



INFORMATION RETRIEVAL

Período 2: Recuperação de Informação

Recuperação de Informação

- Recuperação de informação é o nome dado ao processo ou **método pelo qual um potencial usuário de informação é capaz de converter a sua necessidade de informação em uma lista real de citações a documentos** em um acervo contendo informações úteis para ele.
- Recuperação de informação abrange os aspectos intelectuais da **descrição da informação e sua especificação para a busca**, e também quaisquer **sistemas, técnicas ou máquinas que são utilizadas para realizar a operação**.
- O assunto de **cada documento** ou outra unidade de informação é **caracterizado ou descrito por meio de um conjunto de "descritores"** tirado de um vocabulário formal de tais termos. Uma "lista de cabeçalho de assuntos" remeterá a uma aproximação grosseira do seu significado.

(MOOERS, 1951, p.25)

Recuperação de Informação

- **O usuário** de um sistema de recuperação de informação **está interessado em recuperar “informação” sobre um determinado assunto e não em recuperar registros de dados que satisfazem sua expressão de busca**, nem tampouco documentos, embora seja nestes que a informação estará registrada.
- **Essa característica é o que diferencia os sistemas de recuperação de informação dos sistemas gerenciadores de bancos de dados**, estudados e implementados no âmbito da Ciência da Computação.

Recuperação de Informação

- Para Saracevic (1999), a Recuperação de Informação pode ser considerada a vertente tecnológica da Ciência da Informação e é resultado da relação desta com a Ciência da Computação.
- Na Ciência da Computação a Recuperação de Informação (*Information Retrieval*) se firmou como uma área de pesquisa autônoma cujo interesse está centrado no desenvolvimento de ferramentas para o tratamento de fontes de informação não estruturadas e semiestruturadas. É tema de interesse de uma imensa comunidade de pesquisadores de todas as partes do mundo e abriga uma grande quantidade de vertentes, abordagens e metodologias para os problemas dessa área.



[Biblioteca de Alexandria](#)

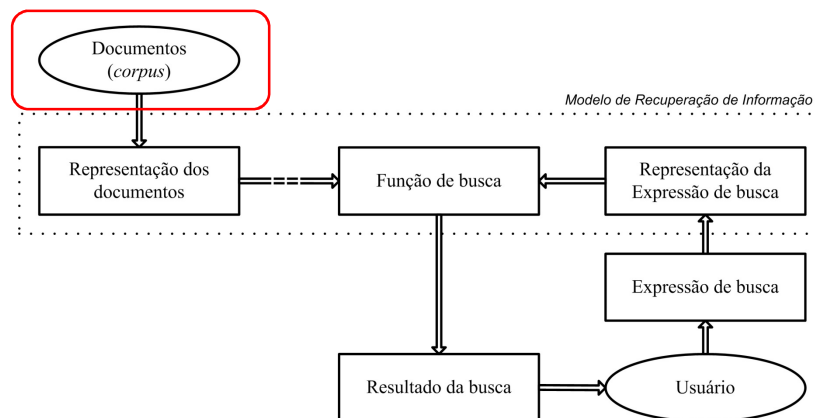
Video

[Eratóstenes](#)



**O processo de
recuperação de
informação**

O processo de recuperação de informação



Documento

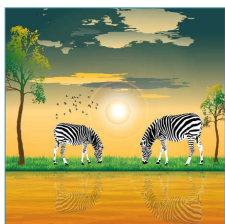
- Buckland (1991):
 - **o termo *informação* é utilizado na maioria das vezes vinculado a um objeto que contém informação: um documento.** Assim, o termo *informação* poderia também designar “algo atribuído a um objeto, tal como dado e documento que se referem à informação, porque deles se espera que sejam informativos”.
- Le Coadic (2004, p.5):
 - **documento é o termo genérico que designa os objetos portadores de informação.** Um documento é todo artefato que representa ou expressa um objeto, uma ideia ou uma informação por meio de signos gráficos e icônicos (palavras, imagens, diagramas, mapas, figuras, símbolos), sonoros e visuais (gravados em suporte de papel ou eletrônicos).

Documento

- Suzanne Briet (2006, p.10):
 - Uma estrela é um documento? Um seixo rolando na torrente de um rio é um documento? Um animal vivo [na natureza] é um documento? **Não.**
 - Mas as fotografias e os catálogos de estrelas, as pedras em um museu de mineralogia, e os animais que são catalogados e apresentados em um zoológico **são documentos.**

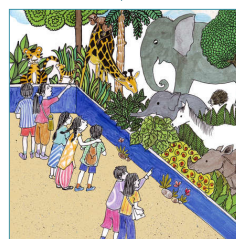
“qualquer signo físico ou simbólico, preservado ou registrado, com a intenção de representar, reconstruir ou demonstrar um fenômeno físico ou abstrato”

Documento



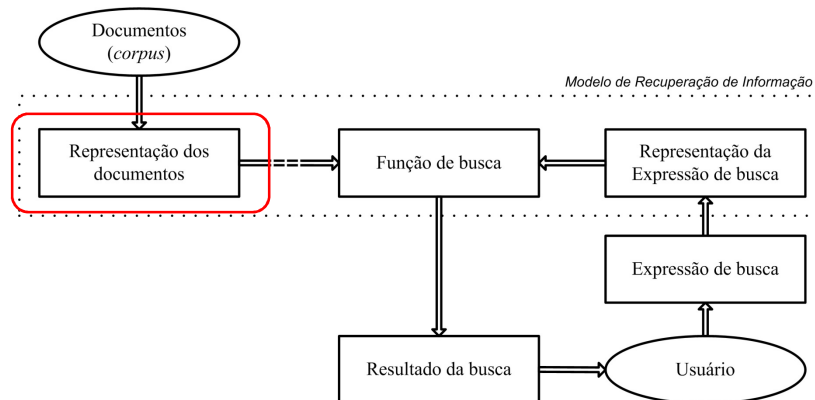
Animal na natureza
Não é documento

Suzanne Briet, 2006, p.10



Animal no zoológico
É documento

O processo de recuperação de informação



Representação dos documentos

- Ciência da Informação
 - Representação Descritiva
 - Representação Temática

Representação dos documentos

- Representação Descritiva
 - Representação descritiva (ou Catalogação) tem como objetivo propor parâmetros e normas para representar/descrever um item informacional.
 - Propiciar a elaboração de catálogos nos bancos de dados que representem o itens dos estoques informacionais pesquisados.
 - Para fins de recuperação da informação utilizam-se os metadados, ou seja, dados que se referem a descrição e identificação dos recursos. Os mesmos são usados para tornar a informação identificável, localizável e acessível.

Representação dos documentos

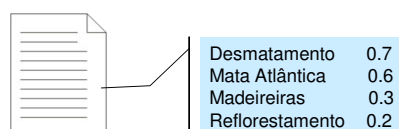
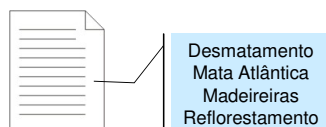
- Representação Descritiva (Catalogação)

<p>A How-To-Do-It Manual for Librarians</p> <p>PLANNING FOR AUTOMATION</p> <p>001 26132811 008 320608s1992 nyu bf 001 0 eng 010 \$a92021087 020 \$a1555701205 040 \$aDLC\$cDLC\$dAGL 049 \$aHNKA 050 00 \$a2678.9\$b.C58 1992 082 00 \$a025/.00285\$220 100 1 \$aCohn, John M 245 10 \$aPlanning for automation : \$ba how-to-do librarians / \$cJohn M. Cohn, Ann L. Kels Fields 260 \$aNew York : \$bNeal Schuman Publishers, \$c1992 300 \$aui, 116 p. : \$c28 cm 440 0 \$aHow-to-do it manuals for libraries : \$aIncludes bibliographical references 504 \$aLibrary planning \$xAutomation \$xManagement \$xHandbooks, manuals, etc. 650 0 \$aLibrary planning \$xAutomation \$xManagement \$xHandbooks, manuals, etc. 700 10 \$aKelsey, Ann L 700 10 \$aFields, Keith Michael</p>	<p>W443n Weinberg, David, A nova desordem Mussi Araujo. – Rio de Janeiro : Tradução de: E. V. digital disorder ISBN 978-85-351-1667-8</p>	<pre><?xml version="1.0" encoding="UTF-8" ?> <collection xmlns="http://www.loc.gov/MARC21/slim"> <record> <leader>01023nam #2200229 a 4500</leader> <controlfield tag="001">bica0000047874</controlfield> <controlfield tag="005">20091011</controlfield> <controlfield tag="007">aj canz</controlfield> <controlfield tag="008">120514s2004 sp ger</controlfield> <datafield tag="010" ind1="1" ind2="1"> <subfield code="a">W 16678-2004</subfield> </datafield> <datafield tag="034" ind1="1" ind2="1"> <subfield code="b">000025000</subfield> </datafield> <datafield tag="040" ind1="1" ind2="1"> <subfield code="a">SpMaBN</subfield> </datafield> <datafield tag="080" ind1="1" ind2="1"> <subfield code="a">(460.41 Caldera de Taburiente, parque nacional)</subfield> </datafield> <datafield tag="130" ind1="0" ind2="1"> <subfield code="k">Mapas topográfico-excursionistas</subfield> <subfield code="a">Caldera de Taburiente (Parque nacional)</subfield> </datafield> <datafield tag="245" ind1="1" ind2="0"> <subfield code="a">Caldera de Taburiente</subfield> <subfield code="b">parque nacional : landkarte</subfield> </datafield> <datafield tag="246" ind1="0" ind2="1"> <subfield code="a"></subfield> </datafield> <datafield tag="255" ind1="1" ind2="1"> <subfield code="a">Escala 1:25.000</subfield> <subfield code="b">proyек. U.T.M., ellipsoid GRS80 (WGS84). Datum REGCAN9</subfield> <subfield code="c">0004-01-01</subfield> </datafield> <datafield tag="260" ind1="1" ind2="1"> <subfield code="b">Centro Nacional de Información Geográfica Nacionales Parques</subfield> <subfield code="a">[Madrid]</subfield> </datafield> </record> </collection></pre>
--	---	---

Representação dos documentos

- Representação Temática
 - Análise de assunto de um documento e a colocação do resultado desta análise numa expressão linguística;
 - **Indexação**
 - Por extração
 - Por atribuição
 - Necessidade de controle do vocabulário
 - Vocabulários controlados

Representação dos documentos



Representação dos documentos

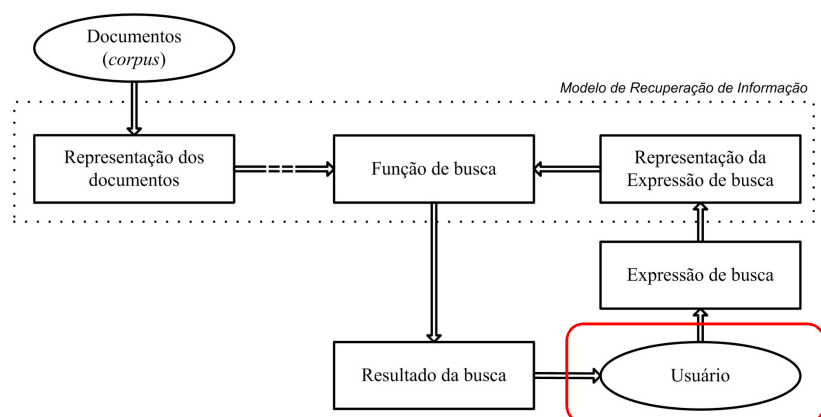


?



?

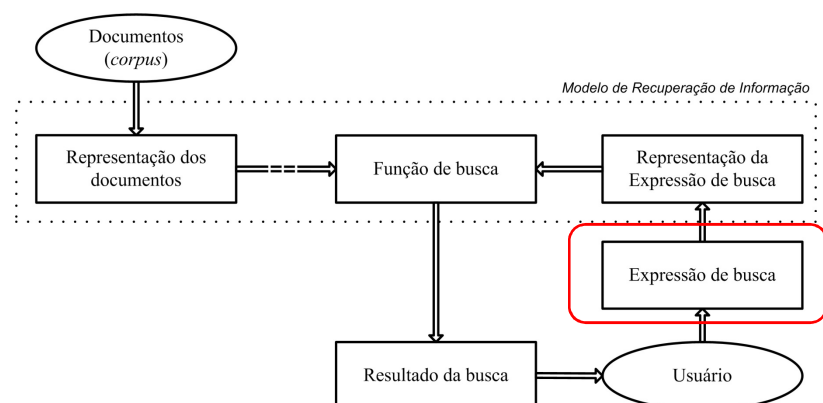
O processo de recuperação de informação



Usuário

- Segundo Riecken (2006):
 - **O usuário e os conteúdos estão no centro das preocupações da CI** [Ciência da Informação] e não as tecnologias. Entretanto, a CI pode valer-se de práticas de modelagem e notação oriundas da engenharia de software e da computação visando a facilitar a operacionalização em sistemas computacionais, suas bases de dados e redes, a partir dos modelos e representações conceituais.

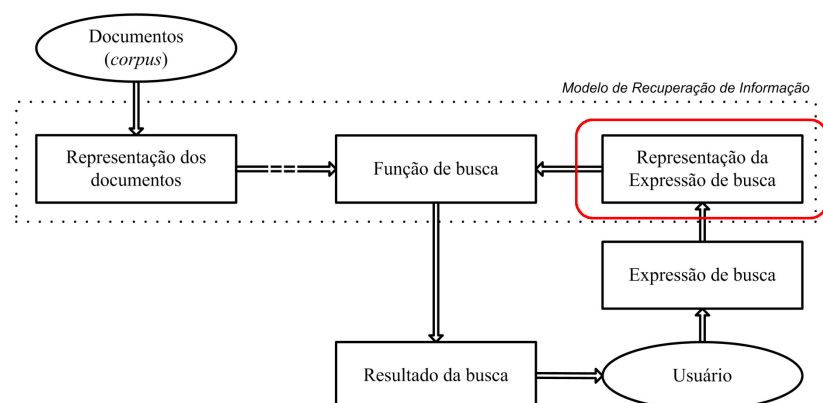
O processo de recuperação de informação



Expressão de Busca

- **Meio que o usuário emprega para comunicar a sua necessidade informacional para o sistema.**
- Podem ser especificada em linguagem natural ou por meio de uma linguagem artificial, dependendo dos recursos oferecidos pelo sistema.
- **Os usuários necessitam ter um mínimo de conhecimento do tema de interesse e do seu vocabulário do domínio.**
- Principal dificuldade: prever quais os termos que foram usadas para representar os documentos que satisfarão sua necessidade.

O processo de recuperação de informação

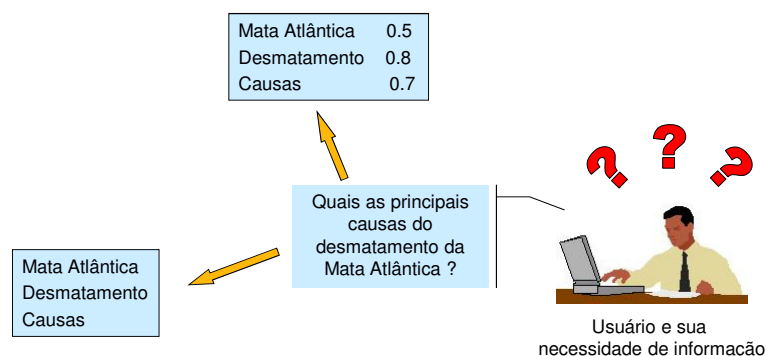


Representação da Expressão de Busca

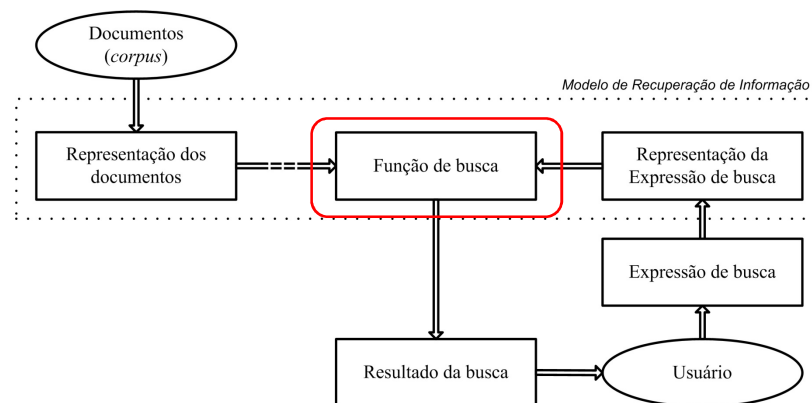
- **É necessário que a expressão de busca seja representada de forma similar à utilizada na representação dos documentos.**
- Essa homogeneidade permitirá a comparação entre a busca e todos os documentos do *corpus* do sistema por meio da função de busca.

[Musipedia](#)

Representação da Expressão de Busca



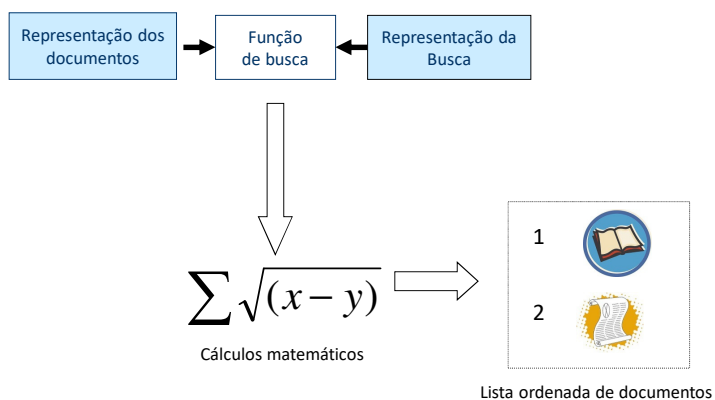
O processo de recuperação de informação



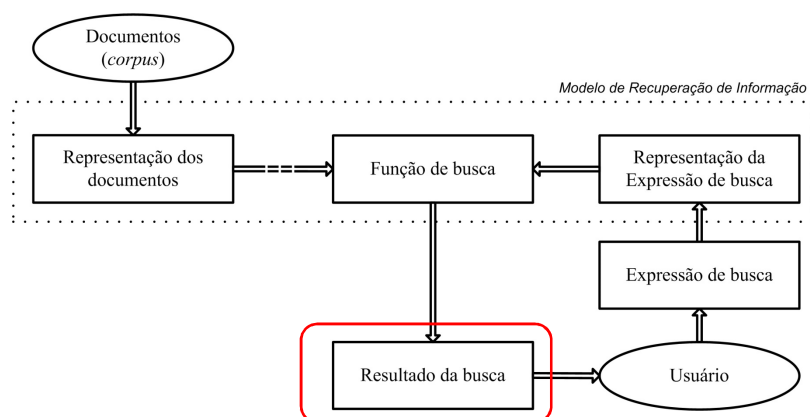
Função de Busca

- **compara as representações dos documentos com a representação da expressão de busca** e recupera os itens que supostamente fornecerão a informação que o usuário procura.
- calcula o **grau de similaridade** entre a expressão de busca e cada um dos documentos do *corpus*.
- O grau de similaridade, na maioria das vezes definido por um valor numérico, pretensamente define o quão relevante é um determinado documento para satisfazer a necessidade de informação do usuário;
- O grau de similaridade é utilizado para ordenar (**ranquear**) os documentos resultantes da busca.

Função de Busca



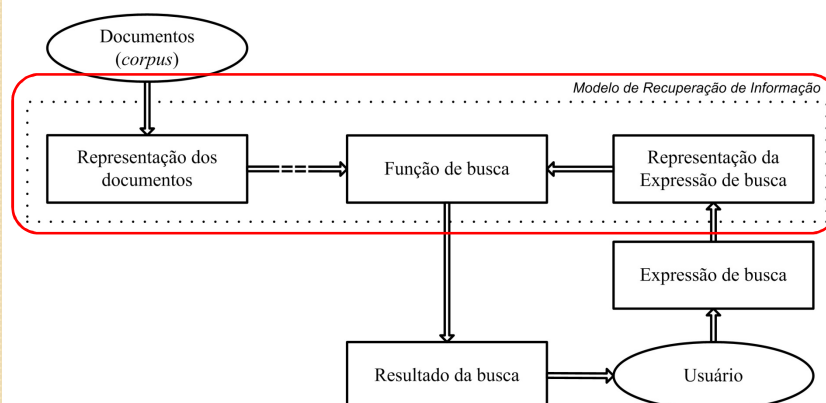
O processo de recuperação de informação



Resultado da Busca

- Geralmente é composto por um **conjunto de documentos que supostamente serão úteis (relevantes) para o usuário.**
- É apresentado na forma de uma **lista ordenada** pelo grau de similaridade (relevância) calculada pela *função de busca*.
- O fato de um termo utilizado na expressão de busca aparecer na representação de um documento não significa que este documento seja relevante para a necessidade do usuário.

O processo de recuperação de informação



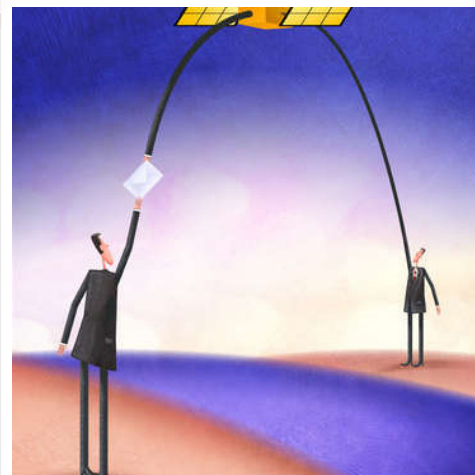
Modelo de Recuperação de Informação

- É a especificação formal de três elementos:
 - a **representação dos documentos**;
 - a representação da necessidade de informação por meio de uma **expressão de busca**;
 - como estes dois elementos serão comparados: a **função de busca**.
- A eficiência de um sistema de recuperação de informação está diretamente ligada ao modelo que ele utiliza, influenciando diretamente em seu modo de operação.



Recuperação de Informação como processo de comunicação

O conceito de relevância



Comunicação

Comunicação

MICHAELIS

Português

Inglês

Espanhol

Alemão

Italiano

Português Brasileiro

Digite o termo desejado

comunicação

co-mu-ni-ca-ção

sf

1 Ato ou efeito de comunicar(-se).

2 **LING** Ato que envolve a transmissão e a recepção de mensagens entre o transmissor e o receptor, através da linguagem oral, escrita ou gestual, por meio de sistemas convencionados de signos e símbolos.

3 O conteúdo da mensagem transmitida.

4 Transmissão de uma mensagem a outrem.

5 Exposição oral ou escrita sobre determinado assunto, geralmente de cunho científico, político, econômico etc.

6 Ato de conversar ou de trocar informações verbais.

7 Nota, carta ou qualquer outro tipo de comunicado através da linguagem escrita.

8 Comunicado oral ou escrito sobre algo; aviso.

9 Aquilo que permite acesso entre dois lugares; passagem.

10 União ou ligação entre duas ou mais coisas.

11 **ELETRÔN** Transmissão de informações de um ponto para outro, usando-se sinais em fios ou ondas eletromagnéticas.

<http://michaelis.uol.com.br/busca?id=wWIL>

Comunicação

Sua língua na Internet

Dicionário Caldas Aulete

Gramática básica

Dicionário

Página principal O que é Palavra do dia Downloads Convi



Verbetes Atualizado Verbetes Original

comunicação

(co.mu.ni.ca.ção)

A A A A

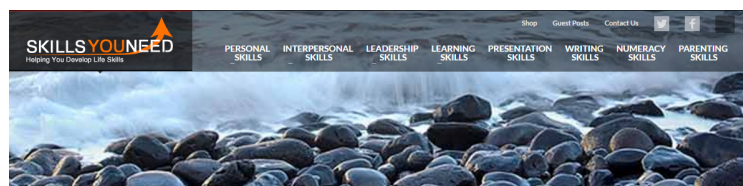
sf.

1. Conceito, capacidade, processo e técnicas de transmitir e receber ideias, mensagens, com vistas à troca de informações, instruções etc.: *A comunicação é um pré-requisito para a formação e consolidação de uma sociedade.*

2. Ação ou resultado de comunicar(-se), de transmitir e receber mensagens: *Esta empresa precisa melhorar a comunicação com as filiais.*

<http://www.aulete.com.br/Comunicação>

Comunicação



Search for SkillsYouNeed:

Google™ Custom Search Search

Start improving your life in just 5 minutes a day!

Get our 5 free 'One Minute Life Skills' and our weekly newsletter:

First Name *

Email Address *

Get Free Life Skills Advice from SkillsYouNeed

The SkillsYouNeed Guide to

What is Communication?

See also: [Interpersonal Communication Skills](#)

Communication is simply the act of transferring information from one place to another.

Although this is a simple definition, when we think about how we may communicate the subject becomes a lot more complex. There are various categories of communication and more than one may occur at any time.

The different categories of communication include:

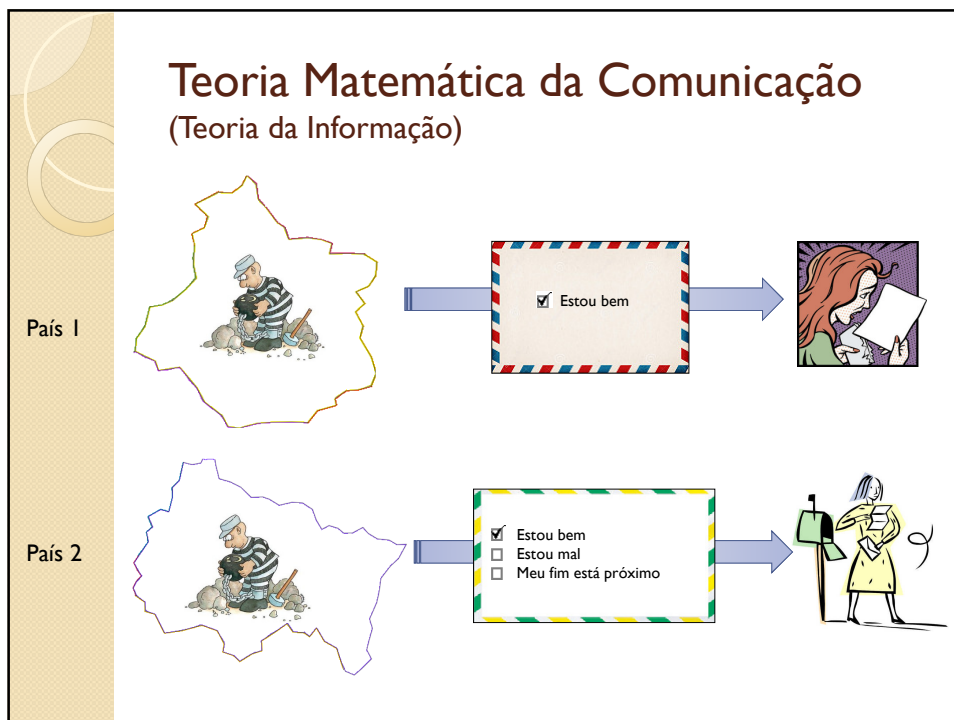
Spoken or Verbal Communication: face-to-face, telephone, radio or television and other media.

Non-Verbal Communication: body language, gestures, how we dress or act - even our scent.

Written Communication: letters, e-mails, books, magazines, the Internet or via other media.

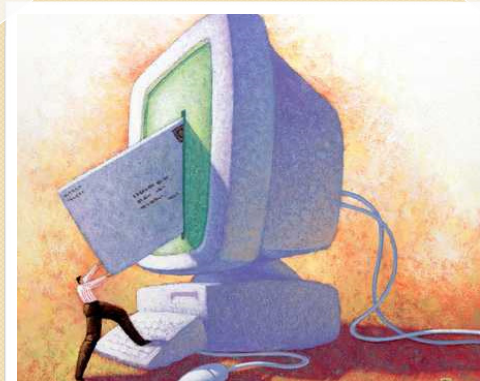
Visualizations: graphs and charts, maps, logos and other visualizations can communicate messages.

<http://www.skillsyouneed.com/general/what-is-communication.html>



Ciência da Informação e a Comunicação

- LARA, M. L. G. Linguagens documentárias, instrumentos de mediação e comunicação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 26, n. 1/2, p. 72-80, 1993.
- JANUARIO, S. B. B. A relação interdisciplinar entre a ciência da informação e a ciência da comunicação: o estudo da informação e do conhecimento na biblioteconomia e no jornalismo. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 7, n. 2, p. 151-165, 2010.
- CIBANGU, S.K. Communication Science and Informartion Science: convergences and divergences. **The Emporia Research Studies**, v.50, n.1, 2015



**Recuperação de
informação como
processo de
comunicação**

RI como processo de comunicação

*La recuperación de información es un proceso de comunicación donde se relacionan emisor y receptor con el fin de cubrir una necesidad de información. **Al hacer la pregunta al sistema el hombre funciona como emisor y el ordenador como receptor. En contrapartida el ordenador al presentar su respuesta pasa a ser el emisor y el hombre el receptor. Esta interacción solo es viable a través del uso del lenguaje.***

(VIEIRA, 1994, p.6)

Recuperação de informação é um processo de comunicação em que o emissor e o receptor se relacionam para cobrir uma necessidade de informação. Ao fazer uma pergunta ao sistema o homem funciona como emisor e o computador como receptor. Em contrapartida, o computador, ao apresentar a sua resposta passa a ser o emissor e o homem o receptor. Esta interação só é possível através do uso da linguagem.

RI como processo de comunicação

Recuperação de informação é um processo de comunicação. Em certo sentido é **um meio pelo qual autores e criadores de registros se comunicam com os leitores**, mas indiretamente e possivelmente com um longo intervalo de tempo entre a criação de uma mensagem ou texto e a sua entrega para o usuário de um sistema de recuperação de informação. [...] **As linguagens e os canais de tal sistema de comunicação** são bastante diferentes de outros modelos bem conhecidos, tais como a radiodifusão ou a comunicação ponto-a-ponto.

(MEADOW et al, 2007, p.3)

RI como processo de comunicação

Considerando que uma das mais comuns formas de codificação de ideias é a palavra, então um **documento pode ser visto como um canal de comunicação** [...] **Desse modo, o ciclo documentário tem similaridades com um sistema de comunicação**, sobretudo no que se refere às etapas finais do ciclo, ou seja, processamento da informação condensada, produtos do processamento, interrogação e busca e recuperação de informação. Além disso, é possível estudar esse sistema a partir do referencial da análise do problema técnico da comunicação, evidentemente com as adaptações conceituais pertinentes.

(GUEDES; ARAÚJO JÚNIOR, 2014)

RI como processo de comunicação



Um sistema de recuperação de informação é um ambiente linguístico mediador da comunicação entre um estoque de informação e os seus requisitantes.



O Conceito de Relevância

O Conceito de Relevância

O conceito de relevância é crucial na área de Recuperação de Informação, sendo muitas vezes utilizado na própria enunciação dos objetivos dessa área
(COOPER, 1971; MIZZARO, 1998)

O Conceito de Relevância

- **iDicionário Aulete:**
(re.le.vân.ci:a)
sf.
1. Qualidade ou característica do que é relevante; **IMPORTÂNCIA:** *temas da maior relevância.* [Antôn.: irrelevância.]
[F.: relevar+ -ância.]
- **Michaelis:**
re le vâ n ci a
sf
1 Qualidade do que é relevante.
2 Aquilo que tem **importância** ou relevo num contexto determinado; pertinência:
“Os temas transversais constituíram um conjunto de temas de grande relevância para uma educação que visa a formação de sujeitos realmente capazes de conviver em harmonia” (NG).
3 Parte saliente de uma superfície plana.
- **Aurélio on-line:**
Relevo. **Importância:** lado vantajoso de alguma coisa.
O necessário, o indispensável. Que releva. Que sobressai. **Importante.**

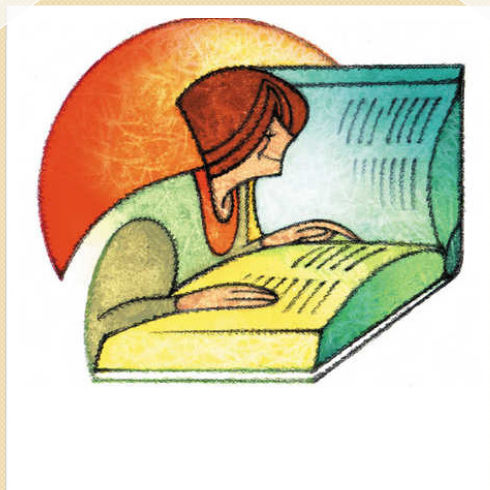
O Conceito de Relevância

- Saracevic (1975):
 - Neste contexto de comunicação e problemas de comunicação, podemos considerar a relevância como uma medida da eficácia de um contato entre uma fonte e um destino em um processo de comunicação. E, uma vez que uma medida é uma relação, a relevância é também uma relação;
 - A comunicação do conhecimento é efetiva quando e se a informação que é transmitida pelo emissor resulta em mudanças no receptor. Relevância é a medida dessas mudanças.



[de cérebros à biblioteca](#)

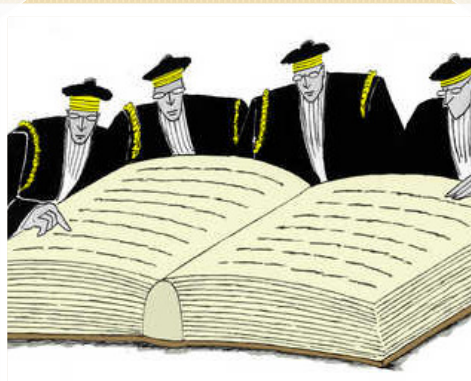
Vídeo



Sugestão de Leituras

Sugestão de Leituras

- KHAPRE, SHAILESH; BASHA, SALEEM. A Theoretical Paradigm of Information Retrieval in Information Science and Computer Science. International Journal of Computer Science Issues, v.9, n.2, 2012.



Referências

Referências

- BRIET, Suzanne. **What is Documentation?**: English Translation of the Classic French Text. Oxford, UK: Scarecrow Press, 2006.
- BUCKLAND, M.K. Information as thing. **Journal of the American Society of Information Science**, v.42, n.5, 1991. p.351-360.
- COOPER, W.S. A Definition of Relevance for Information Retrieval. **Information Storage and Retrieval**, v.7, pp.19-37, 1971.
- GUEDES, W.; ARAÚJO JÚNIOR, R. H. Estudo das Similaridades entre a Teoria Matemática da Comunicação e o Ciclo Documentário. **Informação e Sociedade: estudos**, v.24, n.2, 2014.
- LE COADIC, Y-F. **A Ciência da Informação**. 2.ed. Brasília: Briquet de Lemos, 2004.
- MEADOW, C.T.; BOYCE, B.R.; KRAFT, D.H.; BARRY, C. **Text Information Retrieval System**. 3rded. London UK: Elsevier, 2007.
- MOOERS, C. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v.2, n.1, 1951, p.20-32.
- RIECKEN, R. F. Frame de temas potenciais de pesquisa em Ciência da Informação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 3, n. 2, jan./jun. 2006. p.43-63.
- SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v.50, n.12, 1999.
- SARACEVIC, T. Information science: origin, evolution and relations. In: **International Conference on Conceptions of Library and Information Science**: historical, empirical and theoretical perspectives, 1991, Finland. Proceedings... Helsinki: 1991.
- SARACEVIC, T. Relevance: A review of and a framework for the thinking on the notion of information science. **Journal of American Society for Information Science**, v.26, n.6, p. 321-343, 1975.
- VIEIRA, S.B. **La recuperación automática de información jurídica**: metodología de análisis lógico-sintáctico para la lengua portuguesa. 1994. 382 f. Tese (Doutorado em Ciência da Informação)- Universidad Complutense de Madrid, Madrid, 1994.