

Teoría de la Información

Extractado de "Introducción a la psicología de la comunicación" Alejandro López, Andrea Parada, Franco Simonetti, Ediciones Universidad Católica de Chile, Santiago, 1995.

A partir de la acelerada difusión y especialización que experimentan los medios de comunicación en el procesamiento y transmisión de información durante la primera mitad de nuestro siglo, se desarrolla el primer modelo científico del proceso de comunicación conocido como la Teoría de la Información o Teoría Matemática de la Comunicación. Específicamente, se desarrolla en el área de la telegrafía donde surge la necesidad de determinar, con la máxima precisión, la capacidad de los diferentes sistemas de comunicación para transmitir información.

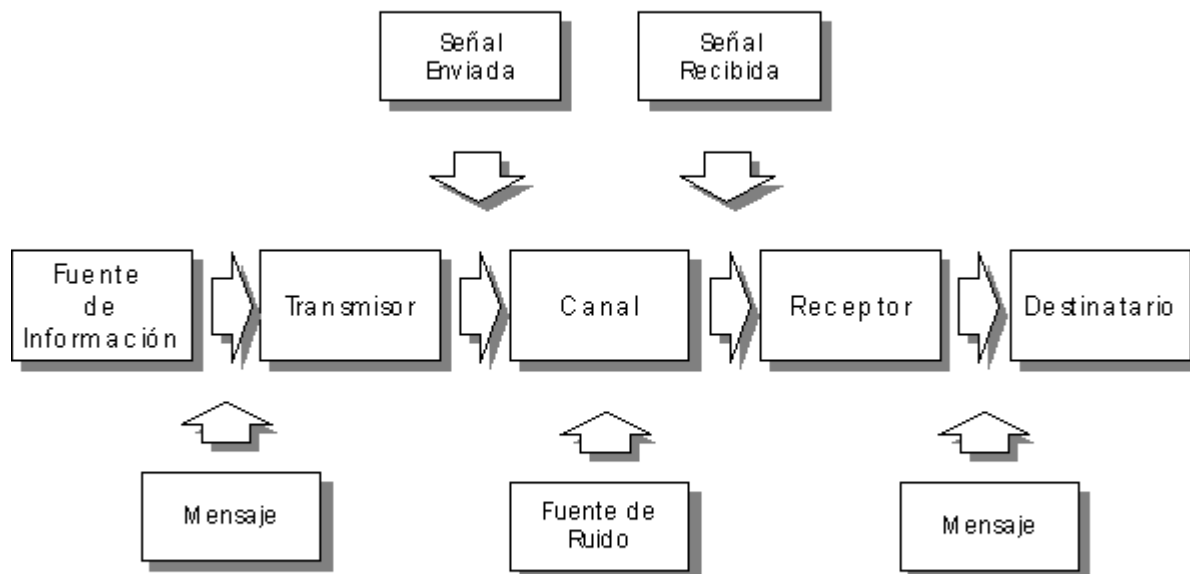
La primera formulación de las leyes matemáticas que gobiernan dicho sistema fue realizada por Hartley (1928) y sus ideas son consideradas actualmente como la génesis de la Teoría de la Información. Posteriormente, Shannon y Weaver (1949) desarrollaron los principios definitivos de esta teoría. Su trabajo se centró en algunos de los siguientes problemas que surgen en los sistemas destinados a manipular información: cómo hablar los mejores métodos para utilizar los diversos sistemas de comunicación; cómo establecer el mejor método para separar las señales del ruido y cómo determinar los límites posibles de un canal.

El concepto de comunicación en el contexto de la Teoría de la Información es empleado en un sentido muy amplio en el que "quedan incluidos todos los procedimientos mediante los cuales una mente puede influir en otra". De esta manera, se consideran todas las formas que el hombre utiliza para transmitir sus ideas: la palabra hablada, escrita o transmitida (teléfono, radio, telégrafo, etc.), los gestos, la música, las imágenes, los movimientos, etc.

En el proceso de comunicación es posible distinguir por lo menos tres niveles de análisis diferentes: el técnico, el semántico y el pragmático. En el nivel técnico se analizan aquellos problemas que surgen en torno a la fidelidad con que la información puede ser transmitida desde el emisor hasta el receptor. En el semántico se estudia todo aquello que se refiera al significado del mensaje y su interpretación. Por último, en el nivel pragmático se analizan los efectos conductuales de la comunicación, la influencia o efectividad del mensaje en tanto da lugar a una conducta. Es importante destacar que la Teoría de la Información se desarrolla como una respuesta a los problemas técnicos del proceso de comunicación, aun cuando sus principios puedan aplicarse en otros contextos.

Modelo de comunicación

El modelo comunicacional desarrollado por Shannon y Weaver se basa en un sistema de comunicación general que puede ser representado de la siguiente manera:



FUENTE DE INFORMACION: selecciona el mensaje deseado de un conjunto de mensajes posibles.

TRANSMISOR: transforma o codifica esta información en una forma apropiada al canal.

SEÑAL: mensaje codificado por el transmisor.

CANAL: medio a través del cual las señales son transmitidas al punto de recepción.

FUENTE DE RUIDO: conjunto de distorsiones o adiciones no deseadas por la fuente de información que afectan a la señal. Pueden consistir en distorsiones del sonido (radio, teléfono), distorsiones de la imagen (T.V.), errores de transmisión (telégrafo), etc.

RECEPTOR: decodifica o vuelve a transformar la señal transmitida en el mensaje original o en una aproximación de este haciéndolo llegar a su destino.

Este sistema de comunicación es lo suficientemente amplio como para incluir los diferentes contextos en que se da la comunicación (conversación, T.V., danza, etc.). Tomemos como ejemplo lo que ocurre en el caso de la radio. La fuente de información corresponde a la persona que habla por el micrófono. El mensaje son las palabras y sonidos que esta persona emite. El micrófono y el resto del equipo electrónico constituyen el transmisor que transforma este mensaje en ondas electromagnéticas, las cuales corresponden a la señal. El espacio que existe entre las antenas transmisoras y receptoras es el canal, mientras que lo que altera la señal original constituye la fuente de ruido. El aparato de radio de cada hogar es el receptor y el sonido que éste emite corresponde al mensaje recobrado. Las personas que escuchan este mensaje radial son los destinatarios.

También podemos ejemplificar esto mediante este artículo que usted está leyendo en este momento. En este caso, nuestros cerebros son la fuente de información y nuestros pensamientos, el mensaje. La máquina de escribir constituye el transmisor que transforma

nuestros pensamientos en lenguaje escrito, el cual corresponde a la señal. El papel es el canal y cualquier error de tipeo o puntuación, manchas, espacios en blanco, etc., constituyen la fuente de ruido. Por último, usted que está leyendo este ejemplo es a la vez el receptor y destinatario, que a través de la lectura recobra el mensaje por nosotros enviado.

Es importante considerar que el problema del significado del mensaje no es relevante en este contexto. El interés principal de la Teoría de la Información lo constituye todo aquello relacionado con la capacidad y fidelidad para transmitir información de los diferentes sistemas de comunicación. En el ejemplo anterior, el mensaje podría haber consistido en una secuencia de letras carentes de todo significado e igualmente el problema de cuánta información es transmitida estaría presente. En un sentido amplio, la Teoría de la Información trata acerca de la cantidad de información que es transmitida por la fuente al receptor al enviar un determinado mensaje, sin considerar el significado o propósito de dicho mensaje. No interesa tanto la pregunta: "¿Qué tipo de información?" sino más bien, "¿Cuánta información?" es la que transmite la fuente.

Información

Antes de analizar lo que se refiere a la capacidad y fidelidad de un canal determinado para transmitir información, es necesario que precisemos los alcances de este último concepto. El concepto de información es definido en términos estrictamente estadísticos, bajo el supuesto que puede ser tratado de manera semejante a como son tratadas las cantidades físicas como la masa y la energía. La palabra "información" no está relacionada con lo que decimos, sino más bien, con lo que podríamos decir. El concepto de información se relaciona con la libertad de elección que tenemos para seleccionar un mensaje determinado de un conjunto de posibles mensajes. Si nos encontramos en una situación en la que tenemos que elegir entre dos únicos mensajes posibles, se dice, de un modo arbitrario, que la información correspondiente a esta situación es la unidad. La Teoría de la Información, entonces, conceptualiza el término información como el grado de libertad de una fuente para elegir un mensaje de un conjunto de posibles mensajes.

El concepto de información supone la existencia de duda o incertidumbre. La incertidumbre implica que existen diferentes alternativas que deberán ser elegidas, seleccionadas o discriminadas. Las alternativas se refieren a cualquier conjunto de signos construidos para comunicarse, sean estas letras, palabras, números, ondas, etc. En este contexto, las señales contienen información en virtud de su potencial para hacer elecciones. Estas señales operan sobre las alternativas que conforman la incertidumbre del receptor y proporcionan el poder para seleccionar o discriminar entre algunas de estas alternativas.

Se asume que en los dos extremos del canal de comunicación -fuente y receptor- se maneja el mismo código o conjunto de signos. La función de la fuente de información será seleccionar sucesivamente aquellas señales que constituyen el mensaje y luego transmitir las al receptor mediante un determinado canal.

Existen diversos tipos de situaciones de elección. Las más sencillas son aquellas en que la fuente escoge entre un número de mensajes concretos. Por ejemplo, elegir una entre varias postales para enviarle a un amigo. Otras situaciones más complejas son aquellas en que la fuente realiza una serie de elecciones sucesivas de un conjunto de símbolos elementales tales como letras o palabras. En este caso, el mensaje estará constituido por la sucesión de símbolos elegidos. El ejemplo más típico aquí es el del lenguaje.

Al medir cuánta información proporciona la fuente al receptor al enviar un mensaje, se parte del supuesto que cada elección está asociada a cierta probabilidad, siendo algunos mensajes más probables que otros. Uno de los objetivos de esta teoría es determinar la cantidad de información que proporciona un mensaje, la cual puede ser calculada a partir de su probabilidad de ser enviada.

El tipo de elección más simple es el que existe entre dos posibilidades, en que cada una tiene una probabilidad de 1/2 (0,5). Por ejemplo, al tirar una moneda al aire ambas posibilidades -cara y sello- tienen la misma probabilidad de salir. El caso del lenguaje e idioma es diferente. En éstos la elección de los símbolos que formaran el mensaje dependerá de las elecciones anteriores. Por ejemplo, si en el idioma español el último símbolo elegido es "un", la probabilidad que la siguiente palabra sea un verbo es bastante menor que la probabilidad que sea un sustantivo o un adjetivo. Asimismo, la probabilidad que a continuación de las siguientes tres palabras "el esquema siguiente" aparezca el verbo "representa" es bastante mayor que la probabilidad que aparezca "pera". Incluso se ha comprobado que, en el caso del lenguaje, es posible seleccionar aleatoriamente letras que luego son ordenadas según sus probabilidades de ocurrencia y éstas tienden a originar palabras dotadas de sentido.

Principios de la medición de información

De acuerdo a estas consideraciones probabilísticas es posible establecer un primer principio de la medición de información. Este establece que mientras más probable sea un mensaje menos información proporcionará. Esto puede expresarse de la siguiente manera:

$$I(x_i) > I(x_k) \text{ si y sólo si } p(x_i) < p(x_k)$$

donde

$I(x_i)$: cantidad de información proporcionada por x_i

$p(x_i)$: probabilidad de x_i

De acuerdo a este principio, es la probabilidad que tiene un mensaje de ser enviado y no su contenido, lo que determina su valor informativo. El contenido sólo es importante en la medida que afecta la probabilidad. La cantidad de información que proporciona un mensaje varía de un contexto a otro, porque la probabilidad de enviar un mensaje varía de un contexto a otro.

Un segundo principio que guarda relación con las elecciones sucesivas establece que si son seleccionados los mensajes X e Y, la cantidad de información proporcionada por ambos mensajes será igual a la cantidad de información proporcionada por X más la cantidad de información proporcionada por Y, dado que X ya ha sido seleccionada. Esto puede ser expresado así:

$$I(x_i \text{ e } y_j) = f p(x_i) + f p(y_j/x_i)$$

Donde:

$I(x_i \text{ e } y_j)$: cantidad de información proporcionada por los mensajes x_i e y_j

f : función

$p(x_i)$: probabilidad de x_i

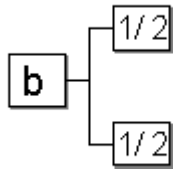
$p(y_j/x_i)$: probabilidad de y_j dado que x_i ha sido seleccionado.

Unidad de información

Una vez que hemos seleccionado el mensaje expresado en un lenguaje determinado es posible transcribirlo a un código de tipo binario. Este consta de sólo dos tipos de señales que indican Si o No, y que generalmente se codifican como 1 o 0. La cantidad de

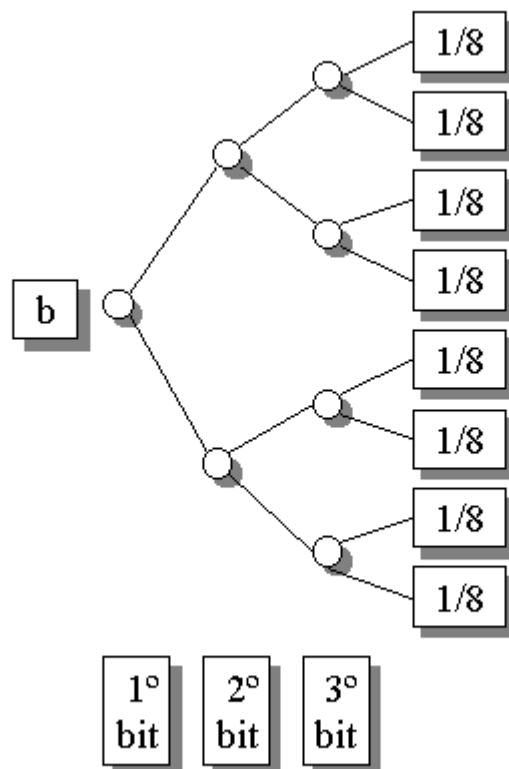
información proporcionada por cada elección entre dos alternativas posibles constituye la unidad básica de información, y se denomina dígito binario, o abreviadamente bit.

La elección existente al tener un bit de información puede ser esquematizada de la siguiente manera:



En la elección (b) tanto la línea superior como la inferior, es decir ambas posibilidades, pueden ser elegidas con la misma probabilidad de $r/2$.

Si existen N posibilidades, todas igualmente probables, la cantidad de información será igual a Log_2N . Es, entonces, el Log_2N la función matemática que nos indicará la cantidad de bits de información de una situación determinada. Esto puede esquematizarse de la siguiente manera:



La figura nos muestra una situación con 8 posibilidades, cada una con una misma probabilidad de $1/8$. Para poder determinar una posibilidad específica de estas 8, la elección requiere como mínimo 3 etapas, cada una de las cuales arroja un bit de información. El primer bit corresponde a la elección entre las primeras cuatro o segundas cuatro posibilidades. El segundo bit corresponde al primer o segundo par de las 4 posibilidades ya elegidas. El último bit determina el primer o segundo miembro del par y especifica la

posibilidad elegida. Como vemos, el primero de bits que se requieren en esta situación para determinar una posibilidad específica es de 3, lo que corresponde al $\text{Log}_2 8$.

Veamos ahora algunos ejemplos de lo recién expuesto:

	Elecciones		
Signo	1º	2º	3º
A	1	1	1
B	1	1	0
C	1	0	1
D	1	0	0
E	0	1	1
F	0	1	0
G	0	0	1
H	0	0	0

Esta figura nos muestra un alfabeto compuesto por sólo 8 signos. Pensemos que una fuente de información selecciona un signo y de alguna manera se lo señala al receptor. La pregunta sería entonces, ¿cuánta Información deberá conocer el receptor para identificar correctamente el signo escogido?

Asumamos que a partir de elecciones anteriores sabemos que cada uno de los 8 signos tiene la misma probabilidad de ser seleccionado. La incertidumbre, entonces, se ha repartido uniformemente sobre nuestro "alfabeto", o lo que es lo mismo, las probabilidades *a priori* de los signos son iguales; en este caso $1/8$.

Las señales que llegan al receptor representan instrucciones para seleccionar alternativas. La primera instrucción responde a la pregunta ¿está en la primera mitad del alfabeto, si o no? (en la figura, si = 1 y no = 0). La respuesta nos proporciona un bit de información y reduce el rango de incertidumbre exactamente a la mitad. Luego, una segunda instrucción divide cada mitad nuevamente en la mitad y, una tercera instrucción, otra vez en la mitad. En este caso, bastan tres simples instrucciones Si-No (1-0) para identificar un signo cualquiera de un total de ocho. La letra F, por ejemplo, podría ser identificada de la siguiente manera: 010. La respuesta a nuestra pregunta es entonces, ¡el receptor deberá obtener tres bits de información para identificar correctamente el signo escogido!

El típico juego de las "Veinte Preguntas" ilustra también algunas de las ideas mencionadas. Este juego consiste en que una persona piensa en un objeto mientras el resto de los jugadores intenta adivinar de que objeto se trata, haciendo no más de veinte preguntas que sólo pueden ser respondidas Si o No. De acuerdo a la Teoría de la Información, cada pregunta y su respuesta pueden proporcionar desde ninguna información hasta un bit de información ($\text{Log}_2 2$), dependiendo de si las probabilidades de obtener resultados Si o No son muy desiguales o casi iguales, respectivamente. Para obtener la mayor cantidad de información posible los jugadores deberán hacer preguntas que dividan el conjunto de posibles objetos en dos grupos igualmente probables. Por ejemplo, si mediante preguntas previas se ha establecido que se trata de una ciudad de Chile, una buena pregunta sería "¿Está al sur del río Maipo?". Así se dividen las ciudades posibles en dos grupos aproximadamente iguales. La segunda pregunta podría ser "¿Está al sur del río Bio-Bio?". Y

así sucesivamente hasta determinar de que ciudad se trata. Si fuera posible hacer preguntas que tuvieran la propiedad de subdividir las posibilidades existentes en dos grupos relativamente iguales, sería posible identificar mediante veinte preguntas un objeto entre aproximadamente un millón de posibilidades. Esta cifra corresponde a los 20 bits que se requieren para identificarla ($\log_2 1.000.000$).

Redundancia

No obstante lo anterior, la mayoría de las fuentes de información producen mensajes que no consisten en una única elección entre posibilidades de igual probabilidad, sino en elecciones sucesivas entre posibilidades de probabilidad variable y dependiente. A este tipo de secuencias se les denomina procesos estocásticos. Como ya lo mencionamos, el caso más típico son las letras y palabras que conforman el lenguaje. El escribir en español constituye un proceso de elecciones dependientes. Por ejemplo, al formar una palabra se elige una primera letra de todas las posibles primeras letras con diferentes probabilidades; luego, se elige la segunda letra cuya probabilidad depende de la primera letra seleccionada, y así sucesivamente hasta formar la palabra deseada. Lo mismo ocurre en el caso de las palabras para formar oraciones.

Lo importante aquí es señalar el hecho de que, en la medida que se avanza en la formación de una palabra u oración, el rango de posibles letras o palabras a ser seleccionadas va disminuyendo y la probabilidad de que ciertas letras o palabras específicas sean seleccionadas va aumentando. Dicho de otra forma, tanto la incertidumbre como la información de las últimas letras de una palabra o de las últimas palabras de una oración es menor comparada con las primeras.

La mayoría de los mensajes se constituyen a partir de un número limitado de posibilidades, por ejemplo, sólo 29 letras en el caso de nuestro idioma. Como vimos, la probabilidad de ocurrencia de una de estas posibilidades dentro de un mensaje depende de las posibilidades seleccionadas previamente; por ejemplo, la probabilidad de que ocurra la letra "q" luego de una "p" es 0. Son estos dos hechos los que en conjunto determinan que todo mensaje contenga cierto grado de redundancia. En otras palabras, la redundancia se refiere a que las posibilidades dentro de un mensaje se repiten, y se repiten de una cierta manera predecible. Mientras mayor sea, entonces, la redundancia de un mensaje, menor será su incertidumbre y menor la información que contenga.

El inglés escrito es un tipo de fuente de información que ha sido ampliamente estudiado. Se ha llegado a determinar que la redundancia de la lengua inglesa está muy próxima al 50%. Es decir, al escribir inglés aproximadamente la mitad de las letras y palabras que se emplean dependen de la libre elección de quien escribe, mientras que la otra mitad está determinada por la estructura probabilística del idioma.

La redundancia de los idiomas permite que si se pierde una fracción de un mensaje sea posible completarlo en forma muy aproximada al original. Este hecho se puede observar al eliminar varias letras de una oración sin que ello impida al lector completar las omisiones y rehacer la oración. Por ejemplo, en la siguiente frase han sido omitidas las vocales:

CMPLT ST FRS

Otra función importante de la redundancia es que nos permite ahorrar tiempo en la decodificación de los mensajes. Generalmente, no leemos cada una de las letras y palabras que conforman un texto, sino que vamos adivinando lo que viene. En el caso del telégrafo, por ejemplo, podríamos ahorrar tiempo ideando un código poco redundante y transmitiendo el mensaje a través de un canal sin ruido. Sin embargo, cuando el canal utilizado tiene ruido es conveniente no emplear un proceso de codificación que elimine toda la redundancia, pues

la redundancia nos ayuda a combatir el ruido. Si se pierde parte del mensaje por el ruido que afecta al canal, la redundancia nos permite rehacer en forma aproximada el mensaje. Por el contrario, la fracción de un mensaje no redundante que se pierde por el ruido es imposible de ser recuperada. La redundancia de los mensajes nos permite, entonces, corregir con facilidad los errores u omisiones que hayan podido ocurrir durante la transmisión.

Capacidad del canal

Ahora que ya hemos precisado el concepto -de información y los conceptos relacionados con él (incertidumbre, bit, redundancia) podemos volver a plantearnos el problema inicial de definir la capacidad de un canal determinado para transmitir información. Dado un canal con una capacidad de C unidades por segundo que recibe señales de una fuente de información de H unidades por segundo, la pregunta es ¿cuánto es el máximo número de bits por segundo que puede ser transmitido a través de este canal? Por ejemplo, un teletipo consta de 32 símbolos posibles que supondremos son empleados con igual frecuencia. Cada símbolo representa entonces 5 bits ($\log_2 32$) de información. De esta forma, si en ausencia total de ruido podemos enviar N símbolos por segundo a través de este canal, entonces podremos enviar $5N$ bits de información por segundo a través de dicho canal.

Son estas dos cantidades, la tasa de transmisión H por la fuente de información y la capacidad C del canal, las que determinan la efectividad del sistema para transmitir información. Si $H > C$ será ciertamente imposible transmitir toda la información de la fuente, no habrá suficiente espacio disponible. Si $H \leq C$ será posible transmitir la información con eficiencia. La información, entonces, puede ser transmitida por el canal solamente si H no es mayor que C .

El teorema fundamental para un canal sin ruido que transmite símbolos discretos afirma que si se emplea un procedimiento adecuado de codificación para el transmisor es posible conseguir que el ritmo medio de transmisión de símbolos por el canal sea muy próximo a C/H . Por muy perfecto que sea el procedimiento de codificación, dicho ritmo nunca podrá ser mayor de C/H .

Sin embargo, el problema de calcular la capacidad del canal se complica por la presencia de ruido. La presencia de ruido durante la transmisión provocará que el mensaje recibido contenga ciertos errores que contribuirán a aumentar la incertidumbre. Recordemos que la información es una medida del grado de libertad de elección que poseemos al momento de seleccionar un mensaje. Cuanto mayor sea la libertad de elección, mayor será la falta de seguridad en el hecho de que el mensaje enviado sea uno determinado. La incertidumbre será mayor y mayor la cantidad de información posible. De esta forma, si el ruido aumenta la incertidumbre, aumentará la información. Esto parecería indicar que el ruido es beneficioso, puesto que cuando hay ruido, la señal recibida es seleccionada a partir de un mayor conjunto de señales que las deseadas por el emisor. Sin embargo, la incertidumbre originada por la libertad de elección del emisor es una incertidumbre deseable; la incertidumbre debida a errores por la influencia del ruido es una incertidumbre no deseable.

Para extraer la información útil de la señal recibida es necesario suprimir la ambigüedad introducida por el ruido. Para ello se recurre a un factor de corrección matemático que no entraremos a analizar. El teorema para la capacidad de un canal con ruido se define como el ritmo máximo a que la información útil (incertidumbre total menos la incertidumbre debida al ruido) puede ser transmitida a través del canal.