

O CONCEITO DE RELEVÂNCIA E SUAS IMPLICAÇÕES

Laura Maia de Figueiredo
IBICT/CTR

Em 1934, S.C. Bradford publicou "Sources of information on specific subjects" que atrairia, mais tarde, a atenção da comunidade de profissionais da informação. Em torno da Lei de Bradford, uma série de estudos começou a ser desenvolvida, incorporando estudos anteriores como o de Lotka(4). Estes estudos, agora, reunidos sob o nome "Bibliometria", procuram medir a produtividade das fontes de informação relacionando um conjunto de "produtores" (periódicos, autores, termos) a um conjunto de "produtos" (documentos, autoria, freqüência).

Naquele documento, Bradford afirmou que "há periódicos de âmbito, obviamente e *a priori*, relevantes a assuntos investigados". Assim, além do aspecto bibliométrico, há também um outro ponto importante abordado no documento: pela primeira vez o conceito de "relevância" apareceu dentro do contexto de medida de uma fonte de informação e com uma conotação probabilística - Bradford também afirmou que existem fontes que "esperamos sejam mais produtivas"... e que "um grande número de referências são produzidas por fontes que, *a priori*, são insatisfatórias".

Saracevic, mais tarde, atentou para o fato de que distribuições bibliométricas são, de fato, "distribuições associadas à relevância".

Como o conceito de "significado", "relevância" é também uma noção intuitiva, implicando "contato

RESUMO

Como o conceito de significado, relevância é uma noção intuitiva e uma medida determinada não somente pela relação entre documento e pergunta mas também em termos de relações entre os documentos — é uma relação comparativa. A Ciência da Informação tem tentado entender e explicar o fenômeno através de diversas abordagens (matemática, estatística, filosófica! — não podendo ser estudada como um fenômeno isolado mm dependente de muitos fatores do processo de comunicação. (HB)

Descritores:

Relevância; Significado; Medidas de Informação; Processo de Comunicação.

produtivo" em um processo de comunicação. Os semanticistas abordam as medidas de informação esclarecendo que a informação com conteúdo significativo é aquela na qual existe um contato produtivo entre um "thesaurus" (estado do conhecimento de cada indivíduo) e uma proposição *T*.

Assim, neste contexto, informação com conteúdo significativo é equivalente à informação relevante. Desconhecemos, ainda, o grau do envolvimento, o ambiente, influências e propriedades da relevance, mas tornando-a equivalente a significado preocupamo-nos com apenas um fenômeno. Do momento em que entendermos o que é relevância poderemos lidar melhor com esta noção e com seus efeitos sobre o indivíduo — e isto é realmente, o que os semanticistas tentaram fazer. Quiseram medir o significado de informação de acordo com a transformação sofrida pelo receptor do conhecimento — e este "nível influencial" da informação requer maior entendimento do processo de comunicação.

Relevância é considerada como uma propriedade em um processo de comunicação — seqüência de eventos onde algo chamado informação é transmitido de um objeto (fonte) a outro (destinatário). O processo é freqüentemente reiterativo ou de retroalimentação.

Relevância, a mais importante propriedade da informação, é a medida de contato efetivo entre

O conceito de relevância e suas implicações
Laura Maia de Figueiredo

a fonte e o destinatário, Muitas outras definições existem, mas a de Saracevic (6,7) é a que melhor expressa este conceito do fenômeno.

Mooers, Perry e Taube, foram os primeiros a falar sobre relevância *per se* 1950-55, como pioneiros da Recuperação da Informação; mas a primeira exploração extensa do conceito relevância apareceu em 1958, durante a "International Conference for Scientific Information" (ICSI) com Fairthorne, Vickery, Bar-Hillel e outros.

Alguns trabalhos interessantes apresentados nesta conferência foram apontados por Saracevic (6,7). Somente Fano sugeriu o aspecto do julgamento de relevância (mas de maneira indireta); seu ponto de vista era de que *ele* poderia dizer que documentos seriam satisfatórios e que documentos seriam insatisfatórios ao seu pedido. A comunidade de informação presente à Conferência concordou em determinados pontos:

- relevância não é, exclusivamente, uma propriedade de documentos;
- relevância não é uma propriedade dicotômica;
- existe uma "relevância para o usuário" que deve ser julgada.

Mais tarde, por volta de 1960, com Maron e Kuhns, apareceu a primeira teoria de relevância — a derivação de uma medida indicando a probabilidade de um documento satisfazer a um determinado pedido: dada uma probabilidade aproximada, um conjunto de documentos satisfazendo um pedido pode ser ordenado de acordo com sua relevância. Maron e Kuhns, desta maneira, explicaram relevância como uma noção quantitativa; da mesma forma que Shannon, abordaram a operação "medida" do ponto de vista probabilístico.

Hillman tentou isolar o fenômeno relevância de seu ambiente, independente de, e logicamente anterior a, alguma noção de relevância como determinada por, e assim restrita a um sistema particular de armazenagem e recuperação.

O grande mérito de Hillman foi o de examinar o conceito do ponto de vista lógico, pois aplicou construções lógicas para definir relevância e suas relações.

W. Goffman(3) explicou o fenômeno através de um ponto de vista matemático formal.

Relevância foi definida como uma medida de informação apresentada por um documento em relação a uma pergunta.

As propriedades matemáticas exigem que uma medida:

- a) tenha valor real e não-negativo (i.e. expressa por um número real positivo);
- b) seja totalmente aditiva (i.e. a soma das medidas é a medida da soma: partes medidas, tanto

independentemente ou como um todo, tem que ter a mesma medida);

- c) tenha ordem assegurada (conjunto menor tem sobre si uma medida menor);

- d) tenha um zero absoluto.

Goffman desenvolveu um teorema provando que a relevância determinada simplesmente na base de relações entre a pergunta e o documento não satisfaz o conceito de medida. Goffman também provou que relevância é determinada não somente pela relação pergunta-documento mas, também, em termos de relações entre documentos. Em ambos os casos relevância tem as propriedades requeridas por uma medida. Como uma medida, relevância, requer relações entre a pergunta e documentos, e entre os próprios documentos — relevância é uma relação comparativa.

Goffman(2), mais tarde desenvolveu a teoria epidêmica da comunicação que é aplicada ao estudo dos processos de comunicação nos quais a informação é transmitida dentro de uma população.

Iniciando com a teoria da comunicação de Shannon, define um processo de comunicação como uma disseminação de idéias fazendo, assim, uma analogia com a disseminação de uma doença.

A população tem três componentes - infectados, suscetíveis e removidos — e pode estar em estado crescente, decrescente ou estável. O fluxo de informação na população pode infectar, ou não, elementos suscetíveis, dentro de determinadas condições. Este contato é efetivo quando a informação é *relevante* ao elemento suscetível — i.e., quando resulta em acumulação de conhecimento.

A teoria epidêmica é inteiramente explicada e provada em linguagem matemática formal e sofisticada, mas até os não matemáticos podem ver alguns aspectos importantes, principalmente se colocarmos a teoria epidêmica no contexto da Teoria Geral da Comunicação. Do ponto de vista da relevância, foi demonstrado que:

- a) relevância deve ser considerada sob o aspecto *tempo*. Pode-se ver, facilmente, que o que é relevante hoje pode não ser amanhã;
- b) cada processo de comunicação tem seu limite de relevância de acordo com as condições da população, o material infeccioso e o componente tempo — realmente não existe a "relevância absoluta";
- c) relevância é também documento-dependente, no sentido de que documentos não são elementos isolados, mas fortemente inter-relacionados — a relevância de um documento é dependente da relevância de outros.

Como Saracevic ressaltou, a incorporação da Relevância em uma teoria da Comunicação é extremamente significativa, pois pela primeira vez.

numa definição matemática totalmente formal, podemos localizar a relevância num processo de comunicação.

Os testes de sistemas de recuperação da informação, no auge, no início da década de sessenta — tornaram-se uma disputa, em várias ocasiões, sobre o "sistema mais relevante". O desempenho dos sistemas era medido em termos de revocação, precisão (relevância), sensibilidade e vários outros parâmetros capazes de definir o sistema "perfeito". Logo tornou-se claro que não seria fácil medir-se a relevância: quem seria, em primeiro lugar, o melhor "juiz em relevância"? O usuário? Um substituto? Ambos? Um grupo?

Se a fase de teste é vista por alguns como uma total perda de tempo, pelo menos teve o mérito de atrair a atenção para vários problemas de comunicação, principalmente relevância. As conclusões da Conferência organizada pela National Science Foundation, em 1964, mostrou quão importante é o fenômeno — a falta de conhecimento sobre ele foi considerado o maior obstáculo ao progresso na área dos sistemas de recuperação de informação.

Na metade da década de sessenta um conjunto de experiências foi realizado a fim de lidar, de forma mais efetiva, com o problema relevância. Os resultados destas experiências não são generalizáveis mas esclareceram, de uma certa maneira, vários aspectos do comportamento de variáveis em relação à relevância; -demonstraram que relevância é também fonte-dependente; e que condições de julgamento desempenham um papel importante na sua determinação.

De todos os trabalhos realizados sobre relevância, há pelo menos um aspecto que chama a atenção sobre o conceito em si: existem vários pontos de vista sobre o fenômeno relevância, como nos mostra Saracevic (7)

- ponto de vista do conhecimento do assunto;
- ponto de vista lógico;
- ponto de vista do sistema;
- ponto de vista do destinatário;
- ponto de vista do conhecimento do destinatário - pertinência;
- ponto de vista pragmático.

Estes não são mutuamente exclusivos mas inter-relacionados e até superpostos.

Saracevic demonstra que existem no entanto algumas propriedades comuns que pertencem a todos estes sistemas:

- a) a existência de um certo corpo de conhecimento, de alguma forma, representado;
- b) o processo de seleção que acompanha o conhecimento mencionado, em seus elementos e estrutura;
- c) um processo de inferência que dirige a seleção;
- d) uma função de mapeamento do componente do conhecimento;
- e) interações dinâmicas entre propriedades existentes no sistema;

- f) associação entre as propriedades dinâmicas e estáticas (estrutura interna de elementos do conhecimento);
- g) redundância de critérios do ponto de vista da satisfação:
há mais de uma forma de inferência, estrutura, mapeamento etc. que poderia ser considerada adequada.

A Ciência da Informação tem tentado entender e explicar o fenômeno relevância através de diversas abordagens, mas ainda existem várias falhas, como as que se relacionam com valores humanos envolvidos na comunicação do conhecimento. Já foi provado que relevância não pode ser estudada como um fenômeno isolado; visto que depende de muitos fatores do processo de comunicação, parece razoável admitir que o melhor entendimento de componentes do processo poderá esclarecer a noção de relevância. Uma das áreas a ser explorada é certamente a relacionada com a estrutura do conhecimento e sua representação.

Por vários anos a função de esquemas de classificação foi vista como um mero instrumento de arquivar itens. O trabalho do Classification Research Group (Londres) e alguns especialistas na área têm mostrado que existe muito mais a ser envolvido. A noção de classificação multi-dimensional hierárquica, como a apresentada por P. Richmond(5) e outros pode conduzir a perspectivas interessantes.

CITAÇÕES BIBLIOGRÁFICAS

- 1 - BRADFORD, S.C. Sources of information on specific subjects. *Engineering*: 85-6,1934.
- 2 — GOFFMAN, W. A General theory of communication. In: Saracevic, T. ed. *Introduction to information science*. New York, Bowker, 1970. p. 727-47.
- 3 - GOFFMAN, W. An indirect method of information retrieval. In: Saracevic, T. ed. *Introduction to information science*. New York, Bowker, 1970. p. 485-92.
- 4 - LOTKA, A.J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*: 317-23,1926.
- 5 - RICHMOND, P.A. Hierarchical definition. *American Documentation*, 11 (2): 91-6,1960.
- 6 - SARACEVIC, T. The concept of "Relevance" in information science: a historical review. In:———. *Introduction to information science*. New York, Bowker, 1970. p. 111-51.
- 7 - SARACEVIC, T. Relevance: a review of and a framework for the thinking on the notion in information science. *JASIS*, 26(6): 321 -43, Nov./Dec. 1975.

O conceito de relevância e suas implicações
Laura Maia de Figueiredo

ABSTRACT

As the concept of meaning, relevance is an intuitive notion and a measure determined not only in the basis of the relationships between query and documents, but among documents themselves—relevance is a comparative relation. Information Science has been trying to understand and explain the relevance phenomenon from several approaches (mathematical, philosophical, statistical). It cannot be studied as an isolate phenomenon since it depends on so many factors of the communication process. (HB)