

LUCAS MACHADO ALF

**ANÁLISE E DESENVOLVIMENTO DE UM REPOSITÓRIO INSTITUCIONAL DE
TRABALHOS ACADÊMICOS.**

Três de Maio
2022

LUCAS MACHADO ALF

**ANÁLISE E DESENVOLVIMENTO DE UM REPOSITÓRIO INSTITUCIONAL DE
TRABALHOS ACADÊMICOS.**

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

Orientadores:

Dr. FirstName LastName

Dr. FirstName LastName

Co-orientador:

Dr. FirstName LastName

Dr. FirstName LastName

Três de Maio

2022

TERMO DE APROVAÇÃO

LUCAS MACHADO ALF

Análise e desenvolvimento de um Repositório Institucional de Trabalhos Acadêmicos.

Relatório aprovado como requisito parcial para obtenção do título de **Bacharel em Sistemas de Informação** concedido pela Faculdade de Sistemas de Informação da Sociedade Educacional Três de Maio, pela seguinte Banca examinadora:

Orientador: Prof. Name, Dr.

Faculdade de Sistemas de Informação da SETREM

Name, Dr.

Faculdade de Sistemas de Informação da SETREM

Name, M.Sc.

Faculdade de Sistemas de Informação da SETREM

Profa. Vera Lúcia Lorensen Benedetti, M.Sc.

Coordenação do Curso Bacharelado em Sistemas de Informação

Faculdade de Sistemas de Informação da SETREM.

Três de Maio, 08 de Agosto de 2019.

RESUMO

O resumo vai aqui...

Palavras-Chave: Sistemas de Informação. Deep Learning. Agricultura. Revisão Sistemática da Literatura.

ABSTRACT

The abstract goes here ...

Key Words: Information Systems. Deep Learning. Agriculture. Systematic Literature Review.

LISTA DE FIGURAS

Figura 1: Simple diagram.....	17
-------------------------------	----

LISTA DE QUADROS

LISTA DE CÓDIGOS

Código 1:	Exemplo de código em Python.	22
Código 2:	Exemplo de código em C++.	23

LISTA DE SIGLAS

ACM	Association for Computing Machinery
AWS	Amazon Web Services
IBGE	Brazilian Institute of Geography and Statistics
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things

SUMÁRIO

1 PLANO DE ESTUDO E PESQUISA	13
1.1 TEMA	13
1.1.1 Delimitação do Tema	13
1.2 OBJETIVO GERAL	13
1.2.1 Objetivos Específicos	13
1.3 JUSTIFICATIVA	14
1.4 PROBLEMA	15
1.5 HIPÓTESES	16
1.6 METODOLOGIA	16
1.6.1 Abordagem	16
1.6.2 Procedimentos	16
1.6.3 Técnicas	16
1.6.4 Validação das Hipóteses	16
1.7 ORÇAMENTO	16
1.8 CRONOGRAMA DE ATIVIDADES	16
2 FUNDAMENTAÇÃO TEÓRICA	17
2.1 ÁREA DOS NEGÓCIOS	17
2.2 FUNDAMENTOS DA ÁREA	17
2.3 TRABALHOS RELACIONADOS	17
2.3.1 BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados	18
2.3.2 Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: Relato de Experiência	19
2.3.3 Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital	20
3 RESULTADOS	22
3.1 APRESENTAÇÃO DA EMPRESA E HISTÓRICO	22
3.2 DESENVOLVIMENTO	22
3.3 EXPERIMENTOS	23

INTRODUÇÃO

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

LARCC (2019)

De acordo com Vogel; Griebler et al. (2017), bla bla ...

Alguns autores preferem incluir figuras e outros não (GRIEBLER; VOGEL; MARON; MALISZEWSKI; SCHEPKE; FERNANDES, 2018).

Existem alguns trabalhos de conclusão de curso desenvolvidos no LARCC (HENTGES; THOMÉ, 2013, MALISZEWSKI; BAUM, 2017, ROCKENBACH; ANDERLE, 2017, STEIN, 2018)

How to PhD thesis appear, simples test (GRIEBLER, 2016, p. 23).

1 PLANO DE ESTUDO E PESQUISA

1.1 TEMA

Análise e desenvolvimento de um Repositório Institucional de Trabalhos Acadêmicos.

1.1.1 Delimitação do Tema

A delimitação do tema se dará como a Análise e desenvolvimento de um Repositório Institucional de Trabalhos Acadêmicos, tendo como objetivo reunir e preservar as publicações acadêmicas e científicas produzidas em âmbito universitário, além de unificar o processo de publicações e correções por parte dos orientadores em uma única plataforma.

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

1.2 OBJETIVO GERAL

Desenvolver um repositório institucional de trabalhos acadêmicos que envolva o processo de correções e envio de revisões de publicações dentro da própria plataforma.

1.2.1 Objetivos Específicos

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

4. Investigar mecanismos de busca e análise textual de documentos. (Ex. ElasticSearch e Full Text Search).
5. Elicitar os requisitos para o desenvolvimento do repositório acadêmico.
6. Definir as tecnologias que serão utilizadas para o desenvolvimento (Ex: banco de dados, linguagens e frameworks).
7. Desenvolver uma aplicação web de repositório acadêmico.
8. Avaliar a usabilidade do repositório acadêmico desenvolvido.
9. 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 processo de divulgação científica é essencial para a propagação do conhecimento, e para a motivação dos acadêmicos envolvidos, tendo isto em mente, é visto a necessidade da existência de plataformas, ou meios onde acadêmicos possam interagir de forma a publicar seus artigos e projetos de pesquisa para que a comunidade, ao menos interna a instituição, possa ter acesso ao material desenvolvido, permitindo um maior engajamento entre os pesquisadores, e possíveis leitores das suas pesquisas, de preferência de forma gratuita e com fácil acesso.

Conforme (PORTO, 2009, p. 15), muitas instituições de ciência e tecnologia da informação acabam por não dispor de uma estrutura profissionalizada de comunicação, suporte e divulgação. Muitos institutos de pesquisas e universidades não contemplam a comunicação em seu organograma funcional, recorrendo a improvisações na hora da comunicação ou disponibilização de meios de acesso e divulgação de seus projetos.

Segundo o mesmo autor, em geral os sites de divulgação de pesquisa científica ainda se caracterizam pela apresentação de conteúdo estático, não tendo grande enfoque em recursos que permitam a interação entre os usuários, muito menos entre os pesquisadores e seus orientadores.

Tendo isto em mente, este trabalho busca desenvolver uma plataforma de repositório acadêmico que possibilite a interação entre os acadêmicos e orientadores, englobando o processo de correção e envio de revisões de publicações dentro da própria plataforma, de forma a abstrair o processo tradicional de correção e envio de revisões, onde o acadêmico precisa realizar a troca de diversos *emails* com seus orientadores, enviando em anexo diferentes revisões de sua pesquisa.

Desta forma, o repositório acadêmico proposto não será apenas um site de exibição de conteúdo estático, mas sim uma plataforma que englobe diferentes processos necessários para o desenvolvimento de pesquisas científicas, como o envio de revisões e correções, o armazenamento e versionamento dos arquivos, a publicação de artigos, e a pesquisa por documentos publicados. Além de possuir outros recursos como a possibilidade de interação entre os usuários da plataforma, por meio de troca de comentários e reações as pesquisas.

1.4 PROBLEMA

Como problema, foram ressaltados dois pontos principais que motivaram a escolha do tema desta pesquisa, sendo o primeiro a necessidade de reunir as publicações acadêmicas produzidas dentro de instituições de ensino superior em um repositório institucional, que fornece rápido acesso às publicações científicas produzidas dentro da instituição, tanto de forma interna para os estudantes da instituição, quanto de forma externa caso a instituição opte por disponibilizar o seu acervo de forma pública.

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

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

1.5 HIPÓTESES

1. O recurso de Full Text Search do banco de dados PostgreSQL pode ser utilizado como uma alternativa viável para realizar as consultas por publicações dentro do repositório acadêmico (menos de 3 segundos por consulta).
2. O processo de 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 relatórios com mais de 200 páginas.

1.6 METODOLOGIA

1.6.1 Abordagem

1.6.2 Procedimentos

1.6.3 Técnicas

1.6.4 Validação das Hipóteses

1.7 ORÇAMENTO

1.8 CRONOGRAMA DE ATIVIDADES

2 FUNDAMENTAÇÃO TEÓRICA

2.1 ÁREA DOS NEGÓCIOS

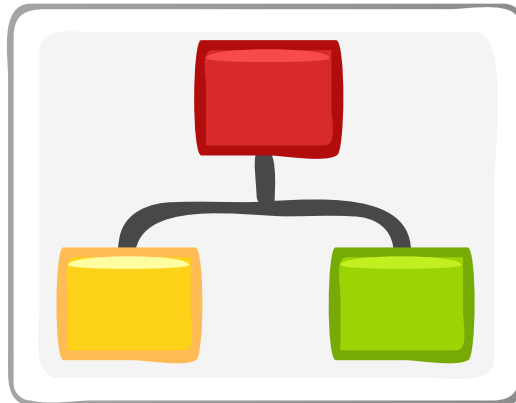
2.2 FUNDAMENTOS DA ÁREA

Equação 2.1 é um exemplo de uma equação em Latex:

$$h_t = f(W^{(hh)}h_{t-1} + W^{(hx)}x_t). \quad (2.1)$$

Figura1 é um exemplo para incluir uma figura.

Figura 1: Simple diagram



Fonte: Retirado de LARCC (2019).

Griebler; Hoffmann et al. (2018)

McCool, Robison e Reinders (2012)

2.3 TRABALHOS RELACIONADOS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3 RESULTADOS

Os experimentos irão nesta seção.

3.1 APRESENTAÇÃO DA EMPRESA E HISTÓRICO

3.2 DESENVOLVIMENTO

Código 1 apresenta um exemplo de código em Python.

Código 1: Exemplo de código em Python.

```
1 import numpy as np
2
3 def incmatrix(genl1,genl2):
4     m = len(genl1)
5     n = len(genl2)
6     M = None #to become the incidence matrix
7     VT = np.zeros((n*m,1), int) #dummy variable
8
9     #compute the bitwise xor matrix
10    M1 = bitxormatrix(genl1)
11    M2 = np.triu(bitxormatrix(genl2),1)
12
13    for i in range(m-1):
14        for j in range(i+1, m):
15            [r,c] = np.where(M2 == M1[i,j])
16            for k in range(len(r)):
17                VT[(i)*n + r[k]] = 1;
18                VT[(i)*n + c[k]] = 1;
19                VT[(j)*n + r[k]] = 1;
20                VT[(j)*n + c[k]] = 1;
21
22            if M is None:
23                M = np.copy(VT)
24            else:
25                M = np.concatenate((M, VT), 1)
26
27            VT = np.zeros((n*m,1), int)
28
29    return M
```

Código 2 apresenta um código em C++.

Código 2: Exemplo de código em C++.

```
1 #include <stdio.h>
2 #include <iostream>
3 // A comment
4 int main(void)
5 {
6     printf("Hello World\n");
7     return 0;
8 }
```

3.3 EXPERIMENTOS

CONCLUSÃO

As conclusões vão aqui ...

REFERÊNCIAS

- CERVI, C.; MANICA, E.; DORNELES, C.; GALANTE, R. 2009. **BDTC - Uma Biblioteca Digital de Trabalhos Científicos com Serviços Integrados**. Revista Brasileira de Computação Aplicada [S.l.] v.1 n.1 p 65–76.
- GARCIA, R. M. 2019. **Desenvolvimento da nova Biblioteca Digital da Biblioteca Brasileira USP: relato de experiência**. PragMATIZES [S.l.] n.16 p 111.
- GRIEBLER, D. 2016. **Domain-Specific Language & Support Tool for High-Level Stream Parallelism**. 2016. Tese (Doutorado em Ciência da Computação) — Faculdade de Informática - PPGCC - PUCRS Porto Alegre, Brazil.
- GRIEBLER, D.; HOFFMANN, R. B.; DANELUTTO, M.; FERNANDES, L. G. 2018. **High-Level and Productive Stream Parallelism for Dedup, Ferret, and Bzip2**. International Journal of Parallel Programming Switzerland v.47 n.2 p 1–19.
- GRIEBLER, D.; VOGEL, A.; MARON, C. A. F.; MALISZEWSKI, A. M.; SCHEPKE, C.; FERNANDES, L. G. 2018. **Performance of Data Mining, Media, and Financial Applications under Private Cloud Conditions**. In: IEEE SYMPOSIUM ON COMPUTERS AND COMMUNICATIONS (ISCC), 23. Natal, Brazil. **Anais**. IEEE.
- HENTGES, E.; THOMÉ, B. 2013. **Análise e Comparação de Ferramentas Open Source de Computação em Nuvem para o Modelo de Serviço IaaS**. [Online]. Três de Maio, RS, Brazil.
- LARCC. 2019. **Laboratory of Advanced Researches for Cloud Computing**

(Home Page). [Online]. Available on: <<http://larcc.setrem.com.br>>. Access date: 15 February.

MALISZEWSKI, A. M.; BAUM, W. 2017. **Performance Characterizations of IaaS Private Clouds for Scientific and Enterprise Workloads.** [Online]. Três de Maio, RS, Brazil.

MCCOOL, M.; ROBISON, A. D.; REINDERS, J. 2012. **Structured Parallel Programming: patterns for efficient computation.** MA, USA: Morgan Kaufmann. ISBN 978-0124159938.

PEREIRA, C. M.; MOREIRA, W.; SEGUNDO, J. E. S. 2021. **Classificação face-tada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital.** Encontros Bibli [S.l.] v.26 p 1–21.

PORTO, C. M. 2009. **Difusão e cultura científica: alguns recortes.** Salvador, BA: EDUFBA. ISBN 9788523206192.

ROCKENBACH, D. A.; ANDERLE, N. 2017. **Análise e Avaliação Comparativa do Desempenho de Bancos de Dados NoSQL.** [Online]. Três de Maio, RS, Brazil.

STEIN, C. 2018. **Programação Paralela para GPU em Aplicações de Processamento Stream.** [Online]. Três de Maio, RS, Brazil.

VOGEL, A.; GRIEBLER, D.; SCHEPKE, C.; FERNANDES, L. G. 2017. **An Intra-Cloud Networking Performance Evaluation on CloudStack Environment.** In: EUROMICRO INTERNATIONAL CONFERENCE ON PARALLEL, DISTRIBUTED AND NETWORK-BASED PROCESSING (PDP), 25. St. Petersburg, Russia. **Anais.** IEEE. p 5.