
Mejora de la Comprensión Lectora mediante Analogías para la Inclusión



Trabajo de Fin de Grado
Curso 2017–2018

Autor

Irene Martín Berlanga
Pablo García Hernández

Director

Virginia Francisco Gilmartín
Gonzalo Rubén Mendez 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

9 de diciembre de 2018

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

9 de diciembre de 2018

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	4
1.3. Estructura de la memoria	5
2. Estado de la Cuestión	7
2.1. Lectura Fácil	7
2.2. Procesamiento del Lenguaje Natural	9
2.2.1. ConceptNet	10
2.2.2. Thesaurus	14
2.2.3. Thesaurus Rex	15
2.2.4. Metaphor Magnet	15
2.2.5. Wordnet	15
2.2.6. EuroWordnet	16
2.3. Figuras retóricas	16
2.4. Servicios Web	17
2.4.1. Arquitectura Servicios Web	18
2.4.2. Ventajas de los Servicios Web	19
2.4.3. Desventajas de los Servicios Web	19
3. Trabajo Realizado	27
3.1. Trabajo realizado por Irene	27
3.2. Trabajo realizado por Pablo	28
4. Conclusiones y Trabajo Futuro	31
4. Conclusions and Future Work	33

A. Título	35
B. Título	37
Bibliografía	39

Índice de figuras

2.1. Logo Lectura Fácil	8
2.2. Ejemplo Red Semántica	10
2.3. Página principal Concepnet	11
2.4. Resultados de búsqueda de una palabra	12
2.5. Vista de JSON	13
2.6. Resultados búsqueda Thesaurus con la palabra <i>casa</i>	14
2.7. Ejemplo formato XML Thesaurus	14
2.8. Ejemplo formato JSON Thesaurus	15
2.9. Resultados búsqueda Thesaurus Rex con la palabra <i>house</i>	20
2.10. Ejemplos formato XML Thesaurus Rex	21
2.11. Resultados búsqueda Metaphor Magnet con la palabra <i>house</i>	22
2.12. Ejemplos formato XML Metaphor Magnet	23
2.13. Resultados búsqueda Wordnet con la palabra <i>house</i>	24
2.14. Resultados búsqueda EuroWordnet con la palabra <i>casa</i>	25
2.15. Estructura de un mensaje SOAP	25

Índice de tablas

Chapter 1

Introduction

Introduction to the subject area.

Capítulo 1

Introducción

En este capítulo vamos a introducir nuestro trabajo, de por qué es importante que se lleve a cabo y los resultados que esperamos conseguir, para ello, este apartado lo hemos dividido en varias partes, en el punto 1.1 se tratará la motivación del trabajo, es decir, los problemas que hay actualmente que queremos ayudar a solucionar, también, queremos alcanzar una serie de objetivos, tanto a nivel personal como técnico, esto lo trataremos en el apartado 1.2. Por último queremos describir la estructura que tendrá este documento para que sea más fácil su lectura, esto lo haremos en el apartado 1.3.

1.1. Motivación

En nuestra sociedad actual, existen ciertos colectivos como pueden ser las personas con algún tipo de trastorno cognitivo, inmigrantes, ancianos, analfabetos funcionales, niños, etc... que tienen dificultad para aprender conceptos complejos o no tan complejos. Existen multitud de palabras cuyo significado es bastante complicado de explicar de una manera sencilla, por lo que una solución para que cualquier persona lo pueda comprender es hacer uso de metáforas o analogías. De esta forma, se puede asimilar el concepto de una manera más rápida y sencilla.

Esto supone una serie de limitaciones en su vida cotidiana, en la forma de relacionarse con otros individuos, en su vida profesional e incluso su vida personal. Por ejemplo, suponemos que un inmigrante necesita realizar gestiones para legalizar su estancia o firmar documentación. Si no tiene una buena comprensión del lenguaje no sabe las consecuencias de firmar los documentos porque no los entiende. Otro ejemplo que podríamos añadir sería cuando una persona firma un contrato, como por ejemplo, una póliza de seguros o un contrato de trabajo, en el cual añaden ciertas cláusulas que son

incomprensibles.

Para ayudar principalmente a estas personas a que puedan entender el significado de cualquier palabra, y de esta forma superar algunas de sus limitaciones, vamos a desarrollar una aplicación que permita definir palabras complejas mediante comparaciones con otras más fáciles ya conocidas por los usuarios. Por ejemplo, si queremos explicar una palabra compleja como puede ser *piraña*, podemos describir sus características utilizando conceptos simples para facilitar su entendimiento de la siguiente manera:

Una piraña nada como un pez y es agresiva como un león

Mediante esta comparación, alguien que desconozca completamente el significado de *piraña*, puede hacerse una idea muy aproximada de lo que es.

1.2. Objetivos

Nuestro objetivo principal es crear una aplicación que dada una palabra compleja para el usuario, obtenga una definición clara y sencilla mediante símiles, analogías o metáforas. Esto se puede ver claramente en el ejemplo de la *piraña* del apartado anterior, ya que de esa manera, cualquier persona que no sepa lo que es una *piraña*, puede hacerse una idea de como es dicho animal y asimilar el nuevo concepto. Además, utilizaremos técnicas centradas en el usuario para diseñar una interfaz lo más usable posible y que así el usuario tenga una experiencia de uso satisfactoria. Por último, queremos que nuestro producto se encuentre al alcance de todas aquellas personas que lo necesiten. Para que todo ello ocurra, los objetivos principales a alcanzar son:

- Desarrollar una aplicación que esté al alcance de todos los usuarios, es decir, ninguna persona puede quedarse excluida de su uso por ningún motivo, debe ser una aplicación sencilla, fácil de usar y entender.
- La aplicación estará construida con servicios web que la doten de funcionalidad.
- Los servicios web desarrollados estarán disponibles en una API pública para que todo el mundo pueda utilizarlos.
- La aplicación funcione con un amplio número de palabras complejas para que tenga la mayor utilidad posible.
- La aplicación se construirá de manera incremental, añadiéndole valor al producto poco a poco.
- Construir una aplicación que esté centrada en el usuario, que sea lo más amigable posible y que solucione problemas reales de una forma usable.

- La aplicación tenga una apariencia atractiva y amigable con el usuario.

Alcanzando los objetivos anteriormente descritos, conseguiremos obtener un producto de calidad, con una gran utilidad tanto social como académica, que puede ayudar a mucha gente a aprender nuestro idioma de una manera más fácil.

Por último, no nos tenemos que olvidar de los objetivos académicos de este trabajo: Poner en práctica los conocimientos adquiridos durante el Grado, y ampliar nuestros conocimientos gracias a la utilización de herramientas, lenguajes y metodologías nuevas.

1.3. Estructura de la memoria

La memoria se dividirá en varios capítulos. El contenido de los mismos es el siguiente:

- El **capítulo uno** (capítulo actual) contiene la introducción.
- El **capítulo dos** es 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, ventajas y desventajas de su uso.
- El **capítulo tres** describe el trabajo realizado por cada uno de los autores.

Capítulo 2

Estado de la Cuestión

En este capítulo vamos a abordar los fundamentos principales sobre los que se construirá nuestro trabajo, como lo que nosotros queremos es una herramienta que ayude a colectivos con dificultades con el aprendizaje, lo primero que trataremos en el primer apartado será la lectura fácil(2.1) ya que está directamente relacionada con nuestro proyecto. Como además, necesitamos aplicar Inteligencia Artificial, para buscar palabras relacionadas con las nuestras, trataremos el tema del Procesamiento del Lenguaje Natural en el siguiente punto (2.2) así como algunas herramientas que hemos valorado utilizar. A continuación, vamos a hablar de las figuras retóricas(2.3) que utilizaremos para construir nuestras comparaciones cuando el usuario quiera buscar la definición de una palabra compleja. Y para terminar hablaremos de que és nuestra aplicación (2.4) en este caso un servicio web, describiendo los tipos que hay, como se utilizan y que ventajas y desventajas tiene su uso.

2.1. Lectura Fácil

Se llama lectura fácil 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 la adaptación de textos, ilustraciones y maquetaciones que permite una mejor lectura y comprensión. Nosotros nos vamos a centrar en la lectura fácil aplicada a textos.

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



Figura 2.1: Logo Lectura Fácil

en la web. En el año 2013, en México se produce la primera sentencia judicial en lectura fácil ¹.

En la actualidad, podemos distinguir los documentos en lectura fácil gracias al logo de la figura 2.1:

Los documentos escritos en Lectura Fácil (Wikipedia, 2018a) son documentos de todo tipo que siguen las directrices internacionales de la IFLA (International Federation of Library Associations and Institutions) y de Inclusion Europe en cuanto al contenido y la forma. Algunas pautas a seguir para escribir correctamente un texto en Lectura Fácil son (Óscar García Muñoz, 2012):

- Evitar mayúsculas fuera de la norma, es decir, escribir en mayúsculas sólo cuando toca según 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 /.
- 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.

¹<https://dilofacil.wordpress.com/2013/12/04/el-origen-de-la-lectura-facil/>

- 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.
- Evitar frases superiores a 60 caracteres.
- Si el documento está paginado, incluir la paginación claramente y reforzar el mensaje de que la información continúa en la página siguiente.

Estos son solo algunos de los aspectos que se deben tener en cuenta para realizar correctamente un documento en lectura fácil.

Debemos también hacer hincapié en la distinción entre palabras fáciles y complejas, puesto que son de gran importancia para la lectura fácil. Las palabras complejas son aquellas que no se utilizan a menudo en nuestra sociedad, como por ejemplo: Meliflúo o Inefable.

Es por ello que este tipo de palabras deben estar totalmente descartadas en la lectura fácil, y en su lugar debemos introducir palabras fáciles, que son aquellas que se utilizan asiduamente. La RAE dispone de un documento con las mil palabras más usadas ².

2.2. Procesamiento del Lenguaje Natural

El Procesamiento del Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial que se encarga de investigar la manera de comunicar máquinas con personas mediante el uso de lenguajes naturales (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). El Procesamiento del Lenguaje Natural se ayuda de las redes semánticas, puesto que estas son una forma de representación del conocimiento lingüístico en la que los conceptos y sus interrelaciones se representan mediante un grafo. Se utilizan para representar mapas conceptuales y mentales (Wikipedia, 2018b). Los nodos están representados por el elemento lingüístico, y la relación entre los nodos sería la arista, como podemos ver en la figura 2.2. Existen principalmente tres tipos de redes semánticas ³:

²<http://corpus.rae.es/lfrecuencias.html>

³<http://elies.rediris.es/elies9/4-3-2.htm>

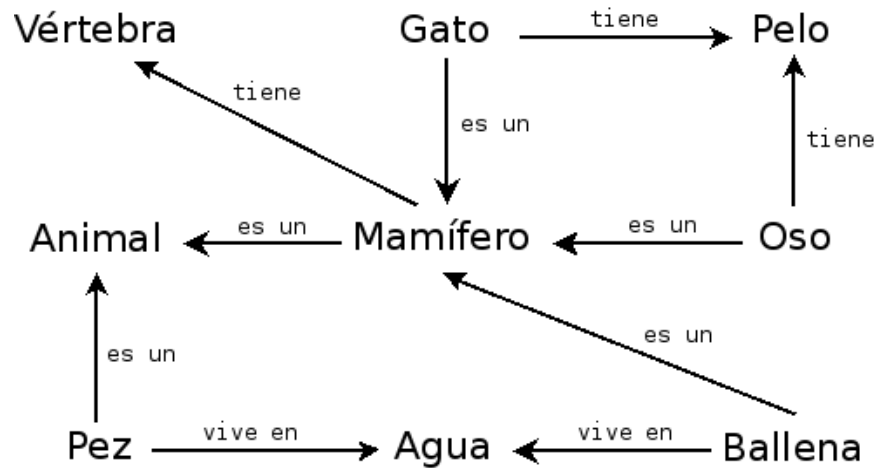


Figura 2.2: Ejemplo Red Semántica

- 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.
- Grafos Conceptuales: Existen dos tipos de nodos: nodos de conceptos y de relaciones. 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. Es por ello que pueden ser de dos tipos:
 - Nodos de conceptos: Pueden representar una entidad, así como un estado o proceso.
 - Nodos de relación: Indican como se relacionan los nodos.
- 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.

Para el trabajo que queremos realizar, existen multitud de páginas web que son redes semánticas capaces de procesar el Lenguaje Natural. A continuación, hablaremos de algunas de ellas.

2.2.1. ConceptNet

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. Esta disponible en múltiples idiomas, como el español, el inglés o el chino. ConceptNet dispone de una aplicación web ⁴,

⁴<http://conceptnet.io/>

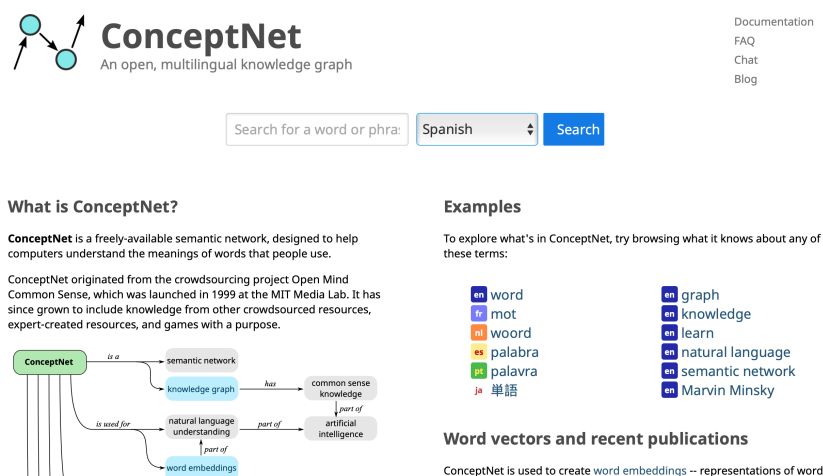


Figura 2.3: Página principal Conceptnet

como se puede ver en la figura 2.3, donde seleccionas el idioma deseado y añades la palabra a buscar. En la figura 2.4, podemos ver que si añadimos la palabra Chaqueta nos devuelve una cantidad de columnas, distinguiéndolas por sinónimos, términos relacionados, símbolos, etc... Principalmente, lo que a nosotros nos interesa son los sinónimos y términos relacionados. Por otro lado, Conceptnet dispone de un servicio web ⁵ que devuelve los resultados en formato JSON. Siguiendo con el mismo ejemplo anterior, podemos ver en la figura 2.5 los resultados en dicho formato:

Este consta de cuatro campos ⁶:

- @context: URL enlazada a un archivo de información del JSON para 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 nuestro caso `/a/[r/Synonym]/c/es/chaqueta/n/,/c/es/americana/`, nos indica que la palabra *americana* es un sinónimo de *chaqueta*.

⁵<http://api.conceptnet.io>

⁶[principaleshttps://github.com/commonsense/conceptnet5/wiki/AP](https://github.com/commonsense/conceptnet5/wiki/AP)

es chaqueta
A Spanish term in ConceptNet 5.6
Sources: English Wiktionary, French Wiktionary, and Open Multilingual WordNet
View this term in the API

Documentation
FAQ
Chat
Blog

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

Figura 2.4: Resultados de búsqueda de una palabra

- @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.
- edges: Representa una estructura de datos devueltos por conceptnet compuesta por:
- 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: Frase más completa, donde adquiriera 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 por qué 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


```

{
  "@id": "/a[/r/Synonym/,/c/es/chaqueta/n/,/c/es/americana/]",
  "@type": "Edge",
  "dataset": "/d/wiktionary/fr",
  "end": {
    "@id": "/c/es/americana",
    "@type": "Node",
    "label": "americana",
    "language": "es",
    "term": "/c/es/americana"
  },
  "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
},

```

Figura 2.5: Vista de JSON

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: Indica de que frase del lenguaje natural se han extraído los datos que estan guardados en conceptnet. En nuestro caso es 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.
- 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

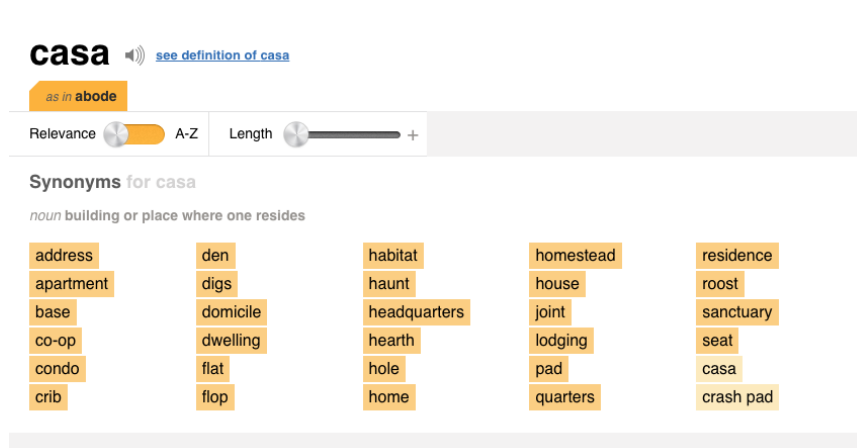


Figura 2.6: Resultados búsqueda Thesaurus con la palabra *casa*

```
<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> tratado de paz | pacification | acuerdo | pact | accord </synonyms> </list>
</response>
```

Figura 2.7: Ejemplo formato XML Thesaurus

2.2.2. Thesaurus

Es una aplicación web ⁷ la cuál se define como principal diccionario de sinónimos de la web. Su nombre proviene de la palabra griega Tesauro, cuyo significado es *Trabajo de referencia que enumera las palabras agrupadas según la similitud del significado*. Esta página ofrece la posibilidad de introducir una palabra para poder conocer sus sinónimos pero como podemos ver en la imagen 2.6, solamente devuelve resultados en inglés. A parte del listado de sinónimos, Thesaurus te dice que tipo de palabra es y una definición de la misma. Dicha aplicación⁸ proporciona una API tipo RESTful que obtiene los sinónimos de una palabra según el idioma específico y 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 |. Podemos ver en la imagen 2.7 un ejemplo de como sería el resultado de una petición en formato XML y en la imagen 2.8 en formato JSON.

⁷<https://www.thesaurus.com/>

⁸<http://thesaurus.altervista.org/>

```
{
  "response": [
    {
      "list": {
        "category": "(noun)",
        "sinónimos": "order | war (antonym)"
      },
      "list": {
        "category": "(noun)",
        "sinónimos": "armonía | concordia | concordancia"
      },
      "lista": {
        "categoria": "(sustantivo)",
        "sinónimos": "seguridad pública | seguridad"
      },
      "lista": {
        "categoria": "(sustantivo)",
        "sinónimos": "tratado de paz | pacificación | tratado | pacto | acuerdo"
      }
    ]
  }
}
```

Figura 2.8: Ejemplo formato JSON Thesaurus

2.2.3. Thesaurus Rex

Thesaurus Rex ⁹ es una distribución de la aplicación Thesaurus comentada en el apartado anterior. Es una red semántica en la cuál puedes introducir una palabra y te devuelve palabras relacionados o bien, puedes introducir un par de palabras (separadas por el operador &) y te devolverá las categorías que comparten ambos términos. Thesaurus Rex solo admite palabras en inglés por lo que en nuestro proyecto no puede ser utilizado. En la imagen 2.9 podemos ver los resultados de haber introducido la palabra *house*. La aplicación devuelve estos resultados en formato XML, y este como se puede ver en la imagen 2.10 los divide en distintos campos: *Categories*, *Modifiers* y *CategoryHeads*.

2.2.4. Metaphor Magnet

Otra aplicación web ¹⁰, en la cuál introduces una palabra pero esta vez devuelve metáforas. El objetivo (Veale.T y Li.G, 2012) de dicha aplicación es encontrar metáforas comunes en los n-gramas de Google y utilizar estos mapeos para interpretar metáforas, pero esta aplicación está limitada ya que solamente se puede buscar palabras en inglés como podemos observar en la imagen 2.11 introduciendo la palabra *house*. Esta consulta devuelve un fichero XML como el expuesto en la figura 2.12 que puede ser utilizado para otras aplicaciones de Procesamiento de Lenguaje Natural.

2.2.5. Wordnet

Es una base de datos léxica solamente en inglés donde agrupa las palabras en función de su significado (University, 2010). Utiliza distintos tipos de palabras tales como sustantivos, verbos, adjetivos y adverbios ignorando preposiciones, determinantes y otras palabras funcionales. Agrupando estos tipos en conjuntos de sinónimos cognitivos o también llamados *synsets*. Estos, aparte contienen una breve definición y en muchas ocasiones oraciones cortas que explican el significado. En el caso de que la palabra tenga distintos significados se representará en tantos *synsets* como significados haya. Podemos ver un ejemplo en la figura 2.13 realizando una búsqueda con la palabra

⁹<http://ngrams.ucd.ie/therex3/>

¹⁰<http://ngrams.ucd.ie/metaphor-magnet-acl/>

house. WordNet dispone de una API HTTP donde permite buscar información en la base de datos de WordNet mediante métodos GET y POST a través de la URL raíz */graph/v1*.

2.2.6. EuroWordnet

Es una base de datos multilingüe que se basa en relaciones semánticas básicas entre palabras. Es muy parecido a WordNet pero en vez de usar *synsets* ellos lo llaman *WordNet*. Cada *WordNet* representa un sistema único de lexicalización y además estos están conectados con el Índice Inter-Lingual (ILI) (Wikipedia, 2017b). Podemos ver en la figura 2.14 un ejemplo de búsqueda con la palabra *casa*, donde la aplicación devuelve los sinónimos, una pequeña definición así como un ejemplo para que el usuario pueda entender mejor el concepto.

2.3. Figuras retóricas

Las figuras literarias (o retóricas) son 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. La metáfora, el símil y la analogía se basan en la comparación de dos conceptos: El de origen, que es el término literal (al que la metáfora se refiere) se llama tenor y el de destino, que es el término figurado, llamado vehículo. 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 (Azaustre Galiana, 1994).

En este trabajo vamos a trabajar con tres tipos de figuras retóricas (Galván, 2016-2017):

- Metáfora: Se refiere a una cosa mencionando otra, utiliza el desplazamiento de características similares entre dos conceptos con fines estéticos o retóricos. Por ejemplo, cuando una persona tiene muy buena memoria, se dice que tiene memoria de elefante. Ya que una de las características de los elefantes es que tienen buena memoria.
- Símil: Realiza una comparación entre dos términos. A pesar de que los símiles y las metáforas son similares, los símiles utilizan explícitamente conectores (por ejemplo, como, cual, que, o verbos). Por ejemplo, cuando nos referimos a una persona que es muy corpulenta, se dice que 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*.

2.4. 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 plataforma (Wikipedia, 2018c). Según el W3C(*World Wide Web Consortium*)¹¹, 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 (Wikipedia, 2017a):

- 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. Deberemos tener algún mecanismo que nos permita encontrar un Servicio Web que realice una determinada función. De esta forma tendremos la posibilidad de que una aplicación localice el servicio que necesite de forma automática, sin tener que conocerlo previamente el usuario.

Por otro lado, los servicios web pueden definirse tanto a nivel conceptual como a nivel técnico, es por ello que mediante este último podemos diferenciar dos tipos distintos de servicio web (Wikipedia, 2017a):

- Servicios web SOAP (Simple Object Access Protocol): es un protocolo basado en XML para el intercambio de información entre ordenadores. Normalmente utilizaremos SOAP para conectarnos a un servicio e invocar métodos remotos ¹². Los mensajes SOAP tienen el siguiente formato representado en la figura 2.15:

¹¹<https://www.w3.org/>

¹²<https://www.ibm.com/support/knowledgecenter/es>

- **<Envelope>**: Elemento raíz de cada mensaje SOAP y contiene dos elementos: **<Header>** que es opcional y **<Body>** que es obligatorio, ambos elementos los describiremos a continuación.
 - **<Header>**: Es un elemento que se utiliza para indicar información acerca de los mensajes SOAP.
 - **<Body>**: Elemento que contiene información dirigida al destinatario del mensaje.
 - **<Fault>**: Elemento en el que se notifican los errores.
-
- **Servicios Web RESTful**: es un protocolo que suele integrar mejor con HTTP que los servicios basado 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.

2.4.1. Arquitectura Servicios Web

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

- **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 capas ¹³:

- **Service Discovery**: Responsable de centralizar los servicios web en un directorio común, de esta forma es más sencillo buscar y publicar.
- **Service Description**: Como ya hemos comentado con anterioridad, los servicios web se pueden definir así mismos, por lo que una vez que los localicemos el Service Description nos dará la información para saber que operaciones soporta y como activarlo.
- **Service Invocatio**: Invocar a un Web Service 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.

¹³<https://diego.com.es/introduccion-a-los-web-services>

- Transport: Todos los mensajes han de ser transmitidos de alguna forma entre el servidor y el cliente. El protocolo elegido para ello es HTTP ((HyperText Transfer Protocol)).

2.4.2. Ventajas de los Servicios Web

Las principales ventajas del uso de los servicios web son las siguientes (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.
- 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.

2.4.3. Desventajas de los Servicios Web

El uso de servicios web también tiene algunas desventajas ¹⁴:

- 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.

¹⁴<http://fabioalfarocc.blogspot.com/2012/08/ventajas-y-desventajas-del-soap.html>



Figura 2.9: Resultados búsqueda Thesaurus Rex con la palabra *house*

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼ <MemberData>
  ▶ <Categories kw="house">...</Categories>
  ▶ <Modifiers kw="house">...</Modifiers>
  ▼ <CategoryHeads kw="house">
    <CategoryHead weight="86"> point </CategoryHead>
    <CategoryHead weight="3320"> object </CategoryHead>
    <CategoryHead weight="26"> measure </CategoryHead>
    <CategoryHead weight="2340"> structure </CategoryHead>
    <CategoryHead weight="1"> layer </CategoryHead>
    <CategoryHead weight="6"> force </CategoryHead>
    <CategoryHead weight="69"> fixture </CategoryHead>
    <CategoryHead weight="1186"> building </CategoryHead>
    <CategoryHead weight="1078"> item </CategoryHead>
    <CategoryHead weight="6"> protection </CategoryHead>
    <CategoryHead weight="41"> sector </CategoryHead>
    <CategoryHead weight="2"> atmosphere </CategoryHead>
    <CategoryHead weight="17"> enterprise </CategoryHead>
    <CategoryHead weight="22"> ceremony </CategoryHead>
    <CategoryHead weight="713"> thing </CategoryHead>
    <CategoryHead weight="43"> set </CategoryHead>
    <CategoryHead weight="10"> action </CategoryHead>
    <CategoryHead weight="302"> location </CategoryHead>
    <CategoryHead weight="636"> area </CategoryHead>
    <CategoryHead weight="8"> article </CategoryHead>
    <CategoryHead weight="21"> communication </CategoryHead>
    <CategoryHead weight="21"> group </CategoryHead>
    <CategoryHead weight="7"> part </CategoryHead>
    <CategoryHead weight="43"> institution </CategoryHead>
```

Figura 2.10: Ejemplos formato XML Thesaurus Rex

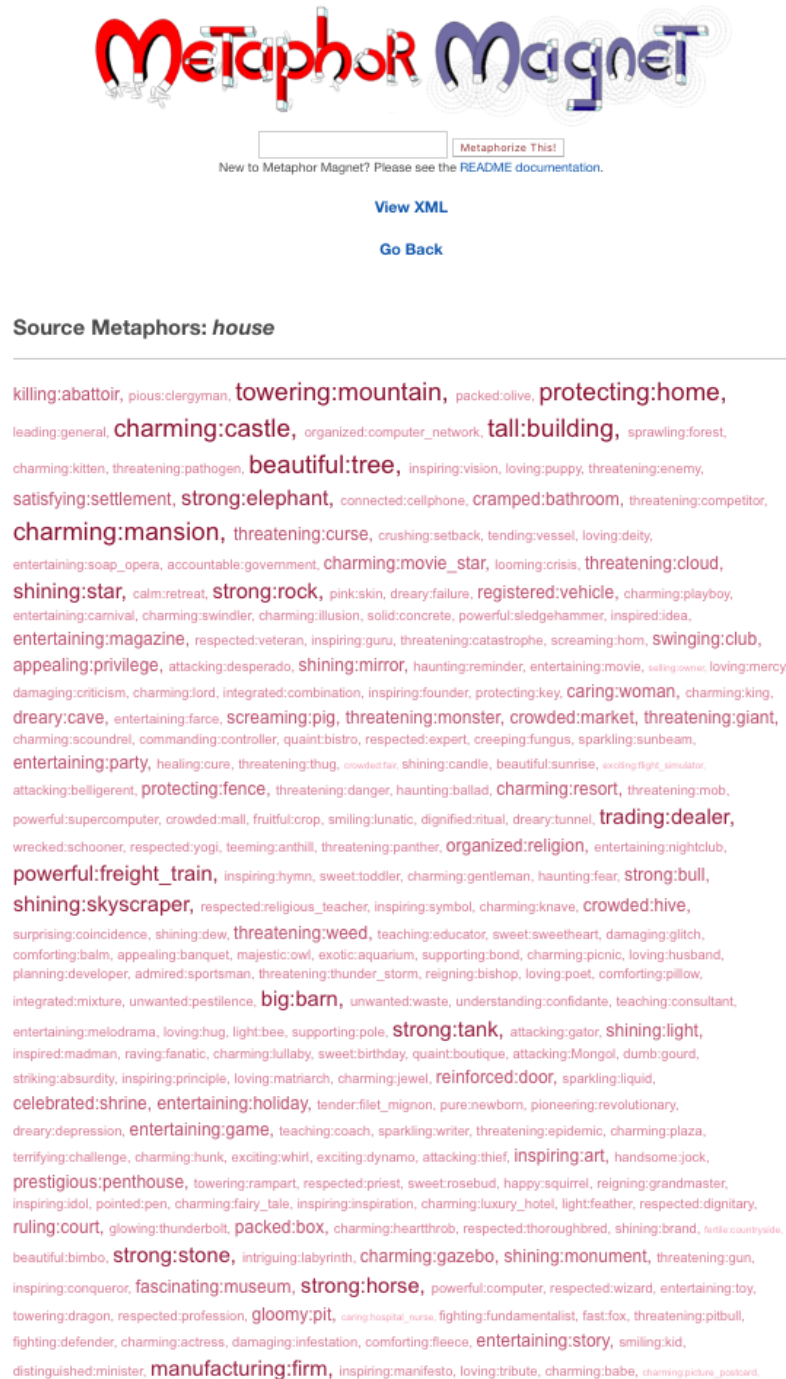


Figura 2.11: Resultados búsqueda Metaphor Magnet con la palabra *house*

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<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> shining:star </Text>
    <Score> 80 </Score>
  </Source>
  ▼<Source Name="house">
    <Text> shining:skyscraper </Text>
    <Score> 79 </Score>
  </Source>
  ▼<Source Name="house">
    <Text> towering:fortress </Text>
    <Score> 79 </Score>
  </Source>
  ▼<Source Name="house">
    <Text> strong:tank </Text>
    <Score> 79 </Score>
  </Source>
  ▼<Source Name="house">
    <Text> majestic:whale </Text>
    <Score> 79 </Score>
  </Source>
  ▼<Source Name="house">
    <Text> big:barn </Text>
    <Score> 77 </Score>
  </Source>
```

Figura 2.12: Ejemplos formato XML Metaphor Magnet

WordNet Search - 3.1
[- WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
 Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) [home](#), [place](#) (where you live at a particular time) *"deliver the package to my home"; "he doesn't have a home to go to"; "your place or mine?"*
- [S:](#) (n) [dwelling](#), [home](#), [domicile](#), [abode](#), [habitation](#), [dwelling house](#) (housing that someone is living in) *"he built a modest dwelling near the pond"; "they raise money to provide homes for the homeless"*
- [S:](#) (n) [home](#) (the country or state or city where you live) *"Canadian tariffs enabled United States lumber companies to raise prices at home"; "his home is New Jersey"*
- [S:](#) (n) [home plate](#), [home base](#), [home](#), [plate](#) ((baseball) base consisting of a rubber slab where the batter stands; it must be touched by a base runner in order to score) *"he ruled that the runner failed to touch home"*
- [S:](#) (n) [base](#), [home](#) (the place where you are stationed and from which missions start and end)
- [S:](#) (n) [home](#) (place where something began and flourished) *"the United States is the home of basketball"*
- [S:](#) (n) [home](#) (an environment offering affection and security) *"home is where the heart is"; "he grew up in a good Christian home"; "there's no place like home"*
- [S:](#) (n) [family](#), [household](#), [house](#), [home](#), [menage](#) (a social unit living together) *"he moved his family to Virginia"; "It was a good Christian household"; "I waited until the whole house was asleep"; "the teacher asked how many people made up his home"; "the family refused to accept his will"*
- [S:](#) (n) [home](#), [nursing home](#), [rest home](#) (an institution where people are cared for) *"a home for the elderly"*

Verb

- [S:](#) (v) [home](#) (provide with, or send to, a home)
- [S:](#) (v) [home](#) (return home accurately from a long distance) *"homing pigeons"*

Adjective

- [S:](#) (adj) [home](#) (used of your own ground) *"a home game"*
- [S:](#) (adj) [home](#) (relating to or being where one lives or where one's roots are) *"my home town"*
- [S:](#) (adj) [home](#), [interior](#), [internal](#), [national](#) (inside the country) *"the British Home Office has broader responsibilities than the United States Department of the Interior"; "the nation's internal politics"*

Figura 2.13: Resultados búsqueda Wordnet con la palabra *house*

Figura 2.14: Resultados búsqueda EuroWordnet con la palabra *casa*

Figura 2.14: Resultados búsqueda EuroWordnet con la palabra *casa*

Figura 2.15: Estructura de un mensaje SOAP

Capítulo 3

Trabajo Realizado

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

3.1. 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, integramos 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.

3.2. 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.

El diseño de la Interfaz debe cumplir las Ocho Reglas de Oro del diseño de interfaces:

- **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 que 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.

Capítulo 4

Conclusiones y Trabajo Futuro

Conclusiones del trabajo y líneas de trabajo futuro.

Chapter 4

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 cerebro de manera que vino
a perder el juicio.*

Miguel de Cervantes Saavedra

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

GALVÁN, P. Generación de recursos lingüísticos mediante la extracción de relaciones entre conceptos. 2016-2017.

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

LEBRÚN, C. A. V. Integración de herramientas de tecnologías de información como soporte en la adm del conocimiento. 2005.

UNIVERSITY, P. About wordnet. 2010.

VEALE.T y LI.G. *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity*. Bllomsbury Academic, 2012.

WIKIPEDIA. Definición soap. 2017a.

WIKIPEDIA. Eurowordnet — wikipedia, la enciclopedia libre. 2017b.

WIKIPEDIA. Lectura fácil — wikipedia, la enciclopedia libre. 2018a.

WIKIPEDIA. Red semántica — wikipedia, la enciclopedia libre. 2018b.

WIKIPEDIA. Servicio web — wikipedia, la enciclopedia libre. 2018c.

*—¿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*

