

LUCAS MACHADO ALF

**DESENVOLVIMENTO DE UM REPOSITÓRIO INSTITUCIONAL DE
PERIÓDICOS ACADÊMICOS BASEADO EM NUVEM**

**Três de Maio
2022**

LUCAS MACHADO ALF

**DESENVOLVIMENTO DE UM REPOSITÓRIO INSTITUCIONAL DE
PERIÓDICOS ACADÊMICOS BASEADO EM NUVEM**

Trabalho de Conclusão de Curso do
Bacharelado em Sistemas de Infor-
mação - Faculdade Três de Maio -
SETREM

Orientadores:

Dr. FirstName LastName

Dr. FirstName LastName

Co-orientador:

Dr. FirstName LastName

Dr. FirstName LastName

Três de Maio

2022

LISTA DE QUADROS

Quadro 1: Orçamento.....	17
Quadro 2: Cronograma.....	18
Quadro 3: Trabalhos Relacionados	31

LISTA DE SIGLAS

SETREM	Sociedade Educacional Três de Maio
BPMN	<i>Business Process Model and Notation</i>

SUMÁRIO

1 PLANO DE ESTUDO E PESQUISA	10
1.1 TEMA	10
1.1.1 Delimitação do Tema	10
1.2 OBJETIVO GERAL	10
1.2.1 Objetivos Específicos	10
1.3 JUSTIFICATIVA	11
1.4 PROBLEMA	12
1.5 HIPÓTESES	13
1.6 METODOLOGIA	13
1.6.1 Abordagem	14
1.6.2 Procedimentos	14
1.6.2.1 <i>Pesquisa Bibliográfica</i>	15
1.6.2.2 <i>Pesquisa Ação</i>	15
1.6.3 Validação das Hipóteses	15
1.6.4 Técnicas	16
1.7 ORÇAMENTO	17
1.8 CRONOGRAMA DE ATIVIDADES	17
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 GESTÃO DO CONHECIMENTO	19
2.1.1 Gestão do Conhecimento Científico	19
2.1.2 Repositórios Institucionais	19
2.1.3 Padrão Dublin Core para metadados descritivos	19
2.1.4 Open Archival Information System (OAIS)	19
2.2 FERRAMENTAS UTILIZADAS	19
2.2.1 Cloud Computing	19
2.2.2 Cloud Object Storage	19
2.2.3 Microsoft .NET	19
2.2.4 ReactJS	19
2.2.4.1 <i>Vite</i>	19
2.2.5 PostgreSQL	19
2.2.5.1 <i>FullText Search</i>	19
2.2.5.2 <i>GIN Index</i>	19

2.3	TRABALHOS RELACIONADOS	19
2.3.1	BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados	20
2.3.2	Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência	21
2.3.3	Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital	22
2.3.4	Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca	23
2.3.5	O mapeamento dos repositórios institucionais brasileiros: perfil e desafios	24
2.3.6	Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis	25
2.3.7	The Emergence of Institutional Repositories: A Conceptual Understanding of Key Issues through Review of Literature	27
2.3.8	Why So Many Repositories? Examining the Limitations and Possibilities of The Institutional Repositories (IR) Landscape	28

INTRODUÇÃO

Os repositórios acadêmicos dentro de instituições de ensino podem ser vistos como uma ferramenta essencial para a preservação e propagação do conhecimento, visto que tais repositórios reúnem o acervo das pesquisas realizadas durante passagem dos acadêmicos pela instituição, e em teoria disponibilizam fácil acesso a este material.

Todavia, conforme Porto (2009) muitas instituições de ciência e tecnologia acabam por não dispor de uma estrutura profissionalizada de comunicação, suporte e divulgação dos materiais produzidos, não contemplando a comunicação em seu organograma funcional, e recorrendo a improvisações na hora da disponibilização de meios de acesso a divulgação de seus projetos.

Tendo isto em mente, o tema desta pesquisa surgiu como uma continuação direta a um projeto de prática profissional desenvolvido pelo próprio autor, durante o sétimo semestre da graduação em Sistemas de Informação, na Sociedade Educacional Três de Maio - SETREM, que possuía como objetivo analisar o processo de armazenamento e acesso as publicações realizadas dentro da instituição.

Durante a realização da prática profissional, foi constatado que a faculdade não possuía um sistema repositório acadêmico que possibilitasse aos alunos realizar a publicação de seus artigos e pesquisas. Também foi verificado por meio da aplicação de um questionário, que mais da metade dos acadêmicos (cerca de 62,1% das respostas) apresentavam interesse máximo em uma escala de 1 a 5, pela utilização de um repositório acadêmico institucional.

Ao pesquisar sobre repositórios acadêmicos existentes em plataformas agregadoras como o OpenDOAR¹, foi constatado que a maioria dos repositórios nacionais utilizam como base o DSpace, um *software open source* utilizado para a criação de repositórios acadêmicos institucionais.

Este *software* foi cogitado como um dos principais candidatos para uma possível implantação dentro da instituição, porem ao verificar trabalhos relacionados como o de Garcia (2019), foi relatado que por se tratar de um *software open source*, fica a critério das instituição todas as questões pertinentes a atualização de versões, personalização, segurança, privacidade dos dados e *backups* dos arquivos, oque resulta em um ambiente fragmentado onde algumas instituições utilizam versões antigas do DSpace, que por exemplo não funcionam de forma adequada em dispositivos moveis, conforme relatado no trabalho relacionado de Fernandes (2019).

Possuindo como inspiração os tópicos anteriormente citados, este trabalho de conclusão de curso possui como objetivo a utilização de tecnologias baseadas em nuvem para realizar a análise e desenvolvimento de uma nova plataforma de repositório acadêmico, que possibilite a interação entre os alunos e orientadores, englobando o processo de correção e envio de revisões de publicações dentro da própria plataforma, e que seja uma alternativa as atuais plataformas de repositórios existentes.

Com a realização deste projeto de pesquisa, buscou-se expor a importância da utilização de repositórios acadêmicos para realizar a preservação digital dos materiais produzido, evitando assim a perda ou esquecimento da produção acadêmica realizada dentro da instituição. Além de trazer contribuição tanto para os acadêmicos da faculdade SETREM, quanto para outras instituições que possam optar pela utilização do repositório acadêmico desenvolvido.

Este trabalho foi estruturado da seguinte forma, no capítulo 1 é abordado o plano de estudo, elencando tópicos como a delimitação do tema da pesquisa, objetivos, justificativa, metodologia, dentre outros. No capítulo 2 é apresentado o embasamento teórico e conceitual visto como necessário para realização deste

¹ <https://v2.sherpa.ac.uk/opensoar/>

trabalho.

No capítulo 3 são apresentados os resultados obtidos com a realização desta pesquisa, protótipos e apresentação do repositório acadêmico proposto. Em ultima parte, é apresentado a conclusão da pesquisa, detalhando os objetivos alcançados, validação das hipóteses, dificuldades e limitações encontradas durante o desenvolvimento, e propostas futuras de pesquisa.

1 PLANO DE ESTUDO E PESQUISA

1.1 TEMA

Desenvolvimento de um Repositório Institucional de Periódicos Acadêmicos baseado em nuvem.

1.1.1 Delimitação do Tema

A delimitação do tema se dará como o desenvolvimento de um repositório institucional de trabalhos acadêmicos, utilizando tecnologias baseadas em nuvem para realizar o armazenamento e acesso aos dados, e que envolva o processo de correções e envio de revisões de publicações dentro da própria plataforma.

O desenvolvimento deste projeto de pesquisa foi realizado durante o período de maio a julho de 2022, pelo acadêmico Lucas Machado Alf do curso bacharelado em Sistema de Informação na Sociedade Educacional Três de Maio – SETREM.

1.2 OBJETIVO GERAL

Desenvolver um repositório institucional de periódicos acadêmicos baseado em nuvem, tendo como objetivo reunir e preservar as publicações acadêmicas e científicas produzidas em âmbito universitário, além de unificar o processo de publicações e correções por parte dos orientadores em uma única plataforma.

1.2.1 Objetivos Específicos

1. Explorar ferramentas existentes de repositório acadêmico.
2. Pesquisar sobre trabalhos e artigos relacionados.
3. Pesquisar sobre formas de armazenamento de documentos em nuvem.

4. Investigar mecanismos de busca e análise textual de documentos. (Ex. ElasticSearch e Full Text Search).
5. Elicitar os requisitos para o desenvolvimento do repositório acadêmico.
6. Definir as tecnologias que serão utilizadas para o desenvolvimento (Ex: banco de dados, linguagens e frameworks).
7. Desenvolver uma aplicação web de repositório acadêmico.
8. Realizar testes no repositório acadêmico desenvolvido, verificando o desempenho da aplicação em relação ao crescimento da quantidade de registros publicados na plataforma.

1.3 JUSTIFICATIVA

O tema desta pesquisa surgiu como uma continuação direta a um projeto de Prática Profissional desenvolvido pelo próprio autor, durante o sétimo semestre da graduação em Sistemas de Informação, na Sociedade Educacional Três de Maio – SETREM, que tinha como objetivo analisar o atual processo de armazenamento e acesso as publicações produzidas dentro da faculdade, e realizar a representação gráfica dos processos por meio de diagramas BPMN (*Business Process Model and Notation*).

Conforme o próprio autor, durante o desenvolvimento da pesquisa foi constatado que a faculdade não possuía um sistema de repositório acadêmico que possibilitasse aos alunos realizar a publicação de seus artigos, práticas profissionais, interdisciplinares, TCCs e demais pesquisas realizadas na instituição. Também foi verificado por meio da aplicação de um questionário, que mais da metade dos acadêmicos (cerca de 62,1% das respostas) apresentavam interesse máximo em uma escala de 1 a 5, por um repositório acadêmico institucional.

Ao pesquisar sobre repositórios acadêmicos existentes em plataformas agregadoras como o OpenDOAR¹, foi constatado que a maioria dos repositórios nacionais utilizam como base o DSpace, um *software open source* escrito em Java

¹<https://v2.sherpa.ac.uk/opensoar/>

para criação de repositórios acadêmicos institucionais, sendo minoria as instituições que utilizam um sistema de repositório institucional diferente deste.

No trabalho relacionado de Garcia (2019) é possível verificar que por se tratar de um *software open source*, fica a critério das instituições as questões pertinentes a personalização do DSpace, segurança e privacidade dos dados, *backup* dos arquivos, e a realização das atualizações de versão, o que resulta em um ambiente fragmentado onde algumas instituições utilizam versões antigas do *software*, que por exemplo não funcionam de forma adequada em dispositivos móveis como relatado no trabalho relacionado de Fernandes (2019).

Possuindo como inspiração os tópicos anteriormente citados, este trabalho de conclusão de curso possui como objetivo e principal contribuição científica, a utilização de tecnologias baseadas em nuvem para realizar a análise e desenvolvimento de uma nova plataforma de repositório acadêmico, que possibilite a interação entre os alunos e orientadores, englobando o processo de correção e envio de revisões de publicações dentro da própria plataforma, e que seja uma alternativa as atuais plataformas de repositórios existentes.

Com a realização desta pesquisa, buscou-se trazer contribuição tanto para os acadêmicos da faculdade SETREM, quanto para outras instituições que possam optar pela utilização do repositório acadêmico desenvolvido. Também houve o propósito de expor a importância da utilização de repositórios acadêmicos para realizar a preservação digital dos materiais produzidos, evitando assim a perda ou esquecimento da produção acadêmica e científica realizada dentro da instituição.

Com o aprendizado adquirido durante o desenvolvimento deste trabalho de conclusão de curso, foi possível realizar o aprofundamento do conhecimento nas áreas de tecnologia da informação, nuvem computacional, repositórios acadêmicos, gestão de conhecimento, linguagens de programação e demais tecnologias utilizadas durante o desenvolvimento da pesquisa.

1.4 PROBLEMA

Como problema, foram ressaltados dois pontos principais que motivaram a escolha do tema desta pesquisa, sendo o primeiro a necessidade de reunir as

publicações acadêmicas produzidas dentro de instituições de ensino superior em um repositório institucional, que forneça rápido acesso às publicações científicas produzidas dentro da instituição, tanto de forma interna para os estudantes da instituição, quanto de forma externa caso a instituição opte por disponibilizar o seu acervo de forma pública.

Já o segundo ponto, envolve a forma como ocorre o processo de publicação, correção e revisão por parte dos orientadores, que em geral envolve a troca de vários emails entre os orientadores e os orientandos, processo o qual poderia ser melhorado por meio de um repositório acadêmico que suporte um esquema de publicação de revisões e correções.

Levando em consideração estes pontos, o problema desta pesquisa foi definido como: Como projetar um repositório de trabalhos acadêmicos baseado em serviços de nuvem, que envolva o processo de correções e envio de revisões das publicações dentro da própria plataforma?

1.5 HIPÓTESES

1. O recurso de Full Text Search do banco de dados PostgreSQL pode ser utilizado como uma alternativa viável para realizar as consultas por publicações dentro do repositório acadêmico (menos de 5 segundos por consulta), mesmo em bases de dados com mais de 30 mil publicações.
2. O processo desenvolvido para extração do texto das publicações para pesquisa e indexação é rápido o suficiente para não necessitar de processamento em segundo plano (menos de 5 segundos), mesmo em publicações com mais de 200 páginas.
3. É possível manter um ambiente em nuvem contendo os serviços de backend, frontend, banco de dados e armazenamento das publicações em nuvem, com um orçamento máximo de R\$ 500,00 por mês.

1.6 METODOLOGIA

Conforme (LOVATO, 2013, p. 15) a metodologia de pesquisa consiste em um ramo da filosofia da ciência, que estuda os métodos que os cientistas podem

utilizar para chegar nos resultados de seus estudos. Em outras palavras, tem como objetivo estudar os métodos que visam conduzir um aumento do conhecimento sobre o tema pesquisado, preocupando-se com o raciocínio, procedimentos e técnicas que podem ser utilizadas para dar credibilidade aos resultados.

1.6.1 Abordagem

Os métodos de abordagem segundo (LOVATO, 2013, p. 29) podem ser divididos em dois grupos, o primeiro consiste no tipo de raciocínio que é utilizado para se chegar aos resultados e conclusões. Já o segundo se relaciona com a utilização, ou não, de análise numérica e estatística.

Segundo o mesmo autor, a abordagem quantitativa é utilizada quando as conclusões são obtidas por meio de um conjunto de dados numéricos e análise estatística, busca compreender melhor situações onde é possível estabelecer relação entre variáveis, correlações ou causa-efeito.

Já segundo Gil (2008) a abordagem dedutiva, conforme a definição clássica, é o método em que parte-se de um conceito geral, onde parte dos conceitos são conhecidos como verdadeiros ou falsos, e se chega a uma conclusão particular, por meio da aplicação da lógica.

Para o desenvolvimento deste trabalho, foram utilizadas as abordagens dedutiva e quantitativa, respectivamente a primeira abordagem foi utilizada de forma a utilizar o raciocínio lógico sobre pesquisas e repositórios acadêmicos já existentes, para melhor compreender o escopo da pesquisa, e assim propor soluções mais adequadas ao problema.

Já a segunda abordagem foi utilizada para mensurar e analisar os dados numéricos referentes ao tempo de consultas e arquivamento das publicações no repositório acadêmico proposto, além dos custos envolvendo o armazenamento das publicações em nuvem.

1.6.2 Procedimentos

São apresentados neste tópicos os procedimentos utilizados durante a realização deste trabalho, como a pesquisa bibliográfica e o método de pesquisa

ação.

1.6.2.1 Pesquisa Bibliográfica

Conforme (LAKATOS, 2003, p. 183), a pesquisa bibliográfica abrange toda a bibliografia que se refere ao tema da pesquisa, incluindo desde publicações avulsas, boletins, jornais, revistas, livros, monografias, até meios de comunicação via rádio, gravações e filmes, tendo como finalidade realizar o contato entre o pesquisador e o material já existente sobre o tema pesquisado.

Este procedimento foi utilizado por meio da realização de pesquisas em livros, artigos e revistas, com a finalidade de se obter conhecimentos sobre o tema pesquisado, e outros sistemas já existentes com propósitos semelhantes.

1.6.2.2 Pesquisa Ação

Como elicitado por (LOVATO, 2013, p. 45), a pesquisa ação diferente de outras abordagens em que o principal embasamento consiste na literatura, o problema é real, e não existe um plano previamente definido e inflexível, mas sim um refinamento constante entre planejar, agir avaliar e refletir.

Por se tratar do desenvolvimento de um *software*, o procedimento de pesquisa ação foi utilizado na forma da pesquisa, aplicação e avaliação de tecnologias existentes, e que possam ser utilizadas durante o desenvolvimento do repositório acadêmico proposto.

1.6.3 Validação das Hipóteses

Durante a elaboração do presente trabalho, foram elencadas três hipóteses, sendo a primeira "O recurso de Full Text Search do banco de dados PostgreSQL pode ser utilizado como uma alternativa viável para realizar as consultas por publicações dentro do repositório acadêmico (menos de 5 segundos por consulta), mesmo em bases de dados com mais de 30 mil publicações". Esta hipótese pode ser validada por meio da preparação de uma base de dados de testes, que contenha mais de 30 mil publicações, e em seguida, realizar uma mesma consulta diversas vezes, realizando a média de tempo de execução.

A segunda hipótese é apresentada como "O processo desenvolvido para extração do texto das publicações para pesquisa e indexação é rápido o suficiente para não necessitar de processamento em segundo plano (menos de 5 segundos), mesmo em publicações com mais de 200 páginas". Esta segunda hipótese pode ser validada por meio de um teste de desempenho, que consiste realizar a publicação de uma mesma pesquisa com mais de 200 páginas diversas vezes, e mensurar o tempo médio para conclusão do processo.

Já a terceira hipótese, consiste em "É possível manter um ambiente em nuvem contendo os serviços de backend, frontend, banco de dados e armazenamento das publicações em nuvem, com um orçamento máximo de R\$ 500,00 por mês". Esta hipótese pode ser validada por meio da comparação dos preços de diferentes provedores de nuvem, com base nos requisitos mínimos estipulados para o repositório acadêmicos proposto.

1.6.4 Técnicas

Conforme (LAKATOS, 2003, p. 174) as técnicas podem ser definidas como um conjunto de procedimentos que servem a uma ciência ou arte, sendo utilizados para trazer tais conceitos a parte prática, tendo em mente que toda ciência utiliza de incontáveis técnicas para a obtenção de seus resultados.

Durante o desenvolvimento desta pesquisa, por se tratar do desenvolvimento de um *software*, foi utilizado da técnica de prototipação de telas por meio de *mockups* realizados com a ferramenta Figma.

Em tradução livre de Uzayr (2022) um *software mockup* consiste em um desenho de uma página ou aplicação web, que é desenvolvida para trazer vida a uma ideia, permitindo que um designer possa examinar como diferentes elementos visuais trabalham juntos. Os *mockups* também permitem que as partes interessadas ou *Stakeholders* do projeto possam verificar como a interface irá parecer, enquanto sugerem mudanças apropriadas em relação a cores, imagens e estilos.

Também foram utilizadas as técnica de modelagem ER (Entidade Relacionamento) por meio da ferramenta pgModeler para o desenvolvimento do banco de dados, e testes para verificar o funcionamento do *software* desenvolvido.

1.7 ORÇAMENTO

De acordo com (LAKATOS, 2021, p. 128) o orçamento tem como objetivo responder a questão "quanto será necessário despende?", distribuindo os gastos com o projeto em vários itens.

Quadro 1: Orçamento

Item	Quantidade	Valor Unitário	Valor Total
Impressões	1000	R\$ 0,30	R\$ 300,00
Capa	2	R\$ 25,00	R\$ 50,00
Espiraais	3	R\$ 5,00	R\$ 15,00
Serviços de Designer	1	R\$ 350,00	R\$ 350,00
Serviços de Banco de Dados em nuvem	1	R\$ 300,00	R\$ 300,00
Serviços de hospedagem em nuvem	2	R\$ 100,00	R\$ 200,00
Serviços de armazenamento em nuvem	1	R\$ 30,00	R\$ 30,00

1.8 CRONOGRAMA DE ATIVIDADES

Em conformidade com (LAKATOS, 2021, p. 128) o cronograma visa responder a pergunta “quando?”, sendo elaborado de forma a representar a previsão do tempo necessário para realização da pesquisa.

Quadro 2: Cronograma

Atividades	2022						
	Mai	Jun	Jul	Ago	Set	Out	Nov
Desenvolvimento do projeto							
Entrega e validação do projeto							
Organização de Dataset de trabalhos acadêmicos							
Desenvolvimento de protótipos							
Desenvolvimento do repositório acadêmico							
Realização de testes no repositório acadêmico							
Documentação dos resultados							
Entrega do Relatório Final							
Apresentação do Relatório Final							

	Proposto
	Realizado

2 FUNDAMENTAÇÃO TEÓRICA

2.1 GESTÃO DO CONHECIMENTO

2.1.1 Gestão do Conhecimento Científico

2.1.2 Repositórios Institucionais

2.1.3 Padrão Dublin Core para metadados descritivos

2.1.4 Open Archival Information System (OAIS)

2.2 FERRAMENTAS UTILIZADAS

2.2.1 Cloud Computing

2.2.2 Cloud Object Storage

2.2.3 Microsoft .NET

2.2.4 ReactJS

2.2.4.1 Vite

2.2.5 PostgreSQL

2.2.5.1 FullText Search

2.2.5.2 GIN Index

2.3 TRABALHOS RELACIONADOS

São apresentados neste tópico, os trabalhos relacionados, que possuem algum ponto em comum com o que se quer realizar no presente projeto. Pretende-se aqui, identificar estes pontos, e compará-los com a proposta deste projeto.

2.3.1 BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados

O trabalho desenvolvido por Cervi; Manica et al. (2009), possui como premissa a apresentação de uma proposta de biblioteca digital de trabalhos científicos, que foi denominada como BDTC, que possui como objetivo prover o suporte a três pontos fundamentais: auto-arquivamento do conteúdo, extração de metadados e busca de similaridade.

Como um dos principais diferenciais, foi desenvolvido um mecanismo de busca por similaridade para as pesquisas realizadas na BDTC, que permite ao usuário encontrar trabalhos relacionados, mesmo que parte da palavra pesquisada não seja exatamente igual ao conteúdo presente no documento.

Para o desenvolvimento deste mecanismo de busca, foi utilizado um recurso denominado *n-gram*, que permite quebrar uma palavra em conjuntos de letras de tamanhos variados, é retratado como exemplo o *3-gram* do termo "Juci", que podem ser quebrado em dois conjuntos de 3 letras, ou seja, "*Juc*" e "*uci*".

Estes conjuntos são armazenados em uma tabela de índices no banco de dados, de forma que ao realizar uma consulta a partir de uma palavra, o sistema retorna todos os trabalhos que contenham algum dos conjuntos de letras que compõem a palavra pesquisada.

Em relação com o presente projeto, pode ser realizado uma comparação com a forma como o mecanismo de busca foi desenvolvido na BDTC. Ao invés de utilizar o recurso de *n-gram*, o sistema proposto neste projeto utilizara o recurso de *tokenization* presente na ferramenta de *Full Text Search* do banco de dados PostgreSQL, que possui um resultado final semelhante, porém não idêntico, visto que o *tokenization* remove o gênero das palavras, os espaços em branco, palavras comuns e palavras que não são consideradas relevantes.

2.3.2 Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência

O relato de experiência desenvolvido por Garcia (2019) apresenta o desenvolvimento da nova plataforma de Biblioteca Digital da BBM, a Biblioteca Brasileira Guita e José Mindlin, em forma de retrospectiva desde o projeto-piloto, relatando os principais problemas, êxitos e desafios encontrados durante o desenvolvimento do projeto, que envolvia a digitalização e desenvolvimento de uma coleção digital para a biblioteca.

A Biblioteca Brasileira Guita e José Mindlin foi inaugurada em março de 2013, sendo um órgão e entidade acadêmica da Pró-Reitoria de Cultura e Extensão da USP (Universidade de São Paulo). Esta biblioteca envolve o projeto Brasileira USP, que foi iniciado em 2005, e tem como objetivo abrigar a coleção Brasileira, doada por José Mindlin. Dentro do escopo deste projeto, em 2008 é iniciado o projeto-piloto da Biblioteca Brasileira Digital, que visa a preservação do acervo e democratização do acesso ao material.

Para o desenvolvimento da biblioteca digital, foi optado por realizar uma customização do sistema DSpace, um *software open source* de repositório digital, com recursos como o Djatoka (servidor de imagens) e visualizadores de livros como IIPImage e BookReader.

Como principais problemas e êxitos, foi ressaltado a forma como as customizações foram realizadas no DSpace, sendo muitas delas realizadas diretamente no código fonte do programa, tornando extramente difícil ou até impossibilitando a atualização para novas versões da plataforma. Resultado em inconsistências na visualização dos documentos digitalizados, lentidão do sistema, dentro outros problemas que vieram a surgir ao longo do tempo.

Além disto, também foi relatado a rotatividade das equipes como um fator de impacto para a continuidade do desenvolvimento da biblioteca, que em sua maioria era constituída por bolsistas, estagiários e poucos profissionais terceirizados contratados por tempo determinado. Também foi constatado que as máquinas digitalizadoras adquiridas para a biblioteca não eram adequadas para o manuseio dos

documentos do acervo, visto que os documentos eram obras raras, e que necessitavam de diversos cuidados para a preservação e conservação do material.

Com o tempo parte dos problemas foram resolvidos, sendo até adquirido novas máquinas digitalizadoras, mais modernas e adequadas para o manuseio do material bibliográfico.

Comparando o trabalho relacionado com o presente projeto de pesquisa, é possível ressaltar que o atual projeto não tem como intenção a digitalização de um acervo físico, porém o sistema DSpace que foi utilizado no trabalho relacionado foi identificado como um sistema muito popular para o desenvolvimento de repositórios acadêmicos e bibliotecas digitais de universidades, sendo possível utilizar a experiência adquirida na implantação desse sistema durante o desenvolvimento do repositório acadêmico proposto.

2.3.3 Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital

O artigo desenvolvido por Pereira, Moreira e Segundo (2021) tem como proposta a apresentação de categorias fundamentais, baseadas nos trabalhos do matemático e bibliotecário indiano S. R. Ranganathan e do *Classification Research Group*, que pode ser utilizadas durante o desenvolvimento de interfaces de navegação facetada de bibliotecas digitais e repositórios de dissertações e teses.

Neste trabalho é relatado que é comum em bibliotecas digitais de teses e dissertações, encontrar problemas referente a facilidade de encontrar e recuperar documentos neles armazenados, visto que a principal forma de pesquisa tende a ser por um campo textual, onde é possível efetuar buscas simples, e em alguns casos utilizar a combinação de operadores lógicos, como *AND*, *OR* e *NOT*, para buscas mais complexas.

Porém esta forma de pesquisa, por meio de um único campo textual, exige que o usuário possua conhecimento prévio de noções de lógica, sigla dos cursos, áreas e linhas de pesquisa, e que seja capaz de utilizar essas informações para construir uma busca mais complexa, geralmente fazendo com que as pesquisas realizadas retornem resultados vazios, ou não exibam o total potencial de docu-

mentos contidos na plataforma, levando em consideração questões semânticas dos termos utilizados na pesquisa.

Desta forma, a abordagem de pesquisa facetada permite que o usuário navegue pela estrutura conceitual das informações armazenadas no repositório, além de combinar conceitos de diferentes perspectivas ou facetas (janelas ou menus), sendo uma abordagem de pesquisa mais eficiente, auxiliando o usuário a encontrar o que procura, de forma visual e intuitiva a partir de palavras chaves modificáveis em um vocabulário controlado.

Como resultado de sua pesquisa, foram obtidos as seguintes categorias fundamentais, seguidas de um exemplo: Documento (Trabalho de conclusão), Tipo (Dissertação, Tese); Curso (Meteorologia); Linha de Pesquisa (Sensoriamento); Tema (Anomalias climáticas); Especialização do tema (Conservação de Energia); Localização (Oceano Atlântico); e Ano de publicação (2020).

Em relação ao atual projeto de pesquisa, as categorias fundamentais encontradas no trabalho relacionado poderiam ser utilizadas para a organização dos documentos dentro do repositório de trabalhos acadêmicos proposto, podendo ser utilizados em recursos de filtragem das publicações dentro da plataforma.

2.3.4 Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca

A pesquisa realizada por Nascimento, Queiroz e Araujo (2020) tem como objetivo apresentar o desenvolvimento do plano de ação de preservação digital para o Arca - Repositório Institucional da Fiocruz, visando descrever as ações necessárias para garantir a preservação dos documentos, bem como a adoção de padrões, procedimentos e tecnologias que possam ajudar a garantir a preservação do seu acervo digital para o futuro.

O repositório institucional da Fundação Oswaldo Cruz (Fiocruz) denominado Arca, foi criado em 2007 e lançado oficialmente como repositório institucional em 2011, utilizando como base o *software open source* Dspace, e tendo como intuito reunir, hospedar, disponibilizar e dar visibilidade a produção intelectual e cultural produzida na fundação.

Como padrão referência, o artigo cita o *Open Archival Information System* (OAIS), presente na norma ISO 14721:2003, e adaptado na norma brasileira NBR 15472:2007. O OAIS define um modelo para configuração e operação de um repositório digital de documentos confiável, descrevendo como deve funcionar a estrutura e fluxo das informações, desde a inserção dos documentos digitais e metadados, até a forma como ocorre o seu armazenamento e acesso.

Dentre as repensabilidades obrigatórias para atender o modelo OAIS está a documentação das políticas e procedimentos para garantir a preservação dos documentos a longo prazo, bem como o plano de ação de sucessão caso o repositório seja desativado ou substituído por outro.

Como resultado de sua pesquisa, foi elaborado uma estrutura básica do Plano de Ação de Preservação Digital, onde em primeiro momento é descrito os elementos essenciais para a preservação, como: o cenário institucional, a descrição da coleção, avaliação de riscos e ameaças, e o planejamento das estratégias para prevenção de obsolescência. E em segundo momento, é optado por uma combinação de estratégias de preservação a serem aplicadas, como a normalização de formatos de arquivos, e a verificação periódica do formatos dos arquivos em uso, que possam apresentar riscos de obsolescência tecnológica.

Como contribuição ao atual projeto de pesquisa, pode ser citado a apresentação ao padrão OAIS, que pode ser utilizado durante o desenvolvimento do repositório acadêmico, visando a conformidade com padrões internacionais de desenvolvimento de repositórios acadêmicos confiáveis, além da apresentação de um plano de ação de preservação digital, que envolve a sucessão das obras armazenadas em caso de desativação ou substituição do serviço.

2.3.5 O mapeamento dos repositórios institucionais brasileiros: perfil e desafios

O artigo realizado por Weitzel (2019) tem como objetivo mapear os repositórios institucionais brasileiros até o período de maio de 2017, com a finalidade de identificar a atual situação de conformidade com a estratégia do Acesso Verde Aberto proposto pela BOIA (*Budapest Open Access Initiative*), além de contribuir

com a orientação de diretrizes nacionais e internacionais para implementação e desenvolvimento de repositórios institucionais, ou sua integração em rede.

A BOIA (*Budapest Open Access Initiative*) estabelece duas estratégias de Acesso Aberto, sendo a primeira o Acesso Aberto Dourado, que se baseia nos esforços da comunidade científica para criar um ambiente ideal onde os periódicos eletrônicos são disponibilizados sem cobrança de assinaturas ou taxas impostas pelas editoras, como as APCs (*Article Processing Charges*). A segunda estratégia é o Acesso Aberto Verde, onde o próprio autor realiza a publicação de seu periódico, seja a versão inicial ou final, em um repositório institucional.

Para a realização do estudo, foi realizado um levantamento de dados em fontes como o OpenDOAR, ROARMAP, ROAR, Lista de Repositórios do IBICT, edital da FINEP, lista de usuários do DSpace e os repositórios listados no *The Ranking Web of World Repositories*. Como indicador de alinhamento com o Acesso Aberto Verde, foi realizada a observação direta sobre cada um dos repositórios, de forma a verificar se o mesmo disponibiliza o acesso aos artigos de periódicos.

Como resultado, foi constatado que cerca de 54,5% dos 101 repositórios analisados estão alinhados com o Acesso Aberto Verde, concentrando 97,5% do total de artigos disponíveis em repositórios brasileiros. Também foi identificado que os sites agregadores de repositórios mundiais não expressam a realidade, contendo informações mal catalogadas, *links* quebrados, dentre outros problemas.

Em relação ao atual projeto de pesquisa, é possível citar a contribuição por meio da apresentação dos conceitos de Acesso Verde Aberto e Acesso Dourado, que visam disponibilizar os artigos de periódicos em repositórios institucionais, sem cobrança de taxas ou assinaturas. Também é possível ressaltar a apresentação de agregadores como OpenDOAR, onde é possível verificar que majoritariamente os repositórios brasileiros são desenvolvidos em cima do *software* DSpace.

2.3.6 Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis

A dissertação de mestrado elaborada por Fernandes (2019) possui como objetivo compreender a forma como ocorre a encontrabilidade da informação em

repositórios institucionais a partir do uso de dispositivos moveis, tendo como base o Repositório Institucional da UNESP.

Para o desenvolvimento da pesquisa foi utilizado o método quadripolar, onde no polo epistemológico foi realizado a definição do objetivo da pesquisa, no espaço da Ciência da Informação; no polo teórico é realizada a fundamentação conceitual sobre repositórios digitais, encontrabilidade da informação e dispositivos moveis; no polo técnico foi utilizado um conjunto de metodologias com *checklist*, teste com *eye tracking*, e entrevistas; e no polo morfológico é realizada a apresentação dos resultados.

Por meio da pesquisa realizada, foi constatado que grande parte dos repositórios são criados com base no *software open source* DSpace, que em suas versões mais recentes já possui uma interface responsiva, sendo adequada a diferentes tamanhos de telas. Porém, visto que as atualizações do *software* DSpace dependem da equipe técnica das instituições, muitos repositórios acadêmicos se encontram desatualizados, tornando difícil a navegação via dispositivos moveis.

Neste ponto é possível traçar uma correlação ao trabalho desenvolvido por Garcia (2019), onde é constatado que as personalizações realizadas pelas instituições no DSpace, podem acabar dificultando ou até impedindo a atualização do *software* para novas versões.

Como resultados de sua pesquisa, foi constatado que o Repositório Institucional UNESP já possui algum nível de preocupação com a responsividade para dispositivos moveis em seu *website*, entretanto com a pesquisa realizada foram encontrados alguns pontos que podem ser aprimorados.

Entre os tópicos que podem ser aprimorados no repositório, um dos principais seria a falta da caixa de busca no corpo da página, que se encontra oculto dentro de um menu. A sugestão seria que esta caixa de busca esteja visível logo no primeiro momento, pois é um dos elementos principais dentro do repositório.

Sendo assim, foi possível afirmar que no repositório analisado, elementos não relevantes em dispositivos moveis possuem maior destaque do que outros elementos que possuem maior importância, como por exemplo a função de auto-

arquivamento, sendo um recursos muito pouco utilizado a partir de dispositivos moveis, e que possui uma visibilidade maior que a caixa de pesquisa.

Outros recursos que podem ser aprimorados no repositório seriam o recurso de autocompletar na caixa de busca, facilitando a localização dos conteúdos pesquisados. E o recurso de *wayfinding* poderia ser melhorado, mudando a cor de *links* já visitados, componentes de *breadcrumb* alternativos para dispositivos moveis, e *link* visível para retornar a pagina inicial.

Como contribuição ao trabalho atual, é possível ressaltar o resultados obtidos com a análise de *eye tracking* realizado pela pesquisa, podendo ser utilizada para a criação de uma interface gráfica para o repositório acadêmico, que exiba conteúdos mais relevantes para os usuários de dispositivos moveis.

2.3.7 The Emergence of Institutional Repositories: A Conceptual Understanding of Key Issues through Review of Literature

Na pesquisa realizada por Saini (2018) é realizado uma revisão de literatura sobre o surgimento dos repositórios institucionais, buscando a compreensão conceitual dos principais problemas que envolvem a criação e uso de repositórios institucionais. A revisão de literatura consiste em uma organização conceitual da combinação de resultados que provem contexto para pesquisas. Ajudando a refinar ideias, conhecer especificações do método de pesquisa, e trazer clareza sobre diversos fatores que cercam o surgimento, motivação, custos, seleção de *software* e uma perspectiva global de sobre os repositórios institucionais.

Em sua pesquisa, foi constatado que o surgimento dos repositórios institucionais não consiste em um fenômeno recente em instituições acadêmicas, e que este tema já vem existindo por quase uma década, surgindo primeiramente em bibliotecas de instituições de grande porte, que realizaram o papel de *early adopters* no desenvolvimento de repositórios institucionais mais focados na acumulação, preservação e disseminação da pesquisa realizada na instituição, de uma forma acessível e aberta.

Como motivação para a adoção de repositórios institucionais, é considerado que estes fazem parte da plataforma de serviços oferecidos pela instituição,

sendo um recurso que pode ser utilizado para vender e divulgar o nome da instituição, de forma a propagar as suas pesquisas e facilitar o intercâmbio de conhecimento.

Já sobre os fatores de custos, é possível elencar um elevado número de considerações, como a utilização de *software open source* ou o licenciamento de *software* proprietário, a abertura de acesso fora do campus, ou o fornecimento de acesso total, ou limitado do conteúdo. Também foi relatado que dentre os maiores está o custo operacional da biblioteca, e que ao optar pelo auto arquivamento digital, boa parte dos custos podem ser evitados.

Sobre a atuação de *softwares open source*, foi relatado que os dois sistemas mais predominantes são o Eprints e o DSpace. O Eprints foi lançado em 2000, desenvolvido por Stephen Harnad e sua equipe na Universidade de Southampton, já o DSpace foi lançado 2002 pelo Instituto de Tecnologia do Massachusetts (MIT). O Eprints foi projetado para suportar formais mais tradicionais de publicações, como periódicos, conferências, artigos, capítulos de livros, e outros. Já o DSpace tem como intenção suportar uma variedade muito maior de documentos, como comunicados formais da instituição, e outras categorias de literatura.

Por fim, sobre a perspectiva global de repositórios institucionais, foi relatado que a maior parte se encontra na Europa, contribuindo com 47.92% do total, seguido pela América do Norte com 29.28%, Ásia com 11.04%, Austrália com 5.84%, a América do Sul com 4.40%, e África com 1.52% dos repositórios mundiais, tendo como fonte dados de março de 2018 presentes no OpenDoar.

Este artigo pode ser visto como relacionado, visto que atua como uma revisão de literatura sobre o surgimento dos repositórios institucionais, que pode ser utilizada como embasamento para a atual pesquisa, trazendo dados referentes a motivação, custos e *softwares* existentes.

2.3.8 Why So Many Repositories? Examining the Limitations and Possibilities of The Institutional Repositories (IR) Landscape

No trabalho realizado por Arlitsch; Grant (2018) possuiu como objetivo examinar as limitações e possibilidades encontradas no cenário de repositórios ins-

titucionais. Em sua pesquisa foi identificado que o número de repositórios vinha crescendo lentamente até meados de 2005, período em ocorreu um grande aumento no número de repositórios existentes, de 128 em dezembro de 2005 para 2,253 em dezembro de 2012 conforme dados do OpenDoar, representando um aumento de 1660% durante o período.

Todo via, grande parte dos repositórios que surgiram ao longo do tempo são hospedados e mantidos de forma local pela própria instituição, o que até possui alguns aspectos positivos, como a flexibilidade, controle, customização, e a tendência de vários sistemas distribuídos possuírem menor risco de falha massiva que um único sistema centralizado.

Porem também existem aspectos negativos sobre a existência de múltiplas instâncias de repositórios acadêmicos sendo hospedados em vários locais, por diferentes instituições, como por exemplo a pesquisa e encontrabilidade do conteúdo, problemas com a coleta e armazenamento dos metadados, e a variedade de diferentes versões de *software* sendo utilizadas.

É relatado que a fragmentação da utilização de diferentes versões de *software* pela comunidade é um tópico dramático, citando como exemplo o *DuraSpace Registry* que realiza a listam dos repositórios que utilizam como base o DSpace, foi identificado cerca de 2,004 repositórios que estavam rodando a plataforma DSpace até dezembro de 2017, destes cerca de 800 (quase 40%) estavam utilizando a versão 1.8 ou menor, uma versão lançada em 2011, sendo que qualquer versão abaixo de 5.x não é mais suportada desde janeiro de 2018.

As implicações negativas desta situação não podem ser ignoradas pelas instituições, visto que a cada nova versão de *software* que é lançada, fica mais desafiador para as instituições realizar a atualização destes sistemas. Além da sucessão a riscos de segurança e ataques de ransomware, que comprometem a confiança que a comunidade possui em depositar suas publicações nestes repositórios.

Em sua pesquisa é citado que a ideia das instituições hospedarem seu próprio repositório institucional de forma local já está obsoleta, e deveriam ser

substituídos por alternativas viáveis, fazendo uma relação ao tempo que as instituições também hospedavam os seus próprios servidores de email, onde hoje em dia são utilizados serviços baseados em nuvem como os fornecidos pela Google e Microsoft.

Porem o conceito de alternativa viável pode cobrir muitas possibilidades, e não existe um acordo comum sobre uma escolha correta, algumas discussões focam na adoção de versões mais recentes de sistemas como o DSpace e Fedora Commons, outros focam em soluções comerciais baseadas em nuvem, que oferecem escalabilidade e redundância.

Como conclusão, é estimado que a proliferação de repositórios locais por bibliotecas de instituições acadêmicas, criam mais problemas do que trazem benefícios, considerando que existem algumas vantagens de possuir um controle local, porem a existência de centenas de repositórios diferentes não servem mais ao usuário comum, e acabam trazendo diversos problemas e custos.

Este trabalho pode ser visto como relacionado a presente pesquisa, visto que aborda tópicos relacionados as vantagens e desvantagens de possuir um repositório local, ou investir em soluções baseadas em nuvem que podem ser fornecidas como serviços a instituição.

Quadro 3: Trabalhos Relacionados

Trabalho	Objetivo	Metodologia	Resultados	Conclusão
Pereira, Moreira e Segundo (2021)	Apresenta uma proposta de categorias fundamentais a serem utilizadas para a construção de uma interface de navegação facetada para uma biblioteca digital	Qualitativa e exploratória.	Foram criadas 8 categorias fundamentais para a navegação facetada.	As categorias fundamentais propostas neste trabalho podem contribuir significativamente para a construção de um mecanismo de busca facetada.
Nascimento, Queiroz e Araujo (2020)	Construção do Plano de Ação de Preservação Digital para o Arca – Repositório Institucional da Fiocruz	Qualitativa e descritiva	O Plano de Ação de Preservação Digital do repositório Arca estabelece os padrões que visam garantir que a produção científica seja preservada de forma permanente, em um ambiente confiável e seguro.	Algumas das recomendações resultantes dos diagnósticos não estão diretamente relacionadas ao Plano de Ação de Preservação Digital, mas podem ser incorporadas ao Plano Operativo do repositório Arca.
Garcia (2019)	Desenvolvimento da na nova plataforma de Biblioteca Digital da BBM (Biblioteca Brasileira Guida e José Mindlin)	Descritiva e relato de experiência.	Foi optado por realizar uma customização do sistema DSpace, um <i>software open source</i> de repositório digital.	Como principais problemas e êxitos, foi ressaltado que a forma como as customizações foram realizadas no DSpace tornam difícil para novas versões da plataforma.
Weitzel (2019)	Mapear os repositórios institucionais brasileiros até o período de maio de 2017 a fim de retratar a situação atual e contribuir com subsídios para orientar as ações e diretrizes para implementação de repositórios.	Levantamento exaustivo de repositórios brasileiros por meio de fontes específicas e pela observação direta.	Foi verificado que cerca de 54,5% dos repositórios concentram 97,5% do total de artigos dentre os 101 repositórios identificados no país.	O estudo recomenda que o governo brasileiro estabeleça diretrizes nacionais para fomentar as boas práticas em repositórios no país com o objetivo de fortalecer o desenvolvimento do Acesso Aberto Verde
Fernandes (2019)	Compreender de que forma ocorre a encontrabilidade da informação em repositórios institucionais a partir do uso de dispositivos móveis, com ênfase no Repositório Institucional UNESP	Método quadrípolar.	No repositório analisado elementos não relevantes em dispositivos móveis possuem maior destaque do que outros elementos que possuem maior importância.	O Repositório Institucional UNESP já possui algum nível de preocupação com a responsividade para dispositivos móveis em seu <i>website</i> , entretanto com a pesquisa realizada foram encontrados alguns pontos que podem ser aprimorados.
Cervi; Manica et al. (2009)	Apresentação da proposta de biblioteca digital de trabalhos científicos, que foi denominada como BDTC.	Qualitativa	Apresentação uma biblioteca digital de trabalhos científicos, a BDTC, que faz uso de serviços como autoarquivamento de conteúdo, extração de metadados dos objetos digitais e busca por similaridade.	Com relação ao processo de extração automática dos metadados dos objetos digitais, a ocorrência de erros é minimizada, ficando a critério do autor realizar apenas a confirmação dos metadados extraídos.

REFERÊNCIAS

- ARLITSCH, K.; GRANT, C. 2018. **Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape.** Journal of Library Administration [S.l.] v.58 n.3 p 264–281.
- CERVI, C.; MANICA, E.; DORNELES, C.; GALANTE, R. 2009. **BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados.** Revista Brasileira de Computação Aplicada [S.l.] v.1 n.1 p 65–76.
- FERNANDES, W. M. 2019. **Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis.** [Online]. Marília, SP, Brazil.
- GARCIA, R. M. 2019. **Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: relato de experiência.** PragMATIZES [S.l.] n.16 p 111.
- GIL, A. C. 2008. **Métodos e técnicas de pesquisa social.** São Paulo, SP: Atlas. ISBN 9788522451425.
- LAKATOS, E. M. 2003. **Fundamentos de metodologia científica.** São Paulo, SP: Atlas. ISBN 8522433976.
- LAKATOS, E. M. 2021. **Metodologia do Trabalho Científico.** 9.ed. São Paulo, SP: Atlas. ISBN 9788597026542.
- LOVATO, A. 2013. **Metodologia da Pesquisa.** [S.l.]: SETREM. ISBN 9788599020050.

- NASCIMENTO, A. G. do; QUEIROZ, C. F. de; ARAUJO, L. D. de. 2020. **Garantindo acervos para o futuro: plano de preservação digital para o repositório institucional arca**. Ciência da Informação [S.l.] v.48 n.3.
- PEREIRA, C. M.; MOREIRA, W.; SEGUNDO, J. E. S. 2021. **Classificação face-tada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital**. Encontros Bibli [S.l.] v.26 p 1–21.
- PORTO, C. M. 2009. **Difusão e cultura científica: alguns recortes**. Salvador, BA: EDUFBA. ISBN 9788523206192.
- SAINI, O. P. 2018. **The Emergence of Institutional Repositories**: a conceptual understanding of key issues through review of literature. [Online]. Lincoln, Nebraska, United States.
- UZAYR, S. 2022. **Mastering UI Mockups and Frameworks**: a beginner's guide. [S.l.]: CRC Press. (Mastering Computer Science). ISBN 9781000583106.
- WEITZEL, S. d. R. 2019. **O mapeamento dos repositórios institucionais brasileiros: perfil e desafios**. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação [S.l.] v.24 n.54 p 105–123.