

O conceito de relevância na Recuperação de Informação

The concept of relevance in Information Retrieval

Jorge Janaite Neto

Doutor em Ciência da Informação pela Universidade Estadual Paulista, UNESP, Brasil; Analista de Informática na Universidade Estadual Paulista, UNESP, campus de Marília, SP, Brasil.
ORCID: <https://orcid.org/0000-0002-2269-5371>
E-mail: jorge.janaite@unesp.br

Edberto Ferneda

Doutor em Ciências da Comunicação pela Universidade de São Paulo, USP, Brasil; Professor Associado da Universidade Estadual Paulista, UNESP, campus de Marília, SP, Brasil.
ORCID: <https://orcid.org/0000-0002-8808-1217>
E-mail: edberto.ferneda@unesp.br

Resumo

Este trabalho visa apresentar um levantamento bibliográfico sobre o conceito de relevância, suas propriedades e características, assim como sistematizar suas manifestações e seu uso na Ciência da Informação, particularmente no processo de recuperação de informação. Para isso, a pesquisa baseou-se no levantamento bibliográfico de diversos autores que têm a relevância como objeto de estudo. Verificou-se que o conceito de relevância apresenta várias definições formais, podendo ser interpretado de diversas maneiras. A relevância também está fortemente ligada aos sistemas de recuperação de informações. Nesse contexto, pode ser classificada sob dois enfoques: a relevância do usuário e a relevância do sistema. A partir dos resultados, concluiu-se que a relevância desempenha um papel fundamental na Ciência da Informação, com destaque para o processo de recuperação de informação, que muitas vezes utiliza esse conceito como métrica para avaliar a eficácia de algorítmicos e sistemas.

Palavras-chave: recuperação de informação; Ciência da informação; relevância; modelos e teorias de relevância.

Abstract

This work aims to present a bibliographical survey on the concept of relevance, its properties and characteristics, as well as systematizing its manifestations and its use in Information Science, particularly in the process of information retrieval. To achieve this, the research was based on a bibliographical survey of several authors who have relevance as an object of study. It found that the concept of relevance has several formal definitions and can be interpreted in different ways. Relevance is also strongly linked to information retrieval systems. In this context, it can be classified under two approaches: user relevance and system relevance. From the results, we conclude that relevance plays a fundamental role in Information Science, with emphasis on the information retrieval process, which often uses this concept as a metric to evaluate the effectiveness of algorithms and systems.

Keywords: information retrieval; Information science; relevance; relevance models and theories.

1. Introdução

O conceito de relevância é de grande importância para a Ciência da Informação, mas é particularmente crucial nas pesquisas em Recuperação de Informação, sendo muitas vezes utilizado na própria enunciação dos objetivos dessa área. É uma questão fundamental e uma preocupação central para o funcionamento e avaliação dos sistemas de recuperação de informação (Saracevic, 2017; Mizzaro, 1997; Cooper, 1971).

O objetivo de todo sistema de recuperação de informações é fornecer informações relevantes aos usuários, em resposta às suas necessidades de informação. Garcia-Molina, Georgia Koutrika e Parameswaran referem-se a esses sistemas com o nome genérico de “mecanismos de provisão de informação” e concluem que:

Assim, um dos problemas fundamentais da Ciência da Computação tornou-se ainda mais crítico hoje: como identificar objetos que satisfaçam as necessidades de informações de um usuário. O objetivo é apresentar ao usuário apenas informações que sejam de interesse e relevância, no lugar e no tempo certos (Garcia-Molina; Koutrika; Parameswaran, 2011, p. 121, tradução nossa).

O objetivo principal de toda pesquisa em Recuperação de Informação é melhorar a relevância dos resultados apresentados pelos sistemas. Nesse sentido, relevância é a noção básica subjacente a todo sistema de recuperação de informação e às ferramentas de busca da web.

Os sistemas de recuperação de informação usam uma variedade de métodos, técnicas e algoritmos com o objetivo de representar e organizar objetos de informação para que possam ser recuperados posteriormente; este tipo de relevância, calculado por algoritmos e programas, pode ser denominado de “relevância de sistema”. Após realizar uma busca, o usuário de um sistema de recuperação avalia e julga os resultados obtidos utilizando a sua própria concepção de relevância, em função de sua necessidade de informação; essa relevância pode ser denominada de “relevância humana” ou “relevância do usuário”. Tanto a “relevância humana” quanto a “relevância do sistema” tratam esse conceito como uma relação entre um objeto informacional e uma entidade que irá julgar esse objeto. Porém, cada um desses mundos pode ter premissas diferentes para avaliar essa relação.

Na Ciência da Informação, as preocupações com a noção de relevância decorrem de problemas operacionais relacionados à recuperação de informação. Segundo Saracevic (1999), a Ciência da Informação é um campo de prática profissional e de investigação científica que aborda o problema da comunicação efetiva de registros de conhecimento entre seres humanos

nos contextos social, organizacional e individual. A questão principal está relacionada ao problema da necessidade e ao uso da informação, envolvendo registros de conhecimento de todos os tipos e em todos os suportes. Para atender a essa necessidade, a Ciência da Informação lida com técnicas, procedimentos, sistemas e tecnologias. No entanto, para a compreensão da Ciência da Informação, assim como de qualquer outro campo, além da definição, é importante uma descrição dos problemas por ela abordados.

O objetivo geral deste trabalho é definir o conceito de relevância, como esse conceito evoluiu ao longo dos anos e apresentar as suas propriedades e manifestações em diferentes contextos e na Ciência da Informação. Para tal, foi realizada uma pesquisa de cunho bibliográfico em que foram analisadas diferentes áreas em que o conceito de relevância tem papel primordial.

Este trabalho é resultado de pesquisa bibliográfica, exploratória e descritiva realizada a partir de teses, dissertações, livros e artigos de periódicos na área da Ciência da Informação, que têm o conceito de relevância como objeto central de estudo no contexto dessa área.

Na seção 2, são apresentadas as definições do conceito de relevância e as suas manifestações, sendo que uma manifestação diz respeito à existência ou à presença de algo. Considerando que um modelo é uma abstração ou uma representação da realidade, na seção 3, são apresentados alguns modelos de relevância, tendo como classificação básica os pontos de vista do sistema e do usuário. Na seção 4, são apresentadas algumas teorias da relevância com base nas áreas da Filosofia e da Comunicação. Por fim, na seção 5, são traçadas algumas considerações sobre o trabalho, enfatizando que, apesar da importância do conceito de relevância para a Ciência da Informação, ainda não existe uma base teórica sedimentada a respeito desse conceito.

2. Relevância e suas manifestações

Em seu uso comum, o termo “relevância” é geralmente utilizado para identificar “o que releva”, que “sobressai” ou “ressalta”, “de grande valor, conveniência ou interesse”, “aquilo que importa ou é necessário” (Relevância, c2009). Relevância é uma característica atribuída a tudo aquilo que seja de fundamental importância para algo ou alguém. Essa característica está relacionada, em termos comparativos, a um elemento que se destaca em relação a outros.

Em muitos casos, a relevância adquire significado físico e objetivo, podendo se referir a uma parte protuberante ou uma saliência em uma superfície lisa; situações nas quais se quer ressaltar a parte que não está de acordo com a superfície plana. Porém, na maioria das vezes, o termo está ligado a julgamentos de valor ou atribuição de importância de uma determinada entidade, seja ela física (um livro, um autor ou, uma obra, por exemplo) ou abstrata (como um assunto, uma notícia ou uma opinião). Nesse aspecto, o conceito relevância adquire um caráter subjetivo e pessoal, próprio de um indivíduo em particular.

A relevância sempre envolve uma relação. Sempre existe um “para” associado à relevância que se refere a um contexto, um assunto em questão. Algo é relevante **para** alguém e **para** um determinado contexto.

Ao se comunicarem na busca por informações, na consulta a objetos que potencialmente transmitem informações e em muitas outras trocas interativas, as pessoas usam a relevância. Elas a aplicam para: filtrar, avaliar, determinar utilidade, inferir, classificar, aceitar, rejeitar, associar, classificar e para outros processos similares. Algumas informações ou objetos de informação podem ter diferentes níveis de relevância para uma determinada pessoa, em determinada situação. Portanto, o conceito de relevância não é necessariamente binário, existindo gradações que se modificam à medida que as intenções e os horizontes cognitivos mudam, ou quando o assunto em questão é alterado (Saracevic, 2007). Segundo Sperber e Wilson (2005, p. 224), a relevância não se manifesta apenas em níveis de presença ou ausência, mas pode atingir maior ou menor grau: “intuitivamente, relevância não é uma questão de tudo ou nada, mas uma questão de graus”.

2.1 As manifestações de relevância

Uma manifestação é a indicação da existência, presença relativa a algo ou alguma coisa. Como muitas outras noções ou fenômenos, a relevância tem várias manifestações. Os esforços para especificar as manifestações de relevância concentram-se na identificação de quais objetos estão relacionados por um determinado tipo de relevância. Diferentes manifestações são apresentadas por diferentes objetos, sendo relacionados por diferentes propriedades usadas para uma relação.

Brian C. Vickery (1959a) foi o primeiro a reconhecer que a relevância apresenta diferentes manifestações. Em seu trabalho, Vickery identificou uma dualidade de manifestações de relevância: a relevância do usuário e a relevância do assunto. Segundo o autor, a relevância do usuário, por um lado, e a relevância do assunto ou tópico, por outro, representariam as manifestações básicas de relevância. Já Mizzaro (1997; 1998) sugeriu uma classificação que acomodasse todas essas manifestações de relevância, propondo que elas pudessem ser classificadas em um espaço quadridimensional: (1) recursos de informação: documentos, substitutos (representações) de documentos e informações; (2) representação do problema do usuário (necessidade real de informação, necessidade percebida, solicitação (*request*) e consulta (*query*)); (3) tempo (interação entre outras dimensões alteradas ao longo do tempo); e (4) componentes (tópico, tarefa e contexto). Tais dimensões representariam as várias manifestações de relevância.

Muitos outros trabalhos abordaram a mesma questão das manifestações de relevância, mas foi o trabalho de Saracevic (1996) que atingiu um certo consenso dentro da literatura. O autor propôs um modelo de manifestação de relevância baseado nos modelos anteriores e que hoje são utilizados quase que unanimemente pela Ciência da Informação. Podemos, com base em seu modelo, distinguir entre as seguintes manifestações de relevância:

- *Relevância sistêmica ou algorítmica*: relação entre uma consulta (expressão de busca) e a informação ou objetos de informação armazenados em um sistema. Cada sistema tem maneiras pelas quais os objetos são representados, organizados e comparados a uma consulta. Eles abrangem uma suposição de relevância, pois a intenção é recuperar um conjunto de objetos que o sistema calculou como relevante para uma determinada consulta. A eficácia comparativa em inferir relevância é o critério para a relevância do sistema;

- *Relevância tópica ou de assunto*: relação entre o assunto ou tópico expresso em uma consulta e o tópico ou assunto coberto por objetos de informação recuperados. Supõe-se que as consultas e os objetos de informação possam ser identificados como sendo de um tópico ou assunto;
- *Relevância ou pertinência cognitiva*: relação entre o estado de conhecimento do usuário e as informações ou objetos de informação. Correspondência cognitiva, informatividade, novidade e qualidade da informação são critérios pelos quais a relevância cognitiva é inferida;
- *Utilidade ou relevância situacional*: relação entre a situação, tarefa ou problema em questão e os objetos de informação. A utilidade na abordagem de uma determinada tarefa, na tomada de decisões, na adequação da informação, na resolução de um problema e na redução da incerteza é o critério pelo qual a utilidade ou a relevância situacional são inferidas;
- *Relevância afetiva*: relação entre as intenções, objetivos, emoções e motivações de um usuário e os objetos de informação. A satisfação, o sucesso, a realização e outros fatores são critérios para inferir a relevância motivacional.

Conforme cita Saracevic (1996, p. 206), classificar manifestações de relevância não diz nada sobre o processo de inferência, a dinâmica da interação entre diferentes manifestações e as variáveis e efeitos envolvidos. Nenhuma classificação pode fazer isso. No entanto, o autor afirma que “a classificação das manifestações de relevância contribui para uma compreensão mais profunda da relevância”.

3. Modelos de relevância

Um modelo é uma versão simplificada de uma dada realidade; é uma representação ou mesmo uma idealização. Os modelos também podem ser abstrações, formando ideias gerais a partir de exemplos específicos, apresentados como uma descrição de uma entidade, processo ou estado de coisas. Modelos e modelagem são indispensáveis para alguns campos científicos. A ciência cognitiva, por exemplo, repousa na suposição de que as pessoas têm modelos mentais. Grande parte dos avanços na Ciência da Computação são fruto da construção e manipulação de modelos.

Uma mesma coisa pode ser representada por diferentes pontos de vista e por meio de diferentes modelos, que podem sugerir ações diferentes e se basear em diferentes suposições. Isso vale também para os modelos que envolvem relevância. Surgiram vários modelos de relevância que evidenciaram que a relevância é “modelável”.

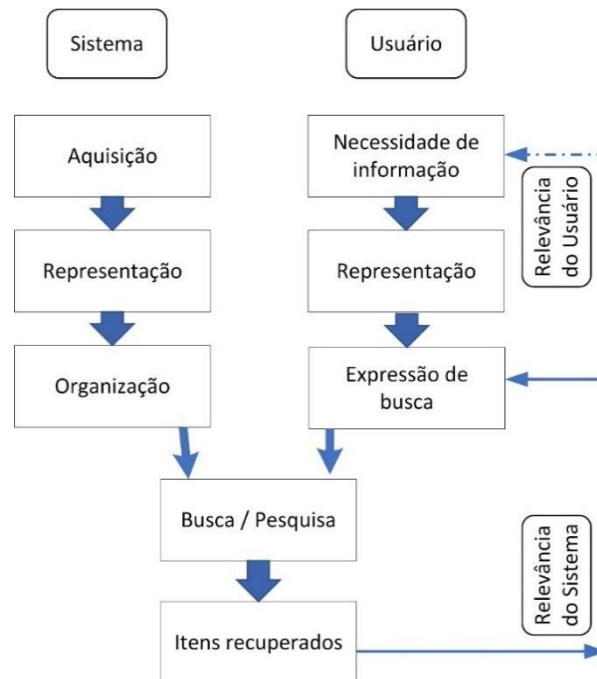
3.1 Visão sistêmica

Um sistema é um conjunto de elementos (às vezes chamados de componentes) em interação, formando um todo complexo. Os sistemas criados pelos seres humanos são um conjunto de elementos que executam certas funções para um determinado objetivo.

Os sistemas de recuperação de informações são um conjunto de elementos que interagem com os usuários com o objetivo de recuperar informações (ou objetos de informação) que sejam relevantes para estes. Vários modelos de sistemas de recuperação foram propostos, cada um enfatizando diferentes funções desse processo, mas todos tomam como certo que a relevância está envolvida em seu propósito.

As principais funções dos sistemas de recuperação de informações são: aquisição de objetos de informação, sua representação e organização e fornecimento de um mecanismo de pesquisa em resposta a uma consulta. Uma expressão de busca (consulta) é a representação de uma determinada necessidade de informação relacionada a um determinado problema em questão ou situação encontrada por um usuário. Em relação aos resultados da busca, pode-se distinguir entre uma probabilidade de relevância atribuída pelo sistema e uma relevância absoluta atribuída pelo usuário.

Figura 1 – Modelo clássico de sistema de recuperação de informação



Fonte: adaptado de (Saracevic, 2017)

Como ilustrado na Figura 1, um sistema de recuperação de informação tem a tarefa de adquirir, representar e organizar objetos de informação. Por outro lado, um usuário precisa representar a sua necessidade de informação em uma expressão de busca. Por meio de uma comparação entre a representação dos objetos de informação e a expressão de busca é recuperado um conjunto de itens que supostamente atenderão à necessidade do usuário. Ao final temos um embate entre a relevância do sistema e a relevância do usuário. Na maioria das vezes nem todos os itens considerados relevantes pelo sistema serão considerados relevantes pelo usuário, sendo necessário, por parte do usuário, um processo de seleção.

Com algumas variações, os elementos básicos e as funções do modelo clássico estão presentes desde o modelo de van Rijsbergen (1979) até o modelo de Baeza-Yates e Ribeiro Neto (2011), além de muitos outros trabalhos sobre Recuperação de Informação nas últimas décadas. Assim, o modelo clássico de recuperação de informação, formulado na década de 1960, depois de mais de meio século de pesquisas, ainda é largamente utilizado nos atuais sistemas de recuperação de informação e mecanismos de busca da web. A implementação de várias funções melhorou drasticamente esses sistemas ao longo do tempo, mas o conceito de relevância sempre permaneceu como elemento principal de seus objetivos.

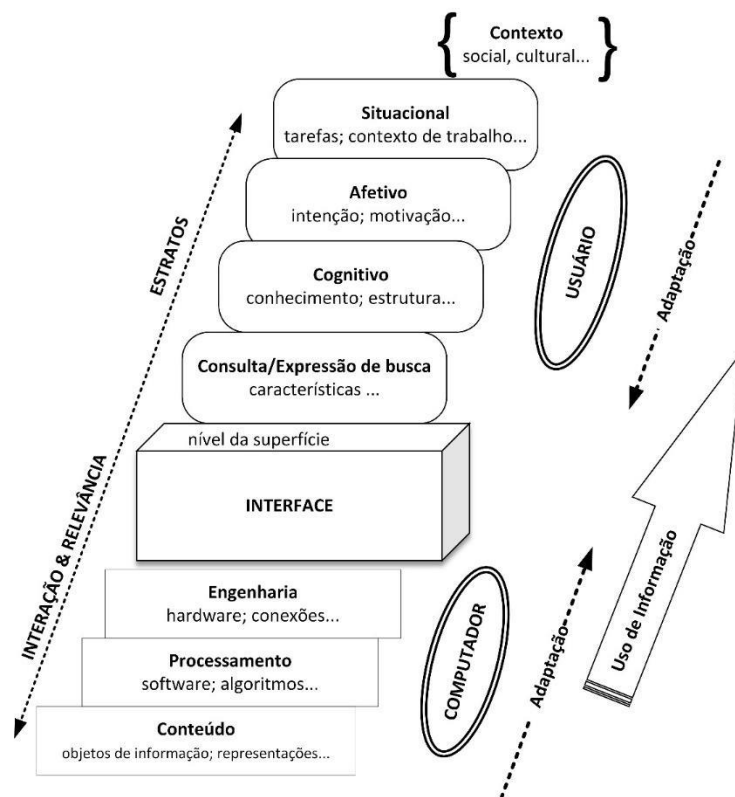
3.2 Ponto de vista dos usuários

Ao longo dos anos a visão de relevância do usuário foi defendida em vários trabalhos, produzindo alguns modelos, dois dos quais são revisados a seguir.

3.2.1 Modelo estratificado de relevância

Um modelo estratificado de relevância, conforme ilustrado na Figura 2, foi proposto e elaborado, posteriormente, por Saracevic (1997; 2007). Modelos que visualizam um objeto complexo e entrelaçado (processo, estrutura, sistema, fenômeno, noção) de maneira estratificada foram sugeridos em vários campos, da linguística à medicina, à meteorologia, à estatística e muito mais. “Estratificado” significa que o objeto modelado é considerado em termos de um conjunto de camadas interdependentes e em interação; é decomposto e composto de volta em termos de camadas ou estratos.

Figura 2 – Modelo estratificado de interações de relevância



Fonte: adaptado de (Saracevic, 2017)

O modelo ilustrado na Figura 2 é um modelo integrador, o que significa que ele incorpora um ponto de vista do sistema e do usuário. Nesse modelo, as interações são descritas

como envolvendo um conjunto de camadas ou estratos. O modelo inicia com as seguintes suposições (Saracevic; Kantor, 1988):

1. os usuários interagem com os sistemas de recuperação para usar informações;
2. o uso da informação está conectado à cognição e depois à aplicação situacional;
3. o contexto está relacionado à relevância.

Os principais elementos do modelo estratificado são usuário e computador, cada um com uma série de variáveis próprias interagindo por meio de uma interface. Os estratos não estão necessariamente embutidos um no outro, nem formam uma hierarquia. As relações e interações entre estratos são complexas e dinâmicas.

O lado do usuário opera em relação a um contexto e apresenta três níveis: cognitivo, afetivo e situacional. Os níveis relacionados ao computador são engenharia (hardware), processamento (software, algoritmos) e conteúdo (objetos e/ou representações de informação). Um determinado nível pode ser mais bem definido do que outros. Alguns níveis podem ser adicionados, dependendo do conjunto de condições ou da ênfase na análise. Além disso, o contexto situacional ou geral pode mudar, resultando em mudanças ou adaptações em vários estratos. O conteúdo e outros aspectos do lado do computador também podem mudar, gerando adaptações novamente. A direção do uso de informação é governada por aspectos contextuais cognitivos, afetivos, situacionais ou mais amplos.

Embora o contexto seja reconhecido como um aspecto importante que afeta a busca por informação em geral e a relevância em particular, a própria noção do que constitui o contexto na Ciência da Informação é relativamente ambígua e não existe consenso (Kelly, 2006). No modelo estratificado, o contexto é tratado da mesma forma que nos estratos do contexto situacional ou de trabalho específico para os contextos sociais mais gerais (incluindo organizacional, institucional e comunitário) e cultural (histórico). Em grande medida, os contextos determinam o problema/situação em questão. Isso corresponde à noção de “situação problemática”, conforme conceituada por Schutz e Luckman (1973, p. 116).

Quando uma busca é realizada, uma variedade de interações é instanciada na interface ou no nível da superfície, mas a interface não é o foco das interações, apesar de poder, por si só, oferecer suporte ou frustrar efetivamente outras interações. Podemos pensar na interação como uma sequência de processos que ocorrem em vários níveis ou estratos conectados.

A interação é, então, um diálogo entre os participantes – usuário e computador – por meio de uma interface, com o objetivo principal de afetar o estado cognitivo do usuário para o uso eficaz de informações relevantes em conexão com um aplicativo, incluindo um contexto. O diálogo pode ser reiterativo, incorporando vários tipos de *feedbacks*, e pode exibir vários padrões.

Cada estrato/nível envolve diferentes elementos ou processos específicos. Do lado humano, os processos podem ser fisiológicos, psicológicos, afetivos e cognitivos. Do lado do computador, eles podem ser físicos e simbólicos. A interface fornece uma interação no nível da superfície em que:

1. Os usuários mantêm um diálogo, realizando declarações (por exemplo, comandos) e recebendo respostas por meio de uma interface com um computador, não apenas para pesquisar, mas também para envolver-se em vários outros processos, tais como: entender e obter os atributos de um determinado componente de computador ou recurso de informação; navegar por entre os recursos de informação; determinar o estado de um dado processo; visualização de resultados; obter e fornecer vários tipos de *feedbacks*; fazer julgamentos.
2. Os computadores interagem com os usuários com determinados processos e “entendimentos” próprios e fornecem respostas; eles também podem solicitar ações ou respostas por parte do usuário.

Na Recuperação de Informação temos um sistema dinâmico e interdependente de relevância. Essa pluralidade foi descrita por Schutz (1970), que denominou “sistema de relevância”, e por Sperber e Wilson (1995), que citam os “princípios de relevância”. No processo de recuperação de informação, a relevância se manifesta em diferentes estratos. Embora muitas vezes possa haver diferenças nas inferências em estratos distintos, essas inferências ainda são interdependentes. O ponto principal da avaliação da recuperação de informação é (ou deveria ser) comparar inferências de relevância de diferentes níveis.

3.2.2 Modelo integrado

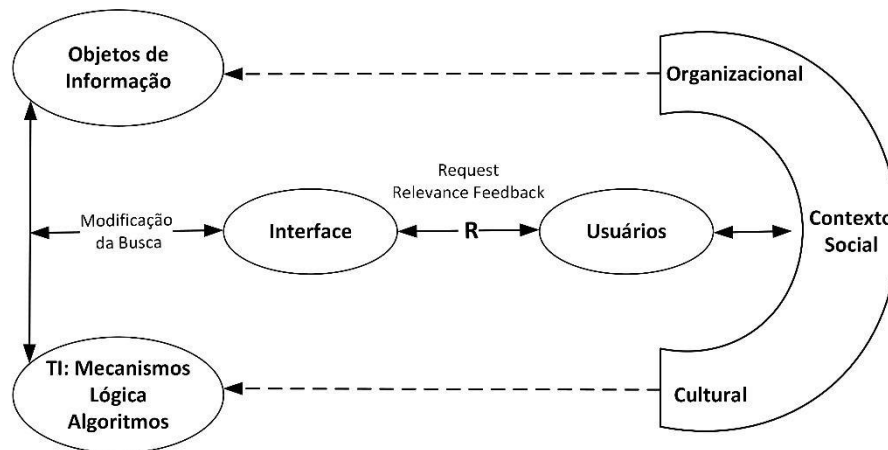
Vários trabalhos tentaram reconciliar os dois pontos de vista, sistemas e usuários, sugerindo modelos de relevância integrados como uma solução para o problema. Partindo de

um ponto de vista cognitivo (usuário), Ingwersen e Järvelin (2005) produziram muitos trabalhos que descrevem a integração de abordagens. Segundo esses autores:

É hora de olhar para trás e esperar desenvolver uma nova visão integrada da busca e recuperação de informações: o campo deve desligar seus caminhos estreitos separados de pesquisa e construir uma nova avenida (Ingwersen; Järvelin, 2005, p. vii, tradução nossa).

Esses autores revisaram todo e qualquer modelo usado na recuperação de informação e produziram seu próprio modelo, integrando aspectos cognitivos e sistêmicos. O modelo integrativo de Ingwersen-Järvelin é ancorado na cognição, refletindo a complexidade do processo. O modelo é composto por cinco componentes centrais (Figura 3).

Figura 3 – Modelo integrado de busca e recuperação de informação (IS&R).



Fonte: adaptado de (Ingwersen; Järvelin, 2005)

Ingwersen e Järvelin (2005) apresentam o modelo e discutem as particularidades de cada um dos componentes. Diversas variáveis, relacionamentos e interações são reconhecidos. Os contextos desempenham um papel importante, e cada um dos componentes tem um contexto associado ou aninhado. Além disso, cada componente está envolvido em inúmeras e variadas interações, dependendo, entre outros fatores, de tarefas, interesses e contexto de trabalho.

3.3 Críticas

Todo modelo pode ser alvo de críticas. Afinal, modelos são simplificações que os deixam vulneráveis a questionamentos. O modelo clássico de recuperação de informação não é exceção. Dois tipos diferentes de críticas são sintetizados aqui: o primeiro lida com a natureza do modelo em si, o segundo com a aplicação do modelo em testes de recuperação de informação.

3.3.1. Estático vs Dinâmico

Várias críticas foram elaboradas a respeito da deficiência do modelo clássico considerar a relevância como um conceito estático, enquanto, na realidade, os processos que envolvem relevância são altamente dinâmicos e interativos. A primeira crítica importante e amplamente citada foi feita por Schamber, Eisenberg e Nilan (1990), que reexaminaram o pensamento sobre relevância na Ciência da Informação e abordaram o seu papel no comportamento da informação humana e na avaliação dos sistemas. Esses autores resumiram ideias e experimentos importantes e chegaram à conclusão de que a relevância, e com ela a recuperação de informação, deve ser modelada como dinâmica e situacional, e não estática. Inúmeros artigos seguiram elaborando a mesma crítica (Saracevic, 2017).

Assim, a interação tornou-se objeto de pesquisa, como relatado por Borlund (2003), Ju e Gluck (2011) e muitos outros. As propriedades dinâmicas da relevância foram discutidas e demonstradas em experimentos e prontamente reconhecidas pelos autores. As ideias desses trabalhos inspiraram uma nova onda de experimentos sobre relevância. Belkin (2015) argumenta que a interação é o que as pessoas realmente fazem ao usar sistemas de recuperação de informação e apresenta um longo conjunto de críticas ao modelo clássico.

3.3.2 Sistema vs Usuário

O modelo clássico de recuperação de informação mostra duas ramificações que se reúnem no processo de busca por informação: uma orientada ao sistema e a outra ao usuário. O lado do sistema concentra-se nas funções necessárias para preparar os objetos de informação para a busca e a condução da busca em si. A relevância do sistema é o resultado de um conjunto de premissas e algoritmos para classificar objetos de informações em relação à expressão de busca (consulta) enunciada pelo usuário. O lado do usuário se concentra no contexto, no problema que produziu uma necessidade de informação subjacente à pergunta, que por sua vez é traduzida em uma expressão de busca.

Essa dupla orientação produz duas relevâncias distintas: relevância do sistema e relevância do usuário. A relevância do sistema é, na atualidade, determinada por algoritmos

processados por máquinas; a relevância do usuário é resultado do comportamento humano e do contexto local.

Os testes de sistemas de recuperação de informação, geralmente praticados em laboratórios, concentram-se exclusivamente no lado do sistema, a relevância do sistema, ignorando completamente a relevância do usuário. Esse se tornou um ponto importante de crítica e até de discórdia entre estudiosos do assunto. Repetidamente e de várias maneiras foi apontado que há uma desconexão entre como os sistemas (máquinas, algoritmos) determinam o que pode ser relevante e como os humanos decidem o que é relevante.

3.4 Os pontos de vista do sistema e do usuário

A relevância é participante de uma disputa mais ampla que começou nos anos 80 e ainda prossegue. Envolve duas visões ou modelos opostos: sistemas e usuários. Segundo Saracevic (2017), os defensores da visão focada no usuário criticaram veementemente os defensores da visão focada no sistema. Alguns autores reconciliadores (como Ingwersen e Järvelin, 2005) tentaram resolver essas diferenças. O ponto de vista dos sistemas é baseado no modelo clássico de recuperação de informação, enfatizando o processamento de objetos de informação pelo sistema. O usuário é representado por uma expressão de busca e não é considerado em nenhum outro aspecto. Em outras palavras, o usuário é considerado um dado a ser ignorado. Além disso, a interação entre sistema e usuário não é considerada.

O ponto de vista do usuário não considera o lado do sistema, considerando-o apenas um dado que pode ser ignorado. O usuário é considerado muito além da consulta, buscando incorporar uma série de dimensões cognitivas e sociais, além da interação. O ponto de vista do usuário não tem um contraponto ao modelo de recuperação clássico amplamente aceito, embora alguns tenham sido propostos.

Historicamente, a defesa mais enfática da visão focada no usuário veio em uma revisão crítica por Dervin e Nilan (1986). Ao revisar abordagens alternativas para a avaliação das necessidades de informação, eles fizeram um apelo por uma mudança significativa de paradigma, ressaltando que a abordagem de sistemas é inadequada aos propósitos da Ciência da Informação. A revisão, considerada um ponto de inflexão nos estudos de usuários, foi muito citada como uma espécie de manifesto. A revisão de Dervin e Nilan não considerou a relevância per se, mesmo ela sendo predominante. Obviamente, os estudos do comportamento humano da

informação (que incluem a busca de informações e os estudos do usuário) podem investigar aspectos que não envolvem relevância. No entanto, ao considerar qualquer aspecto de recuperação e uso, a relevância pode estar explicitamente presente ou não.

Os estudos de usuários têm uma longa história, remontando à Segunda Guerra Mundial. Nos anos 1960, os estudos com usuários tornaram-se fortemente orientados para a recuperação de informação; nos anos 2000 para a web e na década atual volta-se para as mídias/redes sociais. Ao longo do tempo, esses estudos evoluíram para uma área mais ampla do comportamento humano. Surgiram vários estudos a respeito do comportamento informacional, definido por Wilson (2000, p. 49) como “a totalidade do comportamento humano em relação às fontes e canais de informação, incluindo a busca ativa e passiva de informações e o uso da informação”.

A relevância tornou-se a característica subjacente e uma premissa fundamental desses estudos e modelos. Os estudos do comportamento informacional surgiram em uma área própria, com pouca ou nenhuma conexão com o estudo de sistemas de recuperação de informação. De um lado, esses estudos são realizados sob os auspícios da Ciência da Informação e muitas disciplinas em que o uso da informação é considerado. Do outro lado estão os estudos empreendidos sob a iniciativa da Ciência da Computação. Duas visões olhando para um mesmo processo.

Em um estudo maciço de padrões de cocitação em Ciência da Informação para o período 1972-1995, White e Mc Cain (1998) mapearam a estrutura do campo, mostrando dois grandes grupos ou subdisciplinas que chamaram de “análise de domínio” e “recuperação de informação”:

As especialidades podem ser agregadas em duas grandes subdisciplinas: (1) O estudo analítico de literaturas aprendidas e seus contextos sociais, incluindo análise de citações e teoria das citações, bibliometria e comunicação em ciências e P&D; e (2) o estudo da interface humano-computador-literatura, compreendendo recuperação experimental e prática, teoria geral dos sistemas de bibliotecas, teoria do usuário, OPACs e teoria da indexação. [Autores] são essencialmente ‘pessoas de literatura’ ou ‘pessoas de recuperação’ (White; Mc Cain, 1998, p. 337, tradução nossa).

Acontece que os dois lados da batalha estão errados. Dervin, Nilan e seguidores de décadas estão errados ao insistir na primazia ou exclusividade da abordagem do usuário. Os desenvolvedores de sistemas estão errados ao ignorar o lado do usuário e tornar o modelo tradicional de recuperação de informação como base exclusiva de suas pesquisas. No fundo, a questão não é a contraposição sistema versus usuário. A questão central é: como podemos fazer com que os lados trabalhem juntos para o benefício de ambos? A resposta de Belkin (2015) está

em considerar a interação entre pesquisa teórica e desenvolvimento da prática. Quando os sistemas falham, o principal motivo é uma falha na representação do conceito de relevância, sendo esse um motivo para advogar a resolução do problema do usuário do sistema de maneira integrativa.

Vários trabalhos tentaram conciliar os dois pontos de vista, sugerindo modelos de relevância integrativos como uma solução para o problema. Um exemplo é o trabalho de Ingwersen e Järvelin (2005) descrito acima.

Na mesma linha, Ruthven (2005) revisa várias abordagens de relevância, de sistemas situacionais e cognitivos, e defende uma abordagem que integra Recuperação de Informação e pesquisa em busca de informações. Enquanto ele parte do ponto de vista dos sistemas, reconhece completamente a natureza limitada da definição de relevância resultante nesse modelo. O referido autor analisa diferentes tipos de avaliações de relevância e sugere que permitir que os usuários de sistemas de recuperação de informação façam avaliações de relevância diferenciadas parece ser apenas um recurso adicional a ser implementado nas interfaces desses sistemas. Ruthven também lida com a dinâmica da relevância – a questão de alterar as avaliações de relevância dos usuários ao longo do tempo – e comenta como os sistemas responderam mal a esse fenômeno. Ruthven (2005, p. 76) conclui: “A metodologia pela qual investigamos a relevância e incorporamos relevância em nossos sistemas afeta a forma como pensamos sobre relevância”.

Apesar das tentativas de unir as duas culturas, elas ainda permanecem estranhas uma à outra.

4. Teorias da relevância

Uma teoria é um conjunto de suposições, asserções e proposições organizadas em uma forma de explicação ou interpretação da natureza, forma ou conteúdo de um fenômeno, ação ou noção. Um modelo não é uma teoria. Uma teoria explica; um modelo representa uma realidade (Saracevic, 2017).

Claramente, a noção de relevância é uma ótima candidata para a teorização. A relevância foi tratada extensivamente na Ciência da Informação, mas ainda não foram feitas tentativas para formular uma teoria da relevância orientada para essa área. Embora a forma das teorias em potencial tenha sido debatida por um longo tempo, não existe uma teoria da relevância na

Ciência da Informação. No entanto, relevância é uma noção universal, e, portanto, o trabalho acadêmico sobre esse conceito estendeu-se a vários outros campos. Isso incluiu a formulação de teorias sobre relevância que apareceram na Lógica, Filosofia e Comunicação.

As teorias de relevância baseadas na Lógica não foram diretamente usadas na Ciência da Informação, embora tenham sido utilizadas na elaboração dos algoritmos de Recuperação de Informação. Uma teoria de relevância trazida da Filosofia foi usada até certo ponto na Ciência da Informação. Por fim, uma teoria de relevância na Comunicação, formulada na década de 1980 e refinada nas décadas seguintes, teve algum impacto no pensamento sobre a relevância na Ciência da Informação.

4.1. Filosofia: relevância no mundo da vida

Vários filósofos, particularmente na área da fenomenologia, têm se interessado pela relevância. De particular interesse para a Ciência da Informação são dois trabalhos: de Alfred Schutz (1970) e Schutz e Luckman (1973). Este último é um resumo das ideias ao longo da vida de Alfred Schutz, completadas postumamente por seu colaborador, Thomas Luckman.

Resumidamente, Schutz caracterizou a estrutura e o funcionamento do “mundo da vida” – situações que as pessoas enfrentam na realidade da vida cotidiana. Essas situações formam camadas; o mundo da vida é estratificado. Schutz especificou que relevância é o princípio para estratificação e interação dinâmica entre estratos. Ele acreditava que não havia apenas uma relevância, mas um sistema interdependente de relevâncias, e propôs uma tipologia com três categorias principais:

1. relevâncias temáticas (na obra de Schutz, de 1970, denominada “Tópica”), envolvendo a percepção de que algo é problemático;
2. relevâncias interpretativas, envolvendo o estoque de conhecimento disponível para apreender o significado daquilo que é percebido; e
3. relevâncias motivacionais, referentes ao curso de ação a ser adotado.

As categorias se interagem. Por exemplo, os resultados da ação adotada podem motivar o processo de obtenção de material interpretativo adicional, e percepções de relevância temática também podem ser afetadas nessa interação dinâmica (Saracevic, 2017).

Conceitos relacionados à relevância ecoam em muitos trabalhos da Ciência da Informação e a estrutura proposta por Schutz é citada várias vezes como sendo apropriada aos seus interesses. Mesmo quando não há uma citação direta a seu trabalho, seu ponto de vista se reflete nos trabalhos sobre manifestações de relevância. Por exemplo, é totalmente aceito na Ciência da Informação que não há um, mas vários tipos de relevância, isto é, que estamos lidando com um “sistema de relevância” ou com diferentes manifestações. As categorias de relevância elaboradas por Schutz também se refletem em diferentes tipos de relevância na Recuperação de Informação.

4.2 Comunicação: relevância e cognição

Segundo Saracevic (1975), no contexto da comunicação podemos considerar a relevância como uma medida da eficácia do contato entre a fonte da informação e o seu destinatário. Uma vez que toda medida é uma relação, a relevância é também uma relação. A comunicação do conhecimento é efetiva quando a informação que é transmitida pelo emissor resulta em mudanças na estrutura cognitiva do receptor. Relevância é a medida dessas mudanças.

A contribuição mais abrangente e ambiciosa para teorizar sobre a relevância em uma estrutura de comunicação foi feita por Sperber e Wilson (1995; 2005). Sua Teoria da Relevância é baseada em um modelo inferencial de comunicação que visualiza a comunicação em termos de intenções, em oposição ao modelo mais tradicional e amplamente aceito de origem-mensagem-destino. O modelo inferencial considera que o recurso crítico da maioria das comunicações humanas – verbais ou não verbais – é a expressão e o reconhecimento de intenções.

A Teoria da Relevância estava originalmente associada à fala cotidiana ou à comunicação verbal, mas depois foi estendida para abranger processos cognitivos mais amplos. Tem o objetivo ambicioso de ser uma teoria da cognição e da comunicação, unindo-as com base na relevância. No entanto, o problema básico abordado na teoria é como a relevância é criada nos diálogos entre as pessoas.

Comunicar é reivindicar a atenção de alguém e, portanto, implica que as informações comunicadas são relevantes. Tanto Wilson quanto Sperber ancoram firmemente a relevância em um determinado contexto e falam sobre efeitos contextuais – a relevância é contextual. Eles

também consideram a avaliação da relevância como comparativa, não quantitativa – isto é, a relevância é o resultado de uma comparação.

No centro de sua teoria, os autores postulam dois princípios que afirmam refletir tendências universais:

1. Princípio Cognitivo de Relevância – a afirmação de que a cognição humana tende a ser voltada para a maximização da relevância, ou seja, se esforça para maximizar a relevância dos insumos;
2. Princípio Comunicativo de Relevância – a afirmação de que todo estímulo ostensivo transmite uma presunção de sua própria relevância.

Em outras palavras, a cognição humana é orientada à relevância, assim como a comunicação. Esses dois princípios levam à especificação de como a relevância (de uma entrada (*input*) para um indivíduo) pode ser avaliada em termos de dois componentes: efeitos cognitivos e esforço de processamento. Isto é:

- Quanto maiores os efeitos cognitivos positivos alcançados pelo processamento de um *input*, maior a relevância desse *input* para o indivíduo naquele momento;
- Quanto maior o esforço de processamento despendido, menor a relevância do insumo para o indivíduo naquele momento (Wilson; Sperber 2006).

Os dois princípios da relevância e os dois componentes da avaliação estão no cerne da Teoria da Relevância, sendo o primeiro explicativo e o segundo preditivo. Assim, a força da teoria reside em propor uma série de explicações e princípios preditivos operacionais sobre cognição e comunicação em termos de relevância.

Sperber e Wilson constroem a Teoria da Relevância considerando a relevância como uma propriedade psicológica humana que faz parte dos processos mentais. A ideia subjacente a essa noção é “a de que os indivíduos prestam atenção apenas aos fenômenos ou às informações que lhes parecem relevantes” (Silveira, 1997, p. 15). Isso significa dizer que o ouvinte, através de sua atenção, irá procurar a informação que é mais relevante numa situação comunicativa, a fim de interpretar a intenção do falante.

A Teoria da Relevância de Sperber e Wilson serviu como uma teoria emprestada em dois trabalhos, primeiro por Harter (1992) e 15 anos depois por White (2007a, 2007b). Ambos

aplicaram a Teoria da Relevância à Ciência da Informação em geral e à Recuperação da Informação em particular, mas de maneiras muito diferentes.

O trabalho de Harter começa com uma rejeição enfática da relevância tópica. Como solução, ele adota a noção de relevância como sendo exclusivamente relacionada a estados cognitivos que mudam dinamicamente, chamando isso de “relevância psicológica”. Relevância é o que causa mudanças cognitivas em um determinado contexto. A essência da proposta de Harter é considerar um determinado tipo ou manifestação de relevância como o aspecto primário ou mesmo exclusivo da relevância.

A força da noção de relevância psicológica de Harter é que ele tenta basear o conceito em uma base teórica mais elaborada: a Teoria da Relevância. O ponto fraco é que, na verdade, ele não faz isso, além de emprestar alguns conceitos e terminologia. No entanto, Harter discute a dificuldade de testá-la e aplicá-la na prática. Infelizmente, ele não chegou lá, mas apontou o caminho e abriu uma discussão ampla. Ainda assim, o valor de sua tentativa de ganhar alguma base teórica para o conceito de relevância na Ciência da Informação é, por si só, inovadora.

Uma segunda e muito mais abrangente tentativa de transferir a Teoria da Relevância de Sperber e Wilson para a Ciência da Informação foi feita por White (2007a, 2007b). Neste trabalho, White restringe a Teoria da Relevância à aplicação dos efeitos cognitivos e do esforço de processamento, mas ele não usa seus dois princípios de relevância. Em um esforço para integrar a Teoria da Relevância com a recuperação de informação e à bibliometria, ele propõe que os efeitos cognitivos e o esforço de processamento também são componentes nas avaliações de relevância, e podem ser usados como mecanismos preditivos para a avaliação operacional da relevância. White estendeu significativamente sua interpretação da Teoria da Relevância às circunstâncias e interesses da Ciência da Informação, com os pontos fortes e fracos da presente teoria.

Os trabalhos de Harter e White são importantes em seus esforços para adaptar uma teoria da relevância. A Ciência da Informação deve ser estimulada a pensar em tais adaptações, mas permanece a questão de saber se vale a pena esse esforço.

5. Considerações finais

Os autores que discutem a relevância na Ciência da Informação não apresentaram nenhuma teoria sobre esse conceito nem adaptaram com sucesso teorias de outros campos, apesar de algumas tentativas. Ainda estamos em busca de uma base conceitual, um conjunto de princípios e proposições testáveis para explicar a noção de relevância aplicável à prática da Ciência da Informação. Um exemplo dessa busca é o artigo de White (2010), apropriadamente intitulado “Relevância na Teoria” (*Relevance in Theory*). Porém, mesmo na ausência de uma teoria, a prática pode ser realizada com sucesso. As técnicas e tecnologias criadas no campo da Recuperação de Informação são um exemplo disso.

As tentativas de emprestar e adaptar teorias têm um efeito positivo no conhecimento empírico e no entendimento sobre o conceito de relevância na Ciência da Informação. A referência de Schutz a “sistemas de relevância” (plural) sugere o reconhecimento de um conjunto já conhecido de manifestações. Os efeitos cognitivos e os esforços de Sperber e Wilson sugerem que as dimensões usadas na avaliação da relevância, incluindo sua natureza dinâmica, também são bem reconhecidas.

Este trabalho possibilitou entender o conceito de relevância e sua utilização na Ciência da Informação, apresentando as suas diferentes propriedades e manifestações nessa área e em diferentes contextos. Com isso, pode-se notar a necessidade de serem realizadas pesquisas que considerem a relevância como sendo um de seus objetos de estudo.

Para alcançar uma compreensão dessa realidade, este trabalho buscou sistematizar as manifestações e o uso do conceito de “relevância” em diversas áreas do conhecimento, apresentando as diferentes interpretações desse conceito na Ciência da Informação.

Assim, a relevância deveria ser vista como um sistema de interação integrativo, entre usuário e sistema, que se manifesta em diferentes níveis, comumente referidos como tipos de relevância.

Embora o conceito de relevância tenha sido tratado extensivamente na Ciência da Informação, ainda não existe uma teoria orientada para a área. Isto pode ser explicado, entre outros fatores, pelo fato de não ter sido desenvolvida nenhuma teoria culturalmente apropriada sobre essa noção e de não se ter conseguido adaptar com sucesso teorias de outros campos.

Referências

- BAEZA-YATES, R. A.; RIBEIRO-NETO, B. A. **Modern information retrieval: the concepts and technology behind search**. [S. l.]: Addison Wesley, 2011.
- BELKIN, N. J. People interacting with information. **SIGIR Forum**, v. 49, n. 2, p. 13-27, 2015.
- BORLUND, P. The concept of relevance in IR. **Journal of the American Society for Information Science and Technology**, v. 54, n. 10, p. 913-925, 2003.
- COOPER, W. S. A definition of relevance for information retrieval. **Information Storage and Retrieval**, v. 7, p. 19-37, 1971.
- DERVIN, B. NILAN, M. S. Information needs and uses: a conceptual and methodological review. **Annual Review of Information Science and Technology**, v. 21, p. 3-33, 1986.
- GARCIA-MOLINA, H.; KOUTRIKA, G.; PARAMESWARAN, A. Information seeking: convergence of search, recommendations, and advertising. **Communications of the ACM**, v. 54, n. 11, p. 121, 2011.
- HARTER, S. P. Psychological relevance and information science. **Journal of the American Society for Information Science**, v. 53, n. 4, p. 257-270, 1992.
- INGWERSEN, P.; JÄRVELIN, K. **The turn: integration of information seeking and retrieval in context**. Dordrecht: Springer, 2005.
- JU, B.; GLUCK, M. Calibrating information users' views on relevance: a social representations approach. **Journal of Information Science**, v. 37, n. 4, p. 429-438, 2011.
- KELLY, D. Measuring online information seeking context. Part 1: background and method. **Journal of the American Society for Information Science and Technology**, v. 57, n. 3, p. 1729-1739, 2006.
- MIZZARO, S. Relevance: the whole history. **Journal of the American Society for Information Science**, v. 48, n. 9, p. 810-832, 1997.
- MIZZARO, S. How many relevances in information retrieval? **Interacting with Computers**, v. 10, n. 3, p. 303-320, 1998.
- RELEVÂNCIA. *In*: DICIO: dicionário online de português. Porto: 7Graus. c2019. Disponível em: <https://www.dicio.com.br/relevancia/>. Acesso em: 24 out. 2022.
- RUTHVEN, I. Integrating approaches to relevance. *In*: SPINK, A.; COLE, C. (Eds.) **New directions in cognitive information retrieval**. Berlin: Springer, 2005. p. 61-80.
- SARACEVIC, T. Relevance reconsidered. *In*: INGWERSEN, Peter; PORS, Niels Ole. **Proceedings COLIS 2: second conference on conceptions of Library and Information Science**. Copenhagen: Royal School of Librarianship, 1996. p. 201-208.

SARACEVIC, T. The stratified model of information retrieval interaction: extension and applications. **Proceedings of the American Society for Information Science**, n. 34, p. 313-327, 1997.

SARACEVIC, T. Information science. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051, 1999.

SARACEVIC, T. Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. **Journal of the American Society for Information Science and Technology**, v. 58, n. 3, p. 1915-193, 2007.

SARACEVIC, T. **The notion of relevance in information science: everybody knows what relevance is: but, what is it really?** [S. l.]: Morgan & Claypool, 2017.

SARACEVIC, T.; KANTOR, P. A study of information seeking and retrieving: III. searchers, searches, and overlap. **Journal of the American Society for Information Science**, v. 39, n. 3, p. 197-216, 1988.

SARACEVIC, T. Relevance: a review of and a framework for the thinking on the notion of information science. **Journal of American Society for Information Science**, v. 26, n. 6, p. 321-343, 1975.

SCHAMBER, L.; EISENBERG, M. B.; NILAN, M. S. A re-examination of relevance: Toward a dynamic, situational definition. **Information Processing & Management**, v. 26, n. 6, p. 755-776, 1990.

SCHUTZ, A. **Reflections on the problem of relevance**. New Haven: Yale University Press, 1970.

SCHUTZ, A.; LUCKMAN, T. **The structures of the life-world**. Evanston: Northwestern University Press, 1973.

SILVEIRA, J. R. C. **Teoria da relevância: uma resposta pragmático-cognitiva à comunicação inferencial humana**. Rio Grande do Sul: PUCRS/BCE, 1997.

SPERBER, D.; WILSON, D. **Relevance: communication and cognition**. Cambridge: Blackwell, 1995.

SPERBER, D.; WILSON, D. Teoria da relevância. **Linguagem em (Dis)curso**, Santa Catarina, v. 5, p. 221-268, 2005.

VAN RIJSBERGEN, C. J. **Information retrieval**. 2. ed. London: Butterworths, 1979.

VICKERY, B. C. Subject analysis for information retrieval. *In: INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION*, 2., 1959, Washington. **Proceedings** [...]. Washington: National Academy of Sciences, 1959a. p. 855-866.

VICKERY, B. C. The structure of information retrieval systems. *INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION*, 2., 1959, Washington. **Proceedings** [...]. Washington: National Academy of Sciences, 1959b. p. 1275-1290.

WHITE, H. D.; MC CAIN, K. W. Visualizing a discipline: an author cocitation analysis of information science 1972-1995. **Journal of the American Society for Information Science**, v. 49, n. 4, p. 327-355, 1998.

WHITE, H. D. Combining bibliometrics, information retrieval, and relevance theory: first examples of a synthesis. **Journal of the American Society for Information Science and Technology**, v. 58, n. 4, p. 536-559, 2007a

WHITE, H. D. Combining bibliometrics, information retrieval, and relevance theory: some implications for information science. **Journal of the American Society for Information Science and Technology**, v. 58, n. 4, p. 583-605, 2007b.

WHITE, H. D. Relevance in theory. *In*: BATES, M. J.; MAACK, M. N. (Eds). **Encyclopedia of Library and Information Sciences**. 3. ed. New York: Taylor and Francis, 2010. p. 4498-4511.

WILSON, T. D. Human information behavior. **Informing science: the international journal of an emerging transdiscipline (special issue on information science research)**, v. 3, n. 2, p. 49. 2000.

WILSON, D.; SPERBER, D. Relevance theory. *In*: WARD, G.; HORN, L. (eds.) **Handbook of pragmatics**. Oxford: Blackwell, 2006. p. 607-632.

Artigo submetido em: 12 jan. 2023
Artigo aceito em: 20 dez. 2023