

Sistemas de Recuperação da Informação

Information Retrieval Systems

MARIA AUGUSTA DA NÓBREGA CESARINO *

Os sistemas de recuperação da informação podem ser definidos como um conjunto de operações consecutivas executadas para localizar, dentro da totalidade de informações disponíveis, aquelas realmente relevantes. Para isso, executam as funções de seleção, análise, indexação e busca das informações. Em todas essas etapas a interação usuário x sistema é fundamental, embora tenha se apresentado com muitas falhas. A formação do bibliotecário no Brasil não o capacita adequadamente para a adoção de políticas na área técnica e é discutível na sua visão social da recuperação da informação como sistema de comunicação.

A delimitação do tema «Recuperação da Informação» é tarefa das mais difíceis. A observação do conteúdo de livros textos e artigos assim intitulados mostra uma

* Professora da Escola de Biblioteconomia da UFMG

relação de tópicos tão abrangente e diversificada quanto a própria Biblioteconomia. Em geral, esses trabalhos estudam: teoria da informação, canais de comunicação, o usuário da informação, seleção e aquisição de documentos, análise de assunto, linguagens de indexação, armazenagem da informação, formação de base de dados, estratégias de busca, disseminação da informação, planejamento e avaliação de sistemas de informação.

Nesses grandes temas discute-se, preferentemente: a teoria da comunicação, onde o famoso esquema de Shannon e Weaver constitui o paradigma; as redes de comunicação em ciência e tecnologia, sobretudo as redes informais; as necessidades e o comportamento do usuário face aos diferentes canais de informação, suas preferências e frustrações, as possibilidades de educá-lo e/ou treiná-lo em relação aos sistemas de informação; o uso de técnicas bibliométricas para a compreensão da literatura e o uso de filtros de qualidade no processo de seleção; as contribuições da lingüística para os vocabulários controlados, o aperfeiçoamento dos sistemas de classificação e dos tesauros; o uso da tecnologia mais avançada, como a telecomunicação, computadores, audiovisuais e outros na armazenagem e transferência da informação; a utilização de base de dados bibliográficos on line e a montagem de estratégias de busca eficientes para tais sistemas; a disseminação seletiva da informação que atende menos a um problema específico da pessoa e mais a sua pressuposta necessidade de constante atualização; o modelo sistêmico norteando o planejamento de sistemas de informação; as análises dos custos em serviços de informação.

Os sistemas de recuperação da informação podem ser estudados sob dois aspectos:

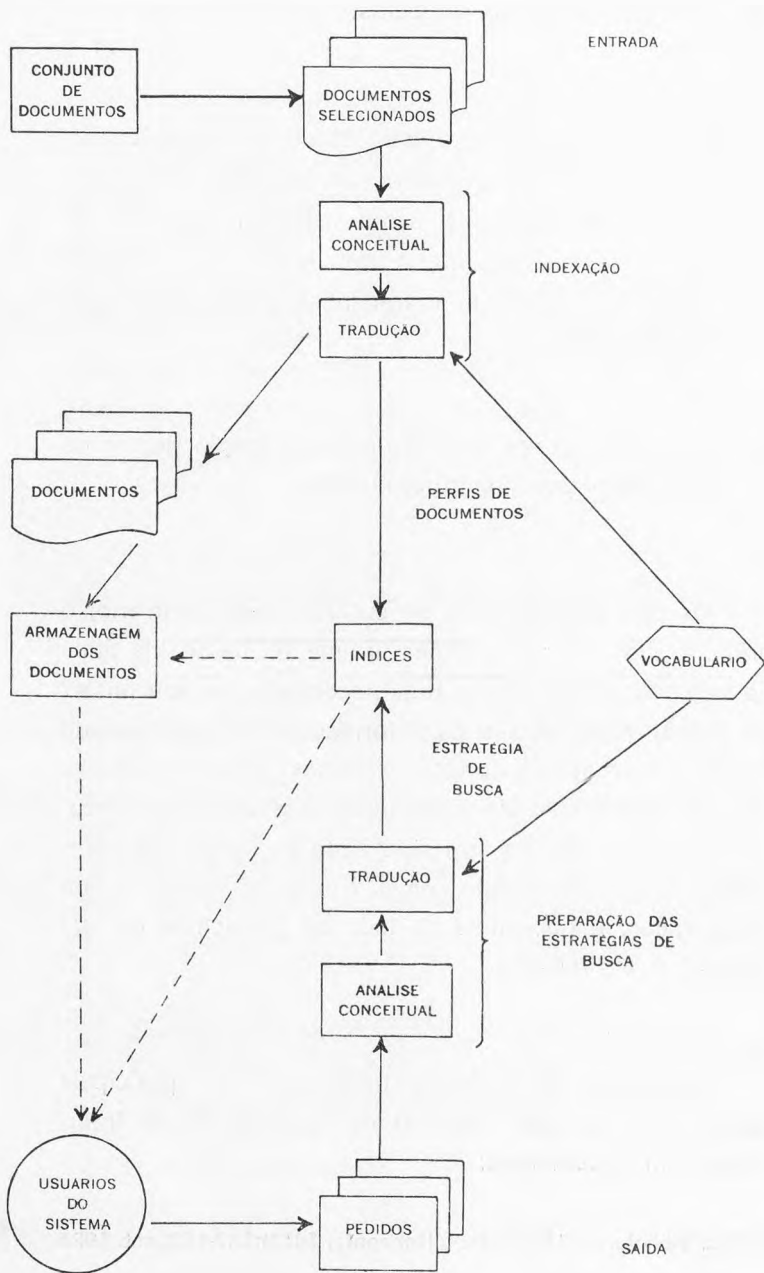
- a) como um conjunto de operações consecutivas, executadas para localizar, dentro da totalidade de informação disponível, aquelas que seriam necessárias ao usuário, ou seja, aumentar a transmissão de informação relevante e diminuir a transmissão de informação não relevante.

- b) como parte de um modelo de comunicação dentro de um contexto social-cultural-histórico.

Os dois aspectos não se excluem, pelo contrário, o primeiro deve ser visto, necessariamente, dentro da ótica do segundo. Entretanto, a Biblioteconomia, no seu mister de formar profissionais da informação, tem privilegiado bastante o detalhamento técnico dessas operações básicas para a recuperação da informação registrada em documentos. Pouca ênfase tem sido dada ao estudo da informação, enquanto produto social. Busca-se otimizar a sua recuperação, e deixam-se de lado os problemas da sua geração e assimilação.

✓ Lancaster (3) sintetiza muito bem as principais funções técnicas dos Sistemas de Recuperação de Informação em documentos:

FIGURA 1 — Principais Funções Desempenhadas nos Sistemas de Recuperação de Informação



Os subsistemas da recuperação da informação podem ser assim estudados:

- 1º) A seleção e aquisição de documentos
- 2º) A indexação, incluindo o processo de análise conceitual dos documentos e a tradução do resultado dessa análise para o vocabulário do sistema
- 3º) Organização e manutenção dos arquivos
- 4º) Estratégia de busca, que também envolve o processo de análise conceitual das questões propostas pelo usuário e a sua tradução para a linguagem do sistema
- 5º) A interação usuário x sistema de recuperação da informação.

O início de um Sistema de Recuperação de Informação se dá pela formação de um conjunto de documentos previamente selecionados e adquiridos dentro dos critérios estabelecidos pela Instituição que o mantém e de acordo com os objetivos a que ele se propõe atingir. Essa etapa é fundamental para o bom desempenho de qualquer sistema. Seria óbvio dizer que armazenar, indexar, recuperar e disseminar informações irrelevantes representam desperdício de tempo, de recursos humanos e de dinheiro. No estabelecimento de uma política de seleção das informações que alimentarão o sistema é fundamental que o usuário possa se manifestar e seja ouvido.

A eficiência de um Sistema de Recuperação de Informação depende muito da qualidade da análise conceitual tanto dos documentos quanto das questões. Grande parte das falhas na recuperação da informação se deve a erros ou omissões na interpretação do conteúdo

dos documentos e na percepção da demanda das pessoas a que se destina o sistema. O estágio de análise conceitual é anterior e independe do estágio de tradução para o vocabulário do sistema. Entretanto, a formação do bibliotecário tem dado ênfase ao estudo das linguagens de indexação. Como bem lembra C. R. Cavalcanti (1) «os trabalhos sobre a elaboração de linguagens de indexação são numerosos mas os documentos relativos ao processo, à forma de identificação do conteúdo temático do item são escassos. Pouco se encontra sobre a maneira de trabalhar dos classificadores e dos indexadores».

É importante para o bibliotecário compreender a linguagem das pessoas com que o sistema interage. Sem isso, é impossível «indexar corretamente» e «se comunicar com os usuários» dois pontos fundamentais para o bom funcionamento de um SRI - Sistema de Recuperação da Informação.

Entender essa linguagem significa compreender o significado de cada termo e perceber as relações entre eles: de equivalência, hierárquicas e associativas. Tais relações estão expressas continuamente nos documentos e nas questões dos usuários e devem encontrar a representação mais adequada no vocabulário adotado pelo sistema, seja através dos termos preferenciais (os símbolos de classificação, os descritores, as palavras-chave, etc.) seja através da rede de remissivas e referências.

A análise conceitual do documento e da questão do usuário vai resultar em termos que representam a chave para a recuperação da informação. Esse conjunto de termos preferenciais constitui a linguagem do sistema.

Há hoje um número muito grande de linguagens de indexação sendo utilizadas em Sistemas de Recuperação de Informação. São variações das linguagens naturais x controladas; linguagens alfabéticas x classifi-

cidas; linguagens pré x pós-coordenadas, etc. Não se pode dizer, simplesmente, quais as melhores ou as piores. Não há linguagem perfeita. Existe sim, a linguagem que melhor se adapta a um Sistema de Recuperação de Informação, com determinadas características.

Cabe ao bibliotecário tomar a decisão acertada: escolher, adaptar ou criar a linguagem que mais se ajusta ao sistema, que melhor atenda às necessidades do usuário.

Este é outro ponto crítico da formação do bibliotecário: a «tomada de decisões» ou «política» em relação aos diversos subsistemas da recuperação da informação, vistos de uma maneira integrada. Não se pode isolar cada componente e pensar numa «política de seleção», «política de indexação», «determinação dos catálogos», etc. Os subsistemas da recuperação da informação afetam uns aos outros. Por exemplo:

1. uma seleção inadequada corta a eficiência do sistema como um todo;
2. o aspecto qualitativo e quantitativo dos recursos humanos afeta a análise conceitual dos documentos e questões, bem como a utilização mais proveitosa do vocabulário adotado. Em que circunstâncias é mais conveniente colocar pessoal melhor qualificado na etapa de indexação ou no momento de busca?
3. os recursos financeiros para acesso, utilização e manutenção de equipamentos tecnologicamente mais avançados influenciam a escolha da linguagem de indexação, o arquivamento dos documentos e a busca da informação;

4. ao se escolher uma linguagem de indexação deve se ter em mente qual ou quais arquivos representativos ela vai gerar;
5. as linguagens de indexação pós-coordenadas levam o sistema a grande precisão mas serão eficientes e econômicas quando usadas em catálogos tradicionais de biblioteca (arquivos sequenciais manuais)?
6. os artifícios contra falsas coordenações encarecem a indexação e dificultam a busca;
7. os arquivos manipulados diretamente pelo usuário tornam o sistema mais econômico, mas exigem grande clareza, e, de preferência, simplicidade;
8. pode-se generalizar dizendo que todo Sistema de Recuperação de Informação deve buscar maior precisão mesmo que perca em revocação?
9. quanto maiores forem as diferenças entre a linguagem controlada do sistema e a linguagem do usuário, maiores as dificuldades na consulta e maiores as possibilidades de erro;
10. as linguagens controladas eliminam a inconsistência e a subjetividade da indexação?
11. em quais circunstâncias a utilização das linguagens que representam menor esforço na entrada do sistema (tais como o índice KWIC) são vantajosas, considerando que menor esforço na indexação representa, inevitavelmente, uma estratégia de busca mais elaborada?

Estas e muitas outras questões devem ser discutidas durante a formação do bibliotecário. Entretanto, a literatura sobre tomada de decisões em Recuperação de Informação é escassa e dispersa.

A política de indexação que é uma decisão administrativa só pode ser estabelecida depois de observados alguns aspectos, tais como:

- identificação das características do usuário (áreas de interesse, nível, experiência, atividades que exercem);
- volume e características da literatura a ser integrada ao sistema;
- volume e características das questões propostas pelo usuário;
- número e qualidade dos recursos humanos envolvidos;
- determinação dos recursos financeiros disponíveis para criação e manutenção do sistema;
- Determinação dos equipamentos disponíveis, etc.

A partir daí, o Sistema de Recuperação de Informação definirá suas características principais:

- se o sistema visará maior revocação ou precisão;
- qual o procedimento em relação à estratégia de busca (busca delegada ou não delegada);
- qual o tempo de resposta aceitável para o usuário;
- quais os serviços e produtos a serem fornecidos.

Com a análise dos dados de identificação e definidas as principais características, pode-se estabelecer a política do sistema:

- critérios de seleção;
- escolha da linguagem de indexação e sua forma de controle;
- determinação do nível de especificidade e de exaustividade da indexação;
- alocação e treinamento dos recursos humanos;
- determinação das características dos arquivos, etc.

Nem sempre se pode proceder de uma forma sequencial ou se pode obter todos os dados, antes de se implantar no Serviço de Recuperação de Informação. Entretanto, os aspectos mencionados devem ser colhidos e analisados para realimentar o sistema e possibilitar que as adequações necessária se façam gradualmente.

Um Sistema de Recuperação de Informação precisa justificar e comprovar a qualidade dos seus serviços. Os estudos tradicionais de desempenho dessas instituições bem como os esforços para o seu aperfeiçoamento, tem se orientado para o desenvolvimento das técnicas de processamento da informação: os meios sendo considerados mais importantes do que os fins. Entretanto eles são sistemas sociais que devem ser avaliados à luz do processo de comunicação de um lado, a informação; de outro o usuário; como canal, o Sistema de Recuperação de Informação.

Nesse sentido, a literatura afirma, com ênfase, que os bons resultados dos Sistemas de Recuperação de Informação dependem muito da maneira como o usuário interage com o sistema, isto é, a interpretação que o usuário faz da capacidade e das limitações do sistema, a habilidade em expressar as próprias necessidades de informação numa linguagem adequada, a divulgação que o Sistema de Recuperação de Informação faz da sua própria organização e serviços, e pela orientação e assistência que dá ao usuário.

Voltamos ao ponto inicial. A Biblioteconomia não tem analisado com mais rigor o aspecto social-cultural-histórico dos Sistemas de Recuperação de Informação. Se essa falta de visão é séria quando se observa que o único usuário que se tem em mente ao se planejar os Sistemas de Recuperação de Informação é aquele que

lida com informações científicas e tecnológicas, gravíssima se torna ao encarar o fato de que modernos Sistemas de Recuperação de Informação podem e devem atender ao usuário-cidadão comum.

Os Sistemas de Recuperação de Informação não tem por área definida apenas a ciência e a tecnologia, ápice da educação formal, mas também o lazer, a educação permanente, a cidadania, enfim, a cultura no seu sentido mais amplo.

Mas para os Sistemas de Recuperação de Informação de todas as áreas vale uma mesma constatação: o discurso da interação usuário x sistema é unilateral e reflete uma visão discutível dos Sistemas de Recuperação de Informação como sistemas sociais ou como modelo de comunicação.

Há uma lacuna e várias falhas que a nova formação do Bibliotecário no Brasil deve procurar sanar.

Information retrieval systems can be defined as a set of consecutive operations aiming to locating, inside the totality of available, information really relevant for an effective information retrieval. Fundamental for an effective information retrieval service are activities such as selection, information analysis, indexing and search for information. Furthermore, the absence of an interaction between users and systems will certainly hinder the success of the service. Brazilian library education does not qualify efficient professionals to elaborate policies in the technical area. Besides library professionals does not possess the perception of information retrieval service as a communication system.

REFERÊNCIAS BIBLIOGRÁFICAS

1. CAVALCANTE, C. R. Indexação In: MACHADO, V. D. ed. **Estudos avançados em Biblioteconomia e Ciência da Informação**. Brasília, ABDF, 1982 v. 1 p. 211-31.
2. GONZALEZ DE GÓMES, M. N. Informação e conhecimento, **Ciência da Informação**, Brasília, 13(2): 107-14 Jul./Dez. 1984.
3. LANCASTER, F. W. **Information Retrieval Systems**; characteristics, tests and evaluation. 2. ed. New York, John Wiley, 1979. 381 p.
4. ————. **Vocabulary control for information retrieval**. Washington, Information Resources Press, 1972. 233 p.
5. OBERHOFER, C. A. Conceitos e princípios para avaliação de sistemas de informação. **Ciência da Informação**, Brasília 12(1): 45-52, 1983.