



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 8 – Informação e Tecnologia
Comunicação Oral

**A EFICÁCIA DOS MODELOS DE RECUPERAÇÃO DE
INFORMAÇÕES: UM ESTUDO PARTICULARIZADO NA
COMUNICAÇÃO CIENTÍFICA NA WEB¹**

***THE EFFECTIVENESS OF INFORMATION RETRIEVAL MODELS: A
INDIVIDUALIZED STUDY IN SCIENTIFIC COMMUNICATION ON THE
WEB***

Oswaldo de Souza, UFC
osvsouza@gmail.com

Hamilton Rodrigues Tabosa, UFC
hrtabosa@gmail.com

Resumo: Objetiva analisar a eficácia dos recursos de recuperação de informação nos sites dos periódicos com Qualis A1 e A2 na área de Ciências Sociais Aplicadas, onde geralmente são publicados os artigos de pesquisadores na área de Ciência da Informação, concernentes aos seus modelos de recuperação (MR) e de seus sistemas de recuperação da informação (SRI). O estudo, caracterizado como descritivo e quantitativo, parte de uma revisão de literatura sobre os termos acessibilidade, usabilidade, recuperação da informação e modelos de recuperação de informação e de dados, seguida de estudo empírico nas páginas WEB dos referidos periódicos. Como principais resultados, os achados da pesquisa revelam que há falhas nos SRI dos periódicos analisados em relação a: a) da existência de metadados descritivos presentes nas bases de dados, mas que não são disponibilizados para o usuário nos resultados de busca, e b) do não oferecimento de recursos de busca e recuperação através de metadados cadastrados nas bases, mas que não se propõem a serem utilizados como chaves de busca. Apresentam-se questões quanto à razão de ser do cadastramento desses metadados como também quanto à eficácia dos SRI. Possíveis soluções são cogitadas, no sentido de maximizar o potencial de recuperação da informação por meio da ampliação dos recursos de busca que levem à plena utilização dos metadados descritivos, tanto para busca e recuperação quanto para a apresentação dos registros aos usuários de informação científica.

Palavras-chave: Recuperação de dados. Recuperação de informação. Modelos de recuperação de informação. Comunicação científica. Acessibilidade.

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

Abstract: It aims to analyze the effectiveness of resources of information retrieval in websites of journals classified as Qualis A1 and A2 in the area of Social Sciences, on which researchers in the field of Information Science usually published their articles. The analysis is concerning their recovery models (RM) and its information retrieval systems (IRS). The study, characterized as descriptive and quantitative, relies on a review of literature on the terms accessibility, usability, information retrieval and, information & data retrieval models, followed by empirical study on the Web pages of these journals. As main results, the survey findings reveal that there are flaws in IRS of analyzed journals, concerning to: a) the existence of descriptive metadata in the database but are not available to the user as search results, and b) of not offering search and recovery features through registered metadata on the bases. It also analyzed issues as the reason for the registration of such metadata as well as the effectiveness of SRI. Solutions are presented in order to maximize the potential of information retrieval leading to the full use of descriptive metadata, both for search and retrieval as for submission of records to scientific information users.

Keywords: Data Retrieval. Information Retrieval. Information Retrieval Models. Scientific Communication. Accessibility.

1 INTRODUÇÃO

Há quase meio século, Borko publicou um conceito de Ciência da Informação (CI) que tem reverberado bastante até os dias atuais, talvez por ser um dos primeiros, ou por ser bastante abrangente, senão vejamos:

Ciência da Informação é que a disciplina que investiga as propriedades e o comportamento da informação, as forças que regem o fluxo de informação e os meios de processamento da informação para melhor acessibilidade e usabilidade. Ela está preocupada com o corpo de conhecimento relacionado com a origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. (BORKO, 1968, p. 3. Tradução nossa).

A despeito das críticas que podem ser feitas a esse conceito, gostaríamos de frisar que dada à amplitude dos termos nele empregados, Borko chega a contemplar questões que hoje se fazem ainda atuais, ou até mais presentes na área do que antes, como por exemplo, a relação da CI com questões que perpassam a acessibilidade e a usabilidade. Nossa argumentação parte do conhecimento inequívoco de que o desenvolvimento tecnológico tem transformado sobremaneira o trabalho informacional que se tornou, e ainda se torna, mais e mais complexo à medida que crescem e se disponibilizam registros de informação, sejam eles científicos ou não. Tal transformação tem colaborado para a necessidade de um olhar mais atento da CI para temas e pesquisas que garantam a acessibilidade a toda essa gama documental, bem como para a usabilidade da informação relevante existente em meio ao crescente volume de registros de informação.

A recuperação da informação, no espaço digital, e em particular na Web, se dá pela interação com sistemas de recuperação da informação (SRI). Tais sistemas são construídos através de elementos técnicos que envolvem linguagens de programação de computadores, design de interfaces e mecanismos de busca e recuperação. Tais elementos, e em especial o design da interface e os mecanismos de busca e recuperação, inserem na atividade de recuperação da informação, problemas ligados à acessibilidade e à usabilidade.

Note-se que Borko (1968) aponta a acessibilidade e a usabilidade como atividades fim, ao seja, como o objetivo final, como a razão de ser da CI. Os termos “acessibilidade”, “usabilidade” e “utilização da informação” remetem, portanto, a todo o processo que envolve a coleta de informação, o processamento e os fluxos de informação, todos culminando na recuperação e uso da informação. A recuperação e uso da informação requerem a interação entre o usuário e o sistema, sendo pertinente observar que essa interação envolve questões de design, facilidade de uso e facilidade na obtenção de respostas do sistema. Tais questões são abordadas geralmente em âmbito da ergonomia e interface humano-máquina. Um sistema que responda positivamente a essas questões concorre para um desempenho adequado quanto à acessibilidade.

A percepção de que todos os esforços culminam na recuperação justifica o especial interesse no que diz respeito à acessibilidade, usabilidade e recuperabilidade da informação científica, que é a base para o desenvolvimento científico, tecnológico e de inovação.

Partindo-se desse interesse, empreendeu-se este estudo com o intuito de averiguar a eficácia dos modelos de recuperação de informações, particularizando-se o estudo na comunicação científica acessível pela Web.

Desse modo, tem-se como objetivo geral deste estudo: analisar a eficácia dos recursos de recuperação de informação nos sites dos periódicos com Qualis A1 e A2 na área de Ciências Sociais Aplicadas I², onde geralmente são publicados os artigos de pesquisadores na área de Ciência da Informação, concernentes aos seus Modelos de Recuperação (MR) e de seus Sistemas de Recuperação da Informação (SRI).

Para subsidiar a análise, empreendeu-se uma investigação que se pautou em uma revisão de literatura seguida de uma parte empírica, na qual objetivou-se:

- a) identificar o MR adotado no periódico;
- b) identificar a quantidade dos campos de busca suportados pelo SRI do periódico;

² Consulta realizada no Sistema Integrado CAPES (SiCAPES) / WebQualis. Disponível em: <http://qualis.capes.gov.br/webqualis/principal.seam>

c) identificar a quantidade de metadados solicitados e integrados ao banco de dados do Periódico.

Justifica-se a decisão de limitar o trabalho ao modelo Web, devido aos mesmos encontrarem-se disponíveis a todos e serem amplamente conhecidos pela comunidade acadêmica. No entanto, deve-se observar que as discussões sobre os MR e suas aplicações nos SRI sejam aplicáveis a todos os tipos de softwares, sejam na Web ou não.

A fim de estabelecer o entendimento dos autores sobre o tema, abordam-se os termos fundamentais à compreensão deste artigo, tais como acessibilidade, usabilidade e arquitetura da informação (AI).

Conforme Souza e Tabosa (2014), a acessibilidade na Web possibilita que qualquer pessoa, utilizando qualquer tipo de tecnologia, navegue sem restrições, sem enfrentar impedimentos ou dificuldades de nenhuma ordem, sendo capaz de interagir com o site e obter a informação que busca.

A acessibilidade no espaço digital consiste, para Torres, Mazzoni e Alves (2002, p. 85) em:

tornar disponível ao usuário, de forma autônoma, toda a informação que lhe for franqueável (informação para a qual o usuário tenha código de acesso ou, então, esteja liberada para todos os usuários), independentemente de suas características corporais, sem prejuízos quanto ao conteúdo da informação.

Assim, para garantir a acessibilidade à informação, é necessário eliminar barreiras arquitetônicas e quaisquer problemas de comunicação, de acesso físico ou de equipamentos, bem como empregar programas (softwares) adequados, promover o competente processamento técnico e otimizar a apresentação da informação.

O termo usabilidade é apresentado por Bevan (1995) como o termo técnico usado para descrever a qualidade de uso de uma interface. Para Pagani (2011), a usabilidade é oriunda da Interação Humano-Computador (IHC) e está contida nas atribuições da Engenharia de Software, carregando esse legado ao longo de sua evolução, podendo ser definida como o grau de facilidade com que o usuário consegue interagir com uma determinada interface.

Souza (2015) argumenta que a usabilidade volta-se para o estabelecimento não só de conceitos, mas também de práticas, para que as interfaces sejam desenvolvidas de forma que a concepção de aplicativos e sistemas traga, como consequência, um meio operacional fácil de usar e mais fácil ainda de ser compreendido. A qualidade de uma interface, ou seja, a sua facilidade de uso, está associada, segundo Nielsen (1993), aos seguintes princípios:

- facilidade de aprendizado;

- facilidade de lembrar como realizar uma tarefa após algum tempo;
- rapidez no desenvolvimento de tarefas;
- baixa taxa de erros;
- satisfação subjetiva do usuário.

Quanto ao termo arquitetura da informação, Wurman (1991) afirma que ele diz respeito à organização de padrões inerentes aos dados de modo a transformar o que é complexo em algo mais claro por meio da criação de estruturas ou mapas informacionais que viabilizem o alcance do conhecimento.

Já Lara Filho (2003) percebe a AI como “um conjunto de procedimentos metodológicos (e *ecológicos*) que permitem criar ordens num hipertexto visando abrir possibilidades de leituras para um conjunto de documentos”.

Dessa forma, temos que a AI se preocupa com a construção e com a otimização das macro e microestruturas sobre as quais são erigidas as ferramentas que viabilizam a eficaz recuperação de informação, garantindo acessibilidade e usabilidade a bancos e bases de dados bem como a suas interfaces como páginas da web.

Naturalmente, todos esses termos, que remetem necessariamente a ideias abstratas e a operações procedimentais, têm como finalidade o atendimento das necessidades do usuário da informação, que precisa estar satisfeito com os produtos e serviços da Ciência da Informação como um todo, circunstância que justifica toda a agenda de pesquisas na área.

2 A RECUPERAÇÃO DA INFORMAÇÃO E DE DADOS

Ao se discutir a recuperação da informação, oportuna-se lançar um olhar ao fundamento epistemológico pertinente, algo que nos remete imediatamente ao trabalho seminal de Shannon (1948) e ao trabalho de Baeza-Yates e Ribeiro-Neto (1999). Shannon, em seu conhecido modelo e Teoria da Informação (também referida por Teoria Matemática da Comunicação) nos diz que:

Frequentemente as mensagens possuem significado, isto é, elas se referem ou são relacionadas a algum sistema de entidades físicas ou conceituais. Esses aspectos semânticos da comunicação são irrelevantes ao problema de engenharia. A faceta significativa é aquela em que a mensagem real tenha sido selecionada entre um grupo de possíveis mensagens. O sistema deverá ser desenhado de modo a operar com cada uma das possíveis seleções a serem efetuadas, e não unicamente com aquela correntemente escolhida, posto que tal possa ser desconhecida quando concebemos ou desenhamos o sistema. (SHANNON, 1948, p. 1. Tradução nossa).

A aludida teoria expressa a preocupação relacionada à capacidade do sistema em

recuperar e entregar adequadamente as possíveis seleções a serem feitas. Shannon, ao referir-se as possíveis seleções, compreendidas como o conjunto de informações que possam ser recuperadas, expressa despreocupação com a semântica da resposta selecionada.

Compete ao MR, como artefato técnico, apenas garantir que a informação selecionada seja entregue corretamente e completamente. Por *corretamente e completamente* deve-se compreender que não compete ao aparato tecnológico utilizado na recuperação da informação, reter parte da informação ou considerá-la desnecessária. A informação selecionada pelo usuário deve ser fornecida em sua totalidade, isto é, todo o conjunto de metadados que compõem a informação deve ser entregue e não apenas fragmentos desse conjunto. Do contrário, qual a finalidade de manter metadados armazenados se não forem para servir ao usuário?

Qual a implicação dessa compreensão? Isso implica que os SRI (não importando qual seja seu MR) devem entregar ao usuário o registro completo, inclusive os metadados administrativos e estruturais, se existirem.

Isso implica em conceber sistemas capazes de responder a questões elaboradas posteriormente a sua concepção, bem como suportarem a inclusão de metadados não previstos no momento da concepção do sistema. Tais sistemas devem, portanto, ser capazes de produzir novas respostas, na medida em que o *corpus* seja modificado.

Deve-se, nesse ponto, fazer uma anotação relevante: compreenda-se por modificação que um *corpus* possa sofrer, não somente a inclusão de novos registros bibliográficos (tomando-se, por exemplo, o caso de um *corpus* relativo a registros bibliográficos), mas a inclusão de novos metadados. Tome-se, por exemplo, o *Digital Object Identifier System* (DOI), que é o elemento técnico adequado para o registro e identificação inequívoca de documentos digitais: são poucos os sistemas de periódicos que o suportam, mesmo tratando-se de periódicos unicamente digitais.

É disso que nos fala Shannon (1948, p. 1, grifo nosso) quando diz que “O sistema deverá ser desenhado de modo a operar com cada uma das possíveis seleções a serem efetuadas, e não unicamente com aquela correntemente escolhida, posto que tal possa ser desconhecida quando concebemos ou desenhamos o sistema”.

Portanto, os SRI, independente do MR nos quais sejam baseados, devem prever e suportar a inclusão de novos metadados, devendo obviamente também permitir a recuperação de informações a partir de novos metadados, ainda que eles venham a ser incorporados posteriormente, isto é, não faziam parte do *corpus*, foram acrescentados posteriormente ao banco de dados original.

Cabe aqui uma distinção, observada na literatura, sobre sistemas de recuperação, categorizando-os oportunamente em:

- a) de Recuperação de Informações (RI);
- b) de Recuperação de Dados (RD).

A RI segundo Baeza-Yates e Ribeiro-Neto (1999, p. 1) ocorre quando se manipula um *corpora* composto por elementos, geralmente textuais (e, portanto não estruturados), tais como livros e documentos diversos. A RD, por outro lado, ocorre quando se manipulam conjuntos de bancos de dados, fortemente estruturados e matematicamente definidos e expressos em tabelas (entidades), campos (atributos) e seus relacionamentos.

Na RI, usualmente o documento é representado por um conjunto pequeno de termos de indexação, por meio dos quais geralmente se viabiliza o acesso ao documento completo. Por outro lado, na RD o documento é decomposto em suas mínimas frações (metadados) de forma que o conjunto de termos de indexação é tão grande quanto o próprio documento. Normalmente na RD também há a indicação e o acesso ao documento completo.

Para que a necessidade informacional do usuário possa ser satisfeita, é preciso que ocorra uma combinação entre a RI e a RD. É disso que nos falamos Baeza-Yates e Ribeiro-Neto (1999, p. 1):

Data retrieval, while providing a solution to the user of a database system, does not solve the problem of retrieving information about a subject or topic. To be effective in its attempt to satisfy the user information need, the IR system must somehow 'interpret' the contents of the information items (documents) in a collection and rank them according to a degree of relevance to the user query.

A partir dessa combinação resultam os diversos tipos de modelos de recuperação, dentre os quais ocorrem, com maior frequência, os chamados booleano, vetorial e probabilístico, que são brevemente analisados a seguir.

2.1 MODELOS DE RECUPERAÇÃO DE DADOS E DA INFORMAÇÃO

Os sistemas de recuperação de dados e os de recuperação de informações são ambos frequentemente baseados nos modelos de recuperação: a) booleano e booleano estendido; b) vetorial; e c) probabilístico, conforme os argumentos de Souza (2006). O ponto comum entre esses modelos é a dependência da chave de busca.

Booleano - o modelo baseia-se no uso dos operadores booleanos *or*, *and* e *not* para estabelecer relações específicas de ocorrência com as chaves de busca, de forma a especificar os documentos a serem recuperados, baseados na presença ou não da chave de busca. Embora

o autor o considere simples e elegante, o modelo não oferece grande eficácia. Conforme Souza, a maior desvantagem da lógica de busca booleana é o fato de ela trabalhar de forma binária, ou seja, os documentos são analisados sob o critério dualista relevante/não relevante, não sendo criado nenhum tipo de ordenação dos resultados que atendam às condições de consulta. O modelo booleano estendido combina aspectos do modelo booleano e do modelo vetorial. Conforme Souza (2006), no modelo booleano estendido, busca-se a superação do problema das decisões binárias do modelo booleano simples por meio da atribuição de pesos aos termos, aproximando o modelo original do modelo vetorial.

Vetorial - Nessa tipologia de modelo, que é não binário, pode-se calcular um grau de similaridade a ser satisfeito pelos documentos para serem considerados relevantes (ex: que as palavras apareçam ao menos duas vezes, etc.) e determinar o grau de similaridade, com vistas a construir um *ranking*.

Probabilístico - nesse modelo, conforme Souza (2006) supõe-se que haja um conjunto ideal de itens que satisfaz a cada uma das possíveis consultas ao sistema e que essa massa documental pode ser recuperada. Através de tentativa inicial com um conjunto de documentos (onde podem ser utilizadas técnicas de outros modelos, como o vetorial ou o booleano) e da avaliação do usuário em sucessivas interações, busca-se aproximar cada vez mais desse conjunto ideal. O mérito do modelo probabilístico está em considerar a avaliação do usuário como um caminho para refinar o resultado continuamente. Esse modelo de recuperação em particular é raramente encontrado em uso.

A depender do sistema de recuperação adotado, de dados ou de informações, a chave de busca será utilizada em comparações entre um vasto conjunto de metadados ou um conjunto pequeno de termos de indexação, respectivamente.

Parece óbvio que seja melhor comparar a chave de busca a um vasto conjunto de metadados ao invés de um conjunto pequeno (termos de indexação), **todavia**, não tem sido essa a escolha dos periódicos científicos no que concerne aos seus sistemas de recuperação, tal afirmação é subsidiada a partir do desdobramento da pesquisa.

3 DESDOBRAMENTO METODOLÓGICO DA PESQUISA

Este estudo iniciou-se com um levantamento bibliográfico e revisão da literatura sobre o tema da pesquisa, de modo que fosse possível delinear o estado da arte e eleger as categorias de análise que embasam a argumentação aqui proposta.

Quanto à parte empírica do estudo, que fundamentou sua tipificação como descritivo e quantitativo, procedeu-se à consulta no SiCAPES - WebQualis - para identificar os periódicos classificados com o Qualis A1 e A2 na área de Ciências Sociais Aplicadas, os quais compuseram o *corpus* deste estudo. O site de cada um dos periódicos foi acessado e observado quanto aos recursos de busca básica e busca avançada, enquanto os dados eram tabulados em uma planilha com os campos de busca utilizados por cada periódico para a recuperação de informação.

Para estabelecer qual o MR adotado no periódico procedeu-se à instalação, em laboratório, de cada um dos softwares utilizados pelos respectivos periódicos, quando se tratava de software livre ou software que permitia uma instalação para testes. Quando não foi possível proceder a essa instalação, informações sobre o MR foram solicitadas ao periódico. A quantidade dos campos de busca suportados pelo SRI do periódico foi obtida utilizando-se as opções de busca oferecidas pelo periódico. Por fim, para identificar a quantidade de metadados solicitados e integrados ao banco de dados do Periódico, foi utilizada a funcionalidade de submissão de textos do software usado no periódico e, em seguida, pela contagem de quantos metadados eram integrados ao banco de dados usado no periódico.

No escopo da pesquisa define-se por “recursos básicos de busca” o mecanismo de recuperação que permite consulta pelos campos de autor, título, palavras-chave, resumo e texto completo. Qualquer outro atributo aceito pelo site como entrada para uma busca como, por exemplo, o ISSN ou o DOI, foi identificado como “algo mais” além do básico. Na sequência, efetuou-se o cadastro na categoria de “autor” e foi simulado um processo de submissão. A partir desse procedimento, obteve-se a quantidade de metadados descritivos requeridos para a submissão. Tais metadados passam a incorporar o *corpus* do periódico após o processamento do documento submetido.

É salutar esclarecer que nem todos os periódicos possuem site exclusivo para buscas, uma vez que alguns compõem bases de dados que abrigam outros títulos e têm um recurso de busca integrado, o qual normalmente é denominado de Portal. Nesses casos, foi o banco de dados do Portal que analisamos, já que é através dele que se realizam buscas nos títulos que estão contidos na base.

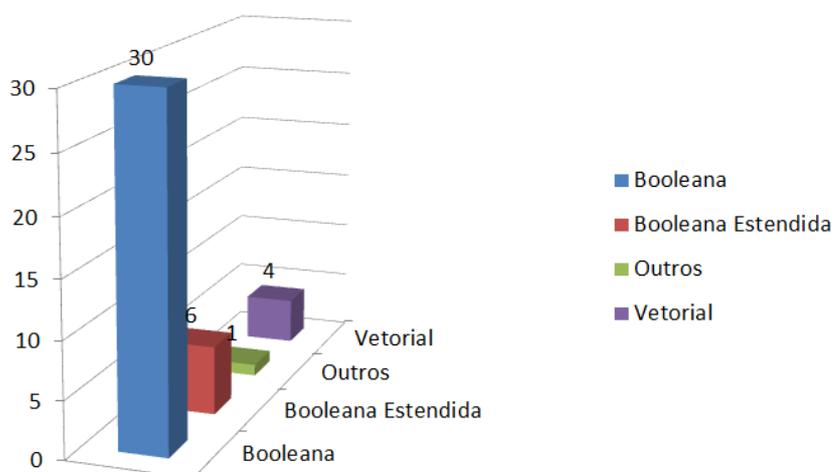
Procedida à análise dos dados planilhados, elaboramos gráficos e realizamos interpretações dos achados da pesquisa, tal como apresentamos na seção seguinte.

4 ANÁLISE E DISCUSSÃO

Observando a Figura 1, nota-se que a grande maioria dos SRI dos periódicos A1, 30 entre o total de 41 pesquisados, são baseados no modelo de recuperação booleano. Como afirma Souza (2006), o modelo booleano é o mais simples, razão que provavelmente explique sua adoção por grande parte dos sistemas, porém, ele não é o mais eficaz. A implementação de um recurso com tamanha simplicidade pode, falsamente, fazer acreditar que ele será conveniente e sabiamente utilizado pela maioria dos usuários. No entanto, devemos atentar para o fato de que sua ineficácia pode acarretar a insatisfação do consultante e inevitável desuso da ferramenta.

A categoria “outros” refere-se a um periódico que adota os recursos do Google para a realização das pesquisas.

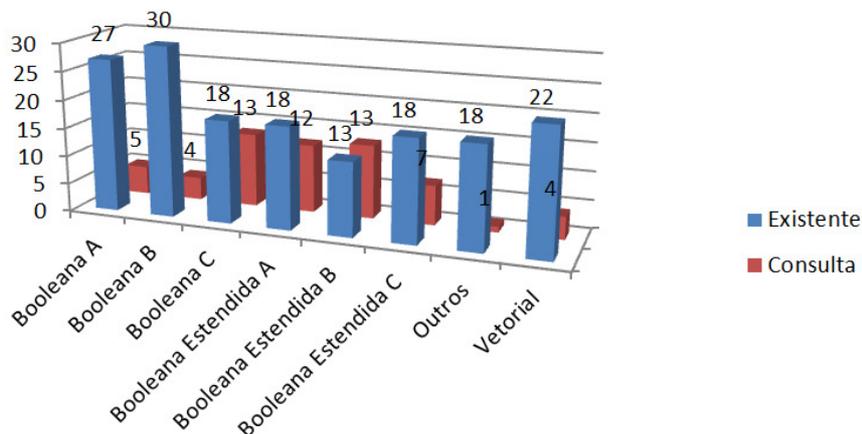
Figura 1 - Distribuição de modelos de recuperação - Qualis A1



Percebe-se, pela Figura 1, que foram encontradas quatro categorias de modelos de recuperação em uso nos 41 periódicos com Qualis A1 verificados na pesquisa.

Os 30 SRI baseados no “modelo booleano” encontrados foram divididos nas categorias A, B e C. Essa subcategorização foi necessária devido a diferenças entre a quantidade de metadados descritivos disponíveis para consulta e a quantidade de metadados descritivos existentes na base de dados. Os dados dessas subcategorias podem ser vistos na Figura 2. Procedimento semelhante foi necessário para a categoria “booleana estendida” dos periódicos Qualis A1, e também são mostrados na Figura 2. A estratégia de subcategorização também foi utilizada na análise do Qualis A2.

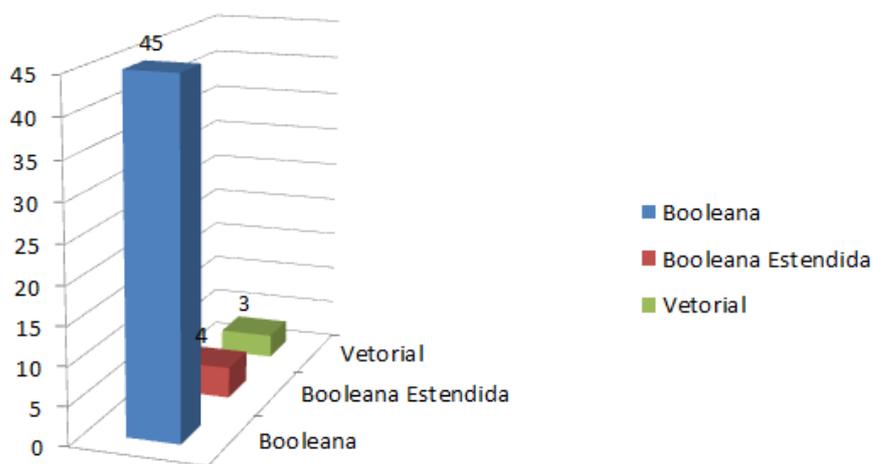
Figura 2 - Qtd. de metadados existentes X disponíveis para consulta - Qualis A1



Dentre os periódicos verificados, o de maior número de metadados no *corpus* é baseado no modelo booleano e dispõe de trinta metadados e apenas quatro metadados disponíveis para consulta. Um dos periódicos baseado no modelo booleano estendido disponibiliza treze metadados para busca e o universo de resposta envolve também treze metadados.

Pela Figura 3 pode se perceber que dos 52 periódicos Qualis A2 pesquisados, a maioria adota o modelo booleano, confirmando a tendência observada nos periódicos Qualis A1. Ao todo foram observadas três categorias de modelo de recuperação em uso.

Figura 3 - Distribuição de modelos de recuperação - Qualis A2

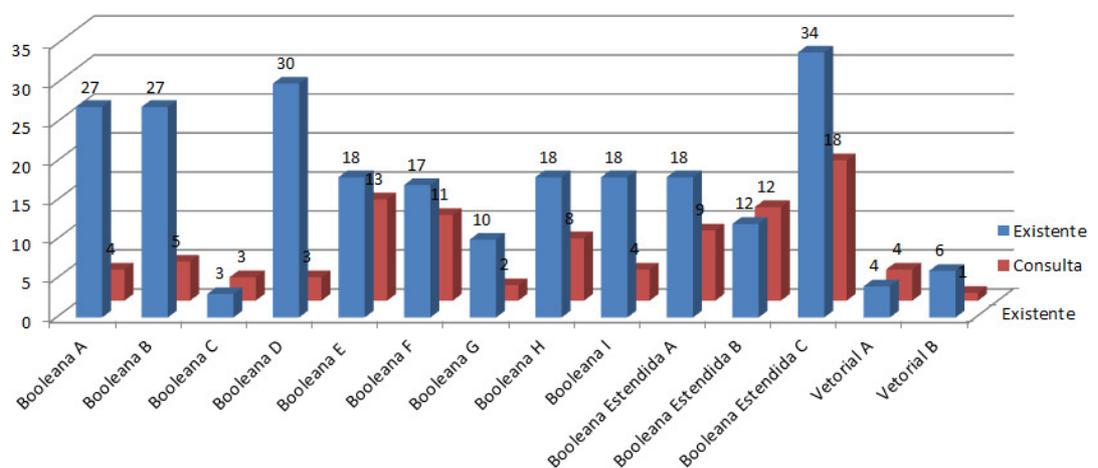


Todavia, ao observamos a Figura 4 podemos perceber que existe uma dispersão bem maior em termos de subcategorias de modelos de recuperação.

Nos periódicos Qualis A1 (Figura 2) observa-se que o máximo de campos de pesquisa permitidos por um modelo é de treze campos, enquanto no Qualis A2 o máximo atinge dezoito com um modelo booleano estendido C.

Uma possível interpretação indica que os periódicos do extrato A2 tem lançado mão de estratégias para aumentarem a disseminação de seus documentos, com consequente aumento do fator de impacto de suas publicações.

Figura 4 - Quantidade de metadados existentes X disponíveis para consulta - Qualis A2



4.1 AVALIAÇÃO DA EFICÁCIA DOS MODELOS DE RECUPERAÇÃO DA INFORMAÇÃO E DE DADOS

Tradicionalmente avaliam-se os MR a partir de indicadores de silêncio e ruído (CHAUMIER, 1988) e revocação e precisão (CARNEIRO, 1985), todavia, neste trabalho, avaliamos os modelos de recuperação da informação quanto ao conjunto de metadados pelos quais um usuário pode proceder a uma busca e quantos metadados ele recupera. Avalia-se que quanto mais metadados puderem ser utilizados como chave de busca em uma base de dados, mais robusto será seu poder de recuperação. Também consideramos que, após a realização da busca, quanto mais metadados forem apresentados como resultado, maior será a eficácia da resposta dada.

Na Figura 5, apresentam-se a melhor e a pior relação entre a quantidade de metadados disponíveis no banco de dados e a quantidade que é fornecida ao consulente,

subdivididas nos Qualis A1 e A2. Na Figura 6 e 7 apresentam-se todos os dados dessa natureza.

Figura 5 - Destaques na eficácia na entrega dos metadados

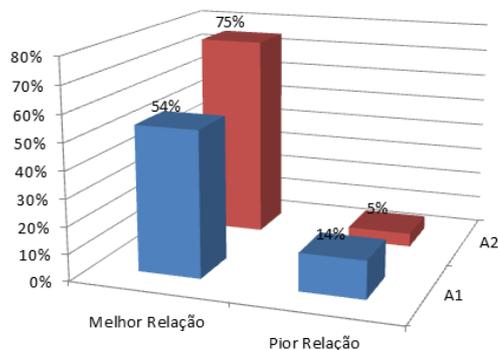
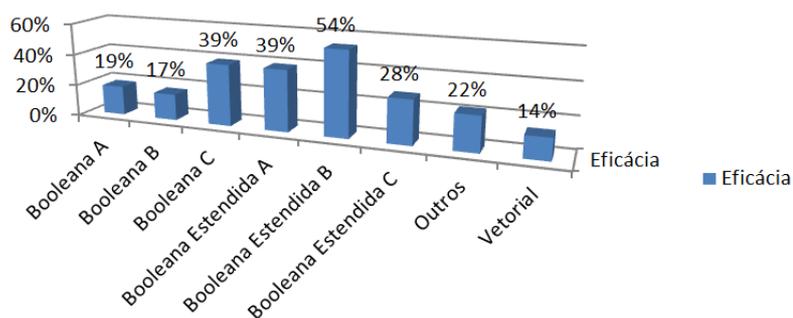
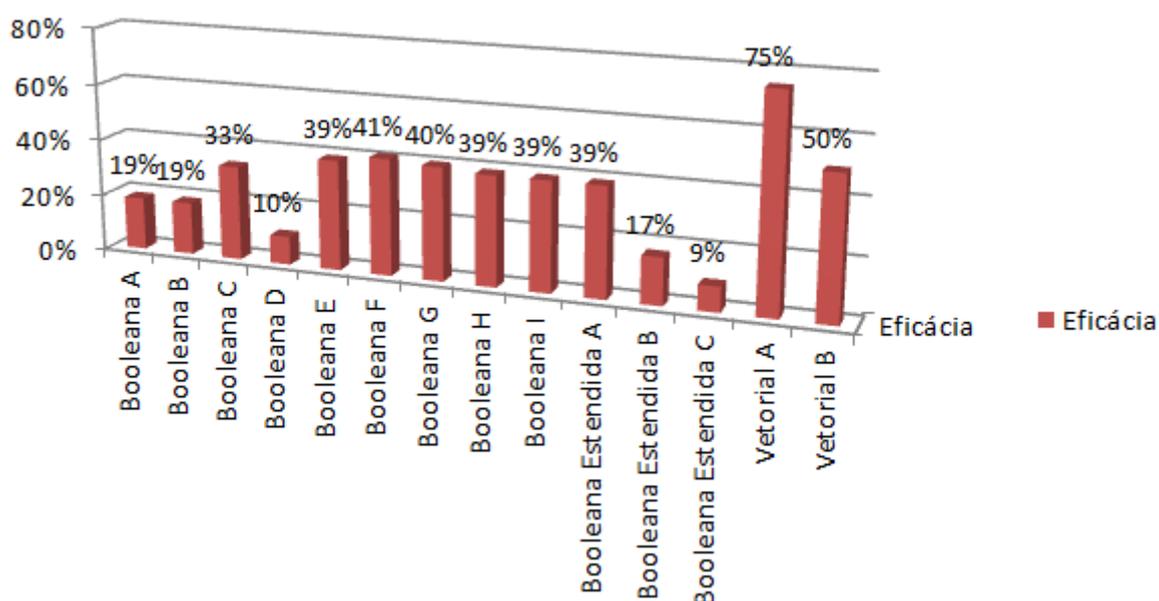


Figura 6 - Eficácia na entrega dos metadados - Qualis A1



Em ambos os extratos A1 e A2 pode-se observar uma variação muito grande entre a eficácia observada nos SRI dos periódicos. Tal variação pode ser explicada pela adoção de sistemas diferentes, ou de versões diferentes do mesmo sistema.

Figura 7 - Eficácia na entrega dos metadados - Qualis A2



Dessa forma, percebemos que, conforme a Figura 2 e Figura 4, dentro do *corpus* analisado, existem interfaces que oferecem dezoito opções diferentes de consulta, enquanto outras fontes trabalham com apenas uma. É óbvio que a primeira tem potencial de atender e satisfazer a um maior número de usuários, dada sua qualidade superior no que se refere à usabilidade e recuperabilidade de informação, mensuradas em função do número de opções de consulta.

Percebemos dois problemas quanto a essa fragilidade dos SRI: a) o primeiro diz respeito à quantidade de atributos através dos quais o usuário pode realizar uma busca e recuperar um documento, e b) a quantidade de metadados que o sistema mostra para o usuário como resultado de uma busca.

Ou seja, acreditamos que se o sistema permite a inclusão de vários metadados descritivos no momento do cadastro de um novo item na base de dados, o ideal seria que o usuário pudesse realizar buscas por todos (ou pelo menos quase todos) os atributos vinculados a um item. Reconhecemos que nem todos os metadados são, a priori, convenientes na montagem de uma estratégia de busca e que alguns deles são mais significativos como refinadores das consultas, como o metadado relativo ao idioma ou mesmo ao ano de uma publicação, por exemplo.

O que não se pode admitir é o fato de o sistema oferecer uma exaustividade na descrição do item e tal descrição não ser utilizada para nenhuma finalidade em termos de recuperação da informação.

O outro ponto digno de nota diz respeito ao que é entregue ao usuário como resultado das buscas: por qual razão os sistemas não apresentam as descrições completas dos materiais, mas apenas uma lista na qual se verificam registros incompletos, quando sabemos que outros tantos metadados dos documentos foram inseridos na base? Determinados dados descritivos podem ser fundamentais para que o usuário decida pela leitura ou abandono de um determinado item bibliográfico.

Como exemplo, veja-se o caso do *Open Journal System*, onde um único documento envolve 27 metadados descritivos. Tais metadados são fornecidos pelo autor durante a submissão, pelos editores do periódico e pelo próprio sistema durante o processo de publicação de uma edição.

Desse conjunto de 27 metadados, o usuário pode recuperar apenas cinco. Observe-se que 22 metadados não são fornecidos ao usuário. Isso representa sonegar 81,48% do registro relativo ao documento.

As Figuras 6 e 7 confirmam esse comportamento generalizado para os periódicos avaliados na pesquisa. Note-se que não consideramos, nesse exemplo, os metadados estruturais e os administrativos, uma vez que consideramos apenas os metadados descritivos do item. Isso revela que parte dos metadados somente são acessíveis aos operadores do sistema e não ao usuário final, ou seja, são dados que são registrados (o que demanda tempo de um operador humano e espaço de armazenamento) apenas para arquivo e não para consulta e uso. Disso decorrem os seguintes questionamentos: se há metadados ocultos no banco de dados e que não são entregues ao usuário, para que servem? Que outros dados são interessantes ao usuário mas não estão disponíveis para recuperação?

A partir dessas indagações naturalmente evoluímos para outras, tais como: Quem elegeu os atuais metadados disponíveis para consulta como os melhores para a recuperação? Com base em que critérios? Quais metadados devem ser acrescentados ao banco de dados?

Possíveis soluções para maximizar a recuperação da informação certamente deverão considerar os problemas e indagações apresentadas. Um possível caminho apresenta-se na aproximação dos SRIs aos Sistemas de Recuperação de Dados. Tendo em vista que os atuais SRIs dos periódicos Qualis A1 e A2 não possuem funcionalidades de recuperação semântica, não há prejuízo algum em aproximarem-se dos sistemas de recuperação de dados.

Com relação aos sistemas de recuperação, quer sejam os utilizados em periódicos científicos, ou na gestão de acervos nas Unidades de Informação, a plena utilização dos metadados descritivos (e dos demais tipos de metadados), visando a uma melhoria na qualidade dos serviços ofertados, somente será efetiva com o rompimento do atual estado de acomodação que se percebe na CI com relação ao tópico.

5 CONCLUSÕES

Se é verdade que a CI, em seu alvorecer (e a partir de então), esteve preocupada com a coleta, armazenamento, processamento técnico e recuperação da informação científica e tecnológica, diferentemente da Biblioteconomia e da Documentação, conforme afirmam Dias (2000), Pinheiro (1999) e Siqueira (2010), é de se esperar que esse campo do conhecimento tivesse construído e incorporado métodos e técnicas avançados e eficazes de organização e recuperação de informação em bases de dados de informação científica e tecnológica, principalmente dentro de sua própria área.

O que percebemos é que, talvez, os cientistas da informação tenham se debruçado mais detidamente sobre elementos teóricos e questões de pesquisa mais voltadas para fenômenos informacionais de cunho, majoritariamente, cognitivo e social, deixando a relação de proximidade da CI com a tecnologia um pouco carente de investigação e inovação.

Buckland (2012) parece ser mais radical em seu ponto de vista sobre esse tema, ao afirmar que a CI tem procurado ganhar visibilidade fazendo *marketing* de si mesma utilizando os termos “informação”, “sociedade” e “tecnologia”, sem conseguir consolidar conhecimentos coerentes sobre a natureza e o escopo do campo e considera essa prática um oportunismo imprudente.

A crítica parece partir de sua preocupação com a dificuldade, por parte da CI, de compreensão de conceitos e fenômenos, bem como com a falta de domínio de técnicas, teorias e artefatos importados de outras áreas e adaptados para seus interesses de estudo. O oportunismo imprudente a que se refere Buckland repousa no fato de a CI procurar atrair interesse por meio da propaganda, a algo que, de fato, ela não consegue dar conta.

Quanto a essa argumentação do autor, criticamos sua postura cética ao asseverarmos que a CI tem realmente atraído o interesse de pesquisadores de áreas correlatas (entre elas a Ciência da Computação) ou com interesses de pesquisa comuns sobre fenômenos e problemas informacionais, mas essa convivência acadêmico-científica tem gerado resultados de pesquisa em que se verificam esforços de uma interdisciplinaridade bastante

produtiva. Uma prova disso é a heterogeneidade dos que formam os programas de pós-graduação em CI, bem como os avanços teóricos e práticos que se observam nos eventos da área.

De fato, este estudo mostrou falhas na recuperabilidade de informação em quase a totalidade do *corpus* analisado, o que nos leva a questionar sobre as razões das inconsistências, que podem ser de várias naturezas, tais como: limitações de ordem técnica no projeto de arquitetura da informação, timidamente manifestados em limitados bancos de dados; o tratamento dos metadados descritivos; questões relacionadas a custo-benefício; negligência quanto à realização de estudos mensuradores do nível de satisfação da comunidade usuária quanto à eficácia do produto, entre outros fatores.

Assim, são necessários outros estudos para o efetivo diagnóstico de cada caso apontado, para que se planejem e se operacionalizem soluções para a otimização das ferramentas de busca, pois, ao contrário de Buckland (2012), acreditamos que a CI tem potencial para dominar o conhecimento necessário à implementação criativa e eficiente de expedientes que resolvam quaisquer questões que gravitem o fenômeno “informação” e seus problemas. E para os casos em que se perceba a necessidade de conhecimentos específicos de outras áreas, a CI terá condições de estabelecer frutíferas relações interdisciplinares para a elaboração conjunta de soluções sob demanda.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. [S.l.]: ACM Press, 1999.

BEVAN, N. Usability is quality of use. In: INTERNATIONAL CONFERENCE ON HUMAN COMPUTER INTERACTION, 6., **Proceedings...** [S.l.]: Elsevier, 1995. Disponível em: <<http://www.usability.serco.com/papers/usabis95.pdf>>. Acesso em: 12 ago. 2015.

BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, jan., 1968.

BUCKLAND, M. K. What kind of science can Information Science be? **Journal of Information Science and Technology**, v. 63, n. 1, 2012. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/whatsci.pdf>>. Acesso em: 13 ago. 2015.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CHAUMIER, Jacques. Indexação: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal**, v. 27, 1948.

DIAS, E. W. Biblioteconomia e Ciência da Informação: natureza e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 5, n. especial, p. 67-80, jan./jun. 2000.

LARA FILHO, D. O fio de Ariadne e a arquitetura da informação na WWW. **DataGramaZero: Revista de Ciência da Informação**, v. 4, n. 6, dez. 2003.

NIELSEN, J. **Usability engineering**. Boston: Academic Press, 1993.

PAGANI, T. **O que é usabilidade**. 2011. Disponível em: <<http://tableless.com.br/o-que-e-usabilidade/>>. Acesso em: 11 ago. 2015.

PINHEIRO, L. V. R. Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. In: _____. **Ciência da Informação, ciências sociais e interdisciplinaridade**. Brasília: IBICT, 1999.

SHANNON, C. E.; WEAVER, W. **Teoria matemática da comunicação**. Tradução de Orlando Agueda. Rio de Janeiro: Difel, 1975.

SIQUEIRA, J. C. Biblioteconomia, Documentação e Ciência da Informação: história, sociedade, tecnologia e pós-modernidade. **Perspectivas em Ciência da Informação**, v. 15, n. 3, p. 52-66, set./dez. 2010.

SOUZA, O. A usabilidade na perspectiva do uso da informação: estatísticas das pesquisas sobre o tema no Brasil. **Informação & Sociedade: Estudos**, João Pessoa, v.25, n.1, p. 159-172, jan./abr. 2015

SOUZA, O.; TABOSA, H. R. Virando a página: um novo conceito de acessibilidade na web para deficientes visuais. **Informação & Sociedade: Estudos**, João Pessoa, v. 24, n. 1, p. 145-161, jan./abr. 2014.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v. 11, n. 2, p. 161 - 173, mai./ago. 2006.

TORRES, E. F.; MAZZONI, A. A.; ALVES, J. B. M. A acessibilidade à informação no espaço digital. **Ciência da Informação**, Brasília, v. 31, n. 3, p. 83-91, set./dez. 2002.

WURMAN, R. S. **Ansiedade de informação**. São Paulo: Cultura Editores Associados, 1991.