

# TEORIA DE LA INFORMACION

## ORIGEN

- Teoría de la Información
  - 1948 Shannon "Una teoría Matemática de la Comunicación"
  - 1929 L. Szilar: Información-Paradoja Física
  - 1928: Hartly: Método de comparación de los distintos métodos de transmisores de la información
- 1948: Wiener: Cibernetica
- 1953: McMillan- Fuente de Información , canal de transmisión
- 1956: Khintchine- Tratamiento completo T.I. para canales ergódicos.
- Resumen
  - 1953: Winograd: Estableció un lazo entre T.C. de Shannon y la teoría de autómatas

## Documento de Shannon

- Escrito por Shannon en 1948.
- En él se desarrolla buena parte de la actual teoría llamada de la información
- Concepto de información
- Medida de "cantidad de información".
- Tratamiento matemático de la información

## INFORMACION

- **Terminología**

- **Señal**
- manifestación física ( de orden electromagnética , onda sonora...) capaz de propagarse en un medio dado. Es la definición más amplia del concepto de señal.



# Terminología

- **Mensaje:** Señal que corresponde a una realización particular del conjunto de señales dadas
- **Fuente:** proceso por el cual, entre todos los mensajes posibles, es escogido de una manera imprevisible un mensaje particular, destinado a ser transmitido a un receptor (observador)



# Terminología

- **Observador:** Destinatario final del mensaje.
- **Canal:** Totalidad de los medios destinados a la transmisión de la señal.
- 



## Terminología

- **Modulación** Transformación de un mensaje en una señal, al efecto de facilitar y aumentar la eficacia de la transmisión y reducir los errores de la misma.
- **Demulación** Operación inversa de la modulación.

## Terminología

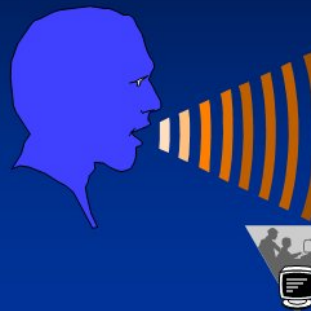
- **Codificación:** Transformación de un mensaje en una señal discreta, cuyo principal objetivo es aumentar la eficacia de la transmisión
- **Decodificación** Operación inversa de la codificación
- **Perturbación:** Señal que modifica una señal aleatoria útil, disminuyendo la cantidad de información que circula por el canal

## a) Teoría de la Información

- **Información:**
  - Conjunto de datos o mensajes inteligibles creados con un lenguaje de representación y que debemos proteger antes las amenazas del entorno, durante su transmisión o almacenamiento, con técnicas criptográficas.
- La **Teoría de la Información** mide la **cantidad de información** que contiene un mensaje a través del número medio de bits necesario para codificar todos los posibles mensajes con un **codificador óptimo**.

## CLASES DE INFORMACION

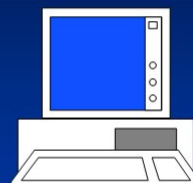
- **Voz:** Mecanismo primario para la comunicación humana. Es de naturaleza acústica.



- **Imágenes:**  
mecanismo primario  
para la  
comunicación  
humana. Es de  
naturaleza óptica.



- **Datos:**  
Información en  
forma numérica. Es  
de naturaleza  
electromagnética.





## DATOS-INFORMACIÓN Y CONOCIMIENTO

- **DATOS:** Secuencias de números, letras, etc. presentados sin un contexto
- **INFORMACIÓN.** Datos organizados, tablas, estadísticas de ventas, una charla (chat) bien presentada (Conjunto coherente de datos que transmite un mensaje)
- **CONOCIMIENTO.** Información organizada junto con la comprensión de lo que significa dentro de un contexto, que se puede utilizar

## INFORMACION-CONOCIMIENTO

Conocimiento: Información integrada en las estructuras

No podemos transmitir conocimiento, sólo información que el receptor puede o no convertirla en conocimiento

## GESTION DEL CONOCIMIENTO

- **Conocimiento** es la capacidad para transformar datos e información en acciones efectivas (eficaces).  
(J.D.Edwards)
- **Gestión del conocimiento** significa entregar a las personas los datos e información necesarias para ser eficientes (eficaces) en sus trabajos





### a) Teoría de la Información

- La **Teoría de la Información** mide la **cantidad de información** que contiene un mensaje a través del número medio de bits necesario para codificar todos los posibles mensajes con un **codificador óptimo**.



### Representación de la Información

Númérica, alfabética, simbólica, lenguaje.

24/01/03	24-01-03	24-1-03	24/01/2003
01/24/03	01-24-03	1-24-03	01-24-2003 ...

- Todos son el día 24 de enero del año 2003 -

Vitaminas: B<sub>12</sub>, C, ...

Grupo sanguíneo: A2 Rh+

Elementos: Fe, Si, Hg

Compuestos químicos: H<sub>2</sub>O, CO<sub>2</sub>

Más común — Lenguaje con código: "Hoy hace calor"

### En función de la extensión del mensaje

- Ante una pregunta cualquiera, una respuesta concreta y extensa nos entregará mayor información sobre el tema en particular, y diremos que estamos ante una mayor "cantidad de información".

Pregunta: ¿Hace calor allí? (*una playa en particular*)

- Respuesta 1: Sí, hace mucho calor.
- Respuesta 2: Cuando no sopla el viento, el calor allí es inaguantable pues supera los 42 grados a la sombra.

### ¿Dónde hay una mayor cantidad de información?

- Respuesta 2: Cuando no sopla el viento, el calor allí es inaguantable pues supera los 42 grados a la sombra. ☹

## ¿Qué es la información?

Veremos qué información nos entrega un mensaje dependiendo del contexto en que nos encontremos:

- En función de la extensión del mensaje recibido.
- En función de la utilidad del mensaje recibido.
- En función de la sorpresa del mensaje recibido.
- Dependiendo del entorno de esa sorpresa.
- En función de la probabilidad de recibir un mensaje.

## Cantidad de información (Caso 1)

### En función de la extensión del mensaje

- Ante una pregunta cualquiera, una respuesta concreta y extensa nos entregará mayor información sobre el tema en particular, y diremos que estamos ante una mayor “cantidad de información”.
- Pregunta: ¿Hace calor allí? (*una playa en particular*)
  - Respuesta 1: Sí, hace mucho calor.
  - Respuesta 2: Cuando no sopla el viento, el calor allí es inaguantable pues supera los 42 grados a la sombra.

– Respuesta 2: Cuando no sopla el viento, el calor allí es inaguantable pues supera los 42 grados a la sombra. @

## Cantidad de información (Caso 2)

### En función de la utilidad del mensaje

- Ante una pregunta cualquiera, una respuesta más útil y clara nos dejará con la sensación de haber recibido una mayor “cantidad de información”.
- Pregunta: ¿Hace calor allí? (*una playa en particular*)
  - Respuesta 1: Sí, bastante calor. ☝
  - Respuesta 2: Si no hay viento de poniente, es normal que la temperatura suba.

¿Dónde hay una mayor cantidad de información?

### Cantidad de información (Caso 3)

#### En función de la sorpresa del mensaje

- Ante una pregunta cualquiera, una respuesta más inesperada y sorprendente, nos dará la sensación de contener una mayor “cantidad de información”.

- Pregunta: ¿Hace calor allí? *(Finlandia en primavera)*

– Respuesta 1: Sí, muchísimo. Es insoportable. ☺

- Respuesta 2: En esta época del año, la temperatura es más suave y el tiempo muy agradable.

¿Dónde hay una mayor cantidad de información?

### Cantidad de información (Caso 4)

#### Dependencia del entorno (sorpresa)

- Ante una pregunta cualquiera, una respuesta más inesperada y sorprendente, nos dará la sensación de contener una mayor “cantidad de información”.

- Pregunta: ¿Hace calor allí?

*(ahora las mismas respuestas hablan de la temperatura en un horno)*

- Respuesta 2: En esta época del año, la temperatura es más suave y el tiempo muy agradable. ☺?

¿Dónde hay una mayor cantidad de información?



## Cantidad de información (Caso 5)

### En función de la probabilidad de recibir un mensaje

- Este enfoque probabilístico es el que nos interesará en cuanto a la definición de Cantidad de información.

¿Dónde le da alegría a su cuerpo Macarena?

- Respuesta 1: En un país de Europa.
- Respuesta 2: En una capital de provincia de España.
- Respuesta 3: En el número 7 de la calle de la Sierpes de Sevilla.

¿Dónde hay una mayor cantidad de información?

## Incertidumbre e información

Ante varios mensajes posibles, en principio todos equiprobables, aquel que tenga una menor probabilidad será el que contenga una mayor cantidad de información.

### • En el ejemplo anterior:

- Al ser más extenso el número de calles en una ciudad que el número de provincias en España y, esto último mayor que el número de países en Europa, el primero de ellos tendrá una **mayor incertidumbre**. Suponiendo todos los estados equiprobables, la **cantidad de información** será la mayor.

## INFORMACION(continuación)

- Algunas veces es conveniente expresar esta incertidumbre con relación a la que teníamos antes de conocer la información:

$$\bullet I_a/I_d$$

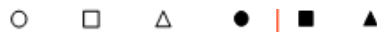
$$\bullet I_a/I_d$$

- Siendo  $I_a$  la incertidumbre antes de conocer el mensaje, e  $I_d$  la que tenemos después de dicho conocimiento.

## Grado de incertidumbre

$$c_i = \frac{\text{Grado de incertidumbre previo } I_a}{\text{Grado de incertidumbre posterior } I_d}$$

**Ejemplo :** En una bolsa hay un círculo, un cuadrado y un triángulo: negros o blancos.



¿Qué cantidad de información tiene cada uno de los estados?



Combinación nº 1	○	□	△
Combinación nº 2	○	□	▲
Combinación nº 3	○	■	△
Combinación nº 4	○	■	▲

Combinación nº 5	●	□	△
Combinación nº 6	●	□	▲
Combinación nº 7	●	■	△
Combinación nº 8	●	■	▲

Si hay equiprobabilidad  
entonces  $p(x_i) = 1/8$

Combinación nº 1	○	□	△
Combinación nº 2	○	□	▲
Combinación nº 3	○	■	△
Combinación nº 4	○	■	▲

Combinación nº 5	●	□	△
Combinación nº 6	●	□	▲
Combinación nº 7	●	■	△
Combinación nº 8	●	■	▲

Los 8 estados serán equiprobables:  $p(x_i) = 1/8$

- Las figuras no son del mismo color:  $I_a$  baja de 8 a 6 al descartarse las combinaciones 1 y 8.
- El círculo es blanco:  $I_a$  baja de 6 a 3 (descarte 5, 6 y 7).
- Hay dos figuras blancas:  $I_a$  baja de 3 a 2 (descarte 4).
- El cuadrado es negro:  $I_a$  baja de 2 a 1 (descarte 2.)

Se acaba la incertidumbre pues la solución es la combinación 3.

### Solución matemática

- Las figuras no son del mismo color.  $I_a$  baja de 8 a 6:  
 $c_{i1} = \log (8/6) = \log 8 - \log 6$
- El círculo es blanco.  $I_a$  baja de 6 a 3:  
 $c_{i2} = \log (6/3) = \log 6 - \log 3$
- Hay dos figuras blancas.  $I_a$  baja de 3 a 2:  
 $c_{i3} = \log (3/2) = \log 3 - \log 2$
- El cuadrado es negro.  $I_a$  baja de 2 a 1:  
 $c_{i4} = \log (2/1) = \log 2 - \log 1$

Todas las magnitudes se pueden sumar como

$$c_i = c_{i1} + c_{i2} + c_{i3} + c_{i4} = \log 8 - \log 1 = \log 8$$

### Base del logaritmo

Sean:  $I_a$  la incertidumbre inicial

$I_d$  la incertidumbre final

$$c_i = \log (I_a / I_d) = \log I_a - \log I_d$$

La cantidad de información tiene como unidad de medida la de un fenómeno de sólo dos estados, un fenómeno binario. Luego:

$$c_i = \log_b (2/1) = \log_b 2 - \log_b 1$$

– Si  $\log_b 2$  debe ser igual a 1 entonces la base  $b = 2$ .

– Precisamente a esta unidad se le llama **bit** (binary digit)

– Ejemplo anterior:  $c_i = \log_2 8 = 3$  ¡Sólo 3 preguntas!

## Con sólo tres preguntas...

Con sólo tres preguntas “*más o menos inteligentes*” podemos pasar de la incertidumbre total a la certeza:

**Pregunta 1:** ¿Está entre la opción 1 y la 4?  $\Rightarrow$  Sí

**Pregunta 2:** ¿Está entre la opción 1 y la 2?  $\Rightarrow$  No

**Pregunta 3:** ¿Es la opción 4?  $\Rightarrow$  No Se acaba la indeterminación

Combinación nº 1	○	□	△	Combinación nº 5	●	□	△
Combinación nº 2	○	□	▲	Combinación nº 6	●	□	▲
Combinación nº 3	○	■	△	Combinación nº 7	●	■	△
Combinación nº 4	○	■	▲	Combinación nº 8	●	■	▲

## INFORMACION(continuación)

- Ejemplos: -Cuando nos dicen que una persona es mujer, la incertidumbre antes era de 2 (número posible de estados), siendo la incertidumbre posterior 1 (ya sabemos que es mujer)
- Si el ordenador que genera letras al azar nos dice que ha salido una vocal, la incertidumbre antes del dato era 27 (número de letras), y ahora es 5 (número de vocales)

## INFORMACION(continuación)

- **Definición:** Sea un suceso A que puede presentarse con probabilidad  $p(A)$ , cuando dicho suceso tiene lugar se ha recibido una información  $I(A) = \log 1/p(A)$
- **Unidades**
  - Bit (base 2)
  - Dit (base 10)
  - Nit (base n)

Esto es cantidad  
de información

## continuación

- BIT =0.30 DIT =0.69 NIT
- DIT 0 3.32 BIT= 2.3 NIT
- NIT =1.44 BIT =0.43 DIT

## INFORMACION(continuación)

- La información más elemental que puede recibirse es la que indica la verificación entre dos sucesos igualmente probables. En este caso se dice que se ha recibido un bit de información.

## INFORMACION(continuación)

- Es muy importante distinguir entre bit como unidad de información y los símbolos 0 y 1 que representa las señales binarias. Estos símbolos se suelen llamar impropriamente bits, pero pueden contener o no 1 bit de información. Para distinguir, a los símbolos 0 y 1 se denominan binitos.



## INFORMACION(continuación)

- Si la fuente dispone de 10 símbolos igualmente probables, la emisión de uno de ellos proporciona una cantidad de información de un Hartley o Dit ( decimal digit ).
- Si se elige un símbolo entre  $e$  (base de logaritmos neperianos ) equiprobables, la información recibida será de 1 Nit.



