



UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior de Jaén

Trabajo Fin de Grado

PARAFRASEO. CÓMO ENGAÑAR A TURNITÍN.

Alumno: Alejandro Juan Rubio Guzmán

Tutor: Manuel Carlos Díaz Galiano
Pilar López Úbeda

Dpto: Departamento de informática

Junio, 2020



Universidad de Jaén
Escuela Politécnica Superior de Jaén
Departamento de Informática

Don Manuel Carlos Díaz Galiano , tutor del Proyecto Fin de Carrera titulado: Parafraseo: Como engañar a Turnitin, que presenta Alejandro Juan Rubio Guzmán, autoriza su presentación para defensa y evaluación en la Escuela Politécnica Superior de Jaén.

Jaén, Mayo de 2020

El alumno:

Los tutores:

Alejandro Juan Rubio Guzmán

Manuel Carlos Díaz Galiano

Pilar López Úbeda

Índice

1. Introducción	7
2. Motivación.....	11
3. Objetivos del proyecto.....	16
3.1. Objetivos generales.....	16
3.2. Objetivos específicos.....	16
4. Metodología.....	18
4.1. Planificación de la iteración.....	18
4.2. Ejecución de la iteración.....	19
4.3. Inspección y adaptación.....	19
5. Análisis de técnicas de parafraseo.....	20
5.1. Cambios morfo-léxicos.....	20
5.1.1. Cambios morfológicos.....	21
5.1.2. Cambios léxicos.....	22
5.2. Cambios semánticos.....	24
5.3. Cambios estructurales.....	25
5.3.1. Cambios sintácticos.....	25
5.3.2. Cambios discursivos.....	26
6. Planificación del proyecto.....	28
6.1. Definición de tareas.....	28
6.2. Estimación de tiempos.....	29
6.3. Diagrama de Gantt.....	30
6.4. Análisis de costes.....	31
7. Análisis del proyecto.....	32
7.1. Casos de uso.....	32
7.2. Análisis de requisitos.....	33
7.2.1. Requisitos funcionales.....	33
7.2.2. Requisitos no funcionales.....	33
7.3. Prototipo.....	34
8. Diseño e implementación.....	36
8.1. Tecnologías y librerías.....	36
8.2. Modelo de datos.....	37
8.3. Implementación.....	39

8.4. Pruebas y validación.....	42
9. Conclusión.....	44
10. Futuras mejoras.....	45
11. Manual de usuario.....	46
12. Referencias.....	51

Índice de ilustraciones

Figura 1: Diagrama PERT.....	29
Figura 2: Diagrama Gantt.....	30
Figura 3: Casos de uso.....	32
Figura 4: Prototipo.....	34
Figura 5: Modelo de datos.....	37
Figura 6: Estructura del proyecto.....	39
Figura 7: Transformación pasiva-activa.....	41
Figura 8: Manual de usuario 1.....	46
Figura 9: Manual de usuario 2.....	46
Figura 10: Manual de usuario 3.....	47
Figura 11: Manual de usuario 4.....	47
Figura 12: Manual de usuario 5.....	48
Figura 13: Manual de usuario 6.....	48
Figura 14: Manual de usuario 7.....	49
Figura 15: Manual de usuario 8.....	50

Índice de tablas

Tabla 1.....20
Tabla 2.....27

1. Introducción

La paráfrasis, llamado coloquialmente parafraseo, es la explicación con palabras propias del contenido de un texto para facilitar la comprensión de la información que contenga dicho texto. En otras palabras, la paráfrasis consiste en decir con palabras más sencillas y con menos palabras técnicas las ideas propias obtenidas de un texto predeterminado. Así, se facilita su comprensión, ya que se expresa la información original extraída de una forma diferente bajo el mismo código de comunicación. La paráfrasis es también una forma rápida y sencilla de adquirir conocimientos, ya que parafrasear un texto facilita recordar lo leído. Puede entenderse también como una “traducción” que da al texto una visión clara, precisa y didáctica para otros lectores. No se trata, de una traducción a otro idioma, sino una forma de resumir un texto con las palabras propias de un individuo.

Generalmente existen dos tipos de paráfrasis hasta el momento, los cuáles son descritos a continuación:

- La paráfrasis mecánica: es la sustitución simple de expresiones que aparezcan en un texto por sinónimos, con cambios sintácticos mínimos. Por ejemplo, la frase original dice “Los dramaturgos se adentran en las dinámicas creativas que les son propias con la confianza de quienes gozan de un público cautivo” podríamos parafrasearla mecánicamente de la siguiente manera “Los escritores de teatro se sumergen en sus métodos de creación literaria con la confianza que les da tener siempre un mismo público”.
- La paráfrasis constructiva: es la reelaboración del enunciado, dando origen a otro con palabras o características que pueden ser muy distintas, pero que conservan el mismo significado. Por ejemplo, la frase original “Los dramaturgos se adentran en las dinámicas creativas que les son propias con la confianza de quienes gozan de un público cautivo”, una paráfrasis constructiva podría ser “A la hora de crear, los escritores de teatro poseen una ventaja que otro tipo de escritores no, y es que tienen un público cautivo para darles mayor confianza”.

La palabra española «paráfrasis» procede de la latina *paraphrasis*, que a su vez procede de la griega *παράφρασις* (imitación del texto original), empleando para ello otro lenguaje, normalmente más sencillo y práctico [7]. Se trata así de un método de resumen ya que la paráfrasis tiende a ser más breve, más simple y más sencilla que su original.

Elaborar una paráfrasis es muy sencillo, pero ante todo requiere de la comprensión del texto original, por este motivo, el primer paso es su lectura e interpretación. Posteriormente, se podrán extraer las ideas principales del texto y luego las ideas secundarias. Una vez conseguida dicha información, se podrá proceder a explicar con el propio lenguaje lo que el texto original decía, sabiendo que la idea principal deberá ser la misma y las ideas secundarias deberán tener la misma relación en el texto. Las ideas terciarias o contextuales podrán descartarse [6].

Los métodos para parafrasear un texto deben operar en una gran mayoría de casos con plantillas. Es decir, dependiendo del tipo de paráfrasis que apliquemos se usará una plantilla u otra que nos llevarán a un resultado en concreto tras introducir un texto origen en el sistema. Pongamos los siguientes ejemplos:

(1) *X wrote Y*

(2) *Y was written by X*

(3) *X is the writer of Y*

Aquí, X e Y en (1) pueden ser transformadas con frases nominales arbitrarias a (2) y (3). Dichas plantillas especifican la relación sintáctica que debe existir entre X e Y para que las diferentes oraciones, en este caso (1), (2) y (3), digan lo mismo con diferentes palabras, es decir, para que se produzca la paráfrasis. Por otro lado, se reconocen, generan o extraen pares (T, H), siendo T la frase origen y H la frase parafraseada, de expresiones del lenguaje natural, de modo que un humano que lee T, inferiría que H es probablemente cierto. Por ejemplo, la frase (4) implica textualmente (5) pero (6) no implica textualmente (7).

(4) *Los medicamentos que ralentizan la enfermedad de Alzheimer funcionan mejor cuanto antes se administren.*

(5) *La enfermedad de Alzheimer puede ralentizarse usando medicamentos.*

(6) *Drew Walker, director de salud pública del Tayside, dijo: “Es importante enfatizar que esto no es una confirmación de caso de rabia”.*

(7) *Se confirmó un caso de rabia.*

En general, no podemos juzgar si dos expresiones del lenguaje natural son paráfrasis o un correcto par (T , H) de vinculación textual sin seleccionar lecturas particulares de las expresiones, entre las que puede ser posible, debido a múltiples sentidos de las palabras, ambigüedades sintácticas, etc. Por ejemplo, la frase (8) implica a (9) con el sentido financiero de “banco”, pero no cuando (8) se refiere a la orilla de un río (bank en inglés significa orilla).

(8) *A bomb exploded near the French bank.*

(9) *A bomb exploded near a building.*

La paráfrasis ofrece la posibilidad de examinar las expresiones del lenguaje sólo en contextos particulares que aclaran sus lecturas previstas. Alternativamente, podemos tratar como correcto cualquier par de vinculación textual (T , H) para el que hay posibles lecturas de T y H; entonces, si un sistema informa que (8) implica (9), su respuesta debe contarse como correcta, independiente del sentido previsto de “bank”. De este modo, las paráfrasis tendrían lecturas posibles que transmiten casi la misma información.

Actualmente existe una conferencia de ámbito internacional llamada SEMEVAL (International Workshop on Semantic Evaluation) que contiene una tarea de sustitución léxica de SEMEVAL [3]. SEMEVAL es una serie de evaluaciones de análisis semántico en sistemas computacionales, las evaluaciones están destinadas a explorar la naturaleza del significado en el lenguaje. Esta serie de evaluaciones proporcionan un mecanismo para caracterizar en términos más precisos exactamente lo que es necesario para calcular el significado. El propósito de los ejercicios de SEMEVAL es evaluar los sistemas de análisis semántico [9]. En la tarea de sustitución léxica de SEMEVAL se requiere que los sistemas encuentren un sustituto apropiado para una palabra en particular en el contexto de una oración dada, puede verse como un caso especial de parafraseo o vinculación textual, restringido a pares de palabras. La tarea de SEMEVAL, sin embargo, incluye el requisito de que debe de ser posible usar las dos palabras (original y de reemplazo) exactamente en

el mismo contexto. De manera similar, uno podría adoptar una definición más estricta de paráfrasis, lo que requeriría que no solo tuvieran el mismo significado, sino también ser expresiones que se pueden usar indistintamente en oraciones gramaticales. En ese caso, aunque la frase (10) y (11) son paráfrasis, sus partes subrayadas no lo son, dado que no se pueden intercambiar en las dos oraciones; las oraciones resultantes serían gramaticalmente incorrectas, es decir, no quedan bien formadas sintácticamente (la (10) sí pero la (11) no).

(10) Edison invented the light bulb in 1879, providing a long lasting source of light.

(11) Edison's invention of the light bulb in 1879 provided a long lasting source of light.

Una definición similar más estricta de implicación textual impondría el requisito adicional de que H y T pueden reemplazarse entre sí en oraciones gramaticales [1].

2. Motivación.

La importancia de la paráfrasis se puede apreciar en diferentes campos [6]:

- En literatura y lenguaje: Se hace uso de la paráfrasis en la exégesis de textos literarios. También se consideran paráfrasis la prosificación del verso y la versificación de la prosa. Este también un recurso que se basa en el uso de sinónimos para evitar repeticiones. En el caso de palabras sin un equivalente, estas pueden ser reemplazadas por una locución, por una frase o traducciones de un idioma.
- En aprendizaje: También es un recurso didáctico y que facilita la adquisición de conocimiento y su memorización. Entre otros casos, encontramos ventajas como las mostradas a continuación:
 - Se aprovecha esta figura retórica para ejercitar la redacción.
 - Los profesores y estudiantes hacen uso de la paráfrasis cuando asimilan primero el contenido de una lección y después lo expresan con palabras distintas, tratando de que nada esencial sea omitido en la exposición y de que esta sea perfectamente comprendida por el auditorio.
- En resolución de conflictos: Por otro lado, en el método alternativo de resolución de conflictos, uno de los principales recursos del mediador o conciliador es la paráfrasis que repite la idea que ha transmitido una de las partes en conflicto, con el mismo criterio o punto de vista aunque con otras palabras.

El Procesamiento del Lenguaje Natural, abreviado PLN, es un campo de la informática, inteligencia artificial y lingüística que estudia las interacciones entre los ordenadores y el lenguaje. El PLN se ocupa de investigar mecanismos eficaces para la comunicación entre personas y máquinas por medio del lenguaje natural [8]. Así, es interesante notar cómo la paráfrasis influye en cada uno de los siguientes ámbitos del PLN [1]:

1. Las expresiones del lenguaje natural que los métodos de parafraseo consideran no son

siempre estamentos. De hecho, muchos de estos métodos se desarrollaron teniendo sistemas de preguntas y respuestas (QA - Questions & Answers) en mente. En sistemas de control de calidad para colecciones de documentos, una pregunta puede estar redactada de manera diferente en un documento que contiene la respuesta y tomar en cuenta tales variaciones puede mejorar el desempeño del sistema significativamente. Por ejemplo, un sistema de control de calidad puede recuperar documentos o pasajes relevantes utilizando la pregunta de entrada como una consulta a un sistema de recuperación de información o motor de búsqueda en la web, y luego que verifique si alguno de los textos recuperados implica textualmente una respuesta candidata. Por ejemplo, si la pregunta de entrada es (12) y el motor de búsqueda devuelve el pasaje (13), el sistema puede verificar si (13) textualmente implica alguna de las respuestas candidatas de (14), donde hemos reemplazado el interrogativo “who” de (12) con todas las expresiones de (13) y que un reconocedor de entidades ha reconocido como nombres de personas.

(12) *Who sculpted the Doryphoros?*

(13) *The Doryphoros is one of the best known Greek sculptures of the classical era in Western Art. The Greek sculptor Polykleitos designed this work as an example of the “canon” or “rule”, showing the perfectly harmonious and balanced proportions of the human body in the sculpted form. The sculpture was known through the Roman marble replica found in Herculaneum and conserved in the Naples National Archaeological Museum, but, according to Francis Haskell and Nicholas Penny, early connoisseurs passed it by in the royal Bourbon collection at Naples without notable comment.*

(14) *Polykleitos/Francis Haskell/Nicholas Penny sculpted the Doryphoros,*

La pregunta de entrada también puede ser parafraseada para permitir obtener más pasajes potencialmente relevantes. La paráfrasis de preguntas es también útil cuando se asignan preguntas de usuarios a listas de preguntas frecuentes que van acompañadas de sus respuestas; y las interfaces de lenguaje natural a las bases de datos a menudo generan paráfrasis de preguntas para permitir a los usuarios comprender si sus consultas se han

entendido.

2. En el resumen de texto, una etapa de procesamiento importante es la extracción de oraciones, que identifica las oraciones más importantes de los textos que se van a resumir. Durante esa etapa, especialmente cuando se genera un resumen único de varios documentos es importante evitar seleccionar oraciones de diferentes artículos de noticias sobre el mismo evento, que transmiten la misma información que otras oraciones que ya han sido seleccionadas u oraciones cuya información se deriva de otras oraciones ya seleccionadas.
3. En compresión de oraciones a menudo también se puede ver una etapa de procesamiento del resumen de texto como un caso especial de parafraseo de oraciones con la restricción adicional de que la oración resultante debe ser más corta que la original y aún gramatical; por ejemplo, una oración que coincida con (2) y (3) podría acortarse en una paráfrasis de la forma de (1). Sin embargo, la mayoría del trabajo de compresión de oraciones permite descartar información menos importantes de la oración original. Por otro lado, la frase (16) es una forma comprimida de (15) producida por un humano [4].

(15) *Mother Catherine, 82, the mother superior, will attend the hearing on Friday, he said.*

(16) *Mother Catherine, 82, the mother superior, will attend.*

Cuando la oración comprimida no es necesariamente una paráfrasis de la original, primero podemos producir compresiones candidatas (gramaticales) que conlleva textualmente a la oración original; por lo tanto, un mecanismo para generar oraciones textualmente implicadas es útil. Sin embargo, se necesitan mecanismos adicionales para clasificar a los candidatos según el espacio que ahorran y el grado en que mantienen información importante.

4. Los sistemas de extracción de información a menudo se basan en patrones diseñados manual o automáticamente para localizar fragmentos de texto que informan de tipos particulares de eventos e identifican las entidades involucradas; por ejemplo, patrones como las frases (17), (18) y (19) o patrones similares que operan en árboles de sintaxis, podrían usarse para localizar fragmentos que se refieran a incidentes de bombardeo e identificar sus objetivos.

Los métodos de parafraseo o de implicación textual se pueden utilizar para generar patrones de extracción semánticamente equivalentes adicionales (en el caso de parafrasear) o patrones que impliquen textualmente los originales. mantienen información importante.

(17) *X was bombed*

(18) *bomb exploded near X*

(19) *explosion destroyed X*

5. En la traducción automática, las ideas de la investigación de parafraseo e implicación textual se han integrado en medidas y procesos que evalúan automáticamente las traducciones generadas por máquinas contra las de autoría humana que pueden usar diferentes frases. Los métodos de parafraseo también se han utilizado para generar automáticamente traducciones de referencia adicionales de las creadas por humanos al entrenar sistemas de traducción automática. Finalmente, se han empleado métodos de parafraseo e implicación textual para permitir que los sistemas de traducción automática hagan frente a palabras de lenguaje fuente y frases más largas que no se han encontrado en corpus de entrenamiento. Supóngase un sistema de traducción automática, basado en frases, que nunca ha encontrado la expresión "*filed a lawsuit against*", pero que sabe que el patrón (20) implica textualmente (21), puede ser capaz de producir una traducción más aceptable al convertir (22) a (23), y luego traduciendo (23). Alguna información se perdería en la traducción, ya que (23) no es una paráfrasis de (22), pero la traducción aún puede ser preferible al resultado de traducir directamente (22) [5].

(20) *X filed a lawsuit against Y for Z*

(21) *X accused Y of Z.*

(22) *Cisco filed a lawsuit against Apple for patent violation.*

(23) *Cisco accused Apple of patent violation.*

6. En la generación del lenguaje natural, al producir textos que describen las entidades de una ontología formal, el parafraseo se puede usar para evitar repetir las mismas frases, por ejemplo, al expresar propiedades de entidades similares, o para producir expresiones

alternativas que mejoren la coherencia del texto, se adhieran al estilo de escritura (por ejemplo, eviten pasivos) o satisfagan otras restricciones. Entre otras posibles aplicaciones, se pueden emplear métodos de parafraseo e implicación textual para simplificar los textos.

3. Objetivos del proyecto

En este apartado se definen los objetivos generales y específicos del proyecto. Los objetivos generales del proyecto se refiere a qué se pretende conseguir con este proyecto. Por otro lado, los objetivos específicos engloban cada una de las tareas que se ha llevado a cabo para realizar este proyecto.

3.1. Objetivos generales

Los objetivos generales de este trabajo son elaborar una aplicación web, junto con la documentación que justifique este trabajo, que contenga diferentes técnicas de paráfrasis a poder usar. Así, con la aplicación, un usuario podrá parafrasear un texto de entrada dado usando una o varias opciones de paráfrasis sobre dicho texto, obteniendo como resultado un texto de salida que estará propiamente parafraseado con las características propias de la opción u opciones que se usen.

3.2. Objetivos específicos

Cada una de las tareas realizadas para la elaboración de este proyecto se enumeran de la siguiente manera:

- A) Realizar un análisis bibliográfico de lo existente en parafraseo: Se ha llevado a cabo una investigación previa sobre el material textual existente sobre la paráfrasis. Así, con este objetivo se pretende entender el estado del arte en que se encuentra el motivo de estudio de este proyecto.
- B) Estudiar las distintas técnicas de parafraseo: Tras investigar sobre el estado del arte de la paráfrasis, el siguiente paso ha sido estudiar cada una de las técnicas propuestas por los diferentes autores respecto a lo que paráfrasis se refiere.
- C) Seleccionar la metodología a desarrollar: La metodología escogida y con la que se ha desarrollado el proyecto ha sido una metodología ágil, concretamente la metodología scrum. Dada sus características es la más elocuente para llevar a cabo este trabajo.
- D) Desarrollar e implementar una librería de parafraseo: Se han escogido alguna de las

técnicas estudiadas y se ha implementado una librería. Las técnicas implementadas en este trabajo y que dan lugar a la librería desarrollada son:

- Traductor: Con esta opción, un texto de entrada será traducido a varios idiomas para finalmente ser nuevamente traducido al español. En la aplicación propiamente se podrán escoger 3 idiomas a los que traducir el texto de entrada, de esta manera, el texto de entrada se traduce primero a un idioma, luego al segundo, luego al tercero y por último al español, obteniendo así la interpretación del mismo texto por 3 idiomas diferentes.
- Sinónimos: Con esta opción, cada sustantivo y adjetivo encontrado en el texto se sustituirá por una lista con los sinónimos disponibles para cada sustantivo y adjetivo dando al usuario una alternativa visual rápida para cambiar las palabras más relevantes del texto.
- Pasiva a activa y viceversa: Con esta opción, una frase pasiva se pasa a pasiva y viceversa.
- Resumen: Por último, esta opción generará un resumen con las frases con más relevancia del texto de entrada.

E) Desarrollar e implementar una interfaz gráfica para el proyecto: Tras implementar cada una de las técnicas mencionadas en el apartado anterior se ha llevado a cabo la elaboración de una aplicación web que funciona como interfaz gráfica para poder usar la librería elaborada y así poder hacer pruebas, implementando así de paso un camino más fácil para usar el trabajo implementado.

F) Formalizar el proyecto mediante la elaboración de una documentación: Por último, se ha formalizado y terminado una documentación en la que se justifica el trabajo realizado además de explicar en qué consiste el proyecto.

4. Metodología.

La metodología seguida en este proyecto es una metodología ágil, concretamente la metodología scrum. Scrum es un proceso en el que se aplican un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y así obtener así el mejor resultado posible. Se realizan entregas parciales y regulares del producto final, priorizadas por el beneficio que aportan al receptor del proyecto. Por ello, Scrum está especialmente indicado para proyectos en entornos complejos, donde se necesita obtener resultados pronto, donde los requisitos son cambiantes o poco definidos, donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

También se utiliza para resolver situaciones en que no se está entregando al cliente lo que necesita, cuando las entregas se alargan demasiado, los costes se disparan o la calidad no es aceptable, cuando se necesita capacidad de reacción ante la competencia, cuando la moral de los equipos es baja y la rotación alta, cuando es necesario identificar y solucionar ineficiencias sistemáticamente o cuando se quiere trabajar utilizando un proceso especializado en el desarrollo de producto.

El proceso parte de la lista de objetivos/requisitos priorizada del producto, que actúa como plan del proyecto. En esta lista el cliente prioriza los objetivos balanceando el valor que le aportan respecto a su coste y quedan repartidos en iteraciones y entregas.

4.1. Planificación de la iteración

El primer día de la iteración se realiza en la reunión de planificación de la iteración. Tiene dos partes:

1. Selección de requisitos: El cliente presenta al equipo la lista de requisitos priorizada del producto o proyecto. El equipo pregunta al cliente las dudas que surgen y selecciona los requisitos más prioritarios que prevé que podrá completar en la iteración, de manera que puedan ser entregados si el cliente lo solicita.
2. Planificación de la iteración: El equipo elabora la lista de tareas de la iteración necesarias

para desarrollar los requisitos seleccionados. La estimación de esfuerzo se hace de manera conjunta y los miembros del equipo se auto-asignan las tareas, se autoorganizan para trabajar incluso en parejas (o grupos mayores) con el fin de compartir conocimiento o resolver juntos objetivos especialmente complejos.

4.2. Ejecución de la iteración.

El equipo inspecciona el trabajo que el resto está realizando para poder hacer las adaptaciones necesarias que permitan cumplir con la previsión de objetivos a mostrar al final de la iteración.

Durante la iteración, el cliente junto con el equipo refinan la lista de requisitos y, si es necesario, cambian o replanifican los objetivos del proyecto con el objetivo de maximizar la utilidad de lo que se desarrolla y el retorno de inversión.

4.3. Inspección y adaptación.

El último día de la iteración se realiza la reunión de revisión de la iteración. Tiene dos partes:

1. Revisión (demostración): El equipo presenta al cliente los requisitos completados en la iteración, en forma de incremento de producto preparado para ser entregado con el mínimo esfuerzo. En función de los resultados mostrados y de los cambios que haya habido en el contexto del proyecto, el cliente realiza las adaptaciones necesarias de manera objetiva, ya desde la primera iteración, replanificando el proyecto.
2. Retrospectiva: El equipo analiza cómo ha sido su manera de trabajar y cuáles son los problemas que podrían impedirle progresar adecuadamente, mejorando de manera continua su productividad.

5. Análisis de técnicas de parafraseo.

Se podría decir que los diferentes métodos de paráfrasis [2] se pueden recoger en la siguiente tabla:

1. Cambios morfo-lexicos	1.1. Cambios morfológicos	1.1.1. Cambio de flexión. 1.1.2. Cambio de derivación. 1.1.3. Cambio de composición/descomposición.
	1.2. Cambios léxicos	1.2.1. Sustitución por sinónimos. 1.2.2. Sustitución por aproximación numérica. 1.2.3. Sustitución por una sigla o acrónimo. 1.2.4. Cambio de forma verbal. 1.2.5. Sustitución de un verbo por un conjunto de elementos equivalentes. 1.2.6. Inserción de palabras. 1.2.7. Eliminación de palabras. 1.2.8. Cambio de orden de palabras.
2. Cambios semánticos		2.1. Sustitución por sinónimos. 2.2. Sustitución por hiperónimos. 2.3. Sustitución por hipónimos. 2.4. Sustitución por holónimos. 2.5. Sustitución por merónimos. 2.6. Sustitución por antónimos. 2.7. Sustitución por acción-actante. 2.8. Sustitución por agente-instrumento.
3. Cambios estructurales	3.1. Cambios sintácticos	3.1.1. Transformación de pasiva/activa. 3.1.2. Repetición/Elipsis. 3.1.3. Inserción/Eliminación de oraciones de relativo
	3.2. Cambios discursivos	3.2.1. Transformación de discurso directo/indirecto. 3.2.2. Fusión de oraciones.

Tabla 1. Tabla con las diferentes técnicas de paráfrasis.

Teniendo en cuenta la tabla anterior, pasemos a analizar cada uno de los métodos indicados atendiendo a la clase de cambios que presentan dichos métodos:

5.1. Cambios morfo-léxicos.

Estos cambios se refieren a los cambios morfológicos o léxicos que se pueden realizar en las

oraciones. Dentro de este grupo podemos apreciar los siguientes cambios:

5.1.1. Cambios morfológicos.

Los cambios morfológicos consisten en las variaciones de forma en las palabras. Así, dentro de estos cambios podemos encontrar los siguientes cambios:

5.1.1.1. Cambio de flexión.

La flexión es la alteración que experimentan las palabras mediante morfemas constituyentes según el significado gramatical o categórico para expresar sus distintas funciones dentro de la oración y sus relaciones de dependencia o de concordancia con otras palabras o elementos oracionales [10]. En gramática tradicional, la flexión suele recibir nombres diferentes según se aplique a diferentes clases de palabras:

- La flexión verbal: suele denominarse conjugación.
- La flexión nominal: suele denominarse declinación y se suele aplicar a sustantivos, pronombres y adjetivos.

Por ejemplo, se puede escribir gat-o, gat-a, gat-os, gat-as... cuya raíz es gat. Entonces para usar esta técnica de paráfrasis habría que escoger el lema de las palabras y añadir sufijos. La principal diferencia con la derivación es que con la flexión no se cambia la categoría gramatical de la palabra.

5.1.1.2. Cambio de derivación.

La derivación es uno de los procedimientos de formación de palabras y permite a las lenguas designar conceptos relacionados semánticamente con otros en cierto sentido considerados como primitivos, añadiendo afijos. La derivación con frecuencia conlleva cambio de categoría gramatical mientras que en la flexión siempre se mantiene la categoría gramatical [11].

Ejemplo: cuchillada de cuchillo. Esta técnica se diferencia de la flexión en que la derivación puede conllevar cambios gramaticales (lo que es una palabra derivada de un verbo puede salir como sustantivo, así, por ejemplo, el verbo materializar sale del sustantivo material).

5.1.1.3. Cambio de composición/descomposición.

La composición o descomposición es un procedimiento morfológico de las lenguas para crear neologismos, esto es, nuevas palabras. Consiste en coordinar dos o más lexemas o raíces, esto es, partes invariables de palabras, para formar una nueva que constituye una unidad semántica y sintáctica [12].

En español es posible la composición de diferentes categorías léxicas: verbo+sustantivo, sustantivo+adjetivo, adjetivo+sustantivo, etc. Así por ejemplo, si tenemos sacacorchos podríamos descomponer en saca y corchos y viceversa como técnica de paráfrasis.

5.1.2. Cambios léxicos.

El léxico es el conjunto de palabras que conforma un determinado lecto y, por extensión, también se denomina así a los diccionarios que los recogen. En un sentido amplio, el concepto es extensible a los lenguajes de programación. Así, teniendo en cuenta esto, podemos encontrar los diferentes métodos de paráfrasis en lo que al cambio de léxico se refiere.

5.1.2.1. Sustitución palabra-definición.

Este tipo de cambios léxicos trataría de sustituir una palabra por su definición o viceversa. Así, por ejemplo, si tenemos la oración: “En la botella había leche” podríamos cambiarla por: “En el recipiente de vidrio había leche” y viceversa.

5.1.2.2. Sustitución por aproximación numérica.

Esta transformación, en textos con números, aproximaría las cifras eliminando decimales como el propio nombre sugiere. Hay diferentes tipos de aproximación numérica por lo que se debería de escoger un método de aproximación y modificar las cifras que se vayan encontrando aproximándolas. Así si tenemos la cifra 1,9999 podríamos aproximarla al número 2.

5.1.2.3. Sustitución por una sigla o acrónimo.

Propiamente consiste en cambiar una sigla/acrónimos por su significado o viceversa. Como ejemplo de esta sustitución podría ser ADN (Ácido desoxirribo nucleico).

5.1.2.4. Cambio de forma verbal.

En esta transformación se trata de modificar la forma verbal de un verbo, manteniendo la concordancia con el sujeto. Por ejemplo, la frase: “Iba a afeitarme” podría cambiarse por: “Iba a ir a afeitarme”.

5.1.2.5. Sustitución de un verbo por un conjunto de elementos equivalentes.

Consistiría en cambiar un verbo por un conjunto de elementos que se comporten igual que el verbo en cuestión. Un ejemplo podría ser la siguiente frase: “Estaba viendo la TV” por “Miraba la TV”.

5.1.2.6. Inserción de palabras.

Consiste en insertar palabras en una oración tal que su significado no variara.

5.1.2.7. Eliminación de palabras.

Al contrario que el anterior, consistiría en eliminar palabras tal que el significado de una oración no variara.

5.1.2.8. Cambio de orden de palabras.

Consiste en modificar el orden de las palabras obteniendo una frase que dice lo mismo. Ejemplo: “Un sábado noche salí de fiesta” por: “Salí de fiesta un sábado noche”.

5.2. Cambios semánticos.

La semántica es la parte lingüística que estudia el significado de las expresiones lingüísticas. Por tanto los cambios semánticos implicarán un cambio en ciertas palabras procurando que el significado de un texto sea el mismo. Estas sustituciones se refieren sobretodo a cambiar una o varias palabras de un texto por las sustituciones a que se refieren las subsecciones siguientes.

5.2.1. Sustitución por sinónimos.

Un sinónimo es una palabra que tiene el mismo significado que otra en cuestión. Así, se sustituiría una palabra por alguno de sus sinónimos ya que una palabra puede tener varios. Por ejemplo, un sinónimo de “texto” puede ser “escrito”.

5.2.2. Sustitución por hiperónimos.

Se denomina hiperónimo a aquel término general que puede ser utilizado para referirse a la realidad nombrada por un término más específico. Por ejemplo, “ser vivo” es hiperónimo para los términos “planta” y “animal” [13].

5.2.3. Sustitución por hipónimos.

Se denomina hipónimo a la palabra que posee todos los rasgos semánticos, o semas, de otra más general pero que en su definición añade otras características semánticas que la diferencian de ésta. Por ejemplo, los hipónimos de día son: lunes, martes, miércoles, etc. Es decir, son palabras que poseen todos los rasgos semánticos y añaden otras características para diferenciarlas de esta [14].

5.2.4. Sustitución por holónimos.

La holonimia es una noción semántica que se opone a la meronimia, del mismo modo que se oponen el todo y la parte. Así, por ejemplo, “bicicleta” es un holónimo mientras que “sillín”, “pedal” y “aro” son merónimos [15].

5.2.5. Sustitución por merónimos.

La meronimia es una relación semántica no simétrica entre los significados de dos palabras dentro del mismo campo semántico. Se denomina merónimo a la palabra cuyo significado constituye una parte del significado total de otra palabra. Por ejemplo, “dedos” es un merónimo de “manos” y “manos” es un merónimo de “brazos” [16].

5.2.6. Sustitución por antónimos.

Los antónimos son palabras que tienen significados opuestos o contrarios entre sí. Deben pertenecer a la misma categoría gramatical. Por ejemplo, antónimos de “alegría” son “tristeza”, “depresión”, etc.

5.2.7. Sustitución acción-actante.

El actante es quien realiza o el que realiza el acto, independientemente de cualquier otra determinación. Entonces, esta sustitución consistiría en sustituir una acción por el actante y viceversa [17].

5.2.8. Sustitución agente-instrumento.

Un agente es el participante que deliberadamente realiza la acción. Un instrumento es el participante que es usado para llevar a cabo la acción. Entonces consistiría en sustituir el agente por el instrumento y viceversa.

5.3. Cambios estructurales.

Los cambios estructurales se refieren a los cambios en la estructura de las oraciones. Así, podemos diferenciar entre cambios sintácticos y discursivos:

5.3.1. Cambios sintácticos.

Los cambios sintácticos se refieren a los cambios en la gramática del lenguaje, de tal forma

que una oración sea dicha con diferente estructura pero con igual significado.

5.3.1.1. Transformación de pasiva/activa.

Esta transformación consiste propiamente en transformar una oración de pasiva a activa y viceversa. Así, la frase “Juan escribió unas cartas” que está en activa se podría pasar a pasiva de la siguiente forma “Unas cartas fueron escritas por Juan”.

5.3.1.2. Repetición/Elipsis.

La elipsis o construcción elíptica es una figura retórica que consiste en la omisión de una o más palabras en una cláusula que, aunque sean necesarias para la correcta construcción gramatical, se sobre entienden por el contexto. Por ejemplo, la frase “Yo llevaba las flores y ellos llevaban el incienso” se puede cambiar a “Yo llevaba las flores y ellos, el incienso” [18].

5.3.1.3. Inserción/Eliminación de oraciones de relativo.

Las oraciones de relativo son aquellas estructuras subordinadas que se insertan en la oración introducidas por un pronombre relativo. Por ejemplo “El viaje que prometiste nunca sucedió” se podría cambiar a “El viaje que me prometiste nunca sucedió”. Así, se podrían añadir/eliminar este tipo de oraciones para hacer paráfrasis [19].

5.3.2. Cambios discursivos.

Una forma discursiva sería la herramienta lingüística compuesta por una semántica compleja y bien formada en un discurso verbal, en otras palabras, es aquel escrito que trata de generar una comunicación eficaz evitando los datos ambiguos, las impresiones personales, la emoción... . Por tanto, en el discurso se pueden aplicar las siguientes transformaciones.

5.3.2.1. Transformación de discurso directo/indirecto.

El discurso directo es la manera de referirse textualmente a un mensaje. Se marca con líneas

de diálogo o con comillas, por ejemplo: Manuel dijo <<Me gusta Dana>>

El indirecto es la descripción del mensaje de otro: Manuel dijo que le gustaba Dana. Puesto que no se repite la cita textualmente esta sufre ciertos cambios a tener en cuenta en esta transformación. La primera persona se convierte en tercera, al igual que los pronombres posesivos, adjetivos posesivos y pronombres objetivos de primera persona. Del mismo modo, la segunda persona se convierte en primera. Los tiempos verbales se actualizan y los adverbios de lugar y tiempo se modifican como corresponda [20].

5.3.2.2. Fusión de oraciones.

Esta transformación trata de unir oraciones mediante conectores. Un conector discursivo es una unidad lingüística que une partes de un texto y le da una relación lógica a las oraciones que siguen; la existencia de estos elementos es vital para integrar las ideas de mejor manera y para hacer inteligible la información. Por ejemplo si tenemos el texto: “Ayer compré dos sobres. También compré cuatro cartas.” se podría cambiar a: “Ayer compré dos sobres, además, también compré cuatro cartas.” dónde la partícula “además” es el conector que fusiona las oraciones.

6. Planificación del proyecto

En esta sección describe cómo se planificará el proyecto, aportando las tareas que son necesarias para la realización del proyecto, una estimación del tiempo necesario para llevarlo a cabo, un diagrama de Gantt donde se visualice de forma gráfica esa definición de tareas así como el tiempo que llevaría cada una y por último el análisis del coste asociado a dicho proyecto.

6.1. Definición de tareas

La siguiente tabla muestra la definición de las tareas necesarias para la ejecución de este proyecto.

Tarea	Duración (en días)
A - Realizar un análisis bibliográfico de lo existente en parafraseo.	14 días.
B - Estudiar distintas técnicas de parafraseo.	7 días.
C - Seleccionar la metodología a desarrollar.	3 días.
D - Desarrollar e implementar una librería de paráfrasis con las siguientes técnicas: Traductor, ActivaPasiva, Resumen, Sinónimos y juntar todas las técnicas anteriores en una librería usable.	30 días.
E - Desarrollar e implementar una interfaz gráfica para el proyecto.	10 días.
F - Formalizar el proyecto mediante la elaboración de una documentación.	21 días.

Tabla 2. Tabla de tareas del proyecto.

6.2. Estimación de tiempos

Dadas las actividades anteriores, el diagrama de PERT permite estimar el tiempo del proyecto así como las relaciones que hay entre las diferentes actividades, quedaría como sigue:

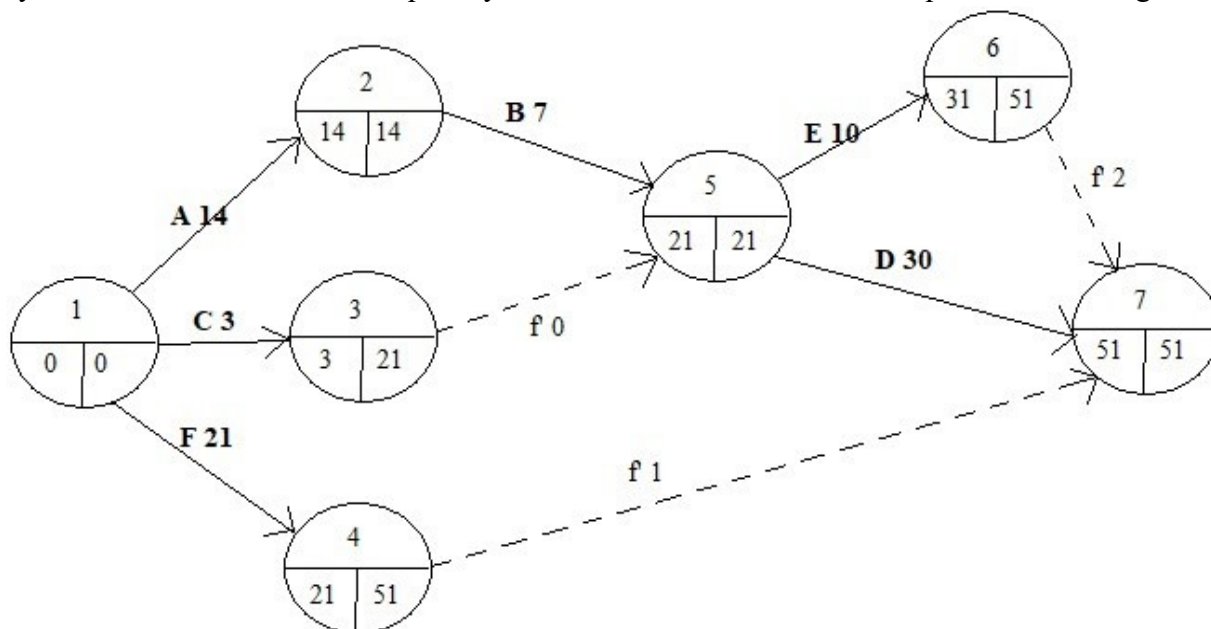


Figura 1: Diagrama PERT

De la imagen anterior, se puede apreciar que las tareas F, desarrollar la documentación, se puede desarrollar en paralelo a las demás tareas. La tarea A, estudiar la bibliografía existente, es necesaria para poder desarrollar la tarea B, estudiar distintas técnicas de parafraseo. La tarea C, elegir la metodología, se puede desarrollar de forma paralela a A y B.

Tras desarrollar las tareas A, B y C, se puede proceder a realizar las tareas E y D de forma paralela que propiamente son la implementación de la interfaz gráfica y la librería respectivamente.

Finalmente, la estimación de tiempo para la finalización del proyecto es de 51 días.

6.3. Diagrama de Gantt

Un diagrama de Gantt es una herramienta útil para planificar proyectos. Al proporcionarse una vista general de las tareas programadas, todas las partes implicadas sabrán qué tareas tienen que completarse y en qué fecha. Un diagrama de Gantt muestra: la fecha de inicio y finalización de un proyecto, qué tareas hay dentro del proyecto, la fecha programada de inicio y finalización de las tareas, una estimación de cuánto llevará cada tarea y cómo se superponen las tareas y/o si hay una relación entre ellas [21] . Con todo esto presente, el diagrama de Gantt de este proyecto queda como sigue:



Figura 2: Diagrama de Gantt

6.4. Análisis de costes

El análisis de costes se define como la medida de la relación coste-producción. Es decir, se preocupa por determinar el coste en el que se incurre al contratar los insumos, y qué tan bien se pueden reorganizar para aumentar la productividad de una empresa. En otras palabras, el análisis de costes se refiere a la determinación del valor monetario de los insumos (mano de obra, materia prima...) denominado como coste general de producción, que ayuda a decidir el nivel óptimo de producción [22]. Por tanto, el análisis de costes es fundamental en la toma de decisiones de negocios, ya que debe entenderse cuidadosamente el coste incurrido en la entrada y salida de producción, antes de planificar la capacidad de producción de la empresa. El análisis de costes promueve el conocimiento de la estructura de costes involucrada con los productos y servicios de una compañía. Se utiliza para fines de evaluación de costes cuando existe una falta de competencia u ofertas comparables en el mercado.

Para realizar un análisis de costes en este proyecto, lo primero será calcular el total de días, tanto de investigación como de implementación, que se han llevado a cabo para formalizar el proyecto.

Investigación y documentación: 45 días.

Implementación: 40 días.

NOTA: En el diagrama PERT y en el Gantt se puede apreciar que el proyecto en total conlleva 51 días, sin embargo esto es una aproximación suponiendo que las tareas se realizan paralelamente unas con otras como se ha expuesto en los gráficos anteriores. Por tanto, en la estimación de costes supondremos que las tareas se realizan de forma secuencial, y por ello salen más días en este apartado que los esperados.

Suponemos que el día se paga a unos 60 euros por lo que tenemos la siguiente aproximación de costes para este proyecto:

$$\text{Coste} = 85 \times 60 = 5100 \text{ euros.}$$

Así, se podría estimar que el precio del proyecto y trabajo realizado podría rondar sobre los 5100 euros.

7. Análisis del proyecto

En esta sección se intenta dar una aproximación del proyecto que se quiere realizar, es decir, se muestran los casos de uso que definen el sistema de paráfrasis implementado, se definen los requisitos necesarios para la elaboración del proyecto y se realiza un prototipo que explica por encima como será y funcionará el sistema planteado.

7.1. Casos de uso

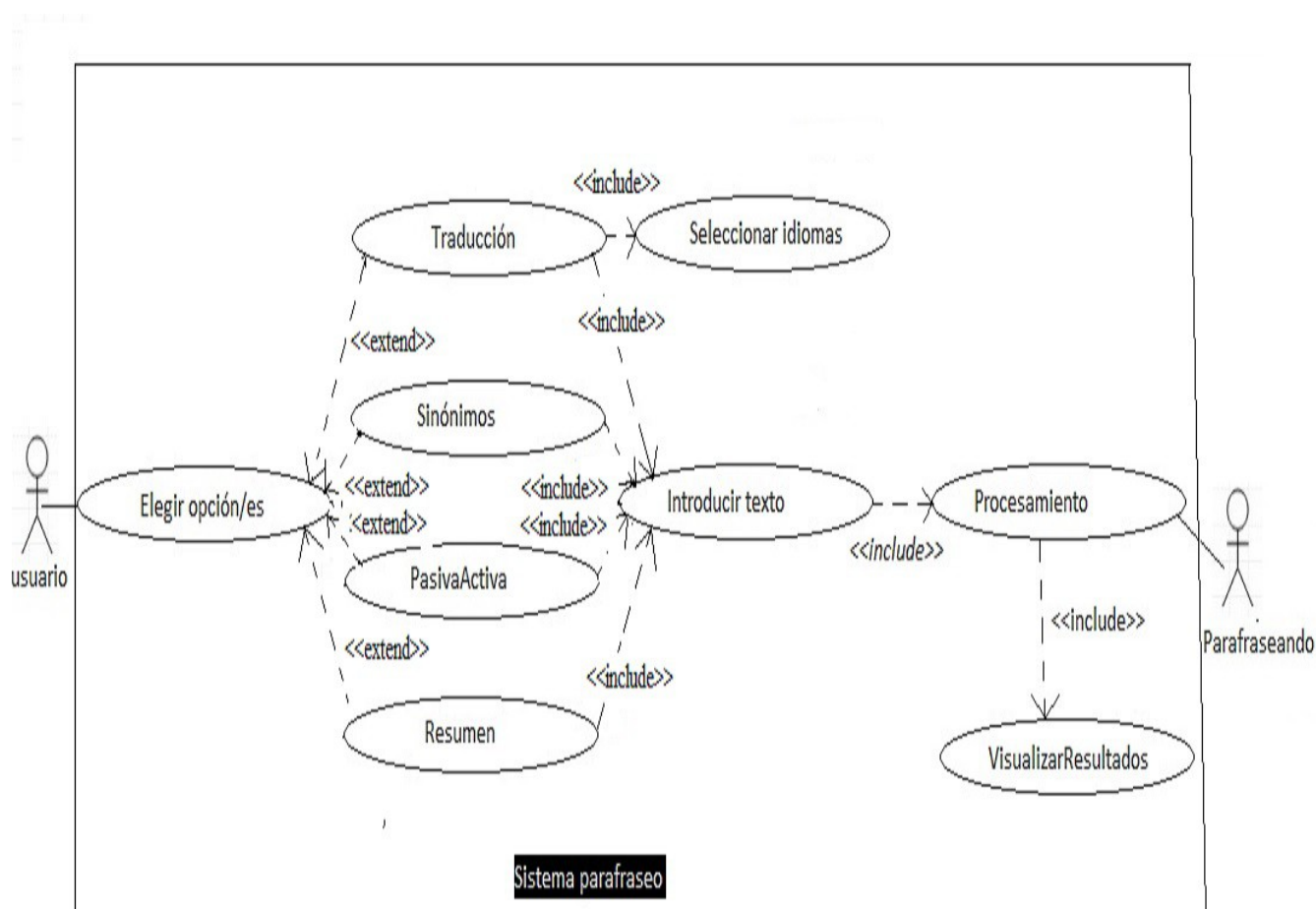


Figura 3: Casos de uso

7.2. Análisis de requisitos

7.2.1. Requisitos funcionales

- El sistema proporcionará cuatro opciones con las que poder parafrasear un texto de entrada.
- Las cuatro opciones dadas podrán ser inclusivas entre sí.
- Las opciones elegibles por el sistema son: traducción, sinónimos, pasiva y activa y/o resumen.
- Al seleccionar la opción traducción se deben especificar 3 idiomas con los que se desea traducir el texto en cuestión. El idioma por defecto es el inglés.
- Una vez seleccionadas las opciones, se deberá introducir un texto de entrada.
- Cuando el usuario haya seleccionado las opciones deseadas e introducido el texto de entrada, deberá pulsar el botón “Parafrasear” para llevar a cabo el procesamiento y la visualización de resultados.

7.2.2. Requisitos no funcionales

- El ordenador donde se ejecute el sistema debe tener instalado python.
- Además deben instalarse las librerías indicadas en el apartado siguiente para que funcione el sistema.
- El sistema debería de funcionar bajo los sistemas operativos Windows, Linux y MacOS.
- El sistema debería de funcionar bajo los navegadores Chrome, InternetExplorer y Firefox.
- La librería implementada (API.py) requiere del fichero moduloActivaPasiva.py y de los diccionarios verbosActiva.txt y verbosPasiva.txt para el correcto funcionamiento de la opción PasivaActiva.
- Es importante respetar la jerarquía de directorios (archivos .py en un lado y en la carpeta templates los archivos que dan lugar a la aplicación web) para que la aplicación funcione.

7.3. Prototipo

Desde el punto de vista de la informática, un prototipo de un sistema informático es una “muestra” más simplificada de un sistema. El prototipo nos permite ver cómo será un sistema. La idea del prototipo es entregar un resultado rápido de como se verá o será el sistema a diseñar, por lo tanto no habrá que esperar a que gran parte del proceso de desarrollo se termine para verlo.

Por lo general el prototipo se muestra al cliente/usuario para lograr cumplir con todos los requisitos necesarios. Sin la existencia de un prototipo, sólo se le podría mostrar modelos, códigos de programación, diagramas, etc, algo que no es fácil de entender por todos.

PARAFRASEANDO

☐ Traducción

Idioma1

Idioma2

Idioma3

☐ Sinónimos

☐ ActivaPasiva

☐ Resumen

Introduzca un texto a parafrasear...

Resultados obtenidos

Texto que se introdució para parafrasear antes de darle al botón "parafrasear"

Texto resultado de parafrasear un texto con la/s opción/es indicadas.

Figura 4: Prototipo

Se puede apreciar que dado el prototipo hay 4 opciones que elegir: Traducción (la cual se ve que mantiene 3 listas desplegables de idiomas), sinónimos, ActivaPasiva y Resumen. A continuación hay un cuadro de texto donde introducir un texto de entrada y más abajo el botón Parafrasear. Al pulsar el botón parafrasear con un texto de entrada arbitrario y alguna opción escogida, en Resultados obtenidos se hará una vista del texto que se introdujo (primera caja de texto) y el texto resultado de parafrasear dicho texto entrada con las opciones escogidas (segunda caja de texto).

8. Diseño e implementación

En esta sección se encontrarán los diferentes aspectos relacionados con la implementación del proyecto y su diseño. Se comenzará hablando de las tecnologías y librerías usadas. Posteriormente se expondrá el modelo de datos que se ha seguido para el desarrollo del proyecto, se comentarán algunos aspectos importantes de la implementación realizada y finalmente, se expondrán algunas pruebas y validaciones.

8.1. Tecnologías y librerías

El lenguaje de programación y las tecnologías que se han usado para desarrollar el proyecto han sido: Python, Html, Javascript y Css.

Las librerías que se han utilizado para desarrollar la librería de parafraseo son:

- Flask: Es un framework minimalista escrito en Python que permite crear aplicaciones web rápidamente y con un mínimo número de líneas de código.
- googletrans: Es una api de google para Python que permite la traducción de un texto en varios idiomas.
- Nltk: Es un conjunto de bibliotecas y programas para el PLN para el lenguaje Python.
- StanfordDependencyParser: Es un analizador de dependencias. Se usa para identificar las diferentes partes de una oración en este proyecto a la hora de convertir una oración pasiva en activa y viceversa, también se utiliza para localizar sustantivos y adjetivos en la opción sinónimos.
- urllib: Con esta librería se pueden obtener y manejar webs con Python. Concretamente, esta librería es usada en la opción de sinónimos, dada una palabra en concreto, con urllib se buscan los diferentes sinónimos de dicha palabra en una Web concreta.

8.2. Modelo de datos

El modelado de datos es el proceso de documentar un diseño de sistema software complejo como un diagrama de fácil comprensión, usando texto y símbolos para representar la forma en que los datos necesitan fluir. El diagrama se puede utilizar como un mapa para la construcción de un nuevo software.

El diagrama UML que se va a exponer sobre la aplicación desarrollada será un diagrama UML basado en componentes, puesto que la idea original es desarrollar una API con diferentes componentes a poder usar. Así, el diagrama UML de componentes quedaría como sigue:

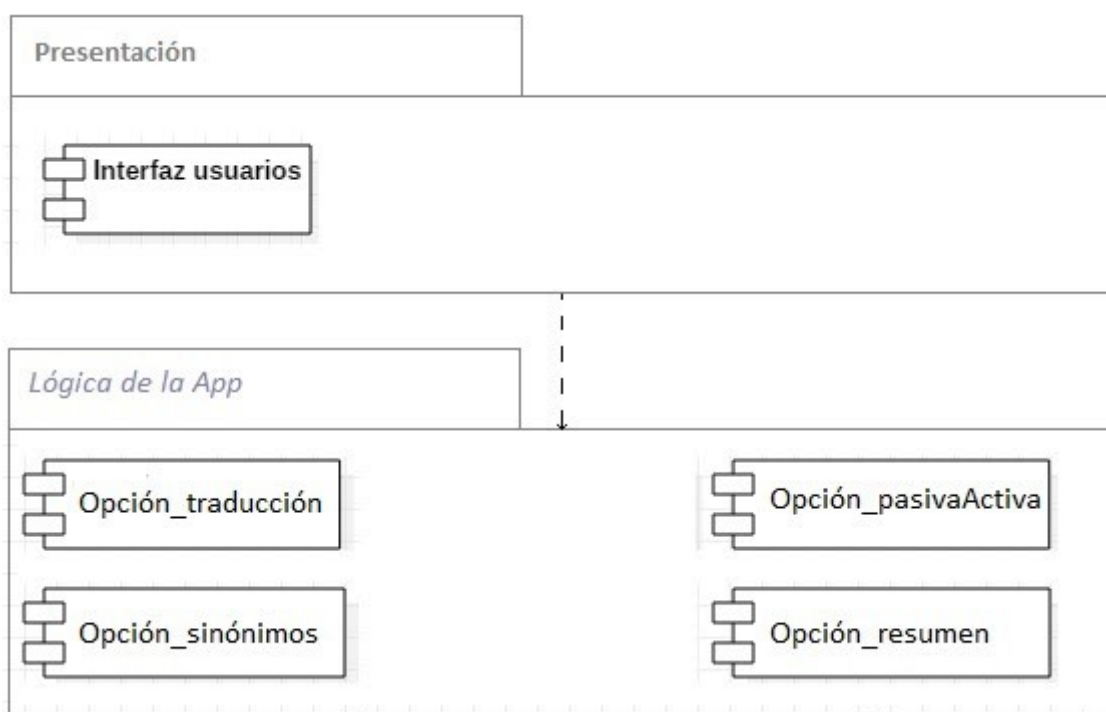


Figura 5: Modelo de datos

Como podemos apreciar, en la sección de presentación, la interfaz de usuarios es la interfaz web y gráfica que se ha implementado para poder usar y probar este proyecto.

Dentro de la lógica de la App podemos encontrar cuatro opciones para parafrasear un texto que se han implementado:

- Opción_traducción: Con esta opción un texto de entrada se traduce a 1, 2 o 3 idiomas diferentes y luego se vuelve a traducir al español.
- Opción_sinónimos: Con esta opción se busca en un texto de entrada todos los sustantivos y

adjetivos que se encuentren y dicha palabra se sustituye por un desplegable que contenga diferentes sinónimos de la palabra en cuestión.

- Opción pasivaActiva: Con esta opción, un texto de entrada se divide en oraciones simples. Tras separar las oraciones, esta opción permite pasar a pasiva una oración escrita en activa y viceversa..
- Opción resumen: Con esta opción, se hace un resumen del texto de entrada. Es decir, se divide el texto en oraciones y se cogen las aquellas más relevantes, usando el TF (término-frecuencia, las veces que sale un término en un texto) como medida para dar peso a cada una de las oraciones (se suman el TF de todos los términos de una oración y así se obtiene el peso de una oración en cuestión, luego se hace la media de los diferentes pesos de las oraciones y se establece un umbral con dicha media. Las oraciones que superen el umbral se consideran las más relevantes) de dicho texto de entrada quedando como resultado un resumen del texto original.

8.3. Implementación

Partiendo de la carpeta principal del proyecto, la estructura del proyecto será la siguiente:

```
AppWeb  
|  
|-requirements.txt  
|-configuration  
|+stanford-parse-full-2018-10-17  
|+-----  
|+templates  
|  |-index.html  
|  |-layout.html  
|-API.py  
|-init.py  
|-moduloActivaPasiva.py  
|-verbosActiva.txt  
|-verbosPasiva.txt
```

Figura 6: Estructura del proyecto.

Como se puede apreciar en la imagen anterior, existen dos carpetas en el proyecto: la carpeta que contiene el parser de stanford y la carpeta “templates” que contiene propiamente la estructura de la página web. A continuación se enumeran los archivos que contienen la implementación del proyecto:

- El archivo requirements.txt es el fichero donde se encuentran las librerías (junto con sus versiones) necesarias para ejecutar el proyecto. Se pueden instalar dichas librerías usando el comando: `pip install -r requirements.txt`
- El archivo configuration es el fichero dónde establecerse el `java_path`, es decir, un archivo de configuración dónde escribir la ruta dónde se encuentra java en el pc que ejecuta el proyecto.
- El archivo init.py es el fichero que se debe iniciar para poder iniciar el proyecto y en él se encuentra la definición de la aplicación web gracias al framework Flask.
- El archivo API.py contiene las diferentes funciones implementadas para parafrasear un texto. Dichas opciones como se ha dicho a lo largo del documento son: `opción_traducción`, `opción_sinonimos`, `opción_PasivaActiva` y `opción_resumen`. Cabe destacar que la opción `pasivaActiva` requiere de los ficheros `moduloActivaPasiva.py`, `verbosPasiva.txt` y

verbosActiva.txt que serán explicados más adelante en esta misma sección. Algunos comentarios a tener en cuenta sobre las diferentes opciones implementadas son:

- Opción_traducción: Se hace uso de la biblioteca googletans que permite, dado un texto de entrada, traducirlo a diferentes idiomas. En esta opción introducimos como parámetro el texto de entrada y los tres idiomas con los que se desea parafrasear (si se escogen los 3 idiomas iguales, por ejemplo, inglés en los 3 idiomas, entonces el parafraseo es como si se hiciera únicamente bajo el idioma inglés). Tras la traducción a los 3 idiomas se vuelve a traducir al español obteniendo así el texto resultado.
- Opción_sinonimos: Se hace uso de la biblioteca urllib con la que se buscan los sinónimos de los adjetivos y sustantivos del texto de entrada. La página dónde se buscan los sinónimos es www.sinonimos.es. Cabe destacar que el texto resultado se procesa de una forma especial para que en la aplicación web la palabra en cuestión se pueda sustituir por un desplegable que indique los posibles sinónimos por los que se puede sustituir.
- Opción_resumir: Dado un texto de entrada, coger las frases con más relevancia dentro del texto (como se ha explicado en el apartado anterior) y crear un resumen con dichas oraciones.
- Opción_pasivaActiva: En esta opción se transforma una frase de activa a pasiva y viceversa. Esta opción, además hará uso de los siguientes archivos:

Archivo verbosPasiva.txt: fichero de texto que se usa como diccionario para el moduloActivaPasiva.py donde se encuentra una colección de verbos en todas las formas de indicativo para todas las personas en forma pasiva.

Archivo verbosActiva.txt fichero de texto que se usa como diccionario para el moduloActivaPasiva.py donde se encuentra la misma colección de verbos que en el archivo anterior con todas las formas de indicativo para todas las personas pero esta vez en forma activa.

Archivo moduloActivaPasiva.py: Este módulo es el encargado de cargar los dos

diccionarios anteriores y en una oración, tras extraer el verbo gracias al parser de Stanford, se comprueba primero si el verbo está en pasiva o activa (en las oraciones pasivas el verbo va precedido de “por” en las activas no, más abajo se especifica un ejemplo de esto) y tras comprobar si es pasiva o activa se busca en el diccionario que corresponda el verbo y la forma del verbo en cuestión. Tras encontrarlo, se mapea ese verbo en el diccionario contrario para encontrar su forma activa o pasiva dependiendo de la transformación que se esté llevando a cabo. Finalmente, se crea la frase en su forma correspondiente teniendo en cuenta el siguiente razonamiento:

Supongamos que una frase se puede dividir en 4 partes:

Parte1 + verbo + Parte2 + resto.

Si la oración es pasiva, se transformaría a activa de la siguiente forma (una vez encontrado el verbo en la forma activa correspondiente):

TransformacionActiva: Parte2 + verboActiva + Parte1 + resto.

Si la oración es activa se transformaría a pasiva de la siguiente forma (una vez encontrado el verbo en la forma pasiva correspondiente):

TransformaciónPasiva: Parte2 + verboPasiva + por + Parte1 + resto.

Por ejemplo, supongamos que tenemos la frase: “La carta fue escrita por Juan el Lunes” tenemos así: Parte1 = La carta; verbo = fue escrita; Parte2 = Juan; resto = el Lunes. (NOTA: si es pasiva, la partícula “por” únicamente se tiene en cuenta a la hora de detectar el verbo)

Por tanto, la transformación a activa quedaría: Juan + escribió (se mapea el verbo en verbosActiva.txt) + la carta + el Lunes.

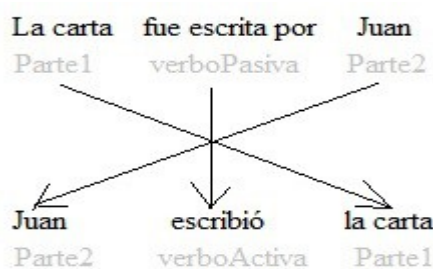


Figura 7: Ejemplo transformación pasiva-activa.

8.4. Pruebas y validación

En esta sección, se comentarán las pruebas realizadas. Esta parte es importante ya que a la hora de crear software, si no se hacen pruebas sobre el mismo no se pueden detectar errores propios de implementación así como de estructura del proyecto. Estas pruebas han permitido descubrir varios bugs en cuanto a la funcionalidad de la aplicación implementada se refiere, que han sido arreglados a posteriori. A continuación comentaré algunos problemas encontrados:

- Búsqueda de sinónimos en la web: este apartado resultó ser complejo, puesto que se buscaba una manera fácil de obtener sinónimos y extraerlos automáticamente de una web. De esta manera se decidió usar la web www.sinonimos.es ya que los resultados se obtenían en un bloque DIV de la plantilla HTML extraída y facilitaba el procesamiento de los resultados. Otra solución planteada, fue buscar una página web en la que los resultados se expresarán en formato JSON, pero tras buscar varias páginas me ha sido bastante difícil intentar encontrar una página que de los resultados de manera correcta en este formato.
- Traducción con googletans: la traducción de texto con googletans se hacía tedioso pues daba problemas de traducción de caracteres de otros idiomas. Sin embargo esto se trataba de un uso inadecuado de la librería lo que pude solventar releyendo el manual de la biblioteca en cuestión.
- Transformación de pasiva y activa: Las pruebas realizadas en la transformación pasiva a activa y viceversa es lo que más problemas ha dado ya que cada transformación tiene a la par que características comunes, características propias, como por ejemplo que de la transformación de activa a pasiva después del verbo se añada un “por” como se ha explicado antes, también tener en cuenta el género y la persona de los verbos, etc.
- Aplicación web: Respecto a la aplicación web cabría decir que el único problema que se ha encontrado y se ha solventado era el de mezclar el idioma javascript con el framework Flask. Se hacía imposible el paso de variables de python a javascript sin embargo encontré un camino intermedio en el que se podían compartir variables con el framework y javascript

dentro de la aplicación web y de esta manera se solucionó el problema

Un problema general que se ha encontrado en el uso de la aplicación es que solo admite un párrafo como texto de entrada. Es decir, si se tiene un texto que sea de la estructura:

parrafo1

espacio

parrafo2

entonces la aplicación no funciona. Esto es debido a que el framework Flask no es perfecto y tiene algunos fallos a la hora de intercambiar variables entre la aplicación web con la programación de la lógica de la aplicación en Python.

Tras hacer diferentes pruebas con la librería creada, y una vez terminada la API implementada el siguiente paso era crear una aplicación web gráfica con la que poder usar estas opciones de manera relativamente fácil. Gracias al framework Flask el desarrollo ha sido bastante más fácil que el de montar una página web desde cero y un servidor. La aplicación, una vez se inicia el archivo init.py es accesible de forma local y funciona en los siguientes navegadores: Google Chrome versión 81.0.4044.138 y Mozilla Firefox versión 75.0. Debido a la desactualización de mi navegador Internet Explorer, no se ha podido incluir este navegador como válido.

9. Conclusión.

La paráfrasis puede ser usada en diversos ámbitos informáticos específicos como sistemas de respuestas, sistemas de opinión, inteligencia artificial, etc. Esta técnica proporciona una salida a la mera reproducción de un texto en crudo proporcionando alternativas a lo que es la comprensión de un texto, haciendo de esta manera un texto incluso más legible.

Se han estudiado diferentes técnicas de paráfrasis que se pueden aplicar y como se ha visto estas técnicas se basan en el PLN, lo que puede resultar curioso para estudiar un texto desde diferentes puntos de vista, pasando por ejemplo un texto de alta complejidad técnica a un texto más comprensible por el lector.

En aplicaciones de inteligencia artificial, la paráfrasis puede resultar bastante interesante pues un mecanismo de paráfrasis bien formado podría sustituir al almacenamiento de respuestas de un robot en una base de datos por un sistema de parafraseo. Por ejemplo, si un sistema de inteligencia artificial está programado para decir buenos días en diez idiomas, se podrían sustituir ese “buenos días” en los 10 idiomas por el sistema de paráfrasis, así cada vez que se inicie una conversación con un bot éste dispondrá de diferentes mecanismos para alterar el lenguaje que utiliza sin tener que usar tanta memoria en el almacenamiento de frases arbitrarias. Bastaría con almacenar unos cientos de frases y el sistema de paráfrasis se encargaría de modificar cada vez que se requiera una frase en cuestión, en lugar de tener almacenadas en memoria diferentes frases para un mismo cometido. Lo dicho en este párrafo es aplicable a diferentes contextos informáticos, incluso un sistema de paráfrasis bien diseñado podría calcar de una manera distinta un código de programación.

En conclusión, se ha visto que esta técnica es bastante útil en diferentes campos informáticos e implementar un sistema de paráfrasis no es una tarea ardua ya que únicamente se toma un texto de entrada como parámetro y la técnica a utilizar. Son sistemas eficientes, eficaces y de bajo coste que hacen que un texto pueda tener más lecturas aparte de la original.

10. Futuras mejoras

En este proyecto tan sólo se han implementado cuatro técnicas de paráfrasis, pero se podrían implementar muchas más técnicas de las que se han visto dando lugar a una colección o API preparada para afrontar textos complejos y convertirlos de alguna manera en textos más simples. Se podría así diseñar incluso nuevas técnicas para parafrasear un texto de las mencionadas en este proyecto. También se podría mejorar la opción ActivaPasiva para que reconociera y trabajase con oraciones complejas, creando un diccionario con los diferentes conectores que existen y así poder dividir una oración compleja en frases simples y poder llevar a cabo esta transformación en cualquier tipo de oraciones.

Por otro lado, la aplicación web es bastante sencilla, se podría optar por una mejora en la misma para que se añadiera un grabador de voz. La grabación se transcribiría a texto, posteriormente se procesaría con técnicas elegidas de paráfrasis y se podría devolver un audio, con una voz preparada para tal fin, con el texto parafraseado en cuestión.

Por último, se podrían mezclar técnicas de paráfrasis con técnicas del lenguaje natural para detectar de una forma más sencilla y sutil negaciones, afirmaciones, preguntas, respuestas, etc, esta sería una gran mejora sobre todo cuando se trata de textos complejos. Así, daría lugar a un sistema en el que la búsqueda de ciertos patrones resulte más fácil de abordar que el texto primigenio.

11. Manual de usuario

En esta sección se encontrará un manual de usuario que define los pasos necesarios para ejecutar la aplicación. Se deberán seguir los pasos para un correcto uso de la aplicación implementada. Además, es importante destacar que para un correcto funcionamiento de algunas funciones, la estructura de carpetas mencionada en la sección 8.c deberá ser la misma, respetando así la posición y la versión del parser de Stanford usada. También es importante aclarar que al introducir texto, el texto introducido debe ser un único párrafo, esto es debido a que Flask tiene algunos errores respecto al paso de variables entre la aplicación web y Python.

1. Lo primero será instalar las librerías necesarias. Esto se consigue ejecutando “pip install x” dónde x es la librería en cuestión. En nuestro caso necesitamos las siguientes librerías: Flask, nltk, y googletrans.

```
IPython 7.8.0 -- An enhanced Interactive Python.
In [1]: pip install googletrans
Requirement already satisfied: googletrans in c:\programdata\
\anaconda3\lib\site-packages (2.4.0)
Requirement already satisfied: requests in c:\programdata\
\anaconda3\lib\site-packages (from googletrans) (2.22.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\
\programdata\anaconda3\lib\site-packages (from requests-
>googletrans) (2019.9.11)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\
\programdata\anaconda3\lib\site-packages (from requests-
>googletrans) (3.0.4)
```

Figura 8: Manual de usuario 1.

Este proceso se repetiría para todas las librerías antes mencionadas. Otra opción es usar "pip install -r requirements.txt". Además en el archivo configuration hay que escribir el path dónde se encuentre java.

2. El siguiente paso será ejecutar el archivo init.py, para ello es posible iniciar la aplicación desde la línea de comandos o usando algún intérprete de Python como Spider. Tras ejecutarlo, en pantalla saldrá algo parecido a lo siguiente:

```
In [2]: runfile('C:/Users/AJ/Desktop/TFG Parafrasear/App web/
init.py', wdir='C:/Users/AJ/Desktop/TFG Parafrasear/App web')
* Serving Flask app "init" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a
production deployment.
Use a production WSGI server instead.
* Debug mode: off
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Figura 9: Manual de usuario 2.

3. El siguiente paso será abrir GoogleChrome y poner en el buscador la siguiente ULR:
<http://127.0.0.1:5000/parafraseo>

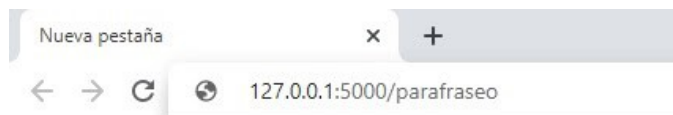


Figura 10: Manual de usuario 3

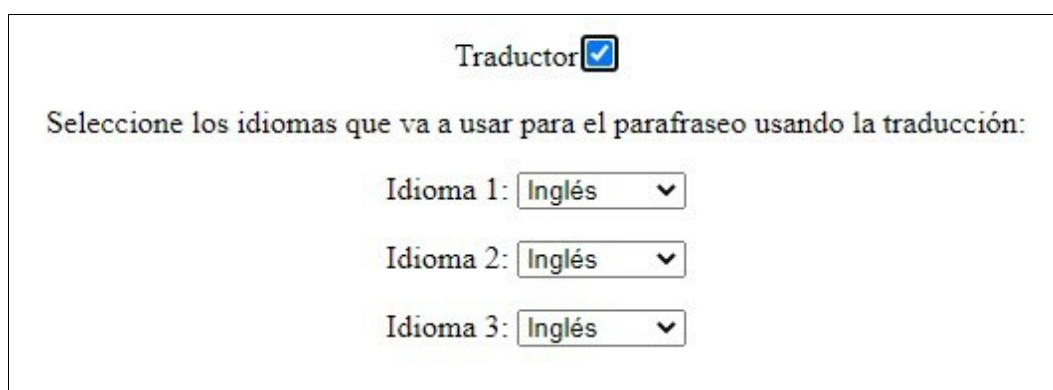
4. Tras ejecutar el paso anterior nos aparecerá la interfaz gráfica de la aplicación web implementada que es la siguiente:

A screenshot of a web application interface. At the top, the word 'PARAFRASEO' is displayed in a green, stylized font. Below it, a message reads 'Aquí va algo de información e introducción a la aplicación web.' In the center, there is a list of options: 'Traductor', 'PasivaActiva', 'Sinónimos', 'Resumir', and 'Otra', each followed by a small square checkbox. Below this list is a large text input area with the placeholder text 'Introduzca el texto a parafrasear aquí'. At the bottom of the input area is a button labeled 'Parafrasear'. Below the input area, the text 'RESULTADOS TRAS PULSAR "PARAFRASEAR"' is displayed in a green, stylized font. Underneath, it says 'Usted quiso parafrasear el siguiente fragmento' and 'Ha obtenido el siguiente resultado'.

Figura 11: Manual de usuario 4.

5. Ya estamos en la aplicación web. Ahora vamos a ver un ejemplo de uso. Se seleccionan en las checkbox las opciones que se desean usar en el parafraseo. Supongamos que en mi ejemplo de uso quiero usar la opción de sinónimos y la opción de PasivaActiva. (Se pueden usar las opciones individualmente, o en conjunto, en este caso yo uso dos opciones).

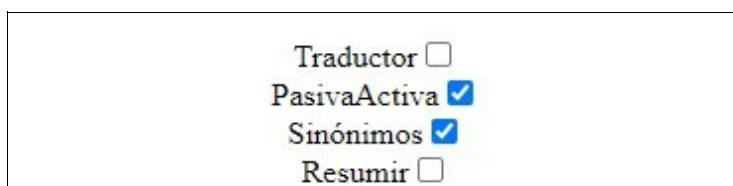
NOTA: Si se usa la opción Traductor, se verán tres desplegables donde se puede elegir cada idioma al que se desea parafrasear. Si en los tres desplegables se elige el mismo idioma, por ejemplo, el Inglés en los tres desplegables, contará como que se traduce al inglés y luego al español. La gracia es combinar diferentes idiomas para el parafraseo.



The screenshot shows a user interface for the 'Traductor' (Translator) option. At the top, the word 'Traductor' is followed by a checked checkbox. Below it, the text 'Seleccione los idiomas que va a usar para el parafraseo usando la traducción:' is displayed. Underneath, there are three dropdown menus labeled 'Idioma 1:', 'Idioma 2:', and 'Idioma 3:'. Each dropdown menu currently shows 'Inglés' and has a downward arrow icon.

Figura 12: Manual de usuario 5.

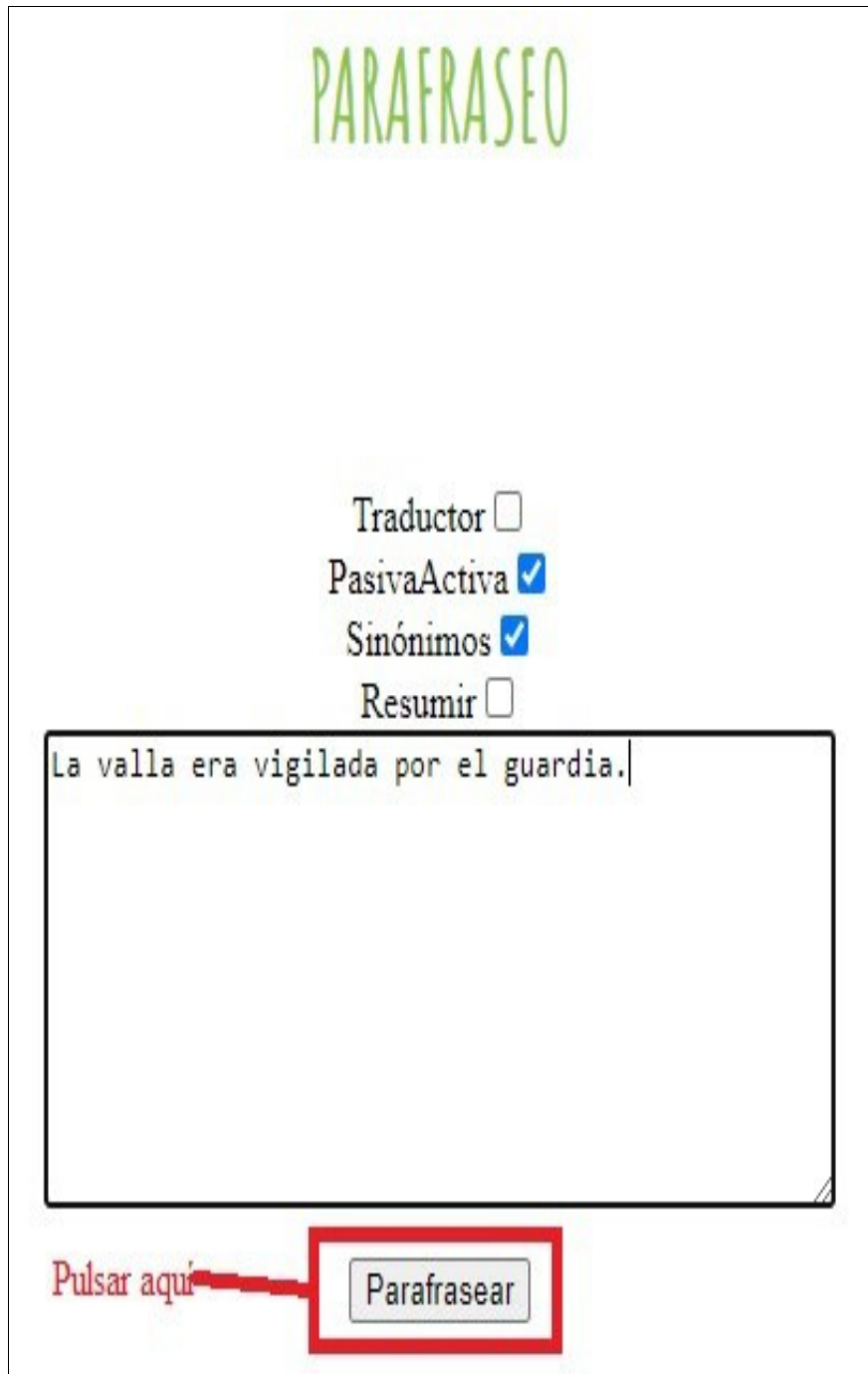
La imagen anterior muestra que pasa cuando se selecciona la opción traductor (para hacer más clara la nota anterior), sin embargo, en nuestro ejemplo de uso la configuración inicial queda como sigue en la siguiente imagen:



The screenshot shows a list of four options with checkboxes: 'Traductor' (unchecked), 'PasivaActiva' (checked), 'Sinónimos' (checked), and 'Resumir' (unchecked). The checkboxes for 'PasivaActiva' and 'Sinónimos' are blue and contain a white checkmark, while the others are empty.

Figura 13: Manual de usuario 6.

6. Ahora, una vez señaladas las opciones, se introduce el texto a parafrasear. Se le da al botón parafrasear y se obtienen los resultados. (Imagen antes de dar al botón parafrasear). El texto debe ser un único párrafo por problemas encontrados con Flask:



PARAFRASEO

Traductor ☐

PasivaActiva ☒

Sinónimos ☒

Resumir ☐

La valla era vigilada por el guardia.

Pulsar aquí  Parafrasear

Figura 14: Manual de usuario 7.

7. Tras pulsar el botón parafrasear, se refrescará la página y los resultados se pueden encontrar abajo de la página. En la sección, Resultados tras pulsar parafrasear.

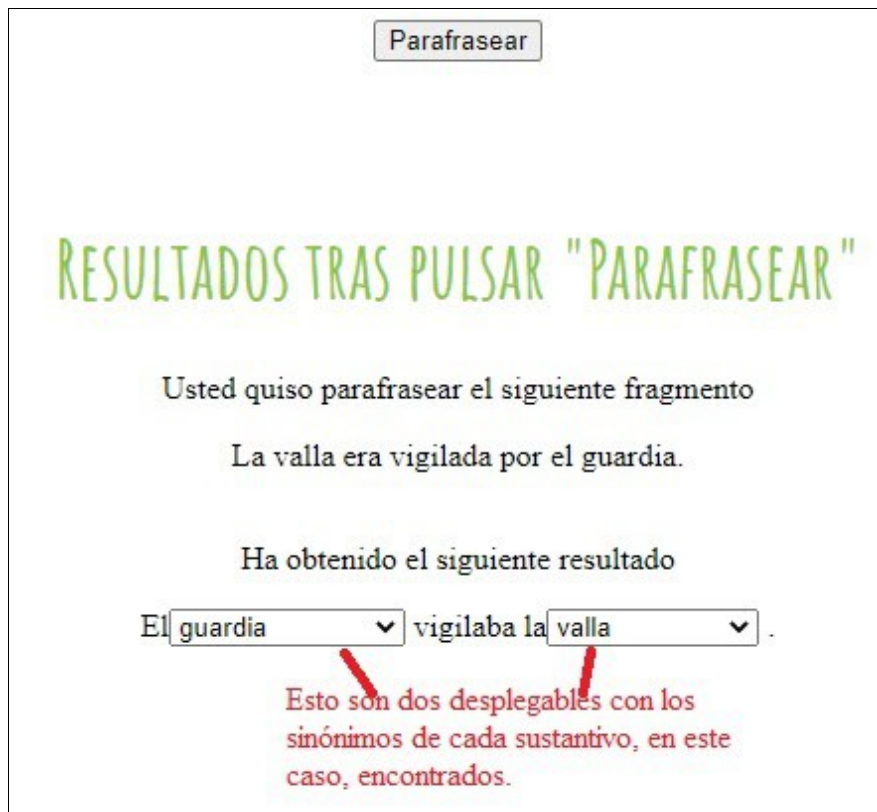


Figura 15: Manual de usuario 8.

Aquí han de hacerse algunas aclaraciones sobre la obtención de resultados de dos opciones:

- La opción pasivaActiva funciona con frases simples, de un solo verbo. No se contempla la posibilidad de transformar una frase con varios verbos o varias frases nominales separadas por conectores. Además, en algunos casos, la parte del resto puede no encontrarse ya que para ello haría falta un diccionario con todos los determinantes, pronombres, conectores intrafrases que existen.
- La opción resumen, como se ve en la imagen, genera una lista desplegable con los sinónimos de la palabra en cuestión. Se buscan sinónimos de sustantivos y adjetivos de la oración origen.

12. Referencias

- [1] Androutsopoulos, I., & Malakasiotis, P. (2010). A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38, 135-187.
- [2] MOTA MONTOYA, M. A., Da Cunha, I., & Lopez-Escobedo, F. (2016). Un corpus de paráfrasis en español: metodología, elaboración y análisis. *RLA. Revista de lingüística teórica y aplicada*, 54(2), 85-112.
- [3] McCarthy, D., & Navigli, R. (2009). The English lexical substitution task. *Language resources and evaluation*, 43(2), 139-159.
- [4] Hu, L., Ying, S., Jia, X., & Zhao, K. (2009, December). Towards an approach of semantic access control for cloud computing. In *IEEE International Conference on Cloud Computing* (pp. 145-156). Springer, Berlin, Heidelberg.
- [5] Mirkin, S., Specia, L., Cancedda, N., Dagan, I., Dymetman, M., & Szpektor, I. (2009, August). Source-language entailment modeling for translating unknown terms. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2* (pp. 791-799). Association for Computational Linguistics.
- [6] Paráfrasis - Concepto, tipos de paráfrasis y ejemplos. Disponible en línea: <https://concepto.de/parafrasis/>
- [7] Paráfrasis - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Par%C3%A1frasis>

[8]Procesamiento de lenguajes naturales - Wikipedia, la enciclopedia libre. Disponible en línea : https://es.wikipedia.org/wiki/Procesamiento_de_lenguajes_naturales

[9] SemEval - Wikipedia. Disponible en línea: <https://en.wikipedia.org/wiki/SemEval>

[10]Flexión (lingüística) - Wikipedia, la enciclopedia libre. Disponible en línea: [https://es.wikipedia.org/wiki/Flexi%C3%B3n_\(ling%C3%BC%C3%ADstica\)](https://es.wikipedia.org/wiki/Flexi%C3%B3n_(ling%C3%BC%C3%ADstica))

[11]Derivación (lingüística) - Wikipedia, la enciclopedia libre. Disponible en línea: [https://es.wikipedia.org/wiki/Derivaci%C3%B3n_\(ling%C3%BC%C3%ADstica\)](https://es.wikipedia.org/wiki/Derivaci%C3%B3n_(ling%C3%BC%C3%ADstica))

[12]Composición (lingüística) - Wikipedia, la enciclopedia libre. Disponible en línea: [https://es.wikipedia.org/wiki/Composici%C3%B3n_\(ling%C3%BC%C3%ADstica\)](https://es.wikipedia.org/wiki/Composici%C3%B3n_(ling%C3%BC%C3%ADstica))

[13] Hiperónimo - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Hiper%C3%B3nimo>

[14] Hipónimo - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Hip%C3%B3nimo>

[15] Holonimia - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Holonimia>

[16] Meronimia - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Meronimia>

[17] Actante - Wikipedia, la enciclopedia libre. Disponible en línea: <https://es.wikipedia.org/wiki/Actante>

[18] Elipsis (lingüística) - Wikipedia, la enciclopedia libre. Disponible en línea:
[https://es.wikipedia.org/wiki/Elipsis_\(ling%C3%BC%C3%ADstica\)](https://es.wikipedia.org/wiki/Elipsis_(ling%C3%BC%C3%ADstica))

[19] Oración de relativo - Wikipedia, la enciclopedia libre. Disponible en línea:
https://es.wikipedia.org/wiki/Oraci%C3%B3n_de_relativo

[20] Discurso directo e indirecto - Wikipedia, la enciclopedia libre. Disponible en línea:
https://es.wikipedia.org/wiki/Discurso_directo_e_indirecto

[21] ¿Qué es y para qué sirve un diagrama de Gantt?. Disponible en línea:
<https://blog.teamleader.es/diagrama-de-gantt>

[22] Análisis de costos: cómo se hace, para qué sirve y ejemplos - Lifeder. Disponible en línea:
<https://www.lifeder.com/analisis-de-costos/>