
Mejora de la Comprensión Lectora mediante Analogías para la Inclusión



Trabajo de Fin de Grado
Curso 2018–2019

Autor

Irene Martín Berlanga
Pablo García Hernández

Director

Virginia Francisco Gilmartín
Gonzalo Rubén Méndez Pozo

Grado en Ingeniería de Software

Facultad de Informática

Universidad Complutense de Madrid

Mejora de la Comprensión Lectora mediante Analogías para la Inclusión

Trabajo de Fin de Grado en Ingeniería de Software
Departamento de Ingeniería de Software e Inteligencia
Artificial

Autor
Irene Martín Berlanga
Pablo García Hernández

Director
Virginia Francisco Gilmartín
Gonzalo Rubén Mendez Pozo

Dirigida por el Doctor
Virginia Francisco Gilmartín
Gonzalo Rubén Mendez Pozo

Grado en Ingeniería de Software
Facultad de Informática
Universidad Complutense de Madrid

5 de mayo de 2019

Autorización de difusión

Los abajo firmantes, matriculados en el Grado de Ingeniería de Software de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores el presente Trabajo de Fin de Grado: "Mejora de la Comprensión Lectora mediante Analogías para la Inclusión", realizado durante el curso académico 2018-2019 bajo la dirección de Virginia Francisco Gilmartín y Gonzalo Rubén Mendez Pozo en el Departamento de Ingeniería de Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Nombre Del Alumno

5 de mayo de 2019

Dedicatoria

Texto de la dedicatoria...

Agradecimientos

Texto de los agradecimientos

Resumen

Resumen en español del trabajo

Palabras clave

Máximo 10 palabras clave separadas por comas

Abstract

Abstract in English.

Keywords

10 keywords max., separated by commas.

Índice

1. Introduction	1
1. Introducción	3
1.1. Motivación	3
1.2. Objetivos	5
1.3. Estructura de la memoria	6
2. Estado de la Cuestión	7
2.1. Lectura Fácil	8
2.2. Figuras retóricas	10
2.3. Servicios Web	11
2.3.1. Tipos de Servicios Web	11
2.3.2. Arquitectura Servicios Web	13
2.3.3. Ventajas de los Servicios Web	13
2.3.4. Desventajas de los Servicios Web	14
2.4. Procesamiento del Lenguaje Natural	14
2.4.1. ConceptNet	17
2.4.2. Thesaurus	20
2.4.3. Thesaurus Rex	22
2.4.4. Metaphor Magnet	23
2.4.5. WordNet	25
3. Herramientas Utilizadas	29
3.1. Django	29
3.2. SpaCy	30
4. Metodología de gestión del Proyecto	33
5. Servicios Web para la obtención de conceptos fáciles relacionados	35

5.1.	Diseño de la evaluación	35
5.2.	Resultados cuantitativos	36
5.3.	Servidor de Base de Datos	36
5.4.	Servicio Web para obtener sinónimos fáciles	36
5.5.	Servicio Web para obtener hipónimos fáciles	37
5.6.	Servicio Web para obtener hiperónimos fáciles	38
5.7.	Servicio Web para obtener una metáfora	39
5.8.	Servicio Web para obtener un símil	40
5.9.	Servicio Web Para ConceptNet	40
5.10.	Servicio Web para WordNet	42
5.11.	Pruebas Estadísticas	42
5.11.1.	Estadística para ConceptNet	43
5.11.2.	Estadística para WordNet	44
5.11.3.	Conclusiones	44
6.	Aprende Fácil	57
6.1.	Diseño de la Aplicación	57
6.1.1.	Primera Iteración: Iteración Competitiva	58
6.1.2.	Segunda Iteración: Evaluación con Expertos	61
7.	Trabajo Realizado	75
7.1.	Trabajo realizado por Irene	75
7.2.	Trabajo realizado por Pablo	76
8.	Conclusiones y Trabajo Futuro	77
8.	Conclusions and Future Work	79
A.	Título	81
B.	Título	83
Bibliografía		85

Índice de figuras

2.1. Logo Lectura Fácil	8
2.2. Ejemplo de Red Semántica	15
2.3. Ejemplo de Red de Marco	16
2.4. Ejemplo de Red IS-A	16
2.5. Ejemplo de Grafo Conceptual	17
2.6. Resultados de ConcepNet para la palabra <i>chaqueta</i>	18
2.7. Resultados de ConcepNet para la palabra polisémica <i>book</i> . .	21
2.8. Resultados búsqueda Thesaurus Rex con la palabra <i>house</i> . .	24
2.9. Resultados búsqueda Metaphor Magnet con la palabra <i>house</i> . .	26
2.10. Resultados de la búsqueda en EuroWordNet para la palabra <i>casa</i>	27
3.1. Ejemplo de clasificación de palabras	31
4.1. Tablero de tareas al inicio del proyecto	34
5.1. JSON devuelto al buscar los sinónimos fáciles de inmueble .	38
5.2. JSON devuelto al buscar los hipónimos fáciles de inmueble .	46
5.3. JSON devuelto al buscar los hiperónimos fáciles de inmueble .	47
5.4. JSON devuelto al buscar las metáforas de la palabra inmueble	48
5.5. JSON devuelto al buscar los símiles de la palabra inmueble .	49
5.6. Interfaz del Servicio Web para ConceptNet	50
5.7. Resultados en formato JSON para el Servicio Web de WordNet	50
5.8. Campo de texto para introducir una palabra	50
5.9. Resultados obtenidos para el Servicio Web de WordNet	51
5.10. Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 1000 palabras	51
5.11. Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 5000 palabras	51

5.12. Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 10000 palabras	52
5.13. Resultados obtenidos para el Servicio Web de WordNet con la lista de 1.000 palabras	52
5.14. Resultados obtenidos para el Servicio Web de WordNet con la lista de 5.000 palabras	52
5.15. Resultados obtenidos para el Servicio Web de WordNet con la lista de 10.000 palabras	53
5.16. Gráfica de los resultados obtenidos para las 1000 palabras con ConcepNet	53
5.17. Gráfica de los resultados obtenidos para las 1000 palabras con WordNet	54
5.18. Gráfica de los resultados obtenidos para las 5000 palabras con ConcepNet	54
5.19. Gráfica de los resultados obtenidos para las 5000 palabras con WordNet	55
5.20. Gráfica de los resultados obtenidos para las 10000 palabras con ConcepNet	55
5.21. Gráfica de los resultados obtenidos para las 10000 palabras con WordNet	56
6.1. Prototipo de Pablo mostrando resultado más común para la palabra vehículo	62
6.2. Prototipo de Pablo mostrando resultado más común para la palabra vehículo y un ejemplo	63
6.3. Prototipo de Pablo mostrando todos los resultados para la palabra vehículo	64
6.4. Prototipo de Pablo mostrando todos los resultados para la palabra vehículo junto con pictogramas	64
6.5. Prototipo de Irene mostrando resultado más común para la palabra vehículo	65
6.6. Prototipo de Irene mostrando resultado más común para la palabra vehículo y una definición	66
6.7. Prototipo de Irene mostrando todos los resultados para la palabra portero	67
6.8. Prototipo de Irene mostrando todos los resultados para la palabra portero junto con pictogramas	68
6.9. Prototipo mostrando resultado más común para la palabra portero	69
6.10. Prototipo Versión 1 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo separados	70

6.11. Prototipo Versión 2 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo a simple vista	71
6.12. Prototipo Versión 3 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo en un mismo botón	72
6.13. Prototipo mostrando todos los resultados para la palabra portero	73
6.14. Prototipo mostrando todos los resultados para la palabra portero incluyendo los pictogramas	74

Índice de tablas

Índice de Listados

2.1.	Estructura de un mensaje SOAP	12
2.2.	JSON devuelto por la API de ConceptNet para la palabra chaqueta	19
2.3.	XML devuelto por Thesaurus para la palabra <i>peace</i>	21
2.4.	JSON devuelto por Thesaurus para la palabra <i>peace</i>	21
2.5.	XML devuelto por Thesaurus Rex para la palabra <i>house</i> . . .	23
2.6.	XML devuelto por Metaphor Magnet para la palabra <i>house</i> .	25

Chapter 1

Introduction

Introduction to the subject area.

Capítulo 1

Introducción

Actualmente, si pensamos en el día a día de cualquier persona en nuestra sociedad, creemos que no tiene ningún inconveniente para realizar cualquier acción. Es más, podemos pensar que hoy en día existen grandes avances que hacen que la vida resulte más fácil. Pero esto no es del todo cierto, ya que existen ciertos colectivos como pueden ser inmigrantes, personas con alguna discapacidad cognitiva, ancianos, analfabetos funcionales, etc... que para poder realizar cualquier acción cotidiana les supone un gran esfuerzo o incluso no pueden realizarla y estos avances no están pensados para ellos.

En especial, este trabajo se va a centrar en personas con discapacidad cognitiva y en la sección 1.1 se explicará con mayor detalle los problemas a los que se tienen que enfrentar, que solución queremos aportar y en qué ayudaría dicha solución. Para poder ofrecer la mejor solución posible se necesita desgranar el problema e ir formando dicha solución en función de sus necesidades y limitaciones cumpliendo pequeños objetivos como puede ser la distinción entre palabras fáciles y palabras difíciles, o como obtener los términos relacionados de un concepto específico. En la sección 1.2 se explicarán cuáles han sido estos objetivos así como la forma en la que se llevarán a cabo.

Por último, en la sección 1.3 vendrá explicada la estructura de dicho documento explicando en cada capítulo lo que se contará de una manera breve.

1.1. Motivación

El español, hoy en día es la segunda lengua más hablada del mundo y actualmente más de 90000 palabras forman el castellano. Se trata de una lengua con multitud de términos, y que dependiendo del contexto en el que se encuentren, pueden tener múltiples significados. Por ejemplo, la palabra

gato puede hacer referencia a un gato de animal o un gato como herramienta para elevar un coche. Si esto puede suponer una complicación para cualquier persona, para ciertos colectivos de la sociedad afectados por algún trastorno cognitivo lo es aún mucho más, afectándoles en su vida cotidiana, profesional o personal. Por ejemplo, el simple hecho de leer un periódico es una acción bastante difícil para ellos ya que muchas palabras no saben lo que significan. Otro ejemplo podría ser leer un manual de instrucciones de una función más técnica, donde se encuentran en la misma situación de no poder entender ciertos conceptos.

Una de las soluciones que se podrían pensar en un primer momento, es buscar su significado en un diccionario. Pero esto no les sirve, puesto que las definiciones que aparecen en muchos casos no utilizan términos o frases sencillas. Por ejemplo, la palabra computadora tiene la siguiente definición en el diccionario: *Máquina electrónica que, mediante determinados programas, permite almacenar y tratar información, y resolver problemas de diversa índole*¹. Esta definición puede ser bastante complicada de entender por una persona con discapacidad cognitiva ya que utiliza términos más técnicos y es una frase bastante larga, por lo que una solución que se ha pensado para poder solventar dicho problema, es ofrecer un resultado más sencillo basado en figuras retóricas.

Dentro de las figuras retóricas, se pueden encontrar multitud de tipos pero para este trabajo se utilizarán únicamente la metáfora, la analogía y el símil. Por ejemplo, para el mismo concepto (computadora) se podría obtener los siguientes resultados:

- Haciendo uso de las metáforas: *Una computadora es un ordenador.*
- Haciendo uso de las analogías: *Una computadora es como una máquina.*
- Haciendo uso de los símiles: *Una computadora es fuerte como una piedra.*

Gracias al uso de las figuras retóricas, se puede obtener conceptos mucho más sencillos y que facilitan el entendimiento del concepto al usuario. En estas definiciones se utilizan frases cortas y conceptos sencillos de entender. Si a esto le sumamos que se añada un pictograma para la palabra ordenador, máquina y piedra así como al propio concepto de computadora, ayudaría todavía más al fácil entendimiento del concepto. Por ello, nosotros crearemos una aplicación que dada una palabra compleja devuelva los resultados mediante comparaciones con conceptos más sencillos y haciendo uso de las figuras retóricas. Esta aplicación, ayudaría a que cualquier palabra que les resulte complicada de entender, puedan buscarla en el momento y así ayudarles en todos los ámbitos de su vida cotidiana. Así, a la hora de poder leer un

¹<https://dle.rae.es/?id=A4hIGQC>

periódico, un libro o un manual técnico no les resultaría un impedimento para ellos.

1.2. Objetivos

El objetivo principal de este Trabajo de Fin de Grado es crear una aplicación web basada en servicios que dada una palabra compleja para el usuario devuelva una definición de dicha palabra mediante símiles, analogías o metáforas que empleen palabras más sencillas y conocidas para el usuario. Para poder obtener esta definición, habrá que estudiar como obtener los términos relacionados de un concepto, así como distinguir de estos resultados cuales son palabras fáciles y difíciles. También habrá que saber que tipos de figuras retóricas existen y cuál utilizar según la relación entre el concepto inicial y los conceptos fáciles, de esta forma se podrá facilitar al usuario final un resultado claro y correcto.

La aplicación estará construida con servicios web que doten de funcionalidad a la aplicación y que sean reutilizables en otras aplicaciones, haciendo así que se puedan adaptar a las distintas necesidades de los usuarios finales. Los servicios web desarrollados estarán disponibles en una API pública para que todo el mundo pueda utilizarlos y puedan servir para que otros desarrolladores integren nuestros servicios en sus aplicaciones.

La aplicación se construirá de manera incremental, añadiendo valor al producto poco a poco. De este modo se podrá testear las distintas hipótesis de trabajo poco a poco y realizar modificaciones de una manera simple para así conseguir una aplicación que se adecúe a las necesidades de los usuarios.

Una vez que se disponga de lo comentado anteriormente, se deberá crear una interfaz. Siempre es necesario que el diseño de la interfaz esté centrada en el usuario, y para ello se debe obtener la mayor información posible sobre el usuario final. De esta forma se realiza un diseño basado en sus necesidades y se obtiene una mayor satisfacción de este al utilizar la aplicación, reduce el tiempo de desarrollo, etc... Como dicho trabajo está enfocado para personas con discapacidad cognitiva, se debe realizar un diseño que se adapte aún más a sus necesidades y limitaciones.

Por último, no se deben olvidar los objetivos académicos de este trabajo, como puede ser poner en práctica los conocimientos adquiridos durante el Grado y ampliar nuestros conocimientos en distintas áreas.

Alcanzando los objetivos anteriormente descritos, se conseguirá obtener un producto de calidad, con una gran utilidad tanto social como académica, que pueda ayudar a muchos usuarios a aprender ciertos conceptos de nuestro idioma de una manera más sencilla.

1.3. Estructura de la memoria

En el **capítulo dos** se presenta el Estado de la Cuestión, en el que se explicará que es la Lectura Fácil y como se aplica y se introducirán los conceptos de Procesamiento del Lenguaje Natural (PLN) y algunas herramientas que sirven para PLN, además se hablará de figuras retóricas y servicios web, en especial qué son, su arquitectura y las ventajas y desventajas de su uso.

En el **capítulo tres** se explicarán las herramientas utilizadas para la creación de este trabajo, como pueden ser Django para el desarrollo de la aplicación y SpaCy para el etiquetado de palabras, donde se explicará que son ambas herramientas, para que se utilizan y sus características principales.

En el **capítulo cuatro** queda detallado como ha sido la gestión del proyecto, que herramientas se han utilizado para la asignación de tareas, como se ha realizado la distinción de si una tarea hace referencia al código o a la memoria y el tratamiento de cada una de ellas, las reuniones establecidas con los directores así como la utilización de un repositorio para el control de versiones de dicho trabajo.

En el **capítulo cinco** se explicarán los Servicios Web implementados por los integrantes del trabajo para dotar de funcionalidad a la aplicación.

El **capítulo seis** está enfocado en el diseño de la aplicación, se explicará el proceso de diseño de la interfaz de la aplicación, desde la creación de los primeros bocetos hasta la implementación del diseño final pasando por la evaluación de los prototipos por parte de los expertos.

En el **capítulo siete** se describe el trabajo realizado por cada uno de los autores de dicho trabajo.

Capítulo 2

Estado de la Cuestión

En primer lugar, se deben buscar los conceptos relacionados de una determinada palabra para que de esta forma se pueda obtener una mayor cantidad de resultados, y así conseguir que este sea lo más correcto y apropiado posible. Para ello, se ha utilizado una rama de la Inteligencia Artificial llamada Procesamiento del Lenguaje Natural (PLN) y en la sección 2.4 se explicará en qué consiste y las distintas herramientas disponibles para la búsqueda de conceptos relacionados de una palabra específica.

Una vez obtenidos dichos resultados, se debe hacer una distinción de cuales pertenecen al grupo de palabras fáciles y cuales al grupo de palabras difíciles, es decir, distinguir entre aquellas palabras que se utilizan asiduamente y las que no se utilizan a menudo en nuestra vida cotidiana. Esta distinción es necesaria ya que uno de nuestros objetivos es trabajar únicamente con las palabras fáciles para facilitar la compresión del significado de un concepto devolviendo un resultado lo más sencillo posible. Dado que este trabajo está enfocado para personas con discapacidad cognitiva, el concepto de lectura fácil nos ayuda a comprender como debemos trabajar con las palabras fáciles y por ello en la sección 2.1 se explicará que es la lectura fácil y algunas pautas que se pueden seguir para escribir correctamente un texto con esta adaptación, así como la definición de palabras fáciles y difíciles y los documentos que se utilizarán como referencia para las palabras fáciles.

Tras conocer que palabras se identificarán como palabras fáciles y saber como trabajar con ellas , se utilizarán figuras retóricas para comparar conceptos complejos con otros más sencillos. De esta manera el usuario se podrá hacer una idea de sus características principales. En la sección 2.2 se hablará de las figuras retóricas y se explicarán los tres tipos fundamentales que se van a utilizar para la realización de este trabajo.

Por último, se necesitará hacer uso de servicios web para dotar a la aplicación de funcionalidad. Como se ha comentado en el anterior capítulo, uno de los objetivos tecnológicos es construir la aplicación con servicios web y



Figura 2.1: Logo Lectura Fácil

en la sección 2.3 se explicará detalladamente que es un servicio web, los tipos que existen, sus características principales, su arquitectura y las ventajas de ser utilizados así como sus desventajas.

2.1. Lectura Fácil

Se llama lectura fácil¹ a aquellos contenidos que han sido resumidos y reescritos con lenguaje sencillo y claro, de forma que puedan ser entendidos por personas con discapacidad cognitiva o discapacidad intelectual. Es decir, es la adaptación de textos, ilustraciones y maquetaciones que permite una mejor lectura y comprensión.

La lectura fácil surgió en Suecia en el año 1968, donde se editó el primer libro en la Agencia de Educación en el marco de un proyecto experimental. A continuación, en 1976, se creó en el Ministerio de Justicia un grupo de trabajo para conseguir textos legales más claros. En 1984 nació el primer periódico en lectura fácil, titulado "8 páginas", que tres años más tarde, en 1987, se publicó de forma permanente en papel hasta que empezó a editarse en la web. En el año 2013, en México se produce la primera sentencia judicial en lectura fácil². En la actualidad, podemos distinguir los documentos en lectura fácil gracias al logo de la Figura 2.1.

Los documentos escritos en Lectura Fácil (Nomura et al., 2010) son documentos de todo tipo que siguen las directrices internacionales de la IFLA³ y de Inclusion Europe⁴ en cuanto al contenido y la forma. Algunas pautas a seguir para escribir correctamente un texto en Lectura Fácil son (García Muñoz, 2012):

- Evitar mayúsculas fuera de la norma, es decir, escribir en mayúsculas

¹<https://www.discapnet.es/areas-tematicas/diseno-para-todos/accesibilidad-de-comunicacion/lectura-facil>

²<https://dilofacil.wordpress.com/2013/12/04/el-origen-de-la-lectura-facil/>

³International Federation of Library Associations and Institutions

⁴Una asociación de personas con discapacidad intelectual y sus familias en Europa

sólo cuando lo dicten las reglas ortográficas, como por ejemplo, después de un punto o la primera letra de los nombres propios.

- Deben evitarse el punto y seguido, el punto y coma y los puntos suspensivos. El punto y aparte hará la función del punto y seguido.
- Evitar corchetes y signos ortográficos poco habituales, como por ejemplo: %, & y /.
- Evitar frases superiores a 60 caracteres y utilizar oraciones simples. Por ejemplo, la oración *Caperucita ha ido a casa de su abuela y ha desayunado con ella* es mejor dividirla en dos oraciones simples: *Caperucita ha ido a casa de su abuela* y *Caperucita ha desayunado con ella*.
- Evitar tiempos verbales como: futuro, subjuntivo, condicional y formas compuestas.
- Utilizar palabras cortas y de sílabas poco complejas. Por ejemplo: casa, gato, comer o mano.
- Evitar abreviaturas, acrónimos y siglas.
- Alinear el texto a la izquierda.
- Incluir imágenes y pictogramas a la izquierda y su texto vinculado a la derecha.
- Evitar la saturación de texto e imágenes.
- Utilizar uno o dos tipos de letra como mucho.
- Tamaño de letra entre 12 y 16 puntos.
- Si el documento está paginado, incluir la paginación claramente y reforzar el mensaje de que la información continúa en la página siguiente.

Se debe también hacer hincapié en la distinción entre palabras fáciles y complejas (García Muñoz, 2012), puesto que son de gran importancia para la lectura fácil. Las palabras complejas son aquellas que no se utilizan a menudo, como por ejemplo: melifluo o inefable. Las palabras complejas deben descartarse en la lectura fácil, y en su lugar se deben introducir palabras fáciles, que son aquellas que se utilizan asiduamente. La RAE (Real Academia Española) dispone de un corpus⁵ donde se encuentran varios documentos en los que se incluyen las mil palabras más usadas, las cinco mil palabras más usadas y las diez mil palabras más usadas por los hablantes de

⁵<http://corpus.rae.es/lfrecuencias.html>

lengua española. En estos listados se incluyen verbos, sustantivos, preposiciones, pronombres, adjetivos, etc... Para este trabajo, es fundamental la distinción entre palabras fáciles y palabras complejas, los listados que se acaban de mencionar son imprescindibles ya que se identificarán como palabras fáciles y por lo tanto son las palabras con las que se trabajarán para obtener los resultados.

2.2. Figuras retóricas

Las figuras literarias (o retóricas) se podrían definir como formas no convencionales de utilizar las palabras, de manera que, aunque se emplean con sus acepciones habituales, se acompañan de algunas particularidades fónicas, gramaticales o semánticas, que las alejan de ese uso habitual, por lo que terminan por resultar especialmente expresivas (Galiana y Casas, 1994). Según la RAE (Real Academia Española)⁶, “*la retórica es el arte de bien decir, de dar al lenguaje escrito o hablado eficacia bastante para deleitar, persuadir o conmover*”. Las tres principales figuras retóricas son la metáfora, el símil y la analogía. Estas figuras se basan en la comparación de dos conceptos (Galiana y Casas, 1994): el origen (o tenor), que es el término literal (al que la metáfora se refiere) y el de destino (o vehículo), que es el término figurado. La relación que hay entre el tenor y el vehículo se denomina fundamento. Por ejemplo, en la metáfora *Tus ojos son dos luceros, ojos* es el tenor, *luceros* es el vehículo y el fundamento es la belleza de los ojos.

En este trabajo se van a utilizar los tres tipos de figuras retóricas mencionados anteriormente (Calleja, 2017):

- Metáfora: Utiliza el desplazamiento de características similares entre dos conceptos con fines estéticos o retóricos. Por ejemplo, cuando el tiempo de una persona es muypreciado se dice: “Mi tiempo es oro”.
- Símil: Realiza una comparación entre dos términos usando conectores (por ejemplo, como, cual, que, o verbos). Por ejemplo, cuando nos referimos a una persona que es muy corpulenta, se dice: “Es como un oso”, ya que los osos son muy grandes.
- Analogía: Es la comparación entre varios conceptos, indicando las características que permiten dicha relación. En la retórica, una analogía es una comparación textual que resalta alguna de las similitudes semánticas entre los conceptos protagonistas de dicha comparación. Por ejemplo: “Sus ojos son azules como el mar”.

En dicho trabajo, las figuras retóricas van a permitir comparar el concepto complejo introducido por el usuario con otros conceptos más sencillos.

⁶<https://dle.rae.es/?id=WISC3uX>

2.3. Servicios Web

Para definir el concepto de servicio web de la forma más simple posible, se podría decir que es una tecnología que utiliza un conjunto de protocolos para intercambiar datos entre aplicaciones, sin importar el lenguaje de programación en el cual estén programadas o ejecutadas en cualquier tipo de plataforma⁷. Según el W3C (*World Wide Web Consortium*)⁸, “*un servicio web es un sistema software diseñado para soportar la interacción máquina-a-máquina, a través de una red, de forma interoperable*”.

Las principales características de un servicio web son (Torres, 2017):

- Es accesible a través de la Web. Para ello debe utilizar protocolos de transporte estándares como HTTP, y codificar los mensajes en un lenguaje estándar que pueda ser accesible por cualquier cliente que quiera utilizar el servicio.
- Contiene una descripción de sí mismo. De esta forma, una aplicación web podrá saber cual es la función de un determinado Servicio Web, y cuál es su interfaz, de manera que pueda ser utilizado de forma automática por cualquier aplicación, sin la intervención del usuario.
- Debe ser localizado. Debe tener algún mecanismo que permita encontrarlo. De esta forma tendremos la posibilidad de que una aplicación localice el servicio que necesite de forma automática, sin tener que conocerlo previamente.

2.3.1. Tipos de Servicios Web

Los servicios web pueden definirse tanto a nivel conceptual como a nivel técnico. A nivel técnico se pueden diferenciar dos tipos de servicios web (Torres, 2017):

- Servicios web SOAP (Simple Object Access Protocol): SOAP es un protocolo basado en XML para el intercambio de información entre ordenadores. Normalmente utilizaremos SOAP para conectarnos a un servicio e invocar métodos remotos⁹. Los mensajes SOAP tienen el formato representado en el Listado 2.1, donde podemos ver un ejemplo para reservar un vuelo y está formado por los siguientes campos:
 - <Envelope>: elemento raíz de cada mensaje SOAP. Contiene dos elementos:

⁷<http://www.jtech.ua.es/j2ee/publico/servc-web-2012-13/sesion01-apuntes.html>

⁸<https://www.w3.org/>

⁹<https://www.ibm.com/support/knowledgecenter/es>

- <Header>: es un elemento opcional que se utiliza para indicar información acerca de los mensajes SOAP. En el ejemplo del Listado 2.1 dentro del campo Header estarían los campos de reservas y pasajeros.
- <Body>: es un elemento obligatorio que contiene información dirigida al destinatario del mensaje. En el ejemplo del Listado 2.1 se puede ver los campos asociados a un itinerario, teniendo este el lugar de partida, de llegada, la fecha de llegada y la preferencia de asiento.
- <Fault>: es un elemento opcional para notificar errores. En el Listado 2.1 podemos ver que no se encuentra presente, pero en caso de ser utilizado deberá aparecer dentro del elemento *Body* y no puede aparecer más de una vez.

Listado 2.1: Estructura de un mensaje SOAP

```
<?xml version='1.0' Encoding='UTF-8'?>
<env:Envelope xmlns:env="http://www.w3.org/2003/05/soap-envelope">
  <env:Header>
    <m:reservation xmlns:m="http://travelcompany.example.org/
      reservation"
      env:role="http://www.w3.org/2003/05/soap-envelope/role/
      next">
      <m:reference>uuid:093a2da1-q345-739r-ba5d-pqff98fe8j7d</m:
      reference>
      <m:dateAndTime>2007-11-29T13:20:00.000-05:00</m:dateAndTime>
    </m:reservation>
    <n:passenger xmlns:n="http://mycompany.example.com/employees"
      env:role="http://www.w3.org/2003/05/soap-envelope/role/
      next">
      <n:name>Fred Bloggs</n:name>
    </n:passenger>
  </env:Header>
  <env:Body>
    <p:itinerary xmlns:p="http://travelcompany.example.org/reservation
      /travel">
      <p:departure>
        <p:departing>New York</p:departing>
        <p:arriving>Los Angeles</p:arriving>
        <p:departureDate>2007-12-14</p:departureDate>
        <p:departureTime>late afternoon</p:departureTime>
        <p:seatPreference>aisle</p:seatPreference>
      </p:departure>
      <p:return>
        <p:departing>Los Angeles</p:departing>
        <p:arriving>New York</p:arriving>
        <p:departureDate>2007-12-20</p:departureDate>
        <p:departureTime>mid-morning</p:departureTime>
        <p:seatPreference></p:seatPreference>
      </p:return>
    </p:itinerary>
  </env:Body>
</env:Envelope>
```

- Servicios Web RESTful: RESTful es un protocolo que suele integrar mejor con HTTP que los servicios basados en SOAP, ya que no requieren mensajes XML. Cada petición del cliente debe contener toda la información necesaria para entender la petición, y no puede aprovecharse de ningún contexto almacenado en el servidor.

2.3.2. Arquitectura Servicios Web

Hay que distinguir tres partes fundamentales en los servicios web (Torres, 2017):

- El proveedor: es la aplicación que implementa el servicio y lo hace accesible desde Internet.
- El solicitante: cualquier cliente que necesite utilizar el servicio web.
- El publicador: se refiere al repositorio centralizado en el que se encuentra la información de la funcionalidad disponible y como se utiliza.

Por otro lado, los servicios web se componen de varias capas¹⁰:

- Descubrimiento del Servicio: responsable de centralizar los servicios web en un directorio común, de esta forma es más sencillo buscar y publicar.
- Descripción del Servicio: como ya hemos comentado con anterioridad, los servicios web se pueden definir a sí mismos, por lo que una vez que los localicemos el Service Description nos dará la información para saber qué operaciones soporta y cómo activarlo.
- Invocación del Servicio: invocar a un Servicio Web implica pasar mensajes entre el cliente y el servidor. Por ejemplo, si utilizamos SOAP (Simple Object Access Protocol), el Service Invocation especifica cómo deberíamos formatear los mensajes request para el servidor, y cómo el servidor debería formatear sus mensajes de respuesta.
- Transporte: todos los mensajes han de ser transmitidos de alguna forma entre el servidor y el cliente. El protocolo elegido para ello es HTTP.

2.3.3. Ventajas de los Servicios Web

Las principales ventajas del uso de los servicios web son las siguientes (Vega Lebrún, 2005):

- Permiten la integración “justo-a-tiempo”: esto significa que los solicitantes, los proveedores y los agentes actúan en conjunto para crear sistemas que son auto-configurables, adaptativos y robustos.
- Reducen la complejidad por medio del encapsulamiento: un solicitante de servicio no sabe cómo fue implementado el servicio por parte del proveedor, y éste, a su vez, no sabe cómo utiliza el cliente el servicio. Estos detalles se encapsulan en los solicitantes y proveedores. El encapsulamiento es crucial para reducir la complejidad.

¹⁰<https://diego.com.es/introduccion-a-los-web-services>

- Promueven la interoperabilidad: la interacción entre un proveedor y un solicitante de servicio está diseñada para que sea completamente independiente de la plataforma y el lenguaje.
- Abren la puerta a nuevas oportunidades de negocio: los servicios web facilitan la interacción con socios de negocios, al poder compartir servicios internos con un alto grado de integración.
- Disminuyen el tiempo de desarrollo de las aplicaciones: gracias a la filosofía de orientación a objetos que utilizan, el desarrollo se convierte más bien en una labor de composición.
- Fomentan los estándares y protocolos basados en texto, que hacen más fácil acceder a su contenido y entender su funcionamiento.

2.3.4. Desventajas de los Servicios Web

El uso de servicios web también tiene algunas desventajas¹¹:

- Al apoyarse en HTTP, pueden esquivar medidas de seguridad basadas en firewall cuyas reglas tratan de bloquear.
- Existe poca información de servicios web para algunos lenguajes de programación.
- Dependen de la disponibilidad de servidores y comunicaciones.

2.4. Procesamiento del Lenguaje Natural

El Procesamiento del Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial que se encarga de la comunicación entre máquinas y personas mediante el uso del lenguaje natural (entendiendo como lenguaje natural el idioma usado con fines de comunicación por humanos, ya sea hablado o escrito, como pueden ser el español, el ruso o el inglés).

Para ello, una de las tareas principales en el Procesamiento del Lenguaje Natural es interpretar un texto escrito en lenguaje natural y entender su significado, entendiendo como significado la relación entre una palabra o una frase con el mundo. Para realizar dicha acción no solo es necesario el conocimiento del propio lenguaje en que está escrito el texto sino que también es necesario un conocimiento del mundo. Por tanto, uno de los grandes retos del Procesamiento del Lenguaje Natural es la representación del conocimiento. Se deben de buscar técnicas que permitan representar conceptos y relaciones semánticas entre ellos.

¹¹ <http://fabioalfarocc.blogspot.com/2012/08/ventajas-y-desventajas-del-soap.html>

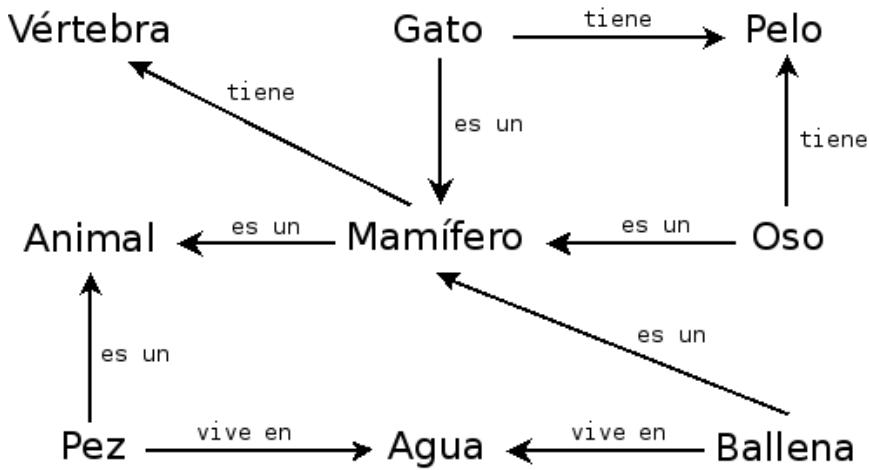


Figura 2.2: Ejemplo de Red Semántica

Una de las principales técnicas de representación de conocimiento son las redes semánticas, en ellas los conceptos que componen el mundo y sus relaciones se representan mediante un grafo. Las redes semánticas se utilizan para representar mapas conceptuales y mentales (Quillian, 1968). Los nodos están representados por el elemento lingüístico, y la relación entre los nodos sería la arista. Se puede ver un ejemplo en la Figura 2.2, donde el nodo *Oso* representa un concepto, en este caso un sustantivo que identifica a un tipo de animal, y otro nodo *Pelo* el cual también es un sustantivo. La relación entre ambos se ve representada por la arista con valor *tiene*, dando lugar a una característica de este animal: *Oso tiene pelo*.

Existen principalmente tres tipos de redes semánticas (Moreno Ortiz, 2000):

- Redes de Marcos: los enlaces de unión de los nodos son parte del propio nodo, es decir, se encuentran organizados jerárquicamente, según un número de criterios estrictos, como por ejemplo la similitud entre nodos. En la Figura 2.3, se muestra un ejemplo de Red de Marco donde por ejemplo el nodo ave tiene como características que vuela, que tiene plumas y que pone huevos, pero en cambio el nodo avestruz que es un tipo de ave, no puede volar. Por lo que los nodos de ave son las características principales de un ave, aunque no todas tienen por que cumplirlo.
- Redes IS-A: los enlaces entre los nodos están etiquetados con una relación entre ambos. Es el tipo que habitualmente se utiliza junto con las Redes de Marcos. En la Figura 2.4 se muestra una red IS-A en la que se representa que: mujer y hombre son personas, y perro y

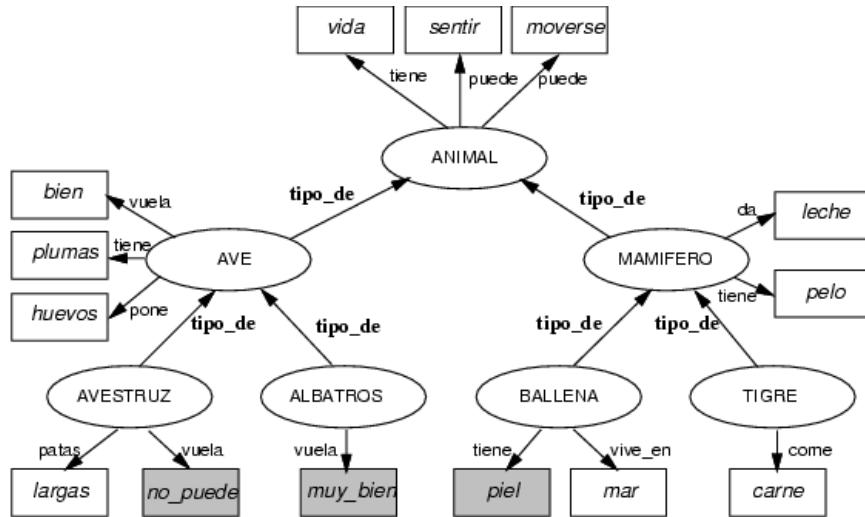


Figura 2.3: Ejemplo de Red de Marco

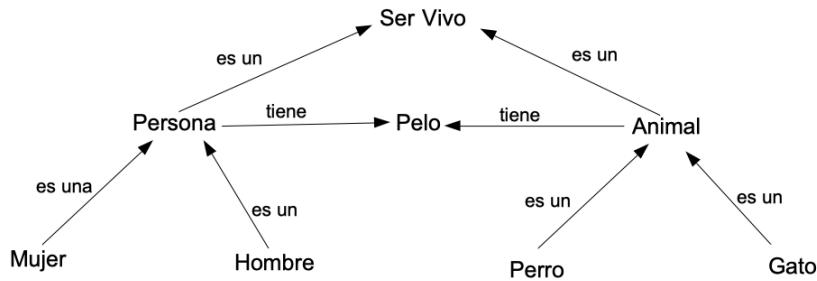


Figura 2.4: Ejemplo de Red IS-A

gato son animales. Por último, tanto personas como animales son seres vivos y una de sus características en común es que tienen pelo.

- **Grafos Conceptuales:** existen dos tipos de nodos en estas redes: nodos de conceptos, los cuales representan una entidad, un estado o un proceso y los nodos de relaciones, que indican como se relacionan los nodos de concepto. En este tipo de red semántica no existen enlaces entre los nodos con una etiqueta, sino que son los propios nodos los que tienen el significado. Se puede ver un ejemplo en la Figura 2.5 en la cual la frase "*Man biting dog*" quedaría representada (Sowa, 1983). Los cuadrados implican el concepto y el círculo la relación entre ambos, por lo que en el caso de *man* y *bite*, la acción de morder la realiza *man* siendo éste el agente, y la relación entre *bite* y *dog* sería el objeto.



Figura 2.5: Ejemplo de Grafo Conceptual

Para el trabajo que queremos realizar, existen varias aplicaciones web de redes semánticas que disponen de una representación del conocimiento y que necesitaremos para relacionar el concepto difícil introducido por el usuario con otros conceptos. A continuación, hablaremos de las más representativas.

2.4.1. ConceptNet

ConcepNet es una red semántica creada por el MIT (*Massachusetts Institute of Technology*) en 1999, diseñada para ayudar a los ordenadores a entender el significado de las palabras. ConceptNet ofrece la posibilidad de obtener de una palabra un listado de sinónimos, términos relacionados, términos derivados, el contexto de la palabra, resultados etimológicamente relacionados, símbolos, etc...

ConcepNet dispone de una aplicación web¹² donde se pueden buscar palabras en distintos idiomas, como el español, el inglés o el chino. En la Figura 2.6 se puede ver la información que devuelve la aplicación ConcepNet para la palabra chaqueta, por ejemplo, se puede ver una columna con los sinónimos del concepto, pero en distintos idiomas (americana, *jacket* o *blouson*). A continuación, aparece otra columna con los términos relacionados (saco o chaquetear) también en distintos idiomas, etc... Para algunos de los resultados de ConcepNet además aparece a la derecha de la palabra y entre paréntesis, una sola letra que indica si es un verbo, un sustantivo o un adjetivo.

En cuanto a si la palabra introducida es polisémica, ConcepNet no realiza ninguna agrupación de los resultados en función del significado, si no que muestra todo el listado de palabras y en algunas de ellas aparece el contexto al cual se refieren, por ejemplo, para la palabra arco algunos de los resultados que devuelve son arco de arquitectura y arco de geometría.

Por otro lado, ConcepNet dispone de un servicio web¹³ que devuelve los resultados en formato JSON. Siguiendo con el mismo ejemplo anterior utilizando la palabra chaqueta, se puede ver en el Listado 2.2 los resultados devueltos por el servicio web. El JSON devuelto en este caso consta de cuatro campos principales¹⁴:

- @context: URL enlazada a un archivo de información del JSON para

¹²<http://conceptnet.io/>

¹³<http://api.conceptnet.io>

¹⁴<https://github.com/commonsense/conceptnet5/wiki/AP>

The screenshot shows the search results for the Spanish term 'chaqueta' in ConceptNet 5.6. At the top, there's a yellow box with the text 'es chaqueta' and a link to 'A Spanish term in ConceptNet 5.6'. Below it, there are links to 'Sources: English Wiktionary, French Wiktionary, and Open Multilingual WordNet' and 'View this term in the API'. On the right side, there are links to 'Documentation', 'FAQ', 'Chat', and 'Blog'. The main content area is divided into four sections: 'Synonyms', 'Related terms', 'Etymologically derived terms', and 'Word forms'. Each section contains a list of terms with small colored icons next to them.

Synonyms	Related terms	Etymologically derived terms	Word forms
en jacket (n) → es jacket (n) → es americana (n) → es americana → es casaca → es casaca (n) → fr jaqueta → es chumpa (n) → es jaqueta (n) → fr blouson (n) → fr veste (n) → fr veston (n) →	en clothing → es chaquetear (v) → es saco (n) → en jacket → es chaquetear → fr branlette → fr veste → fr veston →	es jaketo →	es chaquetas (n) →

Figura 2.6: Resultados de ConcepNet para la palabra chaqueta

comprender la API. También puede contener comentarios que pueden ser útiles para el usuario.

- @id: concepto que se ha buscado y su idioma. En nuestro caso, aparece de la siguiente manera: `/c/es/chaqueta`, donde `c` significa que es un concepto o término, `es` indica el lenguaje, en este caso, el español y por último `chaqueta` que es la palabra buscada.
- edges: representa una estructura de datos devueltos por Conceptnet compuesta por:
 - @id: describe el tipo de relación que existe entre la palabra introducida y la devuelta. En el Listado 2.2 se indica que la palabra *americana* es un sinónimo de *chaqueta*.
 - @type: define el tipo del id, es decir, si es una relación (edge) o un término (nodo).
 - dataset: URI que representa el conjunto de datos creado.
 - end: nodo destino, que a su vez se compone de:
 - @id: coincide con la palabra del id anterior.
 - @type: define el tipo de id, como se ha explicado anteriormente.
 - label: puede ser la misma palabra buscada o una frase más completa, donde adquiera significado la palabra obtenida.
 - language: lenguaje en el que está la palabra devuelta de la consulta.
 - term: enlace a una versión mas general del propio término. Normalmente, suele coincidir con la URI.

- license: aporta información sobre como debe usarse la información proporcionada por conceptnet.
- rel: describe la relación que hay entre la palabra origen y destino, dentro del cual hay tres campos: @id, @type y label, descritos anteriormente.
- sources: indica porqué ConceptNet guarda esa información, este campo como los anteriores, es un objeto que tiene su propio id y un campo @type, A parte, hay un campo *contributor*, en el que aparece la fuente por la que se ha obtenido ese resultado y por último un campo *process* indicando si la palabra se ha añadido mediante un proceso automático.
- start: describe el nodo origen, es decir, la palabra que hemos introducido en ConceptNet para que haga la consulta, este campo esta compuesto por elementos ya descritos como son: @id, @type, label, language y term.
- surfaceText: algunos datos de ConceptNet se extraen de texto en lenguaje natural. El valor de surface text muestra lo que era este texto, puede que este campo tenga valor nulo.
- weight: indica la fiabilidad de la información guardada en ConceptNet, siendo normal que su valor sea 1.0. Cuanto mayor sea este valor, más fiables serán los datos obtenidos.
- view: describe la longitud de la lista de paginación, es un objeto con un id propio, y además, aparecen los campos *firstPage* que tiene como valor un enlace a la primera pagina de los resultados obtenidos, y *nextPage* que tiene un enlace a la siguiente página de la lista.

Listado 2.2: JSON devuelto por la API de ConceptNet para la palabra chaqueta

```
{
  "@context": [
    "http://api.conceptnet.io/ld/conceptnet5.6/context.ld.json"
  ],
  "@id": "/c/es/chaqueta",
  "edges": [
    {
      "@id": "/a/[r/Synonym/,/c/es/chaqueta/n,/c/es/american/a/]",
      "@type": "Edge",
      "dataset": "/d/wiktionary/fr",
      "end": {
        "@id": "/c/es/american/a",
        "@type": "Node",
        "label": "american/a",
        "language": "es",
        "term": "/c/es/american/a"
      },
      "license": "cc:by-sa/4.0",
      "rel": {
        "@id": "/r/Synonym",
        "@type": "Relation",
        "label": "Synonym"
      }
    },
    "sources": [
    ]
  }
}
```

```
{
  "@id": "/and/[ / s/ process / wikiparsec / 1 / , / s/ resource / wiktionary / fr / ]",
  "@type": "Source",
  "contributor": "/s/resource/wiktionary/fr",
  "process": "/s/process/wikiparsec/1"
},
{
  "start": {
    "@id": "/c/es/chaqueta/n",
    "@type": "Node",
    "label": "chaqueta",
    "language": "es",
    "sense_label": "n",
    "term": "/c/es/chaqueta"
  },
  "surfaceText": null,
  "weight": 1.0
},
]

"view": {
  "@id": "/c/es/chaqueta?offset=0&limit=20",
  "@type": "PartialCollectionView",
  "comment": "There are more results. Follow the 'nextPage' link for more.",
  "firstPage": "/c/es/chaqueta?offset=0&limit=20",
  "nextPage": "/c/es/chaqueta?offset=20&limit=20",
  "paginatedProperty": "edges"
}
}
```

2.4.2. Thesaurus

Thesaurus¹⁵ es una aplicación web que se autodefine como el principal diccionario de sinónimos de la web. Esta página ofrece la posibilidad de obtener los sinónimos de una palabra, pero solamente devuelve resultados en inglés. Aparte del listado de sinónimos, Thesaurus indica que tipo de palabra es y una definición de la misma así como un listado de antónimos y un listado de palabras relacionadas con dicho concepto.

Si la palabra introducida es polisémica, Thesaurus devuelve los resultados por grupos según su connotación. En la Figura 2.7, se puede ver los resultados devueltos por Thesaurus para la palabra *book*, donde este puede utilizarse como un sustantivo o como un verbo. Y dentro de si es un tipo u otro, los divide según el significado, ya sea para utilizarlo como un documento, un diario o como una reserva. Y cada uno se muestran en pestañas distintas para facilitar al usuario la distinción de conceptos.

Por otro lado, esta aplicación proporciona una API¹⁶ tipo RESTful que obtiene los sinónimos de una palabra mediante una petición HTTP GET a la url <http://thesaurus.altervista.org/thesaurus/v1>. Este devuelve los resultados en formato XML o JSON.

El contenido de la respuesta es una lista y cada elemento de esta lista contiene un par de elementos: categoría y sinónimos. Este último a su vez contiene una lista de sinónimos separados por el carácter |. Se puede ver en el Listado 2.3 el resultado devuelto por Thesaurus para la palabra *peace*

¹⁵<https://www.thesaurus.com/>

¹⁶<http://thesaurus.altervista.org/>

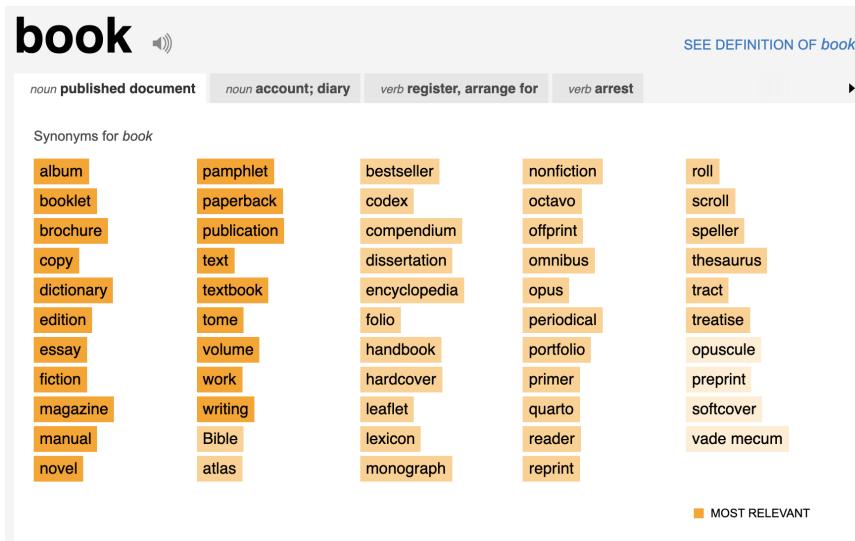


Figura 2.7: Resultados de ConcepNet para la palabra polisémica *book*

en formato XML y en el Listado 2.4 para la misma palabra, *peace*, pero en formato JSON. Ambos son muy similares, por ejemplo en formato XML se puede ver que devuelve el tipo de categoría de las palabras, en este caso son sustantivos y a continuación aparecen los sinónimos. En caso de que alguna palabra sea un antónimo aparecerá entre paréntesis al lado de la misma, como ocurre con la palabra *war*. Por otro lado, el formato JSON devuelve dentro del campo *category / categoría* todos los sinónimos, y en caso de ser un antónimo aparecerá de la misma forma que en el formato XML.

Listado 2.3: XML devuelto por Thesaurus para la palabra *peace*

```
<response>
<list>
  <category>(noun)</category>
  <synonyms> order|war (antonym) </synonyms>
</list>
<list>
  <category>(noun)</category>
  <synonyms> harmony|concord|concordance </synonyms>
</list>
<list>
  <category>(noun)</category>
  <synonyms> public security|security </synonyms>
</list>
<list>
  <category>(noun)</category>
  <synonyms> peace treaty|pacification|treaty|pact|accord </synonyms>
</list>
```

Listado 2.4: JSON devuelto por Thesaurus para la palabra *peace*

```
{ "response":
```

```
[{"list": {"category": "(noun)", "synonyms": "order|war (antonym)" }}, {"list": {"category": "(noun)", "synonyms": "harmony|concord|concordance" }}, {"lista": {"categoria": "(noun)", "synonyms": "public security|security" }}, {"lista": {"categoria": "(noun)", "synonyms": "peace treaty|pacification|treaty|pact|accord" }}]
```

2.4.3. Thesaurus Rex

Thesaurus Rex¹⁷ es una red semántica que solo admite palabras en inglés y que permite obtener las palabras relacionadas con una palabra o las categorías que comparten dos palabras.

En Thesaurus Rex las palabras tienen categorías y modificadores. Las categorías son sustantivos que definen a la palabra introducida por el usuario, como se puede ver en la Figura 2.8, la palabra *house* tiene asociadas las categorías *structure*, *object*, *item* o *building*. Los modificadores son adjetivos que describen a la palabra, siguiendo con el mismo ejemplo, *house* tiene como modificadores *permanent*, *fixed* o *wooden*. Por último, dispone de un listado de las categorías más utilizadas por los hablantes de dicha lengua, como pueden ser *permanent-structure*, *inanimate-object*, *everyday-object*, etc...

En el caso de introducir dos palabras en la aplicación, por ejemplo *coffee* y *cola*, la aplicación devuelve las categorías que comparten dichos conceptos, en este caso serían *cold-beverage*, *dark-beverage* o *stimulating-beverage*. Si se introduce una palabra polisémica, como puede ser *book*, Thesaurus Rex no realiza ninguna distinción entre que resultados corresponden a los distintos significados del concepto buscado.

La aplicación también dispone de una API pública que devuelve los resultados en formato XML. En el Listado 2.5 se muestra el XML devuelto por la aplicación para la palabra *house*. El XML devuelto consta de tres grandes bloques: *Categories*, *Modifiers* y *CategoryHeads*. Todos los resultados tienen un peso (*weight*) asignado, esto significa que cuanto mayor sea el peso mayor es la similitud con el concepto dado. Los campos que se encuentran dentro del apartado *categories* son los resultados más utilizados en ese momento por los hablantes y que Thesaurus Rex ha encontrado, como por ejemplo *permanent-structure*, los que se encuentran dentro de *modifiers*, como se ha comentado anteriormente son atributos del concepto a buscar, como por ejemplo *fixed* y por último los que se encuentran en *categoryHeads* son las categorías más simples que se han encontrado para dicho concepto, como por

¹⁷ <http://ngrams.ucd.ie/therex3/>

ejemplo *structure*.

Thesaurus Rex utiliza el contenido de la web para generar sus resultados, con lo cual la información disponible no es fija, sino que varía según los datos actuales de la web. La ventaja de utilizar esta herramienta es que se encuentra en continua actualización, pero el inconveniente es que en algunos casos la información puede resultar un poco extraña dado que se crea semiautomáticamente desde contenido de la web (Veale y Li, 2013). Por ejemplo, para la palabra *house*, uno de los resultados es *sphere*, que en un primer momento no tiene ninguna relación un concepto con otro.

Listado 2.5: XML devuelto por Thesaurus Rex para la palabra *house*

```
<MemberData>
  <Categories kw="house">
    <Category weight="91"> large : object </Category>
    <Category weight="307"> inanimate : object </Category>
    <Category weight="261"> everyday : object </Category>
    <Category weight="154"> sensitive : area </Category>
    <Category weight="318"> permanent : structure </Category>
    <Category weight="194"> permanent : construction </Category>
    <Category weight="148"> permanent : installation </Category>
    <Category weight="98"> fixed : object </Category>
  </Categories>
  <Modifiers kw="house">
    <Modifier weight="8"> recognizable </Modifier>
    <Modifier weight="9"> relevant </Modifier>
    <Modifier weight="863"> permanent </Modifier>
    <Modifier weight="15"> moderate </Modifier>
    <Modifier weight="477"> fixed </Modifier>
    <Modifier weight="5"> odd </Modifier>
    <Modifier weight="7"> archaeological </Modifier>
    <Modifier weight="5"> electrical </Modifier>
  </Modifiers>
  <CategoryHeads kw="house">
    <CategoryHead weight="6"> protection </CategoryHead>
    <CategoryHead weight="40"> obstruction </CategoryHead>
    <CategoryHead weight="5"> whole </CategoryHead>
    <CategoryHead weight="3320"> object </CategoryHead>
    <CategoryHead weight="2340"> structure </CategoryHead>
    <CategoryHead weight="98"> commodity </CategoryHead>
    <CategoryHead weight="713"> thing </CategoryHead>
    <CategoryHead weight="2"> theatre </CategoryHead>
  </CategoryHeads>
</MemberData>
```

2.4.4. Metaphor Magnet

Metaphor Magnet es una aplicación web¹⁸, que crea metáforas a partir de una palabra y que solo está disponible para el inglés. Metaphor Magnet permite al usuario introducir palabras y, ayudándose de los n-gramas de Google (Veale y Li, 2012), busca e interpreta las distintas metáforas que existan sobre dicha palabra.

Se entiende como n-grama (Manning y Schütze, 1999) a una subsecuencia de n elementos consecutivos en una secuencia dada y a su vez estos pueden ser bigramas, trigramas, etc... Metaphor Magnet lo que hace es utilizando dichos n-gramas busca en la Web cuantas veces aparece el concepto introducido y

¹⁸<http://ngrams.ucd.ie/metaphor-magnet-acl/>

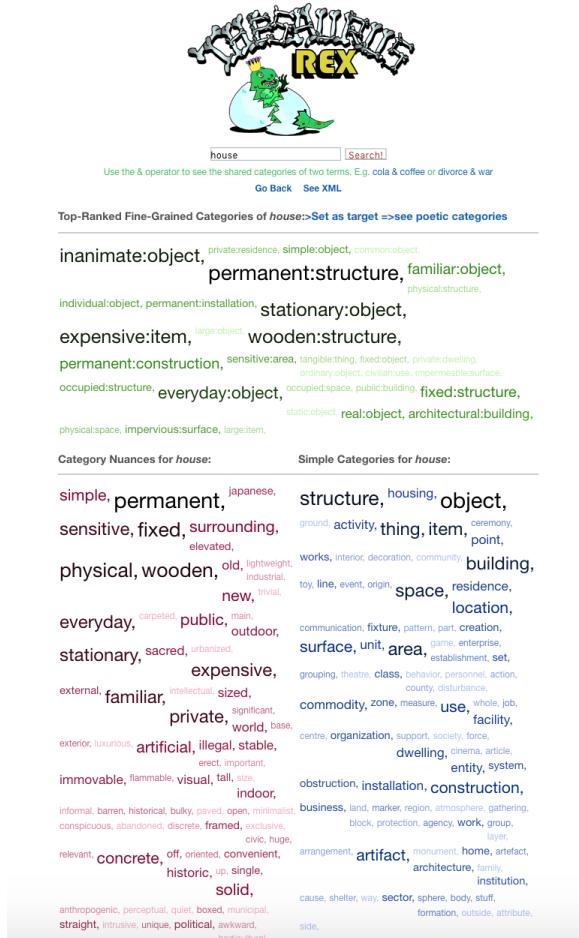


Figura 2.8: Resultados búsqueda Thesaurus Rex con la palabra *house*

cuantas veces aparece junto con otra palabra. De esta forma, se obtienen los resultados más utilizados por los hablantes, por ejemplo, en la Figura 2.9 se puede ver el resultado obtenido en la aplicación para la palabra *house*. En este caso la aplicación devuelve metáforas propias del concepto, por ejemplo *protecting:home*. En el caso de introducir una palabra polisémica, por ejemplo *book*, Metaphor Magnet no realiza ninguna agrupación según los distintos significados, si no que muestra todos los resultados juntos.

Metaphor Magnet también dispone de una API pública que devuelve los resultados en formato XML. En el Listado 2.6 se puede ver el resultado devuelto por la API para la palabra *house*. En el XML devuelto aparece la etiqueta *<Source Name>* seguido de la palabra a buscar, en este caso *house*. Otra etiqueta *<Score>* que muestra un número, cuanto mayor sea más acertado será respecto al concepto introducido, por ejemplo “*tall:building*”

tiene una etiqueta *score* de 86 mientras que “*intuitive:natural*” tiene una etiqueta *score* de 25, indicando así que es uno de los resultados menos acertados a la palabra *house*. Por último, la etiqueta *<Text>* contiene la metáfora, por ejemplo “*protecting:home*”.

Listado 2.6: XML devuelto por Metaphor Magnet para la palabra *house*

```

<Metaphor>
  <Source Name="house">
    <Text> towering:mountain </Text>
    <Score> 88 </Score>
  </Source>
  <Source Name="house">
    <Text> protecting:home </Text>
    <Score> 86 </Score>
  </Source>
  <Source Name="house">
    <Text> tall:building </Text>
    <Score> 86 </Score>
  </Source>
  <Source Name="house">
    <Text> charming:castle </Text>
    <Score> 85 </Score>
  </Source>
  <Source Name="house">
    <Text> beautiful:tree </Text>
    <Score> 84 </Score>
  </Source>
  <Source Name="house">
    <Text> charming:mansion </Text>
    <Score> 83 </Score>
  </Source>
  <Source Name="house">
    <Text> strong:rock </Text>
    <Score> 80 </Score>
  </Source>
  <Source Name="house">
    <Text> strong:elephant </Text>
    <Score> 80 </Score>
  </Source>
  <Source Name="house">
    <Text> powerful:locomotive </Text>
    <Score> 57 </Score>
  </Source>
  <Source Name="house">
    <Text> inspiring:manifesto </Text>
    <Score> 41 </Score>
  </Source>
  <Source Name="house">
    <Text> intuitive:natural </Text>
    <Score> 25 </Score>
  </Source>
</Metaphor>

```

2.4.5. WordNet

WordNet es un *corpus*¹⁹ perteneciente a NLTK (*Natural Language Toolkit*)²⁰ que almacena sustantivos, verbos, adjetivos y adverbios ignorando preposiciones, determinantes y otras palabras funcionales. Además está disponible en varios idiomas como el español, el inglés o el francés. Los conceptos se agrupan en conjuntos de sinónimos cognitivos llamados *synsets*, es decir, se agrupan según su significado. Por ejemplo, si se introduce la palabra coche, un grupo de sinónimos cognitivos sería: “auto, automóvil, carro, máquina, turismo

¹⁹Colección de documentos de texto

²⁰Conjunto de bibliotecas y programas para el Procesamiento del Lenguaje Natural

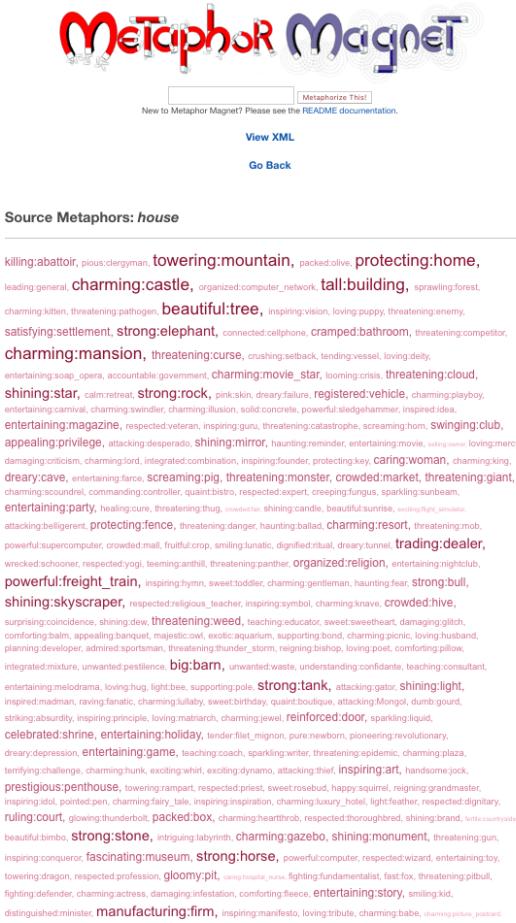


Figura 2.9: Resultados búsqueda Metaphor Magnet con la palabra *house*

y vehículoz otro grupo de sinónimos cognitivos sería: "carruaje, coche de caballos y vagón". A su vez, cada *synset* contiene:

- Hiperónimos: Palabras cuyo significado está incluido en el de otras²¹. Por ejemplo, mamífero es hiperónimo de gato y de perro ya que los gatos y los perros pertenecen al conjunto de los mamíferos.
- Hipónimos: Palabras cuyo significado incluyen el de otra²². Por ejemplo, gato es hipónimo de mamífero ya que está incluido dentro del conjunto de los mamíferos.
- Holónimos: Palabras que representan el todo respecto a una parte. Por ejemplo, coche es el holónimo de rueda, volante y acelerador ya que

²¹ <https://dle.rae.es/?id=KRW1qe2>

²² <https://dle.rae.es/?id=KU5UAn5>

The screenshot shows the EuroWordNet search interface. In the top right corner, there are several checkboxes for language options: Gloss (checked), English_3.0, Catalan_3.0, Score, Basque_3.0, Portuguese_3.0, Rels (checked), Spanish_3.0 (checked), Full, and Galician_3.0. Below these, there are dropdown menus for 'Look up' set to 'Spanish_3.0' and 'Search' also set to 'Spanish_3.0'. The main search bar contains the word 'casa'. To the left of the search bar is a sidebar with a tree view of semantic relations: casa -> near_anonym -> Synonyms -> near_synonym (which is selected and highlighted in orange). Other branches include Antonyms, Hyperonyms, and Hyponyms. The main results area shows two entries:

- spa-30-02913152-n** 298 **edificio_1** una estructura que tiene un techo, paredes y se encuentra más o menos inmueble_1 casa_1 construcción_1 permanente en un solo lugar: *había un edificio de tres pisos en la esquina; se trataba de un edificio imponente*;
- spa-30-03544360-n** 68 **casa_2** una vivienda que sirve de residencia para una o más familias: *el tiene una casa en Cape Cod*;

Figura 2.10: Resultados de la búsqueda en EuroWordNet para la palabra *casa*

forman parte de un todo, que es el coche.

- **Antónimos:** Palabras cuyo significado es totalmente opuesto al concepto inicial. Por ejemplo, el antónimo de alegre sería triste.
- **Definición de la palabra:** Oraciones cortas que explican su significado. Esta información no aparece siempre.

Existen varias aplicaciones web que utilizan dicho *corpus*, una de las más completas es EuroWordNet, ya que está disponible en varios idiomas y permite extraer sinónimos, antónimos, hiperónimos, hipónimos y holónimos. Además de algunas oraciones de ejemplo, y algunas definiciones de los *synsets* obtenidos. En la Figura 2.10 aparece la información devuelta por EuroWordNet para la palabra *casa*. Los datos obtenidos son:

- **Offset del *synset*:** es un código que lo identifica inequívocamente y que finaliza con una letra, la cual describe la categoría gramatical de los sinónimos devueltos. Por ejemplo, un *synset* de la palabra *casa* tiene asociado el *offset* spa-30-02913152-n, donde la "n" hace referencia a *noun*, esto quiere decir que es un sustantivo.
- **Sinónimos:** como por ejemplo, *edificio*, *inmueble*, *construcción* y *edificación*. Siendo todos estos resultados similares unos a otros respecto al significado del concepto a buscar.

- Definición: sirve para contextualizar el concepto. Por ejemplo, "una estructura que tiene un techo, paredes y se encuentra más o menos permanente en un solo lugar", da más información sobre su significado, haciendo así que resulte más fácil y rápida su comprensión.
- Ejemplo: en el caso de que la definición no sea suficientemente clara para el usuario, el ejemplo ayuda a comprender mejor el significado del concepto. Por ejemplo, había un edificio de tres pisos en la esquina.

Si se introduce una palabra polisémica, como portero, EuroWordNet devuelve en los distintos grupos de sinónimos cognitivos o *synsets* los resultados. En este caso devolverá en un *synset*: "conserje, guardabarrera y ostiario" haciendo referencia a la profesión y en otro *synset*: "arquero, guardameta, guardavalla y golero" que son los resultados para el significado de portero como jugador. Además, en el desplegable de la izquierda se muestra la información que se puede consultar de dicho concepto. Por ejemplo, si se selecciona *has hyperonym* se obtendrán los hiperónimos de casa, o seleccionando *has hyponym* se obtendrán los hipónimos.

Capítulo 3

Herramientas Utilizadas

En este capítulo se van a explicar las herramientas utilizadas para el desarrollo de este trabajo. En el apartado 3.1 se explicará Django que es el *framework* utilizado para el desarrollo de la aplicación y en el apartado 3.2 se explicará SpaCy, que es la herramienta que se ha utilizado para la clasificación semántica de las palabras fáciles. Las herramientas y recursos que se han usado para la gestión del proyecto se explicarán en un capítulo aparte.

3.1. Django

Para construir un servicio web se necesita una manera de gestionar los elementos propios de estos servicios, como pueden ser el procesamiento de formularios y el mapeo de urls, para satisfacer estas necesidades, se requieren *frameworks* web que son estructuras que contienen los componentes necesarios para el desarrollo de aplicaciones web¹. Django es un *framework* de alto nivel que permite el desarrollo rápido de sitios web seguros y mantenibles y se basa en el patrón MVC². Fue desarrollado entre los años 2003 y 2005 por un grupo de programadores que se encargaban de crear y mantener sitios web de periódicos. Es gratuito y de código abierto y dispone de una gran documentación actualizada así como muchas opciones de soporte gratuito y de pago.

Algunas de las razones por las que se ha elegido este *framework* han sido las siguientes³:

- Seguridad: Implementa por defecto algunas medidas de seguridad para evitar SQL Injection(adición de consultas SQL malignas que puedan

¹<https://tutorial.djangogirls.org/es/django/>

²<https://docs.djangoproject.com/en/2.0/>

³<https://openwebinars.net/blog/que-es-django-y-por-que-usarlo/>

alterar la base de datos) o Clickjacking(Que el usuario haga click en un enlace oculto, para obtener información del mismo sin su permiso o incluso tomar el control de su ordenador) por JavaScript.

- Escalabilidad: Se puede pasar de una aplicación sencilla a otra más compleja rápidamente, ya que es muy fácil añadir nuevos módulos al *framework*.
- Fácil acceso a bases de datos: Mediante ORM(Object Relational Mapper), que es una biblioteca para el acceso de datos, generando clases a partir de las tablas de la base de datos para poder realizar las consultas. Django tiene su propio ORM, con el que se pueden hacer consultas de manera muy intuitiva.
- Además, es muy popular por lo que para resolver cualquier problema que surja, como se ha explicado anteriormente, hay mucha documentación disponible y muchos hilos en foros de programación en donde encontrar posibles soluciones.

3.2. SpaCy

Cuando se obtuvieron las palabras fáciles de la RAE solo se necesitaban adverbios, nombres, verbos y adjetivos, por lo que se requería un análisis semántico de cada una de ellas para obtener solo las palabras que pertenecían a uno de estos grupos. Para ello se utilizó SpaCy. SpaCy⁴ es una biblioteca de código abierto para el Procesamiento del Lenguaje Natural en Python. Soporta más de 34 idiomas, entre ellos el español.

El analizador semántico de SpaCy, según un estudio realizado en 2015 por la universidad Emory⁵, es el más rápido y según Medium Corporation tiene un índice de acierto mucho mayor que el analizador sintáctico de NLTK⁶. Para utilizar en analizador, hay que importar la biblioteca de idioma correspondiente (en este caso español: “es_core_news_sm”) y pasar como parámetro las palabras que se deseen clasificar. El resultado para cada palabra introducida en el analizador, serán una serie de etiquetas, de todas ellas, la utilizada es la etiqueta “pos”, que contiene el grupo semántico de la palabra analizada, como se puede ver en la Figura 3.1, se ha introducido la frase: “el coche es rojo” y se han obtenido los siguientes resultados:

- el: ha obtenido la etiqueta DET, haciendo referencia a que es un determinante.
- coche: ha obtenido la etiqueta NOUN, que significa nombre en inglés.

⁴<https://spacy.io/>

⁵<https://www.aclweb.org/anthology/P15-1038>

⁶<https://medium.com/@pemagrg/private-nltk-vs-spacy-3926b3674ee4>

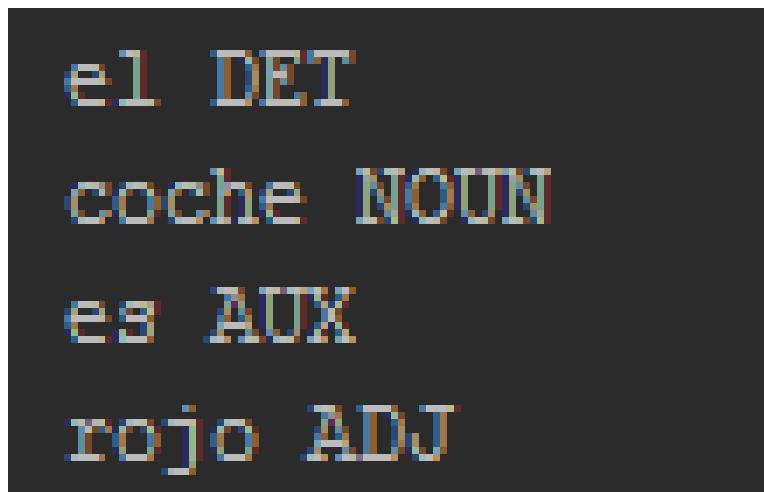


Figura 3.1: Ejemplo de clasificación de palabras

- es: ha obtenido la etiqueta AUX, haciendo referencia a que es un verbo auxiliar.
- rojo: ha obtenido la etiqueta ADJ, haciendo referencia a que es un adjetivo.

Se ha elegido ya que ha sido el que mejor resultado ha dado. Inicialmente se utilizó el POS-tagger (clasificador sintáctico de palabras) de NLTK, pero su índice de acierto no era muy bueno, por lo que se descartó su uso. A continuación se probó con SpaCy y ,además de ser más rápido, su índice de acierto ha sido prácticamente del 100 % ,por lo que fue lo finalmente utilizado.

Capítulo 4

Metodología de gestión del Proyecto

Desde el inicio del proyecto se ha buscado la eficacia y la mejora continua, es por ello que se han tenido reuniones asiduamente con los directores de este trabajo, cada dos o tres semanas en donde se corregían los fallos, se indicaban las siguientes tareas por hacer y se buscaban soluciones a los problemas y dudas que pudieran surgir o plantearse. Por otro lado, ha habido comunicación con ambos directores vía email para consultar dudas o para concretar tutorías. En relación a la gestión de configuración se ha utilizado la plataforma de desarrollo online GitHub para llevar el control de versiones del proyecto.

Se ha hecho uso de un gestor de tareas para emplearlo como radiador de información y que así todos los integrantes del proyecto puedan conocer en cada momento el estado de este. Para ello, se ha elegido Trello, ya que dispone de una interfaz simple, amigable y que no lleva a confusión a la hora de crear nuevas tareas. Existen dos tipos de tareas en el tablero: las relacionadas con código y las relacionadas con la memoria. Se ha realizado una distinción entre ambas, ya que la forma de cambiar su estado en el tablero varía significativamente. Para hacer esta distinción se ha escrito la palabra CÓDIGO o la palabra MEMORIA según corresponda delante de la descripción de la misma. El tablero de tareas tiene tres columnas:

- TO DO: Tareas a realizar, desgranadas al mayor detalle posible e intentando que éstas sean lo más independientes las unas de las otras. De esta forma, se asegura que cada integrante del equipo trabaja en una tarea específica que no influye en el trabajo del otro compañero.
- En proceso: En el momento en el que un integrante del grupo se asigna una tarea, esta se pasa de la columna TO DO a la columna “En proceso”, lo que indica que se encuentra en proceso de realización y que ningún otro compañero puede ponerse a trabajar en ella.

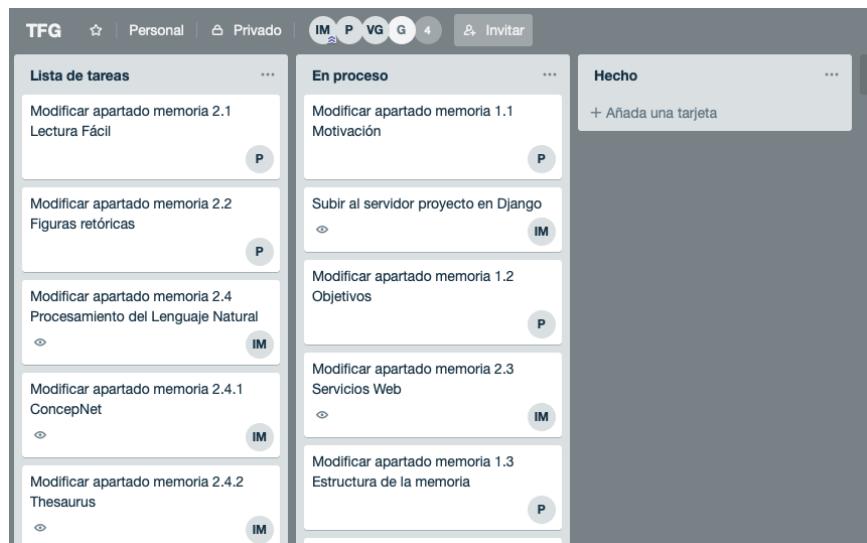


Figura 4.1: Tablero de tareas al inicio del proyecto

- Hecho: En esta columna se encuentran las tareas terminadas y validadas. La validación depende del tipo de tarea: si la tarea es de código, la validación la deben hacer los miembros del equipo, y si es de memoria la deben hacer los directores.

En la Figura 4.1 se muestra el estado del tablero al inicio del proyecto.

Capítulo 5

Servicios Web para la obtención de conceptos fáciles relacionados

Como se ha comentado en anteriores capítulos existen varias aplicaciones web que son redes semánticas y facilitan sinónimos, términos relacionados, metáforas, hipéronimos, etc... de un concepto dado por el usuario. Para este trabajo, algunas de ellas son de gran utilidad ya que de esta forma podemos obtener las palabras fáciles para un concepto más complicado, pero hay que saber exactamente cuál es la que mejor conviene y la que mejores resultados ofrece. Hay que corroborar que los conceptos devueltos son correctos y que disponen de un significado claro y parecido respecto al concepto buscado. Para ello se han implementado dos servicios web, uno utilizando la aplicación de ConceptNet y el otro utilizando WordNet, en la sección 5.1 se explicará el servicio web para ConceptNet con los métodos creados y su explicación y en la sección 5.2 aparece lo mismo pero utilizando WordNet. Por último, una vez implementados dichos servicios, se han realizado una serie de pruebas para comprobar que aplicación era más necesaria para la aplicación final, en la sección 5.3 queda reflejado dichas pruebas, como se han realizado, con que cantidad de palabras se han hecho las pruebas y cuales han sido los resultados obtenidos.

5.1. Diseño de la evaluación

Para decidir que red semántica se iba a usar se decidió hacer una prueba con palabras que fuesen lo más heterogéneas posible, para ello se eligieron una serie de artículos periodísticos de distintos temas: medioambiental, tecnológico, deportivo y político. Estos artículos se filtraron para utilizar únicamente los verbos, artículos, sustantivos y adverbios. Para cada una de estas palabras se obtuvieron en WordNet y ConceptNet sus sinónimos y términos relacionados

1000 palabras

Resultados	Lista 1000 palabras			Lista 5000 palabras			Lista	
	% nada	% noTerm	% noSin	% nada	% noTerm	% noSin	% nada	
D1	2.1 %	2.1 %	2.1 %	2.1 %	2.1 %	2.1 %	2.1 %	
D2	11.6 %	11.6 %	11.6 %	11.6 %	11.6 %	11.6 %	11.6 %	
D3	5.5 %	5.5 %	5.5 %	5.5 %	5.5 %	5.5 %	5.5 %	

(en WordNet se entiende como término relacionado a los hiperónimos e hipónimos) y se analizaron, tanto la cantidad de términos relacionados y sinónimos que generaban cada una de las redes semánticas que coincidían con alguna de las palabras fáciles de la RAE (prueba cuantitativa), como la calidad de los mismos, es decir, si las palabras que generaban tenían alguna relación aceptable con la palabra origen(prueba cualitativa).

5.2. Resultados cuantitativos

5.3. Servidor de Base de Datos

Para este proyecto, se ha utilizado una base de datos para la persistencia de los datos. El sistema encargado de la gestión de la base de datos corresponde con MariaDB y esta constituida por varias tablas:

- Se ha hecho uso de MCR (*Multilingual Central Repository*), utilizando la versión 3.0. MCR es una base de datos de código abierto que integra distintas versiones de WordNet para seis lenguajes diferentes: Inglés, Español, Catalán, Vasco, Gallego y Portugués.
 - Tres tablas que guardan tanto las 1.000, 5.000 y 10.000 palabras más usadas de la RAE.
 - Otra tabla que almacena los pictogramas de ARASAAC.
- TERMINAR!!!

5.4. Servicio Web para obtener sinónimos fáciles

La implementación de dicho servicio web se basa en que introduciendo una palabra y un nivel de búsqueda, este devuelve todos los sinónimos fáciles correspondientes en formato JSON. Para ello se realiza la siguiente petición GET:

`http://127.0.0.1:8000/easySynonym/json/word=palabra&level=nivel`

Donde *palabra* es el concepto a buscar y *nivel* es el grado de búsqueda, el cual puede tomar los siguientes valores:

- Nivel 1 (Nivel sencillo): Se realizará la búsqueda en las 1.000 palabras más usadas de la RAE.
- Nivel 2 (Nivel medio): Se realizará la búsqueda en las 5.000 palabras más usadas de la RAE.
- Nivel 3 (Nivel avanzado): Se realizará la búsqueda en las 10.000 palabras más usadas de la RAE.

Una vez introducida la palabra y el nivel, se realiza una consulta a la base de datos de MCR 3.0 a través de una *queryset* para obtener todos los *offsets* cuya palabra sea igual que la introducida. Por cada *offset* volveremos a buscar en la tabla *Variant* de la base de datos de MCR 3.0 para obtener las palabras que comparten dicho identificador y posteriormente, mediante un cursor se realizará una búsqueda en una de las tres tablas de la RAE (en función del nivel introducido) y buscará si alguna de estas palabras se encuentra en dicha tabla. Si el resultado es positivo se añadirá al JSON.

Por ejemplo, si se realizará una búsqueda para la palabra inmueble con un nivel de búsqueda 2, la petición GET sería de la siguiente manera:

`http://127.0.0.1:8000/easySynonym/json/word=inmueble&level=2`

Y en la Figura 5.1 se puede ver el resultado obtenido en formato JSON. En primer lugar aparece la palabra buscada, en este caso inmueble y a continuación un objeto que incluye tanto el *offset* como la lista de sinónimos fáciles .

5.5. Servicio Web para obtener hipónimos fáciles

La implementación de dicho servicio web se basa en que introduciendo una palabra y un nivel de búsqueda, este devuelve todos los hipónimos fáciles correspondientes en formato JSON. Para ello se realiza la siguiente petición GET:

`http://127.0.0.1:8000/easyHyponym/json/word=palabra&level=nivel`

Donde *palabra* es el concepto a buscar y *nivel* es el grado de búsqueda, igual que en el caso explicado anteriormente para buscar sinónimos fáciles.

Una vez introducida la palabra y el nivel, se realiza una consulta a la base de datos de MCR 3.0 a través de una *queryset* para obtener todos los *offsets* cuya palabra sea igual que la introducida. Por cada *offset* volveremos a buscar en la tabla *Relation* de la base de datos de MCR 3.0 para obtener los offsets que se encuentran en la columna *targetSynset*. Después, con cada *offset* obtenido de esta *queryset*, se realizará la búsqueda en la tabla *Variant* para obtener las palabras cuyo identificador sea igual que el offset de *targetSynset*. Mediante un cursor se realizará una búsqueda en una de las tres tablas de la RAE (en función del nivel introducido) y buscará si alguna de estas palabras se encuentra en dicha tabla. Si el resultado es positivo se añadirá al JSON.

```
GET http://127.0.0.1:8000/easySynonym/json/word=inmueble&level=2

Response headers 5 Request headers 0

date: Sun, 05 May 2019 19:47:33 GMT
server: WSGIServer/0.2 CPython/3.7.0
content-type: application/json
x-frame-options: SAMEORIGIN
content-length: 119

[Array[2]
-0: {
    "word": "inmueble"
},
-1: {
    "offset": "spa-30-02913152-n",
    "synonyms": [Array[4]
        0: "casa",
        1: "construcción",
        2: "edificio",
        3: "edificios"
    ],
}
]
```

Figura 5.1: JSON devuelto al buscar los sinónimos fáciles de inmuble

Por ejemplo, si se realizará una búsqueda para la palabra inmuble con un nivel de búsqueda 2, la petición GET sería de la siguiente manera:

`http://127.0.0.1:8000/easyHyponym/json/word=inmuble&level=2`

Y en la Figura 5.2 se puede ver el resultado obtenido en formato JSON. En primer lugar aparece la palabra buscada, en este caso inmuble y a continuación un objeto que incluye el *offset* de hipónimo fácil , la lista de hipónimos fáciles así como el *offsetFather*, este identificador corresponde al *offset* de uno de los *synsets* de inmuble. De esta forma, si una palabra buscada dispone de varios *synsets*, se podrá mostrar sus correspondientes sinónimos e hipónimos.

5.6. Servicio Web para obtener hiperónimos fáciles

La implementación de dicho servicio web se basa en que introduciendo una palabra y un nivel de búsqueda, este devuelve todos los hiperónimos

fáciles correspondientes en formato JSON. Para ello se realiza la siguiente petición GET:

`http://127.0.0.1:8000/easyHyperonym/json/word=palabra&level=nivel`

Donde *palabra* es el concepto a buscar y *nivel* es el grado de búsqueda, igual que en los casos explicados anteriormente para buscar sinónimos e hipónimos fáciles.

Una vez introducida la palabra y el nivel, se realiza una consulta a la base de datos de MCR 3.0 a través de una *queryset* para obtener todos los *offsets* cuya palabra sea igual que la introducida. Por cada *offset* volveremos a buscar en la tabla *Relation* de la base de datos de MCR 3.0 para obtener los offsets que se encuentran esta vez y con diferencia de la búsqueda de hipónimos fáciles, en la columna *sourceSynset*. Después, con cada *offset* obtenido de esta *queryset*, se realizará la búsqueda en la tabla *Variant* para obtener las palabras cuyo identificador sea igual que el offset de *sourceSynset*. Mediante un cursor se realizará una búsqueda en una de las tres tablas de la RAE (en función del nivel introducido) y buscará si alguna de estas palabras se encuentra en dicha tabla. Si el resultado es positivo se añadirá al JSON.

Por ejemplo, si se realizará una búsqueda para la palabra inmueble con un nivel de búsqueda 2, la petición GET sería de la siguiente manera:

`http://127.0.0.1:8000/easyHyperonym/json/word=inmueble&level=2`

En la Figura 5.3 se puede ver el resultado obtenido en formato JSON. En primer lugar aparece la palabra buscada, en este caso inmueble y a continuación un objeto que incluye el *offset* de hiperónimo fácil, la lista de hiperónimos fáciles así como el *offsetFather*, ya explicado en el apartado anterior.

5.7. Servicio Web para obtener una metáfora

Este servicio web dado un offset y una profundidad obtiene las metáforas de un concepto, es decir, se obtienen tanto los sinónimos como los hiperónimos fáciles formando la metáfora y este servicio devuelve dichos resultados en formato JSON. Para ello se realiza la siguiente petición GET:

`http://127.0.0.1:8000/metaphor/json/word=palabra&level=nivel`

Donde *palabra* es el concepto a buscar y *nivel* es el grado de búsqueda, igual que en los casos explicados anteriormente para buscar sinónimos, hipónimos e hiperónimos fáciles.

La implementación de dicho servicio es idéntico al ya explicado en el servicio web para obtener sinónimos fáciles e hiperónimos fáciles, con la única diferencia que para poder formar la metáfora se llama a otro servicio implementado, el cual utiliza Spacy para definir si el concepto es masculino o femenino y si está en singular o plural.

Por ejemplo, si se realizará una búsqueda para la palabra inmueble con

un nivel de búsqueda 2, la petición GET sería de la siguiente manera:

`http://127.0.0.1:8000/metaphor/json/word=inmueble&level=2`

En la Figura 5.4 se puede ver el resultado obtenido en formato JSON. En primer lugar aparece la palabra buscada, en este caso inmueble y a continuación un objeto que incluye el *offset* del sinónimo o hiperónimo fácil, el *offsetFather* ya explicado en apartados anteriores y la lista de metáforas (es una casa, es una construcción, son unos edificios, es un edificio).

5.8. Servicio Web para obtener un símil

Este servicio web dado un offset y una profundidad obtiene los símiles de un concepto, es decir, se obtienen los hipónimos fáciles formando así el símil y devolviendo los resultados en formato JSON. Para ello se realiza la siguiente petición GET:

`http://127.0.0.1:8000/simil/json/word=palabra&level=nivel`

Donde *palabra* es el concepto a buscar y *nivel* es el grado de búsqueda, igual que en los casos explicados anteriormente. Para la implementación de dicho servicio, se utiliza exactamente el mismo código explicado en el apartado de obtención de hipónimos fáciles pero ncluyendo la llamada al servicio de Spacy para volver a definir si el concepto es masculino o femenino y si es singular o plural.

Por ejemplo, si se realizará una búsqueda para la palabra inmueble con un nivel de búsqueda 2, la petición GET sería de la siguiente manera:

`http://127.0.0.1:8000/simil/json/word=inmueble&level=2`

En la Figura 5.5 se puede ver el resultado obtenido en formato JSON. En primer lugar aparece la palabra buscada, en este caso inmueble y a continuación un objeto que incluye el *offset* del hipónimo fácil, el *offsetFather* ya explicado en apartados anteriores y la lista de símiles (es como una biblioteca es como un colegio, es como un restaurante).

5.9. Servicio Web Para ConceptNet

El primer Servicio Web implementado obtiene los resultados de ConceptNet, una red semántica que al introducir una palabra devuelve sinónimos y términos relacionados de la misma (dicha red semántica ha sido descrita detalladamente en el apartado 2.4.1).

Este servicio web consta de tres servicios, como se puede ver en la Figura 5.6. El primero, cuando se introduce en el campo de texto una palabra en castellano, se realiza una consulta a la API de ConceptNet y se muestran únicamente los sinónimos obtenidos para dicha palabra. El segundo, realiza otra consulta a la API de ConceptNet pero esta vez en lugar de obtener los sinónimos, devuelve solamente los términos relacionados.

Y por último, el tercer servicio hace uso de los servicios 1 y 2 para obtener tanto los sinónimos como los términos relacionados del concepto introducido. Con los resultados obtenidos, se realiza una nueva búsqueda en la lista de “las 1000 palabras más utilizadas del castellano según la RAE” y en caso de que algún resultado se encuentre en dicha lista, se le mostrará al usuario y si no, se mostrará un mensaje informativo indicando que no hay resultados. Aparte, este servicio consta de un selector numérico donde se indica la profundidad de la búsqueda que por defecto, su valor es 1. Esto quiere decir que, por ejemplo, con un nivel de profundidad de nivel 1 se realizará el proceso descrito anteriormente una única vez. Sin embargo, si se introduce una profundidad mayor, por ejemplo 2, se repetirá el proceso dos veces pero esta vez buscando los sinónimos y los términos relacionados de los resultados y no de la palabra introducida en un primer momento. Por ejemplo, si el usuario introduce la palabra Gato, ConcepNet devuelve como sinónimos: feria, madrileño y felino pero ninguna de estas palabras se encuentran en la lista de “las 1000 palabras más utilizadas del castellano según la RAE”, por lo que el servicio web implementado buscará los sinónimos y términos relacionados de cada una de las palabras, es decir, de feria, madrileño y felino. Si alguno de los resultados se encuentra en la lista, entonces se mostrarán y si no, aparecerá el mensaje informativo indicando que no hay resultados.

Si por otro lado, se añade un nivel de profundidad aún mayor pero la coincidencia de los resultados con la lista se encuentra en un nivel anterior, la búsqueda se detendrá y se indicará al usuario en qué nivel ha sido obtenido con éxito dicho resultado. Ahora bien, dependiendo de qué tipo de palabra haya coincidido y en qué nivel, se generará el resultado utilizando un tipo de comparación distinta. Estas pueden ser:

- Si la palabra que coincide es un sinónimo de la palabra original y se ha encontrado en un nivel de profundidad 1, entonces se entiende que existe una similitud alta entre ambos conceptos por lo que el resultado que se mostrará tendrá el siguiente aspecto: “A **es** B”. Por ejemplo, si se ha introducido “hogar” y el resultado obtenido con éxito es “casa”, se mostrará: “hogar **es** casa”.

- En cualquier otro caso, se entiende que la similitud entre los conceptos es más baja por lo que el mensaje que aparecerá será: “A **es como** B”. Por ejemplo, si se ha introducido “casa” y se ha encontrado como término relacionado en el nivel 1 de profundidad o en cualquier otro nivel la palabra “edificio”, la comparación que se mostrará será: “casa **es como** edificio”.

5.10. Servicio Web para WordNet

Para la implementación y desarrollo del servicio web, se ha hecho uso de MCR (*Multilingual Central Repository*), el cuál es una base de datos de código abierto que integra distintas versiones de WordNet para seis lenguajes diferentes: Inglés, Español, Catalán, Vasco, Gallego y Portugués. Lo que se consigue al utilizar MCR es poder acceder a la base de datos de WordNet y obtener los resultados que dicha aplicación. Como se puede ver en la Figura 5.8, el usuario dispone de un campo de texto donde puede introducir únicamente una palabra. Cuando se pulsa el botón de enviar, se capta dicho concepto mediante una petición POST a través de un formulario. Al tener la base de datos de WordNet, podemos realizar consultas directamente para obtener los resultados en función de si se buscan sinónimos, hiperónimos o hipónimos, y aunque la base de datos es mysql, al estar utilizando el framework Django este no admite la realización de *querys* si no que utiliza *querysets*. Las *querysets* son listas de objetos de un modelo determinado y permiten leer los datos de la base de datos, filtrarlos y ordenarlos. Por lo que el siguiente paso, es realizar una *queryset* para obtener todas las filas filtrando por el nombre, posteriormente se busca por separado los sinónimos, los hiperónimos y los hipónimos. Para ello, se vuelve a realizar una *queryset* donde el filtro de búsqueda es que tengan el mismo *offset*, el *offset* es un “identificador” del *synset* que lo tienen todas las palabras que lo forman. Una vez obtenidos los sinónimos, hiperónimos e hipónimos de la base de datos, se realiza una búsqueda en un fichero csv que tiene las mil palabras de la RAE quedándose con las palabras que se encuentren en dicho fichero. Los resultados se guardan en un diccionario cuyas claves son “sinónimos”, “hipónimos” e “hiperónimos” y se convierten a formato JSON. Una vez convertidos se pasan al HTML, donde se recorre y se muestra como en la Figura 5.9 donde los resultados se han dividido en 6 campos: Sinónimos que coinciden con las palabras que forman las mil palabras de la RAE, Sinónimos que devuelven la aplicación de WordNet, y de la misma manera para los hipónimos y los hiperónimos. Por último, el usuario dispone de un botón como el de la Figura 5.7 para ver los resultados que coinciden con las mil palabras de la RAE en formato JSON.

5.11. Pruebas Estadísticas

Una vez implementados los distintos servicios web, los integrantes del grupo deben elegir cuál de ellos devuelve más resultados fiables y cuantos se pueden utilizar, ya que puede ser que los resultados obtenidos difieran mucho del concepto inicial a buscar. Es por ello que se realizaron varias pruebas introduciendo una cantidad de palabras para medir cuantos de los sinónimos y términos relacionados generados, coincidían con las listas de

“las 1000 palabras más utilizadas del castellano según la RAE”, “las 5000 palabras más utilizadas del castellano según la RAE” y “las 10000 palabras más utilizadas del castellano según la RAE”. Lo primero que se buscó fue un artículo periodístico que constaba de un total de 2.542 palabras, pero que tras un proceso de filtrado se eliminaron aquellas palabras que no fueran verbos, sustantivos, adverbios o adjetivos así como las palabras que apareciesen repetidas, quedando finalmente con un total de 717 palabras válidas para realizar la prueba. A continuación, se introdujeron en el prototipo de ConceptNet y de WordNet generando unos resultados que se describirán y se analizarán a continuación.

5.11.1. Estadística para ConceptNet

Para esta prueba se tuvieron en cuenta principalmente tres parámetros:

- Cantidad de palabras que tuviesen al menos un sinónimo.
- Cantidad de palabras que tuviesen al menos un término relacionado.
- Cantidad de palabras de las que no se obtuvo ningún sinónimo ni ningún término relacionado.

Como se puede observar en la Figura 5.10 no se obtuvo ningún resultado para un total de 565 palabras, es decir, un 78,8 % de las palabras introducidas no disponen de ningún sinónimos ni término relacionado que aparezca en la lista de las 1000 palabras de la RAE. Del 21,20 % restante que se atribuye a un total de 173 coincidencias que si disponen de resultados coincidentes en dicha lista, 51 palabras son sinónimos y 122 términos relacionados, es decir, el 6,25 % pertenece a los sinónimos y el 14,95 % a términos relacionados.

Se realizó una segunda prueba pero comparando esta vez con la lista de las 5000 palabras de la RAE, los resultados mejoran con respecto a los resultados anteriores, pasando de 173 a 362 las palabras que tienen al menos un sinónimos o un término relacionado en la lista, como se puede comprobar en la Figura 5.11, lo que supone un aumento del 109 %. Los sinónimos encontrados son 94 (aumento del 84 % respecto a los resultados anteriores) y los términos relacionados son 268 (aumento del 120 %), por otra parte las palabras de las que no se obtuvo ninguna coincidencia pasaron a ser 428 (una reducción del 24 %).

Por último, se probó con la lista de 10.000 palabras obteniendo unos resultados de 423 coincidencias totales como se puede comprobar en la Figura 5.12 (aumento del 16 % con respecto a la lista de 5.000 palabras), de estas 423 coincidencias 112 corresponden a sinónimos (aumento del 19 %) y 311 términos relacionados (aumento del 16 %). Las palabras que no obtuvieron ningún resultado fueron 387 (disminución del 10 %).

5.11.2. Estadística para WordNet

Para esta prueba se valoraron los siguientes parámetros:

- Cantidad de palabras que tuviesen al menos un sinónimo.
- Cantidad de palabras que tuviesen al menos un hipónimo.
- Cantidad de palabras que tuviesen al menos un hiperónimo.
- Cantidad de palabras de las que no se obtuvo ningún sinónimo, ni hipónimo ni hiperónimo.

Al contrastar los resultados de las consultas a WordNet con la lista de las 1.000 palabras, se obtuvieron como se puede observar en la Figura 5.13 un total de 430 coincidencias (131 sinónimos, 112 hipónimos y 187 hiperónimos) y un total de 488 palabras no tienen ningún resultado, lo que supone un 68 % sobre el total de palabras introducidas.

Al aumentar la lista de palabras a 5.000(Figura 5.14), se obtuvieron un total 631 coincidencias(aumento del 47 %) de las cuales 234 eran sinónimos (aumento del 78 %), 164 hipónimos (46 % más) y 233 hiperónimos (aumento del 25 %). El número de palabras sin coincidencia, fue de 427(disminución del 13 %).

Para finalizar la prueba, se probó con la lista de las 10.000 palabras fáciles y los resultados como se pueden apreciar en la Figura 5.15 fueron de un total de 685 coincidencias, lo que supone un aumento del 9 % respecto a los resultados obtenidos con la lista de 5.000 palabras, de las cuales 262 eran sinónimos (aumento del 12 %), 177 hipónimos (aumento del 8 %) y 246 hiperónimos(aumento del 6 %). Por último el número de palabras sin coincidencia fue de 413 sobre el total de 717, lo que supone una disminución del 3 % con respecto a la lista anterior.

5.11.3. Conclusiones

A continuación, se muestran unas gráficas para poder analizar mejor los resultados y obtener una conclusión más clara. En cuanto a los resultados obtenidos utilizando la lista de las 1000 palabras de la RAE podemos observar en la Figura 5.16 que un 78,80 % de las palabras introducidas en ConcepNet no tiene ningún sinónimo ni ningún término relacionado frente a un 68,06 % de las palabras introducidas en WordNet que tampoco tienen ningún resultado coincidente en la lista, como se puede ver en la Figura 5.17 Y los resultados que si disponen de un algún concepto, ya sea un sinónimo, hiperónimo o hipónimo en WordNet es mayor que en ConcepNet, por lo que en un primer momento el servicio web de WordNet sería más acorde a utilizar para nuestra aplicación.

Si nos fijamos en los resultados obtenidos utilizando la lista de las 5000 palabras de la RAE, podemos ver en la Figura 5.18 el porcentaje de palabras de ConcepNet que no tienen ningún resultado coincidente con la lista es similar al de la Figura 5.19 de WordNet, siendo estos un 59,7% y un 59,55% respectivamente. Respecto a los sinónimos, podemos ver que WordNet tiene un porcentaje superior al de ConcepNet, siendo este un 13,10% frente a un 32,63%. Por lo que teniendo en cuenta tanto estos resultados como los anteriores, seguiríamos decantandonos más por WordNet que por ConcepNet.

Y por último, utilizando la lista de las 10000 palabras de la RAE, podemos ver en la Figura 5.20 que el porcentaje de palabras no encontradas por ConcepNet es esta vez menor que en WordNet, como se puede ver en la Figura 5.21, siendo estos un 54% y un 57,60% respectivamente. Incluso el porcentaje de términos relacionados encontrados en ConcepNet es también superior a los hiperónimos e hipónimos de WordNet, pero en cambio el porcentaje de sinónimos encontrados en WordNet con un 36,54% es el doble de los encontrados en ConcepNet, con un 15,62%.

Teniendo también en cuenta la calidad¹ de los sinónimos y términos relacionados generados por ambos servicios decidimos utilizar WordNet para la aplicación definitiva ya los resultados son más satisfactorios y las palabras no difieren demasiado del concepto buscado.

¹Entendiendo por calidad la relación de las palabras devueltas por las dos aplicaciones con la palabra buscada

```
GET http://127.0.0.1:8000/easyHyponym/json/word=inmueble&level=2

Response headers 5 Request headers 0

date: Sun, 05 May 2019 20:03:33 GMT
server: WSGIServer/0.2 CPython/3.7.0
content-type: application/json
x-frame-options: SAMEORIGIN
content-length: 318

[Array[2]
-0: {
    "word": "inmueble"
},
-1: {
    "offsetFather": "spa-30-02913152-n",
    "offset": "spa-30-80000135-n",
    "hyponyms": [Array[19]
        0: "arquitectura",
        1: "centro",
        2: "club",
        3: "residencia",
        4: "hotel",
        5: "casa",
        6: "casa",
        7: "biblioteca",
        8: "ministerio",
        9: "dependencia",
        10: "templo",
        11: "comedor",
        12: "restaurante",
        13: "pista",
        14: "colegio",
        15: "escuela",
        16: "torre",
        17: "templo",
        18: "teatro"
    ],
}
]
```

Figura 5.2: JSON devuelto al buscar los hipónimos fáciles de inmueble

```
GET http://127.0.0.1:8000/easyHyperonym/json/word=inmueble&level=2

Response headers 5
Request headers 0

date: Sun, 05 May 2019 20:12:35 GMT
server: WSGIServer/0.2 CPython/3.7.0
content-type: application/json
x-frame-options: SAMEORIGIN
content-length: 139

[  <> ]
[Array[2]
-0: {
  "word": "inmueble"
},
-1: {
  "offsetFather": "spa-30-02913152-n",
  "offset": "spa-30-04341686-n",
  "hyperonyms": [Array[2]
    0: "construcción",
    1: "estructura"
  ],
}
],
```

Figura 5.3: JSON devuelto al buscar los hiperónimos fáciles de inmueble

GET http://127.0.0.1:8000/metaphor/json/word=inmueble&level=2

Response headers 5 Request headers 0

date: Sun, 05 May 2019 20:23:39 GMT
server: WSGIServer/0.2 CPython/3.7.0
content-type: application/json
x-frame-options: SAMEORIGIN
content-length: 297

[Array[3]]

- 0: {
 "word": "inmueble"
},
- 1: {
 "offsetFather": "",
 "offset": "spa-30-02913152-n",
 "-metaphor": [Array[4]]
 0: "es una casa",
 1: "es una construcción",
 2: "es un edificio",
 3: "son unos edificios"
},
- 2: {
 "offsetFather": "spa-30-02913152-n",
 "offset": "spa-30-04341686-n",
 "-metaphor": [Array[2]]
 0: "es una construcción",
 1: "es una estructura"
},

Figura 5.4: JSON devuelto al buscar las metáforas de la palabra inmueble

```
GET http://127.0.0.1:8000/simil/json/word=inmueble&level=2

Response headers 5 Request headers

date: Sun, 05 May 2019 20:30:00 GMT
server: WSGIServer/0.2 CPython/3.7.0
content-type: application/json
x-frame-options: SAMEORIGIN
content-length: 440

□ □ <> ||| [Array[2]
-0: {
    "word": "inmueble"
},
-1: {
    "offsetFather": "spa-30-02913152-n",
    "offset": "spa-30-80000135-n",
    "simil": [Array[19]
        0: "es una arquitectura",
        1: "es un centro",
        2: "es un club",
        3: "es una residencia",
        4: "es un hotel",
        5: "es una casa",
        6: "es una casa",
        7: "es una biblioteca",
        8: "es un ministerio",
        9: "es una dependencia",
        10: "es un templo",
        11: "es una comedor",
        12: "es una restaurante",
        13: "es una pista",
        14: "es un colegio",
        15: "es una escuela",
        16: "es una torre",
        17: "es un templo",
        18: "es un teatro"
    ],
}
```

Figura 5.5: JSON devuelto al buscar los símiles de la palabra inmueble

The figure shows three separate service interfaces, each with a black header bar containing the service name in white text. Below each header is a red-bordered input field with a placeholder message, a text input field, an 'Enviar' button, and a 'SERVICIO' label. A horizontal line separates the first two services, and another horizontal line separates the second from the third.

SERVICIO 1: Sinónimos y Términos Relacionados
A continuación, introduzca la palabra para obtener sus sinónimos:
Palabra:
Enviar SERVICIO 1
RESULTADOS:

SERVICIO 2: A continuación, introduzca la palabra para obtener sus términos relacionados:
Palabra:
Enviar SERVICIO 2
RESULTADOS:

SERVICIO 3: A continuación, introduzca la palabra para obtener la palabra fácil:
PalabraABuscar:
Profundidad: 1 SERVICIO 3
Enviar SERVICIO 3
RESULTADOS:

Figura 5.6: Interfaz del Servicio Web para ConceptNet

The figure shows a JSON response for the WordNet service. It includes a dropdown menu labeled 'Ver resultados en formato JSON' and the actual JSON data which lists words and their semantic relations.

```
{"sinonimos": ["construcción", "casa"], "hiponimos": ["casa", "centro", "teatro", "ministerio", "club", "escuela"], "hiperonimos": ["construcción", "estructura"]}
```

Figura 5.7: Resultados en formato JSON para el Servicio Web de WordNet

The figure shows a single input field for the WordNet service. It features a black header bar with the service name in white text. Below the header is a red-bordered input field with a placeholder message, a text input field, and an 'Enviar' button.

Sinónimos y Términos Relacionados
A continuación, introduzca la palabra para obtener la palabra fácil:
Word: Enviar

Figura 5.8: Campo de texto para introducir una palabra

Sinónimos y Términos Relacionados

A continuación, introduzca la palabra para obtener la palabra fácil:

Word:

Resultados para la palabra edificio:
SINÓNIMOS QUE COINCIDEN CON LAS MIL PALABRAS RAE:

edificio es construcción
edificio es casa

SINÓNIMOS EN LA APLICACIÓN WORDNET:

edificio es casa	edificio es construcción	edificio es edificación	edificio es edificios	edificio es inmueble
------------------	--------------------------	-------------------------	-----------------------	----------------------

Figura 5.9: Resultados obtenidos para el Servicio Web de WordNet

C ① 127.0.0.1:8000 ☆

Sinónimos y Términos Relacionados

RESULTADOS:

Se han encontrado: 51 sinónimos en 717
Se han encontrado: 122 términos relacionados en 717
No se han encontrado: 565 palabras en 717

Figura 5.10: Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 1000 palabras

C ① 127.0.0.1:8000 ☆

Sinónimos y Términos Relacionados

RESULTADOS:

Se han encontrado: 94 sinónimos en 717
Se han encontrado: 268 términos relacionados en 717
No se han encontrado: 428 palabras en 717

Figura 5.11: Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 5000 palabras

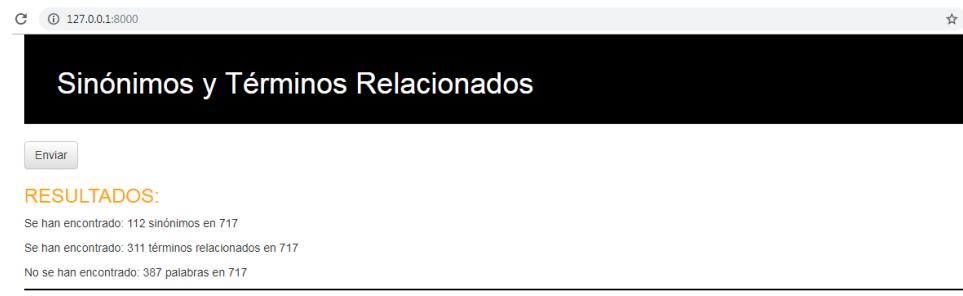


Figura 5.12: Resultados obtenidos para el Servicio Web de ConceptNet con la lista de 10000 palabras

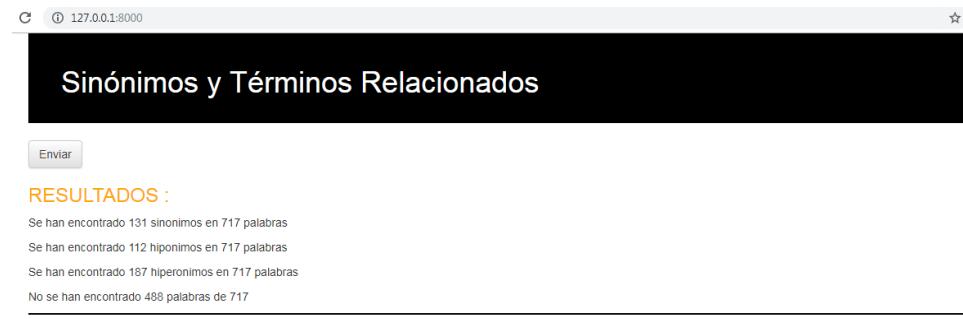


Figura 5.13: Resultados obtenidos para el Servicio Web de WordNet con la lista de 1.000 palabras

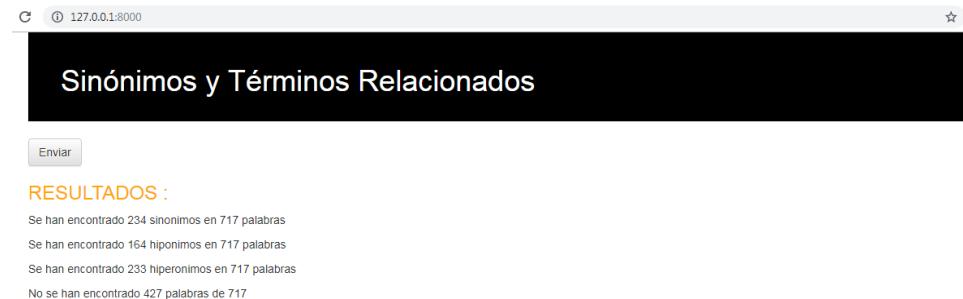


Figura 5.14: Resultados obtenidos para el Servicio Web de WordNet con la lista de 5.000 palabras

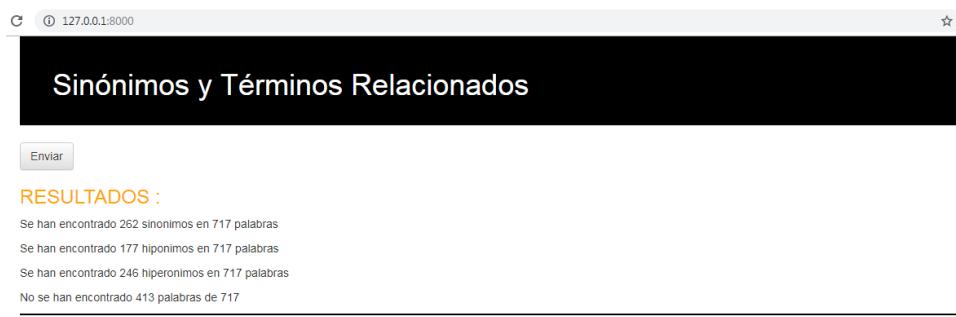


Figura 5.15: Resultados obtenidos para el Servicio Web de WordNet con la lista de 10.000 palabras

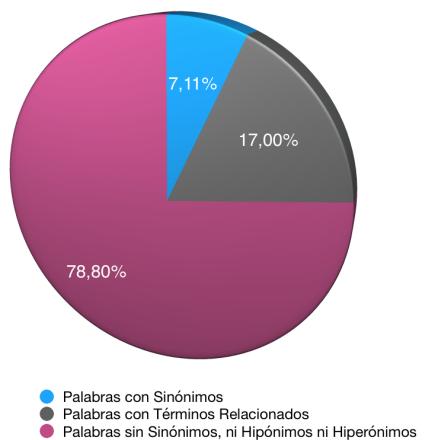


Figura 5.16: Gráfica de los resultados obtenidos para las 1000 palabras con ConcepNet

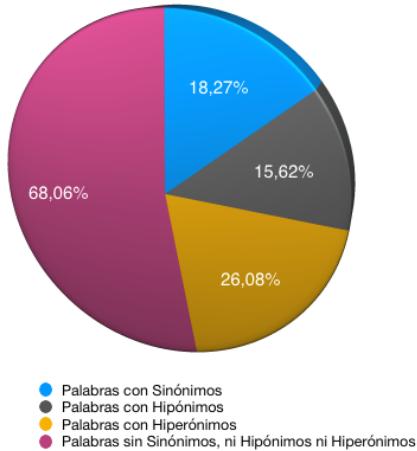


Figura 5.17: Gráfica de los resultados obtenidos para las 1000 palabras con WordNet

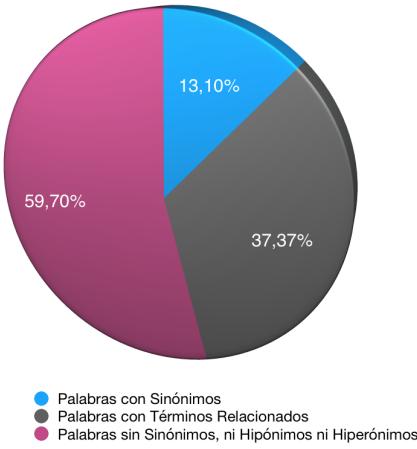


Figura 5.18: Gráfica de los resultados obtenidos para las 5000 palabras con ConcepNet

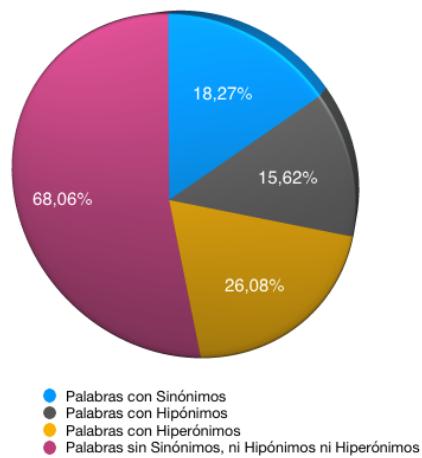


Figura 5.19: Gráfica de los resultados obtenidos para las 5000 palabras con WordNet

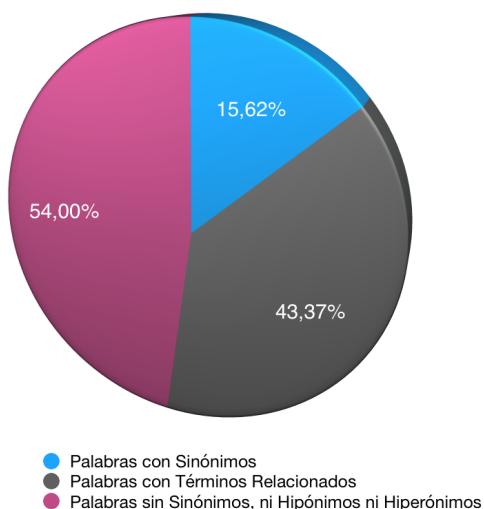


Figura 5.20: Gráfica de los resultados obtenidos para las 10000 palabras con ConcepNet

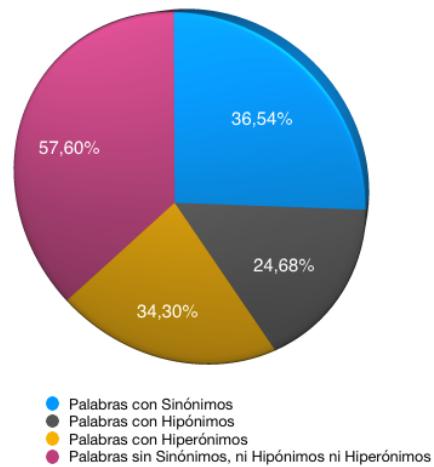


Figura 5.21: Gráfica de los resultados obtenidos para las 10000 palabras con WordNet

Capítulo 6

Aprende Fácil

6.1. Diseño de la Aplicación

Como ya se ha explicado en capítulos anteriores, esta aplicación está creada por y para personas que tienen alguna discapacidad cognitiva, por lo que el diseño de la Interfaz debe estar centrado para este tipo de usuarios, es decir, debe ser lo más sencilla posible, su uso no debe llevar a confusión ni debe hacer sentir al usuario frustrado al usarla. En definitiva, debe cumplir con las Ocho Reglas de Oro del diseño de interfaces, pero simplificándola aún más de lo que podría ser un diseño para otro tipo de aplicación destinada para otro tipo de usuario. Las Ocho Reglas de Oro del diseño de interfaces consisten en que el diseño tenga (Shneiderman y Plaisant, 2006):

- **Consistencia:** La funcionalidad debe ser similar a otras aplicaciones que el usuario está acostumbrado a utilizar. En cuanto a la interfaz debe tener los mismos colores, iconos, formas, botones, mensajes de aviso... Por ejemplo, si el usuario está acostumbrado a que el botón de eliminar o cancelar sea rojo, no debemos añadirle uno de color verde.
- **Usabilidad Universal:** Debemos tener en cuenta las necesidades de los distintos tipos de usuario, como por ejemplo, atajos de teclado para un usuario experto o filtros de color para usuarios con deficiencias visuales.
- **Retroalimentación activa:** Por cada acción debe existir una retroalimentación legible y razonable por parte de la aplicación. Por ejemplo, si el usuario quiere guardar los datos obtenidos de la búsqueda, la aplicación debe informarle de si han sido guardados o no.
- **Diálogos para conducir la finalización:** El usuario debe saber en qué paso se encuentra en cada momento. Por ejemplo, en un proceso de compra que conlleva varios pasos hasta la finalización de la misma, se le debe informar donde se encuentra y cuánto le queda para terminar.

- Prevención de errores: La interfaz debe ayudar al usuario a no cometer errores serios, y en caso de cometerlos se le debe dar una solución lo más clara y sencilla posible. Por ejemplo, deshabilitando opciones o indicando en un formulario el campo en el cual se ha producido el error sin perder la información ya escrita.
- Deshacer acciones fácilmente: Se debe dar al usuario la capacidad de poder deshacer o revertir acciones de una manera sencilla.
- Sensación de control: Hay que dar al usuario la sensación de que tiene en todo momento el control de la aplicación, añadiendo contenidos fáciles de encontrar y de esta forma no causarle ansiedad o frustración por utilizar nuestra aplicación.
- Reducir la carga de memoria a corto plazo: La interfaz debe ser lo más sencilla posible y con una jerarquía de información evidente, es decir, hay que minimizar la cantidad de elementos a memorizar por el usuario.

Teniendo en cuenta estas ocho reglas, la creación del diseño de la Interfaz se ha realizado en dos iteraciones distintas. Una primera iteración competitiva entre los integrantes del grupo y una segunda iteración con expertos del Colegio Estudio3 Afanias.

6.1.1. Primera Iteración: Iteración Competitiva

En esta primera iteración cada integrante del grupo ha realizado su propio diseño de la aplicación. En la realización de estos diseños, los integrantes no podían hablar entre ellos ni comentar las diferentes ideas que tenían para la implementación. De esta forma se consigue que los diseños sean totalmente dispares y que las ideas de uno no provoquen la modificación del diseño del otro y que surgan ideas distintas. Cada integrante realizó cuatro prototipos distintos:

- Un primer prototipo mostrando un único significado, siendo este el más común.
- Un segundo prototipo mostrando el significado más común, añadiendo una definición del concepto.
- Un tercer prototipo mostrando todos los significados de la palabra buscada.
- Un cuarto diseño mostrando todos los significados de la palabra y añadiendo pictogramas¹ de las palabras que facilitan aún más la comprensión del concepto.

¹Se entiende como pictograma un dibujo, imagen o figura que representa el significado de una palabra

En las Figuras 6.1, 6.2, 6.3 y 6.4 se muestran los prototipos creados por Pablo y en las Figuras 6.5, 6.6, 6.7 y 6.8 los prototipos creados por Irene. Una vez que los prototipos estaban terminados, los integrantes de este trabajo se juntaron para hacer una puesta en común y analizar los resultados. A continuación, se explicarán los resultados de los análisis indicando las semejanzas y diferencias que se encontraron. Por lo general, los prototipos de los dos integrantes del grupo eran bastante similares. Las principales diferencias han sido:

- En cuanto al diseño de la interfaz, ambos prototipos integran los mismos elementos: un campo de texto para introducir la palabra, un botón de búsqueda y los resultados se muestran en forma de lista, agrupando los distintos resultados en rectángulos, que a partir de ahora denominaremos fichas. Pablo ha optado por formas rectangulares, tanto para las fichas como para el campo de texto, mientras que Irene utiliza formas redondeadas en todos los elementos. Por otro lado, Pablo implementó un diseño orientado a niños por lo que para el color utilizó azules muy suaves y fondos juveniles con lápices de colores, gomas de borrar, etc... E Irene utilizó un color mostaza, siendo un diseño más simple pero intentando abarcar a un usuario de cualquier edad. Por último en los resultados mostrados en el prototipo de Irene, se muestran en color mostaza indicando de esta forma que pueden clickar sobre él, y se realizará la búsqueda del concepto.
- El orden a la hora de mostrar los resultados es distinto, en los prototipos de Pablo se muestran primero varios sinónimos e hiperónimos (*Un vehículo es una máquina o Un vehículo es un transporte*), después los hipónimos (*Un vehículo es como un taxi*) y finalmente los hipónimos pero añadiendo un adjetivo que representa una característica del resultado devuelto (*Un vehículo es rápido como un caballo*). Sin embargo, en los prototipos de Irene se muestra únicamente un resultado para cada tipo y en el siguiente orden: Primero un hipónimo pero añadiendo un adjetivo que representa una característica del resultado devuelto (*Un vehículo es tan rápido como un caballo*), después un sinónimo o hipérónimo (*Un vehículo es una coche*) y por último un hipónimo (*Un vehículo es como una ambulancia*).
- Los prototipos de Pablo además incluyen un ejemplo siempre visible y justo debajo de los resultados mostrados ("Él necesita un coche para ir a trabajar"), para facilitar al usuario su comprensión, mientras que en los prototipos de Irene muestra únicamente la definición ("Vehículo motorizado de cuatro ruedas por lo general impulsado por un motor de combustión interna"), y se encuentra dentro de un botón, dando a elegir al usuario la decisión de poder verla o no.

- Pablo decidió mostrar la lista de los resultados comenzando con “Un X es” englobando así los resultados para un mismo significado. En caso de existir varias acepciones del concepto, por cada uno de ellos añade “O también puede ser....”, dejando así claramente dividido los distintos significados que pudieran existir para el concepto buscado. En cambio Irene, añade un título al principio (“Resultados para la palabra X”) donde queda claro que los resultados que se obtienen son para dicho concepto y además añade delante de cada resultado “Un X es” seguido del sinónimo, hiperónimo o hipónimo.
- Para la numeración de los resultados obtenidos, Pablo incluye dentro de cada ficha un listado numérico e Irene únicamente añade un número a la ficha.
- En el prototipo donde se incluyen también los pictogramas, Pablo añade un pictograma que hace referencia al concepto según en qué contexto se utilice. Por ejemplo, un vehículo puede hacer referencia a un coche o puede ser un medio para llegar o lograr un fin, e Irene además de incluir el mismo pictograma que Pablo, añade pictogramas a cada resultado. Por ejemplo, en la frase “Un portero es un vigilante” Irene añadió el pictograma de vigilante.

Tanto en el prototipo diseñado por Pablo como en el de Irene se han utilizado los pictos de ARASAAC².

Una vez analizados los prototipos de ambos integrantes nos reunimos con los directores de dicho trabajo y se decidió crear un prototipo con las siguientes características:

- Para el diseño de la interfaz se eligió el implementado por Irene.
- Todos los prototipos deben tener el mismo ejemplo, por lo que se cambiaron para la palabra Portero.
- El orden de los resultados se modificó, haciendo que primero aparezcan los sinónimos e hiperónimos (“Un portero es un vigilante”), después los hipónimos (“Un portero es como un guardia) y por último hipónimos que además tuviesen un adjetivo que permita realizar una comparación (“Un portero es tan ágil como un pájaro”).
- Eliminar la palabra *tan* en las analogías.
- Modificación del título de la aplicación por uno que estuviese en español, por lo que se cambió el nombre de “*Easy Words*” por “Aprende Fácil”.
- Añadir más de tres resultados por ficha.

²<http://www.arasaac.org/>

- Si solo aparece un resultado, eliminar el número de la ficha.
- Añadir un pictograma a la característica que relaciona el concepto con el resultado. Por ejemplo, en la frase *Un portero es tan fuerte como un gorila* añadir el pictograma correspondiente a fuerte.
- Añadir el ejemplo del prototipo de Pablo junto con la definición, en este caso *Mi tío trabaja de portero en una discoteca*.
- Crear un nuevo prototipo donde la definición y el ejemplo se muestre a primera vista, y el usuario no tenga que estar pinchando en ningún botón.
- Crear otro prototipo donde la definición y el ejemplo estén separados. Así el usuario puede elegir lo que quiere ver, si solo una cosa o ambas.

Con todas estas decisiones, se crearon los prototipos que se muestran en las Figuras 6.9, 6.10, 6.11, 6.12, 6.13, 6.14 y que fueron los que se usaron para en la siguiente iteración donde nos reunimos con los expertos para que nos dieran su opinión sobre la aplicación, y nos ayudarán a decidir sobre los siguientes aspectos:

- Mostrar únicamente el *synset* más común o que en su lugar apareciesen todos los resultados.
- Añadir definición y ejemplo o no. Y en caso de añadirlos, mostrarlos juntos en un mismo botón, en distintos botones o sin botón y que aparezcan debajo de los resultados.
- Añadir pictogramas y en caso de añadirlos, solamente el pictograma global o también en cada resultado.
- La forma de mostrar los resultados es correcta, o prefieren otra disposición.

6.1.2. Segunda Iteración: Evaluación con Expertos

El día 26 de Marzo de 2019 a las 09:00h nos reunimos con la directora y los profesores del colegio Estudio3 Afanias³ situado en la Comunidad de Madrid. Este colegio atiende a niños y jóvenes con discapacidad intelectual entre los 3 y 21 años de edad.

Una vez expuestos los distintos prototipos, nos dieron la enhorabuena y nos hicieron partícipes de la gran ayuda que supondría esta herramienta. A continuación nos dieron su opinión sobre distintos aspectos que se podrían modificar de los prototipos:

³<https://afanias.org/que-hacemos/educacion/colegio-estudio-3/>

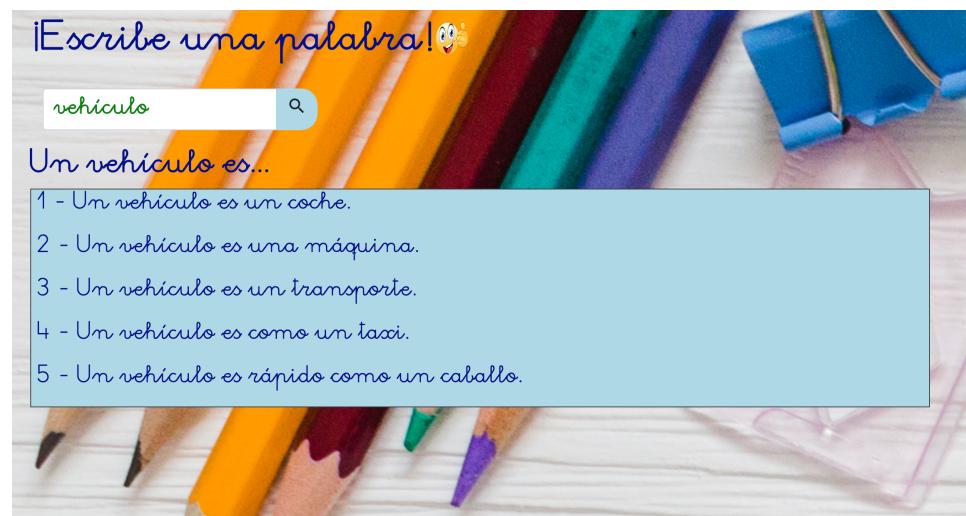


Figura 6.1: Prototipo de Pablo mostrando resultado más común para la palabra vehículo

- Modificar el color amarillo-mostaza de los textos por un color más oscuro que contraste más con el fondo blanco.
- El tipo de letra debe ser Arial o Script, ya que son las letras con las que los alumnos están familiarizados y las que mejor entienden.
- Añadir un reproductor que lea la frase, haciendo así que en caso de no poder leer los resultados, pueda ayudarles la voz.
- Incluir la opción de poder ver un video por si los pictos no ayudan al usuario a entender el concepto.
- Introducción de distintas personalizaciones:
 - Búsqueda en tres niveles: sencillo, medio y amplio. El nivel sencillo sería realizando la búsqueda de las palabras fáciles en las 1.000 palabras más usadas de la RAE, el medio con las 5.000 palabras más usadas de la RAE y el amplio con las 10.000 palabras más usadas de la RAE.
 - Uso de mayúsculas en todos los textos para hacer la aplicación más accesible a aquellos alumnos que no entienden los textos en minúsculas.
 - Que el usuario elija si quiere que aparezca la definición y el ejemplo o no.
 - Poder elegir si aparecen los pictos o no. En caso de que el usuario decida que sí deben aparecer, estos deben situarse debajo de la

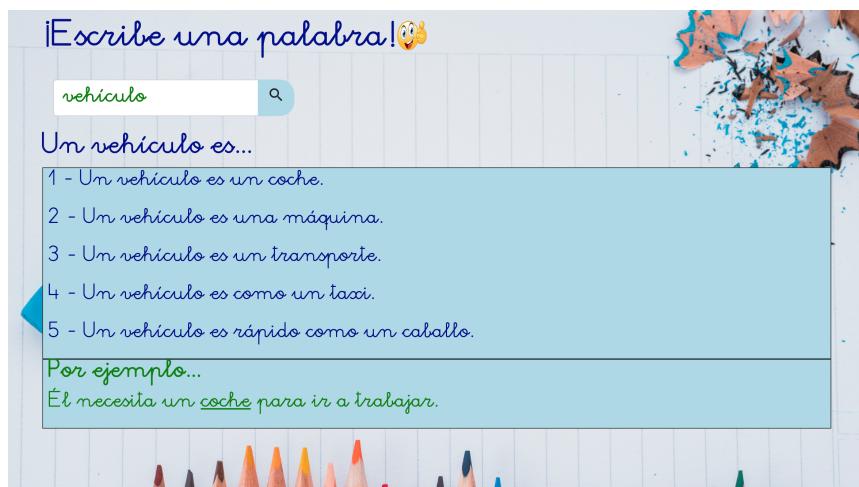


Figura 6.2: Prototipo de Pablo mostrando resultado más común para la palabra vehículo y un ejemplo

palabra y no al lado, ya que los expertos comentaron que esto puede llevar a confusión a los alumnos pensando que sería otra palabra más para leer.

Teniendo en cuenta las observaciones de los expertos se creó el diseño final de la aplicación.

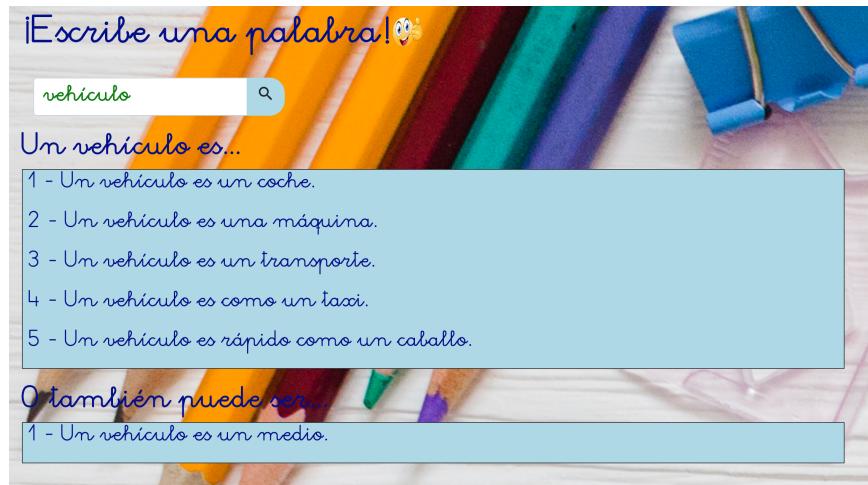


Figura 6.3: Prototipo de Pablo mostrando todos los resultados para la palabra vehículo



Figura 6.4: Prototipo de Pablo mostrando todos los resultados para la palabra vehículo junto con pictogramas



Figura 6.5: Prototipo de Irene mostrando resultado más común para la palabra vehículo

Easy Words

Introduzca una palabra: Vehículo

Resultados para la palabra Vehículo:

1.

Un vehículo es tanto rápido como un caballo

Un vehículo es un coche

Un vehículo es como una ambulancia

Definición ▾

Vehículo motorizado de cuatro ruedas por lo general impulsado por un motor de combustión interna.

Figura 6.6: Prototipo de Irene mostrando resultado más común para la palabra vehículo y una definición



Figura 6.7: Prototipo de Irene mostrando todos los resultados para la palabra portero

Easy Words

Introduzca una palabra: Portero 

Resultados para la palabra Portero:

1.  Un portero es tán fuerte como un gorila 

Un portero es un vigilante 

Un portero es como un guardia 

2.  Un portero es tán ágil como un pajaro 

Un portero es un jugador 

Un portero es como un futbolista 

Figura 6.8: Prototipo de Irene mostrando todos los resultados para la palabra portero junto con pictogramas



Figura 6.9: Prototipo mostrando resultado más común para la palabra portero

Aprende Fácil

Introduzca una palabra: Portero

Resultados para la palabra Portero:

Un portero es un **vigilante**
Un portero es como un **guardia**
Un portero es:
 fuerte como un gorila
 grande como un oso

Definición ▾ Ejemplo ▾

Z

Figura 6.10: Prototipo Versión 1 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo separados



Figura 6.11: Prototipo Versión 2 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo a simple vista

The screenshot shows a user interface titled "Aprende Fácil" in a yellow header bar. Below it is a search bar with the placeholder "Introduzca una palabra:" and a search button containing a magnifying glass icon. The search term "Portero" is entered into the search bar. The results section is titled "Resultados para la palabra Portero:". It displays several definitions and examples. One definition is expanded, showing a blue button labeled "Definición y Ejemplo ▾". Below this button, the expanded content includes the definition "Un portero es alguien que protege una entrada." and the example "Mi tío trabaja de portero en una discoteca.". Large, stylized letters "a" and "Z" are visible on either side of the results area.

Aprende Fácil

Introduzca una palabra: Portero

Resultados para la palabra Portero:

Un portero es un **vigilante**
Un portero es como un **guardia**
Un portero es:
 fuerte como un gorila
 grande como un oso

Definición y Ejemplo ▾

Definición: *Un portero es alguien que protege una entrada.*
Ejemplo: *Mi tío trabaja de portero en una discoteca.*

a Z

Figura 6.12: Prototipo Versión 3 mostrando resultado más común para la palabra portero junto con una definición y un ejemplo en un mismo botón



Figura 6.13: Prototipo mostrando todos los resultados para la palabra portero

Aprende Fácil

Introduzca una palabra: Portero 

Resultados para la palabra Portero:

1.

Un portero es un **vigilante** 

Un portero es como un **guardia** 

Un portero es:
fuerte  **como un gorila** 
grande  **como un oso** 

2.

Un portero es un **jugador** 

Un portero es como un **futbolista** 

Un portero es:
rápido  **como un caballo** 
delgado  **como un flamenco** 

Figura 6.14: Prototipo mostrando todos los resultados para la palabra portero incluyendo los pictogramas

Capítulo

7

Trabajo Realizado

En este capítulo vamos a describir que trabajo hemos hecho cada uno

7.1. Trabajo realizado por Irene

Primero investigué las bibliotecas que utilizaremos para el procesado de las palabras, al principio encontramos una biblioteca para el procesado de texto en Python, que es la nltk pero vimos que las etiquetas que ponía a las palabras no eran del todo correctas por lo que buscamos otra biblioteca y encontramos Spacy, con esta ya pudimos etiquetar bien todas las palabras diseñando un programa inicialmente en el Jupyter. A continuación, investigué que tecnologías utilizar para la realización del prototipo tecnológico, encontramos como entorno de desarrollo Pycharm y como framework Django. Una vez seleccionadas las tecnologías, investigamos como se utilizaban y nos pusimos a trabajar en el prototipo tecnológico.

Yo me encargué de conectar las vistas html con la lógica en Python, a continuación vimos como se implementaba un formulario y como se hacia una redirección a vista. Cuando supimos como se hacia todo esto, integraron el código desarrollado en Jupyter en nuestro servicio web finalizando el prototipo tecnológico.

En cuanto a la memoria me la dividí a partes iguales con Pablo, intentando que los dos toquemos todo, por lo que ambos redactamos tanto una parte de la introducción (en la que redacté la motivación) como el estado de la cuestión (yo hice el apartado de lectura fácil y Procesamiento del Lenguaje Natural).

La investigación de como funcionaba Conceptnet y su API la hicimos conjuntamente.

7.2. Trabajo realizado por Pablo

Al igual que mi compañera, lo primero que hicimos fue investigar como podíamos etiquetar las palabras, encontramos la librería nltk de Python para hacerlo, pero tras un primer intento nos dimos cuenta de que muchas palabras no estaban etiquetadas como deberían por lo que decidimos buscar alternativas, indagando un poco encontramos Spacy, la probamos y obtuvimos unos resultados mucho mejores que con nltk por lo que decidimos utilizar esta última (todo esto lo hicimos desde el Jupyter).

Cuando terminamos de etiquetar las palabras nos pusimos a investigar herramientas para el desarrollo del prototipo tecnológico y nos decantamos por utilizar Django como framework integrado en Pycharm, que es el entorno de desarrollo.

A continuación empezamos el desarrollo del prototipo tecnológico primero investigando como se utilizaban estas tecnologías(implementar formularios, hacer las redirecciones a vista...). Para finalizar migramos lo hecho desde Jupyter a nuestro servicio web.

Irene y yo nos dividimos la redacción de la memoria de tal manera que los dos hicimos tanto la parte de la introducción como del estado de la cuestión, de la introducción a mí me tocó la parte de los objetivos y del estado de la cuestión la parte de figuras retóricas y servicios web.

La investigación de como funcionaba Conceptnet y su API la hicimos de manera conjunta.

Capítulo **8**

Conclusiones y Trabajo Futuro

Conclusiones del trabajo y líneas de trabajo futuro.

Chapter 8

Conclusions and Future Work

Conclusions and future lines of work.

Apéndice A

Título

Contenido del apéndice

Apéndice **B**

Título

Bibliografía

*Y así, del mucho leer y del poco dormir,
se le secó el celebro de manera que vino
a perder el juicio.*

Miguel de Cervantes Saavedra

CALLEJA, P. G. *Generación de recursos lingüísticos mediante la extracción de relaciones entre conceptos*. Trabajo de fin de máster, Universidad Complutense de Madrid, 2017.

GALIANA, A. y CASAS, J. *Introducción al análisis retórico: tropos, figuras y sintaxis del estilo*. Universidad de Santiago de Compostela, 1994.

GARCÍA MUÑOZ, O. *Lectura Fácil: Métodos de redacción y evaluación*. Real Patronato sobre Discapacidad, Ministerio de Sanidad, Servicios Sociales e Igualdad, 2012.

MANNING, C. D. y SCHÜTZE, H. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999. ISBN 0-262-13360-1.

MORENO ORTIZ, A. *Diseño e implementación de un lexicón computacional para lexicografía y traducción automática*. Facultad de Filosofía y Letras., 2000.

NOMURA, M., SKAT NIELSEN, G. y TRONBACKE, B. *Directrices para materiales de lectura fácil*. Federación Internacional de Asociaciones e Instituciones Bibliotecarias, 2010.

QUILLIAN, M. R. *Semantic Memory*. Cambridge, 1968.

SHNEIDERMAN, B. y PLAISANT, C. *Diseño de interfaces de usuario : estrategias para una interacción persona-computadora efectiva*. Pearson Education, 2006.

- SOWA, J. *Conceptual structures: Information processing in mind and machine.* Addison-Wesley Longman Publishing Co., Inc., 1983. ISBN 0-201-14472-7.
- TORRES, J. *Servicios Web.* Universidad Complutense de Madrid, 2017.
- VEALE, T. y LI, G. *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity.* Bllomsbury Academic, 2012.
- VEALE, T. y LI, G. *Creating Similarity: Lateral Thinking for Vertical Similarity Judgment.* In Proceedings of ACL 2013, the 51st Annual Meeting of the Association for Computational Linguistics, 2013.
- VEGA LEBRÚN, C. A. *Integración de herramientas de tecnologías de información como soporte en la administración del conocimiento.* Trabajo doctorado, Universidad Popular Autónoma del Estado de Puebla, 2005.

*-¿Qué te parece desto, Sancho? – Dijo Don Quijote –
Bien podrán los encantadores quitarme la ventura,
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

*-Buena está – dijo Sancho –; fírmela vuestra merced.
–No es menester firmarla – dijo Don Quijote–,
sino solamente poner mi rúbrica.*

*Primera parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

