

Codificación de datos basada en la fuente

Source-based data encoding

Julio César Bautista Hernández ^a, Dulce Rocío Hernández Jaimes ^b, Raúl H. Palacios ^c

Abstract:

Source encoding is a fundamental technique for data compression, which adapts to the specific context in which you are working. Allows you to choose between lossy or lossless compression, depending on the needs of the system or application. Lossy compression is used in composite multimedia files such as video, audio, and images, where the loss of certain information is generally imperceptible to the user. However, in medical applications or other contexts where data integrity is critical, lossless compression is preferred to ensure that information can be accurately decoded to the original state. For text, where data accuracy is crucial, lossless compression is always chosen. In short, source encoding is based on adapting to the context and uses different algorithms depending on the specific data compression needs.

Keywords:

encoding, decoding, data encoding, redundancy

Resumen:

La codificación fuente es una técnica fundamental para la compresión de datos, la cual se adapta al contexto específico en el que se está trabajando. Permite elegir entre compresión con pérdida o sin pérdida, según las necesidades del sistema o aplicación. La compresión con pérdida se utiliza comúnmente en archivos multimedia como video, audio e imágenes, donde la pérdida de cierta información es generalmente imperceptible para el usuario. Sin embargo, en aplicaciones médicas u otros contextos donde la integridad de los datos es crítica, se prefiere la compresión sin pérdida para asegurar que la información pueda ser decodificada con precisión al estado original. En el caso del texto, donde la exactitud de los datos es crucial, siempre se opta por la compresión sin pérdida. En resumen, la codificación fuente se basa en adaptarse al contexto y utiliza diferentes algoritmos según las necesidades específicas de compresión de datos.

Palabras Clave:

codificación, decodificación, codificación de datos, redundancia

Introducción

La teoría de la información, inaugurada por Claude Shannon en 1948 con su revolucionario Teorema de Codificación de Fuente, establece los cimientos de la compresión de datos sin pérdida. Este teorema postula que una fuente de información, caracterizada por su entropía o tasa de entropía (representada por H), puede ser codificada sin pérdida si la tasa de compresión de la fuente (representada por R) es igual o mayor que su entropía. En esencia, si la tasa de compresión supera la

entropía, es posible codificar la fuente sin error alguno. Sin embargo, si la tasa de compresión es inferior a la entropía, no existe una forma de codificar la fuente de manera que el error esté acotado, independientemente de la complejidad del codificador-decodificador utilizado [1]. Esta teoría es fundamental para comprender y desarrollar técnicas de compresión de datos eficientes. La codificación basada en la fuente, también conocida como *source coding* en inglés, aprovecha estos principios para reducir el tamaño de los datos sin perder información esencial. Al conocer el tipo de datos que se están

^a Autor de Correspondencia, Universidad Autónoma del Estado de Hidalgo, <https://orcid.org/0009-0006-9415-8192>, Email: ba440525@uaeh.edu.mx

^b Universidad Autónoma del Estado de Hidalgo, <https://orcid.org/0009-0004-0924-1684>, Email: he448271@uaeh.edu.mx

^c Universidad Autónoma del Estado de Hidalgo, <https://orcid.org/0000-0002-1131-4545>, Email: raul_palacios@uaeh.edu.mx

comprimiendo, se pueden aplicar algoritmos específicos adaptados a las características inherentes de la fuente de datos, ya sea imágenes, vídeos o audio [2].

El objetivo principal de la codificación basada en la fuente es eliminar redundancias o información innecesaria, lo que resulta en una representación más eficiente de los datos. Esta técnica abarca diversos métodos, desde la codificación por transformación, donde se aplican transformadas matemáticas para representar los datos de manera más compacta, hasta técnicas de codificación sin pérdidas como el Run-Length Encoding (RLE), que comprimen secuencias de datos con repeticiones consecutivas [1].

Además de la codificación por transformación y el RLE, otro algoritmo ampliamente utilizado en la compresión de datos sin pérdida es la codificación Huffman. Este algoritmo, desarrollado por David A. Huffman en 1952, se basa en los principios de la teoría de la información para asignar códigos binarios de longitud variable a cada símbolo de una fuente de datos, de manera que los símbolos más frecuentes tengan códigos más cortos y los menos frecuentes tengan códigos más largos. Al igual que otros métodos de codificación basados en la fuente, la codificación Huffman contribuye a reducir el tamaño de los datos eliminando redundancias y optimizando la representación de la información. Este algoritmo, junto con otros enfoques de codificación sin pérdida, constituye

una herramienta fundamental en la gestión eficiente de la información en diversas aplicaciones [1] [2].

La Figura 1 presenta una descripción de la codificación basada en la fuente, que aborda su origen, atribuido a Claude Shannon, y proporciona una serie de características. Estas incluyen la construcción de un modelo de probabilidad basado en la frecuencia de ocurrencia de los símbolos, el modelado de fuentes y la explicación de sus algoritmos de codificación fuente sin pérdida. Se destaca la posibilidad de diseñar algoritmos de codificación (y decodificación) sin pérdida, lo que implica que la cantidad de información transmitida puede ser igual o menor que el logaritmo de la cantidad de símbolos posibles (N), siempre y cuando la entropía (H) sea igual o menor que el logaritmo de N . Entre los algoritmos mencionados se encuentran la Codificación Shannon, la Codificación Huffman, la Modificación de Fuente, la Codificación Aritmética y la Codificación LZ.

Referencias

- [1] Martínez, José M. Transmisión de Datos Codificación de Fuente. Diss. Universidad Autónoma de Madrid, 2012: 47-5.
- [2] Oliver Gil, José Salvador. "Introducción a la compresión de información multimedia." (2018): 7-5.

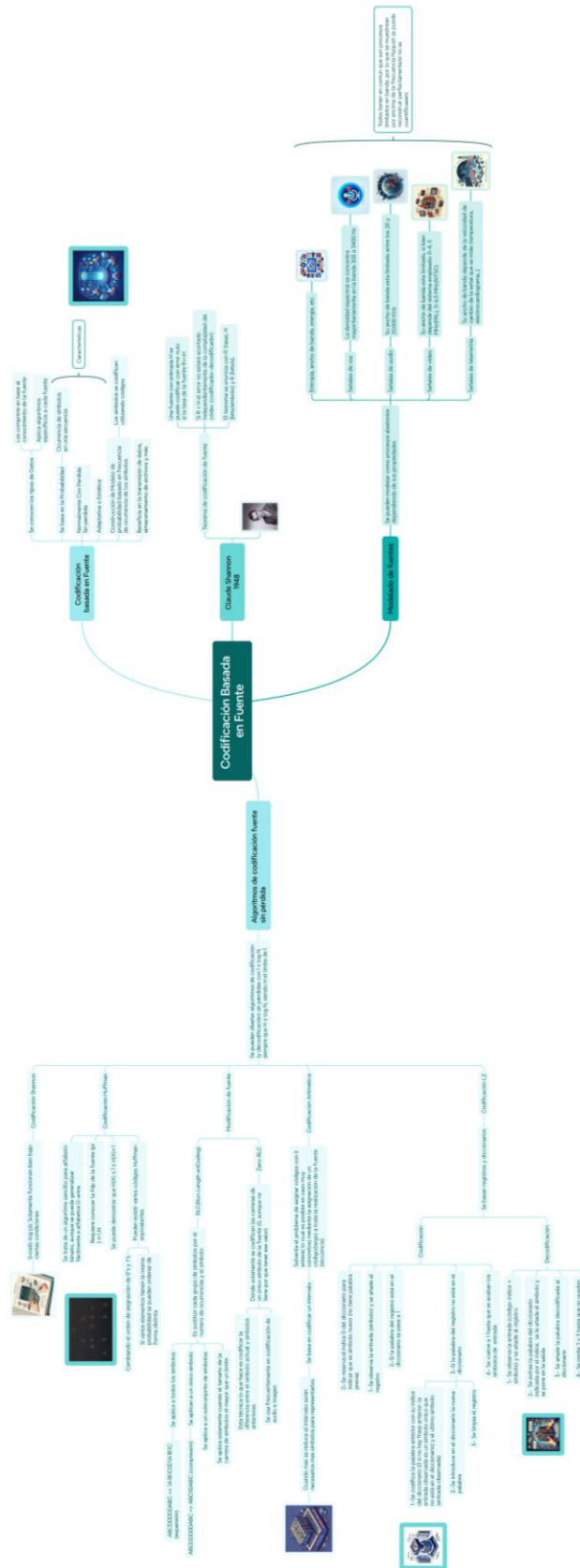


Figura 1. Codificación de datos basada en la fuente.