

Avaliação de Serviços de Indexação e Resumo: Critérios, medidas e Metodologia

**Evaluation of Indexing and Abstracting Services;
criteria, measures and methodology**

EUNICE DE FARIA LOPES *

**Descrição de alguns critérios e medidas para
avaliar a eficácia, o custo/eficácia e o custo/
benefício de serviços de indexação e resumo**

1. INTRODUÇÃO

Os Serviços de Indexação e Resumo (SIR) têm crescido em importância para as bibliotecas que, pressionadas pelo volume da literatura e pela limitação dos recursos financeiros, passaram a preocupar-se com o controle da literatura publicada, antes que com a aquisição das fontes primárias de informação.

Por SIR compreendem-se os serviços de armazenamento, organização e divulgação de dados sobre documentos originais, geralmente apresentados sob a forma de referências bibliográficas, acompanhadas ou não de resumos, disponíveis para recuperação na forma de um banco de dados automatizado ou de publicação

* Bibliotecária da Escola de Veterinária da UFMG.

impresa. Podem registrar informações de uma área mais abrangente do conhecimento ou de uma específica, geradas em um país ou em vários.

Os SIR devem corresponder às necessidades de informação da biblioteca, uma vez que vão servir de intermediários entre a demanda e o documento original, e esta adequação pode ser constatada através de um processo de avaliação.

A literatura no campo da avaliação do SIR é vasta e este artigo está baseado nos trabalhos de CARVALHO (7), que escreve especificamente sobre a metodologia para a avaliação de serviços de **abstracts** e LANCASTER (3, 4, 5), que reúne as principais idéias sobre a conceituação e metodologia para avaliação de serviços de recuperação de informação.

2. PLANEJAMENTO

O processo de avaliação está ligado à observação do desempenho de uma determinada atividade e à obtenção de dados que possam subsidiar decisões administrativas.

Na avaliação de um SIR procura-se determinar a qualidade do produto que é oferecido, o seu nível de desempenho em relação às necessidades do sistema de informação e os custos decorrentes. Os resultados oferecem dados para orientar a seleção de um SIR, a aquisição do produto e o acompanhamento do grau de satisfação às demandas.

Como toda atividade gerencial, a avaliação deve ser precedida de cuidadoso planejamento, a fim de que os resultados possam corresponder aos objetivos que a motivaram.

Inicialmente, deve-se esclarecer o objetivo de um SIR, a saber, o de fornecer aos usuários a indicação de documentos que correspondam às suas necessidades de informação (Faithorne citado por PIEDADE) (6) bem como os objetivos da avaliação: seleção, custo ou verificação do desempenho. Estabelecidos estes conceitos, é possível conduzir um programa de avaliação, para o qual LANCASTER (5) indica as seguintes etapas:

1. definir o campo de avaliação: determinar o que vai ser avaliado, com base nas questões cujas respostas são procuradas;
2. planejar o programa de avaliação: escolher os critérios e medidas que permitam reunir os dados necessários e estabelecidos na definição do campo de avaliação;
3. executar a avaliação: coletar os dados conforme o programa estabelecido;
4. analisar e interpretar os dados: reduzir e manipular os dados obtidos, a fim de que possam responder às questões propostas no programa de avaliação. Esta etapa inclui o relatório dos resultados e as recomendações decorrentes da avaliação;
5. utilizar os resultados e/ou aplicar as recomendações.

Essas etapas são comuns a qualquer programa de avaliação, sendo característico do SIR apenas o item 2, isto é, a escolha dos critérios e das medidas de avaliação.

3. CONCEITO DE RELEVANCIA

A avaliação de um SIR envolve um julgamento do valor da informação recuperada — a relevância, que é a capacidade do sistema em fornecer respostas (referências) que realmente correspondam à questão proposta.

Embora pareça simples à primeira vista, é um julgamento objeto de controvérsias e BOURNE (1) chama a atenção para os principais pontos: «relevância — pode ser medida ou descrita? é uma medida adequada? quem a julga? de que ponto de vista? quantos graus de relevância existem?».

Para LANCASTER (3, 4), relevância é um julgamento do usuário, uma vez que a questão foi por ele proposta e somente ele pode julgar se as suas necessidades de informação foram satisfeitas ou não. Entretanto, observa que cada usuário pode ter diferentes níveis de expectativas e sugere que, em programas de avaliação, o usuário coloque a razão do julgamento para cada grau diferente de relevância, lembrando que:

- o julgamento do usuário vai depender do motivo que originou a demanda de informação;
- o usuário pode mudar de opinião no decorrer da busca;
- entre si, os usuários têm opiniões diferentes.

Já REGAZZI (7) preocupa-se com as medidas de desempenho para sistemas de indexação e o papel dos conceitos de relevância e utilidade no julgamento de seus produtos. Observa que o conceito de relevância tem dois problemas principais: «não há consenso sobre em que bases a relevância é medida e não há um método confiável

para quantificá-la. Por outro lado, o conceito de utilidade contém a noção de relevância e mais, tempo, novidade, criatividade, etc. Conclui que ainda não há segurança sobre um modelo que dê o valor de um item para o usuário e o desempenho eficaz de um sistema em recuperar informações adequadas».

Portanto, na avaliação de um SIR, todas estas variáveis devem ser consideradas no momento de se programar a avaliação e ponderadas no momento de se verificarem os resultados.

4. CRITÉRIOS E MEDIDAS

Um sistema de informação e, conseqüentemente, um SIR, pode ser avaliado em relação à sua eficácia, custo/eficácia e custo/benefício (CARVALHO, 2; LANCASTER, 4, 5).

4.1 Eficácia

Eficácia é a habilidade do SIR em satisfazer as necessidades dos usuários e inclui os critérios e medidas referentes à qualidade do produto e o esforço/tempo despendidos pelo usuário na obtenção da informação desejada.

4.1.1 Autoridade

Autoridade é um critério relacionado com quem é responsável pela elaboração da base de dados do SIR. Pode ser verificada através da reputação do editor/editora/corpo editorial na área de assunto do SIR, da participação de especialistas da área e da qualidade de outros serviços já oferecidos.

4.1.2 Cobertura

Cobertura é um dos mais importantes critérios para a avaliação de um SIR, porque mostra a sua abrangência em relação à literatura do assunto. Indexar todos os materiais publicados é impossível e a cobertura é determinada pela proporção de toda a literatura de um assunto que é incluída no SIR.

Os pontos que devem ser verificados na avaliação da cobertura são:

- a extensão em que é tratado o assunto específico e a inclusão de áreas correlatas;
- a proporção de publicações consideradas relevantes que são incluídas;
- a inclusão de material convencional e não convencional;
- quais os idiomas (dos documentos) incluídos;
- quais as áreas geográficas cobertas.

Um método simples de se verificar a cobertura é avaliar a **core list** do SIR em relação às de outros serviços na mesma área, ou a opinião de especialistas no assunto.

No trabalho de CARVALHO (2) são sugeridas três metodologias, utilizando amostra da população, critérios de revocação e precisão e análise conforme os itens colocados acima.

LANCASTER (3) sugere o uso de uma bibliografia especializada, como um artigo de revisão, e a partir dela, checar cada «única referência» com o SIR, contando quantas são encontradas ou não. A cobertura é medida pela proporção dos itens encontrados, em relação ao total da bibliografia:

$$\text{Tx Co} = \frac{\text{Nº de referências da bibliografia localizadas no SIR}}{\text{Nº total de referências da bibliografia}} \cdot 100$$

4.1.3 Revocação

Revocação é a capacidade do SIR de oferecer, em resposta a uma questão, todas as referências relevantes existentes na base de dados. É medida pela proporção de referências relevantes recuperadas, em relação ao total de referências relevantes existentes no SIR.

$$\text{Tx Re} = \frac{\text{Nº de referências relevantes recuperadas}}{\text{Nº total de referências relevantes existentes no SIR}} \cdot 100$$

Para medir a revocação de maneira ideal, é necessário conhecer previamente quais os dados relevantes existentes na base de dados, o que dificilmente pode ser estabelecido. Assim, na prática, são usados grupos experimentais de documentos, amostras da população ou outro tipo de medida, fazendo inferências a partir dos resultados do teste. É aconselhável ter presente a relação existente entre a cobertura e a revocação: um sistema pode ter uma alta revocação e não estar fornecendo ao usuário toda a informação publicada, já que isto vai depender da cobertura dada ao assunto pelo SIR.

LANCASTER (4) registra as seguintes metodologias:

- 1) Pedir a um grupo de usuários questões baseadas em documentos existentes na base de dados; proceder à busca como se fossem questões reais; levar o resultado ao grupo de usuários para julgamento de

relevância. A taxa de revocação é dada pela proporção de documentos relevantes recuperados no levantamento, em relação ao número total de documentos usados para preparar as questões.

- 2) Comparar os resultados de uma busca real (A) com os de outros levantamentos sobre o mesmo tema (B, C, ...), feitos por outras pessoas mas usando a mesma definição de assunto da busca real; levar os resultados (A, B, C, ...) ao usuário para julgamento de relevância. A taxa de revocação da busca real é dada por:

$$\text{Tx ReA} = \frac{\text{Nº de itens relevantes encontrados na busca A}}{\text{Nº total de referências (únicas) relevantes encontradas nas buscas A, B, C,}} \cdot 100$$

- 3) Estabelecer, por meios alheios ao SIR, um número de documentos relevantes e conhecidos para cada questão-teste. Proceder à busca (questão-teste) e levar os resultados ao usuário para julgamento de relevância. A taxa de revocação é estimada pelo número de itens encontrados na busca-teste, em relação ao número de referências conhecidas pelos meios externos ao sistema:

$$\text{Tx Re} = \frac{\text{Nº de referências relevantes encontradas na busca teste}}{\text{Nº de referências conhecidas previamente}} \cdot 100$$

4.1.4 Precisão

Precisão é a capacidade do SIR em fornecer apenas referências relevantes, eliminando as que não são rele-

vantes para a questão. É uma medida importante para o usuário porque quanto maior a precisão, menor será o esforço do usuário em utilizar o SIR, isto é, os «ruídos» são eliminados.

A precisão é mais fácil de ser medida em buscas delegadas (feitas pelo **staff**), uma vez que, nos levantamentos realizados pelo próprio usuário, a seleção feita durante a busca prejudica a contagem final das referências consultadas.

A precisão é determinada sempre que o resultado de uma busca é submetido a um julgamento de relevância. É medida pela proporção de referências relevantes recuperadas, em relação ao número total de referências recuperadas:

$$\text{Tx Pr} = \frac{\text{N}^\circ \text{ de referências relevantes recuperadas.}}{\text{N}^\circ \text{ total de referências recuperadas}} \cdot 100$$

LANCASTER (4) sugere como metodologia:

- Fazer uma busca bibliográfica de acordo com o teste nº 3 para estimar a revocação;
- levar o resultado da busca ao usuário para determinar a relevância dos itens, usando como critério de julgamento o mesmo estabelecido para a formação da bibliografia/teste usada no estudo de Re. Assim,

$$\text{Tx Pr} = \frac{\text{N}^\circ \text{ de referências relevantes recuperadas}}{\text{N}^\circ \text{ de referências encontradas na busca/teste}} \cdot 100$$

4.1.5 **Novidade**

Novidade é a medida da proporção de referências relevantes recuperadas através do SIR e que o usuário não conhecia anteriormente. É um critério de avaliação importante para os serviços de divulgação corrente (SDI) e está relacionada com o tempo decorrido entre a publicação do original e a sua indexação/divulgação pelo SIR.

CARVALHO (2) sugere tomar uma amostra de referência e verificar o tempo decorrido entre a data de publicação do original e a data do SIR, estabelecendo uma média para o tempo decorrido.

LANCASTER (4) sugere, a partir dos dados obtidos na precisão (referências relevantes recuperadas), estabelecer a proporção de itens não conhecidos anteriormente pelo usuário:

$$\text{Tx No} = \frac{\text{Nº de referências relevantes e novas recuperadas}}{\text{Nº de referências relevantes e recuperadas}} \cdot 100$$

4.1.6 **Esforço do usuário**

É a medida do trabalho que o usuário tem para consultar o SIR, estreitamente relacionado com a precisão: quanto maior for a taxa de precisão, menor será o esforço do usuário em usar o sistema.

O esforço do usuário é calculado em função do tempo que ele depende para executar uma ou várias destas alternativas:

- preparação da questão (formulação, delimitação, teste, etc.);
- diálogo com o **staff** no caso das buscas delegadas;
- compreensão do funcionamento do SIR, nas buscas feitas pelo próprio usuário;
- tempo gasto no levantamento, quando feito pelo próprio usuário;
- exame da lista de referência para extrair os itens relevantes.

LANCASTER (3) indica, para calcular o esforço do usuário em função do tempo gasto na busca:

$$\text{Custo (em tempo)} \\ \text{por referência relevante} = \frac{\text{Tempo total gasto na busca recuperada}}{\text{Nº total de itens relevantes encontrados na busca}}$$

4.1.7 Tempo de resposta

Em levantamentos feitos pelo **staff**, tempo de resposta é o tempo decorrido entre a pergunta e a resposta. Já nas buscas realizadas pelo usuário, é o tempo que ele gasta para realizá-las, coincidindo com a medida de esforço do usuário (LANCASTER, 3).

Em ambos os casos e como no item anterior, a eficácia do SIR é medida em função do custo de tempo, isto é, o quanto custa para o usuário esperar pela informação.

4.1.8 Produtos oferecidos

A base de dados de um SIR pode ser consultada através do computador (**batch** ou **on line**) ou de uma publicação impressa (índice, **abstract**, etc.). Pode oferecer produtos adicionais como: divulgação de literatura corrente (SDI), buscas retrospectivas (compilação de bibliografias curtas), fornecimento de artigos originais, etc.

Os produtos de um SIR e a forma das saídas não são critérios de avaliação quantificáveis, mas afetam a satisfação do usuário (LANCASTER, 3).

Podem ser observadas para conhecimento do SIR e verificação da sua qualidade:

- na consulta através do computador, se o resultado é apresentado em vídeo, listagens, fichas ou microforma;
- nas publicações impressas, a qualidade tipográfica, a existência de índices acumulados;
- os dados apresentados nas referências bibliográficas;
- a existência de resumos.

4.1.9 Linguagem de indexação

A linguagem usada para identificar os assuntos e a qualidade da indexação refletem diretamente na satisfação do usuário, uma vez que o índice é a ponte entre a pergunta e o conteúdo do banco de dados.

Muitos são os trabalhos publicados e os testes realizados com as linguagens de indexação. Aqui são mencionados apenas os pontos que devem ser observados na verificação da qualidade de um índice (LANCASTER, 5):

- a linguagem usada: livre, palavras-chaves, **thesaurus** cabeçalho de assunto, etc.;
- o nível de exaustividade, isto é, a extensão em que todos os assuntos tratados no documento são identificados e traduzidos na linguagem de indexação. É um critério importante porque interfere diretamente na capacidade de revocação e precisão do SIR: quanto maior a exaustividade, mais alta será a revocação e menor a precisão;
- o nível de especificidade do vocabulário usado;
- a qualidade e correção da indexação, incluindo a consistência.

4.2 Custo/Eficácia

Custo/eficácia é o quanto custa para a biblioteca ter um SIR eficaz, isto é, manter um serviço que satisfaça aos usuários. Pode ser avaliado medindo-se o custo de cada referência ou o conjunto de referências relevantes recuperadas (LANCASTER, 5). Um serviço de informação será eficiente ao satisfazer o usuário pelo menor custo.

4.3 Custo/Benefício

Custo/benefício é relação entre os custos de um SIR e o benefício de possuí-lo. É uma medida difícil de ser obtida porque o benefício não é mensurável e é julgado por critérios subjetivos.

CARVALHO (2) recomenda medir o custo médio de cada referência relevante:

$$\text{Custo médio por referência relevante} = \frac{\text{Custo do SIR}}{\text{Nº total de referências relevantes recuperadas}}$$

5. CONCLUSÃO

Os critérios e medidas relacionados são aqueles destinados a medir os sistemas de informação em geral e que aqui são tratados do ponto de vista da consulta a um banco de dados para se obter uma informação, eficientemente.

Um SIR é avaliado quando se querem dados para ajudar no processo de seleção/compra ou verificar o desempenho do mesmo. Sendo uma atividade administrativa, deve ser dada atenção ao planejamento, em todas as suas etapas.

Crítérios e medidas devem ser escolhidas cuidadosamente e de acordo com os objetivos da avaliação. Como o universo estudado é sempre muito grande, o uso de amostras e grupos de teste é bastante difundido, devendo ser observados critérios científicos na sua determinação, a fim de permitir inferências seguras a partir dos resultados.

Os resultados da avaliação podem ser usados diretamente, no caso da ajuda na seleção/compra. Entretanto, quando o objetivo é verificar o desempenho do SIR a sua aplicação restringe quase que ao treinamento dos usuários do serviço, uma vez que dificilmente uma biblioteca pode interferir no processo de coleta/armazenamento/divulgação de um SIR.

Description of some criteria and measures for evaluating effectiveness, cost/effectiveness and cost/benefit of indexing and abstracting services.

BIBLIOGRAFIA

1. BOURNE, C. P. Evaluation of indexing systems. In: CUADRA, C. A., ed. **Annual review of information science and technology**. New York, J. Wiley, 1966. p. 171-90.
2. CARVALHO, M. M. **Methodology for the evaluation of abstract services**. Rio de Janeiro, 1974. 8 p. (Trabalho apresentado ao Curso de Pós-Graduação em Ciência da Informação — IBICT).
3. LANCASTER, F. W. Evaluation of literature searching and information retrieval. In: ————. **The measurement and evaluation of library services**. Washington, Information Resources Press, 1978. p. 140-64.
4. LANCASTER, F. W. Evaluating the operating efficiency of an information retrieval system: designing the test program. In: ————. **Information retrieval systems; characteristics, testing and evaluation**. New York, J. Wiley, 1968. p. 118-29.
5. LANCASTER, F. W. **Information retrieval systems: characteristics, testing and evaluation**. 2. ed. New York, J. Wiley, 1979. cap. 6, 8, 9.
6. PIEDADE, M. A. R. **Estudo comparativo de algumas linguagens de indexação: eficácia e tempo de pesquisa**. Rio de Janeiro, 1976. 101 p. (Dissertação de Mestrado apresentada ao IBICT/UFRJ).
7. REGAZZI, J. J. Evaluating indexing systems: a review after Cranfield. **Indexer**, London, 12 (1): 14-21, Apr. 1980.