un sitio web. Una persona ingresa a un sitio web con un objetivo en particular, como realizar una compra o buscar información específica. En cambio, en este experimento el usuario se ve “forzado” a observar el sitio del programa de MBA y puede no tener interés en realizar un MBA, por lo que puede haber un sesgo entre el comportamiento del usuario en el experimento y en la realidad. Además, los resultados no han sido discriminados por las características de las personas como la edad o género, lo que también genera un sesgo ya que los intereses y objetivos de las personas pueden ser distintos de acuerdo a ciertas características y como consecuencia su comportamiento en la Web.

Por lo tanto, existe una oportunidad de investigación en la que se pueda comprobar si existen diferencias entre segmentos de patrones de dilatación pupilar u otro indicador fisiológico que sean significativas para efectos de lo que el proyecto AKORI quiere lograr.

Dado lo anterior, este proyecto de memoria consistirá en investigar la presencia de diferencias en el comportamiento pupilar y la actividad bioeléctrica cerebral entre usuarios segmentados por género. Esto debido a que es una característica que tiene fuerte relación con los intereses y el comportamiento que la persona puede tener y Además, el género ha sido utilizado durante muchos años como base para la segmentación de mercado, debido a que es fácilmente identificable y sus segmentos son suficientemente grandes [11].

Se han hecho estudios en los que se ha encontrado que la dilatación pupilar tiene relación con los procesos cognitivos de la persona y con la atracción sexual, por ejemplo. Si una persona es homosexual puede experimentar dilatación pupilar al observar a otra persona del mismo género [12]. Además se ha estudiado que la dilatación pupilar puede tener relación con la probabilidad de hacer click, por lo tanto, la probabilidad de que el usuario se comporte como la organización detrás de un sitio web desea.

Para lograr esto, será fundamental diseñar un experimento que se asemeje lo más posible a la realidad para lograr resultados con la validez necesaria. Y a diferencia de los experimentos anteriores, los experimentos se podrán realizar en el Departamento de Ingeniería Industrial de la U. de Chile y el nuevo equipamiento permitirá utilizar un sitio web real.

Por otra parte, en una investigación de mercado hecha previamente para evaluar la propuesta de la herramienta, se encontró que los potenciales clientes valorarían bastante que la herramienta a construir pudiese entregar resultados segmentados [13]. Entonces este proyecto de memoria le interesa a la organización (Web Intelligence Consortium Chile Research Centre) debido a que es un avance en la implementación de la herramienta.

Por lo tanto, la hipótesis de investigación de esta memoria es:

*“Es posible diferenciar por sexo a usuarios de la Web, utilizando su comportamiento pupilar y su actividad bioeléctrica cerebral”.*

Al validar esta hipótesis, existiría evidencia empírica de que la probabilidad de cambios fisiológicos en un sitio web sea distinta para los segmentos a estudiar, por lo que la plataforma informática debiera discriminar entre estos segmentos. Además se abriría un nuevo camino de investigación en el proyecto.

Si no se valida las hipótesis, no se continuaría con esta línea de investigación y se ahorrarían esfuerzos.

### **1.3. Objetivos**

El Objetivo General y Objetivos Específicos de este proyecto se declaran a continuación.

### **1.3.1. Objetivo General**

Examinar diferencias en el comportamiento fisiológico entre hombres y mujeres cuando navegan por la Web, mediante la aplicación de herramientas de eye tracking, electroencefalograma y técnicas de minería de datos.

### **1.3.2. Objetivos Específicos**

1. Investigar y analizar el estado del arte sobre la atención visual, dilatación pupilar, eye tracking, actividad bioeléctrica cerebral, electroencefalograma, diferencias de género y técnicas de minería de datos.
2. Diseñar e implementar experimentos con herramientas de eye tracking y electroencefalograma para la obtención de los datos a analizar.
3. Determinar las diferencias en el comportamiento pupilar y en el potencial eléctrico del cerebro entre usuarios de la Web de distinto género, mediante el análisis de datos de experimentos y técnicas de minería de datos.
4. Determinar la validez de la hipótesis de investigación y concluir.

## **1.4. Metodología**

Para lograr los objetivos específicos declarados, se propone la siguiente metodología.

1. **Investigación del estado del arte:** se deberá investigar y analizar la información existente a la fecha acerca de las materias relacionadas, que corresponden a la atención visual, dilatación pupilar, actividad bioeléctrica cerebral, eye tracking, electroencefalograma y las diferencias que existen acerca de los temas relacionados entre géneros. Es también necesario investigar y analizar las herramientas de minería de datos que existen en la actualidad, para seleccionar los métodos que se adecúen correctamente al problema.
2. **Diseñar experimentos e implementar experimentos:** con la investigación y análisis del estado del arte, se obtendrá el conocimiento necesario para diseñar un experimento que permita validar o rechazar la hipótesis de investigación. Este experimento deberá utilizar herramientas de eye tracking, electroencefalograma y deberá entregar resultados que permitan su análisis.

Una vez que se diseñe el experimento, se deberá llevar a cabo, para la cual se dispone de un laboratorio en el DII de la Universidad de Chile. Para esto se necesita tener conocimiento de la utilización de los recursos y disponer de una cantidad suficiente de personas.

3. **Analizar resultados de experimento:** los experimentos entregarán información que deberá ser tratada y analizada mediante técnicas y modelos de minería de datos. En esta etapa también será fundamental conocer y dominar herramientas computacionales que permitan el análisis.
4. **Concluir:** finalmente se deberá concluir en base al análisis de los resultados y demostrar o rechazar la hipótesis. También se tendrá que realizar una discusión acerca de los resultados y proponer trabajo futuro en el marco del proyecto AKORI.

## 1.5. Contribución de la Memoria y Alcances

De los objetivos específicos se desprenden las siguientes contribuciones esperadas:

- Un marco conceptual que represente el estado del arte de las materias mencionadas y que facilite la investigación.
- Experimento diseñado para el laboratorio del DII, lo que incluye protocolo y recursos necesarios.
- Datos obtenidos de experimentos.
- Métodos y códigos utilizados en análisis y resultados correspondientes.
- Validación o rechazo de hipótesis de investigación.
- Conclusiones y discusión de resultados.
- Trabajo futuro en el marco del proyecto AKORI.

El alcance de esta memoria es el desarrollo de la validación o rechazo de la hipótesis de investigación que se plantea, lo que está sujeto a los resultados de los experimentos. Este proceso se limita a los recursos disponibles para la experimentación en el nuevo laboratorio ubicado en el Departamento de Ingeniería Industrial de la U. de Chile.

Finalmente, este proyecto no debe superar el período correspondiente al Trabajo de Título.

## 1.6. Estructura del Informe

Este informe se organiza en cinco capítulos, los que se describen brevemente a continuación.

El primer capítulo es la Introducción, en el que se contextualiza al lector sobre este proyecto de memoria. Se exponen los antecedentes y la justificación de este tema de memoria. Luego se declaran los objetivos, la metodología, los resultados esperados y los alcances.

El segundo capítulo es el Marco Teórico, en el que se explican al lector los temas estudiados y utilizados para llevar a cabo este proyecto de memoria. Los temas que se exponen son: Movimientos Oculares y su Medición; Actividad Bioeléctrica Cerebral y su Medición; Diferencias de Género; y el proceso KDD.

En el tercer capítulo se presenta lo que tiene relación al diseño e implementación del experimento. Se presenta lo que se requiere para llevar a cabo el experimento y los recursos y métodos implementados.

El cuarto capítulo es el Análisis de Resultados. En este capítulo se explica el tratamiento que se lleva a cabo para obtener resultados válidos, se presentan los resultados finales y se discute en base a ellos.

Finalmente, en el quinto capítulo se concluye en base al trabajo de esta memoria, se exponen recomendaciones para continuar con el trabajo y posible trabajo futuro para continuar con el proyecto AKORI.

# Capítulo 2: Marco Conceptual

En este capítulo se presentan los conceptos necesarios para que este proyecto de memoria se pudiera llevar a cabo y para que el lector tenga los conocimientos suficientes para continuar con la lectura de este informe.

Las secciones de este capítulo son cuatro: Movimientos Oculares y su Medición; Actividad Bioeléctrica Cerebral y su Medición; Diferencias de Género; y Proceso KDD (*Knowledge Discovery in Databases*).

## 2.1. Movimientos Oculares y su Medición

El estudio de la pupila y sus movimientos son parte fundamental de este proyecto de título, por lo que es necesario también comprender el contexto de ello. Es por esto que este capítulo comienza con una breve descripción del ojo humano, sus movimientos y la relación con la atención visual, para finalizar con lo que es la dilatación pupilar y el registro de los movimientos oculares o eye tracking.

### 2.1.1. El Ojo Humano y la Atención Visual

#### 2.1.1.1. Sistema Visual

El sistema visual está compuesto por el globo ocular, los anexos oculares, la cavidad orbitaria, y la vía óptica [14]. La pupila se encuentra en lo que es el globo ocular, que corresponde a una estructura esférica que mide aproximadamente 2,5 centímetros de diámetro en los humanos y está compuesta por tres capas concéntricas: esclerocórnea, úvea y retina. Las componentes del globo ocular se describen a continuación [15], [16].

- **Esclera:** forma parte de la esclerocórnea, que es la capa más externa del glóbulo ocular. Esta componente es una membrana fibrosa que se encarga proteger el globo ocular.
- **Córnea:** también perteneciente a la esclerocórnea. Es una estructura transparente cuya función principal es la óptica, proporcionando gran parte de las cualidades refractivas del ojo.

- **Coroides:** esta se encuentra en la úvea que es la capa media del globo ocular. Es rica en células pigmentarias confiriéndole un rol de pantalla a la luz y tiene una función nutritiva dada su naturaleza vascular.
- **Cuerpo ciliar:** componente de la capa media. Esta está formada por los procesos ciliares, los cuales están encargados de la secreción del líquido que rellena el humor acuoso. También está los músculos ciliares que varían la curvatura del cristalino para un mejor enfoque.
- **Iris:** se encuentra en la zona más anterior de la capa media. Tiene forma de disco y presenta un orificio en su centro llamado pupila.
- **Pupila:** como se mencionó, es un orificio presente en el centro del iris, cuya función principal es regular la cantidad de luz que ingresa en el ojo, variando su tamaño según la intensidad de luz.
- **Cristalino:** se encuentra inmediatamente detrás del iris. Es una lente biconvexa que es capaz de variar su curvatura, por lo que tiene poder de convergencia variable.
- **Retina:** corresponde a la capa más interna del glóbulo ocular. Es donde se inicia el proceso de la visión, ya que es la parte del sistema nervioso destinada a recoger, elaborar y transmitir las sensaciones visuales. Su punto de máxima sensibilidad es una hendidura llamada fovea, que es donde se encuentra una mayor concentración de conos y bastones.

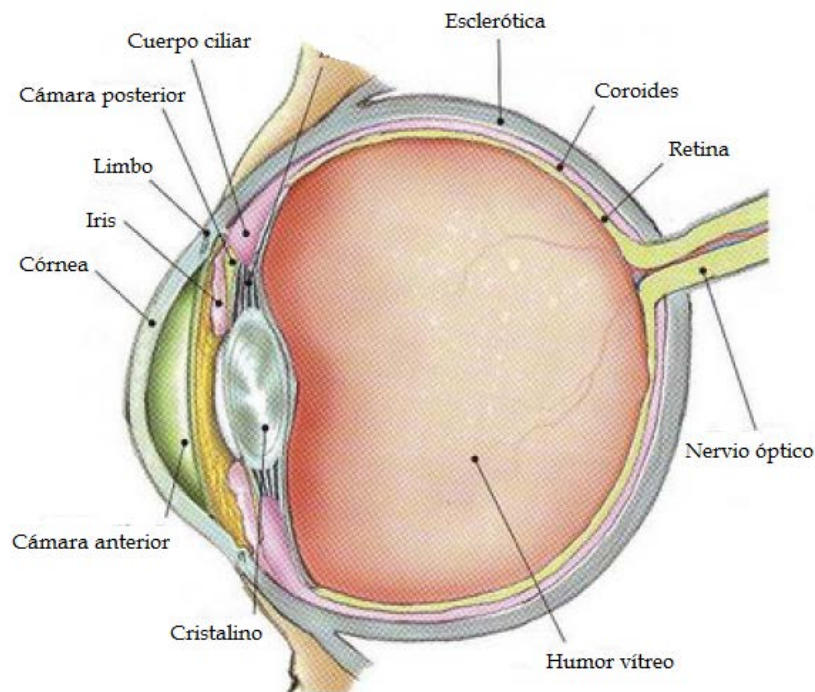


Figura 2.1: Globo Ocular  
Fuente: Imagen de [9]



### **2.1.1.2. Movimientos Oculares y Atención Visual**

El ojo humano realiza distintos tipos de movimientos para poder crear la imagen mental del escenario de acuerdo a los estímulos visuales. Estos movimientos se pueden agrupar en tres categorías principales [4].

La primera categoría corresponde a los movimientos realizados por el ojo para seguir el rastro de un objeto en movimiento. La segunda categoría corresponde a los movimientos de convergencia, que son aquellos movimientos realizados por ambos ojos para mantener la visión binocular. Y la última corresponde a los movimientos sacádicos o sacadas, que ocurren cuando la persona escanea una escena, tratando de identificar las partes más interesantes, para luego enfocarse en ellas.

Cuando una persona escanea una escena, la información visual es generalmente percibida solamente por las fijaciones y no las sacadas. Las fijaciones ocurren cuando una persona observa un elemento de la escena fijamente, por un lapso de tiempo no menor a los 100 milisegundos [17].

En relación a la atención visual, hay que mencionar que la visión corresponde a un proceso cíclico compuesto por los siguientes pasos:

1. Dado el estímulo, el encuadre completo es visto a través de la visión periférica que es en baja resolución. En esta etapa se identifican los elementos que atraen la atención de la persona.
2. Los ojos se dirigen hacia la región donde fue atraída la atención.
3. La atención se centra en esa región y se percibe en alta resolución, lo que es permitido a través de la fovea.

Por lo anterior, se habla de que la frecuencia de fijación en una región puede ser un indicador de importancia, mientras que la duración de la fijación puede reflejar la complejidad y dificultad de la escena visual [18].

### **2.1.1.3. Dilatación Pupilar**

Corresponde al incremento del tamaño de la pupila debido a distintos estímulos. Existe dilatación o contracción de la pupila principalmente cuando se presentan variaciones de la luminosidad en el entorno. También esta respuesta se presenta cuando la persona trata de enfocar la mirada [19].

Los cambios en el tamaño de la pupila son controlados por dos músculos, el dilatador y el esfínter, que son controlados por la rama simpática y parasimpática del sistema nervioso central respectivamente. Un aumento en la actividad simpática activa el músculo dilatador, generando la dilatación de la pupila. Al contrario, un aumento en la actividad parasimpática, activa el esfínter, generando la contracción de la pupila [20]. Estos cambios pueden alterar el tamaño de la pupila desde 1mm hasta los 9 mm [19].

También la pupila puede experimentar cambios en su tamaño cuando se presentan estímulos positivos o negativos, en procesos de aprendizaje o procesamiento de información. Esto demuestra una relación entre la dilatación pupilar y los procesos cognitivos y el comportamiento de la persona [19].

Por ejemplo, Hess y Polt en [21] estudiaron la relación entre la dilatación pupilar y la actividad mental a través de un problema de multiplicación matemática. Ellos solicitaron a los participantes del experimento que calcularan mentalmente una multiplicación matemática, y se observó un incremento hasta el 30% del tamaño de la pupila entre el momento previo a la pregunta y el momento entre la pregunta y la respuesta, disminuyendo su tamaño luego de entregar una respuesta.

Beatty en [22] concluye que la dilatación pupilar es un importante indicador del esfuerzo mental involucrado en resolución de tareas, lo que involucra memoria, procesamiento del lenguaje, razonamiento y percepción.

Hahneman y Beatty en [23] realizaron un experimento en el cual los participantes debían detectar diferencias entre dos tonos. Mientras más parecidos eran los tonos y más difícil era percibir la diferencia, mayor dilatación pupilar se experimentaba.

De igual forma en un contexto web, en un experimento en el que se midió el tamaño de la pupila frente a búsquedas de información en internet con distinta dificultad, se encontró que la dilatación pupilar era mayor en las búsquedas que requerían mayor esfuerzo cognitivo [24].

Por otro lado, la dilatación pupilar se ha relacionado con las preferencias visuales y en particular con las preferencias sexuales. Por ejemplo, Hess y Polt en [25] concluyen que las mujeres experimentan mayor dilatación pupilar cuando observan fotos de hombres y viceversa. Luego Hess, Seltzer y Shlien en [12] encuentran que hombres homosexuales experimentan mayor dilatación pupilar cuando ven fotos de hombres que al ver fotos de mujeres, ocurriendo lo contrario en hombres heterosexuales.

Actualmente, algunos estudios ponen en duda las conclusiones de Hess que concluyen que la dilatación pupilar está dominada por la valencia del estímulo. Estos estudios clasifican a las emociones en un espacio de dos dimensiones, en el que una dimensión representa la valencia de la emoción y la otra representa la excitación provocada. La valencia va desde algo desagradable a placentero y la excitación va desde la tranquilidad a la excitación [26]. Entonces, a diferencia de lo que concluye Hess, se ha encontrado evidencia de que la valencia tiene poca o nula relación con la dilatación pupilar.

Por ejemplo, en [20] concluye que existe dilatación pupilar tanto con estímulos visuales agradables como desagradables. En [27] se concluye que existe mayor dilatación pupilar en estímulos emocionales negativos y positivos que con estímulos neutros. Similares resultados se obtienen en [28], donde el autor afirma que Hess se debió haber enfocado en la excitación y no en la valencia.

Entonces, se ve que la pupila y su relación con los procesos cognitivos ha sido una materia bastante estudiada y consolidada. Sin embargo, conclusiones contradictorias

motivan a continuar con el estudio de este indicador psicofisiológico y las causas de este. Por lo demás, no cabe duda de que sería interesante estudiar su relación con las características de las personas.

### **2.1.2. Eye Tracking**

Consiste en el uso de diversos dispositivos que permiten medir los movimientos oculares de una persona [29]. Por lo tanto se puede saber hacia dónde está mirando una persona durante todo el intervalo de tiempo en el que se estén utilizando los dispositivos. Y es gracias a esta tecnología, que es posible medir el tamaño de la pupila y sus variaciones en el tiempo.

Para el uso de eye tracking, existen distintas técnicas que difieren en la precisión de la medición, y se pueden agrupar en las técnicas que miden la posición del ojo respecto a la posición de la cabeza y las técnicas que miden la orientación del ojo en el espacio. Algunas técnicas son [4], [29]:

- **Electro-oculografía (EOG):** consiste en utilizar electrodos posicionados alrededor de los ojos para medir la diferencia del potencial eléctrico de la piel. Esta técnica mide los movimientos del ojo con respecto a la posición de la cabeza, por lo que no es muy útil para medir la mirada, a menos que se utilicen herramientas para medir además el movimiento de la cabeza.
- **Lente de contacto:** consiste en utilizar un objeto que se monta sobre un lente de contacto. Este es uno de los métodos más precisos para medir los movimientos del ojo, pero a la vez uno de los métodos más invasivos.
- **Foto-oculografía o Video-oculografía:** incluye las técnicas que miden las diferentes características del ojo durante los movimientos de rotación y traslación, como la forma de la pupila o el reflejo de la córnea. Pero generalmente no entregan medidas sobre el punto de atención.
- **Video pupilar:** se caracteriza por proveer una medida para el punto de la mirada. Para esto, se requiere tener la cabeza fija, para que la posición de los ojos coincida con el punto de la mirada. Luego se utilizan cámaras para procesar el punto de la mirada en tiempo real. Se utiliza la posición de la reflexión corneal para estimar la posición de la córnea y la pupila y así obtener el punto de mirada de la persona.

Para este proyecto de título, es esta última técnica la que utilizará para medir el comportamiento pupilar del sujeto en los experimentos.

## 2.2. Actividad Bioeléctrica Cerebral y su Medición

Corresponde a potenciales eléctricos del cerebro generados por la actividad neuronal. Esta es la otra fuente de información utilizada en este trabajo para detectar diferencias entre género y puede ser medida por el electroencefalograma (EEG), la cual es utilizada en los experimentos.

Para entender de mejor manera, en esta sección del marco conceptual se expondrá en primer acerca del cerebro humano y las neuronas, luego sobre los potenciales eléctricos y se finalizará con su medición, que corresponde al electroencefalograma.

### 2.2.1. El Cerebro y las Neuronas

El cerebro humano es el centro del sistema nervioso. Se encarga de realizar el procesamiento de la información recibida por los sentidos y además se encarga de regular las funciones del cuerpo [30].

Este se divide en el hemisferio izquierdo y derecho y está cubierto por la corteza cerebral, la cual a la vez se divide en cuatro lóbulos que se asocian a funciones distintas. Los lóbulos son: temporal, frontal, parietal y occipital. Por ejemplo el lóbulo temporal se asocia a la audición y a la comprensión del lenguaje; el lóbulo frontal se asocia a las cualidades motoras; el lóbulo parietal se asocia al tacto; y el lóbulo occipital se asocia a la visión.

Por otro lado, las neuronas son las células funcionales del tejido nervioso. Se interconectan formando redes de comunicación que transmiten señales por del sistema nervioso. Poseen características morfológicas típicas y las siguientes componentes son las principales:

- **Soma** (o pericarion): es la zona de la célula donde se ubica el núcleo y desde el cuál nacen dos tipos de prolongaciones: las dendritas y el axón.
- **Dendritas**: reciben la información enviada desde los terminales axónicos de otras neuronas.
- **Axón**: proyección de la neurona que se encarga de enviar los impulsos del soma a otras neuronas.

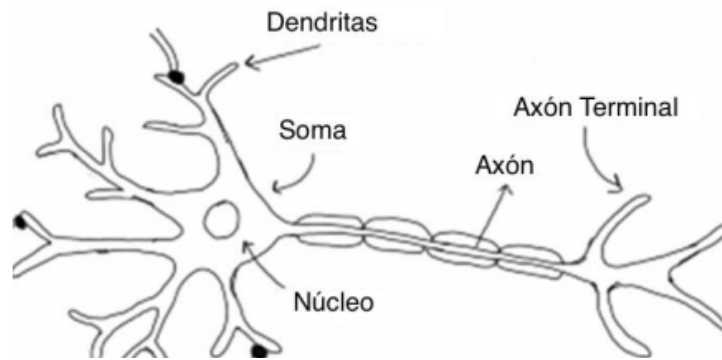


Figura 2.2: Morfología de una Neurona  
Fuente: Imagen adaptada de [30]

Por último, el cerebro posee dos tipos de materias: la blanca y la gris. La materia blanca está compuesta por fibras cubiertas de mielina, que contienen principalmente axones. En cambio la materia gris se compone principalmente de somas neuronales.

### 2.2.2. Potenciales Eléctricos [10]

En el sistema nervioso central, una neurona es activada por otras neuronas a través de potenciales de acción. Cuando esto ocurre, potenciales post-sinápticos excitatorios son enviados hacia sus dendritas apicales. Luego, las dendritas apicales comienzan a despolarizarse y quedan electro-negativamente cargadas, comparados con el soma de la neurona. Como consecuencia de esta diferencia de potencial, la corriente fluye desde el soma no excitado hacia el excitado árbol de dendritas y una negativa polaridad emerge en la superficie. En el caso opuesto, cuando el soma esta excitado, la corriente fluye en la dirección inversa.

Las células ubicadas en la corteza cerebral son paralelas entre ellas y perpendiculares a la corteza cerebral. Gracias a esto, es posible detectar las oscilaciones en la cabellera, que corresponden a la suma de potenciales post-sinápticos, tanto excitatorios como inhibitorios [31].

### 2.2.3. Electroencefalografía

El electroencefalograma o EEG, corresponde a la medición de las diferencias de potencial eléctricos sobre la corteza cerebral y ha sido utilizada ampliamente en el estudio de patologías (como el Alzheimer) y en la relación de diferentes tipos de estímulos con el potencial eléctrico.

Estas mediciones se hacen a través de electrodos colocados principalmente en el cuero cabelludo. Cada electrodo mide la diferencia de potencial con respecto a otro electrodo de referencia [31].

Existe una distribución estándar de la posición de los electrodos sobre el cuero cabelludo, que permite generar estudios comparables. En la Figura 2.3, se muestra la

distribución de electrodos del sistema internacional 10-20, que están representados por los círculos grises. También se puede ver que la notación tiene relación con los lóbulos de la corteza cerebral.

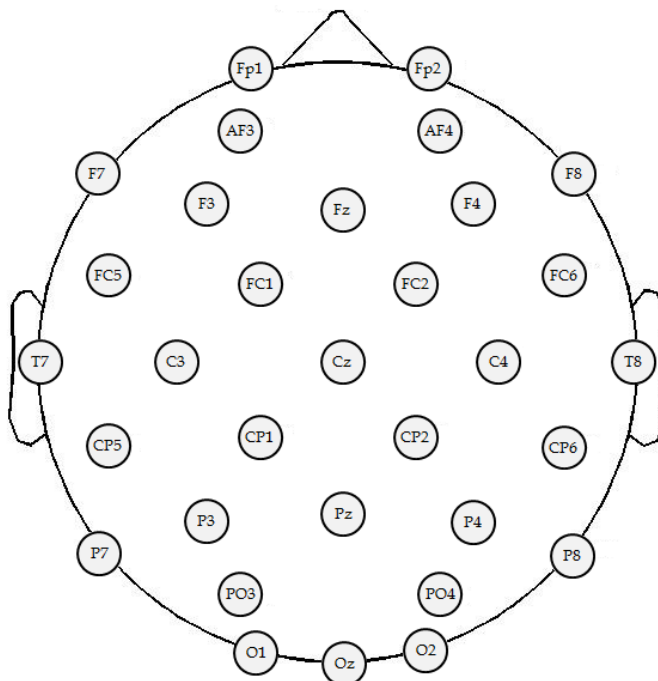


Figura 2.3: Posición de Electrodos del EEG  
Fuente: Imagen de [9]

Dentro de las mediciones del electroencefalograma, se pueden distinguir distintos tipos de ondas, que se diferencian en sus rangos de frecuencia y se relacionan a distintos niveles de atención y procesos cognitivos [32], [33]:

- **Ondas Delta:** su actividad oscila principalmente entre los 0,5 Hz y los 4 Hz, y su amplitud va desde los 20  $\mu$ V hasta los 200  $\mu$ V. Se asocia típicamente al sueño en humanos sanos y a patologías neurológicas.
- **Ondas Theta:** su actividad oscila principalmente entre los 4 Hz y los 7 Hz, y su amplitud va desde los 20  $\mu$ V hasta los 100  $\mu$ V. Se observa prominentemente durante el sueño.
- **Ondas Alpha:** en humanos sanos, su actividad oscila principalmente entre los 8 Hz y los 13 Hz, y su amplitud va desde los 20  $\mu$ V hasta los 60  $\mu$ V. Se observa principalmente en el estado de relajación.
- **Ondas Beta:** en adultos, su actividad oscila principalmente entre los 13 Hz y los 30 Hz [34], y su amplitud va desde los 2  $\mu$ V hasta los 20  $\mu$ V. Estudios demuestran que este tipo de onda se ve principalmente en estado de excitación y concentración.

- **Ondas Gamma:** este tipo de onda oscila principalmente sobre los 30 Hz, pero no existe un consenso claro acerca de su cota superior. Se puede hablar desde los 40 Hz hasta los 100 Hz [34]. Estas ondas se asocian a la atención, excitación, reconocimiento de objetos y al procesamiento sensorial, por lo que junto a otros estudios, se dice que se relacionan a la activación del cerebro.

## **2.3. Diferencias de Género**

Se dice que estas se originan aproximadamente las 8 semanas de gestación, cuando se presentan diferencias en la concentración de hormonas (testosterona). Esto provoca que el desarrollo futuro y el comportamiento del hombre sea distinto al de la mujer [35].

Se ha mostrado que existen diferencias significativas en la anatomía del cerebro de la mujer y del hombre [36]. Por ejemplo, el hombre posee un cerebro de mayor tamaño. Existen diferencias en la materia gris y materia blanca del cerebro, que tiene relación con las conexiones neuronales [37]. Y por otro lado, existen mayores conexiones y asociaciones neuronales entre el hemisferio derecho e izquierdo del cerebro en la mujer [38].

Las diferencias cerebrales entre el hombre y la mujer se traducen en diferencias cognitivas y en la literatura se presentan diferencias bastante consolidadas. Está demostrado que los hombres se desempeñan mejor en habilidades que requieren percepción visual y espacial [37], [39]. En parte, esto se debe a que el hombre posee un lóbulo parietal inferior de mayor tamaño, el que se relaciona con la percepción espacial y la rotación mental en tres dimensiones [38]. En cambio, la mujer se desempeña mejor en habilidades verbales y sociales [37], [39], en parte porque el tamaño del lóbulo temporal y frontal es mayor en la mujer [38].

Existe una gran variedad de estudios que se basan en las diferencias cognitivas que se presentaron anteriormente y que analizan diferencias en el comportamiento de los hombres y las mujeres.

Un resultado relevante de estos estudios, es que los hombres procesan los estímulos de forma selectiva y que se apoyan en heurísticas internas para tomar sus decisiones. En cambio las mujeres evalúan e integran la mayor cantidad de información disponible para tomar sus decisiones [38]. También se ha demostrado que existen diferencias en la reacción de los hombres y las mujeres y que existen distintas preferenciales de estímulos visuales entre ambos géneros [40], [41].

### **2.3.1. Diferencias en el Comportamiento Web**

Basándose en las conclusiones que se han podido obtener en cuanto a las diferencias de género, diversos investigadores han llevado el estudio a la Web, encontrando evidencia que es consistente con el comportamiento de los hombres y las mujeres fuera de la Web.

Estos estudios tienen distintos objetivos, como analizar diferencias en las preferencias de la presentación de contenido de un sitio web, en la forma de navegación, en la búsqueda de información o en el análisis de los factores que pueden afectar la actitud del hombre o la mujer hacia la compra de un producto en la Web.

Por ejemplo, en relación a las preferencias de la presentación de contenido, en [38] a través de encuestas, se concluye que los hombres ponen más atención a la publicidad con más recursos gráficos. También, los hombres mostraron una actitud más favorable a la publicidad mostrada en la parte izquierda de la página web, en cambio las mujeres mostraron una actitud más favorable a la publicidad mostrada en la parte derecha de la página web.

En [42], 30 mujeres y 30 hombres tienen que evaluar la estética 20 páginas web en función de la simetría vertical y otros factores en el diseño. Se concluye que la simetría no influye en la evaluación de las mujeres, pero si influye en la evaluación de los hombres, quienes prefieren páginas web verticalmente simétricas.

En [40], se evalúa la notoriedad de ciertos objetos para mujeres y hombres mientras buscan cierta información solicitada previamente. Estos objetos difieren en el color de fondo (claro y oscuro) y en el uso de imágenes de personas. Y la notoriedad fue medida por la cantidad de fijaciones en estos objetos. El resultado principal es que no hubo diferencias significativas en el número de fijaciones entre hombres y mujeres al variar las características del objeto, lo que indica que no existen diferencias en la notoriedad. Finalmente se concluye que la notoriedad puede no estar relacionada al atractivo. Sin embargo, en este estudio no se evalúa el atractivo del objeto en cuestión, para lo cual hubiese sido interesante analizar el comportamiento pupilar al observar este objeto.

En otros estudios, se ha visto que el hombre muestra diferencias con la mujer al momento de diseñar una página web. Por ejemplo, el hombre prefiere utilizar colores más oscuros como el azul o el negro, utiliza tipografía regular, lenguaje más formal y formas rectas. En cambio las mujeres prefieren utilizar colores claros como el blanco, amarillo o rosado. También utiliza tipografía poco regular, lenguaje más conversacional y utiliza más formas redondas. Los mismos investigadores obtienen evidencia de que los hombres y mujeres responden más favorablemente a sitios diseñados por su mismo género [43], [44], [45].

En [46], se evalúa la percepción y satisfacción en cuanto al diseño de una página web mientras hombres y mujeres realizan una compra a través de esta página. El resultado es que las mujeres valoran más los colores de la página, mientras que los hombres valoran más los recursos más gráficos, como las animaciones.

En cuanto a la navegación en una página web, en [47] se concluye que los hombres prefieren la información presentada de forma simple y directa, mientras que las mujeres son más receptivas a la información presentada de forma compleja. También en [48], se miden las fijaciones de hombres y mujeres al navegar por una página web y se concluye que las mujeres realizan un mayor esfuerzo cognitivo al realizar una navegación más exhaustiva que los hombres. De forma similar, en [17] se concluye que las mujeres observan por más tiempo cada región de una página web al buscar información.



Y en otro ámbito, en [11] se estudia la relación de las emociones con los colores de una página web de venta de boletos de transporte. A través de encuestas a hombres y mujeres, se concluye que los hombres relacionan el azul, verde, gris y negro a emociones positivas y el naranja y el rojo a emociones negativas. En cambio las mujeres relacionan el rojo, rosado y blanco a emociones positivas y el gris y amarillo a emociones negativas.

### ***La Evolución y sus Implicancias en el Comportamiento Web***

Se ha relacionado la evolución de los humanos con las diferencias entre sexos. En particular se dice que el rol de cazador del hombre y el rol de recolector de la mujer en la historia ha originado diferencias entre sexos a través de la evolución. En [49] se mencionan diferencias de género en habilidades espaciales; diferencias en la percepción de color, forma y movimiento; y diferencias verbales entre hombres y mujeres, y sus implicancias en el entorno web. A continuación, se muestran estas diferencias entre hombres y mujeres originadas por la evolución y su relación con la evidencia en el comportamiento web.

<b>Diferencia</b>	<b>Explicación Evolutiva</b>	<b>Evidencia en la Web</b>
Mujeres exploran utilizando puntos de referencia ( <i>landmarks</i> ).	La recolección requería de desplazamientos cortos, en los cuales el hogar sea fácilmente encontrado.	<ul style="list-style-type: none"> <li>• Mujeres indicaron que páginas web con menú desplegable eran más fáciles para navegar.</li> <li>• Mujeres indicaron preferencia por sitios que evitan múltiples niveles de subpáginas en su navegación.</li> </ul>
Hombres exploran utilizando la orientación (“brújula interna”).	La caza requería de desplazamientos de grandes distancias con respecto al hogar.	<ul style="list-style-type: none"> <li>• Hombres demoraron menos tiempo en completar tareas en sitios con el contenido más categorizado en comparación a sitios con el mismo contenido pero menos categorizado.</li> </ul>
Hombres demuestran superior habilidad en la rotación espacial.	La caza requería de precisión en la persecución y en el lanzamiento de proyectiles a la presa en movimiento.	<ul style="list-style-type: none"> <li>• Hombres valoraron más la opción de ver desde varios ángulos un producto en la Web al momento de comprar.</li> </ul>
Mujeres demuestran superioridad en la percepción de colores.	La recolección requería de habilidades para detectar y diferenciar plantas y alimentos entre toda la vegetación.	<ul style="list-style-type: none"> <li>• Mujeres valoraron más los colores en la estética de páginas web.</li> </ul>
Hombres demuestran superioridad en la percepción de movimiento.	La caza requería una mayor atención a objetos en movimiento.	<ul style="list-style-type: none"> <li>• Hombres prefirieron más que las mujeres páginas web que utilizan más gráficos y objetos animados.</li> <li>• Hombres prefirieron las animaciones y la interactividad con páginas web.</li> </ul>

Tabla 2.1: La Evolución y la Evidencia en el Comportamiento Web

Fuente: Tabla adaptada de [49]

### **2.3.2. Electroencefalograma**

Generalmente, en los estudios de medición con electroencefalograma, el género no ha sido considerado como un factor principal en el análisis, si no que se estudia principalmente la influencia de alguna enfermedad. Pero a pesar de esto, se cree que el género si es un factor importante en el potencial eléctrico del cerebro y se han hecho estudios en los que se considera el género cómo factor principal [50].

Uno de los trabajos más importantes en relación al género y al electroencefalograma es el de Matsuura [51], en el que estudia a 1416 sujetos. En este estudio se encuentra que existe mayor actividad en las bandas Beta y Theta en las mujeres, resultado que se repite en otros estudios. También se ha encontrado mayor actividad en la banda Delta en mujeres. Por otro lado se ha encontrado mayor actividad en la banda Alpha en los hombres, sin embargo existen estudios en los que se han encontrado resultados contradictorios [50].

### **2.3.3. Resumen de la Sección**

Se puede concluir que existen diferencias consolidadas entre ambos sexos y que existe evidencia de que estas diferencias se proyectan en el comportamiento web de hombres y mujeres. Principalmente se puede decir que las mujeres navegan más exhaustivamente que los hombres y que existen diferencias en las preferencias del diseño de una página web.

Por otro lado, se observa que la mayoría de los estudios de comportamiento web utiliza encuestas como fuente de información. Existe escasas en la literatura de estudios que utilicen el comportamiento pupilar y la actividad cerebral como fuente de información para analizar el comportamiento web de los usuarios. Y de igual forma, tampoco se utilizan estas fuentes de información para analizar las diferencias entre sexos, y menos aún en el entorno web.

## **2.4. Proceso KDD**

La sigla KDD se refiere a *Knowledge Discovery in Databases*, que corresponde al proceso de extracción de información de gigantes fuentes de datos y es el proceso que será utilizado en esta memoria para el tratamiento y análisis de datos.

El proceso KDD fue definido como el proceso no trivial de identificar patrones válidos, originales, útiles y entendibles sobre los datos [52]. En [53] se define el KDD como un proceso iterativo, pero secuencial, que involucra varios pasos, en los cuales es posible ir y volver de ser necesario. Los pasos de este proceso se describen a continuación.

1. **Selección de Datos:** se seleccionan los datos a utilizar en los siguientes pasos de la muestra total de datos disponibles. Para esto, es necesario tener en consideración los recursos y objetivos del análisis.

2. **Pre-Procesamiento de los Datos:** este paso consiste en resolver anomalías dentro del conjunto de datos seleccionados, como la reducción de ruido, reemplazo de datos faltantes, etc., con el fin de obtener datos limpios.
3. **Transformación de Datos:** en este paso se transforman los datos, aplicando ciertas funciones dependiendo del problema, para que sean posteriormente utilizados en el modelo de minería de datos.
4. **Minería de Datos:** en esta etapa se seleccionan y se aplican los modelos de minería de datos sobre la base de datos limpia y transformada. Esta etapa es la que permite el descubrimiento de patrones en los datos. La selección de los modelos a utilizar es una parte fundamental para el éxito de esta etapa, lo que depende directamente de los objetivos del análisis.
5. **Evaluación e Interpretación:** se evalúan los resultados obtenidos en la etapa anterior y se interpretan. Los resultados pueden ser muy favorables o desfavorables con los objetivos, sin embargo en ambos casos se adquiere conocimiento si se realiza el proceso correctamente.

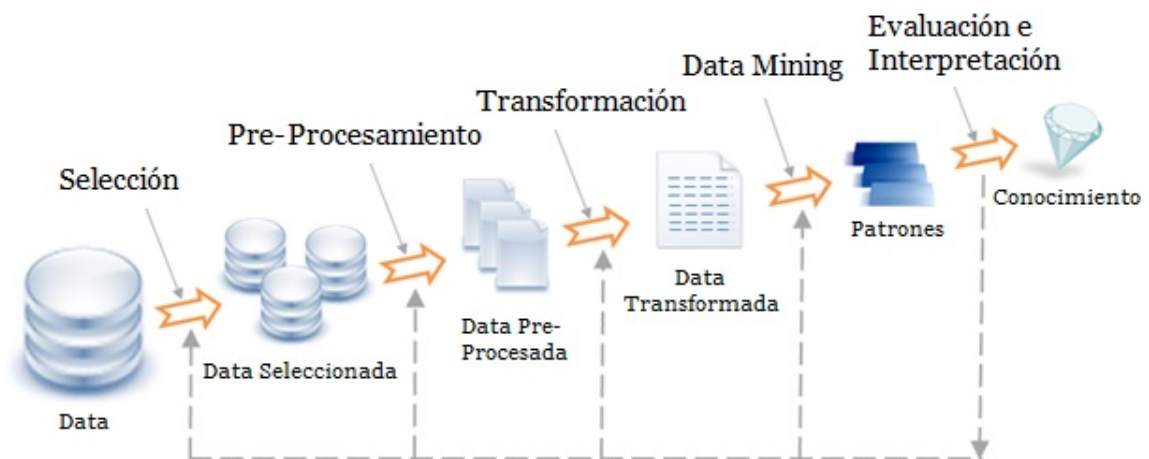


Figura 2.4: Proceso KDD  
Fuente: Imagen adaptada de [52]

### 2.4.1. Minería de Datos

El paso fundamental para la obtención de conocimiento en el proceso KKD, es la minería de datos. La minería de datos consiste de distintas técnicas para la obtención de patrones de los datos. Estas técnicas se pueden diferenciar de acuerdo a los objetivos de la minería de datos, que son: verificación y descubrimiento.

El objetivo de la verificación es la validación de hipótesis a través métodos, como la bondad de ajuste o análisis de varianza. En cambio, el objetivo del descubrimiento es determinar la existencia de patrones y reglas de forma autónoma. A la vez, el

descubrimiento se puede diferenciar en métodos con propósito descriptivo o métodos con propósito predictivo.

El foco de los métodos descriptivos es interpretar y comprender la manera en la que los datos se relacionan con sus partes. En este tipo de métodos se pueden encontrar técnicas como el *clustering*. Por otro lado, los métodos predictivos tienen como objetivo construir un modelo de comportamiento, que pueda obtener nuevos ejemplos y sea capaz de predecir valores en relación al ejemplo. Estos métodos predictivos se pueden dividir en métodos de regresión y métodos de clasificación.

Los métodos de regresión tienen el objetivo de generar un pronóstico cuantitativo a partir de los datos. Y finalmente los métodos de clasificación tienen el objetivo de predecir la clase o categoría de los datos a partir de las variables que los caracterizan. La siguiente figura muestra la taxonomía descrita anteriormente.

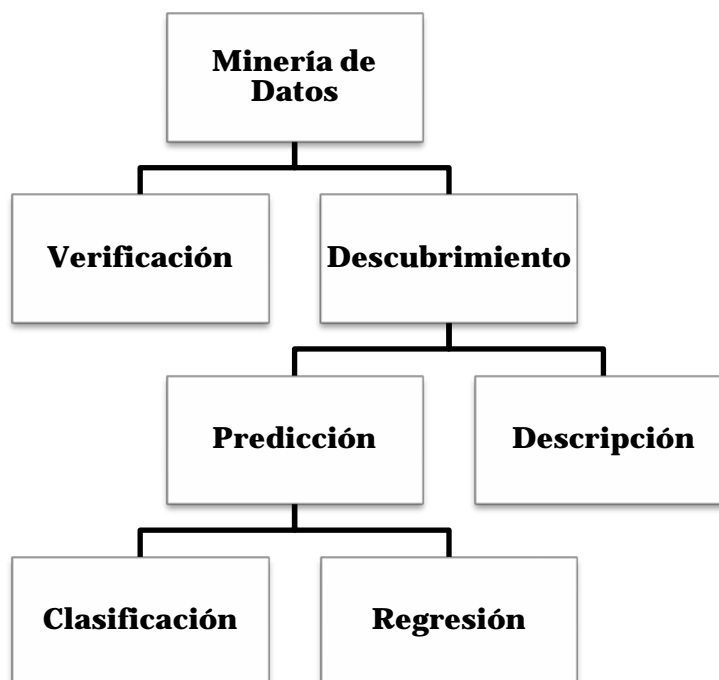


Figura 2.5: Taxonomía de la Minería de Datos  
Fuente: Elaboración propia, basada de [53]

Por otro lado, se pueden distinguir las técnicas de minería de datos en métodos supervisados y métodos no supervisados. Los métodos supervisados intentan descubrir la relación entre las variables de entrada y un atributo predefinido. Estos métodos se asocian a los métodos predictivos. Por otro lado, los métodos no supervisados modelan la distribución de instancias, donde estas no están previamente determinadas. Estos métodos se asocian a los métodos descriptivos.

### **2.4.1.1. Minería de Datos, Electroencefalograma y Eye Tracking**

La minería de datos ha sido implementada en una gran variedad de estudios para analizar los datos obtenidos por el electroencefalograma o el eye tracker. Se ha observado que a diferencia del eye tracker, el electroencefalograma ha sido mucho más utilizado en conjunto con la minería de datos, dadas las características de su señal y por la gran cantidad de canales que se pueden utilizar.

En el caso del electroencefalograma, la minería de datos se ha utilizado para diversos tipos de análisis y problemas, pero en los últimos años se ve principalmente que se estudia la relación de la actividad cerebral con las emociones [54], [55], [56], [57], su relación con la carga cognitiva [58], [59], [60], su relación con las preferencias sensoriales [61], [62] y su relación con patologías [34], [63], [64], [65], [66], [67], en particular con el Alzheimer y Epilepsia.

La señal del electroencefalograma es no estacionaria y no lineal [68], por lo que para transformar la señal para obtener las variables que se utilizarán en los modelos, es bastante utilizado el uso de Transformadas de Wavelet Discreta y ha traído buenos resultados, independientemente del tipo de problema que se esté estudiando.

En relación a los modelos de minería de datos, se observa que se ha utilizado una gran variedad y que se han hecho comparaciones de modelos para distintos tipos de problemas, llegando a conclusiones que destacan algunos modelos sobre otros. En [69], se realiza un estudio de modelos de clasificación con el uso de electroencefalograma en *Brain Computer Interfaces (BCI)* y se realiza una comparación del desempeño de ellos:

- **Linear Discriminant Analysis (LDA):** este modelo ha sido utilizado con éxito en diversos tipos de estudios de *BCI*, sin embargo, puede que no sea una buena opción si se realizan estudios con datos complejos y no lineales del electroencefalograma.
- **Support Vector Machine (SVM):** ha sido aplicado con éxito a una gran cantidad y variedad de problemas. También tiene la ventaja de la utilización de una función Kernel, la cual transforma los datos y permite trabajar en problemas no lineales. El Kernel más utilizado en estos estudios es el Gaussiano o RBF (*Radial Basis Function*), lo que da a lugar al RBF SVM, generando muy buenos resultados.
- **MultiLayer Perceptron (MLP):** es el modelo de clasificación más popular dentro de las redes neuronales. Ha sido utilizado en casi todas las aplicaciones de *BCI*. Sin embargo, el modelo es sensible a datos ruidosos y no estacionarios como los datos del electroencefalograma.
- **Hidden Markov Models (HMM):** son ideales para la clasificación de series de tiempo, por lo que ha sido utilizado para la clasificación de secuencias de tiempo con buenos resultados. Sin embargo, este modelo es poco utilizado.

Finalmente, se concluye que las propiedades del SVM hacen que este modelo sea el más adecuado para una gran variedad de estudios con datos de electroencefalograma. También propone que una combinación de modelos puede traer mejores resultados aún.

En relación a la minería de datos y el eye tracker, una cantidad mucho menor de estudios se ha encontrado. Estos estudios utilizan en los modelos variables basadas en las fijaciones oculares, sacadas y en la pupila. Se enfocan principalmente en la relación de estas variables con la resolución de problemas y tareas visuales con carga cognitiva [70], [71], [72], [73]. Finalmente, al ser pocos estudios, no se ha encontrado una tendencia en la implementación de modelos con eye tracker, sin embargo, se ve que la utilización de RBF SVM ha traído buenos resultados.

#### **2.4.1.2. Support Vector Machine (SVM)**

Como se vio en la sección anterior, el modelo SVM es uno de los modelos de clasificación más populares en el análisis de datos obtenidos por eye tracking y electroencefalograma y ha tenido consistentemente buenos resultados.

Este es un modelo de clasificación y es uno de los más importantes y precisos [67]. Además de la evidencia, se dice que el SVM es adecuado para la clasificación de estados cognitivos de los humanos por tres razones [70]:

1. Es difícil representar estados cognitivos por modelos lineales. SVM puede desempeñarse en problemas no lineales tan eficientemente como en problemas lineales.
2. SVM se puede aplicar sin conocimiento previo antes del entrenamiento. Además tiene buen desempeño con datos ruidosos.
3. Mientras que otros métodos tradicionales sólo minimizan el error de entrenamiento, SVM minimiza el borde superior del error de generalización, haciendo que se produzcan modelos más robustos.

Dado lo anterior, este modelo será utilizado para el análisis de esta memoria será descrito a continuación [10], [74].

SVM es un modelo supervisado que en general clasifica de forma binaria. Para la formulación del SVM, se supone un set de entrenamiento  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , donde los  $x_i$  representan los vectores de datos y los  $y_i$  representan las etiquetas que pueden ser 1 o -1 (clase positiva o negativa), lo que quiere decir si pertenece o no a la clase en estudio respectivamente.

En SVM, se busca una función lineal de la forma de la ecuación 2.1, de tal forma que si el vector  $x_i$  es de la clase positiva, entonces  $f(x_i) \geq 0$  y  $f(x_i) \leq 0$  para la clase negativa.

$$f(x) = \langle w \cdot x \rangle + b \quad (2.1)$$

$$y_i = \begin{cases} 1 & \langle w \cdot x_i \rangle + b \geq 0 \\ -1 & \langle w \cdot x_i \rangle + b < 0 \end{cases} \quad (2.2)$$

El parámetro  $w$  corresponde al vector de pesos y  $b$  es llamado bias.

SVM busca un hiperplano, llamado frontera de decisión, el cual separa los datos de entrenamiento en clases positivas y negativas.

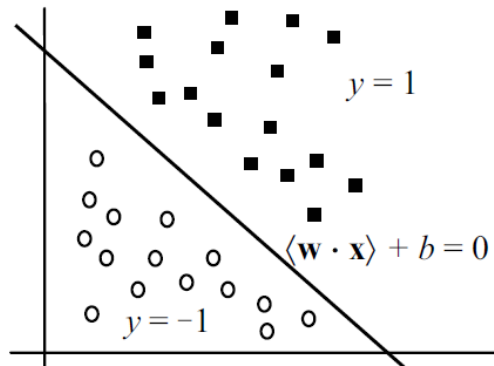


Figura 2.6: Frontera de Decisión  
Fuente: Imagen adaptada de [74]

Para encontrar el hiperplano que divide los datos de entrenamiento en dos clases, se resuelve el problema expresado en la ecuación 2.3, en la que se maximiza el margen entre los datos de distinta clase.

$$\begin{aligned} \text{Min: } & \frac{\langle w \cdot w \rangle}{2} \\ \text{s. a: } & y_i(\langle w \cdot x_i \rangle + b) \geq 1 \end{aligned} \quad (2.3)$$

Podría ocurrir que el hiperplano que cumpla con la ecuación 2.3 no exista y que no se encuentre solución. En este caso, se relaja el problema permitiendo que a algunos datos se les asigne la clase a la cual no pertenecen. Para esto se incluyen variables al problema, quedando como en la ecuación 2.4.

$$\begin{aligned} \text{Min: } & \frac{\langle w \cdot w \rangle}{2} + C \sum_{i=1}^n \xi_i \\ \text{s. a: } & y_i(\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \quad (2.4)$$

El problema descrito anteriormente se puede resolver de forma lineal. Para el caso de problemas no lineales, se utilizan las funciones Kernel ( $\phi$ ), las cuales convierten el espacio de entrada  $X$  en un espacio de características  $F$ , como se muestra en la siguiente figura.

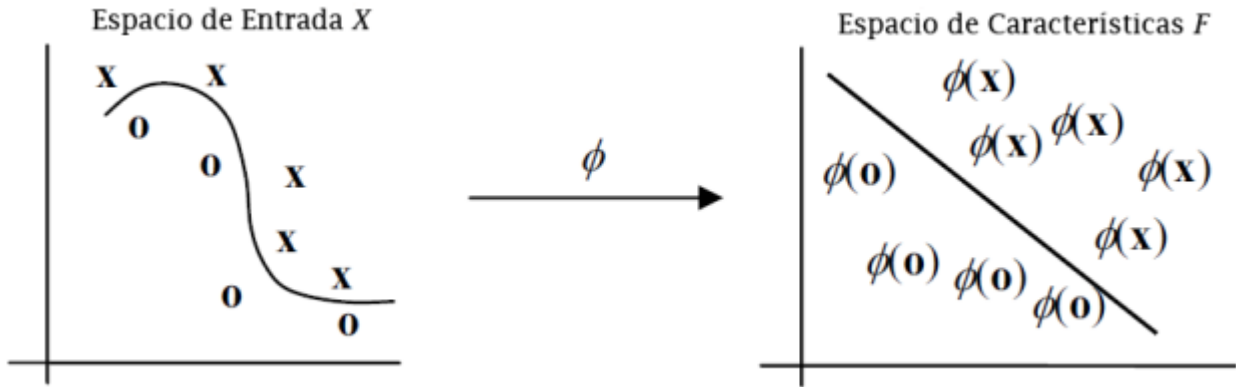


Figura 2.7: Transformación con Función Kernel  
Fuente: Imagen adaptada de [74]

Con esta transformación, el problema queda como 2.5.

$$\begin{aligned}
 \text{Min: } & \frac{\langle w \cdot w \rangle}{2} + C \sum_{i=1}^n \xi_i \\
 \text{s. a: } & y_i (\langle w \cdot \phi(x_i) \rangle + b) \geq 1 - \xi_i \\
 & \xi_i \geq 0
 \end{aligned} \tag{2.5}$$

Como se ha mencionado, la función Kernel más popular en aplicaciones con datos fisiológicos, en particular del electroencefalograma, es el Kernel Gaussiano o RBF. Este Kernel utiliza la distancia Euclidiana entre dos vectores de datos de entrada,  $x_i$  y  $x_j$  y utiliza un parámetro  $\sigma$  que es libre. La siguiente ecuación describe esta función [59].

$$\varphi(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{2.6}$$

Por otro lado, existen otras funciones Kernel que pueden ser menos utilizadas en aplicaciones con electroencefalograma, pero que son globalmente tan populares como el Kernel RBF [75]. Estas funciones son el Kernel Lineal y el Kernel Polinomial, las que se utilizarán de igual forma para comparar su desempeño.



El Kernel Lineal se describe según la siguiente ecuación.

$$\varphi(x_i, x_j) = x_i^T \cdot x_j \quad (2.7)$$

Y el Kernel Polinomial se describe según la siguiente ecuación, con  $d$  el grado del polinomio.

$$\varphi(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (2.8)$$

### **2.4.1.3. Regresión Logística** [10], [76]

Corresponde a otro modelo predictivo, que deriva de las regresiones lineales. En este caso se estima la probabilidad de pertenecer o no a una clase en estudio, por lo que los valores fluctúan entre 0 y 1.

La regresión logística está representada por las ecuaciones 2.9 y 2.10. En la ecuación 2.9,  $x$  es el vector que representa el caso que se quiere predecir y  $\beta$  es el vector de parámetros de la regresión logística. La ecuación 2.10 corresponde a la función logística o sigmoidea y sus valores pueden variar entre 0 y 1, los que representan la probabilidad de pertenecer a la clase en estudio.

$$h_{\beta}(x) = g(\beta^T x) \quad (2.9)$$

$$g(z) = \frac{1}{1 + \exp^{-z}} \quad (2.10)$$

Para encontrar los parámetros de la regresión logística se debe entrenar el modelo, para lo cual se minimiza la ecuación 2.11 llamada función de costos. En esta ecuación, se utiliza el set de entrenamiento  $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , con  $x_i$  vectores de las variables independientes e  $y_i$  la etiqueta de la clase, que puede ser 0 o 1.

$$J(\beta) = -\frac{1}{m} \left[ \sum_{i=1}^m y^{(i)} \log h_{\beta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\beta}(x^{(i)})) \right] \quad (2.11)$$

### **2.4.1.4. Árbol de Decisión** [9], [53]

Este modelo corresponde a un clasificador que se expresa como una partición recursiva de un espacio de datos. Gráficamente se representa por nodos que pueden ser

de 3 tipos: nodo raíz, nodos internos y nodos hojas. El nodo raíz corresponde al nodo inicial del árbol, por lo que no posee bordes entrantes. Los nodos internos están entre el nodo raíz y los nodos hojas, por lo que poseen bordes entrantes y salientes. Por último, los nodos hojas son los nodos terminales, por lo que sólo poseen bordes entrantes.

A cada nodo hoja se le asigna una categoría o clase, que es representada por un valor objetivo. El nodo hoja también puede estar relacionado a un vector que contenga la probabilidad de que el atributo objetivo tenga un cierto valor. En la siguiente figura se muestra un ejemplo de árbol de decisión, en el que el nodo raíz y los nodos internos están representados por círculos y los nodos hojas por triángulos.

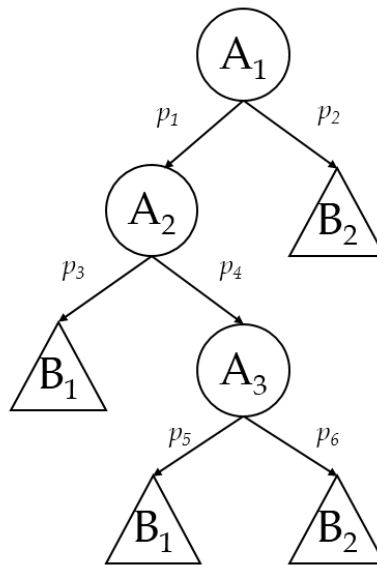


Figura 2.8: Ejemplo de Árbol de Decisión  
Fuente: Imagen de [9]

El árbol de decisión puede incorporar atributos numéricos y nominales, y cada nodo del árbol se encuentra etiquetado con un atributo. Para avanzar al siguiente nodo, se consulta sobre el valor del atributo y su respuesta determinará el camino que se debe seguir. Finalmente para clasificar una instancia, se debe recorrer el árbol desde el nodo raíz hasta un nodo hoja.

El objetivo de los árboles de decisión, es encontrar un algoritmo que minimice el error generalizado, pero además, se pueden definir otras funciones objetivo que busquen minimizar la cantidad de nodos o la profundidad del árbol.

Finalmente, para construir un árbol, típicamente se utiliza el algoritmo *Top-Down Induction*, que utiliza una función que depende de 5 variables.

$$\text{Árbol de Decisión} = f(S, A, y, CDe, CDi) \quad (2.12)$$

La variable  $S$  corresponde al conjunto de datos de entrenamiento,  $A$  corresponde a las características de los datos de entrada,  $y$  corresponde a las características de las variables de salida,  $CDe$  es la condición de detención del algoritmo y  $CDi$  corresponde a la condición de división del árbol.

Este método está construido de manera recursiva, donde en cada iteración, el algoritmo considera el resultado de la función de los atributos de entrada, elige la mejor y selecciona una partición apropiada. Esto se repite hasta que las nuevas particiones no generen una ganancia o hasta que se alcance el criterio de detención predefinido.

#### 2.4.1.5. Evaluación de Modelos

Existen distintos indicadores para evaluar el resultado de un método de clasificación. Para estos indicadores, se utiliza la Matriz de Confusión, representada en la Figura 2.8. Las columnas hacen referencia a las clases predichas por el modelo, mientras que las filas hacen referencia a las clases reales [9].

	Predicción de Positivos	Predicción de Negativos
Actual Positivo	Verdadero Positivo (VP)	Falso Negativo (FN)
Actual Negativo	Falso Positivo (FP)	Verdadero Negativo (VN)

Figura 2.9: Matriz de Confusión  
Fuente: Elaboración propia, basada de [53]

En esta matriz, se identifican los casos:

- **Verdadero Positivo (VP):** casos que el modelo clasificó como positivo y acertó.
- **Falso Positivo (FP):** casos que el modelo clasificó como positivo y no acertó, lo que es conocido como el error de tipo I.

- **Verdadero Negativo (VN):** casos que el modelo clasificó como negativo y acertó.
- **Falso Negativo (FN):** casos que el modelo clasificó como negativo y no acertó, lo que se conoce como el error de tipo II.

A partir de esto, se derivan los indicadores que permitirán evaluar el desempeño de un clasificador. A continuación se exponen tres medidas de evaluación.

- **Recall:** corresponde a la fracción de casos verdaderos positivos, dentro de los actuales positivos.

$$Recall = \frac{VP}{VP + FN} \quad (2.13)$$

- **Precision:** corresponde a la fracción de casos verdaderos positivos, dentro de los predichos positivos.

$$Precision = \frac{VP}{VP + FP} \quad (2.14)$$

- **Accuracy:** corresponde a la fracción de casos correctamente clasificados, dentro del total de los casos.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.15)$$

Para calcular finalmente el *Accuracy* de un clasificador, se utilizan técnicas como el *k-fold cross-validation*, en el que se separan los datos en *k* partes ( *folds*), luego se itera *k* veces, tomando en la iteración *i* el *fold i* como dato de prueba y el resto como dato de entrenamiento. El *Accuracy* se obtiene promediando los *k Accuracy* [8].

# Capítulo 3: Experimento

Para validar o no validar la hipótesis de investigación planteada en esta memoria, se utilizarán datos obtenidos a través de experimentos, los cuales serán analizados posteriormente con herramientas de minería de datos. En este capítulo, se detalla el proceso de diseño e implementación del experimento, con lo que se cumple el objetivo específico número dos de esta memoria.

En primer lugar, se presenta lo que tiene relación al diseño del experimento, seguido por la implementación de este.

## 3.1. Diseño del Experimento

Un correcto diseño es fundamental para que los resultados obtenidos por el experimento sean válidos [77]. En este proceso, se tienen en consideración fundamentalmente los recursos que se poseen y se necesitan, además de la hipótesis que se quiere validar.

En esta sección se describen las componentes del diseño del experimento. Primero se expone sobre la instrumentación, en la que se presentan los recursos que se poseen. Luego se presentan los requerimientos y características que debe tener el grupo experimental. Luego se detalla acerca de la selección de las páginas web a utilizar en el experimento. Y finalmente se presenta el procedimiento que debe realizar el sujeto.

### 3.1.1. Instrumentación

Para la toma de muestra, se utilizó el laboratorio ubicado en el Departamento de Ingeniería Industrial de la Universidad de Chile. Este posee un eye tracker, un equipo de electroencefalograma, un computador con dos pantallas apartadas y las condiciones de luminosidad adecuadas para la utilización del eye tracker.

- **Computador:** en el laboratorio se posee un computador con dos pantallas, una pantalla para que el sujeto visualice el estímulo (23 pulgadas) y otra para monitorear la actividad del sujeto. Ambas pantallas están apartadas y separadas por una cortina que permite las condiciones de luminosidad adecuadas para el sujeto.

El computador posee el sistema operativo Windows 8 y está constantemente conectado a internet. Por lo anterior es posible utilizar Mozilla Firefox como navegador.

- **Eye Tracker:** como se mencionó en el capítulo anterior, esta es la herramienta utilizada para el registro de la mirada y el registro de la pupila. El eye tracker del laboratorio consiste de una cámara que se adhiere a la pantalla. Esta cámara graba un solo ojo y utiliza el reflejo en el ojo de 4 leds que se posicionan en las 4 esquinas de la pantalla.



Figura 3.1: Pantalla del Laboratorio con Eye Tracker  
Fuente: Elaboración propia

Con este eye tracker se pueden registrar secuencialmente las coordenadas de la mirada en la pantalla medidas en pixeles en el tiempo, el área de la pupila en el tiempo, los movimientos sacádicos, los pestañeos, entre otras cosas. Y su tasa de sampleo no está especificada por el fabricante, pero registra aproximadamente 20 datos por segundo (20Hz).

Para el manejo del eye tracker en el computador, se utiliza el software Sofey Eye Tracker. A través de él se observa el ojo del sujeto, se realizan los ajustes para la detección del ojo, se calibra el eye tracker y se registran los datos, entre otras opciones.

- **Equipo de Electroencefalograma:** este equipo es el encargado de medir la actividad bioeléctrica cerebral del sujeto. El equipo que está en el laboratorio es el Neuroheadset EPOC, desarrollado por la empresa Emotiv.



Figura 3.2: Emotiv EPOC Neuroheadset  
Fuente: Imagen de [78]

Este es un equipo inalámbrico que tiene la ventaja de ser considerablemente más fácil de preparar y usar que un equipo de EEG convencional. Este equipo cuenta con 14 electrodos posicionados de acuerdo a la Figura 3.3 y envía las 14 señales recibidas al computador vía bluetooth.

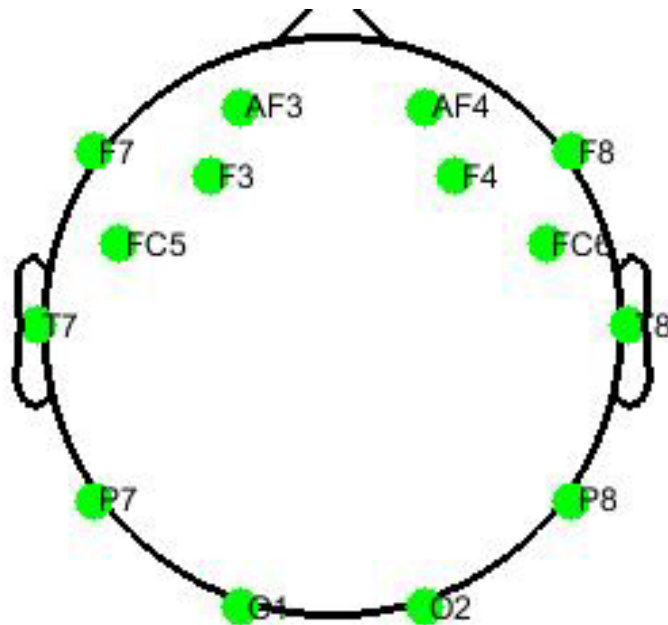


Figura 3.3: Posición de los Electrodos según el Sistema 10-20  
Fuente: Elaboración propia

En la siguiente tabla se muestran las especificaciones técnicas de este equipo.

Número de Canales	14, más 2 de referencia (CMS, DRL)
Posición de Canales (Sistema 10-20)	AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4
Método de Muestreo	Secuencial
Tasa de Muestreo	128 SPS (2048 Hz)
Conectividad	Wireless, banda de 2,4 GHz
Duración de Batería	12 horas

Tabla 3.1: Especificaciones técnicas del Emotiv EPOC Neuroheadset

Fuente: Elaboración propia, adaptada de [78]

De similar forma al eye tracker, este dispositivo cuenta con un software que permite el registro de datos, la visualización de la actividad cerebral en tiempo real, observar el estado del dispositivo, entre otras cosas.

### 3.1.2. Grupo Experimental

El grupo experimental es un factor fundamental en los resultados de la investigación. Los sujetos deben cumplir con ciertas características para que los resultados sean válidos.

En primer lugar, si se quieren estudiar las diferencias de género, se tiene que tener una cantidad considerable y similar de hombres y mujeres. Por otro lado, cada una de las personas, independientemente de su sexo, debe cumplir con condiciones aptas de salud al momento de realizar el experimento. Además, deben poseer los conocimientos básicos para poder realizar una tarea utilizando un computador.

En específico, las condiciones que deben cumplir los sujetos son:

- Tener entre 18 y 45 años de edad.
- Deben estar en condiciones normales de salud y no deben tener un diagnóstico previo de problemas psiquiátricos y/o psicológicos.
- No deben consumir medicamentos que puedan afectar al Sistema Nervioso Central.
- No deben presentar dificultades severas en la visión.
- No deben contar con un historial de abuso de drogas.

Además, cada persona debe estar de acuerdo con el procedimiento del experimento y con el uso que se le pueda dar a los datos obtenidos. Por esto, los sujetos deberán firmar el Consentimiento Informado que se les presentará antes de comenzar con el experimento.

Finalmente, se propone como mínimo tomar muestras de 10 hombres y 10 mujeres [10] y realizar un mínimo 20 experimentos en total [8], [9].



### **3.1.3. Tarea**

Uno de los requerimientos del experimento, es que se replique el comportamiento web del usuario en el experimento lo más posible a cuando navega en la realidad. Para esto, lo ideal es que el sujeto navegue con completa libertad. Sin embargo, al ser así, la variabilidad en el comportamiento dificultaría la obtención de resultados válidos, por lo tanto la mejor opción sería dar la misma instrucción para todos los sujetos [77].

Por otro lado, para representar la realidad, el sujeto debe navegar por la Web teniendo un interés o un objetivo en mente. Pero dada la dificultad de coincidir con los intereses de los sujetos y la dificultad de abarcar una gran variedad para lograr validez, es que se opta por la realización de una tarea en la Web para lograr un objetivo.

Se ha observado que una tarea muy común en el estudio del comportamiento web, es la búsqueda de información para resolver la pregunta a una respuesta, por ejemplo. Además se ha observado que existen diferentes estrategias de navegación para buscar información entre hombres y mujeres. También existe evidencia de que al variar los niveles de dificultad en las tareas de búsqueda de información, se incrementa el esfuerzo cognitivo de la persona, lo que ha sido representado por el tamaño de la pupila [24].

Por último, se ve que esta es una tarea que puede ser de fácil implementación y que está alineada con los objetivos del experimento, por lo que finalmente será este tipo de tarea la utilizada en el experimento.

### **3.1.4. Sitio Web**

Como se ha visto en la literatura, el tamaño de la pupila tiene relación con los estados cognitivos, el nivel de excitación, los intereses y las preferencias de las personas. Y por otro lado, la actividad cerebral medida por el electroencefalograma muestra cambios frente a los distintos estímulos, niveles de atención o estado cognitivos de las personas.

También existe evidencia de que existen diferencias de género a nivel biológico, cognitivo y conductual, las que se han visto reflejadas en el comportamiento web y en las distintas preferencias en la Web.

Por lo anterior, sería interesante medir y comparar el comportamiento fisiológico de interés en esta memoria entre hombres y mujeres, frente a distintos tipos de estímulos. Y dado a que se quiere observar el comportamiento en el entorno web, es que estos estímulos tienen que estar relacionados a este entorno. Es por esto que para el experimento se propone utilizar distintos tipos de sitios web, diferenciados por el sexo del público al cual se enfocan, además de sitios web que no tengan una tendencia de género (neutral). Entonces se tendrán sitios web “masculinos”, “femeninos” y “neutrales”.

En primer lugar, si se quiere tener distintos tipos de sitios web, se descarta la opción de construirlos, ya que requeriría de conocimientos de diseño web que se escapan del alcance de la memoria. Además, si se pudiera diseñar un sitio web por tipo, se

obtendrían resultados aplicables sólo al sitio web diseñado. Es por esto además que se quiere utilizar más de un sitio web por tipo de sitio (masculino, femenino, neutro). Entonces, se buscarán y clasificarán sitios web reales.

Para clasificar el tipo de sitio, se utiliza la evidencia acerca de las preferencias de los hombres y mujeres en un sitio web, las cuales han sido mencionadas en la sección 2.3.1. Principalmente, estas preferencias se pueden dividir en el estilo de navegación, el lenguaje utilizado en la página web y el aspecto visual. Las características que debe poseer idealmente cada tipo sitio web se muestran en la Tabla 3.2.

	Sitio “Masculino”	Sitio “Femenino”	Sitio “Neutral”
Navegación	<ul style="list-style-type: none"> <li>El contenido del sitio debe estar dividido en varias categorías y subpáginas.</li> </ul>	<ul style="list-style-type: none"> <li>El sitio web debe poseer pocas subpáginas y el contenido dividido en pocas categorías.</li> <li>Menús desplegables.</li> </ul>	<ul style="list-style-type: none"> <li>El sitio web debe poseer pocas subpáginas y el contenido debe estar dividido en pocas categorías.</li> </ul>
Lenguaje	<ul style="list-style-type: none"> <li>Formal</li> <li>Tipografía regular</li> </ul>	<ul style="list-style-type: none"> <li>Informal</li> <li>Tipografía especial</li> </ul>	<ul style="list-style-type: none"> <li>Conversacional</li> <li>Tipografía regular</li> </ul>
Visual	<ul style="list-style-type: none"> <li>Color: no más de 3 colores (negro, azul, verde o gris). Si se utiliza el rojo, debe ser con poca intensidad. Evitar el café, naranja, morado, amarillo.</li> <li>Presencia de objetos con gráficos elaborados.</li> <li>Debe tener una organización simple y estructurada.</li> <li>Los objetos y secciones del sitio web deben poseer formas rectas.</li> <li>Debe haber simetría vertical en la distribución del contenido.</li> </ul>	<ul style="list-style-type: none"> <li>Color: varios colores, como el azul, morado, verde o colores más claros y “cremosos”. No usar naranja, café, gris, amarillo.</li> <li>Objetos simples, con poca o nula utilización de gráficos sofisticados o animaciones.</li> </ul>	<ul style="list-style-type: none"> <li>Color: que no destaquen más de 3 colores (azul, verde o blanco). No puede estar presente el morado, gris, café, naranja o amarillo.</li> <li>Objetos simples, con poca utilización de gráficos sofisticados o animaciones.</li> <li>Debe tener una organización simple y estructurada.</li> <li>Los objetos y secciones del sitio web deben mezclar formas rectas y redondas.</li> </ul>

Tabla 3.2: Criterios para la Clasificación de Sitios Web

Fuente: Elaboración propia

Además de estas características, el sitio web debe demostrar que existe trabajo detrás de su diseño y mantención. Además, el contenido y el lenguaje deben ser entendibles y apropiados para este trabajo, lo que será determinado a juicio del autor.

Finalmente, dado que la tarea asignada será buscar información, se seleccionará sólo una subpágina web dentro del sitio web completo, ya que si la persona navega por el sitio web, la variabilidad de los resultados se incrementaría demasiado. Por lo tanto, el criterio de navegación no se consideraría para la selección de sitios web.

## 3.2. Implementación del Experimento

Una vez que el experimento está diseñado, se lleva a cabo. En esta sección se describen los principales aspectos de la implementación del experimento. Se comienza con una descripción del grupo de sujetos que participó en el experimento. Luego se exponen los sitios web utilizados. Se continúa con el procedimiento, que incluye el protocolo a seguir por el investigador y la tarea a realizar por sujeto. Finalmente se muestra el tipo de datos que se obtuvo.

### 3.2.1. Grupo Experimental

El grupo experimental cumple con los requisitos descritos en la sección 3.1.2. Se tomaron muestras de 20 personas, siendo 10 hombres y 10 mujeres. El rango de edad va desde los 24 años a los 37 años y el promedio es de 26,55 años (27,1 años los hombres y 26 años las mujeres). Todo el grupo reside en Santiago de Chile y está compuesto en su mayoría por personas relacionadas a la Ingeniería Civil, ya sea profesionales, memoristas o estudiantes. Sus datos se muestran en la Tabla 3.3.

N°	Sexo	Edad	Ocupación
1	Masculino	25	Ingeniero Civil
2	Masculino	24	Memorista
3	Masculino	34	Ingeniero Civil
4	Masculino	25	Ingeniero Civil
5	Masculino	37	Ingeniero Civil
6	Masculino	25	Estudiante
7	Masculino	26	Ingeniero Civil
8	Masculino	26	Memorista
9	Masculino	25	Memorista
10	Masculino	24	Memorista
11	Femenino	25	Ingeniero Civil
12	Femenino	25	Memorista
13	Femenino	25	Ingeniero Civil
14	Femenino	25	Estudiante
15	Femenino	27	Estudiante
16	Femenino	24	Memorista
17	Femenino	33	Secretaria
18	Femenino	26	Estudiante
19	Femenino	24	Estudiante
20	Femenino	26	Estudiante

Tabla 3.3: Grupo Experimental

Fuente: Elaboración propia

Por último, debido a problemas imprevistos, no se obtuvo el registro del electroencefalograma de 1 hombre y el registro del eye tracker de 1 mujer, por lo que finalmente se obtuvo exitosamente el registro del electroencefalograma de 19 personas y el registro del eye tracker de 19 personas.

### 3.2.2. Sitio Web

Para este trabajo, se seleccionaron distintos sitios web en base a lo descrito en la sección 3.1.4. Cabe decir que las características de la sección 3.1.4 describen el sitio web “ideal” para el hombre o mujer, por lo que han sido una guía para la búsqueda, clasificación y selección. Sin embargo, sería muy complejo encontrar sitios que coincidan con todas las características, y más aún que no cumplan con las características de otro tipo de sitio, por lo que finalmente, ha quedado a criterio del autor la clasificación de los sitios web. Se han seleccionado 15 sitios web, de los cuales 6 se han clasificado como masculinos, 5 como neutros y 4 como femeninos. En la Tabla 3.4 se muestran los sitios web con una breve descripción.

Tipo de Sitio Web	Sitio Web	Descripción
Masculino	gq.com.mx	Revista enfocada en el hombre, que incluye temas como automovilismo, mujeres y moda masculina.
	wired.com	Revista de temas de vanguardia en la economía, diseño, entretenimiento, ciencia y tecnología.
	primermagazine.com	Revista enfocada en hombres de entre 20 y 30 años de edad. Incluye temas como moda masculina, inversiones, cuidado personal y relaciones de pareja.
	businessinsider.com	Sitio web de noticias sobre tecnología, finanzas, estilo de vida, deportes, entre otros temas.
	werd.com	Blog enfocado en el hombre que muestra herramientas, aparatos tecnológicos, automóviles, etc., para distintas finalidades. Incluye una breve descripción y una referencia de cómo adquirirlo.
	uncrate.com	Revista enfocada en el hombre que muestra productos novedosos como herramientas, autos, dispositivos tecnológicos, películas, moda y accesorios masculinos.
Neutro	sciencemag.org	Sitio de contenido científico que contiene publicaciones de una gran variedad de temas, como arqueología, química o economía.
	columbia.edu	Sitio web de la Universidad de Columbia.
	wharton.upenn.edu	Sitio web de la escuela de negocios, Wharton.
	emol.com	Sitio web de noticias enfocado en el entorno de Chile.
	wikipedia.org	Enciclopedia online.
Femenino	elizabethstreet.com	Sitio web enfocado en la mujer adulta, que incluye artículos sobre estilo, belleza, viajes, cocina, diseño, entre otros.
	designsponge.com	Sitio web con temas de diseño de interiores, viajes por el mundo, recetas, estilo de vida, entre otros.
	enfemenino.com	Revista enfocada en la mujer, que incluye temas sobre el horóscopo, sexualidad, belleza, dietas, moda femenina, maternidad, decoración, cocina y vida en pareja.
	divinecaroline.com	Blog enfocado en la mujer, que trata temas de amor, sexualidad, maquillaje, moda, estilo de vida, entre otros.

Tabla 3.4: Sitios Web

Fuente: Elaboración propia

Estos sitios web fueron guardados en el computador para no depender de las condiciones de internet y para no correr el riesgo de que experimenten cambios, lo que aumentaría la variabilidad de los resultados. Y la resolución que se usó para todos los sujetos fue de 1920 x 1080 pixeles.

Las siguientes figuras, muestran ejemplos de los sitios web seleccionados, siendo masculino, neutro y femenino respectivamente.

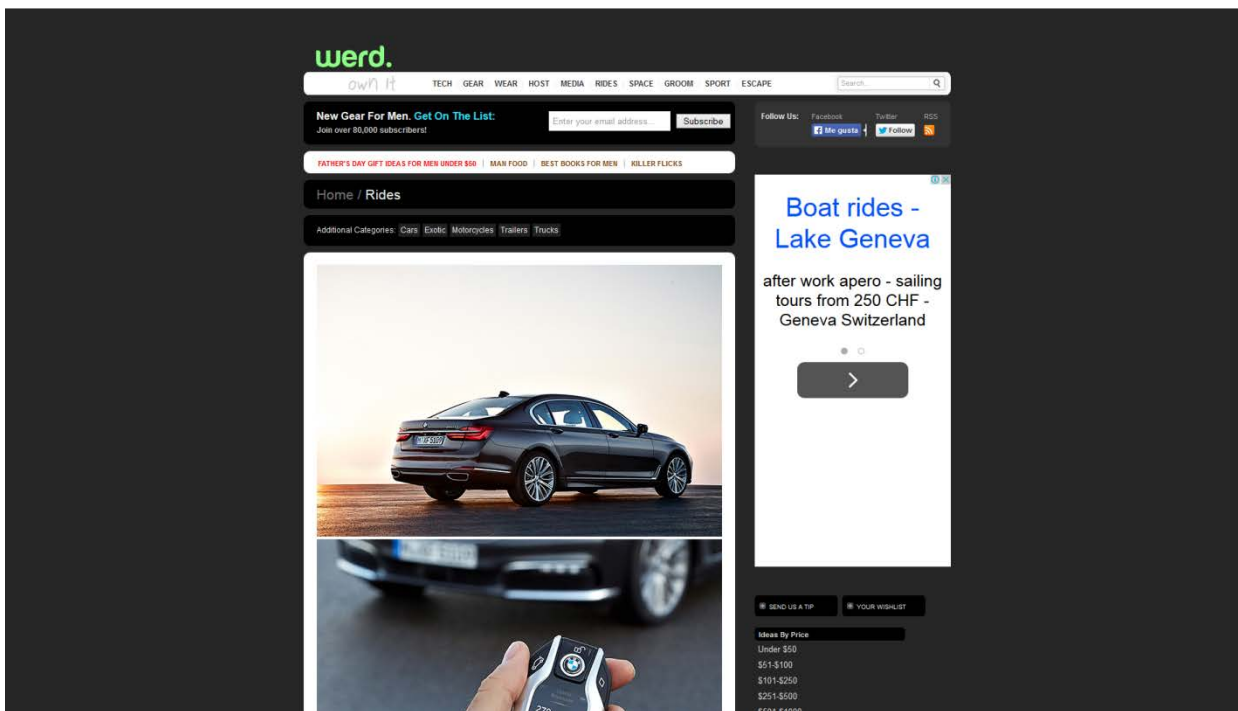


Figura 3.4: werd.com  
Fuente: Elaboración propia

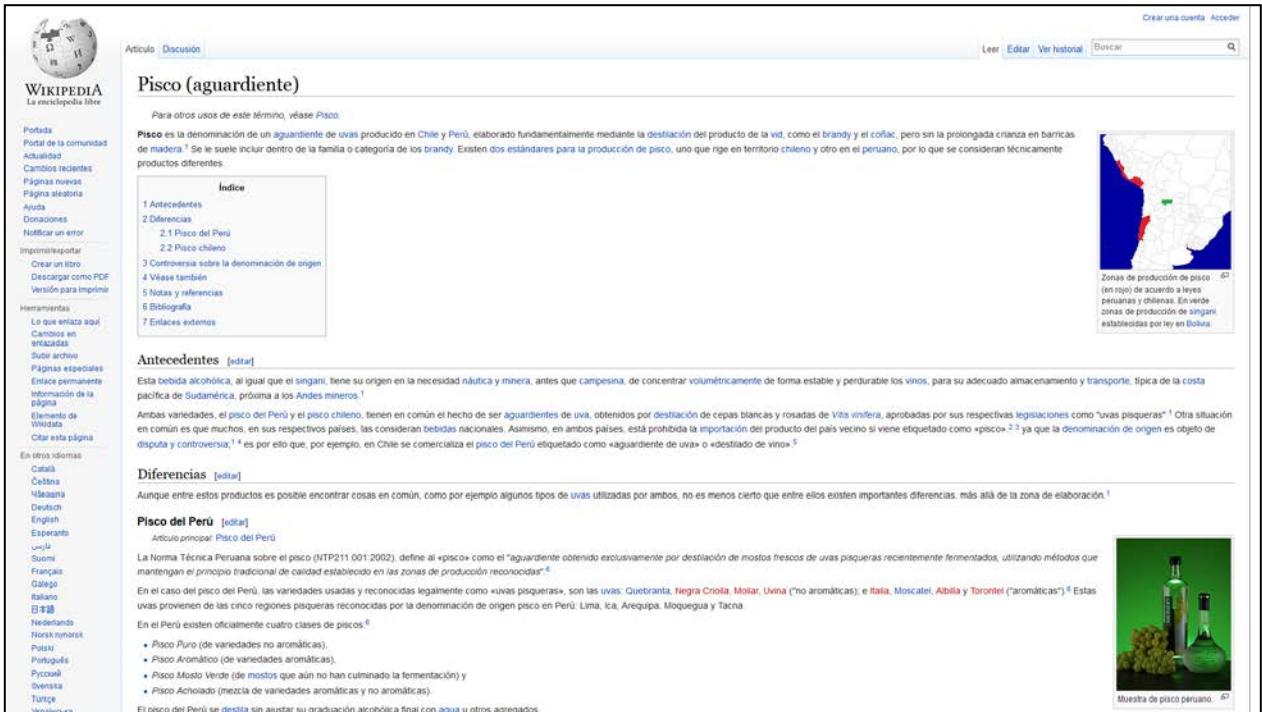


Figura 3.5: wikipedia.org  
Fuente: Elaboración propia

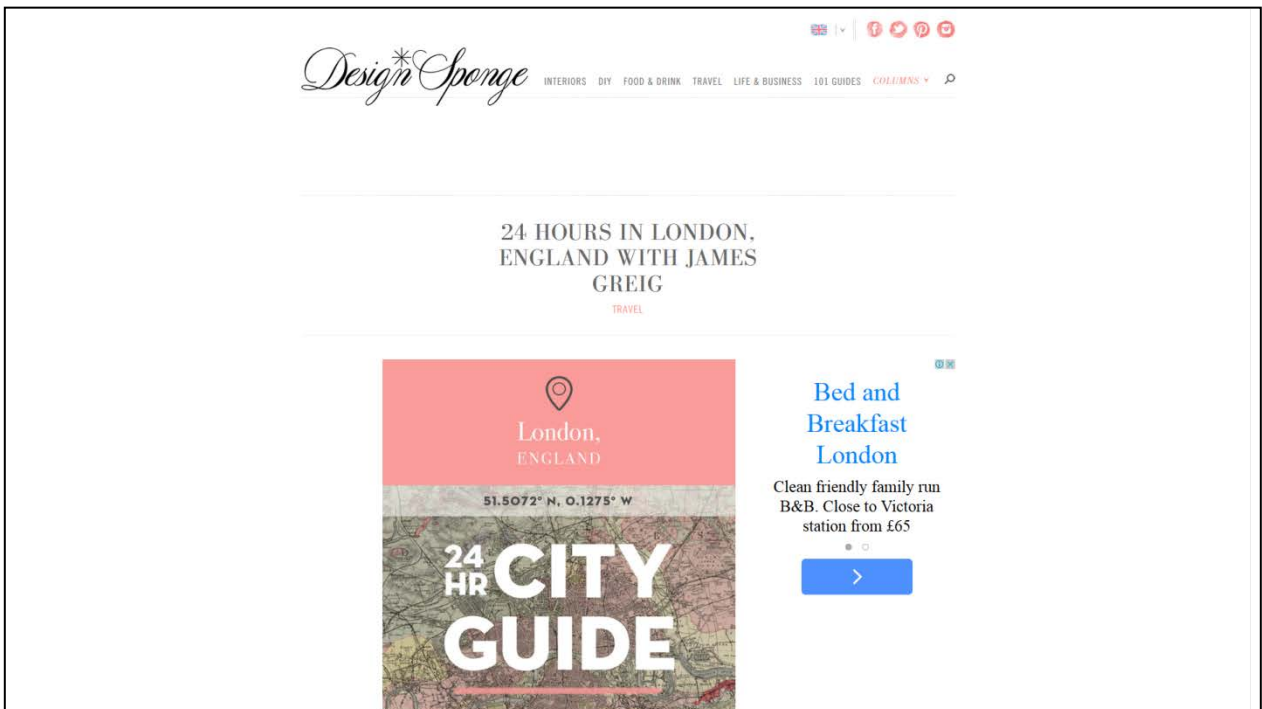


Figura 3.6: designsponge.com  
Fuente: Elaboración propia

### 3.2.3. Procedimiento

En esta sección se describen los pasos a seguir por el investigador y las instrucciones que debe ejecutar el sujeto en el experimento.

Para explicar el procedimiento, hay que mencionar que para el experimento se creó una tabla en Microsoft Excel con las preguntas cuyas respuestas el sujeto debe buscar. Una columna contiene la pregunta y la columna a su derecha contiene la subpágina del sitio web donde se encuentra la respuesta. Como se mencionó en la sección 3.1.4, el sujeto no debe navegar por el sitio web, sino que debe buscar en la subpágina que se la ha dado.

En cada tipo de sitio (masculino, neutro, femenino), el sujeto debe responder 9 preguntas distintas en 9 subpáginas web distintas, distribuidas por todos los sitios web de la categoría. Por ejemplo, 3 preguntas pueden estar en un sitio web masculino, pero cada pregunta se responde en una subpágina web distinta dentro del sitio web. En total, son 27 preguntas (Anexo A). Y todas las preguntas están ordenadas de forma aleatoria en Microsoft Excel. Por último, se ha aplicado un filtro en la tabla de Microsoft Excel para que el sujeto vaya leyendo sólo una pregunta a la vez. Esta tabla tiene que estar hecha para cada sujeto antes de su llegada.

También es muy importante que la instrumentación y el laboratorio estén en condiciones para realizar el experimento. Por ejemplo, el ruido o la suciedad podrían causar alteraciones en los resultados. Los equipos tienen que estar en perfectas condiciones, en particular el equipo de electroencefalograma debe tener todos sus electrodos y la batería debe ser suficiente para que dure todo el proceso. Así mismo, el computador no debe tener características que perjudiquen el experimento.

Teniendo lo anterior en consideración, el procedimiento es así:

1. **Recibir al sujeto.** En primer lugar, se debe dar la bienvenida al sujeto y agradecer por participar voluntariamente en el experimento. En esta parte se le pasa al sujeto el Consentimiento Informado, en el que se expresan los objetivos del experimento, los derechos del sujeto, responsables, entre otros (Anexo B). Cabe decir que las tomas de muestra del proyecto, están a cargo en un Ingeniero Civil Industrial, quien es la persona que asume la responsabilidad en el Consentimiento Informado. Además, se realizaron dos experimentos en la misma instancia, una con fines de otros investigadores y la otra con los fines de esta memoria, que es lo que sale además expresado en el Consentimiento Informado.

Luego, el sujeto debe estar de acuerdo con lo expresado en el Consentimiento informado y firmarlo. Si no lo firma, el procedimiento no continúa.

2. **Preparar equipo de electroencefalograma.** Una vez que el sujeto haya firmado el Consentimiento Informado, se procede a preparar el equipo de electroencefalograma. Como se mencionó en la sección 3.1.1 este equipo es inalámbrico de fácil instalación en comparación a otros tipos de equipos. Su preparación consiste de los siguientes pasos:

- 1) Preparar los 16 electrodos. Cada electrodo se ensambla al equipo del electroencefalograma. Estos electrodos se componen principalmente de una placa de metal y una esponja, la cual se humedece con solución salina para permitir la conductibilidad de los potenciales eléctricos. Esto se realiza en el hidratador, como muestra la siguiente figura.



Figura 3.7: Humectación de Electrodos  
Fuente: Imagen adaptada de [79]

- 2) Ensamblar electrodos al equipo.
- 3) Conectar el receptor USB al computador, encender el equipo de electroencefalograma y verificar que haya hecho conexión con el computador.
- 4) Poner el equipo de electroencefalograma en la cabeza del sujeto.
- 5) Ajustar los electrodos y verificar que cada uno de ellos tenga una buena señal. Para esto, se utiliza el software habilitado por Emotiv que complementa al equipo. A través de este se visualiza la posición de cada electrodo y la calidad de su señal, representada con colores (rojo indica muy mala señal; verde indica muy buena señal). En la Figura 3.9 se muestra un ejemplo de todos los electrodos con buena señal. Con esto se acaba la preparación del equipo de electroencefalograma.

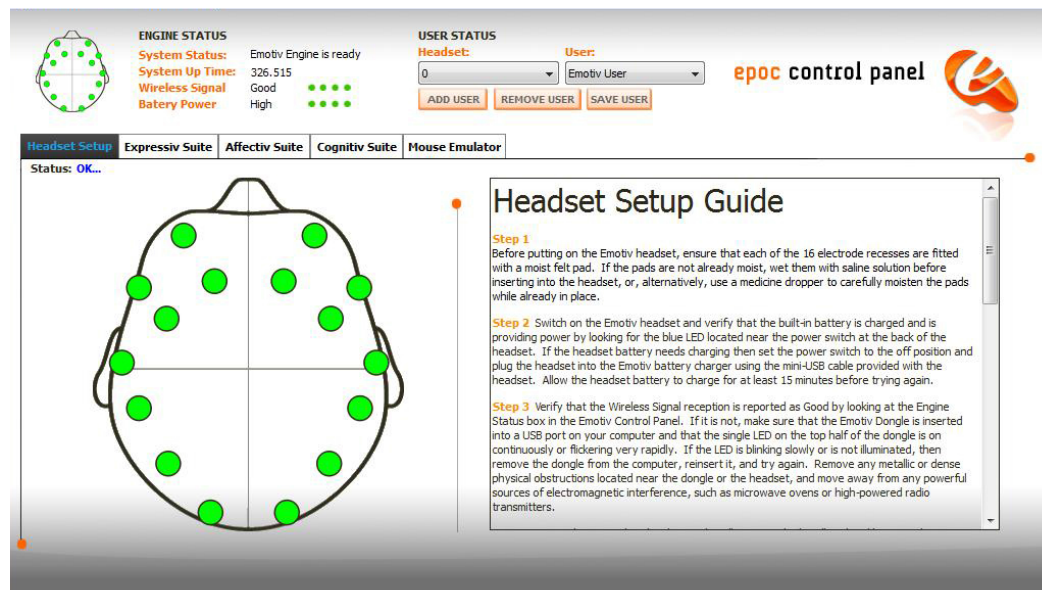


Figura 3.8: Software de Emotiv  
Fuente: Imagen de [79]



### 3. Preparación del eye tracker:

- 1) Acomodar al sujeto. Para que el eye tracker no pierda de vista al ojo, el sujeto debe apoyar su cabeza en un soporte que está a la altura de su cabeza estando sentado. El sujeto debe permanecer así por todo el experimento, por lo que debe estar lo más cómodo posible.
- 2) Una vez que el sujeto se sienta cómodo, se ajusta la cámara del eye tracker a la posición del sujeto para detectar su ojo. El ojo será visto en la pantalla a través del software Sofey Eye Tracker.
- 3) Se configura el software para que detecte la pupila de forma estable. En la siguiente figura se muestra cómo se ve el ojo y la pupila detectada en la ventana del software.

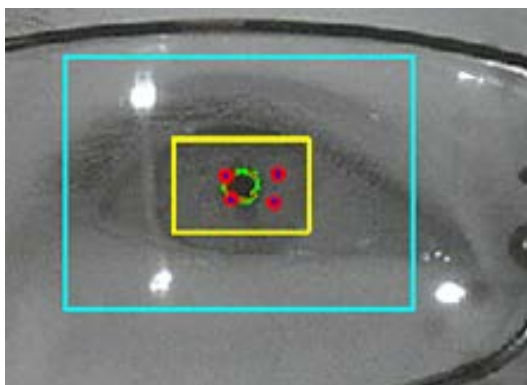


Figura 3.9: Pupila detectada vista en Sofey  
Fuente: Imagen de [80]

- 4) Calibrar el eye tracker, lo que significa que la posición de la mirada en la pantalla coincida con la posición detectada por el eye tracker. Para esto el sujeto debe ver 9 puntos distribuidos por la pantalla. Con esto se calibra el eye tracker. También puede ocurrir que algunos puntos no hayan quedado bien calibrados, por lo que se pueden realizar nuevamente sin afectar a los otros.

Una vez calibrado el eye tracker, está listo para ser utilizado.

4. **Abrir el navegador.** En este caso se utilizó Mozilla Firefox. Este viene con una herramienta habilitada por Sofey que registra a una frecuencia de 30 Hz, la hora, el sitio web en el que se está navegando, el *scroll* realizado, entre otras cosas.
5. **Dar instrucciones al sujeto.** Estas consisten en:
  - 1) Leer detenidamente la pregunta en Microsoft Excel.
  - 2) Ingresar al sitio web a través del hipervínculo adyacente a la pregunta. Al hacer esto, se irá directamente al navegador en una nueva pestaña. Aquí el navegador tendrá dos pestañas: en la que estaba inicialmente (“home”) y la que abrió el sujeto.
  - 3) Buscar la respuesta sin volver a leer la pregunta y sin visitar otras subpáginas del sitio web.

- 4) Cuando la encuentre, seleccionar el objeto que incluye la respuesta, ya sea texto o imagen.
  - 5) Cerrar la pestaña, sin cerrar la otra (“home”) para que no se cierre el navegador.
  - 6) Ir a Microsoft Excel, filtrar la siguiente pregunta, y repetir el procedimiento hasta terminar con las 27 preguntas (Anexo C).
6. **Aislamiento del sujeto y grabación de equipos.** Apagar la luz y cerrar la cortina que separa la pantalla del sujeto y la pantalla de control. Luego se comienza con el registro de datos del equipo de EEG, del eye tracker y de la herramienta de Mozilla Firefox (es por esta herramienta que el navegador tiene que permanecer abierto durante todo el proceso, sino la herramienta deja de grabar).
  7. **Dar inicio al experimento.** El sujeto realiza lo descrito en el paso 5.
  8. **Detener grabación.** Una vez terminado, se detiene la grabación del equipo de EEG, del eye tracker y del navegador.
  9. **Despedir y agradecer al sujeto.**

#### **3.2.4. Datos**

De los experimentos, se obtuvieron 19 registros de electroencefalograma (9 hombres y 10 mujeres) y 19 registros de eye tracker (10 hombres y 9 mujeres).

Cada registro de electroencefalograma consiste en un archivo (formato EDF) que contiene el potencial de todos los canales, desde que se comenzó con la grabación hasta que terminó. Y como se especificó en las características del Emotiv EPOC Neuroheadset, cada canal registra a una tasa de muestreo de 128 Hz.

Cada registro del eye tracker consiste de varios archivos (formato CSV). Uno es el relacionado a la mirada, en el que se muestra el *timestamp* (hora del computador), coordenadas de la mirada en píxeles, área de la pupila, ocurrencia de un movimiento sacádico y ocurrencia de un pestañeo. Otro es el relacionado al mouse, en el que se muestra el *timestamp*, evento del mouse (movimiento, click) y posición del puntero en píxeles. Otro archivo es el relacionado al teclado, en el que se muestra el *timestamp*, evento del teclado y qué tecla se relaciona al evento. El último archivo es el relacionado a la herramienta del navegador, que muestra el *timestamp*, tamaño de la pantalla en píxeles, tamaño de la vista del navegador en píxeles, *scroll* desde arriba, *scroll* desde la izquierda y el sitio web en el que se está en ese tiempo (*timestamp*). Además entrega *screenshots* de todos los sitios web que se visitaron durante la grabación de la herramienta.

# Capítulo 4: Análisis de Resultados

Una vez realizados los experimentos y haber cumplido con el segundo objetivo específico de esta memoria, se procede a analizar los datos obtenidos para poder rechazar o validar la hipótesis de investigación. Es en esta instancia en la que se implementa el proceso KDD, el cual fue descrito en la sección 2.4, por lo que es en este capítulo en el que se muestran los resultados finales de la investigación.

Hay que recalcar que todos los resultados obtenidos y mostrados en este capítulo se basan exclusivamente en los datos obtenidos de los experimentos, los que están sujetos a los aspectos del experimento descritos en el capítulo anterior (como la calidad de los instrumentos de medición).

También cabe decir que – dado que se requiere un conocimiento acerca del procesamiento de datos fisiológicos para enriquecer el análisis – este proceso fue apoyado por el Ph.D. Pedro Maldonado, quien es además el investigador principal del Laboratorio de Neurosistemas de la Universidad de Chile [81].

Se comienza este capítulo mostrando un análisis exploratorio acerca del comportamiento ocular de los sujetos durante en el experimento. Luego se describe el proceso KDD utilizando los datos de la pupila y la actividad bioeléctrica cerebral. Finalmente se discute acerca de los resultados obtenidos.

## 4.1. Comportamiento Ocular

Consiste principalmente de las fijaciones y sacadas del sujeto al momento de navegar por la Web. En esta sección se analiza el comportamiento ocular de forma exploratoria. Este análisis se divide entre los 3 tipos de sitios, con el objetivo de explorar si existen diferencias en el comportamiento ocular de hombres y mujeres en cada tipo de sitio.

Como se ha mencionado, el eye tracker registra el *timestamp* (hora del computador), ocurrencia de una sacada, posición de la mirada en píxeles y la URL de la página web en el que el sujeto está navegando, secuencialmente. De estos registros se puede obtener la siguiente información relacionada al comportamiento ocular en la Web:

1. Número de fijaciones de cada sujeto en cada subpágina web. Para este análisis se considerará una fijación, aquella observación que dure al menos 300 milisegundos.

2. Duración de cada fijación.
3. Número de sacadas de cada sujeto en cada subpágina web.
4. Tiempo total de navegación en cada subpágina web.

El número de sacadas consiste simplemente de la cantidad de eventos de sacadas registrados por el eye tracker. El tiempo total de navegación en cada subpágina web consiste de la diferencia entre el último *timestamp* registrado en la subpágina web y el primer *timestamp* registrado en la misma subpágina web. Por último, para la obtención de las fijaciones y su duración, se tuvo que realizar un proceso más elaborado, el cual se describe en la siguiente sección.

#### **4.1.1. Detección de Fijaciones**

Como se mencionó en el marco conceptual, una fijación ocurre cuando una persona observa un elemento de la escena fijamente, por un lapso de tiempo no menor a los 300 ms en este caso. Entonces, para saber si existe fijación en un elemento es necesario identificar esos elementos (website objects) y la posición de cada uno de ellos en la pantalla.

Para realizar esta tarea, se consideró un programa elaborado por un memorista anterior, el cual escaneaba un sitio web y arrojaba como resultado una lista de todos los website objects presentes en la página web en base al código HTML con sus posiciones en pixeles (posición de la esquina superior izquierda del objeto, su ancho y su alto en pixeles) y un *screenshot* del sitio web con los objetos delimitados por una línea roja.

Sin embargo, este programa no estaba adecuado a los requerimientos de esta memoria, principalmente porque no funcionaba con los sitios web utilizados en los experimentos y los objetos no eran fácilmente reconocibles en los *screenshots*.

Entonces, el programa tuvo que ser adaptado a los requerimientos de esta memoria para lo cual se tuvo la colaboración de un Ingeniero Civil en Computación de la Universidad de Chile. Con ello se desarrolló un programa que puede escanear un amplio espectro de tipos de sitios web, cuyos resultados son principalmente: una lista de objetos, con identificadores y sus coordenadas en pixeles; y un *screenshot* del sitio web con los objetos delimitados y con su identificador en la esquina superior izquierda. En la siguiente figura se muestra un ejemplo del *screenshot* de una parte de una subpágina web utilizada con sus objetos.

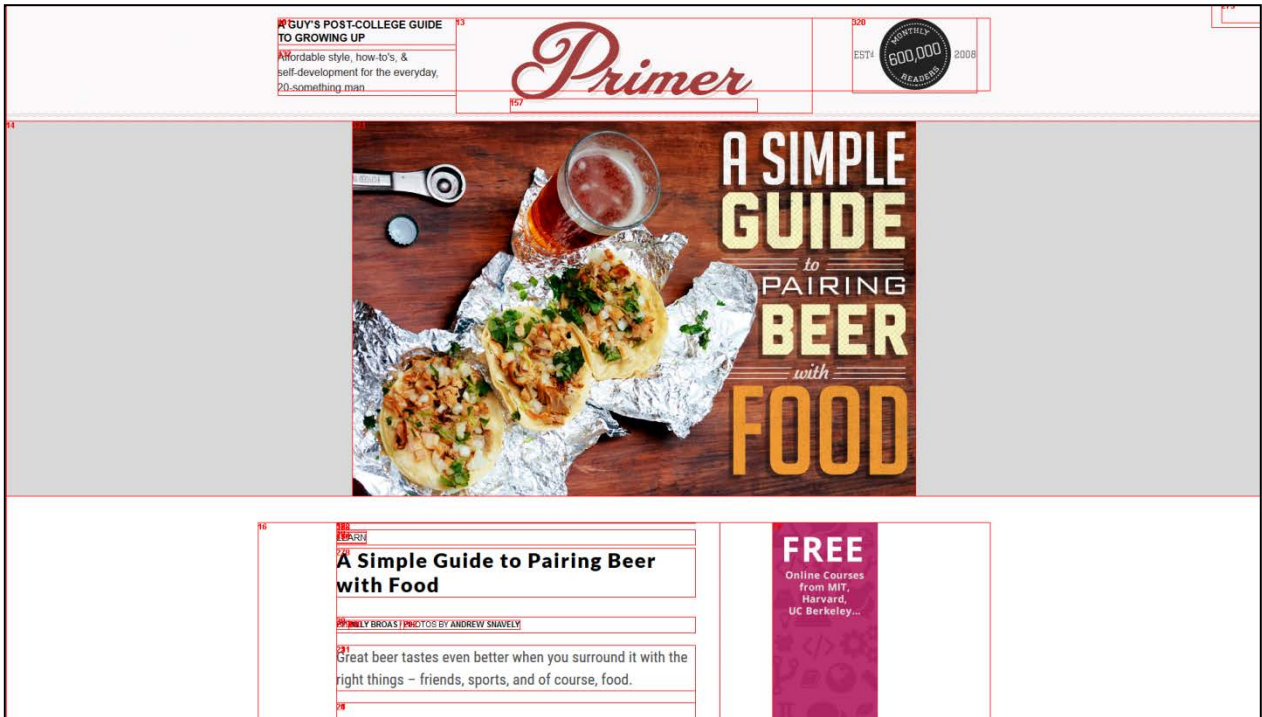


Figura 4.1: Screenshot con Objetos Marcados  
Fuente: Elaboración propia

Entonces, se consideran fijaciones aquellas observaciones en que la posición de la mirada del sujeto permanece al menos 300 ms dentro del cuadrado rojo que representa el objeto. Sin embargo, se observa que algunos de los objetos identificados por el programa poseen otros objetos en su interior, lo que distorsiona los datos. Por ejemplo, si un sujeto observa un objeto que está en el interior de otro por al menos 300 ms, se considerará que existen dos fijaciones y no una como es en realidad.

Para resolver esto, se procedió a seleccionar los objetos a considerar de forma manual, utilizando su identificador ubicado en su esquina superior izquierda. Con esto la cantidad de objetos disminuyó considerablemente y con ello la cantidad de datos, pero si no se realiza esto no se obtendrían resultados válidos. En la siguiente figura se muestra el mismo sitio web de la Figura 4.1 con sus objetos filtrados. Y finalmente, la cantidad de objetos a considerar en las 27 subpáginas web, se muestra en la Tabla 4.1.

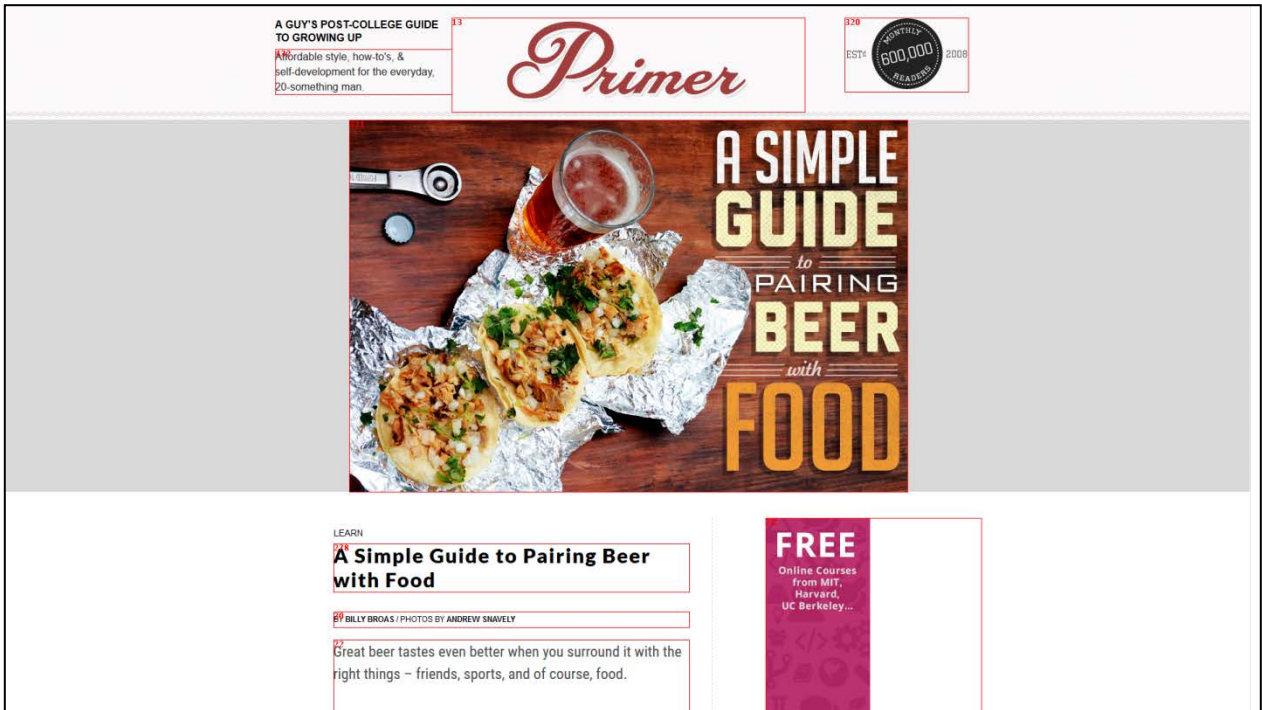


Figura 4.2: Screenshot con Objetos Seleccionados  
Fuente: Elaboración propia

Tipo de Sitio Web	Sitio Web	Subpágina Web	Cantidad de Objetos	Total	Promedio por subpágina web	
Masculino	gq.com.mx	1	78	632	70,22	
		2	50			
		3	109			
	wired.com	4	51			
		primermagazine.com	5			49
			6			54
		businessinsider.com	7			134
		werd.com	8			61
		uncrate.com	9			46
Neutro	sciencemag.org	1	38	379	42,11	
		2	33			
		3	30			
	columbia.edu	4	35			
		5	99			
	wharton.upenn.edu	6	69			
	emol.com	7	44			
	wikipedia.org	8	23			
		9	8			
Femenino	elizabethstreet.com	1	23	417	46,33	
		2	31			
		3	67			
	designsponge.com	4	73			
		5	37			
		6	72			
	enfemenino.com	7	48			
		8	36			
	divinecaroline.com	9	30			

Tabla 4.1: Cantidad de Objetos Seleccionados  
Fuente: Elaboración propia

## 4.1.2. Resultados

En esta sección se muestran los resultados que se han podido obtener en base a los datos del comportamiento ocular. Estos resultados serán divididos por tipo de sitio web y por género. Utilizando el lenguaje Matlab, se han calculado los siguientes valores:

1. **Número total de fijaciones (TF):** corresponde a todas las fijaciones que tuvo el sujeto en un tipo de sitio.
2. **Promedio de fijaciones (PF):** corresponde al promedio de todas las fijaciones que tuvo el sujeto en una pregunta del tipo de sitio. Es el valor anterior dividido por el número de preguntas.
3. **Duración promedio de las fijaciones en el tipo de sitio (DF).**
4. **Número total de sacadas (TS):** corresponde a todas las sacadas que tuvo el sujeto en un tipo de sitio.
5. **Promedio de sacadas (PS):** corresponde al promedio de todas las sacadas que tuvo el sujeto en una pregunta del tipo de sitio (es el valor anterior dividido por el número de preguntas).
6. **Duración total promedio (DTP):** corresponde a la duración promedio de navegación del sujeto en una pregunta en el tipo de sitio.
7. **Promedio de fijaciones por segundo (F/s):** una mujer puede tener muchas más fijaciones que un hombre en un tipo de sitio, pero a la vez pudo haber estado mucho más tiempo, por lo que sería lógico que la mujer tuviera más fijaciones. Este valor refleja la cantidad de fijaciones por unidad de tiempo. Se calcula dividiendo la duración total promedio al promedio de fijaciones.
8. **Promedio de sacadas por segundo (Sac/s):** misma lógica que el valor anterior. Este valor refleja la cantidad de sacadas por unidad de tiempo. Se calcula dividiendo la duración total promedio al promedio de sacadas.

### 4.1.2.1. Sitio Web Neutro

En este tipo de sitio, los resultados son los siguientes:

Hombre	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	241	26,78	861,50	185	23,13	14,78	1,81	1,56
2	202	22,44	1083,00	244	30,50	18,79	1,19	1,62
3	282	31,33	1240,20	394	49,25	26,40	1,19	1,87
4	418	46,44	1361,10	345	43,13	29,77	1,56	1,45
5	244	27,11	1156,80	267	33,38	23,32	1,16	1,43
6	253	28,11	1216,80	271	33,88	21,67	1,30	1,56
7	229	25,44	1267,00	292	36,50	20,29	1,25	1,80
8	309	34,33	1039,90	309	38,63	23,15	1,48	1,67
9	250	27,78	988,77	336	42,00	23,32	1,19	1,80
10	227	25,22	1065,70	218	27,25	17,11	1,47	1,59
<b>Promedio</b>	<b>265,50</b>	<b>29,50</b>	<b>1128,08</b>	<b>286,10</b>	<b>35,76</b>	<b>21,86</b>	<b>1,36</b>	<b>1,64</b>
<b>Desv. Estándar</b>	<b>61,22</b>	<b>6,80</b>	<b>148,71</b>	<b>62,70</b>	<b>7,84</b>	<b>4,39</b>	<b>0,21</b>	<b>0,15</b>

Tabla 4.2: Comportamiento Ocular del Hombre en Sitio Web Neutro  
Fuente: Elaboración propia

Mujer	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	188	00,89	1413,70	241	30,13	16,98	1,23	1,77
2	333	37,00	1055,50	299	37,38	21,89	1,69	1,71
3	363	40,33	1303,60	290	36,25	27,60	1,46	1,31
4	277	30,78	1209,00	290	36,25	21,59	1,43	1,68
5	246	27,33	1167,80	249	31,13	26,02	1,05	1,20
6	420	46,67	805,17	337	42,13	28,34	1,65	1,49
7	463	51,44	1051,70	493	61,63	52,28	0,98	1,18
8	344	38,22	1412,50	329	41,13	28,83	1,33	1,43
9	126	14,00	1246,60	327	40,88	25,08	0,56	1,63
<b>Promedio</b>	<b>306,67</b>	<b>34,07</b>	<b>1185,06</b>	<b>317,22</b>	<b>39,65</b>	<b>27,62</b>	<b>1,26</b>	<b>1,49</b>
<b>Desv. Estándar</b>	<b>108,30</b>	<b>12,03</b>	<b>194,17</b>	<b>73,99</b>	<b>9,25</b>	<b>10,01</b>	<b>0,36</b>	<b>0,22</b>

Tabla 4.3: Comportamiento Ocular de la Mujer en Sitio Web Neutro  
Fuente: Elaboración propia

Los indicadores que permiten comparar la cantidad de fijaciones y la cantidad de sacadas entre hombres y mujeres, son F/s y Sac/s respectivamente. De las tablas se observa que el hombre posee en promedio más fijaciones por unidad de tiempo y más sacadas por unidad de tiempo que las mujeres en un sitio web neutro. Por otro lado la duración de las fijaciones es en promedio mayor en las mujeres y el tiempo en resolver una pregunta también es mayor en las mujeres en un sitio web neutro.

El siguiente paso es ver si estas diferencias son significativas. Para esto, primero se prueba si las variables siguen una distribución normal, para lo cual se puede realizar el test de Lilliefors.



Este es un test de normalidad cuya hipótesis nula señala que la muestra sigue una distribución normal, que a diferencia del test de normalidad de Kolmogorov – Smirnov, este test no requiere que la media y varianza de la población sean parámetros conocidos [82].

Entonces, se aplica el test de Lilliefors utilizando Matlab [83] a los indicadores de interés, que son: promedio de fijaciones por segundo (F/s), promedio de sacadas por segundo (Sac/s) y la duración de las fijaciones (DF).

Los resultados del test de Lilliefors de normalidad muestran que los 3 indicadores para los hombres y las mujeres, siguen una distribución normal con un 5% de significancia. Una vez sabido esto, se procede a aplicar el test de diferencias para muestras normales. Este es el T-test para dos muestras, cuya hipótesis nula señala que las medias de dos muestras son estadísticamente iguales. Este se aplicó a cada indicador utilizando Matlab [84], buscando si este era distinto entre hombres y mujeres en un sitio web neutro.

Finalmente los resultados no rechazan la hipótesis nula con un 5% de significancia, por lo que no se podría decir que la cantidad de fijaciones, cantidad de sacadas y la duración de las fijaciones son estadísticamente distintas entre hombres y mujeres en un sitio web neutro.

#### 4.1.2.2. *Sitio Web Masculino*

En este tipo de sitio, los resultados son los siguientes:

Hombre	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	180	0,00	709,44	284	31,56	16,88	1,19	1,87
2	226	25,11	797,48	356	39,56	20,87	1,20	1,90
3	412	45,78	778,44	594	66,00	34,28	1,34	1,93
4	352	39,11	1094,70	424	47,11	26,16	1,50	1,80
5	429	47,67	957,25	567	63,00	41,34	1,15	1,52
6	351	39,00	1345,20	406	45,11	30,01	1,30	1,50
7	215	23,89	767,56	387	43,00	20,65	1,16	2,08
8	249	27,67	812,65	359	39,89	26,56	1,04	1,50
9	275	30,56	837,30	301	33,44	19,31	1,58	1,73
10	374	41,56	893,81	394	43,78	24,04	1,73	1,82
<b>Promedio</b>	<b>306,30</b>	<b>34,03</b>	<b>899,38</b>	<b>407,20</b>	<b>45,24</b>	<b>26,01</b>	<b>1,32</b>	<b>1,77</b>
<b>Desv. Estándar</b>	<b>88,13</b>	<b>9,79</b>	<b>191,52</b>	<b>101,49</b>	<b>11,28</b>	<b>7,51</b>	<b>0,22</b>	<b>0,20</b>

Tabla 4.4: Comportamiento Ocular del Hombre en Sitio Web Masculino

Fuente: Elaboración propia

Mujer	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	247	27,44	1001,50	304	33,78	19,28	1,42	1,75
2	422	46,89	949,38	512	56,89	29,27	1,60	1,94
3	457	50,78	1379,20	510	56,67	38,68	1,31	1,46
4	320	35,56	893,44	472	52,44	28,09	1,27	1,87
5	575	63,89	900,67	797	88,56	65,12	0,98	1,36
6	369	41,00	942,21	386	42,89	30,37	1,35	1,41
7	278	30,89	613,51	562	62,44	61,75	0,50	1,01
8	463	51,44	1512,30	728	80,89	57,22	0,90	1,41
9	69	7,67	818,93	493	54,78	27,31	0,28	2,01
<b>Promedio</b>	<b>355,56</b>	<b>39,51</b>	<b>1001,24</b>	<b>529,33</b>	<b>58,81</b>	<b>39,68</b>	<b>1,07</b>	<b>1,58</b>
<b>Desv. Estándar</b>	<b>148,22</b>	<b>16,47</b>	<b>277,38</b>	<b>153,62</b>	<b>17,07</b>	<b>17,11</b>	<b>0,44</b>	<b>0,33</b>

Tabla 4.5: Comportamiento Ocular de la Mujer en Sitio Web Masculino

Fuente: Elaboración propia

De las tablas se observa que el hombre posee en promedio más fijaciones por unidad de tiempo y más sacadas por unidad de tiempo que las mujeres en un sitio web masculino. Por otro lado la duración de las fijaciones es en promedio mayor en las mujeres y el tiempo en resolver una pregunta también es mayor en las mujeres en un sitio web neutro.

Los siguientes pasos son los mismos a los aplicados en los tipos de sitios web neutros.

Los resultados del test de Lilliefors de normalidad muestran que F/s, Sac/s y DF para los hombres y las mujeres, siguen una distribución normal con un 5% de significancia. Por lo tanto luego se aplica el T-test.

Los resultados del T-test no rechazan la hipótesis nula con un 5% de significancia, por lo que tampoco se podría decir que la cantidad de fijaciones, cantidad de sacadas y la duración de las fijaciones son estadísticamente distintas entre hombres y mujeres en un sitio web masculino.

#### **4.1.2.3. Sitio Web Femenino**

En este tipo de sitio, los resultados son los siguientes:

Hombre	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	163	18,11	983,27	347	38,56	20,84	0,87	1,85
2	132	14,67	997,53	319	35,44	20,88	0,70	1,70
3	217	24,11	970,71	557	61,89	32,55	0,74	1,90
4	288	32,00	1309,20	717	79,67	44,37	0,72	1,80
5	223	24,78	1015,50	621	69,00	44,52	0,56	1,55
6	157	17,44	1379,10	292	32,44	20,87	0,84	1,55
7	150	16,67	1018,40	442	49,11	24,47	0,68	2,01
8	207	23,00	1071,00	490	54,44	32,99	0,70	1,65
9	177	19,67	1092,50	358	39,78	21,13	0,93	1,88
10	214	23,78	1371,10	458	50,89	28,25	0,84	1,80
<b>Promedio</b>	<b>192,80</b>	<b>21,42</b>	<b>1120,83</b>	<b>460,10</b>	<b>51,12</b>	<b>29,09</b>	<b>0,76</b>	<b>1,77</b>
<b>Desv. Estándar</b>	<b>46,14</b>	<b>5,13</b>	<b>165,45</b>	<b>139,00</b>	<b>15,44</b>	<b>9,35</b>	<b>0,11</b>	<b>0,15</b>

Tabla 4.6: Comportamiento Ocular del Hombre en Sitio Web Femenino  
Fuente: Elaboración propia

Mujer	TF	PF	DF (ms)	TS	PS	DTP (s)	F/s	Sac/s
1	131	14,56	1048,00	301	33,44	19,07	0,76	1,75
2	244	27,11	1214,60	491	54,56	28,90	0,94	1,89
3	218	24,22	1575,80	412	45,78	31,62	0,77	1,45
4	190	21,11	1124,60	504	56,00	30,13	0,70	1,86
5	264	29,33	1122,90	549	61,00	38,32	0,77	1,59
6	332	36,89	987,88	515	57,22	35,12	1,05	1,63
7	221	24,56	702,15	662	73,56	64,07	0,38	1,15
8	405	45,00	1633,10	767	85,22	59,00	0,76	1,44
9	81	9,00	1102,90	472	52,44	27,20	0,33	1,93
<b>Promedio</b>	<b>231,78</b>	<b>25,75</b>	<b>1167,99</b>	<b>519,22</b>	<b>57,69</b>	<b>37,05</b>	<b>0,72</b>	<b>1,63</b>
<b>Desv. Estándar</b>	<b>97,58</b>	<b>10,84</b>	<b>286,53</b>	<b>134,66</b>	<b>14,96</b>	<b>14,92</b>	<b>0,23</b>	<b>0,26</b>

Tabla 4.7: Comportamiento Ocular de la Mujer en Sitio Web Femenino  
Fuente: Elaboración propia

De las tablas se observa que el hombre posee en promedio más fijaciones por unidad de tiempo y más sacadas por unidad de tiempo que las mujeres en un sitio web femenino. Por otro lado la duración de las fijaciones es en promedio mayor en las mujeres y el tiempo en resolver una pregunta también es mayor en las mujeres en un sitio web femenino.

Los siguientes pasos son los mismos a los aplicados en los tipos de sitios web anteriores.

Los resultados del test de Lilliefors de normalidad muestran que F/s, Sac/s y DF para los hombres y las mujeres, siguen una distribución normal con un 5% de significancia. Por lo tanto luego se aplica el T-test.

Los resultados del T-test no rechazan la hipótesis nula con un 5% de significancia, por lo que tampoco se podría decir que la cantidad de fijaciones, cantidad de sacadas y la duración de las fijaciones son estadísticamente distintas entre hombres y mujeres en un sitio web femenino.

### **4.1.3. Conclusiones**

De las tablas, se puede observar que el hombre tiene en promedio más fijaciones por segundo y más sacadas por segundo que las mujeres en los 3 tipos de sitios. La duración de las fijaciones es en promedio mayor en las mujeres y el tiempo en resolver una pregunta también es mayor en las mujeres en los 3 tipos de sitios. Esto último tiene relación con la literatura, en la que se muestra que la mujer realiza una búsqueda más exhaustiva que el hombre.

Se realizó un test de normalidad a tres indicadores de interés para comparar entre hombres y mujeres: promedio de fijaciones por segundo (F/s), promedio de sacadas por segundo (Sac/s) y la duración de las fijaciones (DF). Se encontró que todos seguían una distribución normal.

Finalmente se aplicó un test para detectar diferencias entre dos muestras normales. Con esto se concluye que ningún indicador es significativamente distinto entre hombres y mujeres en los 3 tipos de sitios.

Este análisis ha sido de forma exploratoria y sus resultados no tienen por qué desalentar el análisis posterior. Como se ha mencionado, las fijaciones y las sacadas tienen relación a la atención visual. Un objeto puede llamar la atención pero no se sabe la reacción que provoca este objeto en la persona, por lo que sigue siendo interesante estudiar el comportamiento pupilar y la actividad bioeléctrica cerebral en el entorno web.

## **4.2. Proceso KDD**

En esta sección se describe lo que es el proceso KDD, utilizando los datos de dilatación pupilar y potencial eléctrico del cerebro, los cuales fueron procesados utilizando principalmente Matlab.

El orden de esta sección se relaciona a cada etapa del proceso KDD, es decir, se comienza describiendo la fase de selección de datos de la pupila y actividad bioeléctrica cerebral, hasta finalizar con la fase de evaluación e interpretación de los resultados de la minería de datos.

## **4.2.1. Selección**

En esta sección se describen los datos seleccionados para el análisis posterior y se divide de acuerdo a las 2 herramientas de medición que se utilizan en esta memoria, el eye tracker y el equipo de electroencefalograma.

### **4.2.1.1. Eye Tracker**

Como se mencionó, se obtuvo el registro exitoso del eye tracker de 10 hombres y 9 mujeres. Cada uno de estos registros cuenta con varios archivos (formato CSV) relacionados a: la visión, el mouse, el teclado y navegación en la Web. El único archivo que no se utiliza por completo es el relacionado al teclado. Se recuerda además que el eye tracker registra secuencialmente, a una tasa aproximada de 20 datos por segundo.

Entonces para el análisis posterior, en relación al archivo de visión se selecciona: el *timestamp*, las coordenadas (x, y) de la mirada en pixeles, el área de la pupila, marcador de un evento de sacada y el marcador de un evento de pestañeo. Del archivo del mouse se selecciona: el *timestamp* y el evento del mouse. Y del archivo de navegación se selecciona: el *timestamp*, *scroll* desde arriba del sitio web y el identificador del sitio web.

### **4.2.1.2. Equipo de Electroencefalograma**

Del equipo de electroencefalograma, se obtuvo el registro exitoso de 9 hombres y 10 mujeres. Cada uno de estos registros consiste de un archivo (formato EDF) que contiene el registro de datos de los 14 canales y la hora de inicio de grabación. También cabe recordar que cada canal registra el potencial eléctrico del cerebro a una tasa de 128 datos por segundo.

Como se indicó, el equipo de electroencefalograma tiene sus electrodos posicionados en 14 puntos distintos: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 y AF4 (más los 2 de referencia con los que no se captura información).

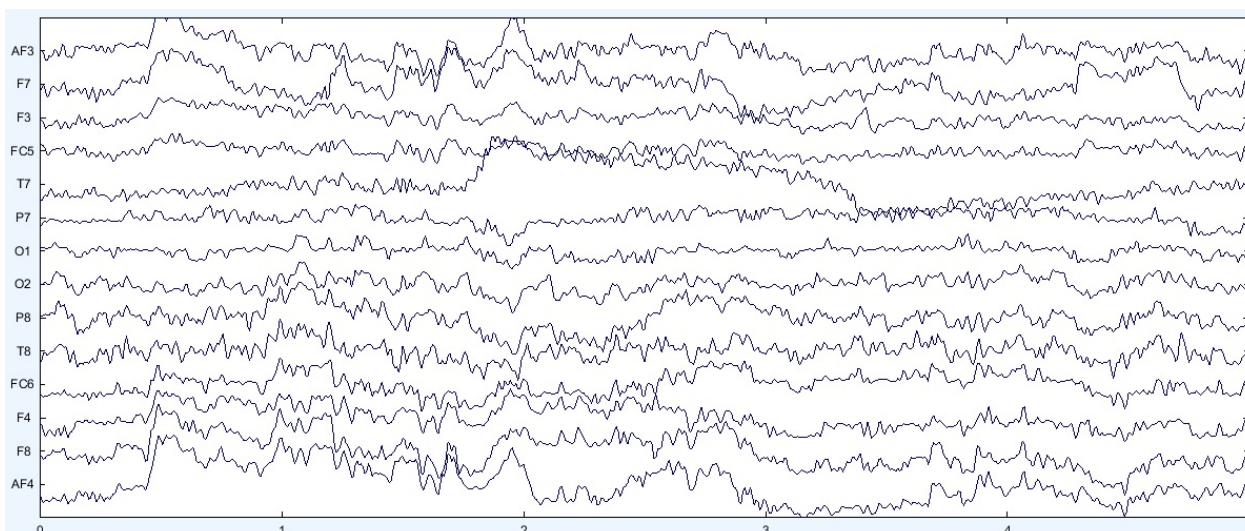


Figura 4.3: Ejemplo de 5 segundos de registro de los 14 canales  
Fuente: Elaboración propia

Debido a que cada canal aporta una gran cantidad de datos y complejidad al análisis, se decide optar por un subconjunto de estos canales como primera instancia de la investigación. Dado que el lóbulo occipital tiene relación con los estímulos visuales en la persona y por recomendación del Ph.D. Pedro Maldonado, se han seleccionado los canales que posee el equipo del electroencefalograma que están relacionados a este lóbulo, que serían el canal O1 y O2 [54]. Por lo tanto, para las siguientes etapas se ha seleccionado el potencial eléctrico del cerebro medido en el canal O1 y O2.

## 4.2.2. Pre – Procesamiento

En esta sección se describe la segunda fase del proceso KKD. Primero se muestra el pre-procesamiento de la dilatación pupilar y luego se muestra el pre-procesamiento del potencial eléctrico del cerebro.

### 4.2.2.1. Dilatación Pupilar

Los datos seleccionados tienen que ser limpiados para continuar con el proceso. Los datos del tamaño de la pupila pueden poseer ruido originado por 3 causas principales:

1. **Parpadeos:** este es un acto fisiológico natural del ojo, pero es una causa de ruido en el registro del eye tracker. Esto porque al ocurrir un parpadeo, el eye tracker no detecta la pupila, por lo que no se registra el tamaño de la pupila en el tiempo en el que ocurrió el parpadeo.

Para corregir esto, los valores perdidos de los parpadeos se obtienen a través de una interpolación lineal.

La Figura 4.4 muestra una señal sin el registro del eye tracker y la Figura 4.5 muestra la misma señal corregida.

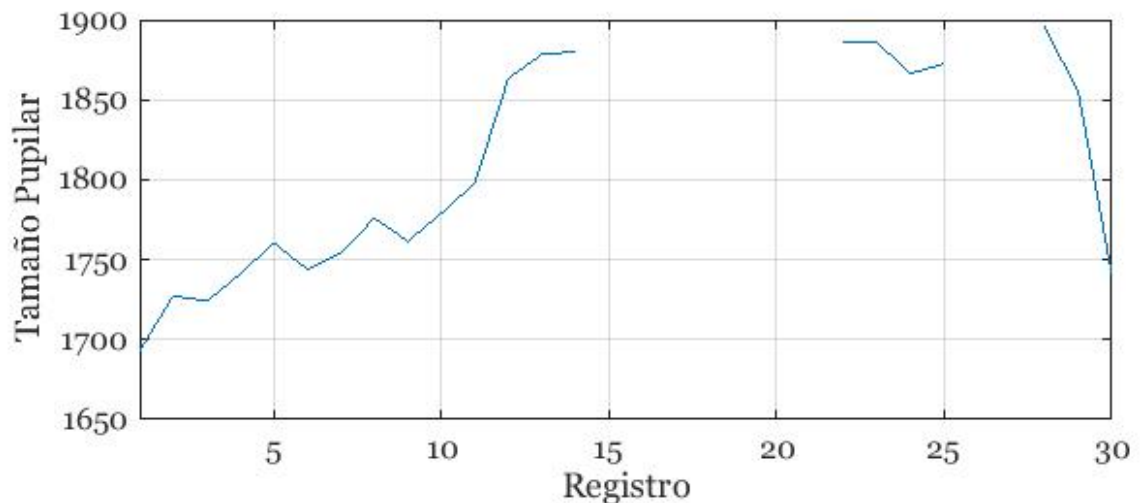


Figura 4.4: Ejemplo de señal sin registro del Eye Tracker  
Fuente: Elaboración propia

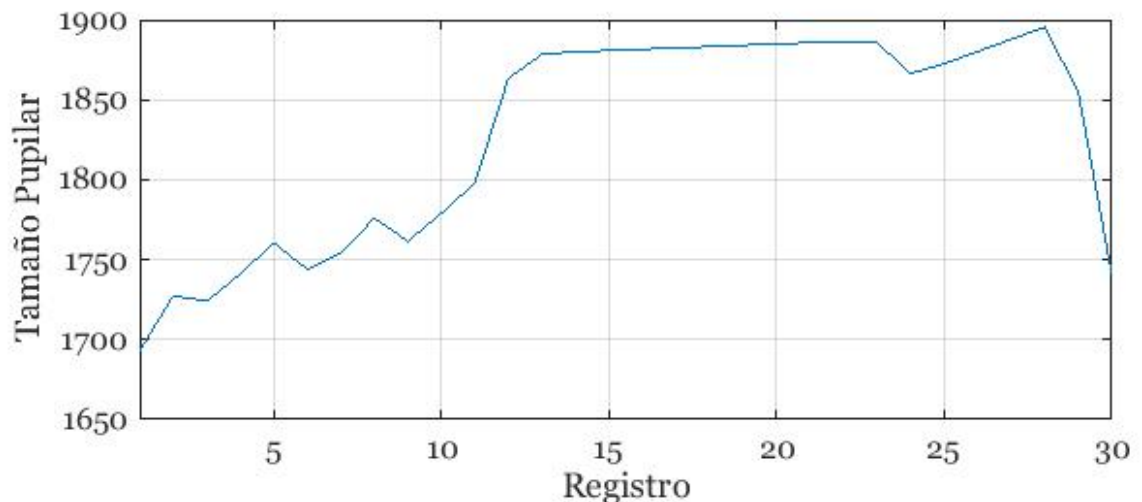


Figura 4.5: Ejemplo de señal corregida con Interpolación  
Fuente: Elaboración propia

2. **Movimientos sacádicos:** las sacadas son movimientos bruscos del globo ocular que son una causa de ruido, ya que al cambiar muy rápidamente la posición de la pupila, el ángulo con el que se captura la señal cambia, pudiendo producir diferencias significativas en el tamaño de la pupila que no reflejan la reacción ante un estímulo.

Para resolver esto, se reemplazan los valores del tamaño de pupila cercanos a una sacada (la cual es detectada por el marcador del eye tracker) por el valor que

precede a este rango. Y los valores siguientes son reemplazados con valores que conservan la diferencia con el valor anterior. La siguiente figura muestra un ejemplo de corrección de sacadas.

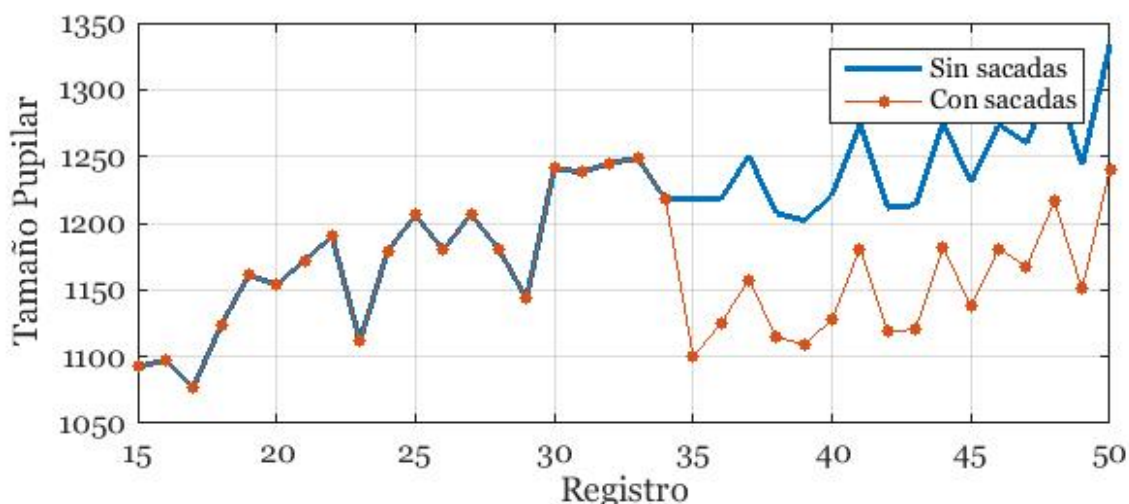


Figura 4.6: Señal con sacadas y señal sin sacadas (corregida)

Fuente: Elaboración propia

3. **Alta frecuencia:** el registro del eye tracker puede contener información de alta frecuencia correspondiente a ruido. Para corregir esto, se aplicó un filtro pasa bajo de 4Hz [71].

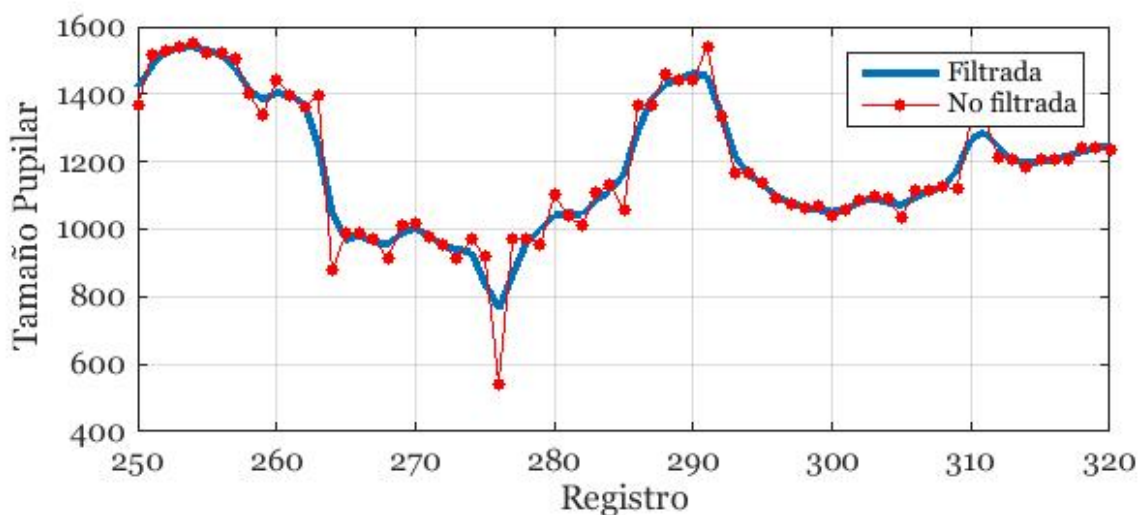


Figura 4.7: Corrección de Alta Frecuencia

Fuente: Elaboración propia



#### **4.2.2.2. Potencial Eléctrico del Cerebro**

Los principales ruidos detectados por el equipo del electroencefalograma son:

1. **Alta y baja frecuencia:** el equipo puede registrar datos que se escapan del rango de actividad de la actividad cerebral.
2. **Movimientos oculares y parpadeos:** los movimientos oculares producen actividad eléctrica que puede ser registrada en el electroencefalograma. El parpadeo también contamina el registro en el electroencefalograma con una mayor magnitud que los movimientos oculares, y principalmente esta actividad es registrada en los electrodos ubicados en la parte frontal, que es son los electrodos más próximos a los ojos.
3. **Electromiogramas:** corresponde a la actividad eléctrica producida por movimientos musculares, principalmente en la cabeza, como tragar saliva o contraer la mandíbula. Esta actividad eléctrica también puede generar ruido en el electroencefalograma y su magnitud depende del grado de la contracción muscular, por lo que este tipo de ruido puede estar presente en un amplio espectro de la medición en el electroencefalograma [85].
4. **Otras fuentes:** también pueden generar ruido otras fuentes menos comunes como toser, los movimientos bruscos de la cabeza, la respiración, la actividad cardiaca, entre otras.

Todas estas fuentes generan ruido al registro en el electroencefalograma y cada una de ellas se ve presente en diferentes rangos de frecuencia. Poder eliminarlas completamente es una tarea casi imposible [85], pero una buena aproximación es utilizar filtros que contemplen sólo el rango de la actividad del cerebro, que tiene relación a las bandas de frecuencia descritas en la sección 2.2.3. Entonces para pre-procesar la actividad bioeléctrica cerebral, se utilizó un filtro pasa alto y pasa bajo de 0,5 Hz y 63 Hz respectivamente [7], [10], [9], [86] (por Teorema de Nyquist, el filtro pasa bajo no puede superar los 64 Hz).

#### **4.2.3. Transformación**

Un vez que se tienen los datos limpios, se procede a transformarlos. Es en esta etapa en la que se obtienen las variables que caracterizan el comportamiento pupilar y la actividad bioeléctrica cerebral y son estas variables, las que serán ingresadas posteriormente a los modelos de minería de datos.

En primer lugar, se procedió a determinar la ventana de tiempo dentro de la navegación de la cual se extraería la información de la pupila y la actividad bioeléctrica cerebral para comparar entre hombres y mujeres. Esta ventana de tiempo tiene que estar relacionada a un evento o marcador en particular dentro de la navegación. El evento considerado corresponde una fijación, es decir, el momento inicial en el que el sujeto comienza a observar un objeto, dentro de aquellos que observó por al menos 300 ms.

Cabe mencionar que los objetos a considerar son los mismos que los seleccionados anteriormente.

Se han considerado las fijaciones porque las fijaciones en los objetos web corresponden a aquellas observaciones que capturan la atención del sujeto, por lo que estudiar la reacción de la pupila y la actividad bioeléctrica cerebral en esta ventana de tiempo representa adecuadamente el comportamiento fisiológico en el entorno web.

Entonces, la ventana de tiempo seleccionada comprende desde el inicio de una fijación hasta un tiempo a determinado posteriormente. Para determinar este tiempo se tomó en cuenta lo siguiente:

- Por recomendación del Ph.D. Pedro Maldonado, este debía ser entre 600 ms y 1000 ms.
- Existe una cierta cantidad de fijaciones que duren al menos un tiempo determinado, pero existirán menos fijaciones que duren al menos un tiempo mayor al anterior, por lo que hay un *trade-off*. Para este punto, se hizo una tabla que muestra el tiempo y la cantidad de fijaciones de al menos ese tiempo de duración. Esta se muestra a continuación.

Duración (ms)	Total de Fijaciones
300	15692
400	13001
500	10971
600	9358
700	7026
800	6155
900	5431
1000	4871
1100	4347
1200	3887
1300	3203
1400	2965
1500	2715

Tabla 4.8: Duración mínima y cantidad de fijaciones  
Fuente: Elaboración propia

Esta tabla muestra que la cantidad de fijaciones decrece rápidamente con el tiempo, pero a la vez la cantidad de fijaciones de al menos 1 segundo de duración sigue siendo considerable.

- La duración promedio de las fijaciones, según las tablas de la sección 4.1.2, va aproximadamente entre los 900 ms y los 1200 ms en los objetos seleccionados.

- La tasa de muestreo de los instrumentos es relativamente baja, por lo que si se quiere obtener una curva representativa, la duración debería ser suficientemente larga.

Por lo tanto, se ha elegido 1 segundo como la duración de la ventana de tiempo a tomar. Entonces, cada observación de la pupila y la actividad bioeléctrica cerebral, será una curva que tendrá una duración de 1 segundo y su inicio estará determinado por el inicio de una fijación que tenga al menos 1 segundo de duración en los objetos seleccionados. Las curvas tendrán en su eje X el tiempo y la curva de la pupila y la actividad bioeléctrica cerebral tendrán en su eje Y el tamaño de la pupila y el potencial eléctrico respectivamente.

#### **4.2.3.1. Dilatación Pupilar**

La transformación se debe realizar teniendo en consideración las características de los datos y el objetivo del análisis. En este caso, se quiere estudiar el comportamiento pupilar que está representado por la curva de 1 segundo de duración que se mencionó anteriormente. Esta curva a la vez puede estar representada o descrita por distintas características, las que podrán ser las variables de entrada a los modelos de minería de datos. Algunas de estas características han sido utilizadas en la literatura o recomendadas por el Ph.D. Pedro Maldonado, y otras han sido propuestas por el autor. Estas características son:

1. **Tamaño máximo de la pupila (PMax):** es el máximo valor del tamaño de la pupila que se encuentra en la curva.
2. **Tamaño mínimo de la pupila (PMin):** es el mínimo valor del tamaño de la pupila que se encuentra en la curva.
3. **Tamaño promedio de la pupila (PMean):** es el tamaño promedio en la curva.
4. **Desviación estándar (PSD):** es la desviación estándar de la curva.
5. **Máxima dilatación (DPMax):** corresponde al mayor incremento en el tamaño de la pupila entre dos instantes consecutivos de la curva.
6. **Máxima contracción (CPMax):** corresponde a la mayor caída en el tamaño de la pupila entre dos instantes consecutivos de la curva.

#### **4.2.3.2. Potencial Eléctrico del Cerebro**

Para la transformación del potencial eléctrico del cerebro, en primer lugar se removió la línea base al canal O1 y canal O2, restando el promedio de la curva a cada valor para crear observaciones comparables. Luego se extraen las características a las curvas de 1 segundo mencionadas anteriormente. Estas características describen estadísticamente la curva, como las utilizadas para la pupila, y otras provienen de la Transformada de Wavelet Discreta. Las descripciones estadísticas son:

1. **Máximo potencial alcanzado en la curva (EEGMax).**
2. **Mínimo potencial alcanzado en la curva (EEGMin).**
3. **Desviación estándar de la curva (EEGSD).**

Las siguientes provienen de la Transformada de Wavelet Discreta, la cual es una herramienta consolidada en el análisis de señales del electroencefalograma. Esta transformada no asume que la señal es estacionaria [58] y es una herramienta que es capaz de descomponer la curva original en el espacio temporal y en niveles de frecuencia (a diferencia de la Transformada de Fourier que sólo descompone en frecuencias) [53]. Cada uno de estos niveles estará representado por coeficientes. El nivel con la frecuencia más baja se dice que posee coeficientes de aproximación, mientras que los siguientes poseen coeficientes de detalle.

Para descomponer la señal con la Transformada de Wavelet Discreta, se pueden utilizar distintas funciones. En este caso se utilizó la función Wavelet Daubechies 4 que igualmente ha sido ampliamente utilizada [57]. Luego, para determinar la cantidad de niveles se tiene que tener en consideración la tasa de muestreo (128 Hz) y el Teorema de Nyquist, que dice que la frecuencia no debe superar la mitad de la tasa de muestreo [54]. Entonces se ha descompuesto la señal en 5 niveles de frecuencia, en donde cada uno representa una banda de frecuencia del electroencefalograma (esto se ha hecho utilizando WaveLab, un *toolbox* adicional y gratuito para Matlab que facilita la implementación de Transformadas de Wavelet [87]). En la siguiente tabla se muestra la descomposición de la señal utilizada en este trabajo.

Nivel de Descomposición	Frecuencia	Tipo de Onda
A4	0 Hz - 4 Hz	Delta
D4	4 Hz – 8 Hz	Theta
D3	8 Hz – 16 Hz	Alpha
D2	16 Hz – 32 Hz	Beta
D1	Mayor a 32 Hz	Gamma

Tabla 4.9: Niveles de Descomposición con Transformada de Wavelet Discreta  
Fuente: Elaboración propia

Luego, las siguientes características se obtienen de los coeficientes de cada nivel de frecuencia. Estas características son [54], [58]:

1. **Energía (EN):** corresponde a la energía de los coeficientes en el nivel  $i$ . En total son 5, un valor por cada nivel y se determina de la siguiente forma:

$$Energía_i = \sum_{k=1}^{n_i} (C_k^i)^2 \quad (4.1)$$

Con  $C_k^i$  representando el coeficiente  $k$  del nivel  $i$  y  $n_i$  el total de coeficientes del nivel  $i$ .

2. **Desviación estándar (STD):** corresponde a la desviación estándar de los coeficientes en cada nivel. En total son 5, un valor por cada nivel.

Entonces, por cada canal del electroencefalograma se tendrán las 3 características descriptivas, un valor de energía por cada nivel (5) y una desviación estándar por cada nivel (5). En total serán 13 características por canal. Dado que se considera el canal O1 y O2, se tendrán finalmente 26 características del potencial eléctrico del cerebro.

#### 4.2.4. Minería de Datos

Teniendo las variables que describen la curva y caracterizan el comportamiento fisiológico, se procede a la implementación de los modelos de minería de datos.

En esta memoria, se realizará un análisis por tipo de sitio, ya que se considera que cada tipo de sitio corresponde a un conjunto de estímulos que se diferencia de los otros tipos de sitios. Entonces si quiere comparar el comportamiento fisiológico entre dos grupos de sujetos, se tiene que comparar frente a un mismo tipo de estímulo.

Para validar o rechazar la hipótesis, se verán si existen diferencias significativas entre cada una de las 32 características definidas en la etapa anterior (6 del comportamiento pupilar, 26 del potencial eléctrico del cerebro). Luego, se utilizarán en el modelo de minería de datos sólo aquellas características que son estadísticamente diferentes, y por lo tanto aquellas características del comportamiento fisiológico que son un atributo diferenciador entre hombres y mujeres.

El modelo de minería de datos correspondería a un clasificador que pueda predecir la clase (hombre, mujer) de las características del comportamiento fisiológico que se le presentan. Como se mencionó en la sección 2.4.1, se utilizará el modelo *Support Vector Machine* con función Kernel RBF, ya que se adecúa a las características del problema y porque ha sido una herramienta bastante consolidada en el análisis de señales biológicas.

Entonces, si se encuentran diferencias significativas entre las características y además el modelo puede predecir con un buen desempeño la clase de las características fisiológicas que se le presentan, se validaría la hipótesis de investigación.

Por último, el tamaño muestral de cada característica correspondería a la misma cantidad de fijaciones de mínimo 1 segundo de duración en el tipo de sitio, ya que es una curva por fijación y por ende una característica por fijación. Las cantidades por tipo de sitio se muestran en la Tabla 4.10.

Tipo de Sitio	Fijaciones de los Hombres	Fijaciones de las Mujeres	Total
Neutro	585	841	1426
Masculino	794	1032	1826
Femenino	591	765	1356

Tabla 4.10: Cantidad de fijaciones a considerar por tipo de sitio  
Fuente: Elaboración propia

Se observa que las mujeres tienen más fijaciones debido a que navegan más tiempo en promedio. También se observa que el total de fijaciones de mínimo 1 segundo de duración en esta tabla es 4608, que es menor a la cantidad de fijaciones que muestra la Tabla 4.8 para 1 segundo de duración (4871). Esto es porque en el modelo de minería de datos se pretende utilizar el comportamiento pupilar junto con la actividad bioeléctrica cerebral. Entonces para esta etapa sólo se consideraron aquellas observaciones que poseen ambas fuentes de información, y como se ha mencionado, hubo dos personas de las cuales se obtuvo el registro de sólo una herramienta. Entonces, el sujeto del cual se obtuvieron las fijaciones pero no su actividad cerebral, no se consideró en esta etapa.

A continuación se presentan los resultados de los test para detectar diferencias estadísticas por tipo de sitio.

#### **4.2.4.1. Sitio Neutro**

Los test que se utilizaron fueron 3. Primero el test de Lilliefors, cuya hipótesis nula señala que la muestra sigue una distribución normal. Si se considera que las muestras siguen una distribución normal, se realiza el T-test para comparar las muestras. En caso contrario, se utiliza el test U de Mann-Whitney (equivalente al test de suma de rangos de Wilcoxon [88]). Y cada uno de estos test, se realizó con un 5% de significancia.

Los resultados se muestran en la siguiente tabla (todas las muestras rechazaron la hipótesis nula del test de Lilliefors, por lo que no se consideran normales).

Feature	Diferencia Estadística
STD A4 O1	No
STD D4 O1	No
STD D3 O1	No
<b>STD D2 O1</b>	<b>Si</b>
<b>STD D1 O1</b>	<b>Si</b>
EN A4 O1	No
EN D4 O1	No
EN D3 O1	No
<b>EN D2 O1</b>	<b>Si</b>
<b>EN D1 O1</b>	<b>Si</b>
<b>STD A4 O2</b>	<b>Si</b>
STD D4 O2	No
STD D3 O2	No
<b>STD D2 O2</b>	<b>Si</b>
<b>STD D1 O2</b>	<b>Si</b>
<b>EN A4 O2</b>	<b>Si</b>
EN D4 O2	No
EN D3 O2	No
<b>EN D2 O2</b>	<b>Si</b>
<b>EN D1 O2</b>	<b>Si</b>
<b>PMax</b>	<b>Si</b>
<b>PMin</b>	<b>Si</b>
<b>PMean</b>	<b>Si</b>
PSD	No
DPMax	No
<b>CPMax</b>	<b>Si</b>
<b>EEGMax O1</b>	<b>Si</b>
EEGMin O1	No
EEGSD O1	No
<b>EEGMax O2</b>	<b>Si</b>
EEGMin O2	No
<b>EEGSD O2</b>	<b>Si</b>

Tabla 4.11: Resultado de diferencias estadísticas, sitio neutro  
Fuente: Elaboración propia

Por lo tanto, las características a considerar son:

1. Tamaño máximo de la pupila (PMax)
2. Tamaño mínimo de la pupila (PMin)
3. Tamaño promedio de la pupila (PMean)
4. Máxima contracción (CPMax)
5. Máximo potencial en el canal O1 (EEGMax O1)
6. Máximo potencial en el canal O2 (EEGMax O2)
7. Desv. estándar del canal O2 (EEGSD O2)
8. Desv. estándar de los coeficientes de la banda Beta en el canal O1 (STD D2 O1)
9. Desv. estándar de los coeficientes de la banda Gamma en el canal O1 (STD D1 O1)
10. Energía de la banda Beta en el canal O1 (EN D2 O1)
11. Energía en la banda Gamma en el canal O1 (EN D1 O1)

12. Desv. estándar de los coeficientes de la banda Delta en el canal O2 (STD A4 O2)
13. Desv. estándar de los coeficientes de la banda Beta en el canal O2 (STD D2 O2)
14. Desv. estándar de los coeficientes de la banda Gamma en el canal O2 (STD D2 O2)
15. Energía en la banda Delta en el canal O2 (EN A4 O2)
16. Energía en la banda Beta en el canal O2 (EN D2 O2)
17. Energía en la banda Gamma en el canal O2 (EN D1 O2)

#### 4.2.4.2. Sitio Masculino

Los resultados se muestran en la siguiente tabla (todas las muestras rechazaron la hipótesis nula del test de Lilliefors, por lo que no se consideran normales).

Feature	Diferencia Estadística
STD A4 O1	No
STD D4 O1	No
STD D3 O1	No
<b>STD D2 O1</b>	<b>Si</b>
<b>STD D1 O1</b>	<b>Si</b>
EN A4 O1	No
EN D4 O1	No
EN D3 O1	No
<b>EN D2 O1</b>	<b>Si</b>
<b>EN D1 O1</b>	<b>Si</b>
<b>STD A4 O2</b>	<b>Si</b>
STD D4 O2	No
STD D3 O2	No
<b>STD D2 O2</b>	<b>Si</b>
<b>STD D1 O2</b>	<b>Si</b>
<b>EN A4 O2</b>	<b>Si</b>
EN D4 O2	No
EN D3 O2	No
<b>EN D2 O2</b>	<b>Si</b>
<b>EN D1 O2</b>	<b>Si</b>
PMax	No
PMin	No
PMean	No
<b>PSD</b>	<b>Si</b>
<b>DPMax</b>	<b>Si</b>
<b>CPMax</b>	<b>Si</b>
EEGMax O1	No
<b>EEGMin O1</b>	<b>Si</b>
EEGSD O1	No
EEGMax O2	No
EEGMin O2	No
EEGSD O2	No

Tabla 4.12: Resultado de diferencias estadísticas, sitio masculino  
Fuente: Elaboración propia



Por lo tanto, las características a considerar son:

1. Desviación estándar de la pupila (PSD)
2. Máxima dilatación (DPM<sub>ax</sub>)
3. Máxima contracción (CPM<sub>ax</sub>)
4. Mínimo potencial en el canal O1 (EEGMin O1)
5. Desv. estándar de los coeficientes de la banda Beta en el canal O1 (STD D2 O1)
6. Desv. estándar de los coeficientes de la banda Gamma en el canal O1 (STD D1 O1)
7. Energía de la banda Beta en el canal O1 (EN D2 O1)
8. Energía en la banda Gamma en el canal O1 (EN D1 O1)
9. Desv. estándar de los coeficientes de la banda Delta en el canal O2 (STD A4 O2)
10. Desv. estándar de los coeficientes de la banda Beta en el canal O2 (STD D2 O2)
11. Desv. estándar de los coeficientes de la banda Gamma en el canal O2 (STD D1 O2)
12. Energía en la banda Delta en el canal O2 (EN A4 O2)
13. Energía en la banda Beta en el canal O2 (EN D2 O2)
14. Energía en la banda Gamma en el canal O2 (EN D1 O2)

#### **4.2.4.3. Sitio Femenino**

Los resultados se muestran en la siguiente tabla (todas las muestras rechazaron la hipótesis nula del test de Lilliefors, por lo que no se consideran normales).

<b>Feature</b>	<b>Diferencia Estadística</b>
<b>STD A4 O1</b>	<b>Si</b>
STD D4 O1	No
STD D3 O1	No
<b>STD D2 O1</b>	<b>Si</b>
<b>STD D1 O1</b>	<b>Si</b>
<b>EN A4 O1</b>	<b>Si</b>
EN D4 O1	No
EN D3 O1	No
<b>EN D2 O1</b>	<b>Si</b>
<b>EN D1 O1</b>	<b>Si</b>
STD A4 O2	No
<b>STD D4 O2</b>	<b>Si</b>
STD D3 O2	No
<b>STD D2 O2</b>	<b>Si</b>
<b>STD D1 O2</b>	<b>Si</b>
EN A4 O2	No
<b>EN D4 O2</b>	<b>Si</b>
EN D3 O2	No
<b>EN D2 O2</b>	<b>Si</b>
<b>EN D1 O2</b>	<b>Si</b>
<b>PMax</b>	<b>Si</b>
<b>PMin</b>	<b>Si</b>
<b>PMean</b>	<b>Si</b>
<b>PSD</b>	<b>Si</b>
<b>DPMax</b>	<b>Si</b>
<b>CPMax</b>	<b>Si</b>
<b>EEGMax O1</b>	<b>Si</b>
<b>EEGMin O1</b>	<b>Si</b>
<b>EEGSD O1</b>	<b>Si</b>
<b>EEGMax O2</b>	<b>Si</b>
<b>EEGMin O2</b>	<b>Si</b>
<b>EEGSD O2</b>	<b>Si</b>

Tabla 4.13: Resultado de diferencias estadísticas, sitio femenino  
Fuente: Elaboración propia

Por lo tanto, las características a considerar son:

1. Tamaño máximo de la pupila (PMax)
2. Tamaño mínimo de la pupila (PMin)
3. Tamaño promedio de la pupila (PMean)
4. Desviación estándar de la pupila (PSD)
5. Máxima dilatación (DPMax)
6. Máxima contracción (CPMax)
7. Máximo potencial en el canal O1 (EEGMax O1)
8. Mínimo potencial en el canal O1 (EEGMin O1)
9. Desviación estándar en el canal O1 (EEGSD O1)
10. Máximo potencial en el canal O1 (EEGMax O2)
11. Mínimo potencial en el canal O1 (EEGMin O2)

12. Desv. estándar del canal O2 (EEGSD O2)
13. Desv. estándar de los coeficientes de la banda Delta en el canal O1 (STD A4 O1)
14. Desv. estándar de los coeficientes de la banda Beta en el canal O1 (STD D2 O1)
15. Desv. estándar de los coeficientes de la banda Gamma en el canal O1 (STD D1 O1)
16. Energía de la banda Delta en el canal O1 (EN A4 O1)
17. Energía de la banda Beta en el canal O1 (EN D2 O1)
18. Energía en la banda Gamma en el canal O1 (EN D1 O1)
19. Desv. estándar de los coeficientes de la banda Theta en el canal O2 (STD D4 O2)
20. Desv. estándar de los coeficientes de la banda Beta en el canal O2 (STD D2 O2)
21. Desv. estándar de los coeficientes de la banda Gamma en el canal O2 (STD D1 O2)
22. Energía en la banda Theta en el canal O2 (EN D4 O2)
23. Energía en la banda Beta en el canal O2 (EN D2 O2)
24. Energía en la banda Gamma en el canal O2 (EN D1 O2)

#### **4.2.4.4. Análisis de Resultados**

Primero se observa que existen diferencias significativas relacionadas a la pupila y al potencial eléctrico del cerebro en los 3 tipos de sitios. Se puede apreciar que en el sitio femenino, 24 de 32 características son estadísticamente distintas, en comparación a las 17 del sitio neutro y a las 14 del sitio masculino. También se observa que algunas características del sitio femenino están relacionadas a la banda Theta y en los otros tipos de sitios no. Pero a la vez se ven menos características relacionadas a la banda Delta que en los otros sitios.

Las características que se ven en los 3 tipos de sitios son:

1. Desviación estándar de la pupila (PSD)
2. Máxima contracción (CPMax)
3. Desv. estándar de los coeficientes de la banda Beta en el canal O1 (STD D2 O1)
4. Desv. estándar de los coeficientes de la banda Gamma en el canal O1 (STD D1 O1)
5. Energía de la banda Beta en el canal O1 (EN D2 O1)
6. Energía en la banda Gamma en el canal O1 (EN D1 O1)
7. Desv. estándar de los coeficientes de la banda Beta en el canal O2 (STD D2 O2)
8. Desv. estándar de los coeficientes de la banda Gamma en el canal O2 (STD D1 O2)
9. Energía en la banda Beta en el canal O2 (EN D2 O2)
10. Energía en la banda Gamma en el canal O2 (EN D1 O2)

De estas características, se puede destacar que la variación de la pupila respecto a su tamaño promedio es significativamente distinta entre hombres y mujeres en los 3 tipos de sitios. En relación al potencial eléctrico del cerebro, se observa que las descripciones estadísticas no logran ser un atributo diferenciador en común, pero sí los coeficientes de la Transformada de Wavelet Discreta, en particular los coeficientes relacionados a las bandas Beta y Gamma.

Por último, se puede observar que en el sitio neutro existe una cantidad considerable de características con diferencias significativas. Esto es en parte contradictorio ya que se esperaría que el comportamiento fisiológico fuese similar. Pero esto se puede deber a distintas razones:

- La clasificación de sitios en neutro, masculino y femenino, se realizó en base a las características que menciona la literatura acerca de las preferencias de los hombres y las mujeres. Puede ocurrir que lo que dice la literatura no se replique al grupo experimental utilizado en este trabajo, ya que la literatura utiliza grupos con diferentes residencias y culturas, lo que puede influir en las preferencias de las personas. Por otro lado, el grupo experimental utilizado en este trabajo está compuesto en su mayoría por personas relacionadas a una profesión en particular que es la Ingeniería Civil, lo que puede tener bastante relación con los intereses o el comportamiento web y en consecuencia se podrían tener brechas en caso de que el grupo experimental presente en la literatura no tenga relación con la Ingeniería Civil.
- Por otro lado, el autor determinó finalmente la clase de los tipos de sitios, lo que agrega subjetividad a la clasificación de estos. Por esta razón, se propone como trabajo futuro desarrollar un modelo que permita clasificar un sitio web y su público objetivamente.
- Otra razón, es que la realización de la tarea en el sitio web neutro puede involucrar otros procesos biológicos que se escapan del alcance de esta memoria.
- Finalmente, se pudo haber dado que temas particulares presentes en los sitios neutros, hayan provocado reacciones significativas en algunas personas y por ende haber alterado su comportamiento pupilar o cerebral.

#### **4.2.5. Evaluación e Interpretación**

En esta sección, se muestran los resultados de la clasificación y una interpretación final de los resultados.

Como se mencionó, esta clasificación se realiza por tipo de sitio, utilizando las características estadísticamente distintas entre hombres y mujeres en el tipo de sitio. Es decir, en el tipo de sitio neutro, masculino y femenino, se utilizan 17, 14 y 24 *features* (características) respectivamente.

Además, para cada tipo de sitio se utilizan las instancias correspondientes a ese tipo de sitio. Cada instancia está representada por los *features* (17 en caso de ser un sitio neutro, por ejemplo) y por el género (clase) de la persona a la cual esa instancia pertenece. Y en cuanto al universo de instancias, dado que cada una de ellas se asocia a una fijación con duración mayor o igual a 1 segundo, la cantidad es la que se muestra en la Tabla 4.10 (1426 en sitio neutro, 1826 en sitio masculino, 1356 en sitio femenino).

Luego, para los 3 tipos de sitios se realiza la predicción de género (Hombre o Mujer) utilizando los *features* correspondientes al tipo de sitio y los modelos descritos en la sección 2.4.1: SVM con Kernel RBF, SVM con Kernel Lineal, SVM con Kernel Polinomial de orden 2, Regresión Logística y Árbol de Decisión. Para cada modelo se realiza *5-fold-cross-validation*, entrenando en cada iteración con el 80% de las instancias correspondientes al tipo de sitio y probando la predicción de género con el

20% restante. Finalmente se obtienen los indicadores de desempeño como se describe en la sección 2.4.1.5. Los resultados se muestran a continuación.

Modelo	Recall	Precision	Accuracy
SVM Kernel RBF	99,64%	96,47%	97,62%
SVM Kernel Lineal	72,45%	66,16%	62,06%
SVM Kernel Polinomial	94,75%	93,76%	93,20%
Regresión Logística	73,17%	65,08%	60,39%
Árbol de Decisión	97,67%	99,43%	98,60%

Tabla 4.14: Resultados en sitio neutro  
Fuente: Elaboración propia

Modelo	Recall	Precision	Accuracy
SVM Kernel RBF	83,45%	85,67%	82,69%
SVM Kernel Lineal	71,13%	69,22%	66,04%
SVM Kernel Polinomial	85,02%	75,32%	75,96%
Regresión Logística	78,59%	68,23%	67,25%
Árbol de Decisión	99,53%	98,13%	98,63%

Tabla 4.15: Resultados en sitio masculino  
Fuente: Elaboración propia

Modelo	Recall	Precision	Accuracy
SVM Kernel RBF	80,92%	71,61%	71,24%
SVM Kernel Lineal	68,68%	68,07%	64,23%
SVM Kernel Polinomial	74,20%	72,52%	69,62%
Regresión Logística	78,50%	66,73%	65,04%
Árbol de Decisión	74,38%	74,86%	71,46%

Tabla 4.16: Resultados en sitio femenino  
Fuente: Elaboración propia

Como se menciona en la sección 2.4.1, *Recall* corresponde a la fracción de casos verdaderos positivos, dentro de los actuales positivos. *Precision* corresponde a la fracción de casos verdaderos positivos, dentro de los predichos positivos. Y *Accuracy* corresponde a la fracción de casos correctamente clasificados, dentro del total de los casos. Este último indicador es el que permite determinar el desempeño del modelo, independientemente de la clase positiva o negativa.

En cuanto al sitio neutro, se puede observar que con los modelos SVM con Kernel RBF y Polinomial y con el Árbol de Decisión se obtiene un excelente desempeño, superando el 93% de *Accuracy*. En cambio, con el modelo SVM con Kernel Lineal y con la Regresión Logística se obtiene un *Accuracy* cercano al 60%. Aquí se puede ver reflejada la diferencia de representar el problema de forma lineal.

En el sitio masculino, se observa que en general el desempeño es superior para los mismos 3 modelos que destacan en el sitio neutro, superando el 75% en *Accuracy*, pero

en este caso el Árbol de Decisión destaca notablemente sobre todos, alcanzando casi un 99% en *Accuracy*. Y nuevamente se quedan atrás los modelos SVM con Kernel Lineal y Regresión Logística, aunque cabe decir que su desempeño es de igual forma considerable (superando el 66%) y que en este tipo de sitio las diferencias entre modelos son menores (sin considerar el Árbol de Decisión).

Y en el sitio femenino, se observa que en general el desempeño es inferior a los otros tipos de sitios y que es relativamente parejo entre todos los modelos. El mayor *Accuracy* se obtiene con el Árbol de Decisión y el más bajo se obtiene con SVM con Kernel Lineal, siendo 71,46% y 64,23% respectivamente. Y de igual forma, en este tipo de sitio el *Accuracy* es superior para los modelos SVM con Kernel RBF, SVM con Kernel Polinomial y el Árbol de Decisión.

Si se compara el desempeño del mismo modelo entre los 3 tipos de sitios, se ve que los modelos SVM con Kernel Lineal y Regresión Logística mantienen un *Accuracy* similar. Por otro lado, se ve que existen diferencias considerables entre los *Accuracy* de los 3 sitios para los otros modelos. Esto se puede deber a distintos factores:

- En primer lugar, la cantidad de características (*features*) utilizadas en el modelo de minería de datos es distinta entre los 3 tipos de sitios. Esta diferencia en la cantidad de características puede generar diferencias en la información proporcionada en relación a aspectos biológicos, que no se puede determinar con los test de diferencia significativa. Por otro lado, una cantidad muy grande de características puede “confundir” al modelo y perjudicar su desempeño.
- En segundo lugar, puede haber sujetos de un sexo que tengan características más cercanas a su sexo opuesto. O también algún tema en particular en alguna de las preguntas hechas en el experimento pudo haber provocado reacciones significativas en el comportamiento fisiológico de una o más personas.

Si se observan los mejores modelos, se aprecia que en los 3 tipos de sitios el modelo con mejor desempeño es el Árbol de Decisión, seguido por el modelo SVM con Kernel RBF. Esto puede indicar que el Árbol de Decisión puede ser una mejor opción que un SVM para la clasificación de género o binaria en aplicaciones con datos fisiológicos.

Finalmente, se ve que los modelos tuvieron un buen desempeño en los 3 tipos de sitios, superando el 60% en *Accuracy*. En particular, el Árbol de Decisión supera el 70% en el sitio femenino y supera el 98% en *Accuracy* de los otros tipos de sitios. Con esto se puede afirmar que es posible diferenciar por sexo a usuarios de la Web, utilizando el comportamiento pupilar y la actividad bioeléctrica cerebral. Por lo tanto, se valida la hipótesis de investigación planteada en esta memoria.

# Capítulo 5: Conclusiones

En este capítulo, se comienza presentando las conclusiones generales que se han podido obtener en este Trabajo de Título y finalmente se presentan recomendaciones y propuestas para el trabajo futuro en el marco del proyecto AKORI.

## 5.1. Conclusiones Generales

En esta memoria se ha realizado una investigación ligada a temas fisiológicos en la que se utilizan herramientas de ingeniería para obtener resultados relevantes para el proyecto en cual se enmarca, el proyecto AKORI.

El objetivo general es examinar diferencias en el comportamiento fisiológico entre hombres y mujeres cuando navegan por la Web, mediante la aplicación de herramientas de eye tracking, electroencefalograma y técnicas de minería de datos. Y su hipótesis de investigación afirma que las diferencias existen y que por lo tanto es posible diferenciar por sexo a usuarios de la Web, utilizando el comportamiento pupilar y la actividad bioeléctrica cerebral.

Para poder cumplir con el objetivo propuesto, se tuvo que analizar el estado del arte de las materias relacionadas, diseñar e implementar un experimento para la obtención de datos y finalmente, analizar los datos. De estas etapas, se pueden mencionar importantes conclusiones.

En primer lugar, del estado del arte se concluye que existen diferencias consolidadas entre ambos sexos y que existe evidencia de que estas diferencias se proyectan en el comportamiento web de hombres y mujeres. Principalmente se puede decir que las mujeres navegan más exhaustivamente que los hombres y que existen diferencias en las preferencias del diseño de una página web. También se concluye que existe escases en la literatura de estudios que utilicen el comportamiento pupilar y la actividad cerebral como fuente de información para analizar el comportamiento web de los usuarios. Y tampoco se utilizan estas fuentes de información para analizar las diferencias entre sexos, y menos aún en el entorno web. Por lo tanto, este trabajo sería una novedad en términos de utilizar herramientas que no han sido aprovechadas en la diferenciación de grupos de personas y en el análisis del comportamiento web.

Cabe decir que el análisis del estado del arte, fue además fundamental para el diseño del experimento, ya que permitió comprender la envergadura de los estudios, los métodos utilizados y seleccionar los tipos de sitios web que representarían el entorno web en el experimento realizado en este trabajo.

En relación al análisis, se realizó en primer lugar un estudio exploratorio del comportamiento ocular de los hombres y las mujeres. De este estudio se concluye principalmente que las mujeres navegan más exhaustivamente que los hombres, que es lo que afirma la literatura analizada en este trabajo. Luego se realiza el proceso KDD utilizando el comportamiento pupilar y la actividad bioeléctrica cerebral. En este proceso fue fundamental el conocimiento compartido por el Ph.D. Pedro Maldonado, que permitió enriquecer el análisis de señales biológicas.

Un tema presente en el análisis que no deja de llamar la atención, es que en los sitios catalogados como neutros, se hayan detectado diferencias significativas en una cantidad considerable de características. Razones de esto pueden ser que lo catalogado como neutro en la literatura, no aplique al grupo experimental de este trabajo. Otra razón es que la clasificación de los sitios web fue en parte subjetiva. Por otro lado, la búsqueda de información puede involucrar procesos biológicos que se escapan del conocimiento de esta memoria y que alteran el comportamiento pupilar y la actividad bioeléctrica cerebral.

En la minería de datos se utilizaron 5 modelos de clasificación distintos: SVM con Kernel RBF, SVM con Kernel Lineal, SVM con Kernel Polinomial de orden 2, Regresión Logística y Árbol de Decisión. Los resultados de estos muestran que los modelos SVM con Kernel RBF y Polinomial y el Árbol de Decisión fueron superiores, lo que se puede deber a que los datos fisiológicos estudiados en esta memoria no se comportan de forma lineal. También se destaca el Árbol de Decisión como el modelo con mejor desempeño en los 3 tipos de sitios, superando el 70% en el sitio femenino y superando el 98% en *Accuracy* de los otros tipos de sitios. Con estos resultados se afirma que es posible diferenciar por sexo a usuarios de la Web, utilizando el comportamiento pupilar y la actividad bioeléctrica cerebral, por lo que se valida la hipótesis de investigación de esta memoria y se cumpliría el objetivo general.

La principal implicancia de estos resultados es que la plataforma informática pretendida por el proyecto AKORI podría incluir en su servicio la segmentación por género. Hasta el momento, gracias a las investigaciones que se han hecho previamente a esta memoria, el servicio de AKORI podría entregar información relevante a sus clientes - como empresas en las cuales el canal online sea fundamental [13] – expresada a través de mapas de calor de los sitios web. Por ejemplo, se pretende ofrecer un mapa de calor que puede mostrar el nivel de atención que capturan los objetos web, a través de la fijación ocular y herramientas de machine learning. Otro ejemplo es ofrecer un mapa que represente la probabilidad de click en una zona, utilizando la dilatación pupilar [8], [89]. Por otro lado, también se pueden ofrecer mapas basados en la información del electroencefalograma, como por ejemplo la emocionalidad que puede provocar el sitio web, pero más se investigación se requiere para esto.

Entonces, esta información expresada en mapas, se podría ofrecer diferenciando el género al cual el cliente se quiera enfocar. Con esto se agregaría valor al cliente, que como además se vio, le gustaría obtener un servicio que pueda segmentar a sus usuarios. Esta podría ser una característica diferenciadora de otros servicios similares que existen en la actualidad, como EyeQuant [90].



También, estos resultados podrían ser una motivación para expandir el proyecto AKORI fuera del entorno web, ya que el género es una de las variables de segmentación de clientes más usadas en todo tipo de organizaciones.

Finalmente, se concluye que el proceso KDD y la minería de datos puede ser una herramienta poderosa para el análisis de señales biológicas y que su utilización puede ser una ventaja competitiva en los grupos de investigación y en las empresas.

## **5.2. Recomendaciones y Trabajo Futuro**

Para los siguientes trabajos enmarcados en el proyecto AKORI, se recomienda y propone lo siguiente:

- Incluir en el análisis los otros canales del electroencefalograma que no han sido utilizados en esta memoria. Como se mencionó, sólo se utilizó el canal O1 y el O2, ya que estaban relacionados a los estímulos visuales y eran adecuados como primer acercamiento a un estudio como este. Pero sigue siendo interesante, estudiar la información que los otros canales pueden entregar.
- La identificación y selección de objetos ha sido un proceso que ha requerido mucho tiempo en las memorias anteriores y en esta memoria. Actualmente se tiene un sistema que puede identificar objetos de los sitios web, pero entrega resultados que requieren ser procesados manualmente para su utilización. Es por esto que se propone desarrollar un sistema que pueda identificar y seleccionar los objetos web de forma automática.
- Como se ha mencionado, la clasificación final de sitios web en neutro, masculino y femenino se hizo bajo criterio del autor. Esto agrega subjetividad el análisis. Por esto, se propone establecer un sistema que pueda clasificar de forma objetiva sitios web según su tipo de usuario objetivo en base al diseño o contenido. Esto debiese ser importante si se quieren entregar recomendaciones con la plataforma informática, ya que cada organización se enfoca en grupos de clientes pero el foco de su canal online puede no estar cumpliendo con las características de diseño adecuadas para esos grupos de clientes.
- También se propone estudiar el comportamiento pupilar y la actividad bioeléctrica cerebral de usuarios de la web segmentados por otras variables, como la edad o características socioeconómicas.
- Por último, en la actualidad existen más herramientas que miden el comportamiento fisiológico de las personas, como el electrocardiograma, medidores de la conductividad de la piel o el reconocimiento facial, las cuales pueden ser integradas en el proyecto AKORI.

# Bibliografía

- [1] International Telecommunication Union, “ICT Facts & Figures: The World in 2015.” 2015.
- [2] L. E. Dujovne and J. D. Velásquez, *Knowledge-Based and Intelligent Information and Engineering Systems*, vol. 5711. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.
- [3] P. Loyola, P. E. Román, and J. D. Velásquez, “Predicting web user behavior using learning-based ant colony optimization,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 5, pp. 889–897, Aug. 2012.
- [4] G. Martínez, “Mejoramiento de una metodología para la identificación de website keyobjects mediante la aplicación de tecnología eye tracking, análisis de dilatación pupilar y algoritmos de web mining,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2013.
- [5] Juan D. Velásquez and L. González, “Una aplicación de herramientas de eye tracking para analizar las preferencias de contenido de los usuarios de sitios web,” *Rev. Ing. Sist.*, vol. 26, no. 1, pp. 95–118, 2012.
- [6] J. Dupré, “Generación de una metodología de detección de website keyobjects basado en métricas de centralidad de teoría de grafos,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [7] G. Slanzi, “Estudio del impacto del uso del electroencefalograma en la identificación de website keyobjects,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [8] J. Jadue, “Incidencia de la dilatación pupilar como variable predictiva del comportamiento de los usuarios en una página web antes de tomar una decisión,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [9] L. Salas, “Análisis de la actividad cerebral, medida a partir de un electroencefalograma, como variable predictiva de la toma de decisión de un sujeto en una página web,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [10] C. Aracena, “Estudio de la relación entre neurodatos, dilatación pupilar y emocionalidad basado en técnicas de minería de datos,” Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [11] J. Prodanova and S. S. Martín, “Estudio sobre el impacto del género y las emociones en el comportamiento de compra de viajes online,” *Rev. Análisis Turístico*, vol. 15, pp. 1–12, 2013.

- [12] E. H. Hess, A. L. Seltzer, and J. M. Shlien, "Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study.," *J. Abnorm. Psychol.*, vol. 70, pp. 165–168, 1965.
- [13] R. Ruiz, "Diseño de un modelo de negocio para un servicio de análisis e indicadores sobre un contenido visual, asociado a proyecto Akori," Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, 2014.
- [14] F. J. Alañón Fernández, S. Ferreiro López, and J. Fernández Pérez, *Oftalmología en atención primaria*. Formacion Alcalá S.L., 2011.
- [15] "Anatomía." [Online]. Disponible: <http://www.sao.org.ar/index.php/informacion-para-pacientes/anatomia>. [Visto: 02-Oct-2015].
- [16] B. Cassin, S. Solomon, and M. L. Rubin, *Dictionary of Eye Terminology*. Triad Pub. Co., 1984.
- [17] G. Buscher, E. Cutrell, and M. R. Morris, "What do you see when you're surfing? Using eye tracking to predict salient regions of web pages," in *Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 21–30.
- [18] B. Pan, H. a Hembrooke, G. K. Gay, L. a Granka, M. K. Feusner, and J. K. Newman, "The determinants of web page viewing behavior: an eye-tracking study," in *Proceedings of the ETRA '04 Symposium on Eye Tracking Research and Applications*, 2004, vol. 1, no. 212, pp. 147–154.
- [19] J. Beatty and B. Lucero-Wagoner, "The pupillary system," in *Handbook of Psychophysiology*, 2nd ed., 2000, pp. 142–162.
- [20] M. M. Bradley, L. Miccoli, M. a Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.
- [21] E. H. Hess and J. M. Polt, "Pupil Size in Relation to Mental Activity during Simple Problem-Solving," *Science (80-. )*, vol. 143, no. 3611, pp. 1190–1192, Mar. 1964.
- [22] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources.," *Psychol. Bull.*, vol. 91, no. 2, pp. 276–292, 1982.
- [23] D. Hahnemann and J. Beatty, "Pupillary responses in a pitch-discrimination task," *Percept. Psychophys.*, vol. 2, no. 3, pp. 101–105, 1967.
- [24] L. Granka, M. Feusner, and L. Lorigo, "Eye Monitoring in Online Search," *Passiv. Eye Monit.*, pp. 347–372, 2008.
- [25] E. H. Hess and J. M. Polt, "Pupil Size as Related to Interest Value of Visual Stimuli," *Science (80-. )*, vol. 132, no. 3423, pp. 349–350, 1960.
- [26] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, *Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals*, vol. 4105. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.
- [27] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Hum. Comput. Stud.*, vol. 59, no. 1–2, pp. 185–198, 2003.
- [28] J. a. Schirillo, "Pupil Dilations Reflect Why Rembrandt Biased Female Portraits

- Leftward and Males Rightward,” *Front. Hum. Neurosci.*, vol. 7, no. January, pp. 1–11, 2014.
- [29] A. T. Duchowski, *Eye Tracking Methodology*. Springer London, 2007.
- [30] J. Kelly, E. Kandel, and J. Schwartz, *Principles of Neural Science*, 4th ed. New York: McGraw-Hill Health Professions Division, 2000.
- [31] E. Niedermeyer and F. H. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed. Lippincott Williams & Wilkins, 2005.
- [32] Y. M. Baraka, H. Bekemeier, A. J. Giessler, and R. Rashid, *Handbook of Psychophysiology*, 3rd ed. Cambridge University Press, 2007.
- [33] P. Barret and P. Sowden, “Psychophysiological Methods,” in *Handbook of Human Factors and Ergonomics Methods*, vol. 8, CRC Press, 2004, pp. 146–159.
- [34] Z. Iscan, Z. Dokur, and T. Demiralp, “Classification of electroencephalogram signals with combined time and frequency features,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10499–10505, 2011.
- [35] M. Hines, “Sex-related variation in human behavior and the brain,” *Trends Cogn. Sci.*, vol. 14, no. 10, pp. 448–456, 2010.
- [36] S. Álvarez, “Neuromarketing: Qué sabe de ti, que los demás no,” Trabajo de Grado, Universidad del Rosario, 2014.
- [37] G. Gong, Y. He, and a. C. Evans, “Brain Connectivity: Gender Makes a Difference,” *Neurosci.*, vol. 17, no. 5, pp. 575–591, 2011.
- [38] K. Goodrich, “The gender gap: Brain-processing differences between the sexes shape attitudes about online advertising,” *J. Advert. Res.*, vol. 54, no. 1, pp. 32–43, 2014.
- [39] T. D. Satterthwaite, D. H. Wolf, D. R. Roalf, K. Ruparel, G. Erus, S. Vandekar, E. D. Gennatas, M. a Elliott, A. Smith, H. Hakonarson, R. Verma, C. Davatzikos, R. E. Gur, and R. C. Gur, “Linked Sex Differences in Cognition and Functional Connectivity in Youth,” *Cereb. Cortex*, pp. 1–12, 2014.
- [40] S. Djamasbi and T. Tullis, “Gender preferences in web design: usability testing through eye tracking,” in *Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS)*, 2007, pp. 1–8.
- [41] S. J. Simon, “The Impact of Culture and Gender on Web Sites: An Empirical Study,” *DATA BASE Adv. Inf. Syst.*, vol. 32, no. 1, pp. 18–37, 2001.
- [42] A. N. Tuch, J. a. Bargas-Avila, and K. Opwis, “Symmetry and aesthetics in website design: It’s a man’s business,” *Comput. Human Behav.*, vol. 26, no. 6, pp. 1831–1837, 2010.
- [43] A. Chadwick-Dias, D. Tedesco, and T. Tullis, “Demographic differences in preferred web site content,” in *Proceedings of Usability Professionals’ Association Annual Conference*, 2004, pp. 1–8.
- [44] G. Moss, R. Gunn, and J. Heller, “Some men like it black, some women like it pink: consumer implications of differences in male and female website design,” *J.*

*Consum. Behav.*, pp. 328–341, 2006.

- [45] G. Moss and R. Gunn, “Gender differences in website design: Implications for education,” *J. Syst. Informatics Cybern.*, vol. 5, no. 6, pp. 38–43, 2007.
- [46] D. Cyr and C. Bonanni, “Gender and website design in e-business,” *Int. J. Electron. Bus.*, vol. 3, no. 6, pp. 565–582, 2005.
- [47] M.-O. Richard, J.-C. Chebat, Z. Yang, and S. Putrevu, “A proposed model of online consumer behavior: Assessing the role of gender,” *J. Bus. Res.*, vol. 63, no. 9–10, pp. 926–934, 2010.
- [48] Q. Li, L. Sun, and J. Duan, “Web page viewing behavior of users: An eye-tracking study,” in *2005 International Conference on Services Systems and Services Management, Proceedings of ICSSSM’05*, 2005, vol. 1, pp. 244–249.
- [49] E. Stenstrom, P. Stenstrom, G. Saad, and S. Cheikhrouhou, “Online hunting and gathering: An evolutionary perspective on sex differences in website preferences and navigation,” *IEEE Trans. Prof. Commun.*, vol. 51, no. 2, pp. 155–168, 2008.
- [50] S. Georgiev, Z. Minchev, C. Christova, and D. Philipova, “Gender event-related brain oscillatory differences in normal elderly population EEG,” *Int. J. Bioautomation*, vol. 15, no. 1, pp. 33–48, 2011.
- [51] M. Matsuura, K. Yamamoto, H. Fukuzawa, Y. Okubo, H. Uesugi, M. Moriiwa, T. Kojima, and Y. Shimazono, “Age development and sex differences of various EEG elements in healthy children and adults — Quantification by a computerized wave form recognition method,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 60, no. 5, pp. 394–406, May 1985.
- [52] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *A I Mag.*, vol. 17, pp. 37–54, 1996.
- [53] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, 2nd ed. Springer, 2005.
- [54] C. E. Valderrama and G. Ulloa, “Combining spectral and fractal features for emotion recognition on Electroencephalographic signals,” *WSEAS Trans. Signal Process.*, vol. 10, pp. 481–496, 2014.
- [55] O. Sourina and Y. Liu, “A Fractal-Based Algorithm of Emotion Recognition from EEG using Arousal-Valence Model,” *Biosignals*, pp. 209–214, 2011.
- [56] M. Murugappan, R. Nagarajan, and S. Yaacob, “Comparison of different wavelet features from EEG signals for classifying human emotions,” in *2009 IEEE Symposium on Industrial Electronics and Applications, ISIEA 2009 - Proceedings*, 2009, vol. 2, pp. 836–841.
- [57] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, D. Hazry, and I. Zunaidi, “Time-Frequency Analysis of EEG Signals for Human Emotion Detection,” in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008 IFMBE*, 2008, pp. 262–265.
- [58] P. Zarjam, J. Epps, and N. H. Lovell, “Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload,” *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 4, pp. 301–310, Dec. 2015.

- [59] N. M. El-kafrawy, D. Hegazy, and M. F. Tolba, "Features Extraction and Classification of EEG Signals Using Empirical Mode Decomposition and Support Vector Machine," *Adv. Mach. Learn. Technol. Appl.*, no. 3, pp. 189–198, 2014.
- [60] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification.," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–4, 2003.
- [61] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3498–3510, 2012.
- [62] R. N. Khushaba, S. Kodagoda, G. Dissanayake, L. Greenacre, S. Burke, and J. Louviere, "A neuroscientific approach to choice modeling: Electroencephalogram (EEG) and user preferences," in *Proceedings of the International Joint Conference on Neural Networks*, 2012.
- [63] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [64] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," *Expert Syst. Appl.*, vol. 36, pp. 1329–1336, 2009.
- [65] E. D. Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," *Digit. Signal Process. A Rev. J.*, vol. 19, no. 2, pp. 297–308, 2009.
- [66] I. Guler and E. D. Ubeyli, "Multiclass Support Vector Machines for EEG-Signals Classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–126, 2007.
- [67] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L. O. Wahlund, Y. Dodge, and T. Dierks, "Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)," *J. Neurosci. Methods*, vol. 161, no. 2, pp. 342–350, 2007.
- [68] N. V Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods.," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 453–495, 2004.
- [69] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces.," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, 2007.
- [70] R. Bednarik, S. Eivazi, and H. Vrzakova, "A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data," *Eye Gaze Intell. User Interfaces*, pp. 111–134, 2013.
- [71] S. Chen and J. Epps, "Automatic classification of eye activity for cognitive load measurement with emotion interference," *Comput. Methods Programs Biomed.*, vol. 110, no. 2, pp. 111–124, 2013.
- [72] M. Qian, M. Aguilar, K. N. Zachery, C. Privitera, S. Klein, T. Carney, and L. W. Nolte, "Decision-level fusion of EEG and pupil features for single-trial visual

- detection analysis,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1929–1937, 2009.
- [73] B. Shi, K. P. Moloney, Y. Pan, V. K. Emery, B. Vidakovic, J. a Jacko, and F. Sainfort, “Classification of High Frequency Pupillary Responses using Schur Monotone Descriptors in Multiscale Domains.” Georgia Institute of Technology, Atlanta, 2004.
- [74] B. Liu, *Web Data Mining*. Springer, 2007.
- [75] S. Maldonado and R. Weber, “Modelos de Selección de Atributos para Support Vector Machines,” *Rev. Ing. Sist.*, pp. 49–70, 2012.
- [76] R. Christensen, *Log-Linear Models and Logistic Regression*. Springer-Verlag, 1997.
- [77] “Applied Human-Machine and Media Interaction Modeling: Optimize Your Web Site Through a Proper Eye-Tracking Study.” [Online]. Disponible: [http://www.ece.uvic.ca/~madhackr/art\\_whybn.htm](http://www.ece.uvic.ca/~madhackr/art_whybn.htm). [Visto: 26-Feb-2015].
- [78] “Emotiv.” [Online]. Disponible: <https://emotiv.com/>. [Visto: 18-Oct-2015].
- [79] “Emotiv EPOC User Manual.” 2003.
- [80] “Sofey.” [Online]. Disponible: <http://sofey.net/>. [Visto: 20-Oct-2015].
- [81] “Miembros del laboratorio | Laboratorio de Neurosistemas.” [Online]. Disponible: <http://neurosistemas.cl/miembros-del-laboratorio/>. [Visto: 21-Oct-2015].
- [82] H. W. Lilliefors, “On the Kolmogorov-Smirnov Test for Normality With Mean and Variance Unknown,” *J. Am. Stat. Assoc.*, vol. 62, no. 318, pp. 399–402, 1967.
- [83] “Lilliefors test - MATLAB.” [Online]. Disponible: <http://www.mathworks.com/help/stats/lillietest.html?refresh=true>. [Visto: 22-Oct-2015].
- [84] “Two-sample t-test - MATLAB.” [Online]. Disponible: <http://www.mathworks.com/help/stats/ttest2.html>. [Visto: 22-Oct-2015].
- [85] J. A. Urigüen and B. Garcia-Zapirain, “EEG artifact removal—state-of-the-art and guidelines,” *J. Neural Eng.*, vol. 12, no. 3, p. 031001, Jun. 2015.
- [86] W. Zheng, B. Dong, and B. L. S. Member, “Multimodal Emotion Recognition using EEG and Eye Tracking Data,” in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 5040–5043.
- [87] “WaveLab.” [Online]. Disponible: [http://statweb.stanford.edu/~wavelab/Wavelab\\_850/index\\_wavelab850.html](http://statweb.stanford.edu/~wavelab/Wavelab_850/index_wavelab850.html). [Visto: 18-Dec-2015].
- [88] “Wilcoxon test - MATLAB.” [Online]. Disponible: <http://www.mathworks.com/help/stats/ranksum.html>. [Visto: 22-Oct-2015].
- [89] J. Jadue, G. Slanzi, L. Salas, and J. D. Velásquez, “Web user click intention prediction by using pupil dilation analysis,” in *2015 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent*

*Technology*, 2015.

- [90] "EyeQuant." [Online]. Disponible: <http://www.eyequant.com/>. [Visto: 26-Oct-2015].



# Anexos

## A Preguntas del Experimento

Tipo de Sitio Web	Sitio Web	Pregunta
Masculino	gq.com.mx	¿De qué año es la película "Aliens"?
		¿Cuál es el rango de precio de BitBite?
		¿Dónde hace su debut el Chevrolet FNR?
	wired.com	¿Qué puntaje asigna el sitio al teléfono Samsung Galaxy Note 4?
	primermagazine.com	¿Cuáles son los consejos para hacer durar más la batería del celular?
		¿Qué tipo de cerveza combina con el chocolate?
	businessinsider.com	¿Cuál es el precio de Vesper?
Neutro	werd.com	¿Cuál es el valor del Mercedes Benz G Class?
		¿Qué hora aparece en la pantalla del Motorola Droid Turbo?
	uncrate.com	¿En qué categoría aparecen jugadores de basketball?
	sciencemag.org	¿A cuál categoría pertenece el artículo "Baboons follow the pack, not the leader"?
		¿Quién es el presidente de la Universidad de Columbia?
columbia.edu	¿Cuántas mujeres hay en la Administración de la Universidad de Columbia?	
	¿Quién ganó el premio Nobel de la paz en el año 2007?	
wharton.upenn.edu	¿Cuál es la edad promedio de los estudiantes que ingresaron al doctorado de Wharton en el 2015?	
	emol.com	¿Cuántos euros cuesta la botella del vino galardonado como el mejor del mundo?
Femenino	wikipedia.org	¿Cuál es la principal uva utilizada en la elaboración del pisco chileno?
		¿En qué año se creó el Parque Nacional Torres del Paine?
	elizabethstreet.com	¿Qué versión del Festival Internacional de Cine de Berlín se celebró en el 2015?
		¿Para cuáles revistas escribe Marnie Hanel?
		¿Cuáles son los artículos populares de la semana en el siguiente sitio?
	designsponge.com	¿Qué galería recomienda el autor a las 11:30 am?
		¿De quién es la cuenta de Instagram, "Mattbg"?
enfemenino.com	¿Quién capturó la tercera fotografía del siguiente artículo?	
	¿Qué alimentos ayudan contra los calambres?	
divinecaroline.com	¿Cuántas personas celiacas se calcula que hay en España?	
	¿Cuántas onzas (oz) de bourbon se necesitan para hacer un "Manhattan"?	

Tabla 1: Preguntas del Experimento

Fuente: Elaboración propia

## B Consentimiento Informado

Comité de ética Departamento de Ingeniería Industrial Facultad de Ciencias Físicas y Matemáticas Universidad de Chile
<b>CONSENTIMIENTO INFORMADO</b>
<b>TITULO</b>
Nombre del investigador principal: Dr. Juan Velásquez S. Institución: Web Intelligence Centre, DII, FCFM, Universidad de Chile. Teléfonos: +56 2 29784834
Se le entregará una copia consentimiento informado completo.
<b>Introducción</b>
Mi nombre es Gino Slanzi. Mi profesión es Ingeniero Civil Industrial y actualmente llevo a cabo el proyecto de investigación al cual usted ha sido invitado, en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. A lo largo de la lectura de este documento usted es libre de manifestar cualquiera de sus inquietudes respecto al procedimiento que se llevará a cabo, tanto hacia mí como con alguien con quien usted se sienta cómodo. Asimismo, puede tomarse el tiempo que requiera para reflexionar respecto a si desea participar del proyecto. Si no entiende alguna de las informaciones contenidas en el presente documento siéntase libre de expresármelo en cualquier momento para explicarle la información en detalle.
<b>Invitación a participar.</b> Le estamos invitando a participar en el proyecto de investigación "AKORI: Plataforma informática basada en Web Intelligence y herramientas de Brain Informatics para la mejora de estructura y contenidos de sitios Web".
Objetivos: Esta investigación tiene por objetivo estudiar las respuestas fisiológicas de los usuarios al momento de navegar por sitios Web. El total de sujetos propuesto para realizar este estudio es de 40 personas.
<b>Procedimientos:</b> Si usted acepta participar de la investigación propuesta usted estará aceptando ser sometido, por una sola vez, al siguiente procedimiento: la medición de la variación del diámetro de su pupila mediante un sistema de cámaras que van adosados a su cabeza, junto con la medición de la actividad que genera el cerebro a través de electrodos que se ubicarán sobre la superficie de su cuero cabelludo (electroencefalograma). Ambos registros son superficiales, esto es, no invasivos, y no producen daño ni efectos adversos. Durante todo el experimento sólo se medirán parámetros provenientes de usted. En ningún momento se le administrará ningún tipo de energía así como ningún tipo de fármaco. <b>El experimento presenta dos tareas por separado. La primera considera una instrucción de búsqueda al usuario, dada una página de un sitio Web determinado. Para la segunda, se selecciona un sitio Web y se le pide al usuario que responda una pregunta de contenido. El usuario debe navegar libremente hasta encontrar la respuesta.</b> Usted es libre de retirarse de la tarea en cualquier momento a lo largo de esta, aunque no haya llegado a su fin. El investigador se encontrará en la misma sala que usted a lo largo de toda la tarea, y usted puede solicitar de él en todo momento cualquier información o expresar cualquier necesidad que estime pertinente.
<b>Riesgos:</b> Bajo los sistemas de registro que utilizaremos no existen riesgos ni efectos

Figura 1: Consentimiento Informado, Parte 1  
Fuente: Elaboración propia

adversos conocidos. Ambos registros mencionados son ampliamente utilizados en el mundo entero para fines de investigación, como en el caso de la tarea a la cual usted será sometido.

**Costos:** Las técnicas utilizadas en este proyecto no tienen costo alguno para Ud.

**Beneficios:** Los beneficios del presente estudio no irán en beneficio directamente de usted. El presente proyecto tiene por objetivo contribuir al conocimiento científico de cómo las emociones son procesadas por el cerebro humano.

**Alternativas:** La decisión de no participar del presente estudio no significará ningún perjuicio para su persona.

**Compensación:** No se considera la entrega de una compensación económica para usted en el presente estudio.

**Confidencialidad:** Toda la información derivada de su participación en este estudio será conservada en forma de estricta confidencialidad, lo que incluye el acceso de los investigadores o agencias supervisoras de la investigación. Cualquier publicación o comunicación científica de los resultados de la investigación será completamente anónima.

**Voluntariedad:** Su participación en esta investigación es totalmente voluntaria y se puede retirar en cualquier momento comunicándolo al investigador.

**Complicaciones:** Aunque no existe riesgo asociado descrito para el procedimiento a realizar, en el caso de que usted presente complicaciones directamente dependientes de la aplicación de las técnicas utilizadas en este estudio, usted recibirá el tratamiento médico completo de dicha complicación, financiado por el proyecto al cual se asocia este estudio, y sin costo alguno para Ud. o su sistema previsional.

**Derechos del participante:** Si Ud. requiere cualquier otra información sobre su participación en este estudio puede llamar a:

Investigador: Gino Sianzi Rodríguez; teléfono +569 75485568  
Autoridad de la Institución: Juan Velásquez Silva; teléfono: +56 2 29784834

**Conclusión:** Después de haber recibido íntegramente y comprendido la totalidad de la información contenida en este documento, no teniendo actualmente ninguna duda respecto a la tarea a realizar, las técnicas de medición, así como los riesgos asociados, otorgo mi consentimiento para participar en el proyecto "AKORI: Plataforma informática basada en Web Intelligence y herramientas de Brain Informatics para la mejora de estructura y contenidos de sitios Web".

Figura 2: Consentimiento Informado, Parte 2  
Fuente: Elaboración propia

Comité de ética  
Departamento de Ingeniería Industrial  
Facultad de Ciencias Físicas y Matemáticas  
Universidad de Chile

_____ Nombre del sujeto	_____ Firma	_____ Fecha
_____ Nombre del informante	_____ Firma	_____ Fecha
_____ Nombre del investigador	_____ Firma	_____ Fecha

Figura 3: Consentimiento Informado, Parte 3  
Fuente: Elaboración propia

# C Microsoft Excel con Preguntas

id	pregunta	URL
1	¿Qué alimentos ayudan contra los calambres?	file:///C:/Users/Gino/Desktop/websites/enfem/Diez%20alimentos%20que%20puedes%20comer%20antes%20de%20hacer%20ejercicio.html
2	¿Cuántas mujeres hay en la Administración de la Universidad de Columbia?	file:///C:/Users/Gino/Desktop/websites/columbia_adm/www.columbia.edu/node/52.html
3	¿Dónde hace su debut el Chevrolet FNR?	file:///C:/Users/Gino/Desktop/websites/ga_autos/www.gq.com.mx/maquinas/autos.html
4	¿De qué año es la película "Aliens"?	file:///C:/Users/Gino/Desktop/websites/ga_secuelas/www.gq.com.mx/actualidad/cine-tv/articulo-aliens-1986.html
5	¿Qué hora aparece en la pantalla del Motorola Droid Turbo?	file:///C:/Users/Gino/Desktop/websites/uncrate_phones/uncrate.com/tech/cell-phones/index.html

Figura 4: Sujeto en quinta pregunta del experimento  
Fuente: Elaboración propia

id	pregunta	URL
1	¿Qué alimentos ayudan contra los calambres?	file:///C:/Users/Gino/Desktop/websites/enfem/Diez%20alimentos%20que%20puedes%20comer%20antes%20de%20hacer%20ejercicio.html
2	¿Cuántas mujeres hay en la Administración de la Universidad de Columbia?	file:///C:/Users/Gino/Desktop/websites/columbia_adm/www.columbia.edu/node/52.html
3	¿Dónde hace su debut el Chevrolet FNR?	file:///C:/Users/Gino/Desktop/websites/ga_autos/www.gq.com.mx/maquinas/autos.html
4	¿De qué año es la película "Aliens"?	file:///C:/Users/Gino/Desktop/websites/ga_secuelas/www.gq.com.mx/actualidad/cine-tv/articulos/cuales-son-las-mejores-secuelas-de-aliens.html
5	¿Qué hora aparece en la pantalla del Motorola Droid Turbo?	file:///C:/Users/Gino/Desktop/websites/uncrate_phones/uncrate.com/tech/cell-phones/index.html
6	¿Quién es el presidente de la Universidad de Columbia?	file:///C:/Users/Gino/Desktop/websites/columbia_about/www.columbia.edu/content/about-columbia.html
7	¿Cuál es el valor del Mercedes Benz G Class?	file:///C:/Users/Gino/Desktop/websites/ward_rides/www.ward.com/category/rides/index.html
8	¿Qué galería recomienda el autor a las 11:30 am?	file:///C:/Users/Gino/Desktop/websites/ds_london/www.designsponge.com/2015/03/24-hours-in-london-england-with-jamie.html
9	¿A cuál categoría pertenece el artículo "Baboons follow the pack, not the leader"?	file:///C:/Users/Gino/Desktop/websites/science_mag/www.sciencemag.org/magazine.html
10	¿Cuáles son los artículos populares de la semana en el siguiente sitio?	file:///C:/Users/Gino/Desktop/websites/ds_berlin/www.designsponge.com/2015/05/berlin-germany-city-guide.html
11	¿Quién ganó el premio Nobel de la paz en el año 2007?	file:///C:/Users/Gino/Desktop/websites/columbia_nobel/www.columbia.edu/content/nobel-laureates.html
12	¿Cuál es el precio de Vesper?	file:///C:/Users/Gino/Desktop/websites/bi_apps/www.businessinsider.com/best-iphone-only-apps-you-cant-get-on-android.html
13	¿Qué versión del Festival Internacional de Cine de Berlín se celebró en el 2015?	file:///C:/Users/Gino/Desktop/websites/es_berlin/www.elizabethstreet.com/travel/berlin-international-film-festival/the-luxe.html
14	¿Quién capturó la tercera fotografía del siguiente artículo?	file:///C:/Users/Gino/Desktop/websites/ds_argentina/www.designsponge.com/2012/12/nafermo-buenos-aires-argentina-city.html
15	¿Cuál es el rango de precio de BitBite?	file:///C:/Users/Gino/Desktop/websites/ga_bitbite/www.gq.com.mx/maquinas/sadets/articulos/bitbite-el-zadzet-ideal-para-los-que-aman-el-bitbite.html
16	¿Cuántas personas celiacas se calcula que hay en España?	file:///C:/Users/Gino/Desktop/websites/enfem/Recetas%20de%20luten%20-%20enfemenino.htm
17	¿En qué año se creó el Parque Nacional Torres del Paine?	file:///C:/Users/Gino/Desktop/websites/wiki_paine/es.wikipedia.org/wiki/Parque_nacional_Torres_del_Paine.html
18	¿En qué categoría aparecen jugadores de basketball?	file:///C:/Users/Gino/Desktop/websites/science_cat/news.sciencemag.org/categories.html
19	¿Cuál es la principal uva utilizada en la elaboración del pisco chileno?	file:///C:/Users/Gino/Desktop/websites/wiki_pisco/es.wikipedia.org/wiki/Pisco_%28guardiente%29.html
20	¿De quién es la cuenta de Instagram, "Mattbg"?	file:///C:/Users/Gino/Desktop/websites/ds_instagram/www.designsponge.com/2014/09/10-travel-instagram-feeds-to-follow.html
21	¿Para cuáles revistas escribe Marnie Hanel?	file:///C:/Users/Gino/Desktop/websites/ds_eggs/www.designsponge.com/2015/04/in-the-kitchen-with-marnie-andrea-and-ivy.html
22	¿Qué tipo de cerveza combina con el chocolate?	file:///C:/Users/Gino/Desktop/websites/primier_beer/www.primiermagazine.com/2012/learn/a-simple-guide-to-pairing-beer.html
23	¿Cuál es la edad promedio de los estudiantes que ingresaron al doctorado de Wharton en el 2015?	file:///C:/Users/Gino/Desktop/websites/wharton_phd/doctoral.wharton.upenn.edu/class-profile/index.html
24	¿Cuáles son los consejos para hacer durar más la batería del celular?	file:///C:/Users/Gino/Desktop/websites/primier_battery/www.primiermagazine.com/2009/field-manual/5-live-ways-to-make-your-phone-battery-last-longer.html
25	¿Qué puntaje asigna el sitio al teléfono Samsung Galaxy Note 4?	file:///C:/Users/Gino/Desktop/websites/wired/Review%20de%20Samsung%20Galaxy%20Note%204%20-%20Wired.html
26	¿Cuántos euros cuesta la botella del vino galardonado como el mejor del mundo?	file:///C:/Users/Gino/Desktop/websites/emol_vino/www.emol.com/tendencias/vino/Noticias/2015/06/09/27386/Expertos-recomiendan-vino-galardonado-como-el-mejor-del-mundo.html
27	¿Cuántas onzas (oz) de bourbon se necesitan para hacer un "Manhattan"?	file:///C:/Users/Gino/Desktop/websites/dc_manhattan/www.divecaroline.com/lifestyle/food-drink/how-much-mannhattan.html

Figura 5: Sujeto con experimento completado  
Fuente: Elaboración propia