



**INDEXAÇÃO E RESUMOS: TEORIA E PRÁTICA**  
F. W. LANCASTER  
Capítulos 2, 3 e 15

Claudia Correa, Diana Torrão, Greicyene Hamaguchi Ueki, Mariana Venturole

CBD0284 - Indexação: Teoria e Métodos

# Capítulo 2 - Princípios da indexação

- **Indexar:** indicar do que trata o documento ou sintetizar seu conteúdo
- Indexação de assuntos e redação de resumos implicam ambas em uma **representação do conteúdo temático**
- Termos atribuídos servem como ponto de acesso para **localização e recuperação**
- Num sistema informatizado, **resumos**, números de **classificação** e **palavras-chave** se complementam

# Etapas da indexação de assuntos

1. **Análise conceitual:** identificação dos **assuntos** estudados ou representados num documento
  - De que trata o documento?
  - Por que foi incorporado ao acervo?
  - Quais de seus aspectos serão de interesse para os usuários?

Há distinção entre o que o documento “trata de” [*about*] e sua “atinência” [*aboutness*]?

Exemplo: imagens *de* morte e imagens *que tratam da* morte

# Etapas da indexação de assuntos

2. **Tradução:** conversão da análise conceitual de um documento num determinado **conjunto de termos de indexação**

**Indexação por atribuição:** termos extraídos de uma fonte que não é o próprio documento, frequentemente de um vocabulário controlado

- Esquemas de classificação
- Listas de cabeçalho
- Tesouro

# Capítulo 3 - A prática da indexação

- Indexador não dispõe de muito tempo para ler os documentos, assim o indexador estará “passando os olhos” pelo texto;
- Oliver et al. (1996) afirma que após verificar as partes mais importantes o indexador “passaria os olhos” para confirmar as partes mais importantes do texto;

# Aonde “passar os olhos”?

- O indexador irá ler com maior atenção as **partes** que forem apresentar a **maior** quantidade de conteúdo no **menor** tempo;

# Quais são essas partes?

- Títulos (cuidado!!!)
- Resumos (confiabilidade??)
- Sinopse
- Conclusões
- Títulos das seções
- Legendas de ilustrações e tabelas
- Introdução as frases e parágrafos de abertura de capítulos e conclusões
- Palavras ou grupos de palavras que apareçam sublinhados ou grafados com tipos diferentes

## Assim é conveniente que ...

- O indexador passe os olhos no restante do texto para confirmar o que creditou como assunto, considerando, “partes lidas e partes que foram vistas de relance” assim como os termos atribuídos;

# Por que se examinar todo o documento?

- Para saber o que colocar na indexação;
- Para se adequar da melhor forma ao usuário daquele catálogo atendendo assim da forma mais plena as suas necessidades de informação;

# Primeira etapa: Análise conceitual

- Primeiro o indexador decide os assuntos que precisam ser representados;
- Nessa fase não se deve ignorar o assunto por imaginar a dificuldade em conseguir expressá-lo adequadamente;
- Autor destaca que nesta fase é possível fazer contribuições para melhora do vocabulário controlado
- ISO5963 também destaca a importância de indexar usando tesouros ou esquemas de classificação, o autor da obra discorda;

## Segunda etapa: Tradução

- Em segundo é preciso se verificar se o vocabulário pode representar adequadamente os termos levantados na análise;

# Exaustividade da indexação

- Política de indexação (controlado pelo gestor) que irá definir o **número médio de termos** atribuídos;
- Exatidão da indexação (controlado pelo indexador)
  - Análise
  - Tradução

# Política de indexação

- Exaustivo

- Maior número de termos para cobrir um conteúdo temático
- Vantagem de que quanto mais termos indexados, maiores as chances do documento ser recuperado
- Mais cara de ser feita do que a seletiva;
- Pode gerar um menor precisão das buscas;

Seletivo

apenas indexa o conteúdo temático principal

# Exemplo

- Arquitetura de catedrais ( assunto)
- Arquitetura religiosa (não é muito específico)
- Catedrais
- Arquitetura da habitação (aumenta a exaustividade)

# Métodos de indexação

- Critério estatístico (contagem de ocorrência de palavras no texto)

Torna possível a aplicação de **densidade**

Número de termos de indexação ou extensão  
Do resumo em relação a extensão do texto

# O que determina se um item será ou não recuperado?

- Quantidade de itens;
- Combinação de busca usando **E** (estratégia de busca);
- Se for indexação seletiva será usado a estratégia **OU** (estratégia de busca);

# Princípio da especificidade

- Um artigo que trate sobre o cultivo das laranjas deve ser indexado sobre o termo Laranja (TE) e não frutas críticas (TG);
- O uso do (TG) remete a redundância, cada vez que se usa o (TG) em vez de um (TE) a localização (diferenciação) do documento fica mais difícil;

# Capitulo 15

**Indexação automática, redação automática de  
resumos e processos afins**

# Indexação por extração automática

Critérios:

- Palavras ou expressões que aparecem no texto são extraídas e utilizadas para representar o conteúdo deste como um todo;
- Adoção de critérios de frequência, posição (título, resumo, legendas), e contexto;
- Combinação de **frequência absoluta e relativa**, critérios posicionais e sintáticos;
- Documento representado pela combinação de palavras e expressões mais importantes;
- Termos que sejam bons discriminantes, mas também que formem classes relacionadas ao documento.

# Indexação por atribuição automática

## Critérios:

- Coincidência, combinação e **co-ocorrência** : desenvolver para cada termo atribuído um perfil de palavra ou expressão, de acordo com a frequência da ocorrência (Ex. Chuva ácida; dióxido de enxofre etc);
- Desenvolvimento de índices: entradas e suas expressões associadas;
- Aplicação de análise sintática na geração de expressões para uso como índices;
- Ampliação dos termos encontrado em busca através da associação de ítems;
- Semelhante às redes de hipertextos e hipermissão.

# Redação automática de resumos

Procedimentos, segundo Luhn (1958):

- Lista de palavras-proibidas para eliminação das não-significantes;
- Seleção de palavras pela ocorrência e sua ordenação por frequência;
- Seleção de palavras de alta frequência ou significativas;
- Buscas de frase com palavras dentro desses critérios.

# Redação automática de resumos

Métodos, de acordo com Edmundson (1969):

- Da chave: similar a frequência;
- Da deixa: valor da significância de acordo com peso positivo ou negativo;
- Do título: títulos e subtítulos como indicadores de conteúdo;
- Da localização: peso às frases de acordo com a posição em que aparecem no documento.

# Abordagens atuais

## Auxílio online:

- Sugestão de termos e advertência ao erro;
- Substituição de termos pelos mais adequados;
- Verificação de como termos já foram usados anteriormente;
- Indexação parcial ou sugestão de termos por meio de comparação com vocabulários controlados.

## Sumarização de textos automática:

- Otimização da seleção de frases para melhor clareza do extrato;
- Formação de resumos por meio de extração de parágrafos ( frase como unidade);
- Resumo de realce: para despertar interesse dos leitores;
- Sumarização aplicada à áudio, vídeo, multimídia, às telas dos celulares e tablets;
- Unidades textuais semânticas: identificação e exibição de palavras-chave e frases mais significativas;

# Abordagens atuais

## Método intrínseco:

- Concordância de frases;
- Qualidade e legibilidade do texto;
- Informatividade por meio do resumo ideal;
- Método baseado na comparação entre resumo e texto completo;
- Fidelidade à fonte.

## Método extrínseco:

- Avaliar resumo em termos de prevenção da relevância do texto;
- Permitir a classificação correta do texto completo;
- Avaliação da compreensão da leitura.

## Método de processamento de linguagem natural:

- Termos de tesouro visando auxiliar a busca dos usuários;
- Processos para tornar mais inteligível a ambiguidade dos usuários.