

Princípios de indexação

Indexing Principles *

UNISIST**

Recomendações do UNISIST com o objetivo de estabelecer os princípios básicos do processo de Indexação visando a facilitar o intercâmbio entre sistemas de informação. Divide a operação de Indexação em dois estágios. O primeiro é o estabelecimento do assunto compreendendo os aspectos de conhecimentos do documento e a identificação e seleção dos conceitos. Considera os fatores exaustividade e especificidade na seleção de conceitos. O segundo estágio compreende a representação dos conceitos por termos de linguagens de indexação, discutindo as características destas linguagens. A qualidade do processo de indexação deve ser testada pela avaliação das recuperações feitas e todo o sistema deve admitir mudanças sempre que necessário

1. OBJETO DO PRESENTE DOCUMENTO

Este documento tem por objetivo o estabelecimento de princípios válidos e consistentes a serem seguidos quando se determina o assunto de um documento. Com o propósito de indexação e recuperação, os conceitos contidos nos documentos podem ser representados por termos selecionados da linguagem natural (ex: palavras-chaves) ou por símbolos (ex: número de classificação).

* UNISIST. *Indexing principles*. Paris, UNESCO, 1976. 11p.

** Traduzido por Maria Cristina Mello Ferreira Pinto, Prof. da Escola de Biblioteconomia da UFMG.

Estes princípios foram desenvolvidos, tanto quanto possível, de maneira a serem independentes de qualquer sistema de informações em particular. Sendo assim, eles constituem um conjunto único de regras ou recomendações que visam a promover:

- a cooperação mais fácil entre diferentes sistemas de informação;
- o desenvolvimento de regras mais específicas, porém compatíveis, no contexto de um sistema de informação em particular.

2. O PROCESSO E O OBJETIVO DA INDEXAÇÃO

A indexação é vista como a ação de descrever e identificar um documento de acordo com seu assunto. Assim, o presente documento não se refere à descrição física do documento como um material (Ex: descrição da forma, número de páginas, língua, data...) apesar, de muitas vezes, a determinação desses fatores por um indexador ser necessária se considerarmos que essa informação permitirá ao usuário determinar de maneira mais precisa se um documento específico será ou não útil para sua pesquisa.

Durante a indexação, os conceitos são extraídos do documento através de um processo de análise, e então traduzidos para os termos de instrumentos de indexação (tais como tesauros, listas de cabeçalhos de assunto, esquemas de classificação, etc.)

Decidida a indexação, os conceitos são registrados como "dados de informação" e são organizados de modo a permitir um fácil acesso na recuperação da informação. Estes registros podem aparecer nas mais variadas formas, ex. índices de livros, índices para catálogos e bibliografias, catálogos em fichas, etc.

Usando esses instrumentos para recuperação (isto é, quando selecionamos um número de documentos, ou partes de um documento relevantes para uma pergunta) a pergunta propriamente dita é tratada de maneira similar, isto é, é analisada no seu conteúdo, e esse conteúdo então é traduzido nos termos da linguagem de indexação.

As técnicas de indexação podem ser usadas de um lado para organizar os conceitos em instrumentos de recuperação da informação, e também por analogia, para analisar e organizar as perguntas em conceitos representados como descritores ou combinações de descritores, símbolos de classificação, etc. (Esta relação estreita existente entre a indexação de um documento e o tratamento de uma pergunta é mostrado num diagrama, no apêndice).

Os princípios gerais estabelecidos por este documento serão igualmente válidos para sistemas manuais ou mecânicos (ou a mistura de ambos), tanto na indexação quanto no estágio de análise da pergunta.

Essencialmente a indexação consiste em dois estágios:

— Estabelecimento dos conceitos tratados num documento, isto é, o assunto;

— Tradução dos conceitos nos termos da linguagem de indexação.

3. PRIMEIRO ESTÁGIO DA INDEXAÇÃO: DETERMINAÇÃO DO ASSUNTO.

O estabelecimento do assunto de um documento pode ser subdividido em 3 estágios:

1. Compreensão do conteúdo do documento como um todo, os objetivos do autor, etc.;

2. identificação dos conceitos que representam este conteúdo, objetivos, etc.;
3. seleção dos conceitos válidos para recuperação;

Na prática esses três estágios se superpõem

3.1. **Compreensão do documento** — A compreensão total depende até certo ponto da forma do documento. Duas diferentes situações podem ser identificadas, isto é, no caso de documentos gráficos e no caso de documentos não-gráficos.

3.1.1. **Documentos gráficos** — Representam o caso mais comum em bibliotecas e centros de informação onde a coleção consiste em sua maioria, de livros, monografias, jornais, periódicos, relatórios, teses... De maneira ideal, a compreensão total desses documentos depende de uma leitura extensiva do texto. Por razões econômicas, no entanto, isto geralmente é impraticável e nem sempre necessário. No entanto, o indexador deve certificar-se que nenhuma informação relevante foi negligenciada.

Partes importantes do texto necessitam ser consideradas cuidadosamente. Atenção especial deve ser dada a:

- título.
- introdução e as primeiras frases de capítulos e parágrafos.
- ilustrações, tabelas, diagrama e suas explicações.
- conclusão.
- palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente.

As intenções do autor são geralmente apresentadas nas seções introdutórias, enquanto as seções finais comunicam o quanto foram alcançadas.

Todos estes elementos devem ser considerados pelo indexador durante o estudo do documento. Não se recomenda a indexação exclusiva e diretamente derivada do título e um resumo, se houver, não deve ser considerado como um substituto satisfatório da leitura e exame do texto. Os títulos podem ser enganadores; tanto o título quanto o resumo podem ser inadequados e em muitos casos nenhum dos dois será uma fonte segura do tipo de informação requerida por um indexador.

3.1.2. Documentos não-gráficos — Uma situação diferente ocorrerá no caso de documentos não-gráficos, tais como áudio-visuais, visuais ou sonoros. Pode não ser possível ao indexador um exame minucioso, profundo, desse tipo de material, de maneira que muitas vezes a indexação a partir do título ou da sinopse torna-se uma necessidade.

3.2 Identificação de conceitos — Depois de examinar o documento, o indexador deve adotar uma abordagem lógica, selecionando os conceitos que melhor expressarão o assunto do documento.

A escolha dos conceitos pode obedecer a um esquema de categorias reconhecidas como importantes no campo coberto pelo documento, ex: o fenômeno, o processo, as propriedades, as operações, o material, o equipamento, etc.

Por exemplo, ao indexar trabalhos sobre "terapia por drogas" o indexador deve observar sistematicamente a presença ou ausência de conceitos relativos a doenças específicas, o nome ou tipo de droga, modo de administrar, resultados obtidos e/ou efeitos colaterais, etc. Do mesmo modo, documentos sobre "a

síntese de compostos químicos" devem ser examinados para determinação de conceitos que indiquem o processo de fabricação, as condições de operação os produtos obtidos, etc...

3.3. A seleção de conceitos — O indexador não necessita obrigatoriamente registrar, como termos indexadores, todos os conceitos identificados durante o exame do documento. A escolha dos conceitos que serão selecionados ou rejeitados depende do objetivo para o qual as informações são indexadas.

Vários tipos de objetivos podem ser identificados, desde a produção de índices alfabéticos impressos até armazenagem mecânica de dados para subsequente recuperação por computadores ou outros meios.

O tipo de documento a ser indexado pode também afetar o produto. Por exemplo a indexação derivada diretamente do texto de livros, artigos de periódicos, é provavelmente diferente da derivada somente do resumo do texto original.

As duas características de um índice provavelmente mais afetadas por estes parâmetros são:

- exaustividade
- especificidade.

3.3.1. Exaustividade — Um indexador, seguindo os procedimentos estabelecidos acima, estaria apto a identificar todos os conceitos de um documento que teriam um valor potencial para os usuários de um sistema de informação. Em alguns casos, dois ou mais assuntos dentro do campo coberto pelo índice aparecerão independentemente num mesmo documento. Estes devem ser tratados separadamente, e se necessário por especialistas diferentes, dentro das suas especialidades.

É importante observar que a dimensão do campo coberto pelo índice não deve ser interpretada muito limitadamente. Com o crescimento das redes de informações, pode acontecer que os dados indexados, criados inicialmente para um grupo de usuários (ex: cientistas e/ou técnicos) possam ser usados por outros grupos (ex: economistas). Com este uso potencial em mente, recomenda-se por exemplo aos indexadores de literatura técnico-científica que não subestimem outros aspectos de um assunto, tais como as implicações sociais e/ou econômicas.

Na seleção de um conceito, o critério principal deve ser sempre seu valor potencial como um elemento de expressão do conteúdo do documento. Ao fazer a escolha dos conceitos, o indexador deve ter sempre em mente as questões que podem ser feitas ao sistema. Com efeito, este critério reassegura a principal função da indexação. Com isto em mente o indexador deve:

- escolher os conceitos que serão provavelmente os mais apropriados a uma determinada comunidade de usuários;
- se necessário, modificar tanto os instrumentos de indexação quanto os procedimentos com base no "feedback" de questões.

Estas modificações não devem, no entanto, ser levadas a um ponto de distorção total da indexação. Não deve existir um limite arbitrário ao número de termos descritores que podem ser fixados a um documento. Isto deve ser determinado pelo montante de informações contidas em cada documento em particular. Nenhum limite arbitrário evitará a perda de objetividade na indexação e o desvio de informações que seriam de valor durante a recuperação. Se, por razões econômicas, o número de termos tem que ser limitado, a sele-

ção dos conceitos deve ficar a cargo do indexador, de acordo com a importância relativa dos conceitos na expressão do assunto geral do documento.

Em muitos casos, o indexador necessita incluir como parte dos dados indexados conceitos que estão somente implícitos no documento, mas que servem para colocar um determinado conceito no contexto apropriado.

3.3.2 Especificidade — Como regra, os conceitos devem ser identificados tão especificamente quanto possível. Conceitos mais genéricos podem ser selecionados em algumas circunstâncias, dependendo dos objetivos do sistema de informações. Em certos casos o nível de especificidade pode ser afetado pela importância dada a um conceito pelo autor. Se o indexador considera que uma idéia não foi completamente desenvolvida, ou foi tratada de forma superficial pelo autor, justifica-se a indexação num nível mais geral.

4. SEGUNDO ESTÁGIO NA INDEXAÇÃO — REPRESENTAÇÃO DE CONCEITOS POR TERMOS DE UMA LINGUAGEM DE INDEXAÇÃO.

Para assegurar a organização de conceitos de uma forma útil e acessível, é necessário o conhecimento profundo dos instrumentos de indexação. O mesmo se aplica ao tratamento das questões propostas pelos usuários.

Os instrumentos mais comumente usados na indexação são de duas categorias:

- os verbais, representados por tesouros, listas de cabeçalhos de assunto, etc.
- os simbólicos, onde os conceitos são representados por símbolos de classificação.

O indexador deve estar familiarizado com estes instrumentos, seus procedimentos e regras. Entretanto, ele deve estar consciente de que esses instrumentos podem, por vezes, impor algumas limitações na prática. Por exemplo, a lista de cabeçalhos de assunto adotada ou as tabelas do esquema de classificação podem não permitir a representação exata de um conceito encontrado num documento.

Se o instrumento de indexação é um tesauro, o número de termos indexadores de um documento e o desdobramento de entradas podem ser reduzidos sem danos, pois as relações genéricas e outras podem ser estabelecidas diretamente do próprio tesauro. Quando se usar um tesauro, deve-se selecionar o descritor mais específico para a representação de um determinado conceito.

Alguns sistemas usam elos, indicadores de função, peso, etc. O indexador deve estar familiarizado com quaisquer regras em uso no sistema.

Se os conceitos são representados por símbolos de um sistema de classificação, deve-se lembrar que estes geralmente indicam um conceito mais amplo (a classe principal) que pode muitas vezes não ser apropriado ao documento em mãos.

Estes dois tipos de instrumentos de indexação (tesauros e sistemas de classificação) podem ser usados simultaneamente para permitir a recuperação por um ou por outro. Um ou outro pode ser mais econômico ou eficiente, dependendo da natureza da questão.

Na prática, o indexador encontrará freqüentemente conceitos que não estão representados nos tesauros ou sistemas de classificação. Dependendo do sistema, esses conceitos poderão ser admitidos imediatamente ou o indexador deverá usar descritores mais genéricos, e os novos conceitos ficam como candidatos para uma nova edição.

5. CONTROLE DE QUALIDADE

— A qualidade da indexação depende de dois fatores:

- as qualificações do indexador
- a qualidade dos instrumentos de indexação.

Em cada sistema de informação, os dados indexados fixados para um determinado documento devem ser consistentes, independente da pessoa do indexador. Além disso, deve existir uma estabilidade durante a vida de cada sistema de indexação. A consistência, a esse ponto, é importante se a informação for permutada entre agências numa rede ou sistema de informação.

Um fator importante no alcance deste objetivo é a completa imparcialidade do indexador. Quase inevitavelmente alguns elementos de julgamento subjetivo irão afetar o processo da indexação; estes devem no entanto, ser minimizados tanto quanto possível. A consistência é mais difícil de ser obtida quando o grupo de indexadores é grande ou quando existem grupos de indexadores trabalhando em locais distintos (como num sistema descentralizado). Nesse caso há necessidade de centralização pelo menos para uma revisão da indexação.

O indexador deve ser de preferência um especialista no campo coberto pelos documentos que está indexando. Ele deve entender os termos encontrados nos documentos, bem como as regras e procedimentos do sistema de indexação em questão.

O controle de qualidade será alcançado mais eficientemente se os indexadores tiverem contato com os usuários. Eles podem juntos, por exemplo, determinar se alguns descritores produzem falsas recuperações ou criam problemas no estágio de saída.

A qualidade da indexação depende também de certas propriedades dos métodos e procedimentos de indexação. É essencial que um índice seja flexível bastante para receber novos desenvolvimentos da terminologia, como também atender a novas necessidades dos usuários: isto é, deve permitir atualização freqüente.

A qualidade de indexação pode também ser testada pela análise dos resultados de recuperação, por exemplo, pelo cálculo das taxas de revocação e precisão.

6. CONCLUSÕES

— Estas recomendações devem pertimir a indexação apropriada para qualquer sistema normal de recuperação de informações. Cada sistema pode, no entanto, estabelecer normas próprias, a fim de alcançar seus objetivos específicos, cuidando para que estas sejam formuladas de acordo com os princípios gerais estabelecidos neste guia.

UNISIST's recommendations to help establish the basic principles of indexing among cooperative information systems. Considers two steps in the indexing process. The first one deals with the identification of the subject including Knowledge of the document and the identification and selection of concepts. Considers the factors exhaustion and specificity in the selection of concepts. The second step deals with the translation of concepts into indexing terminology, discussing the characteristics of those languages. The quality of the indexing process should be evaluated from the point of view of its retrieval capability and the whole system should be able to allow modifications whenever they are necessary.

APÊNDICE

