

Indexação: economizando na entrada e na saída

GLAUCIA HELENA BARBOSA PEREIRA DE SOUZA *

Um índice do tipo KWIC (Key word in context) e outro, construído com base num thesaurus, são soluções práticas adotadas atualmente para uma indexação econômica. Sugere-se, como a opção ideal para os que busquem rapidez, eficácia e economia no processo de indexação de documentos, a introdução do índice SLIC (Selective listing in combination), combinado com idéias básicas de qualquer um dos dois índices acima mencionados.

Nada mais verdadeiro que a afirmação de Faithorne — “a indexação é o problema básico, como também o estrangulamento mais dispendioso na recuperação da informação”. Todos nós bibliotecários conhecemos a complexidade e a conseqüente morosidade que existem na etapa da indexação e/ou classificação de documentos de uma dada coleção.

A escolha do sistema a ser utilizado é uma decisão que requer muito critério, devendo-se pesar prós e contras, bem como inúmeros parâmetros, dentre os quais as características dos usuários, da coleção e o importante aspecto econômico.

Como solução econômica e prática, muitas instituições têm adotado métodos de indexação semi-auto-

* Mestre em Biblioteconomia e Documentação. Rio de Janeiro.

máticos, como os sistemas do tipo KWIC (Key word in context) ou KWOC (Key word out of context), os quais envolvem uma rotação das palavras-chave dos títulos dos documentos e sua subsequente ordenação alfabética. Como para os citados métodos as palavras do título são consideradas os termos indexadores dos documentos, esta solução se torna econômica, pois dispensa a utilização do bibliotecário. Um leigo facilmente preparará os dados para serem entrada no computador, uma vez que seu trabalho consistirá somente em copiar os dados (autor, título, etc.) da página de rosto da obra.

Uma outra solução atual, visando a uma indexação rápida mas que classifique o assunto do documento, é a utilização de um thesaurus de descritores do assunto da coleção para escolha daqueles que se apliquem a cada obra. Esta solução requer um profissional e também especialistas do assunto, não sendo portanto tão econômica e rápida quanto a primeira.

Qualquer uma dessas soluções, entretanto, demandará do pesquisador esforço na etapa da consulta ao índice resultante. Índices KWIC ou índices de descritores isolados requerem que o usuário percorra listas enormes de palavras para coordená-las até chegar ao documento procurado.

Os índices KWIC, mais ainda que os do segundo tipo, sempre criam problemas para o usuário na etapa da busca. Primeiramente, porque apesar de cada entrada estar acompanhada do restante do título dentro ou fora do contexto, percorrer o índice para coordenar duas ou mais palavras-chave não é uma tarefa fácil. Em segundo lugar, as palavras são usadas como fornecidas pelos autores nos títulos, e para assegurar recuperação adequada o pesquisador precisa pensar em todas as alternativas de grafia (ou até mesmo nos prováveis erros de grafia), singulares e

plurais, sinônimos e quase-sinônimos. Resumindo, os termos de busca têm que ser expandidos consideravelmente, multiplicando o esforço do usuário. A economia na entrada, certamente, resultará em trabalho extra na saída.

Por outro lado, a opção que o programa KWIC permite de se pré-coordenar as palavras-chave dos títulos, através da união com um hífen, na etapa da codificação, de duas ou mais palavras consecutivas do título, inverte o problema citado: — a economia na entrada será certamente prejudicada e o trabalho do usuário na saída nem sempre será recompensado, pois a união dos termos sofrerá uma variação extensa, de acordo com os vários títulos da coleção.

O segundo tipo de indexação, ao fornecer descritores ao índice, acarretará também esse mesmo esforço na etapa da busca. Muitas instituições tentam minimizar esse desgaste na saída, coordenando os termos segundo algum critério pré-estabelecido na etapa da entrada. A coordenação, entretanto, quer seja decidida por um indivíduo, quer por uma comissão, nem sempre coincidirá com a coordenação procurada pelo usuário na etapa da busca. Não é, portanto, infalível e torna o sistema mais moroso e dispendioso, voltando ao problema originalmente citado.

Qual, então, o caminho a seguir? Sugerimos a utilização do outro processo automático que permite-nos conservar uma linha econômica e não menos lógica e coerente. Trata-se do SLIC Index (Selective Listing in Combination).

Este método foi idealizado por John Sharp em um trabalho que desenvolveu para a ICI Fibres Ltd. Consiste num programa de computador que combina automaticamente os diversos termos-índice fornecidos para um dado documento, de todas as formas possíveis, arrumando-os numa rígida ordem alfabética.

Para a escolha dos termos que irão indexar uma obra, visualizamos duas opções:

1. Se os usuários não visam grandes especificações de assuntos e se os títulos da coleção são expressivos, que se use as palavras-chave desses títulos como termos-índice;
2. Se, pelo contrário, os usuários forem especialistas, carentes de maiores detalhamentos para seleção dos documentos a consultar, que se retire de um thesaurus altamente especializado e conceituado os termos-índice simples para os documentos.

Na primeira fase de montagem do SLIC Index, o número total de entradas para cada obra é representado pela fórmula 2^{n-1} (onde n é o total de termos de um cabeçalho de assuntos, ou num título, o total de suas palavras-chave). Assim, um título composto de 4 palavras-chave teria o total de 15 entradas no índice combinado, assim distribuídas (termos A, B, C e D):

ABCD
ABC
ABD
ACD
BCD
AB
AC
AD
BC
BD
CD
A
B
C
D

Após esta primeira distribuição das palavras-chave, o SLIC elimina todas as entradas que formam o início de entradas mais longas e o número total final de entradas no índice é então representado pela fórmula $2^{(n-1)}$ (onde n é novamente o número total de palavras-chave). Assim, para os 4 termos: A, B, C, D, teremos no índice o total de 8 entradas, resultantes da eliminação de 7 das 15 combinações anteriormente elaboradas. Assim:

ABCD
 ABD
 ACD (15)
 BCD
 AD
 BD
 CD
 D

resulta em:

ABCD
 ABD
 ACD
 BCD (8)
 AD
 BD
 CD
 D

É importante lembrar que o SLIC Index coordena ou combina os termos relacionados para a etapa da busca, mas não realiza a sua permutação total. Isso seria anti-econômico pois os mesmos 4 termos (ABCD) utilizados no exemplo acima, se totalmente permutados, acarretariam 24 entradas no índice, conforme:

$$4! \text{ (ou) } 4 \times 3 \times 2 \times 1 = 24$$

e não as 8 entradas conseguidas utilizando o SLIC, que, no entanto, abordou os 4 termos de modo a evitar a perda de qualquer um deles em relação a qualquer dos outros 3 da etapa da busca.

É importante, porém, observar-se que mesmo utilizando-se o SLIC, deverá ser limitado o número de termos-chave por documento. Acreditamos que um máximo de 5 é um ponto ótimo, pois $2^{(5-1)}$ geraria 16 entradas no índice para uma única obra enquanto que $2^{(6-1)}$ (supondo-se a utilização de 6 termos) já duplicaria este número para 32 entradas.

Importante, também, é salientar mais uma vez que para uma ação econômica e rápida não basta somente utilizar um programa objetivo e lógico, como o SLIC, mas é sobretudo necessário que o bibliotecário ou documentalista evite a todo custo a complexidade na etapa da determinação dos termos-índice para entrada dos dados no sistema.

A KWIC (Key word in Context Index) index or a thesaurus based index are generally accepted procedures where economical indexing is required. The use of the SLIC index (Selective Listing in Combination Index) jointly with KWIC and thesaurus based indexes is suggested for those seeking speed, efficiency and economy in the document indexing process.

BIBLIOGRAFIA

- KEMP, D.A.; SIMPSON, I.S.; WILSON, T.D. Indexing; permuted, rotated, cycled. *Journal of Documentation*, 28(i): 67-68, Mar. 1972.
- LANCASTER, F.W. *Comunicação Informal*, 1975.
- SHARP, J.R. *Information Retrieval; Notes for Students*. London, Deutsch, 1970. p. 67-9.

———. The SLIC index. *American Documentation*, 17(1): 41-4, Jan. 1966.

———. *Some fundamentals of information retrieval*. London, Deutsch, 1965. p. 205-13.

SOUSA, G.H.B.P. de. *Comparação entre um Índice KWOC (Key word-out-of-context) e um Índice em Cadeia derivado da CDU (Classificação Decimal Universal)*, Rio de Janeiro, 1975. 85 p. (Dissertação apresentada ao IBBB/UFRJ para obtenção do grau de mestre em Biblioteconomia e Documentação).