

O INDEXADOR FACE ÀS NOVAS TECNOLOGIAS DE INFORMAÇÃO

Hagar Espanha Gomes

RESUMO

GOMES, H.E. O indexador face às novas tecnologias de informação. *Trans-in-formação*, 1(2), maio/ago. 1989.

A indexação tradicional, manual, sofre a interferência de muitas variáveis, dentre elas, o propósito do sistema de recuperação de informação, o tipo de linguagem adotada, a política de indexação, a natureza da área de conhecimento e a natureza e propriedade da informação científica e tecnológica. Tais variáveis tornam a indexação uma atividade complexa, sendo o indexador acusado de introduzir subjetivismo e inconsistência no seu trabalho. Pesquisas na área da indexação automática têm mostrado que não existe diferença significativa no desempenho de um e de outro tipo de indexação. Mas a indexação automática vem se sofisticando desde a experiência com o KWIC, com vista a uma recuperação mais eficaz e de custo mais baixo. Com isso, aumenta a participação de um novo tipo de profissional de informação, ainda sem nome.

Unitermos: *Indexação – Indexação automática – Linguagem de indexação.*

Indexação é um tema não muito presente na literatura especializada brasileira, nem nas comunicações apresentadas em congressos ou em dissertações. Também nos textos institucionais sobre os grandes sistemas nacionais se encontram informações sobre procedimentos de indexação/recuperação e sobre linguagens documentárias com a frequência desejável.

Nas atividades de automação de bibliotecas a preocupação, via de regra, se concentra nos procedimentos de controle administrativo – em que o computador vai minimizar o esforço dos bibliotecários – ignorando-se os procedimentos de indexação/recuperação. A catalogação descritiva, por exemplo, tem sido tratada com grande destaque – muito mais do que o necessário – mas, no que se refere à catalogação temática, o silêncio é gritante. O que se vê é muito esforço e dinheiro para produzir fichas, mas isso não é automação de biblioteca.

Ora, a atividade documentária se desenvolve, de fato, sobre dois eixos perpendiculares: no eixo horizontal, correspondendo às atividades de planejamento, administração, avaliação, e no eixo vertical, aí incluídas as atividades de indexação/recuperação e/ou de elaboração de linguagens documentárias.

Nesses dois eixos básicos estão as atividades que exigem estudo e reflexão, não se podendo considerá-las como simples técnicas.

O conjunto dessas atividades se integra e se interrelaciona, sendo sistema; pode-se discutir, na prática, cada componente do sistema de informação, mas não se pode fazê-lo de maneira isolada.

A indexação é vista aqui nessa perspectiva e a partir do lugar comum de sua característica de alta subjetividade. A indexação automática, para os críticos da indexação manual, eliminaria este aspecto negativo, ganhando importância apenas o subsistema de recuperação(1).

Esta dicotomização, no entanto, não ajuda a alcançar ou a perseguir a meta do sistema de informação, que é a melhoria de seu desempenho. Especificamente isso quer dizer: aumentar a revocação sem perder a precisão. O que importa, no momento da decisão sobre a melhor alternativa, é o conhecimento das peculiaridades e do potencial de cada modalidade.

A indexação feita pelo homem pode ser entendida – e muito bem – na perspectiva sistêmica e é isso que se pretende mostrar. Já a indexação automática, isto é, aquela 'feita' pelo computador, tem que ser entendida de alguma outra forma – que cabe averiguar –, pois aí falta o subsistema de indexação no seu significado tradicional.

Embora sem pretender fazer revisão de literatura sobre o assunto, procura-se, aqui, caracterizar cada modalidade, mencionando-se as experiências mais significativas de indexação automática. Fica evidente que um novo tipo de profissional entra em cena, cuja atuação tem que ser considerada, especialmente por causa das novas funções decorrentes das novas tecnologias de informação.

A nosso ver, a importância da indexação tem que ser restabelecida, sob pena de redundar em fracasso os esforços que vêm sendo feitos para automatizar serviços e implantar redes e sistemas de informação.

As questões de indexação levantadas aqui – em sua modalidade manual ou automática – são válidas apenas para a informação científica e técnica, isto é, de natureza semântica-e-lingüística. O adjetivo 'científico' deve ser entendido em seu sentido amplo, em oposição ao 'estético'. O binômio 'semântico-e-lingüístico' deve ser enfatizado porque, embora a uma informação estética se possa adicionar; **a posteriori** é de forma arbitrária, um componente semântico(2), apesar disso, sua indexação apresenta problema distinto daquela que é apenas semântica e lingüística ao mesmo tempo. Dada sua natureza, a informação estética foge ao controle do indexador, já que tal 'leitura' é extremamente individual. Também não se enquadra no tema a

Literatura, por sua dupla natureza: estética e 'semântica-e-lingüística', podendo ser considerados como 'científicos' apenas os textos **sobre** Literatura, os quais, nesta condição, estão sujeitos à mesma problemática dos textos científicos.

INDEXAÇÃO MANUAL

O problema fundamental da indexação manual – e que consiste num desafio para os sistemas de recuperação de informação – é que não existe uma única 'leitura' de um documento e, como consequência, a indexação de qualquer documento pode variar de indexador para indexador e no mesmo indexador, no tempo.

Esse mesmo problema existe do lado do autor e do lado do usuário. O autor escreve para se comunicar com seus pares mas, dada a natureza e as propriedades da informação científica e técnica, seu documento pode ser lido noutro contexto, com outra finalidade, etc.

Além destes fatos, a indexação de um documento está sujeita à interferência não apenas do indexador mas de outras variáveis, dentre elas:

- o propósito do sistema de recuperação de informação (SRI), que é limitado pelo perfil da clientela a que se destina,
- o modelo do SRI: se pré - ou pós-coordenado,
- a política de indexação,
- a natureza da área do conhecimento,
- a natureza e propriedade da informação científica e técnica,

brevemente descritas a seguir.

Propósito do SRI

Cada sistema de recuperação é, num certo sentido, único. Ele pode ser útil a usuários de outros sistemas, por circunstâncias, mas provavelmente a maneira da 'ver' o documento, num dado SRI, pode levar o indexador a representar aspectos que, noutro sistema, seriam irrelevantes e que podem frustrar o usuário de outro sistema.

Pode ocorrer, também, que o usuário de um SRI encontre seu tema de interesse tratado noutro SRI. No entanto, a semelhança de temas não é garantia de uma busca bem sucedida. Por exemplo, um documento sobre 'Carnaval' escrito por um sociólogo tem enfoque diverso daquele, sobre o mesmo tema, escrito por um antropólogo, e essa diferença pode ser fundamental para o usuário. No entanto tal diferença não passa pelo vocabulário de indexação.

O método e o enfoque, principalmente, levam a coleções com características próprias de uma área do conhecimento e a indexação nem sempre registra estas diferenças, que não são temáticas. O enfoque tem reflexos,

também, na estrutura da linguagem documentária, pois um mesmo termo pode representar aspectos diferentes de uma mesma idéia – evidentes na definição e/ou na estruturação – podendo não satisfazer a um usuário de outro sistema.

Modelo de SRI

O tipo de indexação adotado num dado SRI limita os produtos e serviços.

Para busca retrospectiva ou serviço de disseminação seletiva de informação o sistema pós-coordenado é o mais indicado; para os repertórios bibliográficos ou fichas para catálogos o sistema pré-coordenado é melhor.

Se um SRI quiser prover vários desses serviços, um planejamento cuidadoso deverá levar em consideração todos eles, mesmo que sejam aparentemente conflitantes, e identificar alternativas ou dispositivos que permitam produzi-los a partir de uma única entrada de dados.

O modelo do sistema limitará também a escolha da linguagem documentária e, portanto, as decisões do indexador.

Política de indexação

A política de indexação é ditada pelo gestor do sistema. É determinada pelo perfil de necessidades da clientela e refere-se aos aspectos de exaustividade e de especificidade. A linguagem documentária é afetada pela política de indexação e limita, pois, as escolhas do indexador.

Natureza da área do conhecimento

A natureza da atividade científica varia nos diferentes ramos do conhecimento e esta peculiaridade se reflete, de alguma forma, na literatura, seja implícita seja explicitamente. Algumas destas peculiaridades são importantes e merecem ser indexadas. Enquanto nas Ciências Exatas a característica mais importante é o assunto, nas Ciências Sociais outras características são igualmente importantes como, por exemplo, o enfoque dado ao tema ou a escola de pensamento do autor, etc.

No Direito, a forma literária é relevante; um texto de lei e um acórdão podem ser indexados sob o mesmo assunto, mas a forma é também fundamental e decisiva para o usuário. A habilidade do indexador em tratar com estas peculiaridades pode resultar em maior revocação sem perda de precisão.

Natureza e propriedade da ICT

Segundo Mikhailov e colegas(3) a ICT apresenta 12 características, das quais a natureza semântica e a cumulatividade têm implicações na inde-

xação. Ivankin(4) acrescenta outra, a saber, a reproducibilidade que, de certa forma, tem a ver com a natureza da cumulatividade.

Graças a essas propriedades, os textos de documentos sob um tópico específico podem apresentar o mesmo teor de informação, citando as mesmas obras. Isso cria um problema na recuperação, fazendo com que o serviço de informação forneça documentos com informação repetida. É possível evitar isso na indexação? De certa forma, o índice de citação(5) atende a essa necessidade. Por outro lado, não seria útil que um documento fosse indexado pelo que ele tem de diferente e não pelo tópico em que ele não apresenta novidade?

O Indexador

O indexador é mais um elemento de perturbação no SRI. Embora algumas regras de indexação devam ser obedecidas, ainda assim não existe garantia de que se consiga uma única indexação por mais de um indexador nem pelo mesmo indexador.

Estudos de interconsistência e de intraconsistência têm demonstrado que o terreno da indexação é escorregadio, embora não tenham por objetivo provar se a indexação é boa ou não. A qualidade da indexação só pode ser conhecida através de um estudo de avaliação geral do SRI. A respeito de consistência **versus** qualidade da indexação, Cooper(6) enfatiza que se pode ser consistentemente bom e consistentemente mau e ressalta que, num grupo de indexadores, dois podem ser mais consistentes entre si do que com os demais. Um dos indexadores mais 'individualista' nos grupos estudados, – isto é, cuja indexação não era consistente com os demais – provou se mais eficaz porquanto os documentos indexados por ele, quando recuperados, eram julgados mais relevantes, pelos usuários, e sua indexação impedia a recuperação de documentos considerados não-relevantes.

Observações gerais

Como se pode observar essas variáveis fazem da indexação uma atividade complexa. É fácil entender que haja perda de informação; afinal, existe uma redução nesse processo.

A linguagem natural é um instrumento utilizado na representação da realidade, mas não é tal realidade. É possível que, já nesse nível, se verifique perda de informação.

No processo de representação da representação da realidade (assim mesmo, repetido) o indexador pode ser responsável por uma perda maior. Existe, ainda outro fator agravante: na indexação não se procura representar cada documento individualmente, porém, situá-lo numa classe de documentos. Além da perda de informação que tal procedimento acarreta, o que se

recupera contém, ainda, em grande parte, informação repetida. A solução da indexação estaria nos processos automáticos? Afinal, a máquina trabalha no plano concreto dos significantes, não tendo esta ou aquela 'leitura'.

INDEXAÇÃO AUTOMÁTICA(7)

A indexação automática consiste, em tese, na extração, pelo computador, de palavras significativas dos textos. O índice deriva, portanto, dos textos, não se buscando externamente o vocabulário de indexação.

Quando Luhn, na década de '50, desenvolveu o índice KWIC – Key Word In Context – a novidade não estava no aproveitamento das palavras-chave presente nos títulos dos documentos, mas na rapidez com que tal tarefa poderia ser executada.

A idéia era antiga – e a prática também. Encontram-se no século XIV índices alfabéticos das disputas e teses dos escolásticos, formados por palavras-chave dos títulos e do conteúdo; no entanto, o índice alfabético só se torna prática geral quando os códices tomam a forma de livro(8). E nessa época eles ainda não possuíam títulos: eram identificados pelas primeiras palavras do texto, como ainda hoje ocorre com os documentos papais, por exemplo.

No século XIX já era bastante comum o índice alfabético das palavras-chave dos títulos (Schlagwort, Catchword)(9). O Poole's Index to periodical literature, para periódicos de caráter geral, usava como entrada de assunto as palavras-chave dos títulos dos artigos.

É verdade que os índices do século XX se caracterizariam pela atribuição de termos pelos indexadores, sendo conhecidas as atividades de sistematização de Cutter, no final do século XIX, para os catálogos de assunto, que seriam tão populares a partir daí(10). Mas também é verdade que os assuntos dos cabeçalhos – nos catálogos e bibliografias alfabéticos – se baseavam na linguagem natural, a partir dos textos analisados, o que, de certa maneira, os caracterizavam como parentes próximos dos índices derivados.

Embora Taube seja conhecido pela criação da indexação coordenada(11) é bom lembrar que o princípio de indexação adotado por ele no sistema Uniterm era o de extrair dos textos as palavras que formariam o índice e aqui ele está próximo, também, dos índices derivados.

Depois de Luhn, várias têm sido as experiências de indexação automática – derivada – comparando, inclusive, sua eficácia com a indexação manual, que é aquela em que o indexador atribui termos. Tais experiências, no entanto, têm contado com a participação do 'indexador' de forma cada vez maior, como se pode observar nos exemplos a seguir.

A técnica adotada por Luhn consistiu na elaboração de uma lista de palavras não-significativas, isto é, palavras que não poderiam figurar como

entrada no índice. Desde o início, no entanto, verificou ele que outras palavras, fora do título, poderiam ser úteis e essa adição de termos constituiu o chamado 'enriquecimentos de títulos'(12). Tauber, por sua vez percebeu que algumas palavras tinham que vir presas a outras formando uma expressão(13).

Mais tarde, o que se viu foi a preparação dos chamados 'dicionários' que listavam palavras que poderiam figurar nos índices. A preparação destes dicionários exige certo cuidado e uma análise da área do conhecimento. Pode-se dizer que, embora a indexação seja automática, ela se baseia numa lista feita pelo homem.

Este esforço passou a exigir novo tipo de pesquisa. Uma delas, por exemplo, mostrou que uma lista eficaz de palavras significativas não precisa ser longa(14).

No Centro de Documentação do Departamento de Defesa dos Estados Unidos (DoD) desenvolveu-se uma experiência com 'dicionários de palavras'. As frases introduzidas no computador eram comparadas com as palavras aprovadas anteriormente. Aquelas que coincidiam, passavam a ser palavras indexadas. Aquelas que não coincidiam eram submetidas a análise pelos indexadores, para futuras inclusões(15).

Outra sofisticação, no estudo dos princípios para construir índices automaticamente, foi o de identificar palavras significativas que deveriam figurar como 'subdivisões', isto é, em posição secundária(16).

A contribuição do homem nos índices derivados recentes tem sido a de produzir listas de palavras significativas para que no confronto com os textos dos quais derivarão os índices, estes sejam 'corrigidos', isto é, a forma das palavras – grafia, número, etc – seja uniformizada para facilitar futuras recuperações. Um exemplo desta atividade é o BIOSIS, que produz um índice KWIC, no qual sinônimos e formas variantes são verificadas num arquivo pelo computador, e devidamente uniformizados e corrigidos.

Muitos projetos se seguiram permanecendo, porém, em bases experimentais. Um dos mais interessantes foi o projeto de Salton(17), no início dos anos '70, em que ele procurou trabalhar com co-ocorrência de palavras, com o objetivo de criar grupos (clusters) de palavras para ajudar na recuperação, uma vez que tais grupos sugeririam outras palavras associadas de alguma forma. Estas associações são frequentemente estatísticas e ocorrem no plano sintagmático.

Um dos motivos pelos quais tais estudos têm ficado na experimentação pode ser o fato de que os serviços de indexação automática necessitam trabalhar com texto integral, para produzir índices derivados, o que em bases operacionais exigiria enorme capacidade de memória.

A disponibilidade de computadores, no primeiro mundo, tem mostrado que é mais barato trabalhar com índices automáticos, derivados, do que com índices feitos pelo homem. Por outro lado, a máquina pode prover um resul-

tado objetivo e de certa forma consistente, mesmo que o índice deixe um pouco a desejar.

CONSIDERAÇÕES FINAIS

Resumindo, os índices manuais foram derivados, a princípio, e a seguir, sistemáticos ou temáticos. Passaram depois a contar com vocabulários controlados, sob certos aspectos, e até com tesouros.

Com o uso dos computadores retomou-se, em bases modernas, rápidas e mais eficientes, o índice derivado, que passou a contar com alguma forma de auxílio e controle feitos pelo homem.

Desde as primeiras experiências com indexação automática verificou o homem que a seleção automática dos termos de recuperação não é uma atividade tão automática assim, cabendo a ele dar sua contribuição para melhorar o nível de recuperação.

Praticamente inexistem serviços operacionais que comprovem a eficácia das bases de dados indexadas automaticamente. Em relação a índices impressos, é curioso ressaltar que algum tempo após o início da publicação *Current Contents* – da American Chemical Society – que usa o índice KWIC –, os autores melhoraram a formulação de seus títulos(18), colaborando, de certa maneira, com o computador.

A diferença fundamental entre os dois processos – o convencional e o automático – é a forma de atuação do elemento humano, que seria o indexador no primeiro caso e, nos outros, um tipo de profissional ainda sem nome e sem perfil definido.

Não existem normas que assegurem qualidade na indexação automática até o momento visto que as pesquisas tem ficado no plano experimental. Quanto aos índices convencionais, tendo em vista as variáveis abordadas, a atividade continua a ser um desafio. Alguns estudos de avaliação de índices com vocabulário controlado comparados com o desempenho de índices derivados tem mostrado resultados de desempenho semelhantes. Uma das experiências sobre as quais mais se escreveu foi a do teste de Cranfield cujo objetivo foi o de comparar quatro linguagens de indexação: uma classificação facetada, a CDU, o cabeçalho de assunto e o Uniterm. Um dos resultados da fase I é que não foi possível decidir-se sobre o melhor sistema. O uso da linguagem de indexação controlada não se mostrou mais eficaz na recuperação do que o tipo não controlado. Outros estudos(19) tem sido limitados deles não se podendo tirar conclusões generalizadas.

Face a esses dados, não tem sido possível afirmar com segurança quem tem se desempenhado melhor: a linguagem natural ou a linguagem controlada, o indexador ou o computador. Talvez esse não seja o caminho correto.

A natureza da área e sua linguagem podem ser um caminho a investigar. O discurso é uma representação da realidade. Reflete, de certa forma, algumas peculiaridades desta realidade. Se, em algumas áreas, o discurso é mais descritivo e a linguagem especializada melhor estabelecida, noutras áreas, a linguagem reflete ideologias, teorias, sutilezas mais difíceis de serem captadas pelo computador.

Se a experiência com o Current Contents mostrou que a colaboração do autor melhorou o desempenho do sistema, não é de se esperar que, em algumas áreas do conhecimento, o homem venha a disciplinar seu discurso para melhor manipulação pela máquina? Se essa proposta é um tanto esdrúxula, não caberá ao novo profissional executar tal tarefa? De certa forma, não é o que ocorre com experiências como a do BIOSIS?

O maniqueísmo deve ser evitado: cada sistema pode ter melhor desempenho em determinada circunstância. Mais importante é atentar para novas funções do profissional de informação ainda sem nome. Se na indexação convencional as funções são idênticas tanto no sistema manual como no sistema automatizado, ou seja, prover uma boa representação do assunto do documento, na indexação automática – ou derivada – o profissional da informação trabalha com outras variáveis que não a análise do documento. Seu material – o texto – é de natureza lingüística e é aí que ele vai atuar devendo, pois, buscar na Lingüística as bases para seu trabalho. Sendo, porém, idênticos os objetivos – recuperação de informação – alguma base teórica estabelecida para os sistemas convencionais pode ser útil.

Em vez de voltar as costas a uma tecnologia que já está presente e popularizada através do microcomputador, os profissionais de informação devem estar atentos e trabalhar diligentemente para atuar com competência nesta nova função.

De uma coisa, no entanto, pode-se estar certo: uma vez que o que se deseja recuperar é o conteúdo dos documentos, instrumentos de controle de vocabulário – diferentes dos tesouros e similares – precisam ser construídos para permitir que o computador “faça” sua indexação de maneira eficaz. Se, na indexação automática, a ambigüidade da palavra desaparece, pelo contexto, a sinonímia continua presente. Embora a existência do sinônimo absoluto inexista na linguagem natural, essa questão não parece ser insolúvel na linguagem especializada.

Além do controle da sinonímia, técnicas para identificar a função sintática das palavras na representação do conteúdo são, também, úteis para melhorar a revocação. Expressões usuais que auxiliam na identificação destas funções, e não presentes nos sistemas convencionais manuais, são um exemplo de levantamentos necessários(20).

O indexador não pode deixar de estar atento para o fato de que a natureza das diferentes atividades humanas se reflete nos textos e isso tem implicações na indexação – manual ou não – e na forma como o usuário busca

informação. Estas peculiaridades podem ser um estímulo para novas pesquisas na indexação-com-auxílio-do-computador que, afinal de contas, não é tão automática como pode parecer à primeira vista.

Se o indexador tradicional deixa de analisar o conteúdo, ficando o computador com este papel, de outro, ganha importância o profissional de informação que vai prover os "dicionários" com os necessários controles do vocabulário.

REFERÊNCIAS E NOTAS

1. O Prof. Lancaster identifica seis subsistemas no sistema de recuperação de informação: o subsistema de seleção e aquisição, o subsistema de indexação, o subsistema de vocabulário, o subsistema de negociação da pergunta, o subsistema de estratégia de busca e o subsistema de cotejo (match).
2. MOLES, A. **Teoria da informação e percepção estética**. Rio de Janeiro, Tempo brasileiro, 1969
3. MIKHAILOV, A.I., CHERNYI, A.I. GILYAREVSKYI, R.S. Estrutura e principais propriedades da informação científica In: **Ciência da Informação ou Informática**. Coletânea organizada por Hagar Espanha Gomes. Rio de Janeiro, Calunga, 1980. p. 71-89.
4. IVANKIN V.I. Information science in scientific communications. **Nauchno-Technicheskaya Informatsiya**, Ser. 1 n 3: 7-9, 1979.
5. A indexação de documentos via citação bibliográfica é mecânica não trabalhando com a variável assunto. Nesse caso, o desempenho da máquina nada deixa a desejar.
6. COOPER, W.S. Is interindexer consistency a Hobgoblin? **American Documentation** 20; 268-78, 1969.
7. Existem outros sistemas de indexação concebidos para serem realizados pelo homem com auxílio do computador e que podem ser chamados de índices automatizados e não automáticos. Nesta categoria incluem-se o PRECIS (para produzir o índice da British National Bibliography), o índice articulado do Chemical Abstracts, ambos operacionais, o NEPHIS – Nested Phrase Index System variante do PRECIS, ainda em bases experimentais.
Nestes sistemas o indexador desempenha as mesmas funções de análise documentária comuns aos sistemas manuais. Por este motivo tais experiências não são aprofundadas aqui.
8. WITTY F. J The beginnings of indexing and abstracting: some notes towards a history of indexing and abstracting in Antiquity and the Middle Ages. **The Indexer** 8(4): 193-98, Oct. 1973.
9. METCALF, J. **Subject classifying and indexing of libraries and literature**. New York Scarecrow Press, 1959 p. 33, 37. Segundo Metclaf, no século XIX cada obra recebia uma indexação individualizada, já que a entrada do índice era uma palavra-chave (Catchword) do título seguida do resto do título, como modificação. Observe-se que o índice permutado adota o mesmo procedimento. Esse procedimento mudou a partir de Cutter, conforme se percebe na Regra 172 das **Rules for a dictionary catalog**: 'Enter a book under the words which best express cheir subject, **whether it occurs in the title or not**' (O grito é nosso).

10. CUTTER, C A. **Rules for a dictionary catalog**. 4 ed. Washington, GPO, 1904.
11. Hoje chamada de indexação pós-coordenada.
12. O IBBD adotou, durante algum tempo, a prática de enriquecimento dos títulos nas bibliografias com índice KWIC.
13. TAUBE, M Notes on the use of roles and links in coordinate indexing. **American Documentation** 12(2): 98-100, Apr. 1961.
14. FEINBERG, Hilda. **Title derivative indexing techniques: a comparative study**. Metuchen, N.J., Scarecrow Press, 1973. Seu estudo mostrou que 16 palavras significativas eliminavam 29% das palavras dos títulos mas que, para eliminar 48% das palavras, ela precisaria de mais de 1500 palavras.
15. KLINGBIEL, P. H. Machine-aided indexing of technical literatura. **Information Storage and Retrieval** 9:79-84 Feb. 1973.
16. PETRARCA, A.E. & LAY, W.M. The double-KWIC coordinate index. **Journal of Chemical Documentation** 9(4): 256-61, Nov. 1969.
17. SALTON G. & MONG, A. Automatic indexing using term discrimination and term precision measurements. **Information Storage and Retrieval** 12(1): 43-51, 1976. Este é apenas um dos vários trabalhos produzidos por Salton fáceis de serem localizados nos índices especializados.
18. RUHL, Kary J. Chemical documents and their titles: human concept indexing vs KWIC machine indexing. **American Documentation** 15, 136-41, 1964
19. FEINBERG, H., op. cit., p. 297.
20. A **estruturação** dos conceitos continuará a ser imprescindível ao usuário no momento de elaborar sua estratégia de busca, pois não se deve esperar que ele faça, de forma rápida e eficaz, as devidas associações de idéias.

ABSTRACT

GOMES, H. E. *The indexer and the new technology of information*. **Trans-in-formação**, 1(2), maio/ago. 1989.

The traditional manual indexing is under interference of many variables, such as the proposal of the information retrieval system, the language adopted, the indexing policy, the nature of the subject field and the nature and property of the scientific and technology information. Such variables make the indexing a complex activity, the indexer being accused of introducing subjectivism and inconsistency in his job.

Researches in automatic indexing have showed that there are no significant difference in the performance of either one or another type of indexing. The automatic indexing has been improved since KWIC towards a more efficient retrieval and lower cost. Then the participation of a new kind of information professional increases, yet without denomination.