



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

ENCONTRE SEU ORIENTADOR - DESENVOLVIMENTO DE UMA PLATAFORMA
DE DADOS ABERTOS PARA CATEGORIZAÇÃO AUTOMATIZADA DE
TRABALHOS DE CONCLUSÃO DE CURSO NA ÁREA DE TECNOLOGIA

BRENO RIBEIRO CANO E JONATAN DE SOUZA FREIDMON

Orientador
Reinaldo Viana Alvares

RIO DE JANEIRO, RJ - Brasil
DEZEMBRO/2023

Catálogo informatizada pelos autores:

Ribeiro, Breno
RF78R52 ENCONTRE SEU ORIENTADOR - DESENVOLVIMENTO DE UMA
PLATAFORMA DE DADOS ABERTOS PARA CATEGORIZAÇÃO
AUTOMATIZADA DE TRABALHOS DE CONCLUSÃO DE CURSO NA ÁREA DE
TECNOLOGIA / Breno Ribeiro, Jonatan FREIDMON. -- Rio de
Janeiro, 2023.
41

Orientador: Reinaldo Viana.
Trabalho de Conclusão de Curso (Graduação) -
Universidade Federal do Estado do Rio de Janeiro, Graduação
em Sistemas de Informação, 2023.

1. Dados Abertos. 2. Análise de Dados. 3. Categorização
Automática/Web Scraping. I. FREIDMON, Jonatan II. Viana,
Reinaldo, orient. III. Título.

ENCONTRE SEU ORIENTADOR - DESENVOLVIMENTO DE UMA PLATAFORMA
DE DADOS ABERTOS PARA CATEGORIZAÇÃO AUTOMATIZADA DE
TRABALHOS DE CONCLUSÃO DE CURSO NA ÁREA DE TECNOLOGIA

BRENO RIBEIRO CANO E JONATAN DE SOUZA FREIDMON

Trabalho de Conclusão de Curso de
graduação apresentado à Escola de
Informática Aplicada (EIA) da Universidade
Federal do Estado do Rio de Janeiro
(UNIRIO) para obtenção do título de
Bacharel em Sistemas de Informação.

Aprovado por:

Reinaldo Viana Alvares (UNIRIO)

Marcelo Soares Loutfi (UNISUAM)

Paulo Sergio Medeiros (UNIRIO)

RIO DE JANEIRO, RJ - Brasil

DEZEMBRO/2023

Agradecimentos

Expressamos nossa profunda gratidão ao nosso professor e orientador Reinaldo Viana, por toda sua contribuição, paciência e disponibilidade ao longo do desenvolvimento deste projeto. Sua orientação foi inenarrável para harmonizar e sintonizar ideias dispersas, conduzindo-nos com clareza e dedicação para alinhá-las de maneira a tornar possível a concretização deste trabalho.

Também estendemos nossa gratidão à instituição como um todo de uma forma geral, ao BSI, a UNIRIO por proporcionar um ambiente de ensino e aprendizado de qualidade para discentes. Nesse cenário colaborativo que ideias, como as desenvolvidas neste trabalho, encontrando espaço para florescer e se tornar realidade.

Neste momento, ao separar as partes, eu, Breno, desejo inicialmente expressar minha gratidão a Deus, que me proporciona paz e serenidade, acima de todas as coisas. Em segundo lugar, agradeço também a minha família, principalmente ao meu irmão Rafael que me incentivou e foi uma grande fonte de inspiração para seguir na área de tecnologia e a cursar este curso de BSI na UNIRIO. Agradeço a todos os meus amigos - não seria capaz de citar todos - que fiz durante essa trajetória acadêmica. Entretanto, um agradecimento especial aos que estiveram comigo nesse último semestre tanto em trabalhos em grupo e/ou conversas.

Também ressalto e agradeço a todos os professores da UNIRIO. Ao professor Reinaldo Viana, um agradecimento gigantesco por toda sua atenção, preocupação e direcionamento ao longo do nosso TCC que foram fundamentais para a conclusão deste trabalho com qualidade, além dos agradecimentos feitos acima, é claro.

Por fim, mas não menos importante, um agradecimento notável ao Jonatan, um amigo que esteve comigo desde o início quando entrei na faculdade em 2018.1 e com quem compartilho este trabalho. Por toda sua dedicação, esforço e apoio durante este projeto, transformando momentos difíceis e complicados em momentos divertidos e de aprendizado.

Eu, Jonatan, desejo expressar meu profundo agradecimento pela atenção e dedicação incansáveis que nosso orientador dispensou ao longo do

desenvolvimento deste trabalho. Sua orientação foi de extrema importância, marcando a diferença crucial na trajetória deste projeto.

Além disso, é imprescindível destacar o papel fundamental desempenhado por minha namorada ao longo dessa jornada. Seu apoio constante representou um pilar nos momentos mais desafiadores, e suas celebrações a cada etapa concluída foram um estímulo inestimável. Expresso minha sincera gratidão pela presença essencial dela e pelo incentivo constante que me proporcionou.

Não posso deixar de expressar minha profunda gratidão à minha família, principalmente para minha mãe e minha avó, cujo apoio, estímulo e acompanhamento próximo foram fundamentais para minha trajetória acadêmica e pessoal. Sem a presença e o suporte de vocês, nada disso seria possível. Agradeço do fundo do coração por serem uma fonte constante de inspiração e suporte em minha jornada.

Quero também estender meus agradecimentos a todos os amigos que tornaram essa trajetória acadêmica mais tranquila, proporcionando troca de ideias e conversas valiosas. Muitos momentos compartilhados contribuíram para superar todos os desafios. Dedico este agradecimento especial aos amigos de fora da faculdade e aos meus primos Pedro, Eric e Lorrán, cujo apoio constante fez toda a diferença nos momentos mais complicados, sendo essenciais para chegar até aqui.

Um agradecimento especial é reservado ao Breno, meu companheiro desde o início da faculdade. Não poderia ser diferente no trabalho de conclusão. Foram muitos trabalhos realizados ao longo desse percurso, e as trocas positivas foram inúmeras. A contribuição dele neste trabalho foi determinante para tornar o processo mais fluido e leve, transformando um momento que poderia ser difícil em algo mais acessível. Agradeço sinceramente por sua valiosa ajuda.



Resumo: A pesquisa e o conhecimento acadêmico desempenham um papel fundamental na evolução da sociedade, e os Trabalhos de Conclusão de Curso (TCCs) representam uma fonte valiosa de informações e insights. No entanto, o acesso a esses trabalhos muitas vezes é limitado, dificultando a disseminação e a utilização eficiente desses recursos. Nesse contexto, o TCC se concentra no desenvolvimento de uma plataforma web de dados abertos que visa tornar os resumos de TCCs do curso de Bacharelado de Sistemas de Informação (BSI) da UNIRIO acessíveis, organizados e categorizados automaticamente.

Palavras-chave: Dados Abertos; Categorização Automática; Análise de Dados; Web Scraping; ChatGPT.

Abstract: Research and academic knowledge play a fundamental role in the evolution of society, and Undergraduate Thesis Papers (TCCs) represent a valuable source of information and insights. However, access to these papers is often limited, hindering the dissemination and efficient utilization of these resources. In this context, the proposed Undergraduate Thesis Paper focuses on the development of an open data web platform aimed at making abstracts of BSI theses from UNIRIO accessible, organized, and automatically categorized.

Keywords: Open Data; Automatic Categorization; Data Analysis; Web Scraping; ChatGPT.

Índice

1.Introdução	7
1.1 Motivação	7
1.2 Problema	7
1.3 Justificativa	8
1.4 Objetivo	8
1.5 Organização do Texto	9
2. Fundamentação Teórica	10
2.1 Desafios na Utilização de Dados Abertos Não Normalizados	10
2.2 Web Scraping: Uma Solução para a Normalização de Dados Abertos	10
2.3 Categorização de Dados Abertos: Aprendizado de Máquina e PLN como Ferramentas de Transformação	11
2.4 Maximizando a Categorização e Classificação Textual com ChatGPT	12
3. Metodologia	13
4. Site do curso de BSI da UNIRIO	15
5. Desenvolvimento da Plataforma de Dados Abertos para Categorização Automatizada	17
5.1 Descrição do Sistema	17
5.2 Requisitos	17
5.2.1 Requisitos Funcionais	17
5.2.1 Requisitos Não Funcionais	18
5.3 Casos de Uso	18
Caso de Uso 1: Listar TCC's	19
Caso de Uso 2: Ver Detalhe TCC	20
Caso de Uso 3: Listar Professores	20
Caso de Uso 4: Ver Detalhe Professor	21
Caso de Uso 5 : Download de TCCs	21
Caso de Uso 6 : Download Dados Abertos	22
Caso de Uso 7 : Encontre seu orientador	22
Caso de Uso 8 : Ver Estatísticas	23
5.3 Tecnologias Utilizadas	23
5.3.1 Front-End	24
5.3.2 Back-End	25
5.3.3 Arquitetura da solução	25
5.4 Telas da aplicação	26
5.4.1 Home	27
5.4.2 Lista de TCCs	27
5.4.3 Detalhe do TCC	28
5.4.4 Lista de Professores	29
5.4.5 Detalhe do professor	30
5.4.6 Estatísticas	31
5.4.7 Dados abertos	33
5.4.8 Encontre seu orientador	34

6. Discussão	36
6.1 Comparação entre Sistemas de Disponibilização de Trabalhos de Conclusão de Curso (TCC) no Curso de Bacharelado em Sistemas de Informação (BSI)	36
7. Conclusão	37
7.1 Limitações	37
7.2 Dificuldades Encontradas	38
7.3 Trabalhos Futuros	39
Referências Bibliográficas	41

1.Introdução

1.1 Motivação

Na era da informação digital, a propagação do conhecimento acadêmico desempenha um papel fundamental no avanço da sociedade. No cenário acadêmico, os Trabalhos de Conclusão de Curso (TCCs) representam uma fonte rica de pesquisa e inovação, abrangendo uma variedade de temas. No entanto, o acesso a esses trabalhos muitas vezes é limitado, desorganizado ou pouco divulgado.

Este trabalho tem como objetivo propor o desenvolvimento de uma plataforma web de dados abertos. Este sistema irá compilar resumos de TCCs, identificar os nomes dos alunos e orientadores, além do ano de conclusão do TCC. O elemento inovador deste projeto reside na utilização do ChatGPT, uma avançada ferramenta de processamento de linguagem natural, para categorizar cada TCC com base no tema tecnológico abordado.

A plataforma proporcionará uma experiência interativa aos usuários, permitindo que eles naveguem facilmente pelos TCCs, identifiquem tópicos de interesse e acessem informações relevantes. Esta plataforma web estará acessível e disponível ao público geral, promovendo a transparência e o compartilhamento de conhecimento, essenciais para a progressão da pesquisa e da inovação na área da tecnologia.

1.2 Problema

No contexto do trabalho, o problema identificado consiste na limitada e confusa acessibilidade aos valiosos Trabalhos de Conclusão de Curso (TCCs) no curso de BSI da UNIRIO. Muitas vezes, esses documentos, ricos em pesquisa e inovação, permanecem pouco divulgados, desorganizados e de difícil alcance para aqueles que buscam conhecimento ou inspiração. A falta de uma plataforma eficaz que reúna, organize e categorize automaticamente os resumos de TCCs contribui para essa restrição de acesso, dificultando a propagação e a utilização eficiente desses recursos acadêmicos. Portanto, o cerne do problema reside na necessidade

de superar essas barreiras, tornando os TCCs mais acessíveis, organizados e promovendo o compartilhamento de conhecimento na comunidade acadêmica de tecnologia.

1.3 Justificativa

A justificativa para a realização deste trabalho é fundamentada na necessidade de superar as limitações existentes no acesso e na categorização de Trabalhos de Conclusão de Curso (TCCs) no curso de Sistemas de Informação da UNIRIO. Percebe-se algumas limitações na plataforma atual, como por exemplo: a falta de uma plataforma dedicada que compile e categoriza automaticamente os resumos de TCCs, a disponibilidade dos TCCs somente em ordem cronológica, a existência de mais de 200 trabalhos defendidos sem nenhuma categorização ou classificação por tema ou assunto, são alguns fatores que dificultam a propagação do conhecimento acadêmico e a colaboração entre pesquisadores, estudantes e profissionais da área. Ao desenvolver uma plataforma de dados abertos que utiliza tecnologias avançadas como o ChatGPT, busca-se promover a transparência, o compartilhamento de informações e, conseqüentemente, a inovação no cenário acadêmico tecnológico. Essa iniciativa visa eliminar as barreiras que restringem o acesso a esses valiosos recursos, proporcionando uma ferramenta eficaz para a comunidade acadêmica explorar, aprender e colaborar de maneira mais ampla.

1.4 Objetivo

Este estudo concentra-se em validar a viabilidade técnica e a aplicabilidade prática de uma plataforma inovadora de dados abertos destinada aos Trabalhos de Conclusão de Curso (TCCs), com especial ênfase na automação da categorização de temas. O objetivo primordial é contribuir significativamente para a comunidade acadêmica, simplificando o acesso a um vasto conjunto de conhecimentos anteriormente dispersos e de difícil alcance.

Ao longo desta pesquisa, explora-se minuciosamente as diversas fases de desenvolvimento da plataforma, desde a metódica coleta e organização dos dados dos TCCs até a implementação de tecnologias avançadas de processamento de linguagem natural, como o ChatGPT, para realizar a categorização automática dos

temas. Busca-se evidenciar os potenciais benefícios desta plataforma para a pesquisa acadêmica, aprimorando a experiência educacional e impactando, principalmente, os estudantes do curso de BSI como um todo.

1.5 Organização do Texto

O presente trabalho foi desenvolvido em sete capítulos, incluindo este tópico que faz parte do primeiro, e foi dividido da seguinte forma:

Capítulo 2: apresenta os conceitos fundamentais para o entendimento deste trabalho, além de correlacionar o TCC com trabalhos relacionados.

Capítulo 3: expõe o processo utilizado para coletar, organizar e classificar os resumos de TCCs do curso BSI da UNIRIO. Essa etapa envolve desde a coleta automática dos dados até a validação da categorização e a criação de uma plataforma com foco na interatividade e acessibilidade.

Capítulo 4: descreve sobre o site atual de publicações de TCC de BSI da UNIRIO.

Capítulo 5: aborda com detalhes todo o desenvolvimento da plataforma web com uma rápida descrição do sistema, casos de uso e tecnologias utilizadas tanto de Front-End quanto de Back-End.

Capítulo 6: é realizada uma comparação entre como são disponibilizados atualmente as publicações de TCC e como estão disponíveis no sistema desenvolvido neste trabalho.

Capítulo 7: reúne as conclusões, compilando as limitações, dificuldades encontradas e conjecturando possibilidades futuras de continuidade deste projeto.

2. Fundamentação Teórica

2.1 Desafios na Utilização de Dados Abertos Não Normalizados

Conforme Alcantara et al. (2020), a publicação de dados na internet sem prévia normalização acarreta um cenário de dispersão e desorganização. Os autores destacam que "esses dados estão, predominantemente, em formato não estruturado - o que limita sua descrição e reutilização por outras aplicações e pessoas". Essa predominância do formato não estruturado impõe significativas barreiras à descrição, interpretação e reutilização eficaz dessas informações, dificultando a análise e a aplicação em diferentes contextos ou por diferentes usuários. A ausência de um padrão ou formato consistente emerge, portanto, como um desafio preeminente na gestão e no aproveitamento eficiente de dados disponibilizados online.

2.2 Web Scraping: Uma Solução para a Normalização de Dados Abertos

Rozado et al. (2017) em seu artigo *Estudo de Métodos de Acesso Automatizado para Classificação de Conteúdo em Websites de Acordo com Dados Abertos Governamentais*, enfatizam a ineficiência na disponibilização de registros públicos de forma aberta pelo governo. Os autores ressaltam a importância do acesso automatizado para assegurar a qualidade e a aderência aos princípios de dados abertos. A pesquisa sugere que a adoção de métodos automáticos no cenário governamental pode fomentar práticas mais eficazes na categorização de Trabalhos de Conclusão de Curso (TCCs), contribuindo para uma maior transparência e acesso desimpedido às informações acadêmicas na plataforma desenvolvida.

A relevância do Web Scraping é destacada ao observar a vasta quantidade de dados espalhados e desorganizados na internet, muitas vezes sem qualquer forma de normalização. Utilizando técnicas de web scraping, é viável reunir esses dados caóticos em um conjunto coeso e uniformemente estruturado. Esse processo de extração automatizada e subsequente refinamento dos dados - incluindo etapas de limpeza e padronização - simplifica significativamente a comparação e análise de

informações. Essa metodologia não só facilita o acesso à informação atualizada, mas também assegura sua consistência contínua, abordando efetivamente a problemática da falta de normalização.

Esta técnica tem uma aplicação direta em contextos acadêmicos, como demonstrado pelo interesse em desenvolver sistemas automatizados para avaliar e categorizar conteúdos - uma necessidade claramente ilustrada por Rozado et al. (2017).

2.3 Categorização de Dados Abertos: Aprendizado de Máquina e PLN como Ferramentas de Transformação

Rangel et al. (2020) em *Uso de Aprendizado de Máquina para Categorização Automática de Conjuntos de Dados de Portais de Dados Abertos*, ressaltam a importância e complexidade dos dados abertos. O estudo investiga a disponibilização e os desafios na integração de dados de portais governamentais abertos, focando na falta de padronização das categorias de dados. Paralelamente, um Trabalho de Conclusão de Curso associado visa aumentar a transparência e facilitar o compartilhamento de conhecimento acadêmico, propondo a organização dos resumos de TCCs como dados abertos.

Os autores enfatizam a importância do aprendizado de máquina na categorização automática de conjuntos de dados, envolvendo a extração de informações textuais para determinar categorias apropriadas e melhorar a organização e integração em portais governamentais. Eles apontam desafios similares na organização e categorização de dados abertos tanto em portais governamentais quanto no âmbito acadêmico, destacando o aprendizado de máquina como uma solução promissora para superar essas dificuldades e melhorar a eficiência na integração e acesso às informações.

Adicionalmente, o estudo *Categorização automática de conjuntos de dados de portais de dados abertos utilizando aprendizado de máquina supervisionado* por Rangel (2019), oferece insights sobre a aplicação de técnicas avançadas como Processamento de Linguagem Natural e Aprendizado de Máquina na categorização de dados governamentais. Este trabalho resalta a importância da categorização automática para melhorar a gestão de portais de dados abertos, compartilhando a aplicação dessas técnicas avançadas com o TCC mencionado.

A metodologia de categorização automática, utilizando aprendizado de máquina e processamento de linguagem natural, é um ponto em comum entre o artigo e o TCC. Ambos procuram automatizar a categorização de conjuntos de dados específicos, facilitando a identificação rápida de informações pertinentes pelos usuários. Estas investigações sublinham a necessidade de métodos inovadores para abordar a complexidade dos dados abertos e destacam o potencial de técnicas avançadas para melhorar o acesso e a gestão de informações.

2.4 Maximizando a Categorização e Classificação Textual com ChatGPT

Fraçkiewicz, Marcin (2023), em seu artigo "Como o ChatGPT ajuda a melhorar a categorização e classificação de texto", aborda a contribuição do ChatGPT, desenvolvido pela OpenAI, no campo do Processamento de Linguagem Natural (PLN). Destaca-se que essa tecnologia avançada oferece notável potencial para a categorização e classificação de texto, acelerando e aprimorando esses processos por meio de sua capacidade de análise de padrões textuais e geração de conteúdo baseado em parâmetros específicos.

O artigo ressalta a importância do ChatGPT na interpretação de dados não estruturados, especialmente no contexto de dados abertos. Através de modelos de aprendizado profundo e algoritmos sofisticados, o ChatGPT enriquece a categorização de informações, tornando-se uma ferramenta valiosa para a análise precisa de dados provenientes de diversas fontes.

No contexto específico de um Trabalho de Conclusão de Curso (TCC), o artigo menciona a aplicação do ChatGPT para aprimorar a compreensão e a geração de texto natural, contribuindo para uma categorização automática mais eficaz dos resumos de TCCs, com um enfoque particular em temas tecnológicos. A incorporação do ChatGPT visa melhorar a organização e o acesso aos resumos.

3. Metodologia

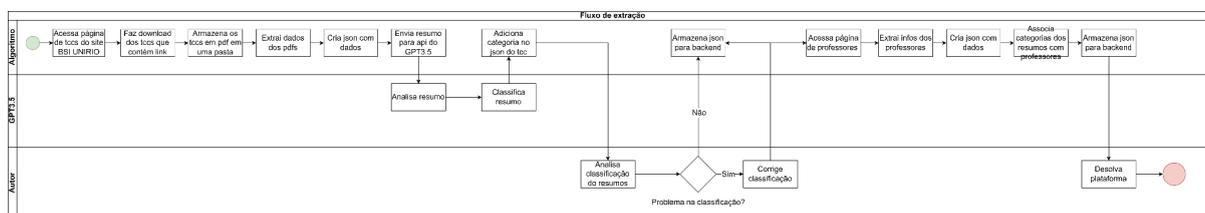
A metodologia empregada neste projeto para a coleta, organização e categorização de resumos de Trabalhos de Conclusão de Curso (TCCs) do curso de Sistemas de Informação da UNIRIO envolveu várias etapas cruciais, descritas a seguir em ordem cronológica:

1. **Coleta de Dados dos TCCs:** Inicialmente, realizou-se a coleta automática de dados dos resumos de TCCs do Bacharelado em Sistemas de Informação da UNIRIO. Esta coleta foi efetuada através de técnicas de web scraping na seção de publicações de TCCs do site oficial do programa (<https://bsi.uniriotec.br/publicacoes-de-tcc/>), durante o mês de outubro de 2023. Esta etapa essencial reuniu uma base de dados sobre os trabalhos acadêmicos realizados.
2. **Extração e Organização de Dados dos Resumos:** Após a obtenção dos TCCs em formato PDF, um algoritmo específico analisou estes arquivos para identificar e extrair os resumos, bem como informações adicionais como autor, orientador, ano e título. As informações extraídas foram organizadas em um arquivo JSON, facilitando o armazenamento e a organização dos dados.
3. **Classificação dos TCCs:** Utilizando a API do GPT-3.5, os TCCs foram classificados em áreas relacionadas à tecnologia. Esse processo permitiu atribuir categorias específicas a cada resumo, que foram adicionadas ao arquivo JSON, enriquecendo o banco de dados.
4. **Validação e Ajuste da Categorização:** Para assegurar a precisão da classificação, foram realizados testes comparativos e ajustes nos prompts enviados à API do GPT-3.5, com intervenção manual quando necessário para garantir a melhor categorização possível.
5. **Scraping de Informações dos Orientadores:** Seguiu-se o acesso ao site dos docentes do BSI da UNIRIO para coletar informações adicionais sobre os orientadores, incluindo fotos, e-mails e nomes completos, organizando-as em um arquivo JSON.
6. **Associação de Categorias aos Orientadores:** Um algoritmo associou cada categoria de TCC aos respectivos orientadores, agregando ao JSON dos

professores informações sobre os temas já orientados por cada um, enriquecendo assim o panorama dos interesses e especializações dos orientadores.

7. **Desenvolvimento de uma Plataforma Web:** Com as etapas de extração e associação concluídas, desenvolveu-se uma plataforma web especializada para apresentar os resumos, incluindo informações pertinentes como detalhes dos estudantes, orientadores e datas de conclusão. A plataforma foi desenhada com ênfase na interatividade, consistência, feedback visual e sonoro, acessibilidade, além de estética e minimalismo.

Figura 1: Página de publicações do site do BSI



Fonte: Elaborado pelo autor

4. Site do curso de BSI da UNIRIO

O portal online do curso de Bacharelado em Sistemas de Informação (BSI) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) foi projetado para centralizar informações para os alunos, oferecendo diversas funcionalidades essenciais. Contudo, destaca-se uma funcionalidade específica para este trabalho: a seção de publicações de Trabalhos de Conclusão de Curso (TCCs). A inspiração para o desenvolvimento deste projeto originou-se desta seção, que se responsabiliza por apresentar uma compilação abrangente de todos os TCCs desenvolvidos desde o início do curso.

Essa seção expõe uma diversidade de soluções já propostas ao longo do tempo, servindo como fonte de inspiração para futuros trabalhos acadêmicos. A seguir na figura 1, é fornecido uma imagem da página, onde é possível observar uma listagem organizada cronologicamente dos artigos. Cada entrada exibe informações cruciais, incluindo título, autor, ano e orientador(a) de cada trabalho.

Figura 2: Página de publicações do site do BSI

BSI – UNIRIO
Bacharelado em Sistemas de Informação

DISCIPLINAS ▾ ALUNOS ▾ PROFESSORES INSTITUCIONAL ▾ SECRETARIA

NOTÍCIAS

Home » Publicações

Publicações

Lista dos trabalhos já defendidos. Os TCCs são publicados em nosso site para servirem de exemplos do que se espera de um TCC e de possíveis temas, além de constituem uma boa fonte de informação sobre tópicos específicos e atualizados na área de Sistemas de Informação. Uma dica: se você gostar de algum TCC, procure o professor-orientador daquele TCC e informe-se da possibilidade de dar continuidade ao trabalho ou elaborar um novo projeto seguindo a mesma linha.

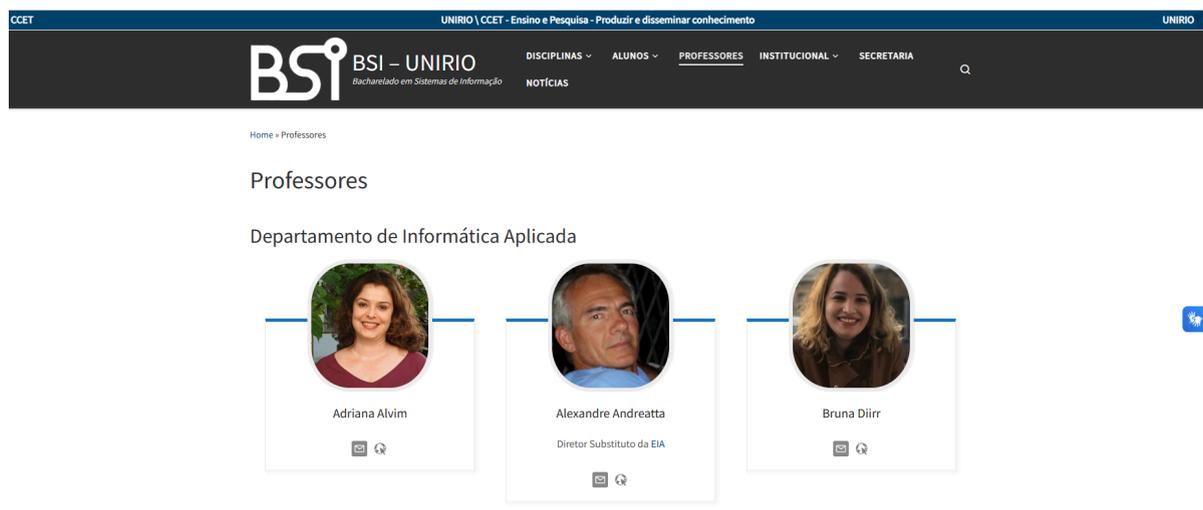
- **Análise da qualidade do design da biblioteca Joda-Time**
João Gabriel Passos Cosendey Brouck e Lucas Carvalho de Vasconcellos. Orientador: Márcio de Oliveira Barros. 08/2023
- **Travel Partner: um Aplicativo para Organização de Viagens**
Luiz Carlos Queiroz Cunha Filho. Orientador: Márcio de Oliveira Barros. 07/2023
- **Acessibilidade e usabilidade da computação em nuvem na terceira idade: um estudo de caso a partir da ferramenta Google Drive**
Pedro Maia. Orientador: Simone Bacellar Leal Ferreira. 07/2023
- **OndaBeat: Um jogo de ritmo para aprendizado rítmico**

Fonte: Site do curso BSI

Uma outra contribuição para o desenvolvimento do sistema foi a integração da seção dedicada aos professores no site do Bacharelado em Sistemas de Informação (BSI). Nesta seção, como é mostrado na figura 2, são apresentados todos os docentes que estão atualmente lecionando na faculdade, exibindo suas informações relevantes, tais como nome, foto, currículo Lattes e endereço de e-mail. Este conjunto de informações está exemplificado na figura a seguir.

Essa área oferece uma oportunidade para os alunos iniciarem conversas com os professores e ampliem seu conhecimento sobre os educadores que os orientam.

Figura 3: Página de professores do Departamento de Informática Aplicada



Fonte: Site do curso BSI

As páginas específicas disponíveis no site do curso de Bacharelado em Sistemas da Informação desempenharam um papel crucial ao fornecer os dados fundamentais que serviram como base para o nosso trabalho. Essa fonte de informações foi essencial para a realização do projeto.

5. Desenvolvimento da Plataforma de Dados Abertos para Categorização Automatizada

5.1 Descrição do Sistema

O site construído no âmbito do TCC, oferece uma interface amigável para a exploração eficiente dos Trabalhos de Conclusão de Curso (TCCs). Destaca-se pela usabilidade, permitindo aos usuários acessar, explorar e navegar intuitivamente pela plataforma. Apresenta funcionalidade de busca avançada para refinar pesquisas por categorias específicas, além da opção de download de TCCs em PDF.

Possui categorização automatizada, empregando tecnologias avançadas como o ChatGPT para analisar os resumos dos TCCs. O sistema monitora continuamente o uso da plataforma, gerando estatísticas para análise, contribuindo para a melhoria constante da experiência do usuário e eficácia do sistema.

5.2 Requisitos

5.2.1 Requisitos Funcionais

Para garantir que o projeto possa abordar efetivamente o problema proposto, foram definidos os seguintes requisitos funcionais:

RF1 - O sistema deve exibir a lista de TCCs disponíveis com informações detalhadas.

RF2 - O sistema deve permitir que o usuário visualize informações detalhadas de um TCC selecionado.

RF3 - O sistema deve permitir que o usuário filtre TCCs por categoria.

RF4 - O sistema deve exibir uma lista de professores.

RF5 - O sistema deve permitir que o usuário visualize informações detalhadas de um professor selecionado.

RF6 - O sistema deve permitir que o usuário faça o download de um TCC em formato PDF.

RF7 - O sistema deve permitir que o usuário faça o download de arquivos de dados abertos.

RF8 - O sistema deve permitir que o usuário insira sua ideia de TCC.

RF9 - O sistema deve exibir uma lista de possíveis orientadores com base na ideia do usuário.

RF10 - O sistema deve exibir gráficos relacionados à quantidade de TCCs concluídos por ano, desenvolvidos por professor e realizados por categoria.

5.2.1 Requisitos Não Funcionais

Para assegurar a excelência do produto final a ser desenvolvido e garantir uma experiência satisfatória para os usuários, é necessário que a aplicação atenda aos seguintes critérios não funcionais:

RNF1 - O sistema deve ser intuitivo e fácil de usar para usuários.

RNF2 - O sistema deve responder rapidamente às solicitações do usuário, especialmente durante a navegação e exibição de detalhes.

RNF3 - O sistema deve garantir a segurança dos dados dos usuários, especialmente durante o download de TCCs e dados abertos.

RNF4 - O sistema deve ser compatível com diferentes navegadores web e dispositivos.

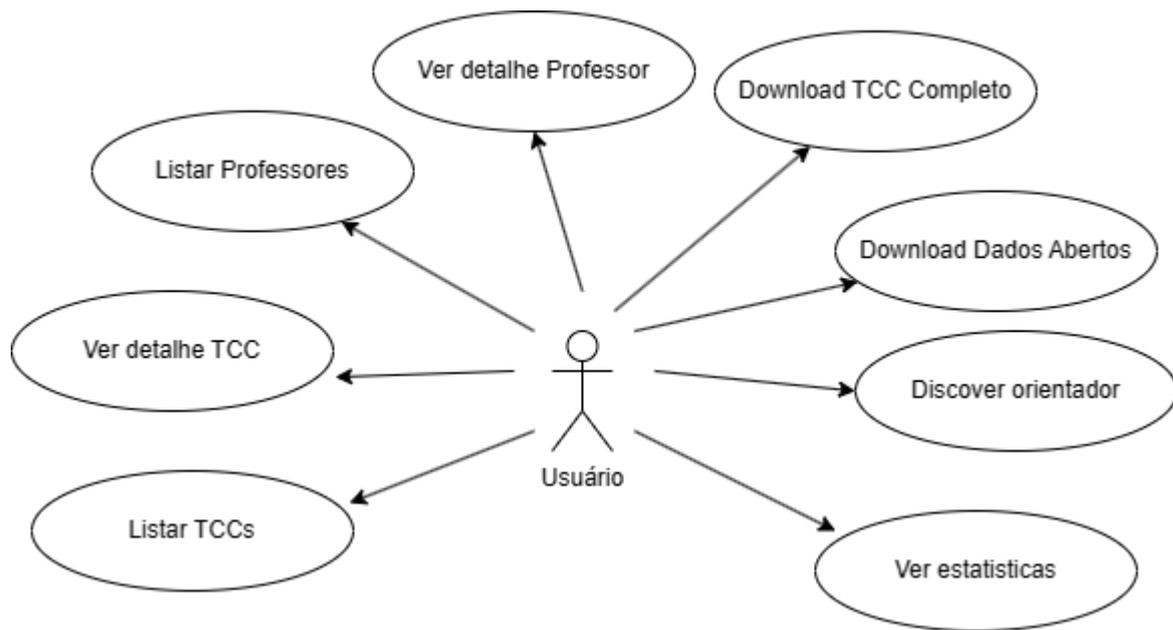
RNF5 - O sistema deve ser capaz de lidar com um grande volume de TCCs e usuários simultaneamente, sem comprometer o desempenho.

RNF6 - O código do sistema deve ser organizado e documentado para facilitar futuras atualizações e manutenções.

5.3 Casos de Uso

O sistema adota uma abordagem unificada, permitindo que tanto alunos quanto professores atuem como usuários, sem distinção entre os papéis. O diagrama foi estruturado de forma abrangente para representar todas as ações realizadas pelo usuário no sistema, pode ser visto na figura 3.

Figura 4: Casos de uso do usuário do sistema



Fonte: Elaborado pelo autor

Caso de Uso 1: Listar TCC's

Ator Principal: Usuário

Pré-condições:

- O usuário acessa a plataforma por meio de um dispositivo compatível.

Pós-condições:

- O usuário está visualizando a página da plataforma com os TCCs.

Fluxo Principal:

- O usuário acessa a plataforma por meio de um navegador web.
- O sistema exibe a página principal com várias opções selecionáveis de botão no menu para o que o usuário desejar, tais como: "TCCs", "Professores", "Estatísticas", "Dados abertos" e "Encontre seu orientador".
- O usuário seleciona a opção "TCCs".
- O sistema redireciona o usuário para a página com a lista de TCCs disponíveis.
- O usuário navega pela página que exibe informações sobre os TCCs, incluindo título, autor, ano de publicação, categoria, orientador e resumo.

Este caso de uso destaca a interação do usuário ao explorar os TCCs no sistema, enfatizando a facilidade de navegação.

Caso de Uso 2: Ver Detalhe TCC

Ator Principal: Usuário

Pré-condições:

- O usuário tem acesso à plataforma.

Fluxo Principal:

- O usuário visualiza a lista de TCCS.
- O usuário seleciona um TCC específico para obter mais detalhes.
- O sistema exibe o TCC escolhido, apresentando informações detalhadas correspondentes.

Fluxo Alternativo:

- O usuário seleciona as opções de categoria no filtro de busca avançado.
- O sistema realiza a pesquisa com base nos filtros selecionados.
- O sistema exibe uma lista de TCCs correspondentes aos critérios de busca.
- O usuário seleciona um TCC específico para ver mais detalhes.

Este caso de uso destaca a funcionalidade de visualizar detalhes de um TCC em específico no sistema.

Caso de Uso 3: Listar Professores

Ator Principal: Usuário

Pré-condições:

- O usuário acessa a plataforma por meio de um dispositivo compatível.

Pós-condições:

- O usuário está visualizando a página da plataforma com os TCCs.

Fluxo Principal:

- O usuário visualiza a página home.
- O usuário seleciona a opção "Professores".
- O sistema redireciona o usuário para a página com a lista de professores.

- O usuário navega pelos professores exibidos na página.

Este caso de uso destaca a interação do usuário ao explorar os professores na plataforma, enfatizando a facilidade de navegação.

Caso de Uso 4: Ver Detalhe Professor

Ator Principal: Usuário

Pré-condições:

- O usuário tem acesso à plataforma.

Fluxo Principal:

- O usuário visualiza a lista de Professores.
- O usuário seleciona um professor específico para obter mais detalhes.
- O sistema exibe o professor escolhido, apresentando informações detalhadas correspondentes.

Este caso de uso destaca a funcionalidade de visualizar detalhes de um professor em específico no sistema.

Caso de Uso 5 : Download de TCCs

Ator Principal: Usuário

Pré-condições:

- O usuário acessou a plataforma e realizou uma busca por TCCs.

Fluxo Principal:

- O usuário visualiza a lista de TCCs exibidos.
- O usuário seleciona um TCC específico para obter mais detalhes.
- O sistema exibe o TCC escolhido, apresentando informações detalhadas correspondentes.
- O usuário encontra a opção de "Ler TCC Completo".
- O sistema redireciona o usuário para a página do TCC no formato PDF.
- O usuário tem a opção de fazer o download do TCC em PDF.

Este caso de uso destaca a funcionalidade crucial de permitir que os usuários baixem os TCCs em formato PDF.

Caso de Uso 6 : Download Dados Abertos

Ator Principal: Usuário

Pré-condições:

- O usuário acessou a plataforma e realizou uma busca por TCCs.

Fluxo Principal:

- O usuário visualiza a página home.
- O usuário seleciona a opção "Dados abertos".
- O sistema redireciona o usuário para a página com os dados abertos do sistema.
- O usuário tem a opção de realizar o download do(s) arquivo(s) desejados.

Este caso de uso destaca a funcionalidade crucial de permitir que os usuários baixem os dados abertos em formato JSON.

Caso de Uso 7 : Encontre seu orientador

Ator Principal: Usuário

Pré-condições:

- O usuário acessou a plataforma e realizou uma busca por TCCs.

Fluxo Principal:

- O usuário visualiza a página home.
- O usuário seleciona a opção "Encontre seu orientador".
- O sistema redireciona o usuário para a página selecionada.
- O sistema exibe uma caixa de texto, a qual o usuário deve colocar sua ideia de TCC.
- O usuário após colocar sua ideia, deve selecionar a opção "Enviar".
- O sistema exibe uma lista de possíveis professores que tenham relação com a ideia descrita pelo usuário.

Este caso de uso destaca a funcionalidade principal do sistema, a qual permite que o usuário descreva sua ideia e encontre o orientador que tenha mais correlação com o tema proposto.

Caso de Uso 8 : Ver Estatísticas

Ator Principal: Usuário

Pré-condições:

- O usuário acessou a plataforma e realizou uma busca por TCCs.

Fluxo Principal:

- O usuário visualiza a página home.
- O usuário seleciona a opção "Estatísticas".
- O sistema redireciona o usuário para a página selecionada.
- O sistema exibe três gráficos:
 - O primeiro gráfico exibe a correlação da quantidade de TCC concluídos por ano.
 - O segundo apresenta a correlação entre a quantidade de TCC desenvolvidos por professor(orientador).
 - O terceiro expõe a relação entre quantidade de TCC realizados por categoria.

Neste caso de uso é ressaltada a funcionalidade do sistema para visualizar gráficos a respeito do TCC.

5.3 Tecnologias Utilizadas

Este segmento do texto tem como objetivo detalhar o processo de construção da solução proposta. Inicialmente, será apresentada uma descrição dos recursos e tecnologias utilizados no desenvolvimento do frontend, incluindo linguagens, bibliotecas e frameworks adotados, bem como sua justificativa de uso no contexto do projeto. Similarmente, para o backend, será discutida a seleção de tecnologias, fornecendo uma visão das escolhas técnicas realizadas.

Adicionalmente, será exposta a arquitetura da solução, detalhando os componentes arquitetônicos, a organização e a interação entre eles, assim como o fluxo de dados e a lógica de processamento. Isso inclui uma descrição das camadas de aplicação, serviços, dados e como estas se interconectam para formar um sistema.

5.3.1 Front-End

A implementação de uma interface gráfica eficiente desempenha um papel crucial na apresentação formalizada dos dados abertos, buscando proporcionar uma experiência de usuário intuitiva e agradável. Nesse contexto, optou-se por integrar tecnologias específicas que contribuem para o uso e a estética da interface.

Para o desenvolvimento da interface, foi escolhido o framework Angular 12, uma plataforma de desenvolvimento Front-End amplamente reconhecida por sua estrutura modular e orientada a componentes. O Angular facilita a criação de SPAs (Single Page Applications), onde a navegação entre diferentes seções ocorre de forma dinâmica, sem a necessidade de recarregar a página inteira. Isso não apenas otimiza o desempenho, mas também proporciona uma experiência de usuário mais fluida.

No aspecto visual, foi incorporado o Bootstrap, um framework de design responsivo que oferece um conjunto abrangente de estilos e componentes pré-construídos. Com o Bootstrap, garantimos uma apresentação consistente e atrativa em diferentes dispositivos, proporcionando uma interface adaptável e de fácil leitura.

Quanto à representação gráfica das estatísticas, foi adotada a biblioteca Chart.js. Essa biblioteca JavaScript oferece uma variedade de opções para a criação de gráficos dinâmicos e interativos, permitindo a visualização clara e intuitiva de dados estatísticos. A escolha do Chart.js adiciona um componente visual dinâmico à interface, contribuindo para uma compreensão mais eficaz dos padrões e tendências presentes nos dados.

Assim, a combinação dessas tecnologias - Angular 12 para estruturação, Bootstrap para design responsivo e Chart.js para representação gráfica - resulta em uma interface que não apenas atende aos requisitos funcionais, mas também proporciona uma experiência visualmente agradável e eficiente para os usuários, enfatizando a importância da apresentação adequada de dados abertos.

5.3.2 Back-End

O desenvolvimento do Back-End foi cuidadosamente planejado e executado em duas fases distintas, cada uma tirando proveito de tecnologias específicas para otimizar a eficiência e a funcionalidade do sistema. Na primeira fase, o foco foi na coleta de informações do site de publicações do BSI. Para esta tarefa, optou-se pelo uso do Python 3.8 como linguagem principal, devido à sua vasta gama de bibliotecas de apoio e ao conhecimento prévio dos desenvolvedores envolvidos na solução. Por meio da biblioteca BeautifulSoup, foi possível realizar um web scraping preciso nos TCCs, extraindo dados de maneira estruturada e organizada. Adicionalmente, para conferir uma categorização mais aprofundada e contextualizada aos TCCs após a coleta, implementou-se a API do GPT-3.5. Essa escolha estratégica possibilitou uma categorização inteligente, aproveitando as avançadas capacidades de processamento de linguagem natural oferecidas pelo GPT-3.5.

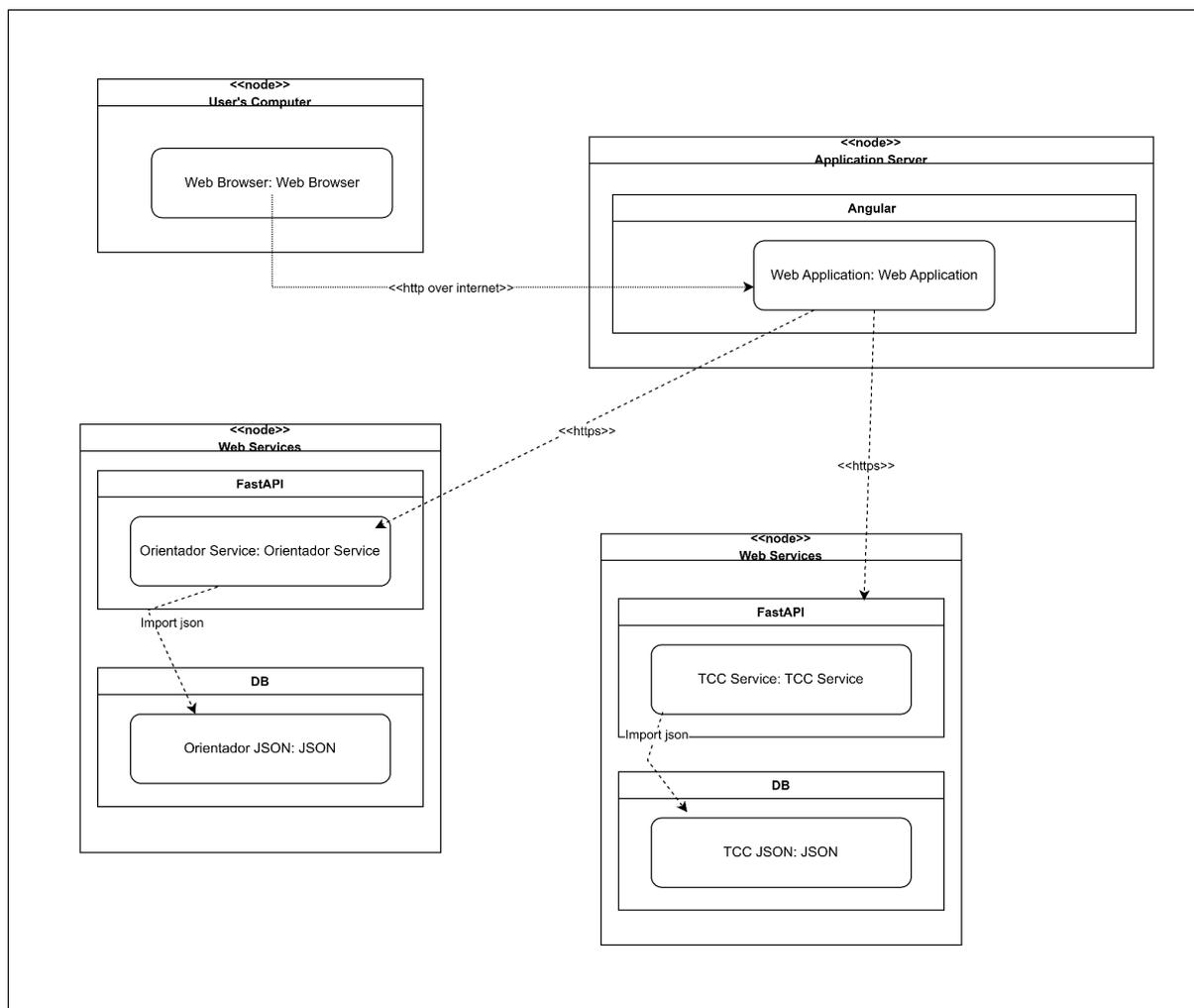
Na segunda fase do desenvolvimento, o foco concentrou-se na disponibilização eficiente dessas informações para o Front-End. Neste contexto, adotou-se o framework FastAPI para a construção de endpoints em Python. O FastAPI não apenas simplifica a criação de uma API, mas também promove a adoção de padrões modernos, como tipagem de dados e documentação automática.

5.3.3 Arquitetura da solução

Ao observar a Figura 7, a solução adota uma arquitetura de microsserviços, na qual os componentes são modularizados em serviços independentes que se comunicam entre si por meio de interfaces bem definidas. Essa abordagem favorece

a coesão e o baixo acoplamento, pois cada microsserviço pode ser desenvolvido, implantado e dimensionado de forma independente, facilitando a manutenção e a evolução do sistema. Dessa forma, a escolha da arquitetura de microsserviços faz sentido em termos de coesão e acoplamento, pois contribui para um sistema mais modular, escalável e fácil de manter.

Figura 5: Arquitetura da aplicação e fluxo de dados



Fonte: Elaborado pelo autor

5.4 Telas da aplicação

Esta seção tem o propósito de ilustrar o funcionamento do aplicativo, utilizando uma série de imagens capturadas e narrativas descritivas das funcionalidades, oferecendo assim uma perspectiva integral sobre a experiência do usuário e a mecânica geral do aplicativo.

5.4.1 Home

A página inicial, onde é apresentada na figura 5, tem como propósito apresentar o conteúdo e a finalidade do site, destacando seu principal objetivo: auxiliar os alunos do curso de Bacharelado em Sistemas de Informação (BSI) a localizar seus orientadores por meio dos dados disponibilizados de forma acessível e intuitiva.

Figura 6: Página inicial

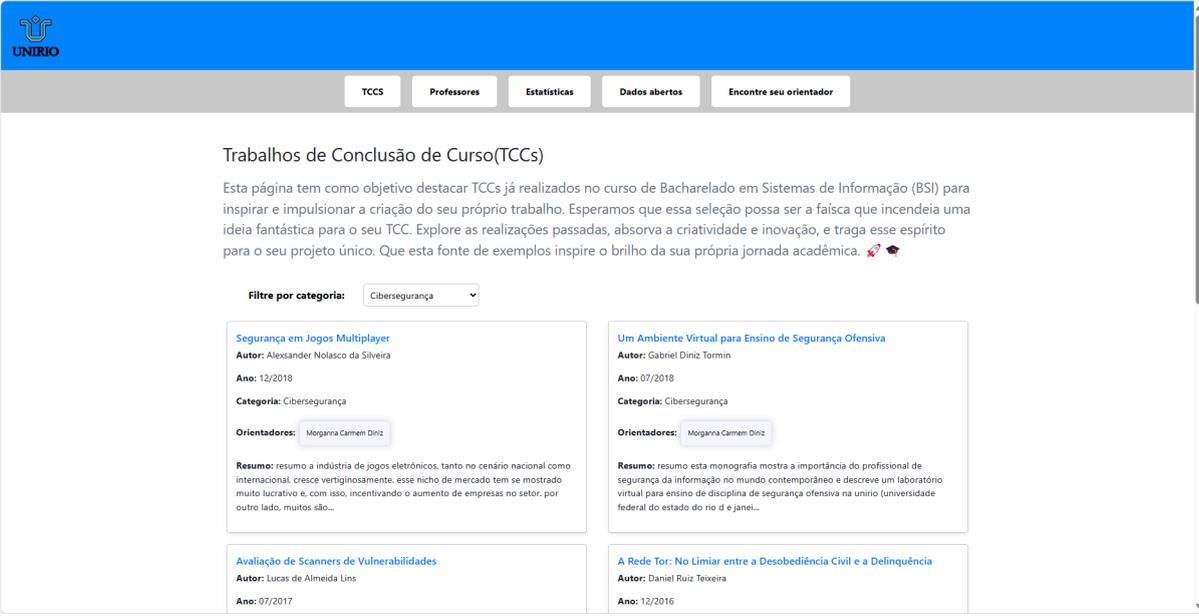


Fonte: Aplicação “Encontre seu orientador”

5.4.2 Lista de TCCs

A página de listagem de Trabalhos de Conclusão de Curso (TCCs) tem como intuito apresentar uma visão abrangente de todos os trabalhos realizados, oferecendo a capacidade de filtrar por categoria na área de tecnologia. A seguir na figura 6, pode ser vista a funcionalidade projetada para aprimorar a experiência do usuário, facilitando a navegação e a busca específica pelos trabalhos desejados.

Figura 7: Página de TCCs



Trabalhos de Conclusão de Curso (TCCs)

Esta página tem como objetivo destacar TCCs já realizados no curso de Bacharelado em Sistemas de Informação (BSI) para inspirar e impulsionar a criação do seu próprio trabalho. Esperamos que essa seleção possa ser a faísca que incendeia uma ideia fantástica para o seu TCC. Explore as realizações passadas, absorva a criatividade e inovação, e traga esse espírito para o seu projeto único. Que esta fonte de exemplos inspire o brilho da sua própria jornada acadêmica. 🚀 🎓

Filtre por categoria: Cibersegurança

<p>Segurança em Jogos Multiplayer Autor: Alexander Nolasco da Silveira Ano: 12/2018 Categoria: Cibersegurança Orientadores: Morgana Carmen Diniz</p> <p>Resumo: resumo a indústria de jogos eletrônicos, tanto no cenário nacional como internacional, cresce vertiginosamente, esse nicho de mercado tem se mostrado muito lucrativo e, com isso, incentivando o aumento de empresas no setor, por outro lado, muitos são...</p>	<p>Um Ambiente Virtual para Ensino de Segurança Ofensiva Autor: Gabriel Diniz Tormin Ano: 07/2018 Categoria: Cibersegurança Orientadores: Morgana Carmen Diniz</p> <p>Resumo: resumo esta monografia mostra a importância do profissional de segurança da informação no mundo contemporâneo e descreve um laboratório virtual para ensino de disciplina de segurança ofensiva na unirio (universidade federal do estado do rio de janeiro...</p>
<p>Avaliação de Scanners de Vulnerabilidades Autor: Lucas de Almeida Lins Ano: 07/2017</p>	<p>A Rede Tor: No Limiar entre a Desobediência Civil e a Delinquência Autor: Daniel Ruiz Teixeira Ano: 12/2016</p>

Fonte: Aplicação “Encontre seu orientador”

5.4.3 Detalhe do TCC

Na página de detalhes do TCC, o usuário tem a oportunidade de visualizar um resumo do trabalho, proporcionando uma prévia do tema antes de decidir realizar o download completo. Essa abordagem visa agilizar a navegação do usuário, permitindo uma análise inicial antes de tomar uma decisão mais aprofundada. Se o tema despertar interesse, o usuário pode então optar por acessar o TCC completo.

Além dessas funcionalidades, o usuário tem a facilidade de clicar no nome do professor associado ao TCC, sendo direcionado para detalhes adicionais sobre o docente. Adicionalmente, a página oferece a capacidade de visualizar TCCs semelhantes ao selecionado, permitindo ao usuário explorar trabalhos relacionados. Essa funcionalidade, apresentada na figura 7, é projetada para proporcionar uma experiência de navegação contínua e enriquecedora, onde o usuário pode explorar temas similares com apenas um clique e atualizar a página para ler trabalhos correlacionados.

Figura 8: Página sobre um TCC

Segurança em Jogos Multiplayer
Autor: Alexander Nolasco da Silveira
Ano: 12/2018
Categoria: Cibersegurança
Orientadores: Morgana Carmen Diniz

Resumo:

resumo a indústria de jogos eletrônicos, tanto no cenário nacional como internacional, cresce vertiginosamente, esse nicho de mercado tem se mostrado muito lucrativo e, com isso, incentivando o aumento de empresas no setor. por outro lado, muitos são os desafios de se manter servidores de jogos como: problemas jurídicos, limitações tecnológicas, manutenção da infraestrutura, recursos humanos, ataques ao software entre outros. nesse contexto, este projeto visa explorar as falhas de segurança em jogos online (multi player) e abordar algumas regras que possam ser aplicadas a esses softwares para aumentar o nível de segurança e confiabilidade do jogo, este trabalho propõe um pequeno algoritmo (trainer) para simular um ataque direto a um jogo multiplayer de computador e, inclusive, ao próprio utilizador deste, para tanto é desenvolvido um código malicioso dentro do — trainer || transformando -o em um cavalo de troia a fim de obter informações sensíveis do usuário, palavras -chave : trainer , hacker , aplicação, segurança, ma licioso , jogo 5

Ler TCC completo

Título do TCC: Um Ambiente Virtual para Ensino de Segurança Ofensiva
Professor Orientador: Morgana Carmen Diniz
Ano: 07/2018

Título do TCC: Avaliação de Scanners de Vulnerabilidades
Professor Orientador: Letia Cristina Vasconcelos de Andrade
Ano: 07/2017

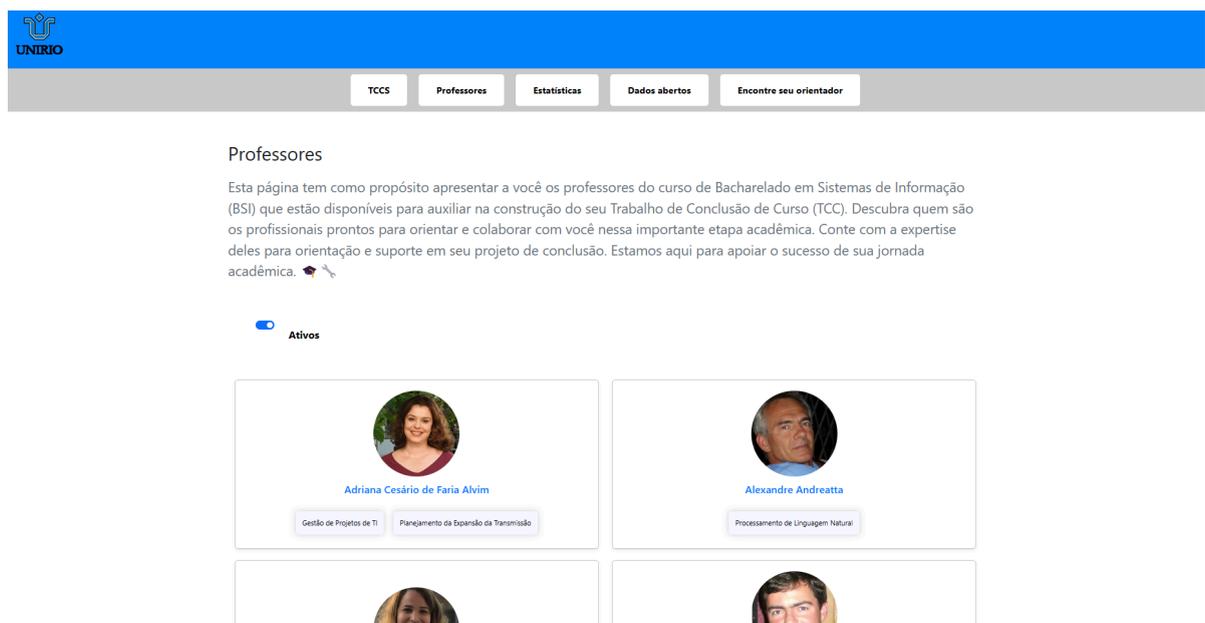
Fonte: Aplicação “Encontre seu orientador”

5.4.4 Lista de Professores

A finalidade da página de listagem de professores é apresentar os docentes que já desempenharam a orientação de trabalhos na faculdade. Ao acessar a página, são exibidos inicialmente apenas os professores ativos na instituição. No entanto, há um filtro disponível para verificar aqueles que já orientaram, mas que não estão mais presentes na grade de professores do site do BSI.

Um aspecto notável desta página, vista na figura 8, é a exibição das áreas da tecnologia nas quais os professores já atuaram em trabalhos de conclusão de curso (TCC). Isso oferece aos usuários uma perspectiva sobre a expertise e o histórico de orientação dos docentes.

Figura 9: Página de professores



Fonte: Aplicação “Encontre seu orientador”

5.4.5 Detalhe do professor

Nesta página, são apresentados detalhes sobre o professor, incluindo seu currículo Lattes e o endereço de e-mail para facilitar o contato direto. Além disso, ao lado dessas informações, encontramos uma lista abrangente de áreas da tecnologia em que o professor tem experiência. A figura 9 mostra que a página também oferece uma relação completa de todos os Trabalhos de Conclusão de Curso (TCCs) nos quais o professor colaborou, proporcionando um acesso fluido e organizado ao conteúdo acadêmico que ele já contribuiu. Isso permite que os interessados explorem e compreendam melhor o impacto e a expertise do professor no âmbito acadêmico.

Figura 10: Página sobre um professor

UNIRIO

TCCS Professores Estatísticas Dados abertos Encontre seu orientador

Colaborou nos tccs:

Título do TCC: Sistema de Cadastro de Informações Museológicas
Autor: Bernardo Cirne Severo
Categoria: Gestão de Projetos de TI
Ano: 07/2018

Título do TCC: Planejamento da Expansão de Sistemas de Transmissão: Um Estudo de Métodos
Autor: Paulo Fernando de Matos Araujo
Categoria: Planejamento da Expansão da Transmissão
Ano: 01/2014

Adriana Cesário de Faria Alvim

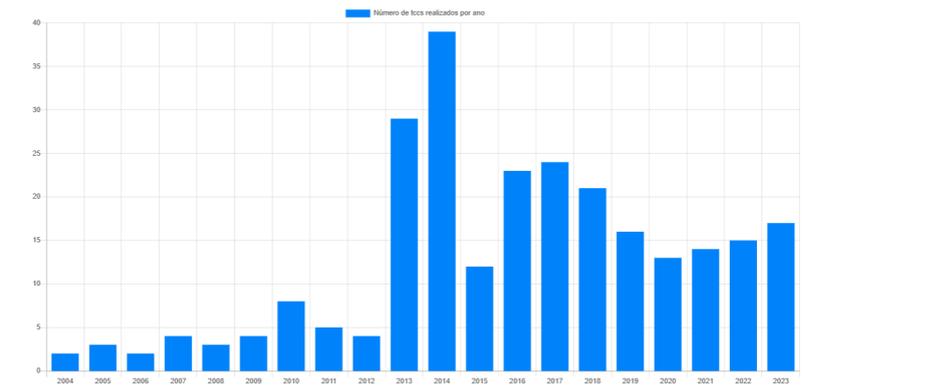
Gestão de Projetos de TI Planejamento da Expansão da Transmissão

Fonte: Aplicação “Encontre seu orientador”

5.4.6 Estatísticas

O objetivo principal desta página é fornecer estatísticas relevantes derivadas de dados extraídos, entregando insights valiosos aos usuários. A seção inclui três gráficos distintos: a Figura 10 ilustra a distribuição temporal dos Trabalhos de Conclusão de Curso (TCCs), a Figura 11 ressalta a distribuição de TCCs por orientador, e a Figura 12 examina a incidência desses trabalhos nas dez principais áreas de tecnologia, com a categoria "Outros" agregando as demais áreas não listadas no top 10. Esta configuração visual e analítica visa oferecer uma interpretação clara e direta das tendências e padrões nos dados, enriquecendo a experiência do usuário com informações precisas e envolventes.

Figura 11: Quantidade de TCC por ano.

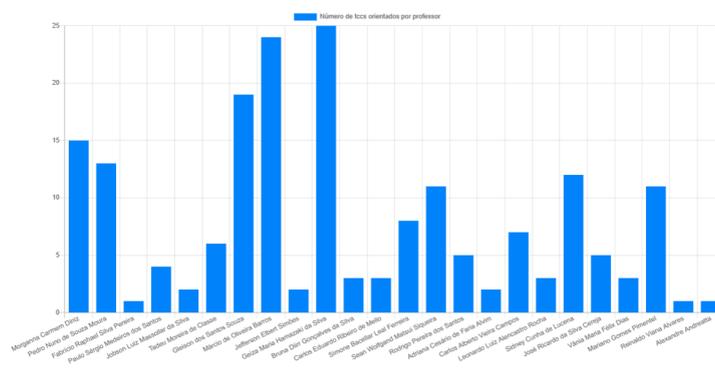


Fonte: Aplicação “Encontre seu orientador”

Figura 12: Gráfico sobre a relação entre TCC e orientador

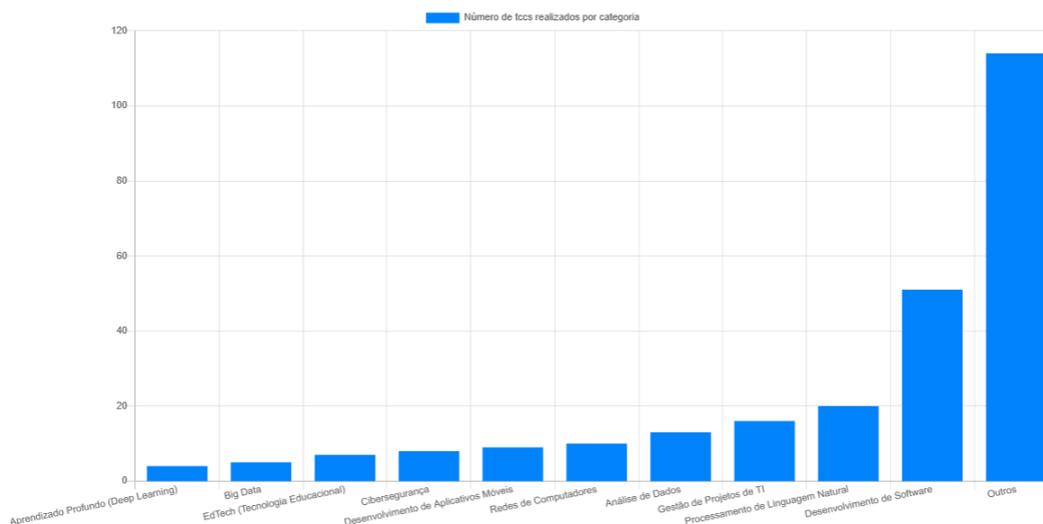
Estatísticas

Esta página tem como objetivo fornecer gráficos com o intuito de incentivar reflexões a partir dos dados disponíveis na plataforma. Pretendemos enriquecer sua compreensão por meio de representações visuais, estimulando análises mais aprofundadas. Explore os gráficos disponíveis e deixe que eles guiem seus insights e interpretações sobre os dados apresentados. Apreciamos sua dedicação à exploração e interpretação dos dados disponíveis. 🇧🇷 🚀



Fonte: Aplicação “Encontre seu orientador”

Figura 13: Gráfico sobre a relação entre TCC e área da tecnologia



Fonte: Aplicação “Encontre seu orientador”

5.4.7 Dados abertos

O intuito desta página, que é mostrada na figura 13, é fornecer aos usuários acesso aos dados normalizados utilizados pela aplicação. Dessa maneira, os usuários têm a possibilidade de realizar o download de uma das seis bases disponíveis: TCCs, professores, categorias, gráfico de TCCs por ano, gráfico de TCCs por professor e gráfico de TCCs por área da tecnologia. Essa funcionalidade visa proporcionar uma experiência flexível e personalizada, permitindo que os usuários explorem e utilizem os dados conforme suas necessidades específicas.

Figura 14: Página dos dados abertos utilizados na plataforma

Dados abertos

Esta página tem como propósito disponibilizar as bases de dados utilizadas no trabalho de conclusão de curso, visando fomentar o uso de dados abertos e potencializar pesquisas e trabalhos futuros. Sintam-se à vontade para utilizar os recursos disponíveis com parcimônia e dedicação. O conhecimento está à disposição de todos. Baixem com responsabilidade e que esses dados inspirem novas descobertas e contribuições acadêmicas. Agradecemos o interesse e o comprometimento com a pesquisa e análise de dados abertos. [📊](#) [#DadosAbertos](#) [#PesquisaAcadêmica](#) [#ComprometimentoCientífico](#)

- [Base de dados de TCCS](#)
- [Base de dados de Orientadores](#)
- [Base de dados de Categorias](#)
- [Base de dados usado para gráfico ano/tccs](#)
- [Base de dados usado para gráfico categoria/tccs](#)
- [Base de dados usado para gráfico orientadores/tccs](#)

Fonte: Aplicação “Encontre seu orientador”

5.4.8 Encontre seu orientador

O propósito desta página é facilitar a busca de um orientador que se alinhe à ideia do usuário. Ao inserir a descrição da sua proposta, o site utiliza a API do GPT para categorizar a ideia, identificando os três professores que mais orientaram temas semelhantes. A seguir na figura 14, pode ser visto esse processo que visa auxiliar o usuário no momento de escolher um orientador, proporcionando uma abordagem eficiente e personalizada na seleção do docente mais adequado para a sua proposta.

Figura 15: Página “Encontre seu orientador” mostrando a funcionalidade em ação.

Discover

Deixe suas ideias fluírem e encontre quem possa ajudá-lo a transformá-las em realidade!

gostaria de trabalhar com dados abertos para construir uma plataforma inovadora

Enviar

Orientadores recomendados para iniciar sua jornada



Morganna Carmem Diniz

Desenvolvimento de Software Cibersegurança Análise de Dados

Processamento de Linguagem Natural Redes de Computadores Tecnologia Educacional

Internet das Coisas Computação em nuvem, segurança Cloud Computing



Reinaldo Viana Alvares

Análise de Dados

Fonte: Aplicação “Encontre seu orientador”

6. Discussão

6.1 Comparação entre Sistemas de Disponibilização de Trabalhos de Conclusão de Curso (TCC) no Curso de Bacharelado em Sistemas de Informação (BSI)

A distinção entre a plataforma do site do BSI e a inovação "Encontre o seu orientador" é notável, evidenciando avanços significativos na experiência do usuário e na acessibilidade às informações acadêmicas.

No site do curso, a organização dos TCCs por ordem cronológica demanda que os usuários naveguem pela interface gráfica através de rolagens para encontrar o que procuram. Em contraste, o novo website foi estruturado para simplificar a busca, permitindo que os usuários encontrem TCCs com base em suas áreas de interesse, como a área da tecnologia, otimizando assim a navegação. Além disso, a funcionalidade que permite a visualização de TCCs semelhantes ao selecionado proporciona uma experiência contínua e enriquecedora, algo que não estava disponível na plataforma anterior.

Outro ponto que destaca a diferença é a maneira como os TCCs orientados por professores são apresentados. No site BSI, essa informação demanda uma busca manual, percorrendo individualmente os professores. Por outro lado, o sistema "Encontre o seu orientador" apresenta uma interface que lista todos os professores ativos do curso, permitindo aos usuários visualizar o histórico acadêmico de cada professor, incluindo os temas em que estão mais envolvidos. Essa abordagem simplificada facilita a avaliação do potencial de cada professor como orientador, fornecendo um panorama mais abrangente e eficiente.

Além disso, a introdução da disponibilização de dados abertos, estatísticas e a funcionalidade "Discover" representam avanços substanciais que preenchem lacunas existentes na plataforma do BSI. Essas adições não apenas aprimoram a experiência do usuário, mas também fornecem ferramentas valiosas para a comunidade acadêmica, tornando a nova interface uma plataforma mais completa e eficaz.

7. Conclusão

Este TCC busca contribuir significativamente para a disseminação do conhecimento acadêmico na área de tecnologia como um todo, porém mais especificamente para todo o corpo docente de BSI da UNIRIO, promovendo a transparência e o compartilhamento de informações. A plataforma de dados abertos proposta tem o potencial de se tornar uma valiosa ferramenta para estudantes, pesquisadores e profissionais interessados em explorar TCCs relacionados a área acadêmica no que tange a assuntos tecnológicos.

Espera-se que este projeto possa facilitar o acesso a informações acadêmicas relevantes, inspirar novas pesquisas e promover a colaboração na comunidade acadêmica e tecnológica. De forma a impulsionar a inovação e o avanço da tecnologia por meio do compartilhamento de conhecimento.

7.1 Limitações

Este projeto, embora almeje uma melhoria substancial na acessibilidade e organização dos resumos de TCCs, enfrenta, por natureza, algumas limitações que são fundamentais para delinear as fronteiras e considerações no desenvolvimento da plataforma. Foi abordado a seguir algumas dessas limitações, reconhecendo sua importância no contexto do projeto.

A primeira delas diz respeito à dependência de dados disponíveis, uma vez que a qualidade da categorização automática e dos resultados da plataforma estar intrinsecamente ligada à disponibilidade e qualidade dos resumos de TCCs obtidos do site de BSI da UNIRIO. Limitações na variedade ou quantidade desses dados podem impactar diretamente a precisão da categorização, destacando a necessidade de uma fonte robusta e diversificada de informações.

Outra limitação relevante é a diversidade de temas abordados nos TCCs. Se a quantidade de temas for limitada, a plataforma pode não cobrir totalmente a diversidade de áreas na tecnologia, o que por sua vez pode restringir a utilidade da plataforma para usuários com interesses específicos. É crucial reconhecer a

importância da abrangência e representatividade dos temas tratados para garantir a relevância da plataforma em contextos diversos.

Além disso, a eficácia da categorização automática está fortemente ligada à tecnologia utilizada, como é o caso do ChatGPT. Limitações nesta tecnologia podem impactar a precisão da categorização, ressaltando a necessidade contínua de atualizações e melhorias tecnológicas para garantir resultados precisos e confiáveis.

Finalmente, a exclusão de outras áreas acadêmicas é uma consideração importante. Como o escopo da plataforma está restrito à área de tecnologia, há uma limitação na colaboração potencial ou no interesse de estudantes e pesquisadores de outras disciplinas. A possibilidade de ampliar o alcance da plataforma para abranger uma variedade mais ampla de áreas acadêmicas pode enriquecer a experiência e atrair uma gama mais diversificada de usuários.

Dessa forma, ao reconhecer e abordar essas limitações de maneira proativa, o projeto busca aprimorar continuamente sua eficácia e utilidade, mantendo-se consciente das fronteiras inerentes ao seu escopo e das oportunidades de desenvolvimento futuro.

7.2 Dificuldades Encontradas

A primeira dificuldade enfrentada revelou-se no fato de que o portal de publicações do BSI não se encontrava devidamente configurado para a partilha eficiente de dados referentes aos Trabalhos de Conclusão de Curso (TCCs). Observa-se que a referida plataforma não foi projetada com a intenção específica de disponibilizar informações de maneira estruturada, com vistas a facilitar sua utilização posterior na concepção de novos sistemas. Em virtude dessa lacuna de padronização, fez-se necessário estabelecer diretrizes específicas para tornar possível a extração dos TCCs, impondo, assim, a realização de uma análise minuciosa e individualizada de cada cenário. Este procedimento, por sua vez, adiciona uma camada de complexidade ao processo, culminando em um incremento no tempo de processamento. Em resumo, a existência de uma estruturação

adequada dos dados abertos proporciona uma simplificação notável a essa operação.

Outro obstáculo enfrentado refere-se à extração de resumos de alguns TCCs, os quais não estavam em conformidade com o padrão estabelecido, consistindo em não se encontrarem entre as seções de resumo e abstract. Em decorrência disso, alguns resumos apresentavam-se vazios ou englobavam mais texto do que o estritamente designado como resumo. Nessas circunstâncias, o modelo ChatGPT encontrava dificuldades em categorizar, resultando em saídas vazias. Consequentemente, foi necessário empreender ações manuais para extrair corretamente os resumos desses TCCs e submetê-los ao ChatGPT. No entanto, é relevante destacar que a grande maioria dos trabalhos seguiu o padrão proposto, facilitando assim o processo de extração automática de resumos.

7.3 Trabalhos Futuros

À medida que o propósito é aprimorar continuamente a plataforma, foi identificado diversas oportunidades para desenvolvimentos futuros que podem ampliar significativamente seu impacto e utilidade. Estas diretrizes direcionam áreas promissoras para trabalhos futuros, visando enriquecer a experiência dos usuários e expandir ainda mais o alcance da plataforma.

Uma consideração crucial para o futuro envolve a expansão da plataforma para outras áreas acadêmicas. Essa abordagem não apenas ampliaria a base de usuários, mas também fomentaria colaborações interdisciplinares, enriquecendo a diversidade de tópicos disponíveis na plataforma.

Explorar a incorporação de algoritmos mais avançados de processamento de linguagem natural para aprimorar ainda mais a precisão da categorização automática. Isso pode envolver o uso de modelos de linguagem mais avançados ou abordagens específicas para lidar com termos técnicos.

Buscar enriquecer a interação entre os usuários, promovendo a colaboração e a troca de conhecimentos. A integração de recursos de colaboração online, como fóruns de discussão, oferecerá um ambiente dinâmico para que estudantes, pesquisadores e profissionais compartilhem ideias e experiências, fortalecendo a comunidade em torno da plataforma.

Explorar a possibilidade de integrar a plataforma com outras plataformas acadêmicas existentes para criar um ecossistema mais amplo de recursos, informações e fornecer aos usuários acesso a uma gama mais diversificada de informações e oportunidades acadêmicas. Essa colaboração fortalecida beneficia a comunidade acadêmica como um todo.

Referências Bibliográficas

J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniec, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieleszczenko-Kowszewicz, P. Miłkowski, M. Oleksy, M. Piasecki, Ł. Radliński, K. Wojtasik, S. Woźniak, P. Kazienko. ChatGPT: Jack of all trades, master of none. **Elsevier B.V.**, 2023. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S156625352300177X>. Acesso em: 20/11/2023.

P. P. Ray. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. **Elsevier B.V.**, 2023. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S266734522300024X>. Acesso em: 20/11/2023.

M. Fraçkiewicz. Explorando os benefícios do ChatGPT para categorização e classificação de texto. **TS2 SPACE**, 2023. Disponível em:

<https://ts2.space/pt/como-o-chatgpt-ajuda-a-melhorar-a-categorizacao-e-classificacao-de-texto/#gsc.tab=0>. Acesso em: 10/11/2023.

M. M. Rangel. Categorização automática de conjuntos de dados de portais de dados abertos utilizando aprendizado de máquina supervisionado. **Portal IDUFF**, 2019. Disponível em:

<https://app.uff.br/riuff/bitstream/handle/1/13072/Mateus%20Rangel.pdf?sequence=1&isAllowed=y>.

Acesso em: 01/12/2023.

A. P. ROZADO, R. M. D. SOUZA, A. S. CORRÊA. ESTUDO DE MÉTODOS DE ACESSO AUTOMATIZADO PARA CLASSIFICAÇÃO DE CONTEÚDO EM WEBSITES DE ACORDO COM DADOS ABERTOS GOVERNAMENTAL. **Andreivid**, 2023. Disponível em:

https://andreivid.com/support/2017_8conict/8conict_raul.pdf. Acesso em: 10/11/2023.

W. Alcantara, J. Bandeira, A. Barbosa, A. Lima, T. Ávila, I. Bittencourt, S. Isotani. Vista do Desafios no uso de Dados Abertos Conectados na Educação Brasileira. **ResearchGate**, 2020. Disponível em:

[\(PDF\) Desafios no uso de Dados Abertos Conectados na Educação Brasileira \(researchgate.net\)](#).

Acesso em: 25/11/2023.

M. Rangel, F. Bernardini, J. Viterbo, R. Monteiro, E. Seixas, Higor dos Santos Pinto. Uso de Aprendizado de Máquina para Categorização Automática de Conjuntos de Dados de Portais de Dados Abertos. **Semantic Scholar**, 2020. Disponível em:

[Uso de Aprendizado de Máquina para Categorização Automática de Conjuntos de Dados de Portais de Dados Abertos | Semantic Scholar](#). Acesso em: 25/11/2023.

M. Rangel, F. Bernardini, J. Viterbo, R. Monteiro, E. Seixas, Higor dos Santos Pinto. Uso de Aprendizado de Máquina para Categorização Automática de Conjuntos de Dados de Portais de Dados Abertos. **Semantic Scholar**, 2020. Disponível em:

[Uso de Aprendizado de Máquina para Categorização Automática de Conjuntos de Dados de Portais de Dados Abertos | Semantic Scholar](#). Acesso em: 25/11/2023.

CCET - Centro de Ciências Exatas e Tecnologia. **BSI-UNIRIO**, 2023. Disponível em: <https://bsi.uniriotec.br/>. Acesso em 23/10/2023.