

## Aspectos conceituais da indexação

Isidoro Gil Leiva

**Como citar:** LEIVA, Isidoro Gil. Aspectos conceituais da indexação. *In:* LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti Lopes (org.). **Política de indexação.** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2012. p. 31-106. DOI: <https://doi.org/10.36311/2012.978-85-7983-199-7.p31-106>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Marília



**CULTURA  
ACADÊMICA**  
*Editora*



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).

## CAPÍTULO 2

### ASPECTOS CONCEITUAIS DA INDEXAÇÃO<sup>1</sup>

*Isidoro Gil Leiva*

#### 1 O PROCESSO COGNITIVO E A INDEXAÇÃO

A indexação gera palavras-chave, índices ou os cabeçalhos de assunto de um documento. Para obtê-los, previamente foi desencadeada uma sucessão interativa e simultânea dos processos mentais que têm a ver com a percepção da informação, da memória e da compreensão. As operações mentais realizadas pelos seres racionais para a recepção seletiva de informação, a sua codificação simbólica e seu armazenamento e recuperação, denomina-se processo cognitivo.

A Psicologia Cognitiva é a disciplina que estuda os processos cognitivos como a percepção sensorial da informação, a aprendizagem (linguagem, leitura e escrita), a memória ou a capacidade de raciocínio. Então, vamos fazer uma breve passagem por essas atividades mentais do processo cognitivo realizadas durante a indexação. Embora estas atividades mentais estejam interligadas e sejam concorrentes durante a realização da indexação, aqui na FIG. 3 são apresentadas de forma sequencial para uma maior clareza expositiva.

---

<sup>1</sup> Tradução autorizada pelo autor, de síntese adaptada pelo próprio autor dos capítulos “El proceso cognitivo y la indización” e “La indización” do livro “Manual de indización. Teoría y práctica. Gijón: Trea, 2008”

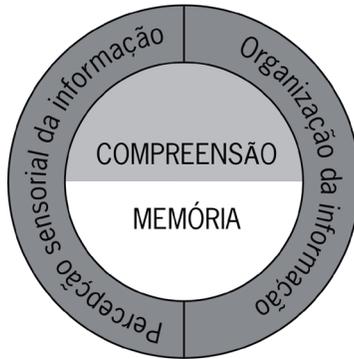


Figura 5 – Processo cognitivo na indexação

Fonte: GIL LEIVA, 2008

## 1.1 ORGANIZAÇÃO DA COMUNICAÇÃO

A comunicação entre as pessoas ocorre principalmente por meio dos discursos oral, textual e audiovisual. Para decodificar uma mensagem visual através da leitura, a exibição de um relatório ou um sinal de trânsito, ou uma mensagem auditiva (uma conversa ou uma sirene) deve haver pelo menos um código conhecido pelas partes (emissor e receptor) e, em alguns casos, que a mensagem esteja de acordo com as regras estabelecidas no código. Caso contrário, a comunicação não é possível.

Das três principais formas de comunicação humana, vamos nos deter na do discurso textual. A palavra *discurso*, usamos aqui, como unidade de comunicação do que fazemos, pensamos ou sentimos em relação ao texto (livro, artigo, tese, legislação, receita culinária, etc.)

### 1.1.2 DISCURSO TEXTUAL

#### 1.1.2.1 CONCEITO DE TEXTO

O conceito de texto que vamos utilizar é o proposto pela linguística textual, também conhecido como *ciência do texto*, *semiótica do texto*, *semiótica discursiva* ou *linguística do texto*, entre outras denominações. A linguagem textual é decorrente da linguística funcional surgida na década de 1960 para o estudo teórico e prático da estrutura e função dos textos. A maioria dos linguistas empregam *texto* e *discurso* como sinônimos. Portanto, a partir de agora nós entendemos “*texto*” e “*discurso*” como equivalentes e como uma unidade eficaz de comunicação, escrita ou oral, feita por regras estabelecidas em uma língua.

### 1.1.2.2 CRITÉRIOS DE TEXTUALIDADE

Os textos devem cumprir certas condições ou critérios de textualidade visíveis tanto para o falante quanto para o leitor, permitindo-lhe determinar se um conjunto de palavras, frases e parágrafos formam um texto ou não, isto é, se possui textualidade, ou o que é o mesmo, uma unidade de comunicação global. Os critérios que conferem textualidade a um texto são:

Quadro 1 – Resumo de critérios de textualidade de Beaugrande e Dressler (1997)

<b>Crítérios</b>	<b>Explicação</b>
COESÃO (centrada no texto)	As frases são ligadas por meio de elementos superficiais (repetição, elipse ou conectores) para manter uma ligação sequencial e, dessa forma, podem ser entendidas umas em relação às outras.
COERÊNCIA (centrada no texto)	Unidade subjacente que confere a união das frases e dos parágrafos para que o leitor ou ouvinte entendam o significado global do texto.
INTENÇÃO (centrada no emissor)	O autor tem um plano para cumprir um objetivo (transmitir conhecimento).
ACEITABILIDADE (centrada no receptor)	Uma série de sequências que constituem um texto coeso e coerente é aceitável para um determinado receptor se ele percebe que tem alguma relevância (adquirir conhecimentos novos ou cooperar com seu interlocutor para conseguir um objetivo discursivo determinado).
SITUACIONAL (centrada no contexto)	Fatores que fazem com que um texto seja relevante dependendo da situação ou do contexto em que aparecem.
INFORMATIVO (centrada no receptor)	Serve para avaliar até que ponto as sequências de um texto são previsíveis ou inesperadas, se transmitem informação conhecida ou nova.
INTERTEXTUALIDADE (centrada no receptor)	Fatores que fazem com que a utilização e compreensão adequada de um texto dependam do conhecimento de outros textos.

Fonte: BEAUGRANDE; DRESSLER, 1997, p. 33-45.

Beaugrande e Dressler (1997, p. 46) apontam, e os aceitamos literalmente, que esses critérios funcionam como “princípios fundamentais” da comunicação textual, portanto, a violação de qualquer um destes sete pontos estabelecidos atenta contra o processo de comunicação. Junto a estes princípios fundamentais existem os “princípios reguladores” que controlam a comunicação textual, ou seja, a *eficácia*, a *efetividade* e a *adequação* dos textos. A *eficácia* depende se os participantes usam um mínimo de esforço na comunicação, a *efetividade* de

um texto depende se gera uma forte impressão no receptor e se cria condições mais favoráveis para que o autor possa alcançar o objetivo comunicativo que havia estabelecido, e, finalmente, a *adequação* de um texto depende se ele estabelece um equilíbrio entre o uso do texto numa determinada situação e o modo como as regras da textualidade são respeitadas.

### 1.1.2.3 ESTRUTURA DO TEXTO

Para explicar a composição interna dos textos, vamos seguir literalmente a teoria da macroestrutura elaborada por Teun A. van Dijk e exposta em seus livros *Some Aspects of Text Grammars* (1972), *Text and Context* (1977) e *Ciencia Del Texto* (1996).

Segundo o autor, as relações produzidas entre as proposições (utilizaremos indistintamente proposição ou frase) de uma sequência proporcionam uma coerência e esta sequência é coerente quando cada frase da sequência pode ser compreendida em relação à interpretação de outras proposições da sequência ou das frases implícitas. Ao mesmo tempo, duas frases apresentam uma ligação semântica quando há uma série de condições como um pressuposto, uma condição causal, uma comparação ou as frases se relacionam por meio de uma conexão parte-todo.

Van Dijk propôs as noções de *micro* e *macroestrutura* para distinguir os dois níveis textuais. O linguista observa que, se uma frase é “mais” do que uma série de palavras, podemos analisar os textos num nível que supera a estrutura das sequências das frases. Desse modo, existem conexões baseadas no texto como um todo ou, pelo menos, em unidades textuais maiores. E são essas estruturas de texto mais globais que são chamadas de *macroestrutura*. Assim, as macroestruturas representam a estrutura global de significado do texto. Dessa forma, enquanto as sequências de frases devem satisfazer as condições de coerência linear (relação semântica entre as frases em cadeia), os textos não devem apenas atender a essas condições, mas à coerência global (a percepção do significado e sentido pelo receptor).

As macroestruturas ao serem portadoras da coerência global (significado do texto) estão baseadas num nível superior ao das frases. Uma sequência completa ou parcial de um grande número de proposições pode formar uma unidade de sentido num nível mais global. Portanto, o autor chama de macroestrutura do texto, a macroestrutura mais geral e global de um texto completo, enquanto que certas partes do texto podem ter caminhos de macroestruturas formadas por proposições. Finalmente, devemos dizer que cada

macroestrutura deve satisfazer as mesmas condições de coesão e coerência que os níveis microestruturais (proposições).

A redução das proposições a macroestruturas seria:

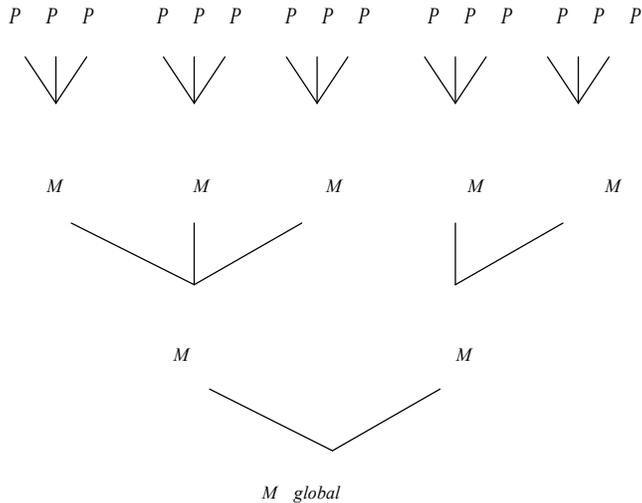


Figura 6 – Redução de proposições a macroestrutura

FONTE: VAN DIJK, 1996, P. 56

**Exemplo**

Apliquemos esta teoria a uma parte do resumo de um artigo de revista selecionado aleatoriamente:

Texto (1)

Quando falamos de depressão nos referimos a um esgotamento emocional, embora haja também uma significativa falta de realização pessoal e também se detecta um processo de despersonalização constante. Assim, esses últimos seriam os sintomas depressivos mínimos para detectar a presença de mobbing. Em contraste, a ansiedade causada por assédio psicológico no trabalho se manifesta pela presença de intrusões obsessivas; o surgimento de sonhos repetitivos relacionados com o abuso; irritabilidade com as pessoas ao redor; e um desenvolvimento progressivo do comportamento de esquiva.



Vamos traduzir essa noção de macroestrutura do linguista ao que nos interessa, a indexação de documentos. Durante a leitura, devemos nos concentrar na localização e na seleção das proposições que contêm o significado ou então construí-las, já que uma frase com o significado global pode ser reduzida a uma ou mais formas substantivas, como por exemplo: PRODUTOS DE CONFEITARIA; AMÊNDOAS; PASTELARIA, ou formas mais complexas, como BASES DE SCHIFF; MÉTODO DE ELEMENTOS DE CONTOURNO ou ESPECTROSCOPIA RMN. No exemplo anterior Texto (1) o significado global “Os sintomas que caracterizam o mobbing são a depressão e a ansiedade” nos remeteria às seguintes palavras-chave: ANSIEDADE; DEPRESSÃO; SINTOMAS; MOBBING.

Van Dijk incorporou à esta teoria da macroestrutura uma metodologia para chegar até as macroestruturas de um texto. Criou um conjunto de regras para unir as proposições em macroestruturas. Este tipo de norma transforma um conjunto de proposições em outro conjunto de proposições (diferentes ou iguais). Estas regras são chamadas de *macroregras*. Cada feixe de linhas de proposições que se junta para formar uma macroestrutura (M<sub>i</sub>) de um nível superior representa uma macroregra, conforme mostrado no exemplo anterior de “Redução de proposições a macroestruturas”.

Portanto, no nível cognitivo podemos considerar as macroregras como operações para reduzir a informação semântica.

A noção de macroestrutura é apresentada por Van Dijk para trabalhar o conceito de tema textual ou oral, pois sem as macroestruturas, ao ouvir ou ler uma série de frases, o falante / leitor deveria sempre perguntar “o que é isso?” “Onde o autor quer chegar?”. A macroestrutura permite a um falante ou ao leitor responder perguntas como “o que foi falado na conferência?”, “Do que se trata o livro?”, incluindo textos longos e complicados. Um falante também pode responder quando o assunto ou o objeto em si não são mencionados de maneira total e explícita. Assim, o tema de um discurso (textual ou oral) é o mesmo que a macroestrutura, e por sua vez, as macroregras são a reconstrução formal desta “dedução” do tema. Discutiremos este assunto adiante com mais profundidade.

#### 1.1.2.4 TIPOS DE TEXTO

Antes de começarmos esta parte é necessário esclarecer dois pontos. Por um lado, continuaremos utilizando texto e discurso como sinônimos e, por outro, começaremos a falar de tipos de texto, que como veremos agora, não existe uma correlação plena com o conceito de tipologia documental como utilizado em Biblioteconomia e Documentação. Interessa-nos aprofundar o assunto

das tipologias textuais porque os indexadores trabalham com documentos que contém informação organizada de diferentes maneiras.

A partir de 1970 com o desenvolvimento da linguística textual, muitos especialistas estudaram os textos e fizeram a proposta de diversas tipologias textuais. Algumas das mais importantes são:

#### Quadro 2 – Tipologias textuais

Autor	Tipologias propostas
Werlich (1975)	Descritivo, narrativo, expositivo, argumentativo e instrutivo.
Grosse (1976)	Normativo, de contato, de indicação de grupo, poético, de auto-manifestação, exortativos, de transferência de informação, de transição.
Van Dijk (1972 y 1977)	Argumentativo, científico, narrativo, jornalístico e outros possíveis como conferência, sermão, atestado, petição, discurso público, entre outros.
Adam (1992)	Narrativo, descritivo, argumentativo, explicativo e dialógico.

Fonte: Elaborado pelo autor.

Conforme descrito a seguir, num texto podem aparecer um ou vários tipos de uma vez, daí a dificuldade para estabelecer o gênero de um texto em determinadas ocasiões. Embora não se tenha chegado a uma teoria geral dos tipos de texto, existe um consenso sobre alguns deles. Vejamos os tipos mais estudados pelos linguistas e em que tipologias textuais se manifestam:

- argumentativo** raciocínio argumentativo, julgamentos ou opiniões para persuadir ou atrair o interlocutor diante de determinadas teses, usando para isso argumentos intelectuais próprios do emissor que sejam valores, crenças ou experiências.  
Tipologias documentais resultantes: artigo científico, resenha
- narrativo** eventos narrativos ou temporais principalmente de pessoas.  
Tipologias documentais: relato, novela, conto, fábula, reportagem de imprensa, atestado, relato autobiográfico
- descritivo** dados espaciais de objetos, lugares, sentimentos (monumentos, cidades, paisagens) ou algo organizado como um todo (pessoas, animais).  
Tipologias documentais: dicionário, descrição científico-técnica

<b>expositivo</b>	organiza as ideias através da síntese e da análise para explicar um determinado assunto de maneira compreensível. Tipologias documentais: artigo científico, manuais didáticos, relatórios, teses.
<b>instrutivo</b>	Informa sobre como proceder. Tipologias documentais: normas, manuais de procedimentos, receitas culinárias.
<b>dialogado</b>	informação sequencial construída entre dois ou mais falantes de um assunto. Tipologias documentais: conversas ou reproduções de conversas por monólogo.
<b>jornalístico</b>	classifica a informação de acordo com sua relevância. Tipologias documentais: artigo informativo, artigo de opinião

Podemos dizer que existem muitas classificações de tipos de textos ou discursos. Em seguida, reproduzimos a proposta de Brewer (1980) e complementamos, assim, o Quadro 3.

Quadro 3 – Tipos de textos

DISCURSO		FORÇA DO DISCURSO		
		Informar	Entreter	Persuadir
(Estrutura subjacente)				
DESCRIÇÃO (Espacial)	Descrição Técnica Botânica Geografia	Descrição Comum	Anúncio de uma casa	Descrição poética
NARRAÇÃO (Tempo/Evento)	Conto História Instruções Receitas Biografia	Novela de mistério Novela do oeste Biografia Drama	Novela-mensagem Parábola Fábula Anúncios Drama	Novela literária Conto Drama
EXPOSIÇÃO (Lógica)	Artigo científico Filosofia Definição abstrata		Sermão Propaganda Editorial Anúncios Ensaio	

Fonte: AMAT DE BETANCOURT, 1991, p. 166.

Em suma, uma estrutura e alguns conteúdos determinados configuram as tipologias textuais, por isso que elas abrem o caminho da compreensão. E a compreensão de um texto abre o caminho da indexação, algumas pesquisas confirmam que quanto mais estruturado um resumo, mais ele contribue para a indexação.

Aqui está um exemplo para comprovar a informação que pode fornecer as estruturas internas de várias tipologias de texto. Exemplo:

Imaginemos que um indexador especialista esteja em sua mesa de trabalho pronto para começar a trabalhar na análise de documentos jurídicos (artigos de revistas, notícias de imprensa, atas, livros, relatórios e acórdãos). O indexador não sabe qual documento vai indexar até abrir a pasta em sua frente. Na memória de longo prazo deste profissional está armazenada a informação que do título de um artigo científico pode obter dados do tema (Que), do período ou tempo (quando) ou da localização geográfica do assunto tratado (onde), e assim sucessivamente, como demonstrado no seguinte quadro:

Quadro 4 – Relação entre tipologia documental, estrutura textual e paradigma de Lasswel

Artigo	Artigo de jornal	Ata	Informe	Sentença judicial
<b>Título:</b> O que, Quem, Quando, Onde	<b>Título:</b> O que, Quem, Quando, Onde	<b>Assistentes e acusados:</b> Quem	<b>Título:</b> O que, Quem, Quando, Onde	<b>Cabeçalho:</b> Quem, O que, Quando, Onde
<b>Resumo:</b> O que, Quem, Quando, Onde, Como	<b>Cabeçalho:</b> O que, Quem, Quando, Onde Por que, Como	<b>Local e data:</b> Onde, Quando	<b>Índice:</b> O que, Quem, Quando, Onde	<b>Antecedentes:</b> O que, Como
<b>Palavras chave:</b> Conceitos relevantes propostos pelo autor	<b>Texto:</b> O que, Quem, Quando, Onde Por que, Como	<b>Acordos adotados e deliberações:</b> Que, Por que	<b>Epígrafes:</b> O que, Quem, Quando, Onde, Por que, Como	<b>Fundamentos do Direito:</b> O que, Por que
<b>Introdução:</b> O que, Por que		<b>Notário:</b> Quem		<b>Decisão judicial:</b> O que
<b>Material e método:</b> Quando, Onde, Como				
<b>Resultados e discussão:</b> O que				
<b>Conclusões:</b> O que				

Fonte: Elaborado pelo autor.

Cada uma das questões tratadas (paradigma de Lasswell) refere-se a:

O que	→ Assunto, ação, objeto	→ descritor temático → embasamento de assunto → subcabeçalho de assunto
Como	→ instrumento, técnica, método	→ descritor temático → embasamento de assunto → subcabeçalho de assunto
Quem	→ nome próprio da pessoa → nome próprio do objeto	→ Descritor onomástico → embasamento de assunto → identificador → embasamento de assunto
Quando	→ tempo	→ Descritor cronológico → subcabeçalho de tempo
Onde	→ lugar	→ Descritor topográfico → subcabeçalho de lugar

Continuando o exemplo acima, quando o indexador abre a pasta reconhece a tipologia textual que aparece diante de seus olhos: uma sentença judicial. De maneira automática e inconsciente recupera da sua memória de longo prazo informações deste tipo de texto e descarta as informações de outros tipos. Então, sabe que o “cabeçalho” da frase contém lugar e data, um resumo do objeto de litígio e os nomes dos participantes, em “antecedentes” estão os fatos e ações judiciais que deram origem ao litígio, em “Fundamentos da lei” estão desenvolvidos os argumentos e os fundamentos legais subjacentes à decisão do juiz e, finalmente, a “decisão judicial” contém a solução do litígio.

Em resumo, a informação que recupera de sua memória em relação à estrutura das frases indica por onde deve começar a trabalhar e a classe de conceitos que podem ser encontrados nesta tipologia (conceitos relativos ao tema - descritores temáticos, ao tempo – descritores cronológicos, ao lugar - descritores onomásticos, aos nomes próprios de pessoa ou objetos - identificadores).

## 1.2 PERCEPÇÃO SENSORIAL DA INFORMAÇÃO

A informação pode chegar ao indexador por três sentidos: visão, audição e tato (por exemplo: indexador cego que usa a leitura tátil). Pela visão, recebe o texto escrito (livro, artigo, lei etc.), a imagem fixa (fotografia, anúncio publicitário etc.) e a imagem em movimento (vídeo, filme), bem como objetos físicos sujeitos à indexação (escultura). Pela audição, recebe os sons articulados por sistemas linguísticos (discurso), sons por composição artística (música), sons pela interação social (ruídos urbanos, domésticos, festas etc.), sons mecânico-industriais (cadeia de produção, perfurações etc.), sons da natureza (ruídos da

floresta, ondas etc.) ou sons apelativos (relógios, timbres, sinos chamando para a oração etc.) No momento em que a informação é percebida por algum dos sentidos, são ativados os processos da memória.

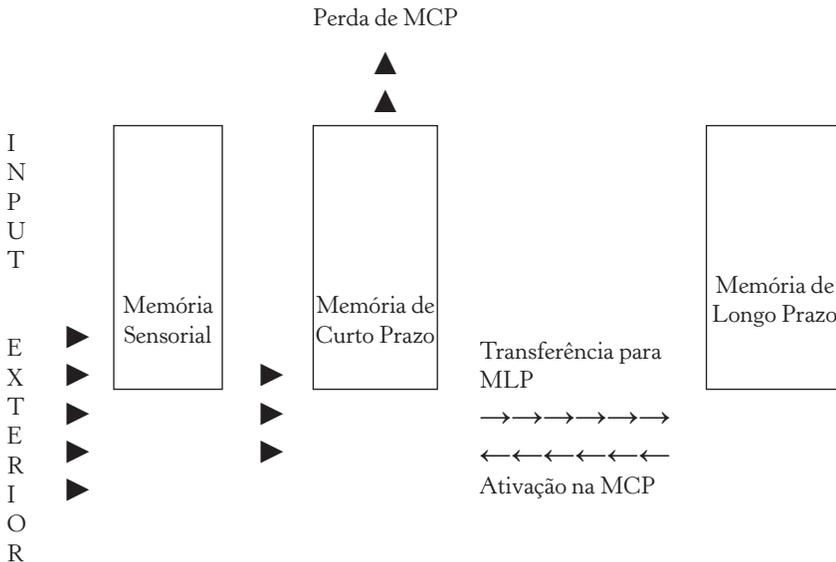
### 1.3. ATIVAÇÃO DA MEMÓRIA

Durante a indexação, a memória entra em jogo constantemente, uma vez que o indexador tem muitos conhecimentos armazenados e os utiliza a cada momento. Armazena dados relativos ao mesmo processo de indexação (tanto teóricos quanto práticos), adquiridos durante sua formação e através da experiência, dados da política de indexação da instituição aplicáveis sobre o uso de linguagem de indexação ou em relação aos usuários; ou dados da área que trabalha (Química, Direito, Meio ambiente etc.), entre outros.

Segundo Vega (1998, p. 59) as pesquisas sobre este assunto giram em torno de três áreas:

- a) Estruturas da memória. São postulados três dispositivos básicos: memória sensorial, memória de curto prazo e memória de longo prazo. Cada uma dessas estruturas tem propriedades funcionais específicas sobre o tipo de informação que armazena, a capacidade de armazenamento, a persistência temporal da informação e o formato simbólico da informação.

Figura 7 – Modelo estrutura da memória



Fonte: VEGA, 1998, p. 61.

- b) Processos da memória. A memória se apresenta como uma sequência de operações bem diferenciadas, pelo menos conceitualmente:
- 1) Codificação: ocorre quando o fluxo de estímulos que chega aos receptores são analisados, organizados ou processados gerando um determinado código simbólico que constitui uma réplica cognitiva do *input*.
  - 2) Armazenamento: a informação codificada é armazenada e retida por um tempo variável.
  - 3) Retenção: a disponibilidade da informação armazenada acontece desde alguns milésimos de segundos até meses ou anos, dependendo, entre outros fatores, do tipo de codificação que ocorreu.
  - 4) Recuperação de informação: a informação pode ser recuperada com diversos propósitos.
- c) Representações da memória. As pesquisas são direcionadas para determinar se a forma simbólica da informação que se armazena, trata ou recupera é uma linguagem interna, se é um código abstrato, verbal ou de imagens mentais.

### 1.3.1 MEMÓRIA SENSORIAL

A memória sensorial ou dos sentidos registra dados de duração escassa (milésimos de segundos) que não são analisados e podem ser visuais (memória icônica), auditivos (memória ecóica), táteis (memória tátil) ou de movimento (memória cinestésica).

### 1.3.2 MEMÓRIA DE CURTO E LONGO PRAZO

Para Vega (1998, p. 88) a memória de curto prazo (MCP) e a memória de longo prazo (MLP) são estruturas propostas por psicólogos cognitivos nas décadas de 1960 e 1970, baseadas em algumas observações empíricas que sugerem a existência de dois mecanismos de memória diferentes. As distinções entre a MCP e a MLP que o autor coloca estão descritas no quadro a seguir:

## Quadro 5 – Diferenças entre a MCP e a MLP

MEMÓRIA	PROPRIEDADES DA MEMÓRIA
MCP	<p><i>Persistência limitada:</i> A informação recebida parece desaparecer entre 15 e 30 segundos após o recebimento. A maioria das pessoas pode repetir um número de 6 ou 7 dígitos imediatamente após ou transcorrido alguns segundos de ter escutado pela primeira vez; podemos repetir uma lista de 6 ou 7 palavras ou letras; podemos repetir a última frase de um interlocutor. No entanto, em todas estas situações, passados alguns segundos, a informação parece desaparecer por completo.</p>
	<p><i>Capacidade limitada:</i> Se alguém nos lê 15 dígitos, não seremos capazes de repeti-los sem erros nem que seja logo em seguida. Os limites da MCP estão estabelecidos em sete unidades de informação. Quando se excede esta quantidade, a MCP recebe uma sobrecarga que se manifesta por uma escassa retenção.</p>
	<p><i>Ativação:</i> Por meio de <i>inputs</i> sensoriais.</p>
	<p><i>Busca ou recuperação da informação:</i> Acesso quase instantâneo da informação armazenada.</p>
MLP	<p><i>Persistência ilimitada:</i> A informação armazenada permanece em estado inativo ou latente e apenas são recuperados fragmentos da informação, quando uma determinada tarefa assim o exige. Diante da pergunta “Quem descobriu a América?”, respondemos de forma rápida sem erro. O segmento de informação “Colombo descobriu a América” estava armazenado num estado inativo até que a pergunta anterior desencadeou o processo de recuperação dessa informação.</p>
	<p><i>Capacidade ilimitada:</i> Podemos armazenar enormes quantidades de informação que permanecem conosco enquanto vivemos.</p>
	<p><i>Ativação:</i> Opera a partir de <i>inputs</i> procedentes tanto do exterior como do próprio sistema cognitivo.</p>
	<p><i>Busca ou recuperação da informação:</i> Devido ao enorme repertório de conhecimentos que armazena, requer processos sistemáticos de busca e de inferência.</p>

Fonte: Elaborado pelo autor, a partir de Vega (1998, p. 89).

A MCP e a MLP são dispositivos intimamente relacionados, apesar de suas diferenças funcionais. Há uma interação dupla entre as duas estruturas. Por um lado, a informação retida na MCP, a partir de *input* sensorial, por vezes, é transferida para MLP (por exemplo, se o número de telefone que ouvimos pela

primeira vez nos interessa, podemos repeti-lo várias vezes e, conseqüentemente, transferi-lo para a MLP para lembrá-lo para sempre). Por outro lado, quando queremos recuperar alguma informação da MLP para uso imediato, tal informação é ativada na MCP. A MCP é, portanto, um dispositivo que opera a partir de *inputs* vindos tanto do exterior como do próprio sistema cognitivo.

Na MCP se combina a informação do *input* sensorial com os conhecimentos permanentes do sujeito, executando operações complexas de codificação e processamento da informação. Devido a essa característica importante, a MCP também é chamada de “memória ativa” ou “memória operativa”. Exemplifiquemos com uma operação própria da indexação:

Exemplo:

Caso 1: Quando um indexador novato seleciona como palavra-chave uma parte do texto, por exemplo, “empréstimo entre bibliotecas” deve manter alguns momentos em sua memória de curto prazo (MCP), enquanto tecla, anota ou procura o termo autorizado na linguagem de indexação. Comprovará que o descritor para essa expressão é “EMPRÉSTIMO INTERBIBLIOTECÁRIO”. Ao realizar esta operação várias vezes, armazenará em sua memória de longo prazo que “empréstimo entre bibliotecas” é um descritor de “EMPRÉSTIMO INTERBIBLIOTECÁRIO”.

Caso 2: Quando um indexador experiente lê várias vezes em um texto “desigualdade entre os sexos” e “sexismo” já sabe que o termo de indexação que corresponde a essas expressões é “DISCRIMINAÇÃO SEXUAL”, porque em sua memória de longo prazo tem armazenado que as duas primeiras noções não são descritores, portanto, atribuirá como termo de indexação “discriminação sexual” para tal texto.

Quadro 6 – Memórias participantes de acordo com o objeto indexado

Organização da Comunicação	Memórias
Texto	Memória Icônica / MCP / MLP
Imagem fixa	Memória Icônica / MCP / MLP
Audiovisual (Imagem movimento+som)	Memória Icônica / MCP / MLP
Som (Conversa, música, outros)	Memória Ecóica / MCP / MLP

Fonte: Elaborado pelo autor.

#### 1.4 COMPREENSÃO

A Psicologia Cognitiva nasceu da Psicologia como uma área de pesquisa na década de 1950 centrada no estudo teórico e prático de tudo o que se relaciona com a percepção, a memória, a aprendizagem e o raciocínio. Desde a década de 1970, os psicólogos têm procurado compreender os processos envolvidos na leitura e na compreensão de textos a partir da perspectiva cognitiva.

### 1.4.1 PROCESSOS NA COMPREENSÃO

Sanchez Miguel (1993, p. 44) que seguimos literalmente, explica em que consiste compreender um texto, bem como as estratégias e processos utilizados para isso. Entende que compreender um texto leva-nos a fazer seu significado, para o qual temos que:

1. Desvendar as ideias que contêm as palavras do texto; ou em outras palavras, construir ideias com as palavras do texto.
2. Ligar as ideias entre si, isto é, compor, em ordem, o fim condutor entre elas.
3. Assumir e/ou construir a hierarquia que tem ou que podemos conceber entre essas ideias, isto é, a macroestrutura ou tema global.
4. Reconhecer o padrão das relações que ligam as ideias globais e que produz uma estrutura textual ou tipologia textual determinada.

Para compreender um texto são realizadas muitas atividades cognitivas. Essas atividades podem ser agrupadas em três grupos de estratégias segundo Sánchez Miguel:

1. Estratégias para desvendar a informação do texto

Os tipos de conhecimento envolvidos e as estratégias textuais são esquematizadas pelo autor da seguinte maneira:

Quadro 7 – Tipos de conhecimentos e estratégias textuais

Conhecimentos	Níveis de processamento	Resultado
Significado das palavras. Forma ortográfica das palavras. Forma de letras e sílabas. Regras que relacionam grafemas com fonemas	Reconhecer as palavras  ↓ ↓ ↓	Acesso ao significado das palavras ou significado lexical.
Esquema proposicional. Conhecimentos da língua. Conhecimentos do mundo físico e social.	Construir proposições  ↓ ↓ ↓	Organização dos significados das palavras em termos de predicado e vários argumentos.

Conhecimentos dos textos (sinais da progressão temática do texto). Conhecimento sobre o mundo físico e social.	Integrar as proposições ↓ ↓ ↓	As proposições se relacionam entre si, tanto tematicamente como de maneira causal, motivacional ou descritiva.
Conhecimentos dos textos (sinais que indicam a macroestrutura no texto). Conhecimentos sobre o mundo físico e social	Construir a macroestrutura ↓ ↓ ↓	Derivam-se do texto e dos conhecimentos do leitor as ideias globais que individualizam, dão sentido e diferenciam as proposições derivadas do texto.
Conhecimentos dos textos (sinais de superestruturas). Conhecimentos do mundo físico e social	Inter-relação global das ideias	As ideias globais se relacionam entre si em termos causais, motivacionais, descritivos, comparativos ou temporais.

Fonte: SÁNCHEZ MIGUEL, 1993, p. 53, p. 89; SÁNCHEZ MIGUEL, 1998, p. 81.

## 2) Estratégias para a utilização do conhecimento prévio na leitura

A quantidade de conhecimento prévio, ou seja, armazenada na memória do leitor que pode ser acessada durante a leitura e a compreensão é enorme, como é também grande o número de inferências (referências, elaborações etc.) feitas durante a leitura e a compreensão. Algumas dessas inferências servem para formar os papéis semânticos, comprovar a continuidade temática, estabelecer relações causais entre as ideias ou criar o significado global, entre outras. E acredita-se que durante a leitura e compreensão só são ativados os conhecimentos prévios que o texto evoca e estão relacionados com a construção gradual do tema global do texto pelo leitor.

## 3) Estratégias para auto-regular o curso da interpretação. Se subdividem em:

⇒ Planejamento: A leitura de um texto depende muito do objetivo ou do propósito com que lemos. Tem sido demonstrado que o tempo de leitura de uma frase varia dependendo de qual é o objetivo da leitura. Isso fornece evidências de que os processos e estratégias específicos de leitura são executados de forma muito diferente, dependendo de qual é o objetivo que perseguimos.

⇒ Avaliação: É necessário avaliar em todo o momento se o grau de compreensão é satisfatório em relação ao objetivo. Para isso, usamos critérios léxicos

(se compreendemos o significado das palavras), os critérios sintáticos (compreensão de frases) e critérios semânticos (verificamos se o que é afirmado no texto é coerente com o nosso conhecimento do mundo, consistente e coerente entre si ou suficiente para alcançar os nossos propósitos).

⇒ Regulamento: Além de detectar falhas de compreensão de acordo com qualquer um dos critérios mencionados, os leitores devem decidir se é necessário adotar alguma medida como releitura do texto para esclarecer o significado; saltar no texto para encontrar informação útil; resumir o que é conhecido; ou determinar o tipo de palavra problemática.

Depois desta breve introdução sobre a compreensão dos textos, vejamos agora alguns aspectos da leitura, uma vez que a leitura é a base da indexação. Durante a leitura são detectados diversos níveis de processamento. Para essa abordagem, esquematizamos da seguinte maneira o que foi proposto por dois especialistas:

Quadro 8 – Níveis de processamento durante a leitura

Antonini e Pino (1991, p. 140)	Vega (1998, p. 423)
<ul style="list-style-type: none"> <li>- Identificação das letras.</li> <li>- Lista de letras com sons.</li> <li>- Identificação das palavras.</li> <li>- Identificação de frases.</li> <li>- Identificação da estrutura gramatical.</li> <li>- Atribuição de significado às palavras e frases.</li> <li>- Estabelecimento de relações entre as frases do texto.</li> <li>- Utilização do conhecimento prévio para prever a informação e adivinhar o significado das palavras desconhecidas.</li> <li>- Realização de inferências com base no contexto da leitura e esquemas cognitivos do leitor.</li> </ul>	<ul style="list-style-type: none"> <li>- Reconhecimento de letras e integração de sílabas.</li> <li>- Codificação das palavras: as letras e sílabas se agrupam em palavras.</li> <li>- Codificação sintática: As palavras se relacionam entre si por regras sintáticas.</li> <li>- Codificação das proposições: As frases do texto incluem proposições elementares, que o leitor abstrai automaticamente.</li> <li>- Integração temática: O leitor não apenas lê e interpreta as sentenças, mas constrói um modelo coerente e integrado do texto. Isso implica o uso ativo de esquemas temáticos e formais (“texto experimental”, “narração” etc.).</li> </ul>

Fonte: Elaborado pelo autor a partir desses autores

#### 1.4.2 ELEMENTOS PARA A COMPREENSÃO

Os textos ou discursos (textuais, orais e audiovisuais, mas principalmente os dois primeiros), para serem compreendidos devem ter coesão e coerência. Portanto, todos os potenciais documentos textuais, orais e audiovisuais, objeto de indexação, devem ser coesos e coerentes em um momento. Vejamos

dois exemplos. Trata-se de um discurso textual e um discurso oral em que ocorre coesão, mas não a coerência.

Exemplo:

Texto (2)

Discurso Textual	Discurso Oral
<p>A Península Ibérica é descrita como uma das principais regiões mediterrâneas em relação ao número de espécies de anfíbios que apresenta. <b>Neste sentido</b>, os hospitais espanhóis dispõem de um profissional, preferencialmente alguém da enfermagem, para cuidar dos casos de cardiopatia crônica que não podem ser tratados nos ambulatórios devido à complexidade. Assim, se prestaria atenção especial à preservação e conservação e restauração do patrimônio bibliográfico brasileiro. No entanto, continuaríamos sem conhecer a fundo os três períodos religiosos dominantes de cada época na Índia, a saber, o período védico, o bramismo e o hinduísmo.</p>	<p><i>Duas pessoas estão conversando sobre a necessidade de planejamento familiar nos países em desenvolvimento e um deles (falante 1) pergunta:</i></p> <p><b>Falante 1.</b> O que você acha sobre o Vaticano ter obrigado os bispos brasileiros a retratarem-se por terem apoiado o uso do preservativo?</p> <p><b>Falante 2.</b> E em segundo lugar, não está ainda claro que as operações de combate em 2003 no Iraque foram um grande êxito dos Estados Unidos.</p> <p><b>Falante 1.</b> Não, eu te perguntava sobre o Vaticano e os bispos brasileiros.</p> <p><b>Falante 2.</b> Com certeza, o mais apropriado para ocupar o cargo é Jesse Wolfgang, Presidente da Federação de Diabéticos da Alemanha.</p> <p><b>Falante 1.</b> Continuo sem te entender.</p> <p><b>Falante 2.</b> Com certeza, concordou com a construção de casas para as pessoas que perderam suas casas nas inundações recentes.</p> <p><b>Falante 1.</b> Não estou compreendendo nada. Não entendo o que quer dizer.</p>

Como veremos, esses discursos cumprem a condição da coesão, mas não a da coerência. Portanto, para que um discurso seja aceito (entendido, compreendido) pelo leitor/ouvinte deve ser coeso, e também ser coerente. Vimos que apesar destas proposições estarem construídas sintaticamente corretas e terem uma boa coesão (ligação) entre elas, não há uma coerência linear entre as proposições e, portanto, nenhuma coerência global, de modo que não se consiga compreender o significado do discurso textual nem do oral apresentados como exemplo.

#### 1.4.2.1 COESÃO DISCURSIVA

A coesão tem sido estudada desde a década de 1970 por autores como Dressler (1972), Halliday e Hasan (1976) e Beaugrande e Dressler (1981) e Mederos Martín (1988) quanto à coesão em espanhol.

A coesão se manifesta na superfície do discurso, ou seja, sob a forma da construção. No processamento de fragmentos de textos longos, muitas vezes costuma-se utilizar mecanismos para a reutilização, a modificação ou a compreensão das estruturas e dos padrões utilizados anteriormente. Estes mecanismos, que contribuem tanto para estabilizar o sistema como para poupar esforço de processamento, são os seguintes (BEAUGRANDE; DRESSLER, 1981, p. 90). Os exemplos foram preparados para esclarecer cada um dos mecanismos enumerados pelos autores:

- ⇒ Repetição: Consiste na reutilização direta de elementos ou de padrões formalmente idênticos.  
*A temperatura média da Terra é agora de 0,7 graus Celsius mais elevada do que um século atrás. Essa temperatura média da Terra de 0,7 continuará aumentando se continuarmos...*
- ⇒ Repetição parcial: permite a transcategorização de um elemento usado anteriormente em outro tipo de elemento diferente (por exemplo, um verbo como “andar”, pode tornar-se um substantivo como “os andares”).*Os estudiosos das universidades e do Conselho Superior de Investigações Científicas de Valência podem cobrar em breve. Esses trabalhadores da Administração Pública além dos seus salários exigem...*
- ⇒ Paralelismo: a repetição é construída sobre uma estrutura enriquecida pelo aporte de novos elementos.  
O experimento confirmou nossas expectativas, reafirmou nosso método e consolidou nossa linha de pesquisa.
- ⇒ Paráfrase: Repetição do mesmo conteúdo, mas transmitidos através de expressões linguísticas diferentes.  
*A fonte de informação utilizada são os relatórios sobre A renda nacional da Espanha e sua distribuição regional, preparados pelo Banco de Espanha. O parecer da instituição monetária central do país foram consultados para estabelecer...*
- ⇒ Pré-formas: permite substituir elementos independentes de conteúdo por formas dependentes menores. Ou seja, certas palavras (pronomes, demonstrativos, possessivos, sintagma nominal, números, advérbios como ele, ela, esta, se, ali, primeira etc.) assumem o significado de uma determinada parte do discurso, substituindo os substantivos e os sintagmas nominais com os que estabeleceram uma relação de correferência. Quando essas palavras aparecem depois da expressão correlata denomina-se anáfora e, quando aparecem antes, catáfora.  
Exemplo de anáfora:

Existem diferenças entre a indexação manual e a indexação automática. Essa é executada por um programa e a primeira por uma pessoa.

O pronome “essa” tem o significado de “indexação automática”, enquanto “a primeira” refere-se a “indexação manual”.

Exemplo de catáfora:

Ela defende que a desertificação é a degradação das terras nas zonas áridas, semi-áridas e subúmidas causada por vários fatores. Monique Maingued já indicou várias vezes que esses fatores são as variações climáticas e as atividades humanas, principalmente.

No entanto, aqui o pronome ‘Ela’ tem o significado de algo que vem depois, nesse caso, “Monique Maingued”.

⇒ Elipse: omissão de alguma das expressões superficiais originais sem distorcer o significado.

A aldeia tem duas igrejas, a Igreja de Santa Margarita, interna e a de San Leandro, externa.

Omite-se ‘Igreja’ ao dizer ‘a de San Leandro’

⇒ Conexão: É o mecanismo mais óbvio para a sinalização das relações entre as frases. Os quatro processos principais são: a conjunção, a disjunção, a adversativa e a subordinação. A ligação é realizada através de *conectores* ou *marcadores* que são unidades de superfície que unem as proposições. Entonação: Nos textos falados permite marcar a importância ou a novidade do conteúdo linguístico.

Acabamos de observar que a ligação textual é realizada por meio de *marcadores textuais* (também chamados de *conectores* ou *conectivos*). São unidades superficiais que norteiam as relações entre as proposições como um todo, ou seja, a comunicação. Como visto no exemplo anterior (Texto 2) as sequências de frases podem ser conectadas sem serem coerentes, de modo que a conexão não é uma condição suficiente para a aceitação do discurso. Para Van Dijk (1998, p. 83) deve-se apresentar uma série de condições para que se produza a ligação semântica entre as frases e as sentenças; que os fatos denotados pertençam a mundos relacionados; que se trate de feitos compatíveis; que os conceitos envolvidos não estejam muito distantes; e, finalmente, menciona que deveria haver um tópico oral, ou seja, um tema que vai unindo as proposições que compõem o discurso.

A questão dos conectores produziu uma abundante bibliografia. Numa revisão pequena, identificamos várias monografias, inúmeros artigos e muitas teses sobre esse assunto. E, cada autor que o estuda propõe uma classificação, mas na nossa opinião, uma das mais completas é:

## Quadro 9 – Classificação dos marcadores de discurso

<b>ESTRUTURADORES DA INFORMAÇÃO</b>	COMENTADORES	Pois, pois bem, assim etc.
	ORDENADORES	Em primeiro lugar, em segundo lugar; por um lado/ por outro lado, etc.
	DIGRESSORES	Na verdade, a propósito, etc.
	CONECTORES ADITIVOS	Também, à parte, mesmo assim, etc.
<b>CONECTORES</b>	CONECTORES CONSECUTIVOS	Portanto, conseqüentemente, daí, então, pois, assim, etc.
	CONECTORES CONTRARGUMENTATIVOS	Por outro lado, embora, não obstante, etc.
	REFORMULADORES EXPLICATIVOS	Ou seja, isto é, a saber, etc.
<b>REFORMULADORES</b>	REFORMULADORES DE RETIFICAÇÃO	Melhor dizendo, melhor ainda, etc.
	REFORMULADORES DE DISTANCIAMENTO	De qualquer caso, em todo caso, de todo modo, etc.
	REFORMULADORES RECAPITULATIVOS	Em suma, em conclusão, finalmente, etc.
	<b>OPERADORES ARGUMENTATIVOS</b>	OPERADORES DE REFORÇO ARGUMENTATIVO
OPERADORES DE CONCRETIZAÇÃO		Por exemplo, em particular, etc.
DE MODALIDADE EPISTÊMICA		Claro, desde logo, pelo visto, etc.
DE MODALIDADE DEONTICA		Bem, bom, etc.
<b>MARCADORES ORAIS</b>		FOCADORES DA ALTERIDADE
	METADISCURSIVOS ORAIS	Bom, este, etc.

Fonte: MARTÍN ZORRAQUINO; PORTOLÉS LÁZARO, 1999, p. 4081.

#### 1.4.2.2 A COERÊNCIA DISCURSIVA

A coerência é um elemento que deve estar presente no discurso (textual ou oral) para que o discurso aconteça, caso contrário teríamos um discurso sem sentido, como no exemplo do Texto 2. Em suma, a coerência surge de uma continuidade de sentido que leva à compreensão global do discurso e, fornece, assim, a intenção comunicativa. A coerência se manifesta na interrelação entre as partes, na presença de um tema a ser desenvolvido gradualmente e, finalmente,

quando o destinatário tem uma ideia global do que é visualizado (lido ou visualizado) e ouvido (oral).

Charolles (1978) especificou quatro meta-regras para conseguir a coerência textual: regra da recorrência, que indica a continuidade de sentido por meio da repetição, a anáfora e catáfora, a substituição léxica, a paráfrase e a elipse; regra da progressão temática que orienta o progresso gradual da comunicação; regra da não-contradição, que são os textos coerentes sem elementos que se contradizem entre si; e, finalmente, a regra de relação ou da conexão para marcar a coerência entre as frases pelas relações semânticas e os marcadores.

A coerência textual ou discursiva é construída da parte para o todo, isto é, na mesma frase, entre umas frases e outras e entre uns parágrafos e outros. Portanto, primeiro vai se tecendo a coerência local, entre as frases e, depois, os parágrafos relacionados semanticamente configuram a coerência global ou textual.

Os indexadores só analisam os discursos (sejam textuais, orais ou audiovisuais) coerentes. E o principal elemento que permeia a coerência do discurso é a existência de um tema como veículo comunicativo entre emissor e receptor. Ou seja, estar na presença de um discurso ou texto com coerência global é como estar diante de um discurso com um significado compreensível pelo receptor.

O produto da indexação de documentos pode ser um ou vários conceitos gerais como por exemplo: Medicina; Língua espanhola; Novela espanhola; Biologia (indexação por assunto); ou por conceitos mais específicos pertencentes a uma linguagem especializada de uma área temática (indexação com descritores). Assim, no âmbito da Medicina, podemos encontrar descritores como: Laparoscopia; Trisomia 21; Apendicectomia; Rinite; etc Para isso, vamos analisar diferentes teorias linguísticas que apareceram na Europa desde a década de 1950, embora, em alguns casos, as bases são anteriores. Mas primeiro vamos olhar para este exemplo:

Exemplo:

O resultado da análise de um mesmo texto seria diferente se fosse indexado usando uma lista de assunto (indexação de assuntos) ou um tesauro (indexação com descritores):

Indexação por assunto:

Espanha-História-S. XX                      ► cabeçalho de assunto – subcabeçalho de lugar - subcabeçalho cronológico

Indexação com descritores:

História Política, Políticos                ► Descritores temáticos

Espanha    ► Descritores topográficos

## Século XX ► Descritores cronológicos

Monarquia Parlamentar;

Diretório Militar

Diretório Civil;

II República; Guerra Civil Espanhola ;

Ditadura

franquista; Democracia

## ► Identificadores

## 1.4.2.2.1 TÓPICO FRASAL

Para falar de coerência textual global deve-se ter a coerência da frase e a coerência entre os parágrafos. Vamos, portanto, do menor para o maior. Vilém Mathesius fundador do *Círculo Linguístico de Praga*, em meados de 1920, foi o pioneiro da teoria dos elementos informativos de frase. Dos trabalhos de Mathesius (1928) e de seus seguidores, principalmente Firbas (1964 e 1971), nasceu a noção de *tema* e *rema* como elementos articuladores das frases. Também conhecidos por *tópico* e *comentário* ou *comentários* de outras tendências linguísticas.

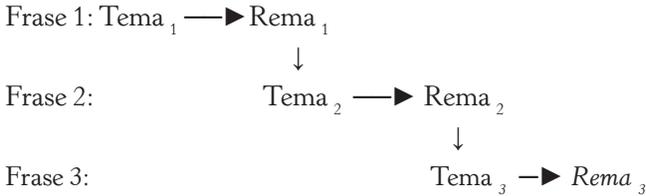
Na Escola de Praga, trabalharam a função comunicativa da linguagem a partir da perspectiva funcional da frase, para explicar que o desenvolvimento informativo de texto tem como ponto de partida a introdução de um elemento, que serve de início e eixo da comunicação. Assim, defenderam que a carga informativa que transmitem as frases é baseada em dois componentes das frases: em primeiro lugar, o *tema* (elementos estáveis compartilhados pelos interlocutores, ou seja, o velho); e, por outro lado, a *rema* (novos elementos que estão contribuindo, ou seja, o novo).

Esta teoria do *tema e rema* foi adotada pela linguagem textual a partir dos anos 1970. Em particular, František Danes (1974), outro linguista da Escola de Praga, articulou essa dicotomia como o elemento que fornece a coerência textual. Para ele, a noção de “progressão temática” é entendida como o desenvolvimento temático gradual do discurso que ultrapassa o âmbito da frase. Assim, a teoria da *progressão temática* de Danes permite a detecção dos *temas* (conhecido pelos interlocutores) e das *remas* (o novo) de cada uma das frases. Portanto, o tópico frasal é uma das expressões contidas na frase.

A teoria da progressão temática de Danes foi esboçada por Combettes (1983, p. 90) conforme ilustrado aqui. Os exemplos que acompanham cada tipo foram preparados para facilitar a compreensão.

**PROGRESSÃO TEMÁTICA LINEAR**

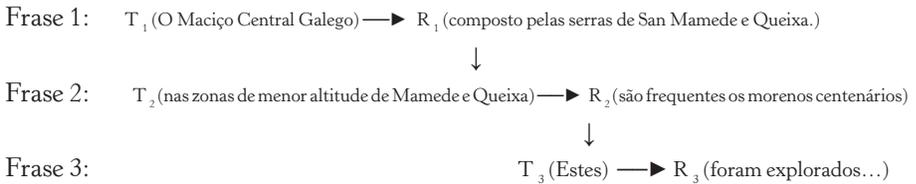
A rema da primeira frase é a origem do tema da frase seguinte e assim por diante.



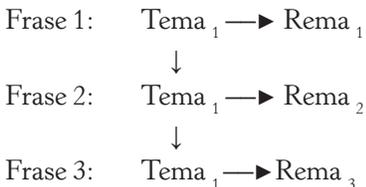
A seta horizontal ( $\longrightarrow$ ) indica a ligação entre o tópico frasal e a rema; enquanto que as setas verticais ( $\downarrow$ ) expressam a relação contextual entre as frases.

**Exemplo**

O Maciço Central Galego é composto pelas serras de São Mamede e Queixa. Nas zonas de menor altitude de Mamede e Queixa são comuns os morenos centenários. Estes foram explorados coletivamente pelos moradores das aldeias vizinhas.

**PROGRESSÃO COM TEMA ÚNICO OU CONSTANTE**

Esta progressão temática é mais comum do que linear, porque sua estrutura é mais simples. Na progressão com tema constante é repetido em cada frase ou sentença o ponto de partida, ou seja, o tema, enquanto incorporam remas novas.



Exemplo:

Quando falamos de *alcoolismo*, nos referimos a uma pessoa dependente do álcool, cujo consumo proporciona problemas físicos, psicológicos e/ou sociais. O *alcoolismo* gera sérias dificuldades que afetam tanto o consumidor como seu meio. A dependência do álcool é alcançada pela interrelação de fatores relacionados com a personalidade individual, como o ambiente familiar, educacional, trabalhista, cultural, social, etc. O *alcoolismo* aparece, portanto, como um dos principais desafios colocados pelo atual Plano Nacional de Saúde.

**PROGRESSÃO COM TEMAS DERIVADOS**

É uma progressão temática mais complexa do que as precedentes. Inicialmente, se apresenta um hipertema que aponta para os subtemas que propiciam a evolução temática progressiva.

Hipertema

Frase 1: Tema<sub>1</sub> —► Rema<sup>1</sup>

Frase 2: ... .. Tema<sub>2</sub> —► Rema<sub>2</sub>

Frase 3: ... .. Tema<sub>3</sub> —► Rema<sub>3</sub>

Exemplo

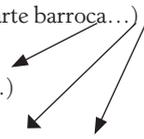
*A arte barroca foi desenvolvida entre os séculos XVII e XVIII e logo se tornou um veículo de propaganda para a Igreja da Contra-Reforma, dos estados absolutistas e da burguesia protestante. A pintura barroca reflete a realidade com limites imprecisos, as formas que vêm e vão, os objetos em primeiro plano intrascendente, os escorços e as atitudes violentas que dão à obra grande dinamismo. A escultura barroca caracteriza-se por sua força e imponência, seu movimento de composição, seu dinamismo, suas composições diagonais, sua expressão e seu tratamento do vestuário. A arquitetura barroca é resumo e paradigma do espírito da “civilização do Barroco”, onde se adotam as linhas curvas antes das retas para gerar uma maior agilidade e expressividade.*

Frase 1: Hipertema (a arte barroca...) —► Rema (se desenvolve nos séculos XVII e XVIII...)

F2: Tema<sub>1</sub> (a pintura...) —► Rema<sub>1</sub> (reflete a realidade...)

F3: Tema<sub>2</sub> (a escultura...) —► Rema<sub>2</sub> (se caracteriza por sua força...)

F4: Tema<sub>3</sub> (a arquitetura...) —► Rema<sub>3</sub> (é resumo e paradigma...)

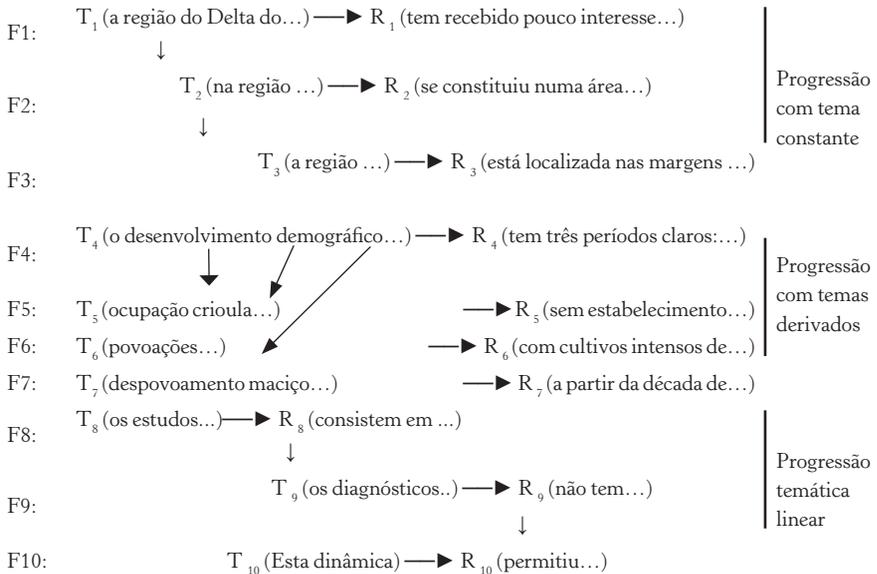


**PROGRESSÃO COMBINADA**

Em ambos os textos aparecem a progressão linear, a progressão constante e a progressão com temas decorrentes, nesta ou em outra ordem.

Exemplo:

*A região do delta do Paraná é uma área que tem recebido pouco interesse para os pesquisadores, embora a área tornou-se relativamente grande em importância econômica e social. A região está situada às margens de uma vasta área metropolitana de Buenos Aires. O desenvolvimento demográfico tem três períodos claros: ocupação crioula sem estabelecimento permanente de importância, baseada na extração direta de recursos naturais; povoações com fruticultura intensa por pequenas unidades familiares e, um êxodo maciço a partir da década de 1950 devido ao declínio da produção de frutos. Os estudos referentes à essa área, consistem de diagnósticos sincrônicos das últimas décadas. Os diagnósticos sincrônicos não abordaram a trajetória histórica da dinâmica social. Essa dinâmica social permitiu uma economia flutuante.*



Em suma, enquanto a rema promove a progressão da comunicação, o tema tece a rede semântica que transmite a coerência do discurso.

### 1.4.2.2.2 TEMA TEXTUAL

O tema textual ou discursivo tem sido chamado pelos linguistas de diferentes maneiras, assim, tem sido utilizadas expressões como *alusão*, *plano global*, *tópico do discurso* ou *macroestrutura textual*.

Vimos como identificar o tópico frasal, agora veremos como se apresenta o tema textual, ou seja, nas unidades mais amplas, os parágrafos ou grandes trechos. O *tema* do texto transmite a mensagem desejada, é a informação principal frente a secundária, é do que se trata o texto. Às vezes, o *tema* de uma passagem ou de uma conversa se reduz a uma frase ali presente, mas na maioria das vezes deve fazer um esforço de abstração para representá-lo. E como já mencionado, a indexação consiste nisso, na localização e representação do assunto ou assuntos tratados no objeto analisado.

No início deste capítulo, falamos sobre os diferentes modelos da *linguística textual* de acordo com as diferentes concepções de linguistas. Já apontamos que um dos modelos mais conhecidos é o de Van Dijk (1972; 1977; 1978). O pesquisador desenvolveu toda uma teoria em que designou denominações e propôs regras para obter o *tema textual*.

Van Dijk (1998, p. 200) observa que um conceito ou uma estrutura conceitual (uma proposição) pode se tornar o tema do discurso se hierarquicamente organiza a estrutura conceitual da sequência de frases. Portanto, a noção abstrata de *tema* pode ser explicitada em estruturas de frase e, também deve estar relacionada com o restante das frases como um todo. Para designar o *tema* do texto, o linguista designou a noção de *macroestrutura*. Entende-se por *macroestruturas* as representações semânticas que visam a organização da informação por meio de processos redutores de informação, chamadas de *macroregras*. As macroregras destroem a informação não significativa, generalizam e incluem informação em macro-sentenças de um nível superior. São geradas pelo leitor ou ouvinte, porém podem estar presentes no texto ou no diálogo. Para obter as macroestruturas, o autor aplica quatro regras: omissão, seleção, generalização e construção ou integração.

Para entender as macroregras, vamos seguir literalmente as explicações de Van Dijk, em seu livro *Ciencia del texto* (1996, p. 54). Como ponto de partida, dizemos que as regras de omissão e seleção são processos de anulação de informação e as de generalização e integração são de substituição de informação, e podem ser formuladas da seguinte forma:

- (I)  $\langle \alpha, \beta, \gamma \rangle \rightarrow \beta$   
 (II)  $\langle \alpha, \beta, \gamma \rangle \rightarrow \delta$

As quatro macroregras devem respeitar o princípio da implicação semântica, ou seja, que cada macroestrutura conseguida através de macroregras deriva de microestruturas, ou seja, de proposições ou de outra macroestrutura inferior.

### OMISSÃO

Significa que toda a informação de pouca importância e que não seja essencial pode ser omitida. Quando temos uma série de proposições, podemos eliminar algumas se estas não têm uma “função” para o texto.

### SELEÇÃO

Essa regra exige que a proposição selecionada contenha, de alguma forma, a informação das proposições omitidas.

### GENERALIZAÇÃO

Essa regra omite informações essenciais, mas faz de maneira que se percam (como na omissão). São omitidos elementos essenciais de um conceito ao substituir uma proposição por outra nova mais geral, de acordo com o esquema  $\langle \alpha, \beta, \gamma \rangle \rightarrow \delta$

- (zz) (i) Os bibliotecários selecionam assuntos.  
 (ii) Os bibliotecários catalogam os assuntos.  
 (iii) Os bibliotecários difundem a informação.

Estas três proposições podem ser substituídas por uma nova já que as proposições (zz) implicam conceitualmente (yy). Substitui uma série de conceitos por um sobreconceito superior.

- (yy) Os bibliotecários executam processos técnicos documentários.

### CONSTRUÇÃO OU INTEGRAÇÃO

Assemelha-se à regra de seleção, mas também opera de acordo com o esquema  $\langle \alpha, \beta, \gamma \rangle \rightarrow \delta$ , de modo que a informação passa a ser substituída por uma nova e não é omitida nem selecionada. O texto pode incluir uma série de aspectos (condições normais, circunstâncias, componentes, conseqüências,

etc.), para que juntos formem uma ideia global, produto da compreensão como no texto a seguir, que localizamos num artigo de jornal:

*Na Galiza foram encontrados caracóis-fêmea Nucella lapillus com um pênis milimétrico; enquanto que em Huelva foram encontrados caracóis Bolinas brandaris com um pênis que chega a vagina. No rio Ebro surgiram carpas macho com ovários. E em estuários contaminados do Mediterrâneo foram encontradas salmonetes com problemas semelhantes. Os resultados de numerosas investigações feitas desde 1970 até à data nas diferentes partes do mundo e estudando diferentes espécies sugerem que a causa é o uso de tributilo de estanho nas tintas antiincrustantes usadas para proteger os cascos dos barcos de algas e crustáceos.*

O texto acima pode ser transformado em uma frase no presente que resulta da compressão do mesmo pelo leitor/ouvinte. “A contaminação química provoca distúrbios hormonais nos seres vivos”.

Continua apontando em seu “*Ciencia del texto*” que o tema não tem porque ser nomeado no texto. Se aparece, estamos diante de uma palavra temática ou frase temática e ambas têm a importante função cognitiva de colocar o leitor ou ouvinte em condições de fazer uma macrointerpretação “correta” do texto. Este recebe uma ajuda para sua “suposição” do que pode se tratar o texto. Em resumo, o linguista diz que estas regras permitem-nos decidir de uma forma mais ou menos exata o que é principal e secundário, dependendo do contexto de cada texto; apesar de que diferentes falantes podem fazer diferentes aplicações das regras. Para uma pessoa um texto “significa” globalmente M; enquanto que para outro pode significar M’I, dependendo de vários fatores, tais como interesse, conhecimentos, desejos ou opiniões. De qualquer forma, o falante/escritor utiliza recursos para fazer a macroestrutura correta ou que tem intenção de expressar por meio de sumários, títulos ou frases temáticas. Além disso, o leitor/ouvinte também tem um modelo cognitivo do falante/escritor para tornar possível a comunicação.

Agora vamos aplicar essa teoria redutora de informação de Van Dijk a alguns exemplos para ver como funciona:

Exemplo:

Quadro 10 – Estratégias redutoras de informação

<p><b>SEQUENCIA DE PROPOSIÇÕES</b></p>	<p><b>CASO 1</b> A chegada dos fenícios pressupõe o surgimento das técnicas como o filigrana e a granulação. Desaparecem as peças maciças grandes e a maioria tem uma base oca. Para a filigrana empregam-se tipos de fios diferentes isolados ou em combinação.</p>	<p><b>CASO 2</b> Os bonsais podem ser afetados pela clorose, que é a perda da cor verde nas folhas; enquanto o ataque de fungos causa o mofo branco ou o apodrecimento das raízes, entre outros males.</p>	<p><b>CASO 3</b> - Juan, onde está?  - Acabo de chegar à estação rodoviária de Madri. Vou de metrô até o aeroporto. Espero chegar a tempo, uma vez que o vôo para o Rio de Janeiro sai em uma hora.</p>
<p><b>MACROREGRAS</b></p>	<p>↓ [Omissão- Seleção]</p> 	<p>↓ [Generalização]</p> 	<p>↓ [Integração]</p> 
<p><b>MACROES-TRUTURAS RESULTANTES</b></p>	<p>↓ A chegada dos fenícios pressupõe o surgimento das técnicas filigrana e granulação.</p>	<p>↓ Pragas nos bonsais.</p>	<p>↓ Juan está viajando.</p>
<p><b>INDEXAÇÃO</b></p>	<p>↓ - FENICIOS - FILIGRAMA - GRANULAÇÃO</p>	<p>↓ - PRAGAS - BONSAIS</p>	<p>↓ - VIAGENS - MEIOS DE TRANSPORTE</p>

Fonte: Elaborado pelo autor.

Para complementar o exposto, podemos ver os exemplos de Moraes, Guimarães e Guarido (2007) para análise de conteúdo temático de textos narrativos de ficção.

Explicada uma maneira de identificar e, portanto, de selecionar o tema de um texto, vejamos agora um exemplo para esclarecer as diferenças entre tópico frasal, tema do parágrafo e tema textual.



**TÓPICO FRASAL, TEMA DO PARÁGRAFO E TEMA TEXTUAL**

A ferrovia foi inaugurada com a linha aberta por George Stephenson Entre Stockton e Darlington em 1825.

Tópico frasal:  
**Ferroviária**  
**George Stephenson**

Posteriormente, em 1829 Stephenson e seu filho alcançaram 47 km / h com sua locomotiva Rocket.

Tópico frasal:  
**Velocidade**  
**locomotiva**  
**Locomotiva the**  
**Rocket**

Tema parágrafos:  
**Transporte**  
**terrestre**

Outro meio de transporte importante é o automóvel como conhecemos atualmente, foi inventado na Alemanha por Carl Benz em 1886.

Tópico frasal:  
**Automóvel**  
**Carl Benz**

No entanto, o transporte mais seguro de todos é o aéreo. Os avanços da navegação aérea, das telecomunicações e das facilidades eletrônicas permitiram que a aviação tenha progredido muito. O primeiro vôo de um avião foi realizado em 1904 pelos Irmãos Wright e cem anos depois, o homem pode ir para o espaço e visitar outros planetas com naves não tripuladas.

Tópico frasal:  
**Avanços da aviação**

Tópico frasal:  
**Vôos**  
**Irmãos Wrigth**

Tema parágrafo:  
**Transporte aéreo**

Tema textual:  
**MEIOS DE TRANS-**  
**PORTE**

Na atualidade, o transporte marítimo é utilizado por mais de três quartos do comércio mundial. Sua internacionalização nos últimos anos, têm se manifestado com as mudanças de propieades e gestão de navios, a criação de novos registro, o progresso técnico e a contratação de tripulações de diferentes culturas. Entretanto, o transporte fluvial não sofreu grandes mudanças nas últimas décadas e ainda é usado para o transporte de mercadorias pesadas (granéis sólidos e líquidos) que não exigem transporte rápido.

Tópico frasal:  
**Transporte marítimo**

Tópico frasal:  
**Internacionaliza-**  
**ção do transporte**  
**marítimo**

Tema parágrafo:  
**Transporte por**  
**água**

Tópico frasal:  
**Transporte fluvial**

Tópico frasal:  
**Uso transporte fluvial**

Para terminar, esquematizamos no quadro a seguir a ligação entre os diferentes elementos do processo cognitivo com a indexação.

Quadro 11 – Relação entre o processo cognitivo e a indexação

INDEXAÇÃO		PROCESSO COGNITIVO
<ul style="list-style-type: none"> <li>Conhecimentos prévios do indexador:                             <ul style="list-style-type: none"> <li>- Teoria e prática da indexação</li> <li>- Contexto da indexação a executar (necessidades dos usuários, políticas de indexação, condicionantes externos, etc.)</li> <li>- Assunto a ser indexado (História, Física, Direito...)</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>Memória de curto prazo</li> </ul>
<ul style="list-style-type: none"> <li>Percepção sensorial da informação:                             <ul style="list-style-type: none"> <li>- Visual: texto, objeto físico, imagem</li> <li>- Auditiva: palavra, música, som ambiente</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>Memória sensorial:                             <ul style="list-style-type: none"> <li>- Visual (memória icônica)</li> <li>- Auditiva (memória ecóica)</li> </ul> </li> </ul>
<ul style="list-style-type: none"> <li>Análise do documento</li> </ul>	Aplicação dos conhecimentos prévios Reconhecimento de tipologias textuais  Identificação de temas	<ul style="list-style-type: none"> <li>Memória de curto prazo</li> <li>Memória de curto prazo</li> <li>Memória de longo prazo</li> <li>Memória de curto prazo</li> <li>MLP</li> <li>Compreensão</li> <li>Estratégias redutoras de informação</li> <li>Tema-remã</li> </ul>
<ul style="list-style-type: none"> <li>Conversão da linguagem natural para a linguagem controlada</li> </ul>		<ul style="list-style-type: none"> <li>Memória de curto prazo</li> <li>Memória de longo prazo</li> </ul>
<ul style="list-style-type: none"> <li>Armazenamento da indexação resultante na base de dados</li> </ul>		<ul style="list-style-type: none"> <li>Memória de curto prazo</li> <li>Memória de longo prazo</li> </ul>

Fonte: Elaborado pelo autor.

## 2 A INDEXAÇÃO

O processo técnico documentário (FIG. 8) é o conjunto de operações realizadas para a seleção, a aquisição, o registro e o tratamento dos documentos a fim de possibilitar seu armazenamento e, posterior recuperação.

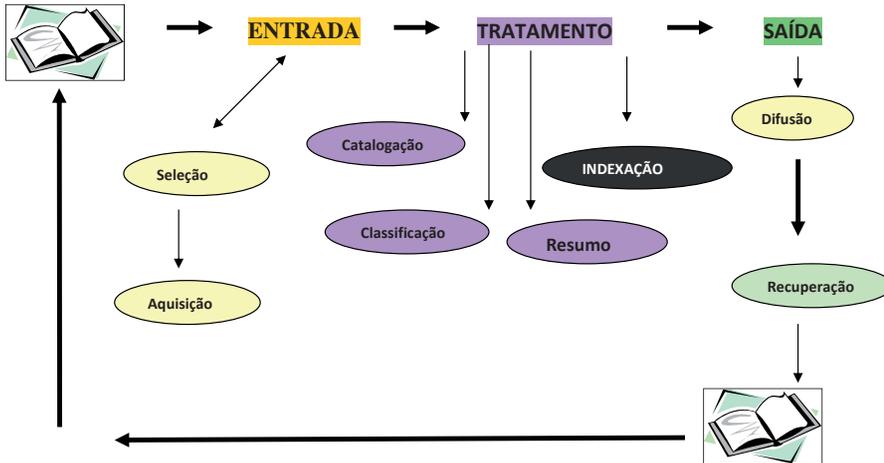


Figura 8 – Esquema básico do Processo Técnico Documental

Fonte: Elaborado pelo autor.

Em geral, a entrada dos documentos em um centro ocorre por meio de duas etapas sucessivas: a seleção e a aquisição. A política de aquisições deve ser concebida em relação ao pressuposto, aos recursos e aos serviços prestados. Os fundos vão aumentando pela compra ou pela troca de documentos entre as instituições ou por doação de fundos privados. Além disso, como o nome sugere, selecionar é escolher a documentação que deve ser incorporada. Portanto, a seleção do material e sua posterior aquisição é a primeira etapa do processo que conduz a constituição dos fundos documentários. Normalmente, cada instituição tem pessoal encarregado para essa tarefa e, uma questão chave no processo de seleção é verificar as necessidades, as características e as preferências dos usuários.

Outra fase do processo documentário é o tratamento técnico que recebem os documentos para que possam ser utilizados. Este tratamento é dividido em análise da forma e análise do conteúdo. A análise da forma de um documento também é conhecida como descrição bibliográfica ou catalogação, enquanto que a análise de conteúdo inclui a classificação, o resumo e a indexação.

A última fase do processo documentário é a saída da informação. Todas as operações desenvolvidas na fase de entrada e de tratamento da informação têm um objetivo principal: a divulgação da informação. A razão das operações descritas é a divulgação da informação que foi selecionada, adquirida ou recebida e analisada anteriormente.

## 2.1 CONCEITO DE INDEXAÇÃO

As origens da indexação se encontram nas tarefas realizadas pelos antigos escribas da Mesopotâmia. Naquela época, começaram a ter salas para a cópia das tábuas de argila, a elaboração de etiquetas e a conservação das placas. Os textos eram armazenados em prateleiras de madeira, colocados em nichos nas paredes ou eram dispostos em caixas de madeira. Para saber o que continham, colocavam uma pequena etiqueta anexada na lateral, onde escreviam o conteúdo dos documentos. Nessas tarefas rudimentares, vemos os primeiros passos do que hoje conhecemos como a indexação (GIL LEIVA; RODRÍGUEZ MUÑOZ, 1996, p. 53). Posteriormente, nas bibliotecas, foram aparecendo os catálogos para controlar e localizar os livros numa coleção. Aos catálogos tradicionais de autor ou título se juntaram o catálogo de assunto no final do século XIX, com a finalidade de localizar todos os livros sobre um determinado assunto.

Desde a década de 1970 até a atualidade, foram realizados muitos esforços para estabelecer os fundamentos teóricos da indexação a partir de perspectivas cognitivas ou linguísticas (JONES, 1976; BORKO, 1977; FUGMANN 1979; FARROW, 1911; BERTRAND; CELLIER, 1995, AMAR; 2000, ANDERSON; PÉREZ-CARBALLO, 2001), para complementar a teoria com a prática (LANCASTER, 1996; FUGMANN, 1993) ou para propor regras para a indexação dos documentos (COOPER, 1978; FROHMANN, 1990).

Antes de entrarmos no processo de indexação é importante mencionar que qualquer objeto pode ser indexado, ou seja, reduzido a representações conceituais que facilitem seu armazenamento e recuperação em bases de dados. Se aceitarmos essa premissa, podemos indexar o texto impresso ou o digital, áudio (música, discurso ou som ambiente), imagem fixa (fotografia, cartaz, quadro, etc.), imagem em movimento (filme, spot de publicidade, etc.), obra de arte (escultura, cerâmica), arquitetura (ponte, fábrica, igreja, etc.) ou um produto industrial (selo). Vejamos agora a indexação de vários objetos realizada em VRA (Visual Resources Association – Associação de Recursos Visuais) e, mais especificamente, em seu projeto CCO (Cataloging Cultural Objects – Catalogação de Objetos Culturais)<sup>1</sup>:

<sup>1</sup> A indexação dos objetos estava disponível no ano de 2008 no sítio web da Fundação VRA ([http://cco.vrafoundation.org/index.php/toolkit/index\\_of\\_examples/](http://cco.vrafoundation.org/index.php/toolkit/index_of_examples/)), entretanto na atualidade os exemplos de indexação estão presentes sem as figuras.

	<p><b><u>DATE</u></b></p> <p>Display: c. 625-600 BCE</p> <p><b><u>DESCRIPTION</u></b></p> <p>Display: Two handled neck-amphora with five animal friezes plus neck decoration from Rhodes</p> <p><b><u>LOCATION</u></b></p> <p>Display: London: British Museum name: Rhodes</p> <p><b><u>MATERIAL</u></b></p> <p>Display: painted terracotta</p> <p><b><u>MEASUREMENTS</u></b></p> <p>Display: 35 cm (H)</p> <p><b><u>STYLEPERIOD</u></b></p> <p>Display: Early Corinthian, Animal Style</p>	<p><b><u>SUBJECT</u></b></p> <p>Display: animals; birds; friezes (ornamental bands); zoophori</p> <p>Index:</p> <p>subject:</p> <p>term: animals</p> <p>type: descriptiveTopic</p> <p>vocab: ICONCLASS</p> <p>term: birds</p> <p>type: descriptiveTopic</p> <p>vocab: ICONCLASS</p> <p>term: friezes (ornamental bands)</p> <p>type: descriptiveTopic</p> <p>vocab: AAT</p> <p>term: zoophori</p> <p>type: descriptiveTopic</p> <p>vocab: AAT</p> <p><b><u>TECHNIQUE</u></b></p> <p>Display: painting</p> <p><b><u>TITLE</u></b></p> <p>Display: Corinthian Black-figure neck amphora</p>
	<p><b><u>AGENT</u></b></p> <p><b>Display:</b> Francisco Goya (Spanish, 1746-1828) name: Goya, Francisco ; vocab: ULAN dates: earliestDate: 1746 latestDate: 1828</p> <p><b><u>DATE</u></b></p> <p><b>Display:</b> first published in 1799; 2nd edition published c. 1803</p> <p><b><u>LOCATION</u></b></p> <p><b>Display:</b> Herbert F. Johnson Museum of Art, Cornell University, Ithaca, NY, USA (63.108); name: Ithaca ; type: geographic ; vocab: TGN</p> <p><b><u>MATERIAL</u></b></p> <p><b>Display:</b> ink on paper</p> <p><b><u>STYLEPERIOD</u></b></p> <p><b>Display:</b> Romantic ; vocab: AAT</p>	<p><b><u>SUBJECT</u></b></p> <p><b>Display:</b> satires; caricatures term: satires ; type: conceptTopic ; vocab: AAT term: caricatures ; type: conceptTopic ; vocab: AAT</p> <p><b>Display:</b> dreams term: dreams ; type: conceptTopic ; vocab: AAT</p> <p><b><u>TECHNIQUE</u></b></p> <p><b>Display:</b> etching and aquatint</p> <p><b><u>DESCRIPTION</u></b></p> <p>description: Man, asleep at a table, surrounded by demonic-looking animals and birds. Originally intended as the frontispiece for the series.</p> <p><b><u>INSCRIPTION</u></b></p> <p><b>Display:</b> El Sueño de la Razon Produce Monstruos</p> <p><b><u>MEASUREMENTS</u></b></p> <p><b>Display:</b> 7 1/8 x 4 3/4 in, (18.1 x 12.2 cm)</p> <p><b><u>TITLE</u></b></p> <p><b>Display:</b> El Sueño de la Razon Produce Monstruos</p>

Fonte: VRA Foundation (Visual Resources Association – Associação de Recursos Visuais) e, mais especificamente, em seu projeto CCO (Cataloging Cultural Objects – Catalogação de Objetos Culturais), 2008 - [http://cco.vrafoundation.org/index.php/toolkit/index\\_of\\_examples/](http://cco.vrafoundation.org/index.php/toolkit/index_of_examples/)

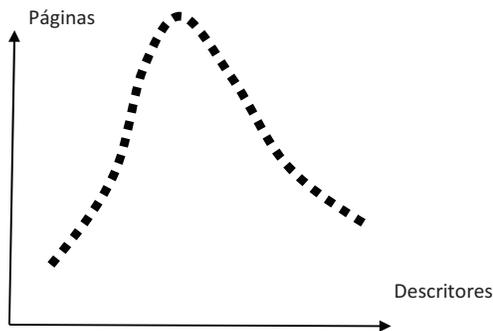
Da mesma forma, podemos indexar parágrafos, títulos, resumos, artigos de revistas, livros, relatórios, sons de guerra, da floresta, comerciais

de rádio e filmes. Além disso, em geral, o tamanho do objeto indexado não tem relação direta com o número de palavras-chave, assuntos ou descritores empregados em sua representação.

Tanto autores como algumas políticas de indexação de instituições têm tentado fazer a correspondência dos tamanhos de texto ou tipologias documentais com o número de termos de indexação atribuídos. Por exemplo, Lancaster (1991, p. 23, tradução de Elsa E. Barber) disse que a instituição deve formar uma faixa de termos para cada item, assim, por exemplo, o centro de informação de uma grande empresa pode estabelecer a seguinte política: para relatórios técnicos 10-15; para patentes 15-20; ou artigos de revista 5-10 termos e assim por diante. Por outro lado, as diretrizes para a indexação na base de dados ERIC recomenda aos indexadores que para uma monografia destinam-se 8-10 descritores; para um artigo crítico 6-8; para um poema, 3-4 descritores.

No entanto, o estudo das bases de dados de artigos científicos ou de imprensa não tem nenhuma relação entre tamanho e número de descritores. De qualquer modo, parece lógico que a indexação de um título gere menos termos que a de um resumo e, a de um resumo menos do que de um texto completo. Embora, quando se analisa um documento com título, resumo e texto completo, de início se produza um aumento proporcional entre número de páginas analisadas e número de conceitos selecionados, mas chega um momento em que a progressão da leitura/análise do documento não somam novos conceitos, ainda que possa variar em função do objeto indexado (FIG. 9).

Figura 9 – Relação entre páginas e descritores



Fonte: Elaborado pelo autor.



A análise das Bases de Dados do Conselho Superior de Pesquisas Científicas Espanhol IME (Biomedicina), ISOC (Ciências Sociais e Humanas), ICYT (Ciência e Tecnologia), todas indexadas por profissionais, permitiu detectar quatro realidades (GIL LEIVAY RODRÍGUEZ MUÑOZ, 1997, p. 162):

- Artigos com poucas páginas e poucos descritores (2 páginas e 2 descritores)
- Artigos com muitas páginas e poucos descritores (mais de 30 páginas e 3 descritores);
- Artigos com poucas páginas e uma quantidade importante de descritores (4 páginas e 9 descritores)
- Artigos com um número elevado de páginas e importante de descritores (mais de 30 páginas e 9 descritores).

Uma vez apresentados alguns aspectos básicos da indexação, vamos ao processo.

Na literatura sobre indexação é habitual encontrarmos vocábulos distintos para o mesmo conceito. Não indo muito longe, na mesma definição de *indexação* é surpreendente a variedade de verbos empregados para descrever essa ação: *reter, extrair, captar, resumir, descrever, caracterizar, escolher, analisar, identificar, traduzir, indexar, indicar, interpretar, enumerar, etc.*

#### Quadro 12 – Algumas definições de indexação

Analisar o conteúdo informacional dos registros do conhecimento e expressar o conteúdo informacional na linguagem do sistema de indexação.	Borko; Bernier (1978, p. 8)
Representação pelos elementos de uma linguagem documental ou natural, das noções resultantes da análise do conteúdo de um documento para facilitar sua localização	NF Z 47-102 1978
Descrever o conteúdo de documentos ou demandas documentárias para possibilitar a elaboração de estratégias de recuperação mediante conceitos ou assuntos.	García Gutiérrez (1984, p. 105)
Ação de descrever ou identificar um documento em relação ao conteúdo.	ISO 5963-1985
Identificar informação numa entidade de conhecimento (que seja texto ou não) e organizá-la para que esteja disponível num sistema de recuperação.	Cleveland, D.B.; Cleveland, A.D. (2001, p. 97)
Determinar o assunto temático dos documentos e expressar em índices (por exemplo, descritores, cabeçalhos de assunto, números de chamada, códigos de classificação ou índice) para tornar possível a recuperação temática.	Mai (2005, p. 599)

Fonte: Elaborado pelo autor

Como se observa no Quadro 12, para definir a indexação, várias vezes, faz-se referência ao documento como única fonte de análise. Isto significa efetuar uma delimitação tendenciosa do processo, uma vez que se deixa de lado a indexação do pedido do usuário ou da questão documentária.

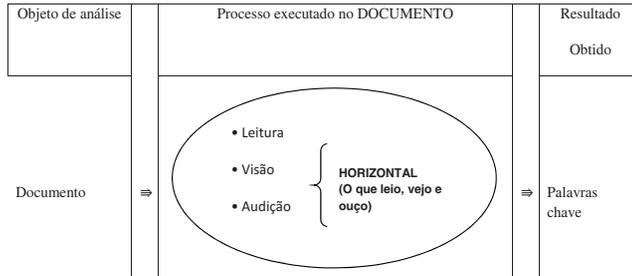


Figura 10 – Seleção de conceitos presentes no documento  
 Fonte: Elaborado pelo autor

Na análise do documento (leitura, audição ou visão) a seleção dos conceitos se realiza elegendo entre os conceitos presentes (FIG. 10) e em algumas ocasiões, também se recorre a conceitos implícitos (que não estão presentes, mas “flutuam” ou são mencionados de alguma forma no documento (FIG. 11).

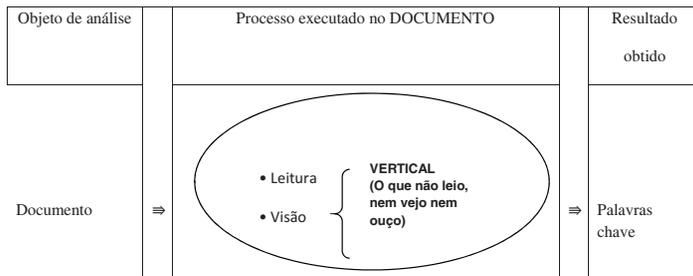


Figura 11 – Seleção de conceitos que não estão presentes no documento  
 Fonte: Elaborado pelo autor

Um processo similar se repete para indexar a necessidade de informação materializada em uma pergunta.

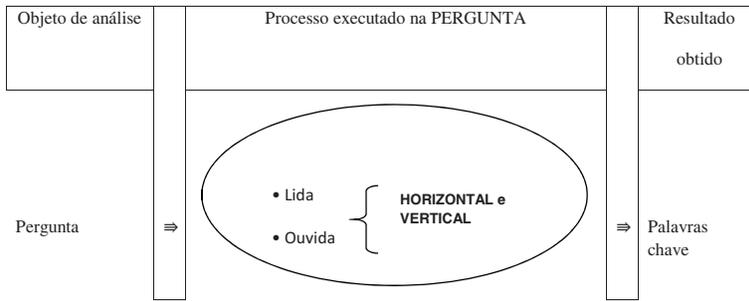


Figura 12 – Seleção de conceitos da pergunta documentária

Fonte: Elaborado pelo autor

Posteriormente, tanto os conceitos da linguagem natural extraídos quanto os atribuídos (se for necessário), seguem um desses caminhos: o armazenamento desses conceitos, ou seja, com palavras da linguagem natural extraídas do próprio texto, ou a conversão em termos padronizados e controlados com a ajuda de uma lista de descritores, uma lista de cabeçalhos de assunto ou de um tesouro.

As diferentes maneiras que os indexadores lidam com a indexação de documentos podem ser agrupadas em três correntes presentes na literatura. A indexação centrada no documento, no usuário e no domínio.

### INDEXAÇÃO CENTRADA NO DOCUMENTO

Os indexadores realizam uma descrição precisa e fiel do documento, sem considerar o contexto ou as necessidades de informação dos usuários que se utilizam dela. (SOERGEL, 1985; LANCASTER, 1991).

### INDEXAÇÃO CENTRADA NO USUÁRIO

Os indexadores selecionam os conceitos e os convertem em termos controlados por uma linguagem de indexação, levando em conta o conhecimento que tem dos usuários e suas possíveis necessidades de informação (ALBRECHTSEN, 1993; FIDEL, 1994). A mesma norma ISO 5963:1985 estabelece que os indexadores podem selecionar os conceitos que eles consideram úteis para um grupo de usuários.

## INDEXAÇÃO CENTRADA NO DOMÍNIO

Jens-Erik Mai (2005) propôs esta terceira possibilidade chamada *indexação centrada no domínio*. Este método é baseado no conhecimento profundo da organização (histórico, objetivos, metas, pessoas e relacionamentos, fluxos de informação, etc). E isso faz com que o autor subdivida a indexação em quatro processos: análise do domínio, das necessidades dos usuários, papéis desempenhados ou adotados pelos indexadores e, finalmente, a análise do documento, tendo em conta os elementos acima referidos. Estes tópicos não são repetidos para cada documento, exceto o último, quando vários deles mudam significativamente. Portanto, as diferenças de abordagem entre a indexação centrada no documento e a centrada no domínio consistem em que enquanto a primeira tem com única referência a informação contida no documento, a indexação centrada no domínio gerencia outros elementos além do documento. Como, às vezes, uma imagem é mais didática do que uma longa explicação, temos a figura a seguir para ilustrar esta dicotomia: (MAI, 2005, p. 607).

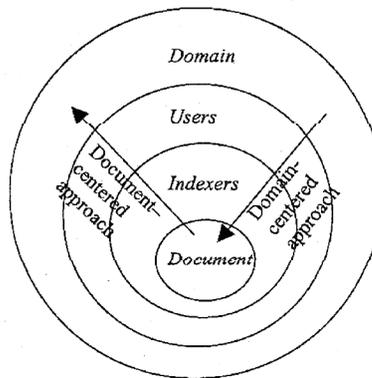


Figura 13 – Indexação centrada no documento versus indexação centrada no domínio

Fonte: MAI, 2005, p. 607.

A FIG. 11 mostra um esquema com a nossa visão particular de todo o processo de indexação. Dessa observação podemos ver tanto o objetivo como o processo:

- O objetivo da indexação dos documentos é o de permitir o armazenamento, enquanto que a indexação das questões encaminha a recuperação de documentos. Assim, o objetivo geral de indexação é o armazenamento da informação para atender às necessidades de informação. Portanto, a indexação e recuperação são duas faces da mesma moeda.

- As etapas de indexação são a análise dos documentos e as questões para a seleção dos conceitos explícitos ou implícitos, e o armazenamento destas palavras-chave como estão, ou sua conversão numa linguagem controlada.

Em resumo, a indexação é um processo executado nos objetos suscetíveis de serem representados e nas solicitações dos usuários para, em última análise, satisfazer as necessidades de informação.

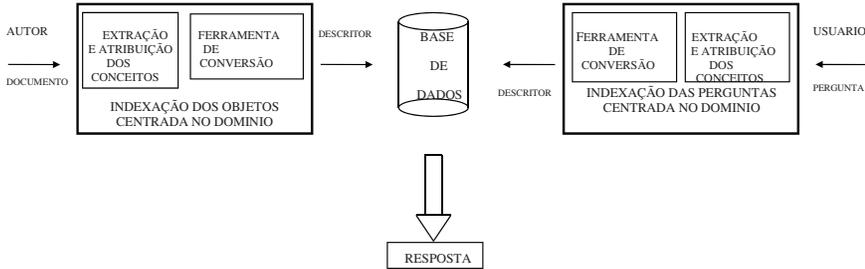


Figura 14 – Processo total da indexação

Fonte: Elaborado pelo autor.

### 2.1.1 ÍNDICE VERSUS INDEXAÇÃO<sup>2</sup>

Um índice serve para ligar um objeto à sua localização. Quando falamos de documentos, a elaboração de um índice é extrair palavras ou frases do seu contexto, classificá-las e indicar o local do documento onde se localiza cada palavra ou frase. Por isso, procura-se estabelecer uma conexão exata entre dois pontos.

Exemplo:

Índice para a página

Facilidade de localização da informação —► baixa

<sup>2</sup> Serão mantidas as palavras *indización* e *indexación* em espanhol no texto considerando-se os diferentes significados; no decorrer do texto a palavra *indexação*, em português, tem o significado de *indización*.

**Índice Onomástico**

Abbri, Ferdinando, 155, 159  
 Acosta, José, 63, 64, 65  
 Agrícola, Giorgio (Georg Bauer), 26, 33, 41, 45, 46, 48  
 Agrippa, Cornélio, 31, 32  
 Alberti, Leon Battista, 41  
 Aldrovandi, Ulisse, 57, 194  
 Alejandro Magno, 28  
 Alembert, Jean d', 49  
 Altieri Biagi, Maria Luisa, 209  
 Ammannati, Giulia, madre de Galileo Galilei, 84  
 Ammonio Sacca, 32  
 Anaximandro, 236  
 Andrea del Castagno, 42  
 Antal, Frederick, 42  
 ...  
 ...

ROSSI, P. *O nascimento da ciência moderna na Europa*.  
 Crítica: Barcelona, 1997. p. 265.

**ÍNDICE PARA A EPÍGRAFE**

Facilidade de localização da informação —► média

**Índice**

...  
 ...  
 Wall, E., L2.1.2  
 Ward, J., B1.2  
 Web thesaurus compendium, B3.1  
 websites, *see* Internet  
 weights and weighting  
 in automatic thesaurus construction, F4  
 as precision devices, B3.2.2i, G6, figure 7  
 West, L., K4.2.3  
 Whitehead, C., L2.1  
 whole entibies, *see* entibies  
 whole-part relationship  
 associative, F1.3.2<sup>a</sup>  
 hierchical, F1.2.2  
 Will, L., J, K3  
 word-by-word alphabetization, H2.4b  
 word distance devices, B3.2.2E, G2.3  
 Yan, P.F., A1.4  
 Zoological record online thesaurus on the Internet, H1, figure 8

---

AITCHISON, J., GILCHRIST, A. y BAWDEN, D. *Thesaurus construction and use: a practical manual*. 4<sup>th</sup>. ed. London: Aslib, 2000, p. 218.

## ÍNDICE PARA PARÁGRAFO

Facilidade da localização da informação —► máxima

**Índice temático**

Os algarismos romanos referem-se aos capítulos; os números arábicos remetem aos parágrafos; os números de índice indicam o número de nota que devem ser consultados. Os números em itálico indicam os fragmentos que explicam cada termo. Os conceitos mais importantes dos procedimentos que são apresentados neste manual estão destacados em negrito.

Abstração, IX.28, 30

acesso, IV.1, V.30, VII.23, X<sup>25</sup>

**ação**, I.6, V.25, V<sup>11</sup>, VI.7, 10, 28; ver ação discursiva

– **discursiva**, IV.54, VI.11, VIII.10, VIII<sup>10</sup>

Aceitabilidade, Prefácio.6, I.14-16, 1<sup>10</sup>, II<sup>3</sup>, III.16, VI.1

**aceitação**, V.I.28-33

**acontecimento**, V.25, V<sup>11</sup>, IX.25

atitudes, I.13-16, III.16, VI.1

**ativação**, I.6, III.29, V.4, 10, 24

- **de plano**, VI.13

- **generalizada**, V.12, 29, 32, VII.23, IX.31, 37.6

ato de enunciação, VI.7

- ilocutivo, VI.7-8

- proposicional, VI.7

atos perlocutivos, VI.7

**atualização**, II.23, III.12, III<sup>10</sup>, VI:23, VII.1.8.7, X.5, 19, 24, V.34, IX<sup>17</sup>

BEAUGRANDE, Robert; DRESSLER, W. *Introdução à linguística do texto*. Barcelona: Ariel, 1997. p. 332.

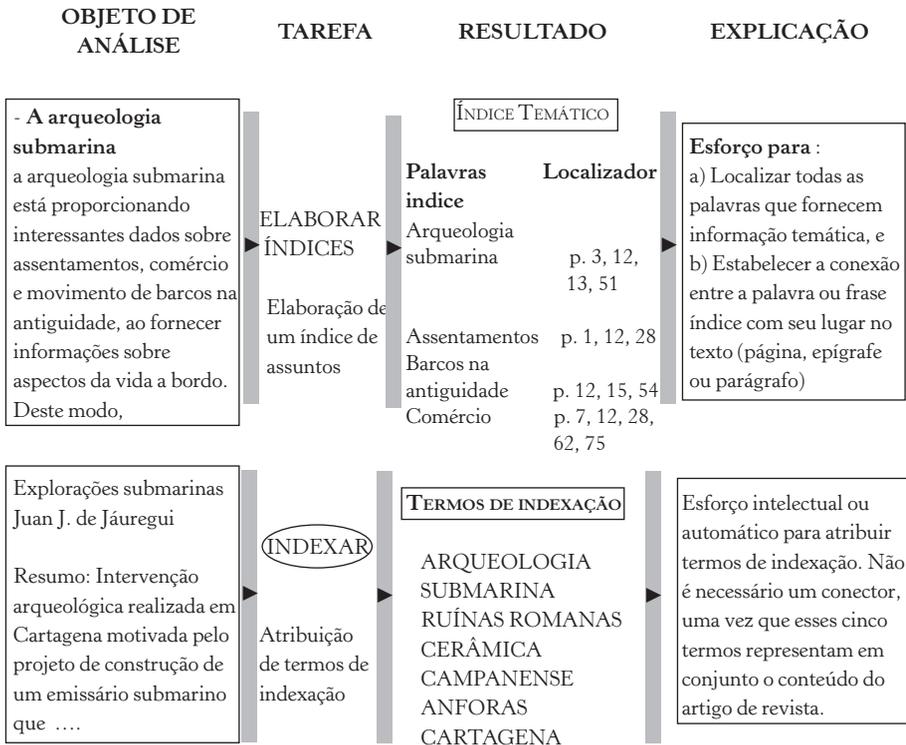
Como vimos, os índices são listas que servem para conectar uma palavra ou frase do texto ao lugar exato do mesmo. Por outro lado, o produto resultante da *indización* (palavra-chave, descritor ou cabeçalho de assunto) não está associado ao lugar exato do documento. A *indización* implica um esforço de avaliação e condensação do conteúdo implícito e explícito, enquanto que na elaboração de um índice, a tarefa se reduz a selecionar e localizar palavras ou frases do texto. Além disso, cada processo tem sua norma ISO. Para os índices, a ISO 999-1975: *Guidelines for the content, organization and presentation of indexes*, que deram origem às normas nacionais: ANSI Z39.4-1984 (Estados Unidos, BS 3700-1988 (Reino Unido), UNE 50 -111 a 89 (Espanha), NBR6034-2004 (Brasil), para citar alguns exemplos, enquanto para a *indización* a conhecida ISO 5963-1985 e suas variações que veremos mais tarde. Para saber mais sobre índices, consultar Moreira González (2004, p. 58). No esquema “*indizar X*

elaborar índices” (Quadro 13) se explica, graficamente, as diferenças entre a indexação e a elaboração de índices.

Em espanhol, convivem as palavras *indexación* e *indización* e, às vezes, são entendidas como sinônimos. Esta substituição pode ser devido à literatura francesa e depois à inglesa, que exerceram grande influência sobre a nossa ciência. Em latim, existe a palavra ‘index’ (índice), daí a “indexação” francesa, o “indexing” inglês e a “indexação” espanhola. Na Informática é usada “indexação” nos dois casos. Em primeiro lugar, para se referir a elaboração automática de entradas ou índices em um documento diante da marcação de palavras que não sejam vazias (artigos, preposições, pronomes demonstrativos, conjunções, etc.) ou vice-versa, de modo que se pode usar qualquer vocábulo que não seja uma palavra vazia para recuperar a informação de um sistema. E, um segundo uso é para se referir a todos os itens obtidos manual ou automaticamente, que representam o conteúdo de um documento. Vejamos um exemplo real deste segundo uso que apareceu na seção “Cartas ao Editor” do jornal espanhol El País em março de 2007: “Para realizar este estudo foram utilizados diferentes processos e metodologias baseadas na Indexação, Relevância e Popularidade. A indexação é o conjunto de ações a serem executadas na página da web para que o navegador possa acessar todas as informações que ele contém”.

Quadro 13 – Processo total da indexação





Fonte: Elaborado pelo autor.

Analisemos agora *indización* versus *indexación*. Conforme observado anteriormente, a indexação (*indexación*) é um processo intelectual ou automático que visa obter um conjunto de unidades conceituais que representam o objeto analisado integralmente. Também foi dito que os itens que representam um objeto a ser indexado podem ser explicados em linguagem natural. Quando isso ocorre, essas unidades conceituais são chamadas de *Palavras-chave*. Isso significa que elas foram tomadas literalmente dos objetos analisados e, portanto, são o resultado da indexação (*indización*). No entanto, na maioria dos casos as palavras ou frases selecionadas para representar o conteúdo do objeto sofrem uma transformação na normalização e controle antes de serem armazenadas no banco de dados. Esta modificação é feita para termos sistemas de informação coerentes e mais eficazes. Assim, evita-se falar dentro do sistema de formas para designar a mesma coisa (*Méjico* e *México*; *software* e *programa de computador*; *OMS* e *Organização Mundial da Saúde*). Porque se isso ocorresse, as buscas feitas com *OMS*, *Software* e *Méjico* não recuperariam os documentos da base de dados indexados por “Organização Mundial da Saúde”, “programa de computador” e “México”.

Em continuidade, veja como proceder para obter palavras-chave de um texto, sua normalização e controle. Também relatamos a diferença entre *indizar* e *indexar*.

*INDIZACIÓN* Versus *INDEXACIÓN*

TEXTO:

Durante os meses de janeiro e fevereiro de 1997, houve um episódio de mortalidade de peixes no rio Fox. A pesquisa analítica realizada por HRGC / MS na água, sedimentos, peixes e lodos de esgoto permitiu investigar pesticidas (principalmente clorpirifos e endosulfan) como agentes causadores do episódio, possivelmente devido a um vazamento de uma indústria de produtos farmacêuticos.

Processos Executados para Obter

<p><b>Palavras-chave</b>                  Selecionar do próprio texto as palavras ou frases que melhor representam o conteúdo do mesmo e armazená-las da mesma maneira no campo destinado para esse fim na base de dados.</p>	<p><b>Controle do vocabulário</b>                  Uma vez selecionadas as palavras ou frases que melhor representam o conteúdo do texto, são “filtradas” para uma linguagem de indexação para evitar sinonímia, polissemia, etc. na base de dados. Dessa maneira, se consegue um vocabulário controlado e normalizado para ser introduzido na base de dados correspondente.</p>	<p><b>Indexación</b>                  Selecionar ou marcar automaticamente todas as palavras do texto que não estejam num antídicionário, ou seja, as que não são, por exemplo, verbos, adjetivos, preposições, artigos, conjunções, pronomes, etc. para que possam ser utilizadas mais tarde na recuperação do documento.</p>	<p><b>Indización</b>                  Selecionar as palavras ou frases que melhor representam o conteúdo do texto e, normalmente, são “filtradas” por uma linguagem de indexação para conseguir um controle de vocabulário apresentado na base de dados.</p>
---	--	--	--



Palavras chave		Vocabulário controlado (Tesouro Eurovoc)	<i>Indexación</i>	<i>Indización com o tesouro Eurovoc</i>
----------------	--	---	-------------------	---

Mortalidade de peixes	▶	- MORTALIDADE - PEIXES	Meses janeiro	MORTALIDADE PEIXES
Rio Fox	▶	RIO FOX	Fev.	RIO FOX
Análise de águas	▶	ANÁLISE DA AGUA	Episódio	ANÁLISE DA AGUA
Análise de sedimentos	▶	QUÍMICA ANALÍTICA	Mortalidade	QUÍMICA ANALÍTICA
Análise de lodos de esgoto	▶	-QUÍMICA ANALÍTICA - RESIDUO	Peixes Rio	RESIDUO
Pesticidas	▶	PESTICIDA	Fox	PESTICIDA
Contaminação industrial	▶	- CONTAMINAÇÃO INDUSTRIAL	Investigação analítica	CONTAMINAÇÃO INDUSTRIAL
Contaminação de rios	▶	CONTAMINAÇÃO FLUVIAL	HRGC/MS Agua	CONTAMINAÇÃO FLUVIAL
Vazamento de resíduos	▶	ELIMINAÇÃO DE RESIDUOS	Sedimentos Lodos	ELIMINAÇÃO DE RESIDUOS
Clorpirifos	▶	CLORPIRIFOS	Depuradora	CLORPIRIFOS
Endosulfan	▶	ENDOSULFAN	Pesticidas	ENDOSULFAN
			Clorpirifos	
			Endosulfan	
			Agentes	
			Vazamento	
			Industria Produtos Farmacêuticos 1997 ...	

## 2.2 QUALIDADES DA INDEXAÇÃO

Elementos que caracterizam tanto o processo quanto o resultado da indexação são a exaustividade, a consistência, a especificidade e a correção.

### 2.2.1 EXAUSTIVIDADE NA INDEXAÇÃO

Tanto a norma ISO 5963-1985 quanto sua tradução em espanhol (UNE 50-121-91) indicam que a exaustividade na indexação está relacionada com a quantidade de conceitos que caracterizam todo o conteúdo do documento

e não com o número de descritores atribuídos a um objeto indexado. Parecem similares, mas não são.

### Exemplo

Imaginemos duas unidades de informação diferentes que indexaram o mesmo documento:

	<b>UNIDADE DE INFORMAÇÃO A</b>	<b>UNIDADE DE INFORMAÇÃO B</b>
	<b>Descritores atribuídos</b>	<b>Descritores atribuídos</b>
	ISO 15489	ISO 15489
Documento h	NORMALIZAÇÃO	NORMALIZAÇÃO
	GESTÃO DOCUMENTÁRIA PRIVADA	NORMAS ISO
	GESTÃO DOCUMENTÁRIA PÚBLICA	GESTÃO DOCUMENTÁRIA
		GESTÃO DOCUMENTÁRIA PRIVADA
		GESTÃO DOCUMENTÁRIA PÚBLICA
		ARQUIVOS UNIVERSITÁRIOS

Se considerarmos a exaustividade na indexação com o número de descritores que são atribuídos ao documento, vemos que a Unidade de Informação B parece ser mais exaustiva. Nada poderia estar mais longe da realidade, porque com o documento em mãos, comprovamos que o descritor “Arquivos universitários” é erroneamente atribuído, ou seja, não deve ser proposto como um descritor para esse documento e, em segundo lugar, que os termos “Norma ISO” e “Gestão Documentária” são dois termos gerais de “ISO 15489” e, em terceiro lugar, “Gestão documentária privada e Gestão documentária pública” foram atribuídos seguindo as diretrizes estabelecidas pela instituição, mas esses conceitos não abrangem a parte conceitual do documento. Portanto, se retirarmos, “Normas ISO”, “Gestão documentária” e “Arquivos universitários”, as Unidades de informação A e B têm o mesmo número de termos, mas com um olhar superficial poderia parecer outra coisa diferente.

Soergel (1994, p. 591) observou que em alguns sistemas, uma vez que foram selecionados os termos de indexação, aumenta o seu número com a atribuição de outros relacionados ou genéricos. Assim, por exemplo, se um

documento é sobre a “depressão clínica” pode ser indexado pelo descritor “Depressão Clínica”, que foi talvez um dos mais precisos, mas também por outros termos mais gerais como “Psicose”, “Psicose afetiva”, ou “Doença mental”. No entanto, esses três termos não fornecem um maior grau de exaustividade em termos de conceitos presentes no documento. Então, este fator deverá ser considerado para uma análise comparativa entre vários sistemas ou bases de dados que analisam o mesmo documento para verificar o grau de exaustividade. Existem várias razões pelas quais os sistemas de informação ou bases de dados ampliam o número de descritores sem considerar mais conceitos contidos no documento: pela generalidade ou especialização da informação tratada, pelas exigências dos usuários, e porque quanto mais descritores, maior é a possibilidade de recuperação, embora, obviamente, diminua o grau de precisão.

### 2.2.2 ESPECIFICIDADE

Charles Ammi Cutter em seu *Rules for a printed dictionary catalogue* de 1876 falou da especificidade na atribuição de assuntos. Ele aconselha que os documentos sejam indexados por cabeçalhos de assunto específico, evitando um cabeçalho mais geral. Diz, por exemplo, que um livro intitulado “A Arte da Pintura” seja indexado por “Pintura”; que outro chamado “O ferro” seja pelo assunto “Ferro”, mas não “Metais” ou “Metalurgia”, e um livro de cacto, “Cacto” e não “Botânica” (CUTTER, 1876, p. 15).

Desde a metade da década de 1970 foram publicados trabalhos sobre a especificidade dos termos de indexação (SPARCK JONES, 1972; ROBERTSON, 1974, YU; SALTON, 1976; WU; SALTON, 1981; ROWLEY, 1988; OU WONG; YAO, 1992). A própria norma ISO 5963-1985 sugere que a especificidade tem a ver com a precisão com que um termo de indexação representa fielmente um conceito particular que aparece no documento em análise. Portanto, é preciso ter linguagens de indexação adequadas ao grau de especificidade que se deseja conseguir e, em seguida, que o indexador localize esses termos adequados durante o processo.

Em suma, o padrão de especificidade é definido tanto pela experiência dos indexadores como pelos descritores ou assuntos escolhidos nas linguagens de indexação utilizadas, assim como a política de indexação da unidade de informação, mas ambos aspectos devem estar pré-estabelecidos. Em geral, a indexação em uma biblioteca não especializada será mais geral do que em uma especializada.

### 2.2.3 CORREÇÃO

A correção na indexação ou a ausência de erros é de suma importância, porque a indexação pode ter dois tipos de falhas: por omissão (quando um termo é omitido) e por inclusão (adição de um termo sem necessidade). Por outro lado, a ausência de um termo que deveria estar presente e a atribuição de um mais genérico, específico ou afins, é um tipo especial de erro de omissão e de inclusão. Para determinar esse tipo de erro é importante saber quais descritores ou assuntos são mais apropriados para um documento, para isso deve-se ter um consenso entre vários indexadores e usuários. Portanto, a perfeição da indexação estabelece uma relação entre a presença de descritores corretos e a ausência de omissões (SOERGEL, 1994, p. 593).

### 2.2.4 CONSISTÊNCIA

A consistência na indexação foi definida por Zunde e Dexter (1969, p. 259), dois pioneiros nesse assunto, como “[...] o grau de concordância na representação da informação essencial de um documento por meio de um conjunto de termos de indexação selecionados por cada um dos indexadores de um grupo”. A consistência na indexação pode ser estudada como uma referência a um ou vários indexadores. Quando um profissional indexa o mesmo documento em momentos diferentes falamos de intra-consistência ou consistência intra-indexador.

#### Intra-indexador

Janeiro 2008		Junho 2010
		
Indexação A		Indexação B

E em segundo lugar, quando vários profissionais indexam um mesmo documento de maneira diferente, falamos de inter-consistência ou consistência inter-indexador.

## Inter-indexador

Indexador h		Indexador j
		
Indexação de h		Indexação de j

O resultado da indexação resulta da conjunção de vários elementos que devemos levar em consideração na hora de estudar a consistência. Esses elementos são:

1. A formação, os conhecimentos do assunto, o grau de profissionalismo e a motivação do indexador;
2. As características do objeto indexado, e
3. As condições que a indexação é realizada.

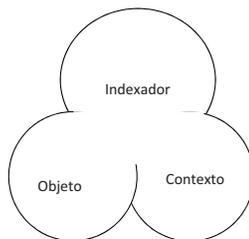


Figura 15 – Âmbitos que intervêm na consistência da indexação

Fonte: Elaborado pelo autor

Em contrapartida, na indexação automática não ocorrem fatores externos que podem afetar o próprio sistema. Um programa de computador indexará sempre igual, bem ou mal, um documento sem que intervenha no contexto. A indexação variará somente quando forem feitas alterações nos parâmetros de análise do sistema.

A consistência propiciou abundante bibliografia desde a década de 1960 até o presente. Por um lado, pode ser classificada em literatura teórica que investiga os motivos que causam os diferentes resultados na indexação e, por outro lado, uma literatura mais experimental, que procura quantificar a similaridade entre várias indexações. No primeiro grupo, os pesquisadores têm trabalhado em:

- fatores envolvidos no processo;
- habilidades e metodologias de leitura dos indexadores;
- relação entre a seleção de conceitos e a recuperação;
- as causas que levam os indexadores a escolherem ou rejeitarem termos de acordo com suas propriedades, ou;
- aspectos psicológicos que mediam durante a indexação.

A outra parte importante da literatura sobre a consistência visa sua quantificação, ou seja, a obtenção de índices de consistência variando de 1 a 100 por cem, mediante fórmulas diferentes da semelhança entre indexações, apoiando em diferentes aspectos como a experiência (indexadores principiantes ou experientes) ou tipos de documentos (livros, patentes, artigos de jornal, fotografias, etc.). Fazer comparações entre indexações é um assunto complicado. Quando a intenção é comparar a indexação de uma instituição com outra é necessário controlar o número máximo de elementos que, num maior ou menor grau, afetam o resultado.

Elementos a serem considerados na comparação de indexações:

<b>INDEXADOR</b>	Formação e experiência em indexação: indexador experiente x novato Conhecimento do assunto Domínio das ferramentas de indexação (linguagem de indexação) Profissionalismo
<b>CONTEXTO</b>	Políticas de indexação da instituição Objetivo da indexação: temas principais X especificidade Tipos e necessidades dos usuários Carga de trabalho e tempo dedicado
<b>OBJETO</b>	Complexidade do objeto indexado: livro infantil X patentes Características e propriedades do objeto indexado: texto X material gráfico ou audiovisual Tamanho: indexação de textos curtos X textos longos
<b>MOMENTO</b>	A comparação é executada com as palavras-chave tiradas diretamente do texto, uma vez convertidas em descritores depois de passarem pelo vocabulário controlado
<b>FORMULA</b>	São muitas as fórmulas matemáticas utilizadas para conseguir os índices de consistência

Da revisão dos resultados obtidos nos experimentos realizados nos últimos anos, vemos que a média dos índices oscila entre 25% e 60% de coincidência. Uma análise cuidadosa do que foi escrito nos parágrafos anteriores, juntamente com os dados mencionados, faz com que assumamos que a inconsistência é uma característica inerente à indexação e não uma anomalia esporádica.

Também é verdade que a subjetividade não está presente apenas na indexação, pois, como observou Cleverdon (1984), se duas pessoas ou grupos constroem um tesouro pode ser que concordem com um pouco mais do que a metade dos termos incluídos, se dois profissionais questionam uma base de dados com a mesma questão, não chega a cinquenta por cento a informação comum que é recuperada; e, finalmente, se dois pesquisadores são questionados pela relevância de documentos para uma dada questão, o acordo entre os dois não passa de sessenta por cento.

### 2.3 AS INDEXAÇÕES DE UM DOCUMENTO

Desde o final do século XIX até metade do XX, a indexação tem se mantido mais ou menos inalterada, uma vez que a atribuição de assuntos nas bibliotecas consistia em condensar em um ou dois cabeçalhos o conteúdo de um livro (Informática – Tratados e manuais, ou Fueros-Navarra-S. XIX-XX). Entretanto, com o surgimento das bibliotecas especializadas e dos centros de documentação, ainda que principalmente das grandes bases de dados, o conteúdo dos documentos foi sendo parcelado cada vez mais em campos para acomodar a indexação, para facilitar a recuperação.

Indexação de um livro

Cabeçalho de assunto +  
 subcabeçalho de assunto  
 subcabeçalho de lugar  
 subcabeçalho de tempo  
 subcabeçalho de forma

*exemplo:*

- Arquitetura–História–S.XX–Estudos, ensaios
- Edifícios industriais–Construção

Indexação de um artigo



- Descritores temáticos:
- Descritores onomásticos:
- Descritores topográficos:
- Séculos:
- Épocas:
- Períodos:
- Identificadores:
- Códigos temáticos:
- Palavras chave:
- Indexação química:
- Indexação numérica:
- Indexação objeto astronômico:



LIVRO

ARTIGO

Do Paneuropeísmo a União Européia: 1923-1957		Do Paneuropeísmo a União Européia: 1923-1957
▼		▼
União Européia– História–1923-1957	Descritores temáticos: Revisão histórica ; Instituições; História das instituições; Políticos ; Biografias ; Fotografias ; Documentos ; Estatísticas Identificadores: Movimento paneuropeu ; Conselho da Europa; Comunidade Européia do Carvão e do Aço; Tratado de Roma; Comunidade Econômica Européia ; União Européia Descritores onomásticos: Altiero Spinelli; Jules Gueron; Jean Monnet Descritores topográficos: Paris ; Roma ; Benelux Períodos: 1923-1957 Séculos: XX	

A análise de um documento proporciona conceitos explícitos ou implícitos de diferente índole. Vejamos como surgem os conceitos a partir dos títulos:

TÍTULO	Conceitos de tema	Conceitos de lugar	Conceitos de tempo	Conceitos de nomes próprios
O robô a serviço da arte	<ul style="list-style-type: none"> <li>• Robo</li> <li>• Arte</li> </ul>			
a restauração de Castillo de San Juan de las Águilas da vila de Águilas	<ul style="list-style-type: none"> <li>• Restauração</li> <li>• Castillos</li> </ul>	Águilas		Castillo de San Juan de las Águilas
Níveis de alfabetização em Salamanca de 1748 a 1860	<ul style="list-style-type: none"> <li>• Alfabetização</li> <li>• Dados estatísticos</li> </ul>	Salamanca	<ul style="list-style-type: none"> <li>• XVIII</li> <li>• XIX</li> </ul>	
as inscrições romanas de Mengíbar	Inscrições	Mengíbar	Época romana	
os vales dos Pirineus franceses	Vales	<ul style="list-style-type: none"> <li>• Pirineus</li> <li>• França</li> </ul>		
Entrevista a Miguel Delibes	<ul style="list-style-type: none"> <li>• Entrevistas</li> <li>• Escritores</li> <li>• Miguel Delibes</li> </ul>			Miguel Delibes

Assim, a reunião de cada classe de conceitos dá lugar ao que poderemos denominar “classes de indexação” distintas:

**INDEXAÇÃO TEMÁTICA** → Palavras chave temáticas; Descritores temáticos; Cabeçalhos de assunto

Robô; Arte; Restauração; Castillos; Alfabetização; Dados estatísticos; Inscrições; Vales; Entrevista; Escritores; Miguel Delibes

**INDEXAÇÃO DE LUGAR** → Topônimos; Descritores topográficos; Cabeçalhos de lugar ou subcabeçalhos de lugar

Classes de lugares

Unidades tribunais: países, principados, departamentos, estados, províncias, condados, povos, distritos, cidades, bairros, dioceses, arquidioceses	→	Suíça; Andorra; Arizona; Murcia; México D.F.; Carabanchel; Águilas; os Arejos; Diocese de Cartagena
--	---	---

Unidades supranacionais	→	Países Árabes; União Européia; Organização para a Unidade Africana; Organização dos Estados Americanos; Mercosul, Organização do Tratado do Atlântico Norte
Unidades não tribunais: - lugares arqueológicos, cidades antigas, reinos - áreas, regiões - aquedutos, canais - pântanos, bosques, selvas - acidentes geográficos: cavernas, desertos, ilhas, lagos, montanhas, planícies, oceanos, rios, mares - bacias geológicas - barragens - estradas, túneis	→	Depósito de Atapuerca; Reino de Granada; Canal Isabel II; Selva Negra; Lago Victoria; Everest; Rio Amazonas; Mar de Bering; Túnel de Viella

**I**

**INDEXAÇÃO DE TEMPO** → Datas; Descritores cronológicos; subcabeçalho de tempo

Séculos (XVIII, XXI); Idades (Idade da Pedra; Idade Contemporânea); Épocas (Época romana); Períodos (1914-1918; 1939-1945); Décadas (1990-1999); Anos (2008; 2009; 2010)

**INDEXAÇÃO DE NOMES PRÓPRIOS** → Nomes próprios de pessoas (Descritores onomásticos, cabeçalho ou subcabeçalho de assunto) e nomes próprios de objetos (identificadores)

Nomes próprios de pessoas: Miguel Delibes; Cristobal Colón; Fidel Castro; Charles Ammi Cutter; Robert Redford

Nomes próprios de objetos: Aos nomes próprios de objetos, preferimos denominá-los de “identificadores”. As classes de identificadores são numerosas, assim mostramos algumas sem exaustividade:

## Classes de identificadores

Nomes de famílias, dinastias, casas reais etc.	→	Família de Vélez; Dinastia dos Austrias;
Personagens mitológicos, legendários ou de ficção	→	Afrodite; Robin Hood; Tarzan; Sancho Panza; Indiana Jones
Guerras e batalhas	→	II Guerra Mundial; Guerra de Iraque; Batalha do Ebro

Tratados	→	Tratado de Versalhes; Tratado de Maastricht
Organismos	→	Conselho Superior de Investigações Científicas; Liga Árabe; Anistia Internacional; Organização Mundial da Saúde
Projetos	→	Projeto Homem; Projeto MACS
Testes, experimentos, ensaios	→	Teste visomotor de Bender; Teste de Kolmogorov-Smirnov; Teste de Spearman-Brown
Empresas, marcas e produtos	→	Zara; Microsoft; Kodak; Aspirina; Danone;
Obras do homem (literárias, artísticas, musicais, arquitetônicas, etc.)	→	Crime e castigo; As Meninas; A Traviatta; Castelo de San Juan de las Águilas; Catedral de Reims; Praça de São Pedro
Acontecimentos histórico-político-culturais	→	Revolução francesa; Revolução industrial; Crise de 29; Renascimento; Romantismo
Correntes de pensamento	→	Humanismo; Darwinismo; Positivismo; Antiglobalização
Eventos periódicos ou eventuais (Festivais, Encontros, Congressos, Reuniões, etc.)	→	Bienal de Veneza; Festival de Cinema de San Sebastián; Congresso Internacional de Matemáticos; Carnaval do Rio de Janeiro; Tour de França
Instituições	→	União Européia; Defensor do Povo; Instituição Livre de Ensino; Fundação Cajamurcia; Tribunal Superior de Justiça
Organizações e estruturas políticas	→	Monarquia; República; Ditadura; Ministério da Educação; Comunidade Autônoma; Direção Geral do Meio ambiente
Reconhecimentos e prêmios	→	Gran Cruz de Isabel a Católica; Medalha ao Mérito no Trabalho; Prêmio Príncipe de Astúrias; Premio Nobel
Nomes científicos de animais e plantas	→	Octopus vulgaris (pulpo); Apis mellifera (abelha); Malus domestic (maça)
Nomes de planetas, estrelas, constelações, etc.	→	Saturno; Orión; Andrómeda; Osa menor.

Cada unidade documentária ou base de dados acaba construindo suas listas de identificadores autorizados para garantir a máxima coerência na indexação e na recuperação e, em última instância, na qualidade do sistema.



Listas de identificadores de uma base de dados e de um catálogo coletivo

Base de dados Eric	Autores e Entidades Red Bibliotecas CSIC
<b>AAS Degrees</b> USE Associate Applied Science Degrees	<b>Astour, Michael C.</b> U.p. Astur, Michal
<b>AASCU ERIC Model Programs Inventory</b>	<b>ASTR</b> V. American Society for Therapeutic Radiology And Oncology
<b>AAUP Contracts</b>	<b>Astrana Marín, Luis</b>
<b>ABA Inventory</b> USE Alcohol Behavior Altitude Inventory	<b>Astrofizicheskiy Institut (Alma-Ata)</b> T.r. Akademia Nauk Kazajskoi SSR
<b>Abacuses</b>	<b>Astrofyzikalni Observator (Praga)</b> T.r. Astronomicky Ustav (Praga) Ceskollovenská Akademie Ved Ustredni Ustav Astronomicky (Checoslovaquia)
<b>Abbey Wood Schools</b>	<b>Astronomical Contact Group</b> U.p. ACG
<b>Abbott House</b>	<b>Astronomical Institute</b> V. Astronomicky Ustav (Praga)
<b>ABC (Approach by Concept) Method</b>  <b>ABC (Act for Better Child Care) Bill</b> USE Act for Better Child Care Services	

## INDEXAÇÃO DE COMPOSTOS QUÍMICOS

Compostos, fórmulas ou símbolos químicos (GaAIAs ; gap ; Co)

## INDEXAÇÃO DE TRATAMENTO

Forma em que foi tratado o tema objeto de investigação (Aplicação; Bibliográfico; Experimental; Revisão, etc.)

## INDEXAÇÃO DE DADOS NUMÉRICOS

Dados numéricos relativos à temperatura, frequência, longitude de onda, massa, velocidade, etc. (1.8E+04 ; 9.5E-01 ; tamanho 3.0E-01 m)

## INDEXAÇÃO DE OBJETOS ASTRONÔMICOS

Para a representação de um nome, de uma localização no espaço, etc. (HD 200120; 59 Cyg ; NGC 204)

### 2.4 ÁREAS DE EXTRAÇÃO DE CONCEITOS E TEMPO GASTO

Tratando-se de texto, a norma espanhola UNE 50-121-91 que traduz a ISO 5963-1985 *Methods for examining documents, determining their subjects, and selecting index terms* determina a importância no momento de indexar os títulos, resumos, sumários ou as tabelas de conteúdos, apresentações, frases de começo de capítulos e parágrafos, conclusões, ilustrações, palavras ou grupos de palavras com uma tipografia incomum.

Na indexação de artigos científicos, os resumos costumam ter um maior número de conceitos do que os títulos (GIL LEIVA; RODRÍGUEZ MUÑOZ, 1997). Da mesma forma, também é comum que os indexadores lidem num maior ou menor grau, com as palavras-chave fornecidas pelos autores dos artigos para a atribuição dos assuntos ou dos descritores, conforme encontramos nos dois estudos que realizamos nas bases de dados nacionais espanholas (ISOC, ICYT e IME) e internacionais (CAB, LISA, INSPEC e ISTA) (GIL LEIVA; ALONSO ARROYO, 2005; 2007):

Base de dados	% de coincidência entre as palavras chave dos autores e os descritores atribuídos
ISOC	60,48
ICYT	58,18
IME	64,96
=====	
CAB	60,58
LISA	42,52
INSPEC	41,62
ISTA	37,89

É difícil definir o tempo médio gasto, ou que deveria ser dedicado à indexação de um documento. Primeiro, porque depende das orientações definidas pela instituição, que por sua vez, deve-se a vários fatores, incluindo os tipos e as necessidades do usuário ou simplesmente, o número de profissionais dedicados a esta operação. E se a instituição não tem um padrão, o tempo gasto em cada documento é baseado no profissionalismo dos indexadores. Em segundo lugar, o tempo está sujeito às características próprias da indexação como a profundidade ou a perfeição pretendidas, bem como outros fatores como a classe de documentos analisados, o tipo e a complexidade das informações contidas nos mesmos, ou até mesmo o tamanho dos documentos, embora nem sempre este último se cumpra.

A política de indexação de algumas bases de dados estabelece algumas indicações de tempo. Assim, na base de dados ERIC recomenda-se 30 minutos para um romance; para uma biografia ou autobiografia, 30 minutos; para poesia, 5 minutos; e para um artigo de crítica, 20-30 minutos.

## 2.5 NORMAS

As instituições internacionais, tais como a IFLA, a FID e a UNESCO têm colaborado com a ISO na produção de recomendações para a execução de tarefas comuns em unidades documentárias. Os assuntos objetos de normalização são divididos e são criados comitês técnicos compostos por representantes de vários países. Atualmente, existem mais de 200 comitês técnicos criados para trabalhar em diversos âmbitos. O comitê técnico *TC 46 Información y documentación* trabalha na normalização da prática própria das bibliotecas, centros de informação e documentação, arquivos, documentação de museus, indexação e resumos e ciências da informação. Desde a sua criação até hoje, o TC 46 deu origem a quase uma centena de normas.

A normalização nas tarefas de análise de conteúdo documentário é complexa devido aos elementos envolvidos nesta operação: aspectos cognitivos, educacional-intelectuais e subjetivos. Não é em vão que desde meados dos anos de 1960 existem trabalhos que tentam explicar como indexar os documentos. Alguns desses trabalhos vêm de iniciativas privadas ou instituições estatais de grande porte (UNESCO, Centro Nacional de Informação Científica e Técnica russo, Centro de Documentação da NASA, Biblioteca Nacional de Medicina dos Estados Unidos, etc.). Estas políticas perseguem, entre outros aspectos, a aplicação de critérios uniformes nas regras de indexação, a indexação de todos aqueles conceitos presentes nos documentos, ou a indexação de todos os conceitos no seu grau correto de especificidade.

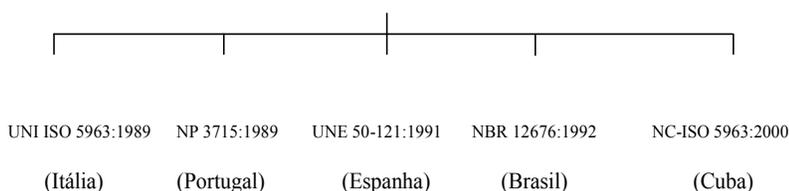
Na década de 1970 começaram os esforços de normalização na indexação. Na França, foi publicada a NF Z 47-102 1978 *Principles généraux pour l'indexation des documents* (Princípios gerais para a indexação de documentos), a partir dos trabalhos feitos pelo Comitê Técnico “Documentación” da ISO e da UNESCO. Esta norma está estruturada da seguinte forma: 1. Objeto e domínio de aplicação; 2. Processo e finalidade; 3. Reconhecimento de conceitos contidos na informação (captação do conteúdo do documento, identificação dos conceitos que representam o conteúdo e seleção dos conceitos considerando a exaustividade e a especificidade desejada); 4. Representação de conceitos na linguagem documentária; 5. O controle de qualidade da indexação, Anexo 1: Esquema da indexação na cadeia documentária e Anexo 2: Glossário.

Vários anos mais tarde apareceu no Reino Unido, numa iniciativa do *British Standards Institute*, a norma BS 6529:1984 *Recommendations for examining documents, determining their subjects and selecting indexing terms*. Desse mesmo ano também é a norma cubana NC 39-22: 1984 *Indización manual de documentos*. No ano seguinte foi publicada a norma ISO 5963:1985 relativa à análise de conteúdo dos documentos para selecionar os termos de indexação (*Methods for examining documents. Determining their subjects and selecting indexing terms*). Tomando como ponto de partida, principalmente, a francesa NF Z 47-102 1978 e a britânica BS 6529:1984, a ISO 5963, 1985 ficou articulada desta forma: 1. Objeto e âmbito de aplicação; 2. Normas para consulta; 3. Definições; 4. Processo de indexação; 5. Análise do documento; 6. Identificação dos conceitos, 7. Seleção dos termos de indexação; 8. Controle de qualidade; Anexo: Organograma do processo de indexação utilizando um tesauro.

Na ISO 5963:1985 foram modificados aspectos pontuais, inclui um mini-glossário no corpo da norma e foi incluído como anexo um fluxograma explicando como indexar um documento com um tesauro.

Desde a publicação da norma ISO 5963-1985 foram elaboradas ou atualizadas normas equivalentes em diferentes países que são traduções ou interpretações mais ou menos fiéis.

#### ISO 5963-1985



## 2.6 RELAÇÃO ENTRE INDEXAÇÃO E RECUPERAÇÃO

No tópico 1 *Conceito de indexação* dissemos que a indexação e a recuperação são duas faces da mesma moeda. Agora vamos nos deter nesta relação.

O modelo usado pelos sistemas na comparação entre a pergunta e os documentos determina como eles devem ser indexados. Assim, pode-se falar de diferentes modelos de sistemas de recuperação da informação segundo Peña, Baeza-Yates e Rodriguez Muñoz (2002, p. 218):

1. O modelo de busca por texto livre, que não exige a indexação dos documentos, portanto, não se mantem nos índices. A informação está diretamente acessível tal como foi apresentada e as questões na base de dados são feitas através de combinações de cadeias de caracteres.
2. O modelo booleano é utilizado por quase todos os sistemas de recuperação de informação comerciais. O sistema combina os termos utilizados pelo usuário na pergunta através dos operadores Y (e) O (ou) e NO (não) aos termos (não vazios) presentes na base de dados. Não dispõe de mecanismos para especificar em que grau um termo é mais representativo para um documento do que outro.
3. O modelo booleano estendido pode atribuir um peso aos termos no momento da pergunta ou na indexação dos documentos. Pode ser representado da seguinte maneira:  $A_{p1} OB_{p2}$ ;  $A_{p1} YB_{p2}$ ;  $A_{p1} \dots P_2 B$  onde A e B representam os conjuntos indexados pelos respectivos termos, enquanto que P1 e P2 são os pesos aplicados a cada conjunto.
4. O modelo de espaço vetorial é representado no sistema por duas listas ordenadas numericamente. Estas listas correspondem, por um lado, aos pesos atribuídos ao documento (em função da soma dos valores atribuídos aos termos de indexação que representam o documento), e por outro lado, aos valores da pergunta do usuário (em função dos pesos atribuídos aos termos de indexação da pergunta).
5. O modelo de lógica difusa se baseia na inexistência de um limite definido que permita agrupar algo a um único grupo, mas sempre existem características que podem incluí-lo em algum lado. O conjunto difuso D é definido como:  
 $D = \{ x \in U, f(x) / (x) > 0 \}$  em que x são os elementos do conjunto,  $\in$  é o símbolo que indica pertinência, U é o conjunto universal e f(x) a função de pertinência que define o grau de pertinência do indivíduo ao grupo.

Outra classificação dos modelos de recuperação foi feita pelo alemão Kuroпка (2004). Faz uma divisão em função de seu fundamento matemático (modelos teóricos, modelos algébricos e modelos de probabilidade) e de suas propriedades (modelos sem interdependência entre seus termos e modelos com interdependência.).

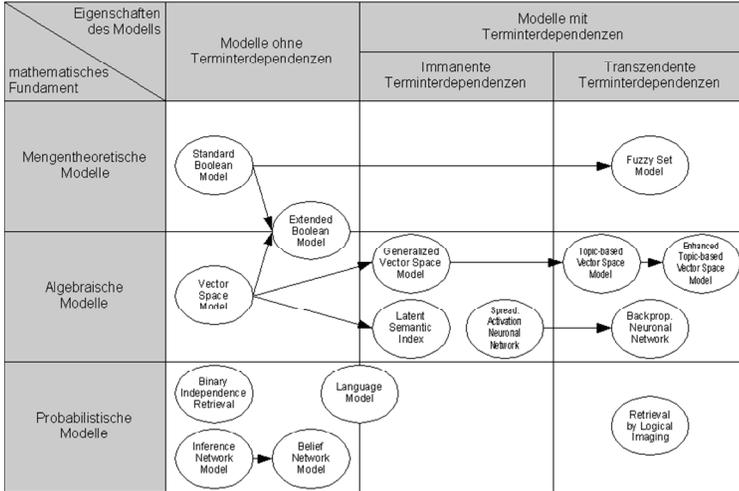


Figura 16 – Modelos de recuperação

Fonte: KUROPKA, 2010.

Para medir todos os aspectos relativos à recuperação tem se recorrido a critérios diferentes e variados. Aqui está uma compilação de diversas medidas para a avaliação da recuperação de informação aglutinadas no Quadro 14 e literalmente copiado de Martínez Méndez (2002, p. 70).

Quadro 14 – Medidas para a avaliação da recuperação de informação

Medidas baseadas na Relevância	
Precisão	Documentos relevantes recuperados divididos entre o total de documentos recuperados
Revocação	Documentos relevantes recuperados divididos entre o total de documentos relevantes
Média da efetividade E-P	Média da efetividade em pares de valores de <i>exaustividade e precisão</i>
Medidas baseadas no Processo	

Seleção	Mede quantos documentos têm na base de dados, o grau de sobreposição com outras relacionadas, o que se espera da base de dados antes da busca
Conteúdo	Tipo de documentos da base de dados, temática dos documentos, frequência de atualização
Tradução de uma consulta	Verifica se o usuário pode fazer a consulta diretamente ou precisa de intermediação
Erros no estabelecimento da consulta	Média de erros sintáticos na escrita da busca que propiciam a recuperação de conjuntos vazios ou errôneos
Tempo médio de realização da busca	Tempo médio de realização de uma estratégia de busca
Dificuldade na realização da busca	Identifica problemas que usuários inexperientes podem encontrar
Número de comandos precisos para uma busca	Média de instruções necessárias para realizar uma busca
Custo da busca	Custos diretos e indiretos na sua realização
Nº docs recuperados	Extensão do resultado de uma busca
Número de documentos revisados pelo usuário	Média de documentos que os usuários estão dispostos a revisar
<b>Medidas baseadas no Resultado</b>	
Precisão	-- já definida anteriormente --
Exaustividade	-- já definida anteriormente --
Média efetividade E-P	-- já definida anteriormente --
Medidas da satisfação do usuário	Medidas que pretendem medir a reação dos usuários diante do resultado de uma busca

Fonte: MARTÍNEZ MÉNDEZ (2002, p. 70).

Quando se realiza uma busca numa base de dados, a coleção de documentos armazenados se divide em dois grupos: os documentos recuperados e os não recuperados. Os documentos recuperados se subdividem, segundo o critério do usuário, em documentos relevantes e irrelevantes. Isto também se aplica para aqueles que não foram recuperados.

Quadro 15 – Documentos recuperados e não recuperados

Documento	Relevante	Irrelevante	Total
Recuperado	a	B	a+b
Não recuperado	c	D	c+d
Total	a+c	b+d	a+b+c+d

Fonte: LANCASTER, 1978, p. 171

Uma busca é perfeita quando o usuário julga que foram recuperados da base de dados todos os documentos relevantes ( $a+c$ ), e foram omitidos aqueles que considera irrelevantes ( $b+d$ ). Neste caso, estamos diante de uma resposta com o maior grau tanto de exaustividade como de precisão na recuperação. Para determinar os graus de exaustividade e precisão na recuperação se utilizam as seguintes fórmulas:

$$\text{Exaustividade} = \frac{\text{n}^\circ \text{ total de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos relevantes na coleção}}$$

ou seja,

$$\text{Exaustividade} = \frac{a}{a+c}$$

isto é,

$$\text{Precisão} = \frac{\text{n}^\circ \text{ total de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos relevantes na coleção}}$$

$$\text{Precisão} = \frac{a}{a+b}$$

Os quocientes de abrangência e precisão de uma resposta documentária são variáveis. Em alguns casos, há respostas com uma exaustividade e uma precisão muito baixas; uma exaustividade muito baixa e uma precisão muito alta; ou uma exaustividade e precisão muito altas. No entanto, estes dois parâmetros aparecem inversamente relacionados na prática, porque se há uma melhora em algum dos dois, está associado com um pior resultado do outro. Além disso, a precisão pode ser considerada como um tipo de fator de custo do tempo que o usuário precisa para separar as referências relevantes das irrelevantes.

Finalmente, o ruído e o silêncio documentários estão relacionados com a exaustividade na recuperação. Por ruído documentário entendemos a informação resultante de uma busca que não interessa ao usuário.

Em contraste, o *silêncio documentário* é a informação que interessa ao usuário, mas não foi recuperada do sistema.

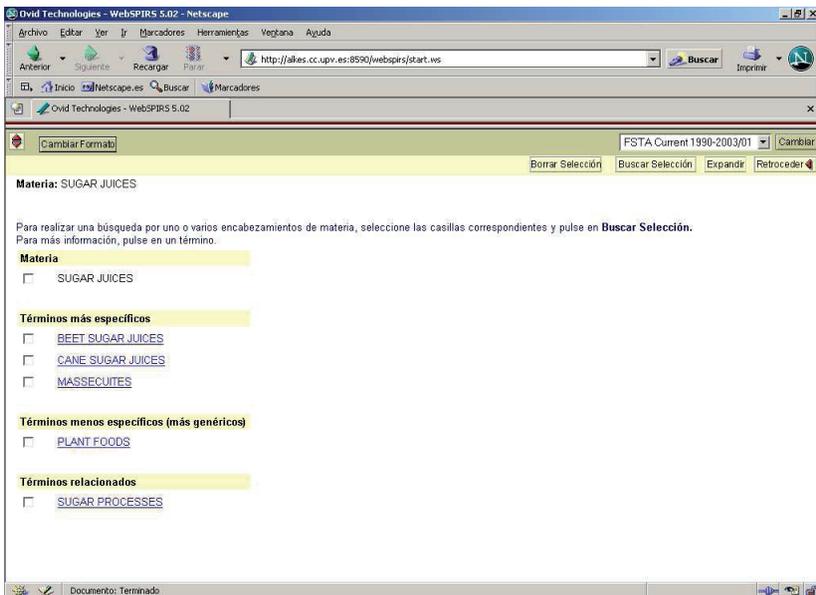
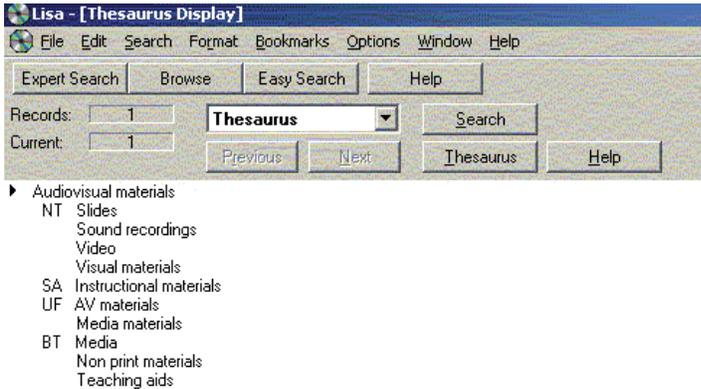
Já foi dito anteriormente que a indexação e recuperação são duas faces da mesma moeda, uma vez que a indexação não tem razão de existir se não for para recuperar a informação. Durante a indexação é produzida uma fragmentação e representação do conteúdo de um documento em determinados campos da base de dados, que serão empregados posteriormente para localizar de maneira precisa a informação fragmentada.

A maneira mais fácil de comprovar a relação entre a indexação e a recuperação é indo para as bases de dados documentárias, já que estas proporcionam caixas de diálogo para facilitar as buscas para os usuários e, finalmente, a recuperação da informação. Na base de dados Medline, por exemplo, há um campo denominado Age Tags (Etiquetas de idade) que servem para abrigar as idades dos seres humanos objeto de um estudo (recém-nascido, criança, adolescente, meia-idade, etc.) Posteriormente, na fase de interrogatório, a própria base de dados oferece várias caixas de diálogo para facilitar a busca dos usuários, e neste caso das idades é a seguinte:

Outros exemplos

Término:	Número de Registros
<input type="checkbox"/> INTERNET	7768
<input type="checkbox"/> INTERNET-	4290
<input type="checkbox"/> INTERNET--A	1
<input type="checkbox"/> INTERNET--AN	1
<input type="checkbox"/> INTERNET--ARE	2
<input type="checkbox"/> INTERNET--AT	1
<input type="checkbox"/> INTERNET--CONNECTING	1
<input type="checkbox"/> INTERNET--FROM	1
<input type="checkbox"/> INTERNET--HOW	1
<input type="checkbox"/> INTERNET--ONE	1
<input type="checkbox"/> INTERNET--ORIGINALLY	1
<input type="checkbox"/> INTERNET--SUCH	1
<input type="checkbox"/> INTERNET--SURVEY	1
<input type="checkbox"/> INTERNET--THE	3
<input type="checkbox"/> INTERNET--USING	1
<input type="checkbox"/> INTERNET--WHATS	1
<input type="checkbox"/> INTERNET-ACCESS	4
<input type="checkbox"/> INTERNET-ACCESSED	2

A primeira tela dá acesso aos usuários da Base de Dados ISOC a lista de descritores que podem empregar para a preparação de suas equações de busca. A segunda mostra parte do índice da palavra “internet” da base de dados ISA, que os usuários lidam para suas consultas.



As duas telas anteriores mostram tesouros de bases de dados, o tesouro da base de dados LISA e o tesouro da base de dados FSTA. Como se observa, as bases de dados documentárias costumam dispor de um vocabulário controlado

para a indexação de documentos. Mais frequentemente, este mesmo vocabulário está à disposição dos usuários das bases de dados para usarem em suas pesquisas.

Temos dito que o resultado da indexação é distribuído através de várias secções, para ser utilizado na recuperação da informação. Trata-se de campos tais como:

- Descritores temáticos (o assunto do documento)
- Descritores cronológicos (períodos, datas, épocas, etc.)
- Descritores onomásticos (nomes próprios de pessoas)
- Descritores topográficos (nomes de lugares)
- Identificadores (nomes próprios das coisas)

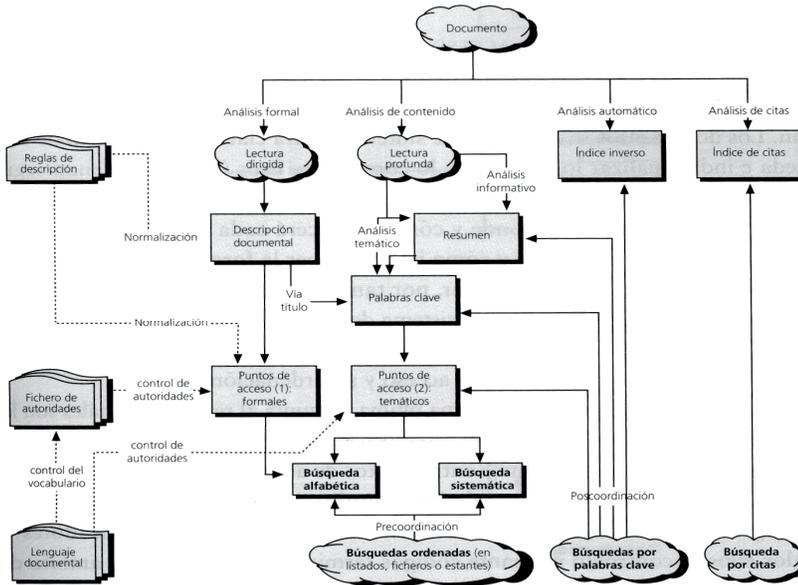
Exemplo:

Registro de um artigo de revista da base de dados ISOC:

*BASE:* ISOC  
*DOCUMENTO:* 277385  
*AUTOR:* Vilar Ramírez, Juan Bautista; Egea Bruno, Pedro María  
*TÍTULO:* Uma experiência protestante na Espanha de Franco, através de um epistolário inédito. A Igreja Evangélica Batista de Águilas (Murcia), 1950-1957  
*LUGAR-TRABALHO:* Univ. Murcia, Fac. Letras, Espanha; UNED, Fac. História, Murcia, España  
*REVISTA:* Carthaginensia  
*DADOS-FONTE:* 1996, 12 (21-22): 391-424  
*ISSN:* 0213-4381  
*IDIOMA:* Espanhol  
*TIPO-DOCUMENTO:* Artigo de revista  
*CLASSIFICAÇÃO:* 508205 História religiosa. Espanha. Século XX  
*DESCRITORES:* Franquismo; Minorias; Religiosas; Protestantismo; Epistolário  
*IDENTIFICADORES:* Igreja Evangélica Batista  
*TOPÓNIMOS:* Águilas; Região de Murcia  
*PERÍODO-HISTÓRICO:* 1950:1957  
*SÉCULO:* 20

*RESUMO:* O restabelecimento das relações diplomáticas com EUA foi decisivo para a reabertura do culto protestante na Espanha. Com os convênios hispano-norte-americanos de 1953 e, em especial, com a lei da liberdade religiosa de 1967, foi se consolidando a prática cristã protestante proibida em 2 de fevereiro de 1939. Nesse trabalho é apresentado um manuscrito inédito sobre a incidência de todos esses acontecimentos, além de outros de caráter interno, vividos pela Missão Batista de Águilas (Murcia). (A.)

Finalmente, no diagrama seguinte também se vê claramente a relação da indexação com a recuperação da informação.



*Tipos de análisis documental y modos de búsqueda*

Figura 17 – Tipos de análises documental e modos de busca  
 Fonte: PINTO MOLINA; GARCÍA MARCO; LACRUZ, p. 179, 2002.

## 2.8 CRONOLOGIA DA INDEXAÇÃO

No quadro seguinte, compilamos momentos que são importantes para configurar a teoria e a prática da indexação no nosso ponto de vista.

Quadro 16 – Cronologia da indexação

<p>3 0 . 0 0 0 anos AC</p>	<p>Etiquetas de barro</p>	<p>Os antigos escribas da Mesopotâmia para saberem o que continham as cestas de vime onde estavam depositados os documentos (tabuletas de barro) anexavam uma pequena etiqueta de barro com uma frase ou uma palavra. Desta maneira, conheciam o conteúdo sem abrir a cesta.</p>
--------------------------------	---------------------------	--

Egito	As cartelas do Egito	Os egípcios introduziram o papiro como suporte documentário. Ele era enrolado em uma vareta de madeira ou metal. Desta maneira, para saber o que continha ou ler era necessário desenrolar, então colocaram numa das extremidades da vareta uma etiqueta ou cartela onde escreviam as primeiras frases do documento ou umas palavras.
1876	<i>Rules for a dictionary catalog</i> de A. Cutter	Princípios para a atribuição de assuntos de Amis Cutter que apareceram em sua obra <i>Rules for a dictionary catalog</i> .
1895	<i>List of Subject Headings for Use in Dictionary Catalogs</i>	A <i>American Library Association (ALA)</i> publicou em 1895 a <i>List of Subject Headings for Use in Dictionary Catalogs</i> como ferramenta de indexação para bibliotecas de pequeno e médio porte com acervos não especializados.
1909	Subject Headings Used in the Dictionary Catalogues of the Library of Congress	Aparece a primeira <i>Subject Headings Used in the Dictionary Catalogues of the Library of Congress</i> tendo como principais referentes a lista mencionada da ALA e as regras feitas por Ammi Cutter. A partir desse momento, a lista de cabeçalhos se torna referência para as bibliotecas do mundo todo que a traduzem em outros idiomas (português, Brasil, 1948; Francês, 1946 e 1980; espanhol, árabe, etc.). Em 1975, passa a ser denominada <i>Library of Congress Subject Headings (LCSH)</i> .
1923	List of Subject Headings for Small Libraries	Minnie Earl Sears foi a autora da Lista de cabeçalhos de assunto <i>SEARS</i> . Trata-se de uma versão reduzida da <i>LCSH</i> para bibliotecas pequenas.
1934	Guia de cabeçalho de assunto para os catálogos dicionários	Juana Manrique de Lara publicou no México a primeira lista de cabeçalhos em espanhol. Trata-se de uma tradução e adaptação da <i>SEARS</i> , da lista da <i>ALA</i> e da <i>LCSH</i> .
1946	<i>Répertoire de vedettes-matière (RVM)</i>	Aparece a primeira edição do <i>Répertoire de vedettes-matière (RVM)</i> da Biblioteca da Universidade de Laval (Canadá) com onze mil cabeçalhos de assunto.
1949	SEARS traduzida em espanhol	Aparece na Argentina a primeira tradução da <i>SEARS</i>
1952	Unitermo	Mortimer Taube propõe e explica o termo unitermo.

1951	Descritor	Calvin Mooers propõe o termo descritor para especificar o tema de uma determinada informação em um contexto de recuperação de informação.
1957	Indexação automática	Hans Meter Luhn começa trabalhar com indexação automática, aplicando o método da frequência que a palavra aparece (relativa e total).
1957	NF Z 44-070 : 1957	É publicada a norma francesa Z 44-070 : 1957 <i>Catalogue alphabétique de matières</i> (1ª edição)
1960	Projetos de compatibilidade e integração de vocabulários controlados	Nos primeiros anos da década de 1960 iniciam os primeiros projetos de conversão entre vocabulários controlados por meio de tabelas de equivalência para ampliar as buscas em várias bases de dados.
1961	Sistema SMART	Gerald Salton desenvolve o sistema SMART para a análise automática de textos (indexação automática) e a recuperação da informação.
1967	<i>Guidelines for the Development of Information Retrieval Thesauri</i>	Diretrizes para a elaboração de tesouros feitos pelo <i>U.S. Federal Council for Science and Technology, Committee on Scientific and Technical Information (COSATI)</i> . Washington: Government Printing Office, 1967.
1967	Lista de cabeçalho de assunto para bibliotecas	Lista de cabeçalhos em espanhol para grandes bibliotecas gerais e especializadas compilada por Carmen Rovira e Jorge Aguayo sob a supervisão da União Panamericana.
1969	<i>ISO/R 919:1969 Guide pour l'elaboration des vocabulaires systématiques (exemple de méthode)</i> .	Esta norma posteriormente foi revisada e substituída pela ISO 10241: 1992 Terminology standards -Preparation and layout.
1970	Guidelines for the Establishment and Development of Monolingual Scientific and Technical	Encarregada e publicada pela UNESCO.
1974	Sistema de indexação PRECIS	Derek Austin desenvolve o sistema de indexação PRECIS na <i>British National Bibliography</i> .
1974	ISO 2788 : 1974	<i>Guidelines for the establishment and development of monolingual thesauri</i> (1ª edição).

1974	ANSI Z39.19-1974	<i>American national standard guidelines for thesaurus structure, construction and use.</i>
1978	NF Z 47-102 : 1978	Norma francesa sobre indexação denominada <i>Principles généraux pour l'indexation des documents.</i>
1980	<i>Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU)</i>	Como iniciativa da Biblioteca Nacional francesa aparece a primeira edição do <i>Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU)</i> . Foram tomadas como referência a RVM Laval e a LCSH.
1981	ISO 5127 / 3A – 1981	Norma ISO chamada <i>Information and documentation. Vocabulary. Acquisition, identification and analysis of documents and data.</i> Nesta norma são definidos conceitos como cabeçalho de assunto, de forma, indexação, extração de termos, indexação pré-coordenadas e pós-coordenada, etc.
1983	ISO 5127 / 1 – 1983	Norma ISO denominada <i>Documentation and information. Part 1: Basic concepts.</i> São definidos conceitos utilizados na indexação como linguagem e terminologia, linguagem natural, linguagem artificial, terminologia, sinonímia, quase-sinonímia, polissemia, etc.
1984	BS 6529 : 1984 NC 39-22 : 1984	Norma britânica na indexação denominada <i>Recommendations for examining documents, determining their subjects and selecting indexing terms</i>  Norma cubana denominada <i>Indización manual de documentos</i>
1983	<i>Bilindex</i>	Lista de cabeçalhos de assunto em espanhol equivalente a LCSH. Em 2007 foi publicada a 15ª edição.
1983	ISO 5127/6	ISO 5127/6-1983: Documentation and information: vocabulary; part 6 (documentary languages). A norma espanhola equivalente é a UNE 50113-6:1997.
1985	NF Z 47-200 : 1985	É publicada a norma francesa Z 47-200 : 1985 <i>Liste d'autorité de matières: Structure et règles d'emploi.</i>

1985	ISO 5963 : 1985	<i>Methods for examining documents. Determining their subjects and selecting indexing terms.</i> Tomando como ponto de partida a francesa NF Z 47-102 1978 e a britânica BS 6529:1984, o comitê técnico da ISO redigiu esta norma que foi ponto de partida para que muitos países a traduzissem ou adaptassem (Itália, UNI ISO 5963:1989; Portugal, NP 3715:1989; Espanha, UNE 50-121-91; Brasil, NBR12676:1992).
1985	ISO 5964 : 1985	A ISO publica <i>Guidelines for the establishment and development of multilingual thesauri</i> (1ª edição).
1986	Abandono dos símbolos tradicionais das listas de cabeçalhos pelos próprios dos tesouros	Com a 10ª edição da <i>LSCH</i> são substituídos os símbolos tradicionais das listas de cabeçalhos (x, See, xx, s.a., v.a.) pelos dos tesouros (SN, USE, UF, BT, NT, RT). A partir daí muitas outras listas de assuntos adotam esta simbologia.
1986	NF Z 44-070 : 1986	<i>Indexation analytique par matière.</i> Trata-se da atualização da norma que apareceu em 1957, mas com uma denominação diferente.
1986	NF Z 44-061 : 1986	<i>Catalogue-Forme et structure des vedettes noms de personne, des vedettes titres, des rubriques de classement et des titres forgés.</i> Prossegue a normalização dos pontos de acesso tanto nos catálogos de assunto como nos cabeçalhos de nomes de pessoas.
1986	ISO 2788-1986	A ISO edita as <i>Guidelines for the establishment and development of monolingual thesauri</i> (2ª edição).
1986	NF Z 44-070 : 1986	Publicação da norma francesa <i>Indexation analytique par matière</i> para atualizar a norma anterior em indexação de assuntos que datava de 1957, a NF Z 44-070.
1986	<i>Unified Medical Language System</i>	O Sistema Unificado de Linguagens em Medicina é um projeto de integração e exploração de vocabulários de ciências da saúde iniciado e mantido pela Biblioteca Nacional de Medicina dos Estados Unidos.

1995-	Universalização da Internet	O desenvolvimento da internet supôs a difusão e a popularização de conceitos (palavras chave, descritores, vocabulário controlado, tesouros, classificações temáticas), de técnicas (classificação, indexação, indexação automática) e das práticas (elaboração de tesouros ou de classificações) até esse momento, próprias de bibliotecários, documentalistas e arquivistas. Definitivamente, tudo isso criou um <i>Universo da indexação na web</i> .
1995	Meta dados	A partir da metade da década de 1990 começa a generalização do uso de meta dados como forma de definir, estruturar e fazer um intercâmbio de dados entre sistemas de informação. Nesses anos aparece também a lista de meta dados de Dublin Core. A maioria de linguagens específicas tem etiquetas para abrigar o resultado da indexação como, por exemplo, <subject>, <keywords>, <corpname>, <persname>, <geoname>, <date>, etc.
1997	Projeto MACS	A iniciativa da <i>Conference of European National Librarians</i> (CENL) inicia o Projeto MACS ( <i>Multilingual Access to Subject - Acesso temático multilíngue-</i> ) com as principais bibliotecas nacionais européias com a intenção de facilitar a recuperação diante da compatibilização de três linguagens de indexação: o <i>SWD/RSWK</i> alemão, o <i>RAMEAU</i> francês e a <i>LCSH</i> usada no Reino Unido e Suíça.

Fonte: Elaborado pelo autor.