



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

REQUISITOS DE RECUPERAÇÃO DA INFORMAÇÃO: IDENTIFICAÇÃO E CLASSIFICAÇÃO COM BASE NOS CONCEITOS DE PRECISÃO E REVOCAÇÃO¹

Mariana Baptista Brandt - Câmara dos Deputados

Silvana Aparecida Borsetti Gregorio Vidotti - Universidade Estadual Paulista (Unesp)

Modalidade: Trabalho Completo

Resumo:

Considerando a importância de uma boa definição de requisitos para mecanismos de busca para que haja recuperação efetiva da informação, objetiva-se identificar, na literatura da Ciência da Informação (CI) prioritariamente e da Computação, se forma secundária, possíveis requisitos para recuperação da informação (RI) e relacioná-los às medidas de precisão e revocação. A necessidade de identificação de requisitos de RI surgiu em pesquisa de doutorado com o objetivo de compor uma das etapas de uma metodologia de arquitetura da informação. Procedeu-se à pesquisa bibliográfica sobre RI e mecanismos de busca para identificação e extração desses requisitos, além de revisão narrativa, conforme bibliografia estudada no decorrer da pesquisa. Como resultado, observa-se que a maior parte dos itens identificados pode ser classificado como dispositivos de precisão ou revocação, o que permite validá-los como requisitos de RI e consolidá-los como um instrumento para identificação de requisitos de busca em sistemas de informação bem como para avaliação de bases de dados quanto à ferramenta de busca.

Palavras-Chave: Recuperação da Informação; Mecanismos de busca; Requisitos para Sistemas de Informação; Arquitetura da Informação.

Abstract:

Considering the importance of defining good requirements for search engines so that there is effective information retrieval, this paper aims to identify, in Library and Information Science (LIS) literature, primarily, and Computer Science, secondly, possible requirements for information retrieval (IR) and relate them to measurements of accuracy and recall. That need came from a doctoral research so as to compose an information architecture methodology that was developed in the research. For this purpose, a bibliographic research on information retrieval and search engines was performed to identify and extract these requirements. Also, a narrative review of literature was conducted considering the bibliography studied during the research. As a result, it was observed that most of the items identified can be classified as precision or recall devices, which allows them to be validated as information retrieval requirements and consolidated as an instrument for the definition of search requirements in information systems as well as for evaluating databases regarding the search tool.

¹ Artigo derivado de tese (BRANDT, 2020)

Keywords: Information Retrieval; Search engine; Requirements for Information Systems; Information Architecture.

1 INTRODUÇÃO

Os profissionais da informação atuam nos ambientes informacionais mais diversos e exercem atribuições que vão além do escopo dos ambientes de bibliotecas, arquivos e museus. A necessidade de atividades voltadas para os ambientes informacionais digitais é cada vez maior, e, nesse contexto, destacamos a definição, modelagem e avaliação de bases de dados e outros repositórios de informações e o levantamento de requisitos para sistemas de informação e ambientes informacionais digitais. Um dos aspectos centrais destes ambientes é a recuperação da informação, em que vários de seus requisitos deverão orientar a forma de organização e armazenamento de dados, metadados, informações e documentos nestes repositórios. Para isso, é importante que haja uma definição de requisitos de recuperação da informação e acredita-se que um artefato que liste os possíveis requisitos seja de grande utilidade para esta tarefa.

A necessidade de uma lista de requisitos de RI surgiu durante a elaboração de uma metodologia de arquitetura da informação (AI) desenvolvida na tese “Modelagem da informação legislativa: arquitetura da informação para o processo legislativo brasileiro” (BRANDT, 2020). Assim, um estudo para identificação de possíveis requisitos de RI para mecanismos de busca foi realizado nesta pesquisa. Além de sua utilidade como etapa da metodologia de AI citada, acredita-se que a publicação dos requisitos encontrados por meio de artigo encontre um público mais abrangente, contribuindo para a disseminação desse resultado específico da pesquisa, o qual poderá ser utilizado também em outros contextos.

Assim, este artigo traz o estudo realizado durante a tese para elaboração de um instrumento que reúne, de forma abrangente, os possíveis requisitos de recuperação da informação e suas respectivas definições. Além disso, os requisitos encontrados foram relacionados aos dispositivos de precisão e revocação, essenciais para a RI. Acredita-se que tal instrumento seja de utilidade para profissionais que atuam nesta área bem como para pesquisadores, que poderão utilizá-lo como ponto de partida para novos estudos e complementá-lo de acordo com suas necessidades acadêmicas. Para isso, buscou-se na literatura da Ciência da Informação (CI), prioritariamente, e da Computação, de forma subsidiária, os requisitos de recuperação da informação elencados pelos autores. Os

requisitos foram analisados e classificados quanto aos dispositivos básicos de RI (precisão e revocação) e, então, foi elaborada uma proposta de sistematização.

Trata-se de uma pesquisa qualitativa que utilizou como metodologia a pesquisa bibliográfica, recorrendo-se à literatura clássica de RI para identificar suas origens e seu desenvolvimento, além de buscar possíveis requisitos. Após contextualizar o histórico da RI, foram buscados os termos “recuperação da informação” e “requisitos”, nos artigos de CI, utilizando-se a base de dados Brapci - Base de dados em Ciência da Informação. As buscas foram realizadas em 2017, de forma ampla (busca em todos os campos) e não foram encontrados registros que atendessem aos critérios. Assim, foram utilizados documentos estudados no decorrer da pesquisa de doutorado que tratavam deste assunto, configurando uma revisão narrativa da literatura. Os itens elencados como requisitos encontrados foram analisados, categorizados e compilados, resultando em uma tabela com as possibilidades de implementações para RI. Esses requisitos foram avaliados com base nos conceitos de precisão e revocação, essenciais na RI e os quais contribuem para a validação da CI como ciência.

O artigo apresenta na seção 2 um breve histórico da Recuperação da informação e os principais conceitos envolvidos, em seguida traz os requisitos de Recuperação da Informação encontrados na pesquisa, na seção 3, e, posteriormente, os resultados são apresentados na seção 4, com uma tabela de compilação dos itens e discussões a respeito dos requisitos encontrados. A seção 5 finaliza com as considerações finais sobre a pesquisa e sua utilidade para a CI.

2 RECUPERAÇÃO DA INFORMAÇÃO

A Recuperação da Informação é uma disciplina nuclear da Ciência da Informação, consumindo grande parte dos esforços e recursos na área, além de ser considerada uma das maiores responsáveis pela interdisciplinaridade da CI (SARACEVIC, 1995). Considera-se a Recuperação da Informação como um dos grandes objetivos da CI, já que os principais processos e técnicas da disciplina são realizados com este propósito. Como exemplo, temos os processos de organização da informação e do conhecimento, que abarcam a elaboração de vocabulários e classificações, a catalogação bibliográfica, a indexação de assuntos e todos os demais processos técnicos operados nos ambientes de informação.

A Arquitetura da Informação é também uma área intimamente relacionada com a Recuperação da Informação. Seus sistemas e elementos básicos como sistemas de organização, taxonomias e metadados são modelados com o objetivo de facilitar a recuperação de informações do ambiente informacional. Além disso, o sistema de busca da AI é, essencialmente, um sistema de recuperação da informação:

A recuperação da informação está fortemente presente nos estudos da ciência da informação, bem como na arquitetura da informação. Os sistemas de recuperação da informação operam por sentenças lógicas analisadas dos documentos (entenda-se aqui como documento todo e qualquer suporte com informações registradas), formando um conjunto de características que indicam o conteúdo dos documentos. Este conjunto de propriedades é ordenado de acordo com a pertinência das necessidades dos usuários. (ROBREDO et al, 2008)

O Dicionário de Biblioteconomia e Arquivologia de Cunha e Cavalcanti (2008, p. 307) traz as seguintes definições para a expressão Recuperação da informação:

1. Restituição de dados constantes do sistema para obtenção de informações específicas ou genéricas. A restituição, ou recuperação, abrange o processo total de identificação, busca, encontro e extração da informação armazenada. Nesta operação, não se incluem, nem a criação, nem a utilização posterior das informações ou dos dados.
2. A recuperação de dados, informações ou documentos de uma coleção ou acervo a partir de um pedido formulado.
3. Ação, métodos e procedimentos que tem por objetivo extrair de um conjunto de documentos informações desejadas. (AFNOR)
4. Área que estuda as técnicas de tratamento, armazenamento e recuperação da informação num arquivo ou acervo de documentos, com o objetivo de atender às necessidades de informação do usuário.

A Recuperação da Informação não deve ser vista com um fim em si mesma, mas como um meio de levar a informação até o usuário. Ou seja, o grande objetivo final da recuperação está na disseminação da informação e na satisfação das necessidades de informação do usuário. O acesso à informação se dá, em grande parte das vezes, a partir do resultado de uma busca em um sistema de recuperação da informação.

O termo Recuperação da Informação foi cunhado por Calvin Mooers, em 1950: "O problema de direcionar o usuário a informações armazenadas, algumas das quais podem ser desconhecidas para ele é o problema da 'recuperação da informação'" (MOOERS, 1950, p. 572, tradução nossa). Mooers (1950, 1951), em uma analogia com a teoria matemática da comunicação, afirma que o usuário (destinatário) ao invés do emissor (remetente) é

participante ativo na recuperação da informação e que o objetivo é encontrar, no menor tempo possível, as mensagens (informações) de interesse do usuário.

Os primeiros estudos de Recuperação da Informação como disciplina começaram nos anos 1960. Vickery (1965) afirma que, em 1961, a RI estava começando a se firmar como uma disciplina unificada e em 1965 é reconhecida como disciplina. A disciplina trouxe consigo os conceitos de relevância, precisão e revocação. A relevância é a avaliação do usuário em relação a quantidade de documentos recuperados que atendem suas necessidades de informação, conforme explica Lancaster (2004, p. 14) “a relação entre um documento e uma necessidade de informação ou entre um documento e um enunciado de necessidade de informação (uma consulta)”.

Na perspectiva da Ciência da Computação, a Recuperação da Informação é classificada como uma área e há dois pontos de vista, um com foco no computador e outro com foco no usuário:

Na visão centrada no computador, a RI consiste principalmente na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento, a fim de melhorar os resultados. Na visão centrada no usuário, a RI consiste principalmente em estudar o comportamento do usuário, entender suas principais necessidades e determinar como esse entendimento afeta a organização e a operação do sistema de recuperação. (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 1).

Para Baeza-Yates e Ribeiro Neto (2013), a área de RI cresceu muito comparada aos objetivos iniciais de indexação de documentos de uma coleção para busca e atualmente inclui modelagem, classificação de textos, arquitetura de sistemas, interface de usuário, visualização de dados, filtragem e linguagem, deixando clara a necessidade de relação entre as áreas de informação e de tecnologia.

A área de Recuperação da Informação foi bastante impactada com o surgimento da *World Wide Web* no início dos anos 1990. O desafio deixou de ser a recuperação de documentos e informações em coleções específicas e controladas como bibliotecas e bases de dados e passou a ser em um ambiente digital que pode ser considerado "infinito", já que consiste em um repositório universal, amplo e acessível pelo mundo todo, tanto para inclusão quanto para busca de informações. Além disso, a existência dos hiperlinks, conectando informações neste grande repositório, mudou completamente a forma como os usuários recuperam e, principalmente, selecionam informações.

Surgem então os chamados motores ou mecanismos de busca na web. "A busca na web é a aplicação mais proeminente da RI e suas técnicas [...] a web teve um grande impacto no desenvolvimento da RI" (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 12). Um dos impactos está relacionado ao armazenamento dos documentos, que é descentralizado, demandando que o mecanismo de busca faça uma coleta prévia para que eles sejam indexados e possam ser recuperados. Além disso, as informações são inseridas na web a todo momento, seu volume aumenta exponencialmente, assim como a quantidade de usuários executando buscas. Isso requer alto desempenho no processamento da busca, em especial no atual contexto do big data, o que trouxe avanços nas tecnologias desses motores. As grandes empresas especializadas em busca na web Yahoo e Google "[...] têm melhorado e contribuído com o desenvolvimento do processo de recuperação, pesquisando e aplicando novos métodos e conceitos que tendem a facilitar, agilizar e tornar eficaz o processo de recuperação da informação na Web". (SANTAREM SEGUNDO, 2010, p. 40).

2.1 Precisão e revocação

A literatura clássica sobre RI aponta os critérios de precisão e revocação como medidas de avaliação da recuperação da informação, e, conseqüentemente, do sistema de recuperação da informação (VICKERY, 1965; SHARP, 1965; LANCASTER, 1968, 2004; BAEZA-YATES; RIBEIRO-NETO, 2013). Para Lancaster (1968), precisão e revocação são as medidas de performance mais importantes para um sistema de informação.

Precisão é a "qualidade informativa dos documentos recuperados por um sistema, para atendimento de pedidos relativos a temas específicos" (CUNHA; CAVALCANTI, 2008, p. 289), e representa uma medida que indica, entre os documentos recuperados em uma busca, a quantidade de documentos relevantes. A revocação indica o número de documentos relevantes recuperados dentre todos os documentos relevantes da coleção. Conforme definido por Lancaster (2004), a revocação é a capacidade de recuperar documentos úteis, enquanto a precisão é a capacidade de evitar documentos inúteis.

A precisão está relacionada, portanto, a buscas mais específicas, que utilizam mais critérios ou argumentos, e, assim, recuperam menos itens não relacionados à necessidade de informação do usuário. Já a revocação deve aumentar o número de itens recuperados, e, para isso, a expressão de busca traz menos argumentos e critérios, ou seja, é mais genérica. As medidas são inversamente proporcionais: quanto maior a precisão, menor tende a ser a

revocação, e vice-versa. Dessa forma, ambas devem ser consideradas importantes na construção de um motor de busca: o sistema deve permitir a construção de argumentos de busca que priorizem a precisão ou a revocação, a depender da necessidade de informação.

[...] o pesquisador é capaz de variar a performance de precisão e revocação para uma busca particular por meio da adoção de estratégias de busca. [...] Dada a habilidade de variar a expressão de busca (com o objetivo de recuperar mais documentos ou menos documentos conforme a situação demandar), com a subordinação e superordenação das categorias, com o uso de sinônimos ou com outra técnica, somos capazes de executar buscas com diferentes níveis de generalidade. [...] Então podemos decidir entre sacrificar a precisão para obter uma alta revocação, ou sacrificar a revocação para obter uma busca mais precisa, ou podemos trabalhar num meio termo. (LANCASTER, 1968, tradução nossa, p. 71)

Esses conceitos permitiram, por meio de fórmulas matemáticas, realizar experiências repetíveis e quantificar os objetivos de recuperação de informação. Svenonius (2000) explica que essa quantificação dos objetivos de RI em termos de variáveis de precisão e revocação torna possível o estabelecimento de proposições sobre o impacto de vários fatores (especificidade da indexação, profundidade da indexação e tamanho do vocabulário) para a efetividade da recuperação. Essas proposições que expressam relações entre as variáveis são científicas, no sentido em que representam generalizações de alto nível sobre o objeto de estudo, o que contribui com o status de ciência da Ciência da Informação.

3. REQUISITOS PARA MECANISMOS DE BUSCA

Os mecanismos de busca (*search engine*) realizam a recuperação da informação pontual nos sistemas de informação e demais ambientes informacionais digitais e devem ser elaborados com base nos requisitos técnicos necessários. Os requisitos devem variar conforme o ambiente informacional, já que alguns terão necessidades mais complexas, como busca booleana e integração com vocabulários controlados, enquanto outros serão mais simples e uma busca nos metadados de assunto ou até mesmo no texto completo já seria o suficiente.

Segundo Vickery (1965), quando surgiam os primeiros sistemas de recuperação da informação, não havia resposta clara sobre as características que deveriam ser adotadas em um sistema de recuperação da informação (SRI), e que a resposta deveria ser baseada na compreensão do ambiente informacional como um todo: uso, organização e equipamentos

de cada tipo de sistema, avaliação de seus propósitos, parâmetros e performance, ou seja, uma análise completa dos sistemas de recuperação. A afirmação do autor continua válida nos dias atuais.

Um dos achados da pesquisa foi o artigo de Cianconi (1994), que identifica requisitos desejáveis para escolha de um SRI com bases textuais e imagens. Para a autora, devem ser levados em conta critérios:

- **relativos ao sistema:** compatibilidade, integração e interoperabilidade, facilidade de uso, customização etc;
- **relativos à informação:** natureza da informação, necessidade de integração de informações estruturadas e não estruturadas, uso da informação, atualização e crescimento da base de dados etc;
- **relativos ao uso:** objetivo de arquivamento, de busca em dados textuais, de recuperação por palavras-chave, etc.

Cianconi (1994) especifica que, quanto às necessidades de busca, devem ser considerados:

- **Busca por comandos:** Permite o uso de operadores booleanos, operadores de proximidade no texto com indicação de intervalo de palavras, operadores de comparação etc.
- **Busca orientada por menus:** Através de opções do menu. Convém observar que o menu é útil para o iniciante, porém à medida que o usuário se torna experiente, acha cansativo ter que ir e vir em telas de menu apenas. Combinar opções de menu e comandos tem sido uma estratégia adotada com sucesso.
- **Busca baseada em hipertexto:** Os sistemas de hipertexto podem ser implementados de diferentes modos. É preciso observar se corresponde às expectativas. Por exemplo: se é possível voltar ao ponto de partida, refazendo o caminho inverso; se permitem associar textos, imagens, arquivos e programas entre si; se existe, além da navegação, a possibilidade de busca por comandos. (CIANCONI, 1994, p. 251)

Apesar do trabalho supracitado datar de um período pré-web, a autora traz vários conceitos relacionados à AI para web e à encontrabilidade da informação. O que Cianconi (1994) denomina como “busca orientada por menus” e “busca baseada em hipertexto” atualmente são conhecidas por “navegação”. A autora considera ainda critérios mais específicos para a linguagem de busca, apresentados no quadro 1 a seguir:

Quadro 1 - Recursos de recuperação da linguagem de busca

Recurso	Descrição
1. Lógica Booleana	Capacidade de efetuar a busca utilizando os operadores lógicos E, OU, NÃO, de modo a precisar melhor o que se deseja recuperar
2. Busca no contexto	Capacidade de especificar o contexto no qual as palavras aparecem, identificando termos próximos no registro. Ou seja, indicar a vizinhança ou intervalo entre termos no campo, na frase, no parágrafo, no documento etc., recuperando palavras precedidas ou seguidas de outras.
3. Por comparação	Capacidade de permitir comparações entre conjuntos, geralmente utilizando operadores como > < = (aplicável principalmente a datas, campos numéricos).
4. Truncamento ou mascaramento	Aumento da capacidade de recuperação através do mascaramento de caracteres no início, meio ou final de uma palavra (geralmente com símbolos como #, *, ?).
5. Busca por listas de sinônimos e/ou tesouro	Facilidade de utilizar vocabulários controlados para auxiliar a consulta.
6. Capacidade de recuperar informações em diferentes formatos de arquivo	Capacidade de acesso a informações em bases de dados textuais (não estruturadas) e em bases de dados estruturadas — relacionais ou não —, em planilhas etc.
7. Capacidade de salvar buscas	Permite guardar as buscas efetuadas para reutilização posterior.
8. Sort	Possibilidade de reorganizar as saídas na tela ou em relatórios impressos.

Fonte: Elaborado pelas autoras com base em Cianconi (1994, p. 252)²

Em uma obra mais recente, de 2010, Morville e Callender afirmam que a busca deve ter como foco a experiência do usuário. Os autores propõem um *checklist* de itens que devem ser verificados na implementação de um sistema de busca, que inclui: arquitetura do sistema, performance, formatos de arquivos, integração, controle de acesso, funcionalidades, implementação, modelo de negócio (preço) e credenciais do vendedor (MORVILLE; CALLENDER, 2010). Dentre esses itens, considerou-se que devem fazer parte dos requisitos de RI os seguintes:

² Extraído da tese (BRANDT, 2020)

- Performance: identificar quantas buscas simultâneas são possíveis, qual o número máximo de fontes e qual o tamanho máximo do repositório de dados;
- Formatos de arquivo: verificar os tipos de conteúdos e formatos de dados (HTML, PDF etc) possíveis, se o sistema lida com dados estruturados e não-estruturados;
- Integração: verificar se segue padrões para interoperabilidade;
- Funcionalidades: Analisar se o sistema suporta texto completo e metadados, operadores booleanos, *wildcards*, *stemming*, *stop words*, busca por proximidade e auto-correção. Verificar qual algoritmo é utilizado para ranqueamento, se há opções para refinamento da busca e se os resultados podem ser salvos, impressos e compartilhados.

Os demais itens do *checklist* dizem respeito a questões essencialmente tecnológicas (nível técnico) e relacionadas a modelos de negócio para aquisição ou desenvolvimento de ferramentas. Morville e Callender (2010) apresentaram também itens a serem avaliados nas ferramentas de busca:

- Velocidade de resposta do motor de busca: deve ser rápido;
- Ajuste de relevância: considerar critérios de ordenação do resultado da busca, como popularidade, conteúdo, tipo, data e diversidade;
- Navegação e filtros: considerar se é possível customizar facilmente a ordenação e filtrar resultados, verificar se há navegação facetada e se é rápido;
- Busca federada: verificar como o sistema lida com busca em múltiplas bases ou índices e o impacto na velocidade;
- Ferramentas linguísticas: verificar se é possível integração com tesouros e cruzamento de vocabulários e se há auto categorização e extração de entidades;
- Análise das buscas: checar se há ferramentas para mensuração e análise de comportamento do usuário e API para compartilhamento e reuso desses dados.

Rosenfeld, Morville e Arango (2015) elencam, no âmbito do sistema de busca da arquitetura da informação para web e demais dispositivos, itens que consideram como blocos construtores da questão de busca. Segundo os autores, além dos algoritmos de busca, há vários outras coisas que afetam os resultados de uma busca e citam como exemplos:

- Corretor ortográfico;
- Ferramentas fonética;
- Ferramentas de derivação ou truncamento (*stemming*);

- Processamento de linguagem natural;
- Vocabulários controlados e tesouros.

4 RESULTADOS E DISCUSSÃO

Cada um dos itens das listas propostas por Morville e Callender (2010) para *checklist* e avaliação e dos itens identificados por Rosenfeld, Morville e Arango (2015) pode ser traduzido em um ou mais requisitos que devem ser implementados em um SRI. Assim, definem-se possíveis requisitos para recuperação da informação, desmembrados de cada um dos itens, com base na proposta de Cianconi (1994) – quadro 1. Tais requisitos, relacionados ao item de avaliação ou *checklist* correspondente, são apresentados no quadro 2, a seguir. Com isso, os requisitos foram classificados como dispositivos de precisão e revocação.

Quadro 2 - Requisitos a serem considerados em um SRI

Item	Requisito(s)	Dispositivo
Performance / Velocidade de resposta	O mecanismo de busca deve retornar resultados em até xx segundos.	N/A
Formatos de arquivo	O sistema deve armazenar e executar busca em arquivos nos formatos xxxx (verificar os formatos necessários) e deve executar busca em informações estruturadas (metadados) e não-estruturadas.	revocação
Integração	O sistema deve permitir integração com instrumentos de controle terminológico e com outros sistemas.	ambos
Funcionalidades	Deve permitir pesquisa textual unificada (busca livre) - campo de pesquisa único, com opção de escolha dos locais de consulta, podendo ser selecionada mais de uma opção ou não selecionar (busca em todos os campos).	ambos
	Deve possuir lógica booleana - uso dos conectores AND, OR, NOT (E, OU e NÃO) indicando para o sistema de busca como deve ser feita a combinação entre os termos ou expressões de uma pesquisa.	AND, NOT: precisão OR: revocação
	Deve permitir pesquisa por proximidade - uso do conector "NEAR" para buscar palavras aproximadas em qualquer ordem. O operador será sempre utilizado com um número, que indica a delimitação de distância entre as palavras, podendo ser escolhido livremente.	precisão

	Permitir truncamento com uso de caracteres “wildcards” (*) em substituição a um ou mais caracteres no argumento de busca, de modo a permitir que se considere na busca um termo e também suas variações singular/plural, masculino/feminino, prefixos/sufixos e grafias próximas	revocação
	Permitir pesquisa por intervalo de parâmetros: o sistema deve aceitar como argumento de pesquisa um intervalo de valores. Ex: ano 2000 a 2005.	precisão
	Deve possuir busca de termo exato com uso de aspas: permite buscar por termos compostos, pequenas frases ou termo exato.	precisão
	Combinação das aspas e do truncamento no argumento de pesquisa para pesquisar variações de um termo composto, como “polici* milit*”.	precisão
	Permitir que o sistema faça buscas por termos similares foneticamente (busca fonética), de modo a se recuperar termos com grafias próximas, singular/plural, com acento e sem acento, com cedilha e sem cedilha etc.	revocação
	Permitir adoção de lista de “stop words”, que são artigos, preposições e outras palavras que devem ser desconsideradas na busca ao serem digitadas pelo usuário.	precisão
	Permitir que a tela de resultados apresente opção para execução de nova pesquisa sem precisar sair da tela.	N/A
	Possibilitar a seleção de itens do resultado para inclusão em “cesta”. Incluir opção para selecionar todos ou desfazer toda a seleção.	N/A
	Possibilitar que os resultados das buscas e seleções de resultados sejam salvos, impressos, compartilhados e exportados.	N/A
	O resultado da pesquisa deve informar quais foram os critérios utilizados para se chegar àquele resultado e a data em que foi feita a pesquisa.	N/A
Ajuste de relevância	Permitir ordenação dos resultados de busca de acordo com o critério definido pelo usuário, como: por popularidade por maior ocorrência do termo pesquisado (relevância) por data	precisão
Navegação e filtros	Incluir filtros definidos pelo usuários do sistema (área de negócio)	precisão
	Permitir navegação entre as páginas de resultados, possibilitando ao usuário navegar de uma página para a	N/A

	seguinte, para a anterior, ou saltar diretamente para uma página especificada	
	Permitir refinamento dos resultados da busca	precisão
Busca federada	O motor de busca deve ser capaz de realizar pesquisa em vários bancos de dado ao mesmo tempo, sem prejudicar o tempo de resposta	revocação
Ferramentas linguísticas	Incluir o recurso autocompletar e / ou autossugestão	revocação
	Deve haver integração com o tesouro ou vocabulário da instituição ou processo de trabalho	ambos
	Sugerir valores similares de parâmetros quando o resultado de uma pesquisa for vazio ou mínimo. Carregar a ferramenta com listas de termos similares	revocação
	Processamento de linguagem natural	precisão
Análise das buscas	A ferramenta deve guardar log de buscas: termos mais buscados, termos com busca negativa, quantidade de buscas por período etc.	N/A

Fonte: elaborado pelos autores³

Verificou-se que 19 dos 26 requisitos correspondem a critérios que contribuem para o aumento da precisão ou da revocação, dependendo de como o requisito for utilizado em cada busca, corroborando a ideia apresentada anteriormente pela literatura de RI, bem como mostrando pertinente a proposta dos autores pesquisados (CIANCONI, 1994; MORVILLE; CALLENDER, 2010; ROSENFELD; MORVILLE; ARANGO, 2015).

Assim, considera-se que o quadro 2 pode ser utilizado como um guia para definição de requisitos de recuperação da informação para sistemas de informação dos variados ambientes informacionais digitais.

4 CONSIDERAÇÕES FINAIS

A definição dos requisitos necessários para recuperação da informação de um SRI é uma etapa crucial para o sucesso de um sistema de informação. Nos ambientes digitais atuais, poucos são os que não necessitam de um mecanismo de busca: até mesmo para encontrar um aplicativo no celular, muitas vezes é necessário utilizar o buscador. A quantidade de informações é cada vez maior, tornando mais difícil sua organização

³ Extraído da tese, com adaptações (BRANDT, 2020)

unicamente por taxonomias navegacionais, e a premissa para que os usuários utilizem e tenham uma boa experiência de uso em um ambiente informacional é que a informação necessária seja encontrada. Com isso, a inclusão do mecanismo de busca é recomendada para esses ambientes, consistindo em um dos sistemas da Arquitetura da Informação: o sistema de busca.

Este trabalho realizou a identificação e a extração de requisitos para ferramentas de busca, compilando-os e sistematizando-os em uma tabela que visa apoiar o trabalho de profissionais da informação, bibliotecários e demais interessados em realizar um levantamento de requisitos para ferramentas de busca para um sistema ou ambiente de informação digital. O compilado de requisitos, classificados em sua maioria como itens que atuam nas dimensões da precisão ou da revocação, pode auxiliar também para avaliação de motores de busca das bases de dados, na etapa de contratação de assinaturas em bibliotecas e demais centros de informação.

Assim, além da utilização prática de seus resultados, espera-se que essa pesquisa possa contribuir com a Ciência da Informação e que possa auxiliar pesquisadores, como uma fonte de informação na temática da recuperação da informação.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação**: Conceitos e Tecnologia das Máquinas de Busca. 2. ed. São Paulo: Bookman, 2013.

BRANDT, M. B. **Modelagem da informação legislativa**: arquitetura da informação para o processo legislativo brasileiro. 2020. 266 p. Tese. (Doutorado em Ciência da Informação) — Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, SP. Disponível em: <https://repositorio.unesp.br/handle/11449/191740>

CIANCONI, R. B. Requisitos mínimos para gerenciamento e recuperação de textos e imagens. **Ci. Inf.** Brasília, v. 23, n. 2, p. 249-253, maio/ago 1994.

CUNHA, M. B.; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008. 451p.

LANCASTER, F. W. **Information retrieval systems**: characteristics, testing, and evaluation. New York: J. Willey & Sons, 1968. 222 p.

LANCASTER, F. W. **Avaliação de serviços de biblioteca**. Brasília: Briquet de Lemos, 2004.

MOOERS, C. Information retrieval viewed as temporal signaling. *In: Proceedings of the International Congress of Mathematicians*. v. 1, p. 572-573, 1950.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, v. 2, n. 2, p. 20-32. 1951.

MORVILLE, P., CALLENDER, J. **Search patterns**. Sebastopol, CA, O'Reilly Media, 2010.

ROBREDO, J. et al. Reflexões sobre Fundamentos da Arquitetura da Informação. IX ENANCIB: Diversidade Cultural e Políticas de Informação, *Anais...* USP, São Paulo. 2008.

ROSENFELD, L.; MORVILLE P; ARANGO, J. **Information Architecture for the web and beyond**. 4. ed. Boston: O'Reilly, 2015.

SANTAREM SEGUNDO, J. E. **Representação Iterativa**: um modelo para repositórios digitais. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - UNESP, Marília. 2010.

SARACEVIC, T. Interdisciplinary nature of information science. *Ciência da Informação*. v. 24, n. 1, 1995.

SHARP, J. R. **Some fundamentals of information retrieval**. London: A. Deutsch, 1965. 224 p.

SVENONIUS, E. **The Intellectual foundation of information organization**. Cambridge: MIT Press, 2000.

VICKERY, B. C. **On retrieval system theory**. 2. ed. London: Butterwoths, 1965. 191 p.