

Capítulo **4**

Sistema de búsqueda para el lenguaje de señas empleando lematización, sinónimos y antónimos

*Alan Ramírez-Noriega¹
Gibrán U. López Coronel¹
Ángeles Quezada²
Juan Carlos Guzmán¹*

<https://doi.org/10.61728/AE24050043>

¹ Universidad Autónoma de Sinaloa

² Instituto Tecnológico de Tijuana

Resumen

Un Sistema de Recuperación de Información se encarga de recuperar, almacenar y proporcionar información de acuerdo con las consultas de los usuarios. Sin embargo, las consultas de los usuarios suelen ser desestructuradas e incompletas, lo que dificulta que un motor de búsqueda satisfaga completamente sus necesidades. Por lo tanto, se utilizan técnicas de Expansión de Consulta (EC) para mejorar el rendimiento de la recuperación de información. En el contexto de un sistema de recuperación de información basado en videos de Lenguaje de Señas Mexicano, la búsqueda se realiza solo a partir de la descripción y el nombre de los videos, lo que complica el proceso. El sistema de búsqueda de contenido propuesto inicia con la corrección ortográfica de las palabras de búsqueda, seguido de la lematización para encontrar la raíz de las palabras. Luego se aplican las técnicas de EC, como la búsqueda por sinónimos, antónimos y palabras relacionadas. Los resultados de la búsqueda se presentan organizados por categorías. Se realizaron experimentos para evaluar el rendimiento del sistema de búsqueda utilizando palabras de prueba. Los resultados mostraron las palabras derivadas del lexema, sinónimos, antónimos y palabras relacionadas para cada palabra de búsqueda. Se evaluó el éxito de la búsqueda en función de la coincidencia de palabras y se obtuvieron porcentajes de éxito adecuados considerando las condiciones del experimento.

Introducción

Un sistema de recuperación de información se define como un conjunto de procesos desarrollados con el fin de recuperar, almacenar y regresar información de acuerdo con las consultas de un usuario (Rey et al., 2014).

Un sistema de recuperación de información funcionaría muy bien si el usuario pudiera transmitir su necesidad de información de manera precisa. Sin embargo, la consulta del usuario rara vez es completa. La consulta proporcionada por el usuario es, a menudo, desestructurada e incompleta. Una consulta incompleta impide que un motor de búsqueda

da satisfaga la necesidad de información del usuario. En la práctica, es necesaria alguna representación que pueda expresar correcta y completamente la necesidad de información del usuario (Kankaria, 2005).

Un método para mejorar el rendimiento de la recuperación de información es complementar una consulta original con términos adicionales, es decir, reformular una consulta para tener mayor éxito en la recuperación de información, a esto se le llama Expansión de la Consulta (EC) (Imran y Sharan, 2010; Tavares da Silva y Maia, 2019). La EC se puede realizar de forma automática o interactiva y puede tener lugar en la formulación inicial de la consulta, o en una etapa de reformulación de la consulta de la búsqueda en línea o en ambas (Tavares da Silva y Maia, 2019).

Las siguientes son algunas de las técnicas simples de expansión de consultas (Kankaria, 2005):

- Encontrar sinónimos de palabras y hacer búsquedas por estos nuevos sinónimos;
- Encontrar todas las diversas formas morfológicas de palabras, aplicando la técnica de Stemming o lematización (Peinado Rodríguez, 2003) en cada palabra en la consulta de búsqueda;
- Corrección de errores ortográficos y búsqueda automática de la forma corregida o sugerirla
- Volver a ponderar los términos en la consulta original;
- Crear un diccionario de términos de expansión para cada término y luego buscar en el diccionario para la expansión en posteriores consultas.

Actualmente, se está desarrollando un sistema de recuperación de información basado en videos, es decir, el retorno de las consultas es para ver videos. Sin embargo, este es un tipo especial de video, los videos se caracterizan por ser cortos, no es necesario el sonido y el video muestra señas de la Lengua de señas mexicana. Por lo que, la consulta se basa solamente en la descripción y nombre que los usuarios hagan de los videos. Esto complica la búsqueda, por cual, para tener un mayor éxito de recuperación de videos se aplican técnicas de EC.

Las técnicas de EC para el sistema en desarrollo se basan en la corrección de ortografía para evitar búsquedas que no existan; lematiza-

ción para buscar la raíz de la palabra en lugar de la palabra completa; búsqueda por sinónimos para ampliar las posibilidades de búsqueda; búsqueda de antónimos para brindar señas opuestas a la búsqueda originalmente y palabras relacionadas para ampliar la frase que una persona pudiera mencionar.

Por tanto, el objetivo del presente artículo es probar el funcionamiento del algoritmo de búsqueda de videos basados en la EC. Dado que el sistema está en sus primeras fases de desarrollo el experimento se simula para determinar el alcance del algoritmo.

Este artículo está organizado de la siguiente forma. En la sección dos se describen los principales temas de la investigación, tales como las personas sordas y el lenguaje de señas. En la sección tres se muestran la forma de funcionar del sistema propuesto, así como el flujo de datos. En la cuarta sección se muestra la experimentación, los resultados y la discusión. Por último, se plasman las conclusiones y las referencias.

Marco teórico

Personas sordas

Los discursos médicos y el sentido común, suele definir a los sordos en términos negativos, en relación con poseer una audición limitada o nula, y se piensa que presentan una discapacidad del lenguaje. Una perspectiva psicosociolingüística se distancia de estas nociones, y entiende a la sordera como “el rastro que deja la diferencia biológica oír/no oír tanto en el ámbito discursivo, psicológico, social e institucional, como en el conjunto de creencias” (Diconca Malaquín, 2016).

Desde una perspectiva antropológica, las personas sordas forman una minoría lingüística y cultural con unas características particulares. A pesar de la gran diversidad de perfiles lingüísticos, subyace un elemento biológico que tiene relevancia en la conformación de su identidad y que sigue una perspectiva de promoción de la lengua de señas, que está detrás de los modelos educativos bilingües y binoculares (CNSE, 2019).

El concepto de sordera se refiere un trastorno auditivo profundo que impide la comunicación a través del lenguaje hablado; sin embargo, actualmente se le reconoce también como un espectro que incluye desde la deficiencia parcial hasta la total para percibir sonidos de manera nítida, ya sean ambientales o del lenguaje oral, debido a alteraciones anatómicas o fisiológicas (Ramírez Jiménez, 2018).

Las causas de la sordera las subdividen en congénitas y adquiridas. La sordera congénita obedece a factores genéticos que impactan sobre la conformación y fisiología del sistema auditivo o de las áreas auditivas del sistema nervioso; estas a su vez pueden identificarse como sindrómicas y no sindrómicas (Ramírez Jiménez, 2018).

Las sorderas adquiridas pueden desarrollarse de manera temprana o tardía durante el ciclo de vida, entre las causas que pueden originar la pérdida auditiva en los periodos pre y perinatal se encuentran: infecciones de la madre durante el embarazo, bajo peso al nacer, meningitis bacteriana, accidente hipóxico-isquémico, ictericia grave, traumatismos obstétricos, etcétera; pero, son las infecciones e inflamaciones de oído crónicas, el uso inadecuado o prolongado de algunos medicamentos, presencia de líquido en los oídos, traumatismos craneoencefálicos o de los oídos, exposición a ruido excesivo, obstrucción del conducto auditivo por cerumen o cuerpos extraños, entre otros (Ramírez Jiménez, 2018).

Los sordos son personas eminentemente visuales, tienen una lengua viso gestual que es propia, llamada la lengua de señas o lenguaje de señas. Presentan una serie de costumbres y valores como la manera de saludarse, despedirse, felicitar y de aplaudir que le son particulares. Poseen sus propias reglas de interacción social, así como clubes y asociaciones donde permiten ser y expresarse, surgidos de su condición visual y de la necesidad humana de socializar entre pares y compartir intereses. Con ello, se afirma el hecho de referirse a los sordos simplemente como “sordos”, rechazando los términos como sordomudo, discapacitado de la comunicación o términos similares. Desde la cultura mayoritaria se define a la persona sorda por lo que no tiene y por lo que no puede ser, desconociendo su capacidad sociocultural y lingüística (Diconca Malaquín, 2016).

Lenguaje de señas

El Diario Oficial de la Federación en el apartado de la Ley General de las Personas con Discapacidad, define a la Lengua de señas como la lengua de una comunidad de sordos, la cual consiste en: “una serie de signos gestuales articulados con las manos y acompañados de expresiones faciales, mirada intencional y movimiento corporal, dotados de función lingüística”. Además, menciona que “forma parte del patrimonio lingüístico de dicha comunidad y es tan rica y compleja tanto en gramática como en vocabulario como cualquier lengua oral” (Instituto para las Personas con Discapacidad de la Ciudad de México, 2017).

Lenguaje de Señas Mexicano (LSM)

El lenguaje de señas mexicano (LSM) es el lenguaje que usan las personas Sordas y no Sordas para comunicarse en México. El LSM se estructura mediante el uso del cuerpo en el espacio. Dentro de este espacio proyectado existen tres límites: el vertical, el horizontal y un tercero que indica la proximidad de las manos con respecto al cuerpo. El límite vertical va de la cintura a la coronilla de la cabeza; el límite horizontal, hasta la altura de los codos, con los brazos doblados. Cualquier movimiento que sobrepase estas dimensiones será interpretado como exageración o énfasis (Instituto para las Personas con Discapacidad de la Ciudad de México, 2017).

La lengua de señas mexicana o la lengua de los signos es una forma de comunicación manual, que depende principalmente de la interacción directa visual para percibir la información lingüística y son utilizadas por las distintas comunidades de hipoacúsicos. Por ser un lenguaje visual, este utiliza dimensiones de espacio y movimiento para transmitir información sobre diversos parámetros espaciales simultáneamente. Los lenguajes de signos son sistemas organizados a partir de movimientos convencionales estructurados gramaticalmente para la comunicación (Real González, 2016).

El uso del LSM permite el desarrollo integral de las personas sordas de México; ya que esta ofrece una cultura rica en historia y valores, ade-

más de que ofrece la posibilidad de acceder a información en diversos contextos y sobre todo a diversas ideologías.

Funcionamiento del sistema de búsqueda de contenido

El esquema general del sistema de búsqueda está representado en la Figura 1. El proceso inicia con el usuario y su necesidad de búsqueda, en este caso serán palabras que representan la consulta.

Las palabras de búsqueda tienen que ser corregidas ortográficamente, ya que todo proceso fracasará si las palabras no están bien escritas. En la Figura 1 se muestra un ejemplo de una sugerencia de corrección, se introduce la palabra “cansión” aunque mal escrita, por lo que se sugiere un cambio en la entrada.

Figura 1: Corrección de ortografía en búsqueda de video



Palabra o Frase

cansión

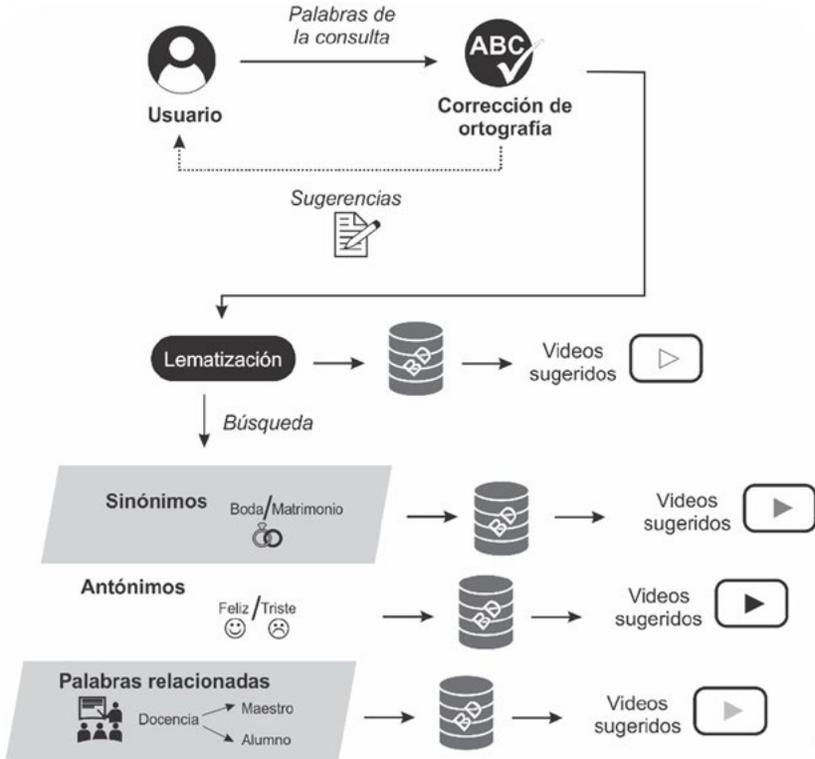
Buscar

¿Quisiste decir **canción**?

Fuente: Elaboración propia.

Posteriormente, el proceso de búsqueda aplica la lematización, el cual es un proceso lingüístico donde se ubica el lema o raíz de la palabra (Peinado Rodríguez, 2003). Por ejemplo, la palabra cantante tiene como raíz el verbo cantar, por lo que el sistema buscaría todos los videos que tengan como lema la palabra cantar; de esta forma, la respuesta a la consulta sería más rica. A su vez, si se busca música, algunas variaciones a música pueden aparecer como músicos o musicales.

Figura 2: Flujo de búsqueda del sistema



Fuente: Elaboración propia.

Posteriormente se emplean los sinónimos, dos palabras son sinónimos en un contexto cuando pueden intercambiarse, es decir son palabras con un significado en común, por ejemplo, comprar y adquirir (Azad y Deepak, 2019). De esta forma, una seña puede ser buscada de acuerdo con su sinónimo, así se amplían las posibilidades de coincidencias. Los sinónimos son más apropiados que las palabras relacionadas para introducir menos ruido y mejorar la eficiencia de la consulta (Wei et al., 2010).

Por otra parte, los antónimos tienen significados opuestos, por ejemplo, blanco y negro (Wang et al., 2010). Aunque en un contexto de señas sería muy útil el uso de verbos, por ejemplo, emplear la palabra amor y odio para representar una frase. Un antónimo negado puede servir para

representar una frase. Es decir, si no sabemos expresar la seña morir podríamos expresar no vivir, completando la idea de la oración con un verbo negado (Bui et al., 2020; Turney, 2008).

La asociación de palabras, es decir, palabras relacionadas por pertenecer a un mismo contexto (Turney, 2008), también pueden ser de utilidad en la búsqueda de señas, ya que brindan información extra sobre el contexto de la búsqueda. De esta forma, al buscar una seña como médico pueden aparecer otras señas relacionadas como hospital o enfermera, estas pueden complementar una oración agregando información extra.

Como se puede apreciar en la Figura 2, las palabras de búsqueda tendrán como resultados videos organizados por categorías. Primeramente, aparecen las palabras con coincidencia exacta, al menos en el proceso de lematización. Posteriormente, aparece videos de acuerdo con una búsqueda por sinónimos, después por antónimos y, finalmente, como información extra de apoyo videos con palabra relacionadas a las palabras iniciales de búsqueda.

Ejemplos de búsqueda

La Tabla 1 muestra los casos para ejemplificar la forma de trabajar el algoritmo, se muestran tres ejemplos organizados por columnas.

De acuerdo con el ejemplo 1, la palabra de búsqueda (entrada) es carros, sin embargo, al obtener su lexema se omiten la ultimas letras de la palabra tomando como raíz “carr”. De esta forma, la palabra de búsqueda es carros, aunque también se busca el lexema carr, así podríamos obtener un video que muestre la seña de carros, aunque también los derivados como carruaje, carro, carrito o carretera. Aunado a esto, podrían obtenerse sinónimos como auto, automóvil y coche. Los antónimos podrían no existir para ciertos casos, simplemente no se muestran videos. Finalmente, se muestran palabras relacionados a la palabra carros como llanta, puerta, motor.

Tabla 1: Ejemplos del sistema de búsqueda

	Ejemplo 1	Ejemplo 2	Ejemplo 3
Entrada	Carros	Alto	Músicos
Lexema	Carr	alt	music
Derivados	Carruaje, Carro, Carrito, carretera	Altitud, altímetro, alta	Música, músico, musicalizar
Sinónimos	Auto, automóvil, coche	Elevado, crecido, arriba	Intérpretes, artis- tas, autores
Antónimos	-	Bajo, pequeño, inferior	-
Palabras relacio- nadas	llanta, puerta, motor	valioso, árbol, largo	Sonido, Guitarra, ritmo

Fuente: Elaboración propia.

El ejemplo 2 y 3 muestran los posibles resultados para las palabras “Alto” y “Músicos”; funcionan de forma similar al ejemplo 1, aunque en este caso para la palabra Alto si pudiera encontrarse antónimos tales como bajo, pequeño e inferior.

De esta forma, se modela un sistema de búsqueda de palabras para emplearse en el uso del lenguaje de señas. Hasta el momento el sistema esta implementado en un sitio web de prueba empleando CSS, HTML, Angular y PHP.

Experimentación

El experimento en esta investigación consiste en buscar videos de señas a través de palabras y evaluar el éxito en los resultados. El objetivo es ver el comportamiento del sistema de búsqueda de Contenido y, si es necesario, ajustarlo.

Descripción del experimento

Para las pruebas se seleccionaron diez palabras del español para buscar en la base de datos del sistema. Dado que estamos en fase de pruebas, se eligió el diccionario de la Real Academia de Española (RAE) (Real

Academia Española, 2022) con 638600 palabras para simular la base de datos de videos, donde cada palabra representa un video. En esta primera etapa se hacen búsquedas por palabras.

Las palabras de búsqueda son: perro, rápido, rojo, viernes, morir, idioma, mano, despertar, bicicleta y hamburguesa. El objetivo es analizar las palabras derivadas por el lexema, los sinónimos, los antónimos y las palabras relacionadas.

Es importante definir el contexto de uso de las palabras del experimento para evitar confusiones de semántica, el significado de las palabras es (Real Academia Española, 2022):

- Perro: Mamífero doméstico de la familia de los cánidos, de tamaño, forma y pelaje muy diversos, según las razas, que tiene olfato muy fino y es inteligente y muy leal a su dueño.
- Rápido: Que se mueve, se hace o sucede a gran velocidad, muy deprisa (adjetivo).
- Rojo: Dicho de un color: Semejante al de la sangre o al del tomate maduro, y que ocupa el primer lugar en el espectro luminoso (adjetivo, sustantivo).
- Viernes: Quinto día de la semana, que es festivo para el islamismo.
- Morir: Llegar al término de la vida (verbo).
- Idioma: Lengua de un pueblo o nación, o común a varios.
- Mano: Parte del cuerpo humano unida a la extremidad del antebrazo y que comprende desde la muñeca inclusive hasta la punta de los dedos.
- Despertar: Cortar o interrumpir el sueño a quien está durmiendo (verbo).
- Bicicleta: Vehículo de dos ruedas, normalmente, de igual tamaño, cuyos pedales transmiten el movimiento a la rueda trasera por medio de un plato, un piñón y una cadena.
- Hamburguesa: Pieza de carne picada aplastada y con forma redondeada, mezclada con diversos ingredientes, que se hace a la plancha, a la parrilla o frita.

El experimento considera comparar la palabra buscada con las palabras devueltas, dependiendo del tipo de palabra, por ejemplo, es neces-

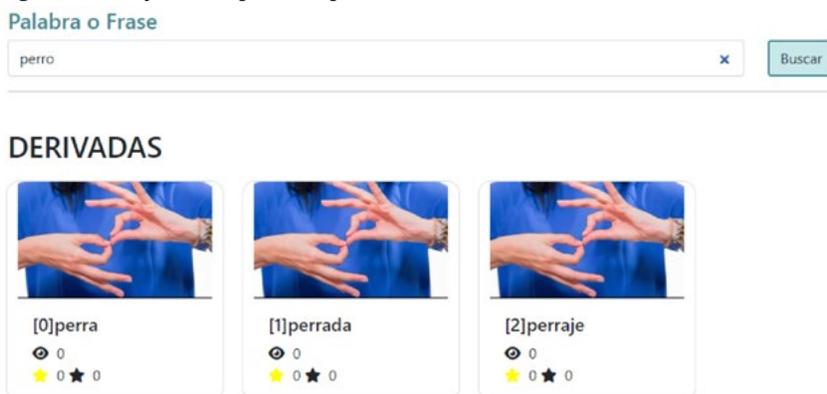
sario definir si una palabra es sinónimo, si es antónimo o si es palabra relacionada. Para definir si coincide con el tipo de palabra se consultó la RAE.

Resultados

La búsqueda de la primera palabra “Perro” arroja los siguientes resultados (ver Figura 3):

- Lexema: perr
- Derivados: perra, perrada, **perraje**, perramenta, **perramente**, **perrera**, **perreaba**, **perreaban**, **perread**, **perreada**, **perreado**, **perran**, **perreando**, **perrear**, **perreara**, **perrearán**, **perreare**, **perrearen**, **perrearia**, **perrearian**, **perrearon**, **perrease**, **perreasen**, **perreaste**, **perreda**, **perree**, **perreen**, **perrengue**, **perreo**, perrera, **perreriam** perrero, perrezno, perrilla, perrillo, perro, perrona, **perrroquete**, perruna, **perrunilla**, perruno. (12/41)
- Sinónimo: can, chucho, gozque (3/3)
- Antónimo: gato (1/1)
- Relacionadas: canesú, perrito, cachorro, canido, canino, cachorra, dogo, caniche, perrería, ventor, podenco, mastín. (12/12)

Figura 3: Búsqueda de palabra perro



Fuente: Elaboración propia.

Los resultados se muestran organizados por categorías, la palabra tachada significa que no coincide con la palabra de búsqueda.

En la búsqueda de la palabra perro, el sistema brinda: 42 palabras derivadas de la raíz, 3 sinónimos, un antónimo y 12 palabras relacionadas. Las palabras derivadas parecen muchas (42), sin embargo, al considerar que es una simulación, el sistema no contendrá tantos videos con tantas palabras que inician con el lexema “perr”, lo que implica una disminución en los resultados de búsqueda. Los sinónimos son adecuados ya que devuelve can, chucho y gozque. El antónimo es muy preciso al encontrar como antónimo de perro a gato.

Es necesario identificar que palabras realmente están relacionadas con el lexema, ya que en las palabras derivadas pueden aparecer muchas palabras que podrían no tener relación con la palabra original, aunque si con el lexema.

Los resultados de la palabra “perro” y del resto de las palabras del experimento están concentradas en la Tabla 2. La tabla contiene las palabras, el lexema derivado de la palabra, las palabras derivadas del lexema, los sinónimos, antónimos y las palabras relacionadas a la palabra de búsqueda. Además, se incluye las palabras con éxito de la búsqueda, el total de palabras y el porcentaje de éxito.

Tabla 2: *Concentrado de resultados del experimento*

#	Búsqueda	lexema	Derivados			Sinónimos			Antónimo			Relacionadas		
			éxito	total	%	éxito	total	%	éxito	total	%	éxito	total	%
1	Perro	perr	12	41	0.29	3	3	1.00	1	1	1.00	12	12	1.00
2	Rápido	Rapid	4	4	1.00	21	27	0.78	0	0	0.00	17	29	0.59
3	Rojos	Roj	0	0	0.00	13	15	0.87	1	1	1.00	23	23	1.00
4	Viernes	Viern	0	0	0.00	0	0	0.00	0	0	0.00	18	22	0.82
5	Morir	Mori	5	5	1.00	12	12	1.00	1	1	1.00	19	20	0.95
6	Idioma	Idioma	2	2	1.00	7	9	0.78	0	0	0.00	14	17	0.82
7	Mano	Mano	52	102	0.51	2	31	0.06	1	1	1.00	18	25	0.72
8	Despertar	Despertar	7	7	1.00	0	0	0.00	0	0	0.00	20	20	1.00
9	Bicicleta	Bibliclet	0	0	0.00	3	3	1.00	0	0	0.00	19	19	1.00
10	Hamburguesa	Hamburgues	1	1	1.00	0	0	0.00	0	0	0.00	13	13	1.00
Totales			83	162	0.51	61	100	0.61	4	4	1.00	173	200	0.87

Fuente: Elaboración propia.

Como se puede apreciar en la Tabla 2, varias búsquedas de palabras no tienen incidencia, dado el tipo de palabra. Por ejemplo, no hay un antónimo para la palabra “viernes”, o también está el caso que no hay palabras derivadas para la palabra “rojo”. Todos estos casos están identificados con cero (0) en la tabla.

Discusión

Las palabras derivadas muestran el rendimiento más bajo de las categorías, ya que muchas palabras no relacionadas pueden tener el mismo inicio de la palabra, tal como se indica en la palabra “mano”. Muchas palabras empiezan con “mano” y no tienen que ver con la parte del cuerpo humano. Lo mismo ocurrió con la palabra “perro”, su lexema “perr” es inicio de muchas palabras que no tienen que ver con el mamífero. Esto ocasiona que el porcentaje de efectividad de las palabras derivadas sea de 0.51. Aunque el resto de las palabras tienen un porcentaje de éxito de 1.

Los sinónimos brindan muy buenos resultados de forma general, por alguna razón la palabra “mano” también arroja muchos sinónimos, aunque la mayoría no coinciden (2 de 32), el rendimiento es de 0.61, pero omitiendo la palabra “mano” es 0.86.

Los antónimos son los más complicados de encontrar, solo fueron encontrados cuatro casos, por ejemplo, para la palabra morir se encontró el antónimo vivir, para mano corresponde pie, para perro se encontró gato, por último, para rojo se encontró azul. Son pocos casos ya que depende del tipo de palabra para tener éxito en el antónimo, por ejemplo, hamburguesa no tiene un antónimo claro, ni idioma, por lo que no se encuentran casos de éxito.

Por último, las palabras derivadas son aquellas que tienen que ver con el contexto de la palabra de búsqueda, hubo 200 palabras encontradas de las cuales 173 fueron de éxito, lo que no permite un 0.87 de rendimiento. Esto indica que la mayoría de las palabras que aparezcan pueden ser empleadas por el usuario en el mismo contexto de la palabra de búsqueda.

De esta manera, es posible apreciar el comportamiento del sistema de búsqueda de Contenido. Aunque es una simulación, podemos apre-

ciar el rendimiento estimado de acuerdo con un sistema con demasiados videos. Al final el funcionamiento está dado por la cantidad de videos y el correcto uso que hagan los usuarios del sistema.

Es importante considerar las características que debe de tener un sistema de búsqueda de este tipo, ya que cubre una necesidad de comunicación a través del lenguaje de señas, por lo que obtener diferentes formas de expresar una palabra puede ser muy útil para facilitar la comunicación.

Conclusiones

El presente artículo explica el funcionamiento de un sistema de búsqueda de videos para el lenguaje de señas. Aunque se define el lenguaje de señas mexicano, el sistema encaja con cualquier lenguaje de señas del mundo. Las búsquedas al sistema devuelven un conjunto de videos relacionados considerando el lexema, sinónimos, antónimos y palabras relacionadas, esto permite ampliar la consulta y tener más opciones para elaborar una oración con lenguaje de señas.

El desarrollo de sistemas de software para resolver problemas de la sociedad siempre es considerar de alto valor. Una herramienta como la propuesta en esta investigación colabora con la comunidad sorda, ya que plantea un sistema de recuperación de información, en este caso videos, para facilitar la comunicación entre las personas sordas. El sistema es empleado para buscar palabras que quieran ser empleadas al momento de comunicarse con el lenguaje de señas. El sistema está diseñado para buscar palabras, pero a su vez devolver una serie de palabras relacionados para tener más riqueza en la comunicación.

Los resultados del experimento indican un rendimiento bueno. El peor caso fue en las palabras derivadas por el lexema, ya que es común encontrar palabras con la misma raíz, pero sin relación. Los sinónimos tuvieron un comportamiento regular, a excepción de la palabra mano que arrojó una lista de casos que no eran sinónimos. Los antónimos como las palabras relacionadas se comportaron de forma adecuada. De esta forma, es posible considerar que cuando el sistema este implementado podría arrojar videos no relacionados a la búsqueda, sin embargo,

estos serán pocos ya que no habrá tantos videos como los considerados en la simulación.

Como trabajo futuro, el sistema de búsqueda será implementado en un Sistema Web Colaborativo basado en videos para traducir palabras o frases del LSM. Es colaborativo porque la comunidad puede hacer sus aportaciones al sistema de su conocimiento de LSM. Para controlar las aportaciones y evitar el uso incorrecto, el sistema contará con un módulo de evaluación de contenido para validar la veracidad de la información subida, este aspecto también será controlado por la comunidad.

Referencias

- Azad, H. K. y Deepak, A. (2019). Query expansion techniques for information retrieval: A survey. *Information Processing and Management*, 56(5), 1698-1735. <https://doi.org/10.1016/j.ipm.2019.05.009>
- Bui, V. T., Dinh, K. Q. y Nguyen, P. T. (2020). Vietnamese Antonyms Detection Based on Specialized Word Embeddings using Semantic Knowledge and Distributional Information. 12th International Conference on Knowledge and Systems Engineering (KSE), 159-164. <https://doi.org/doi:10.1109/KSE50997.2020.9287542>
- CNSE, C. E. de P. S. (2019). Sobre las personas sordas. <https://www.cnse.es/index.php/nuestro-trabajo/publicaciones-y-recursos/item/283-sobre-las-personas-sordas>
- Diconca Malaquín, Y. (2016). Sordos: Una comunidad Lingüística. Universidad de la República.
- Imran, H. y Sharan, A. (2010). A Framework for Automatic Query Expansion. En J. Wang, F. L., Gong, Z., Luo, X., Lei (Ed.), *Web Information Systems and Mining. WISM 2010. Lecture Notes in Computer Science*, vol 6318 (pp. 386–387). https://doi.org/https://doi.org/10.1007/978-3-642-16515-3_48
- Instituto para las Personas con Discapacidad de la Ciudad de México, (INDEPEDI CDMX). (2017). *Diccionario de Lengua de Señas Mexicana* (C. E. Escobedo Delgado, Ed.; 1st ed.).
- Kankaria, A. (2005). Query Expansion techniques. 1-5.
- Peinado Rodríguez, J. (2003). Lematización para palabras médicas complejas: Implementación de un algoritmo en LISP. *Revista Médica Herediana*, 14(3), 223-228.
- Ramírez Jiménez, D. (2018). *Lengua de señas y pensamiento en niños Sordos: propuesta de un programa de estimulación lexical* [Tesis de licenciatura]. Universidad Nacional Autónoma de México.
- Real Academia Española, (RAE). (2022, December 20). *Diccionario de la Lengua Española*. <https://dle.rae.es/>.
- Real González, R. E. (2016). *Modelo de Producción y Curso Virtual de Computación en Lengua de Señas Mexicana (LSM)*. Instituto Politécnico Nacional.

- Rey, M., Kuna, H., Martini, E., Podkova, L., Pautsch, G. y Zamudio, E. (2014). Generación de un Método de Expansión de Consultas Basado en Ontologías para un Sistema de Recuperación de Información. XX Congreso Argentino de Ciencias de La Computación, 1-10.
- Tavares da Silva, F. y Maia, J. E. B. (2019). Query Expansion in Text Information Retrieval with Local Context and Distributional Model. *Journal of Digital Information Management*, 17(6), 313-320. <https://doi.org/10.6025/jdim/2019/17/6/313-320>
- Turney, P. D. (2008). A uniform approach to analogies, synonyms, antonyms, and associations. *Coling 2008 - 22nd International Conference on Computational Linguistics, Proceedings of the Conference*, 905-912. <https://doi.org/10.3115/1599081.1599195>
- Wang, W., Thomas, C. y Chan, V. (2010). Pattern-based synonym and antonym extraction. *Proceedings of the Annual Southeast Conference*, 4-7. <https://doi.org/10.1145/1900008.1900094>
- Wei, X., Peng, F., Tseng, H., Lu, Y., Wang, X. y Dumoulin, B. (2010). Search with synonyms: Problems and solutions. *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, 1318-1326.