

LUCAS MACHADO ALF

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

Três de Maio

2022

LUCAS MACHADO ALF

**DESENVOLVIMENTO DE UM *SOFTWARE* DE REPOSITÓRIO INSTITUCIONAL
DE PESQUISAS ACADÊMICAS 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: Orçamento.....	17
Quadro 2: Cronograma.....	18
Quadro 3: Exemplo de elementos do Dublin Core	22
Quadro 4: Sistemas Relacionados.....	32
Quadro 5: Trabalhos Relacionados	53

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	8
1 PLANO DE ESTUDO E PESQUISA	10
1.1 TEMA	10
1.1.1 Delimitação do Tema	10
1.2 OBJETIVO GERAL	10
1.2.1 Objetivos Específicos	11
1.3 JUSTIFICATIVA	11
1.4 PROBLEMA	12
1.5 HIPÓTESES	13
1.6 METODOLOGIA	13
1.6.1 Abordagem	14
1.6.2 Procedimentos	14
1.6.2.1 Pesquisa Bibliográfica	15
1.6.2.2 Pesquisa Ação	15
1.6.3 Validação das Hipóteses	15
1.6.4 Técnicas	16
1.7 ORÇAMENTO	17
1.8 CRONOGRAMA DE ATIVIDADES	17
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 ÁREA DE NEGOCIO	19
2.1.1 Gestão do Conhecimento	19
2.1.2 Repositórios Institucionais	20
2.1.3 Padrão Dublin Core para metadados descritivos	21
2.1.4 Open Archival Information System (OAIS)	22
2.2 PROCESSOS	22
2.2.1 Engenharia de Software	22
2.2.2 Prototipação de Software	23
2.3 TECNOLOGIAS	23
2.3.1 Cloud Computing	23
2.3.2 Cloud Object Storage	23
2.3.3 Microsoft .NET 6	23
2.3.4 React	24

2.3.5	Vite	25
2.3.6	PostgreSQL	27
2.3.6.1	Full-Text Search	28
2.3.6.2	GIN Index	30
2.4	SISTEMAS RELACIONADOS	31
2.5	TRABALHOS RELACIONADOS	33
2.5.1	BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados	33
2.5.2	Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência	34
2.5.3	Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital	35
2.5.4	Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca	36
2.5.5	O mapeamento dos repositórios institucionais brasileiros: perfil e desafios	37
2.5.6	Encontrabilidade da informação no repositório institucional da Unesp: um estudo de eye tracking em dispositivos móveis	39
2.5.7	The Emergence of Institutional Repositories: A Conceptual Understanding of Key Issues through Review of Literature	40
2.5.8	Why So Many Repositories? Examining the Limitations and Possibilities of The Institutional Repositories (IR) Landscape	42
2.5.9	Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS	43
2.5.10	Using Open Access Institutional Repositories to Save the Student Symposium during the COVID-19 Pandemic	44
2.5.11	Understanding Institutional Repository in Higher Learning Institutions: A Systematic Literature Review and Directions for Future Research	46
2.5.12	Discovery Tools to Enhance Resources Findability in the Institutional Repositories: an overview	47
2.5.13	Crafting Linked Open Data to Enhance the Discoverability of Institutional Repositories on the Web	49
2.5.14	Open Access Scientific Digital Repositories : An Analytical Study of the Open DOAR	50
2.5.15	Evolution of institutional repositories: Managing institutional research output to remove the gap of academic elitism	51
	REFERÊNCIAS	54

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 6.0 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 deste projeto de 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 ultima parte, é apresentado a conclusão da pesquisa, detalhando os objetivos alcançados, validação das hipóteses, dificuldades e limitações encontradas durante o desenvolvimento, e propostas futuras de pesquisa.

1 PLANO DE ESTUDO E PESQUISA

1.1 TEMA

Desenvolvimento de um *software* de repositório institucional de pesquisas acadêmicas 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 6.0 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 projeto 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 acadêmico 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 acadêmico institucional.

Ao pesquisar sobre repositórios acadêmicos existentes em plataformas agregadoras como o OpenDOAR¹, foi constatado que a maioria dos repositórios nacionais utilizam como base o DSpace, um *software open source* escrito em Java para criação de repositórios acadêmicos institucionais, sendo minoria as instituições que utilizam um sistema de repositório institucional diferente deste.

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

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

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

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

¹<https://v2.sherpa.ac.uk/opensoar/>

acesso às publicações científicas produzidas dentro da instituição, 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. É possível manter um ambiente de infraestrutura em nuvem contendo os serviços de *backend*, *frontend*, banco de dados e armazenamento das publicações com um orçamento máximo de R\$ 500,00 por mês.
2. 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 5 segundos por consulta), mesmo em bases de dados com mais de 30 mil publicações.
3. 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 5 segundos), mesmo em publicações com mais de 200 páginas.

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 três hipóteses, sendo a primeira "O recurso de Full Text Search do banco de dados PostgreSQL pode ser utilizado como uma alternativa viável para realizar as consultas por publicações dentro do repositório acadêmico (menos de 5 segundos por consulta), mesmo em bases de dados com mais de 30 mil publicações".

Para realizar a validação desta hipótese será utilizado do arXiv Dataset², 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

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

35 mil publicações.

Após realizar a preparação da base de dados, serão desenvolvidas 5 consultas diferentes, que serão executadas 30 vezes cada, de forma a obter um tempo médio de execução por consulta. Todos os testes realizados serão executados utilizando a versão 14 do PostgreSQL, em ambiente Windows 11, com as seguintes configurações de hardware: Intel Core i3 9100, 16GB RAM DDR4 2400Mhz, 1TB SSD NVME WD Blue SN550.

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

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

1.6.4 Técnicas

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

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

Em tradução livre de Uzayr (2022) um *software mockup* consiste em um desenho de uma página ou aplicação web, que é desenvolvida para trazer vida a uma

ideia, permitindo que um designer possa examinar como diferentes elementos visuais trabalham juntos. Os *mockups* também permitem que as partes interessadas ou *Stakeholders* do projeto possam verificar como a interface irá parecer, enquanto sugerem mudanças apropriadas em relação a cores, imagens e estilos.

Também foram utilizadas as técnicas de análise estatística e comparativa para realizar a análise de preços em diferentes ambientes de nuvem, e a técnica de testes para verificar o funcionamento do *software* desenvolvido.

1.7 ORÇAMENTO

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

Quadro 1: Orçamento

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

1.8 CRONOGRAMA DE ATIVIDADES

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

Quadro 2: Cronograma

Atividades	2022						
	Mai	Jun	Jul	Ago	Set	Out	Nov
Desenvolvimento do projeto							
Entrega e validação do projeto							
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

2 FUNDAMENTAÇÃO TEÓRICA

Ao longo deste capítulo é apresentado o embasamento teórico do projeto, 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 DE NEGOCIO

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 enfeasse 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 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 às 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 de 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 à publicação da *Scholarly Publishing and Academic Resour-*

ces 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 3: Exemplo de elementos do Dublin Core

Identificador	Definição
<i>Title</i>	Nome dado ao 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 referencia 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 referencia 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 3 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)

Descrever...

2.2 PROCESSOS

2.2.1 Engenharia de Software

Descrever...

2.2.2 Prototipação de Software

Descrever...

2.3 TECNOLOGIAS

2.3.1 Cloud Computing

Descrever...

2.3.2 Cloud Object Storage

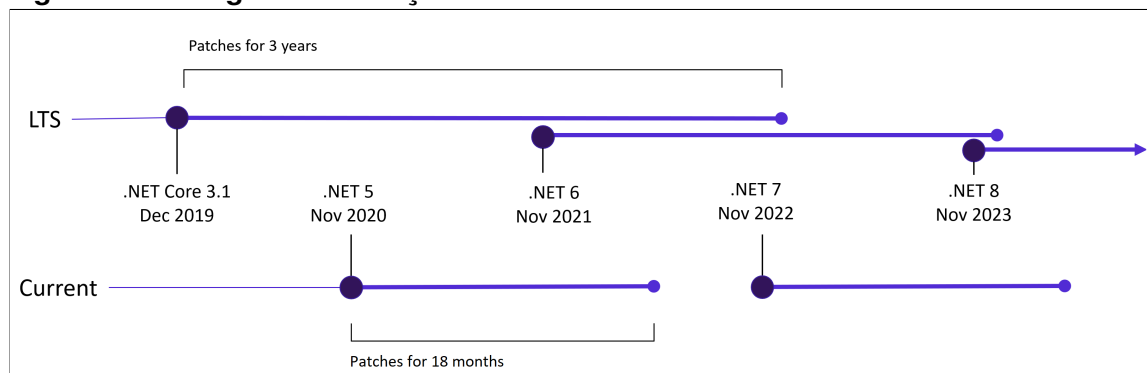
Descrever...

2.3.3 Microsoft .NET 6

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 1: Cronograma de lançamento de versões do .NET

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

A Figura 1 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 6.0, que possui um ecossistema vasto para o desenvolvimento de aplicações web, em nuvem, desktop, IoT e mobile.

O .NET 6.0 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, porem é recomendada pelo própria React.

Figura 2: 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 2 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 dentro de uma tag HTML.

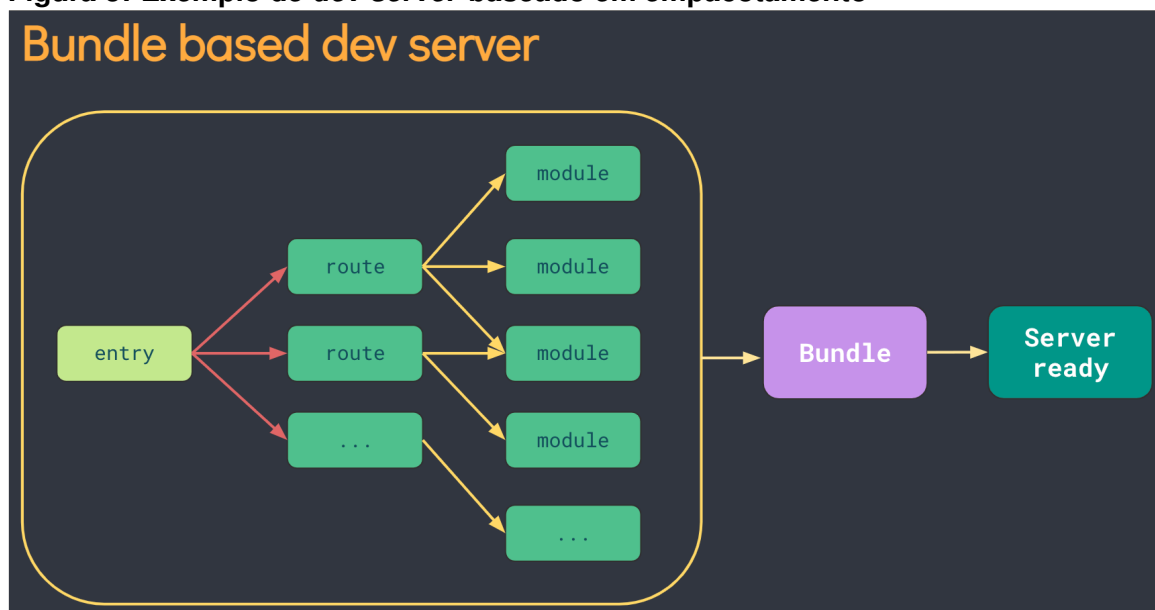
O React foi escolhido como ferramenta para realizar o desenvolvimento da interface gráfica do repositório institucional proposto, pois é uma biblioteca gratuita e de código aberto, e extramente popular tanto no mercado de trabalho quanto no meio acadêmico, além de contar com um vasto ecossistema de bibliotecas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

2.3.5 Vite

Conforme o site oficial da ferramenta VITE (2022), 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 da aplicação.

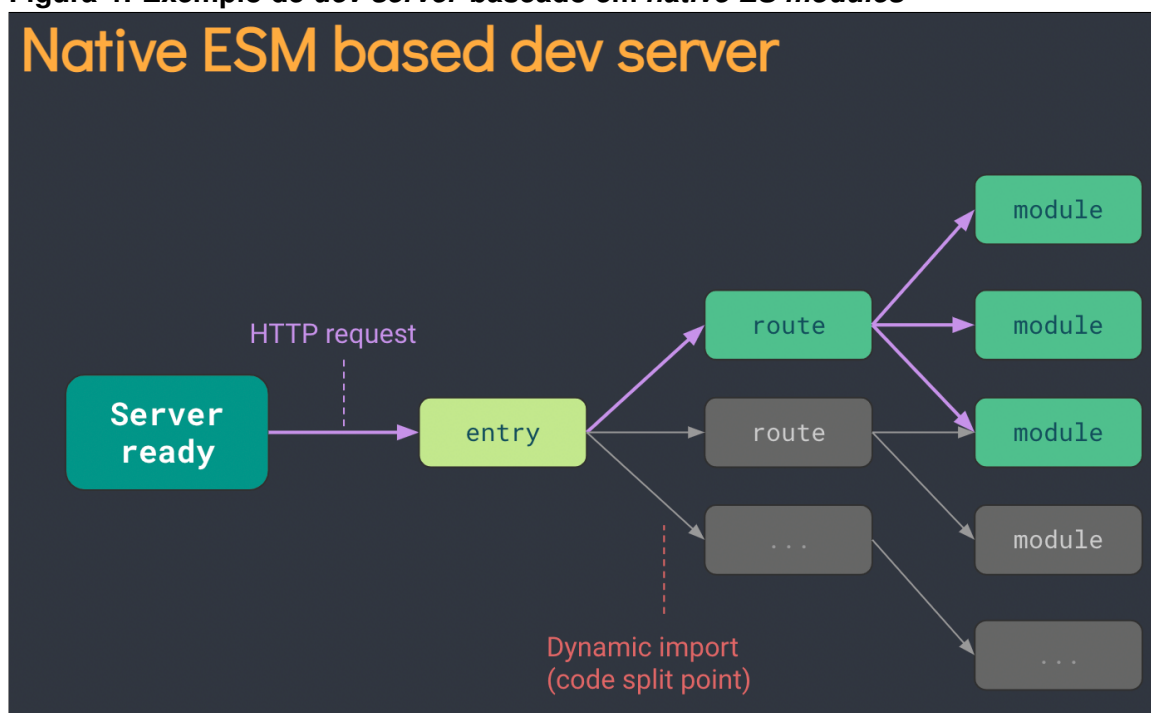
Figura 3: Exemplo de *dev server* baseado em empacotamento



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

Na Figura 3 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 4: Exemplo de *dev server* baseado em *native ES modules*



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

Já a Figura 4 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.

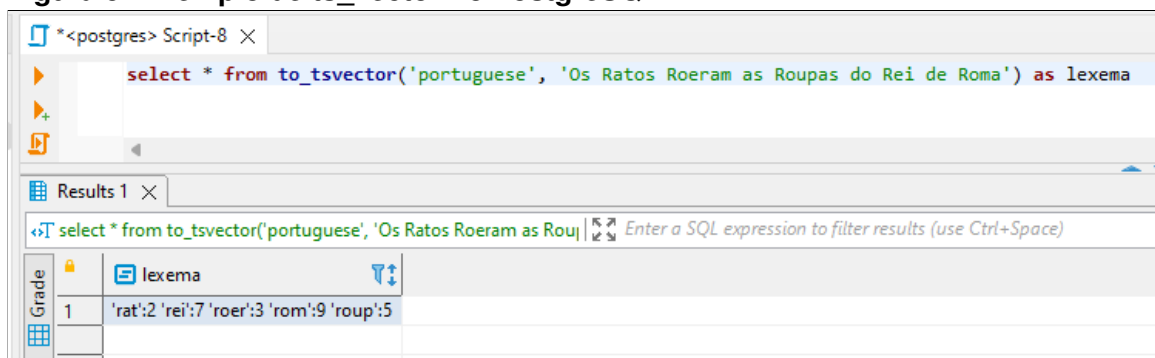
Diferente dos operadores de busca mais comuns presentes em bancos de dados relacionais como o *like* e *ilike*, o recurso de *Full Text Search* fornece suporte a um contexto linguístico, permitindo por exemplo, a busca por palavras derivadas como no caso de *satisfies* e *satisfy* do inglês, algo que 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 pes-

quisa.

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*, porem em um formato normalizado.

Como demonstrado na Figura 5 este processo de normalização envolve deixar todos o texto em minúsculo, 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 5: Exemplo de ts_vector no PostgreSQL



The screenshot shows a PostgreSQL script editor with a query and its results. The query is: `select * from to_tsvector('portuguese', 'Os Ratos Roeram as Roupas do Rei de Roma') as lexema`. The results table has one column named 'lexema' and one row containing the normalized text: `'rat':2 'rei':7 'roer':3 'rom':9 'roup':5`.

Grade	lexema
1	'rat':2 'rei':7 'roer':3 'rom':9 'roup':5

Fonte: Figura do autor

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 do 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 reação 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 4 é apresentado um resumo de sistemas relacionados a esta pesquisa, contendo o nome do sistema, ano de lançamento, tecnologias, empresa mantenedora e licença, e site de acesso.

Quadro 4: 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, porem 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 moveis, 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 moveis; 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 moveis.

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

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

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

no primeiro momento, pois é um dos elementos principais dentro do repositório.

Sendo assim, foi possível afirmar que no repositório analisado, elementos não relevantes em dispositivos moveis possuem maior destaque do que outros elementos que possuem maior importância, como por exemplo a função de auto-arquivamento, sendo um recurso muito pouco utilizado a partir de dispositivos moveis, e que possui uma visibilidade maior que a caixa de pesquisa.

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

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

2.5.7 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 provem contexto para pesquisas. Ajudando a refinar ideias, conhecer especificações do método de pesquisa, e trazer clareza sobre diversos fatores que cercam o surgimento, motivação, custos, seleção de *software* e uma perspectiva global de sobre os repositórios institucionais.

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

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

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

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

confiança que a comunidade possui em depositar suas publicações nestes repositórios.

Em sua pesquisa é citado que a ideia das instituições hospedarem seu próprio repositório institucional de forma local já está obsoleta, e deveriam ser substituídos por alternativas viáveis, fazendo uma relação ao tempo que as instituições também hospedavam os seus próprios servidores de email, onde hoje em dia são utilizados serviços baseados em nuvem como os fornecidos pela Google e Microsoft.

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

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

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

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 vídeo quanto de áudio, 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 áudio ou vídeo. 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 áudio e vídeo, permitindo o upload em massa das publicações de uma única vez.

Como o repositório da instituição era feito em 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 aderiram 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 relevância 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 escopo 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 polícias 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, referencias 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.

No Quadro 5 é apresentado um resumo dos trabalhos relacionados citados nesta pesquisa, contendo o autor, título, objetivo, metodologia e resultados de cada trabalho.

Quadro 5: Trabalhos Relacionados

Autor	Título	Objetivo	Metodologia	Resultados
Symulewicz; 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	Realizar 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 Cyprus University of Technology 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	Retrair 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.
Meital (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.
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 quadrípolar.	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 OpenDOAR.	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.

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.
- 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.
- 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.I.].

BECERRA-FERNANDEZ, I.; SABHERWAL, R. 2014. **Knowledge Management: systems and processes.** [S.I.]: Taylor & Francis. ISBN 9781317503033.

CALLICOTT, B.; SCHERER, D.; WESOLEK, A. 2015. **Making Institutional Repositories Work.** [S.I.]: 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.I.]: American Library Association. (ALA editions). ISBN 9780838908471.

CARVALHO, V. 2017. **PostgreSQL:** banco de dados para aplicações web modernas. [S.I.]: 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.I.] v.1 n.1 p 65–76.

FERNANDES, W. M. 2019. **Encontrabilidade da informação no repositório institucional da Unesp:** um estudo de eye tracking em dispositivos móveis. [Online]. Marília, SP, Brazil.

GARCIA, R. M. 2019. **Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: relato de experiência.** PragMATIZES [S.I.] 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.I.] 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.
- PEREIRA, C. M.; MOREIRA, W.; SEGUNDO, J. E. S. 2021. **Classificação facetada: 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.
- 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.

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

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.; KOUNOUEDES, 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.