

A OTIMIZAÇÃO DA BUSCA NO DSPACE VERSÃO 8: UM ESTUDO PARA IMPLEMENTAÇÃO NO RI UFRN

Lucas Angelo da Silveira

Doutor em Informática. Ibict, Brasília, DF, Brasil.

lucasangelo@ibict.br.

<https://orcid.org/0000-0002-8107-9659>.

Clediane de Araújo Guedes Marques

Doutoranda em Ciência da Informação. UFRN, Natal,

RN, Brasil. clediane.guedes@ufrn.br.

<https://orcid.org/0000-0001-5504-4826>.

Maria Aniolly Queiroz Maia

Doutora em Gestão & Organização do Conhecimento.

Ibict/ UFRN, Natal, RN, Brasil. mariamaia@ibict.br.

<https://orcid.org/0000-0003-4485-1774>.

Milton Shintaku

Doutor em Ciência da Informação. Ibict, Brasília, DF,

Brasil. shintaku@ibict.br.

<https://orcid.org/0000-0002-6476-4953>.

RESUMO

O estudo teve como objetivo otimizar o DSpace – versão 8 – com a implementação do operador booleano E (AND), além de identificar a relevância dessa melhoria para o processo de busca em repositórios institucionais, incluindo o Repositório Institucional (RI) da Universidade Federal do Rio Grande do Norte (UFRN). A otimização proposta contribui significativamente para o aprimoramento do acesso, da recuperação e do uso das informações nesses ambientes. Os resultados evidenciam a necessidade contínua de pesquisas voltadas à recuperação da informação, especialmente nos aspectos de usabilidade e encontrabilidade, no contexto do DSpace 8, de modo a orientar futuras melhorias no software. Metodologicamente, a pesquisa apresenta caráter exploratório, ao investigar o DSpace e sua ferramenta de busca, e configura-se como um relato de experiência, que possibilita a sistematização de conhecimentos iniciais sobre as práticas desenvolvidas.

Palavras-chave: DSpace 8. RIUFRN. Operador booleano. Solr.

SEARCH OPTIMIZATION IN DSPACE VERSION 8: A STUDY FOR IMPLEMENTATION IN THE UFRN INSTITUTIONAL REPOSITORY

ABSTRACT

The study aimed to optimize DSpace – version 8 – through the implementation of the Boolean operator AND, as well as to identify the relevance of this improvement for the search process in institutional repositories, including the Institutional Repository (IR) of the Federal University of Rio Grande do Norte (UFRN). The proposed optimization significantly contributes to enhancing access, retrieval, and use of information within these environments. The results highlight the ongoing need for research focused on information retrieval, particularly regarding usability and findability, in the context of DSpace 8, in order to guide future improvements to the software. Methodologically, the research has an exploratory nature, as it investigates DSpace and its search tool, and it also takes the form of an experience report, enabling the systematization of initial knowledge about the practices developed.

Keywords: DSpace 8; RIUFRN; Boolean operator; Solr.

Recebido em: 10/10/2025

Aceito em: 17/12/2025

Publicado em: 23/12/2025

1 INTRODUÇÃO

O movimento do acesso aberto (open access) transformou profundamente a comunicação científica ao defender que o conhecimento científico deve estar disponível livremente, sem restrições de acesso. Nesse contexto, Harnad et al. (2004) cunharam os termos via dourada (golden road) e via verde (green road) para designar, respectivamente, as revistas de acesso aberto e os repositórios institucionais. Esses repositórios passaram a desempenhar um papel fundamental como facilitadores do acesso à informação científica, conforme destaca Björk (2007), ao possibilitar a disponibilização a partir de autoarquivamento de artigos que muitas vezes permanecem restritos — atuando, portanto, como importantes sistemas de disseminação do conhecimento.

Com o avanço desse movimento, diversos softwares livres foram desenvolvidos para a implementação de repositórios, sendo o DSpace o mais amplamente utilizado, representando mais de 50% dos repositórios registrados no mundo, de acordo com o Registry of Open Access Repositories (ROAR). No Brasil, parte desse êxito pode ser atribuída às iniciativas do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), que tem desempenhado um papel central na promoção e no fortalecimento do acesso aberto no país.

Um exemplo proeminente dessa adoção é o Repositório Institucional (RI) da Universidade Federal do Rio Grande do Norte (UFRN), que tem como propósito "armazenar, preservar e disponibilizar na Internet, textos completos de acesso livre" (Repositório, 2025). Atualmente operando com o DSpace 7.6.2, o RIUFRN, em conjunto com o Ibict, identificou uma oportunidade de aprimoramento na funcionalidade de busca da versão 8 do DSpace. Notavelmente, a busca padrão desta versão utiliza o operador booleano OU (OR), o que pode gerar um aumento de resultados, mas com menor relevância.

Diante disso, a presente pesquisa propõe a otimização do DSpace – versão 8 – para incorporar o operador booleano E (AND) como padrão, buscando resultados de pesquisa mais precisos e contribuindo significativamente para outros repositórios DSpace no Brasil que almejam maior precisão e recuperação em suas buscas.

Nesse sentido, quanto à perspectiva metodológica, a pesquisa apresenta caráter

exploratório, ao investigar o DSpace e sua ferramenta de busca. Trata-se, também, de um relato de experiência, que possibilita a sistematização de conhecimentos iniciais acerca das práticas desenvolvidas.

2 METODOLOGIA

Para atender ao objetivo da pesquisa em foco, o presente estudo tem aspectos da Ciência da Computação, com modelo de relato de experiência. Da mesma forma, pode classificar o estudo como exploratório por trazer maior familiaridade com o objeto estudado, o DSpace e a sua ferramenta de busca (Gil, 2008). Quanto ao relato de experiência, Mussi, Flores e Almeida (2021) defendem que por meio desse formato cria-se conhecimentos, muitas vezes iniciais sobre um tema, sendo extremamente útil no relato de atividades e seus resultados, de forma científica.

Nesse contexto, o processo de estudo iniciou-se na identificação do problema, passando pelo ciclo simplificado de desenvolvimento de *software*, muito utilizado nas Ciências da Computação. Com isso, segue-se um fluxo de análise do problema, levantamento de coleta e análise dos resultados, e baseando-se na identificação desses dados dar-se-á a realização de melhorias, especialmente no contexto da versão DSpace 8 para implementação no RIUFRN.

3 OTIMIZAÇÃO DA BUSCA NO DSPACE VERSÃO 8: IMPLEMENTAÇÃO NO RI UFRN

A proposta de otimização da busca no Dspace 8 visa oferecer uma experiência de busca mais precisa, previsível e limpa, tanto na interface quanto na estrutura das consultas submetidas ao servidor. O DSpace, em sua configuração padrão, realiza buscas com base em todos os campos indexados, aplicando um modelo de recuperação genérico e pouco preciso.

Entre os principais problemas identificados estão: a busca difusa, em que o sistema utiliza o parser *EdisMax*¹ do Solr com configuração de mm flexível, permitindo a

¹ *EdisMax* é um parser avançado de consultas do Apache Solr que permite a construção flexível de buscas por múltiplos campos, com recursos como peso por campo, tolerância a erros e operadores lógicos configuráveis.

recuperação de resultados mesmo quando apenas um ou dois termos da consulta estão presentes, o que compromete a relevância e favorece a ocorrência de falsos positivos; a mistura de metadados, pois não há distinção clara entre os campos pesquisados, de modo que uma consulta simples varre campos como *dc.title*, *dc.subject* e *dc.contributor*, dificultando a compreensão do motivo pelo qual determinado item foi recuperado; o comportamento “iniciar com” nos filtros, já que os campos de título, autor e assunto utilizam lógica de prefixo (*startsWith*²), retornando valores que apenas começam com os termos digitados, o que pode gerar listagens ambíguas e fora do escopo real da pesquisa; e, por fim, as URLs confusas e desorganizadas, pois, a cada nova busca ou modificação (como alteração do campo de pesquisa ou aplicação de filtros), o sistema gera endereços com parâmetros duplicados ou conflitantes — como múltiplas ocorrências de *query=* ou repetições do mesmo filtro — resultando em links extensos, pouco legíveis e de difícil reutilização.

Para superar as limitações apresentadas, foi implementada uma customização completa no front-end do DSpace, sem a necessidade de alterações no back-end ou na configuração do Solr. A solução foi concebida para construir consultas bem definidas, que pudessem ser encaminhadas ao Solr de forma limpa e eficaz pelo back-end, assegurando controle total sobre o comportamento da busca. O principal recurso adicionado à interface (front-end³) foi um menu de seleção, no qual o usuário escolhe explicitamente o metadado a ser utilizado — como *dc.title*, *dc.subject* ou *dc.contributor.author*, entre outros.

A partir dessa escolha, são aplicadas três estratégias complementares: a restrição do escopo da busca ao campo selecionado, a obrigatoriedade de que todos os termos digitados estejam presentes no metadado (com uso do parâmetro *mm=100%*⁴ no Solr) e a eliminação do comportamento de busca por prefixo, exigindo correspondência pela presença total dos termos, independentemente da ordem ou do início das palavras. A lógica adotada é genérica e pode ser aplicada a qualquer metadado indexado no sistema, bastando incluí-lo no menu de seleção. Para exemplificar, ao realizar a busca pelo termo “aborto 20 saúde” no campo título, a consulta gerada é *dc.title:(+aborto +20 +saúde)*, com

² *startsWith* é uma lógica de filtragem que restringe os resultados a valores que começam com o texto informado.

³ *Front-end* é a interface do sistema com a qual o usuário interage diretamente.

⁴ *mm=100%* obriga a presença de **todos os termos** da consulta no campo buscado.

os parâmetros adicionais `spc.sf=score`, que define a ordenação por relevância, e `spc.mm=100%`, que exige a presença de todos os termos. Cada termo é prefixado com o símbolo `+`, que, no contexto do parser *EdisMax* do Solr, indica obrigatoriedade.

Essa estrutura evita que itens sejam retornados com apenas parte da expressão buscada. A construção da consulta ocorre inteiramente no front-end, porém é formatada de modo que o back-end⁵ (API REST do DSpace) a envie ao Solr sem modificações indevidas, garantindo sua correta interpretação. Além da construção das consultas customizadas, foram realizados ajustes importantes na experiência de usuário. O campo de entrada mantém o texto original digitado pelo usuário, ocultando os símbolos técnicos inseridos automaticamente e o valor selecionado no menu de metadado é restaurado a cada navegação. Também foram incorporados dois componentes adicionais à interface — um para ordenação e outro para seleção de quantidade de resultados por página.

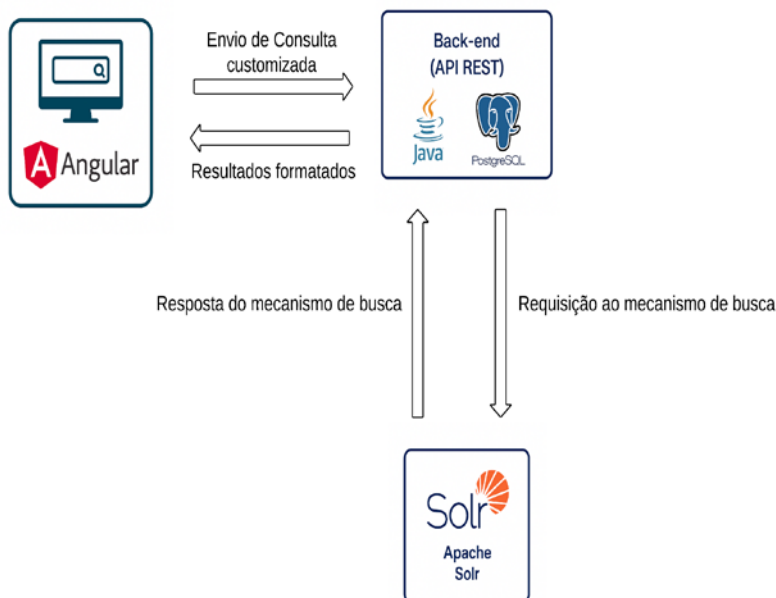
4 RESULTADOS E DISCUSSÃO

O conjunto de alterações propostas representam uma alternativa robusta à busca padrão do DSpace, mais alinhada à intenção real do usuário e mais transparente na comunicação entre a interface, a *API REST* e o mecanismo de busca Solr. O conjunto de alterações propostas constitui uma alternativa robusta em relação ao mecanismo de busca padrão do DSpace, apresentando maior aderência à intenção real do usuário e promovendo maior transparência na comunicação entre a interface, a *API REST* e o mecanismo de busca Solr.

A Figura 1 ilustra o fluxo de comunicação da solução. O front-end (Angular) monta a query customizada com base no metadado escolhido e a envia ao back-end (API REST), que repassa a requisição ao mecanismo de busca (Solr). Os resultados retornam ao back-end, que os entrega ao front-end para exibição.

⁵ Back-end do DSpace é o motor que armazena, processa e recupera todo o conteúdo digital e executa as regras de negócio do repositório.

Figura 1 - Fluxo de comunicação da solução



Fonte: Elaborado pelos próprios autores (2025).

4 CONCLUSÃO

A realização deste estudo, visou otimizar o DSpace – versão 8 – para implementar o operador booleano E (AND), assim como identificou a relevância de tal melhoria para a busca em repositórios institucionais, incluindo o RIUFRN.

Essa otimização facilita significativamente o acesso, a recuperação e o uso das informações nesses ambientes. Além disso, os achados reforçam a necessidade contínua de novas pesquisas em recuperação informacional, como usabilidade e encontrabilidade, especialmente no contexto do DSpace 8, para guiar futuras melhorias no *software*.

REFERÊNCIAS

BJÖRK, Bo-Christer. A model of scientific communication as a global distributed information system. **Information Research**, v. 12, n. 2, p. 1-47, 2007. Disponível em: <https://informationr.net/ir/12-2/paper307.html>. Acesso em: 22 maio 2025.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

HARNAD, Stevan et al. The access/impact problem and the Green and Gold Roads to Open Access. **Serials Review**, London, v. 30, n. 4, p. 310-314, jan. 2004. DOI: 10.1080/00987913.2004.10764930. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00987913.2004.10764930>. Acesso em: 22 maio 2025.

MUSSI, Ricardo Franklin de Freitas; FLORES, Fábio Fernandes; ALMEIDA, Claudio Bispo de. Pressupostos para a elaboração de relato de experiência como conhecimento científico. **Revista práxis educacional**, v. 17, n. 48, p. 60-77, 2021. Disponível em: http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S2178-26792021000500060. Acesso em 23 maio 2025.

REPOSITÓRIO Institucional da UFRN. Página inicial, 2025. Disponível em: <https://repositorio.ufrn.br/home>. Acesso em: 25 maio 2025.