

MIDA SUS PALABRAS, JOVEN MATEMÁTICAS ENTRE CUERVOS Y ESCRITORIOS

MEASURE YOUR WORDS, YOUNG MAN
MATHEMATICS BETWEEN CROWS AND DESKS

Daniel Mocencáhua Mora

ISSN 2448-5829

Año 11, No. 31, 2025, pp. 64 - 71

RD-ICUAP

Facultad de Ciencias de la Electrónica
Benemérita Universidad Autónoma de Puebla

*Autor de correspondencia: daniel.mocencahua@correo.buap.mx

<https://orcid.org/0000-0003-4718-7442>



Resumen

Este artículo explora el clásico acertijo de Lewis Carroll, “¿En qué se parece un cuervo a un escritorio?”, desde una perspectiva matemática, en particular por medio de métricas de similitud y distancia entre palabras. Partiendo del juego literario y sus diversas respuestas, examinamos como conceptos matemáticos como la distancia de Hamming, la similitud de Jaccard, el coseno entre vectores y la distancia de Levenshtein permiten cuantificar relaciones entre palabras. Cada una de estas métricas, al medir diferencias y semejanzas, ofrece una aproximación numérica que aporta nuevos matices a la idea de similitud, a diferencia de las interpretaciones puramente literarias. El análisis se ilustra con una comparación entre las palabras “cuervo” y “escritorio”, demostrando cómo cada métrica arroja distintos grados de similitud. Finalmente, reflexionamos sobre la importancia de estos métodos en aplicaciones tecnológicas actuales, desde procesamiento del lenguaje natural hasta sistemas de recomendación, donde entender la “distancia” entre palabras y conceptos es esencial para personalizar y mejorar la interacción humana con la tecnología. Así, como en el acertijo de Carroll, este artículo muestra que, con el enfoque adecuado, es posible hallar conexiones entre ideas que parecían distantes.

Palabras clave: Similitud de palabras, Procesamiento del lenguaje natural, Análisis de texto, Cuervo y escritorio, Métricas de similitud.

Abstract

This article examines Lewis Carroll's classic riddle, "How does a raven resemble a desk?," from a mathematical standpoint, with a focus on metrics of similarity and distance between words. We will begin by examining the literary play and its various responses. We will then move on to examine how mathematical concepts such as Hamming distance, Jaccard similarity, cosine between vectors, and Levenshtein distance allow us to quantify relationships between words. Each of these metrics provides a numerical approach to measuring differences and similarities, offering a new level of detail and nuance to the concept of similarity beyond what can be achieved through purely literary interpretations. The analysis is illustrated with a comparison between the words "crow" and "desk," demonstrating how each metric yields different degrees of similarity. Finally, we consider the value of these methods in current technological applications, from natural language processing to recommender systems. Understanding the "distance" between words and concepts is essential to personalize and improve human interaction with technology. As demonstrated by Carroll's riddle, a suitable approach can reveal connections between seemingly disparate ideas.

Keywords: Word similarity, Natural language processing, Text analysis, Cuervo y escritorio, Similarity metrics.

Cuervo y escritorio

Alicia en el país de las maravillas es un libro que todos debemos leer algún día. En el capítulo 7 Alicia es cuestionada por el sombrerero con un acertijo que desde 1865 ha llamado la atención: Why is a raven like a writing-desk?

Desafortunadamente Alicia no logra dar la respuesta, y los demás dicen no saberla, ni siquiera el autor del libro, y no es hasta 1869, que Carroll da su solución: Because it can produce very few notes, though they are very flat; and it is never put with the wrong end in front.

Este acertijo se vuelve una desafío que algunos autores aceptaron: Because there's a b in both, and because there's an n in neither (Adous Huxley), Because the notes for which they are noted are not noted for being musical notes (Sam Loyd), las cuales pierden su gracia si se traducen (Gamero, 2014). Pero una respuesta que si se puede traducir es "porque sobre ambos escribió Edgar Allan Poe" (Igalada, 2009). En la figura 1 tenemos una red semántica del poema El cuervo de Poe que muestra una posible asociación entre nuestros palabras de interés.

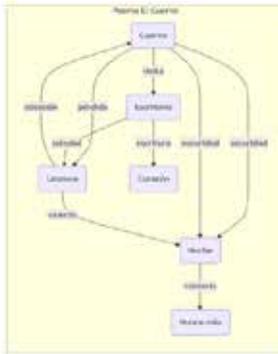


Figura 1. Red semántica del poema El cuervo
Elaboración propia

Y ahora que la veo hasta se antoja una metáfora: "El cuervo es el guardián silencioso de los secretos que yacen en el escritorio."

Pero volvamos a lo nuestro. Si

traducimos la pregunta literalmente obtendríamos algo como ¿por qué un cuervo es como un escritorio? y así cobran sentido las respuestas mostradas. En español se traduce generalmente como "¿en qué se parece un cuervo a un escritorio?" (Carroll, 2003, p. 63)

Con estas dos traducciones podemos abordar la pregunta desde dos enfoques: que tienen en común ambas palabras, o que tanto que se parecen. En lugar de buscar una respuesta puramente ingeniosa, exploremos cómo la matemática puede ofrecernos un enfoque alternativo.

Como matemático no puedo dar una respuesta literaria ingeniosa al acertijo, pero los enfoques mencionados me conducen a considerar los concepto de medidas y similitudes. Me atrevo a pensar en las palabras como puntos en un espacio específico, o como conjuntos, y por lo tanto pienso en medidas de distancia para el primer caso o similitud para el segundo.

Notemos que la distancia es una medida de diferencia, mientras que la similitud es una medida de parecido. Y hay una relación interesante: a mayor distancia, menor similitud. Creo que por ese camino vamos a llegar a algo.

Con la vara que midas...

Empecemos contando las letras: cuervo (6), escritorio (10), la diferencia en longitud es evidente, pero no es lo que buscamos. ¿Cuántas letras distintas tiene cada una? cuervo (6), escritorio (7), siguen siendo diferentes en eso, pero ¿qué tan iguales son? Esto de centrarnos en las letras nos va a servir ahora.

Hamming

Hacemos lo siguiente: alineamos las palabras y vemos donde coinciden las letras. A esto se le llama distancia de Hamming, la podemos apreciar en la Figura 2.

Distancia de Hamming

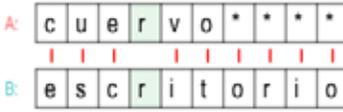


Figura 2. Distancia de Hamming
Elaboración propia

Hay 9 diferencias, marcadas con rojo, es decir, para Hamming, estas palabras son muy diferentes. Hemos completado la cadena de la palabra “cuervo” con asteriscos porque Hamming solo se puede aplicar a cadenas de igual longitud. Nota que, si sus diferencias son 9 de 10, su similitud es $1/9 \approx 0.11$.

Jaccard

Si vemos nuestras palabras como conjuntos podemos usar medidas de similitud como la de Jaccard (j), que definimos para conjuntos A y B por

$$j(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Cabe mencionar que aclarar que $|A \cap B|$ representa el número de letras comunes y $|A \cup B|$ el número de letras únicas combinadas (Kannan et al., 2016, p. 63).

Es decir, contamos cuantas letras (distintas) tienen en común (4) y dividimos entre el total de letras distintas que juntan ambas palabras (9). Si A es el conjunto de las letras de “cuervo” y B el conjunto de letras de la palabra “escritorio” tenemos

$$d(\text{cuervo}, \text{escritorio}) = d(A, B) = \frac{4}{9} \approx 0.44$$

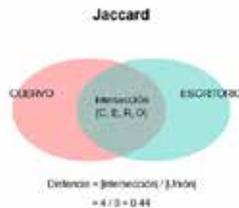


Figura 3. Similitud de Jaccard
Elaboración propia

Coseno

Ya que andamos metiendo números hagamos lo siguiente: asociemos cada letra con su lugar en el abecedario, por ejemplo “a” estaría asociada con el 1, “b” con el 2 y así sucesivamente. Esto nos permite convertir nuestras palabras en vectores, como se ve en la tabla 1.

Palabra	Vector	Vector en 10 lugares
Cuervo	(3, 21, 5, 18, 22, 15)	$x = (3, 21, 5, 18, 22, 15, 0, 0, 0, 0)$
Escritorio	(5, 19, 3, 18, 9, 20, 15, 18, 9, 15)	$y = (5, 19, 3, 18, 9, 20, 15, 18, 9, 15)$

Tabla 1. Las palabras convertidas a vectores, agregamos ceros para igualar longitudes.

Ahora podemos usar la similitud entre vectores del coseno, la cual se define para dos vectores x,y por

$$d(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

Esta fórmula nos mide el coseno del ángulo entre los vectores x,y, (cosas de álgebra lineal, ya sabes).

Este valor está entre -1 y 1, pero en nuestro caso particular es

$$\frac{(3 \cdot 5) + (21 \cdot 19) + (5 \cdot 3) + (18 \cdot 18) + (22 \cdot 9) + (15 \cdot 15) + (0 \cdot 20) + (0 \cdot 18) + (0 \cdot 9) + (0 \cdot 15)}{\sqrt{3^2 + 21^2 + 5^2 + 18^2 + 22^2 + 15^2} \sqrt{5^2 + 19^2 + 3^2 + 18^2 + 9^2 + 20^2 + 15^2 + 18^2 + 9^2 + 15^2}}$$

lo cual da como resultado 0.7106. La figura 4 ilustra esta idea.



Figura 4. Similitud del Coseno
Elaboración propia

El número obtenido es el coseno del ángulo entre los dos vectores, si es cercano a cero los vectores son perpendiculares. Un valor cercano a cero indicaría que los vectores son completamente distintos en dirección, mientras que un valor cercano a uno indica que están casi alineados, o

que tienen alta similitud.

Levenshtein

Esta medida es un poco más complicada. Se le conoce como medida de la edición pues cuenta las operaciones, entre agregados, eliminaciones o cambios, para pasar de una palabra a otra (Rodríguez, 2020). Para eso generamos una matriz con la palabra "cuervo" en las filas y "escritorio" en las columnas. En la imagen figura 5 se muestra el proceso que se sigue a continuación.

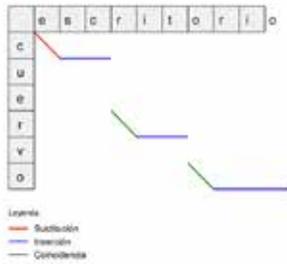


Figura 5. Distancia de Levenshtein
Elaboración propia

La distancia L de Levenshtein es 8, que representa el número mínimo de operaciones (sustituciones, inserciones o eliminaciones) necesarias para transformar "cuervo" en "escritorio".

Por Cáceres (año) sabemos que $0 \leq L \leq |A| + |B|$. En este caso tenemos que $|A|=6$ y $|B|=10$ y con esto podemos valorar la similitud como $8/16=0.5$

Nota que, a diferencia de Hamming, aquí las palabras pueden ser de distinta longitud.

En resumidas cuentas

Determinemos la bondad de los números encontrados con esta escala con la tabla 2.

Valor	Similitud
$v=1$	Perfecta
$0.8 \leq v < 1$	Muy alta

Valor	Similitud
$0.6 \leq v < 0.8$	Alta
$0.4 \leq v < 0.6$	Moderada
$0.2 \leq v < 0.4$	Baja
$0 \leq v < 0.2$	Muy baja
$v=0$	Nula

Tabla 2. Escala de valoración de las medidas de similitud

Ahora recordemos lo que obtuvimos:

Métrica	Valor obtenido	Similitud
Hamming	0.11	Muy baja
Jaccard	0.44	Moderada
Coseno	0.71	Alta
Levenshtein	0.5	Moderada

Tabla 3. Resumen de medidas

Fíjate como, a través de diferentes métricas, observamos distintos grados de similitud entre las palabras, lo que implica que la elección de la métrica afecta nuestra percepción de "similitud".

Así que a la pregunta ¿qué tanto se parece la palabra "cuervo" a "escritorio"? podemos responder un poco como los biólogos y decir "depende de la métrica", ya que la similitud va de moderada a alta. Podemos deducir, tal como lo haríamos con las respuestas literarias, que son palabras que tienen algo en común.

¿Y todo para qué?

Es posible que, al resolver el acertijo, una respuesta matemática parezca menos divertida que una respuesta ingeniosa. Sin embargo, métricas como Hamming, Jaccard, Coseno y Levenshtein tienen un papel fundamental en muchas aplicaciones actuales: cuando interactuamos con nuestros teléfonos, utilizamos servicios de recomendación en redes sociales o buscamos información en internet, estos algoritmos comparan textos, analizan patrones y encuentran similitudes que personalizan nuestra experiencia. Y esto no termina aquí.

Cada distancia que calculamos en este artículo refleja, en alguna medida, las formas en que las máquinas comprenden palabras, significados y hasta nuestras intenciones. Por ejemplo, en la frase “El cuervo es el guardián silencioso de los secretos que yacen en el escritorio”, la distancia semántica entre “cuervo” y “escritorio” es mínima porque están relacionados en un contexto específico. Matemáticamente, se podría calcular esta cercanía usando técnicas de modelado

de lenguaje como word embeddings, que asignan a cada palabra un vector numérico basado en su contexto, permitiendo medir su proximidad en términos de significado. La distancia semántica tiene amplias aplicaciones: desde mejorar motores de búsqueda (Jalal et al., 2022) y asistentes virtuales (Ramos Alvarez et al., 2024), hasta facilitar la traducción automática y el análisis de sentimientos (Nuñez-Prado et al., 2023). Cada una de estas tecnologías utiliza el contexto para acercar significados y ofrecer resultados más relevantes (Larraz et al., 2023).

Aunque “cuervo” y “escritorio” pueden parecer distintos a primera vista, para la inteligencia artificial es esencial identificar las relaciones entre conceptos para hacer asociaciones más humanas.

Al final, como en el acertijo de Carroll, lo importante es que siempre podemos hallar algo en común si miramos desde el ángulo adecuado, y las matemáticas nos dan la perspectiva para hacerlo.

Declaración de privacidad

Los datos de este artículo, así como los detalles técnicos para la realización del experimento, se pueden compartir a solicitud directa con el autor de correspondencia.

Los datos personales facilitados por los autores a RD-ICUAP se usarán exclusivamente para los fines declarados por la misma, no estando disponibles para ningún otro propósito ni proporcionados a terceros.

Conflicto de interés

Los autores de este manuscrito declaran no tener ningún tipo de conflicto de interés.

Agradecimientos

Referencias

- Carrol, L. (2003). Alicia en el país de las maravillas. Ediciones del sur. <https://www.ucm.es/data/cont/docs/119-2014-02-19-Carroll.AliciaEnElPaisDeLasMaravillas.pdf>
- Gamero, A. (2014, abril 4). ¿En qué se parece un cuervo a un escritorio? [Blog]. La piedra se Sísifo. Gabinete de curiosidades. <https://lapiedradesisifo.com/2014/04/06/en-qu%C3%A9-se-parece-un-cuervo-a-un-escritorio/>
- Igualada, J. (2009, enero 31). Solución al acertijo [Blog]. Mil matices de gris. <https://milmaticesdegris.blogspot.com/2009/01/solucion-al-acertijo.html>
- Jalal, A. A., Jasim, A. A., & Mahawish, A. A. (2022). A web content mining application for detecting relevant pages using Jaccard similarity. *International Journal of Electrical and Computer Engineering (IJECE)*, 12(6), 6461. <https://doi.org/10.11591/ijece.v12i6.pp6461-6471>
- Kannan, S., Karuppusamy, S., Nedunchezian, A., Venkateshan, P., Wang, P., Bojja, N., & Kejariwal, A. (2016). Big Data Analytics for Social Media. En *Big Data* (pp. 63–94). Elsevier. <https://doi.org/10.1016/B978-0-12-805394-2.00003-9>
- Larraz, I., Míguez, R., & Sallicati, F. (2023). Semantic similarity models for automated fact-checking: ClaimCheck as a claim matching tool. *El Profesional de la información*, e320321. <https://doi.org/10.3145/epi.2023.may.21>
- Nuñez-Prado, C.-J., Talavera Ortega, C., Chanona-Hernández, L., & Sidorov, G. (2023). Aplicación de algoritmos de aprendizaje automático sobre un corpus depresivo digital. *Research in Computing Science*, 152(8), 89–98. https://racs.cic.ipn.mx/2023_152_8/Aplicacion%20de%20algoritmos%20de%20aprendizaje%20automatico%20sobre%20un%20corpus%20depresivo%20digital.pdf
- Ramos Alvarez, J. I., Torres, R. V., Segarra Faggioni, V., & Pinza Tapia, E. I. (2024). Creación de distractores para preguntas de opción múltiple mediante técnicas de incrustación (Word Embedding). *Revista Tecnológica - ESPOL*, 36(2), 95–110. <https://doi.org/10.37815/rte.v36n2.1161>
- Rodríguez, D. (2020, junio 17). La distancia de Levenshtein [Analytics Lane]. *Ciencia de datos*. <https://www.analyticslane.com/2020/06/17/la-distancia-de-levenshtein>