# Teoria da Informação - Parte 1

Teoria da Informação
Fonte: ft8.pdf (Capítulo 8 do livro Fundamentos das Telecomunicações)
Tags: #FCD #uni #SoftwareEngineering

---

## Introdução

Questão principal da engenharia de comunicações:

> Dada uma fonte de informação qualquer, como é que devem ser representadas as mensagens por ela emitidas, de modo a poderem ser transmitidas fiavelmente através de um canal de comunicação, dadas as inerentes limitações físicas deste?

Ao atacar esta questão Shannon (Matemático dos Laboratórios Bell) concentrou-se na **informação da mensagem em si,** e **não** nos **sinais utilizados para a transmitir**.
Esta abordagem deu origem ao que é hoje conhecido por *Teoria da Informação*, que estuda quatro problema fundamentais:

1. A **medida** da informação produzida por uma fonte;
2. A **codificação** da *fonte* destinada a representar a informação produzida pela fonte com o menor número possível de símbolos;
3. A **capacidade** do canal de comunicação para transmitir informação (ou seja, o máximo da quantidade de informação por unidade de tempo que é possível transmitir num canal);
4. A **codificação do canal** como mecanismo para melhor utilizar a capacidade do canal que normalmente se designa por codificação para controlo de erros.

*Canal* - sistema (ou *via*) de comunicação completo entre a fonte e o destino da informação.
*Codificação* - representação de mensagens quer por formas de onda contínuas quer por formas de onda discretas ou outros símbolos.

Os 4 problemas atrás referidos conjugam-se naquele que constitui o *teorema fundamental* da Teoria da Informação:

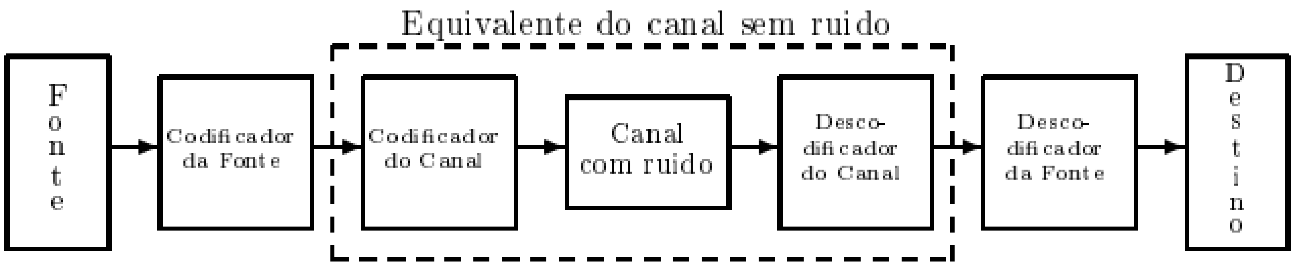
> Dado um canal de comunicação e uma fonte cujo débito de informação não escede a capacidade do canal, existe um código tal que a informação pode ser transmitida através do canal com uma frequência de erros arbitrariamente pequena, apesar da presença de ruido.

O aspeto surpreendente deste teorema é a promessa que faz de transmissão de informação *sem erros* através de um canal ruidoso, o que se torna possível através de operações de *codificação*.

O *codificador* e o *descodificador do canal* executam a tarefa de *controlo de erros*, detetando-os e ventualmente corrigindo-os (através de mecanismos que serão abordados posteriormente) e pelos quais os efeitos do ruído podem ser reduzidos ou eliminados.

O codificador da fonte representa a informação através de símbolos - normalmente binários - tentando fazê-lo com o menor número possível desses símbolos, operação esta que se designa *compressão da fonte*.

O descodificador da fonte executa a operação inversa retirando dos símbolos recebidos a informação neles contida entregando-a ao destino.
Pode dizer-se então que o par *codificador* / *descodificador da fonte* têm o papel de adaptar a fonte ao canal equivalente sem ruído.




(O canal é o bloco tracejado na figura acima)

## Medida de Informação

O ponto de partida da teoria da informação é a medida da *informação*.
(Termo técnico que não deve ser confundido com *dados*, nem com *conhecimento* ou *significado*, conceitos cuja definição e medida são ainda objeto de debate.

> ⚠ Warning
>
> No contexto das comunicações, informação não é mais do que o produto, o bem ou o objeto imaterial útil produzido por uma fonte que tem de ser transferido para um utilizador num destino. Se a informação está previamente disponível no destino a transferência será zero.

É senso comum que quanto menos provável for uma determinada mensagem maior a quantidade de informação nela contida. Pode pois concluir-se que a medida da informação deve relacionar-se com o grau de ***incerteza*** do destinatário quanto à mensagem que vai receber.
(Posto de outra forma, a informação mede a *liberdade de escolha* exercida pela fonte ao selecionar uma mensagem dentro do conjunto universo das possíveis mensagens.)

### Informação Própria

Torna-se evidente, portanto, que a medida da informação deve ser uma função da *probabilidade* de ocorrência da mensagem.

Seja $f()$ essa função e $x_i$ uma mensagem arbitrária tal que a probabilidade do acontecimento $x_i$ ser selecionado para transmissão é $P(x_i) = P_i$. A quantidade de informação, $I_i$, associada à ocorrência de $x_i$, será então $I_i = f(P_i)$.

A função $f()$ deve possuir as seguintes propriedades:

$$(i)\ \ f(P_i) \geq 0 \text{ para } 0 \leq P_i \leq 1$$
$$(ii)\ \ lim_{P_i \to 1}\ \ f(P_i) = 0$$
$$(iii)\ \ f(P_i) > f(P_j) \text{ para } P_i < P_j$$

**(i)** A informação nunca é negativa
**(ii)** A informação é nula se o acontecimento for certo
**(iii)** A informação aumenta com a incerteza

Existem muitas funções que possuem estas 3 propriedades. Contudo, considere-se o caso em que a fonte produz duas *mensagens sucessivas e independentes*, $x_i$ e $x_j$, com a probabilidade conjunta $P(x_i x_j) = P_{ij} = P_i P_j$ (acontecimentos independentes). Pretende-se que a informação total seja igual à soma da informação das mensagens individuais o que exige a seguinte propriedade adicional:

$$(iv)\ \ f(P_i P_j) = f(P_i) + f(P_j)$$

A única função que satisfaz estas quatro propriedades é a **função logarítmica negativa**, $-log_b()$. A base $b$ que se utilizar para o logaritmo define a *unidade* de medida da informação. A convenção adotada na teoria da informação é tomar $b = 2$ e designar a unidade correspondente por *bit*.

### Bit como unidade de medida de informação

> **O bit é a quantidade de informação necessária para escolher uma entre duas alternativas igualmente prováveis** ou, a quantidade de informação contida numa mensagem emitida por uma fonte capaz de emitir apenas duas mensagens distintas e equiprováveis.

Portanto, e por definição, a quantidade de informação, ou informação prórpia,

$$I_i \stackrel{def}{=} log_2 \frac{1}{P_i} \ \ bits$$

(Fica $\frac{1}{P_i}$ de modo a retirar o sinal negativo da função logarítmica)
Se a fonte possuir um alfabeto de apenas duas mensagens, $x_1$ e $x_2$, e se $P(x_1) = P(x_2) = \frac{1}{2}$, tem-se $I_1 = I_2 = log_2(2) = 1 \text{ bit}$.

> ⚠ Warning
>
> Deve ter-se cuidado em distinguir *bits* de informação dos *dígitos binários* utilizados para a representar ou codificar - especialmente porque um dígito binário pode transportar mais ou menos do que um bit de informação.
> De modo a evitar a confusão, o dígito binário é muitas vezes designado por *binit* e, no contexto da Teoria da Informação, o termo *bit* designa a unidade de medida de informação e não o dígito binário.
> Por simplicidade utiliza-se também o termo *símbolo* em alternativa ao termo *mensagem*.

É frequente ter de se converter logaritmos de uma base *b* qualquer, por exemplo a natural ou a decimal, para a base 2. É fácil verificar que:

$$log_2(\alpha) = \frac{log_b(\alpha)}{log_b(2)} = \frac{log_{10}(\alpha)}{log_{10}(2)}$$

### Entropia

Consideraremos agora uma fonte de informação que emite uma sequência de símbolos (mensagens) selecionados de um alfabeto de $m$ símbolos distintos. Seja $X = \{x_1, x_2, \dots, x_m\}$ esse conjunto alfabeto. Podemos tratar cada símbolo $x_i$ como uma mensagem que ocorre com probabilidade $P_i$ e transporta a auto-informação $I_i$. Os valores das probabilidades devem obviamente satisfazer a igualdade $\sum_{i=1}^{m} P_i = 1$.

Suponhamos que os sucessivos símbolos são estatísticamente *independentes* e são produzidos pela fonte a um débito médio de $r_s$ símbolos por segundo. Suponhamos ainda que a fonte é *estacionária*, o que significa que as probabilidades não variam com o tempo. Estas propriedades definem o que se designa por *fonte discreta sem memória*.

A quantidade de informação produzida pela fonte durante um intervalo de símbolo arbitrário é uma variável aleatória discreta que toma valores $I_1, I_2, \dots, I_m$. A informação média por símbolo é então dada pela média estatística,

$$H(x) \stackrel{def}{=} \sum_{i=1}^{m} P_i I_i = \sum_{i=1}^{m} P_i \ \ log_2 \frac{1}{P_i} \ \text{bits/símbolo}$$

grandeza que é chamada *entropia* da fonte. A equação acima pode ser interpretada com representando a *quantidade média de informação* obtida quando uma variável aleatória $X$ toma um valor.

O valor de $H(X)$ para uma dada fonte depende das probabilidades dos símbolos da fonte, $P_i$, e da cardinalidade, $m$, do alfabeto estando limitado por:

$$0 \leq H(X) \leq log_2(m)$$

O limite inferior corresponde à **inexistência de incerteza** o que acontece quando um dos símbolos tem probabilidade $P_j = 1$ e todos os restantes $P_i = 0 \ \ (i \neq j)$, ou seja, quando a fonte emite sempre o mesmo símbolo.
Portanto, a prova do limite inferior na relação acima obtem-se facilmente bastando notar que $\alpha \ \ log_2(\frac{1}{\alpha} \rightarrow 0)$ quando $\alpha \rightarrow 0$. O limite superior corresponde à máxima incerteza que ocorre quando $\forall_i : P_i = \frac{1}{m}$, isto é, quando **todos os símbolos são igualmente equiprováveis**.

### Débito de informação

A equação acima que representa a quantidade de informação média obtida também pode ser interpretada da seguinte maneira: quando uma fonte emite uma sequência de $n \geq 1$ símbolos, a informação total produzida é aproximadamente igual a $H(X)$ bits. Dado que a fonte produz $r_s$ símbolos por segundo, a duração desta sequência é de $\frac{n}{r_s}$ seg. A informação deve pois ser transferida a um débito médio $nH(X)/(nr_s) = r_s H(X)$ bits/seg. O *débito médio de informação* de uma fonte é então definido por

$$R \stackrel{def}{=} r_s H(X) \text{ bits/seg}$$

uma grandeza crítica em sistemas de transmissão. O teorema fundamental da teoria da informação diz que a informação produzida por qualquer fonte discreta sem memória pode ser codificada em dígitos binários e transmitida através de um canal sem ruído a um ritmo binário $r_b$, com $r_b \geq R$.