

LUCAS MACHADO ALF

**DESENVOLVIMENTO DE UM *SOFTWARE* DE REPOSITÓRIO INSTITUCIONAL
DE PESQUISAS PREPARADO PARA AMBIENTES BASEADOS EM NUVEM**

Três de Maio

2022

LUCAS MACHADO ALF

**DESENVOLVIMENTO DE UM *SOFTWARE* DE REPOSITÓRIO INSTITUCIONAL
DE PESQUISAS PREPARADO PARA AMBIENTES BASEADOS EM NUVEM**

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

Orientadores:

Msc. Tiago Cesa Seibel

Três de Maio

2022

LISTA DE QUADROS

| | | |
|------------|--|----|
| Quadro 1: | Exemplo de elementos do Dublin Core | 23 |
| Quadro 2: | Sistemas Relacionados..... | 42 |
| Quadro 3: | Trabalhos Relacionados Parte 1 | 63 |
| Quadro 4: | Trabalhos Relacionados Parte 2 | 64 |
| Quadro 5: | Tela de Login..... | 67 |
| Quadro 6: | Cadastro de usuário..... | 68 |
| Quadro 7: | Recuperação de senha | 69 |
| Quadro 8: | Visão geral..... | 70 |
| Quadro 9: | Minhas publicações | 71 |
| Quadro 10: | Nova publicação | 72 |
| Quadro 11: | Pesquisar publicações..... | 73 |
| Quadro 12: | Visualizar publicação..... | 74 |
| Quadro 13: | Perfil do usuário | 75 |
| Quadro 14: | Requisitos mínimos de <i>backend</i> | 84 |
| Quadro 15: | Requisitos de banco de dados..... | 85 |
| Quadro 16: | Orçamento..... | 92 |
| Quadro 17: | Cronograma | 93 |

LISTA DE SIGLAS

| | |
|----------|--|
| APCs | <i>Article Processing Charges</i> |
| API | <i>Application Programming Interface</i> |
| BBM | Biblioteca Brasileira Guita e José Mindlin |
| BDTC | Biblioteca Digital de Trabalhos Científicos |
| BOIA | <i>Budapest Open Access Initiative</i> |
| BPMN | <i>Business Process Model and Notation</i> |
| BSD | <i>Berkeley Software Distribution</i> |
| FINEP | Financiadora de Estudos e Projetos |
| Fiocruz | Fundação Oswaldo Cruz |
| GPL | <i>GNU General Public License</i> |
| HMR | Hot Module Replacement |
| HTML | <i>HyperText Markup Language</i> |
| IBICT | Instituto Brasileiro de Informação em Ciência e Tecnologia |
| IR | <i>Institutional Repository</i> |
| ISO | <i>International Organization for Standardization</i> |
| NBR | Norma Brasileira |
| OA | <i>Open Access</i> |
| OAIS | <i>Open Archival Information System</i> |
| OAR | <i>Open Access Repositories</i> |
| OpenDOAR | <i>Directory of Open Access Repositories</i> |
| RDF | <i>Resource Description Framework</i> |
| ROAR | <i>Registry of Open Access Repositories</i> |
| ROARMAP | <i>Registry of Open Access Repositories Mandatory Archiving Policies</i> |

| | |
|--------|------------------------------------|
| SETREM | Sociedade Educacional Três de Maio |
| URI | <i>Uniform Resource Identifier</i> |
| USP | Universidade de São Paulo |

SUMÁRIO

| | |
|---|----|
| INTRODUÇÃO | 9 |
| 1 PLANO DE ESTUDO E PESQUISA | 11 |
| 1.1 TEMA | 11 |
| 1.1.1 Delimitação do Tema | 11 |
| 1.2 OBJETIVO GERAL | 11 |
| 1.2.1 Objetivos Específicos | 12 |
| 1.3 JUSTIFICATIVA | 12 |
| 1.4 PROBLEMA | 14 |
| 1.5 HIPÓTESES | 14 |
| 1.6 METODOLOGIA | 14 |
| 1.6.1 Abordagem | 15 |
| 1.6.2 Procedimentos | 15 |
| 1.6.2.1 Pesquisa Bibliográfica | 16 |
| 1.6.2.2 Pesquisa Ação | 16 |
| 1.6.3 Validação das Hipóteses | 16 |
| 1.6.4 Técnicas | 18 |
| 1.7 ORÇAMENTO | 18 |
| 1.8 CRONOGRAMA DE ATIVIDADES | 19 |
| 2 FUNDAMENTAÇÃO TEÓRICA | 20 |
| 2.1 ÁREA DO CONHECIMENTO | 20 |
| 2.1.1 Gestão do Conhecimento | 20 |
| 2.1.2 Repositórios Institucionais | 21 |
| 2.1.3 Padrão Dublin Core para metadados descritivos | 22 |
| 2.1.4 Open Archival Information System (OAIS) | 23 |
| 2.2 PROCESSOS | 25 |
| 2.2.1 Engenharia de Software | 26 |
| 2.2.2 Prototipação de Software | 27 |
| 2.2.3 Gestão Eletrônica de Documentos | 28 |
| 2.3 TECNOLOGIAS | 29 |
| 2.3.1 Cloud Computing | 29 |
| 2.3.2 Object Storage | 32 |
| 2.3.3 Microsoft .NET | 33 |

| | | |
|---------|--|----|
| 2.3.4 | React | 34 |
| 2.3.5 | Vite | 35 |
| 2.3.6 | PostgreSQL | 37 |
| 2.3.6.1 | <i>Full-Text Search</i> | 38 |
| 2.3.6.2 | <i>GIN Index</i> | 40 |
| 2.4 | SISTEMAS RELACIONADOS | 41 |
| 2.5 | TRABALHOS RELACIONADOS | 43 |
| 2.5.1 | BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados | 43 |
| 2.5.2 | Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência | 44 |
| 2.5.3 | Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital | 45 |
| 2.5.4 | Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca | 46 |
| 2.5.5 | O mapeamento dos repositórios institucionais brasileiros: perfil e desafios | 47 |
| 2.5.6 | Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis | 49 |
| 2.5.7 | The Emergence of Institutional Repositories: A Conceptual Understanding of Key Issues through Review of Literature | 50 |
| 2.5.8 | Why So Many Repositories? Examining the Limitations and Possibilities of The Institutional Repositories (IR) Landscape | 52 |
| 2.5.9 | Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS | 53 |
| 2.5.10 | Using Open Access Institutional Repositories to Save the Student Symposium during the COVID-19 Pandemic | 54 |
| 2.5.11 | Understanding Institutional Repository in Higher Learning Institutions: A Systematic Literature Review and Directions for Future Research | 56 |
| 2.5.12 | Discovery Tools to Enhance Resources Findability in the Institutional Repositories: an overview | 57 |
| 2.5.13 | Crafting Linked Open Data to Enhance the Discoverability of Institutional Repositories on the Web | 59 |
| 2.5.14 | Open Access Scientific Digital Repositories : An Analytical Study of the Open DOAR | 60 |
| 2.5.15 | Evolution of institutional repositories: Managing institutional research output to remove the gap of academic elitism | 61 |
| 3 | RESULTADOS | 65 |
| 3.1 | CONCEPÇÃO DO REPOSITÓRIO INSTITUCIONAL | 65 |
| 3.2 | BACKLOG DO PRODUTO | 66 |
| 3.3 | ARQUITETURA DA APLICAÇÃO | 75 |
| 3.3.1 | Estrutura de Backend | 76 |
| 3.3.2 | Estrutura de Frontend | 77 |
| 3.4 | ANÁLISE DOS RESULTADOS | 77 |
| 3.4.1 | Preparação da base de dados | 77 |

| | | |
|--------------------|--|----|
| 3.4.2 | Teste de Consulta por Publicações | 79 |
| 3.4.3 | Teste de Submissão de Publicações | 81 |
| 3.4.4 | Requisitos Mínimos do Sistema | 83 |
| <i>3.4.4.1</i> | <i>Frontend</i> | 83 |
| <i>3.4.4.2</i> | <i>Backend</i> | 84 |
| <i>3.4.4.3</i> | <i>Banco de dados</i> | 85 |
| REFERÊNCIAS | | 86 |
| APÊNDICES | | 91 |
| APÊNDICE A | | 92 |
| APÊNDICE B | | 93 |

INTRODUÇÃO

Os repositórios acadêmicos dentro de instituições de ensino podem ser vistos como uma ferramenta para a preservação e propagação do conhecimento, de forma a reunir o acervo das pesquisas realizadas durante a passagem dos acadêmicos pela instituição, disponibilizando fácil acesso a este material.

Para Lynch (2003) um repositório institucional pode ser definido como um conjunto de serviços disponibilizados por uma universidade aos seus membros, visando a disseminação dos materiais digitais produzidos pela instituição, e pelos membros de sua comunidade.

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 foi definido como o desenvolvimento de um repositório institucional de pesquisas acadêmicas baseado em nuvem, sendo 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.

A pesquisa possui como objetivo desenvolver um repositório institucional de pesquisas acadêmicas projetado para ser executado em ambientes baseado em nuvem, utilizando de tecnologias como React e Vite para o frontend, Microsoft

.NET 7.0 para a criação de APIs, PostgreSQL como banco de dados, e recursos de armazenamento de objetos baseados em nuvem.

A plataforma também possui como premissa envolver recursos referentes 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 expor a importância da utilização de repositórios acadêmicos para realizar a preservação digital dos materiais, 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 trabalho.

No capítulo 3 são apresentados os resultados obtidos com a realização desta pesquisa, protótipos e apresentação do repositório acadêmico proposto. Em última parte, é apresentada 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 *software* de repositório institucional de pesquisas preparado para ambientes baseados em nuvem.

1.1.1 Delimitação do Tema

A delimitação do tema se dará como o desenvolvimento de uma aplicação web de repositório institucional, que envolva o processo de envio de correções e revisões das publicações dentro da própria plataforma, e que seja preparado desde a sua concepção para ser executado em ambientes baseados em nuvem.

Para a criação do repositório institucional será utilizando de tecnologias como React e Vite para o frontend, Microsoft .NET 7.0 para a criação de APIs, PostgreSQL como banco de dados, e recursos de armazenamento baseados em nuvem como o Amazon S3 ou Digital Ocean Spaces para o armazenamento de arquivos.

O desenvolvimento deste relatório de pesquisa foi realizado durante o período de maio a dezembro 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 pesquisas acadêmicas baseado em nuvem, visando 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 e pesquisar sobre trabalhos e artigos relacionados.
2. Pesquisar sobre formas de armazenamento de documentos em nuvem e mecanismos de busca e análise textual de documentos.
3. Elicitar os requisitos para o desenvolvimento do repositório acadêmico e definir as tecnologias que serão utilizadas para o desenvolvimento (Ex: banco de dados, linguagens e frameworks).
4. Desenvolver uma aplicação web de repositório acadêmico, e popular o banco de dados com um *dataset* de publicações acadêmicas.
5. 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 institucional que possibilitasse aos acadêmicos 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 institucional acadêmico.

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 para criação de repositórios 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 de atualização 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 <https://v2.sherpa.ac.uk/opensoar/>

1.4 PROBLEMA

Para a escolha do problema desta pesquisa foi ressaltado como sendo o principal fator motivador, a necessidade de reunir as publicações acadêmicas produzidas dentro de instituições em um repositório institucional, que forneça rápido acesso às publicações científicas, tanto de forma interna para os estudantes da instituição, quanto de forma aberta ao público em geral.

Também foram levados em conta os fatores relatados nos trabalhos relacionados referentes a personalização, compatibilidade com dispositivos móveis e dificuldades na atualização entre versões dos *softwares open source* de repositório institucional existentes.

Sendo consideradas tais premissas, o problema desta pesquisa foi definido: Como projetar um *software* de repositório institucional de pesquisas acadêmicas, utilizando de tecnologias de nível empresarial, e que esteja preparado para operar em ambientes baseados em nuvem?

1.5 HIPÓTESES

1. O recurso de Full Text Search presente no banco de dados PostgreSQL pode ser utilizado como uma alternativa viável para realizar as consultas por publicações dentro do repositório acadêmico proposto, (menos de 1 segundo por consulta), mesmo em bases de dados com mais de 3.500 publicações.
2. O processo desenvolvido para extração do texto das publicações durante o auto arquivamento é rápido o suficiente para não necessitar de processamento em segundo plano (menos de 3 segundos), mesmo em publicações com mais de 12.000 palavras, cerca de 40 páginas de texto em português com fonte tamanho 12, considerando que cada página tenha 300 palavras.

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 so-

bre 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, 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 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 tópico 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 duas 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 1 segundos por consulta), mesmo em bases de dados com mais de 3.500 publicações".

Já a segunda hipótese é apresentada como "O processo desenvolvido para extração do texto das publicações durante o auto arquivamento é rápido o suficiente para não necessitar de processamento em segundo plano (menos de 3 segundos), mesmo em publicações com mais de 12.000 palavras, cerca de 40 páginas de puro texto em português com fonte tamanho 12, considerando que cada página tenha

300 palavras".

Ambas as hipóteses possuem o viés de medir o desempenho da aplicação desenvolvida. O tempo máximo de 1 segundo para consulta, e 3 segundos para uma publicação, foi escolhido com base na teoria de Nielsen (1994), que afirma que o tempo de resposta inferior a 0.1 segundos é considerado como uma reação instantânea pelo usuário, acima de 1 segundo o usuário já percebe o atraso, necessitando de um *feedback* visual, como uma barra de carregamento, e acima de 10 segundos o usuário provavelmente perderá o interesse e vá realizar outras atividades enquanto aguarda.

Para realizar a validação da primeira hipótese será utilizado do *arXiv Dataset*², um *dataset* gratuito com mais de 1.7 milhões de artigos e publicações acadêmicas, para popular a base de dados do repositório institucional. Como o arXiv Dataset possui um grande número de registros, a base de dados será limitada a cerca de 3.500 publicações.

Após realizar a preparação da base de dados, será utilizado da ferramenta Apache JMeter³ para realizar um *benchmark* de desempenho sobre a aplicação. O *benchmark* proposto consiste em preparar uma consulta a API de pesquisa por publicações do repositório desenvolvido, que será executada por mil vezes. Com os resultados obtidos do *benchmark* será possível calcular o tempo médio da realização de uma consulta por publicações no sistema proposto.

A segunda hipótese também pode ser validada por meio de um teste de desempenho ou *benchmark*, que consiste realizar a publicação de uma mesma pesquisa com mais de 12.000 palavras diversas vezes, e mensurar o tempo médio para conclusão do processo.

Este *benchmark* também será realizado através da ferramenta Apache JMeter, por meio da preparação de uma requisição a API do repositório desenvolvido, que será repetida por mil vezes. Esta requisição realizará a publicação de um arquivo PDF contendo dados fictícios puramente textuais, visto que imagens não geram dados textuais a serem extraídos. Com os resultados obtidos através deste

²<https://www.kaggle.com/Cornell-University/arxiv>

³<https://jmeter.apache.org/>

benchmark será possível calcular o tempo médio da realização de uma publicação no sistema proposto.

Todos os testes realizados serão executados utilizando a versão 14 do PostgreSQL, sobre o sistema operacional Ubuntu 22.04 LTS, com as seguintes configurações de hardware: Intel Core i3 9100, 16GB RAM DDR4 2400MHz, 1TB SSD NVME WD Blue SN550.

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 das técnicas de programação para o desenvolvimento do sistema, e de testes para verificar o funcionamento do *software* desenvolvido, conforme descrito na validação da hipóteses.

1.7 ORÇAMENTO

O Quadro 16 - Apêndice A apresenta o orçamento do trabalho, demonstrando de maneira geral os gastos da realização desta pesquisa. O quadro é composto por colunas que descrevem o item, a quantidade, o valor unitário e o valor total.

1.8 CRONOGRAMA DE ATIVIDADES

O Quadro 17 - Apêndice B apresenta o cronograma de atividades, que demonstra quando cada atividade desta pesquisa foi realizada. O quadro é composto por colunas que descrevem a atividade, e quando ela será realizada, entre os meses de maio a novembro de 2022.

2 FUNDAMENTAÇÃO TEÓRICA

Ao longo deste capítulo é apresentado o embasamento teórico do relatório, o qual dispõe de pesquisas já realizadas por outros autores sobre os temas propostos e os conceitos que servirão de base para o desenvolvimento do atual trabalho.

2.1 ÁREA DO CONHECIMENTO

Os tópicos abordados nesta seção abordam questões relacionadas a área do conhecimento vinculada ao tema desta pesquisa, como a definição de gestão do conhecimento, repositórios institucionais, e padrões de arquivamento de publicações.

2.1.1 Gestão do Conhecimento

Em tradução de (Becerra-Fernandez; Sabherwal, 2014, p. 4) a Gestão de Conhecimento, em inglês *Knowledge Management* ou *KM*, pode ser definida como "fazer o que for necessário para tirar o máximo de proveito dos recursos de conhecimento disponíveis". A Gestão de Conhecimento pode ser aplicada para indivíduos, porém vem ganhando atenção por parte das organizações, pois é vista como uma disciplina que promove a criação, compartilhamento e aproveitamento do conhecimento.

Já em tradução de (AWAD, 2007, p. 26) a Gestão de Conhecimento é um modelo de negócios interdisciplinar emergente, que possui o conhecimento como sendo o centro da estrutura da organização. Possuindo raízes em várias disciplinas, incluindo economia empresarial, psicologia e gestão da informação, e envolvendo pessoas, tecnologias e processos.

Conforme o mesmo autor, a Gestão de Conhecimento também pode ser o processo de capturar e fazer uso do conhecimento coletivo da organização, sendo separado em dois tipos, o conhecimento explícito, que é documentado em papel ou em bases de dados, e o conhecimento tácito, que existe somente na mente dos indivíduos.

Seguindo a definição proposta por Becerra-Fernandez; Sabherwal (2014), a Gestão de Conhecimento possui foco em organizar e tornar o conhecimento disponível, onde e quando for necessário. Tendo sua ênfase em conhecimentos que já foram reconhecidos e articulados de alguma forma, incluindo o conhecimento sobre os processos de uma organização, procedimentos, propriedade intelectual, documentação de boas práticas, previsões, lições aprendidas, e soluções para problemas recorrentes.

Aos poucos a Gestão de Conhecimento também vem abordando a ênfase sobre documentar e organizar o conhecimento que muitas vezes só reside dentro das mentes dos colaboradores mais experientes da organização.

2.1.2 Repositórios Institucionais

Em tradução livre de Lynch (2003) repositórios institucionais podem ser definidos como sendo um conjunto de serviços que a universidade pode oferecer aos seus membros, visando o gerenciamento e disseminação dos materiais digitais produzidos tanto pela instituição quanto pelos membros de sua comunidade.

Já em tradução de Callicott, Scherer e Wesolek (2015) a iniciativa de repositórios institucionais consiste em um conjunto de serviços que possuem o objetivo de suportar a preservação, organização e acesso aos conhecimentos produzidos pela instituição. No sentido de infraestrutura de software, são desenvolvidos para solucionar alguns dos problemas que a comunicação acadêmica enfrenta na era digital, oferecendo a possibilidade de publicação imediata, preservação de longo prazo e acesso global as publicações.

Conforme o mesmo autor, a ideia geral de repositório institucional surgiu com o lançamento do repositório de disciplinas específicas arXiv em 1991. Em seu lançamento a comunicação eletrônica de literatura acadêmica foi rapidamente

adotada por físicos, e desde então expandiu-se para outras áreas correlatas da física e matemática, hospedando mais de um milhão de *e-prints*.

No início dos anos 2000, a ideia publicação imediata e aberta em repositório disciplinares começou a ser popularizada e aplicada a nível institucional. Em seguida em 2002, houve o lançamento da primeira versão oficial do *software open source* DSpace, junto a publicação da *Scholarly Publishing and Academic Resources Coalition* (SPARC), e do paper *The Case for Institutional Repositories* escrito por Raym Crow.

Conforme Callicott, Scherer e Wesolek (2015), este dois eventos serviram como base para *softwares* amplamente acessíveis para repositórios institucionais, além de vincular a ideia de que repositórios institucionais atuam sobre a visibilidade e prestígio da instituição.

2.1.3 Padrão Dublin Core para metadados descritivos

Em tradução livre de Caplan; American Library Association (2003) o padrão *Dublin Core Metadata Element Set* ou somente *Dublin Core* pode ser definido como um esquema de propósito geral, utilizado para a descrição de recursos e objetos na internet. O padrão consiste em um conjunto de 50 elementos, contendo ao menos um identificador e uma descrição de sua definição.

Já em tradução de Banerjee; Reese (2018) a história do padrão Dublin Core remonta a segunda edição da *International WWW Conference* em 1994. Surgindo de uma discussão sobre a dificuldade existente no período em encontrar materiais presentes na internet.

A partir desta discussão, em 1995 em Dublin, Ohio, a OCLC (*Online Computer Library Center*) e a NCSA (*National Center of Supercomputers Applications*) lideraram em conjunto um *workshop* que foi chamado de *OCLC/NCSA Metadata Workshop*, tendo como objetivo decidir quais elementos seriam necessários para promover a melhor encontrabilidade de documentos na internet, explorar soluções sobre como criar mecanismos que seriam flexíveis para publicações realizadas na internet no passado, presente e futuro, e como promover que tais soluções sejam utilizadas caso já existam.

Neste encontro, os participantes entraram em acordo sobre um conjunto elementos descritivos, dos quais 50 destes termos poderiam ser virtualmente aplicados a qualquer recurso presente na internet até o momento. Destes 50 termos, surgiu a iniciativa do padrão Dublin Core.

Quadro 1: Exemplo de elementos do Dublin Core

| Identificador | Definição |
|--------------------|---|
| <i>Title</i> | Nome do recurso. |
| <i>Creator</i> | Entidade responsável pela criação do conteúdo do recurso. |
| <i>Subject</i> | Assunto do conteúdo do recurso. |
| <i>Description</i> | Descrição do conteúdo do recurso. |
| <i>Publisher</i> | Entidade responsável por tornar o recurso disponível. |
| <i>Contributor</i> | Entidade responsável por realizar contribuições ao conteúdo do recurso. |
| <i>Date</i> | Data associada a um evento do ciclo de vida do recurso. |
| <i>Type</i> | A natureza ou gênero do conteúdo do recurso. |
| <i>Format</i> | O formato físico ou digital do recurso. |
| <i>Identifier</i> | O identificador único dado ao recurso dentro de um contexto. |
| <i>Source</i> | Uma referência ao recurso ao qual o presente recurso é derivado. |
| <i>Language</i> | O idioma em qual o conteúdo do recurso foi escrito. |
| <i>Relation</i> | Uma referência a um recurso relacionado. |
| <i>Coverage</i> | A extensão ou escopo do conteúdo do recurso. |

Fonte: Traduzido de (CAPLAN; American Library Association, 2003, p. 77)

O Quadro 1 apresenta um exemplo de elementos presentes no padrão Dublin Core, sendo todos os elementos opcionais e passíveis de repetição. O esquema em si é independente de um formato em específico, sendo possível apresentá-lo em diferentes formatos como XML e JSON.

Originalmente, o esquema foi produzido como um método para descrição de documentos na internet, como em *websites*, de forma a utilizar de *meta tags* presentes no cabeçalho dos documentos. Tais *tags* podem ser facilmente utilizadas por ferramentas de pesquisa para explorar e indexar documentos.

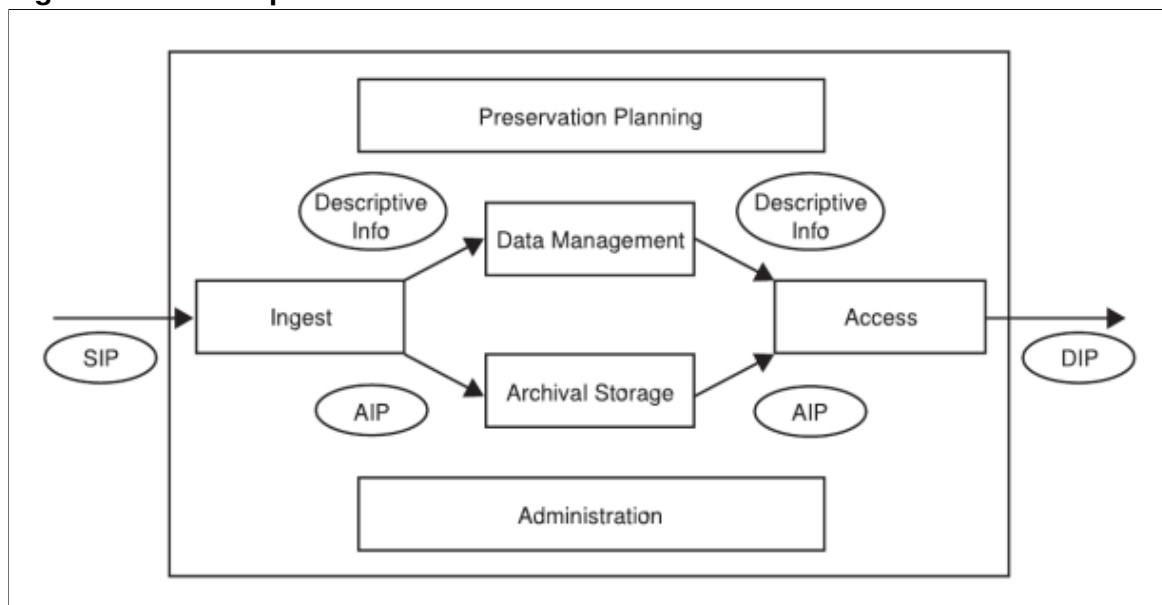
2.1.4 Open Archival Information System (OAIS)

Em tradução livre de Borghoff; Rödiger et al. (2007) o OAIS (*Open Archival Information System*) consiste em um modelo referência para a preservação a longo

prazo de publicações digitais. Este modelo possui como objetivo trazer clareza a termos básicos sobre a preservação a longo prazo, e identificar os processos chave correspondentes, sendo considerado uma base comum para a troca internacional de ideias e experiências na área.

Conforme o mesmo autor, em 1995 o *International Organization for Standardization* (ISO) direcionou ao *Consultive Committee for Space Data Systems* (CCSDS) o desenvolvimento deste padrão. O CCSDS consiste em um comitê subordinado a NASA que foi fundado em 1982 por várias agências espaciais, com o objetivo de simplificar a troca de pesquisas espaciais. Após vários anos de pesquisa, o CCSDS apresentou em maio de 1999 a primeira versão deste padrão, que ainda passaria por diversas revisões, sendo a revisão de fevereiro de 2003 aceita como o padrão ISO (14721:2003) (ISO 2003a) após revisado para a (ISO 14721:2012).

Figura 1: Fluxo de processo do modelo de referência OAIS



Fonte: Retirado de (Borghoff; Rödiger et al., 2007, p. 25)

A Figura 1 representa o fluxo do processo realizado no modelo de referência OAIS, neste fluxo são realizadas as etapas de *Ingest* (Ingestão), *Archival Storage* (Armazenamento de arquivo), *Data Management* (Gestão de dados), *Administration* (Administração), *Preservation Planning* (Planejamento de preservação) e *Access* (Acesso).

O processo de ingestão recebe um SIP (*Submission Information Package*), que consiste no objeto digital a ser armazenado junto aos seus metadados, e o prepara para os processos de armazenamento e administração, transformando o SIP em um AIP (*Archival Information Package*), que contém os metadados descritivos, dados técnicos, informações do projeto, direitos de acesso e uso, e até mesmo pode conter informações referentes a verificação por vírus no arquivo.

Após o processo de ingestão, o AIP gerado irá entrar no processo de Armazenamento de arquivo, que é responsável por distribuir o AIP a diferentes formas de armazenamento, sendo também responsável por garantir a integridade dos dados, e a permitir a reconstrução dos AIPs em caso de catástrofes.

O processo de gestão de dados possui como finalidade realizar o gerenciamento das informações descritivas, normalmente estes dados são armazenados em um banco de dados, e precisam ser atualizados e administrados. Este processo também é responsável pela execução das pesquisas pelos arquivos armazenados.

O processo de administração executa rotinas de regras de negócio sobre o arquivo a ser armazenado, sendo responsável por verificar os pré-requisitos para o envio dos documentos, e se os dados presentes no SIP atendem os padrões pré definidos.

A etapa de acesso consiste na interface gráfica que permite o usuário recuperar os arquivos. Já a etapa de planejamento de preservação envolve as questões referentes a estratégia utilizada para manter a preservação dos arquivos (que deve ser regularmente atualizada), englobando o monitoramento regular por falhas nos discos de armazenamento, verificação de integridade, e outras tarefas deste gênero.

2.2 PROCESSOS

Os tópicos abordados nesta seção abordam questões relacionadas aos processos vinculados ao desenvolvimento do repositório institucional proposto, como a engenharia de *software*, prototipação de *software* e a gestão eletrônica de documentos.

2.2.1 Engenharia de Software

Segundo Valente (2020) a área da engenharia de *software* se preocupa em propor e aplicar os princípios de engenharia na construção de *software*, visando a aplicação de abordagens sistemáticas, disciplinadas e quantificáveis, no processo de desenvolver, operar, manter e evoluir o *software*.

Já Wazlawick (2019) afirma que existem várias possíveis definições para a engenharia de *software*, porém aborda como sendo o processo de estudar, criar e otimizar os processos de trabalho para os desenvolvedores de *software*, envolvendo atividades como observar, avaliar, orientar e alterar os processos produtivos.

Seguindo a definição de Valente (2020) a área da Engenharia de Software teve as suas origens pelo final da década de 1960, quando os computadores passaram a ser mais populares, começando a estarem presentes em várias universidades norte americanas e europeias, além de grandes empresas. Nas décadas anteriores os primeiros computadores modernos já estavam sendo projetados, porém com o foco em construir máquinas para executar programas específicos, e não qualquer tipo de *software* em geral.

Com o progresso nas tecnologias de construção de *hardware*, o desenvolvimento de novos sistemas se tornavam possíveis e ganharam demanda, sendo necessário uma área da computação destinada especificamente a investigar e propor soluções para o desenvolvimento de *software*, que ficavam cada vez mais complexos. Esta nova área vinha a ser chamada de Engenharia de *Software*.

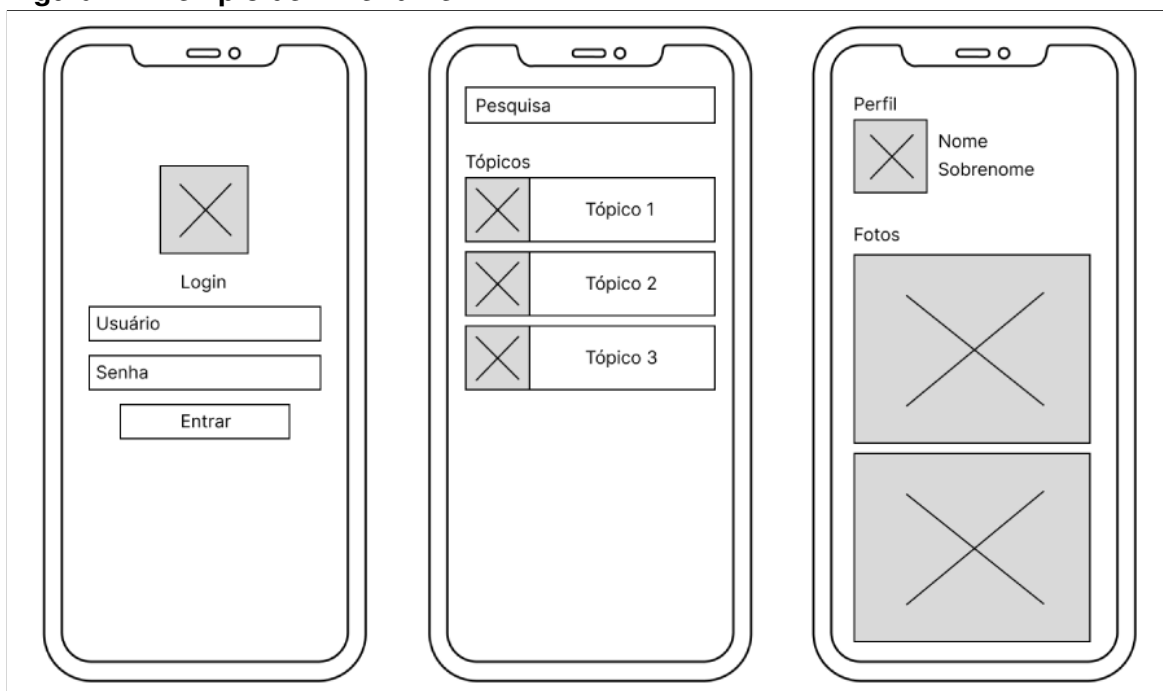
Em especial, o ano de 1968 é considerado o marco histórico em que a área de engenharia de *software* foi criada. Mais especificamente em outubro de 1968, quando um grupo de 50 cientistas da computação se reuniram durante uma semana em Garmisch na Alemanha, para participar de uma conferencia patrocinada pelo comité científico da OTAN, que possuía como objetivo chamar a atenção para a necessidade que *softwares* fossem construídos tendo como base princípios tanto práticos quanto teóricos, da mesma forma como ocorre em outras áreas da engenharia.

2.2.2 Prototipação de Software

Segundo Uzayr (2022) o processo de *design* de *software*, no sentido visual, chega as aplicações *web* por meio cinco de etapas, que podem ser definidas como a idealização, *wireframing*, *mockup*, prototipação e implementação.

A etapa de *wireframing* consiste na criação de um *layout* bruto para uma página, visando o conceito elaborado durante a etapa de idealização do sistema, e usando da teoria do *design* para atingir estes objetivos. Como visível na Figura 2, em geral os *wireframes* são criados em preto e branco e visam expressar aspectos mais funcionais do que visuais. Uma analogia a um *wireframe* seria como uma planta baixa de uma casa.

Figura 2: Exemplo de Wireframe



Fonte: Adaptado de Uzayr (2022)

A etapa de *mockup* é construída com base nos resultados obtidos após o processo de *wireframing*. Os *mockups* são mais robustos, possuem cores, estão voltados a aspectos mais visuais do que funcionais, e trazem substância ao *wireframe* criado na etapa anterior. Ainda sendo apresentados como imagens estáticas, os *mockups* permitem que tanto os *designers* quanto os *stakeholders* visualizem como o resultado final ira ficar, a tempo de realizar sugestões para a mudança de cores, imagens e estilos.

Já a etapa de prototipação possui como objetivo simular a maior parte das funcionalidades presentes no *website*, de forma a entregar aos *stakeholders* uma experiência de usuário (UX) muito semelhante ao resultado final. Em analogia, um *mockup* funciona como uma imagem estática de uma página, enquanto um protótipo funciona como uma simulação do *website* final, permitindo a experiência de interagir e navegar pelas páginas antes do sistema ser de fato desenvolvido.

2.2.3 Gestão Eletrônica de Documentos

Conforme Anjos (2020) o processo de Gestão Eletrônica de Documentos (GED) surgiu como uma solução para uma série de problemas relacionados ao crescimento da geração de documentos físicos pelas empresas, e a necessidade da recuperação rápida das informações. Mundialmente o GED é conhecido como Enterprise Document Management (EDM), e aborda um conjunto de tecnologias utilizadas para os processos de conversão de documentos físicos para digitais (*Document Imaging*), e gerenciamento de fluxos de armazenamento e recuperação dos dados (*Workflow*).

Para AVEDON (2002), o Gerenciamento Eletrônico de Documentos (GED) pode ser caracterizado como uma configuração de equipamentos, software, e recursos de telecomunicação, que automatizam o processo de armazenamento de imagens de documentos, que podem ser lidos por máquinas e recuperados sob solicitação.

Já para CONARQ (2022) o GED consiste em um conjunto de tecnologias utilizadas para a organização da informação não estruturada (aquela que não está armazenada em bancos de dados, como arquivos de texto, imagens, sons e planilhas) de um órgão ou entidade, abordando funcionalidades como a captura, gerenciamento, armazenamento e distribuição. Também podendo envolver tecnologias referentes a digitalização, automação de fluxos de trabalhos (*workflows*), processamento de formulários, indexação, gestão de documentos e repositórios.

O CONARQ (Conselho Nacional de Arquivos) disponibiliza de forma gratuita em seu site o livro e-ARQ Brasil, que define modelos de requisitos para sistemas informatizados de gestão arquivística de documentos. Este livro pode ser

visto como uma leitura essencial para quem deseja ingressar na área de desenvolvimento de sistemas de GED, abordando desde a base conceitual, até requisitos funcionais, não funcionais e questões referentes a metadados dos documentos armazenados.

2.3 TECNOLOGIAS

Os tópicos abordados nesta seção abordam questões relacionadas as tecnologias utilizadas para o desenvolvimento do repositório institucional proposto, como a definição de *Cloud Computing*, *Object Storage*, e outras ferramentas utilizadas para o desenvolvimento do *frontend* e *backend* da aplicação.

2.3.1 Cloud Computing

Em tradução de Sehgal, Bhatt e Acken (2019) o NIST (*National Institute of Standards and Technology*) dos Estados Unidos define o termo *Cloud Computing* como sendo "um modelo para permitir o acesso em rede de forma onipresente, conveniente e sob-demanda a um conjunto de recursos computacionais configuráveis (como redes, servidores, armazenamento, aplicativos e serviços) que podem ser rapidamente provisionados e liberados com o mínimo de esforço do provedor de serviços".

Em sua definição o NIST apresenta que este modelo é composto por cinco características essenciais, três modelos de serviço e quatro modelos de implantação. Dentre as características essenciais estão:

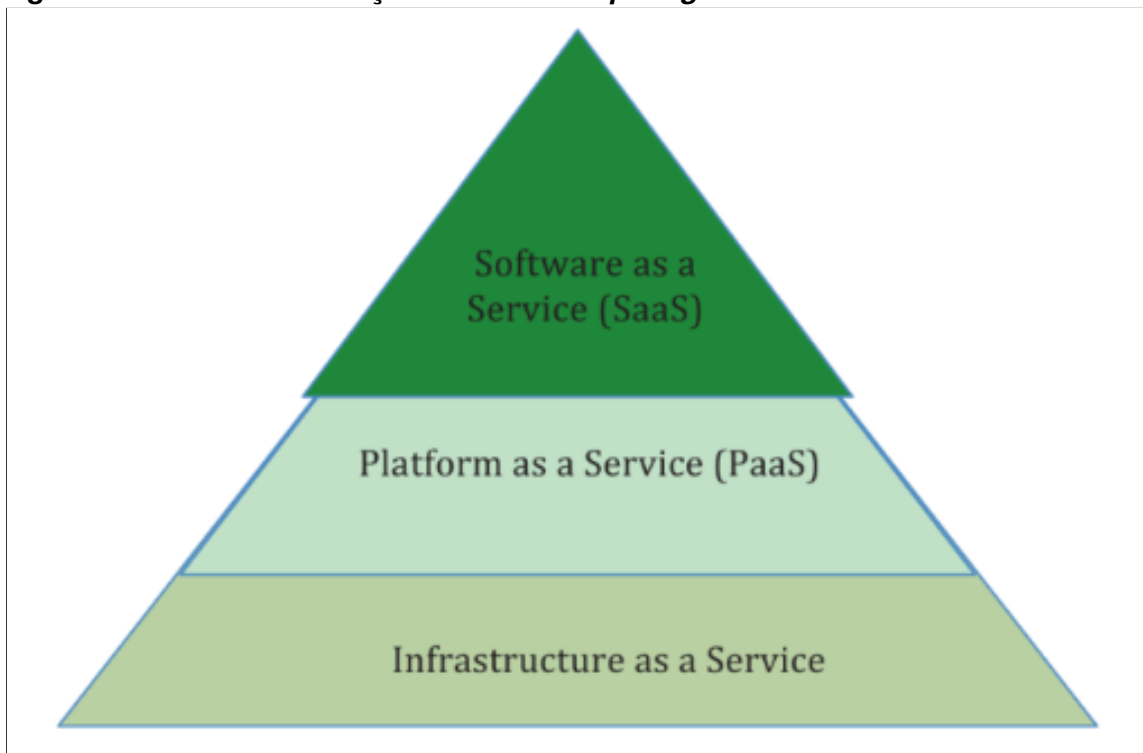
1. *On-demand self-service*: capacidade do consumidor por alterar os recursos de computação conforme o necessário, sem a necessidade de intervenção do prestador de serviço.
2. *Broad network access*: os recursos estão disponíveis em rede, e podem ser acessados por meio de mecanismos padrões por diferentes dispositivos de diversas plataformas (Ex: *smartphones*, *laptops*, *tablets* e *workstations*).
3. *Resource pooling*: os recursos computacionais do provedor de serviços podem ser disponibilizados a múltiplos consumidores por meio de um modelo

multi-tenant, onde diferentes recursos físicos e virtuais são realocados conforme a demanda do consumidor.

4. *Rapid elasticity*: os recursos computacionais são distribuídos de forma elástica, podendo automaticamente ser incrementados ou reduzidos conforme a demanda. Para o consumidor final os recursos computacionais aparentam ser ilimitados, e podem ser escalados a qualquer quantidade a qualquer momento.
5. *Measured service*: Os recursos computacionais são monitorados, controlados, e reportados, de forma a prover transparência tanto para o provedor de serviços, quando para o consumidor. A utilização dos recursos computacionais é controlada e otimizada com base nas métricas coletadas sobre algum nível de abstração.

A Figura 3 representa os três diferentes modelos de serviços de *Cloud Computing* apresentados pelo NIST.

Figura 3: Modelos de serviço de *Cloud Computing*



Fonte: Retirado de (SEHGAL; BHATT; ACKEN, 2019, p. 51)

1. *Software as a Service (SaaS)*: Capacidade provida ao consumidor de utilizar

de aplicações que estão sendo executadas em uma infraestrutura em nuvem. O consumidor não possui acesso ou controle a infraestrutura em nuvem, apenas com a possibilidade de realizar configuração limitadas e específicas a aplicação em execução.

2. *Platform as a Service (PaaS)*: Capacidade provida ao consumidor de hospedar suas próprias aplicações, desenvolvidas ou adquiridas, utilizando de linguagens, bibliotecas, serviços ou ferramentas que são suportadas pelo provedor de nuvem. Neste modelo o consumidor não possui controle sobre a infraestrutura de nuvem, porém possui controle sobre a aplicação que está sendo hospedada.
3. *Infrastructure as a Service (IaaS)*: Capacidade provida ao consumidor de operar sobre os recursos de processamento, armazenamento, rede e outros recursos computacionais, podendo de forma arbitrária hospedar e executar *softwares* incluindo sistemas operacionais. Neste modelo o consumidor não possui controle a infraestrutura física da nuvem, porém possui controle sobre o sistema operacional que deseja utilizar, recursos de armazenamento, e até controle limitado a recursos de rede, como *firewalls*.

Já referente aos quatro modelos de implantação de nuvem definidos pelos NIST, é possível citar:

1. *Private cloud*: a infraestrutura de nuvem é provida para uso exclusivo de uma única organização, que pode possuir vários consumidores. Esta infraestrutura é adquirida, gerenciada e operada pela própria organização, um terceiro, ou uma combinação de ambos.
2. *Community cloud*: a infraestrutura de nuvem é provida para uso exclusivo de uma comunidade específica de consumidores possuem preocupações ou necessidades compartilhadas (como quesitos ou políticas de segurança). A infraestrutura é adquirida, gerenciada e operada por um ou mais membros da comunidade, um terceiro, ou uma combinação de ambos.

3. *Public cloud*: a infraestrutura de nuvem é provida para o uso livre do público em geral, podendo ser adquirida, gerenciada e operada por uma empresa, ou uma organização acadêmica ou governamental.
4. *Hybrid cloud*: a infraestrutura de nuvem é composta por dois ou mais provedores de nuvem distintos (*private*, *community* ou *public*), sendo unificados por meio de uma tecnologia aberta ou proprietária que permite a portabilidade das aplicações.

2.3.2 Object Storage

Em tradução livre de Patil; Rangarao et al. (2020) o *Object Storage* também conhecido como *Object Based Storage* (OBS), consiste em uma tecnologia de armazenamento de dados, sendo uma progressão lógicas aos conceitos de *Block Storage* e *File Storage*.

Dentre as suas premissas, o *Object Storage* surgiu com o objetivo resolver uma série de problemas e desafios encontrados em outros modelos de armazenamento, como por exemplo a dificuldade de realizar o gerenciamento de dados em grande escala, a complexidade em gerenciar os *namespaces* em escala, e a aplicação de esquemas de segurança complexos sem a necessidade de envolvimento humano.

Conforme o mesmo autor, dentre as suas características são utilizados de três conceitos base para realizar o armazenamento e referência aos dados, sendo o primeiro o *Data* que consiste no conteúdo que será armazenado, podendo ser no formato de um texto, binário, ou qualquer formato conhecido. Como segundo conceito está o *Metadata*, que é composto por uma série de atributos descritivos do conteúdo que será armazenado. E como último conceito está o *Unique Identifier* (UUID), que consiste em uma chave única utilizada para referenciar o objeto, podendo ser utilizada para realizar a manipulação ou recuperação dos dados.

Dentre os demais conceitos utilizados no *Object Storage*, estão os *Buckets* e as *Access Control Lists* (ACLs), sendo um *Bucket* uma abstração lógica do local onde os dados serão armazenados, não estando realmente endereçado a um conjunto físico de discos, e podendo inclusive estar espalhado por diferentes regiões

pelo mundo. Já as ACLs consistem em um conjunto de regras de segurança e acesso, que podem ser vinculadas tanto a objetos individuais quanto a *Buckets* inteiros.

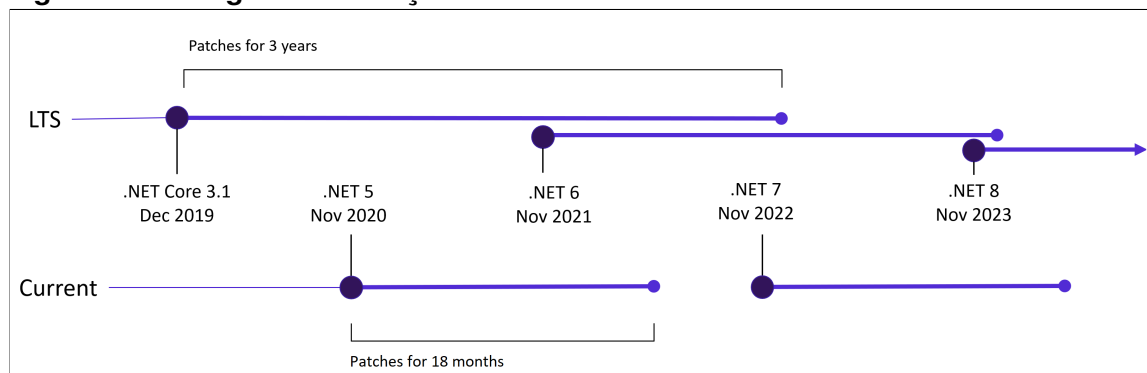
Neste modelo de armazenamento, o acesso aos dados dos arquivos é realizado por meio de uma interface REST, sobre o protocolo HTTP, por meio dos métodos *PUT*, *GET*, *DELETE* e *LIST*, permitindo a recuperação de qualquer dado a qualquer momento por meio de seu identificador único (UUID).

2.3.3 Microsoft .NET

Em tradução livre de Akella; Tamirisa et al. (2022), em 2002 a Microsoft anunciou a primeira versão do .NET Framework, uma plataforma de desenvolvimento de aplicações *web* e *desktop*. O .NET Framework oferece vários serviços, incluindo a execução de código gerenciado, um vasto conjunto de APIs, gerenciamento de memória, um sistema de tipagem comum, e vários frameworks de desenvolvimento como o ADO.NET, ASP.NET, WCF, WinForms e WPF (Windows Presentation Framework), além do suporte a diferentes linguagens de programação como C#, F# e Visual Basic. Inicialmente o .NET Framework era distribuído como um instalador a parte, porém com o tempo a Microsoft passou a embutir junto ao sistema operacional Windows, sendo o .NET Framework 4.8 a última versão lançada.

Todavia, em 2014 a Microsoft anunciou uma implementação *open source* e multi plataforma do .NET, denominada de .NET Core, sendo totalmente reescrita para ser executada em múltiplas plataformas como o Linux, macOS e Windows. O .NET Core é rápido, modular e permite que diferentes versões sejam executadas ao mesmo tempo na mesma máquina, sem afetar a aplicação.

Em 2020 após o lançamento de 3 versões do .NET Core (1.0, 2.0 e 3.0), e outras versões intermediárias, a Microsoft decidiu por unificar o .NET Framework e o .NET Core em uma única plataforma denominada somente de .NET, pulando a versão 4.0 do .NET Core para não causar confusões de nomenclatura com o antigo .NET Framework 4.8.

Figura 4: Cronograma de lançamento de versões do .NET

Fonte: Retirado de <https://dotnet.microsoft.com/en-us/platform/support/policy/dotnet-core>

A Figura 4 demonstra o cronograma de lançamento de novas versões do .NET, sendo todas a partir do .NET Core 3.1 sendo denominadas apenas de .NET seguido do número da versão. A versão mais recente até a realização deste trabalho consiste no .NET 7.0, que possui um ecossistema vasto para o desenvolvimento de aplicações *web*, em nuvem, *desktop*, *IoT* e *mobile*.

O .NET 7.0 foi escolhido como plataforma para o desenvolvimento da API do repositório institucional proposto, por meio da linguagem de programação C#, visto que consiste em uma plataforma popular, *open source*, e multi plataforma, e que possui um grande número de bibliotecas disponíveis por meio do repositório de pacotes Nuget¹.

2.3.4 React

Em tradução direta de (THAKKAR, 2020, p. 30), React ou React.js consiste em uma biblioteca JavaScript de código aberto, criada em maio de 2013 pelo Facebook, sendo utilizada para construção de interfaces de usuário.

Dentre suas principais características, o React é declarativo, ou seja, apenas é necessário criar a camada de visão, e o próprio React irá gerenciar de forma eficiente os estados da aplicação, renderizando somente os componentes necessários, e os atualizando quando ocorre alguma mudança de estado, por meio de uma abordagem de reconciliação denominada *Virtual DOM*. O React também é baseado em componentes, de forma que cada elemento visual da aplicação pode ser encapsulado em um componente, que possui seus próprios controles de estado, e

¹ <https://www.nuget.org/>

que podem ser reutilizados.

Para realizar a criação de uma página ou componente no React, é utilizado a sintaxe JSX, que consiste em uma extensão do JavaScript que permite juntar em um único arquivo a parte lógica da aplicação, ao código que será renderizando em tela. A sintaxe JSX não é obrigatória, porém é recomendada pelo própria React.

Figura 5: Exemplo de sintaxe JSX

```
1  function formatName(user) {  
2    return user.firstName + ' ' + user.lastName;  
3  }  
4  
5  const user = {  
6    firstName: 'Harper',  
7    lastName: 'Perez'  
8  };  
9  
10 const element = (  
11   <h1>  
12     Hello, {formatName(user)}!  
13   </h1>  
14 );
```

Fonte: Retirado de reactjs.org/docs/introducing-jsx.html

A Figura 5 demonstra um exemplo de código em JSX, onde é utilizado uma função para realizar a formatação de um texto contendo um nome e sobrenome, que será renderizando na interface do usuário dentro de uma tag HTML.

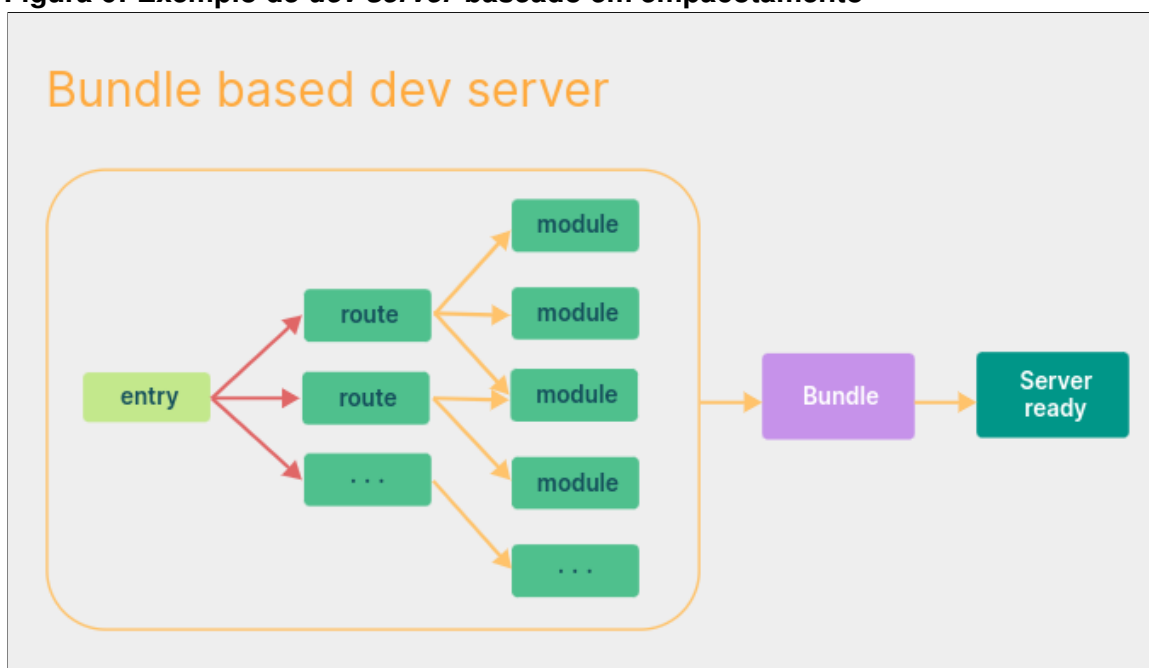
O React foi escolhido como ferramenta para realizar o desenvolvimento da interface gráfica do repositório institucional proposto, pois é uma biblioteca gratuita e de código aberto, e extramente popular tanto no mercado de trabalho quanto no meio acadêmico, além de contar com um vasto ecossistema de bibliotecas e componentes desenvolvidos pela comunidade.

2.3.5 Vite

Conforme VITE (2022) o site oficial da ferramenta, o Vite pode ser definido como sendo uma *build tool* (Ferramenta de Compilação), que possui como objetivo prover uma experiência de desenvolvimento mais rápida e leve para projetos *web* modernos, podendo ser utilizado em conjunto a diferentes ferramentas de desenvolvimento *frontend* como o Vue, React, Preact, Lit, Svelte ou Vanilla.

Dentre os principais diferenciais do Vite, é possível citar a utilização do *native ES modules* no lugar das tradicionais ferramentas de empacotamento de código como o Webpack, Rollup ou Parcel. Desta forma o Vite é capaz de fornecer um ambiente de desenvolvimento com o HMR (*Hot Module Replacement*, que pode ser definido como a capacidade de atualizar elementos visuais em tela durante o desenvolvimento de uma aplicação, sem a necessidade de recarregar a página ou recompilar o projeto) extramente rápido, e que tende não ficar lento com o crescimento do projeto.

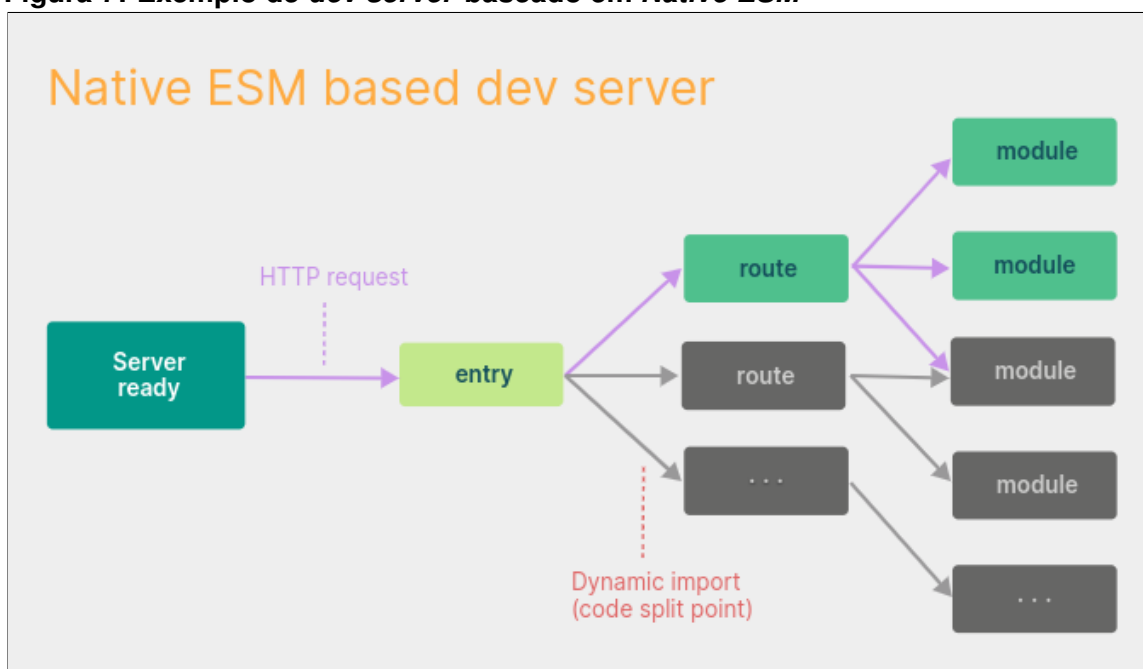
Figura 6: Exemplo de *dev server* baseado em empacotamento



Fonte: Retirado de vitejs.dev/guide/why.html

Na Figura 6 é demonstrado o funcionamento de um *dev server* (servidor de desenvolvimento) baseado no modelo tradicional de empacotamento de código. Neste é possível verificar que ao realizar a alteração em um arquivo, o *dev server* irá realizar novamente a compilação de todos os arquivos do projeto, para gerar o pacote de código que será carregado na interface, esta etapa de empacotamento pode levar um tempo considerável em projetos grandes, onde existem muitos arquivos para serem empacotados.

Figura 7: Exemplo de *dev server* baseado em *Native ESM*



Fonte: Retirado de vitejs.dev/guide/why.html

Já a Figura 7 representa o funcionamento de um *dev server* baseado no modelo de *Native ES Modules*, onde é possível verificar que não existe uma etapa de empacotamento, o código fonte é fornecido diretamente ao navegador, que fica responsável por carregar o código a medida do necessário, realizado a importação de dependências de formas dinâmica para a página que está sendo exibida.

O Vite foi escolhido como ferramenta de *build tool* para o *frontend* do repositório institucional proposto, e será utilizada no lugar da *build tool* padrão do React que utiliza o Webpack para realizar o empacotamento de código, tendo em vista os benefícios do Vite a longo prazo prevendo o crescimento da aplicação. Também vale ressaltar que não é necessário realizar adaptações de código específicas para o Vite, sendo possível trocar para outra ferramenta a qualquer momento sem grandes dificuldades.

2.3.6 PostgreSQL

Conforme Carvalho (2017) O PostgreSQL é um Sistema Gerenciador de Banco de Dados (SGBD) relacional, de código aberto, possuindo a versão 6.0 como sua primeira versão oficial liberada em janeiro de 1997. Hoje sendo uma alternativa de banco de dados popular, de nível corporativo, e com mais de 15

anos de desenvolvimento ativo. O PostgreSQL possui uma arquitetura que entrega confiabilidade, integridade, conformidade a padrões, desempenho e uma grande quantidade de recursos.

Dentre os seus recursos está o suporte a vários tipos de dados sofisticados, como o JSON, XML, objetos geométricos e geoespaciais, hierarquias, tags e matrizes. Além de permitir a inclusão de novos tipos de dados e funções que podem ser escritas em linguagens suportadas como o SQL, C, Python, Perl, dentre outras.

Como principal recurso que motivou a escolha do PostgreSQL como banco de dados para o repositório institucional proposto neste trabalho, está o suporte ao recurso de *full-text search*, que permite uma busca e indexação aprimorada por textos de grande tamanho na base de dados, que neste contexto será utilizado para armazenar e indexar os artigos e publicações presentes no repositório.

2.3.6.1 Full-Text Search

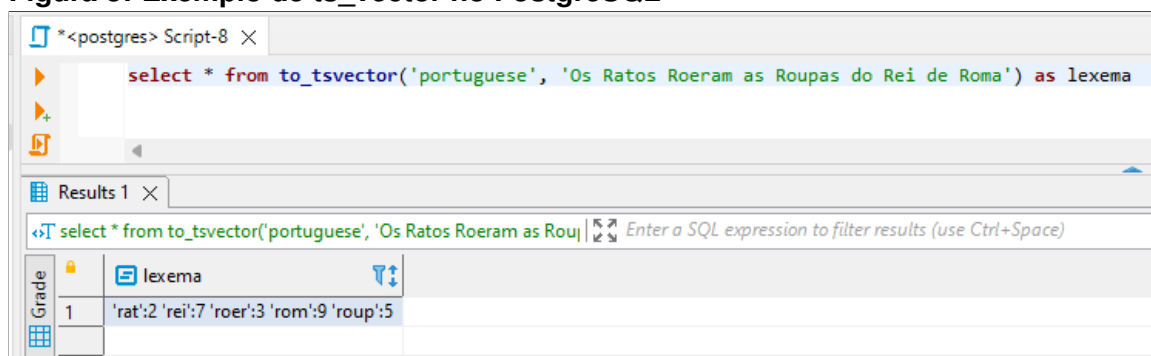
Em tradução livre de PostgreSQL (2022) o recurso de *Full Text Search* ou Pesquisa Textual Completa, fornece a capacidade de identificar linguagem natural em documentos que satisfazem um critério de busca, podendo de forma opcional ordená-los por relevância aos termos pesquisados. Para realizar este tipo de pesquisa, são utilizados os conceitos de critério de busca ou *query*, e o conceito de similaridade. De forma simplificada é possível definir uma *query* como um conjunto de palavras, e a similaridade como sendo a frequência que estas palavras sem repetem no documento.

Diferente dos operadores de busca mais comuns presentes em bancos de dados relacionais como o *like* e *ilike*, o recurso de *Full Text Search* fornece suporte a um contexto linguístico, permitindo por exemplo, a busca por palavras derivadas como no caso de *satisfies* e *satisfy* do inglês, algo que seria extramente difícil de mapear somente com o uso de expressões regulares. Este recurso também fornece capacidade de indexação para uma pesquisa mais rápida, e ranqueamento dos resultados para que os resultados mais relevantes apareçam no topo da pesquisa.

Conforme o mesmo autor, para que este tipo de pesquisa seja possível, o texto presente nos documentos precisam passar por uma etapa de pré-processamento, chamado de extração de dados ou *Tokenization*, onde os textos são convertidos em *tokens* que podem ser separados em diferentes classes, como palavras, números, endereços de email, etc. Após este processamento, os *tokens* podem ser convertidos em *lexemas* ou *lexemes*, que são semelhantes aos *tokens*, porém em um formato normalizado.

Como demonstrado na Figura 8 este processo de normalização envolve deixar todos o texto em minuscuro, remoção de sufixos, pluralidade, e em alguns casos até o gênero das palavras, permitindo que em uma única busca seja possível encontrar por todas as diferentes variantes de uma palavra. Este processo também elimina as chamadas *stop words*, palavras comuns que não possuem relevância dentro do texto, e que não precisam ser indexadas, como artigos, preposições, pronomes, conjunções, etc.

Figura 8: Exemplo de ts_vector no PostgreSQL



Fonte: Adaptado de PostgreSQL (2022)

No banco de dados PostgreSQL tais *lexemas* podem ser armazenados utilizando o tipo de dado denominado de *tsvector*, que possui um conjunto próprio de funcionalidades para pesquisa textual, sendo o mais comum o operador `@@` que permite a busca por textos que contemplem um critério de busca, também existe o suporte a pesquisa com operadores lógicos como *and*, *or* e *not*, e a utilização da função *to_tsquery* para converter um texto simples para um *query* de pesquisa.

2.3.6.2 GIN Index

Em tradução livre de PostgreSQL (2022) o termo GIN consiste em uma abreviação de *Generalized Inverted Index* ou Índice Generalizado Invertido em português, também conhecido pelo nome de Lista Invertida, um tipo de índice de banco de dados projetado para casos onde os itens a serem armazenados consistem em valores compostos, e as pesquisas realizadas devem ser feitas em cada item do conjunto. Por exemplo, os itens a serem armazenados podem ser documentos de texto, e a pesquisa a ser realizada pode ser feita por documentos que contenham uma palavra em específico.

Conforme o mesmo autor o índice GIN armazena um conjunto de pares de chaves (chave e lista de ocorrências), onde a lista de ocorrências é um conjunto de identificadores de linhas onde a chave/valor foi encontrada. Cada chave é armazenada somente uma única vez, fazendo com que o índice tenha um tamanho muito compacto, e ao realizar a pesquisa o índice utilizará o valor chave, no lugar de seu valor real.

Já em tradução de PostgreSQL Global Development Group (2011), o índice GIN não apresenta perdas para grande parte das consultas, porém sua performance depende de forma logarítmica ao número de palavras únicas no texto. Em comparação ao índice GiST (Generalized Search Tree) que pode ser utilizado por diferentes tipos de dados, o índice GIN apenas pode ser utilizado em dados do tipo *tsvector*, e alguns quesitos em relação ao desempenho devem ser considerados ao trabalhar com este tipo de índice.

Como por exemplo segundo o mesmo autor, o índice GIN pode apresentar uma consulta até 3 vezes mais rápidas que o índice GiST, porém o tempo necessário para sua indexação pode ser até 3 vezes mais longo. Os índices GIN podem ser mais lentos de atualizar (cerca de 10 vezes mais lentos do que os índices GiST), se os recursos de atualização rápida de índices estiverem desabilitados, e o tamanho ocupado pelo índice GIN pode ser de duas a três vezes maior do que o índice GiST.

O índice do tipo GIN foi escolhido para realizar a indexação do texto das

publicações presentes no repositório institucional proposto, de forma a converter o texto das publicações em um campo do tipo *tsvector* que pode ser indexado, permitindo a utilização das funções de *full-text search* embutidas no banco de dados PostgreSQL de uma forma rápida e eficiente.

2.4 SISTEMAS RELACIONADOS

No Quadro 2 é apresentado um resumo de sistemas relacionados a esta pesquisa, contendo o nome do sistema, ano de lançamento, tecnologias, empresa mantenedora, licença e site de acesso.

Quadro 2: Sistemas Relacionados

| Nome | Lançamento | Tecnologias | Mantenedor | Licença | Site |
|-----------------|-------------------|---------------------|--|----------------|---|
| Digital Commons | 2004 | Não divulgado | Elsevier/bepress | Comercial | https://bepress.com/products/digital-commons |
| DSpace | 2002 | Java e Angular | DuraSpace/Lyrisis | BSD | https://dspace.lyrisis.org |
| EPrints | 2000 | Perl | University of Southampton | GPL 3.0 | https://www.eprints.org/uk |
| Islandora | 2010 | PHP | Islandora Foundation | GPL 2.0 | https://www.islandora.ca |
| OPUS 4 | 2010 | PHP | Cooperative Library Network Berlin-Brandenburg | GPL | https://www.opus-repository.org |
| WEKO | 2008 | Python e JavaScript | National Institute of Informatics | BDS | https://github.com/RCOSDP/weko |

2.5 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.

2.5.1 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.

2.5.2 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 digi-

talizadoras 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.

2.5.3 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 documentos 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.

2.5.4 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 normal 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.

2.5.5 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.

2.5.6 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 móveis, tendo como base o Repositório Institucional da UNESP.

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

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 móveis 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 móveis possuem maior destaque do que outros elementos que possuem maior importância, como por exemplo a função de auto-arquivamento, sendo um recurso muito pouco utilizado a partir de dispositivos móveis, 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 móveis, e *link* visível para retornar a página 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 móveis.

2.5.7 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 provém 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 acu-

mulação, 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 intercambio 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, conferencias, 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%, Asia com 11.04%, Austrália com 5.84%, a América do Sul com 4.40%, e Africa 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.

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

No trabalho realizado por Arlitsch; Grant (2018) possui como objetivo examinar as limitações e possibilidades encontradas no cenário de repositórios institucionais. 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.

Porém 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.

Porém 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, porém 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.

2.5.9 Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS

O trabalho relacionado de Zervas; Kounoudes et al. (2019) possui como foco a transformação do repositório institucional da *Cyprus University of Technology* (CUT) denominado KTISIS, em um *Current Research Information System* (CRIS), sendo a primeira apresentação de uma implementação europeia baseada no *software open source* DSpace-CRIS.

A motivação para a elaboração da pesquisa consiste em atender as ne-

cessidades dos pesquisadores, permitindo que possam submeter seus trabalhos e perfil acadêmico para o KTISIS. Tendo como alteração do KTISIS mais notável a provisão do Perfil do Acadêmico, onde os pesquisadores terão acesso dedicado a um conjunto de funcionalidades que agregam valor aos seus trabalhos, e ao próprio repositório.

Os trabalhos realizados no KTISIS possuem como objetivo seguir as diretrizes e os comportamentos sugeridos pela *Next Generation Repositories* publicado pelo COAR (*Confederation of Open Access Repositories*), que visam colocar os repositórios em posição para a fundação de uma infraestrutura de rede globalmente distribuída para a comunicação acadêmica.

Como principais benefícios da transformação do repositório KTISIS em um sistema CRIS usando o DSpace-CRIS, está a provisão de um modelo de dados flexível que permite descrever todas os documentos que compõem o ambiente acadêmico e suas ligações, por meio da provisão de identificadores únicos que permitem criar relações entre os documentos do sistema, como por exemplo o ORCID (*Open Researcher and Contributor ID*) que seria o identificador do pesquisador. Também foi realizado o uso da funcionalidade que permite importar registros de bases de dados externas como a Web of Science e o Scopus.

Como contribuição a atual pesquisa, é possível constatar a apresentação dos conceitos de *Next Generation Repositories* que visa fornecer diretrizes e sugestões para o desenvolvimento de repositórios acadêmicos preparados para uma rede global de comunicação acadêmica. Além dos conceitos de identificadores únicos para relacionar e citar documentos e autores no repositório como o ORCID.

2.5.10 Using Open Access Institutional Repositories to Save the Student Symposium during the COVID-19 Pandemic

O trabalho relacionado de Symulevich; Hamilton (2022) tem como objetivo descrever como duas universidades utilizaram de seus repositórios institucionais para adaptar os seus simpósios de pesquisa acadêmica a eventos virtuais em questão de semanas.

O primeiro caso traz o relatado de como a mostra de primavera para pes-

quisa e inquérito criativo da universidade de Longwood teve de ser adaptada para um evento online. A universidade de Longwood consiste em uma universidade pública do centro-sul da Virginia nos Estados Unidos, que desde a primavera de 2018 realiza um simpósio de pesquisa acadêmica de alunos dos cursos de graduação, de forma presencial.

Com o crescimento dos casos de COVID-19, os administradores da instituição começaram a discutir a possibilidade de alterações na forma como os eventos são realizados, sendo constatado que o repositório acadêmico da universidade poderia ser utilizado para hospedar as publicações do evento.

Após a aprovação da ideia, o diretório de pesquisas contatou a biblioteca da universidade sobre a possibilidade da utilização do repositório Bepress Digital Commons², para hospedar as publicações do evento virtual, além de requisitar ferramentas para facilitar a comunicação entre os acadêmicos por meio e de comentários que são realizados durante a apresentação do evento de tres dias.

Dentre as ferramentas, foram elencados o Disqus³ e o Intense Debate⁴ como ferramentas para troca de comentários. Também foram discutidas soluções para o envio de apresentações no formato PowerPoint que continham videos embutidos, visto que o arquivo do video é separado do arquivo da apresentação.

A Universidade de Longwood realizou outros dois eventos no formatado virtual, no outono de 2020 e na primavera de 2021. Os organizadores optaram por migrar para a plataforma Symposium by Forager One⁵ para a realização dos eventos.

O segundo caso aborda a realização do simpósio virtual de pesquisa acadêmica da *University of South Florida St. Petersburg*, sendo uma ramificação da *University of South Florida* (USF). Com a chegada da COVID-19, a universidade teve de se adaptar ao ambiente virtual, e pausar as atividades presenciais no campus. Com isto o escritório de pesquisa começou a entrar em contato com os bibliotecários da *Nelson Poynter Memorial Library* para discutir a possibilidade de um

²<https://bepress.com/>

³<https://disqus.com/>

⁴<https://intensedebate.com/>

⁵<https://symposium.foragerone.com/>

simpósio virtual.

Uma variedade de plataformas foram elencadas como candidatas para realizar a hospedagem do simpósio, como o Website do campus, Canvas, LibGuides, Bepress Digital Commons, e até o Facebook. Porém levando como critério fatores relacionados a moderação, segurança, engajamento e arquivamento, foi optado pela recomendação do próprio repositório da instituição para hospedar o simpósio virtual.

Como o escritório de pesquisa queria manter uma experiência semelhante ao evento presencial, foi requisitado que a plataforma incluísse tanto opções de video quanto de audio, publicação e visualização de posters, e participação por meio de comentários.

A biblioteca teve de realizar upload de 55 projetos de pesquisa, 43 deles contendo uma apresentação em audio ou video. Para realizar o upload dos arquivos foi optado por elaborar uma planilha de metadados como o nome dos autores, títulos das publicações, abstract's e links para os arquivos de audio e video, permitindo o upload em massa das publicações de uma única vez.

Como o repositório da instituição era feito em com base no Bepress Digital Commons, sistema que não possui suporte a troca de comentários, o time de tecnologias optou por integrar a ferramenta Intense Debate ao repositório para poder cumprir com o requisito.

Esta pesquisa pode ser considerada relacionada ao atual trabalho, uma vez que relata os desafios enfrentados pelas instituições na tentativa de hospedar seus simpósios virtuais dentro de repositórios institucionais. Além de citar tecnologias e recursos que foram vistos como necessários para a realização dos eventos.

2.5.11 Understanding Institutional Repository in Higher Learning Institutions: A Systematic Literature Review and Directions for Future Research

A revisão de literatura desenvolvida por Asadi; Abdullah et al. (2019) possui como objetivo proporcionar um melhor entendimento e uma revisão aprofundada sobre o atual estado dos estudos envolvendo IRs. A pesquisa utiliza o método

de revisão sistemática de literatura e segue um protocolo para organizar de forma apropriada os trabalhos relatados a repositórios institucionais.

Os dados foram coletados de estudos publicados entre 2007 e 2018, em seis das maiores bases de dados (ScienceDirect, IEEE Explorer, Springer, ACM, Taylor and Francis e Emerald Insight), seguindo um critério de revisão, aplicação a inclusão da pesquisa ou exclusão, com um total de 115 estudos sendo incluídos na principal parte da pesquisa.

Como resultado da análise dos estudos, foi constatado a falta de conhecimento sobre repositórios de acesso aberto por escolas e instituições de ensino, e a infraestrutura inadequada tanto de informação quanto de comunicação, formam desafios significantes para o desenvolvimento de repositório de acesso aberto. Também identificado que dentre os maiores benefícios da adoção de repositórios institucionais estão a visibilidade aumentada da instituição, o aumento em rankings locais e globais, aumento do prestígio e valor público, a melhora do ensino, aprendizado e pesquisa desenvolvida nas instituições.

Também foi relatado que a maior parte dos estudos sobre esta área de pesquisa focam na “implantação, implementação e adoção” e os “benefícios e desafios” que tais repositórios trazem as instituições.

Este trabalho de revisão de literatura pode ser considerado relacionado a atual pesquisa, pois visa trazer embasamento sobre o atual estado dos repositórios acadêmicos, além de diretrizes para a implementação de IRs em instituições de ensino.

2.5.12 Discovery Tools to Enhance Resources Findability in the Institutional Repositories: an overview

A pesquisa realizada por Mettai (2019) possui como objetivo demonstrar formas como os repositórios institucionais podem melhorar a recuperação e descoberta por recursos, de forma a melhor suprir as necessidades dos usuários, por meio da realização de uma análise das interfaces de um conjunto de repositórios digitais das universidades da Argélia, possuindo como base a literatura documentada para trazer fundamentação teórica a análise realizada.

Para a realização da pesquisa, foi realizada uma revisão da literatura envolvendo temas como o surgimento de ferramentas de descoberta, o comportamento das pesquisas, e validação das ferramentas de pesquisa como parte essencial da biblioteca e repositórios digitais.

No paper é definido uma ferramenta de descoberta (*Discovery Tool*) como sendo um programa de computador que trabalha como um mecanismo de busca compreensivo para melhorar a encontrabilidade de recursos, de forma a coletar todos os metadados de diferentes locais e reuni-los em um único local que permite pesquisas abrangentes. Também é definido que em geral, estas ferramentas consistem de dois componentes principais, o primeiro realiza a unificação e indexação, e o segundo fornece uma camada de descoberta de conteúdo que provém diferentes recursos como o ranking de relevância, interface intuitiva e pesquisa facetada.

Em sua revisão de literatura, foi constatado que a maior parte dos usuários não se sentem confortáveis com a utilização de ferramentas de descoberta de pesquisa fornecidos por bibliotecas, pois muitos usuários já possuem uma ideia formada sobre tais ferramentas de pesquisa. Logo quando o Google Scholar foi lançado, muitos usuários o adotaram de forma imediata, visto que este oferece uma experiência de pesquisa muito melhor que as interfaces providas por bibliotecas. Por outro lado, também foi verificado que em geral os usuários se sentem satisfeitos, e até mesmo preferem utilizar uma base de dados especializada para temas de pesquisa específicos, ou o Google Scholar para pesquisa de âmbito mais geral.

Em relação ao futuro das ferramentas de descoberta, é citado 12 recursos que são vistos como essenciais, sendo um local unificado de pesquisa para toda a informação, uma interface web que esteja no estado da arte, o enriquecimento de conteúdo, navegação facetada, campo de pesquisa por palavras chaves em todas as páginas, ordenação de relevância, mecanismo de correção de texto como “Você quis dizer...” e a recomendação de materiais relacionados.

Como contribuição a atual pesquisa, é possível citar os recursos mencionados como essenciais para ferramentas de descoberta, que também podem ser aplicados no contexto de repositórios institucionais. Além da apresentação da re-

levância de que as pesquisas contidas no repositório devem ser indexadas por mecanismos como o Google Scholar para ganhar maior visibilidade.

2.5.13 Crafting Linked Open Data to Enhance the Discoverability of Institutional Repositories on the Web

O trabalho relacionado de Jin; Sandberg (2019) realiza um estudo de como o BIBFRAME 2.0 (*Bibliographic Framework*) pode ser utilizado para descrever objetos em repositórios institucionais, tendo como objetivo trazer esforços conjuntos entre duas comunidades devotadas ao conhecimento aberto. No estudo é examinado um conjunto de mapeamentos e conversões do padrão Dublin Core para o BIBFRAME 2.0, para verificar se o novo padrão irá trazer ganhos na visibilidade dos documentos digitais na web.

No artigo é definido o conceito de *Linked Open Data* como sendo um método de publicação de dados estruturados, para que estes possam ser interconectados, de forma a utilizar a rede para conectar pedaços de dados, informação e conhecimento dentro da rede semântica (*Semantic Web*), utilizando de URIs e RDF, normalmente usando de uma licença aberta, que não impeça a utilização de forma gratuita.

Já o BIBFRAME é delimitado como sendo um modelo de dados para a descrição bibliográfica, que foi desenvolvido para substituir o padrão MARC, que era utilizado para acomodar comunidades de usuários amplas como as de museus, arquivos, e editoras por meio da utilização de princípios de dados conectados, para tornar os dados bibliográficos mais úteis tanto para a comunidade interna quanto externa. O BIBFRAME e os conceitos de dados conectados permitem que os recursos bibliográficos sejam publicados de forma que a *Web* possa entender, tornando mais fácil encontrar o conteúdo em pesquisas e buscas realizadas na *Web*.

Para sumarizar o processo de migração, foram extraídas 459 teses do repositório institucional, após foi realizado a conversão do padrão Dublin Core para o MARCXML, usando a ferramenta MarcEdit. Em seguida foi convertido o padrão MARCXML para o BIBFRAME, por meio da ferramenta *Library of Congress*

MARCXML to BIBFRAME Transformation software.

Como conclusão, foram relatados problemas ao realizar a transição do padrão Dublin Core para o BIBFRAME 2.0, como por exemplo a baixa qualidade dos metadados no repositório, e outras dificuldades ao mapear registros entre um modelo e outro, citando que o BIBFRAME 2.0 possui algumas dificuldades em mapear registros não baseados no padrão MARC.

E possível dizer que este trabalho se relaciona a atual pesquisa, ao abordar diferentes padrões de armazenamento de metadados que são essenciais para o desenvolvimento do repositório acadêmico proposto, além de trazer algumas comparações entre o padrão Dublin Core e o BIBFRAME 2.0.

2.5.14 Open Access Scientific Digital Repositories : An Analytical Study of the Open DOAR

O trabalho relacionado de Ali, Loan e Mushatq (2018) possui como principal objetivo descobrir a composição dos repositórios de dados científicos presentes no *Directory of Open Access Repositories* (OpenDOAR), baseando-se em parâmetros como a distribuição regional, contribuição por país, tipo de repositório, desenvolvimento de coleções, conteúdos arquivados, diversidade de idiomas, *software* utilizado e políticas de metadados.

A pesquisa possui foco no monitoramento do crescimento de OARS (*Open Access Repositories*) dentre as diversas áreas da ciência mundialmente registradas no *Directory of Open Access Repositories* (OpenDOAR). Este iniciado em 2006 pela universidade de Nottingham (UK), consistindo em um repositório oficial para os OARs, apresentado cerca de 2600 repositórios no momento em que a pesquisa foi realizada.

Para a realização da pesquisa, foram coletados dados na própria plataforma do OpenDoar, tendo o espoco da pesquisa limitado somente a repositórios de pesquisa científica presentes nesta plataforma.

Como resultados, foi revelado que a Europa contribui com o maior número de repositórios por região, enquanto os Estados Unidos possui a maior quantidade de repositórios por país. Boa parte dos repositórios analisados são institucionais,

aceitam conteúdo em inglês, reúnem os conteúdos de periódicos e artigos, alguns possuem coleções, e utilizam o *software* DSpace para gerenciar o conteúdo, não possuindo políticas de metadados bem definidas.

Em relação ao atual trabalho, é possível citar que a pesquisa se relaciona ao trazer um contexto geral do estado dos repositórios acadêmicos pelo mundo, trazendo embasamento sobre tecnologias utilizadas, conteúdos, políticas de metadados, dentro outros.

2.5.15 Evolution of institutional repositories: Managing institutional research output to remove the gap of academic elitism

O trabalho realizado por Bashir; Gul et al. (2021) possui como premissa realçar a evolução e estrutura conceitual dos repositórios institucionais, e seus impactos nos círculos escolar e acadêmico, em termos de melhor visibilidade, aumento de audiência e comunicação de pesquisa.

O estudo se baseia na examinação e avaliação vários de periódicos revisados por pares, mostrando inúmeras dimensões de repositórios institucionais desde sua evolução até aceitação acadêmica.

Para a realização da pesquisa foi realizada uma revisão de literatura em bases de dados como Clarivate Analytic's, Web of Science e Scopus. Foram utilizados termos como repositórios institucionais, resultado de pesquisa institucional, repositórios de acesso aberto, acesso aberto verde, acesso aberto, publicações de acesso aberto, iniciativas de acesso aberto, bibliotecas digitais, diretório de repositórios de pesquisa aberta, OpenDOAR, e comunicação escolar.

Como resultado, o estudo revelou que os benefícios associados aos repositórios institucionais são vários, eles fornecem aos usuários uma informação que de outro modo estaria indisponível, por razões que variam entre a falta da disponibilidade de materiais complementares (como relatórios e papers não publicados, material audio visual, objetos de aprendizado, referências bibliográficas, bases de dados, notas de leitura, dentre outros), a modelos de pagamento ou assinatura adotados por canais de comunicação escolar.

Também é relatado que a utilização dos IRs está gerando um impacto real

aos modelos de cobrança orquestrados pelas editoras comerciais, democratizando a voz dos acadêmicos por meio remoção de barreiras em plataformas estudantis.

Este trabalho pode ser caracterizado como relacionado a atual pesquisa, visto que aborda questões referentes a como a implementação de repositórios acadêmicos de acesso aberto podem democratizar o acesso aos materiais de pesquisa, e remover barreiras impostas pelos valores e assinaturas cobradas por editoras.

Nos Quadros 3 e 4 é apresentado um resumo dos trabalhos relacionados citados nesta pesquisa, contendo o autor, título, objetivo, metodologia e resultados de cada trabalho.

Quadro 3: Trabalhos Relacionados Parte 1

| Autor | Título | Objetivo | Metodologia | Resultados |
|-------------------------------------|---|---|-------------------------------------|--|
| Symulevich; Hamilton (2022) | Using Open Access Institutional Repositories to Save the Student Symposium during the COVID-19 Pandemic | Descrever como duas universidades utilizaram de seus repositórios institucionais para adaptar os seus simpósios a eventos virtuais. | Documental | Relata os desafios enfrentados pelas instituições na tentativa de hospedar seus simpósios virtuais dentro de repositórios institucionais. |
| Pereira, Moreira e Segundo (2021) | Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital | Apresenta categorias fundamentais a serem utilizadas para a construção de uma interface de navegação facetada | Qualitativa e exploratória. | Foram criadas 8 categorias fundamentais para a navegação facetada. |
| Bashir; Gul et al. (2021) | Evolution of institutional repositories: Managing institutional research output to remove the gap of academic elitism | Realçar a evolução e estrutura conceitual dos repositórios institucionais, e seus impactos nos círculos escolar e acadêmico. | Revisão de literatura | Foi constatado os benefícios dos repositórios institucionais, e como os IRs de acesso aberto estão democratizando o acesso a pesquisa. |
| Nascimento, Queiroz e Araujo (2020) | Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca | 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, confiável e segura. |
| Zervas; Kounoudes et al. (2019) | Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS | Documentar a transformação do repositório institucional da <i>Cyprus University of Technology</i> denominado KTISIS, em um CRIS. | Qualitativa e descritiva | Foram relatados os principais benefícios da transformação do repositório KTISIS em um sistema CRIS, utilizando o DSpace-CRIS. |
| Weitzel (2019) | O mapeamento dos repositórios institucionais brasileiros: perfil e desafios | Retratar a situação atual e contribuir com subsídios para orientar as ações e diretrizes para implementação de repositórios. | 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. |
| Garcia (2019) | Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência | Desenvolvimento da na nova plataforma de Biblioteca Digital da BBM (Biblioteca Brasileira Guita 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. |
| Mettai (2019) | Discovery Tools to Enhance Resources Findability in the Institutional Repositories: an overview | Demonstrar formas como os repositórios institucionais podem melhorar a recuperação e descoberta por recursos, de forma suprir as necessidades dos usuários. | Revisão de literatura | Foram ressaltadas questões referentes a utilização de ferramentas de descoberta e pesquisa, e recursos vistos como essenciais nestas ferramentas. |
| Jin; Sandberg (2019) | Crafting Linked Open Data to Enhance the Discoverability of Institutional Repositories on the Web | Realiza um estudo de como o BIBFRAME 2.0 pode ser utilizado para descrever objetos em repositórios institucionais. | Qualitativa e descritiva | Foram relatados problemas ao realizar a transição do padrão Dublin Core para o BIBFRAME 2.0. |

Quadro 4: Trabalhos Relacionados Parte 2

| Autor | Título | Objetivo | Metodologia | Resultados |
|-------------------------------|---|---|--------------------------------|--|
| Fernandes (2019) | Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis | Compreender de que forma ocorre a encontrabilidade da informação em repositórios institucionais a partir do uso de dispositivos móveis. | Método quadripolar. | No repositório analisado elementos não relevantes em dispositivos móveis possuem maior destaque do que outros elementos que possuem maior importância. |
| Asadi; Abdullah et al. (2019) | Understanding Institutional Repository in Higher Learning Institutions: A Systematic Literature Review and Directions for Future Research | Proporcionar um melhor entendimento e uma revisão aprofundada sobre o atual estado dos estudos envolvendo IRs. | Revisão sistemática de leitura | Foi constatado que a falta de conhecimento sobre repositórios de acesso aberto e a infraestrutura inadequada formam desafios para o desenvolvimento de repositório de acesso aberto. |
| Saini (2018) | The Emergence of Institutional Repositories: A Conceptual Understanding of Key Issues through Review of Literature | Buscar a compreensão conceitual dos principais problemas que envolvem a criação e uso de repositórios institucionais. | Revisão de literatura | Foram levantados dados referentes ao surgimento dos repositórios institucionais, motivação, fatores de custo e <i>softwares open source</i> . |
| Ali, Loan e Mushatq (2018) | Open Access Scientific Digital Repositories : An Analytical Study of the Open DOAR | Descobrir a composição dos repositórios de dados científicos presentes no Open DOAR. | Quantitativo | Foram elencados os países com maior número de contribuição, tecnologias utilizadas, conteúdos armazenados e políticas de metadados. |
| Arlitsch; Grant (2018) | Why So Many Repositories? Examining the Limitations and Possibilities of The Institutional Repositories (IR) Landscape | Examinar as limitações e possibilidades encontradas no cenário de repositórios institucionais. | Revisão de literatura | Foi constatado que a existência de centenas de repositórios diferentes não servem mais ao usuário comum, e acabam trazendo diversos problemas e custos. |
| Cervi; Manica et al. (2009) | BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados | Apresentação da proposta de biblioteca digital de trabalhos científicos, que foi denominada como BDTC. | Qualitativa | Apresentação uma biblioteca digital de trabalhos científicos que faz uso de serviços como autoarquivamento de conteúdo, extração de metadados dos objetos digitais e busca por similaridade. |

3 RESULTADOS

Neste capítulo são apresentados os resultados do trabalho desenvolvido, apresentando a criação do repositório institucional proposto, desde a sua concepção, *backlog* do produto, protótipos, resultado final e validação das hipóteses.

3.1 CONCEPÇÃO DO REPOSITÓRIO INSTITUCIONAL

A concepção do repositório institucional proposto foi realizada com base nos desafios e problemas relatados nos trabalhos relacionados, e na análise dos pontos positivos e negativos encontrados em outros sistemas de repositório institucional existentes.

Como um dos principais diferenciais do repositório proposto, é que desde o início de seu desenvolvimento ele foi projetado para ser facilmente implantado e executado em ambientes baseados em nuvem, sendo assim todas as suas configurações podem ser realizadas por meio de variáveis de ambiente, que em geral podem ser facilmente definidas pelo próprio painel ou *dashboard* fornecido pelo provedor de nuvem. A ideia é nunca precisar acessar remotamente a máquina em que a aplicação está sendo executada para realizar qualquer tipo de configuração.

Outro ponto de diferença é que o repositório proposto somente aceita arquivos no formato PDF, visando mitigar problemas relacionados a formatação do documento, e buscando garantir que o usuário ao acessar as publicações, visualize o documento da mesma forma que o autor originalmente a escreveu. Esta redução do escopo de formatos de arquivos aceitos pelo repositório, também garante que todas as publicações estejam normalizadas em um mesmo formato, permitindo a utilização de técnicas mais especializadas para a extração de texto dos documentos.

Para realizar a escolha do nome e identidade visual do repositório, foram buscados por termos que remetessem a "Repositório", "Acesso Aberto" e "Pesquisa", chegando a escolha do nome RESOAR, uma sigla em inglês para *Research Open Access Repository*. A sigla "OAR" que remete a *Open Access Repository* já é bastante difundida, e pode ser encontrada em nomes como o OpenDOAR¹ e RO-ARMAP².

Figura 9: Identidade visual do repositório



Fonte: Do próprio autor

A Figura 9 apresenta a logo ou identidade visual do repositório, que foi elaborada por um designer contratado da região de Três de Maio - RS, que desenvolveu tanto a identidade visual quanto os primeiros protótipos do repositório.

Em sua composição é possível perceber um símbolo de lupa embutido na primeira letra "R" representando a pesquisa. A cor padrão da logo foi escolhida como o laranja com o código HEX #ff5400, porém o símbolo também pode ser exibido em outras variações, como em preto e branco, escala de cinza ou outras cores.

3.2 BACKLOG DO PRODUTO

Nesta seção será apresentado o *backlog* do produto que foi desenvolvido seguindo a metodologia de *User Stories*, contendo uma lista de funcionalidades que o sistema final deve conter, além de *mockups* das telas, servindo como uma base para o desenvolvimento da aplicação.

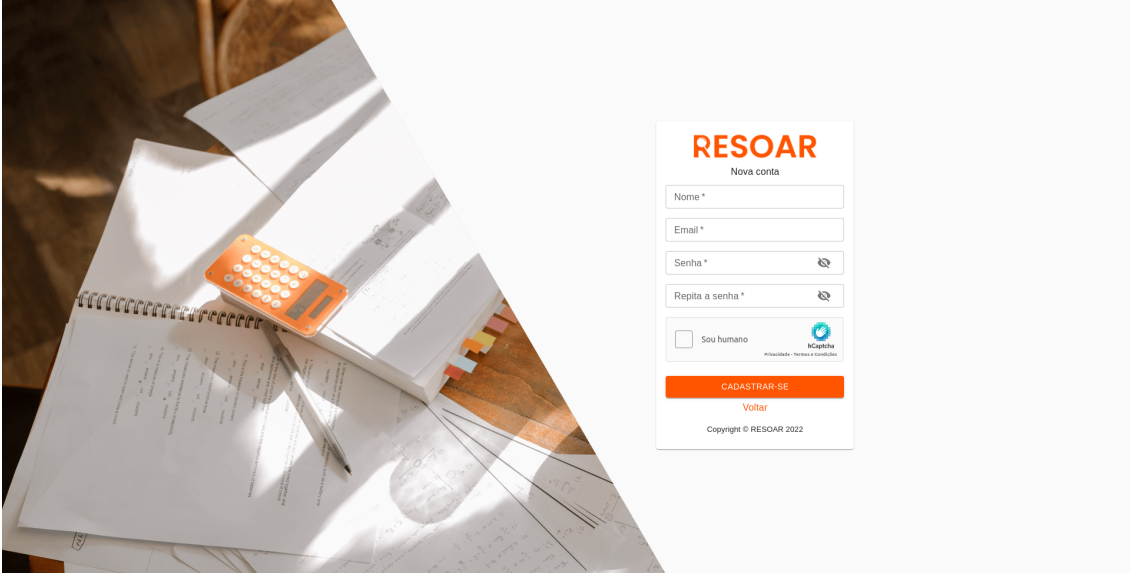
¹<https://v2.sherpa.ac.uk/opensoar/>

²<https://roarmap.eprints.org/>

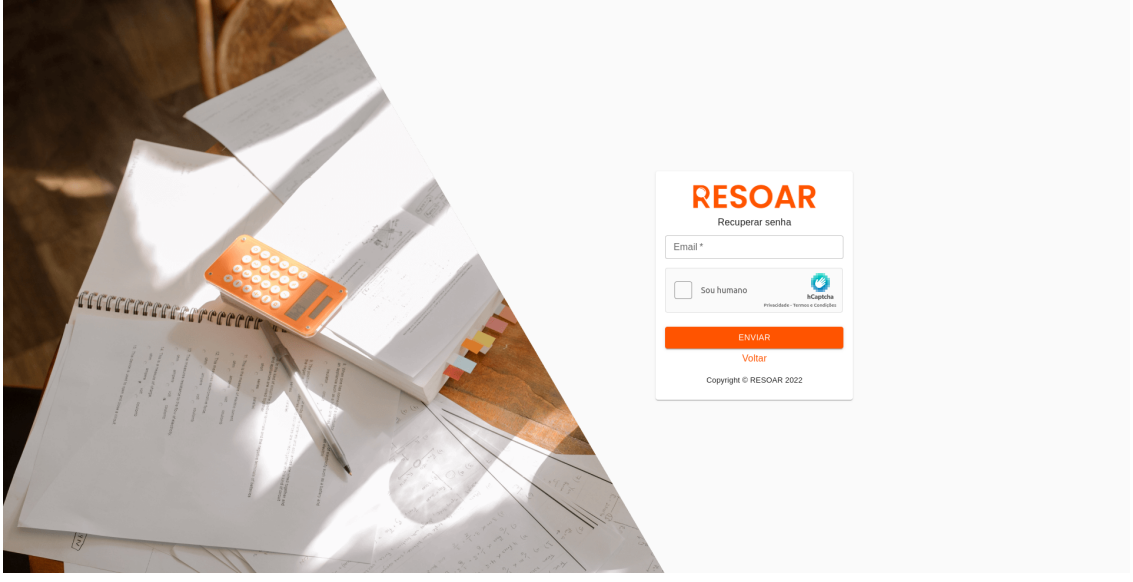
Quadro 5: Tela de Login

| | |
|---|----------------------|
| 01 | Tela de Login |
| Como usuário do sistema | |
| Eu quero realizar o login utilizando meu email e senha | |
| Para que possa utilizar as funções presentes no sistema. | |
| Critérios de aceitação | |
| <ol style="list-style-type: none">1. Deve possuir o botão de exibir/esconder a senha.2. Deve possuir os atalhos para as telas de recuperação de senha e cadastro de usuário.3. Caso o usuário digitar a senha incorretamente mais de 3 vezes, deve solicitar a solução de um desafio captcha. | |
|  | |

Quadro 6: Cadastro de usuário

| | |
|---|----------------------------|
| 02 | Cadastro de usuário |
| Como novo usuário Eu quero me cadastrar no sistema Para que possa entrar no sistema. | |
| Critérios de aceitação | |
| <ol style="list-style-type: none">1. Deve validar se já existe um usuário cadastrado com o mesmo email.2. Deve validar se os dois campos de senha são iguais.3. Deve solicitar a solução de um desafio captcha. | |
|  | |

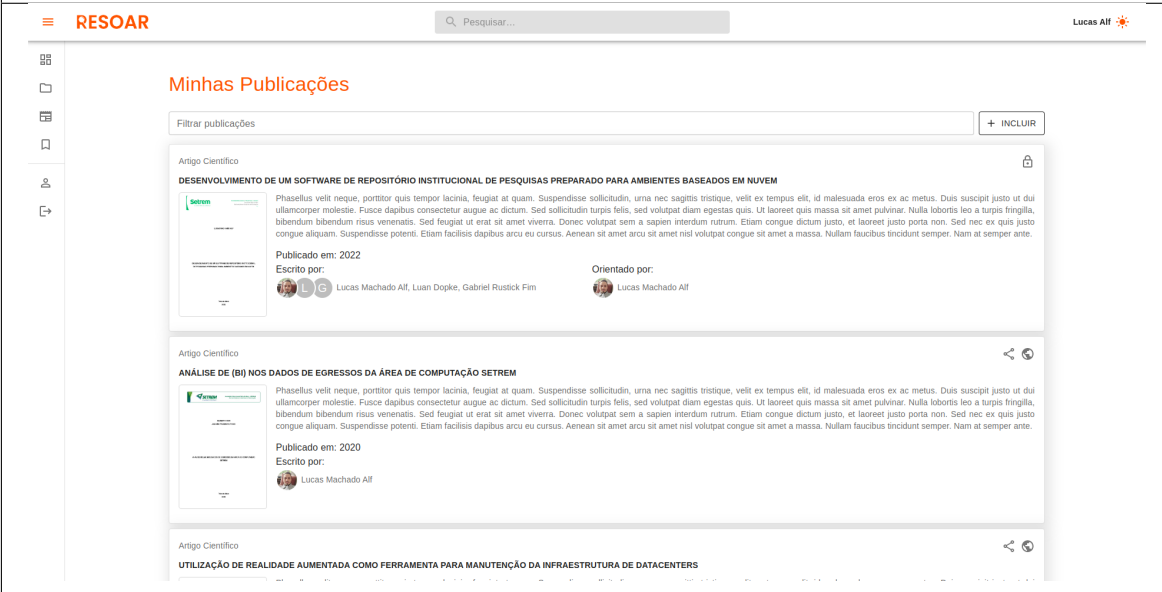
Quadro 7: Recuperação de senha

| | |
|--|-----------------------------|
| 03 | Recuperação de senha |
| Como usuário do sistema | |
| Eu quero poder solicitar um email de recuperação de senha | |
| Para que possa criar uma nova senha para meu usuário. | |
| Critérios de aceitação | |
| <ol style="list-style-type: none">1. Deve validar se o email informado existe no sistema.2. Deve enviar um email de recuperação, com validade máxima de 3 horas.3. Deve solicitar a solução de um desafio captcha. | |
|  | |

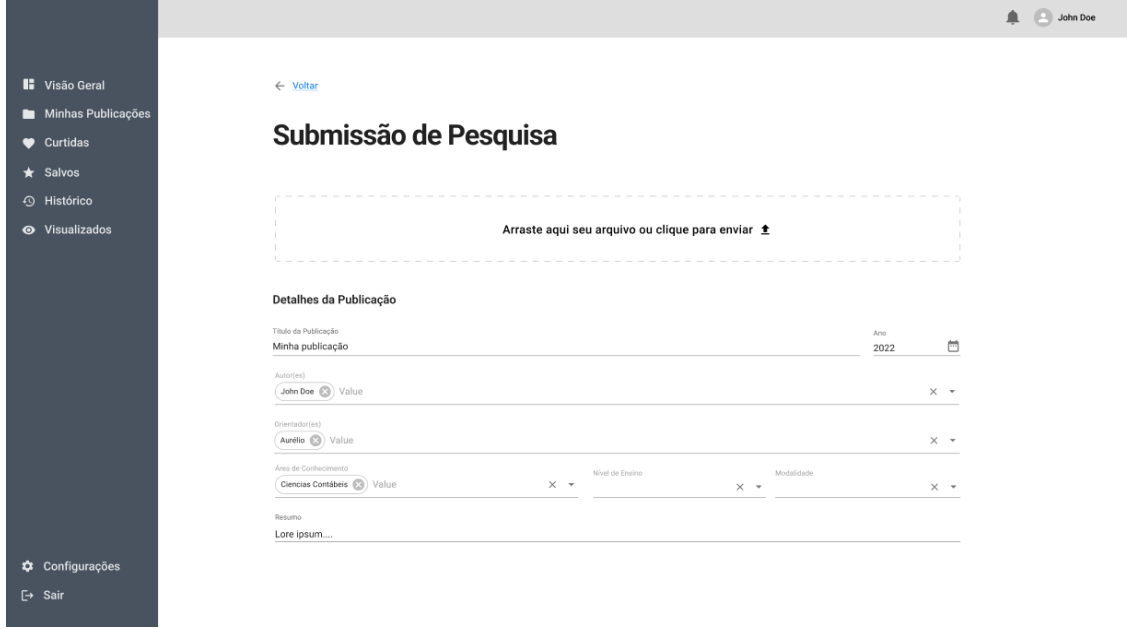
Quadro 8: Visão geral

| | |
|--|--------------------|
| 04 | Visão geral |
| <p>Como usuário do sistema</p> <p>Eu quero acessar uma tela de boas vindas ao entrar no sistema, contendo as minhas publicações mais recentes, publicações salvas, e histórico de publicações acessadas.</p> <p>Para que tenha um ponto de partida.</p> | |
| Critérios de aceitação | |
| <ol style="list-style-type: none"> 1. Deve exibir ao menos as 3 últimas publicações do usuário. 2. Deve exibir ao menos as 3 últimas publicações salvas para leitura. 3. Deve exibir ao menos um histórico das 3 últimas publicações acessadas. | |
|  | |

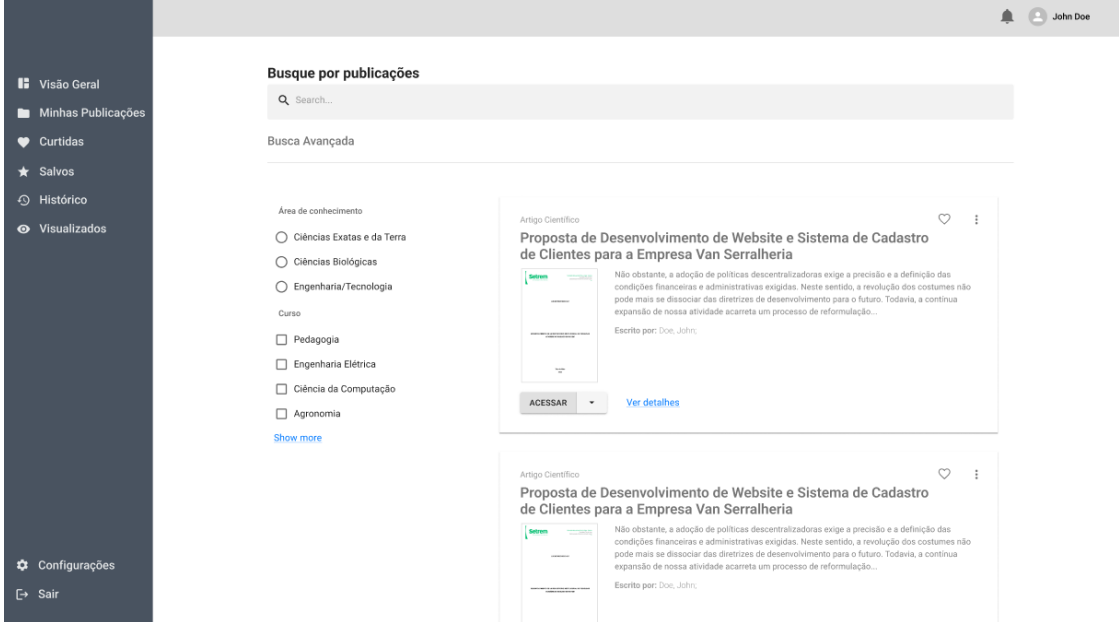
Quadro 9: Minhas publicações

| | |
|----|---|
| 05 | Minhas publicações |
| | <p>Como usuário do sistema</p> <p>Eu quero acessar uma listagem contendo todas as minhas publicações</p> <p>Para que possa navegar pelas publicações que realizei.</p> |
| | Critérios de aceitação |
| | <ol style="list-style-type: none"> 1. Deve permitir a filtragem de publicações pelo título. 2. Deve exibir na listagem o título, a imagem de capa, os autores, orientadores e parte do resumo. 3. Deve possuir um botão para a inclusão de uma nova publicação. 4. Ao clicar em uma publicação, deve redirecionar para a página detalhada da publicação. |
| |  <p>A imagem é uma captura de tela da interface de usuário do sistema RESOAR, especificamente a seção 'Minhas Publicações'. No topo, há uma barra de navegação com o logo 'RESOAR' e uma barra de busca. À esquerda, há um menu lateral com ícones para diferentes funcionalidades. O conteúdo principal mostra uma listagem de publicações com o seguinte layout:</p> <ul style="list-style-type: none"> Filtros: Um campo de texto 'Filtrar publicações' e um botão '+ INCLUIR'. Artigo Científico 1: <ul style="list-style-type: none"> Título: DESENVOLVIMENTO DE UM SOFTWARE DE REPOSITÓRIO INSTITUCIONAL DE PESQUISAS PREPARADO PARA AMBIENTES BASEADOS EM NUVEM Resumo: Phasellus velit neque, portitor quis tempor lacinia, feugiat at quam. Suspendisse sollicitudin, urna nec sagitta tristique, velit ex tempus elit, id malesuada eros ex ac metus. Duis suscipit justo ut dui ullamcorper molestie. Fusce dapibus consetetur augue ac dictum. Sed sollicitudin turpis felis, sed volutpat diam egestas quis. Ut laoreet quis massa sit amet pulvinar. Nulla lobortis leo a turpis fringilla, bibendum bibendum risus venenatis. Sed feugiat ut erat sit amet viverra. Donec volutpat sem a sapien interdum rutrum. Etiam congue dictum justo, et laoreet justo porta non. Sed nec ex quis justo congue aliquam. Suspendisse potenti. Etiam facilisis dapibus arcu eu cursus. Aenean sit amet arcu sit amet nisl volutpat congue sit amet a massa. Nullam faucibus tincidunt semper. Nam at semper ante. Publicado em: 2022 Escrito por: Lucas Machado Alf, Luan Dopke, Gabriel Russtick Fim Orientado por: Lucas Machado Alf Artigo Científico 2: <ul style="list-style-type: none"> Título: ANÁLISE DE (BI) NOS DADOS DE EGRESSOS DA ÁREA DE COMPUTAÇÃO SETREM Resumo: Phasellus velit neque, portitor quis tempor lacinia, feugiat at quam. Suspendisse sollicitudin, urna nec sagitta tristique, velit ex tempus elit, id malesuada eros ex ac metus. Duis suscipit justo ut dui ullamcorper molestie. Fusce dapibus consetetur augue ac dictum. Sed sollicitudin turpis felis, sed volutpat diam egestas quis. Ut laoreet quis massa sit amet pulvinar. Nulla lobortis leo a turpis fringilla, bibendum bibendum risus venenatis. Sed feugiat ut erat sit amet viverra. Donec volutpat sem a sapien interdum rutrum. Etiam congue dictum justo, et laoreet justo porta non. Sed nec ex quis justo congue aliquam. Suspendisse potenti. Etiam facilisis dapibus arcu eu cursus. Aenean sit amet arcu sit amet nisl volutpat congue sit amet a massa. Nullam faucibus tincidunt semper. Nam at semper ante. Publicado em: 2020 Escrito por: Lucas Machado Alf Artigo Científico 3: <ul style="list-style-type: none"> Título: UTILIZAÇÃO DE REALIDADE AUMENTADA COMO FERRAMENTA PARA MANUTENÇÃO DA INFRAESTRUTURA DE DATACENTERS |

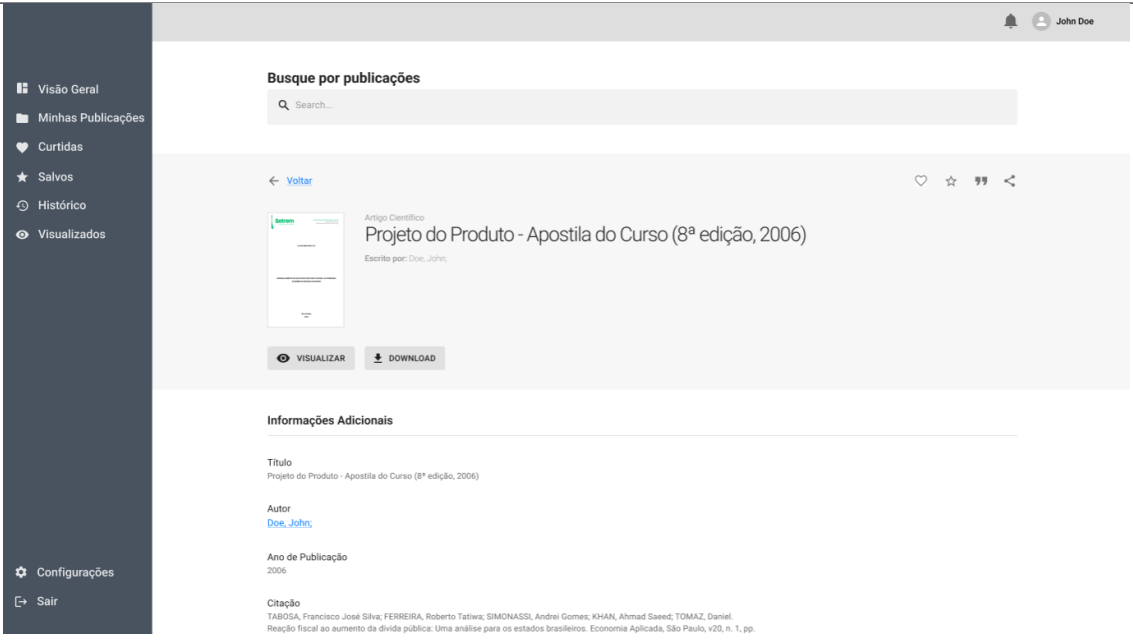
Quadro 10: Nova publicação

| | |
|--|------------------------|
| 06 | Nova publicação |
| <p>Como usuário do sistema</p> <p>Eu quero realizar uma nova publicação</p> <p>Para que possa salvar as minhas publicações no sistema.</p> | |
| Critérios de aceitação | |
| <ol style="list-style-type: none"> 1. Deve solicitar os campos de título, resumo, ano, tipo, visibilidade, idioma, instituição, autores, orientadores e arquivo da publicação. 2. Dentro do campo de autores sempre deve haver no mínimo o próprio usuário. | |
|  <p>The screenshot displays a web interface for submitting a research publication. On the left is a dark sidebar with navigation options: 'Visão Geral', 'Minhas Publicações', 'Curtidas', 'Salvos', 'Histórico', 'Visualizados', 'Configurações', and 'Sair'. The main area is titled 'Submissão de Pesquisa' and features a large dashed box for file upload with the text 'Arraste aqui seu arquivo ou clique para enviar'. Below this is the 'Detalhes da Publicação' section, which contains several input fields: 'Título da Publicação' (with the value 'Minha publicação'), 'Ano' (with the value '2022'), 'Autor(es)' (with the value 'John Doe'), 'Orientador(es)' (with the value 'Aurelio'), 'Área de Conhecimento' (with the value 'Ciencias Contábeis'), 'Nível de Ensino', and 'Modalidade'. Each of these fields has a clear button (X) and a dropdown arrow. At the bottom is a 'Resumo' field with the placeholder text 'Lorem ipsum...'.</p> | |

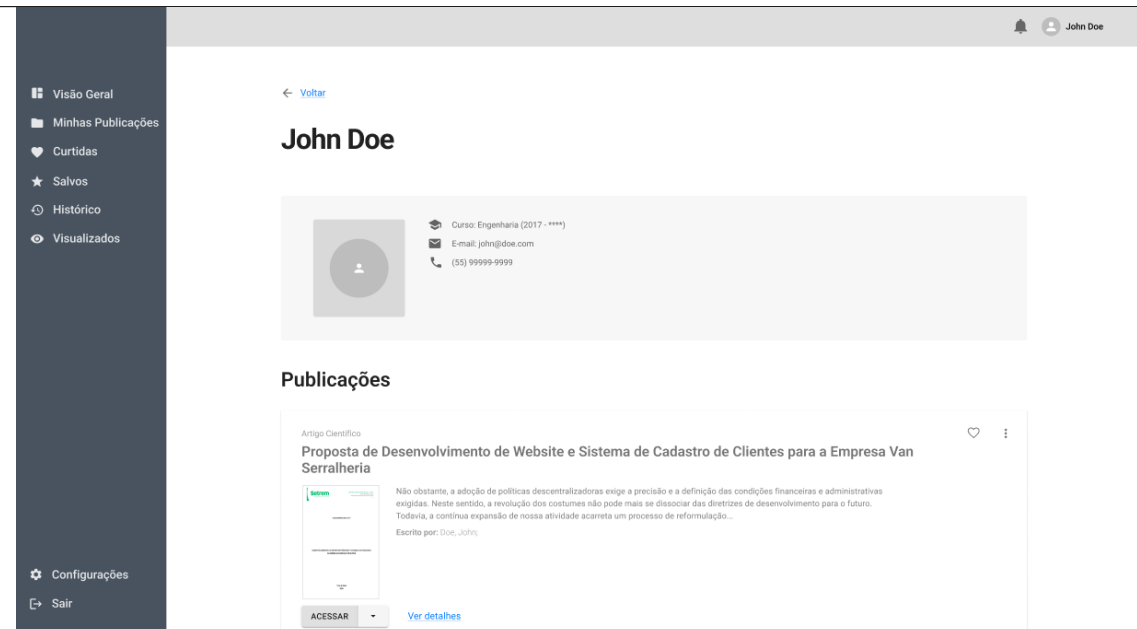
Quadro 11: Pesquisar publicações

| | |
|----|--|
| 07 | Pesquisar publicações |
| | <p>Como usuário do sistema</p> <p>Eu quero pesquisar por publicações, podendo utilizar de filtros</p> <p>Para que possa encontrar as publicações que desejo visualizar.</p> |
| | Critérios de aceitação |
| | <ol style="list-style-type: none"> 1. Deve permitir a busca por termos contidos dentro das publicações. 2. Deve permitir a filtragem por ano, autor, orientador, tipo e idioma. 3. Uma caixa de pesquisa por publicações deve estar sempre visível na interface do usuário. |
| |  |

Quadro 12: Visualizar publicação

| | |
|---|------------------------------|
| 08 | Visualizar publicação |
| <p>Como usuário do sistema</p> <p>Eu quero visualizar uma publicação</p> <p>Para que possa ver os autores, orientadores, resumo e baixar o arquivo da publicação.</p> | |
| Critérios de aceitação | |
| <ol style="list-style-type: none">1. Deve exibir o título da publicação em destaque.2. Deve exibir a capa da publicação, e informações como os autores, orientadores, tipo de publicação, idioma e resumo.3. Deve permitir realizar o <i>download</i> da publicação.4. Deve permitir salvar a publicação para mais tarde.5. Deve permitir gerar um <i>link</i> de compartilhamento da publicação. | |
|  | |

Quadro 13: Perfil do usuário

| | |
|--|--------------------------|
| 09 | Perfil do usuário |
| <p>Como usuário do sistema</p> <p>Eu quero acessar uma página de perfil de usuário</p> <p>Para que possa atualizar minhas informações, e trocar minha senha.</p> | |
| Critérios de aceitação | |
| <ol style="list-style-type: none"> 1. Deve exibir o nome e a foto do usuário. 2. Deve listar as publicações as quais o usuário participa como autor. 3. Caso o usuário esteja acessando o seu próprio perfil, devem existir opções para editar as informações e trocar a senha. | |
|  | |

3.3 ARQUITETURA DA APLICAÇÃO

Nesta seção será apresentada a estrutura utilizada para desenvolver tanto o *frontend* quanto o *backend* do repositório institucional, além das ferramentas e tecnologias utilizadas durante o processo de desenvolvimento.

Para realizar o versionamento de código, tanto do *backend* quanto do *frontend* foi utilizado o GitHub³, em conjunto com o SonarCloud⁴ para a realização de *code review* automatizado, e controle de qualidade de código.

³<https://github.com/>

⁴<https://sonarcloud.io/>

3.3.1 Estrutura de Backend

O *backend* do repositório institucional consiste em uma API REST escrita na linguagem C#, utilizando do .NET 7. Para conexão com o banco de dados foi utilizado do *ORM* (*Object Relational Mapper*) Entity Framework, porém o projeto também oferece suporte ao *Micro ORM* Dapper. Como esquema de autenticação de usuário foi optado pelo JWT (*JSON Web Tokens*).

Este projeto segue a estrutura de *Repository Pattern*, em conformidade com os conceitos do DDD (*Domain-Driven Design*), que permitem a divisão do código fonte em diversas camadas, que neste projeto foram divididas em:

1. **Core:** contém as camadas de *Domain* e *Application*, envolvendo as entidades que serão utilizadas para modelar o banco de dados, e as *Services* que funcionam como uma camada intermediária, que visam integrar somente o que desejamos expor da camada *Infrastructure* a camada *Presentation*.
2. **Infrastructure:** responsável por todo e qualquer acesso ou operação sobre o banco de dados, contendo os arquivos de configuração de conexão, configurações do *ORM*, *Migrations* e *Repositories*.
3. **Presentation:** consiste em uma API REST, responsável pela comunicação entre o usuário final e a aplicação, nesta camada estarão presentes as *Controllers* e outras configurações da API.
4. **Tests:** contém os arquivos referentes a testes unitários, testes de integração, ou outros *scripts* de automatização de testes.

Como banco de dados foi utilizado o PostgreSQL, visto que este possui recursos de *full-text search* embutidos, que foram utilizados para realizar a consulta por publicações. Para realizar a modelagem do banco de dados, em vez de seguir uma estrutura tradicional de modelo ER, foi utilizado de uma abordagem *code first*, por meio do recurso de *Migrations* fornecido pelo *Entity Framework*.

Para realizar o armazenamento dos arquivos das publicações, é utilizado de serviços de *Object Storage* baseados em nuvem, dando suporte ao *Digital Ocean Spaces* e ao *Amazon S3*.

3.3.2 Estrutura de Frontend

O *frontend* do repositório institucional foi desenvolvido majoritariamente na linguagem de programação JavaScript utilizando do React, em conjunto do *build tool* Vite, e dos componentes visuais do Material UI.

Este projeto segue uma estrutura básica, onde os arquivos estão organizados em pastas, que normalmente são encontradas em projetos feitos em React, sendo:

1. **Assets:** contém arquivos estáticos de imagens, sons, arquivos de texto, etc.
2. **Components:** contém os componentes personalizados do projeto, como o *layout* do site, *NavBar*, menu lateral, etc.
3. **Pages:** contém as páginas do site, como a página de *login*, *home*, perfil do usuário, etc.
4. **Services:** contém os arquivos responsáveis pela comunicação com a API, por meio de requisições HTTP.
5. **Themes:** contém os arquivos de estilos do projeto, como o tema claro, tema escuro, cores primárias, secundárias, etc.

3.4 ANÁLISE DOS RESULTADOS

Nesta seção será apresentado os resultados dos testes realizados sobre o repositório institucional desenvolvido, visando corroborar ou refutar as hipóteses de desempenho da aplicação. Também serão apresentados gráficos de métricas de utilização de processador e memória do *backend* da aplicação, visando estabelecer os requisitos mínimos do sistema.

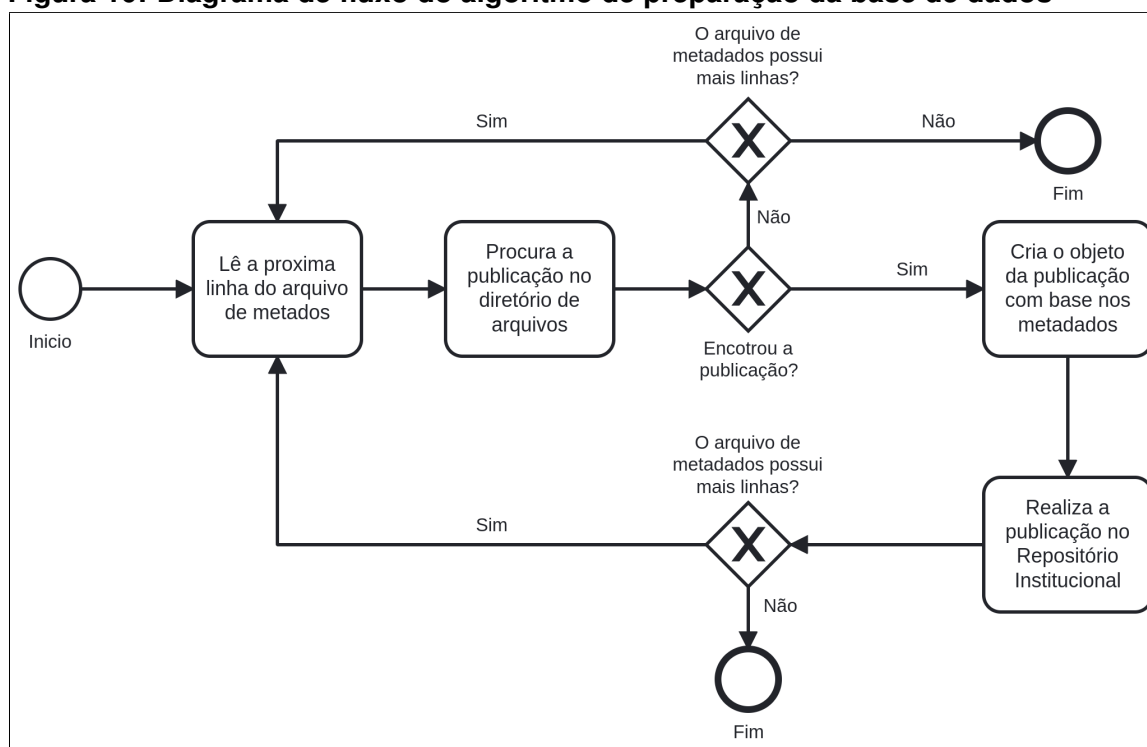
3.4.1 Preparação da base de dados

Para a realização dos testes no repositório institucional desenvolvido, foi realizado a preparação de uma base de dados com 3.500 publicações, tendo como base o *arXiv Dataset*, que consiste em um *dataset* gratuito de publicações acadêmicas.

Para realizar a preparação da base de dados, foi realizado o *download* de 3.500 publicações do *arXiv Dataset* em formato PDF, e desenvolvido um algoritmo na linguagem C# (mesma linguagem que foi desenvolvido o *backend* do repositório institucional), que realiza a leitura do arquivo de metadados disponibilizado pelo *arXiv*, e compara o identificador único presente no arquivo de metadados com os arquivos presentes na pasta em que foi realizado o *download* das publicações.

O arquivo de metadados disponibilizado pelo *arXiv*, consiste em um arquivo no formato JSON contendo uma lista de dados, como o identificador único da publicação, o nome do usuário que publicou o artigo, o nome dos autores, o título da publicação, comentários, informações sobre a revista em que o artigo foi publicado, o número DOI (*Digital Object Identifier*), o resumo, as categoriais e as revisões da publicação.

Figura 10: Diagrama de fluxo do algoritmo de preparação da base de dados



A Figura 10 apresenta o funcionamento do algoritmo, que percorre o arquivo de metadados, e procura pela publicação dentro da pasta onde foi realizado o *download* das publicações. O algoritmo identifica os arquivos por meio de seu nome, visto que os arquivos possuem um padrão de nomenclatura, que consiste no identificador único concatenado ao número da revisão.

Ao encontrar uma combinação, o algoritmo realiza a transformação dos metadados do *arXiv* em um formato que o repositório institucional possa processar, e realiza a submissão da publicação dentro do repositório institucional, passando pelo mesmo processo que uma publicação realizada por um usuário real, porém de forma automatizada.

3.4.2 Teste de Consulta por Publicações

Para obter o tempo médio de consultas por publicações dentro do repositório institucional, foi utilizado da ferramenta Apache JMeter, que consiste em um *software* para a realização de testes de carga e estresse.

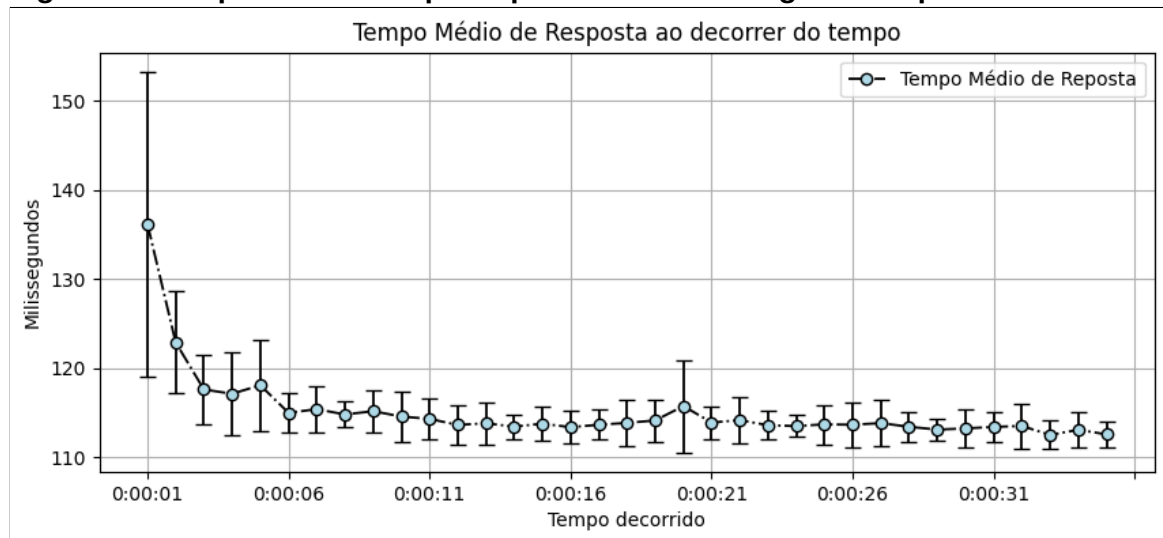
A partir do Apache JMeter, foi preparado um cenário de teste onde uma consulta pelo termo "*Distribution of H2O*" foi realizada por mil vezes a API do repositório institucional. O termo utilizado na pesquisa poderia ser outro, e foi escolhido apenas pelo conhecimento que haveria publicações com termos semelhantes na base de dados, visto que não faria sentido realizar um teste de estresse sobre um termo que não possuísse resultados.

Figura 11: Métricas de consulta por publicações

| Pesquisa Avançada | | | | | | | | | | | |
|-------------------|-------------|-----------------------|--------------|---------------|------------|--------------|--------------|----------------|--------------|---------------|--------------|
| Round | Requisições | Tempo de Reposta (ms) | | | | | | | Throughput | Rede (KB/sec) | |
| | | Média | Mínima | Máxima | Mediana | Percentil 90 | Percentil 95 | Percentil 99 | Transação/s | Recebido | Enviado |
| 1 | 1000 | 114,28 | 109 | 875 | 113 | 116 | 120 | 127 | 8,74 | 177,16 | 7,59 |
| 2 | 1000 | 115,01 | 110 | 1163 | 113 | 116 | 118 | 129,97 | 8,68 | 176,01 | 7,54 |
| 3 | 1000 | 113,36 | 108 | 1213 | 112 | 114 | 116 | 130,99 | 8,8 | 178,52 | 7,65 |
| 4 | 1000 | 115,39 | 110 | 1233 | 114 | 116 | 118 | 126,98 | 8,65 | 175,42 | 7,52 |
| 5 | 1000 | 114,66 | 110 | 1175 | 113 | 116 | 118 | 129,97 | 8,71 | 176,54 | 7,57 |
| Médias | 1000 | 114,54 | 109,4 | 1131,8 | 113 | 115,6 | 118 | 128,982 | 8,716 | 176,73 | 7,574 |

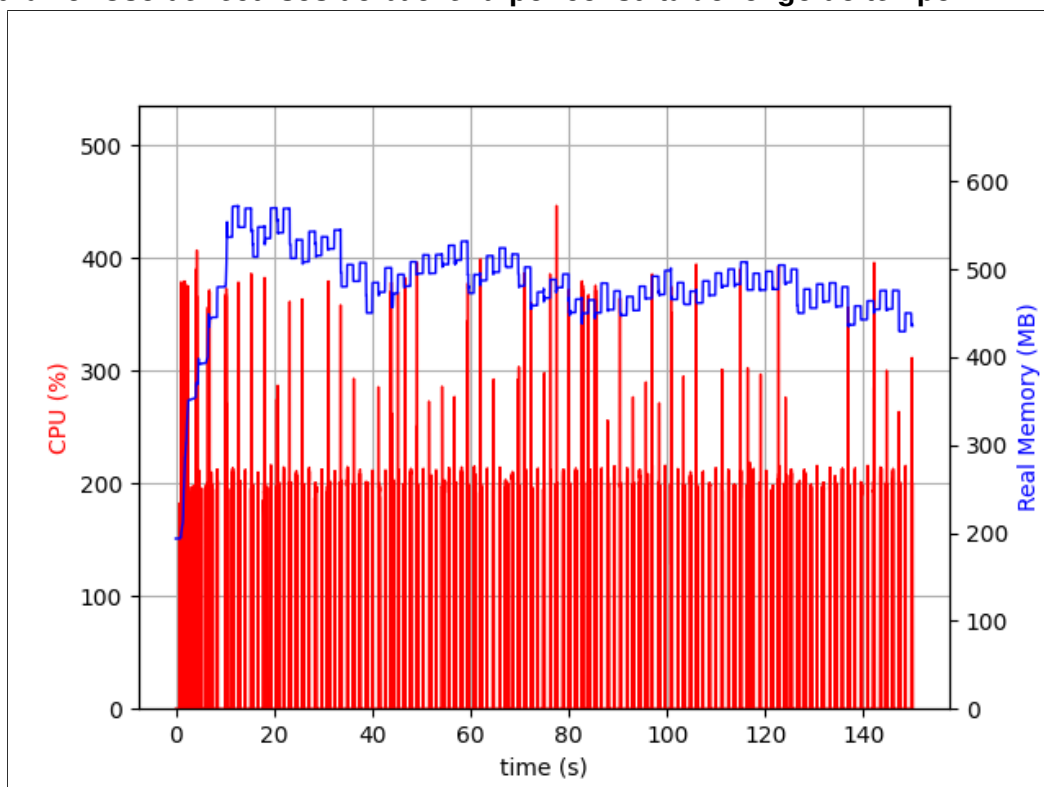
A Figura 11 apresenta as métricas resultantes do teste realizado com o Apache JMeter, nesta figura é possível visualizar que houve um tempo de resposta médio de 114,54 milissegundos por requisição, um tempo mínimo de 109,4 milissegundos, e um tempo máximo de 1131,8 milissegundos. Na figura também é possível visualizar os valores de mediana, percentil, transações por segundo, e informações da rede.

Figura 12: Tempo médio de reposta por consulta ao longo do tempo



A Figura 12 apresenta o tempo médio de respostas em milissegundos, dentro dos primeiros 31 segundos de execução do teste. Nesta figura é possível perceber que o tempo médio de resposta diminui ao longo do tempo, iniciando em 136,16 milissegundos, e reduzindo para valores entre 117 e 112 milissegundos ao longo da execução. É provável que o tempo médio por consulta esteja sendo reduzido devido ao banco de dados realizar o cache das consultas.

Figura 13: Uso de recursos do *backend* por consulta ao longo do tempo



A Figura 13 apresenta o uso de recursos de memória e processador do *backend* do repositório institucional ao longo dos primeiros 150 segundos de execução do teste. Estas métricas foram coletadas por meio da ferramenta *psrecord*⁵.

Nesta mesma figura é possível visualizar que durante o teste houve um pico de 559,62 MB de utilização de memória, mínima de 198,73 MB, e média de 471,09 MB. Já a utilização de processador teve picos acima de 400% visto que o teste foi executado em um processador com quatro núcleos, porém em grande parte do tempo se manteve acima de 200%, demonstrando que o *backend* está aproveitando os múltiplos núcleos do processador.

3.4.3 Teste de Submissão de Publicações

Da mesma forma que o teste de consulta por publicações, o teste de submissão de publicações também foi elaborado a partir do Apache JMeter. Para obter o tempo médio de submissão de publicação, foi preparado um cenário de teste onde uma mesma publicação com 12.000 palavras (cerca de 40 páginas de texto em português com fonte tamanho 12, considerando que cada página tenha 300 palavras) foi realizada por mil vezes a API do repositório institucional.

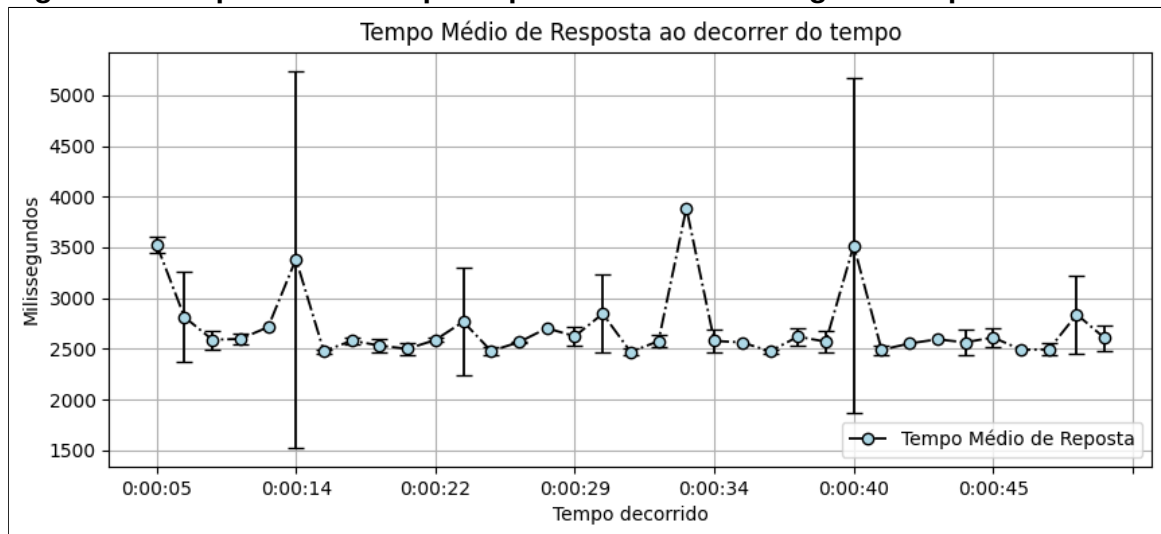
Figura 14: Métricas de submissão de publicações

| Submissão de Pesquisa | | | | | | | | | | | |
|-----------------------|-------------|-----------------------|--------|--------|---------|--------------|--------------|--------------|-------------|---------------|---------|
| Round | Requisições | Tempo de Reposta (ms) | | | | | | | Throughput | Rede (KB/sec) | |
| | | Média | Minima | Máxima | Mediana | Percentil 90 | Percentil 95 | Percentil 99 | Transação/s | Recebido | Enviado |
| 1 | 1000 | 2634,7 | 2085 | 8888 | 2514 | 2874 | 3063 | 3903,82 | 0,38 | 0,09 | 51,91 |
| 2 | 1000 | 2582,67 | 2420 | 7376 | 2455 | 2835 | 3027,95 | 3876,61 | 0,39 | 0,09 | 52,96 |
| 3 | 1000 | 2582,34 | 2547 | 5667 | 2574 | 2601 | 2613 | 2645 | 0,39 | 0,09 | 52,96 |
| 4 | 1000 | 2546,69 | 2321 | 5669 | 2538 | 2567 | 2575 | 2602,93 | 0,39 | 0,09 | 53,71 |
| 5 | 1000 | 2490,66 | 2136 | 5500 | 2480 | 2517 | 2524 | 2556,98 | 0,4 | 0,09 | 54,92 |
| Médias | 1000 | 2567,412 | 2301,8 | 6620 | 2512,2 | 2678,8 | 2760,59 | 3117,068 | 0,39 | 0,09 | 53,292 |

A Figura 14 apresenta as métricas obtidas através do teste, onde é possível visualizar que houve um tempo médio de resposta de 2,56 segundos, com um tempo mínimo de 2,30 segundos, e um tempo máximo de 6,62 segundos. A figura também apresenta os valores de mediana, percentil, transações por segundo, e informações de rede.

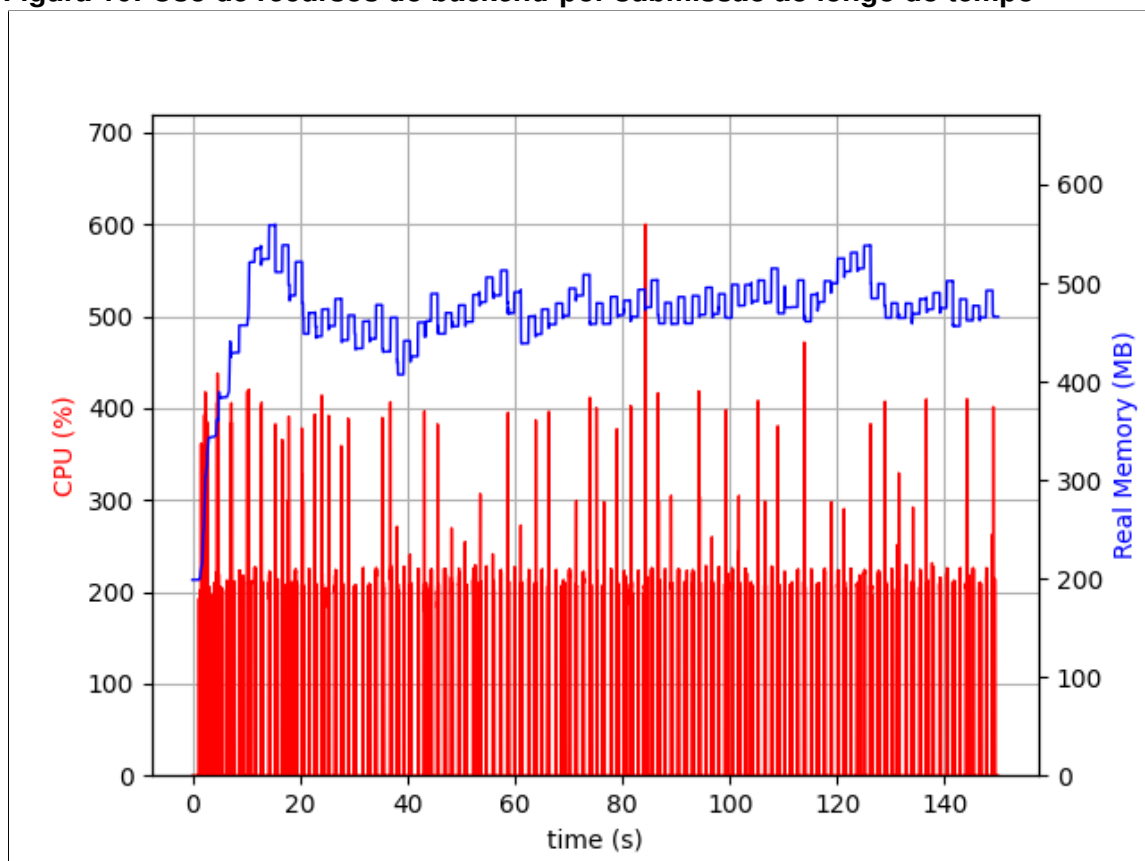
⁵<https://github.com/astrofrog/psrecord>

Figura 15: Tempo médio de resposta por submissão ao longo do tempo



A Figura 15 apresenta o tempo médio de resposta de submissão de publicação ao longo dos primeiros 50 segundos do teste, onde é possível visualizar que as primeiras publicações realizadas levam em média 3,52 segundos para serem concluídas, e este tempo logo em seguida é reduzido, e se mantém entre 2,5 e 2,8 segundos.

Figura 16: Uso de recursos do *backend* por submissão ao longo do tempo



Já a Figura 16 demonstra o uso de recursos de memória e processador do *backend* do repositório institucional, ao longo dos primeiros 150 segundos da execução do teste de submissão de publicações. Da mesma forma que o gráfico de uso de recurso por consultas, estas métricas também foram coletadas por meio da ferramenta *psrecord*.

Nesta mesma figura é possível visualizar que durante a execução do teste houve uma utilização máxima de 559,62 MB de memória, mínima de 198,73 MB, e média de 471,09 MB. Também é perceptível que a curva de utilização de memória parece um pouco quadriculada, isso ocorre devido ao tempo necessário para realizar o *upload* do arquivo PDF aos servidores da Digital Ocean, que é o *Cloud Provider* utilizado para o armazenamento das publicações.

O gráfico também demonstra que o *backend* do repositório institucional está aproveitando dos múltiplos núcleos do processador para processar as requisições, porém na maior parte do tempo a carga parece estar sendo dividida em dois núcleos.

3.4.4 Requisitos Mínimos do Sistema

Com base nas métricas coletadas através dos testes de carga e estresse apresentados no tópico anterior, foram definidos os requisitos mínimos de *hardware* para realizar a execução do repositório institucional desenvolvido.

3.4.4.1 Frontend

Por se tratar de um site estático desenvolvido em React, o *frontend* desenvolvido para o repositório institucional não possui requisitos mínimos de CPU ou memória, podendo ser hospedado em qualquer provedor de nuvem que ofereça a opção de hospedagem de sites estáticos, realizando o cache dos arquivos em CDN (*Content Delivery Network*).

Alguns provedores de nuvem fornecem a opção de realizar este tipo de hospedagem forma gratuita, em conjunto com HTTPS também gratuito, como no caso da Digital Ocean App Platform⁶ e do Vercel⁷.

⁶<https://www.digitalocean.com/products/app-platform>

⁷<https://vercel.com/>

Já no quesito de compatibilidade com navegadores, o *frontend* desenvolvido segue os mesmos padrões de compatibilidade do Vite 3, necessitando de suporte a *JavaScript* moderno, módulos *JavaScript* nativos, importação dinâmica de módulos nativos, e suporte ao **import.meta**.

1. **Chrome** ≥ 87
2. **Firefox** ≥ 78
3. **Safari** ≥ 13
4. **Edge** ≥ 88

O suporte a navegadores legados poderia ser implementado por meio de tecnologias como o Polyfill⁸, porém foi optado por oferecer suporte por padrão somente a navegadores modernos.

3.4.4.2 Backend

Com base nos dados coletados durante os testes de carga e estresse, e nos requisitos do .NET 7, foram elaborados dois conjuntos de requisitos para a execução do *backend* do repositório institucional, sendo os requisitos mínimos, e os recomendados.

Quadro 14: Requisitos mínimos de *backend*

| Recurso | Mínimo | Recomendado |
|---------------|------------------------|-------------|
| CPU | 1 vCPU | 2 vCPU+ |
| Memória RAM | 512 MB | 1 GB+ |
| Armazenamento | 4.5 GB | 4.5 GB+ |
| OS | Debian 10+ x64 | |
| | Ubuntu 18.04+ x64 | |
| | Alpine Linux 3.15+ x64 | |

O Quadro 14 apresenta os requisitos mínimos e recomendados de CPU, memória RAM, armazenamento e sistema operacional. Devido ao uso de algumas

⁸<https://polyfill.io>

bibliotecas como o *SkiaSharp* para a manipulação de imagens, o *backend* somente possui suporte a sistemas operacionais Linux.

Também é necessário ressaltar que os requisitos mínimos apenas devem ser utilizados em ambientes de testes. Em ambientes de produção com usuários ativos utilizando o sistema, deve ser utilizado de requisitos iguais ou superiores aos recomendados.

3.4.4.3 Banco de dados

Como requisito de banco de dados, deve ser utilizado do PostgreSQL 14 ou superior, já os requisitos de *hardware* variam de acordo com o número de usuários ativos utilizando o sistema.

Quadro 15: Requisitos de banco de dados

| Recurso | Mínimo | Recomendado |
|---------------|--------|-------------|
| CPU | 1 vCPU | 2 vCPU+ |
| Memória RAM | 1 GB | 4 GB+ |
| Armazenamento | 10 GB | 25 GB+ |

O Quadro 15 apresenta os requisitos de *hardware* mínimos e recomendados para o banco de dados do repositório institucional. Os requisitos mínimos devem ser utilizados somente em ambientes de testes, onde existam ao máximo 22 conexões simultâneas ao banco de dados.

Em ambientes de produção, deve ser utilizado de requisitos iguais ou superiores aos recomendados, também sendo indicado a utilização de um sistema de *pool* de conexões como o *pgBouncer*⁹ para melhores resultados.

⁹<https://www.pgouncer.org/>

REFERÊNCIAS

- AKELLA, R.; TAMIRISA, A.; KUNANI, S.; MUTHIYALU, B. 2022. **Enterprise Application Development with C# 10 and .NET 6**: become a professional .net developer by learning expert techniques for building scalable applications. [S.l.]: Packt Publishing. ISBN 9781803239125.
- ALI, M.; LOAN, F. A.; MUSHATQ, R. 2018. **Open Access Scientific Digital Repositories : an analytical study of the open doar**. In: INTERNATIONAL SYMPOSIUM ON EMERGING TRENDS AND TECHNOLOGIES IN LIBRARIES AND INFORMATION SERVICES (ETTLIS), 2018. **Anais**. [S.l.: s.n.]. p 213–216.
- ANJOS, C. 2020. **GED - GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS: dificuldades na implantação do ged nas empresas médias e pequenas**. [Online]. Recife, PE, Brazil.
- 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.
- ASADI, S.; ABDULLAH, R.; YAH, Y.; NAZIR, S. 2019. **Understanding Institutional Repository in Higher Learning Institutions: a systematic literature review and directions for future research**. IEEE Access [S.l.] v.7 p 35242–35263.
- AVEDON, D. M. 2002. **GED de A a Z. Tudo sobre gerenciamento eletrônico de documentos**. [S.l.]: Elsevier Editora Ltda.

- AWAD, E. 2007. **Knowledge Management**. [S.l.]: Pearson Education. ISBN 9788131714034.
- BANERJEE, K.; REESE, T. 2018. **Building Digital Libraries, Second Edition**. [S.l.]: American Library Association. (How-To-Do-It Manuals). ISBN 9780838916353.
- BASHIR, S.; GUL, S.; BASHIR, S.; NISA, N. T.; GANAIE, S. A. 2021. **Evolution of institutional repositories: managing institutional research output to remove the gap of academic elitism**. Journal of Librarianship and Information Science [S.l.].
- BECERRA-FERNANDEZ, I.; SABHERWAL, R. 2014. **Knowledge Management: systems and processes**. [S.l.]: Taylor & Francis. ISBN 9781317503033.
- BORGHOFF, U.; RÖDIG, P.; SCHEFFCZYK, J.; SCHMITZ, L. 2007. **Long-Term Preservation of Digital Documents: principles and practices**. [S.l.]: Springer Berlin Heidelberg. ISBN 9783540336402.
- CALLICOTT, B.; SCHERER, D.; WESOLEK, A. 2015. **Making Institutional Repositories Work**. [S.l.]: Purdue University Press. (Charleston Insights in Library, Archival, and Information Sciences). ISBN 9781612494234.
- CAPLAN, P.; American Library Association. 2003. **Metadata Fundamentals for All Librarians**. [S.l.]: American Library Association. (ALA editions). ISBN 9780838908471.
- CARVALHO, V. 2017. **PostgreSQL: banco de dados para aplicações web modernas**. [S.l.]: Casa do Código. ISBN 9788555192562.
- 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.
- CONARQ. 2022. **e-ARQ Brasil: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos**. Rio de Janeiro, RJ: CONARQ. ISBN 978-85-7009-007-2.

- 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.
- JIN, Q.; SANDBERG, J. 2019. **Crafting Linked Open Data to Enhance the Discoverability of Institutional Repositories on the Web**. Qualitative and Quantitative Methods in Libraries [S.l.] v.7 n.4 p 595–606.
- 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.
- LYNCH, C. A. 2003. **Institutional Repositories: essential infrastructure for scholarship in the digital age**. Libraries and the Academy [S.l.] v.3 n.2 p 327–336.
- METTAI khaled. 2019. **Discovery Tools to Enhance Resources Find ability in the Institutional Repositories**: an overview. [Online].
- 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.
- NIELSEN, J. 1994. **Usability Engineering**. [S.l.]: Elsevier Science. (Interactive Technologies). ISBN 9780125184069.

- PATIL, A.; RANGARAO, D.; SEIPP, H.; LASOTA, M.; SANTOS, R. dos; MARKOVIC, R.; CASEY, S.; BOLLERS, S.; GUCER, V.; LIN, A. et al. 2020. **Cloud Object Storage as a Service: ibm cloud object storage from theory to practice - for developers, it architects and it specialists**. [S.l.]: IBM Redbooks. ISBN 9780738442457.
- 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.
- POSTGRESQL. 2022. **Full Text Search**. [Online]. Disponível em: <<https://www.postgresql.org/docs/current/>>. Acessado em: 7 de Agosto.
- PostgreSQL Global Development Group. 2011. **PostgreSQL 9.0 Official Documentation - Volume I. the SQL Language**. [S.l.]: Fultus Corporation. n.v. 1. (Linux Documentation Library). ISBN 9781596822467.
- SAINI, O. P. 2018. **The Emergence of Institutional Repositories: a conceptual understanding of key issues through review of literature**. [Online]. Lincoln, Nebraska, United States.
- SEHGAL, N.; BHATT, P.; ACKEN, J. 2019. **Cloud Computing with Security: concepts and practices**. [S.l.]: Springer International Publishing. ISBN 9783030246129.
- SYMULEVICH, A.; HAMILTON, M. 2022. **Using Open Access Institutional Repositories to Save the Student Symposium during the COVID-19 Pandemic**. Information Technology and Libraries [S.l.] v.41 n.1.
- THAKKAR, M. 2020. **Building React Apps with Server-Side Rendering: use react, redux, and next to build full server-side rendering applications**. [S.l.]: Apress. ISBN 9781484258699.
- UZAYR, S. 2022. **Mastering UI Mockups and Frameworks: a beginner's guide**. [S.l.]: CRC Press. (Mastering Computer Science). ISBN 9781000583106.

VALENTE, M. T. 2020. **Engenharia de Software Moderna**: princípios e práticas para desenvolvimento de software com produtividade. [S.l.]: Independente. ISBN 9786500000771.

VITE. [Online]. <https://vitejs.dev/>. Accessed: 2022-07-03.

WAZLAWICK, R. 2019. **Engenharia de software**: conceitos e práticas. [S.l.]: Elsevier Editora Ltda. ISBN 9788535292732.

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.

ZERVAS, M.; KOUNOUDIS, A.; ARTEMI, P.; GIANNOULAKIS, S. 2019. **Next generation Institutional Repositories: the case of the cut institutional repository ktisis**. Procedia Computer Science [S.l.] v.146 p 84–93. 14th International Conference on Current Research Information Systems, CRIS2018, FAIRness of Research Information.

APÊNDICES

APÊNDICE A

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

Quadro 16: Orçamento

| Item | Quantidade | Valor Unitário | Valor Total |
|-------------------------------------|------------|----------------|-------------|
| Capa | 2 | R\$ 25,00 | R\$ 50,00 |
| Espiraís | 3 | R\$ 5,00 | R\$ 15,00 |
| Horas trabalhadas | 800 | R\$ 30,00 | R\$ 24,000 |
| Impressões | 1000 | R\$ 0,30 | R\$ 300,00 |
| Serviços de armazenamento em nuvem | 1 | R\$ 26,52 | R\$ 26,52 |
| Serviços de Banco de Dados em nuvem | 1 | R\$ 79,55 | R\$ 79,55 |
| Serviços de hospedagem em nuvem | 1 | R\$ 53,03 | R\$ 53,03 |
| Serviços de Designer | 1 | R\$ 350,00 | R\$ 350,00 |

APÊNDICE B

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 17: Cronograma

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

| | |
|--|-----------|
| | Proposto |
| | Realizado |