



A COGNITIVE PROCESS MODEL OF DOCUMENT INDEXING

(“UM MODELO DE PROCESSO COGNITIVO DE INDEXAÇÃO DE DOCUMENTOS”)

JOHN F. FARROW

Anderson

João

Luana

Mayara

A COGNITIVE PROCESS MODEL OF DOCUMENT INDEXING, JOHN F. FARROW

John Farrow faz parte da corrente inglesa da década de 1970, destacada pelos estudos de indexação focados no viés do processo cognitivo e subjetivo em torno do seu desenvolvimento.

O artigo faz observações sobre a **Norma ISO 5963** (1985), a qual recomenda um método para análise de documentos em dois pontos principais: o exame do documento e a identificação de seus conceitos de forma sistemática. O início do artigo discorre sobre essas etapas da indexação e ao longo do texto mostra-se a abordagem do autor sobre o processo cognitivo de leitura que o indexador utiliza em seu trabalho.

JOHN F. FARROW

Indexação

Para John Farrow o processo de indexação consiste na “compreensão do documento a ser indexado”, seguida da “produção de um conjunto de termos de indexação”. A partir de 1970, segundo o autor, a indexação passou a ser pensada através de novos elementos trazidos pela psicologia cognitiva, que, em termos gerais, fazem a correlação entre as representações (linguagem natural e documentária) com os processos cognitivos humanos, que seriam a sua base de formulação (p.149).

Estágios para a Indexação

1. Analisar o documento:

- a. Exame de todos os elementos do texto, onde a leitura completa seria o ideal.
- b. Não reduzir a análise de assunto ao título e/ou o resumo do texto
- c. Tempo estimado quatro minutos

Dificuldades: O tempo ideal de quatro minutos implica em uma leitura superficial (por scanner) do texto. A leitura completa do texto também não é possível dentro da rotina do indexador seja por falta de tempo, indisposição, fluxo de serviço, etc...

Recomendação: fazer um *checklist* das partes principais do texto: título, resumo, lista de conteúdos, introdução, frases que iniciam os capítulos e parágrafos e a conclusão, ilustrações, diagramas e tabelas; palavras que estejam sublinhadas ou impressas em tipos diferentes.

Estágios para a Indexação

2. Identificar os principais conceitos presentes no texto

- a. Estabelecer um checklist com os fatores reconhecidos como importantes no domínio abrangido pelo índice, respondendo perguntas como:
 - O assunto contém um conceito ativo, por exemplo, uma ação ou uma operação?
 - O assunto possui um agente que pratica essa ação ou operação?
 - O texto se refere a métodos, técnicas ou instrumentos especiais?

- b. Na seleção dos conceitos, não interpretar de forma restrita a amplitude do interesse coberto pelo índice:

“quando os termos da literatura científica ou técnica estão sendo indexados seus aspectos sociais ou econômicos não devem ser negligenciados” (p.150)

Estágios para a Indexação

c. Escolha da linguagem:

Cabe ao indexador optar pela escolha entre a linguagem livre, a linguagem controlada e a linguagem pré coordenada ou pós coordenada.

No entanto, o autor indica que, em termos gerais, o indexador deve tentar escolher os conceitos mais adequados a comunidade de usuários, devendo considerar a realização de um feedback para sua avaliação.

Também alerta que no caso da linguagem controlada, Os tesauros embora sejam um grande auxílio para o indexador, em determinados casos, podem não trazer uma boa representação de assunto por fatores como especificidade temática do usuário.



Qualidade e consistência da indexação

Segundo o autor, a qualidade e a consistência da indexação dependem:

- ★ Imparcialidade do indexador
 - Advertência: é quase inevitável que alguns elementos de julgamento subjetivo afetem a realização dessa atividade
- ★ Conhecimento do campo coberto pelos documentos a serem indexados
 - Formação profissional contínua
- ★ O contato direto entre o indexador com os usuários, verificando as suas necessidades ;
- ★ A observação das mudanças na terminologia específica a cada área.



Compreensão do texto para indexação

Os requerimentos para um modelo cognitivo do processo de indexação se relacionam tanto com a compreensão do texto pelo indexador quanto com a produção da representação (índices, resumos).

Modelo cognitivo da indexação: o que acontece na mente do indexador entre a leitura e a produção.



Indexadores compreendem textos essencialmente da mesma maneira que leitores fluentes, com 4 modificações:

1. Indexadores trabalham sob limite de tempo, o que os leva a esquadriñar o texto em vez de lê-lo numa velocidade normal.
2. Indexadores compreendem o texto somente para sua tarefa de indexá-lo. Isso modifica o resultado de experimentos psicológicos que tentam medir a compreensão de um texto após leitura rápida.
3. A compreensão de um texto por um indexador tem como objetivo a produção de resumo, entradas etc. Portanto um modelo de compreensão do texto por indexador tem que levar isso em conta e estar relacionado com esse modelo de produção.
4. Muitos indexadores trabalham com um leque reduzido de tipos de texto e assuntos, e essa repetição leva a um processamento automático que vai além de uma fluência normal de leitura.



Os estudos reunidos sugerem que a compreensão de um texto envolve processamento tanto top-down quanto bottom-up.

O processamento top-down usa informações que não estão contidas no texto, mas que fazem parte do conhecimento de mundo que o autor possui e que ele assume que os leitores também possuirão. Processamento conceitual.

Processamento bottom-up, informação que está presente no texto. Processamento perceptivo.



Leitura dinâmica e esquadramento

Autor cita uma série de estudos: no geral, leitores dinâmicos não atingem níveis de compreensão mais elevados do que leitores comuns.

Similaridade entre leitura dinâmica e esquadramento para indexação:

Para ambos, detalhes e coerência local são distrações, o que importa é encapsular a compreensão em um punhado de palavras-chave isoladas.



Esquadrinhamento do texto para indexação e resumo

Indexadores esquadrinham textos seletivamente, procurando por pistas perceptivas. Algumas delas são sumariamente tipográficas, títulos, palavras em itálico, começo de parágrafos, mas muitas são verbais. As pistas verbais se dividem em 3 grupos:

1. Frequência da palavra: se um indexador nota a palavra *carpete* se repetindo no texto, essa palavra ficará na sua mente até o final da leitura.
2. Redes semânticas: o indexador interage com o processamento conceitual e perceptivo, por isso, uma vez que ele nota e guarda a palavra *carpete*, ele ficará mais sensível a notar palavras semanticamente próximas, como *tapete*.
3. Características estruturais: O indexador sempre carrega consigo um conjunto de palavras e frases que apontam para características estruturais do texto, como “introdução”, “os resultados sugerem” etc.

BASE PARA UM MODELO DE PROCESSO COGNITIVO

O modelo mais conhecido de compreensão de discurso é de Kintsch e van Dijk, que oferece um conjunto de procedimentos (macrorregras) simples para reduzir o texto aos seus elementos essenciais .

Contudo, Beghtol observou que a aplicabilidade desse modelo para a indexação seria problemática, já que na indexação os procedimentos de busca são frequentemente inconsistentes com o tipo de leitura que as macrorregras requerem.

Modelo de 1983 é mais baseado no processo do que nas regras e as macrorregras são substituídas pelas macroestratégias.



BASE PARA UM MODELO DE PROCESSO COGNITIVO

Macroestratégias contextuais

- Limitar as buscas semânticas ao contexto cultural geral do escritor.
- Decidir quais tópicos são característicos do tipo de discurso esperado em determinado contexto.

Macroestratégias textuais

- Estrutura do texto e pistas perceptivas.

Esse modelo admite o processamento do texto e a inferência de uma macroestrutura sem a necessidade de se ler cada palavra e também se adapta facilmente ao processo de indexação e forma a base para o modelo que é proposto.



MODELO DE COMPREENSÃO

Utiliza três divisões convencionais da memória

Registro sensorial

Transmite a informação visual para a memória de curto prazo.

Memória de curto prazo

Processamento e funções de armazenamento limitadas.

Memória de longo prazo

Compreende a totalidade dos conhecimentos e crenças de um indivíduo sobre o mundo.



MODELO DE COMPREENSÃO

No modelo de van Dijk e Kintsch a memória de longo prazo tem três divisões interativas

- **Memória de texto episódica:** o que é lembrado sobre um texto.
- **Conhecimento relevante:** características estruturais, assunto, usuário, conhecimento do mundo.
- **Sistema de controle:** elementos que surgem inconscientemente durante o processo.

As **pistas visuais** entram na **memória de curto prazo** a partir do **registro sensorial**, junto com o **conhecimento relevante** e **sistema de controle** do indexador, ou seja, já com a ideia do indexador sobre o assunto do documento e o objetivo e propósito do processamento.

À medida que o processamento continua, mais pistas passam entre a **memória de curto prazo** e a **memória de texto episódico** como termos candidatos.



MODELO DE COMPREENSÃO

Operações realizadas na memória de curto prazo:

- a. **Reforço:** um termo candidato para a indexação é consistente com os assuntos do documento e/ou corresponde a um termo já processado.
- b. **Modificação:** um termo candidato é modificado conforme o conhecimento ou propósito do indexador.
- c. **Chunking:** vários termos candidatos relacionados são mesclados para formar um único termo.
- d. **Rejeição:** um termo candidato é inconsistente com o assunto, conhecimento relevante ou objetivo do indexador.



MODELO DE PRODUÇÃO

A produção de discursos, de um modo geral, não é comum

- Falta de dados iniciais para comparação
- Necessidade de “manusear” ideias, desejos, vontades e experimentá-los

Indexação, Classificação e Resumos exigem modelos de produção mais simples, uma vez que utilizam-se dos modelos de compreensão em seu resultado final

- A classificação bibliográfica gera um código que, por si só, representa todo o documento
- A indexação gera índices que indicam os temas contidos nos documentos
- Um resumo pode transmitir, em poucas palavras, as informações, os métodos e a conclusão presentes no documento em questão

MODELOS DE PRODUÇÃO

Apesar de parecer mais simples, as três produções (classificação, indexação e resumo) estão sujeitas a restrições do sistema informacional na qual estão presentes

→ As classificações são sistemas fechados. O classificador pode apenas especificar os tópicos presentes no documento. O sistema se encarrega de prever e impor um padrão fixo para cada assunto

- No sistema Decimal de Dewey, por exemplo, um documento que trata de pássaros estaria na hierarquia zoologia-vertebrado-aves

→ A indexação se torna restrita de acordo com o sistema e se ele usa um vocabulário controlado, natural ou livre

- Usar apenas termos permitidos pela instituição/sistema, utilizar apenas termos encontrados no documento ou ter uma infinidade de termos a serem usados para criar relação com o texto

MODELOS DE PRODUÇÃO

→ Os resumos sofrem restrições do ambiente no qual estão inseridos. A editora a que o documento pertence, o público a que se destina, o estilo literário, número de leitores, etc

Apesar de ser comum os 3 processos serem feitos de forma separada, é o ponto em comum entre eles, ou seja, o tópico central do documento que define o modo como o documento será processado.



DISCUSSÃO

- O modelo apresentado é apenas um dos muitos modelos de indexação
- É necessário fazer um processamento visando o conceito dos assuntos e, para tal, é necessário que os indexadores tenham familiaridade com o tema que estão trabalhando
- É necessário uma educação apropriada aos indexadores, para que possam atuar com eficiência e de forma adequada na indexação
- É necessário que sejam feitas mais pesquisas e testes na área da indexação. Há pouco material conhecido sobre uma área tão básica.