

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



Ontologías de dominio para dar soporte al proceso de
creación de diccionarios monolingües

Tesis para obtener el grado académico de Maestro en Informática
con mención en Ciencias de la Computación que presenta:

Carlos Alberto Rebaza Valdivia

Asesores:

César Armando Beltrán Castañón

Ian Paul Brossard Núñez

Lima, 2024


Informe de Similitud

Yo, **Ian Paul BROSSARD NUÑEZ**, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada "Ontologías de dominio para dar soporte al proceso de creación de diccionarios monolingües", de el autor **Carlos Alberto REBAZA VALDIVIA**, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 08%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 24/07/2024.
- He revisado con detalle dicho reporte y la tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

San Miguel, 24 de Julio de 2024.

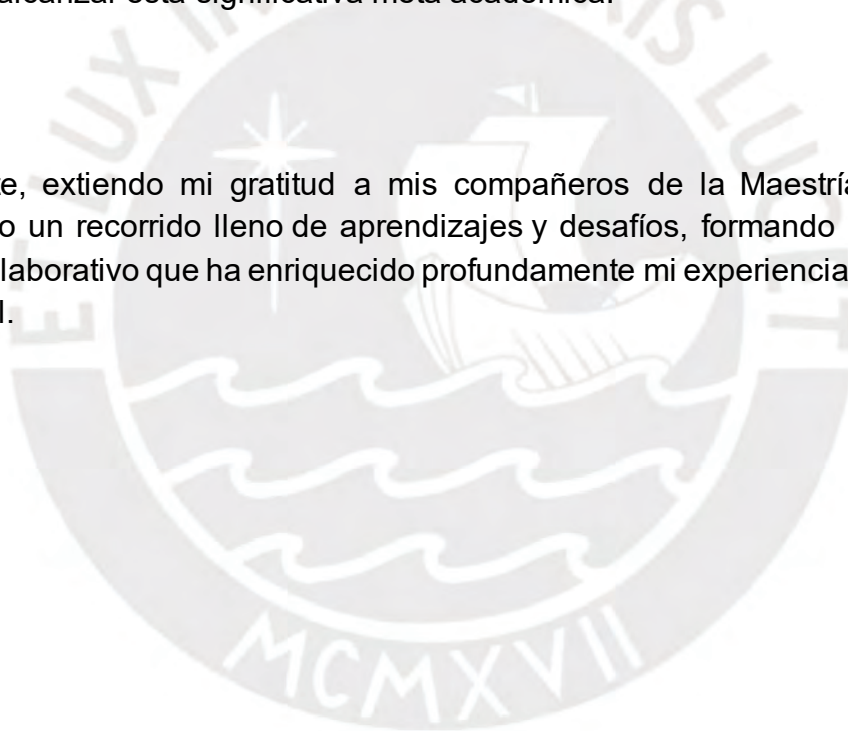
Apellidos y nombres del asesor / de la asesora: <u>BROSSARD NUÑEZ, Ian Paul</u>	
DNI: 46134777	Firma
ORCID: 0000-0003-2073-3829	

AGRADECIMIENTOS

Al reflexionar sobre el camino recorrido en esta maestría, mi primer agradecimiento es para Dios, cuya guía y fortaleza han sido mi faro en este viaje de crecimiento académico y personal. Su presencia ha sido esencial en cada etapa de mi investigación.

Mi familia ha sido el sostén emocional en este proceso. Su apoyo incondicional y confianza constante en mis capacidades me han impulsado a superar cada desafío y alcanzar esta significativa meta académica.

Finalmente, extiendo mi gratitud a mis compañeros de la Maestría. Hemos compartido un recorrido lleno de aprendizajes y desafíos, formando un equipo unido y colaborativo que ha enriquecido profundamente mi experiencia educativa y personal.



RESUMEN

Los diccionarios han sido, durante siglos, fundamentales para el entendimiento y la preservación de las lenguas. Fueron también la última palabra para resolver muchas discusiones, actuando como autoridades definitivas en cuestiones de significado y uso correcto de las palabras. La importancia de los diccionarios radica en su capacidad para proporcionar a los lectores una referencia confiable y precisa que facilita la comunicación efectiva y el aprendizaje. Sin embargo, la creación de un diccionario es una tarea ardua y meticulosa que puede llevar décadas en completarse. Este proceso implica la recopilación exhaustiva de datos léxicos, el análisis detallado de palabras y sus múltiples significados, y la verificación de su uso en diferentes contextos. Cada nueva edición de un diccionario requiere un esfuerzo considerable para incorporar cambios en el lenguaje, incluyendo la adición de nuevas palabras, la modificación de definiciones existentes y la eliminación de términos obsoletos. Esta labor intensa asegura que los diccionarios continúen siendo recursos valiosos y relevantes en un mundo en constante evolución.

La lexicografía, la disciplina dedicada a la elaboración y estudio de diccionarios, enfrenta numerosos desafíos en su práctica. Más allá del diseño de la estructura del diccionario, el lexicógrafo debe lidiar con la complejidad de la lengua, donde uno de los mayores retos es la polisemia. Las palabras con múltiples significados requieren un tratamiento cuidadoso para asegurar que las definiciones sean precisas y relevantes para el contexto en el que se utilizarán. Este proceso implica no solo identificar todos los posibles significados de una palabra, sino también determinar cuál de estos es más adecuado para el lector objetivo del diccionario. Además, el lexicógrafo debe asegurarse de que las definiciones sean claras y comprensibles, evitando ambigüedades y proporcionando ejemplos de uso que ilustren adecuadamente cada significado. Este desafío se amplifica en un entorno lingüístico dinámico donde el lenguaje evoluciona constantemente, haciendo imprescindible la utilización de herramientas avanzadas que apoyen en la toma de decisiones y en la estructuración eficiente de las entradas del diccionario.

Actualmente, tecnologías como TLex y Microsoft Word brindan herramientas que se enfocan en la edición y presentación de las entradas del diccionario, orientadas a detallar las definiciones seleccionadas y mejorar la calidad de su presentación. Sin embargo, el lenguaje es dinámico y evoluciona constantemente, comportándose como un ente vivo en constante evolución que refleja cambios sociales, culturales y tecnológicos. Palabras nuevas emergen, otras caen en desuso y los significados pueden transformarse con el tiempo. Este comportamiento evolutivo del lenguaje presenta un desafío significativo para la lexicografía tradicional.

La propuesta de esta tesis es tratar el corpus lexicográfico como un ente orgánico que evoluciona, integrando ontologías y folksonomías para gestionar y adaptar este dinamismo. Las ontologías proporcionan una estructura jerárquica y organizada del conocimiento, permitiendo representar de manera precisa las relaciones entre los términos. Por otro lado, las folksonomías, que son sistemas de clasificación colaborativa, permiten analizar el uso real del lenguaje de manera más flexible y adaptativa. Al apoyarse en grafos de conocimiento, es posible realizar análisis detallados y visualizaciones que ayudan a identificar tendencias, relaciones y cambios

en el uso del lenguaje. Este enfoque no solo facilita la actualización y mejora continua de los diccionarios, sino que también permite ofrecer definiciones más precisas y relevantes.

Herramientas como las nubes de palabras (wordclouds) pueden proporcionar al lexicógrafo valiosa información sobre las definiciones y su composición léxica. Estas herramientas visualizan la frecuencia de uso de las palabras, sugiriendo qué términos son más comunes y, por ende, más fácilmente comprendidos por los lectores. Esto permite al lexicógrafo identificar rápidamente cuáles definiciones están compuestas por palabras de uso más frecuente, facilitando la creación de entradas más accesibles y relevantes. De igual manera, si algunas palabras dentro de las definiciones pudieran estar sujetas a algún tipo de censura, esta información puede ser comunicada al lexicógrafo, permitiéndole tomar decisiones informadas sobre qué definiciones incluir en el diccionario, acorde al público objetivo. Al integrar estas herramientas en el proceso lexicográfico, se mejora la precisión y relevancia de las definiciones, asegurando que el diccionario cumpla con las expectativas y necesidades de sus lectores.

PALABRAS CLAVE: LEXICOGRAFIA, DICCIONARIO, ONTOLOGIA DE DOMINIO, TAXONOMIA , FOLKSONOMIA, GRAFOS DE CONOCIMIENTO, ALGORITMO TD-IDF



ABSTRACT

Dictionaries have been fundamental for the understanding and preservation of languages for centuries. They have also been the final authority to resolve many discussions, acting as definitive authorities on questions of meaning and correct use of words. The importance of dictionaries lies in their ability to provide readers with a reliable and precise reference that facilitates effective communication and learning. However, creating a dictionary is an arduous and meticulous task that can take decades to complete. This process involves the exhaustive collection of lexical data, detailed analysis of words and their multiple meanings, and verification of their use in different contexts. Each new edition of a dictionary requires considerable effort to incorporate changes in language, including the addition of new words, modification of existing definitions, and elimination of obsolete terms. This intense labor ensures that dictionaries remain valuable and relevant resources in a constantly evolving world.

Lexicography, the discipline dedicated to the creation and study of dictionaries, faces numerous challenges in its practice. Beyond designing the structure of the dictionary, the lexicographer must deal with the complexity of language, where one of the greatest challenges is polysemy. Words with multiple meanings require careful treatment to ensure that definitions are precise and relevant to the context in which they will be used. This process involves not only identifying all possible meanings of a word but also determining which of these is most appropriate for the target reader of the dictionary. Additionally, the lexicographer must ensure that definitions are clear and understandable, avoiding ambiguities and providing usage examples that adequately illustrate each meaning. This challenge is amplified in a dynamic linguistic environment where language constantly evolves, making it essential to use advanced tools that support decision-making and efficient structuring of dictionary entries.

Currently, technologies like TLex and Microsoft Word provide tools that focus on the editing and presentation of dictionary entries, aimed at detailing selected definitions and improving the quality of their presentation. However, language is dynamic and evolves constantly, behaving like a living entity that reflects social, cultural, and technological changes. New words emerge, others fall into disuse, and meanings can change over time. This evolutionary behavior of language presents a significant challenge for traditional lexicography.

The proposal of this thesis is to treat the lexicographic corpus as an evolving organic entity, integrating ontologies and folksonomies to manage and adapt to this dynamism. Ontologies provide a hierarchical and organized structure of knowledge, allowing precise representation of relationships between terms. On the other hand, folksonomies, which are collaborative classification systems, allow for more flexible and adaptive analysis of the actual use of language. By leveraging knowledge graphs, it is possible to perform detailed analyses and visualizations that help identify trends, relationships, and changes in language use. This approach not only facilitates the continuous updating and improvement of dictionaries but also enables the provision of more precise and relevant definitions.

Tools like word clouds can provide lexicographers with valuable information about definitions and their lexical composition. These tools visualize the frequency of word usage, suggesting which terms are more common and therefore more easily understood by readers. This allows lexicographers to quickly identify which definitions are composed of more frequently used words, facilitating the creation of more accessible and relevant entries. Similarly, if some words within definitions could be subject to some form of censorship, this information can be communicated to the lexicographer, allowing them to make informed decisions about which definitions to include in the dictionary, according to the target audience. By integrating these tools into the lexicographic process, the precision and relevance of definitions are improved, ensuring that the dictionary meets the expectations and needs of its readers.

KEYWORDS: LEXICOGRAPHY, DICTIONARY, DOMAIN ONTOLOGY, TAXONOMY, FOLKSONOMY, KNOWLEDGE GRAPHS, TF-IDF ALGORITHM



INDICE GENERAL

Contenido	
CAPITULO 1	12
1. Generalidades.....	12
1.1. Introducción.....	12
1.2. Objetivo General	16
1.3. Objetivos específicos	16
1.4. Resultados Esperados.....	17
1.5. Delimitación del Proyecto	17
CAPITULO 2	20
2. Marco Conceptual	20
2.1. Lexicografía.....	20
2.2. Ontologías en Informática.....	25
2.3. Ontologías de Dominio.....	25
2.4. Ontologías en Lexicografía	26
2.5. Grafos de Conocimiento.....	26
2.6. Taxonomía.....	27
2.7. Folksonomía.....	28
2.8. Tripletas Semántica.....	28
2.9. WordClouds.....	28
2.10. Algoritmo TF IDF.....	28
CAPITULO 3	30
3. Estado del Arte.....	30
3.1. Herramientas para la creación de Diccionarios.....	30
3.1.1. TLEX.....	30
3.1.2. WordNet	30
3.2. Revisión Sistemática	31
3.2.1. Planeamiento	31
3.2.2. Ejecución	31
3.2.3. Reporte de Resultados	31
3.2.4. Protocolo de Investigación	32
CAPITULO 4	36

4.	Metodología en Uso de Ontologías en la Creación de Diccionarios.....	36
4.1.	Jerarquía de Datos de Taxonomía	36
4.2.	Folksonomía: Un Enfoque de Crecimiento Orgánico	37
4.3.	Modelado de la Ontología.....	38
4.4.	Actividades Realizadas y Uso de la Ontología	44
4.5.	Adaptación Dinámica de Contenido.....	55
4.6.	Desarrollo de Matriz Relacional.....	56
4.7.	Implementación de la Ontología	58
CAPITULO 5		70
5.	Conclusiones	70
CAPITULO 6		71
6.	Bibliografía	71



INDICE DE FIGURAS

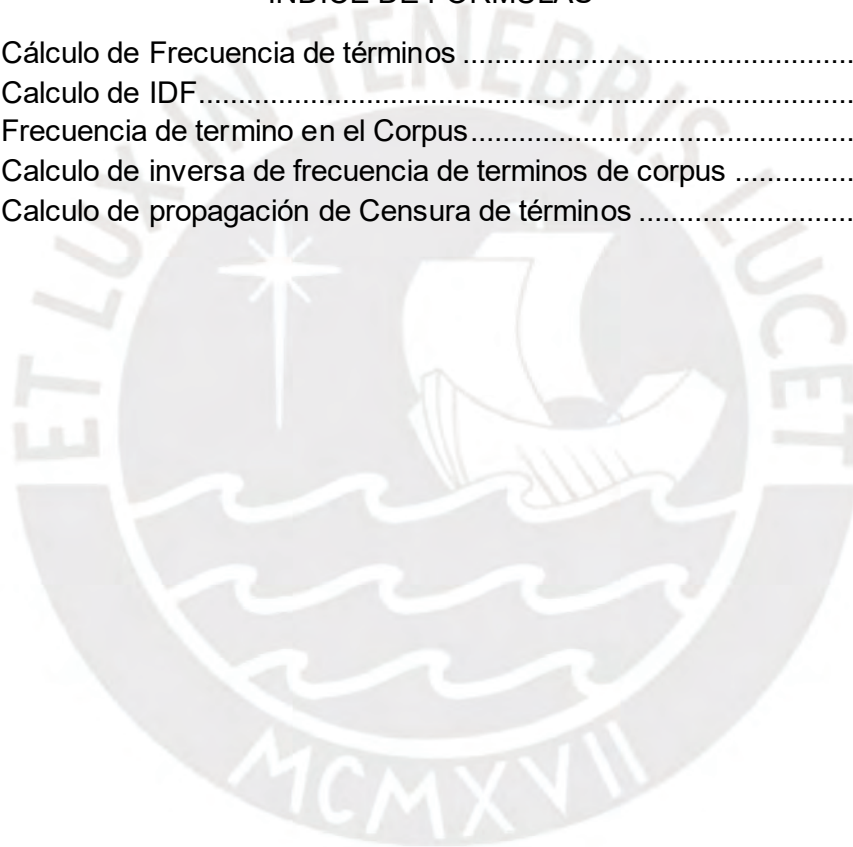
Figura 1 Ejemplo de múltiples definiciones - fuente wikidiccionario.org	13
Figura 2 Ejemplo de wordcloud - fuente elaboración propia	13
Figura 3 Ejemplo de palabra común de diferentes definiciones- elaboración propia ...	14
Figura 4 Wordcloud de corpus	14
Figura 5 Ejemplo de selección de definición para palabra "operación" - elaboración propia	15
Figura 6 Wordcloud de candidatos de definición a "operación "	15
Figura 7 Ejemplo de definición para público objetivo	23
Figura 8 Muestra de una entrada léxica.....	24
Figura 9 Jerarquía de clases de ontología - Elaboración Propia	38
Figura 10 Representación grafica de relaciones de clase PALABRA - Elaboración propia	40
Figura 11 Representación grafica de Relaciones de clase Meta_Frase	42
Figura 12 Representación grafica de relaciones de clase Definición	43
Figura 13 Representación grafica de relaciones de clase Entrada	44
Figura 14 Flujo de alimentación del Corpus	45
Figura 15- Flujo de Creación de Definición – Elaboración propia	46
Figura 16- Flujo de Creación de Entrada- Elaboración propia	48
Figura 17 Prototipo de Wordcloud de definiciones de la palabra Caminar	49
Figura 18 Prototipo de stopwords con alta frecuencia de uso	51
Figura 19 Inclusión de fase de Calculo orgánico.....	52
Figura 20 Flujo de ingreso de datos al corpus con inclusión de cálculo de frecuencia – Elaboración propia	54
Figura 21 Flujo de Ingreso con cálculo de Censura – Elaboración propia.....	55
Figura 22 Prototipo de Ajuste Dinámico de Censura.....	56
Figura 23 Prototipo de contenido adaptado dinámicamente	56
Figura 24 Prototipo de Matriz de relación	57
Figura 25 Flujo de funcionamiento de matriz relacional - Elaboración propia	57
Figura 26 Prototipo de Prompt de ingreso de datos	59
<i>Figura 27 Confirmación de Entrada Prompt</i>	<i>60</i>
<i>Figura 28 Almacenamiento de Metafrase “Esto es una prueba del uso de Prompt”</i>	<i>61</i>
<i>Figura 29- Reporte de confirmación de ingresos de recursos</i>	<i>61</i>
<i>Figura 30 Resultado de almacenamiento de 2 metafrases similares en grafo</i>	<i>63</i>
Figura 31 Ingreso de instrucción para generación de Definiciones en cascada de Matriz con ChatGPT.....	65
Figura 32 Generación de definiciones relacionadas a palabra "Dar" por ChatGPT	66
Figura 33 Inferencia por emparejamiento de definiciones con el uso de Matriz	67
Figura 34 Ejemplo de consulta al grafo NEO4J.....	67
Figura 35 Implementacion de Algoritmo en NextJs , Javascript y Nodejs	68
Figura 36 Score de coherencia con ChatGPT	69

INDICE DE TABLAS

Tabla 1 – Términos de búsqueda	33
Tabla 2 Uso de Términos revisión general.....	33
Tabla 3 Términos de búsqueda revision general	33
Tabla 4 Términos de búsqueda con mayor relevancia	34
Tabla 5 Ejemplo de frecuencia de términos	53
Tabla 6 Valores de Frecuencia	53
Tabla 7 Calculo del TF IDF de los términos	53
Tabla 8 comparación de palabras similares entre dos metafrases	62
Tabla 9 Conteo de aristas después de ingreso de metafrase	63
Tabla 10 porcentaje de resultados de score con ChatGPT.....	69

INDICE DE FORMULAS

Fórmula 1 Cálculo de Frecuencia de términos	53
Fórmula 2 Calculo de IDF.....	53
Fórmula 3 Frecuencia de termino en el Corpus.....	54
Fórmula 4 Calculo de inversa de frecuencia de terminos de corpus	54
Fórmula 5 Calculo de propagación de Censura de términos	55



CAPITULO 1

1. Generalidades

1.1. Introducción

La construcción de diccionarios monolingües y académicos se enmarca en el ámbito de la Lexicografía, disciplina encargada del estudio sistemático de las palabras de un idioma. Esta práctica se centra en la compilación, descripción y ordenamiento de términos, incorporando sus definiciones, usos, etimologías y variantes. Abarca un meticuloso análisis lingüístico que refleja la evolución y riqueza del vocabulario, y es esencial para la enseñanza y preservación de la lengua. Los lexicógrafos, a través de este proceso, no solo documentan la lengua, sino que también contribuyen a su expansión y adaptación a nuevos contextos y realidades culturales.

Desde la década de 1960, el uso de computadoras en la elaboración de diccionarios se ha vuelto cada vez más central. Hoy en día, los buenos diccionarios comienzan con datos de corpus como punto de partida, y los lexicógrafos contemporáneos dependen de una variedad de tecnologías, incluyendo computadoras personales con gran capacidad de almacenamiento y acceso rápido a Internet, datos de corpus procesados con herramientas de software desarrolladas en la comunidad de Procesamiento del Lenguaje Natural, y software para la entrada de texto de diccionarios y bases de datos que almacenan y gestionan este texto a medida que se desarrolla (Atkins & Rundell, 2008).

Es fundamental comprender los diferentes tipos de diccionarios existentes. Los diccionarios monolingües, por ejemplo, ofrecen definiciones, sinónimos, antónimos, ejemplos de uso y otra información relevante relacionada con las palabras de un solo idioma. Estos diccionarios son una herramienta esencial para entender y utilizar correctamente el vocabulario de un idioma específico. Por otro lado, los diccionarios bilingües proporcionan traducciones entre dos idiomas, facilitando la comunicación y el aprendizaje de idiomas extranjeros.

Además de los tipos de diccionarios, también es necesario considerar la importancia de la meta información o metadata en la construcción de diccionarios. La meta información incluye datos como el origen de las definiciones, el nivel de registro (formal o informal), el campo de aplicación (ciencia, medicina, derecho, etc.) y la especialización temática. Estos elementos permiten a los lectores contextualizar y utilizar adecuadamente las definiciones, ya que cada término puede tener diferentes acepciones y matices dependiendo del contexto y, en especial, del público objetivo.

Un desafío particular en la lexicografía es la gestión de la polisemia, la existencia de múltiples significados para una sola palabra. La palabra "banco", por ejemplo, puede referirse tanto a una institución financiera como a un asiento. La polisemia requiere un tratamiento cuidadoso para asegurar que las definiciones sean precisas y relevantes para el contexto en el que se utilizarán. Las herramientas actuales permiten a los lexicógrafos analizar y editar estas múltiples acepciones, pero en última instancia, la decisión sobre qué definición incluir y cómo presentarla recae en el criterio experto del lexicógrafo.

dar 65 idiomas

Entrada Discusión Leer Editar Ver historial Herramientas

Entradas similares: -dar, DAR, Dar, dar-, dař, dár, dâr, dâr, dâr

Español [editar]

Etimología 1 [editar]
Del castellano antiguo *dar* ("dar"), y este del latín *do*.

Verbo transitivo [editar]

- 1
Traspasar algo a otra persona de forma gratuita.
 - **Sinónimo:** donar
- 2
Hacer entrega de algo.
- 3
Otorgar alguna cosa.
 - **Ejemplo:** Le han dado el diploma de la carrera.
- 4
Considerar que algo es de una determinada manera. Se usa junto con la preposición *por*.
 - **Ejemplo:** Doy *por* terminado el contrato.
- 5
Proporcionar frutos o beneficios.
 - **Ejemplo:** Estas acciones me han dado mucha rentabilidad.
- 6
Mostrar en público un espectáculo como por ejemplo emitir una película o programa de televisión.
 - **Ejemplo:** ¿Qué dan hoy por la tele?
- 7
Marcar el reloj una hora, especialmente referido a las campanadas.
 - **Ejemplo:** Casi han dado las doce y el fontanero aún no ha venido.

dar

pronunciación (AFI)	[ˈdar]
silabación	dar
acentuación	monosílaba
longitud silábica	monosílaba
rima	ar

Figura 1 Ejemplo de múltiples definiciones - fuente wikidiccionario.org

Como se observa en la figura 1, la palabra "dar" es una de las que más definiciones tiene en la lengua española. Con el proceso de construcción tradicional, el lexicógrafo puede elegir qué definiciones utilizar basándose en diferentes factores, como su experiencia y el público objetivo al que va dirigido el diccionario. Sin embargo, este conjunto de definiciones suele estar compuesto por frases estáticas almacenadas en bases de datos que sirven como fuente de las entradas o lemas del diccionario. Revisar y actualizar estas definiciones es un proceso arduo, lo que provoca que nuevas versiones de diccionarios tomen décadas en completarse.

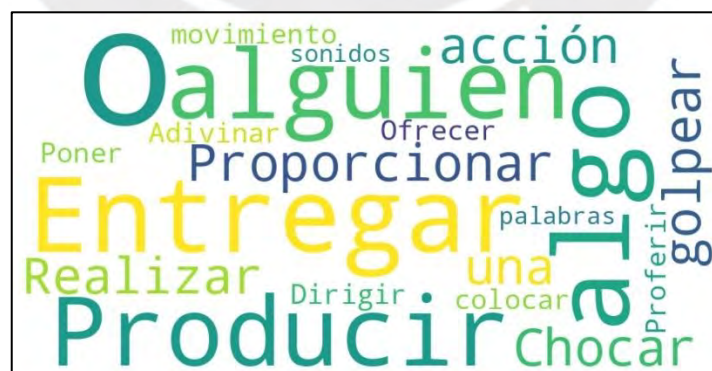


Figura 2 Ejemplo de wordcloud - fuente elaboración propia

Como se muestra en la figura 2, un wordcloud muestra la frecuencia de palabras. Esta propiedad puede ser explotada en el proceso de construcción de diccionarios, ya que utilizar términos de uso más frecuente en una definición brinda **una decisión**

más informada para el lexicógrafo al momento de escoger las entradas de un diccionario. De esta manera, en forma de cascada, las palabras marcadas con uso frecuente generan una definición comprensible que, a su vez, retroalimenta el corpus indicando qué palabras son de uso más frecuente en las propias definiciones.

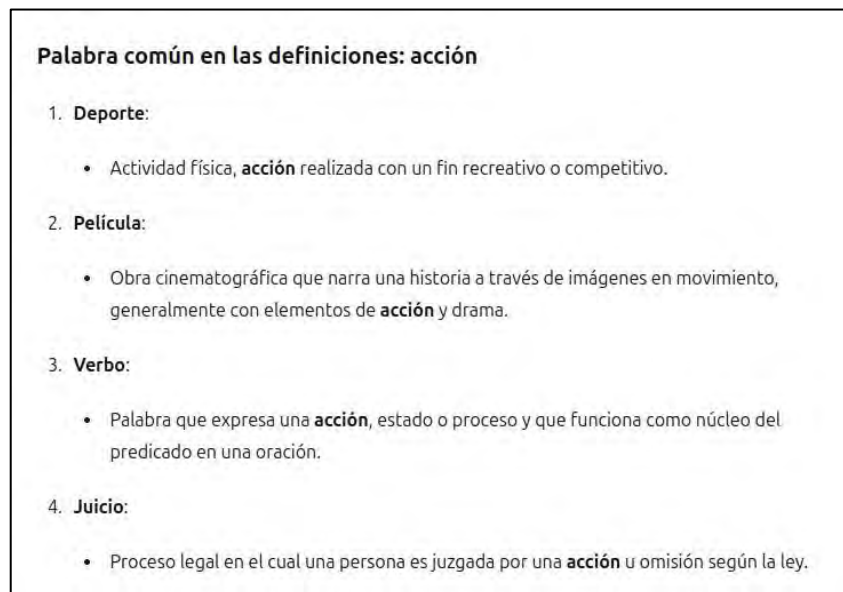


Figura 3 Ejemplo de palabra común de diferentes definiciones- elaboración propia

Como se puede observar en la figura 3, la palabra “**acción**” aparece en múltiples definiciones en diferentes entradas para diferentes palabras. Por lo que el corpus debe entenderlo como una palabra de uso frecuente y posiblemente de mejor comprensión.



Figura 4 Wordcloud de corpus

Como se puede ver en la figura 4, el estado del corpus estaría apuntando a que la palabra acción es de uso mas frecuente, por lo que aquellas definiciones que contengan esta palabra podrían heredar esta frecuencia de uso de forma que el lexicógrafo podría tomar una mejor decisión en cuanto a que definición usar para una entrada del diccionario.

El Capítulo 1 de esta tesis establece el marco fundamental del proyecto, presentando el objetivo principal y los objetivos específicos que guían la investigación. En este capítulo se delimita claramente el alcance del proyecto, definiendo tanto su enfoque como sus limitaciones. También se aborda la justificación del uso de la ontología como herramienta principal para el estudio, resaltando por qué es la metodología elegida para abordar los desafíos y las necesidades identificadas en el campo de la lexicografía digital.

El Capítulo 2 presenta un marco conceptual fundamental, detallando los conceptos clave y las teorías subyacentes que son esenciales para la comprensión y el desarrollo de esta tesis. En este capítulo se expone una explicación fundamentada sobre la construcción de diccionarios, incluyendo sus distintos tipos y las metodologías empleadas en su elaboración. También se abordan temas críticos como la ontología, la taxonomía y la folksonomía, examinando cómo estos conceptos interactúan y contribuyen significativamente al campo de la lexicografía digital.

El Capítulo 3 se dedica a presentar el estado del arte, proporcionando una revisión sistemática exhaustiva de los avances y desarrollos clave en los campos relacionados con la tesis. Este capítulo emplea una revisión rigurosa para examinar la literatura existente, identificando y analizando hitos importantes en la evolución de temas como la lexicografía digital, la ontología, la taxonomía y la folksonomía. A través de esta revisión se destaca cómo han progresado estos campos a lo largo del tiempo.

El Capítulo 4 se centra en detallar las actividades y la metodología adoptadas en el desarrollo de la tesis. Este capítulo comienza con una descripción de cómo los insights obtenidos de la revisión de la literatura informan y guían la metodología del proyecto. Se explica el proceso de modelado de la ontología, destacando cómo se estructura y se integra dentro del marco de la investigación. También se describen las actividades específicas llevadas a cabo para cumplir con los objetivos de la tesis, incluyendo el desarrollo de herramientas, la recopilación y el análisis de datos, y la implementación de la ontología en contextos prácticos.

El Capítulo 5 describe las conclusiones y recomendaciones derivadas de esta investigación. Se presentan las principales conclusiones del estudio, destacando el impacto positivo de la implementación de ontologías y folksonomías en la lexicografía digital, mejorando la precisión y adaptabilidad de las definiciones en los diccionarios. Además, se abordan los desafíos enfrentados, como la gestión de la polisemia y la variabilidad en la calidad de las definiciones, y se explican las soluciones implementadas para mitigar estos problemas.

1.2. Objetivo General

Establecer un marco basado en ontologías de dominio para la creación y manejo de un corpus lexicográfico, empleando la folksonomía para capturar definiciones y enriquecer el criterio lexicográfico, asegurando una organización altamente escalable que brinde soporte de decisión añadida dinamismo al contenido.

1.3. Objetivos específicos

1.3.1. Optimizar el Manejo de Polisemia (OE1)

- 1.3.2. Adaptar Dinámicamente Contenido (OE2)
- 1.3.3. Diseñar una Matriz Relacional de Definiciones (OE3)

1.4. Resultados Esperados

- 1.4.1. Para el Objetivo Específico 1 se plantea diseñar una ontología que facilite la parametrización de definiciones, aprovechando métodos de minería de texto y algoritmos de frecuencias de términos. La expectativa es que este enfoque proporcione una selección más precisa y contextualizada de definiciones, destacando aquellas que mejor representen el uso y el significado actual de palabras polisémicas. Este resultado no solo mejorará la calidad de las definiciones en el diccionario, sino que también ofrecerá una herramienta valiosa para lexicógrafos y lectores, facilitando una comprensión más profunda y matizada del lenguaje.
- 1.4.2. Para el Objetivo Específico 2 se plantea adecuar el contenido del diccionario de acuerdo con las necesidades específicas del público objetivo. Empleando las propiedades de las ontologías de dominio, se determinará el peso semántico y la pertinencia de los términos y definiciones, facilitando así una mejor comprensión de los términos y convenciones lexicográficas. Esta adaptación también permitirá implementar una censura adecuada, asegurando que el contenido sea apropiado y respetuoso para diferentes sectores del público objetivo. El resultado esperado es un diccionario que no solo proporcione una representación precisa y actualizada del lenguaje, sino que también se adapte de manera dinámica a las variaciones culturales y normativas, convirtiéndose en un recurso lexicográfico invaluable y sintonizado con su audiencia específica.
- 1.4.3. Para el Objetivo Específico 3 Se plantea crear una Matriz Relacional innovadora que permita asociar eficazmente los corpus lexicográficos. Esta matriz, diseñada mediante un algoritmo, facilitará la vinculación de definiciones y el uso de ontologías de dominio para identificar de manera orgánica posibles polisemias o para la creación de nuevas definiciones. Al analizar y comparar definiciones, la matriz podrá inferir relaciones semánticas complejas entre palabras, permitiendo así una mejor comprensión de las conexiones y matices del lenguaje. Este enfoque promoverá un análisis lexicográfico más profundo y una expansión del corpus, ofreciendo una herramienta poderosa para el enriquecimiento y la precisión en la lexicografía.

1.5. Delimitación del Proyecto

1.5.1. Alcance

El alcance de este proyecto se enfoca en el manejo de léxicos y corpus lexicográficos, orientado específicamente hacia el desarrollo de diccionarios monolingües en español. El proyecto pone un énfasis particular en diccionarios académicos escolares, destinados a un público escolar. La capacidad para adaptar dinámicamente el contenido es clave, ya que permite ajustar el nivel de censura y la relevancia de las definiciones. Esto será crucial para filtrar aquellas definiciones que no sean adecuadas según criterios preestablecidos, lo cual es especialmente valioso para la creación

de diccionarios de bolsillo, donde la concisión y pertinencia del contenido son esenciales. La adaptabilidad de la ontología facilitará la actualización y refinamiento continuos del corpus lexicográfico, asegurando que el diccionario se mantenga actualizado y relevante para su audiencia objetivo, y se adapte a las variaciones culturales y normativas.

1.5.2. Riesgos

Uno de los principales riesgos del proyecto radica en la naturaleza orgánica y colaborativa de la folksonomía, que, aunque enriquece el corpus lexicográfico, también puede introducir un elemento de informalidad y variabilidad en las definiciones. Esto podría llevar a inconsistencias o a una calidad variable en las entradas del diccionario. Además, el manejo de stopwords representa otro desafío significativo: si no se trata adecuadamente, podría sesgar la relevancia de ciertas palabras, afectando la precisión y la objetividad del análisis lexicográfico. Estos riesgos requieren una atención meticulosa para asegurar que el enfoque colaborativo y el procesamiento del lenguaje se optimicen y alineen con los objetivos académicos y educativos del proyecto, manteniendo un equilibrio entre la riqueza de datos y la rigurosidad lexicográfica.

1.5.3. Justificación

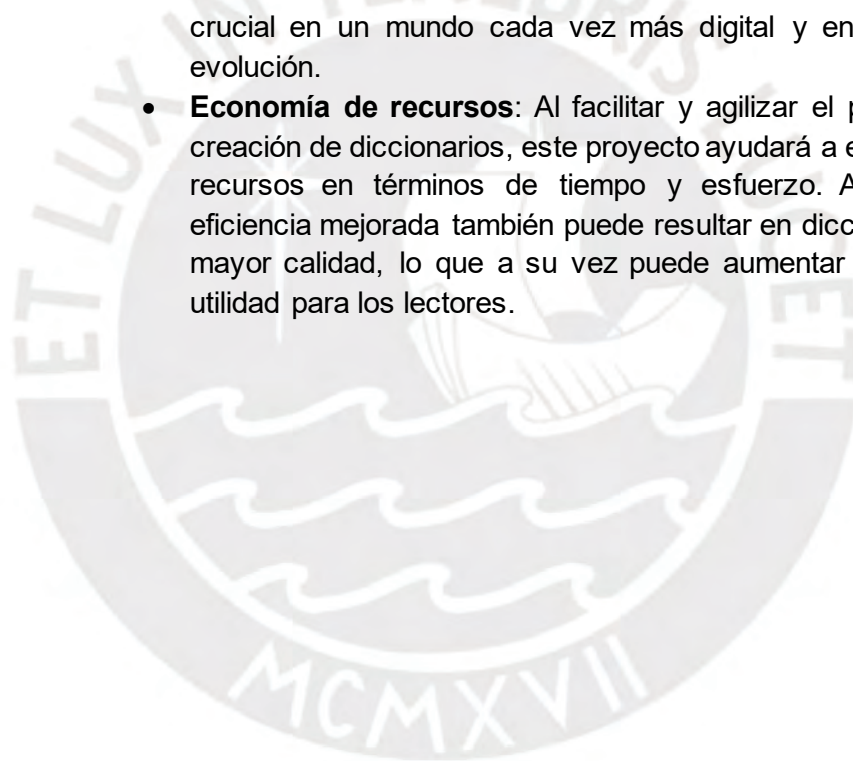
1.5.3.1. Motivación

La elaboración de diccionarios, un proceso central en la lexicografía es históricamente un proceso complejo y prolongado. Esto queda evidenciado en la creación de ediciones monumentales como el Diccionario de Oxford, cuya primera edición requirió 70 años y la segunda más de 60 años para su completitud. Incluso diccionarios de alcance más limitado, como los regionales, enfrentan desafíos similares debido al uso de métodos tradicionales, convirtiendo su elaboración en una tarea ardua y prolongada. En este contexto, la principal motivación de este proyecto es introducir enfoques y herramientas contemporáneas que agilicen significativamente este proceso. La integración de ontologías de dominio y folksonomía tiene el potencial de mejorar considerablemente la eficiencia en la lexicografía, facilitando la creación de diccionarios más dinámicos y actualizados en un plazo más reducido. Esta innovación representa un avance técnico importante y una respuesta necesaria a la rápida evolución del lenguaje y a las demandas de agilidad y eficiencia en la era digital. (*History of the OED*, s. f.)

1.5.3.2. Uso de Ontologías de Dominio

Las ontologías son una representación formal y sistemática de un conjunto de conceptos y las relaciones entre ellos. Su uso en este proyecto permitirá un manejo más eficiente y dinámico del contenido del diccionario, lo que a su vez facilitará la adaptación del diccionario a diferentes públicos y contextos. Además, las ontologías permitirán una mejor comprensión y representación de los símbolos y marcas lexicográficas con lo que se puede resaltar:

- **Aporte a la lexicografía:** Este proyecto ayudará a modernizar y optimizar el proceso de creación de diccionarios. Al utilizar una plataforma digital y ontologías, se mejora la eficiencia, la precisión y la flexibilidad del proceso de creación de diccionarios, lo que es particularmente valioso dada la creciente complejidad y dinamismo de las lenguas modernas.
- **Desafío de la polisemia:** La polisemia es uno de los desafíos más grandes en lexicografía. El uso de una estructura ontológica facilitará el manejo de la ambigüedad semántica y mejorará la precisión y claridad de las definiciones del diccionario. Esto permitirá a los lectores del diccionario comprender y utilizar el lenguaje de manera más efectiva.
- **Adaptabilidad y agilidad:** Este proyecto permitirá que los diccionarios sean más adaptables y ágiles, lo que significa que podrán ser actualizados y personalizados más fácilmente para satisfacer las necesidades cambiantes de los lectores. Esto es crucial en un mundo cada vez más digital y en constante evolución.
- **Economía de recursos:** Al facilitar y agilizar el proceso de creación de diccionarios, este proyecto ayudará a economizar recursos en términos de tiempo y esfuerzo. Además, la eficiencia mejorada también puede resultar en diccionarios de mayor calidad, lo que a su vez puede aumentar su valor y utilidad para los lectores.



CAPITULO 2

2. Marco Conceptual

En esta sección se muestran los conceptos relacionados a la presente investigación para asegurar su comprensión.

2.1. Lexicografía

Según Emilia Anglada (Universidad de Barcelona, 1991), la lexicografía es "el arte de crear léxicos". Según Almansa Ibáñez (2022), la lexicografía es el estudio de los diccionarios y su compilación, organización y presentación de información lexical sobre un idioma. Esto incluye el análisis de significados, uso, etimologías, propiedades fonéticas y gramaticales, y el desarrollo histórico de las palabras. La lexicografía también abarca el estudio de vocabularios especializados, como vocabularios científicos y técnicos, y el tratamiento de marcas humorísticas en los diccionarios. La historia de la lexicografía está estrechamente vinculada con la presentación y tratamiento de las terminologías, ya que los repertorios lexicográficos han incluido, desde sus orígenes, aquellas palabras de especialidad que han ido apareciendo a lo largo de los siglos en función de los avances científicos y técnicos.

2.1.1. Problemas que afronta un Lexicógrafo

Según (Atkins & Rundell, 2008) alguno de los mayores desafíos que afronta un lexicógrafo son:

- Selección de palabras: Es deber del lexicógrafo decidir qué palabras incluir en un diccionario. Esto puede ser especialmente complicado para diccionarios de idiomas en evolución constante, como el inglés, donde se crean nuevas palabras y frases regularmente.
- Determinar definiciones precisas: El lexicógrafo debe ser capaz de proporcionar una definición precisa y concisa de cada palabra. Esta es una tarea desafiante, ya que muchas palabras tienen múltiples significados dependiendo del contexto.
- Uso correcto de las palabras: El lexicógrafo también debe proporcionar ejemplos del uso correcto de una palabra. Esto puede ser un desafío, especialmente para palabras con múltiples significados.
- Estar al día con el cambio lingüístico: Los idiomas están en constante evolución, y el lexicógrafo debe mantenerse al día con estos cambios. Esto incluye el seguimiento de nuevas palabras y frases, así como cambios en el significado o uso de las palabras existentes.
- Decisiones sobre la inclusión de jerga y lenguaje ofensivo: El lexicógrafo debe tomar decisiones sobre si incluir o no la jerga, las palabras ofensivas y los términos despectivos. Esto puede ser un desafío, ya que la decisión de incluir estas palabras puede ser controvertida.

- Factores culturales y regionales: Los lexicógrafos también deben tener en cuenta las variaciones culturales y regionales en el uso del lenguaje. Una palabra puede tener diferentes significados en diferentes culturas o regiones.

Según las palabras de Sven Tarp, "la distinción fundamental entre lexicografía y lingüística radica en que abordan campos de estudio completamente diferentes: el objeto de estudio de la lingüística es el lenguaje, mientras que la lexicografía se ocupa de los diccionarios y las obras lexicográficas en general" ("Más allá de la lexicografía" en *Lexicography at a Crossroads*, 2009).

En 1971, Ladislav Zgusta, un lingüista histórico y lexicógrafo, publica el "Manual de Lexicografía" el primer manual internacional importante sobre lexicografía, considerado hasta la actualidad por muchos como la referencia estándar en el campo.

2.1.2. Publicación de un Diccionario

La historia de los diccionarios en Europa ofrece una perspectiva fascinante sobre el desarrollo de la lexicografía moderna, iniciando en Italia y avanzando a través de Francia hasta llegar a España, con cada país contribuyendo significativamente a la evolución de esta disciplina.

En Italia, destaca el "Vocabolario degli Accademici della Crusca", cuya primera edición fue publicada en 1612 por la Accademia della Crusca. Esta obra es reconocida como el primer diccionario de una lengua moderna en Europa.

En Francia, la Academia Francesa marcó un hito con la publicación en 1694 del "Dictionnaire de l'Académie française", un proyecto que buscaba normalizar y regular el francés.

Finalmente, en España, la Real Academia Española introdujo el "Diccionario de autoridades" en 1726, un diccionario influenciado por la autoridad de escritores clásicos y contemporáneos, esencial para establecer normas en el uso del español.

2.1.3. Tipos de Diccionario

Existen diversos tipos de diccionarios que abarcan desde sinónimos, antónimos, diccionarios inversos, ideológicos, de vocabulario técnico, de jergas, de dudas e incorrecciones, etimológicos, de dialectalismos, e incluso diccionarios de crucigramas, nombres de pila o insultos. Sin embargo, los dos tipos más relevantes para el lector común son los diccionarios generales y los diccionarios bilingües. En ambos casos, se trata de obras que prometen al lector una cobertura exhaustiva del vocabulario de un idioma y proporcionan información detallada en cada entrada.

Diccionario General: El Diccionario de la Real Academia Española (DRAE) es considerado el diccionario general de referencia de la lengua española. Otros diccionarios se basan en él, ya sea para seguirlo o imitarlo, o para marcar

diferencias. El DRAE adopta una filosofía normativa en la que solo se incluye en el léxico de la lengua lo que aparece en su última edición. Esta postura normativa puede resultar irritante para algunos hablantes, quienes consideran que las palabras son creadas por la gente en la calle y no por académicos. Si bien el léxico de cualquier lengua no puede ser infinito debido a las limitaciones cognitivas del cerebro, la última edición del DRAE contiene alrededor de 83,500 entradas. Aunque se puedan agregar palabras de otros diccionarios y dominios terminológicos, nunca se acercará al infinito. Por lo tanto, es más preciso decir que el léxico de una lengua tiene un límite incierto y nadie puede determinar con objetividad el número exacto de palabras. Esta incertidumbre contrasta con la actitud autosuficiente del gramático normativo, que pretende tener conocimiento sobre lo que forma parte de la lengua y lo que queda fuera.

Diccionario Escolar: Un diccionario escolar es un diccionario diseñado específicamente para niños en edad escolar. Tiene un vocabulario limitado que incluye las palabras más comunes utilizadas en español y ofrece definiciones simples y fáciles de entender. Además, se proporcionan ejemplos para ayudar a los niños a comprender el significado de las palabras en contexto, y a menudo se incluyen ilustraciones para visualizar el significado de las palabras. Los diccionarios escolares son herramientas valiosas que ayudan a los niños a ampliar su vocabulario, comprender nuevas palabras y mejorar sus habilidades de escritura y comprensión de lectura.

Diccionarios bilingües: Un diccionario bilingüe brinda palabras en dos idiomas. Cada idioma se agrupa alfabéticamente en mitades separadas del libro, con traducciones al otro idioma.

- Un diccionario bilingüe unidireccional, como su nombre lo indica, va "en una dirección". Por ejemplo, un diccionario bilingüe inglés-francés contiene un único texto en el cual el idioma de origen sería el inglés y el idioma objetivo sería el francés.
- Un diccionario bilingüe bidireccional contiene dos textos y funciona "en ambas direcciones". En un diccionario bilingüe inglés-francés hay un texto en el que el idioma de origen es el inglés y el idioma objetivo es el francés, y un segundo texto en el que el idioma de origen es el francés y el idioma objetivo es el inglés.

Diccionarios de crucigramas: Un diccionario de crucigramas agrupa las palabras según el número de letras en la palabra para ayudar a las personas a encontrar palabras de una cierta longitud y así completar sus crucigramas.

Diccionario de rimas: Un diccionario de rimas es aquel en el que las palabras se agrupan según sus sonidos finales. Cuando dos palabras terminan con el mismo sonido, riman, y se utilizan con mayor frecuencia en poesía que en prosa para crear efectos sonoros.

Tesoro: Un tesoro es un libro que organiza las palabras por categorías y conceptos, de manera que los sinónimos y los casi sinónimos se agrupan.

2.1.4. Público Objetivo

La creación de un diccionario implica la toma de decisiones: se deben tomar decisiones importantes en la etapa de planificación y decisiones más pequeñas a medida que avanza el proyecto. Muchas de estas decisiones implican alguna forma de selección, ya que cada diccionario contiene un subconjunto de toda la información disponible sobre el idioma objetivo y su vocabulario. Por ejemplo, en cualquier momento dado del proceso editorial, se puede tener que decidir si se incluye una palabra específica y, de ser así, cuánta información se proporciona al respecto. (He & Yiu, 2022)

<p>clam-ber / .../ v [] always + adv/prep] to climb or move slowly somewhere, using your hands and feet because it is difficult or steep: [+over/across etc] <i>They clambered over the slippery rocks.</i> [] <i>We all clambered aboard and the boat pulled out.</i></p> <p>LDOCE-4 (2003)</p>	<p>clamber / .../ (clambers, clambering, clambered) If you clamber somewhere, you climb there with difficulty, usually using your hands as well as your feet. □ <i>They clambered up the stone walls of a steeply terraced olive grove ...</i></p> <p>VERB = scramble V prep/adv</p> <p>COBUILD-5 (2006)</p>
---	---

Figura 7 Ejemplo de definición para público objetivo

Como se observa en la figura 2, los diccionarios LDOCE-4 (2003) y COBUILD-5 (2006) están diseñados para el mismo público objetivo: estudiantes de inglés avanzado. Ambos diccionarios muestran definiciones ilustrativas usando diferentes formatos o marcas y ambos indican que la palabra "clamber" suele ir seguida de un complemento adverbial o preposicional.

- LDOCE sostiene que sus usuarios no necesitan que se les informe sobre la morfología (regular) de este verbo, pero COBUILD proporciona un conjunto completo de flexiones.
- LDOCE enumera dos partículas que suelen seguir a "clamber" ("over" y "across"). COBUILD no lo hace, pero, en cambio, ofrece un sinónimo cercano ("scramble").
- LDOCE utiliza un estilo de definición convencional, mientras que COBUILD opta por una definición en forma de oración completa.

Esto brinda un panorama de las notables diferencias que se pueden hallar entre diccionarios que ocupan el mismo nicho de mercado bien definido. Todas estas diferencias reflejan elecciones editoriales tomadas durante la fase de planificación. Existen dos métodos para comprender al usuario: la investigación de usuarios y creación de perfiles de usuarios. El proceso no siempre es científico, pero el único punto de partida posible es el grupo objetivo de usuarios. Es fundamental comprender de forma clara quiénes utilizarán el diccionario, con qué propósito lo utilizarán y qué habilidades aportarán a la tarea. Si se obtienen respuestas a todas estas interrogantes, se establece una base sólida para tomar decisiones bien informadas sobre el contenido y la presentación.

2.1.5. Estructura de un Diccionario

La **macroestructura** se refiere a la organización general del diccionario y cómo se presentan las entradas y la información relacionada. Esto incluye la disposición de las palabras, la estructura de las definiciones, la inclusión de ejemplos y otros elementos en cada entrada, así como la organización alfabética o temática de las palabras en el diccionario.

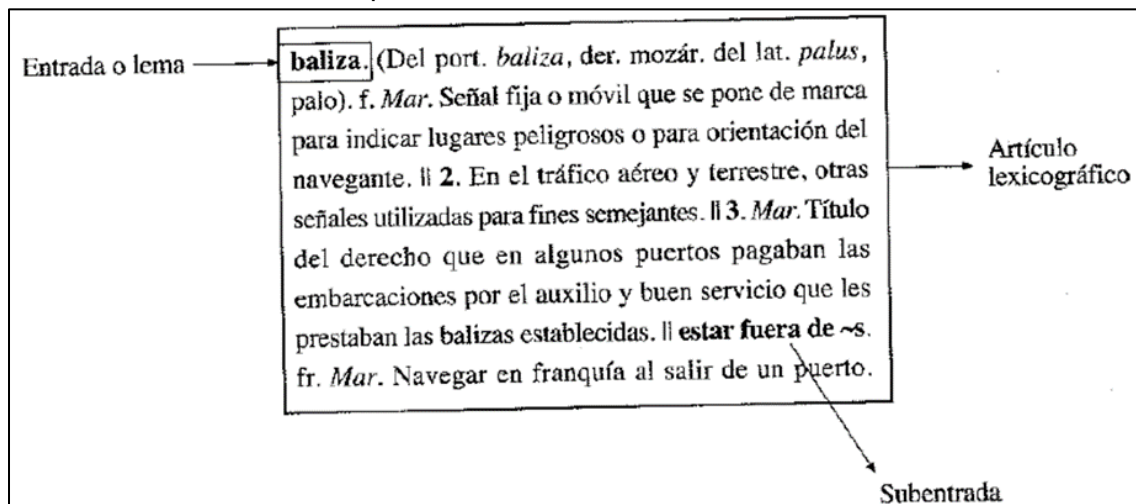


Figura 8 Muestra de una entrada léxica

Como se muestra en la figura 3, la entrada léxica contiene:

- En sentido estricto: unidad que es objeto de un artículo lexicográfico independiente en el diccionario.
- En sentido lato: cualquier unidad léxica sobre la que el diccionario, sea en su macroestructura o microestructura, ofrece información.

El término "macroestructura" se aplica a una lista de entradas, que es solo parte de la estructura general de un diccionario. Dado que un diccionario puede contener más de una lista de palabras, también puede tener más de una macroestructura. Los principios de ordenación pueden ser sistematizados de la siguiente manera:

- Por propiedades intrínsecas del signo del lenguaje
 - Por propiedades del significante
 - Propiedades ortográficas: orden alfabético
 - Dirección de Orden
 - Orden ascendente
 - Orden descendente
 - Propiedades fonológicas
 - Propiedades estructurales: por ejemplo, estructura morfológica
 - Por propiedades del significado: orden semántico
- Por criterios extrínsecos: orden por frecuencia

La **microestructura** se refiere a los detalles y la organización interna de cada entrada individual en el diccionario. Esto incluye la forma en que se presentan los diferentes elementos dentro de una entrada, como la posición de la definición, los ejemplos de uso, los sinónimos y otros detalles gramaticales o lingüísticos relacionados con la palabra.

2.1.6. Corpus o Corpora

Según John Sinclair, un pionero en el uso de corpora para la lexicografía, un corpus es "una colección de piezas de texto en forma electrónica seleccionadas según criterios externos para representar, en la medida de lo posible, un idioma o variedad lingüística como fuente de datos para la investigación lingüística" (Sinclair 2005: 16). Esta definición subraya que un corpus es una herramienta de investigación compuesta por textos seleccionados que buscan representar un idioma o variedad lingüística para el análisis lingüístico.

2.2. Ontologías en Informática

En ciencia de la computación y ciencia de la información, una ontología es una definición formal de tipos, propiedades y relaciones entre entidades que real o fundamentalmente existen para un dominio de discusión en particular. Es una aplicación práctica de la ontología filosófica, con una taxonomía. Las ontologías en informática son utilizadas en diversas áreas, como la web semántica, la inteligencia artificial y la ingeniería del conocimiento, entre otras. Permiten la interoperabilidad entre sistemas y la integración de datos de diferentes fuentes, lo que facilita el desarrollo de aplicaciones y servicios que requieren una comprensión más profunda de los datos y la información. (Cagle, s. f.)

2.3. Ontologías de Dominio

Desde el punto de vista informático, las ontologías son teorías que especifican un vocabulario relativo a un cierto dominio. Este vocabulario define entidades, clases, propiedades, predicados, funciones y las relaciones entre estos componentes. Las ontologías juegan un papel clave en la resolución de la interoperabilidad semántica entre sistemas de información y su uso. Según Sandoval Cantor (2007), "una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema, así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario." Las ontologías definen relaciones entre un vocabulario específico con el que podemos hacer referencia al conocimiento organizacional. Las relaciones entre estos conceptos deben ser de jerarquía y sinonimia, donde aquellos temas más genéricos deberían contener a los más específicos, creando una relación de hiperonimia e hiponimia. Debido a la naturaleza flexible del vocabulario natural en el ser humano, existe la posibilidad de referirse a un concepto de diferentes maneras, por lo cual la relación de sinonimia también es necesaria.

2.4. Ontologías en Lexicografía

Las ontologías tienen la capacidad de cumplir un papel importante en la creación de diccionarios al proporcionar una estructura semántica subyacente y un marco conceptual para organizar y relacionar el conocimiento léxico. Las ontologías son modelos formales que representan conceptos, sus propiedades y las relaciones entre ellos. (*¿Qué es la ontología en informática? – Aplicaciones Informáticas, s. f.*)

En el contexto lexicográfico, las ontologías pueden contribuir de las siguientes maneras:

- Organización y estructuración del léxico: Las ontologías permiten organizar y clasificar el vocabulario de manera jerárquica, estableciendo relaciones de subordinación y categorización entre los términos. Esto facilita la navegación y la búsqueda eficiente de información en un diccionario.
- Definiciones más precisas y claras: Al utilizar una ontología, se pueden establecer definiciones más precisas y claras para los términos, ya que se basan en un marco conceptual formalizado. Esto ayuda a evitar ambigüedades y a proporcionar una comprensión más profunda de los significados de las palabras.
- Relaciones semánticas y sinónimos: Las ontologías permiten establecer relaciones semánticas entre los términos, como sinónimos, antónimos, hiperónimos e hipónimos. Esto enriquece la información léxica al mostrar las conexiones entre las palabras y permitir una búsqueda más avanzada y precisa.
- Integración de diferentes fuentes de información: Las ontologías pueden integrar diferentes fuentes de información léxica, como diccionarios especializados, tesauros y recursos lingüísticos, en un único marco estructurado. Esto facilita la creación de diccionarios más completos y coherentes.
- Interoperabilidad y reutilización: Al utilizar ontologías, se promueve la interoperabilidad y la reutilización de recursos léxicos. Esto significa que los datos lexicográficos pueden ser compartidos, enlazados y utilizados en diferentes contextos y aplicaciones, lo que facilita la colaboración y la expansión del conocimiento léxico.

2.5. Grafos de Conocimiento

Un grafo es una estructura matemática que representa un conjunto de objetos, denominados vértices o nodos, junto con las relaciones que existen entre ellos, llamadas aristas o conexiones. Los grafos se emplean ampliamente en diversos campos, como la informática, las matemáticas y el análisis de datos, debido a su capacidad para modelar relaciones y estructuras complejas.

Dentro del contexto de las ontologías, los grafos se utilizan para representar los conceptos y las relaciones en un dominio específico. Una ontología es una representación formal del conocimiento que define los

conceptos, las propiedades y las relaciones en un dominio particular. Los grafos son herramientas útiles para visualizar y analizar ontologías, lo que facilita su comprensión y mantenimiento.(Thorn et al., 2005)

Algunas de las formas en que los grafos están relacionados con las ontologías incluyen:

- Utilización de algoritmos de grafos para descubrir patrones en las ontologías: Los investigadores han explorado la aplicación de algoritmos de la teoría de grafos para descubrir partes y patrones reutilizables en las ontologías. Sin embargo, encontrar patrones en las ontologías puede ser un desafío debido a la infinidad de formas en que se pueden asignar nombres a los conceptos y estructurar la ontología.
- Herramientas de visualización basadas en grafos para ontologías: Las herramientas de visualización basadas en grafos pueden ayudar a los desarrolladores de ontologías a crear y mantener definiciones de términos en formato de texto. Estas herramientas ofrecen una representación visual de la estructura de la ontología, lo que permite a los desarrolladores comprender mejor las relaciones entre conceptos y propiedades.
- Algoritmos de agrupación basados en grafos para la inducción de significados de palabras: En el campo del procesamiento del lenguaje natural, se han empleado algoritmos de agrupación basados en grafos para la inducción de significados de palabras, lo que implica la detección automatizada de los posibles significados de una palabra en diferentes contextos. Estos algoritmos utilizan grafos para modelar las relaciones entre palabras y sus contextos, lo que ayuda a identificar los diferentes significados de una palabra

2.6. Taxonomía

Una taxonomía es un sistema de clasificación jerárquica utilizado para organizar y categorizar objetos, conceptos, organismos u otros elementos en grupos basados en sus similitudes y características compartidas. El propósito principal de una taxonomía es proporcionar una estructura organizativa que permita la identificación y la disposición sistemática de elementos dentro de un dominio específico. El papel de un taxonomista es determinar en qué clase se ajusta mejor una entidad, como un animal o un libro. Esto se relaciona con la labor de los bibliotecarios, quienes clasifican nuevos libros en una taxonomía.

Las taxonomías se utilizan a menudo como un componente de las ontologías para proporcionar una estructura jerárquica a los conceptos y términos que se están representando. Por ejemplo, una ontología de dominio puede incluir una taxonomía de conceptos que agrupa conceptos relacionados en categorías más amplias. Además, las taxonomías pueden utilizarse para ayudar a organizar y estructurar los datos que se utilizan para construir una ontología. Por ejemplo, LiReWi, un sistema para la

obtención de relaciones para ontologías educativas, combina métodos basados en gramática, coocurrencia y taxonomía junto con varias bases de conocimiento para obtener relaciones entre temas del ámbito de aprendizaje. Por lo tanto, el uso conjunto de taxonomía y ontología puede proporcionar una representación más completa y estructurada de la información que se está organizando. (Conde et al., 2019)

2.7. Folksonomía

La folksonomía es un sistema colaborativo de clasificación y etiquetado que se utiliza en contextos digitales. En este sistema, los usuarios asignan etiquetas a diversos recursos, como imágenes, videos, páginas web y otros contenidos. Estas etiquetas se emplean posteriormente para organizar y buscar ese contenido.

La principal distinción entre la folksonomía y las taxonomías y ontologías tradicionales radica en su flexibilidad y en su falta de una estructura jerárquica predefinida. Mientras que las taxonomías y ontologías siguen una organización más estructurada y jerárquica, la folksonomía permite a los usuarios etiquetar los recursos de manera libre, reflejando una diversidad de perspectivas y enfoques en la clasificación de la información

2.8. Tripleta Semántica

Es un concepto fundamental en la Web Semántica y la representación del conocimiento. Es una declaración que consta de tres partes: sujeto, predicado y objeto. El sujeto representa la entidad que se está describiendo, el predicado representa la relación entre el sujeto y el objeto, y el objeto representa el valor o entidad que está relacionado con el sujeto a través del predicado. Los triples semánticos se utilizan con frecuencia para representar conocimiento en forma de ontologías y grafos de conocimiento. Pueden almacenarse en un almacén de triples, que es una base de datos optimizada para el almacenamiento y recuperación de triples semánticos.

2.9. WordClouds

Un wordcloud, también conocido como nube de palabras o tagclouds, es una representación visual de las palabras más frecuentes en un texto. En un wordcloud, las palabras se muestran con diferentes tamaños, donde las palabras más grandes indican una mayor frecuencia en el texto. Esta herramienta visual permite identificar rápidamente los términos más relevantes o importantes en un conjunto de datos, como un artículo, un discurso o un libro. Los wordclouds se utilizan comúnmente en análisis de texto, visualización de datos y presentaciones para resaltar los conceptos clave de manera intuitiva y atractiva.

2.10. Algoritmo TF IDF

El algoritmo TF-IDF (Frecuencia de Término - Frecuencia Inversa de Documento) es una herramienta estadística clave para analizar la importancia

de las palabras en un corpus. En esta propuesta, TF-IDF permite identificar términos significativos en el corpus, diferenciándolos de las stopwords o palabras de bajo peso semántico. Al aplicar TF-IDF, se realza la relevancia de palabras únicas o raras en el corpus, mientras que se minimiza la importancia de términos comunes.



CAPITULO 3

3. Estado del Arte

3.1. Herramientas para la creación de Diccionarios

3.1.1. TLEX

TLex es un software avanzado y especializado diseñado para la creación y gestión de diccionarios. Aunque no utiliza ontologías para la organización de datos lexicográficos, ofrece una amplia gama de funcionalidades que lo hacen una herramienta poderosa, especialmente para lexicógrafos con experiencia técnica. Una de las características distintivas de TLex es su capacidad para conectarse a bases de datos externas, lo que permite a los usuarios acceder a un vasto conjunto de recursos lingüísticos. Sin embargo, a diferencia de los sistemas basados en ontologías que se enfocan en el modelado de relaciones y conceptos dentro de un dominio específico, TLex se centra más en la edición y presentación de entradas o lemas léxicos.

La curva de aprendizaje para TLex puede ser alta, y su interfaz no siempre es intuitiva, lo que sugiere que está diseñado pensando en usuarios que tienen un conocimiento técnico previo en lexicografía y edición de diccionarios. Su amplio conjunto de funciones está claramente orientado a la edición detallada y la personalización de las entradas del diccionario, ofreciendo una gran flexibilidad para el manejo de datos lexicográficos complejos. TLex, por tanto, se destaca por su enfoque en la edición y formateo avanzado de entradas lexicográficas, más que en el análisis semántico profundo o el modelado de conocimiento que caracteriza a los sistemas basados en ontologías. Esta orientación lo hace particularmente valioso para proyectos que requieren una alta precisión en la presentación de datos lexicográficos y una gestión detallada de las entradas del diccionario.

3.1.2. WordNet

WordNet es una base de datos léxica ampliamente reconocida para el idioma inglés, conocida por su estructura única basada en "synsets", conjuntos de sinónimos que representan conceptos léxicos. Cada synset contiene palabras o frases con significados similares, acompañadas de definiciones y ejemplos. Estos synsets están interconectados mediante relaciones semánticas como antonimia, hiponimia y meronimia, lo que facilita la comprensión de la estructura y el significado del lenguaje. WordNet se utiliza extensamente en el procesamiento del lenguaje natural para tareas como la desambiguación de palabras y análisis semánticos, así como en la educación y la investigación lingüística.

Aunque WordNet a menudo se confunde con una ontología debido a su estructura y funcionalidad, en realidad es una base de datos léxica. No está diseñada para modelar conocimiento de un dominio específico, sino para representar relaciones léxicas y semánticas entre palabras. A pesar de su utilidad, WordNet tiene limitaciones, incluyendo inconsistencias semánticas, según se señala en fuentes como Wikipedia. Estas inconsistencias pueden surgir

de la complejidad inherente al lenguaje y su representación en una estructura de base de datos. WordNet, mantenido por el Cognitive Science Laboratory de la Universidad de Princeton, ha sido influyente en el desarrollo de otras bases de datos léxicas y proyectos semánticos en varios idiomas.

3.2. Revisión Sistemática

3.2.1. Planeamiento

Definición de Objetivos: En esta etapa inicial, se establecen los objetivos y el alcance de la revisión sistemática. Se determinan las preguntas de investigación a abordar y se definen los criterios de inclusión y exclusión de los estudios que se revisarán.

Diseño del Protocolo: Se desarrolla un protocolo de revisión que detalla los pasos a seguir, las estrategias de búsqueda, las herramientas de evaluación de la calidad y los métodos de extracción y análisis de datos. Este protocolo actúa como una guía para todo el proceso de revisión.

Búsqueda de Literatura: Se planifica la búsqueda exhaustiva de literatura científica y otros recursos relevantes. Esto incluye la identificación de palabras clave, bases de datos a consultar y estrategias de búsqueda que aseguren la recopilación de estudios pertinentes.

3.2.2. Ejecución

Búsqueda y Selección: En esta fase, se ejecuta la búsqueda de literatura de acuerdo con el protocolo previamente establecido. Se seleccionan estudios basándose en los criterios definidos, revisando títulos, resúmenes y textos completos de los artículos.

Evaluación de Calidad: Los estudios seleccionados se someten a una evaluación de calidad. Esto puede implicar el uso de escalas de calidad específicas o listas de verificación para determinar la validez de los estudios incluidos en la revisión.

Extracción de Datos: Se recopilan datos de los estudios incluidos. Esto implica extraer información relevante, como resultados, metodologías y conclusiones. Los datos se registran de manera sistemática.

3.2.3. Reporte de Resultados

Análisis y Síntesis de Resultados: En esta etapa, se analizan los datos extraídos de los estudios y se realiza una síntesis de los resultados. Se buscan patrones, tendencias o discrepancias en la literatura revisada para responder a las preguntas de investigación.

Presentación de Resultados: Se presentan los hallazgos de manera clara y coherente, utilizando tablas, gráficos u otros formatos visuales si es necesario. Los resultados se presentan de manera que sean comprensibles y útiles.

3.2.4. Protocolo de Investigación

Preguntas de Investigación

Pregunta Principal

- ¿Cómo se está aplicando el uso de ontologías de dominio para la generación de diccionarios y lexicografía?

Preguntas específicas:

- ¿Cuáles son los principales desafíos que se afrontan en la lexicografía y cómo se ha abordado el uso de ontologías?
- ¿Cuáles son las características de las ontologías utilizadas en las propuestas?
- ¿Cómo se planteó o modeló la ontología para mejorar la semántica de las palabras y agilizar el proceso de generación?
- ¿Cómo se implementa una ontología en estructuras de almacenamiento semántico como grafos?
- ¿Cómo se resuelven las colisiones o se maneja el contenido, propiedades y relaciones de una ontología?
- ¿Qué otras áreas utilizan ontologías, como el análisis de texto?

Definición de términos de búsqueda

Para llevar a cabo una revisión sistemática exhaustiva y efectiva, se inició el proceso con la utilización de términos de búsqueda preliminares. Estos términos se seleccionaron cuidadosamente para capturar generalidades y tendencias relevantes en la literatura existente en el área de estudio.

Los términos de búsqueda preliminares se eligieron de manera estratégica, teniendo en cuenta palabras clave amplias y conceptos fundamentales relacionados con el tema de investigación. Estos términos permitieron identificar una amplia gama de estudios y documentos que podrían proporcionar información clave sobre el estado actual del campo.

La búsqueda inicial se centró en recuperar estudios y literatura que ofrecieran una visión panorámica de las tendencias, teorías predominantes y enfoques comunes. Esta fase de búsqueda exploratoria permitió identificar áreas de interés y definir áreas temáticas específicas que requerirían un análisis más detallado.

	TERMINO DE BUSQUEDA PRINCIPAL	TERMINOS DE BUSQUEDA ALTERNATIVOS	
T1	LEXICOGRAPHY	LEXICOLOGY	LEXICAL ANALYSIS
T2	ONTOLOGIES	KNOWLEDGE ONTOLOGIES	ONTOLOGICAL MODELS
T3	KNOWLEDGE GRAPHS	SEMANTIC GRAPHS	KNOWLEDGE GRAPH DATABASES
T4	CHALLENGES IN LEXICOGRAPHY	OBSTACLES IN LEXICOGRAPHY	DIFFICULTIES IN LEXICOGRAPHY
T5	SEMANTIC WEB	WEB SEMANTICS	SEMANTIC TECHNOLOGIES
T6	NLP (NATURAL LANGUAGE PROCESSING)	COMPUTATIONAL LINGUISTICS	TEXT ANALYSIS
T7	VOCABULARY	LEXICON	WORD BANK
T8	TAXONOMIES	CATEGORIZATION	CLASSIFICATION SYSTEMS
T9	SEMANTIC SIMILARITY	WORD SIMILARITY	CONCEPTUAL OVERLAP
T10	INFORMATION RETRIEVAL	SEARCH ALGORITHMS	RETRIEVAL TECHNIQUES
T11	CORPUS ANALYSIS	TEXTUAL DATA ANALYSIS	CORPUS LINGUISTICS
T12	MULTILINGUAL LEXICOGRAPHY	CROSS-LANGUAGE LEXICOGRAPHY	MULTILINGUAL DICTIONARIES

Tabla 1 – Términos de búsqueda

Bases de Datos

Se seleccionaron las bases de datos más relevantes con el mayor número de artículos en la revisión panorámica de artículos, siendo estas SCOPUS y EBSCO.

TERMINO DE BUSQUEDA	SCOPUS	EBSCO
LEXICOGRAPHY		
ONTOLOGY	3,282	5134
ONTOLOGY TAXONOMIES	36,692	10.041
TEXT ANALYSIS	90000	484.392
DICTIONARY CONSTRUCTION	22032	90000
CORPUS ANALYSIS	2,899	697.97

Tabla 2 Uso de Términos revisión general

La búsqueda general de los términos resultó en cantidades que exceden el umbral de cantidad, lo cual muestra una alta tendencia en el uso de la terminología de la lexicografía y ontologías. Se procede a buscar términos en los títulos y refinar los términos.

TERMINO DE BUSQUEDA	SCOPUS	EBSCO
KNOWLEDGE GRAPH ONTOLOGY	149	2794
taxonomies AND ontology	57	10041
dictionary AND ontology	118	484.392
corpus and ontology	125	3374
graph AND ontology	468	20055

Tabla 3 Términos de búsqueda revisión general

Finalmente se itera con esta cadena de búsqueda:

TERMINO DE BUSQUEDA	RESULTADOS
FOLKSONOMY ONTOLOGY	31
GRAPH LEXICOGRAPHY	6
DICTIONARY ONTOLOGY	47

Tabla 4 Términos de búsqueda con mayor relevancia

Después de una revisión de los resúmenes, se eligen 36 artículos. Se extrae la siguiente información:

En "La Construcción del Diccionario" (Lara, s. f. 2012), Luis Fernando Lara explora la evolución de la lexicografía, desde sus raíces en listas primitivas de palabras para comerciantes y traductores en la antigua Mesopotamia y Egipto, hasta su desarrollo en Europa, donde las necesidades académicas y religiosas dieron origen a diccionarios más complejos. Lara destaca el impacto de la imprenta en la transformación de la lexicografía en un negocio editorial y señala la evolución hacia diccionarios monolingües en los Estados nacionales europeos, con ejemplos como el "Tesoro de la lengua castellana o española" de Covarrubias, que marcó un cambio significativo en la lexicografía de la época.

El autor Lara también aborda la situación actual y futura de la lexicografía, destacando la influencia de la tecnología informática en la mejora de la representatividad y recopilación de usos reales de palabras. Lara sugiere la necesidad de un cambio radical en los enfoques lexicográficos, abogando por la creación de corpus de datos detallados y un análisis semántico original por equipos de lingüistas nativos, especialmente en el ámbito hispánico. Este enfoque cultural es fundamental para entender los matices sutiles en el significado de las palabras que reflejan las experiencias sociales de las comunidades lingüísticas.

Lara ofrece una perspectiva integral sobre la historia y la evolución de la lexicografía, subrayando su continua adaptación y la necesidad de enfoques innovadores en respuesta a los desafíos contemporáneos, tanto en el contexto hispánico como global.

El artículo "Knowledge Graph Comparative Reasoning for Fact Checking" de Liu et al. (2022) presenta métodos innovadores para el razonamiento comparativo en grafos de conocimiento, enfocándose en detectar inconsistencias y comunalidades en información comparativa. El trabajo introduce algoritmos basados en redes neuronales y kernel gráfico para abordar este desafío. La investigación propone una serie de técnicas, incluyendo el uso de caminatas aleatorias con reinicio y modelos de lenguaje como BERT (Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019)), para aprender la representación semántica de los nodos en un grafo de conocimiento. Estos métodos permiten identificar elementos clave (nodos o aristas) en segmentos de conocimiento y tomar decisiones basadas en estos elementos y la información relacionada en el grafo de conocimiento.

Es importante destacar que el análisis de grafos de conocimiento en la lexicografía, con su aproximación tecnológica y metodológica, podría ofrecer insights valiosos para el análisis semántico y la gestión de datos en proyectos relacionados con ontologías y lexicografía. Las técnicas presentadas en el documento, como la captura de la semántica y la estructura de los nodos en un espacio de incrustación y la detección de inconsistencias, podrían ser aplicables en este contexto. Esto es especialmente relevante en la creación y gestión de diccionarios basados en ontologías, donde el análisis semántico y la detección de relaciones son cruciales.

Como mencionan los autores en "Structure-Augmented Text Representation Learning for Efficient Knowledge Graph Completion" (Wang et al., 2021), se presenta un enfoque híbrido para mejorar la eficiencia y precisión en la completación de grafos de conocimiento. Este enfoque combina técnicas de codificación textual con aprendizaje estructurado, utilizando algoritmos avanzados como encoders de estilo Siamese y modelos pre-entrenados tipo BERT. Especialmente relevante para esta investigación es la aplicación de estas técnicas en el contexto de grafos de conocimiento, lo que puede aportar ideas sobre cómo mejorar la representación y manipulación de ontologías en diccionarios, especialmente en términos de estructuración de datos y análisis semántico. Este enfoque refleja la tendencia actual hacia métodos más integrados y eficientes en el manejo de grandes volúmenes de datos en estructuras complejas como los grafos de conocimiento.

El artículo "The Folksonomy Tag Cloud: When is it Useful?" de James Sinclair y Michael Cardew-Hall (Sinclair & Cardew-Hall, 2008) investiga la utilidad de los wordclouds en folksonomías. Se enfoca en cómo estas nubes de etiquetas pueden facilitar la búsqueda de información, dependiendo de la naturaleza de la tarea de búsqueda. El estudio encuentra que los wordclouds son preferidos para tareas de búsqueda general, mientras que las interfaces de búsqueda tradicionales son más efectivas para tareas específicas. Estos hallazgos son relevantes para este estudio, ya que exploran cómo las folksonomías y los métodos de organización de datos pueden influir en la eficiencia de la búsqueda y recuperación de información, aspectos cruciales en el desarrollo de recursos lexicográficos basados en ontologías.

CAPITULO 4

4. Metodología en Uso de Ontologías en la Creación de Diccionarios

Las ontologías de dominio son estructuras de conocimiento que desempeñan un papel fundamental al definir y organizar los conceptos específicos de un dominio particular, como la lingüística en el caso de un diccionario. Estas ontologías permiten una representación detallada y sistemática de términos, sus relaciones y atributos, lo que resulta esencial para una comprensión precisa de los conceptos lingüísticos.

En el ámbito de los diccionarios, las ontologías de dominio son esenciales para estructurar la información de manera coherente, permitiendo una navegación más eficiente y una búsqueda precisa. Al definir y conectar los términos en el dominio lingüístico, estas ontologías facilitan la creación de diccionarios más completos y fáciles de usar. Además, al utilizar ontologías, se posibilita la interoperabilidad de datos lingüísticos, lo que resulta especialmente beneficioso en el contexto de la creación de diccionarios digitales y su integración en aplicaciones web.

En el contexto de la creación de diccionarios, los grafos de conocimiento se presentan como estructuras de datos interconectadas que permiten representar de manera efectiva los términos lingüísticos y sus relaciones semánticas. Estos grafos facilitan la organización y la representación de información lingüística de una manera que resalta las interconexiones entre los conceptos.

La incorporación de grafos de conocimiento en la construcción de diccionarios puede mejorar la calidad de la información presentada y ofrecer a los usuarios una experiencia de búsqueda más eficiente y significativa al permitir la exploración de relaciones entre términos lingüísticos. Como desarrollador con experiencia en tecnologías web, considerar la implementación de tecnologías como RDF (Resource Description Framework) y consultas SPARQL puede enriquecer la funcionalidad de la aplicación de diccionarios.

4.1. Jerarquía de Datos de Taxonomía

La lexicografía, la disciplina que se encarga de la creación de diccionarios, se beneficia enormemente del uso de la taxonomía para diseñar una jerarquía de datos que organice de manera eficiente las palabras y sus significados. Esta estrategia no solo es útil, sino fundamental, para construir diccionarios precisos y de fácil acceso.

4.1.1. Clasificación y Organización de Palabras

La base de la taxonomía en lexicografía es la clasificación y organización sistemática de las palabras y sus definiciones. Podemos dividir el lenguaje en categorías o clases semánticas que reflejen las relaciones entre las palabras.

4.1.2. Jerarquía de Datos en un Diccionario

La taxonomía crea una jerarquía de datos en la que cada palabra o concepto tiene su lugar específico. Esta jerarquía permite organizar las

palabras de manera que reflejen su relación con otras palabras en términos de significado y uso. Por ejemplo, si consideramos la palabra "perro", esta podría ubicarse en la categoría "sustantivos", subcategoría "animales domésticos". Dentro de esta subcategoría, podría haber otras palabras relacionadas como "gato", "conejo" y "loro".

4.1.3. Facilitar la Búsqueda y Recuperación de Información Lexicográfica

El uso de la taxonomía en lexicografía también facilita significativamente la búsqueda y recuperación de información en un diccionario. Los usuarios pueden navegar de manera más rápida y precisa a través del diccionario cuando las palabras están organizadas en una jerarquía de datos. Además, la taxonomía permite realizar búsquedas temáticas, lo que significa que los usuarios pueden buscar palabras relacionadas dentro de una categoría o subcategoría específica, como buscar sinónimos de un sustantivo en particular.

4.1.4. Mejorar la Comprensión del Contexto Lexicográfico

Otra ventaja importante de la taxonomía en lexicografía es que mejora la comprensión del contexto. Al organizar las palabras en categorías y subcategorías, se resaltan las relaciones semánticas y se proporciona un contexto más rico para cada término. Esto es esencial para los usuarios que desean entender cómo una palabra se utiliza en diferentes contextos y cómo se relaciona con otras palabras en un diccionario.

4.2. Folksonomía: Un Enfoque de Crecimiento Orgánico

La folksonomía, en su esencia, representa una práctica que permite a los usuarios etiquetar y categorizar palabras y términos de manera abierta y flexible. En el contexto de la lexicografía, esta técnica ofrece un crecimiento orgánico de recursos. Las palabras ingresadas por los usuarios se convierten en elementos esenciales para la expansión y mejora continua del diccionario, aprovechando cada entrada del sistema como un posible recurso futuro.

Una de las ventajas clave de la folksonomía es su capacidad para reflejar el uso real del lenguaje en tiempo real. Al permitir que los usuarios etiqueten términos según su propio entendimiento y uso, se crea una representación dinámica y actualizada del vocabulario. Esto es especialmente útil en el manejo de la polisemia, donde una palabra puede tener múltiples significados dependiendo del contexto.

El enfoque propuesto en esta tesis utiliza la frecuencia de uso de las palabras como un respaldo en la toma de decisiones al seleccionar definiciones en escenarios de polisemia. Herramientas como los wordclouds pueden visualizar estas frecuencias, proporcionando a los lexicógrafos una representación clara de qué significados son más prevalentes en el uso cotidiano. Esto no solo facilita la selección de definiciones más relevantes y comprensibles, sino que también asegura que el diccionario refleje de manera precisa el uso contemporáneo del lenguaje.

Al integrar la folksonomía con ontologías y grafos de conocimiento, se puede crear una estructura lexicográfica que no solo es precisa y relevante, sino

también adaptativa y reflejante de la evolución del lenguaje. Este enfoque permite una actualización continua y una mejora constante del diccionario, asegurando que siempre esté alineado con las necesidades y el uso del público objetivo.

4.3. Modelado de la Ontología

Nombre: La "Ontología de Dominio para la Creación de Diccionarios" es una herramienta fundamental en el ámbito de la lexicografía y la lingüística. Su enfoque se centra en proporcionar una estructura semántica precisa y organizada para representar y relacionar los términos lingüísticos, permitiendo la creación de diccionarios más completos y eficientes.

Propósito: El propósito primordial de esta ontología es agilizar el proceso de creación de diccionarios al brindar una base sólida para la organización de los términos y la representación de sus interconexiones. Al facilitar la navegación, búsqueda y comprensión de las relaciones semánticas entre palabras y definiciones, esta ontología se convierte en una herramienta esencial para la construcción de diccionarios monolingües de alta calidad, mejorando la precisión y la eficiencia en el proceso de desarrollo.

Dominio: Lingüística y lexicografía. La "Ontología de Dominio para la Creación de Diccionarios" se encuentra intrínsecamente ligada al ámbito de la lingüística y la lexicografía. En este contexto, el enfoque se dirige hacia la comprensión y representación de las complejidades lingüísticas, incluyendo la estructura, el significado y las relaciones entre palabras. La lingüística y la lexicografía son los pilares que sustentan la organización y definición de términos, y esta ontología tiene como objetivo principal mejorar la precisión y eficiencia en la construcción de diccionarios monolingües.

4.3.1. Jerarquía de Clases

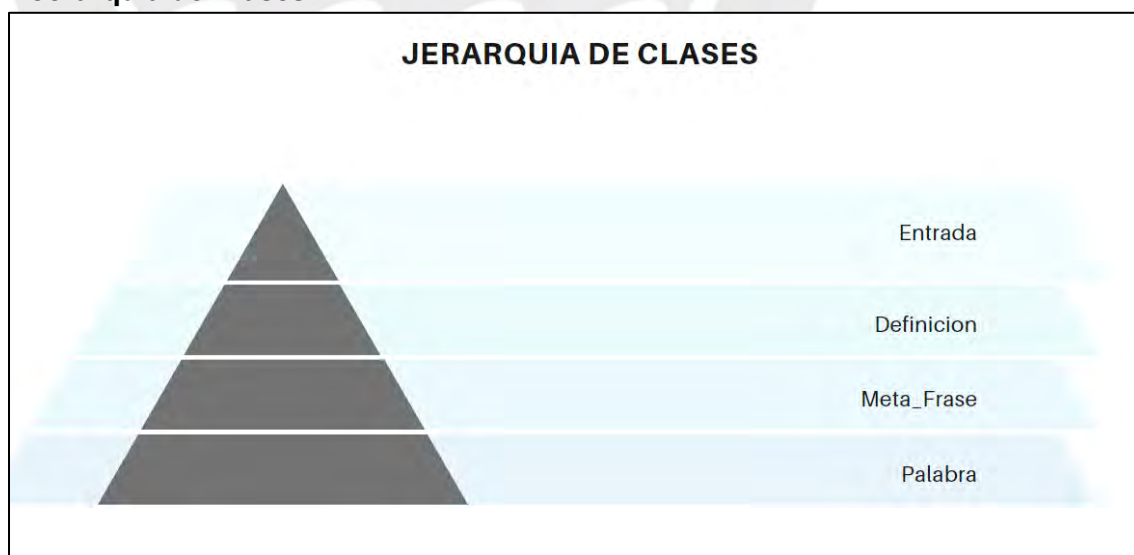


Figura 9 Jerarquía de clases de ontología - Elaboración Propia

Como se muestra en la figura 9, la jerarquía implica estas 4 clases. En este gráfico, la clase de mayor jerarquía y amplitud, el hiperónimo, se encuentra en la

cima de la pirámide, mientras que las clases más específicas o hipónimos ocupan niveles descendentes en la pirámide.

4.3.1.1. Clase Palabra

La clase "Palabra" sirve como piedra angular en la ontología de dominio para la creación de diccionarios. Representa la unidad fundamental del lenguaje y abarca una amplia gama de términos, desde sustantivos y verbos hasta adjetivos y adverbios. En el contexto de la ontología, la clase "Palabra" cumple un principio esencial de no duplicidad, asegurando que cada término sea único y distintivo. Esto se traduce en la incorporación de términos en diversas secciones, como meta-frases y definiciones. La representación precisa de "Palabra" y su interconexión con otras clases dentro de la ontología permiten una organización eficiente y una navegación significativa a través de los diccionarios.

Propiedades:

- **Frecuencia:** Indica la frecuencia con la que cada palabra es utilizada en el ámbito específico abordado en la ontología. Cada vez que una palabra se relaciona o establece una nueva arista con otro nodo en la ontología, esto refleja su mayor uso en ese contexto particular. Esta propiedad permite no solo cuantificar la importancia relativa de las palabras, sino también analizar cómo se relacionan y se entrelazan las palabras en el dominio de conocimiento definido por la ontología.
- **Censura:** Valor numérico, rango 1 – 10. La propiedad "censura" tiene como objetivo cuantificar el grado de censura asociado a cada palabra en el contexto específico que aborda la ontología. Un valor más alto en esta propiedad indica un mayor grado de censura, mientras que un valor más bajo sugiere una menor censura. Esta propiedad ofrece una valiosa perspectiva sobre la sensibilidad de las palabras dentro del ámbito de estudio, permitiendo el análisis de qué términos pueden estar sujetos a restricciones o limitaciones en su uso.
- **Peso Semántico:** Valor numérico, rango 1 – 10. La propiedad "peso semántico" se utiliza para determinar la importancia relativa de las palabras dentro de la ontología. Aquellas palabras que poseen un alto peso semántico se consideran más relevantes en el contexto y pueden desempeñar un papel destacado en el análisis y la inferencia de conocimiento. Por otro lado, las palabras con un peso semántico más bajo pueden clasificarse como stopwords o términos menos significativos en el dominio abordado.

Relaciones:

- **ES_ENTRADA:** Esta relación se establece con la clase "ENTRADA" y obedece a que cada palabra en un contexto de

corpus puede ser una entrada, es decir, una palabra que a su vez contiene definiciones relacionadas.

- **CONTIENE:** Esta relación se establece con la clase "META_FRASE" y enlaza a cada palabra que puede tener la meta_frase.
- **DEFINE:** Esta relación se establece con la clase "DEFINICION" y existe cuando dicha palabra ha sido utilizada para asignarle una definición.

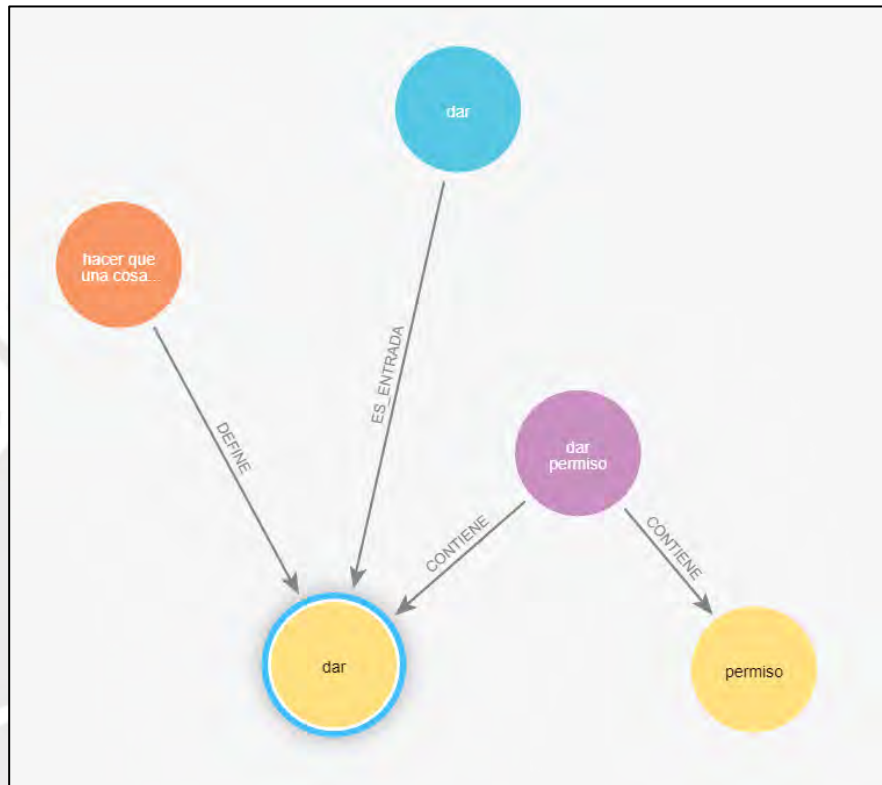


Figura 10 Representación grafica de relaciones de clase PALABRA - Elaboración propia

4.3.1.2. Clase Meta_Frase

La clase "Meta_Frase" desempeña un papel distintivo en la ontología de dominio para la creación de diccionarios al representar un nivel de organización superior. A diferencia de las palabras individuales, las "Meta_Frases" contienen palabras interconectadas que, aunque no necesariamente poseen un contenido semántico fuerte por sí mismas, desempeñan un papel fundamental en la generación de definiciones y el enriquecimiento de la información lingüística. Estas "Meta_Frases" actúan como vínculos contextuales que conectan y proporcionan un marco de referencia para las palabras dentro de un diccionario. Su función es esencial para la generación de definiciones y para facilitar una comprensión más completa de los términos, alimentando el corpus global.

Propiedades:

- **Frecuencia:** Valor numérico. Esta propiedad, similar a la propiedad de frecuencia de la clase "PALABRA", indica la frecuencia con la que cada "META_FRASE" es utilizada en el ámbito específico abordado en la ontología. Cada vez que una "META_FRASE" se relaciona o establece una nueva arista con otro nodo en la ontología, esto refleja su mayor uso en ese contexto particular. Esta propiedad permite no solo cuantificar la importancia relativa de las "META_FRASES", sino también analizar cómo se relacionan y se entrelazan los léxicos en el dominio de conocimiento.
- **Censura:** Valor numérico, rango 1 – 10. La propiedad "censura" tiene como objetivo cuantificar el grado de censura asociado a cada "META_FRASE" en el contexto específico que aborda la ontología. A diferencia de la propiedad "censura" de la clase "PALABRA", esta censura será calculada en base al contenido de las palabras que componen dicha "META_FRASE", siendo resultado de la propagación de censura de las "PALABRAS" que contiene.
- **Peso Semántico:** Valor numérico, rango 1 – 10. La propiedad "peso semántico" se utiliza para determinar la importancia relativa de las "META_FRASES" dentro de la ontología. De manera similar a su propiedad de censura, las "PALABRAS" que poseen un alto peso semántico y que están contenidas en la "META_FRASE" por efecto de propagación le otorgarán mayor peso semántico a dicha "META_FRASE".

Propiedades:

- **DESCRIBE:** Esta relación se establece con la clase "DEFINICION" y se enlaza cuando la "META_FRASE" está contenida en la definición, ya sea de manera parcial o total. Es decir, dicha "META_FRASE" puede ser parte de la definición o constituir toda la definición.
- **CONTIENE:** Como se menciona en la descripción de las relaciones de la clase "PALABRA", esta es la relación inversa y se establece para definir qué instancias de la clase "PALABRA" están contenidas en dicha "META_FRASE".

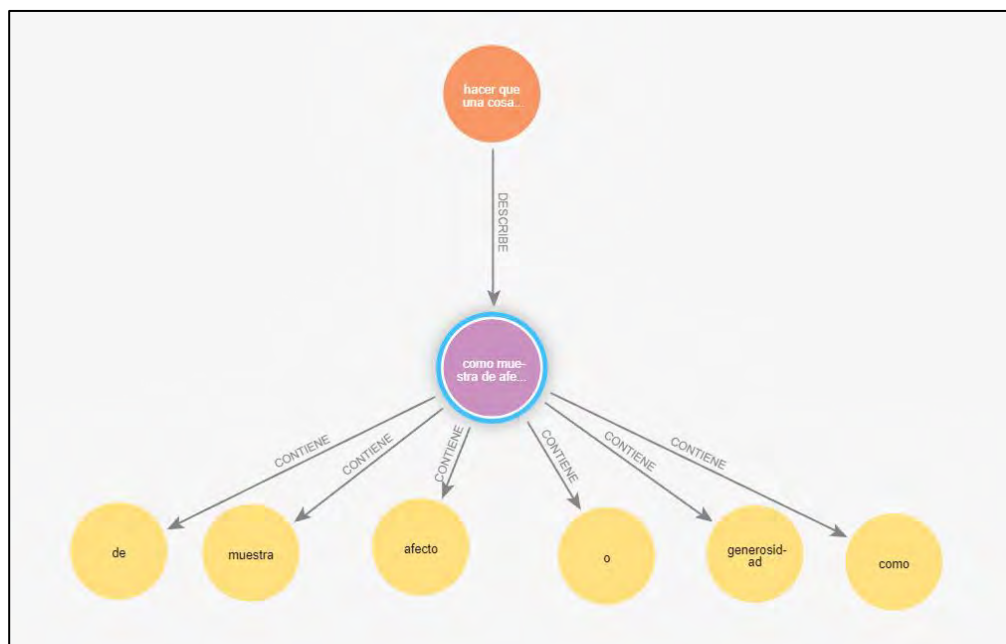


Figura 11 Representación gráfica de Relaciones de clase Meta_Frase

4.3.1.3. Clase Definición

La clase "Definición" es un componente central de la ontología de dominio para la creación de diccionarios. Esta clase se compone de "Meta_Frases" y "Palabras" y mantiene una relación directa y necesaria con una "Palabra". Cada "Definición" se asocia de manera unívoca con una palabra específica, lo que garantiza que las explicaciones estén vinculadas con precisión a su término correspondiente. Las "Meta_Frases" y las "Palabras" trabajan en conjunto para proporcionar una interpretación completa y detallada de cada término, enriqueciendo así el contenido del diccionario. Esta estructura jerárquica asegura que las definiciones sean contextualmente relevantes y precisas.

Propiedades:

- **Censura:** Valor numérico, rango 1 – 10. De igual manera, esta propiedad es la propagación tanto de la clase "PALABRA" como de la clase "META_FRASE". Con esta propagación se puede tener una noción del grado de censura que contiene dicha definición.

Relaciones:

- **CONTIENE_DEF:** Esta relación se establece con la clase "ENTRADA" cuando la definición es relacionada como entrada de una palabra.
- **DESCRIBE:** Esta relación funciona de manera inversa con la clase "META_FRASE" y existe cuando las meta_frases son parte de dicha instancia.
- **DEFINE:** Esta relación se establece con la clase "PALABRA" cuando se asocia una definición a una palabra.

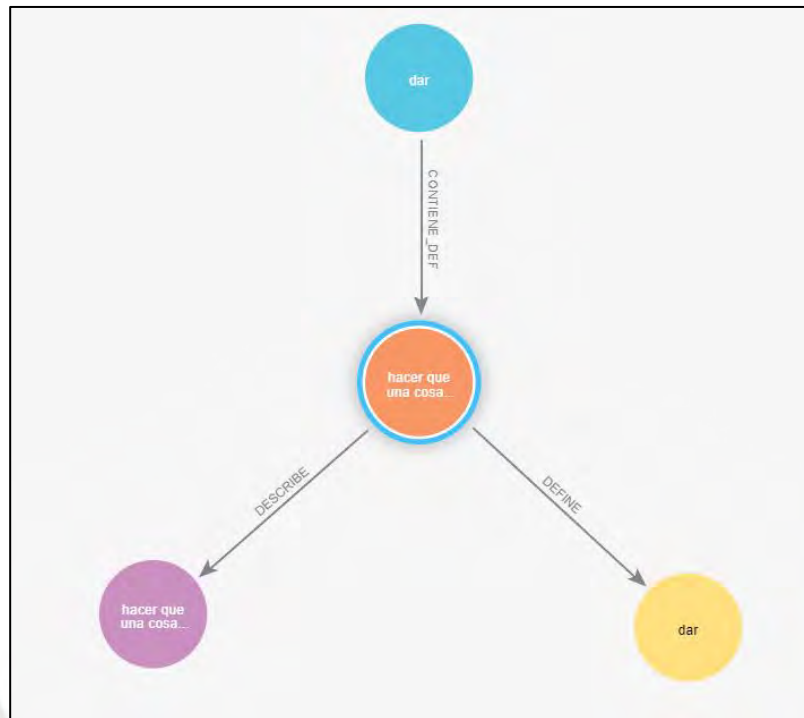


Figura 12 Representación grafica de relaciones de clase Definición

4.3.1.4. Clase Entrada

La clase "Entrada" es la de mayor jerarquía en la ontología. Las instancias de esta clase indican que una palabra en el diccionario ha sido asociada a una o varias definiciones.

Propiedades:

- **Censura:** Valor numérico, rango 1 – 10. De igual manera, esta propiedad es la propagación de sus clases hipónimas.

Relaciones:

- **CONTIENE_DEF:** Esta relación es la inversa de la clase "Definición" y se establece cuando una instancia de definición es asociada a una palabra.
- **ES_ENTRADA:** Esta es la relación inversa con la clase "PALABRA". Con esta relación se forma un enlace circular entre la clase más inferior y la clase superior, aportando un significativo valor semántico a la instancia de la "PALABRA" que en este caso ya tendría una definición asociada.



Figura 13 Representación grafica de relaciones de clase Entrada

4.4. Actividades Realizadas y Uso de la Ontología

En esta sección se detalla el uso, las ventajas y la discusión de la ontología propuesta. Se realizaron diversas actividades clave, incluyendo el diseño y la implementación de la ontología de dominio, que implicó la identificación y clasificación de términos relevantes, la creación de relaciones jerárquicas y la incorporación de propiedades específicas como la frecuencia de uso, el grado de censura y el peso semántico.

4.4.1. Etapas:

4.4.1.1. Alimentación del Corpus

Como se observa en la figura 14, existe un flujograma para el ingreso de datos al corpus:

- **Ingreso de Datos:**

Objetivo: Introducir palabras y metafrases en el sistema.

Proceso: Captura y registro de palabras individuales y combinaciones de palabras (metafrases).

Aspectos Clave: Asignación de peso semántico y censura a cada palabra, manejo de repeticiones para actualizar la frecuencia de las palabras ya existentes.

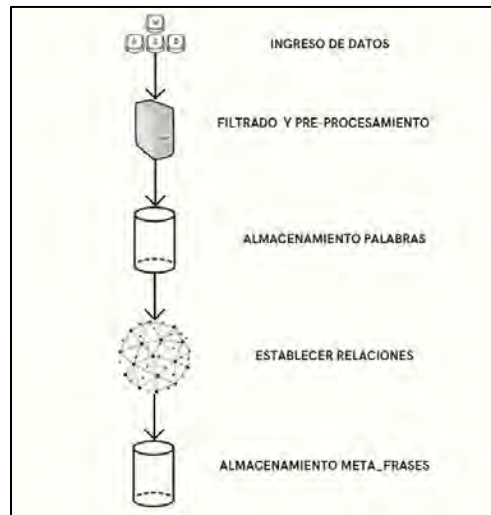


Figura 14 Flujo de alimentación del Corpus

- **Filtrado:**
 Objetivo: Limpiar las palabras ingresadas, eliminando símbolos y caracteres no deseados.
 Proceso: Revisión y eliminación de elementos no relevantes o inapropiados en el corpus.
 Aspectos Clave:
 Mantenimiento de la calidad y coherencia del corpus.
- **Almacenamiento de Palabras:**
 Objetivo: Guardar las palabras limpias y procesadas en la base de datos.
 Proceso: Almacenamiento eficiente de palabras individuales en la estructura de datos.
 Aspectos Clave:
 Organización y optimización para facilitar el acceso y la recuperación de información.
- **Establecimiento de Relaciones con Metafrase:**
 Objetivo: Vincular palabras individuales con metafrases.
 Proceso: Identificación y creación de relaciones entre palabras y las metafrases a las que pertenecen.
 Aspectos Clave:
 Análisis de cómo se combinan las palabras en frases y su contexto semántico.
- **Almacenamiento de Metafrase:**
 Objetivo: Almacenar las metafrases que han sido relacionadas con palabras individuales.
 Proceso: Registro de metafrases en la base de datos, tras establecer sus relaciones con las palabras individuales.
 Aspectos Clave:

Asegurar que la estructura de datos soporte eficientemente el almacenamiento de combinaciones de palabras.
Mantener una organización que permita análisis complejos y recuperación eficiente.

Con esta estructura, el proceso de alimentación del corpus adquiere una profundidad adicional, permitiendo un análisis más detallado tanto de palabras individuales como de su uso en contextos más amplios (metafrases). Esta dualidad de enfoque es esencial para un análisis lexicográfico completo y para el desarrollo de un corpus rico y matizado.

4.4.1.2. Creación de la Definición

En la figura 15 muestra el flujograma de la creación de una definición

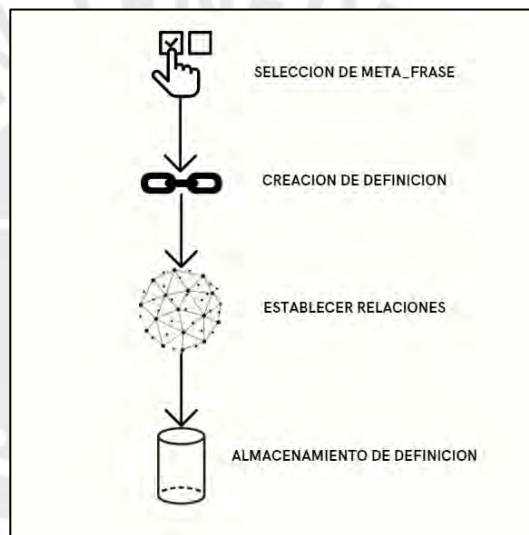


Figura 15- Flujo de Creación de Definición – Elaboración propia

- **Selección de Metafrases:**
Objetivo: Elegir las metafrases adecuadas que formarán parte de una definición.
Proceso: Se revisan las metafrases almacenadas y se seleccionan aquellas que son relevantes para la nueva definición.
Aspectos Clave:
Criterios de selección basados en la relevancia, el contexto y la adecuación de las metafrases para la definición propuesta.
Consideración de la frecuencia y la censura de las palabras en las metafrases para asegurar la coherencia y adecuación de la definición.
- **Creación de la Definición:**
Objetivo: Formular una definición clara y precisa, utilizando las metafrases seleccionadas.

Proceso: Concatenación y ajuste de las metafrases seleccionadas para formar una definición coherente y completa.

Aspectos Clave:

Mantenimiento de la integridad semántica y la fluidez en la transición entre metafrases.

Asegurar que la definición final refleje con precisión el significado y el uso de la palabra o concepto definido.

- **Creación de Relaciones entre Metafrases y Definición:**

Objetivo: Vincular las metafrases seleccionadas con la definición creada.

Proceso: Establecimiento de relaciones semánticas y estructurales entre las metafrases y la nueva definición.

Aspectos Clave:

Identificación de cómo cada metafrase contribuye al significado global de la definición.

Resaltar la interdependencia entre las metafrases y la definición para un análisis más profundo.

- **Almacenamiento de la Definición:**

Objetivo: Registrar la definición en la base de datos.

Proceso: Almacenamiento eficiente de la definición en la estructura de datos, incluyendo sus relaciones con las metafrases.

Aspectos Clave:

Conservar la información sobre la frecuencia y la censura de las metafrases y cómo estas características se propagan a la definición.

Facilitar el acceso y la recuperación de las definiciones para su uso posterior en el corpus.

4.4.1.3. Creación de Entrada Léxica

En la figura 16 se muestra el flujograma de la creación de una entrada léxica

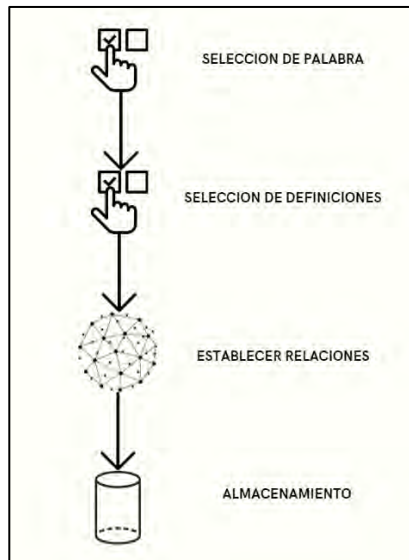


Figura 16- Flujo de Creación de Entrada- Elaboración propia

- Selección de Palabra:**
 Objetivo: Escoger la palabra que se definirá en la entrada lexicográfica.
 Proceso: Revisión de las palabras almacenadas en la base de datos para seleccionar la más adecuada para la nueva entrada.
 Aspectos Clave:
 Considerar la relevancia y la frecuencia de uso de la palabra en el corpus.
 Asegurar que la palabra seleccionada se alinee con los objetivos y el alcance del diccionario.
- Selección de Definición:**
 Objetivo: Elegir una o varias definiciones que explicarán la palabra seleccionada.
 Proceso: Examinar las definiciones disponibles para identificar aquellas que mejor describan la palabra.
 Aspectos Clave:
 Evaluar la precisión, claridad y relevancia de las definiciones.
 Considerar la censura y el peso semántico de las definiciones en relación con la palabra.
- Establecimiento de Relaciones entre Definiciones y Entrada:**
 Objetivo: Vincular las definiciones seleccionadas con la entrada lexicográfica.
 Proceso: Crear conexiones semánticas y estructurales entre las definiciones y la entrada.
 Aspectos Clave:
 Garantizar que las definiciones proporcionen un entendimiento completo y matizado de la palabra.
 Resaltar cómo cada definición contribuye al significado general de la entrada.

- Establecimiento de Relación entre Palabra y Entrada:**
 Objetivo: Asociar la palabra seleccionada con la entrada lexicográfica creada.
 Proceso: Formar una relación directa entre la palabra y su entrada en el diccionario.
 Aspectos Clave:
 Asegurar que la relación refleje adecuadamente la naturaleza y el uso de la palabra en diferentes contextos.
 Considerar la historia y la evolución de la palabra en el corpus para una representación más rica.
- Almacenamiento de la Entrada:**
 Objetivo: Registrar la entrada lexicográfica completa en la base de datos.
 Proceso: Almacenar de manera eficiente la entrada, incluyendo la palabra, sus definiciones y las relaciones establecidas.
 Aspectos Clave:
 Mantener una estructura de datos que facilite el acceso y la consulta de las entradas.
 Conservar la trazabilidad de la información desde las palabras individuales y metafrases hasta las definiciones y la entrada final.

4.4.2. Propagación de Propiedades

4.4.2.1. Discusión

En la propuesta actual como se muestra en la figura etapa de creación de entrada. (Véase sección 4.4.1.3) se puede contar con una paleta de opción más matizada donde no solo el tamaño de la candidata a definición de una palabra para convertirse en entrada léxica mejora la orientación del lexicógrafo, también se puede aprovechar para mostrar un colorado para aquellas definiciones que arrastren alguna censura en su meta contenido que pueden ser las metafrases o palabras.

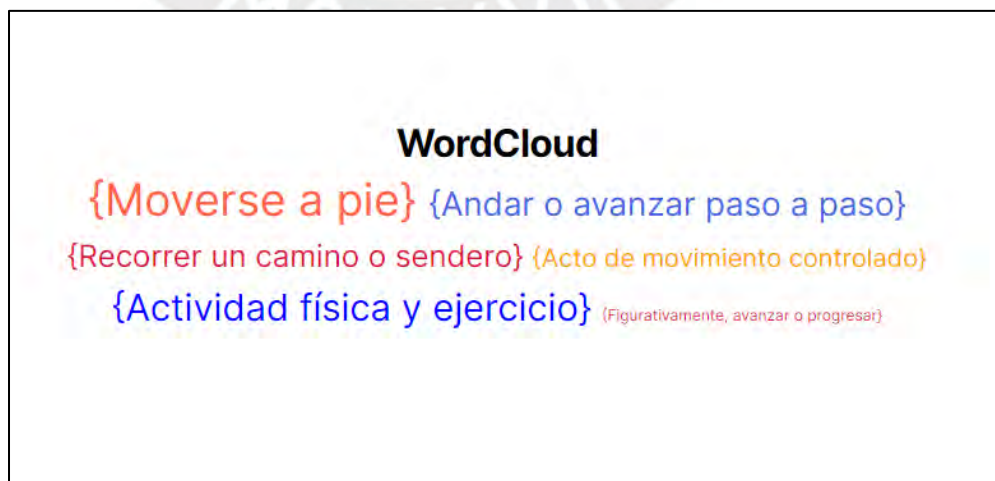


Figura 17 Prototipo de Wordcloud de definiciones de la palabra Caminar

Pero esto a su vez genera un desafío relacionado a la gestión de la censura y la frecuencia de las palabras en el corpus. Al asignar un nivel de censura a cada palabra ingresada en la base de datos, este valor se propaga a través de las estructuras lingüísticas más amplias, como metafrases, definiciones y finalmente entradas lexicográficas. La censura, en este contexto, se establece basándose en criterios subjetivos del lexicógrafo, reflejando así su percepción sobre la adecuación del término en diferentes contextos.

Sin embargo, la dinámica del lenguaje presenta un desafío inherente a este enfoque. Las palabras, aunque inicialmente puedan ser consideradas como inapropiadas o sensibles, pueden adquirir con el tiempo una aceptación más amplia y un uso más generalizado. Este fenómeno se ve reflejado en la frecuencia con la que una palabra aparece en el corpus. Una palabra que originalmente se asignó una alta censura, pero que se utiliza con frecuencia, podría necesitar una reevaluación de su nivel de censura. El enfoque de la folksonomía, sugiere que la censura de una palabra debe ser flexible y adaptarse a su uso actual y percepción en la sociedad.

Por lo tanto, surge la cuestión de cómo equilibrar la censura inicialmente asignada con la evolución lingüística y social de las palabras. Un sistema rígido de censura podría no reflejar adecuadamente el dinamismo y la evolución constante del lenguaje, mientras que un enfoque demasiado laxo podría fallar en proporcionar las guías necesarias para el manejo de contenido sensible o inapropiado. La solución a este dilema requiere un método que no solo considere la censura asignada a nivel de palabra individual, sino que también tenga en cuenta su frecuencia de uso y la evolución de su significado y aceptación dentro de la comunidad lingüística.

Este equilibrio entre censura y representatividad lingüística es crucial para desarrollar un corpus lexicográfico que sea tanto culturalmente sensible como lingüísticamente representativo. En última instancia, el objetivo es crear un recurso que refleje con precisión el uso y la evolución del lenguaje, manteniendo al mismo tiempo un entorno respetuoso y considerado en términos de contenido.

De manera similar el trato del peso semántico y la frecuencia emerge una cuestión importante relacionada con el tratamiento de palabras que, si bien son de uso frecuente, poseen un bajo peso semántico. Estas palabras, comúnmente conocidas como stopwords (palabras vacías), incluyen preposiciones, conjunciones, artículos y otros conectores que, aunque esenciales para la estructura gramatical y la coherencia de las oraciones, a menudo carecen de significado sustantivo por sí mismas.

La relevancia de estas palabras en el corpus lexicográfico presenta un dilema. Por un lado, su presencia es crucial para proporcionar descripciones precisas y contextualmente ricas de las definiciones. Permiten formar frases y oraciones cohesivas, asegurando que las definiciones sean comprensibles y reflejen el uso real del lenguaje. Por otro lado, en términos de análisis semántico y la reutilización de recursos lingüísticos, estas palabras pueden tener una importancia limitada. Por ejemplo, palabras como "a", "y" y "que" pueden aparecer con alta frecuencia en el corpus, pero su contribución al peso semántico general de una frase o definición puede ser mínima. (Ver figura 18)



Figura 18 Prototipo de stopwords con alta frecuencia de uso

En este contexto, se plantea el reto de equilibrar la necesidad de incluir estas palabras para la integridad gramatical y la fluidez del lenguaje, con la necesidad de centrarse en palabras y frases de mayor relevancia semántica para el análisis lingüístico. Una solución podría ser la implementación de un sistema dual en el tratamiento de estas palabras: por un lado, reconocer su importancia en la formación de estructuras gramaticales y, por otro, permitir su filtrado o ponderación reducida en análisis donde el enfoque es el peso semántico.

4.4.2.2. Balance de Propagación

4.4.2.2.1. Balance de Stopwords, Peso semántico y Propagación

En el tratamiento de stopwords una estrategia eficaz puede ser implementar una lista estática de palabras consideradas como stopwords. Estas palabras, generalmente artículos, preposiciones y conjunciones, se identifican por su baja contribución al peso semántico en el análisis del lenguaje. La implementación de tal lista permitiría a la ontología ignorar eficientemente estas palabras durante el análisis y procesamiento del texto, concentrándose en elementos con mayor relevancia semántica.

Además, esta lista de stopwords podría ser dinámica y orgánica, adaptándose a los cambios en el uso del lenguaje. Se podrían incorporar mecanismos para actualizar la lista, añadiendo palabras que, a través del análisis del corpus, se identifiquen como de bajo peso semántico. Este enfoque orgánico permite que la ontología refleje más precisamente el uso actual del lenguaje, adaptándose a su evolución y a los contextos específicos en los que se aplica.

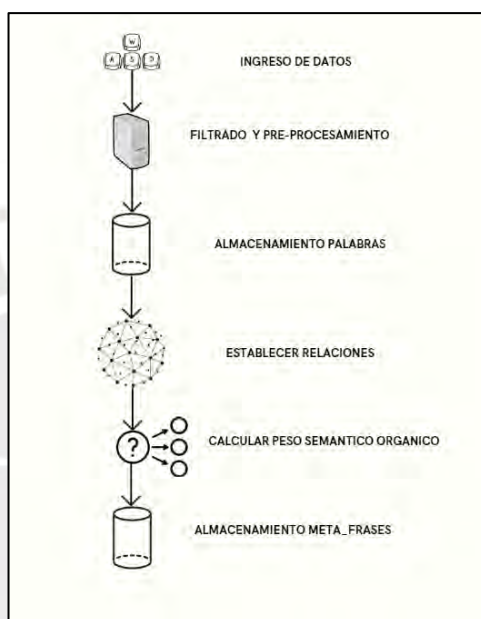


Figura 19 Inclusión de fase de Calculo orgánico

Con esta adición se propone un enfoque dinámico que incluye no solo propiedades descritas en el modelado de la ontología, sino también propiedades adicionales que evolucionan orgánicamente en respuesta al comportamiento del corpus, como “frecuencia orgánica”, “peso semántico orgánico”. Esta dualidad permite que las instancias de las clases en la ontología no solo mantengan una estructura definida por el lexicógrafo, sino que también reflejen cambios emergentes y tendencias dentro del corpus. Algunas propiedades podrían ser ajustadas manualmente por el lexicógrafo, mientras que otras se modificarían automáticamente basándose en patrones de uso, frecuencia y otros factores dinámicos del lenguaje. Este enfoque híbrido asegura que la ontología se mantenga actualizada y relevante, adaptándose continuamente al flujo natural y a la evolución del lenguaje generando un mejor balance.

Uso de Algoritmo TF IDF para relevancia de términos

En esta propuesta, TF-IDF permite identificar términos significativos en el corpus, diferenciándolos de las stopwords o palabras de bajo peso semántico. Al aplicar TF-IDF, se realiza la relevancia de palabras únicas o raras en el corpus, mientras que se minimiza la importancia de términos comunes. Para calcular la frecuencia de términos (TF) se utiliza la siguiente fórmula.

$$tf = \frac{n(\text{palabra aparece en documento})}{\text{numero palabras en documento}}$$

Fórmula 1 Cálculo de Frecuencia de términos

En la fórmula 1, n refiere a la cantidad de veces que una palabra aparece en un documento en este caso en el corpus.

Documentos	Texto	Términos
D1	Huevo huevo gato	Huevo gato
D2	Gato huevo niño gato niño	Gato huevo niño
D3	Pato huevo cosa	Pato webo cosa

Tabla 5 Ejemplo de frecuencia de términos

TF	huevo	gato	niño	pato	Cosa
D1	2/3	1/3			
D2	1/5	2/5	2/5		
D3	1/3			1/3	1/3

Tabla 6 Valores de Frecuencia

En la tabla 5 se muestra el contenido y los términos, que aparecen en cada documento, y en la tabla 6 se muestra los valores de la frecuencia que aparece cada palabra en cada documento.

Para el cálculo del IDF se utilizará la siguiente formula:

$$idf = \log \frac{\text{numero de documentos}}{\text{documento donde aparece palabra}}$$

Fórmula 2 Calculo de IDF

Aplicando las formulas 1 y 2 en el ejemplo del corpus.

TF IDF	huevo	gato	niño	pato	cosa
IDF	log(3/3)	log(3/2)	log(3/1)	log(3/1)	log(3/1)
D1	0	0.06	0	0	0
D2	0	0.07	0.19	0	0
D3	0	0	0	0.16	0.16

Tabla 7 Calculo del TF IDF de los términos

4.4.2.2. Uso de Algoritmo TF-IDF para balance de peso semántico

Para calcular el TF (Term Frecuency se usará como referencia la cantidad de veces que una palabra aparece en las meta_frases de la ontología, con lo que debe contarse las aristas que generen la relación "CONTIENE" entre las clases "Meta_Frase" y "Palabra", quedando para su cálculo, la siguiente formula.

$$tf = \frac{\sum_{i=1}^n NroPM}{palabras}$$

Fórmula 3 Frecuencia de termino en el Corpus

Describiendo la fórmula 3 se tiene:

NroPM: Representa la cantidad de aristas que una clase "Palabra" puede tener con la clase "MetaFrase", indicando el número de veces que una instancia de la clase "PALABRA" es utilizada en todo el corpus.

Para el cálculo del IDF del corpus se puede usar la siguiente formula:

$$idf = \log \frac{NroD}{NroDAT}$$

Fórmula 4 Calculo de inversa de frecuencia de terminos de corpus

Donde **NroD** es el numero total de definiciones de la ontología y **NroDAT** es el numero de definiciones donde el termino o "PALABRA" aparece.

Entonces, aplicando este enfoque, es posible balancear el peso semántico de los términos del corpus. Este cálculo debe incluirse en la parte de ingreso de datos (véase la sección 4.4.1.1). De esta manera, el flujograma sufriría esta adición:

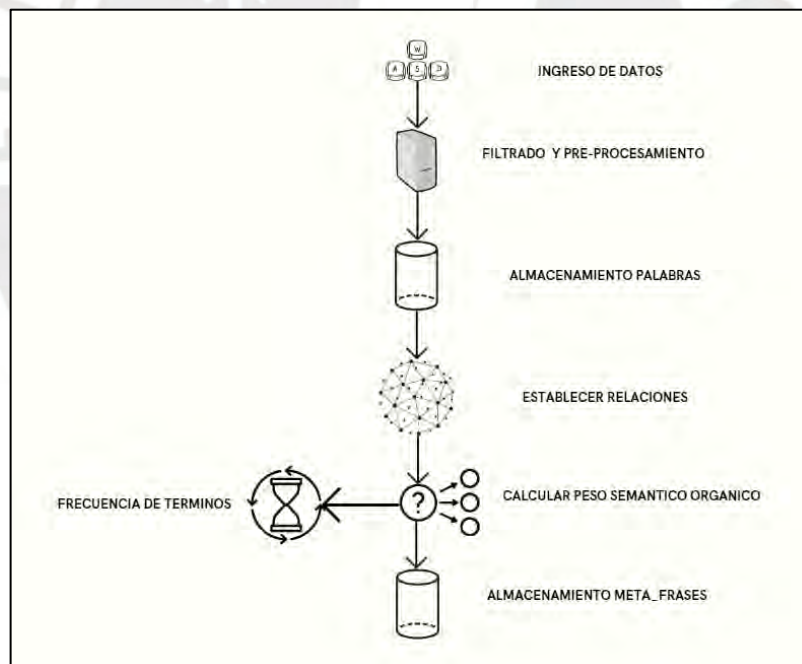


Figura 20 Flujo de ingreso de datos al corpus con inclusión de cálculo de frecuencia – Elaboración propia

4.4.2.2.3. Balance de Censura Propagada

Aplicando el enfoque de la sección anterior Utilizando el valor TF-IDF, que mide la relevancia de un término en el corpus, se recalcula la censura orgánica asignada a las palabras. La lógica detrás de este enfoque es que palabras con un uso frecuente y relevante, reflejado por un bajo valor de TF-IDF, deben tener una censura reducida. La **operación matemática** propuesta para

este ajuste es una función inversa, donde la censura original se divide por un factor derivado del valor TF-IDF. Este método garantiza que la censura de las palabras se alinee más estrechamente con su uso y aceptación actual en la sociedad, proporcionando así un análisis más preciso y contextual del lenguaje.

$$censura\ orgánica = \frac{censura}{1 + (TF - IDF)}$$

Fórmula 5 Cálculo de propagación de Censura de términos

En la fórmula 5, la censura es el valor que le asigna el lexicógrafo a una instancia. A la división se le suma 1 para evitar divisiones entre cero, asegurando que la censura orgánica puede, en el peor de los casos, ser igual pero nunca superior al valor original. Este cálculo debería hacerse cuando se calcula el TF-IDF.

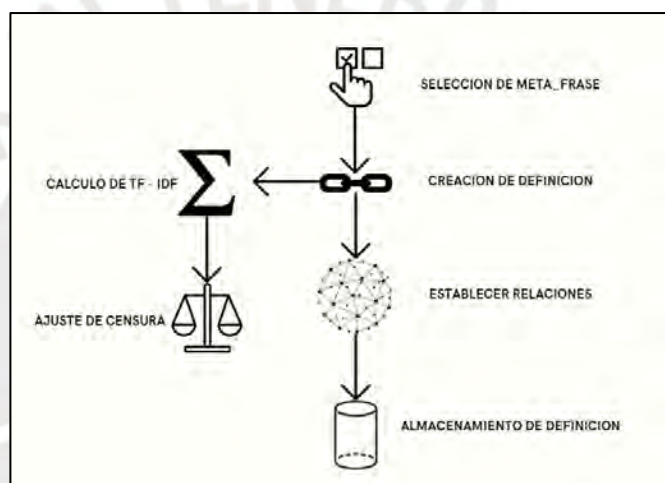


Figura 21 Flujo de Ingreso con cálculo de Censura –Elaboración propia

4.5. Adaptación Dinámica de Contenido

4.5.1. Propiedad de Censura en la Ontología

Cada instancia de las clases 'Palabra', 'Metafrase', 'Definición' y 'Entrada' en la ontología tiene asociada una propiedad de censura. Esta propiedad es inicialmente asignada por el lexicógrafo y luego ajustada orgánicamente en función del uso y la frecuencia de la palabra, reflejando así la censura propagada. La censura propagada en las definiciones es crucial, ya que determina la adecuación del contenido para el público objetivo. Estas propiedades se convierten en parámetros clave para filtrar y presentar contenido en aplicaciones que utilizan la ontología.

4.5.2. Ajuste Dinámico del Umbral de Censura

En la etapa de desarrollo y en la implementación de aplicaciones que utilizan la ontología, se puede establecer y modificar un umbral de censura. Este umbral define qué nivel de censura es aceptable para el público objetivo. Al ajustar este umbral, las aplicaciones pueden incluir o excluir automáticamente definiciones basadas en su nivel de censura, permitiendo una adaptación dinámica del contenido. Por ejemplo, un umbral más bajo

ocultaría definiciones con censura alta, mientras que un umbral más alto permitiría una mayor variedad de contenido.



Figura 22 Prototipo de Ajuste Dinámico de Censura

Esta capacidad de adaptación proporciona a los desarrolladores y gestores de contenido una herramienta poderosa para asegurar que el material presentado sea adecuado y relevante para el público objetivo. Permite una personalización eficiente sin la necesidad de revisar manualmente cada elemento del corpus, lo cual es especialmente valioso en corpus grandes y en constante evolución.

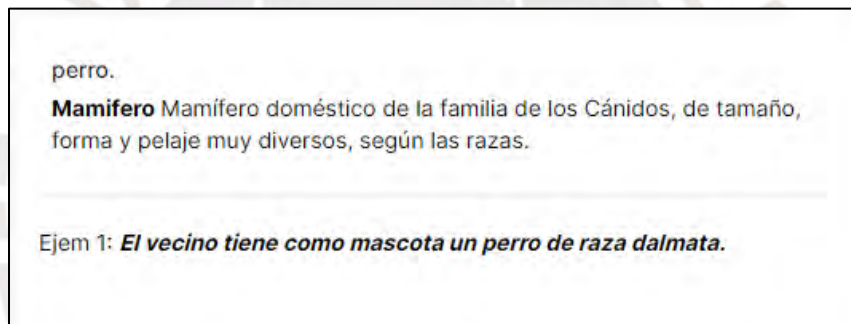


Figura 23 Prototipo de contenido adaptado dinámicamente

4.6. Desarrollo de Matriz Relacional

El valor de esta matriz radica en su capacidad para generar nuevas definiciones y relaciones semánticas entre palabras. Por ejemplo, ver figura. Si "**dar**" se define como "**hacer que una cosa pase a otra persona**" y "**conceder**" incluye esta metafrase además de "**por la autoridad que se tiene**", entonces se puede inferir que "**conceder**" es una forma de "**dar**" pero con la adición de la palabra "**autoridad**". Así, una nueva definición para "**conceder**" podría ser "**Dar por la autoridad que se tiene**", generándose una nueva definición completamente coherente.

Palabra / Lexicon	hacer que una cosa pase a otra persona	transmitiendo responsabilidad	por la autoridad que se tiene	como muestra de afecto o generosidad
dar	+	-	-	-
entregar	+	+	-	-
conceder	+	+	+	-
regalar	+	-	-	+

Figura 24 Prototipo de Matriz de relación

Este enfoque permite una generación dinámica y flexible de definiciones, y al mismo tiempo, revela relaciones semánticas más profundas entre las palabras. La matriz no solo sirve para construir definiciones, sino que también facilita la exploración del lenguaje, mostrando cómo diferentes palabras comparten elementos comunes en sus significados y se diferencian en otros aspectos.



Figura 25 Flujo de funcionamiento de matriz relacional - Elaboración propia

Selección de Palabras

Son instancias de clase "Palabra" a las cuales se les asociará una definición. Las palabras deben ser seleccionadas por su potencial para formar relaciones significativas con otras palabras. Esto implica que la selección pueda basarse en la experiencia y el conocimiento del lexicógrafo, quien puede identificar palabras que tienen una relación inherente o una sinergia semántica. Por ejemplo, palabras que son sinónimos, antónimos, o que comparten un campo semántico común, pueden ser candidatas ideales para esta etapa. Además, es posible

apoyarse en metadatos que estén organizados de alguna forma para facilitar la relación que estas puedan tener.

Selección de Metafrases

Esta es una etapa crucial en la matriz relacional, donde se eligen metafrases coherentes y compatibles para formar definiciones lógicas y significativas. Esta selección, crítica para la cohesión del corpus lexicográfico, depende tanto de la experticia del lexicógrafo como de herramientas avanzadas como el aprendizaje automático y el análisis de texto. El lexicógrafo debe elegir metafrases semánticamente ricas y versátiles, mientras que el machine learning ayuda a identificar patrones y estructuras frasales a partir de grandes conjuntos de datos textuales. Aunque este proceso puede ser complejo y no completamente transparente (una "caja negra"), enriquece la selección de metafrases, facilitando la creación de un corpus dinámico y contextualizado.

Concatenación de MetaFrases

Utilizando la matriz como una interfaz interactiva, el lexicógrafo selecciona y une cuidadosamente las metafrases elegidas, formando definiciones completas y significativas. En esta fase, la concatenación se realiza manualmente, con un enfoque en la coherencia semántica y la correcta construcción gramatical. El lexicógrafo combina las metafrases de manera que reflejen fielmente el significado deseado.

Establecer Relaciones

Implica la vinculación de cada palabra con su definición correspondiente, asegurando que la relación entre ellas sea clara y lógicamente coherente.

Almacenamiento

Registra la definición y las relaciones creadas en el sistema.

Inferir Nuevo Conocimiento

Se pueden utilizar consultas Cypher, que es un lenguaje de consulta de grafos, para explorar relaciones semánticas entre palabras, identificando definiciones que comparten metafrases comunes. Esta técnica permite descubrir conexiones semánticas profundas y sutilezas del lenguaje, enriqueciendo el corpus lexicográfico y facilitando la investigación lingüística avanzada. Es una etapa clave para ampliar la comprensión del lenguaje a través de la estructura y análisis de la ontología.

4.7. Implementación de la Ontología

4.7.1. Herramientas Usadas

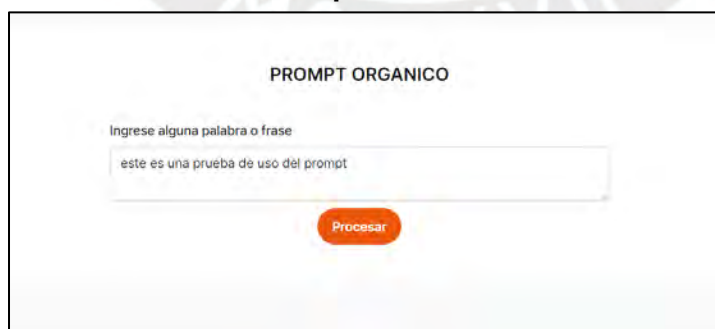
Para la implementación de un prototipo que permita probar la ontología, la coherencia y las relaciones establecidas al ingresar palabras y definiciones, se utilizaron las siguientes tecnologías:

- **JAVASCRIPT**
JavaScript es un lenguaje de programación esencial para el desarrollo web moderno y actualmente es el más utilizado en el

entorno web (JavaScript | MDN, 2023). Considerando la arquitectura web Cliente-Servidor, JavaScript tiene un papel crucial al ejecutarse del lado del cliente, permitiendo manipular no solo la lógica sino también el DOM (Document Object Model). Esto permite actualizar dinámicamente el contenido de una aplicación web sin necesidad de recargarla completamente. Este dinamismo ha sido fundamental para el desarrollo de tecnologías basadas en esta arquitectura, brindando una mejor experiencia de usuario al interactuar con una aplicación web.

- **NODEJS**
NodeJs es un entorno de ejecución para JavaScript que permite ejecutar JavaScript fuera de los navegadores web, donde originalmente era ejecutado. Esto amplía su uso para la generación de lógica del lado del servidor, como servidores y servicios backend, APIs y aplicaciones en tiempo real.
- **NEXTJS**
NextJs es un framework full stack basado en React para desarrollo web en JavaScript, que integra tanto el lado del servidor como el lado del cliente, permitiendo el desarrollo de ambos en un solo entorno.
- **NEO4J**
Neo4j es una base de datos orientada a grafos, ideal para el manejo de ontologías debido a las relaciones que se generan entre las diferentes clases. El uso de grafos para organizar la taxonomía de la ontología es más natural que el uso de otras bases de datos, como las relacionales. Además, Neo4j posee un motor de consultas muy versátil llamado Cypher, que permite la exploración fácil e intuitiva del grafo.

4.7.2. Desarrollo de Prototipo



The image shows a web form titled "PROMPT ORGANICO". It features a text input field with the placeholder text "Ingrese alguna palabra o frase" and a pre-filled example "este es una prueba de uso del prompt". Below the input field is an orange button labeled "Procesar".

Figura 26 Prototipo de Prompt de ingreso de datos

La figura 26 muestra un prompt que será usado para el ingreso de cualquier recurso léxico. Luego, este será grabado en la ontología y las relaciones con las palabras y las meta_frases serán registradas. Como se menciona en el capítulo 4.2, la folksonomía servirá para alimentar el corpus léxico que

contiene los recursos necesarios para elaborar el diccionario. La intención es consumir cualquier entrada e integrarla a la base de datos, estableciendo las relaciones que la entrada genera.

Después del ingreso de la metrafrase esta es tratada de la siguiente manera:

Descomposición:

La Metafrase es descompuesta en las palabras que contiene, ejemplo: "Este es una prueba de uso del Prompt"

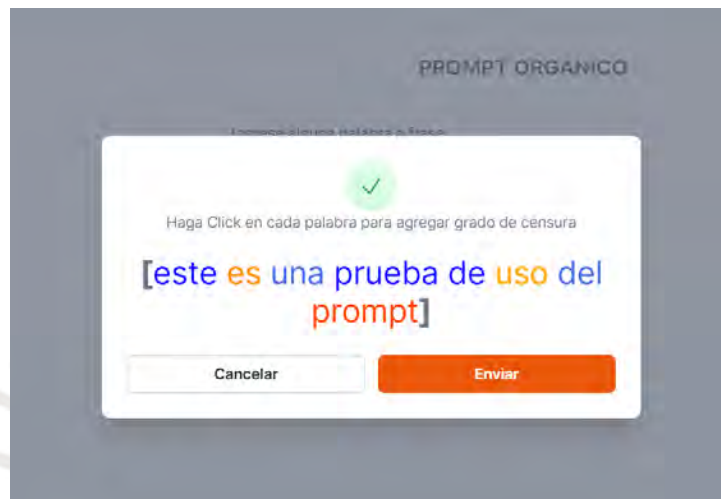


Figura 27 Confirmación de Entrada Prompt

En la figura 27 se muestra una interfaz de confirmación que filtra el contenido de caracteres. En esta parte, el usuario también puede, a través de clics en cada palabra, agregar los niveles de censura cambiando el color de cada palabra de gris a rojo en caso de tener mucha censura. En esta etapa se almacena cada palabra que potencialmente puede convertirse en recurso para alguna definición o entrada léxica de otra palabra. En esta fase se calcula la cantidad de veces que dicha

palabra ha sido ingresada anteriormente, lo que indica una arista hacia la metafrase correspondiente con la que fue ingresada.

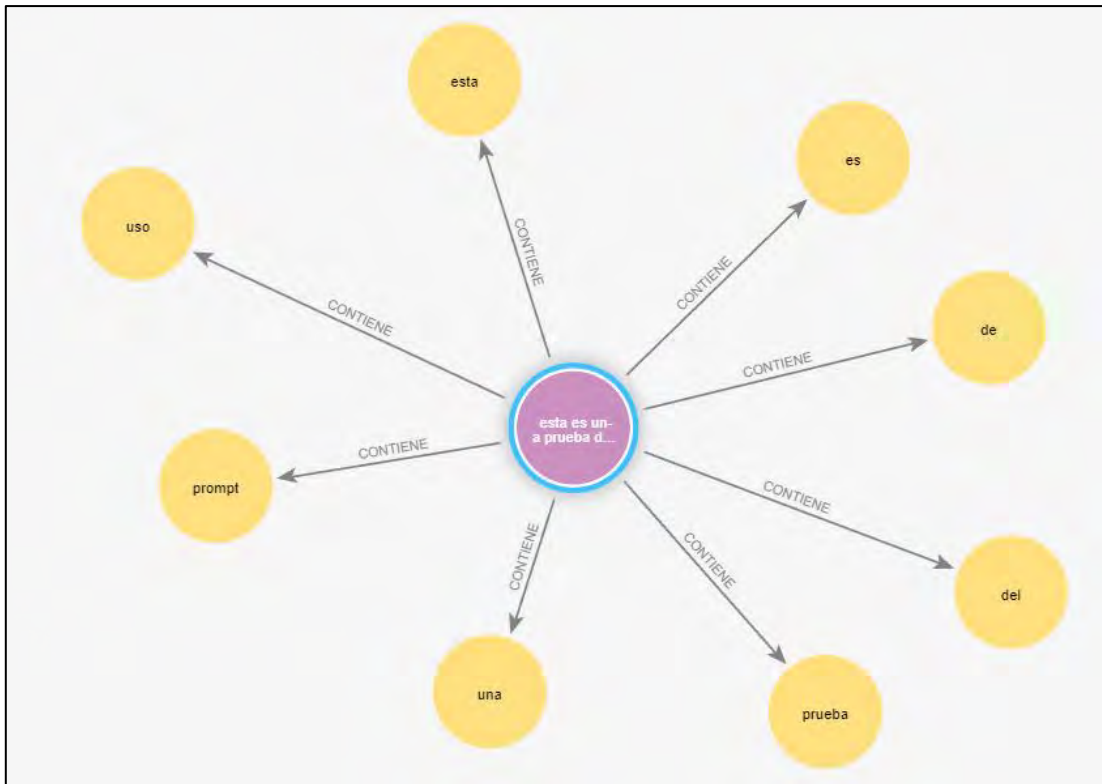


Figura 28 Almacenamiento de Metafrase "Esto es una prueba del uso de Prompt"

WordCloud

este es una prueba
de uso del prompt

Lexicon

"este es una prueba de uso del prompt" nueva

PALABRA	RELACIONES	ORDEN	CENSURA
este nueva	1	0	1
es nueva	1	1	4
una	2	2	2
prueba nueva	1	3	1
de	2	4	1
uso nueva	1	5	3
del nueva	1	6	2
prompt nueva	1	7	6

Figura 29- Reporte de confirmación de ingresos de recursos

En la figura 30 se muestra un reporte sobre el ingreso y la frecuencia de cada palabra que fue ingresada.

En la figura 29 puede verse como queda el almacenamiento de la metafrase “**Este es prueba de uso del Prompt**” de esta manera tanto la metafrase como cada palabra que la contiene quedan almacenadas en la base de datos.

En el caso de ingresar otra metafrase pero si esta contiene alguna palabra ingresada con anterioridad, entonces la ontología incrementa la frecuencia de uso de esa palabra de manera orgánica tan solo agregando la arista correspondiente a la nueva Metafrase, quedando así un manejo de frecuencia de uso tan solo por el conteo de aristas que una palabra pueda tener así alguna metafrase:

En el caso de ingresar la metafrase “Esta es una segunda prueba de uso del prompt”

Considerando:

metafrase 1 : “**Esta es una prueba de uso del Prompt**”

metafrase 2 : “**Esta es una segunda prueba de uso del Prompt**”

METAFRASE	esta	es	una	segunda	Prueba	de	uso	del	prompt
1	X	X	X		X	X	X	X	X
2	X	X	X	X	X	X	X	X	X

Tabla 8 comparación de palabras similares entre dos metafrases

Como se puede ver en la tabla 8, la metafrase 2 solo contiene, como palabra nueva en comparación con la metafrase 1, la palabra "segunda". Así mismo, la ontología conserva el principio de ingreso único, por lo que las palabras que ya fueron ingresadas anteriormente solo se relacionan con la nueva metafrase, y aquellas palabras nuevas son creadas.

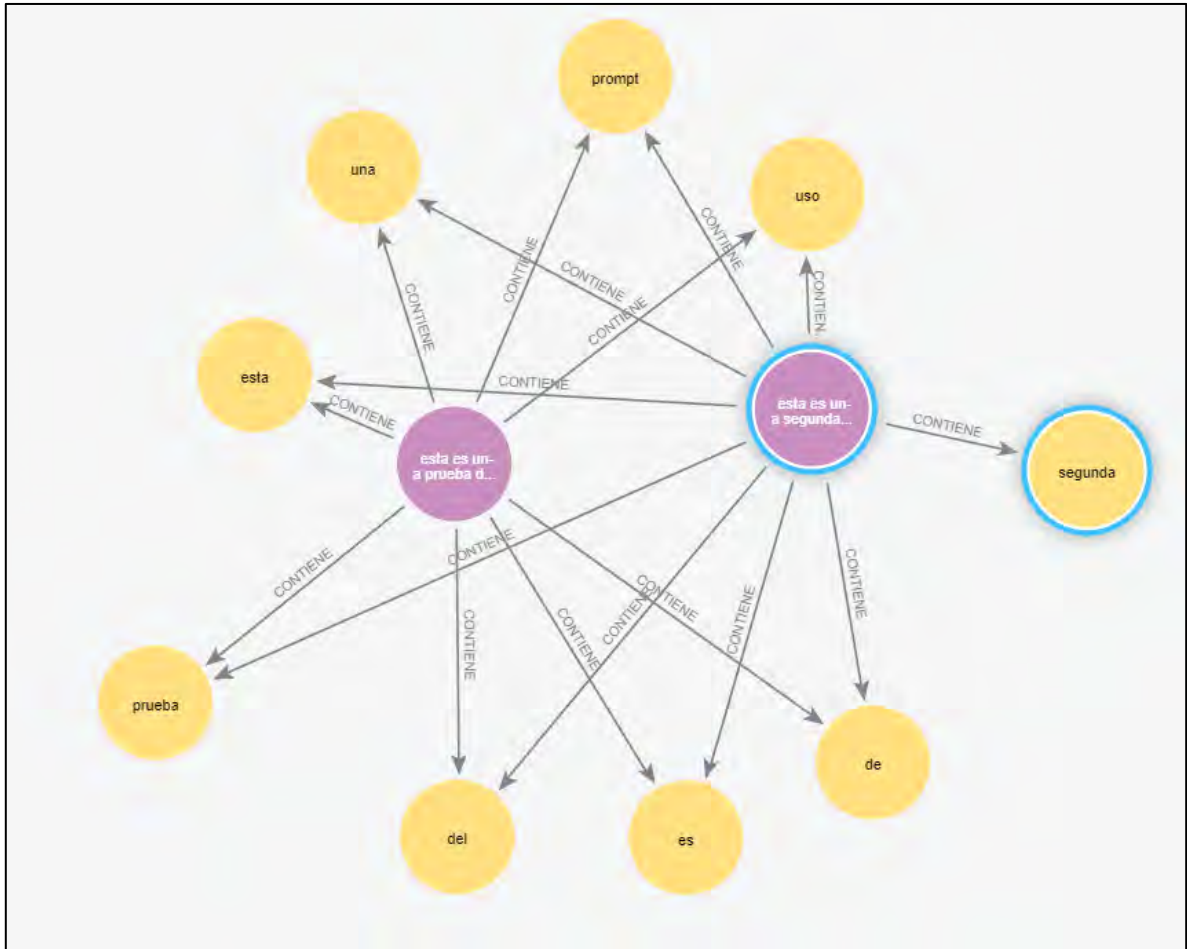


Figura 30 Resultado de almacenamiento de 2 metafrases similares en grafo

Como se muestra en la figura 31, se pueden observar las ventajas de tener un crecimiento orgánico del corpus gracias a la folksonomía, que alimentará la entrada de recursos para la elaboración del diccionario. De esta forma, aquellas definiciones cuyas palabras sean de mayor uso o aparezcan en relaciones con otras metafrases en el corpus léxico pueden ser sugeridas para su uso. Esto proporciona al lexicógrafo, encargado de elaborar el diccionario, una herramienta de soporte para la toma de decisiones al momento de incluir las entradas del diccionario. Esta frecuencia de uso se denotará como un WordCloud al mostrar las palabras de una definición.

PALABRAS	esta	es	una	segunda	Prueba	de	uso	del	prompt
ARISTAS	2	2	2	1	2	2	2	2	2

Tabla 9 Conteo de aristas después de ingreso de metafrase

4.7.3. Desarrollo de Matriz de Definiciones

Como es mencionado en la sección 4.6 la matriz permite relacionar definiciones que en forma de cascada pueden concatenar una nueva definición.

Considerando estas definiciones:

Dar: *“Hacer que una cosa pase a otra persona”*

Conceder: *“Hacer que una algo paso de una persona transmitiendo responsabilidad”*

En este caso reemplazando la definición en la segunda palabra “conceder” se puede inferir que:

Conceder: “Dar transmitiendo responsabilidad”

La ontología podrá inferir este conocimiento debido a que una definición es contenida dentro de otra. Si bien es posible acceder a esta cascada de definiciones mediante un ingreso manual, dependiendo de la experiencia del lexicógrafo, también se puede generar utilizando inteligencia artificial (IA). Técnicas de "few shots" pueden ser empleadas para guiar un prompt de ChatGPT, facilitando así la generación automática y coherente de definiciones.

4.7.3.1.1. GENERACION DE METADATOS CON CHATGPT

Elaboración asistida con Few Shots Learning

Ingreso de Datos:

Se almacena la palabra que debe ser, en este caso, la palabra principal a través de la cual se generarán otras definiciones. En el contexto actual, esta palabra sería la del ejemplo anterior: "dar".

Ingreso de Petición:

A través de un conector se conecta a un modelo generativo de ChatGPT:

Modelo: gpt-3.5 Turbo

```
return `genere una lista de palabras relacionadas con un término específico,
proporcionando una descripción o definición de cada palabra en la lista.
Las definiciones deben mantener una relación semántica coherente entre las palabras de la lista.
A continuación, se muestra un ejemplo utilizando el término "dar":

1. Dar: Hacer que una cosa pase a otra persona.
2. Entregar: Hacer que una cosa pase a otra persona transmitiendo responsabilidad.
3. Conceder: Hacer que una cosa pase a otra persona por la autoridad que se tiene.

Ahora, aplique el mismo enfoque utilizando el término "${palabra}".
A continuación se proporciona un segundo ejemplo con el término "objetivo" para mayor claridad:

1. Meta: Un objetivo específico y medible que se desea alcanzar.
Relación: Un objetivo es una meta, pero no todas las metas son objetivos.

2. Propósito: La razón o intención detrás de un objetivo.
Relación: El propósito guía la elección de un objetivo y le da significado.

3. Finalidad: El objetivo o razón por la cual algo se hace o se lleva a cabo.
Relación: La finalidad de un objetivo es lograr un resultado deseado.

4. Destino: El resultado o logro al que se aspira alcanzar con un objetivo.
Relación: El destino es el punto final hacia el cual se dirige un objetivo.

5. Ambición: Un objetivo o deseo fuerte de lograr algo importante.
Relación: La ambición impulsa a establecer objetivos desafiantes y a trabajar para alcanzarlos.`;
```

Figura 31 Ingreso de instrucción para generación de Definiciones en cascada de Matriz con ChatGPT

Resultado:

Ejecutada la instrucción, se puede observar que la IA genera una lista de definiciones coherentes que pueden ser utilizadas como metafrases de la ontología y que guardan relación con la palabra "dar". Aplicando el principio de ingreso único de la ontología, también es capaz de distinguir las palabras que van siendo utilizadas con más frecuencia. En este ejemplo, las palabras "persona" y "cosa" se repiten con mayor frecuencia. La ontología, a través de WordCloud, otorgaría un mayor tamaño a estas palabras, de manera que serían sugeridas como de uso frecuente para relacionarlas en definiciones según el lexicógrafo.

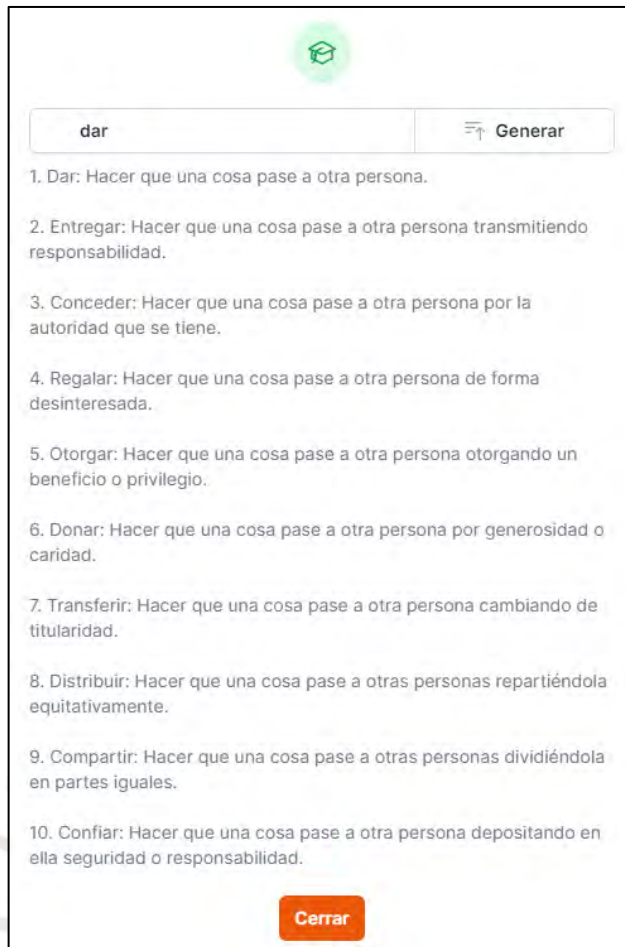


Figura 32 Generación de definiciones relacionadas a palabra "Dar" por ChatGPT

Descripción del Algoritmo:

Como se observa en la figura 32, aquellas definiciones que sean subconjuntos de otras pueden ser reemplazadas por la palabra en cuestión.

Dar: "Hacer que una cosa pase a otra persona"

Entregar: "Hacer que una cosa pase a otra persona" + "transfiriendo responsabilidad"

Entonces: se infiere que "entregar" puede ser:

Entregar: "Dar" + "transfiriendo responsabilidad"

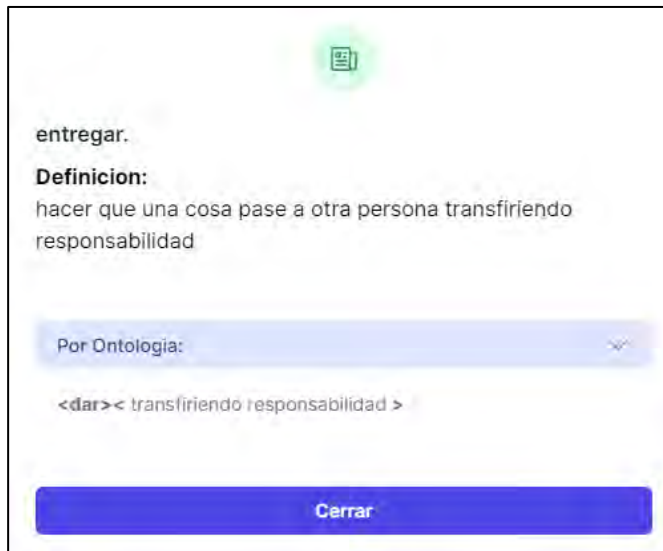


Figura 33 Inferencia por emparejamiento de definiciones con el uso de Matriz

El algoritmo recorre la matriz y compara las relaciones de cada fila ("palabra") y cada columna ("definición"). Al encontrar una coincidencia total con un concepto previo, puede generar una nueva relación, agrupando y clasificando aquellas palabras que comparten definiciones en su totalidad. Estas palabras, al convertirse en subconjuntos de otras definiciones, serán consideradas como candidatas para generar nuevas definiciones. A nivel de aplicación, esto es muy útil para servir como guía de la fuerte relación que pueden tener conceptos que, a simple vista, no parecen tener una fuerte conexión semántica para el lexicógrafo.

De igual manera, la estructura de la ontología permite realizar consultas para mostrar qué otras palabras o definiciones han sido subconjuntos de otras. Esto permite, de manera análoga al ejemplo de "dar", mostrar nuevas opciones de definiciones a nivel de base de datos, haciendo que dicha búsqueda sea intrínseca a su implementación en la aplicación.

```

return
MATCH (d1:definicion {name: "${definicion}"})
MATCH (d1)-[:DESCRIBE]->(mf1:meta_frase)

MATCH (d2:definicion)-[:DESCRIBE]->(mf2:meta_frase)
WHERE d1 <-> d2

WITH COLLECT(DISTINCT mf1) AS metafrasesD1, d2, COLLECT(DISTINCT mf2) AS metafrasesD2

WITH d2, [mf IN metafrasesD2 WHERE mf IN metafrasesD1] AS metafrasesEnComun, metafrasesD2
WHERE SIZE(metafrasesEnComun) = SIZE(metafrasesD2)

MATCH (d2)-[:DEFINE]->(palabra:palabra)

RETURN d2, palabra, metafrasesEnComun

```

Figura 34 Ejemplo de consulta al grafo NEO4J


```

const generarInferencias = () => {
  const palabras = dataSet?.palabras;

  palabras?.forEach((palabra) => {
    palabra.inferencias = []; // Establecer inferencias como un array vacío
    palabra.sinonimos = []; // Establecer sinonimos como un array vacío

    const definicionesPalabra = palabra.definiciones;
    const inferencias = {};

    console.log(`1-Palabra actual: ${palabra.palabra}`);
    console.log(`2-Definiciones palabra actual: ${definicionesPalabra}`);

    palabras.forEach((otraPalabra) => {
      if (otraPalabra.id !== palabra.id) {
        console.log(`3-Otra palabra: ${otraPalabra.palabra}`);

        const definicionesOtraPalabra = otraPalabra.definiciones;
        const definicionesComunes = definicionesPalabra.filter((def) =>
          definicionesOtraPalabra.includes(def)
        );

        console.log(`4-Definiciones comunes: ${definicionesComunes}`);
        console.log(`5-Definiciones otra palabra: ${definicionesOtraPalabra}`);

        if (
          definicionesComunes.length === definicionesOtraPalabra.length &&
          definicionesOtraPalabra.length < definicionesPalabra.length &&
          definicionesPalabra.length > 0 &&
          definicionesOtraPalabra.length > 0
        ) {
          console.log(`6-Entra en el bloque |`);

          const definicionesFaltantes = definicionesPalabra.filter(
            (def) => !definicionesOtraPalabra.includes(def)
          );

          console.log(`7-Definiciones faltantes: ${definicionesFaltantes}`);

          const definicionesCorrespondientes = definicionesFaltantes.filter(
            (def) => !definicionesComunes.some((comun) => def < comun)
          );

          console.log(`8-Definiciones correspondientes: ${definicionesCorrespondientes}`);

          if (
            definicionesCorrespondientes.length > 0 &&
            definicionesCorrespondientes.length === definicionesFaltantes.length
          ) {
            inferencias[otraPalabra.id] = definicionesCorrespondientes;
          }
        }
      }
    });

    if (
      definicionesComunes.length === definicionesPalabra.length &&
      definicionesPalabra.length !== 0 &&
      definicionesComunes.length === definicionesOtraPalabra.length
    ) {
      console.log(`9-Agregando sinonimo: ${otraPalabra.id}`);
      palabra.sinonimos.push(otraPalabra.id);
    }
  });

  if (Object.keys(inferencias).length > 0) {
    palabra.inferencias = inferencias;
  }

  console.log(`10-Inferencias de la palabra actual: ${JSON.stringify(inferencias)}`);
  console.log(`11-Sinonimos de la palabra actual: ${JSON.stringify(palabra.sinonimos)}`);
});

```

Figura 35 Implementacion de Algoritmo en NextJs , Javascript y Nodejs

4.7.3.1.2. Validación de coherencia con ChatGPT



Figura 36 Score de coherencia con ChatGPT

Se realizo pruebas de validación con chatGPT y todas las definiciones superaron el score de 90/100.

PORCETAJE DE MUESTRA	SCORE de coherencia chatGPT
10%	90 a 95
80%	Mas de 95
10%	100

Tabla 10 porcentaje de resultados de score con ChatGPT

CAPITULO 5

5. Conclusiones

El presente trabajo tuvo como objetivo principal mostrar un enfoque innovador sobre el trato del corpus lexicográfico, aplicando principios de la folksonomía para tratar el corpus como un ente que evoluciona en tiempo real según la retroalimentación que la propia construcción del diccionario le brinda al lexicógrafo. Esto permite tomar decisiones más informadas al momento de seleccionar los significados de las entradas del diccionario, aprovechando el uso de wordclouds para informar sobre la frecuencia y comprensión de las palabras. Este enfoque es especialmente útil para diccionarios escolares, donde las definiciones que contienen palabras de uso común pueden ser más comprensibles para el público objetivo.

La implementación de métodos para calcular y propagar propiedades como el peso semántico y la censura sugiere que es posible mantener corpus altamente personalizables y actualizados. Este enfoque no solo podría agilizar la construcción de diccionarios, sino que también permitiría compartir corpus entre lexicógrafos, proporcionando un punto de partida enriquecido para futuras obras lexicográficas. Por ejemplo, un corpus utilizado para crear un diccionario escolar puede ser aprovechado por otro lexicógrafo en un escenario similar, contando ya con información sobre las palabras más frecuentemente utilizadas y su grado de censura.

La adopción de un enfoque folksonómico ha permitido un crecimiento orgánico del corpus lexicográfico, mostrando cómo este se expande y evoluciona a medida que se ingresan más textos y frases. A diferencia de los métodos tradicionales, este enfoque ha mostrado ser más fluido y adaptativo, reflejando en tiempo real los cambios y tendencias en el uso del lenguaje. La ontología no solo facilita la organización del conocimiento lingüístico, sino que también abre nuevas vías para la exploración y comprensión del lenguaje al permitir la inferencia y el descubrimiento de nuevo conocimiento mediante el análisis de las relaciones y patrones dentro del corpus.

Finalmente, la simplicidad del diseño de la ontología y su potencia en la ejecución proporcionan una base sólida para futuras aplicaciones en lexicografía y procesamiento del lenguaje. Se presenta una propuesta de interfaz web para motivar y facilitar la implementación de esta ontología. Este trabajo establece una base para futuros estudios y ofrece vastas posibilidades para expandir y profundizar en los métodos propuestos, contribuyendo significativamente al campo y preparando el camino para futuras innovaciones.

CAPITULO 6

6. Bibliografía

- Almansa Ibáñez, S. (2022). La presencia de las terminologías en la historia de la lexicografía española. Breve panorámica desde los orígenes hasta el siglo XIX. *FRASEOLEX. Revista Internacional de Fraseología y Lexicología*, 1, 45-62.
<https://doi.org/10.5565/rev/fraseolex.41>
- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.
- Battaner Arias, P. (2018). El fenómeno de la polisemia en la lexicografía actual: Otra perspectiva*. *Revista de Lexicografía*, 14, 7-25.
<https://doi.org/10.17979/rlex.2008.14.0.3836>
- Bellandi, A. (2021a). LexO: An open-source system for managing OntoLex-Lemon resources. *Language Resources and Evaluation*, 55(4), 1093-1126.
<https://doi.org/10.1007/s10579-021-09546-4>
- Bellandi, A. (2021b). LexO: An open-source system for managing OntoLex-Lemon resources. *Language Resources and Evaluation*, 55(4), 1093-1126.
<https://doi.org/10.1007/s10579-021-09546-4>
- Bercea, I. O., & Even, G. (2021). Dynamic Dictionaries for Multisets and Counting Filters with Constant Time Operations. En A. Lubiw, M. Salavatipour, & M. He (Eds.), *Algorithms and Data Structures* (Vol. 12808, pp. 144-157). Springer International Publishing. https://doi.org/10.1007/978-3-030-83508-8_11

- Bercea, I. O., & Even, G. (2023). Dynamic Dictionaries for Multisets and Counting Filters with Constant Time Operations. *Algorithmica*, 85(6), 1786-1804.
<https://doi.org/10.1007/s00453-022-01057-0>
- Bergenholtz, H., Nielsen, S., & Tarp, S. (2021). *Lexicography at a Crossroads*.
<https://www.peterlang.com/document/1105942>
- Bergenholtz, H., & Tarp, S. (1995). *Manual of specialised lexicography: The preparation of specialised dictionaries* (Vol. 12). John Benjamins Publishing.
- Bustos Plaza, A., & Wiegand, H. E. (2006). Condensación textual lexicográfica: Esbozo de una concepción integral. *Revista de Lexicografía*, 12, 7-46.
<https://doi.org/10.17979/rlex.2006.12.0.4766>
- Cagle, K. (s. f.). *Taxonomies vs. Ontologies*. Forbes. Recuperado 7 de octubre de 2023, de <https://www.forbes.com/sites/cognitiveworld/2019/03/24/taxonomies-vs-ontologies/>
- Caselli, A., Falquet, G., & Métral, C. (2021). KNOWLEDGE GRAPH CONSTRUCTION FOR SUBSURFACE OBJECTS INCLUDING UNCERTAINTY AND TIME VARIATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4-W4-2021*, 131-136. ISPRS TC IV
16th 3D GeoInfo Conference 2021 - 11–14 October 2021, New York City, USA. <https://doi.org/10.5194/isprs-archives-XLVI-4-W4-2021-131-2021>
- Compastié, M., López Martínez, A., Fernández, C., Gil Pérez, M., Tsarsitalidis, S., Xylouris, G., Mlakar, I., Kourtis, M. A., & Šafran, V. (2023). PALANTIR: An NFV-Based Security-as-a-Service Approach for Automating Threat Mitigation. *Sensors*, 23(3). <https://doi.org/10.3390/s23031658>

Conde, A., Larranaga, M., Arruarte, A., & Elorriaga, J. A. (2019a). A Combined Approach for Eliciting Relationships for Educational Ontologies Using General-Purpose Knowledge Bases. *IEEE Access*, 7, 48339-48355.

<https://doi.org/10.1109/ACCESS.2019.2910079>

Conde, A., Larranaga, M., Arruarte, A., & Elorriaga, J. A. (2019b). A Combined Approach for Eliciting Relationships for Educational Ontologies Using General-Purpose Knowledge Bases. *IEEE Access*, 7, 48339-48355.

<https://doi.org/10.1109/ACCESS.2019.2910079>

Diccionario CLAVE. Lengua española. (2020, noviembre 23). Grupo SM.

<https://www.grupo-sm.com/es/book/diccionario-clave-lengua-espa%C3%B1ola>

Dictionary Types Definition—Grammar Terminology. (s. f.). UsingEnglish.Com.

Recuperado 2 de julio de 2023, de

<https://www.usingenglish.com/reference/dictionary-types/>

España, B. N. de. (2011, septiembre 26). *Diccionarios monolingües.* Biblioteca Nacional de España.

<https://www.bne.es/es/Micrositios/Guias/ObrasReferencia/Diccionarios/Tipologia/Monolingues/>

Ferreyra, D. (2021). Sistemas de organización del conocimiento: Propuesta de un modelo unificado de definición. *Palabra Clave (La Plata)*, 11(1).

<https://www.redalyc.org/journal/3505/350568363008/html/>

He, X., & Yiu, S. M. (2022). Controllable Dictionary Example Generation: Generating Example Sentences for Specific Targeted Audiences. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 610-627. <https://doi.org/10.18653/v1/2022.acl-long.46>

<https://www.facebook.com/alberto.bustos>. (2012, junio 16). *Hiperonimia*. BLOG DE LENGUA. <https://blog.lengua-e.com/2012/hiperonimia/>

<https://www.facebook.com/thoughtcodotcom>. (s. f.-a). *What Is Lexicography?* ThoughtCo. Recuperado 2 de julio de 2023, de <https://www.thoughtco.com/what-is-lexicography-1691229>

<https://www.facebook.com/thoughtcodotcom>. (s. f.-b). *What Is Lexicography?* ThoughtCo. Recuperado 2 de julio de 2023, de <https://www.thoughtco.com/what-is-lexicography-1691229>

I. Introducción. (1989). En H. Hernández, *Los diccionarios de orientación escolar* (pp. 1-14). DE GRUYTER. <https://doi.org/10.1515/9783111340562.1>

JavaScript | MDN. (2023, julio 24). <https://developer.mozilla.org/es/docs/Web/JavaScript>

Kang, S.-J., & Lee, J.-H. (2001). Semi-Automatic Practical Ontology Construction by Using a Thesaurus, Computational Dictionaries, and Large Corpora. *Proceedings of the ACL 2001 Workshop on Human Language Technology and Knowledge Management*. <https://aclanthology.org/W01-1006>

Lara, L. F. (s. f.). *La construcción del diccionario*.

Lexicography. (s. f.-a). Recuperado 4 de julio de 2023, de https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/index.html?https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/microstructure.html

Lexicography. (s. f.-b). Recuperado 4 de julio de 2023, de https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/in

[dex.html?https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/microstructure.html](https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/microstructure.html)

Lipski, J. M., & Lara, L. F. (1998). Teoría del diccionario monolingüe. *Hispania*, 81(3), 568. <https://doi.org/10.2307/345668>

Liu, L., Zhao, R., Du, B., Fung, Y. R., Ji, H., Xu, J., & Tong, H. (s. f.). *Knowledge Graph Comparative Reasoning for Fact Checking: Problem Definition and Algorithms*.

Lubiw, A., Salavatipour, M., & He, M. (Eds.). (2021). *Algorithms and Data Structures: 17th International Symposium, WADS 2021, Virtual Event, August 9–11, 2021, Proceedings* (Vol. 12808). Springer International Publishing. <https://doi.org/10.1007/978-3-030-83508-8>

Luna, J. A. G., Bonilla, M. L., & Torres, I. D. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica*, 2(50), Article 50. <https://doi.org/10.22517/23447214.6693>

Martínez, A. A., Gonzales, S. C., Cruz, E. G., Rueda, C. M., Juárez, T., Salazar, C. T., Lizárraga, C. A., de Piura, U., & Pérez, J. C. (s. f.). *CORRECTORES DE EDICIÓN: Dra. María Clotilde Chavarría Mendoza*.

Molina Sangüesa, I. (2023). Diseño de una ontología aplicada a la lexicografía histórica digital. *Círculo de Lingüística Aplicada a la Comunicación*, 93, 229-242. <https://doi.org/10.5209/clac.72654>

Neo4j Graph Database & Analytics – The Leader in Graph Databases. (s. f.). Graph Database & Analytics. Recuperado 18 de junio de 2024, de <https://neo4j.com/>

Online, TI. (s. f.). *TLex Lexicography, Terminology and Corpus Software*.

Recuperado 14 de julio de 2024, de <http://tshwanedje.com/tshwanelex/>

Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25-29. <https://doi.org/10.5120/ijca2018917395>

¿Qué es la ontología en informática? – *Aplicaciones Informaticas*. (s. f.).

Recuperado 24 de mayo de 2023, de

<https://basicinfoweb.wordpress.com/2016/04/20/que-es-la-ontologia-en-informatica/>

Ravichandran, T., & Rai, A. (2003). Structural analysis of the impact of knowledge creation and knowledge embedding on software process capability. *IEEE Transactions on Engineering Management*, 50(3), 270-284. IEEE Transactions on Engineering Management. <https://doi.org/10.1109/TEM.2003.817278>

Reference Sources: Types of Dictionaries: General Language Dictionaries- (For CBSE, ICSE, IAS, NET, NRA 2024) FlexiPrep. (s. f.-a). Recuperado 2 de julio de 2023, de <https://www.flexiprep.com/NIOS-Notes/Senior-Secondary/Library-Science/NIOS-Library-Science-Unit-7-Reference-Sources-Part-2.html>

Reference Sources: Types of Dictionaries: General Language Dictionaries- (For CBSE, ICSE, IAS, NET, NRA 2024) FlexiPrep. (s. f.-b). Recuperado 2 de julio de 2023, de <https://www.flexiprep.com/NIOS-Notes/Senior-Secondary/Library-Science/NIOS-Library-Science-Unit-7-Reference-Sources-Part-2.html>

Reference Sources: Types of Dictionaries: General Language Dictionaries- (For CBSE, ICSE, IAS, NET, NRA 2024) FlexiPrep. (s. f.-c). Recuperado 2 de julio

de 2023, de <https://www.flexiprep.com/NIOS-Notes/Senior-Secondary/Library-Science/NIOS-Library-Science-Unit-7-Reference-Sources-Part-2.html>

Sandoval, A. M. (1999, septiembre 1). *Diccionarios: Calidad, utilidad y tipos*. RdL - Revista de Libros. <https://www.revistadelibros.com/diccionarios-calidad-utilidad-y-tipos/>

Sandoval Cantor, A. E. (2007). Uso de ontologías y Web Semántica para apoyar la gestión del conocimiento. *Ciencia e Ingeniería Neogranadina*, 17(2), 111-129. <https://doi.org/10.18359/rcin.1079>

Selvakumar, A., & Sudha, R. (2009). An approach of concept similarity computation in taxonomy ontology. *2009 International Conference on Intelligent Agent & Multi-Agent Systems*, 1-4. <https://doi.org/10.1109/IAMA.2009.5228048>

Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: When is it useful? *Journal of Information Science*, 34(1), 15-29. <https://doi.org/10.1177/0165551506078083>

Tejaswi, T., & Murali, G. (2016). Semantic inference method using ontologies. *2016 International Conference on Communication and Electronics Systems (ICCES)*, 1-6. <https://doi.org/10.1109/CESYS.2016.7889858>

Thorn, C., Eriksson, O., Blomqvist, E., & Sandkuhl, K. (2005a). Potentials and Limits of Graph-Algorithms for Discovering Ontology Patterns. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 1, 174-179. <https://doi.org/10.1109/CIMCA.2005.1631261>

Thorn, C., Eriksson, O., Blomqvist, E., & Sandkuhl, K. (2005b). Potentials and Limits of Graph-Algorithms for Discovering Ontology Patterns. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 1, 174-179.
<https://doi.org/10.1109/CIMCA.2005.1631261>

Wang, B., Shen, T., Long, G., Zhou, T., Wang, Y., & Chang, Y. (2021a). Structure-Augmented Text Representation Learning for Efficient Knowledge Graph Completion. *Proceedings of the Web Conference 2021*, 1737-1748.
<https://doi.org/10.1145/3442381.3450043>

Wang, B., Shen, T., Long, G., Zhou, T., Wang, Y., & Chang, Y. (2021b). Structure-Augmented Text Representation Learning for Efficient Knowledge Graph Completion. *Proceedings of the Web Conference 2021*, 1737-1748.
<https://doi.org/10.1145/3442381.3450043>

WordNet. (s. f.). WordNet. Recuperado 14 de julio de 2024, de
<https://wordnet.princeton.edu/homepage>

Zouaq, A., & Martel, F. (2020). What is the schema of your knowledge graph?: Leveraging knowledge graph embeddings and clustering for expressive taxonomy learning. *Proceedings of The International Workshop on Semantic Big Data*, 1-6. <https://doi.org/10.1145/3391274.3393637>

FIN