

Data Insights: uma aplicação web de análise de dados acadêmicos para avaliação e acompanhamento de desempenho docente e de programas de pós-graduação

Renick Muller Teixeira Costa, Luciano Arruda Teran, Marcelle Pereira Mota

¹Instituto de Ciências Exatas e Naturais (ICEN) – Universidade Federal do Pará (UFPA)
R. Augusto Corrêa, 01 - Guamá, Belém - PA, 66075-110

renickmillerteixeiracosta@gmail.com,
luciano.teran@icen.ufpa.br, mpmota@ufpa.br

Resumo. *A crescente necessidade de gestores acadêmicos obterem dados precisos e atualizados sobre a produção acadêmica dos professores motivou a criação desta aplicação. Esses gestores enfrentam desafios significativos para coletar, organizar e analisar informações dispersas em diferentes plataformas, tornando o processo manual e demorado. Nesse sentido, esse trabalho visa desenvolver uma aplicação web automatizada que facilite a coleta, armazenamento e visualização de dados acadêmicos, proporcionando uma visão integrada e prática das informações. Para isso, o Data Insights utiliza técnicas de web scraping implementadas em Node.js para obter dados dos perfis do Google Acadêmico fornecidos por meio de um arquivo CSV. As informações coletadas, incluindo contagem de citações, índice h e índice i10, são armazenadas em um banco de dados MySQL. O backend expõe uma API para o frontend, permitindo que os usuários acessem e visualizem os dados por meio de tabelas e gráficos. Além disso, o Data Insights permite o upload de arquivos ZIP contendo currículos Lattes de professores, extraindo a contagem de publicações e associando os professores aos respectivos programas de pós-graduação dentro do instituto. Como contribuições, o Data Insights oferece uma solução integrada e eficiente para gerenciar dados acadêmicos, incluindo funcionalidades administrativas para excluir registros, importar arquivos CSV e ZIP, e inserir informações específicas de desenvolvimento de programas, como o número de alunos matriculados, formados, em andamento e desistentes em um determinado programa de pós-graduação e ano específico. Logo, essa aplicação automatiza processos anteriormente manuais, economizando tempo e melhorando a precisão na gestão acadêmica.*

1. Introdução

A gestão acadêmica eficiente e a promoção da pesquisa científica são pilares fundamentais para o desenvolvimento de instituições de ensino superior. No entanto, a obtenção de informações detalhadas e atualizadas sobre a produção científica de docentes frequentemente se torna um desafio, devido à dispersão e à diversidade das fontes de dados.

[Costa 2011] conduziu uma análise dos cursos de pós-graduação lato sensu oferecidos pelo Centro de Formação, Treinamento e Aperfeiçoamento da Câmara dos Deputados. O estudo revelou fatores importantes para a relevância acadêmica, incluindo a organização de grupos de pesquisa integrados aos cursos, forte suporte à gestão e um

corpo docente bem preparado. Um dos aspectos positivos destacados foi o processo de acompanhamento dos egressos, que forneceu informações valiosas para o contínuo aprimoramento dos cursos.

As experiências descritas mostram que a qualidade dos cursos é significativamente influenciada pela maneira como a gestão institucional organiza tanto os aspectos pedagógicos quanto administrativos.

Atualmente, gestores acadêmicos, pesquisadores e administradores precisam acessar manualmente múltiplas plataformas, como o *Google Scholar* e o currículo *Lattes*, para compilar e analisar esses dados. Esse processo, além de demorado, é suscetível a erros e inconsistências.

Neste cenário, esse trabalho tem como objetivo desenvolver uma aplicação web para automatizar a coleta, armazenamento e exibição de dados acadêmicos. A motivação para desenvolver o *Data Insights* reside na necessidade de proporcionar praticidade e eficiência na gestão dos dados acadêmicos. Ao sintetizar as informações dispersas em um sistema interativo, a equipe de gestão acadêmica tem facilidade na visualização de indicadores importantes, como o número de publicações, citações e tendências ao longo do tempo.

A aplicação web *Data Insights* foi produzida para atender os objetivos deste trabalho, sendo responsável por centralizar as informações relevantes sobre a produção acadêmica dos docentes do Instituto de Ciências Exatas e Naturais (ICEN) da Universidade Federal do Pará (UFPA). Para desenvolvê-la, foi necessário a elicitação de requisitos para compreender as necessidades e as funcionalidades desejadas. Em seguida, os requisitos foram modelados para estruturar a arquitetura do sistema. O banco de dados foi então projetado utilizando o MySQL Workbench, garantindo a organização e o armazenamento eficiente dos dados coletados. A codificação do back-end foi realizada com Node.js, implementando a lógica de negócio e a API para interação com o banco de dados. O front-end foi desenvolvido com React.js, proporcionando uma interface amigável e interativa para os usuários. Além disso, foram seguidas boas práticas de desenvolvimento, como controle de versão, testes automatizados e padrões de codificação, para garantir a qualidade e a manutenção do sistema.

Como resultados, a *Data Insights* dispõe de um conjunto de recursos que facilita a visualização gráfica dos dados e a capacidade de filtrá-los por diferentes critérios, como ano de publicação e número de citações. Esses recursos favorecem na concepção de uma visão clara e abrangente do desempenho educacional e científico de diversos docentes, beneficiando tanto gestores quanto a comunidade acadêmica em geral.

Esse trabalho contribuiu para a inovações no gerenciamento acadêmico, promovendo uma cultura de uso de Tecnologias Digitais de Informação e Comunicação (TDICs) para suportar o planejamento educacional, científico e administrativo. O *Data Insights* não apenas economiza tempo, mas também aprimora a tomada de decisões estratégicas para a gestão acadêmica de uma determinada instituição. Portanto, o trabalho é necessário para abordar as limitações dos métodos tradicionais de coleta e análise de dados acadêmicos e para demonstrar como a automação e a centralização de informações podem contribuir significativamente para a eficiência administrativa e o avanço da pesquisa científica.

Este artigo está estruturado da seguinte forma: a Seção ??, apresenta a motivação por trás do desenvolvimento do *Data Insights*; a Seção ?? explora a justificativa do estudo; a Seção 3 destaca os conceitos relacionados ao desenvolvimento do *Data Insights*; a Seção 4 apresenta os objetivos deste trabalho, enquanto a Seção 5 destaca a metodologia utilizada para atendê-los; a Seção 6 aborda a construção do *Data Insights* para índices acadêmicos, explicando o ciclo de desenvolvimento e tecnologias utilizadas; a Seção 7 apresenta os recursos interativos da *Data Insights*; por fim, a Seção 8 apresenta as considerações finais deste trabalho de pesquisa.

2. Motivação e Justificativa

A criação do *Data Insights* foi impulsionada pela necessidade de simplificar e aprimorar a obtenção de informações acadêmicas relevantes sobre os docentes do ICEN. Anteriormente, a coleta de dados como citações, publicações e índices de impacto exigia o acesso a diversas plataformas, como o *Google Acadêmico*, e a realização de buscas individuais para cada professor. Este processo não apenas consumia tempo significativo, mas também estava sujeito a erros e omissões, devido à fragmentação das fontes de informação e à ineficiência na coleta manual de dados.

Gestores acadêmicos, pesquisadores e administradores enfrentavam desafios consideráveis para compilar dados atualizados e precisos, essenciais para análises, avaliações de desempenho e tomadas de decisões estratégicas. Conforme mencionado por [Delgado et al. 2017], optar por uma abordagem automatizada de coleta de dados, como o *Web Scraping*, exige um alto grau de conhecimentos técnicos. Esse alto nível de competência técnica envolve a habilidade de programar scripts complexos, compreender a estrutura HTML de páginas web, manipular e processar dados de forma eficaz, e solucionar problemas que possam surgir durante a extração de informações. Teoricamente, *web scraping* é a prática de coletar dados por qualquer meio que não seja um programa interagindo com uma API (ou, obviamente, por um ser humano usando um navegador web). Isso é comumente feito escrevendo um programa automatizado que consulta um servidor web, requisita dados (em geral, na forma de HTML e de outros arquivos que compõem as páginas web) e então faz parse desses dados, que é o processo de analisar e transformar dados brutos, geralmente em formatos como HTML ou JSON, em uma estrutura mais organizada e utilizável, para extrair as informações necessárias [Mitchell 2015]. Embora a implementação dessa técnica apresente desafios técnicos, a capacidade de personalizar e ajustar o processo de extração de dados de acordo com as necessidades específicas do projeto é uma vantagem significativa.

A construção do *Data Insights* não apenas aborda essas dificuldades, mas também traz inúmeros benefícios que justificam sua implementação. Destaca-se a fomentação à ciência e tecnologia, ao promover o desenvolvimento de soluções tecnológicas avançadas. A aplicação web oferece aos gestores acadêmicos das faculdades do ICEN da UFPA uma ferramenta para acessar informações relevantes sobre os docentes, facilitando a gestão acadêmica.

Além disso, o *Data Insights* estimula a comunidade acadêmica a inovar no gerenciamento educacional e científico, propondo novas maneiras de sintetizar dados dispersos na internet. A aplicação permite elencar professores com base no número de trabalhos acadêmicos publicados, visualizar graficamente a quantidade de citações em

artigos científicos e categorizar esses dados por ano. Essas funcionalidades proporcionam uma visão clara e detalhada do desempenho acadêmico dos docentes, auxiliando a administração do instituto em suas pesquisas e tomadas de decisão.

Os principais benefícios gerados a partir do trabalho desenvolvido incluem:

- **Fomentação à Ciência e Tecnologia:** promover o desenvolvimento de soluções tecnológicas avançadas.
- **Facilidade de Acesso a Informações:** proporcionar uma aplicação web prática para os gestores acadêmicos do ICEN obter dados relevantes sobre os docentes.
- **Estímulo à Inovação Acadêmica:** incentivar a comunidade acadêmica a propor novas tecnologias para o gerenciamento escolar.
- **Síntese de Dados Dispersos:** facilitar a compilação e análise de dados que antes estavam dispersos na internet.
- **Melhoria na Gestão Acadêmica:** oferecer recursos para a administração do instituto categorizar e visualizar dados acadêmicos de forma eficiente, auxiliando na tomada de decisões estratégicas.

3. Referencial Teórico

Essa seção apresenta o referencial teórico que engloba os conceitos para o desenvolvimento deste trabalho de pesquisa, sendo eles a Engenharia de Software (Subseção 3.2) e ferramentas de gestão acadêmica (Subseção 3.3).

3.1. Engenharia de Requisitos

Segundo [Sommerville 2011] os requisitos de software são frequentemente divididos em duas categorias:

- **Requisitos funcionais:** São declarações de serviços que o sistema deve fornecer, de como o sistema deve reagir a entradas específicas e de como o sistema deve se comportar em determinadas situações. Em alguns casos, os requisitos funcionais também podem explicitar o que o sistema não deve fazer.
- **Requisitos não funcionais:** São restrições aos serviços ou funções oferecidos pelo sistema. Incluem restrições de timing, restrições no processo de desenvolvimento e restrições impostas pelas normas. Ao contrário das características individuais ou serviços do sistema, os requisitos não funcionais, muitas vezes, aplicam-se ao sistema como um todo.

Segundo [Figueiredo 2011], os requisitos funcionais descrevem de forma clara e concisa das funcionalidades e serviços de uma aplicação, além de documentar como a aplicação web deve reagir a entradas específicas, se comportar em determinadas situações e o que não se deve fazer.

O propósito dos requisitos funcionais é estabelecer quais funcionalidades um sistema deve possuir para ser utilizado pelo usuário. Eles definem de forma clara e específica o que o sistema deve fazer, cobrindo aspectos como operações, cálculos, manipulação de dados e outras atividades essenciais. Além disso, cada requisito funcional possui uma prioridade que determina a importância da funcionalidade para o sistema, o que ajuda a organizar o desenvolvimento em fases e a focar primeiro nas funcionalidades críticas. Esses critérios são essenciais para garantir a eficiência e a qualidade do sistema, ao mesmo

tempo em que permitem uma melhor gestão do processo de desenvolvimento, minimizando o risco de falhas e garantindo que as necessidades dos usuários sejam atendidas de maneira eficaz.

De acordo com [Cysneiros and do Prado Leite 1997], os requisitos não funcionais são definidos como comportamentos e restrições que determinado software deve satisfazer.

Requisitos não funcionais são essenciais para garantir a qualidade e o desempenho da aplicação, abrangendo aspectos como usabilidade, desempenho, disponibilidade e outros. Embora não estejam relacionados diretamente às funcionalidades, esses requisitos são de suma importância para direcionar o desenvolvimento da aplicação e proporcionar uma experiência satisfatória ao usuário. Eles definem critérios que devem ser atendidos, como tempos de resposta rápidos, interface intuitiva e responsiva, segurança dos dados, escalabilidade e disponibilidade do sistema. O cumprimento desses requisitos não funcionais é fundamental para assegurar que a aplicação atenda aos padrões de qualidade e às expectativas dos usuários.

3.2. Web Scraping

Segundo o autor [Mitchell 2019], o conceito de *Web Scraping* abrange a técnica de coletar dados de forma automatizada, por meio de métodos distintos da interação direta com uma *API (Interface de Programação de Aplicação)* ou da intervenção manual de um usuário em um navegador web. A utilização do *Web Scraping* possibilita a extração eficiente e sistemática de informações valiosas de várias fontes online, tornando-se uma abordagem essencial para a obtenção de dados de maneira ágil e precisa. Essa prática tecnológica tem se mostrado cada vez mais relevante em diferentes contextos, desde análises acadêmicas até aplicações empresariais, impulsionando a disponibilização de informações diversificadas e oportunidades inovadoras para a tomada de decisões embasadas em dados confiáveis.

[Zhao 2017] destaca que esta técnica é eficiente para a raspagem de dados distintos na internet, uma vez que uma enorme quantidade de dados heterogêneos são constantemente gerados diariamente e há necessidade de constante atualização.

De acordo com [Zanon 2017], a coleta de informações por meio de *Web Scraping* ocorre da seguinte maneira: um programa, denominado *web robot*, emula a interação comum de um indivíduo ao navegar pelos servidores web durante um processo de extração de informações convencional. Essa abordagem tecnológica, baseada na simulação de interações humanas, possibilita a obtenção automatizada e eficiente de dados de múltiplas fontes online. Ao empregar o *web robot*, torna-se possível agilizar e otimizar a coleta de informações, viabilizando análises detalhadas e a disponibilização de dados valiosos para diversas aplicações, desde pesquisas acadêmicas até análises de mercado.

[Cocón et al. 2023] destaca a importância da técnica de *Web Scraping* como uma ferramenta profissional para coletar dados valiosos da web de forma eficaz e responsável, proporcionando a automatização da coleta de dados, o acesso a dados não estruturados, a personalização da coleta de dados, a atualização contínua de uma sequência de informações, custo-efetividade e amplo alcance.

3.3. Ferramentas de Gestão Acadêmica

Conforme [Carvalho 2010], os sistemas de gestão acadêmica, além de ajudar nas questões de controle administrativo, facilitam a prática do docente, oferecendo recursos para promover a implementação de um ensino de excelência, comprometido com o processo de desenvolvimento pessoal e intelectual.

[Barroca Filho et al. 2013] destacam um exemplo de ferramenta mobile fornecida pela Universidade Federal do Rio Grande do Norte (UFRN), chamada Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA). O SIGAA mobile é um sistema de informação web corporativo que informatiza os procedimentos da área acadêmica através dos módulos de: graduação, pós-graduação (stricto e lato sensu), ensino técnico, ensino médio e infantil, submissão e controle de projetos e bolsistas de pesquisa, submissão e controle de ações de extensão, submissão e controle dos projetos de ensino (monitoria e inovações), registro e relatórios da produção acadêmica dos docentes, atividades de ensino a distância e um ambiente virtual de aprendizado denominado Turma Virtual.

Sendo assim, [Carvalho 2010] enfatiza que os sistemas de gestão acadêmica são de extrema importância para redução de custos e aquisição de ferramentas específicas para auxílio na resolução de demandas institucionais. Nesse sentido, os sistemas de gestão acadêmica desempenham um papel crucial na modernização e eficiência das instituições de ensino. Eles não apenas automatizam processos administrativos, mas também proporcionam uma série de benefícios que impactam diretamente a qualidade e eficácia da gestão institucional.

4. Objetivos

O objetivo geral deste projeto consiste em desenvolver uma aplicação web que proporcione um acesso simplificado às informações dos docentes vinculados ao ICEN da UFPA. Logo, busca-se otimizar o acesso às informações e fortalecer a comunicação entre a comunidade acadêmica, promovendo uma maior integração e visibilidade dos docentes.

Os objetivos específicos do trabalho de pesquisa foram:

- Realizar a coleta de dados de fontes relevantes utilizando técnicas apropriadas.
- Armazenar e organizar os dados adquiridos em um sistema de gerenciamento de banco de dados eficiente.
- Desenvolver APIs para facilitar o acesso e a manipulação dos dados armazenados.
- Estabelecer integrações eficazes entre os componentes do sistema para garantir um fluxo de dados eficiente.
- Implementar recursos de interfaces para facilitar a pesquisa, filtragem e navegação sobre a busca de informações.

5. Metodologia

Nesta seção, será destacado a metodologia adotada no desenvolvimento do projeto (Figura 1).

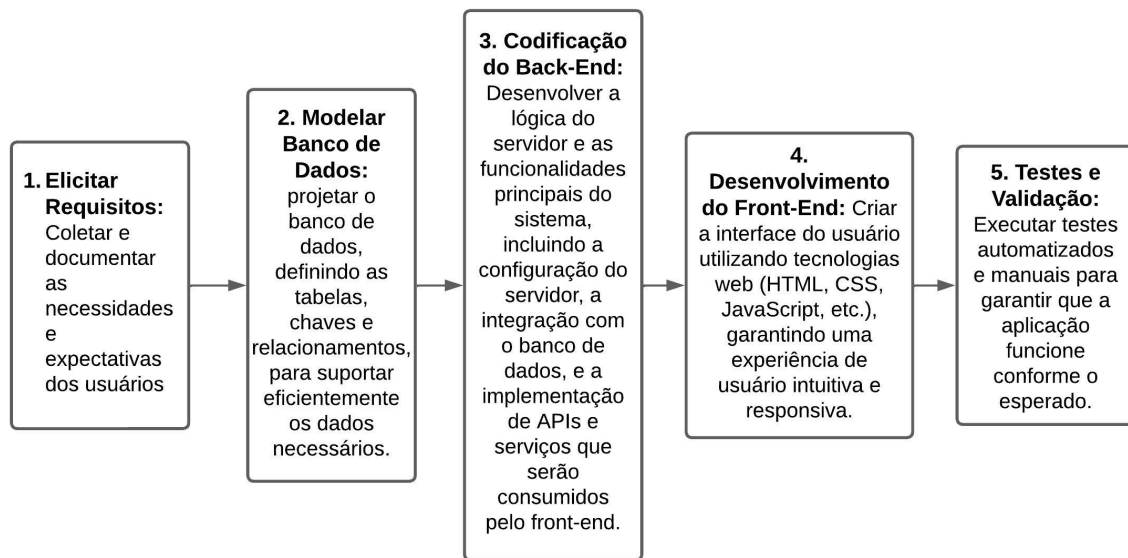


Figura 1. Cronologia da metodologia de criação do Data Insights.

Inicialmente, foram definidos os requisitos funcionais e não-funcionais da aplicação *Data Insights*, os quais serviram como base para direcionar todo o processo de desenvolvimento. Dessa forma, foram aplicadas diversas técnicas, desde o *Web Scraping*, para coleta dos dados, até o desenvolvimento de uma API com Node.js, para construir os serviços e gerenciar o acesso ao banco de dados. Além disso, a construção de uma aplicação web de fácil interação, utilizando o *framework* React, proporcionou uma experiência adequada na busca por informações relevantes. Ao final do processo, foram realizados testes abrangentes para garantir o correto funcionamento dos recursos do *Data Insights*.

No que diz respeito à implementação, foram utilizadas as tecnologias Node.js, React.js e SQL para atender aos requisitos do projeto. A coleta de dados das publicações dos docentes foi realizada por meio da plataforma Google Acadêmico, através de técnicas de *web scraping* com Node.js. Essas informações foram extraídas a partir de um arquivo de extensão CSV, contendo os nomes dos docentes do ICEN da UFPA e os endereços de acesso aos seus perfis do Google Acadêmico.

O *front-end* da aplicação, construído com React.js, possibilitou a interação do usuário por meio do envio de arquivos de extensão .csv e .zip contendo currículos Lattes dos professores. Esses arquivos foram processados pelo *back-end*, que extraiu o número de publicações de cada professor, bem como seu programa de pós-graduação no instituto. Os dados coletados foram armazenadas em um banco de dados MySQL que, posteriormente, foram disponibilizados ao *front-end*. Os dados foram apresentados aos usuários por meio dos recursos de interação do sistema, tais como as tabelas e os gráficos.

Além das funcionalidades mencionadas, a aplicação também inclui uma área administrativa que permite a exclusão de informações da tabela, importação de arquivos .csv e .zip e inserção de informações relacionadas ao programa de desenvolvimento da unidade. Essas informações englobam dados sobre a quantidade de alunos que ingressaram,

concluíram, estão para concluir ou desistiram de determinado programa de pós-graduação do instituto, considerando o ano especificado no formulário de busca.

Para validar a eficácia do *Data Insights*, foram conduzidos testes experimentais utilizando uma base de dados abrangendo o currículo Lattes de professores que compõem os corpos docentes dos Programas de Pós-Graduação no ICEN da UFPA. Esses testes permitiram verificar e validar a funcionalidade e a precisão dos dados fornecidos pela aplicação. Os tipos de testes realizados incluíram:

- **Testes Funcionais:** Estes testes foram realizados para assegurar que todas as funcionalidades da aplicação estavam operando conforme o esperado. Envolveu o upload de um arquivo .csv contendo os dados de vários docentes para testar a capacidade da aplicação em processar essas informações e realizar a varredura no Google Scholar. A verificação focou na precisão dos dados extraídos e na capacidade da aplicação de lidar com diferentes formatos e variações nos dados de entrada.
- **Testes de Integração:** Foram realizados para garantir que todos os componentes da aplicação estavam integrados corretamente e funcionando de maneira coesa. Estes testes verificaram a interação entre a aplicação e o Google Scholar, assegurando que os dados extraídos fossem corretamente associados aos perfis dos docentes no Google Scholar, sem perda ou corrupção de informações.

Os testes foram conduzidos pelo próprio desenvolvedor do projeto, garantindo uma validação detalhada e prática da aplicação. Foi constatado que a aplicação realizava corretamente a varredura no Google Scholar, fornecendo dados precisos e funcionais, o que valida sua eficácia e utilidade para a instituição.

6. Desenvolvimento do Data Insights

Nesta seção do trabalho, serão apresentadas duas listas de Requisitos definidas para o desenvolvimento do sistema Data Insights, uma de requisitos funcionais na Tabela 1, e outra de Requisitos não funcionais na Tabela 2.

O sistema foi desenvolvido a partir de boas práticas de construção de código, tais como: controle de versão, testes automatizados e padrões de codificação. Essas práticas visam facilitar possíveis alterações e manutenções futuras, além de promover uma base sólida e sustentável para o desenvolvimento do sistema.

6.1. Requisitos Funcionais

Na Tabela 1, foram definidos um total de 10 requisitos funcionais, numerados de RF 01 até RF 10, que são essenciais para o funcionamento adequado do sistema. Além disso, cada requisito possui uma prioridade atribuída, a fim de destacar aqueles mais urgentes durante a implementação.

Dois requisitos funcionais prioritários são o RF 01 e o RF 03, que determinam a necessidade de permitir a inserção de novos professores no sistema. Isso possibilita ao usuário adicionar *links* do Google Acadêmico de professores de interesse através de um arquivo de extensão .csv e também inserir novos currículos lattes como arquivo de extensão .zip na área administrativa, a fim de obter os dados relevantes sobre suas atividades

Tabela 1. Requisitos Funcionais do Data Insights.

Nº	Descrição	Prioridade
RF01	Deve-se permitir a inserção de novos professores no sistema	Alta
RF02	Deve-se permitir deletar professores no sistema	Alta
RF03	Deve-se permitir a inserção de novos currículos lattes no sistema	Alta
RF04	Deve-se permitir a exclusão de dados referentes ao Programa de Desenvolvimento da Unidade por Programa de Pós-Graduação	Alta
RF05	O sistema deve gerar gráficos para melhor visualização dos dados pelo usuário	Média
RF06	O sistema deve conter uma área administrativa para gerência dos dados informados	Média
RF07	Deve-se permitir a navegação por páginas através de um <i>sidebar</i> lateral no sistema	Média
RF08	Deve-se permitir que o usuário classifique por coluna a ordem de apresentação da tabela de docentes do Google Acadêmico	Baixa
RF09	Deve-se permitir o usuário através de botão explícito visitar o link do Google Acadêmico do docente	Baixa
RF10	Deve-se permitir o usuário selecionar dados referentes ao Programa de Pós-Graduação requerido	Média

acadêmicas. Esse requisito é considerado de alta prioridade, pois é fundamental para o funcionamento básico do sistema.

Outros requisitos funcionais importantes são o RF 02 e RF 04, que envolvem a remoção de informações irrelevantes para o sistema. Essa funcionalidade contribui significativamente para a gerência dos dados exibidos pois sua administração é um elemento essencial para o sistema como um todo.

Além disso, há requisitos funcionais menos relevantes, como o RF 07 e RF 08, que tratam de categorização e classificação de professores por meio de colunas e navegação entre páginas através de um *sidebar*. Embora esses requisitos definam funcionalidades importantes, o sistema ainda seria capaz de cumprir seu objetivo principal mesmo que eles não sejam implementados.

Outros requisitos funcionais incluem a possibilidade de visualização dos dados por meio de formas gráficas para o usuário (RF 05), área administrativa que solicita credenciais para realização de ações com potencial alto na aplicação, como inserir e deletar (RF 06), botões de fácil acesso que façam referência a um professor e que tenham os dados referentes ao seu perfil no Google Acadêmico (RF 09) e seleção de dados através de um botão de opções para exibição de dados conforme programa de pós graduação selecionado (RF 10).

Esses requisitos funcionais foram cuidadosamente definidos para garantir que a aplicação atenda as necessidades dos usuários, fornecendo uma aplicação web adequada para a análise dos dados acadêmicos dos docentes do ICEN da UFPA.

6.2. Requisitos Não Funcionais

Foram definidos 6 requisitos não funcionais para o *Data Insights* (Tabela 2), numerados de RNF 01 até RNF 06, com suas respectivas prioridades. O RNF 01 estabelece que o sistema deve ser exclusivamente web e ser compatível com todos os navegadores de *Internet*, garantindo uma ampla acessibilidade aos usuários. A prioridade média do RNF 02 determina que o sistema deve ter sua interface e conteúdo em PT-BR, para melhorar a usabilidade e compreensão pelos usuários.

Os requisitos RNF 03 e RNF 04 definem que o sistema deve ser desenvolvido em Node.js para a criação do back-end e em React.js para a criação do front-end, respectiva-

Tabela 2. Requisitos Não Funcionais do Data Insights.

Nº	Descrição	Prioridade
RNF01	O sistema deve ser necessariamente WEB e deverá ser compatível com todos os navegadores atuais	Alta
RNF02	O sistema deve ter sua linguagem em PT-BR	Média
RNF03	O sistema deve ser desenvolvido em NodeJS para criação de back-end	Alta
RNF04	O sistema deve ser desenvolvido em ReactJS para criação de front-end Desenvolvimento da Unidade por Programa de Pos-Graduação	Alta
RNF05	O sistema deve possuir um banco de dados construído em MySQL pelo usuário	Alta

mente. Essas escolhas tecnológicas são fundamentais para desenvolver uma arquitetura que ofereça estruturas facilmente escaláveis e de alto desempenho.

O RNF 05 estabelece que deve-se utilizar o Sistema Gerenciador de Banco de Dados (SGBD) Mysql para gerenciamento dos dados do *Data Insights*, oferecendo uma estrutura robusta e confiável para a aplicação.

Esses requisitos não funcionais desempenham um papel fundamental no desenvolvimento da aplicação, garantindo que o sistema seja acessível, compatível com as tecnologias mais recentes, seguro e preparado para futuras evoluções.

6.3. Construção da API

O desenvolvimento da camada de *back-end* do *Data Insights* envolve a criação da API responsável por fornecer os dados e serviços necessários à camada de *front-end*. Neste caso, o Node.js foi a linguagem de programação escolhida para o desenvolvimento do *back-end*, devido a sua natureza assíncrona e eficiência em lidar com operações em tempo real. Também foram utilizadas bibliotecas externas devido a dependência dos serviços da aplicação, que são apresentadas na Tabela 3.

Tabela 3. Bibliotecas Utilizadas no Back-End.

Nome	Versão	Descrição
adm-zip	0.5.10	Utilizada para a manipulação de arquivos .zip contendo os currículos Lattes dos professores, permitindo a extração das informações necessárias para análise.
body-parser	1.20.0	Responsável pelo parsing das requisições HTTP, possibilitando o acesso aos dados enviados no corpo das requisições.
cheerio	1.0.0-rc.12	Utilizada para realizar o <i>web scraping</i> dos perfis dos professores no Lattes, extraindo as informações relevantes para análise acadêmica.
cors	2.8.5	Habilita a política de mesmo-origem do CORS (Cross-Origin Resource Sharing) para permitir as requisições HTTP entre diferentes origens.
express	4.18.1	<i>Framework</i> Node.js para a construção de aplicativos web, utilizado para a criação das rotas e lógica de negócio da aplicação back-end.
fast-csv	4.3.6	Utilizada para a leitura e escrita de arquivos .csv contendo informações dos professores, facilitando o processamento e a persistência desses dados.
iconv-lite	0.6.3	Biblioteca utilizada para a conversão de <i>encodings</i> , permitindo a manipulação adequada dos dados importados dos currículos Lattes.
multer	1.4.5-lts.1	<i>Middleware</i> para o gerenciamento de uploads de arquivos, utilizado para receber os arquivos .csv e .zip enviados pelo usuário.
mysql2	2.3.3	Biblioteca para interação com o banco de dados MySQL, utilizada para a criação e execução de consultas no banco de dados da aplicação.
puppeteer	14.2.1	Biblioteca Node.js para automação de navegadores, utilizada para o <i>Web Scraping</i> dos perfis dos docentes no Google Acadêmico.
xml2js	0.5.0	Utilizada para a conversão de arquivos XML dos currículos Lattes em objetos JavaScript, facilitando o acesso às informações contidas nesses arquivos.

6.4. Construção do Front-end

O *front-end* da aplicação *Data Insights*, desenvolvido em React.js, desempenha um papel crucial na interface com o usuário e na apresentação dos dados e informações relevantes aos pesquisadores e profissionais acadêmicos. React.js é uma biblioteca JavaScript moderna que permite criar interfaces de usuário interativas e responsivas, tornando-a uma escolha ideal para a construção da camada de *front-end* dessa aplicação.

O uso de componentes reutilizáveis do React.js permitiu o desenvolvimento de uma arquitetura modular e escalável, facilitando a adição de novos recursos e funcionalidades à medida que a aplicação evolui.

A integração de bibliotecas e ferramentas adicionais enriqueceu a experiência do usuário, permitindo a criação de gráficos interativos, seleção de opções personalizadas e aprimoramento visual geral do *Data Insights*. A Tabela 4 apresenta a lista de bibliotecas e ferramentas adicionais utilizadas *Data Insights*.

Tabela 4. Bibliotecas Utilizadas no Front-End.

Nome	Versão	Descrição
bootstrap	5.1.3	Framework CSS para estilização da aplicação, fornecendo estilos de componentes para fácil integração nos sistemas desenvolvidos.
chart.js	3.8.0	Biblioteca para a criação de gráficos e visualização de dados.
react	18.2.0	Biblioteca para a criação de interfaces de usuário reativas e interativas.
react-bootstrap	2.4.0	Biblioteca que combina o React.js com o Bootstrap, oferecendo componentes estilizados e customizáveis.
react-chartjs-2	4.2.0	Componente para integrar o Chart.js ao React.js, possibilitando a criação de gráficos dinâmicos.
react-dom	18.1.0	Biblioteca para a manipulação do DOM no React.js.
react-google-charts	4.0.0	Biblioteca para a manipulação do DOM no React.js.
react-icons	4.4.0	Conjunto de ícones para uso no React.js.
react-router-dom	6.3.0	Biblioteca para o gerenciamento de rotas no React.js.
react-scripts	5.0.1	Conjunto de scripts para a construção e execução da aplicação React.js.
react-select	5.3.2	Biblioteca para a criação de menus de seleção no React.js.
react-toastify	9.1.1	Biblioteca para a exibição de notificações no React.js.

6.5. Construção do Banco de Dados

No processo de construção do Banco de Dados para o *Data Insights*, utilizou-se a ferramenta de administração de banco de dados MySQL Workbench para gerenciar e modelar a estrutura do banco. A escolha do MySQL se deve à sua ampla aceitação e confiabilidade no mercado, sendo uma solução robusta e segura para o gerenciamento de grandes volumes de dados. O MySQL é conhecido por sua eficiência em operações de leitura e escrita, o que é crucial para garantir o desempenho do sistema. Além disso, o MySQL Workbench oferece uma interface intuitiva que facilita a modelagem e a visualização da estrutura do banco de dados, permitindo uma criação eficiente das tabelas necessárias para armazenar as informações essenciais da aplicação, com suporte para SQL avançado e ferramentas de design que ajudam a evitar erros de configuração. A combinação dessas características fez do MySQL a escolha ideal para suportar as demandas do *Data Insights*, garantindo escalabilidade e flexibilidade futuras.

O Banco de Dados foi projetado para atender às necessidades específicas da aplicação, garantindo a eficiência no armazenamento e recuperação de dados. Ele é composto por três tabelas fundamentais:

- **Tabela Lattes:** esta tabela desempenha um papel central no armazenamento de informações extraídas dos arquivos de extensão xml. Ela foi cuidadosamente projetada para abranger todos os dados relevantes dos currículos Lattes dos pesquisadores. Com campos bem definidos para armazenar informações como o Programa de Pós-Graduação do docente e o número de publicações por ano do mesmo, a Tabela Lattes é a base para disponibilizar dados de pesquisa e currículos dos pesquisadores de forma organizada e acessível.
- **Tabela PDU (Programa de Desenvolvimento da Unidade):** essa tabela é responsável por armazenar todas as informações preenchidas no formulário da área administrativa. O formulário é utilizado para compor a tabela do Programa de Desenvolvimento da Unidade, que por meio dela os dados são estruturados e mantidos para futuras análises e relatórios. Essa tabela desempenha um papel fundamental na avaliação e monitoramento do desenvolvimento da unidade, fornecendo informações importantes para tomada de decisões estratégicas.
- **Tabela Professores:** a terceira tabela é dedicada a conter informações dos docentes do Google Acadêmico. Esses dados são obtidos a partir de um arquivo de extensão csv enviado pela área administrativa. A Tabela de professores é projetada para armazenar detalhes relevantes sobre os docentes, incluindo citações por ano de cada docente, entre outros dados acadêmicos relevantes. Essas informações são cruciais para proporcionar uma visão abrangente do perfil dos docentes, permitindo uma análise mais aprofundada das suas atividades de pesquisa.

Em conjunto, essas três tabelas compõem um Banco de Dados robusto e completo, que serve como a base sólida da *Data Insights*. Com a sua criação cuidadosa e estruturação eficiente, o banco de dados garante que os dados sejam armazenados e organizados de maneira coerente, possibilitando análises precisas e relatórios relevantes que atendam às demandas da comunidade acadêmica e contribuam para o avanço do conhecimento científico.

7. Apresentação do *Data Insights*

Nesta seção, exploraremos as funcionalidades implementadas no *Data Insights*, destacando como cada componente contribui para alcançar os objetivos do projeto e melhorar a experiência dos usuários.

7.1. Página Inicial do *Data Insights*

Inicialmente, pode-se visualizar a tela inicial ao ser executada pela primeira vez (RF07). Ao expandir o *sidebar* (Figura 2), pode-se visualizar, além dos ícones, o significado de cada um (RF07).

Na parte esquerda da imagem (Figura 2), encontramos um menu com várias opções de interação disponíveis ao usuário. Essas opções são descritas a seguir.

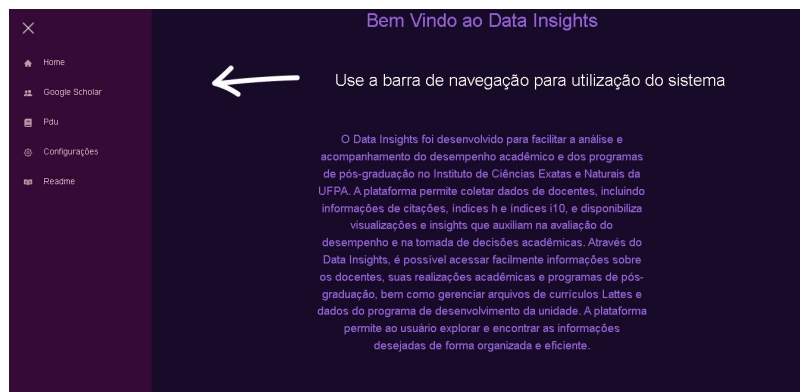


Figura 2. Página Inicial do Data Insights.

7.2. Google Scholar

A opção do Google Scholar (Figura 3) redireciona o usuário para uma tabela com as seguintes informações: **nome do docente**; **subunidade**; **número de citações em todas as publicações**; **número de citações em todas as publicações nos últimos 5 anos**; **índice h, que é o maior número h, sendo que h publicações possuem, no mínimo, h citações**; **índice h, nos últimos 5 anos**; **índice i10, que é o número de publicações com, no mínimo, 10 citações**; e, por fim, o **índice i10, nos últimos 5 anos**.

Ainda na tabela temos ao lado do nome de cada docente a opção de um botão com uma *link* que redireciona o usuário para o Google Acadêmico do referido docente (RF09), no topo da página há um botão com a opção **Atualizar**, que atualiza a informação de todos os docentes contidos na tabela, e por fim, em cada coluna da tabela é possível realizar a ordenação ao clicar em cima da coluna, com a finalidade de filtrar informações (RF08).

#	Docente	Subunidade	Citações	Índice h	Índice i10	Citações desde 2018	Índice h desde 2018	Índice i10 desde 2018
1	Aldabaro Barreto da Rocha Klautau Junior	Computação	4574	22	52	1818	17	25
2	Antônio Jorge Gomes Abelem	Física	1107	17	29	356	8	7
3	Blanchi Senque Meiguins	Química	620	14	17	356	10	12
4	Carlos Gustavo Resque dos Santos	Estatística	227	9	8	211	9	8
5	Carlos Renato Lisboa Frances	Computação	1143	16	33	562	12	17
6	Claudemiro de Souza de Sales Junior	Estatística	682	11	14	567	10	11
7	Cleudson Ronald Botelho de Souza	Computação	4019	31	88	1538	17	40
8	Denis Lima do Rosário	Estatística	2130	24	59	1488	21	47
9	Eduardo Coelho Senqueira	Física	4354	32	117	2525	25	67
10	Eloi Luiz Favero	Computação	632	14	20	176	7	3
11	Filipe de Oliveira Sarava	Estatística	190	7	7	170	7	6
12	Gustavo Henrique Lima Pirtto	Química	2913	29	74	2420	27	70
13	Gustavo Pessin	Computação	3009	24	56	2403	23	46
14	Jefferson Magalhães de Moraes	Estatística	342	11	13	230	9	9
15	Josivaldo de Souza Araujo	Computação	55	4	1	21	3	0

Figura 3. Tabela de Docentes com Informações do Google Acadêmico.

Abaixo da tabela do Google Acadêmico (Figura 3), possuímos 3 gráficos (Citações, Índice h e Índice i10), em formato de barra (Figura 4), para melhor visualização das informações (RF05).

Este é o número de citações de todas as publicações. A segunda coluna tem a versão "recente" desta métrica, que é o número de novas citações nos últimos 5 anos de todas as publicações.

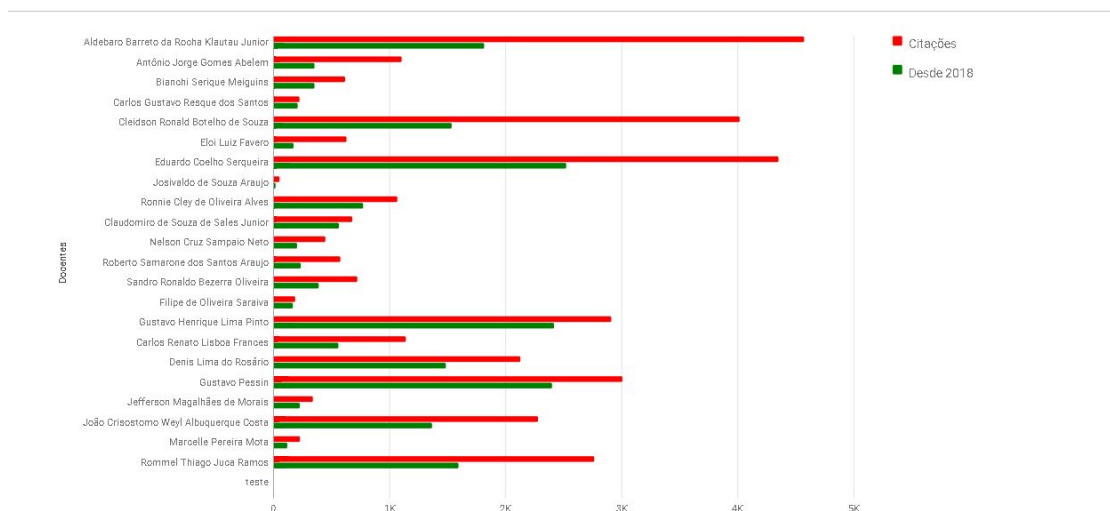


Figura 4. Gráfico de citações para exemplificação.

7.3. Programa de Desenvolvimento da Unidade

Esta opção é responsável por conter informações do Programa de Desenvolvimento da Unidade (PDU), com dados de todos os programas de pós-graduação do ICEN individuais e geral.

Na parte superior da página temos dois botões com a opção de “**Selecionar PPG**” e “**Atualizar**”(Figura 5), o primeiro é responsável por selecionar o programa de pós-graduação a ser exibido (RF10) e o segundo é responsável por atualizar as informações contidas nos currículos Lattes de todos os docentes referentes aos seus respectivos institutos.

Ao selecionar um programa de pós-graduação, a tabela exibirá dados estimados e reais dos 2 anos anteriores e posteriores do atual, das seguintes informações: “**Taxa de Sucesso do curso de Mestrado**”, “**Taxa de Sucesso do curso de Doutorado**”, “**Taxa de Retenção no Mestrado**”, “**Taxa de Retenção no Doutorado**”, “**Taxa de Evasão no Mestrado**”, “**Taxa de Evasão no Doutorado**”, “**Nº de Cursos de Doutorado Profissional**”, “**Média dos Conceitos CAPES/MEC dos cursos de Pós-Graduação**”, “**Nº de Titulados em Programas de Pós-Graduação**”, “**Nº de Artigos Publicados em Periódicos**” e “**Nº de docentes afastados para capacitação no exterior (pós-doutorado, etc)**” (Figura 5).

GERAL		2021	2022	2023	2024	2025
Indicadores						
Taxa de Sucesso do Curso de Mestrado (Estimado)		88%	88%	88%	88%	88%
Taxa de Sucesso do Curso de Mestrado (Real)		26%	26%	26%	%	%
Taxa de Sucesso do Curso de Doutorado (Estimado)		71%	71%	71%	71%	71%
Taxa de Sucesso do Curso de Doutorado (Real)		38%	38%	38%	%	%
Taxa de retenção no Mestrado (Estimado)		61%	61%	61%	61%	61%
Taxa de retenção no Mestrado (Real)		48%	48%	48%	%	%
Taxa de retenção no Doutorado (Estimado)		57%	57%	57%	57%	57%
Taxa de retenção no Doutorado (Real)		49%	49%	49%	%	%
Taxa de evasão no Mestrado (Estimado)		56%	56%	56%	56%	56%
Taxa de evasão no Mestrado (Real)		32%	32%	32%	%	%
Taxa de evasão no Doutorado (Estimado)		56%	56%	56%	56%	56%
Taxa de evasão no Doutorado (Real)		27%	27%	27%	%	%
Nº de Cursos de Doutorado Profissional (Estimado)		31	31	44	37	43
Nº de Cursos de Doutorado Profissional (Real)		41	41	27	25	44
Média dos Conceitos CAPES/MEC dos cursos de Pós Graduação (Estimado)		33	34	35	45	45
Média dos Conceitos CAPES/MEC dos cursos de Pós Graduação (Real)		45	38	63	62	53
Nº de Titulados em Programas de Pós Graduação (Estimado)		53	45	55	45	57
Nº de Titulados em Programas de Pós Graduação (Real)		44	35	72	47	48
Nº de Artigos Publicados em Periódicos (Estimado)		27	39	41	54	58
Nº de Artigos Publicados em Periódicos (Real)		38	23	0	0	0
Nº de docentes afastados para capacitação no exterior (pós-doutorado, etc) (Estimado)		53	53	53	53	53
Nº de docentes afastados para capacitação no exterior (pós-doutorado, etc) (Real)		31	31	31	31	31

Figura 5. Programa de Desenvolvimento da Unidade.

Após a tabela exibindo todos os dados do Programa de Desenvolvimento da Unidade, há diversos gráficos para visualizar as mesmas informações de outra perspectiva (Figura 6).

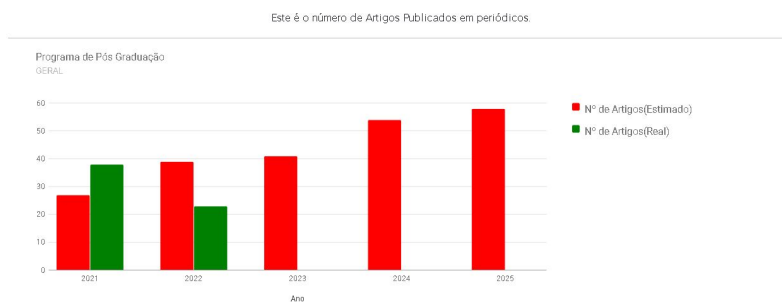


Figura 6. Gráfico com informações do Programa de Desenvolvimento da Unidade para exemplificação.

7.4. Configurações

Esta área (Figura 7), é destinada à administração dos dados que o sistema contém. Para ter acesso a esse recurso, é necessário possuir as credenciais de administrador (RF06).

Figura 7. Configurações.

Quando credenciado, o administrador possuirá as opções (Figura 8): “**Importar Lista Google Acadêmico**”, “**Preencher Informações do PDU**”, “**Remover Docente**”, “**Remover Informações do PDU**” e “**Sair**”.

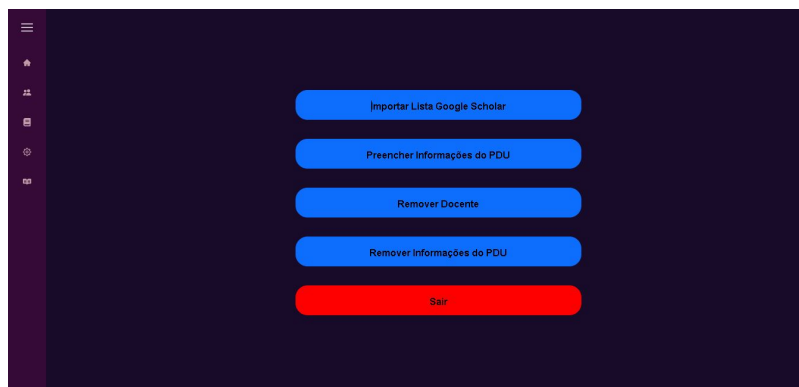


Figura 8. Opções do administrador.

Ao selecionar a opção “**Importar Lista Google Acadêmico**”(Figura 9), o administrador terá a opção de realizar o upload de um arquivo .csv contendo obrigatoriamente as seguintes informações: 3 (três) colunas, a primeira com os nomes dos docentes, a segunda com seus respectivos *links* do Google Acadêmico e a terceira com a subunidade do docente inserido (RF01).

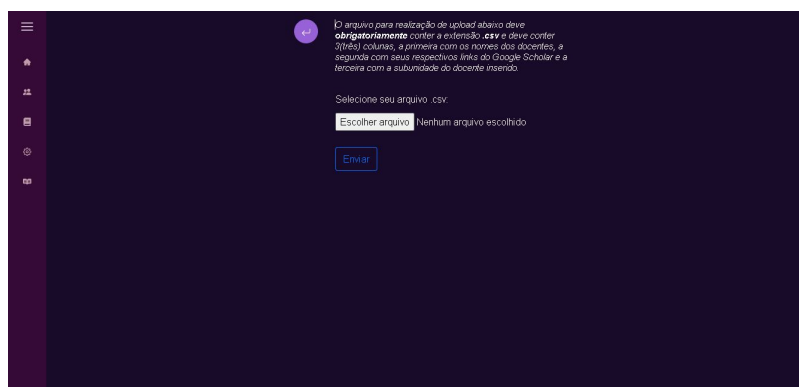


Figura 9. Importação de arquivo .csv.

Selecionando a opção “**Preencher Informações de PDU**”(Figura 10), o administrador deverá selecionar o Programa de pós-graduação e o ano que as informações preenchidas serão alocadas e ao final deverá submeter essas informações clicando no botão de “**Enviar**”.

Informe o Ano e PPG que você deseja inserir informações sobre o PDU

2025 PPGCC

Dados Estimados

Qual a taxa de sucesso ESTIMADA no mestrado no ano de 2025?

Qual a taxa de retenção ESTIMADA no mestrado no ano de 2025?

Qual a taxa de evasão ESTIMADA no mestrado no ano de 2025?

Qual a taxa de sucesso ESTIMADA no doutorado no ano de 2025?

Qual a taxa de retenção ESTIMADA no doutorado no ano de 2025?

Qual a taxa de evasão ESTIMADA no doutorado no ano de 2025?

Qual o número de cursos de Doutorado Profissional ESTIMADO no ano de 2025?

Qual a Média dos Conceitos CAPES/MEC dos cursos de Pós-graduação ESTIMADA no ano de 2025?

Qual o Nº de Trabalhos em Programas de Pós-graduação ESTIMADO no ano de 2025?

Qual o Nº de docentes atuando para capacitação no exterior (pós-doutorado, etc) ESTIMADO no ano de 2025?

Qual o Nº de Artigos Publicados em Periódicos ESTIMADO no ano de 2025?

Dados Reais

Quantos alunos ingressaram em 2025 no mestrado?

Quantos alunos concluíram em 2025 o mestrado?

Quantos alunos estão ativos no 2025 e ainda não defenderam o mestrado?

Quantos alunos no 2025 foram designados do mestrado (desistiram ou cancelaram)?

Quantos alunos ingressaram em 2025 no Doutorado?

Quantos alunos concluíram em 2025 o Doutorado?

Quantos alunos estão ativos no 2025 e ainda não defenderam o Doutorado?

Quantos alunos no 2025 foram designados do Doutorado (desistiram ou cancelaram)?

Qual o número de cursos de Doutorado Profissional REAL no ano de 2025?

Qual a Média dos Conceitos CAPES/MEC dos cursos de Pós-graduação REAL no ano de 2025?

Qual o Nº de Trabalhos em Programas de Pós-graduação REAL no ano de 2025?

Qual o Nº de docentes atuando para capacitação no exterior (pós-doutorado, etc) REAL no ano de 2025?

Finalizar

Figura 10. Formulário do Programa de Desenvolvimento da Unidade.

Caso o administrador selecione a opção “**Remover Docente**”(Figura 11), será exibida uma tabela com os nomes dos docentes da tabela Google Acadêmico, e na última coluna um botão com o sinal negativo (-), possibilitando a remoção do professor (RF02).

#	Professor	Citações	Índice h	Índice I10	Citações desde 2017	Índice h desde 2017	Índice I10 desde 2017	Opções
1	Aldebara Barreto da Rocha Klautau Junior	4574	22	52	1819	17	25	
2	Antônio Jorge Gomes Abelem	1107	17	29	356	8	7	
3	Blanchi Senique Meiguins	620	14	17	356	10	12	
4	Carlos Gustavo Resque dos Santos	227	9	8	211	9	8	
5	Claudson Ronald Beteito de Souza	4019	31	88	1538	17	40	
6	Eloi Luiz Favero	632	14	20	176	7	3	
7	Eduardo Coelho Serqueira	4354	32	117	2525	25	67	
8	Josivaldo de Souza Araújo	55	4	1	21	3	0	
9	Ronnie Clei de Oliveira Alves	1070	19	28	775	16	28	
10	Claudomiro de Souza de Sales Junior	682	11	14	567	10	11	
11	Nelson Cruz Sampaio Neto	450	11	14	206	7	5	
12	Roberto Samaron dos Santos Araújo	579	11	14	238	9	9	
13	Sandro Ronaldo Bezerra Oliveira	725	11	14	393	9	8	

Figura 11. Remoção de docente.

Ao selecionar a opção “**Remover Informações do PDU**”(Figura 12), o administrador deverá selecionar o Programa de pós-graduação a qual deseja remover todos os seus dados do banco, e após a seleção clicar no botão “**Deletar**”.

Informe o PPG que você deseja DELETAR informações sobre o PDU

PPGCC

Deletar

Figura 12. Remoção de informações do programa de desenvolvimento da unidade.

7.5. Readme

Esta opção é responsável por fornecer um arquivo `readme.txt` que contém todas as informações de uso do sistema ao usuário.

8. Considerações Finais

Neste trabalho, foi desenvolvido uma aplicação web capaz de extrair e apresentar informações relevantes dos currículos Lattes e do Google Acadêmico de pesquisas realizadas por docentes para apoiar na tomada de decisões acadêmico-pedagógicas do ICEN da UFPA. A automatização desse processo representa uma significativa contribuição para análises acadêmicas, pois reduz consideravelmente o tempo despendido pelos gestores acadêmicos na obtenção dessas estatísticas, que frequentemente são coletadas manualmente.

Embora o *Data Insights* ainda tenha espaço para melhorias, ela já se mostra um facilitador para a comunidade acadêmica e outros usuários que dela necessitem. A simplicidade e a intuição na utilização, com apenas o requisito de conhecimento básico em computação, tornam-na acessível a diversos perfis de usuários.

Para trabalhos futuros, é relevante buscar melhorias na estilização das telas do sistema para torná-lo visualmente mais agradável ao usuário. Além disso, a hospedagem do *Data Insights* na Internet é um passo importante para ampliar seu alcance, uma vez que a versão atual só é executada localmente. Explorar outras áreas para extração de dados relevantes dos currículos Lattes e apresentá-los visualmente também é uma oportunidade de aprimoramento. Outro aspecto a ser considerado é a implementação de um banco de dados NoSQL, que pode oferecer maior flexibilidade e escalabilidade na gestão dos dados extraídos, permitindo lidar com grandes volumes de informação de forma mais eficiente e adaptável às futuras necessidades do sistema.

Entre as possibilidades de expansão, pode-se considerar a implementação de novas funcionalidades para atender as necessidades específicas dos gestores acadêmicos, como um filtro para selecionar anos específicos de publicações de pesquisadores sem a necessidade de selecionar todos, pois a página exibe desde o ano atual até os 4 anteriores, totalizando 5 anos.

Em resumo, o sistema de *Data Insights* desenvolvido neste trabalho demonstrou ser uma aplicação web promissora para a área acadêmica, com potencial para oferecer análises mais eficientes e auxiliar pesquisadores e gestores em suas atividades acadêmicas. As melhorias e futuras implementações podem impulsionar ainda mais o seu alcance, e melhorar positivamente a gestão acadêmica à tomada de decisão.

Referências

- Barroca Filho, I., Aquino, G., and Santa Rosa, J. G. (2013). Sigaa mobile—o caso de sucesso da ferramenta de gestão acadêmica na era da computação móvel. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 24, page 92.
- Carvalho, R. S. (2010). Sistemas de gestão da aprendizagem e sistemas de gestão acadêmica: avaliados pela ótica do docente. Master's thesis, Universidade Federal de Pernambuco.

- Cocón, F., Pérez-Cruz, D., Pérez-Rejón, J. Á., Zavaleta-Carrillo, P., Barradas-Arenas, U., Gómez-Ramón, R., and Cruz, J. A. P. (2023). Web scraping: Uso de plataformas de extracción de datos aplicadas a un sitio web sobre profesiones en México. *RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação*, (52):61–73.
- Costa, P. A. L. (2011). Educação superior na câmara dos deputados: um estudo sobre a pós-graduação lato sensu.
- Cysneiros, L. M. and do Prado Leite, J. C. S. (1997). Definindo requisitos não funcionais. In *Anais do XI Simpósio Brasileiro de Engenharia de Software*, pages 49–64. SBC.
- Delgado, A., Paulino, P., and Lagarto, S. (2017). É possível substituir os censos por informação administrativa? *XXIV Jornadas de Classificação e Análise de Dados (JOCLAD 2017)*, page 23.
- Figueiredo, E. (2011). Requisitos funcionais e requisitos não funcionais. *Icex, Dcc/Ufmg*, 14.
- Mitchell, R. (2015). *Web Scraping com Python: Coletando dados na web moderna*. Novatec Editora.
- Mitchell, R. (2019). *Web Scraping com Python: Coletando mais dados da web moderna*. Novatec Editora.
- Sommerville, I. (2011). Software engineering (ed.). *America: Pearson Education Inc.*
- Zanon, C. F. (2017). Aplicação de um programa de extração de dados através da técnica de web scrapin para determinação do perfil epidemiológico dos pacientes da clínica de estomatologia da faculdade de odontologia ufrj cadastrados no software estomato web.
- Zhao, B. (2017). Web scraping. *Encyclopedia of big data*, pages 1–3.