

RDBCIRevista Digital de Biblioteconomia e Ciência da Informação
Digital Journal of Library and Information Science

Sistema de Armazenamento e Recuperação da Informação: uma análise do impacto das variáveis e medidas visando à organização e recuperação de informação centrado no usuário

Gercina Angela de Lima¹, Maria Luiza Almeida Campos²

RESUMO

Introdução: O efetivo desempenho de Sistema de Recuperação da Informação depende da qualidade com a qual a organização da informação é realizada, o que implicará em uma recuperação da informação mais relevante e pertinente, visto que esses procedimentos são condicionados um ao outro fazendo uma ponte entre a entrada e a saída da informação. **Objetivo:** Avaliar o impacto das variáveis exaustividade e especificidade e das medidas de revocação e precisão, assim como os conceitos de relevância e pertinência, em Sistemas de Recuperação de Informação. **Metodologia:** Caracteriza-se como um estudo descritivo e exploratório, baseado em revisão de literatura narrativa visando apresentar as diversas conceituações, os seus pontos convergentes e divergentes. **Resultados:** Como contribuição apresenta-se uma proposta de fluxo para um Sistema de Armazenamento e Recuperação de Informação, centrado no usuário, reunindo diversos aspectos relacionados às medidas de revocação e precisão, de relevância e pertinência. **Conclusão:** Considera-se como contribuição final deste estudo evidenciar a importância de uma visão sistêmica, na qual todos os elementos de um Sistema de Armazenamento e Recuperação de Informação estão em inter-relação, tendo o usuário como elemento principal; e apresentar as fundamentais atividades que são importantes para a formação de profissionais aptos à construção de Sistemas consistentes.

PALAVRAS-CHAVE

Sistema de armazenamento e recuperação da informação. Revocação. Precisão. Pertinência. Relevância. Avaliação do impacto de medidas em SRIs.

Information Storage and Retrieval System: an analysis of the impact of variables and measures aimed at the organization and retrieval of information centered on the user

ABSTRACT

Introduction: The effective performance of an Information Retrieval System depends on the quality with which the organization of information is performed, which will imply a retrieval of the most

Correspondência do autor

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil / e-mail: limagercina@gmail.com

²Universidade Federal Fluminense, Niterói, RJ, Brasil / e-mail: marialuizalmeida@gmail.com

relevant and pertinent information, since these procedures are conditioned to each other, creating a bridge between input and output of information. **Objective:** To evaluate the impact of the exhaustiveness and specificity variables and the recall and precision measures, as well as the concepts of relevance and pertinence, in Information Retrieval Systems. **Methodology:** It is characterized as a descriptive and exploratory study, based on a narrative literature review aiming to present the different concepts, their converging and divergent points. **Results:** As a contribution, we present a proposal for a flow for an Information Storage and Retrieval System, centered on the user, bringing together several aspects related to measures of recall and precision, of relevance and pertinence. **Conclusion:** It is considered as the final contribution of this study to highlight the importance of a systemic view, in which all elements of an Information Storage and Retrieval System are in interrelation, having the user as the main element; and present the fundamental activities that are important for the training of professionals able to build consistent Systems.

KEYWORDS

Information storage and retrieval system. Recall. Precision. Relevance. Pertinence. Assessing the impact of measures on SRIs.

CRediT

- **Reconhecimentos:** A primeira coautora agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de Produtividade em Pesquisa (PQ-D1).
- **Financiamento:** Este estudo foi financiado pela agência brasileira Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) para as bolsas.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não é aplicável.
- **Disponibilidade de dados e material:** Não aplicável.
- **Contribuições dos autores:** Conceituação, Curadoria de Dados, Análise Formal, Aquisição de Financiamento, Investigação, Metodologia, Recursos, Supervisão, Validação, Visualização, Redação – rascunho original: LIMA, G.A.; CAMPOS, M.L.A. Redação – revisão & edição: LIMA, G.A.; CAMPOS, M.L.A.



JITA: ID. Knowledge representation.

Artigo submetido ao sistema de similaridade



Submetido em: 15/12/2021 – Aceito em: 05/05/2022 – Publicado em: 17/05/2022

1 INTRODUÇÃO

A partir da revolução técnico-científica posterior à Segunda Guerra Mundial, o grande volume de informações gerado no crescente número de áreas do conhecimento passou a demandar um nível maior de organização informacional, pois a informação deve ser ordenada, estruturada ou fixada materialmente, tornando-se um documento, senão permanecerá amorfa e inutilizável. Pode-se afirmar que os avanços que ocorreram desde a década de 1950 até os dias atuais foram relevantes e marcaram o desenvolvimento tanto na forma de armazenamento e representação quanto na recuperação da informação.

Assim as necessidades dos usuários passam a ser o ponto central de estudos na área, já que a recuperação da informação é o principal objetivo de toda a área da Organização da Informação. Com isso, os bibliotecários começaram a ter novos desafios com as mudanças na conceituação e na forma de entrega e acesso dos serviços de biblioteca, tendo, pois, que assumir vários papéis, além do já previsto.

O armazenamento da informação, antes feito somente nos discos rígidos dos computadores com grande capacidade de processamento, agora se dá de maneira diversa, de modo que há a possibilidade de se acessarem arquivos, dados e aplicativos em qualquer lugar e a qualquer hora, utilizando-se tanto um computador como dispositivos móveis, desde que haja conexão com a Internet, por meio do armazenamento nas chamadas “nuvens”.

No âmbito dos processos que envolvem a organização e a recuperação da informação, no contexto da Biblioteconomia e Ciência da Informação (BCI), considera-se a materialidade desses processos na esfera dos Sistemas de Recuperação de Informação (SRI). Para Lancaster (1978), a principal função de um SRI é atuar como interface entre uma população particular de usuários e o universo dos recursos informacionais em forma impressa ou outra. É nesse ambiente que os processos catalogação, indexação e classificação subsidiam a organização e a recuperação da informação dos diversos suportes informacionais. O produto dessas atividades é a elaboração de catálogos de um dado acervo físico, ou de banco de dados de uma biblioteca digital, ou mesmo de um catálogo online e de repositórios digitais.

O efetivo desempenho de SRI depende da qualidade com a qual a organização da informação é realizada, o que implicará uma recuperação da informação mais relevante e pertinente, visto que esses procedimentos são condicionados um ao outro, fazendo uma ponte entre a entrada e a saída da informação. Nesse contexto, deve-se atentar para as variáveis exaustividade e especificidade na realização da indexação, bem como para o nível de revocação e precisão que o sistema propõe a atender aos seus usuários. Isso influenciará a capacidade de recuperação da informação do sistema, resultando em documentos relevantes e pertinentes para o usuário. A relevância consiste no grau de similaridade entre termos que compõem as expressões de buscas de usuários e a ocorrência em documentos da coleção ou nos termos de indexação. Já a pertinência é a relação que existe entre a informação obtida em uma busca que responde à necessidade ou demanda de informação do usuário, ou seja, a informação que é útil para o usuário.

Neste artigo, discute-se o papel da organização da informação e a recuperação no âmbito de um Sistema de Recuperação da Informação (SRI), com o intuito de avaliar o impacto das variáveis exaustividade e especificidade e das medidas de revocação e precisão, assim como os conceitos de relevância e pertinência nos SRIs. Como contribuição, apresenta-se uma proposta de fluxo para um Sistema de Armazenamento e Recuperação de Informação, centrada no usuário, reunindo os diversos aspectos apresentados.

A partir de tais discussões, considerou-se trazer, de forma didática, conceitos importantíssimos para o fazer profissional e que poderão auxiliar nos procedimentos relacionados à elaboração de políticas de indexação, procedimentos fundamentais para a formação de profissionais que se dedicam ao tratamento informacional. Nesse sentido, apresenta-se um fluxo no qual reuniram-se os diversos aspectos de variáveis e medidas dentro de uma proposta de um Sistema de Armazenamento e Recuperação da Informação (SARI).

Nas próximas seções, descreve-se, primeiramente, a metodologia utilizada para realização deste estudo, e apresenta-se a organização e a recuperação da informação como estão sendo consideradas, no âmbito deste estudo. Posteriormente, descreve-se a respeito dos SRI e seus fluxos, e os processos para auxiliar os usuários a alcançar relevância e pertinência em suas buscas; logo após tais discussões são trazidas e consolidadas por meio da proposta de um SARI e, por último, as considerações finais são apresentadas.

2 METODOLOGIA

Este estudo se caracteriza como descritivo e exploratório, com o intuito de compreender, por meio de um fluxograma, como os processos de organização e recuperação da informação (ORI) impactam nas variáveis exaustividade e especificidade e, conseqüentemente, nas medidas revocação e precisão, com o intuito de atingir uma maior relevância e pertinência nos resultados da consulta dos usuários no âmbito de um SRI.

Para estudar como essas variáveis e medidas são evidenciadas nas estruturas de um fluxograma dos processos de SRI no contexto da ORI, propôs-se, primeiramente, elaborar uma revisão narrativa da literatura, já que a pesquisa é conduzida por questões mais abertas, para: (1) mapear os principais fluxogramas de SRI presentes na literatura e (2) coletar insumos iniciais para embasar a proposta teórico-metodológica de um SRI que englobasse todos esses elementos. Esses critérios se justificam porque os fluxogramas que representam os processos de um SRI não situam explicitamente essas variáveis e medidas nesse contexto específico. No caso em questão, selecionaram-se cinco fluxogramas a partir dos seguintes critérios: (1) os autores seminais da área e (2) os mais citados na literatura.

Para tanto, fez-se uma busca exploratória, sem definição temporal, no Google Acadêmico, e em cinco bases especializadas: *Library and Information Science Abstracts* (LISA), *Information Science & Technology Abstracts* (ISTA), *Library, Information Science & Technology Abstracts with Full Text* (LISTA), *Scopus* e *Web of Science*. Essas bases foram escolhidas considerando-se a relevância e a relação que possuem com a área e com a subárea do conhecimento delimitadas para esta revisão, utilizando-se as seguintes expressões de busca, apresentadas no Quadro 1.

Quadro 1. Expressões utilizadas para busca na literatura

Sigla	Expressões de busca
E1	"Sistema de Recuperação da Informação" OR "SRI"
E2	"Information Retrieval Systems" OR "IR system"
E3	"Política de indexação" AND "Organização da Informação" AND "Recuperação da Informação"
E4	"Indexing Policy" AND "Organization of Information" AND "Information Retrieval"
E5	"Exaustividade" AND "Especificidade" AND "Revocação" AND "Precisão" AND "Relevância" AND "Pertinência"
E6	"Exhaustivity and Specifity" AND "Precision and Recall" AND "Relevance and Pertinence"

Fonte: Elaborado pelas autoras.

A princípio, selecionaram-se 48 documentos dos recuperados, utilizando-se três critérios: (1) documentos que tratavam do sistema de recuperação da informação e política de indexação; (2) documentos que apresentavam estudos sobre as variáveis exaustividade e especificidade, sobre as medidas revocação e precisão, e sobre relevância e pertinência no âmbito do SRI; (3) documentos que traziam esses termos no título ou nas palavras-chaves. Desses 48 documentos selecionados, foram utilizados 33 documentos que tratam, especificamente, sobre a temática, dentre os quais se encontram os cinco fluxogramas utilizados para embasar uma nova estrutura que contextualizasse a posição de cada uma dessas variáveis

e medidas dentro de uma proposta de um Sistema de Armazenamento e Recuperação da Informação (SARI), como está apresentado na seção 6 deste artigo.

3 ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

Nesta seção, apresentam-se os conceitos organização da informação e recuperação da informação, que são considerados relevantes para o contexto deste artigo.

3.1 Organização da Informação

A área da Organização da Informação (OI) compreende todos os estudos relacionados aos processos e instrumentos utilizados na organização de recursos informacionais de qualquer natureza, no intuito de viabilizar o atendimento às necessidades de informação de uma determinada comunidade de usuários.

Dahlberg (2006) define organização da informação como a ordenação de objetos de forma a criar um elo entre o objeto de uma área e sua própria atividade. Novellino (1996) afirma que o processo de representação da informação se caracteriza, principalmente, pela substituição do conteúdo descritivo e temático de um documento por uma descrição abreviada, que será armazenada para posterior recuperação.

Assim, a OI objetiva a representação, o armazenamento e a recuperação da informação. De acordo com Barreto (2002), o objetivo do processo de organização da informação é possibilitar e facilitar o acesso à informação, que, por sua vez, tem a competência e a intenção de produzir conhecimento. Nessa medida, compreende-se que a entrada de dados inconsistentes implicará a saída de dados também inconsistentes. Nesse processo, são ativados mecanismos cognitivos que influem tanto na entrada como na saída do sistema de recuperação da informação, porque são dependentes da maneira como usamos nossa mente para realizar abstrações.

No contexto da Biblioteconomia e Ciência da Informação, a representação da informação é realizada por meio dos processos de catalogação, de indexação e de classificação em um SRI, conforme ilustrado na Figura 1.

| 5

Figura 1. Sistema de Recuperação da Informação



Fonte: Adaptado de Cesarino (1985, p. 161).

No processo de catalogação, também conhecido como representação descritiva, descreve-se um item bibliográfico com o intuito de torná-lo único entre os demais em um determinado acervo, permitindo identificá-lo, localizá-lo e representá-lo nos catálogos. Mey (1995, p. 5) considera que “catalogação é o estudo, preparação e organização de mensagens codificadas, com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de

forma a permitir interseção entre as mensagens contidas nos itens e as mensagens internas dos usuários”.

A indexação é outro importante processo de representação que ocorre dentro de um SRI, no qual se espera que o indexador leia o documento e faça a distinção entre a informação relevante e a periférica para melhor representá-lo, para a posterior recuperação. De acordo com a norma *ISO 5963-1985* (1985), a indexação é vista como “[...] a representação do conteúdo dos documentos por meio de símbolos especiais, quer retirados do texto original, quer escolhidos numa linguagem de informação ou de indexação”. Esse processo é realizado em duas etapas: a primeira é a da análise do documento para identificação de seu conteúdo informacional; a segunda, a de tradução dos conceitos nos termos de uma linguagem de indexação, utilizando-se

os sistemas de organização do conhecimento, do tipo tesouros e sistema de classificação bibliográfico.

Enquanto a catalogação descreve as características físicas de um item bibliográfico e a indexação preocupa-se com as questões que envolvem o conteúdo intelectual do documento, a classificação, como processo, envolve a atribuição ordenada e sistemática de cada entidade a apenas uma classe dentro de um sistema de classes mutuamente exclusivas e não sobrepostas, com base em semelhanças e diferenças. Segundo Tristão *et al* (2004, p. 163), “a classificação é um processo mental mediante o qual podemos distinguir coisas, seres ou pensamentos pelas suas semelhanças ou diferenças. ” Trata-se de uma atividade fundamental da mente humana que processa as ideias e as distingue a partir das características em comum. De acordo com Lima (2021), outras conceituações ainda podem ser admitidas para a palavra classificação:

Dependendo do ponto de vista, a classificação é considerada uma disciplina, mas também pode ser o produto que resulta do ato de classificar e, simultaneamente, é a ferramenta utilizada para realizar o processo de classificação (LIMA, 2021).

Na área da Biblioteconomia e Ciência da Informação, pelo menos quatro conceituações podem ser atribuídas à palavra classificação: classificação tem sido estudada como uma disciplina, como um processo de agrupar e ordenar o conhecimento, como o produto do processo de agrupamento e ordenamento, e como instrumento de representação da informação.

3.2 Recuperação da Informação

A Recuperação da Informação (RI) é uma área originária da Ciência da Computação (CC), e a expressão foi atribuída ao engenheiro estadunidense Calvin Mooers, em 1951, que a definiu, à época, como processo que “[...] engloba os aspectos intelectuais de descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação” (MOOERS, 1951, p. 51). Saracevic (1999), no âmbito da Biblioteconomia e Ciência da Informação, afirma que essa concepção de recuperação da informação trazida por Mooers estava centrada na construção de sistemas, porém, a partir do final de 1970, ampliou-se para abordagem centrada no usuário, levando em consideração os modelos mentais dos usuários.

Para Ferneda (2003), o processo de RI consiste em “identificar, no conjunto de documentos (corpus) de um sistema, que atendem à necessidade de informação do usuário”. Assim, a RI é considerada uma operação importante em um Sistema de Recuperação da Informação, que objetiva relacionar a busca do usuário com os itens armazenados na base de dados, por meio de um conjunto de elementos interligados de rotinas de processamento de registros informacionais, visando atender às necessidades de informação de uma comunidade de usuários. No entanto alguns autores definem a RI a partir de diferentes abordagens.

Para Saracevic (1999), a recuperação da informação envolve aspectos intelectuais de descrição de informações e as especificações de busca, além de os sistemas, técnicas e

equipamentos que são aplicados para realizar todo o processo. Enquanto Baeza-Yates e Ribeiro-Neto (2011, p. 1) apontam que:

A recuperação de informação trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas da Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia. A representação e organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse.

Já Salton (1968) considera a RI como uma área de pesquisa que se preocupa com a estrutura, a análise, a organização, o armazenamento, a recuperação e a busca de informação. Por sua vez, Lancaster (1993) a considera como um processo de pesquisar uma coleção de documentos a fim de identificar aqueles textos que tratam de um determinado assunto.

De acordo com Rowley (1994, p. 113), RI é o “processo de localizar documentos e itens de informação que tenham sido objeto de armazenamento”. Para a autora, o processo é composto por três elementos: a consulta, a comparação e o resultado. A consulta é a questão do usuário transformada em estratégia de busca; a comparação é a ação de verificar se a questão formulada combina com os itens armazenados; e o resultado é a lista ou itens que coincide(m) com a busca do usuário (ROWLEY, 2002). No entanto o processo de recuperação da informação depende muito das etapas realizadas na entrada (catalogação, indexação, classificação) e do armazenamento, que impactam diretamente as buscas realizadas em um Sistema de Recuperação da Informação.

4 SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO (SRI): UMA ANÁLISE A PARTIR DOS FLUXOS

Para Rowley (2002), os SRIs e os computadores quase foram usados como sinônimos, porém, antes do surgimento de qualquer computador e da própria informática, os sistemas de fichas e arquivos baseados em papel já existiam. O SRI é uma parte integrante de um sistema de comunicação e ou sistema de informação. Esses sistemas permitem que os usuários procurem informações em uma coleção de documentos (ou outras fontes de informação) por meio de consultas geralmente formatadas como um conjunto de metadados, e as obtenham de modo a atender suas necessidades com relevância e pertinência. Mas os SRIs lidam, pelo menos, com dois problemas diferentes em relação às necessidades de informações ou consultas dos usuários: (1) eles devem distinguir e identificar as informações relevantes relacionadas à consulta e (2) devem obter a resposta rapidamente.

Salton e McGill (1983, p. xi) definem um SRI como “um sistema que trata da representação, do armazenamento, da organização e do acesso aos itens de informação, e isso pode ser em um acervo físico ou digital.” Os autores o consideram como um conjunto interligado de rotinas de processamento de registros informacionais, com propósitos e critérios próprios, visando atender às necessidades de informação de uma comunidade de usuários. Para Silva, Santos e Fereda (2013, p. 29), os sistemas de recuperação de informação “têm por função representar o conteúdo dos documentos do corpus e apresentá-los ao usuário de uma maneira que lhe permita uma rápida seleção dos itens que satisfazem total ou parcialmente a sua necessidade de informação [...]”.

Para Lancaster (1986, p. 1), todo SRI é composto de dois subsistemas: a entrada de informação (*input*) e a saída de informação (*output*), representado na forma de um ciclo, realizado por basicamente três estágios: a representação, o armazenamento e a recuperação, caracterizando-se como um processo contínuo e retroalimentado.

Quando os documentos são selecionados para compor uma coleção, seja física, seja digital, inicialmente passa-se por uma organização levando em consideração as necessidades da comunidade a ser atendida. Essa organização é realizada pelos processos de catalogação,

indexação, classificação e resumo. Sendo assim, o SRI engloba desde os processos de organização até a recuperação da informação pelo usuário.

Existem na literatura várias propostas para representar o fluxo desses subsistemas da entrada da informação, a representação e a recuperação em um SRI. Neste artigo, vamos apresentar alguns desses esquemas, com base na literatura, com o intuito de verificar a estrutura e os processos de seus subsistemas. Para isso, selecionaram-se os esquemas a seguir.

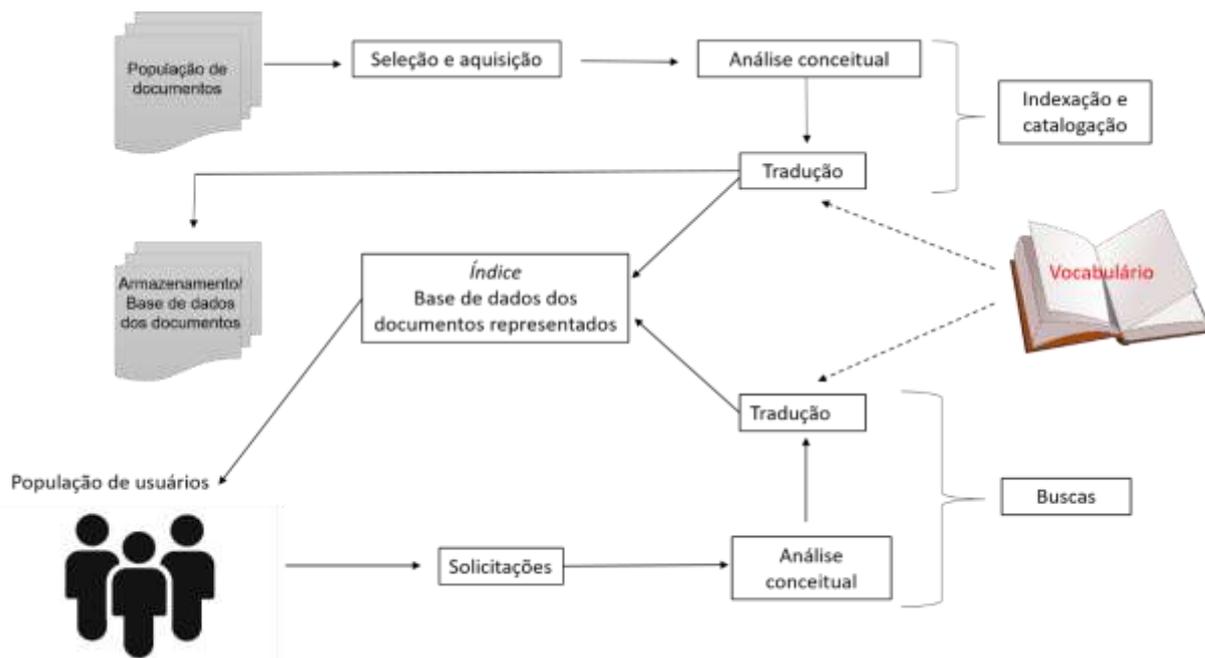
Lancaster, já em 1978, em seu livro *Information Retrieval Systems* (LANCASTER, 1978), identifica os elementos que compõem o SRI. Os subsistemas identificados por ele são: seleção e aquisição; indexação; vocabulário; busca; interação usuário-sistema (negociação da pergunta); “match”¹. Ao chamar todos esses elementos de subsistemas, ele traz no bojo o conceito de interrelação, no qual existe uma relação entre usuários, acervos, vocabulários, indexação e recuperação. Isso significa que eles se afetam mutuamente. A visão de conjunto de um SRI apresentado por Lancaster, como veremos adiante, é uma visão do todo, na qual é de fundamental importância a compreensão do comportamento de cada elemento para que um SRI possa atingir a sua função, ou seja, possibilitar recuperações consistentes aos seus usuários.

Para maior compreensão dessa interrelação, e apontando que todas as ações partem, sempre, da identificação do usuário, como veremos adiante, podemos apresentar as seguintes questões: qual a sua área de atuação (assunto), que atividade desenvolve (professor, aluno, pesquisador)?; que meios de comunicação utiliza?; que linguagem usa (científica, técnica, não é especialista)?; e assim por diante. As respostas vão fornecer elementos para se determinar princípios para formar o acervo, para tratar o documento, para determinar sua classificação e até mesmo ações que, aparentemente, não têm relação, como a quantidade de documentos que um usuário pode levar por empréstimo e prazo permitidos (GOMES; CAMPOS, 1998).

A partir dessa concepção, Lancaster propõe dois esquemas em suas publicações (1986, 1993). No primeiro esquema, em seu livro *Vocabulary Control for Information Retrieval* (1986, p. 3), o autor apresenta os componentes de um SRI. Conforme se pode observar, em sua proposta de fluxo, a realização da representação de um documento centrou-se somente nos processos indexação e catalogação. O autor não inseriu detalhes dos elementos da catalogação, deixando as etapas da indexação no mesmo nível de compreensão. Após essa organização, os documentos são armazenados em uma base de dados impressa ou digital, na qual se podem realizar as buscas para atenderem às solicitações dos usuários, conforme pode ser observado na Figura 2.

¹ Ainda não temos um termo em português para esta ação. Trata-se do encontro do perfil da pergunta com o perfil dos documentos.

Figura 2. Os componentes de um Sistema de Recuperação da Informação

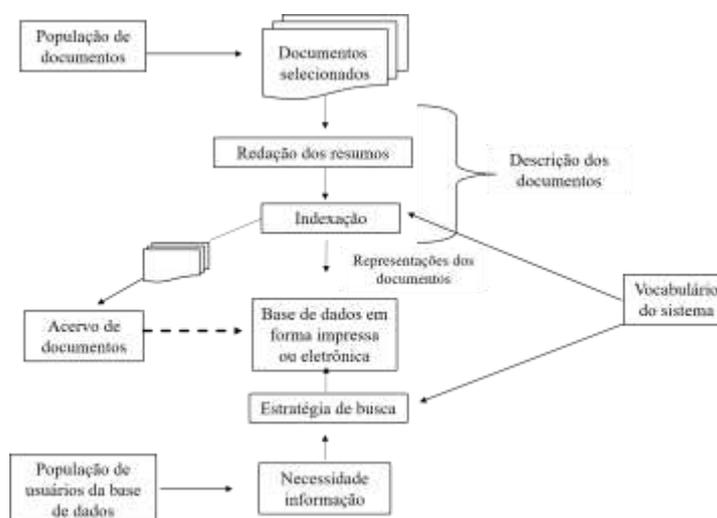


Fonte: Adaptado de Lancaster (1986, p. 2).

O autor considera que a catalogação possui dois elementos importantes, que devem ser considerados: (1) a descrição física de um documento e a (2) escolha do ponto de acesso para representá-lo. Em relação à indexação, o autor a considera como um processo intelectual composto de duas etapas: análise conceitual e tradução, ou seja, faz-se a análise de assunto de um documento e, posteriormente, utiliza-se de um vocabulário controlado para padronizar a representação desse assunto. Porém, nessa sua proposta de fluxo, o autor considera as etapas análise conceitual e a tradução dentro dos processos de indexação e catalogação, apesar de tê-las diferenciado no momento de suas descrições.

Lancaster, em seu livro *Indexação e Resumos: teoria e prática* (1993, p. 2), traz outra proposta, mais ampliada, na qual ele considera, também, a elaboração de índice e resumos, conforme se pode notar na Figura 3. Essas modificações também são consideradas devido aos avanços tecnológicos à época.

Figura 3. Função da elaboração de índice e resumos no quadro mais amplo da recuperação da informação



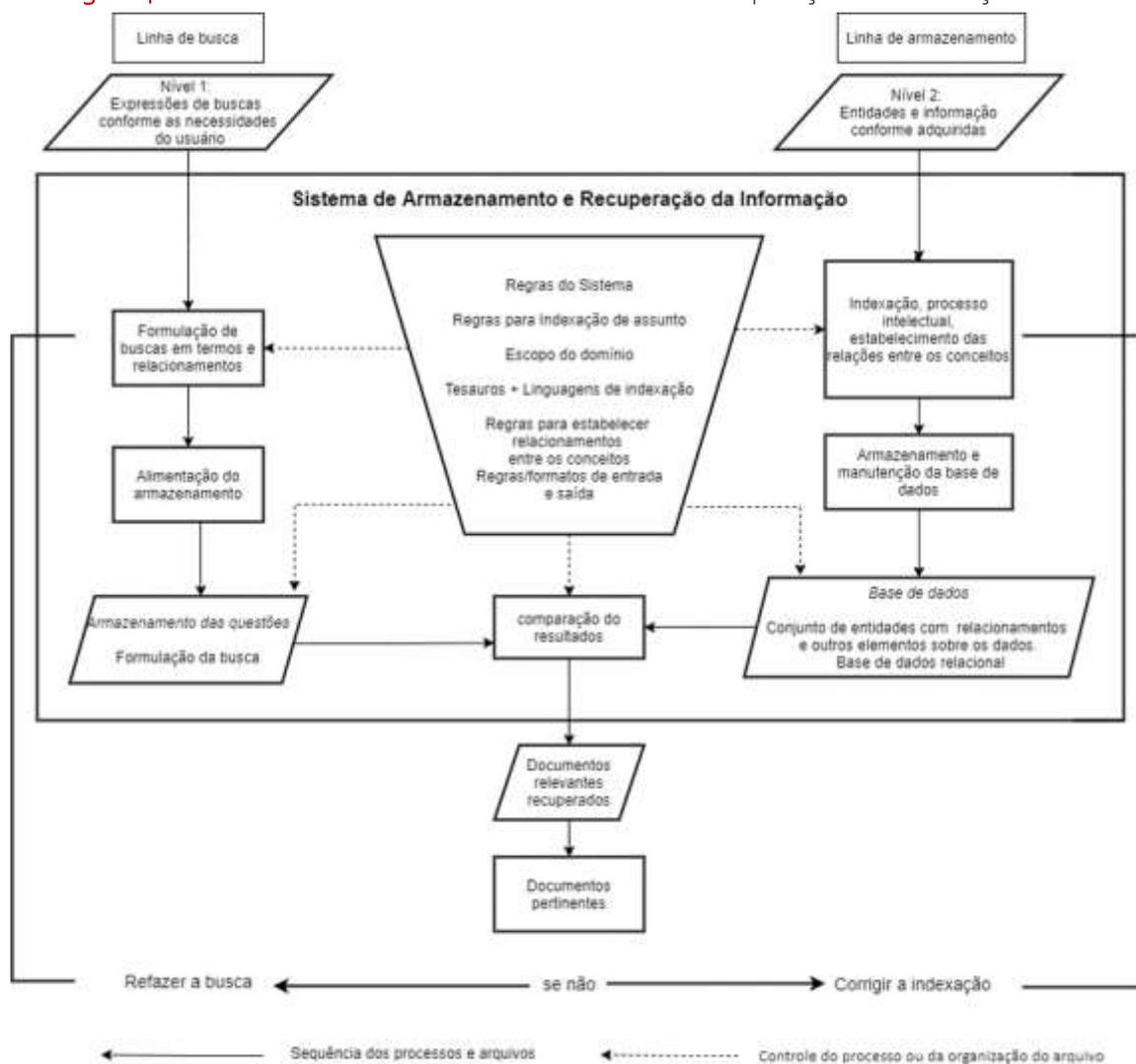
Fonte: Traduzido de Lancaster (1993, p. 2).

Conforme se pode observar, nesse esquema, Lancaster antecede a redação de resumos à indexação como forma de auxiliar na condensação da representação do conteúdo do documento, e os considera como o processo de descrição dos documentos. Por outro lado, o autor aponta a necessidade de planejar a estratégia de busca de acordo com as necessidades dos usuários. Em ambos os subsistemas, o autor sugere o uso do vocabulário controlado para padronizar a terminologia utilizada pelo autor com a do usuário em sua busca. Esses termos atribuídos pelo indexador tornam-se pontos de acesso utilizados para a recuperação dos itens bibliográficos. Uma característica comum nesses esquemas são os processos de análise de assunto e tradução, que podem ocorrer tanto na entrada, ou seja, na representação, quanto na saída, que é a recuperação da informação. Além desses aspectos, nota-se, também, nesse fluxo de 1993, a inclusão da especificação e diferenciação de bases de dados em forma impressa ou eletrônica.

Verifica-se que os fluxos de Lancaster nos alertam para a decisão de quais aspectos de um documento serão representados em um SRI, assim como qual o nível de especificidade ou exaustividade serão atribuídos aos descritores e quais relacionam-se ao conjunto de decisões adotadas pela política de indexação do SRI. Essa decisão perpassa pelo tipo de vocabulário adotado no Sistema, pois ambas, a política de indexação e a atualização do vocabulário, caminham juntas.

Soergel (1985) agrega outro aspecto a ser observado em SRI, a questão que envolve o armazenamento de informação, na qual itens de informação precisam ser processados, pesquisados, recuperados e disseminados para várias comunidades de usuários. Nesse sentido, Soergel (1985) considera que um Sistema de Armazenamento e Recuperação da Informação (*Information Storage and Retrieval – ISAR System*) é um subsistema de um sistema de informação no todo. O autor apresenta uma estrutura de Sistema de Armazenamento e Recuperação da Informação, na qual dá ênfase não somente aos subsistemas de entrada de informação e da saída de informação, como também ao subsistema de armazenamento, conforme pode ser observado na Figura 4.

Figura 4. A estrutura do Sistema de Armazenamento e Recuperação da Informação.



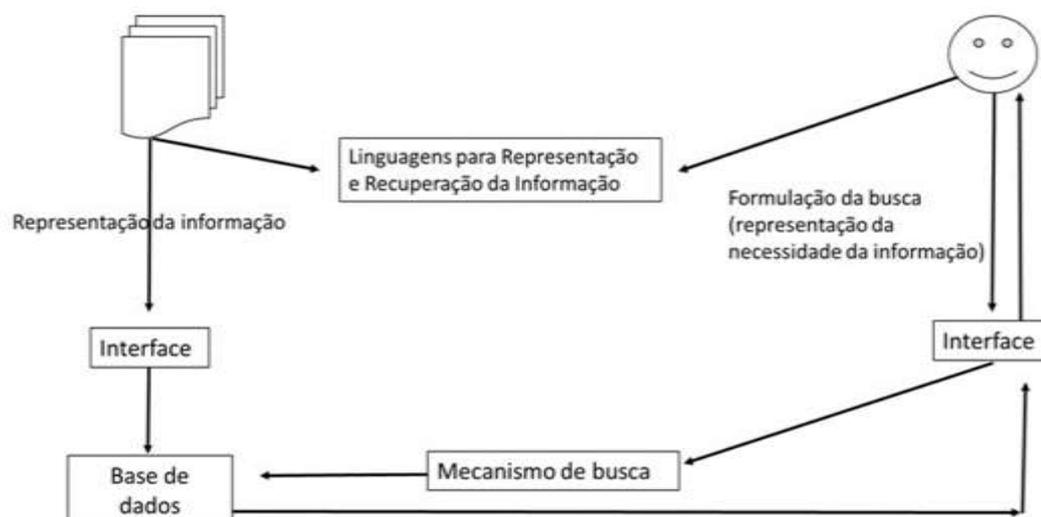
Fonte: Adaptado de Soergel (1985, p. 19).

Diferentemente dos esquemas apresentados por Lancaster (1986, 1993), Soergel começa apresentando, primeiramente, o sistema de saída da informação, ao qual o autor chama de Linha de busca, em que se estudam os perfis dos usuários e formulam-se as buscas a partir de termos e relacionamentos para se chegar aos resultados representados no nível 1. No outro extremo, encontra-se sistema de entrada da informação, com os documentos e dados, no qual ocorre os processos de representação da informação a partir dos processos de indexação (descritiva e de assunto), para que os itens bibliográficos possam ser armazenados e sejam disponibilizados para atender às necessidades de sua comunidade. O autor acrescenta nesse fluxo uma etapa de comparação dos resultados, quando ocorre a avaliação da pertinência dos itens recuperados em relação à estratégia de busca realizada, e o potencial da relevância na ocorrência em documentos do acervo.

Outro ponto importante nesse esquema apresentado por Soergel (1985) é a sugestão a ser realizada, a partir da potencial relevância ou não da análise dos documentos, que seja refeita a busca ou que corrija o processo de indexação realizado para aquele documento. Essa análise, quando realizada, contribui para a manutenção da eficiência e para a precisão do sistema.

Chu (2005) apresenta os principais componentes de um SRI, o qual ele chama de processo de Representação e Recuperação da Informação (RRI). Para o autor, esses componentes são: a base de dados, o mecanismo de busca, a linguagem e a interface, conforme ilustrado na Figura 5.

Figura 5. Processo de Representação e Recuperação da informação.



Fonte: Traduzido de Chu (2005, p. 18).

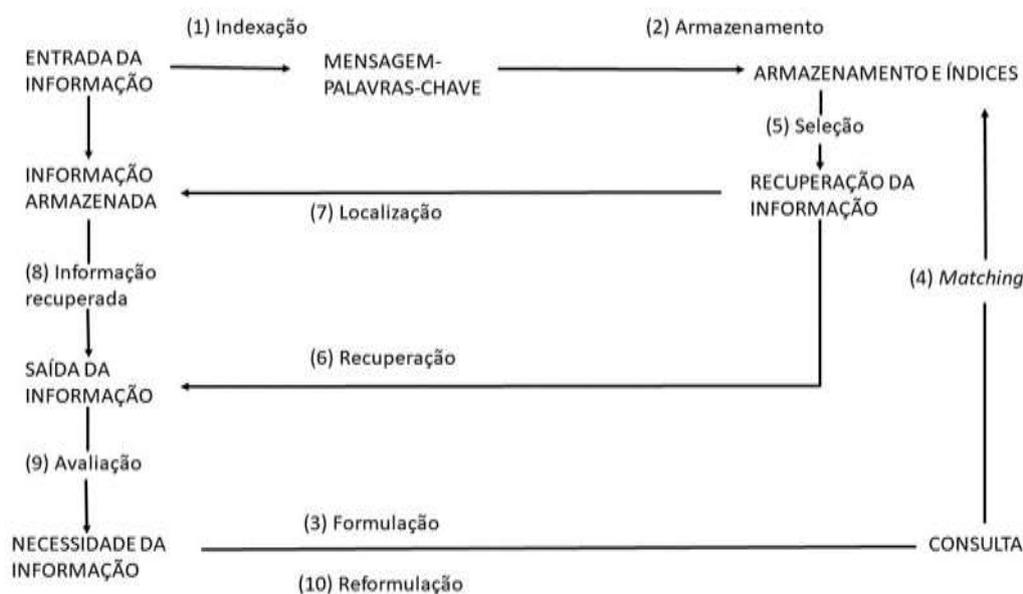
Nesse fluxo, vemos que aparece com ênfase o processo de representação da informação em vez de nominar os processos conforme os esquemas anteriores. O autor também apresenta em um mesmo nível as linguagens e o processo de recuperação da informação, e destaca a interface de acesso à base de dados e a interface para formulação de busca do usuário. Outro ponto importante, evidenciado por Chu (2005), é o papel do profissional da informação na representação da informação com o uso do vocabulário controlado, visto que podem ocorrer discrepâncias durante esse processo, o que pode causar problemas no retorno das buscas, com baixa relevância e pertinência.

Chu (2005) aponta que três problemas podem ocorrer nos processos no âmbito de um SRI: (1) a tipologia documental, que pode não estar em um formato que possa ser arquivado; (2) a dificuldade de encontrar um termo que representa exatamente o conteúdo do documento com o descritor do tesauro; e (3) a inconsistência da representação da informação, que pode ocorrer quando mais de um indexador realiza os processos de representação.

Por outro lado, na formulação da busca, o usuário tem que lidar com a busca em linguagem natural que, na maioria das vezes, não coincide com os descritores utilizados no vocabulário controlado. Nesse caso, o autor sugere que a busca poderá obter mais sucesso se houver uma padronização entre as tomadas de decisão e os instrumentos utilizados na representação e na recuperação da informação (CHU, 2005).

O último esquema que é apresentado, neste trabalho, é o de Vickery e Vickery (2004). Nesse esquema, os autores apresentam o fluxo a partir de 10 processos (1) indexação, (2) armazenamento, (3) formulação da estratégia de busca, (4) *matching* dos resultados, (5) seleção dos documentos pertinentes, (6) recuperação dos documentos, (7) localização dos documentos, (8) informação recuperada, (9) avaliação dos resultados. Se os resultados tiverem sido satisfatórios, termina-se nessa etapa. Caso os resultados não tenham sido satisfatórios, realiza-se a (10) reformulação da busca, conforme apresentado na Figura 6.

Figura 6. Armazenamento e Recuperação da Informação.



Fonte: Traduzido de Vickery e Vickery (2004, p. 118).

Nota-se que, na entrada da informação, nesse esquema, os autores dão ênfase somente ao processo de indexação, sem mencionar os outros processos, passando diretamente para o armazenamento e, com isso, descrevem-se mais detalhadamente as atividades da recuperação da informação, o que atualmente pode ser realizado por meio da interface de um computador. Além disso, diferentemente dos sistemas anteriormente mencionados, este não apresenta a questão do vocabulário controlado como um elemento do sistema.

De forma geral, o que se observa nesses esquemas apresentados é que não existe uma padronização no fluxo das atividades apresentadas pelos autores. Todos os fluxos apresentam os estágios de entrada, armazenamento e recuperação; alguns especificam os processos, e outros, não. Porém, todos têm o objetivo de melhorar a eficácia e a eficiência na recuperação, pois essas medidas impactam diretamente a relevância e a pertinência dos resultados das buscas realizadas pelos usuários. Assim infere-se que a qualidade da informação na entrada de um SRI determina a qualidade da informação na saída.

O resultado das análises realizadas nesses cinco fluxogramas trouxe subsídios para embasar uma nova estrutura que contextualizasse a posição de cada uma das variáveis e medidas dentro de uma nova proposta de um Sistema de Armazenamento e Recuperação da Informação (SARI), apresentado na seção 5 deste artigo, porém se fez necessário apresentar, primeiramente, a importância desses procedimentos e como eles impactam a satisfação do usuário em um SRI.

5 UMA ANÁLISE DO IMPACTO DAS VARIÁVEIS DA POLÍTICA DE INDEXAÇÃO E DAS MEDIDAS DE RECUPERAÇÃO DE INFORMAÇÃO EM UM SRI: EM BUSCA DA RELEVÂNCIA E PERTINÊNCIA

O planejamento de critérios de indexação é um procedimento importante em um SRI, que contribui para o desempenho de recuperação da informação. A consolidação desse planejamento deve ser materializada em um documento – A Política de Indexação – que apresenta os princípios e critérios que servirão para nortear as tomadas de decisões para alcançar o objetivo do SRI, ou seja, um tratamento consistente para uma recuperação de qualidade das

informações solicitadas. A política de indexação pode ser considerada como uma decisão administrativa, da gestão, tendo em vista a proposta de um SRI, em relação às características do sistema, tais como o tipo de informação armazenada, os tipos de buscas e o tipo de usuário.

Carneiro (1985, p. 221), em seu artigo Diretrizes para uma política de indexação, salienta que uma política de indexação

[...] deve servir como um guia para tomada de decisões; deve levar em conta os seguintes fatores: características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

Assim a política de indexação tem por objetivo principal planejar os procedimentos para alcançar a eficiência na recuperação da informação de um SRI, ou seja, ela se constitui um conjunto de princípios de análise do conteúdo do documento para sua representação em um registro.

Todos os usuários esperam que um SRI seja capaz de responder às suas buscas com um ou mais documentos pertinentes às suas necessidades. Para saber a eficiência de um SRI, Lancaster (1986, p. 131) aponta três critérios, que são essenciais para essa avaliação: (1) qualidade, (2) esforço, e (3) tempo de resposta do sistema. Em relação à qualidade, o autor sugere observar “a cobertura da base de dados, a habilidade do sistema em recuperar os documentos relevantes em resposta à pergunta do usuário, a habilidade de reter os documentos não relevantes ao mesmo tempo”. A avaliação da performance de um SR, em relação à sua eficiência, é medida a partir da satisfação do usuário, ou seja, da capacidade do sistema em responder às questões de sua comunidade.

Para isso, na representação da informação, devem-se observar os aspectos do nível de exaustividade e especificidade no processo de indexação, para conseguir a qualidade e as medidas de precisão e revocação do sistema na recuperação da informação, para atingir a relevância e pertinência no resultado da busca realizada pelos usuários. O impacto e a relação entre essas variáveis são retratados na Figura 7.

Figura 7. As variáveis que impactam a satisfação do usuário em um SRI.



Fonte: Elaborado pelas autoras.

Dias e Naves (2013, p. 22) definem exaustividade como “uma decisão prévia tomada pelo sistema, de reconhecer, além do assunto principal, todos os assuntos secundários contidos no documento que está sendo indexado”; já a especificidade “refere ao quanto se pode ser preciso ao se especificar o assunto de um documento que está sendo indexado”. Nesse sentido,

consideramos que a exaustividade na indexação está relacionada à decisão, por parte do indexador, quanto ao número de descritores atribuídos a um documento; e a especificidade tem relação com o grau de coextensividade entre o descritor atribuído ao documento e o conceito nele tratado.

Assim podemos definir ‘exaustividade’ como a medida da extensão em que os diferentes tópicos tratados em um documento são reconhecidos e representados na indexação. Quando isso ocorre, dizemos que o documento foi indexado exaustivamente. Dessa forma, ele pode ser recuperado por muitos termos de indexação ou por muitas combinações de termos, o que aumenta a possibilidade de sua recuperação. Um documento pode ser considerado sob múltiplos aspectos, entretanto apenas aqueles que vão ao encontro das finalidades do serviço devem ser representados. Por exemplo, num catálogo de uma biblioteca especializada em hotelaria, num documento sobre serviços de bebidas, não há necessidade de indexar os aspectos agrícolas das plantações e/ou regiões em que tais bebidas se originam, embora o documento forneça tais informações. Num acervo especializado em bebidas, ao contrário, todos os aspectos são úteis. O conceito oposto à exaustividade é a seletividade. Quando poucos termos de indexação são selecionados para representar o assunto de um documento, diminui, também, seu potencial de recuperação. Na política de indexação, deve ficar claro qual o critério adotado (GOMES; CAMPOS, 1998).

A especificidade, por sua vez, em síntese, é um princípio relacionado ao nível de representação de um dado num registro. Ser específico significa representar o conteúdo do dado no mesmo nível do documento. Se um documento trata das qualidades do tecido sintético, é nesse termo que ele deve ser indexado e não em ‘indústria têxtil’, por exemplo, que seria um termo amplo e que não representaria o assunto tratado no documento. O conceito oposto é generalidade. Esse termo deve ser entendido como termo genérico, ou seja, o termo mais geral dentro da categoria a que pertence o termo específico. Assim o termo genérico de ‘Tecido sintético’ é ‘Tecido’ (GOMES; CAMPOS, 1998).

Deve-se considerar sempre que uma alta exaustividade, ou seja, um grande número de descritores atribuídos a um documento, pode provocar relacionamentos inadequados e, conseqüentemente, recuperações falsas. Além disso, fatores de suma importância e que devem ser levados em consideração nessas decisões são: o perfil do usuário e o nível de especificidade do assunto tratado no documento.

Dois novos conceitos entram em cena e eles são decorrentes da exaustividade e da especificidade. Eles têm implicações na recuperação. São eles: Revocação e Precisão.

Quando um serviço decide adotar uma política de exaustividade, existe a possibilidade de recuperar mais itens em resposta a uma solicitação de busca, mas esses são recuperados em meio a outros que não satisfazem a busca. A extensão em que isso ocorre chama-se revocação.

Lancaster (2004, p. 4) define revocação como “a capacidade de recuperar documentos úteis”, e a precisão “a capacidade de evitar documentos inúteis”. Dito de outra forma, consideramos que a revocação expressa o quanto o sistema é capaz de recuperar todos os itens relevantes, ou seja, a relação entre a quantidade de registros não pertinentes recuperados e o total de registros recuperados em uma busca. Já a precisão expressa o quanto o sistema é capaz de recuperar apenas itens relevantes, ou seja, a relação entre a quantidade de registros pertinentes recuperados e o total de registros recuperados em uma busca.

Lancaster (1986, p. 133) propõe uma tabela de critérios para avaliar a eficiência do sistema no retorno da busca do usuário, na qual o autor sugere que (a) sejam os documentos relevantes, (b) documentos irrelevantes, (c) quais são relevantes e (d) quais não são, de acordo com a Tabela 1, a seguir.

Tabela 1. Proposta de julgamento da relevância na busca de informações pelo usuário

JULGAMENTO DE RELEVÂNCIA DO USUÁRIO				
PREVISÃO DE RELEVÂNCIA DO SISTEMA	Documento	Relevante	Irrelevante	TOTAL
	Recuperado	a (êxitos)	b (ruído)	a+b
	Não recuperado	c (perdidos)	d (rejeitados)	c+d
	TOTAL	a+c	b+d	a+b+c+d

Fonte: Adaptado de Lancaster (1986, p. 133).

A capacidade de revocação e a precisão do sistema, como aqui citado, estão relacionadas com a exaustividade e a especificidade. Essas medidas podem ser calculadas pelas expressões:

$$\text{Revocação} = \frac{\text{n}^{\circ} \text{ de docs relevantes recuperados}}{\text{n}^{\circ} \text{ total de docs relevantes no sistema}} \rightarrow R: \frac{a}{a+c} \times 100$$

$$\text{Precisão} = \frac{\text{n}^{\circ} \text{ de docs relevantes recuperados}}{\text{n}^{\circ} \text{ total de docs recuperados}} \rightarrow P: \frac{a}{a+b} \times 100$$

Sabe-se que a exaustividade resulta em uma alta revocação e, conseqüentemente, em uma baixa precisão. Essa relação é inversamente proporcional, uma alta especificidade resulta em uma alta precisão e baixa revocação.

Quando o usuário faz a busca num serviço de recuperação de informação, se ele seleciona poucos termos para representar sua pergunta, existe probabilidade de alta revocação; por exemplo, se o usuário faz a busca apenas pelo termo ‘Plantas ornamentais’ num serviço especializado sobre Cultivo de Plantas ornamentais, a revocação será muito elevada e o número de documentos irrelevantes também. No entanto, se ele incluir um ou mais aspectos de seu interesse, por exemplo, ‘Enxertia’, então a revocação será menor, mas a probabilidade de recuperar documentos relevantes aumenta. Portanto, quando a revocação é grande, aumenta a probabilidade de documentos indesejados, ou seja, ela introduz imprecisão. Assim Revocação e Precisão estão em relação inversa: quando maior a revocação, menor a precisão.

Mas a exaustividade não é a única responsável pela imprecisão. O indexador pode, também, introduzir imprecisão na indexação por meio de:

- a) omissão de um descritor ou de descritores importantes para a representação de um assunto. Por exemplo, o indexador deixou de indexar algum aspecto relevante; e,
- b) uso de um termo inadequado para representar uma ideia. Por exemplo, o indexador não conseguiu identificar corretamente o significado do termo atribuído a um documento. Nesse caso, o vocabulário contribuiu para a falha do indexador (GOMES; CAMPOS, 1998).

Os níveis de exaustividade e especificidade estão relacionados ao sistema; as medidas de revocação e precisão, ao processo de recuperação; além disso, existem as medidas de satisfação dos usuários - a relevância e a pertinência.

Segundo Lancaster (2004), os termos relevância e pertinência são empregados para referir a itens úteis, porém definidos de diferentes formas, demonstrando que existe uma controvérsia na literatura em relação à definição desses termos.

Cooper (1971, p. 19), apesar de considerar o conceito relevância inexplicável, salienta que “relevância é um dos mais fundamentais, senão o fundamental, conceito encontrado na

teoria da recuperação da informação [...] seja o que for, está no cerne do problema dos acessos intelectuais”. Também, Silva, Santos e Ferneda (2013) consideram que o conceito de relevância é subjetivo e inexato, não podendo ser definido por fórmulas matemáticas e implementadas em sistemas computacionais. Para os autores, a relevância consiste em “[...] mostrar os resultados possivelmente mais relevantes em forma de ranque (ranking), do mais relevante ao menos relevante”².

Uma das primeiras definições sobre relevância na literatura foi apresentada por Cuadra e Katter (1967, p. 51), na qual os autores destacam que “relevância é a correspondência no contexto entre um requisito de [informação] e um artigo, ou seja, até que ponto o material de cobertura do artigo é apropriado para a declaração de requisitos.”

Vickery e Vickery (2004, p. 265) apresentam uma definição mais abrangente de relevância, retratando-a como “uma medida da eficácia do contato entre uma fonte e um destino (destinatário) em um processo de comunicação”. Já Dias e Naves (2013) ressaltam que a relevância é o julgamento feito pelo usuário em relação ao resultado da busca em um SRI. Em suma, a relevância é o grau de similaridade entre termos que compõem as expressões de buscas dos usuários e a ocorrência em documentos da coleção ou nos termos de indexação, sendo considerada uma relação comparativa entre pergunta e documento.

Ingwersen e Järvelin (2005) definem relevância como uma avaliação da atualidade, pertinência, ou utilidade das fontes de informação, realizada por um (ns) ator(es) cognitivo(s) ou dispositivos algorítmicos, de acordo com a necessidade informacional em uma situação; sendo uma situação percebida como uma tarefa de trabalho, podendo ser um problema ou formulação de uma necessidade de informação, em um tempo específico. Esse processo é dinâmico, os resultados podem mudar com o tempo, mesmo que seja para o mesmo ator. A relevância pode ser de natureza objetiva de ordem inferior ou ordem superior, ou seja, de natureza multidimensional subjetiva, sendo sua medição binária ou graduada.

Enquanto a relevância está relacionada com a resposta do sistema à pergunta do usuário em um SRI, a pertinência é a relação desses documentos relevantes recuperados com a pergunta do usuário. Lancaster (2004) vê a pertinência como a utilidade da informação na recuperação de um item na biblioteca, que visa suprir a necessidade de informação do usuário. Similarmente, Kemp (1974, p. 37) resalta que a pertinência de um determinado documento para uma determinada necessidade é algo que só pode ser decidido pela pessoa com a necessidade daquela informação. Para Fosket (1972), pertinência significa agregar novas informações às já armazenadas na mente do usuário, que lhe são úteis às necessidades informacionais, as quais motivaram essa busca.

Saracevic (1975) demonstra, por meio de um diagrama, Figura 8, a diferença entre relevância e pertinência. Como se pode observar, a relevância se relaciona aos resultados da busca à pergunta realizada, enquanto a pertinência relaciona os resultados à necessidade de informação daquele usuário específico.

² Nesta definição, poderíamos substituir o termo relevante por apropriado.

Figura 8. Relevância versus pertinência



Fonte: Adaptado de Saracevic (1975, p. 331).

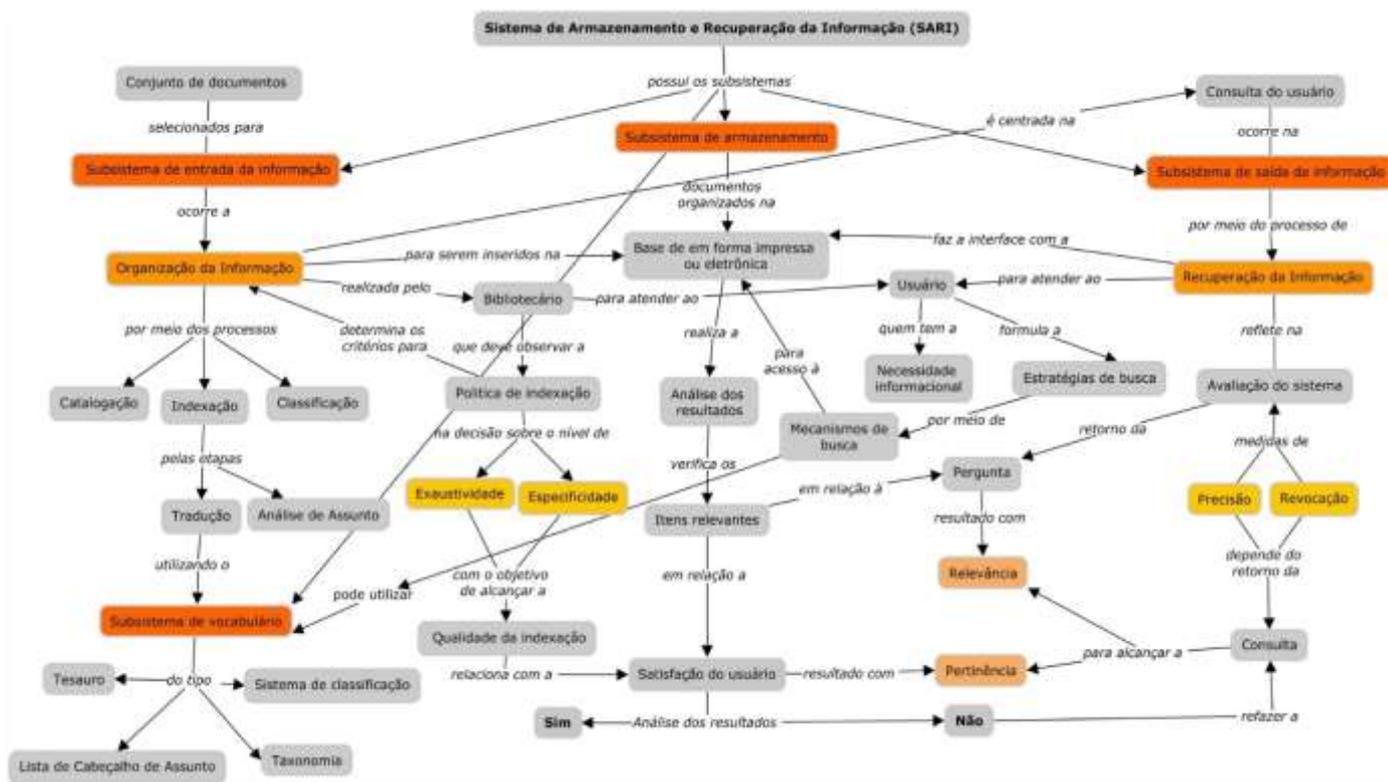
Ressalta-se que os SRI não podem avaliar a pertinência dos resultados, visto que somente os usuários são capazes de realizar essa avaliação. Por outro lado, fica a cargo do SRI responder às perguntas realizadas pelos usuários, calculando a relevância entre a pergunta do usuário e o retorno dado pelo sistema. Dessa forma, infere-se que a decisão sobre a pertinência dos resultados é feita pelo usuário; nesse caso, os documentos considerados impertinentes em uma pesquisa específica não podem ser considerados como falta do SRI.

Na próxima seção, apresentamos uma proposta de um modelo de SRI, o Modelo Sistema de Armazenamento e Recuperação da Informação, com a finalidade de incorporar todas as variáveis e medidas apresentadas acima.

6 PROPOSTA DE UM MODELO DE SARI A PARTIR DO MAPA CONCEITUAL CONSIDERANDO AS VARIÁVEIS E AS MEDIDAS

A partir da análise do impacto das variáveis da política de indexação e das medidas de recuperação de informação em um SRI apresentada na seção anterior, considerou-se contextualizar a posição de cada uma dessas variáveis e medidas dentro de uma proposta de um Sistema de Armazenamento e Recuperação da Informação. Apresenta-se, a seguir, a representação gráfica dos relacionamentos desses conceitos, por meio de um mapa conceitual, conforme Figura 9.

Figura 9. Proposta de um modelo de SARI a partir do mapa conceitual



Fonte: Elaborado pelas autoras.

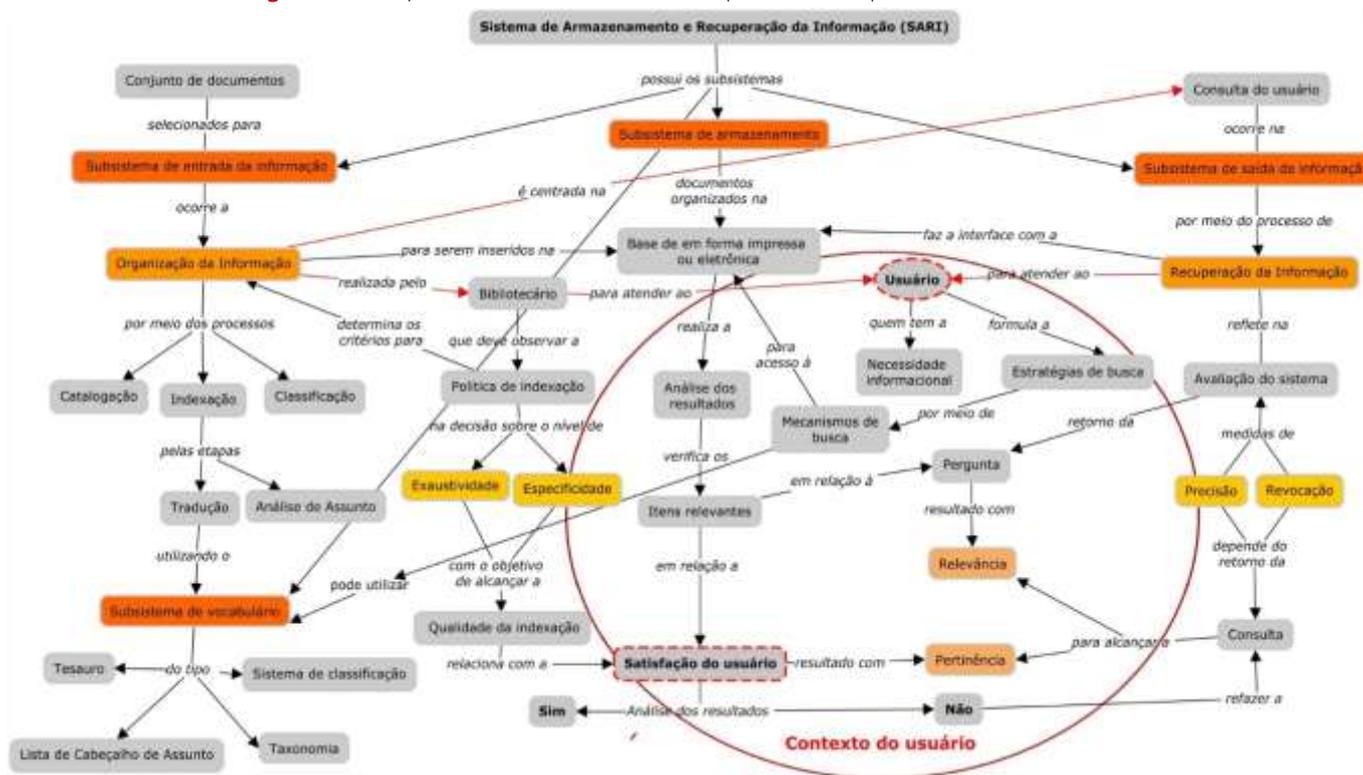
Nesse mapa, as proposições explicitam os relacionamentos entre os conceitos que compõem um SARI, posicionando a exaustividade e a especificidade no subsistema de entrada, a revocação e a precisão como medidas de avaliação; a relevância e a pertinência entre os subsistemas armazenamento e recuperação da informação, relacionando com as análises das solicitações realizadas pelos usuários. Nota-se que essas variáveis e medidas estão intrinsecamente relacionadas.

Existem vários modelos de fluxogramas de SRI que retratam as atividades e os processos que ocorrem no âmbito de um sistema de informação. Nesse caso, são chamados os modelos centrados no sistema. Esses sistemas exploram os relacionamentos entre as técnicas e os processos, mas excluem muitas variáveis que são relacionadas com a recuperação da informação, as quais são essenciais para que os SRI sejam eficientes. Entre elas, encontram-se o conhecimento da comunidade a ser atendida, a real necessidade do usuário e o contexto no qual esse SRI está inserido.

A estrutura cognitiva é diferente para cada indivíduo, pois cada pessoa tem seu modelo de mundo e processa a informação de forma variável. Assim, a estrutura do SARI, como é proposta neste estudo, leva em consideração o modelo de mundo dos seus usuários e como eles processam a informação, e como o sistema interage com o usuário no que tange às suas emoções, intuições e experiências que representam o modelo de mundo do usuário. Logo o profissional da informação que faz a intermediação entre o conteúdo informacional dos documentos e o usuário que necessita da informação são vistos como partes dos sistemas RI, não apenas os processos e os componentes técnicos.

Nessa perspectiva, marcou-se no mapa a seguir, Figura 10, o início do SARI centrado no Sistema, porém com uma interrelação com a parte da Recuperação da Informação, que se encontra em ambos. No entanto circulam-se os procedimentos, no contexto dos usuários, em relação à busca no sistema e o seu grau de satisfação, no que concerne ao resultado atingindo. Nesse caso, culminando na análise do grau de relevância ou pertinência em seus resultados.

Figura 10. Proposta de modelo de SARI a partir do mapa conceitual centrado no usuário



Fonte: Elaborado pelas autoras.

Apoiados na perspectiva de Lancaster (1978) como apresentado anteriormente (seção 3), infere-se, nesse caso, que os quatro subsistemas estão relacionados; não existe a possibilidade de um existir sem o outro, e cada um tem sua funcionalidade e seu papel dentro de um SARI. Percebe-se que a atuação do usuário se concentra entre a proposta do subsistema centrado no usuário e no subsistema de armazenamento, nos quais eles atuam como atores principais, assim como no subsistema de vocabulário³, pois, se a linguagem utilizada não estiver em consonância com o discurso do usuário, haverá um hiato na comunicação sistema/usuário. O profissional da informação tem um papel importante em relação à transferência da informação da entrada à saída, pois tem o compromisso de fazer chegar às ideias dos autores mais fidedignamente possível aos seus usuários.

Assim pressupõe-se que o tipo de sistema, suas características, como e quais componentes e estruturas que compõem um SARI devem estar contextualmente relacionadas entre as partes. Por isso a necessidade de decidir-se inicialmente qual tipo de variável quer se dar ao sistema: se o sistema vai adotar uma maior exaustividade em nível de representação ou se o vocabulário controlado adotado permite uma representatividade dos conceitos com maior especificidade, o que tem relação com o grau de coextensividade entre o descritor atribuído ao documento e o conceito nele tratado. Isso impactará as medidas de revocação e precisão e, conseqüentemente, afetará a avaliação do resultado das buscas dos usuários. Relevância, se o sistema atendeu à sua necessidade (a partir da análise de sua pergunta e o retorno do sistema); e Pertinência, a partir da análise desses resultados em relação à sua pergunta.

³ Com as tecnologias atuais, o subsistema vocabulário não possui somente o “documento vocabulário”, mas todo um ambiente computacional em que esses vocabulários estão inseridos, no qual é possível o estabelecimento de definições e relações conceituais, descoberta de conhecimento, entre outros, dependendo do tipo de SOC a ser utilizado.

7 CONSIDERAÇÕES FINAIS

O tratamento e a recuperação de informações são etapas fundamentais para o entendimento do complexo fazer informacional. Tais processos possuem, além de uma série de etapas, uma gama de aspectos que se inter-relacionam, como: a perspectiva do contexto informacional, o papel do usuário dentro desse contexto, o papel do profissional da informação e sua visão sistêmica, a questão do tipo de produção de conhecimento e sua materialização em documentos, o tipo de vocabulário controlado adotado, as questões tecnológicas agregadas e, conseqüentemente, a forma de tratar e recuperar as informações. Se a indexação é uma atividade que ocorre na etapa de entrada de dados, e a recuperação, na saída, numa visão sistêmica, a entrada afeta a saída. Por isso são de fundamental importância discussões que envolvem os fatores, medidas e variáveis que devem ser pensados em uma política de indexação.

Neste artigo, consideramos destacar essas questões mostrando como nos fluxos informacionais discutidos esses aspectos podem estar agregados em um todo consistente, consubstanciando-se no que se denomina de Sistema de Armazenamento e Recuperação de Informação (SARI). Essa visão sistêmica em que todos os elementos de um SARI estão em interrelação e na qual o usuário é o centro das atenções, em nossa concepção, é fundamental para a formação de profissionais aptos à construção de Sistemas que objetivem um tratamento com qualidade em suas representações e que implicará em uma recuperação da informação mais relevante e pertinente.

Assim, evidenciamos, como contribuição final deste estudo, a importância de uma visão sistêmica, na qual todos os elementos de um Sistema de Armazenamento e Recuperação de Informação (SARI) se encontram relacionados, tendo o usuário como elemento principal; bem como as atividades que são consideradas importantes para a formação de profissionais aptos à construção de Sistemas consistente.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. New York: Addison Wesley, 2011.

BARRETO, A. A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, jul./set. 2002.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, p. 221-241, set. 1985. Disponível em: <https://periodicos.ufmg.br/index.php/reb/article/view/36523/28575>. Acesso em: 26 nov. 2021.

CESARINO, M. A. N. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, 1985. Disponível em: <https://periodicos.ufmg.br/index.php/reb/article/view/36507/28553>. Acesso em: 24 nov. 2021.

CHU, H. **Information representation and retrieval in the digital age**. Melford: Information Today, Inc., 2005.

COOPER, W. S. A definition of relevance for information retrieval. **Information Storage and Retrieval**, v. 1, n. 7, p. 19-37, 1971.

CUADRA, C. A.; KATTER, R. V. **Experimental studies of relevance judgements**. Santa Monica: Systems Development Corporation, 1967. (NSF Rep. TM-3520/001, 002, 003, 3 volumes).

DAHLBERG, I. Knowledge organization: a new science? **Knowledge Organization**, v. 33, n. 1, p. 11-19, 2006. Disponível em: https://www.ergon-verlag.de/isko_ko/downloads/ko3320061c.pdf. Acesso em: 26 nov. 2021.

DIAS, E. W.; NAVES, M. L. **Análise de assunto: teoria e prática**. Brasília: Briquet de Lemos, 2013. 115 p.

FERNEDA, E. **Recuperação da Informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. 2003. 147 f. Tese (Doutorado em Ciência da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003.

FOSKET, D. J. A note on the concept of “relevance”. **Information Storage and Retrieval**, v. 2, n. 8, p. 77-78, 1972.

GOMES, H. E.; CAMPOS, M. L. A. **Política de indexação**. SESC, 1998. (Material didático apresentado no Curso de Capacitação na área de indexação.)

INGWERSEN, P.; JÄRVELIN, K. **The turn: Integration of information seeking and retrieval in context**. Dordrecht: Springer, 2005. 448 p.

INTERNATIONAL STANDARD ORGANIZATION. *ISO 5963-1985 – Documentation: methods for examining documents, determining their subjects, and selecting indexing terms*. Suíça: ISO, 1985.

KEMP, D. A. Relevance, pertinence and information systems development. **Information Storage and Retrieval**, v. 10, n. 2, p. 37-47, 1974.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 1993.

LANCASTER, F. W. **Information retrieval systems**. 2. ed. New York: Wiley, 1978.

LANCASTER, F. W. **Vocabulary control for information retrieval**. 2. ed. Arlington: Information Resources Press, 1986. 270 p.

LIMA, G. A. Gênese da classificação: uma análise de conteúdo a partir da definição. **Perspectivas em Ciência da Informação**, v. 26, n. 1, p. 197-237, mar. 2021. DOI: <https://doi.org/10.1590/1981-5344/32686>. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/4402/2463>. Acesso em: 26 nov. 2021.

MEY, E. S. A. **Introdução à catalogação**. Brasília: Briquet de Lemos, 1995.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 1, p. 20-32, 1951. DOI: <https://doi.org/10.1002/asi.5090020107>.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37-45, jul./dez. 1996. Disponível em: https://www.brapci.inf.br/_repositorio/2010/05/pdf_0e3cc20139_0010458.pdf. Acesso em: 26 nov. 2021.

ROWLEY, J. **A biblioteca eletrônica**. 2. ed. Brasília: Briquet de Lemos, 2002. 399 p.

ROWLEY, J. **Informática para bibliotecas**. Brasília: Briquet de Lemos, 1994.

SALTON, G. **Automatic information organization and retrieval**. New York: McGraw-Hill, 1968.

SALTON, G.; MCGILL, J. M. **Introduction to modern information retrieval**. New York: McGraw-Hill, 1983.

SARACEVIC, T. Information science. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051-1063, 1999. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:12<1051::AID-ASI2>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1051::AID-ASI2>3.0.CO;2-Z).

SARACEVIC, T. Relevance: A review of and a framework for the thinking on the notion in information science. **Journal of the American Society for Information Science**, v. 26, n. 6, p. 321-343, 1975. DOI: <https://doi.org/10.1002/asi.4630260604>.

SILVA, R. E.; SANTOS, P. L. V. A. C.; FERNEDA, E. Modelos de recuperação de informação e web semântica: a questão da relevância. **Informação & Informação**, v. 18, n. 3, p. 27-44, 2013. DOI: <http://dx.doi.org/10.5433/1981-8920.2013v18n3p27>. Disponível em: https://www.uel.br/revistas/uel/index.php/informacao/article/view/12822/pdf_3. Acesso em: 26 nov. 2021.

SOERGEL, D. **Organizing Information: principles of data base and retrieval systems**. California: Academic Press, 1985.

TRISTÃO, A. M. D. *et al.* Sistema de classificação facetada: instrumento para organização da informação sobre cerâmica para revestimento. **Informação e Sociedade: Estudos**, v. 14, n. 2, 2004. Disponível em: <https://www.proquest.com/docview/1494045851>. Acesso em: 26 nov. 2021.

VICKERY, B. C.; VICKERY, A. **Information science in theory and practice**. 3. ed. rev. aum. Munique: KG Saur, 2004. 400 p.