

# Privacidad en las búsquedas, generalizando consultas

**César Augusto Goya Putpañe**

*Plan de estudios: Grado de Ingeniería Informática*

*Área del trabajo final: Seguridad Informática*

*Jorge Miguel Moneo*

*Helena Rifà Pous*

*Entrega: 10 junio 2021*



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Descripción del trabajo</i>
<b>Nombre del autor:</b>	<i>César Augusto Goya Putpañe</i>
<b>Nombre del consultor/a:</b>	<i>Jorge Miguel Moneo</i>
<b>Nombre del PRA:</b>	<i>Helena Rifà Pous</i>
<b>Fecha de entrega (mm/aaaa):</b>	<i>Junio/2021</i>
<b>Titulación:</b>	<i>Grado de Ingeniería Informática</i>
<b>Área del Trabajo Final:</b>	<i>Seguridad Informática</i>
<b>Idioma del trabajo:</b>	<i>Español</i>
<b>Palabras clave</b>	<i>Generalización, Búsquedas, Privacidad</i>
<p><b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i></p>	
<p>Realizas una consulta en un buscador de internet y tu información queda automáticamente procesada, analizada y almacenada en sus servidores, información que les permite crear un perfil tuyo bastante preciso. La falta de transparencia sobre qué ocurre con esos datos limita la percepción de privacidad que tenemos los usuarios. ¿Y si hubiera alguna manera de evitar que los buscadores puedan crear un perfil nuestro basándose en nuestras búsquedas?</p> <p>Este trabajo tiene como objetivo la realización de una extensión para el navegador <i>Google Chrome</i> con la misión principal de obstaculizar la labor de los buscadores al crear un perfil preciso de sus usuarios. La extensión utiliza la técnica de generalización de palabras utilizando la base de datos léxico <i>WordNet</i>.</p> <p>Tras una actividad de investigación se da a conocer la efectividad de la herramienta de generalización de palabras, concluyendo con un veredicto alternativo, gracias a nuevos reglamentos a favor de la privacidad, donde tomar el control de tus datos, haciendo uso de las herramientas que los propios tomadores de datos ponen a la disposición del usuario, fomentando así la educación digital.</p>	

**Abstract (in English, 250 words or less):**

You make a search in an internet search engine and your information is automatically processed, analyzed, and stored on their servers, information that allows them to create a fairly accurate profile of you. The lack of transparency about what happens with these data limits the perception of privacy that users have. What if there was some way to prevent search engines from creating a profile of us based on our searches?

The objective of this work is to create an extension for the Google Chrome browser with the main mission of hindering the work of search engines by avoiding creating an accurate profile of their users. The extension uses word generalization method using the WordNet lexical database.

After a researching a proposed activity, the effectiveness of the word generalization tool is revealed, concluding with an alternative verdict, thanks to new regulations in favor of privacy, where you can take control of your data, making use of the tools that own data takers make available to the user, thus promoting digital education.

# Tabla de Contenido

1.	Introducción.....	7
1.1.	Contexto y justificación del Trabajo.....	7
1.2.	Objetivos del trabajo.....	8
1.3.	Enfoque y método seguido .....	9
1.4.	Planificación del Trabajo .....	9
1.5.	Breve resumen del producto obtenido.....	11
2.	WordNet en acción .....	11
2.1.	Acerca de WordNet.....	11
2.2.	Uso de WordNet en otros trabajos.....	11
2.3.	Uso de WordNet en este trabajo final de grado .....	12
3.	Trabajos relacionados.....	13
3.1.	Búsquedas generalizadas .....	13
3.2.	Búsquedas apoyadas por redes sociales.....	13
4.	Estado del arte.....	13
5.	Creación de la extensión.....	14
5.1.	Requisitos.....	14
5.2.	Proceso de diseño .....	16
5.3.	Proceso de desarrollo .....	17
5.4.	Instalación de la extensión.....	21
5.5.	Validación de uso .....	23
6.	Probando la extensión .....	25
6.1.	Configurando el escenario .....	25
6.2.	Resultados.....	26
6.3.	Limitaciones .....	35
6.3.1.	Limitaciones de la solución a nivel técnico .....	35
6.3.2.	Limitaciones de la solución a nivel funcional.....	37
7.	El control sobre nuestros datos.....	38
7.1.	El reglamento GDPR.....	38
7.2.	Google y el control de la información.....	39
8.	Conclusiones.....	44

9.	GLOSARIO .....	45
10.	Anexos .....	47
	Anexo 1: PEC 1 Plan de trabajo .....	47
	Anexo 2: PEC 2 Investigación .....	47
	Anexo 3: PEC 3 Desarrollo.....	48
	Anexo 4: PEC 4 Memoria y producto resultante.....	48
	Anexo 5: Código fuente resultante del fichero manifest.json .....	48
	Anexo 6: Código fuente resultante del fichero index.html.....	49
	Anexo 7: Código fuente resultante del fichero style.css.....	49
	Anexo 8: Código fuente resultante del gs.js .....	50
11.	Bibliografía .....	51

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

Internet, tal y como lo conocemos hoy en día, difiere mucho a como era en sus inicios donde el acceso a contenidos era viable sin prestar demasiada atención a cómo se realizaba y si ésta cumplía ciertos criterios básicos de seguridad. Tanto así, que es sabido que Internet no se creó pensando en la seguridad [1] y debido a ello se ha visto a lo largo de su vida, implementando y mejorando protocolos existentes para que estos cumplan un nivel mínimo requerido de seguridad.

Podemos fácilmente recordar que no hace muchos años acceder a una web mediante el protocolo de hipertexto *http* era lo más habitual, para ahora recibir un aviso de seguridad si no se está utilizando el protocolo de hipertexto seguro *https*, tal y como lo adelantó *Google* en un artículo publicado por su equipo de seguridad dedicado a su navegador *Google Chrome* en el 2016 [2].

Con el avance incesante de esta red de redes, aparecieron los primeros buscadores que permitían indexar los sitios webs que por aquel entonces solo eran visibles si habían sido manualmente publicados en directorios webs, una especie de sitio web al estilo páginas amarillas que servía como punto de partida. Estos buscadores, hoy en día imprescindibles, tenemos grandes conocidos como *Google* que es el buscador de referencia, *Bing* de *Microsoft* y el cada vez más valorado por su apuesta por la privacidad *DuckDuckGo* [3], entre otros.

Diariamente se realizan millones de consultas en estos buscadores. Los usuarios buscan acceder a la información de forma eficiente y veloz, por lo que los algoritmos de indexación y ranking de los motores de búsqueda son revisados y mejorados periódicamente. Una de las técnicas utilizadas para ofrecer resultados de búsqueda personalizados consiste en almacenar información del usuario que se podrían categorizar como información sensible, como por ejemplo almacenar el historial de las consultas realizadas, los enlaces visitados, la localización desde donde realiza las consultas, etc.

Este tipo de técnicas de personalización, sin embargo, plantean un problema importante de privacidad y en definitiva de seguridad. La falta de transparencia de los motores de búsqueda sobre qué información es almacenada y cómo es utilizada, limita la percepción de privacidad que tienen los usuarios sobre el servicio de búsqueda.

Cada año que transcurre, los usuarios son cada vez más conscientes del asalto a su privacidad que incurren varios sitios web que manejan datos personales, así como convertirse en víctimas al ser expuestos por una filtración de datos de sitios webs en los que los mismo usuarios hayan depositado su confianza, como podrían ser las redes sociales, pasando por el modelo de negocio de empresas que hacen dinero a cambio de compartir datos personales que no les pertenecen y que venden a otras empresas, a esto sumarle el desconocimiento que dicha empresa ajena pueda hacer con esos datos.

¿Y si hubiera alguna manera de evitar que los buscadores puedan crear un perfil nuestro basándose en nuestras búsquedas? ¿sería esto una manera correcta de contrarrestar la invasión a nuestra privacidad? ¿A qué renunciaríamos si fuera posible limitar la creación de perfiles? y finalmente, ¿aceptaríamos renunciar a ello?

## 1.2. Objetivos del trabajo

Con la intención de focalizar los esfuerzos en la realización de este trabajo, se decide seleccionar el buscador de *Google* para el análisis, pruebas y fijación de objetivos. *Google*, no requiere presentación, domina el mercado de las búsquedas y de la publicidad, por lo que resulta conveniente centrarse en cómo la compañía realiza sus perfiles de usuarios en base a las búsquedas que realizamos en él.

Una de las primeras reacciones que llama la atención, es cómo *Google* ha aprendido a ofrecer a sus usuarios mayor control sobre su cuenta de usuario, claro que en parte esto ha sido promovido por la regulación GDPR [4] que entró en vigor el 25 de mayo del 2018, a la cual la empresa californiana está obligada a cumplir.

La generalización de consultas, como bien define la palabra “generalización”, es una técnica que permite omitir una consulta con contenido específico a un concepto más amplio y genérico, por ejemplo, si se busca la palabra coche, una generalización de esta sería buscar la palabra vehículo. Esta técnica es utilizada como alternativa para intentar mitigar la creación de perfiles precisos por parte de los buscadores. Este procedimiento en ningún momento evita la creación de perfiles por parte de los buscadores, más bien proporciona un perfil con información poco precisa, evitando así que el buscador obtenga un perfil ajustado y real del usuario, esta propuesta trae consecuencias negativas en la precisión de los resultados esperados, sin embargo, este trabajo intenta formalizar la creación de un pequeño aplicativo que nos ayude a generalizar estas búsquedas para poder ser posteriormente analizadas.

- 1.2.1. Este Trabajo de Final de Grado (TFG) propone, en primer lugar, estudiar los retos de implementación de un mecanismo sencillo de generalización de consultas que opere en modo local, es decir, en el lado del usuario. El nivel de desarrollo de la propuesta planteada, está acotada en su complejidad puesto que el objetivo es estudiar la viabilidad técnica de una herramienta que generalice consultas en tiempo real; implementar una extensión de este tipo requiere que haya acceso a una base de datos léxica a partir de la cual realizar generalizaciones de palabras, almacenar una base de datos de este calibre en el lado del usuario plantea problemas de almacenamiento, de computación y de rendimiento bastante evidentes cuando la protección debe garantizarse en tiempo real y sin la ayuda de infraestructura. El primer objetivo de esta propuesta es estudiar todos estos aspectos.
- 1.2.2. Este TFG tiene como objetivo investigar medidas de privacidad de perfiles en el caso concreto de la generalización de consultas. Aunque existen varias métricas en la literatura basadas en teoría de la información [5] [6] [7] [8] [9], este TFG dará prioridad a medidas intuitivas y fácilmente configurables.
- 1.2.3. Comprobar la idoneidad de realizar búsquedas generalizadas en comparación con búsquedas tradicionales y menos privadas, se busca responder qué tan beneficioso es realizar este tipo de búsqueda que tiene como efecto colateral la pérdida de precisión en los resultados obtenidos.
- 1.2.4. Proporcionar orientación del control actual que *Google* permite a sus usuarios, al haber sido impulsado por el nuevo reglamento de la Unión Europea. Este objetivo intenta recabar información de calidad para entender el nivel de familiaridad de los usuarios con



los distintos controles, personalización y administración que son actualmente posible con el buscador *Google*.

### 1.3. Enfoque y método seguido

Los perfiles que los buscadores crean de nosotros, los usuarios, es sustancial y proporciona grandes beneficios, ya que los resultados a nuestras búsquedas son más eficientes y precisas, lo que realmente es preocupante es la cantidad de datos personales que pueden llegar a recolectar sin nosotros ser conscientes y por ello la importancia de remarcar el equilibrio entre privacidad y calidad de los resultados.

Tras una investigación preliminar del tema, se han encontrado trabajos realizados en la misma materia con un enfoque basado totalmente en el desarrollo de una extensión para el navegador *Google Chrome*. Esto ha dado pistas de las posibilidades de la extensión y de los posibles motivos por el cual no ha sido ya implementado abiertamente.

En este trabajo se utiliza *WordNet*. *WordNet* es una gran base de datos léxica del inglés [10], de la que se habla en detalle el capítulo siguiente, el trabajo se apoya en esta base de datos de conocimientos para conseguir generalizar palabras introducidas por el usuario en el buscador. Para la consecución de esta tarea, se estudiará la viabilidad de generar una extensión para el navegador *Google Chrome*, dicha extensión facilitará la conexión con la base de datos *WordNet* y con ello recuperar el hiperónimo de la palabra que se desea buscar, así generalizando la búsqueda, de tal forma que si un usuario busca algo relacionado con “*computer program*”, la extensión transformará el término a buscar a una palabra más genérica, accederá a la base de datos de *WordNet* y recuperará su hiperónimo, como en este ejemplo sería la palabra “*software*”.

A diferencia del trabajo anterior, en el que también se usa una extensión de *Google Chrome*, en este trabajo se aborda el análisis y explicación lógica de las limitaciones tecnológicas. Analizaremos la infraestructura necesaria para mantener una extensión de este calibre y la probabilidad de una implementación sin la ayuda de ella. Averiguaremos la eficiencia de este tipo de búsquedas para conocer su precisión y posible uso o no uso de la herramienta en el mercado.

### 1.4. Planificación del Trabajo

Este trabajo de final de grado implica realizar el trabajo en seis bloques diferentes, siendo el penúltimo bloque dedicado a la presentación visual del TFG y el último bloque la defensa de este. Para la planificación principal, enfocamos el trabajo en los primeros cuatro bloques a los que denominamos: PEC 1, PEC 2, PEC 3 y PEC 4.

La PEC 1 se basa enteramente en la planificación que se va a seguir en el resto de los bloques (véase anexo 1), para completar el plan de trabajo, debemos dar respuestas a lo siguiente:

- Explicación detallada del problema a resolver
- La enumeración de los objetivos que se pretenden alcanzar con la realización del TFG
- La descripción de la metodología que se seguirá durante el desarrollo del TFG

- El listado de las tareas a realizar para alcanzar los objetivos descritos
- La planificación temporal detallada de estas tareas y sus dependencias
- Análisis de la herramienta WordNet
- Una revisión del estado del arte

La PEC2 se centra en la investigación y recolección de datos necesarios para la elaboración posterior del trabajo (véase anexo 2). Completar este bloque implica haber abordado los siguientes puntos:

- Entender qué es *WordNet* y qué trabajos se han realizado con esta herramienta en relación con la privacidad al momento de realizar búsquedas.
- Investigar proyectos de generalización de consultas/búsquedas o similares para entender el contexto en el que nos movemos, comprobar ventajas y desventajas de estas para intentar evitarlas.
- Implicaciones y limitaciones de una extensión de generalización de búsqueda, entender porque no se usa aún. Esta tarea pretende responder si la extensión es realmente eficaz y precisa a la hora de realizar las búsquedas generalizadas.
- Entender cómo funcionan los buscadores a nivel técnico. Se pretende buscar alternativas, anonimizar una búsqueda, evitar la recopilación de los datos del usuario cuando este realiza una búsqueda, por ello la importancia de entender cómo funcionan los buscadores hoy en día.
- Indagar en el nivel de control que un usuario tiene sobre su cuenta y qué puede hacer al respecto para evitar la recopilación de información desmedida.

La PEC3 se centra en la elaboración del contenido recopilado en el bloque anterior, esto implica que la dirección del proyecto pueda tomar un rumbo distinto al esperado, por lo que la metodología a usar debe ser lo suficientemente flexible para adaptarse a nuevas ideas que pueden aparecer durante el proceso de investigación (véase anexo 3).

- Realización de la extensión o en consecuencia motivos y razones para su descarte
- Documentar especificaciones técnicas de la extensión
- ¿Existen limitaciones? ¿cuales?
- Detallar motivos y razones que expliquen porque una extensión de estas características no se ha desarrollado ya o porque no se ha le ha dado tanta publicidad. ¿Sopesan más los inconvenientes de este?
- Redactar información recopilada en el bloque anterior respecto al control actual que tienen los usuarios sobre sus datos
- Poner en perspectivas las posibles alternativas a la extensión que se hay podido encontrar
- Evaluar posibilidad de desarrollar dichas alternativas

La PEC4 queda reservada para la finalización en las redacciones pendientes del bloque anterior, no se esperan modificaciones del trayecto marcado, más bien completar (véase anexo 4), finalizar las tareas pendientes, así como dejar el producto del trabajo final en un estado presentable.

Para acceder a la información de la planificación con línea de tiempo de este trabajo, véase anexo 1.

## 1.5. Breve resumen del producto obtenido

Siguiendo la planificación mencionada anteriormente, esta ha guiado el proyecto hacia el desarrollo de una extensión para el navegador *Google Chrome*, que una vez instalado facilita la conversión en tiempo real de palabras pertenecientes al vocabulario inglés hacia términos más amplios y genéricos, con lo cual se espera permita en cierto modo evadir la creación precisa de perfiles de usuarios de los motores de búsqueda, con el objetivo de salvaguardar la privacidad del usuario.

Esta extensión recibe el nombre de *Generalized Search*.

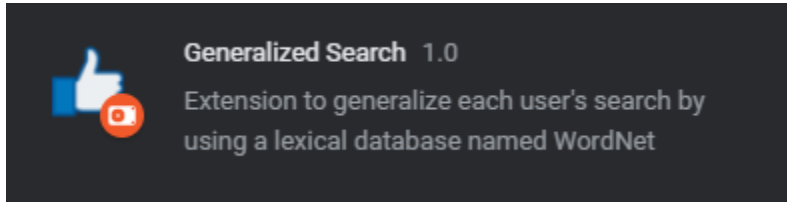


Figura 1. Descripción de la extensión instalada en Google Chrome.

## 2. WordNet en acción

### 2.1. Acerca de WordNet

*WordNet* es una gran base de datos léxica del inglés. En esta base de datos se pueden encontrar sustantivos, verbos, adjetivos y adverbios agrupados en conjunto de sinónimos cognitivos llamados *synsets*. Esta red resultante de palabras y definiciones es accesible mediante un navegador web, lo cual lo convierte en una herramienta útil de acceso para la lingüística computacional y el procesamiento de lenguaje natural [10].

### 2.2. Uso de WordNet en otros trabajos

#### 2.2.1. Idoneidad de aplicar *WordNet* para medir la privacidad

Este artículo tiene como objetivo analizar la idoneidad de aplicar *WordNet* para medir la similitud semántica o la relación de palabras en el campo de la privacidad. Argumenta que en plena era de la *Big Data*, la protección de la privacidad es un tema fundamental y que, para personalizar la protección de la privacidad, primero se debe medir la cantidad o el grado de datos privados filtrados. Se desea determinar si *WordNet* puede ser útil para tal labor.

Para cada palabra, *WordNet* lista sinónimos y/o diferentes definiciones. Dentro del dominio de la privacidad, se intenta determinar, para una palabra, si *WordNet* es ideal para obtener otras palabras similares semánticamente o definiciones que siguen estando relacionadas con la privacidad. El análisis incluye un experimento donde se pide a personas calificar un listado de pares de palabras relacionadas con la privacidad para luego compararla con lo que ya existe en la base de datos *WordNet*, la tarea de la persona es puntuar si estas palabras siguen teniendo relación con la privacidad o si esta se pierde.

### 2.2.2. Usando *WordNet* para refinar las consultas a Google

Este trabajo basó su argumento en que la gran mayoría de búsquedas en *Google*, eran de palabras genéricas, en la que según *Search Engine Journal*, en el 2004 más del 50% de las consultas era de una o dos palabras. El problema principal que este trabajo plantea dar solución era evitar la ambigüedad en las consultas, ya que esto suponía una fuerte limitante del uso de la tecnología de búsqueda web [11].

*WordNet* jugaba un papel importante al proporcionar los hiperónimos de la palabra a buscar, por ejemplo, si se busca la palabra Plutón (pluto en inglés), este podría ser un planeta, un dios griego o un personaje de *Disney*, dando así las siguientes posibilidades como sinónimos candidatos:

- sinónimo 1: planet, sun, orbit
- sinónimo 2: greek, god, mythology
- sinónimo 3: cartoon, character, Walt Disney

Luego tras aplicar el algoritmo de *Hearst Patterns*, el cual aplica patrones léxico sintáctico con la intención de reconocer hiponimia (palabra de definición más específica) se empezaba a eliminar la ambigüedad y se conseguían parejas de palabras. Si escogemos *pluto* como planeta (sinónimo 1), este podría ser:

- Pluto is a planet
- Pluto is a sun
- Pluto is a orbit

Se realizaba entonces una búsqueda por cada una de estas frases y de esta manera, contar los resultados obtenidos, teniendo en cuenta la popularidad se pretende deshacer la ambigüedad.

### 2.3. Uso de *WordNet* en este trabajo final de grado

Los *synsets* mencionados anteriormente están interconectados en base a su relación léxica y semántica lo cual permite realizar búsquedas jerárquicas de palabras más genéricas llamadas hiperónimos o palabras más específicas llamadas hipónimo, a partir de la introducción de una palabra.

Para este Trabajo de Final de Grado, se desea generalizar las búsquedas, con la intención que ello aumente la privacidad del usuario al no proporcionar información específica al buscador, lo cual evita que puedan crear un perfil preciso del usuario. Al generalizar las palabras a buscar, lo que realmente buscamos en la base de datos de *WordNet*, es el hiperónimo de la palabra introducida por el usuario.

## 3. Trabajos relacionados

### 3.1. Búsquedas generalizadas

Uno de los trabajos encontrados, es de una estudiante de la Universidad Oberta de Catalunya que realizó este mismo tema a inicios del año 2020 [12]. En él, se estudia la implementación de un mecanismo para generalizar consultas web que opere en modo local. El producto resultante es una extensión para el navegador *Google Chrome* que ha centrado su funcionamiento en el buscador *Google*, donde dicha extensión detecta el texto introducido en el buscador de *Google*, un *script* realizar una selección de una palabra más genérica, buscando en la base de datos de *WordNet*, devuelve la nueva palabra y esta reemplaza la que haya introducido el usuario en el buscador. Este método tiene como consecuencia una búsqueda poco precisa y se limita a realizar búsquedas de una única palabra ya que, al utilizar varias palabras, formando una frase, se corre el peligro de perder aún más la precisión sobre lo que se intenta buscar, ya que cada palabra independientemente realizaría una selección de su hiperónimo, entonces difícilmente se podría elegir el hiperónimo correcto en consonancia con las demás palabras.

### 3.2. Búsquedas apoyadas por redes sociales

Este otro proyecto [13], con similar objetivo, indaga en la posibilidad de proteger la privacidad de los usuarios al realizar búsquedas mediante el uso de las redes sociales. Este estudio no evita que los buscadores formen un perfil del usuario, sin embargo, el perfil que se genera está distorsionado por el uso de las redes sociales. A grandes rasgos, el producto de este estudio es un aplicativo que se inicia cuando un usuario intenta realizar una búsqueda, dicha palabra o frase a buscar es reenviada a una red social, en la cual habría usuarios que la reciben y a su vez lo reenviarían a otros usuarios dentro de la red social, hasta que uno de ellos decidiese realizar la búsqueda y devolver el resultado. Como se aprecia, en ningún momento se evita que el buscador cree un perfil del usuario, pero sería un perfil anonimizado al haber sido una búsqueda realizada por otro usuario, no el original.

## 4. Estado del arte

Tras revisiones de diferentes trabajos, se ha encontrado una variedad de trabajos relacionados con la privacidad de las búsquedas, no necesariamente utilizando la base de datos léxico *WordNet* que se ha explicado anteriormente, pero sí enfocado en resolver el problema de filtración de información hacia motores de búsquedas que posteriormente usan para la generación de perfiles. Uno de esos trabajos utiliza las redes sociales para crear búsquedas anónimas [13], que no evita la generación del perfil del usuario, pero este se ve afectado por diferentes procedencias (la red de contactos), lo cual provoca la generación de un perfil de usuario que no es real, ni mucho menos precisa, las búsquedas se realizan anónimamente, pero no están relacionadas al haber sido influenciadas por diferentes usuarios de la red de contactos.

Otro trabajo similar al propuesto en este trabajo propone una extensión para el navegador *Google Chrome* que generaliza las búsquedas mediante *WordNet* [12], sin embargo, carece en analizar más

en detalle el motivo por el cual la forma en la que está implementado proporciona una precisión reducida en los resultados de las búsquedas, lo cual desestima la propuesta y uso.

También se ha encontrado una propuesta interesante de otra extensión de *Google Chrome* que limita el envío de ciertas *cookies* de la sesión durante una búsqueda, de tal manera que consigue dos objetivos claves, el primero, impedir que se rastree al usuario y se realice el perfil en base a sus búsquedas, y el segundo, permite continuar utilizando los demás servicios de Google, ya que simula estar con la sesión cerrada, pero sólo para búsquedas, si el usuario se dirige a otros servicios de *Google* como *Gmail* o a *Google Drive*, la sesión de usuario está iniciada y es válida [14].

## 5. Creación de la extensión

Se empieza por hacer un repaso necesario para entender qué es lo que se va a crear en este apartado. ¿Qué es una extensión para *Google Chrome*? Una extensión de *Google Chrome* es un pequeño aplicativo que se añade al navegador *Google Chrome*, de esta manera permiten añadir nuevas funcionalidades al navegador, así como permitir personalizar la experiencia del usuario, mediante modificaciones visuales. La creación de una extensión se desarrolla usando tecnologías *Web*, como HTML, JavaScript y CSS, los cuales se ejecutan en un entorno separado llamado *sandboxed* desde donde la extensión interactúa con el navegador *Chrome* [15].

Algunos ejemplos del tipo de funcionalidades que permiten las extensiones son:

- Herramientas de productividad
- Enriquecimiento de contenido web
- Agregación de información
- Juegos y pasatiempos
- Bloqueo de publicidad
- Optimizador de memoria
- Notas en páginas web
- Gestor de contraseñas

*Google* pone a disposición del público, una plataforma donde cualquier desarrollador puede publicar extensiones, esta plataforma se llama *Chrome Web Store* [16], a pesar del nombre y que existen extensiones de pago, la gran mayoría de extensiones encontradas en esta tienda, son gratuitas.

### 5.1. Requisitos

Una vez se ha dado a conocer lo que es una extensión para el navegador *Google Chrome*, se listan los requisitos necesarios para llevar a cabo la creación de la extensión que da a lugar a este trabajo de final de grado.

Se separan los requisitos en dos tipos, donde el primer tipo es un requisito operacional, necesario para que la extensión funcione. El segundo tipo de requisitos es el estudio de los requerimientos funcionales para la extensión de este trabajo.

#### 5.1.1. Requisitos Operacionales

Todas las extensiones necesitan del fichero manifiesto (*manifest.json*), este fichero de formato *JSON* almacena información esencial, cómo el nombre de la extensión, descripción de la extensión, la versión de la extensión y la versión del manifiesto utilizado, la cual a fecha actual, debe ser la versión 2 en adelante, ya que la versión 1 se dejó de utilizar después de actualizar *Google Chrome* a su versión 18 [17].

El siguiente componente requerido es un fichero HTML que tendrá la estructura de la extensión. Técnicamente hablando, una extensión *Chrome* puede realizar operaciones sin necesidad de ser visualizado por el usuario, es decir en segundo plano, lo que implica que el fichero *HTML* es opcional, sin embargo, para el propósito de este trabajo, es necesario. Un fichero CSS, para personalizar la apariencia visual de la estructura *HTML*. Un fichero JavaScript, que es donde se desarrollará la parte lógica y operacional de la extensión. Opcionalmente, un icono, utilizado para personalizar y representar visualmente la extensión en la lista de extensiones del navegador.

### 5.1.2.Requisitos Funcionales

Aunque existen varias maneras de aproximar la solución, la finalidad es que la extensión recupere la búsqueda que desea hacer el usuario, generalizarlo y enviarla al buscador habitual del usuario. La decisión de focalizar los esfuerzos sobre el buscador *Google*, es acertada por ser el dominante en el sector de búsquedas web, por lo que la extensión una vez recupere el valor a buscar, la ha de enviar al motor de búsqueda *Google* y este mostrar los resultados como de costumbre.

La generalización de la búsqueda debe ocurrir al momento que el usuario le da a buscar, es allí cuando la conversión del valor a buscar es generalizada utilizando la base de datos léxica *WordNet*. Para entender mejor cuando se realiza la conversión, analizamos cómo un usuario realiza búsquedas.

1. Abre el navegador o en su defecto una pestaña nueva
2. Hace clic en la barra de direcciones (*Google Chrome* la llama *omnibox*.)
3. Introduce el texto a buscar
4. Le da a la tecla *ENTER*

Y el buscador predefinido realiza la búsqueda y presenta sus resultados.

Con la intención de ser lo más transparente posibles, la extensión para *Chrome* deberá permitir el mismo proceso de funcionamiento o acercarse lo más posible. Solicitar al usuario que abra la extensión *Chrome* para realizar la búsqueda desde allí es cambiar el flujo habitual, por lo que no sería acertado. Por lo que, la transparencia para el usuario es uno de los requisitos funcionales.

Aunque la base de datos de *WordNet* es bastante completa, carece de muchos términos considerados argot, los cuales también son utilizado por usuarios, la extensión no debe impedir que se realice una búsqueda por no encontrar el término que se intenta buscar, debe siempre devolver un resultado, entonces debe presentar mecanismos de fiabilidad y confiabilidad.

## 5.2. Proceso de diseño

Se plantea un diagrama del diseño de la extensión a crear. Este diagrama muestra los diferentes componentes que se deben desarrollar, se muestra también cómo interactúan estos componentes entre ellos.

El usuario inicia el proceso al realizar una búsqueda desde la extensión de *Chrome*, este mediante un fichero *JavaScript*, que ejecuta código en el lado del usuario, realiza una petición a una función que se encuentra almacenada en la nube de *Azure*, dicha petición contiene la palabra que el usuario ha introducido y para la que desea buscar su alternativa.

La importancia de emplear una función en la nube de *Azure* es debido a la necesidad de esta extensión de hacer uso del intérprete de *Python*, que es un lenguaje de programación que ayudará con la búsqueda de palabras al permitir navegar la interfaz de *WordNet*.

Una vez el intérprete de *Python* se ha inicializado, este procederá a la carga del módulo *NLTK*, este módulo explicado anteriormente permite navegar dentro de la base de datos léxica de *WordNet*, donde realizaremos la búsqueda de la palabra deseada por el usuario.

Si la palabra es encontrada, devolvemos el listado de hiperónimos disponibles al intérprete de *Python*, para que este decida qué valor devolver al *script JavaScript* de la extensión.

El *script JavaScript*, al recibir un valor a su petición inicial, realiza una segunda petición, solicita al navegador web mostrar una dirección web en su pestaña actual, esta dirección web es la página web del buscador *Google*. Junto con la petición de abrir el navegador, entrega la palabra que recibió como valor e indica al buscador que busque dicha palabra.

Como se puede intuir por la explicación del flujo de procesos, existen dos desarrollos independientes y complementarios. Por una parte, se han de diseñar la extensión propiamente dicha, que es un conjunto de ficheros, HTML, CSS y JavaScript, todos funcionales en el lado del usuario. La segunda parte del desarrollo es del lado del servidor, toda la parte lógica que realiza las operaciones de búsquedas y devolución de valores al *script* del lado del cliente.

En la figura 2 se muestra un diagrama de flujo resumiendo el proceso a seguir por la extensión a crear, así como visualizar el diseño e interconexión de los diferentes componentes que entran en juego.



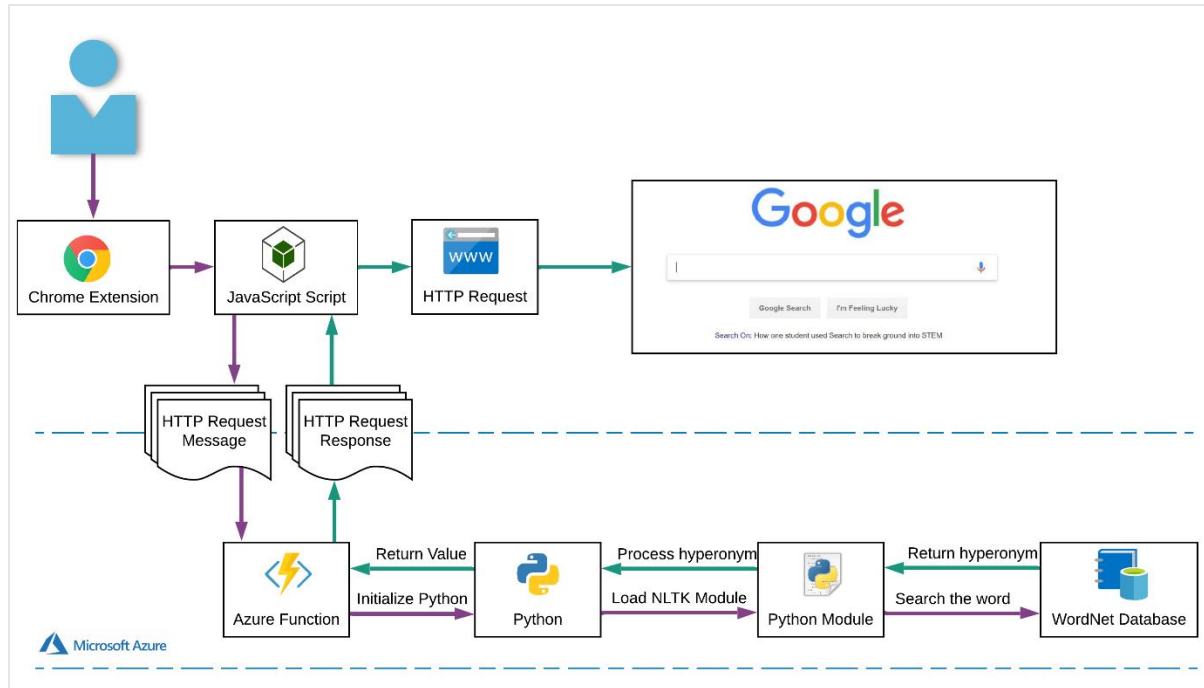


Figura 2. Diagrama de flujo de la extensión.

### 5.3. Proceso de desarrollo

Para iniciar con el desarrollo de la extensión, se inicia el proceso de desarrollo enfocando en el primer componente de la extensión, la parte del lado del servidor. Se empieza por desgranar el proceso de la creación de la extensión de *Google Chrome*. Como bien se explicó anteriormente, debemos generar los siguientes ficheros:

1. Fichero Manifiesto: manifest.json
2. Fichero HTML: index.html
3. Fichero CSS: style.css
4. Fichero JavaScript: gs.js

Y un quinto fichero en formato Python, que contendrá el código necesario para navegar la base de datos *WordNet* y encontrar la palabra buscada. Este fichero no es un requisito para la creación de una extensión *Chrome* en general, pero lo es para este proyecto.

5. Fichero Python: search.js

#### 1. Fichero Manifiesto: manifest.json

Un fichero manifiesto puede llegar a tener una gran cantidad de parámetros a configurar y todo ello dependerá de las necesidades de cada desarrollador y el producto en desarrollo.

Para un listado más completo de parámetros puede visitar el sitio web que *Google* ha reservado para ello, su *Google Chrome Developer Website*: <https://developer.chrome.com/docs/>

Las primeras líneas del fichero contienen información considerada básica para la extensión. Se especifica la versión del manifiesto a utilizar, el nombre que se le da a la extensión, la versión de ella, una pequeña descripción de lo que hace la extensión y por último el nombre del autor.

```
"manifest_version": 2,  
"name": "Generalized Search",  
"version": "1.0",  
"description": "Extension to generalize each user's search by using a lexical database named WordNet",  
"author": "Cesar Goya",
```

En la sección *browser\_action*, se especifica el fichero *HTML* que debe abrirse cuando se accede a la extensión una vez en instalada, por último, se especifica un icono en formato png para la extensión.

```
"browser_action": { "default_popup": "index.html", "default_icon": "img/icon.png" },
```

Toda extensión *Chrome* es accesible desde múltiples lugares y cada lugar requiere un icono de mayor o menor resolución, siendo este auto configurado por el mismo navegador *Chrome* cuando se le incluye un icono con suficiente resolución, lo recomendado es incluir los diferentes tamaños que pueda necesitar.

```
"icons": { "16": "img/icon16.png", "32": "img/icon32.png", "64": "img/icon32.png" },
```

Las siguientes capturas representan algunos lugares donde se aprecia el icono en acción.

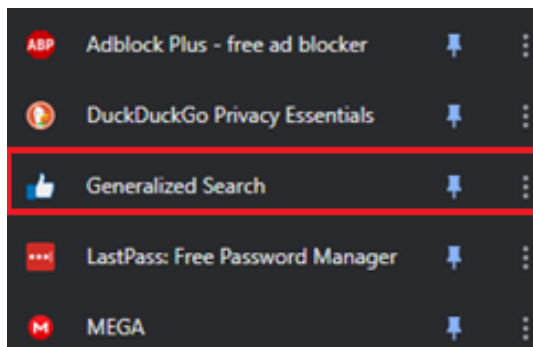


Figura 3. Lateral derecho del navegador, lista de extensiones.

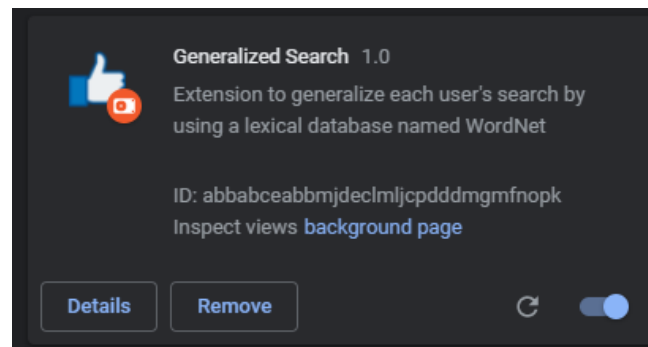


Figura 4. Visible al acceder `chrome://extensions/`

La sección *background* es donde se especifica los ficheros *JavaScript* que utilizará la extensión.

```
"background": { "scripts": ["scripts/gs.js", "scripts/jquery.min.js"], "persistent": false },
```

*Omnibox* es como llama Google Chrome a su barra de direcciones, este apartado es importante para cumplir con el requisito de transparencia para el usuario, requisito definido en el apartado 5.1.2.

```
"omnibox": { "keyword": "gs" },
```

Al configurar este parámetro se permite acceder a la extensión desde la barra de direcciones del navegador, mediante la inclusión de una palabra clave (*keyword*).

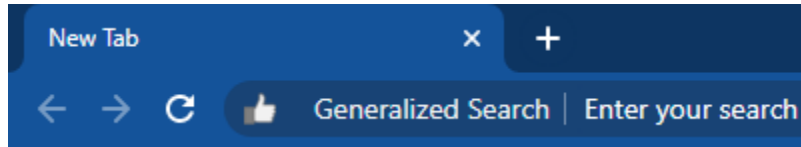


Figura 5. Barra de direcciones tras escribir "gs + barra espaciadora".

Por último, se otorga a la extensión permisos sobre las pestañas del navegador (*tabs*), necesario para que la extensión interactúe con las diferentes pestañas que el usuario tiene abierto. También, se otorgan permisos sobre el sitio web *azurewebsites.net*, donde reside parte del código involucrado con la búsqueda sobre la base de datos *WordNet*, el cual se explica en detalle en la sección de limitaciones de la extensión, sección 6.3.

```
"permissions": [ "tabs", "https://uocwordnet.azurewebsites.net/" ]
```

Para la lectura del código fuente resultante de fichero *manifest.json* véase anexo 5.

## 2. Fichero HTML: index.html

Este fichero es utilizado de modo informativo para proporcionar los pasos necesarios de cómo utilizar la extensión, el código es muy sencillo, no se requiere mayor complejidad en este apartado. Para la lectura del código fuente resultante de fichero *index.html* véase anexo 6.

## 3. Fichero CSS: style.css

Este fichero contiene la parte estética con la que se da una apariencia más amigable al fichero *HTML*. Como el contenido del fichero *HTML* es bastante básico con pocas etiquetas de hipertexto, el contenido de fichero *style.css* también lo es y apenas ocupa unas pocas líneas. Para la lectura del código fuente resultante de fichero *style.css* véase anexo 7.

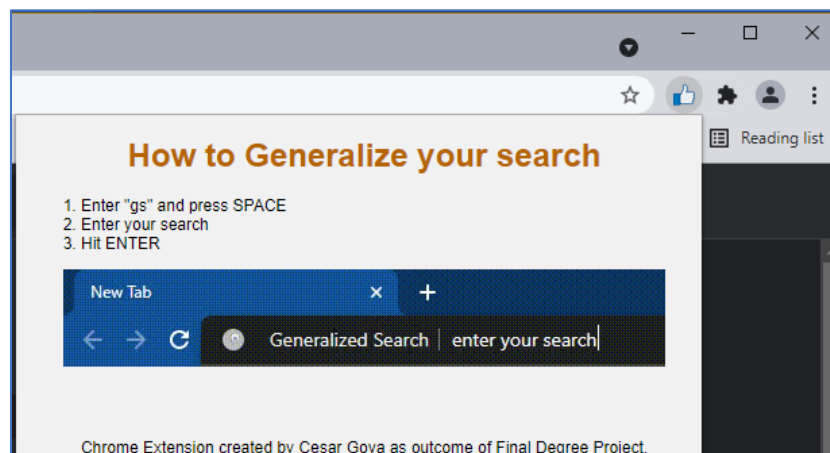


Figura 6. Resultado visual del HTML + CSS. Así se muestra la página web en modo ventana emergente.

#### 4. Fichero JavaScript: gs.js

Este fichero contiene la programación necesaria para detectar lo que el usuario ha escrito en la barra de direcciones (*omnibox*).

Mediante la función, `chrome.omnibox.onInputEntered.addListener()` se detecta el texto introducido y posteriormente envía el texto hacia un *script* almacenado en los servidores de *Azure*, que devolverá su hiperónimo. Por último, este fichero *JavaScript*, una vez ha recibido el hiperónimo, envía la nueva palabra al buscador definido, en este caso a *Google*.

Para la lectura del código fuente resultante de fichero *gs.js* véase anexo 8.

#### 5. Fichero Python: search.py

Este fichero almacenado en *Azure* mediante su tecnología *Azure Function Serverless Compute*, permite almacenar *scripts* que se ejecutan mediante una acción desencadenante (*trigger*). En este proyecto la acción desencadenante es la recepción de la búsqueda iniciada por el fichero *JavaScript* mencionado anteriormente.

El fichero *search.py* primeramente necesita importar un módulo primordial para su funcionamiento. *Natural Language Toolkit (nlk)*.

#### ¿Qué es NLTK?

*“NLTK es una plataforma líder para crear programas Python que funcionen con datos de lenguaje humano. Proporciona interfaces fáciles de usar para más de 50 corpus y recursos léxicos como WordNet”* [18].

Con estas líneas, se indica a *search.py* que debe cargar el módulo *NLTK* y descargar el paquete *WordNet* para su utilización en la ejecución del *script*.

```
import nltk
nltk.download("wordnet")
from nltk.corpus import wordnet as wn
```

*WordNet* agrupa nombres, verbos, adjetivos y adverbios en grupos de significado comunes, llamados sinónimos cognitivos. Estos grupos se llaman *synsets* y la interfaz de *WordNet* proporcionado por el módulo *NLTK* permite buscar palabras con la siguiente instrucción: `wn.synsets("palabra")`, el resultado de esta búsqueda es un listado de sinónimos cognitivos para dicha palabra, de la cual especificamos de la siguiente manera para elegir el sinónimo cognitivo que nos interesa: `wn.synsets("palabra")[i]` donde *i* es un valor numérico entre 0 y *n-1*, siendo *n* el número de elementos que conforman dicho conjunto, *synsets*.

La siguiente captura muestra un ejemplo a partir de la web oficial de *WordNet*, el cual muestra el resultado también en *synsets*, pero de una manera más visible.

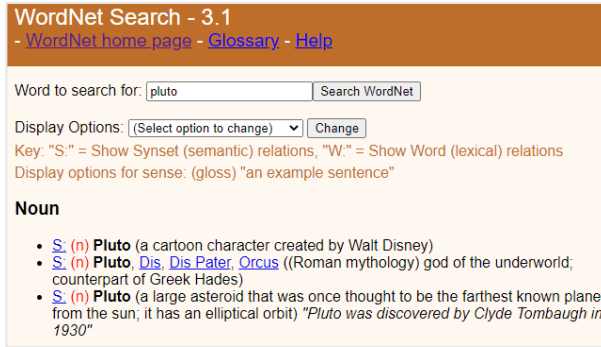


Figura 7. Buscando la palabra plutón en WordNet.

La búsqueda de la palabra pluto: `wn.synsets("pluto")` devuelve 3 sinónimos cognitivos.

- Pluto como personaje de Disney [posición 0]
- Pluto como Dios griego [posición 1]
- Pluto como ex planeta [posición 2]

Para acceder a pluto como dios griego se usaría la siguiente instrucción: `wn.synsets("pluto")[1]`

A partir de la obtención del *synset* deseado, se puede trabajar sobre ello de varias formas:

- Mostrar la definición de la palabra: `wn.synsets("pluto")[1].definition()`
- Mostrar su hipónimo: `wn.synsets("pluto")[1].hyponyms()`
- Mostrar su antónimo: `wn.synsets("pluto")[1].antonyms()`
- Mostrar su hiperónimo: `wn.synsets("pluto")[1].hypernyms()`

A continuación, se define una función en el lenguaje de programación *Python* que se encargará de procesar la palabra recibida desde el fichero *JavaScript* y realizar la búsqueda del hiperónimo de la palabra en cuestión. Una instrucción condicional (*if...else*) nos permite cumplir con uno de los requisitos funcionales comentados en el apartado anterior, el de proporcionar siempre un resultado y no detener la búsqueda al usuario cuando no se encuentre la palabra buscada: Mecanismos de fiabilidad y confiabilidad

```
def gs(s):
    if wn.synsets(s):
        result = [lemma.name() for synset in wn.synsets(s)[0].hypernyms() for lemma in synset.lemmas()]
        return result[0].replace("_", " ")
    else:
        return s
```

*Navegar la interfaz WordNet está explicada en detalle en su página web. [17]*

## 5.4. Instalación de la extensión

Para iniciar el proceso de validación, lo primero que se debe realizar es la instalación de la extensión. La extensión es un desarrollo que no está publicado en la *Chrome Web Store*, el sitio web oficial de *Google* para publicar extensiones para el navegador *Chrome*, por ello no es accesible desde dicho sitio web ni instalable como se haría con las demás extensiones allí publicada, más detalles son proporcionados en la sección 6.3 Limitaciones.

Los pasos para instalar la extensión de este trabajo final de grado, “*Generalized Search*” son los siguientes:

1. Abrir el navegador *Google Chrome*
2. Copiar y pegar la siguiente dirección en la barra de direcciones: *chrome://extensions*
3. Habilitar el modo desarrollador, haciendo clic en icono deslizable.

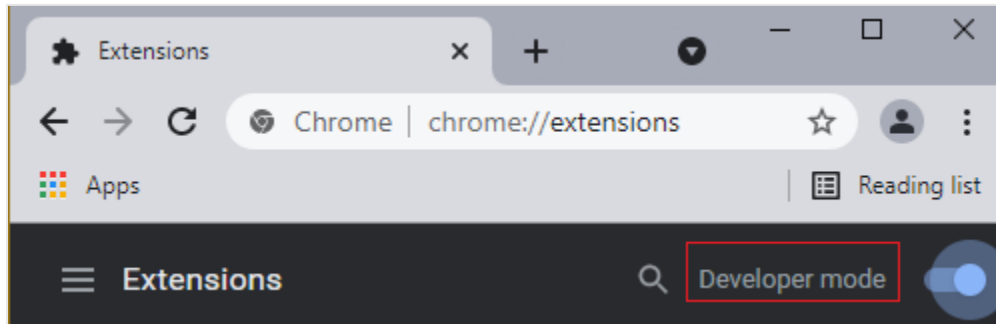


Figura 8. Habilitar el modo desarrollador en Google Chrome.

4. Clic en Cargar descomprimida (*load unpacked*)
5. Selecciona la carpeta raíz de la extensión

El navegador muestra la extensión cargada correctamente, tal y como se muestra en la siguiente imagen.

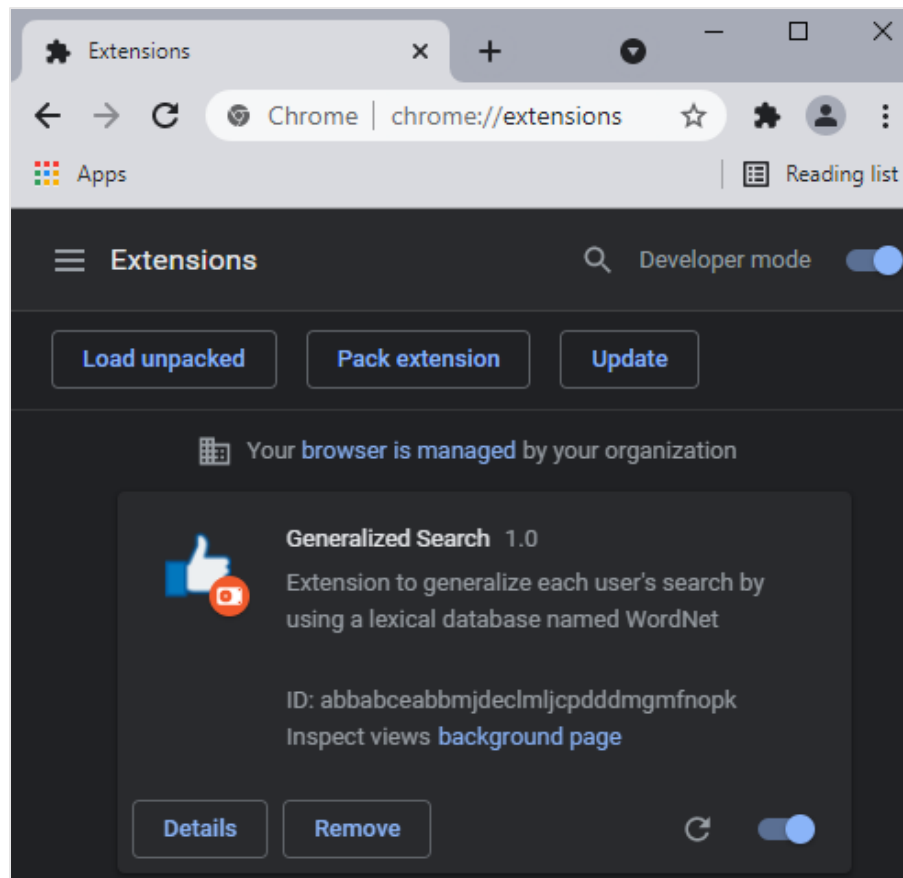


Figura 9. Navegador Google Chrome con la extensión correctamente instalada.

## 5.5. Validación de uso

Con la extensión correctamente instalada, se procede a realizar las primeras pruebas, para ello se debe abrir una pestaña nueva e introducir la palabra clave “gs”, palabra clave configurada en el fichero *manifest.json*, esta palabra activa la extensión para su uso desde la barra de direcciones.

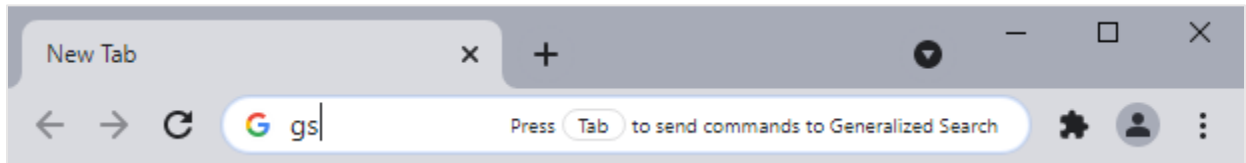


Figura 10. Introducción de la palabra clave que activa la extensión.

Cómo se aprecia en la imagen, el texto informativo “Presione TAB para enviar comandos a la extensión *Generalized Search*”. Presionar TABULAR o BARRA Espaciadora tienen el mismo resultado.

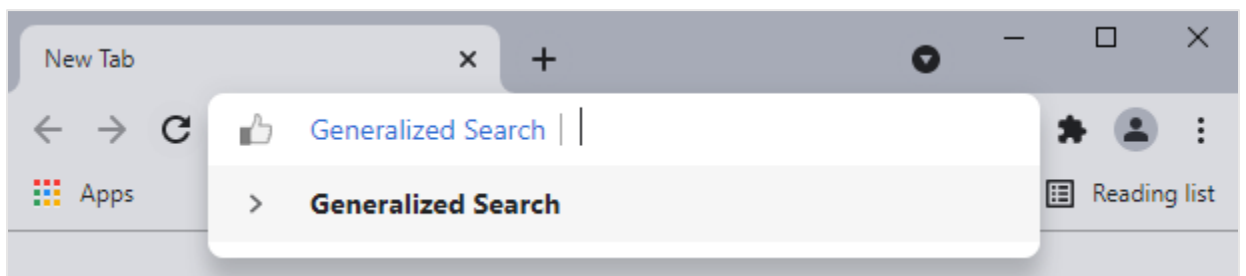


Figura 11. Visualización de la extensión en funcionamiento.

A modo de ejemplo, introducimos la palabra *house* y damos a ENTER. Se aprovecha este momento para recordar que la base de datos *WordNet* es una base de datos léxica del inglés.

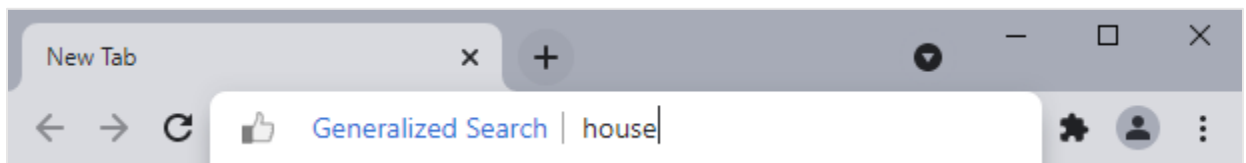


Figura 12. Visualización de la extensión en funcionamiento.

El hiperónimo de *house*, es *Building* y como tal, es lo que la extensión envía al buscador de Google.

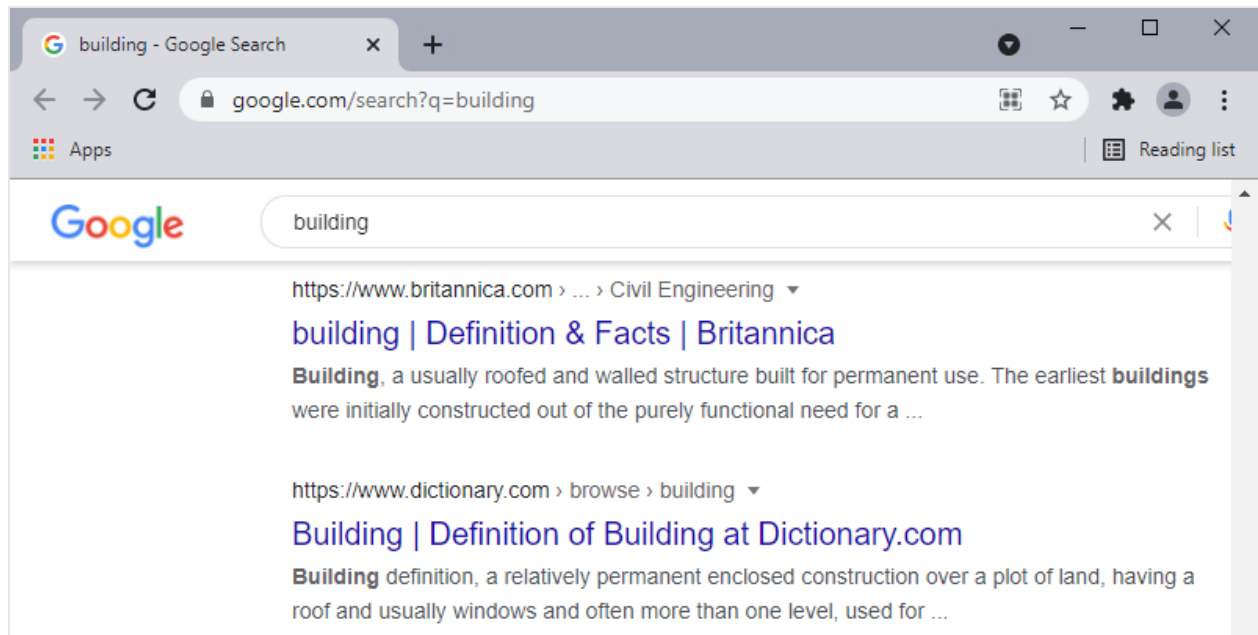


Figura 13. Resultado de la extensión en funcionamiento.

Como se puede apreciar, el desarrollo de esta extensión para *Chrome* no requiere de mayor complejidad, de hecho, no se entra en detalle de todo el código necesario para su funcionamiento, pero sí de las funciones necesarias para que este sea operativo y permita ser utilizado para la investigación siguiente.



## 6. Probando la extensión

### 6.1. Configurando el escenario

Una vez la extensión está funcionando, se desarrolla la siguiente práctica con la intención de dar visibilidad cuantitativa y cualitativa de los datos obtenidos. La intención de la práctica es demostrar la validez de una extensión de estas características en una aplicación de uso real y para ello se ha entregado la extensión a un grupo de gente para que realicen prácticas de búsquedas con la extensión.

A partir de ello responder un pequeño formulario donde recoger información que permita responder las preguntas planteadas en los objetivos iniciales de este trabajo.

#### Configuración

- Se crea un listado cerrado de palabras, dichas palabras están relacionadas en mayor o menor medida con una materia concreta, a la que llamamos categoría.
- Solicitamos al usuario buscar esa palabra mediante la extensión, la extensión realiza la conversión a su hiperónimo y realiza la búsqueda.
- Preguntamos al usuario si la palabra que la extensión ha buscado le parece acertada como sustituta de la primera palabra, dando tres posibles respuestas: sí, no, no estoy seguro.

El riesgo de utilizar hiperónimos es que la capacidad de generalización de una palabra concreta puede alejarse mucho de lo que realmente se intenta buscar. También es posible que aquellas palabras con diferentes definiciones, una vez haya sido “generalizada”, la extensión elija un hiperónimo (dentro del conjunto de hiperónimos disponibles) que cambie el significado inicial.

Para entender este último concepto, supongamos la palabra Plutón una vez más. Se mencionó anteriormente tres posibles significados para Plutón: como dios griego, ex planeta y personaje de Disney. La extensión debe elegir una de esas definiciones antes de proporcionar su hiperónimo, de lo contrario estaría realizando una tarea al azar, lo cual es aceptable a este nivel de complejidad de la extensión.

Lo que marca la diferencia es el grado de acierto que la extensión pueda tener en su elección y eso depende de la percepción humana, por ello categorizamos las palabras en una materia, y solicitamos al usuario que decida si la palabra buscada sigue teniendo relevancia y relación con dicha categoría.

#### Categorías y palabras

Privacy	Security
1. Aloofness	1. Bond
2. Isolation	2. Care
3. Quiet	3. Guarantee
4. Retirement	4. Insurance
5. Separation	5. Surveillance

Illness	Problem
1. Allergies	1. Complication
2. Cold	2. Dilemma
3. Cancer	3. Dispute
4. Flu	4. Issue
5. Conjunctivitis	5. Obstacle

Tabla 1. Lista de categorías y palabras preparadas para la investigación.

## 6.2. Resultados

La práctica realizada contó con la participación de 29 personas. En los resultados, los usuarios tuvieron que responder: sí, no o no estoy seguro, a la pregunta si aceptarían el término propuesto por la extensión como alternativa al primer término, siempre entendiendo que la palabra propuesta por la extensión siga teniendo una definición relacionada con la categoría.

Se busca entender la percepción de un usuario real frente a una hipotética situación en la que se utilice la extensión cotidianamente y cómo este, además de proporcionar mayor privacidad frente a los motores de búsquedas, creadores de perfiles de usuarios, sea a la vez útil y que usarlo permita mantener un balance en el acierto de la mayoría de los términos que buscarían en un buscador.

A modo de ejemplo seleccionamos el término *aloofness* y su hiperónimo proporcionado por la extensión: *indifference*. El 31% de los participantes responde que sí acepta el hiperónimo como alternativa, un 13.8% no está seguro y más de la mitad de encuestados, un 55.2% no lo acepta.

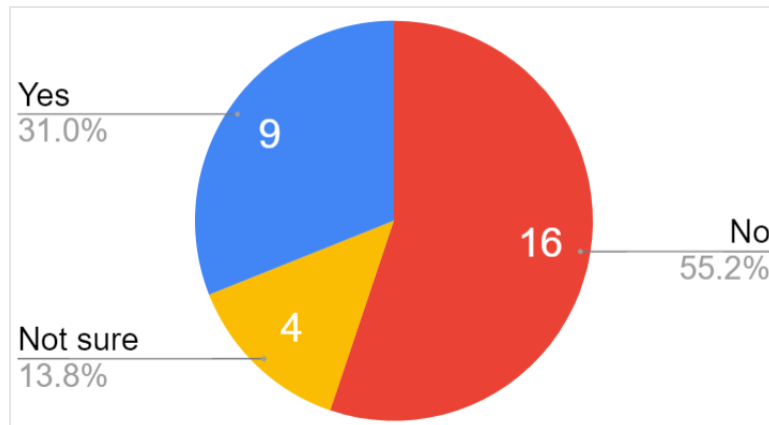


Figura 14. Gráfica de la categoría Privacy [aloofness > indifference]

A continuación, los resultados obtenidos por las distintas categorías proporcionadas.

Categoría: Privacidad (Privacy)

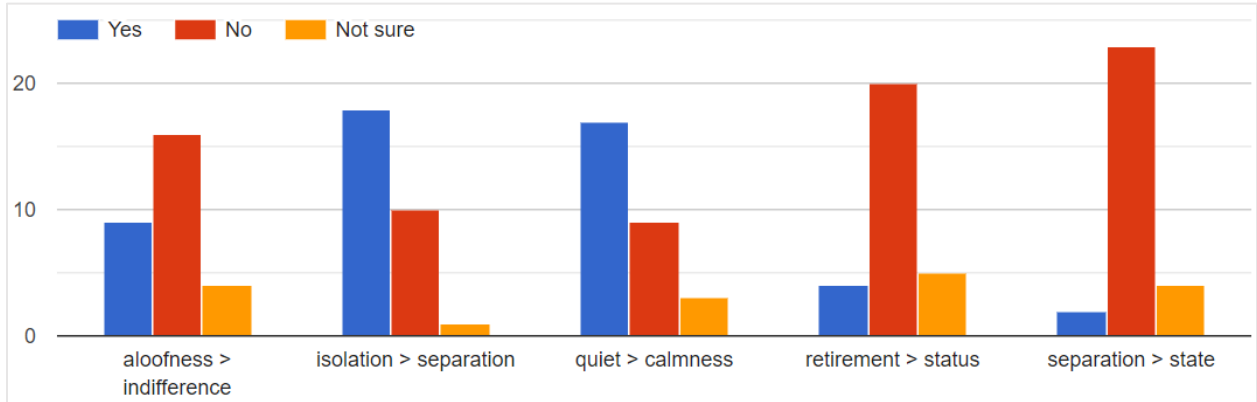


Figura 15. Gráfica en detalle de la categoría privacidad.

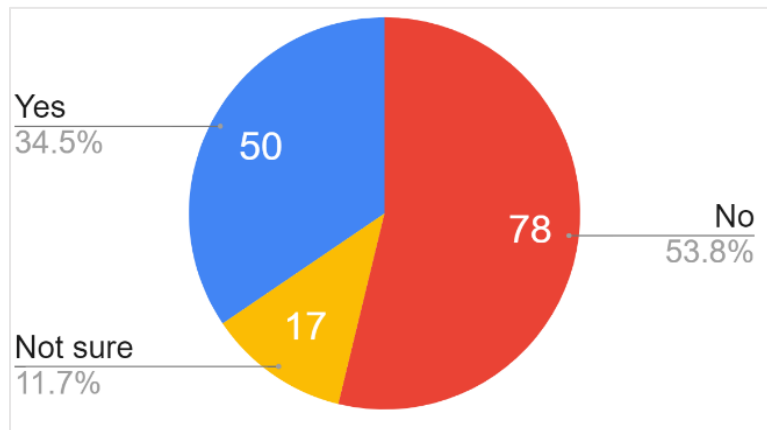


Figura 16. Gráfica resumida de la categoría privacidad.

Observando los resultados totales para la categoría *Privacy*, los participantes que aceptarían el término alternativo propuesto por la extensión es de sólo el 34.5%, frente a un 53.8% que respondieron que no aceptarían la palabra ofrecida por la extensión. El porcentaje de dudas es poco notorio en esta categoría con solo un 11.7%.

Los encuestados dejaron comentarios del estilo:

- “no reemplazaría una palabra negativa con una palabra neutra, positiva o viceversa”
- “no reemplazaría una palabra más específica con una palabra más general”
- “quiet es un adjetivo y calmness es un sustantivo. Usaría calm en todo caso”

Categoría: Seguridad (Security)

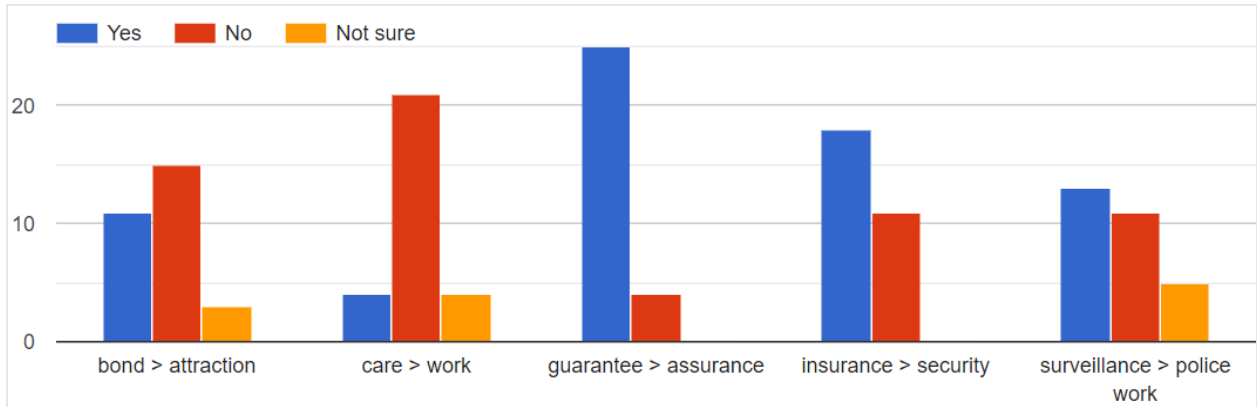


Figura 17. Gráfica en detalle de la categoría seguridad.

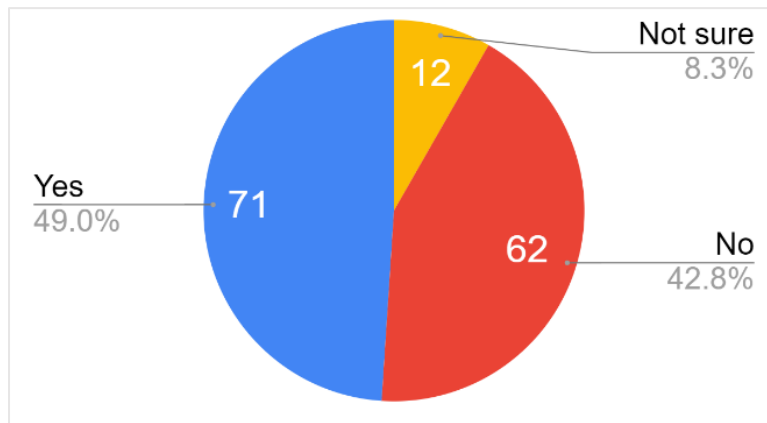


Figura 18. Gráfica resumida de la categoría seguridad.

En la categoría *Security*, se observa mayor aceptación para las alternativas propuestas, siendo del 49% frente a un 42.8% que rechaza la alternativa propuesta por la extensión. En esta categoría los porcentajes totales entre aceptación y rechazo son más cercanos. El porcentaje de dudas es poco notorio en esta categoría con apenas un 8.3%.

Los encuestados dejaron comentarios del estilo:

- “no tengo idea de cómo care está relacionado con work en el contexto de seguridad”
- “care significa work pero solo en situaciones concretas”
- “insurance y security, son palabras que tienen significados diferentes incluso en la misma categoría de seguridad”
- “habría situaciones donde las reemplazaría, pero no en todas, porque no siempre tienen el mismo significado o connotación”

Categoría: Enfermizo (Sick)

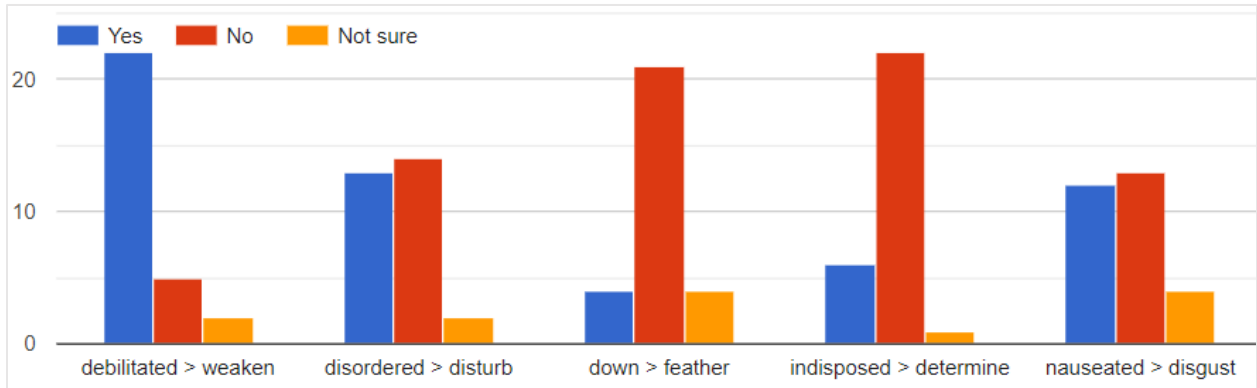


Figura 19. Gráfica en detalle de la categoría enfermizo.

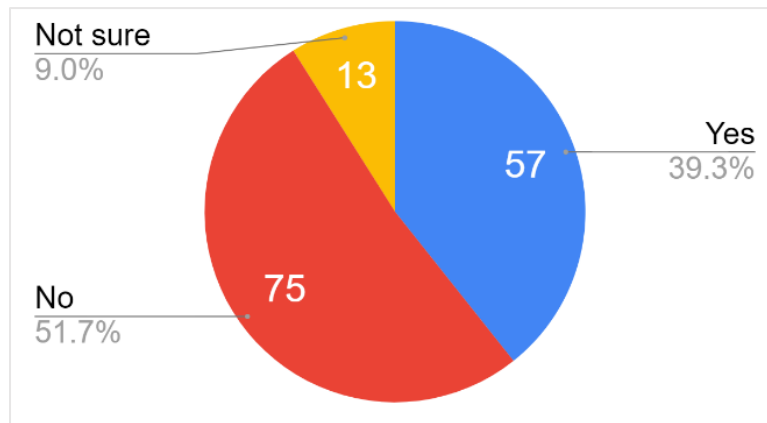


Figura 20. Gráfica resumida de la categoría enfermizo.

En la categoría *Sick*, los participantes que aceptarían el término alternativo propuesto por la extensión es de solo el 39.3%, frente a un 51.7% que respondieron que no aceptarían la palabra ofrecida por la extensión. El porcentaje de dudas es poco notorio en esta categoría con tan solo un 9%.

Los encuestados dejaron comentarios del estilo:

- *“debiliatate y weakness tiene significados muy similares. Disordered debería reemplazarse por disturbed, que es el pasado de disturb, Ocurre lo mismo con nauseated a ser reemplazado por disgusted en vez de disgust.”*
- *“down feather es en si una palabra compuesta en sí misma que describe un tipo de pluma”*

Categoría: Enfermedad (*illness*)

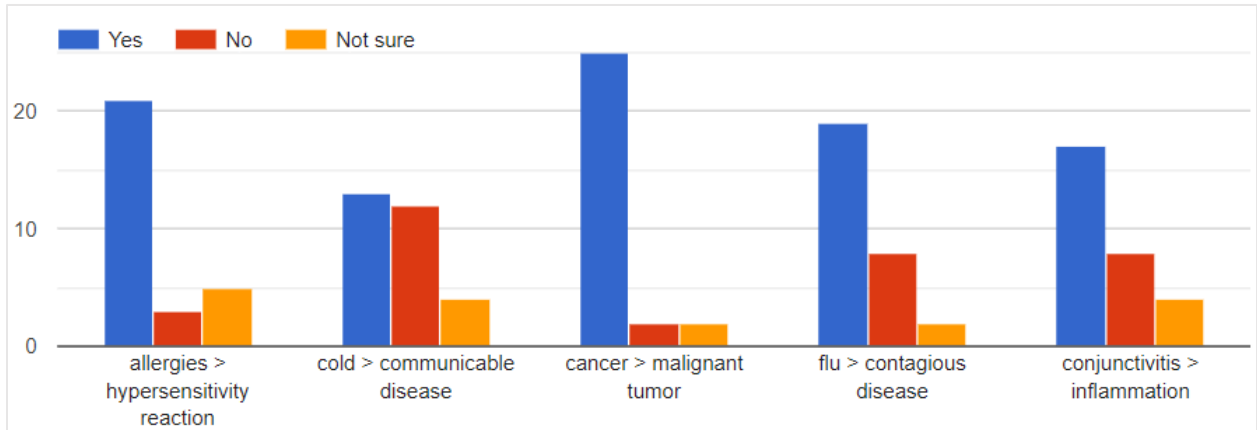


Figura 21. Gráfica en detalle de la categoría enfermedad.

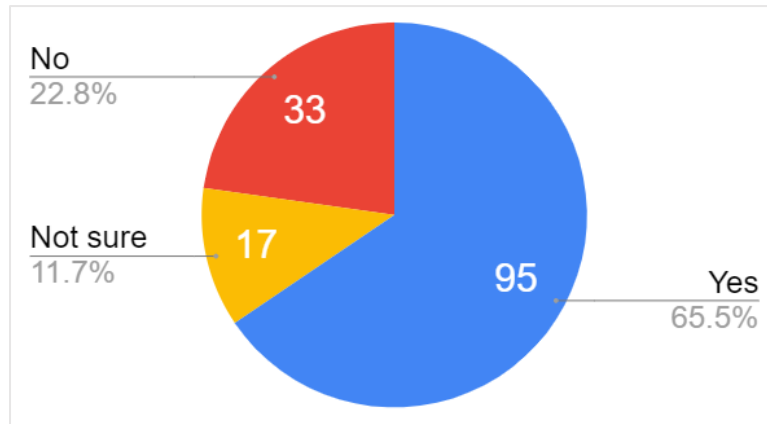


Figura 22. Gráfica resumida de la categoría enfermedad.

En la categoría *Illness*, se observa mayor aceptación en general para los términos descritos y sus alternativas propuestas, siendo del 65.5% frente a un 22.8% que rechaza la alternativa propuesta por la extensión. En esta categoría los porcentajes totales entre aceptación y rechazo se alejan considerablemente. El porcentaje de dudas es poco notorio en esta categoría con un 11.7%.

Los encuestados dejaron comentarios del estilo:

- “se pierde mucha precisión con los términos alternativos en esta categoría”

Categoría: Problema (Problem)

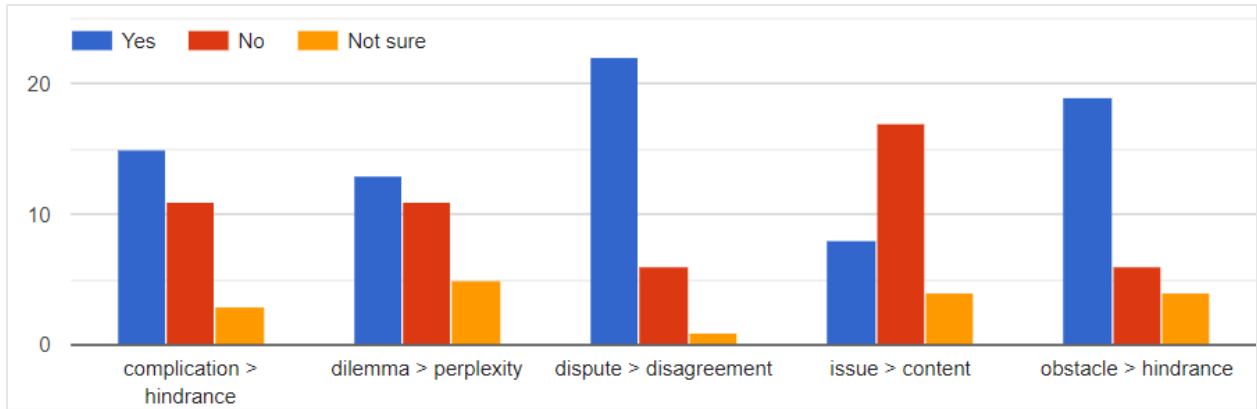


Figura 23. Gráfica en detalle de la categoría problema.

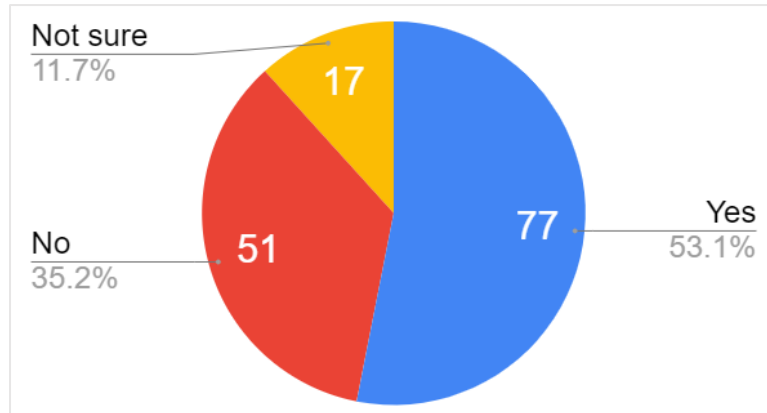


Figura 24. Gráfica resumida de la categoría problema.

En la categoría *Problem*, se observa mayor aceptación en general para los términos descritos y sus alternativas propuestas, siendo del 53.1% frente a un 35.2% que rechaza la alternativa propuesta por la extensión. En esta categoría los porcentajes totales entre aceptación y rechazo se alejan. El porcentaje de dudas es poco notorio en esta categoría estando por debajo del 12%.

Los encuestados dejaron comentarios del estilo:

- “no encuentro *perplexity* intercambiable con *dilemma* en el contexto de la categoría *problema*, pienso lo mismo de la palabra *content*”

Categoría: Animales (Animals)

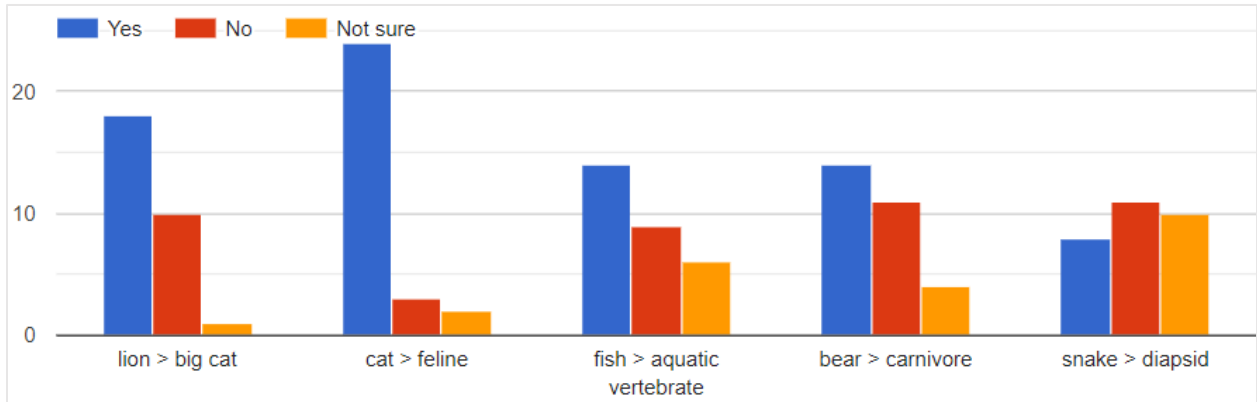


Figura 25. Gráfica en detalle de la categoría animales.

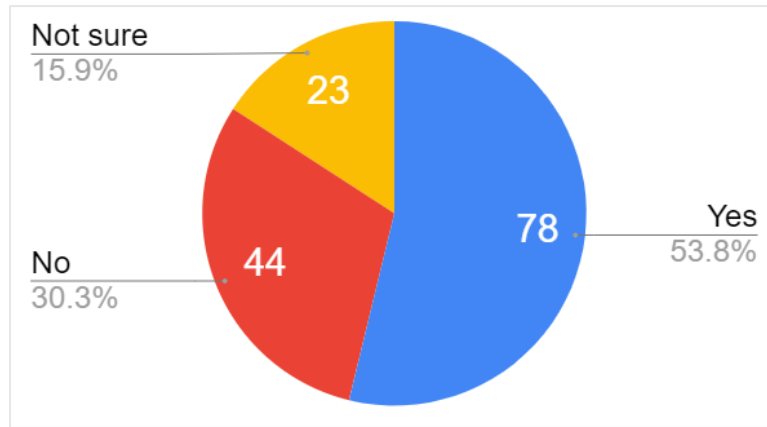


Figura 26. Gráfica resumida de la categoría animales.

En la categoría *Animals*, se observa mayor aceptación en general para los términos descritos y sus alternativas propuestas, siendo del 53% frente a un 30.3% que rechaza la alternativa propuesta por la extensión. En esta categoría los porcentajes totales entre aceptación y rechazo se alejan. El porcentaje de dudas es esta categoría es llamativo rozado el 16%.

Los encuestados dejaron comentarios del estilo:

- “no conocía el término diapsid”
- “feline y aquatic vertebrate, añade incluso mayor precisión por lo que son unas alternativas muy buenas”
- “big cat incluye muchos más especies”



La gráfica siguiente es un resumen general del resultado mostrado en las gráficas anteriores respecto a las categorías y cada una de las palabras incluidas. Entre los 29 participantes de esta actividad, se han recolectado un total de 428 síes que representa casi el 50% de los votos, frente a un, ligeramente por debajo del 40% de votos que han respondido que no aceptan las palabras alternativas, es decir la palabra proporcionada por la extensión. Apenas un 11.4% han dudado sobre si usar o no la palabra proporcionada por la extensión como una alternativa correcta.

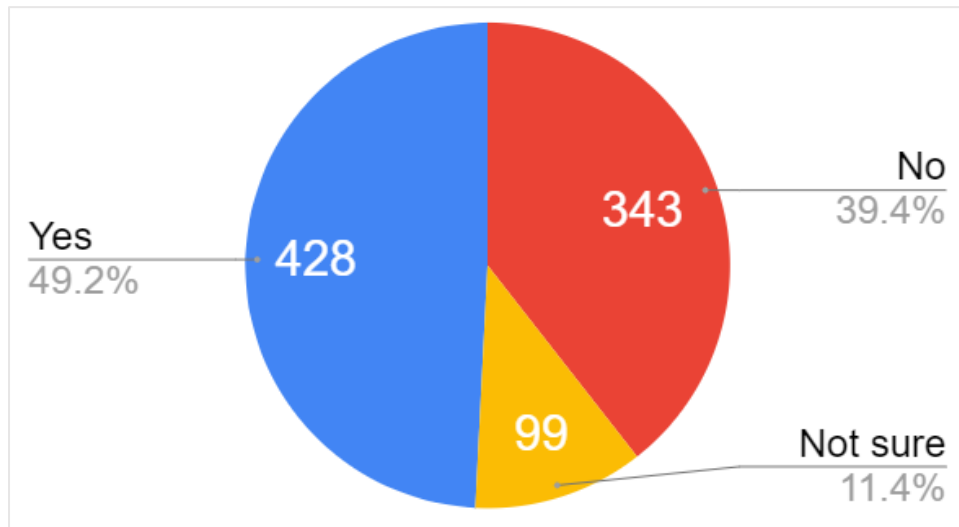


Figura 27. Gráfica resumida de todas las categorías.

Para culminar con los resultados de la actividad, se ha recolectado información demográfica básica de los participantes.

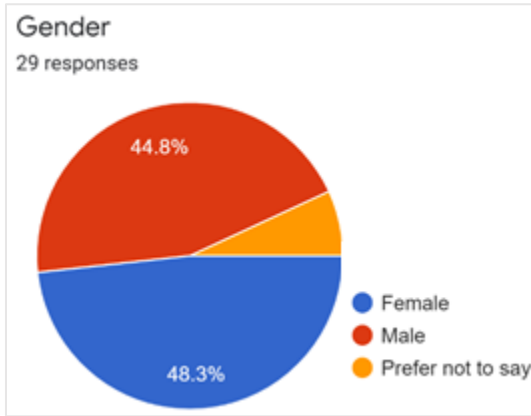


Figura 28. Gráfica en detalle por géneros.

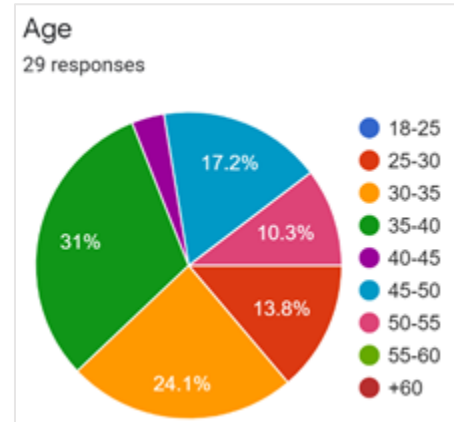
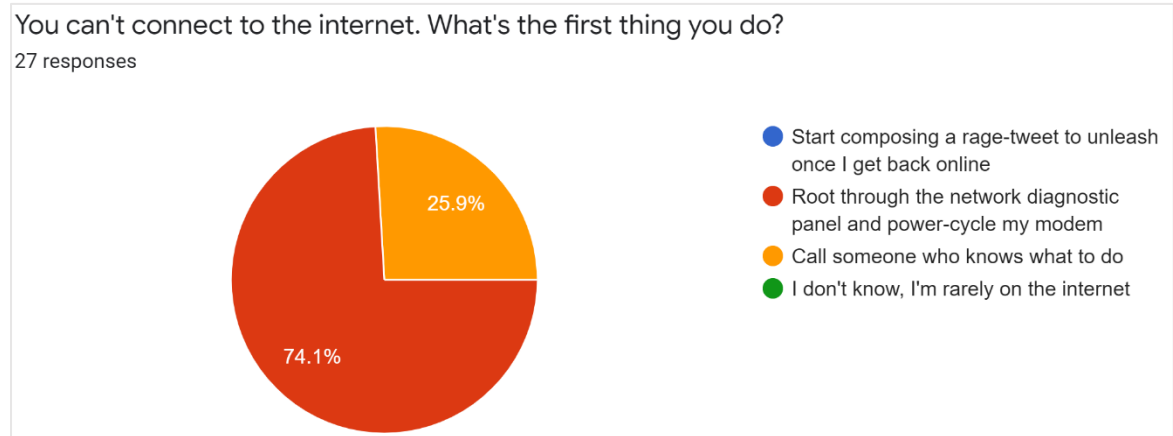
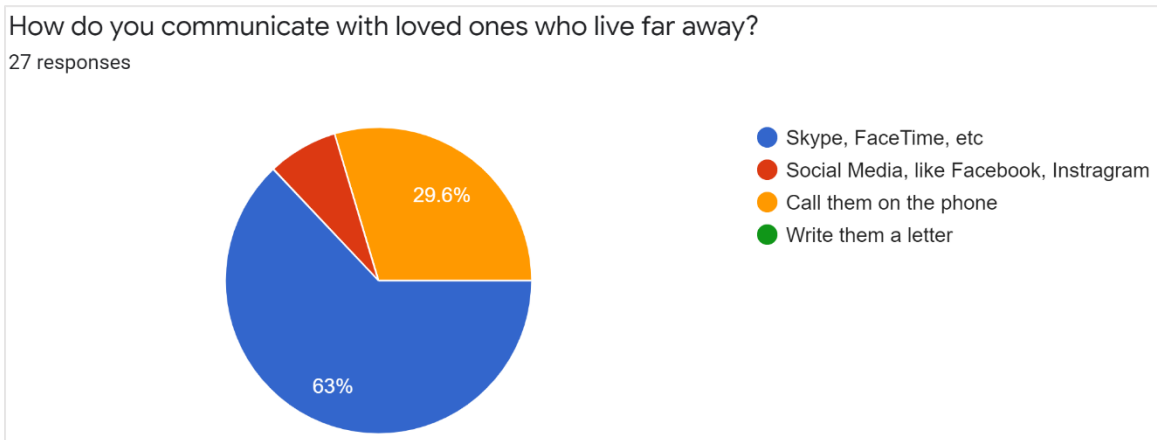
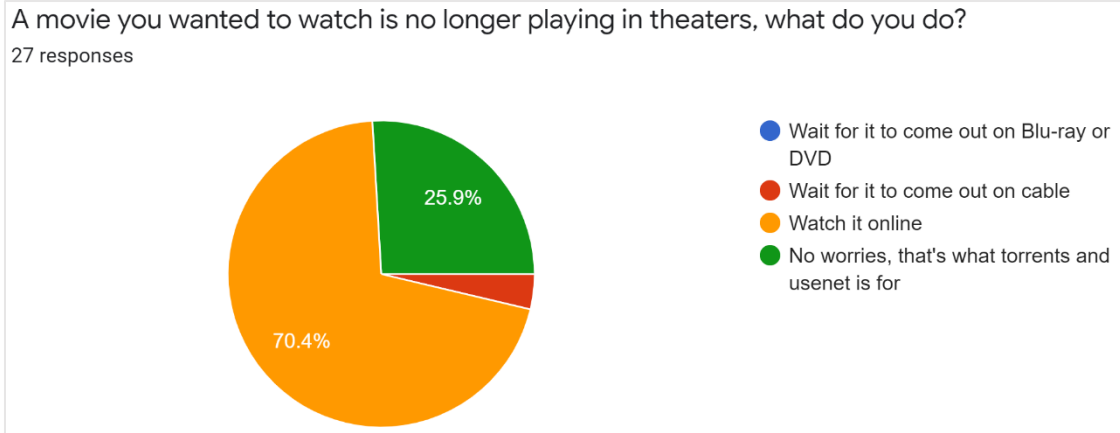


Figura 29. Gráfica en detalle por edades.

Con la intención de conocer un poco a los participantes en relación con su nivel y hábitos con el uso de las tecnologías y ya que este trabajo trata acerca de la generalización de palabras y consultas, se ha intentado respetar su privacidad evitando muchas de las preguntas demográficas comunes como su salario medio, nivel de estudios, etc.





### 6.3. Limitaciones

Hasta aquí los resultados obtenidos por los participantes en las diferentes cuestiones planteadas por cada una de las categorías. Antes de pasar a las conclusiones que han derivado de este modesto estudio, es interesante repasar las limitaciones que ha tenido la extensión durante su prueba.

Se han dividido en dos categorías, las de limitaciones de la solución a nivel técnico, la cual está relacionado con el desarrollo de la aplicación, las tecnologías involucradas y retos encontrados durante su desarrollo. La segunda categoría, limitaciones de la solución a nivel funcional, explica los inconvenientes que tiene un usuario al realizar uso de la extensión y donde la extensión, por su naturaleza, no puede cumplir las expectativas de los usuarios.

#### 6.3.1. Limitaciones de la solución a nivel técnico

Se fijó como objetivo la creación de la extensión de Chrome en el lado usuario, es decir un aplicativo que se instale y se ejecute en su totalidad en el lado usuario, esta técnica tiene sus inconvenientes entre las cuales destaca la imposibilidad de ejecutar código que

requieren de un servicio en ejecución, como puede ser una base de datos, un intérprete de lenguajes de programación o un servidor web, no se puede pretender que los usuarios tengan estos servicios habilitados en sus dispositivos personales. Incluso si los tuvieran activos y en ejecución, sería pretencioso suponer que estos estuvieran configurados en la manera deseada por el creador de la extensión *Chrome*.

La manera en la que se ha solventado la ejecución de código es ayudarse de una infraestructura con servicios en la nube, la cual almacena el código de la extensión que requiere de un servicio en ejecución. En este proyecto se requería acceder al servicio del intérprete del lenguaje de programación *Python*, el cual permite trabajar con el módulo *NLTK*, que a su vez proporciona acceso a la base de datos *WordNet*. Esto también evita pensar en una instalación descentralizada por cada usuario que utiliza la extensión, ya que no habría que instalar el paquete *WordNet* en los equipos usuarios. Este paquete es de alrededor de 400 *Megabytes*, lo cual excede el tamaño medio de las extensiones, los cuales están en la media de la unidad de *kilobytes*.

Esta solución ha sido proporcionada por *Azure Functions Serverless* [19]. Esta solución requiere de una infraestructura mínima lo cual proporciona un ahorro en costes considerables si lo comparamos con soluciones tradicionales.

Otra limitación, viene dada por el mismo uso de *Azure Functions Serverless*, aunque más que una limitación, es más una decisión económica para este trabajo en concreto. Se trata del arranque en frío de las funciones en *Azure (Serverless Cold Start)* [20]. De hecho, entender esta limitación es importante de cara a la demostración del funcionamiento de la extensión. Una función almacenada en *Azure Functions* entra en modo inactivo si no se utiliza en los siguientes 5 minutos después de su último uso, esto es así, ya que *Azure* factura por tiempo de ejecución de las funciones almacenadas, si no se usan, no hay costos, y si no se usan por más de 5 minutos, entran en estado de arranque frío. Entonces, cuando se utiliza una función que está en modo arranque en frío, la primera vez tarda más tiempo en ejecutarse, este tiempo puede ser de hasta unos 10 segundos, las siguientes interacciones con la función a partir de entonces son de apenas pocos milisegundos.

Las soluciones al arranque en frío de las funciones *Azure Function Serverless* encontradas, son dos posibles escenarios. La primera, la más obvia y la más cara es cambiar el plan a un modo de pago 24h activo y así evitar que la función entre en modo arranque frío. La segunda manera es crear otra función con cierto código que se ejecute automáticamente cada 4 min, por ejemplo, siempre menor a los 5 min que provocaría a la primera función entrar en modo arranque frío y esta segunda función que active la primera función, evitando así que la primera función entre en modo arranque frío.

Otra limitante encontrada está relacionada con la forma de instalar la extensión, si se desea una experiencia real, lo normal sería descargar e instalar la extensión desde la *Chrome Web Store*, sin embargo, esta requiere un registro como desarrollador e incurre en un pago. Como alternativa se entrega este trabajo junto a un fichero .zip que contiene todos los ficheros necesarios para que se puedan instalar manualmente y así poder completar las pruebas pertinentes.

### 6.3.2. Limitaciones de la solución a nivel funcional

A pesar de que la extensión hace el trabajo para lo cual ha sido diseñado, se realizan búsquedas de palabras y este realiza la conversión a una palabra más genérica de la introducida y por último la envía al buscador, mostrando así los resultados.

La extensión solo es efectiva cuando se realizan búsquedas de palabras individuales. La base de datos *WordNet* es parecida a un diccionario de términos con características adicionales, pero ninguna de ellas se aplica a frases o palabras compuestas, simplemente no existen en su base de datos, esta no ha sido diseñada para almacenar frases, únicamente palabras. Esta limitación es remarcable, ya que un usuario normal raras veces realiza búsquedas de palabras sueltas, más bien realiza preguntas o introduce frases completas en el buscador.

Otra de las situaciones en las que esta extensión tiene especial dificultad, es cuando el usuario realiza una búsqueda utilizando palabras consideradas argot, estas no están incluidas en un diccionario común y lamentablemente, lo mismo se aplica a la base de datos *WordNet*.

La predicción sobre cuál hiperónimo utilizar es una tarea muy compleja de lograr. Como ocurren en varios idiomas, muchas palabras en el habla inglés tienen diferentes significados y diferentes escenarios donde usarlos, la extensión actualmente carece de un mecanismo que le permita poder o al menos intentar adivinar a cuál de los muchos escenarios debe priorizar para devolver un hiperónimo dentro que mantenga la relación. La configuración realizada en la extensión es la de devolver el hiperónimo que se encuentra en segunda posición, la razón es sencilla, tras realizar distintas pruebas y mediciones, el primer hiperónimo tiende a ser muy genérico y pierda rápidamente precisión.

La precisión de la búsqueda se ve reducida siempre, en algunos casos esto representa algo positivo para la persona realizando la búsqueda, pero en su mayoría de casos tienden a discrepar del nuevo término.

## 7. El control sobre nuestros datos

Si algo se ha ganado con los años, es que cada vez seamos más conscientes de la dinámica y el funcionamiento de la recolección de datos por parte de terceros, tanto así, que se han logrado grandes cambios que permiten ayudarnos a tomar control y mantener la decisión sobre nuestros datos.

### 7.1. El reglamento GDPR

Hace apenas unos años, el 25 de mayo del 2018, se puso en efecto la regulación general de protección de datos en Europa [19]. Este reglamento pone sobre manifiesto la importancia que el usuario debe tener sobre sus datos, trabajando así sobre 6 principios [20]:

**1. Licitud, equidad y transparencia**

Las organizaciones necesitan asegurarse de que sus prácticas de recolección de datos cumplan la ley y que no oculten nada a los interesados.

**2. Limitación de la finalidad**

Las organizaciones deberían recolectar solo información que necesiten para propósitos específicos, los cuales deben estar claramente indicados.

**3. Minimización de datos**

Las organizaciones deben solamente procesar información que necesitan para alcanzar sus propósitos, así minimizar el riesgo de desvelar información que no deberían de disponer en caso de ataques informáticos y facilitar la actualización de datos, cuando estos son mínimos es más fácil mantenerlos al día.

**4. Precisión**

La importancia de tener datos precisos es fundamental para la GDPR. El reglamento establece que se debe tomar todas las medidas necesarias para borrar o rectificar datos que sean inexactos o incompletos.

**5. Límite de almacenamiento**

Las organizaciones deben borrar datos e información que ya no necesiten.

**6. Integridad y confidencialidad**

Este principio especifica que las organizaciones deben procesar los datos en una manera que garantice la seguridad, la protección contra el procesamiento no autorizado o ilegal, así como ante la destrucción o daño accidental.

A pesar de que este trabajo final de grado sea la de investigar una alternativa que ayude al usuario a tener mayor privacidad en Internet, buscando la manera de limitar la cantidad de información que se proporciona a los motores de búsquedas (o al menos de tergiversarlo), con los cuales ellos crean perfiles muy precisos de nuestros hábitos, así como recolectar información que les permiten

fácilmente identificarnos y ser el objetivo de su trabajo de marketing, existen otras herramientas alternativas, que curiosamente son proporcionadas por ellos mismos, los motores de búsquedas.

## 7.2. Google y el control de la información

Tal y como se anunció al inicio de este trabajo, es de acertada elección utilizar al buscador *Google* como referencia, por su dominio en el mercado de las búsquedas y de la publicidad.

Se coloca a *Google* como centro de atención con la intención de desgranar cómo ha ido adaptando su negocio y prácticas respecto a la creación de perfiles de sus usuarios, mientras mantiene en cumplimiento la ley dictada por el reglamento GDPR.

Se inicia esta actividad mostrando la página principal de “Administrar tu cuenta de Google”, en el momento de la realización de esta actividad, mayo del 2021, la dirección es <https://myaccount.google.com/>.

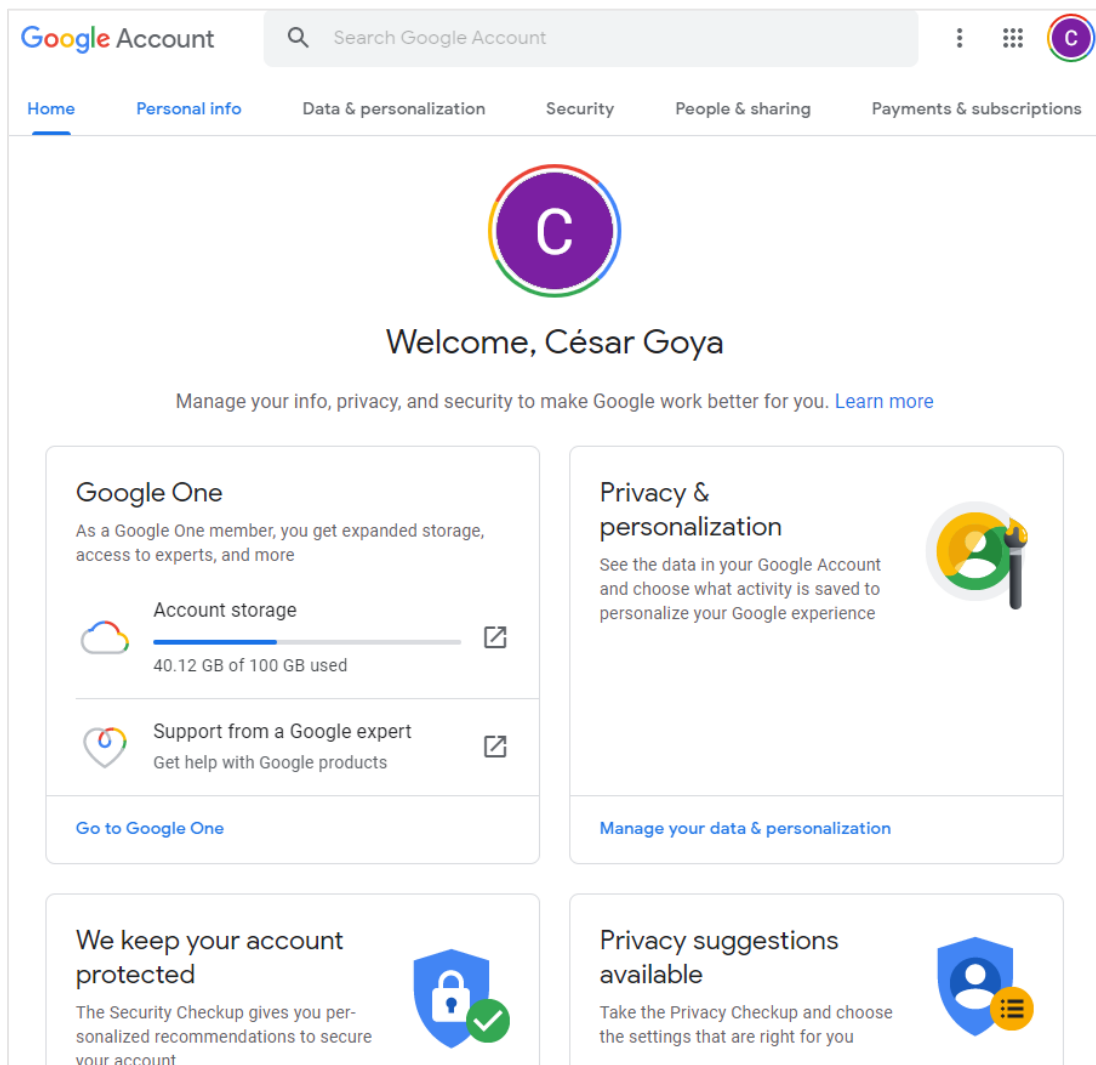


Figura 30. Página principal de administración de tu cuenta en Google.

De entrada, nos llama la atención dos secciones, “privacidad y personalización” y “Sugerencias de privacidad disponibles”. Empezamos por la primera, hacemos clic en “privacidad y personalización” y gratificadamente vemos que *Google* nos vuelve a mostrar la opción de “Seguridad de privacidad disponibles” y como primera opción.

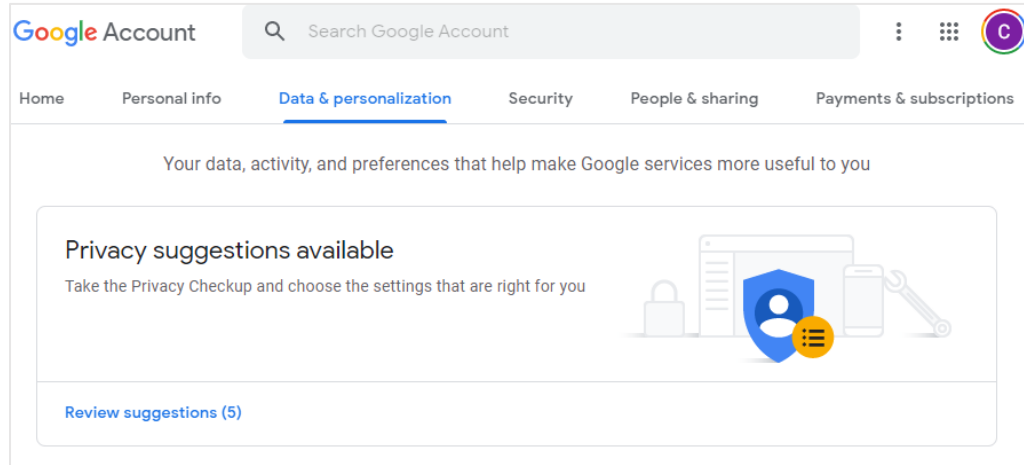


Figura 31. Página de personalización y datos de tu cuenta de Google.

En mi cuenta tengo cinco sugerencias, de las cuales cuatro están relacionadas con la privacidad.

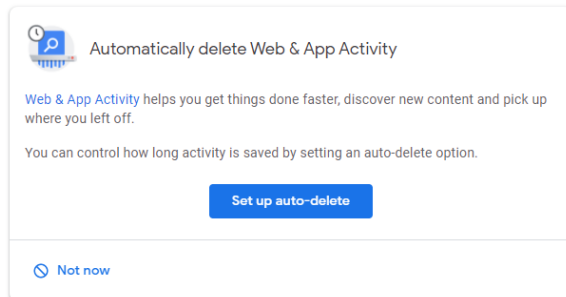


Figura 32. Configurar borrado automático de actividad web y de aplicaciones

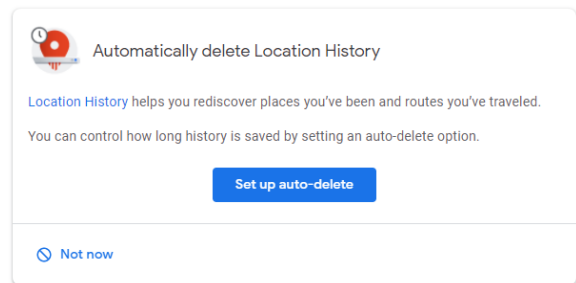


Figura 33. Configurar borrado automático del historial de ubicaciones.

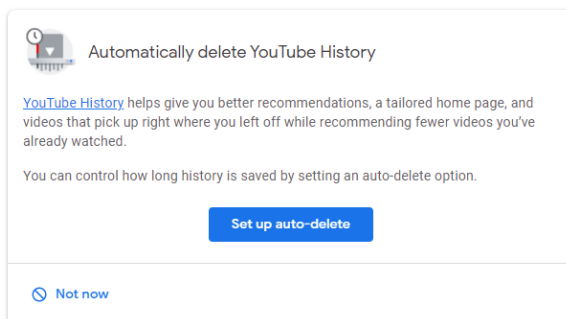


Figura 34. Configurar borrado automático del historial en YouTube.

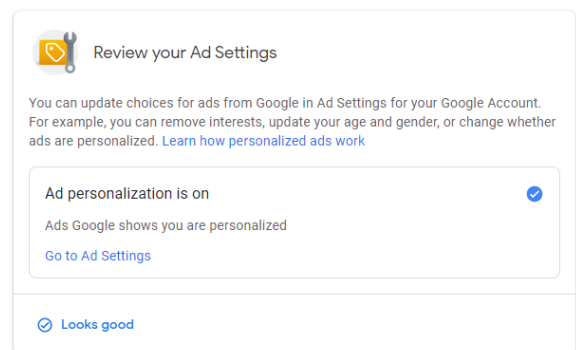


Figura 35. Revisar configuración de publicidad.



Las primeras cuatro opciones se explican por sí solas y *Google* nos da la posibilidad de eliminarlas automáticamente cuando estas son mayores a 3, 18 o 36 meses, tal y como se observan en las figuras 31, 32,33 y 34. *Google* hace su labor de advertir que mantener el historial es “para recibir una experiencia más personalizada”

- “When you choose to keep your Web & App Activity, you can get more personalized experiences, like faster searching and more relevant app and content recommendations.”
- “When you choose to keep your Location History, you can get more personalized experiences, like a private map of where you’ve been and recommendations based on places you’ve visited”
- “When you choose to keep your YouTube History, you can get more personalized experiences, like better content recommendations and the ability to pick up where you left off”

Indagamos más en la cuarta recomendación, figura 34. “Revisar la configuración de publicidad” y nos encontramos con un listado de categorías que han sido asignadas a mi cuenta basado en dos criterios: 1 - proporcionado por mí manualmente al usar/registrarme en sus servicios y 2 - una estimación del interés por dicha categoría basado en mi actividad en los servicios de *Google*.

En mi caso son más de 100 criterios, de los que a modo de ejemplo muestro los primeros.

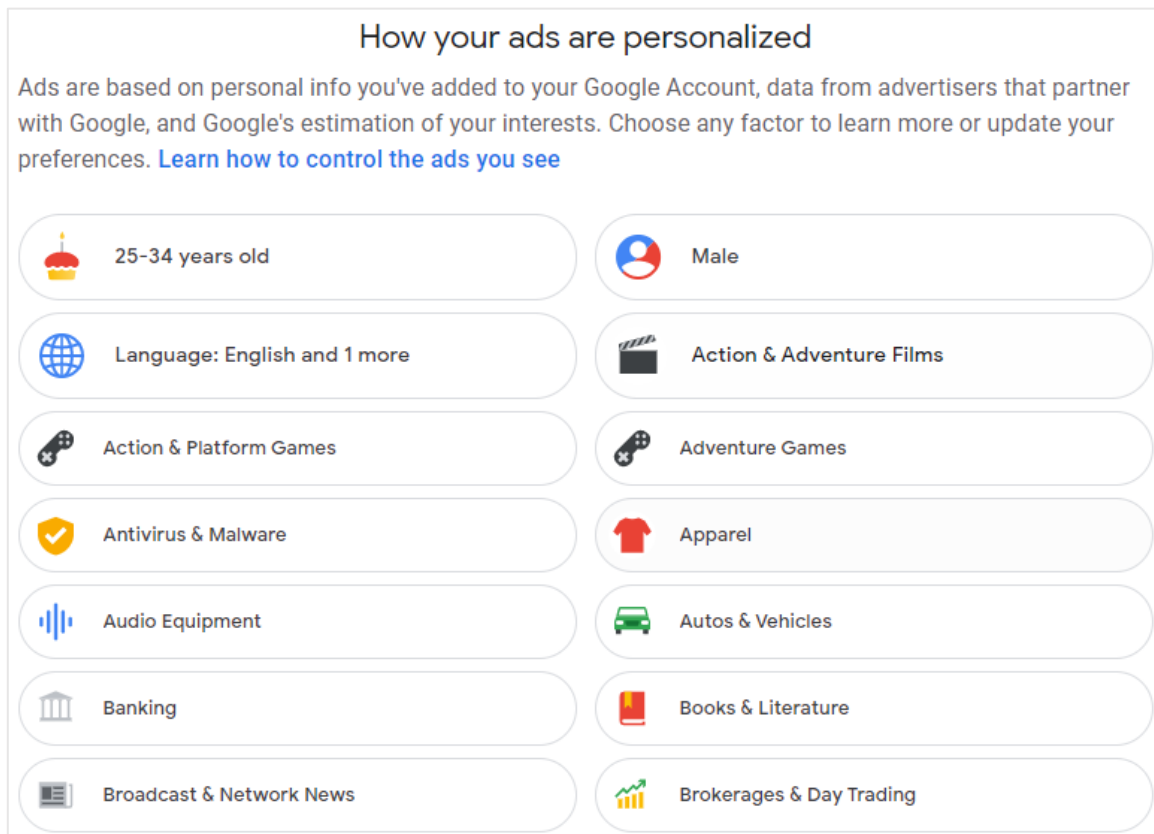


Figura 36. Mis criterios de interés según Google.

Si quisiera desactivar esta personalización automática, solo debería desactivar la opción que para ello Google ha dejado disponible y visible por encima del listado de criterios.

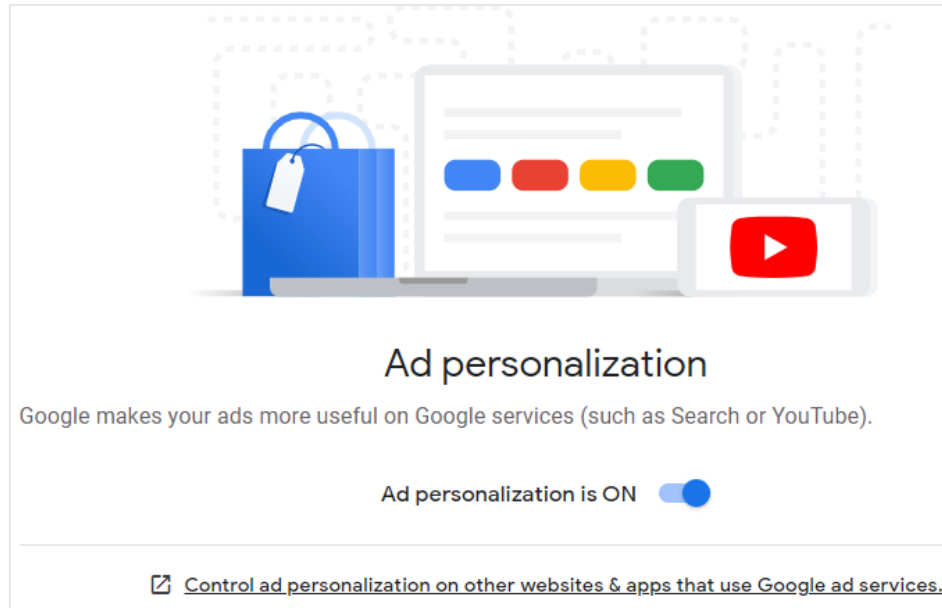


Figura 37. Habilita o deshabilita la opción de recibir publicidad personalizada.

Sin embargo, cabe remarcar que, *Google* te ofrece la posibilidad de deshabilitar la personalización de publicidad en tu cuenta, no el deshabilitar mostrar publicidad en tu cuenta y para ello, lo deja claro en su mensaje de advertencia cuando se intenta desactivar la personalización de publicidad.

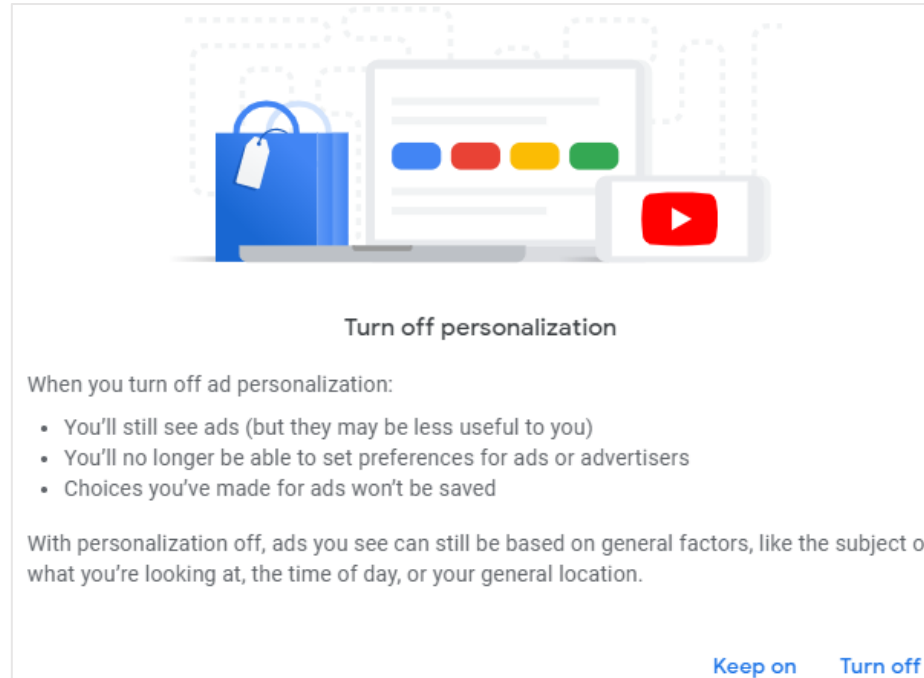


Figura 38. Mensaje de advertencia al intentar desactivar la publicidad personalizada.

Es posible desactivar únicamente los criterios, figura 35, de manera individual. Por lo tanto, existe granularidad al momento de configurar el tipo de publicidad que deseamos ver.

Las categorías por las que *Google* te categoriza tienen restricciones, existen categorías que *Google* mismo las define como inapropiadas, privadas o demasiado complejas como para utilizarlas y obliga a sus anunciantes a respetarlas basada en su política de colección de datos [21].

Dentro de las categorías prohibidas, están las siguientes:

- Alcohol
- Juegos de apuestas
- Reclutamientos para ensayos clínicos
- Medicamentos restringidos

Además, *Google* tiene una política [22] clara contra sus anunciantes que se aprovechen de mostrar publicidad personalizada que afecten o quieran aprovechar:

- Dificultades personales – incluye evitar publicidad relacionada con condiciones de salud, tratamientos, traumas personales, problemas personales.
- Identidad y creencia - incluye evitar publicidad sobre discriminación y acoso, pertenencia a grupos, discriminación o prejuicios.
- Intereses sexuales – incluye evitar publicidad que tenga como objetivo usuarios con un interés sexual particular, categorías de sugerencia sexual o que tengan intención de excitar sexualmente.
- Acceso a oportunidades – incluye evitar publicidad que fomente la desigualdad, prejuicios en la sociedad,

## 8. Conclusiones

Internet, tal y como lo conocemos hoy en día, tendrá poco que ver con el Internet de aquí a unos años. La tecnología da cambios a pasos agigantados. Actualmente realizamos millones de consultas en los buscadores, la recolección de datos no tiende a detenerse y ahora con la tecnología *Big Data* será de magnitudes inimaginables lo que nos depara.

En relación con la extensión que se desarrolló en este trabajo final de grado, proporciona un nivel de privacidad notable, es cierto que generalizar palabras ayuda a evitar que los buscadores tengan información precisa de un usuario, la generalización de palabras es sin duda una opción, sin embargo, es difícil considerar una solución factible debido a las limitantes encontradas en ella. La aceptación de una palabra más genérica como palabra alternativa a utilizar en los buscadores, proporciona inconvenientes previsibles de precisión, en los resultados se ha visto corroborado unos porcentajes de aceptación superiores a los rechazos, sin embargo, a título personal, sería deseable que la aceptación esté por encima del 90% para que la extensión se contemple útil. En este trabajo se observó que, cada 10 palabras buscadas por algún usuario usando la extensión, 5 veces aceptaría la búsqueda y 5 las rechazaría, esto prácticamente se traduce a omitir la extensión y volver a la forma rutinaria de utilizar el buscador.

Algo importante que se ha demostrado es que los valores aceptados tienden a variar mucho dependiendo del tipo de categoría. Con el uso de palabras en categorías como privacidad, apenas se obtenía un 39% de aprobación, sin embargo, en la categoría problemas o enfermedades se ha visto una aprobación superior, en el caso de la categoría enfermedades ha sido casi del 66%. Obviamente existen multitudes de categorías que han quedado fuera de la actividad presentada en este trabajo y muchas más palabras que a su vez pueden pertenecer a muchas otras categorías, pero quizás exista la remota posibilidad que la extensión si sea útil para búsquedas más segmentadas, es decir, la extensión podría no ser útil para un público general, pero lo podrían ser para algunos grupos reducidos.

Alternativamente, tomar el control de nuestros datos puede tener un mayor impacto, cierto es que hace unos años atrás, sería una labor muy compleja de lograr, pero la transparencia de uso de nuestros datos es cada vez mayor, con nuevos reglamentos que apoyan al usuario final y presentan una serie de requisitos a cumplir por parte de los tomadores de datos. Un importante cambio en los requisitos que han de cumplir es la facilidad con la que deben mostrar las opciones que tienen los usuarios sobre su información recolectada y la facilidad de desactivar historiales de uso de servicios, revocar permisos de uso, modificar su información almacenada o incluso darlos de baja eliminándolos indefinidamente.

Obviamente queda mucho camino por delante, pero reglamentos a favor de la protección de datos, están en auge y presionan a empresas tomadoras de datos que desarrollen y proporcionen herramientas fiables con las que devolvemos el control sobre nuestros datos, un aspecto esencial que se debería difundir más y así fomentar la educación digital.

## 9. GLOSARIO

- **Algoritmos de indexación**  
Es un procedimiento para la construcción de una estructura de datos o un índice diseñado para acelerar las búsquedas de una base de datos.
- **Azure**  
Es un servicio de computación en la nube creado por la empresa *Microsoft* para la creación, testeo, desarrollo y manejo de aplicaciones y servicios a través de los centros de datos administrados de *Microsoft* [23].
- **Azure Function Serverless Compute**  
Es un servicio proporcionado por *Azure* que ejecuta código de programación bajo demanda sin la necesidad de hospedar el código en un servidor ni administrar la infraestructura necesaria para ello.
- **Big Data**  
Es una colección de enormes cantidades de datos que crece en volumen exponencialmente con el tiempo. Se trata de datos con un tamaño y complejidad tan grande que ninguna de las herramientas tradicionales de administración de datos puede almacenarlo o procesarlo.
- **Cookies**  
Una cookie HTTP es un fragmento pequeño de datos que un navegador web almacena en el ordenador del usuario que navega por un sitio web. Las cookies fueron diseñadas para ser un mecanismo confiable para que los sitios web recuerden información o registren la actividad de navegación del usuario [24].
- **Corpus**  
En lingüística, un corpus es un recurso lingüístico que consta de un conjunto de textos amplios y estructurados. Se utiliza para realizar análisis estadísticos y prueba de hipótesis, verificando ocurrencias o validando reglas lingüísticas dentro de un territorio lingüístico específico.
- **Formato JSON**  
Es un formato de archivo estándar y abierto de intercambio de datos que utiliza texto legible por humanos para almacenar y transmitir datos que consisten en matrices y pares de atributos-valor. Es un formato de datos muy común, con una amplia gama de aplicaciones, un ejemplo son las aplicaciones web que se comunican con un servidor [25].
- **Hiperónimo**  
Definida por la RAE como “Palabra cuyo significado está incluido en el de otras” [9], aunque la definición es bastante acertada, podemos ayudarnos también de su origen etimológico, donde el vocablo hiper proviene del griego y se puede traducir como “sobre” o “por encima de”, mientras que el vocablo ónimo, también procedente del griego significa “nombre”. Es

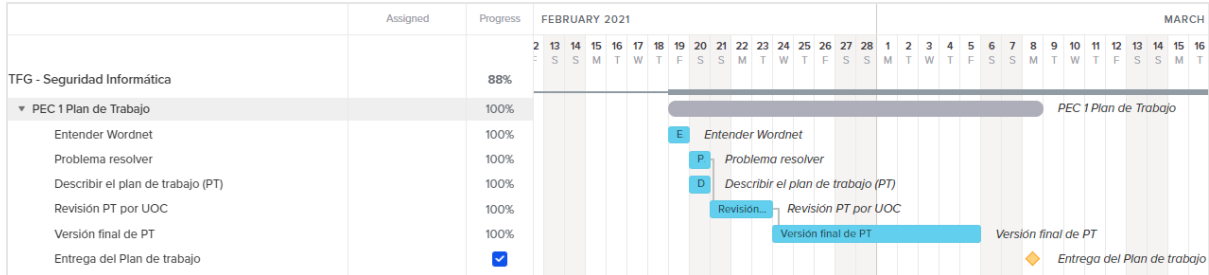
decir, el nombre que está por encima. Así entonces “mamífero” es hiperónimo de “perro” [10].

- **Hipertexto**  
Es un texto que se muestra en la pantalla de un ordenador u otros dispositivos electrónicos, este tipo de texto tiene la característica de estar interconectados por hipervínculos, que generalmente se activan con un clic de ratón, una pulsación de tecla o una pantalla táctil.
- **Hiponimia**  
Relación existente entre un hipónimo con otra palabra en curso significado se encuentra englobado el hipónimo.
- **Interfaz**  
Es una capa aplicativa compartida entre dos o más componentes informáticos que permiten el intercambio de información. Estos componentes pueden ser un programa informático, la parte física del ordenador, periféricos, lenguajes de programación o una combinación de ellos.
- **Omnibox**  
Es la barra de direcciones de un navegador web combinado con la opción de realizar búsquedas, el primero en implementarlo fue Google Chrome, sin embargo, hoy todos los navegadores populares han seguido el mismo camino.
- **Sandbox**  
Es un entorno de pruebas que aísla los cambios que puede producir el programa en desarrollo del sistema en producción, asegurando así un entorno seguro de trabajo [26].

## 10. Anexos

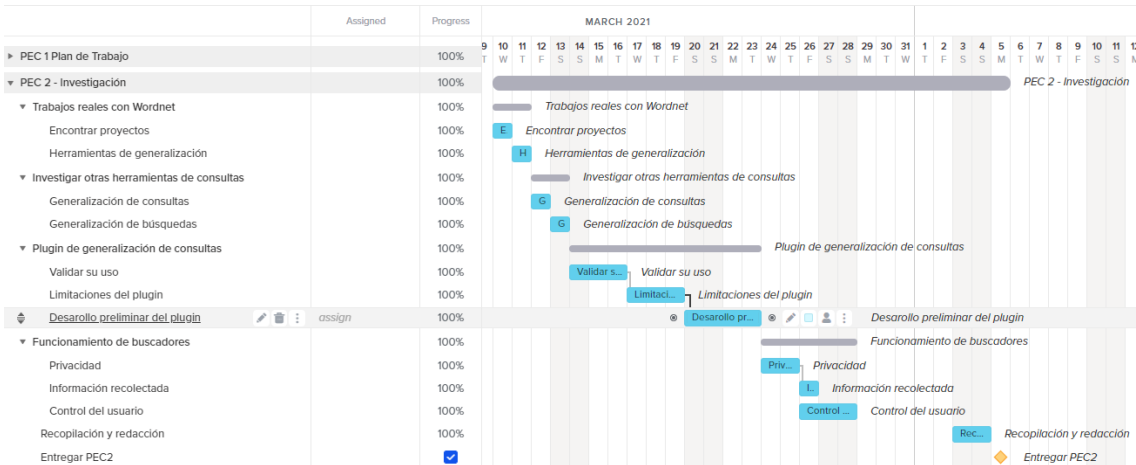
### Anexo 1: PEC 1 Plan de trabajo

Planificación con línea de tiempo del trabajo final de grado para el bloque PEC1 de entrega mediante el uso de diagramas de Gantt.



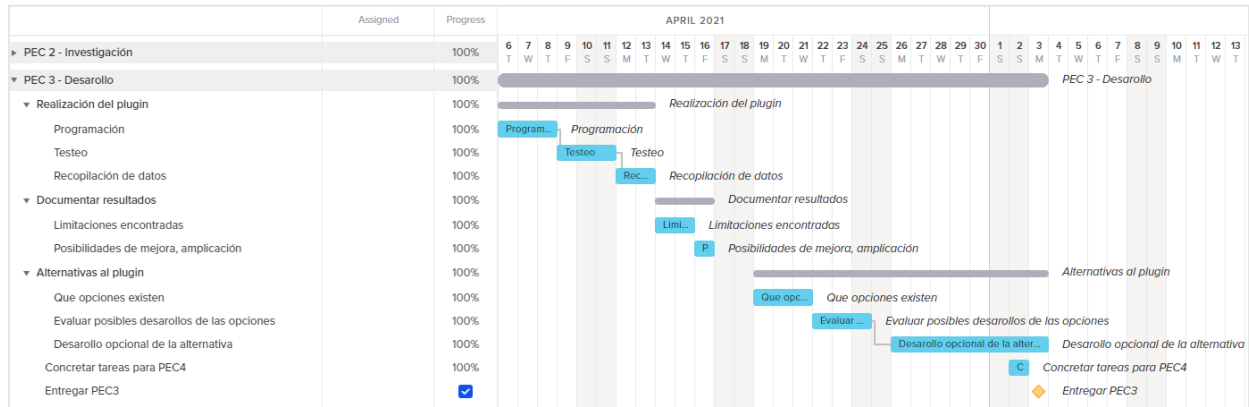
### Anexo 2: PEC 2 Investigación

Planificación con línea de tiempo del trabajo final de grado para el bloque PEC2 de entrega mediante el uso de diagramas de Gantt.



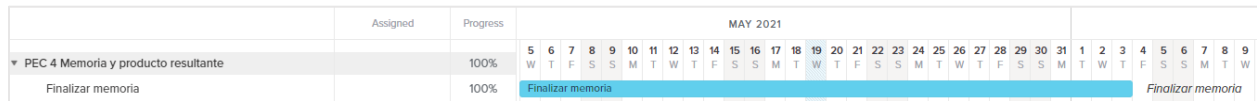
### Anexo 3: PEC 3 Desarrollo

Planificación con línea de tiempo del trabajo final de grado para el bloque PEC de entrega mediante el uso de diagramas de Gantt.



### Anexo 4: PEC 4 Memoria y producto resultante

Planificación con línea de tiempo del trabajo final de grado para el bloque PEC1 de entrega mediante el uso de diagramas de Gantt.



### Anexo 5: Código fuente resultante del fichero manifest.json

```

{
  "manifest_version": 2,
  "name": "Generalized Search",
  "version": "1.0",
  "description": "Extension to generalize each user's search by using a lexical database named WordNet",
  "author": "Cesar Goya",

  "browser_action": { "default_popup": "index.html", "default_icon": "img/icon.png" },

  "icons": { "16": "img/icon16.png", "32": "img/icon32.png", "64": "img/icon32.png" },

  "background": { "scripts": ["scripts/gs.js", "scripts/jquery.min.js"], "persistent": false },

  "omnibox": { "keyword": "gs" },

  "permissions": [ "tabs", "https://uocwordnet.azurewebsites.net/" ]
}

```



## Anexo 6: Código fuente resultante del fichero index.html

```
<!doctype html>
<html>
  <head>
    <title>Generalize Searches</title>
    <link rel="stylesheet" type="text/css" href="style/style.css" media="all" />
  </head>
  <body>
    <div>
      <h1>How to Generalize your search</h1>
      <ol>
        <li>Enter "gs" and press SPACE</li>
        <li>Enter your search</li>
        <li>Hit ENTER</li>
      </ol>
      
    </div>
  </body>
  <footer>Chrome Extension created by Cesar Goya as outcome of Final Degree Project.</footer>
</html>
```

## Anexo 7: Código fuente resultante del fichero style.css

```
body{
  background-color:#f1f1f1;
  width:500px;
  height:220px;
  margin-left: 25;
  margin-right: auto;
  text-align: center;
  font-family:Arial, Helvetica, sans-serif;
}
h1 { color: rgb(182, 100, 0);}
ol {text-align: left;}
img {margin-bottom: 50px;}
```

## Anexo 8: Código fuente resultante del gs.js

```
// Define google URL search variable
search_url = "https://www.google.com/search?q="

// Get text entered in omnibox
chrome.omnibox.onInputEntered.addListener(
  function(text) {
    // Send text to search in Python script
    let scrapeJSON = 'https://uocwordnet.azurewebsites.net/api/wordnethypernym?s='+text

    $.get(scrapeJSON, function(data) {
      // Get returned value from python script
      if (text){
        // Send value to search engine
        chrome.tabs.update({"url" : search_url + data});
      }
    })
  }
);
```

## 11. Bibliografía

- [1] T. Craig, "Net of Insecurity, A flaw in the design," The Washington Post, 30 06 2015. [Online]. Available: <https://www.washingtonpost.com/sf/business/2015/05/30/net-of-insecurity-part-1/>. [Accessed 24 03 2021].
- [2] E. Schechter, «Moving towards a more secure web,» Google Blog, 08 09 2016. [En línea]. Available: <https://security.googleblog.com/2016/09/moving-towards-more-secure-web.html>. [Último acceso: 04 03 2021].
- [3] DuckDuckGo, «DuckDuckGo Privacy,» [En línea]. Available: <https://duckduckgo.com/privacy>. [Último acceso: 23 04 2021].
- [4] Google.com, « Google Cloud y el Reglamento general de protección de datos,» [En línea]. Available: <https://cloud.google.com/security/gdpr/>. [Último acceso: 27 04 2021].
- [5] J. Parra-Arnau, J. Prasad Achara y C. Castelluccia, «MyAdChoices: Bringing Transparency and Control to Online Advertising,» *ACM Transactions on the Web*, vol. 11, nº 1, March 2017.
- [6] J. Parra-Arnau, A. Perego, E. Ferrari, J. Forné y D. Rebollo-Monedero, «Privacy-Preserving Enhanced Collaborative Tagging,» *IEEE Trans. Knowl. Data Eng.*, vol. 26, nº 1, pp. 180-193, Jan 2014.
- [7] J. Parra-Arnau, D. Rebollo-Monedero y J. Forné, «Measuring the Privacy of User Profiles in Personalized Information Systems,» *Elsevier Future Gen. Comput. Syst. (FGCS), Special Issue Data Knowl. Eng.*, vol. 33, pp. 53-63, Apr 2014.
- [8] D. Rebollo-Monedero, J. Parra-Arnau, C. Diaz y J. Forné, «On the Measurement of Privacy as an Attacker's Estimation Error,» *Int. J. Inform. Security*, vol. 12, nº 2, pp. 129-149, Apr 2013.
- [9] J. Parra-Arnau, D. Rebollo-Monedero y J. Forné, «Optimal Forgery and Suppression of Ratings for Privacy Enhancement in Recommendation Systems,» (*MDPI Entropy*), vol. 16, nº 3, pp. 1586-1634, Mar 2014.
- [10] Princeton University, «WordNet A lexical Database for English,» [En línea]. Available: <https://wordnet.princeton.edu/>. [Último acceso: 29 03 2021].
- [11] J. Nemrava, «Using WordNet Glosses to Refine Google,» *Department of Information and Knowledge Engineering*, Jan 2006.
- [12] S. Cobo Gómez, «Generalización de consultas web para protección de privacidad de usuarios,» Universitat Oberta de Catalunya (UOC), Jan 2, 2020.
- [13] J. M. Puga Moreira, «Sistema para proteger la privacidad de los usuarios de los motores de búsqueda de Internet,» Fundacio per la Universitat Oberta de Catalunya, 2011.

- [14] «Searchononymous - Chrome Extension,» [En línea]. Available: <https://chrome.google.com/webstore/detail/searchononymous/onhfdppooafpnnnigbmnpnnjmbajggek/related>. [Último acceso: 10 04 2021].
- [15] Google, «What are extensions?,» Chrome Developers, 21 02 2013. [En línea]. Available: <https://developer.chrome.com/docs/extensions/mv3/overview/>. [Último acceso: 25 04 2021].
- [16] Google, «Chrome Web Store,» [En línea]. Available: <https://chrome.google.com/webstore/category/extensions>. [Último acceso: 25 04 2021].
- [17] Google, «Manifest Version,» Chrome Developers, 17 09 2012. [En línea]. Available: <https://developer.chrome.com/docs/apps/manifestVersion/>. [Último acceso: 25 04 2021].
- [18] NLTK, «Natural Language Toolkit,» [En línea]. Available: <https://www.nltk.org/>. [Último acceso: 02 05 2021].
- [19] GDPR.Eu, «What is GDPR, the EU's new data protection law?,» [En línea]. Available: <https://gdpr.eu/what-is-gdpr/#:~:text=The%20regulation%20was%20put%20into,tens%20of%20millions%20of%20euros..> [Último acceso: 22 05 2021].
- [20] L. Irwin, «The GDPR: Understanding the 6 data protection principles,» IT Governance, 30 07 2020. [En línea]. Available: <https://www.itgovernance.eu/blog/en/the-gdpr-understanding-the-6-data-protection-principles>.
- [21] Google, «Data collection and use,» [En línea]. Available: <https://support.google.com/adspolicy/answer/6020956>. [Último acceso: 22 05 2021].
- [22] Google, «Personalized advertising,» [En línea]. Available: <https://support.google.com/adspolicy/answer/143465?hl=en>. [Último acceso: 22 05 2021].
- [23] Wikipedia, «Microsoft Azure,» [En línea]. Available: [https://en.wikipedia.org/wiki/Microsoft\\_Azure](https://en.wikipedia.org/wiki/Microsoft_Azure). [Último acceso: 20 05 2021].
- [24] Wikipedia, «HTTP cookie,» [En línea]. Available: [https://en.wikipedia.org/wiki/HTTP\\_cookie](https://en.wikipedia.org/wiki/HTTP_cookie). [Último acceso: 20 05 2021].
- [25] Wikipedia, «JSON,» [En línea]. Available: <https://en.wikipedia.org/wiki/JSON>. [Último acceso: 20 05 2021].
- [26] Wikipedia, «Sandbox (software development),» Wikipedia, [En línea]. Available: [https://en.wikipedia.org/wiki/Sandbox\\_\(software\\_development\)](https://en.wikipedia.org/wiki/Sandbox_(software_development)). [Último acceso: 21 05 2021].
- [27] NLTK - WordNet Interface, «WordNet Interface,» [En línea]. Available: <https://www.nltk.org/howto/wordnet.html>. [Último acceso: 2 05 2021].
- [28] Microsoft Build, «Introducción a Azure Functions,» 20 11 2020. [En línea]. Available: <https://docs.microsoft.com/es-es/azure/azure-functions/functions-overview>. [Último acceso: 03 05 2021].
- [29] Microsoft Azure, «Understanding serverless cold start,» 7 02 2018. [En línea]. Available: <https://azure.microsoft.com/en-us/blog/understanding-serverless-cold-start/>. [Último acceso: 03 05 2021].

