



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Curso de Graduação em Sistemas de Informação



Desenvolvimento de um Sistema Web Interativo para análise e visualização dos dados educacionais dos discentes do ICEA/UFOP

Diogo Leite Lucas

João Monlevade, MG
2024

Diogo Leite Lucas

**Desenvolvimento de um Sistema Web Interativo para análise e
visualização dos dados educacionais dos discentes do
ICEA/UFOP**

Trabalho de conclusão de curso apresentado ao curso de graduação em Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto, como parte dos requisitos necessários para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Profa. Dra. Helen de Cássia Sousa da Costa Lima

Coorientador: Prof. Dr. Thiago Augusto de Oliveira Silva

João Monlevade, MG

2024



FOLHA DE APROVAÇÃO

Diogo Leite Lucas

Desenvolvimento de um Sistema Web Interativo para análise e visualização dos dados educacionais dos discentes do ICEA/UFOP

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Sistemas de Informação.

Aprovada em 17 de outubro de 2024.

Membros da banca

Dra. Helen de Cássia Sousa da Costa Lima - Orientadora - Departamento de Computação e Sistemas - UFOP
Dr. Thiago Augusto de Oliveira Silva - Coorientador - Departamento de Engenharia de Produção - UFOP
Dr. George Henrique Godim da Fonseca - Departamento de Computação e Sistemas - UFOP
Dr. Igor Muzetti Pereira - Departamento de Computação e Sistemas - UFOP

Helen de Cássia Sousa da Costa Lima, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 07/11/2024.



Documento assinado eletronicamente por **Helen de Cassia Sousa da Costa Lima, PROFESSOR DE MAGISTERIO SUPERIOR**, em 07/11/2024, às 16:19, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0807586** e o código CRC **85DAB335**.

Este trabalho é dedicado à todas as pessoas que me apoiaram e acreditaram no meu potencial ao longo desta jornada acadêmica.

Agradecimentos

Primeiramente, agradeço a Deus por me guiar e me dar forças em cada etapa desta jornada. A Ele dedico todas as minhas conquistas, sabendo que sem sua graça e proteção, nada disso seria possível.

Gostaria de expressar minha profunda gratidão aos meus pais Ronilda Leite Lucas e José Feliciano Lucas, e também a minha irmã Diovana Aparecida Leite Lucas, pelo amor incondicional, apoio e compreensão em todos os momentos. Vocês sempre foram meu alicerce, oferecendo não apenas o suporte necessário, mas também acreditando em mim mesmo nas horas mais desafiadoras.

Aos meus amigos, que estiveram ao meu lado durante essa caminhada acadêmica, compartilhando risadas, momentos de tensão e palavras de encorajamento. Cada um de vocês fez a diferença em minha vida, e sou imensamente grato por tão grandeza amizade.

Agradeço também a minha orientadora Helen de Cassia Sousa da Costa Lima e ao meu coorientador Thiago Augusto de Oliveira Silva, que com paciência, dedicação e conhecimento, me guiaram ao longo deste trabalho. Suas orientações foram essenciais para que eu pudesse alcançar este resultado, e sou profundamente grato por todo o aprendizado proporcionado.

Por fim, não poderia deixar de agradecer a todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho. Cada gesto de apoio, cada palavra de incentivo e cada momento de ajuda foram fundamentais para que este trabalho se tornasse realidade. A todos, o meu mais sincero e profundo agradecimento.

"Pra ser bom é preciso sentir raiva da mediocridade."

— Filipe Ret

Resumo

Este trabalho apresenta o desenvolvimento de um sistema *web* de visualização de dados educacionais para uma instituição de ensino, utilizando a metodologia *Clean Architecture* como base para a estruturação do *software*. O objetivo principal foi criar uma solução que facilitasse o acesso, a análise e a gestão dos dados referentes à jornada acadêmica dos alunos, identificando padrões comportamentais dos alunos por meio de técnicas de clusterização, gerando *insights* sobre o desempenho acadêmico e possíveis tendências de evasão. Como resultado, concluiu-se que o sistema desenvolvido trouxe melhorias significativas na gestão e no monitoramento dos dados educacionais, promovendo uma visualização clara e acessível das informações acadêmicas.

Palavras-chaves: Evasão, Retenção, Clusterização, Desempenho acadêmico, Sistemas *Web*, ETL, *Clean Architecture*.

Abstract

This work presents the development of a web-based system for visualizing educational data for an academic institution, using the Clean Architecture methodology as the foundation for structuring the software. The main goal was to create a solution that would facilitate access, analysis, and management of data related to students' academic journeys, identifying behavioral patterns through clustering techniques, generating insights into academic performance and potential dropout trends. As a result, it was concluded that the developed system brought significant improvements in the management and monitoring of educational data, promoting a clear and accessible visualization of academic information.

Keywords: Academic dropout, Academic Retention, Clustering, Academic performance, Web Systems, ETL, Clean Architecture.

Lista de ilustrações

Figura 1 – Etapas do KDD	7
Figura 2 – Camadas da Clean Architecture (CA)	9
Figura 3 – Técnica de agrupamento	15
Figura 4 – Modelagem do banco de dados	24
Figura 5 – Arquitetura do Sistema em Alto Nível - <i>back-office</i>	26
Figura 6 – Arquitetura do Sistema em Alto Nível - <i>back-end e front-end</i>	26
Figura 7 – Captura de tela do sistema Back-office	31
Figura 8 – Captura de tela da página de login	32
Figura 9 – Captura de tela do sucesso de login	32
Figura 10 – Captura de tela do erro de login	33
Figura 11 – Container da aplicação	33
Figura 12 – Botão perfil	34
Figura 13 – Captura de tela da página inicial	35
Figura 14 – Captura de tela da página meu perfil	36
Figura 15 – Captura de tela da página de usuários	37
Figura 16 – Captura de tela da página de usuários	38
Figura 17 – Captura de tela do sucesso em cadastrar usuário	38
Figura 18 – Captura de tela do erro em cadastrar usuário	38
Figura 19 – Captura de tela do gerenciamento de acesso do usuário	39
Figura 20 – Captura de tela do sucesso em atualizar o acesso do usuário	39
Figura 21 – Captura de tela do erro em atualizar o acesso do usuário	39
Figura 22 – Captura de tela da confirmação de exclusão do usuário	40
Figura 23 – Captura de tela do sucesso em excluir o usuário	40
Figura 24 – Captura de tela do erro em excluir o usuário	40
Figura 25 – Captura de tela da página de relatório do aluno	45
Figura 26 – Captura de tela da página de relatório do aluno para docentes	46
Figura 27 – Captura de tela da página de clusterização da seção de situação dos alunos	47
Figura 28 – Captura de tela da página de clusterização da seção de coeficiente	48
Figura 29 – Captura de tela da página de clusterização da seção de tempo de curso	49
Figura 30 – Captura de tela da página de admissões	50
Figura 31 – Captura de tela da tabela da página de admissões	51
Figura 32 – Captura de tela de estatísticas gerais página de graduações	52
Figura 33 – Captura de tela do graduados por período de ingresso da página de graduações	53
Figura 34 – Captura de tela da página de evasões	54
Figura 35 – Captura de tela da página de disciplinas	55

Figura 36 – Captura de tela da grade de Sistemas de Informação na página grafo de dependências	56
Figura 37 – Captura de tela das estatísticas de disciplina na página grafo de dependências	56
Figura 38 – Captura de tela do índice da página de referência da API	57
Figura 39 – Captura de tela do módulo de requisição da página de referência da API . .	57
Figura 40 – Diagrama de casos de uso - Login	61
Figura 41 – Diagrama de casos de uso - Meu Perfil	62
Figura 42 – Diagrama de casos de uso - Usuários	63
Figura 43 – Diagrama de casos de uso - Relatório do Aluno	64
Figura 44 – Diagrama de casos de uso - Clusters	65
Figura 45 – Diagrama de casos de uso - Admissões	66
Figura 46 – Diagrama de casos de uso - Disciplinas	67
Figura 47 – Diagrama de casos de uso - Graduações	68
Figura 48 – Diagrama de casos de uso - Evasões	69
Figura 49 – Diagrama de casos de uso - Grafo de Dependências	70
Figura 50 – Diagrama de casos de uso - Referência da API	71
Figura 51 – Diagrama de casos de uso - Back-office	72

Lista de tabelas

Tabela 1 – Dicionário de dados da planilha DADOS INGRESSANTES.csv	12
Tabela 2 – Dicionário de dados da planilha NOTAS.csv	13
Tabela 3 – Dicionário de dados da planilha DADOS TRANCAMENTO.csv	13
Tabela 4 – Descrição dos campos extraídos	14
Tabela 5 – Descrição dos atributos gerados para modalidade Estatísticas de Graduações	17
Tabela 6 – Descrição dos atributos gerados para modalidade Estatísticas de Disciplinas	17
Tabela 7 – Descrição dos atributos gerados para modalidade Estatísticas de Admissões	17
Tabela 8 – Descrição dos atributos gerados para modalidade Estatísticas de Evasões . .	17
Tabela 9 – Descrição dos atributos gerados para modalidade de Clusterização	18
Tabela 10 – Descrição dos dados carregados	19
Tabela 11 – Descrição das principais tecnologias utilizadas no <i>back-office</i>	20
Tabela 12 – Descrição das principais tecnologias utilizadas no <i>back-end</i>	22
Tabela 13 – Descrição das principais tecnologias utilizadas no <i>front-end</i>	23
Tabela 14 – Resumo dos problemas encontrados com relatório consolidado	29
Tabela 15 – Resumo dos problemas encontrados pelos usuários e ações tomadas	30

Lista de quadros

Quadro 3.1 – Estrutura dos atributos acadêmicos	16
-----------------------------------------------------------	----

Lista de Algoritmos

1	Algoritmo k-means	5
---	-------------------------	---

Lista de abreviaturas e siglas

CA Clean Architecture

CD Entrega Contínua

CI Integração Contínua

CSV *Comma Separated Values*

Decea Departamento de Ciências Exatas e Aplicadas

DOM *Document Object Model*

ENEM Exame Nacional do Ensino Médio

ETL *Extract, Transform e Load*

HTTP *Hypertext Transfer Protocol*

ICEA Instituto de Ciências Exatas e Aplicadas

JSON *JavaScript Object Notation*

KDD *Knowledge-discovery in Databases*

PCA Análise de componentes principais

UML *Unified Modeling Language*

Sumário

1	INTRODUÇÃO	1
1.1	Justificativa do Tema	1
1.2	Problema de Pesquisa	2
1.3	Objetivos	2
1.4	Organização do Trabalho	3
2	REVISÃO DA LITERATURA	4
2.1	Sistemas Web	4
2.2	Tomada de Decisão Orientada a Dados	4
2.2.1	Algoritmo <i>k-means</i>	5
2.2.2	<i>Extract, Transform e Load (ETL)</i>	6
2.2.3	<i>Knowledge-discovery in Databases (KDD)</i>	6
2.3	<i>Clean Architecture</i>	8
2.4	Trabalhos Relacionados	9
2.4.1	Desenvolvimento de um Dashboard para Análise e Visualização dos Dados Educacionais dos Discentes do ICEA/UFOP	9
2.4.2	Caracterização do Perfil de Alunos do ICEA: Uma Análise Guiada por Dados e Indicadores de Sucesso	10
3	DESENVOLVIMENTO	11
3.1	Obtenção e Preparação dos Dados	11
3.1.1	Extração	11
3.1.2	Transformação	14
3.1.3	Carregamento	18
3.2	Construção dos Casos de Uso	18
3.3	Estrutura do Sistema	19
3.3.1	<i>Back-office</i>	20
3.3.2	<i>Back-end</i>	21
3.3.3	<i>Front-end</i>	22
3.4	Modelagem do Banco de Dados	24
3.5	Arquitetura do Sistema	25
3.6	Avaliação e Inspeção de Usabilidade do Sistema	27
4	RESULTADOS	30
4.1	Ajustes após Avaliação e Inspeção do Sistema	30
4.2	Interface do <i>Back-office</i>	31

4.3	Interface Login	32
4.4	Layout da Aplicação	33
4.5	Página Inicial	34
4.5.1	Meu Perfil	34
4.6	Usuários	35
4.6.1	Cadastrar Usuário	36
4.6.2	Gerenciar Acesso do Usuário	36
4.6.3	Excluir Usuário	38
4.7	Relatório do Aluno	39
4.8	Clusterização	41
4.9	Admissões	42
4.10	Graduações	42
4.11	Evasões	43
4.12	Disciplinas	43
4.13	Grafo de Dependências	44
4.14	Referência da API	44
5	CONCLUSÃO	58
	REFERÊNCIAS	59
	APÊNDICE A – DIAGRAMAS DE CASOS DE USO	61
A.1	Login	61
A.2	Meu Perfil	62
A.3	Usuários	63
A.4	Relatório do Aluno	64
A.5	Clusters	65
A.6	Admissões	66
A.7	Disciplinas	67
A.8	Graduações	68
A.9	Evasões	69
A.10	Grafo de Dependências	70
A.11	Referência da API	71
A.12	Back-office	72

1 Introdução

A queda no ingresso de novos estudantes em cursos e instituições de nível superior é notória e preocupante. Vinicius de Andrade ¹, fundador do projeto SalvaGuarda, que visa auxiliar alunos da rede pública de educação no Brasil no ingresso ao ensino superior, em entrevista ao Jornal da Unesp, ressalta que a queda no ingresso é motivada pela falta de incentivo dentro das redes públicas de educação, a falta de investimentos e cortes no limitado orçamento destinado às universidades, a eletização do principal meio de ingresso [Exame Nacional do Ensino Médio \(ENEM\)](#) e outros fatores sociais que afetam diretamente alunos em vulnerabilidade social.

O estudo feito pela [SEMESP \(2024\)](#) mostra a queda no ingresso no ensino superior em redes de ensinos públicas e privadas no período de 2015 à 2021, sendo 2.227.841 ingressantes em 2015, contra 1.467.717 em 2021, resultando em uma queda de 760.124 alunos ingressantes. Certamente, a preocupação diante do fato se agrava ainda mais quando em paralelo à queda no ingresso, temos uma crescente e elevada taxa de evasão no mesmo período, ou seja, temos menos estudantes ingressando no ensino superior e mais estudantes abandonando o curso, se agravando ao passar dos anos.

A situação se torna ainda mais crítica durante e após a pandemia, período que mais impactou nos índices da educação. Em atendimento às medidas emergenciais adotadas pelo governo, ocorreu o fechamento das instituições de ensino, ainda em março de 2020 ([COSTA, 2022](#)). A necessidade de manter o distanciamento social em conformidade com os protocolos de segurança internacional trouxe mudanças importantes e inevitáveis nos principais processos de ensino das universidades, como a educação remota e reformulação das avaliações.

1.1 Justificativa do Tema

A tecnologia da informação tornou-se essencial no contexto acadêmico, transformando a forma como o conhecimento é acessado, produzido e compartilhado. Plataformas digitais, ambientes virtuais de aprendizagem e ferramentas de colaboração online facilitaram a integração entre alunos, professores e pesquisadores, rompendo barreiras geográficas e temporais. Elas também possibilitam a gestão eficiente de dados acadêmicos, promovendo o ensino personalizado e viabilizando novas metodologias.

¹ Disponível em: <https://jornal.unesp.br/2023/06/22/por-que-o-numero-de-jovens-que-se-candidatam-a-uma-vaga-no-ensino-superior-gratuito-tem-caido-nos-ultimos-anos/>

A evasão escolar e a retenção de alunos têm sido fatores determinantes na adoção de sistemas de dados educacionais, pela necessidade de monitoramento da forma mais precisa e eficiente do desempenho acadêmico e a frequência dos alunos. Dados educacionais, quando bem estruturados, permitem identificar padrões de evasão, entender suas causas e desenvolver estratégias de intervenção mais direcionadas. A utilização desses dados proporciona uma visão mais clara das dificuldades enfrentadas pelos estudantes e possibilita a criação de políticas públicas focadas na redução da evasão, promovendo um ambiente educacional mais inclusivo e eficaz.

1.2 Problema de Pesquisa

Na universidade, há uma série de dados educacionais dos discentes que são compilados e analisados em relatórios acadêmicos. Contudo, há diversas melhorias a serem implementadas para aprimorar o relatório gerado, tornando-o mais acessível e de fácil compreensão para qualquer pessoa do ambiente acadêmico. O problema da evasão e da retenção de alunos na universidade pode ser melhor compreendido através da análise de dados e a acessibilidade da informação que é de grande relevância, pois permite que o conhecimento estudado seja público, auxiliando na tomada de decisões dos gestores e na compreensão, por parte dos alunos, sobre seu desempenho acadêmico ao longo dos anos. Quando os dados são analisados de forma eficiente, podem fornecer *insights* valiosos sobre diferentes aspectos do perfil dos discentes.

Nos institutos educacionais, decisões podem ser baseadas em má interpretação dos dados que podem gerar de maneira equivocada informações sobre o índice de evasão escolar. Quando os dados sobre desempenho, frequência e outros fatores críticos são mal utilizados, as estratégias de intervenção podem não ser funcionais. Assim, uma melhora da análise e uso dos dados educacionais auxilia para implementar ações mais assertivas em relação a evasão e a retenção.

1.3 Objetivos

O objetivo geral do presente trabalho é o desenvolvimento de um sistema *web* interativo para análise de dados educacionais dos discentes do [Instituto de Ciências Exatas e Aplicadas \(ICEA\)](#), com o intuito de auxiliar na tomada de decisão orientada a dados dentro da instituição. Logo, os objetivos específicos necessários para atender a proposta do trabalho são:

- Integrar as bases de dados disponíveis;
- Disponibilizar API *RESTful* com dados úteis para trabalhos futuros;
- Analisar os dados educacionais do *campus* através do uso de visualizações;
- Elaborar visualizações que auxiliem no monitoramento de indicadores de desempenho da instituição.

1.4 Organização do Trabalho

O presente trabalho está dividido em cinco capítulos, sendo o Capítulo 1 a introdução ao tema, apresentando uma contextualização e definindo o problema de pesquisa além da apresentação dos objetivos geral e específicos a serem alcançados. O Capítulo 2 apresenta os conceitos e fundamentações teóricas que foram utilizadas no desenvolvimento, bem como os trabalhos relacionados que precedem ao tema. O Capítulo 3 apresenta todas as etapas necessárias para a resolução do problema de pesquisa proposto para alcançar possíveis resultados que serão apresentados no Capítulo 4. Por fim, o Capítulo 5 apresenta a conclusão do trabalho, discutindo os objetivos alcançados e discussão de possibilidades de trabalhos futuros.

2 Revisão da Literatura

Este capítulo tem como objetivo fornecer a fundamentação teórica que sustenta o presente trabalho, definindo os termos relevantes para seu desenvolvimento e apresentando trabalhos relacionados que complementam o tema.

2.1 Sistemas *Web*

O conceito de *web* pode ser definido como um sistema que possibilita transferência e conexão entre hipertextos através da internet (PAULA *et al.*, 2020). Hipertextos são arquivos digitais que podem incluir elementos como textos, imagens, vídeos e sons, logo, estão interligados por meio de links, um mecanismo que direciona o usuário para outros hipertextos, também conhecidos como páginas ou sistemas *web*.

Santos (2018) afirma que aplicações *web* disponibilizam ferramentas que permitem ao usuário do sistema executar tarefas antes complexas em um curto espaço de tempo. Certamente, essas aplicações *web* possibilitam que os usuários as realizem de forma mais eficiente e rápida, sugerindo que a evolução das tecnologias *web* facilitou o uso de sistemas, proporcionando maior agilidade e acessibilidade para os usuários.

Os *softwares web* não tinham credibilidade, pois eram acessíveis apenas em redes locais dentro das organizações (RICHARDSON; AMUNDSEN; RUBY, 2013). Assim, existe uma limitação inicial dos *softwares web*, que, em seus primórdios, eram desenvolvidos para funcionar apenas dentro de infraestruturas internas das organizações, fazendo com que essa restrição impossibilite que os usuários externos ou de diferentes localidades acessassem os sistemas, comprometendo a escalabilidade e a confiança em sua adoção para usos mais amplos, em contrapartida, na sociedade contemporânea atual essa limitação foi requisito para melhores avanços nos sistemas *web*, com melhores conexões com a internet e computadores mais acessíveis, viabilizou a utilização dos sistemas *web*.

2.2 Tomada de Decisão Orientada a Dados

Tomada de decisão orientada por dados refere-se à prática de basear as decisões na análise de dados, em vez de confiar apenas na intuição. Por exemplo, uma empresa de varejo pode usar a análise de comportamento de compra dos clientes para decidir quais produtos promover, em vez de depender exclusivamente da experiência ou de palpites sobre o que pode vender melhor.

Para Patrocínio, Wesley S. (2017), atualmente existem dois cenários macro de aplicações de tomadas de decisão orientada por dados, usando os princípios de *Data Science*: (1) as descobertas realizadas a partir de dados e (2) decisões repetitivas e em grande escala. Logo, ambos os princípios se beneficiam de algoritmos avançados e análises preditivas, seja para identificar novas oportunidades e *insights* estratégicos, ou para otimizar processos e operações com base em decisões automatizadas e consistentes.

2.2.1 Algoritmo *k-means*

O *k-means* é um algoritmo de clusterização que pertence aos métodos particionais, se caracterizando devido o uso do erro quadrático como critério de mensuração da convergência das partições em relação aos *clusters*. O algoritmo realoca iterativamente os elementos com base na menor distância. Além disso, o *k-means* forma grupos de dados compactos internamente e bem distintos entre si, assim, o Algoritmo 1 detalha o funcionamento de todo o processo de clusterização (JAIN; MURTY; FLYNN, 1999).

Algoritmo 1: Algoritmo k-means

Entrada: Dados, k

Saída: Agrupamentos

```
1 início
2   Escolha k centros de cluster aleatoriamente ;
3   para cada ponto em Dados faça
4     Calcule a distância euclidiana do ponto para cada centro de cluster;
5     Atribua o ponto ao centro de cluster mais próximo;
6   fim
7   Atualize a posição dos centros de cluster, calculando a média aritmética de todos os
   pontos em cada cluster;
8   repita
9     Recalcule os centros de cluster usando as associações de cluster atuais;
10  até que o critério de convergência for alcançado;
11 fim
```

Amplamente utilizado devido à sua simplicidade e eficiência computacional, o que o torna adequado para grandes conjuntos de dados. Suas vantagens incluem a rápida convergência e a facilidade de interpretação dos resultados. No entanto, ele apresenta algumas desvantagens, como a sensibilidade a *outliers* e a necessidade de definir previamente o número de *clusters*, o que pode não ser trivial em alguns contextos (JAIN; MURTY; FLYNN, 1999).

2.2.2 *Extract, Transform e Load (ETL)*

De acordo com Mukherjee e Kar (2017), o processo de Extração, Transformação e Carregamento, ou *Extract, Transform e Load (ETL)*, é uma abordagem utilizada para extrair, transformar, limpar e carregar os dados em um local apropriado para consumo de outras aplicações e montar visualizações. Comumente é utilizada em mineração de dados, integração entre sistemas, ambientes que lidam com *Big Data* e migração de bases de dados. Assim, cada letra da sigla **ETL** representa uma etapa do processo:

1. **Extração (*Extract - E*)**: etapa responsável por recuperar os dados brutos de uma ou mais fontes de dados;
2. **Transformação (*Transform - T*)**: etapa responsável por limpar, tratar, aplicar métricas e converter os dados brutos extraídos na etapa anterior em um modelo de dados final;
3. **Carregar (*Load - L*)**: etapa responsável por carregar os dados finais gerados na etapa anterior em um banco de dados, *Data Warehouse* e ou *Data Lake*.

Conforme Kimball e Ross (2002), o processo **ETL** pode ser visto como um processo de construção de um *pipeline* de dados. Consequentemente, a extração de dados é a primeira etapa desse *pipeline* e envolve a coleta de informações de diversas fontes, como bancos de dados operacionais, arquivos e outras origens. A segunda etapa é a transformação de dados, que inclui a limpeza, padronização e conversão dos dados extraídos, preparando-os para análise no *Data Warehouse*. A terceira etapa é o carregamento de dados, que consiste na inserção dos dados transformados.

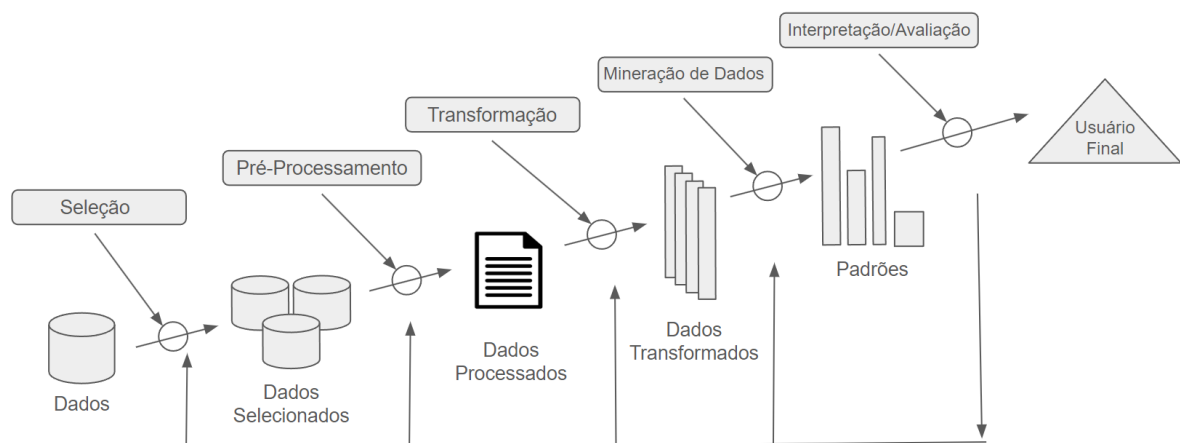
Em suma, os processos de **ETL** são responsáveis pela extração dos dados da fonte e seu transporte para uma área de *staging*, onde ocorrerão as transformações necessárias. Além disso, durante a transformação, são identificadas e isoladas as tuplas problemáticas, que podem causar quebras nas regras estruturais do destino. Logo, o processo é crucial para assegurar que organizações disponham de informações confiáveis e integradas. Analogamente, é indispensável para a tomada de decisões estratégicas e para alcançar uma vantagem competitiva.

2.2.3 *Knowledge-discovery in Databases (KDD)*

A descoberta de conhecimento em banco de dados, ou *Knowledge-discovery in Databases (KDD)*, é uma metodologia proposta por Fayyad, Piatetsky-Shapiro e Smyth (1996) com o objetivo de transformar dados em informações, na qual é composta por várias etapas interconectadas, podendo retornar a qualquer etapa anterior durante a avaliação dos resultados, caso o resultado não atenda o esperado no contexto do problema. A Figura 1 ilustra o fluxo de execução das seguintes etapas:

1. **Seleção dos dados:** etapa responsável pela seleção dos dados relevantes para a análise;
2. **Pré-processamento dos dados:** etapa responsável pelo pré-processamento dos dados brutos;
3. **Transformação dos dados:** etapa responsável por aplicar as técnicas de tratamento de dados necessária;
4. **Mineração de dados:** etapa responsável por aplicar algoritmos e técnicas de mineração de dados, como por exemplo, a clusterização utilizando o algoritmo *k-means*;
5. **Interpretação e avaliação dos dados:** etapa responsável pela análise dos resultados, a validação de sua significância estatística e a interpretação do seu impacto no contexto do problema.

Figura 1 – Etapas do KDD



Fonte: Adaptado de [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#).

O **KDD** é fundamental para transformar dados brutos em informações valiosas que podem orientar decisões estratégicas. Em um cenário onde a quantidade de dados gerados é crescente e complexa, consegue oferecer uma abordagem estruturada para descobrir padrões, tendências e *insights* ocultos que não são evidentes a partir de uma análise superficial.

Assim, o **KDD** possui capacidade de agregar valor aos dados, proporcionando uma compreensão mais profunda dos comportamentos e necessidades dos clientes, otimização de processos, e identificação de oportunidades de negócios. Através de técnicas avançadas de análise, o **KDD** permite que organizações antecipem mudanças de mercado, melhorem a eficiência operacional e criem estratégias mais eficazes. Em suma, no mundo orientado por dados, se torna um diferencial competitivo essencial, ajudando empresas a tomar decisões mais informadas e a adaptar suas estratégias em tempo real.

2.3 Clean Architecture

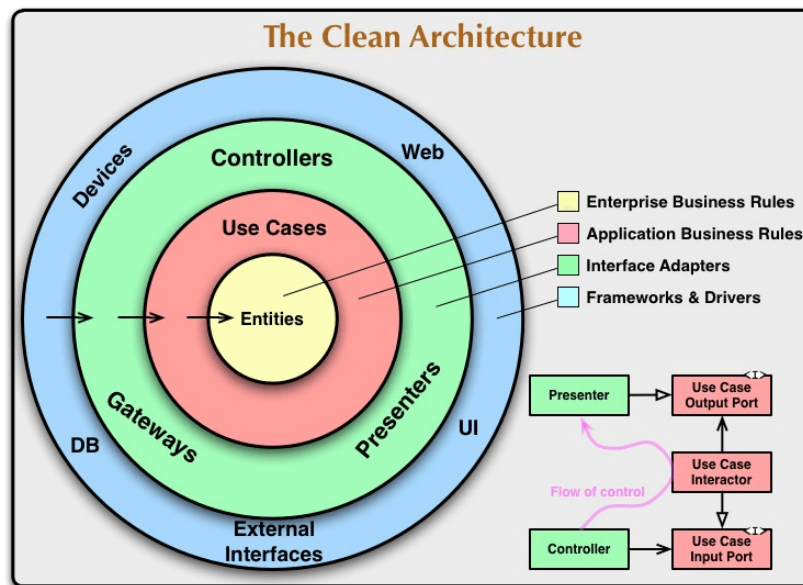
Segundo a definição dada pela *Institute of Electrical and Electronics Engineers* pode-se dizer que: “Arquitetura de *software* pode ser definida como a organização fundamental de um sistema, seus componentes, as relações entre eles e o ambiente que guia seu *design* e evolução”(IEEE, 2000). Seguindo o princípio da Arquitetura Limpa, podemos complementar essa definição ao enfatizar a importância da independência de detalhes de implementação, promovendo a separação clara entre as regras de negócio e os componentes de infraestrutura, tendo como objetivo garantir que o núcleo do sistema seja protegido de mudanças externas, permitindo que o *software* evolua de maneira sustentável, com baixo acoplamento e alta coesão entre os componentes, proporcionando maior flexibilidade e facilitando a manutenção e escalabilidade do sistema ao longo do tempo.

A Arquitetura Limpa, também conhecida como **CA**, é um padrão de arquitetura de software proposto por Martin (2017) em seu livro "*Clean Architecture: A Craftsman's Guide to Software Structure and Design*", afirma que este padrão de arquitetura visa criar sistemas que sejam altamente escaláveis, facilmente mantidos e que permitam a evolução contínua sem que isso comprometa a estrutura do software. As características defendidas pela **CA** são:

- **Independência de *framework*:** a arquitetura não depende de nenhuma ferramenta externa à linguagem;
- **Independência de interface de usuário:** a interface de usuário deve ser independente do restante do sistema, podendo ser alterada sem que afete as regras de negócio;
- **Independência de banco de dados:** as regras de negócio não podem ser impactadas pela troca de banco de dados, não podem estar vinculadas à tecnologias específicas de um banco de dados;
- **Independência de elementos externos:** as regras de negócio não devem estar vinculadas a elementos externos;
- **Testabilidade:** as regras de negócio podem ser testadas sem a necessidade de elementos externos.

A Figura 2 ilustra a divisão em camadas da **CA**. A camada mais interna, **Entities** ou camada de domínio, abriga as regras de negócios centrais e as entidades principais da aplicação. A camada **Application Business Rules** é responsável por coordenar os casos de uso e gerenciar o fluxo da aplicação. A camada **Interface Adapters** é responsável por disponibilizar uma interface de comunicação para a camada mais externa. Por fim, a camada mais externa, **Frameworks & Drivers**, abriga interfaces de usuário, *frameworks* e sistemas externos.

Figura 2 – Camadas da CA



Fonte: Martin (2017)

Portanto, com o objetivo de criar um sistema com componentes bem definidos, de fácil manutenção, testes simplificados e flexibilidade para troca de tecnologias, será adotada a CA no desenvolvimento, aplicando seus principais fundamentos.

2.4 Trabalhos Relacionados

Foram estudados trabalhos relacionados no ICEA, focados em dados educacionais de discentes. Estes estudos fornecem uma base sólida para entender como a análise de dados pode ser aplicada no contexto educacional, ajudando a identificar padrões e tendências no desempenho acadêmico dos alunos. A abordagem e os métodos utilizados nos trabalhos foram cruciais para aprimorar a compreensão sobre como as informações educacionais podem ser utilizadas para melhorar a gestão, os resultados acadêmicos e a acessibilidade dos usuários inseridos nesse contexto.

2.4.1 Desenvolvimento de um Dashboard para Análise e Visualização dos Dados Educacionais dos Discentes do ICEA/UFOP

Em seu trabalho, Paranhos (2021) visa extrair conhecimento a partir dos dados educacionais de um campus universitário, utilizando visualizações. A pesquisa é motivada pela vasta quantidade de dados que, quando adequadamente explorados, podem fornecer informações valiosas para a comunidade acadêmica.

O objetivo principal é empregar a análise e visualização desses dados para apoiar gestores nas tomadas de decisão dentro do campus. Para isso, utilizam-se a metodologia **KDD** e o software de *Business Intelligence Power BI* para análise e criação das visualizações, respectivamente. Como resultado, foram desenvolvidos sete dashboards interativos, abordando os seguintes temas: demografia, admissão, evasão, colação de grau, desempenho dos alunos, desempenho das disciplinas e desempenho dos departamentos.

No presente trabalho, a principal diferença em relação ao trabalho de **Paranhos (2021)** reside no fato de que, além de realizar análises detalhadas dos dados, as visualizações resultantes serão disponibilizadas por meio de um sistema *web* com utilização de tecnologias *open source*, proporcionando uma aplicação acessível e interativa para os usuários. Além disso, as análises de dados foram enriquecidas com a aplicação de técnicas de clusterização, que permitiram identificar padrões e agrupar informações de maneira mais eficiente, aprimorando a compreensão dos dados e contribuindo para uma tomada de decisão mais informada e gerando novos conhecimentos além do trabalho análogo.

2.4.2 Caracterização do Perfil de Alunos do ICEA: Uma Análise Guiada por Dados e Indicadores de Sucesso

Ferreira (2023) desenvolveu um trabalho que consiste em realizar uma caracterização do perfil dos alunos do **ICEA**, com base em uma análise orientada por dados e indicadores de sucesso. O objetivo é identificar os fatores que influenciam o desempenho acadêmico e a evasão dos estudantes, além de destacar as características comuns entre os diplomados, evadidos e matriculados de cada curso. Para isso, foram utilizados dados sociais e acadêmicos fornecidos pela unidade, referentes aos ingressantes entre 2011 e 2022.

Esses dados passaram por tratamento e análise por meio de técnicas estatísticas e de mineração de dados, como análise exploratória, redução de dimensionalidade e clusterização. Os resultados revelam diferenças significativas entre os grupos de alunos, identificados pela clusterização, tanto em variáveis relacionadas ao período pré-universitário quanto em indicadores acadêmicos. Além disso, foram identificados os principais fatores que contribuem para a evasão e o sucesso dos estudantes, resultando em sugestões de possíveis ações para melhorar a qualidade do ensino e da aprendizagem.

A principal diferença entre o presente trabalho e o estudo de **Ferreira (2023)** reside na abordagem adotada para a disponibilização das visualizações de dados e no foco das análises. Enquanto o estudo em questão se concentra em apresentar os dados de forma mais estática, este trabalho propõe o desenvolvimento de uma aplicação tecnológica que torna o acesso à informação mais ágil e intuitivo, permitindo uma maior interação por parte dos usuários.

3 Desenvolvimento

Este capítulo tem como objetivo descrever todo o desenvolvimento do projeto, desde a obtenção dos dados até o desenvolvimento do sistema. O presente trabalho usará a metodologia **KDD** nas etapas de **ETL**, utilizando conceitos da arquitetura de *software* **CA** durante o desenvolvimento do sistema, no qual foi dividido em três partes: *back-office*, *back-end* e *front-end*.

3.1 Obtenção e Preparação dos Dados

A subseção de Obtenção e Preparação dos Dados descreve o processo adotado para coletar e tratar as informações utilizadas neste estudo, informações sobre as fontes utilizadas e os dados no formato bruto, garantindo a integridade e relevância para a análise. Posteriormente, foi realizada a etapa de preparação, na qual os dados brutos passaram por processos de limpeza, sendo essa fase crucial para assegurar a qualidade dos dados.

3.1.1 Extração

A fase de extração dos dados define as demais etapas do projeto. Durante a extração, são selecionados os campos que, posteriormente, serão tratados e carregados no banco de dados para atender ao sistema. As informações foram disponibilizadas por meio de arquivos *Comma Separated Values (CSV)* pela seção de ensino do **ICEA**. A base de dados disponibilizada abrange todos os quatro cursos oferecidos pelo *campus*, do período de 2002 a 2023, sendo eles: Sistemas de Informação, Engenharia de Computação, Engenharia Elétrica e Engenharia de Produção.

As bases de dados foram disponibilizadas em seis planilhas, contendo os quatro cursos ofertados pelo *campus*. Esses arquivos incluem informações sobre os ingressantes dos cursos, notas e trancamentos. Assim, é necessário um dicionário de dados detalhado, contendo as informações pertinentes às planilhas utilizadas no projeto. Esse dicionário visa descrever os campos e os tipos de dados, proporcionando uma visão abrangente.

A planilha **DADOS INGRESSANTES.csv** apresenta informações detalhadas sobre os discentes que ingressaram na instituição. Esse conjunto de dados inclui, entre outros aspectos, dados pessoais, acadêmicos e informações relativas ao curso e período de entrada, permitindo a análise e acompanhamento do perfil e trajetória desses alunos, conforme a Tabela 1.

Foram utilizadas quatro planilhas, intituladas **NOTAS.csv**, correspondentes a cada curso. Essas planilhas contêm informações relevantes sobre o desempenho acadêmico dos discentes, como as notas obtidas em disciplinas, permitindo uma avaliação detalhada do progresso e da situação acadêmica dos estudantes ao longo de seus cursos, conforme a Tabela 2.

Tabela 1 – Dicionário de dados da planilha DADOS INGRESSANTES.csv

NOME DO CAMPO	TIPO	DESCRIÇÃO
MATRICULA	string	Matrícula do aluno ingressante
NOME	string	Nome do aluno ingressante
CIDADE NASCIMENTO	string	Cidade de nascimento do aluno ingressante
ESTADO NASCIMENTO	string	Estado de nascimento do aluno ingressante
PAIS NASCIMENTO	string	País de nascimento do aluno ingressante
ENDERECO FAMILIAR	string	Endereço familiar do aluno ingressante
BAIRRO FAMILIAR	string	Bairro familiar do aluno ingressante
CEP FAMILIAR	string	CEP familiar do aluno ingressante
CIDADE FAMILIAR	string	Cidade familiar do aluno ingressante
ESTADO FAMILIAR	string	Estado familiar do aluno ingressante
PAIS FAMILIAR	string	País familiar do aluno ingressante
ENDERECO ACADEMICO	string	Endereço acadêmico do aluno ingressante
BAIRRO ACADEMICO	string	Bairro acadêmico do aluno ingressante
CEP ACADEMICO	string	CEP acadêmico do aluno ingressante
CIDADE ACADÊMICA	string	Cidade acadêmica do aluno ingressante
ESTADO ACADEMICO	string	Estado acadêmico do aluno ingressante
PAIS ACADEMICO	string	País acadêmico do aluno ingressante
DATA NASCIMENTO	string	Data de nascimento do aluno ingressante
CPF	string	CPF do aluno ingressante
RG	string	RG do aluno ingressante
SEXO	string	Gênero do aluno ingressante
ANO ADMISSAO	number	Ano de entrada do aluno ingressante
SEMESTRE DE ADMISSAO	number	Semestre de entrada do aluno ingressante
COD MODO ADMISSAO	string	Código do modo de entrada do aluno ingressante
DESCRICA O MODO ADMISSAO	string	Descrição do modo de entrada do aluno ingressante
COD CURSO ADMISSAO	string	Código do curso escolhido pelo aluno ingressante
COD SITUACAO ALUNO	string	Código de situação do aluno ingressante
DESCRICA O SITUACAO ALUNO	string	Descrição da situação do aluno ingressante
CARGA HORARIA CURSADA	number	Carga horária que o aluno ingressante já cursou
CURSO	string	Curso escolhido pelo aluno ingressante
DATA INGRESSO	date	Data de entrada do aluno ingressante
MODALIDADE CONCORREN- CIA	string	Modalidade que o aluno ingressante optou
ANO DIPLOMACAO	number	Ano de diplomação do aluno ingressante
SEMESTRE DIPLOMACAO	number	Semestre de diplomação do aluno ingressante

Fonte: Elaborado pelo autor.

Tabela 2 – Dicionário de dados da planilha NOTAS.csv

NOME DO CAMPO	TIPO	DESCRIÇÃO
MATRICULA	string	Matrícula do aluno
ANO	number	Ano da disciplina cursada
SEMESTRE	number	Semestre da disciplina cursada
COD DEPARTAMENTO	string	Código do departamento da disciplina
COD DISCIPLINA	string	Código da disciplina cursada
DESCRICA0	string	Descrição da situação do aluno na disciplina
COD CURSO	string	Código do curso do aluno
CARATER	string	Disciplina obrigatória ou eletiva
MEDIA FINAL	number	Média final da disciplina
EXAME ESPECIAL	number	Nota do exame especial
FALTAS	number	Quantidade de faltas do aluno
SITUACAO	string	Situação do aluno na disciplina
PROFESSOR	string	Professor da disciplina
AULA DADA	string	Quantidade de aulas dadas
EMAIL INSTITUCIONAL	string	E-mail do aluno
DATA NASCIMENTO	date	Data de nascimento do aluno
GRAVACAO	date	Data de gravação do registro
TIPO ESCOLA	string	Sinaliza aluno oriundo de escola pública ou particular

Fonte: Elaborado pelo autor.

A planilha DADOS TRANCAMENTO.csv apresenta informações detalhadas sobre os trancamentos acadêmicos dos discentes no instituto. Esses dados incluem registros como o curso e o período de trancamento, conforme a mostra a Tabela 3.

Tabela 3 – Dicionário de dados da planilha DADOS TRANCAMENTO.csv

NOME DO CAMPO	TIPO	DESCRIÇÃO
MATRICULA	string	Matrícula do aluno
ANO	number	Ano do trancamento
SEMESTRE	number	Semestre do trancamento
COD CURSO	string	Código do curso do aluno
TIPO	string	Tipo do trancamento escolhido pelo aluno
COD DISCIPLINA	string	Código da disciplina trancada pelo aluno
DISCIPLINA	string	Disciplina trancada pelo aluno
DATA TRANCAMENTO	date	Data em que foi realizado o trancamento

Fonte: Elaborado pelo autor.

A Tabela 4 indica os campos extraídos durante a primeira fase do ETL. Campos que não agregam valor para as análises, ou campos que apenas armazenam dados sensíveis, como por exemplo documentos pessoais e endereços, foram removidos. Esse processo faz referência a primeira etapa de seleção do KDD.

Tabela 4 – Descrição dos campos extraídos

QUANTIDADE	PLANILHAS	CAMPOS EXTRAÍDOS
1	DADOS INGRESSANTES	MATRICULA, NOME, SEXO, ANO ADMISSAO, SEMESTRE DE ADMISSAO, COD MODO ADMISSAO, DESCRICAO MODO ADMISSAO, COD CURSO ADMISSAO, COD SITUAÇÃO ALUNO, DESCRICAO SITUAÇÃO ALUNO, CARGA HORARIA CURSADA, CURSO, DATA INGRESSO, MODALIDADE CONCORRENCIA, ANO DIPLOMACAO, SEMESTRE DIPLOMACAO
4	NOTAS SJM, NOTAS CJM, NOTAS EJM, NOTAS PJM	MATRICULA, ANO, SEMESTRE, COD DEPARTAMENTO, COD DISCIPLINA, DESCRICAO, COD CURSO, CARATER, MEDIA FINAL, EXAME ESPECIAL, FALTAS, SITUAÇÃO, PROFESSOR
1	DADOS TRANCAMENTO	MATRICULA, ANO, SEMESTRE, COD CURSO, TIPO, COD DISCIPLINA, DISCIPLINA, DATA TRANCAMENTO

Fonte: Elaborado pelo autor.

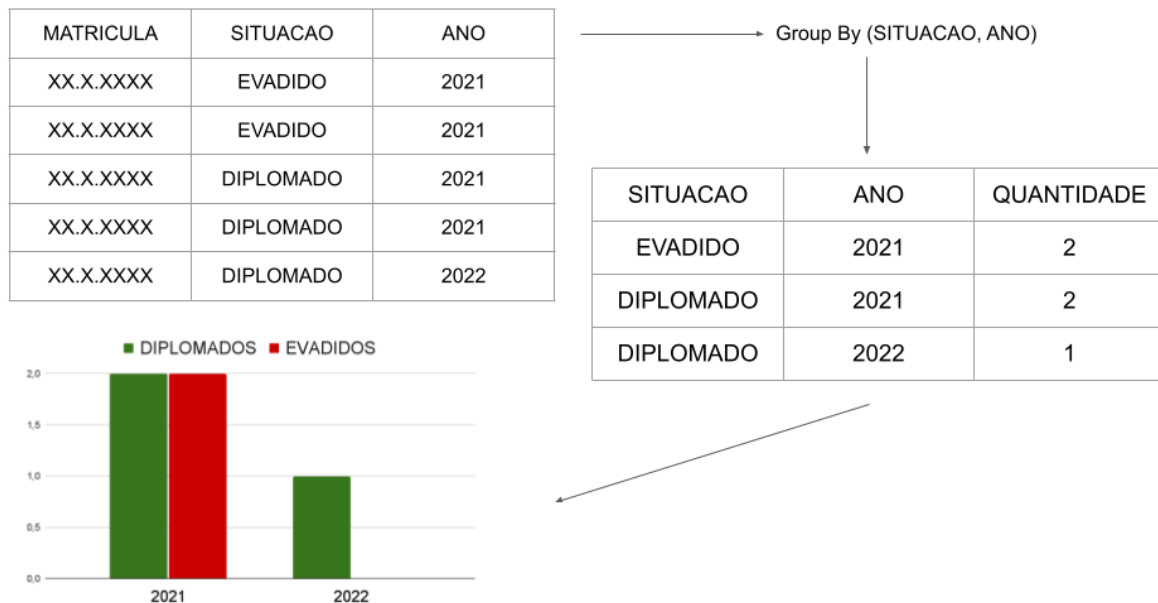
3.1.2 Transformação

A fase de transformação dos dados foi dividida em seis modalidades diferentes: Estatísticas de Graduações, Estatísticas de Disciplinas, Estatísticas de Admissões, Estatísticas de Evasões, Notas e Clusterização. Para cada modalidade, foram realizados o pré-processamento e a transformação dos dados, processos que correspondem às etapas subsequentes à seleção no KDD.

Para as modalidades: Estatísticas de Graduações, Estatísticas de Disciplinas, Estatísticas de Admissões e Estatísticas de Evasões, foram empregadas técnicas de agrupamento dos dados, que, posteriormente, foram convertidos em gráficos nas interfaces do sistema. A técnica de agrupamento permite organizar e sumarizar dados com base em uma ou mais variáveis (colunas). Para a modalidade de Notas, a etapa de transformação consistiu apenas na unificação dos dados dos quatro cursos em uma única base, sem a aplicação de tratamentos especiais.

A Figura 3 ilustra a aplicação da técnica de agrupamento durante a transformação dos dados, mostrando como a interface do sistema se comportará em relação aos dados resultantes. O agrupamento é realizado com base em um ou mais campos; ao inserir um campo para agrupamento, é gerada a quantidade de linhas agrupadas, permitindo uma visualização em gráficos.

Figura 3 – Técnica de agrupamento



Fonte: Elaborado pelo autor.

A modalidade de clusterização foi desenvolvida com base no estudo de [Ferreira \(2023\)](#), constituída em três etapas: Definição dos atributos acadêmicos, aplicação do [Análise de componentes principais \(PCA\)](#) e clusterização utilizando algoritmo *k-means*. A utilização do *k-means* está diretamente relacionada à etapa de mineração de dados do [KDD](#).

A primeira etapa consiste na definição dos atributos acadêmicos, com o objetivo de estabelecer medidas que permitem uma análise quantitativa do desempenho de cada aluno. Para garantir uma análise precisa, foi necessário definir e organizar os atributos relevantes. O [Quadro 3.1](#) exemplifica a estrutura dos atributos definidos para cada aluno, onde i representa aluno e j disciplina, conforme descrito a seguir:

- Quantidade cursada (**QC**): qc_{ij} representa a quantidade cursada da disciplina j pelo aluno i ;
- Nota final (**NF**): nf_{ij} representa a nota final da disciplina j pelo aluno i ;
- Aprovações (**AP**): ap_i representa a quantidade de disciplinas que o aluno i foi aprovado;

- Reprovado por nota (**RPN**): rpn_i representa a quantidade de disciplinas que o aluno i foi reprovado por nota;
- Reprovado por falta (**RPF**): rpf_i representa a quantidade de disciplinas que o aluno i foi reprovado por falta;
- Reprovado por nota e falta (**RNF**): $rnfi$ representa a quantidade de disciplinas que o aluno i foi reprovado por nota e falta;
- Trancamentos (**TRA**): tra_i representa a quantidade de disciplinas que o aluno i trancou;
- Cancelados (**CA**): ca_i representa a quantidade de disciplinas que o aluno i cancelou;
- Taxa de aprovação nas disciplinas do Departamento de Ciências Exatas e Aplicadas (Decea) (**APD**): apd_i representa a taxa de aprovação nas disciplinas do Decea do aluno i , visto que o departamento concentra a maior parte das disciplinas de início dos cursos;

$$APD = \frac{\text{disciplinas aprovadas no DECEA}}{\text{disciplinas cursadas no DECEA}}$$

- Semestres cursados (**CA**): sem_i representa a quantidade de semestres que o aluno i cursou;
- Distorção carga horária período (**CH/P**): ch/p_i representa a distorção carga horária período do aluno i .

$$CH/P = \frac{\text{carga horária completa}}{300 * \text{número de semestres cursado}}$$

Quadro 3.1 – Estrutura dos atributos acadêmicos

mat	QC_j	QC_{j+1}	...	NF_j	NF_{j+1}	...	AP	RPN	RPF	RNF	TRA	CA	APD	SEM	CH/P
...
mat_i	qc_{ij}	qc_{ij+1}	...	nf_{ij}	nf_{ij+1}	...	ap_i	rpn_i	rpf_i	$rnfi$	tra_i	ca_i	apd_i	sem_i	ch/p_i
...

Fonte: Adaptado de Ferreira (2023).

A segunda etapa da clusterização, após o pré-processamento e transformação dos dados, consiste na aplicação do PCA na estrutura de dados resultante da etapa anterior, método estatístico que reduz a dimensão dos dados mantendo a maior parte da informação. Ele transforma as variáveis originais em componentes principais, que são combinações lineares das variáveis originais e são ordenados pela variância (JOLLIFFE, 2002).

A terceira etapa consiste na aplicação da clusterização utilizando o algoritmo *k-means* com o resultado de 23 colunas após a aplicação do PCA. O número de *clusters* para cada curso foi definido com base nas diretrizes estabelecidas por Ferreira (2023). O resultado dos *clusters* serão apresentados no Capítulo 4 e a caracterização fica a critério do usuário durante a análise.

As Tabelas 5, 6, 7, 8, 9 apresentam a descrição e os atributos gerados durante a etapa de transformação para as modalidades Estatísticas de Graduações, Estatísticas de Disciplinas, Estatísticas de Admissões, Estatísticas de Evasões e Clusterização, respectivamente. Utilizando a técnica de agrupamento e clusterização com o algoritmo *k-means*. Vale ressaltar que, para a modalidade Notas, não foi aplicada nenhuma transformação.

Tabela 5 – Descrição dos atributos gerados para modalidade Estatísticas de Graduações

ATRIBUTO	DESCRIÇÃO
COEFICIENTE	Média ponderada das notas dos alunos graduados
TOTAL GRADUADOS	Total de alunos graduados

Fonte: Elaborado pelo autor.

Tabela 6 – Descrição dos atributos gerados para modalidade Estatísticas de Disciplinas

ATRIBUTO	DESCRIÇÃO
ESTUDANTES MATRICULADOS	Total de alunos matriculados
ESTUDANTES REPROVADOS	Total de alunos reprovados
ESTUDANTES APROVADOS	Total de alunos aprovados
ESTUDANTES TRANCADOS	Total de alunos trancados
ESTUDANTES CANCELADOS	Total de alunos cancelados
INDICES REPROVADOS	Índice de reprovação

Fonte: Elaborado pelo autor.

Tabela 7 – Descrição dos atributos gerados para modalidade Estatísticas de Admissões

ATRIBUTO	DESCRIÇÃO
TOTAL ADMISSAO	Total de admissões
TOTAL ADMISSAO GRADUADOS	Total de admissões que graduaram

Fonte: Elaborado pelo autor.

Tabela 8 – Descrição dos atributos gerados para modalidade Estatísticas de Evasões

ATRIBUTO	DESCRIÇÃO
TOTAL EVADIDOS	Total de evadidos

Fonte: Elaborado pelo autor.

Tabela 9 – Descrição dos atributos gerados para modalidade de Clusterização

ATRIBUTO	DESCRIÇÃO
CLUSTER	Cluster atribuído ao aluno durante a etapa de clusterização
TEMPO CURSADO	Número de semestres cursados pelo aluno

Fonte: Elaborado pelo autor.

3.1.3 Carregamento

Após a conclusão das fases de extração e transformação, os dados resultantes são carregados em um banco de dados com as estatísticas, notas e clusterização dos alunos. Cada modalidade é carregada em tabelas distintas no banco de dados, sem vínculo direto entre elas, conforme mostra a Tabela 10.

3.2 Construção dos Casos de Uso

Os diagramas de casos de uso são representações gráficas utilizando a linguagem *Unified Modeling Language (UML)*, com o objetivo de ilustrar as interações entre atores (que podem ser usuários ou sistemas externos) e o caso de uso em questão. Nesses diagramas, os atores são representados por pequenos bonecos em formato palito, enquanto os casos de uso são representados por elipses. Os atores do sistema estão listados a seguir:

- **Auxiliar Administrativo:** usuários responsáveis pela gestão e manutenção dos arquivos necessários para a geração das análises. Interagem diretamente com o *back-office*.
- **Discente:** usuários categorizados como estudantes e alunos. Possuem acesso a relatórios individuais sobre seu desempenho acadêmico, além de estatísticas gerais. Interagem diretamente com o *front-end*.
- **Docente:** usuários categorizados como professores, educadores e colegiado do curso. Possuem a capacidade de consultar relatórios individuais dos discentes, além de estatísticas gerais. Interagem diretamente com o *front-end*.

As relações entre atores e casos de uso são indicadas por setas que conectam um ator a um caso de uso, demonstrando que o ator participa ou realiza a atividade descrita. Além disso, o diagrama pode também mostrar relacionamentos entre casos de uso, utilizando setas para indicar que um estende o outro (VALENTE, 2020).

Para representar os diagramas de casos de uso foi utilizado a ferramenta *Draw.io*¹, os resultados estão listados no Apêndice A.

Tabela 10 – Descrição dos dados carregados

MODALIDADE	PLANILHAS NECESSÁRIAS	CAMPOS CARREGADOS
Estatísticas de Graduações	DADOS INGRESSANTES, DADOS TRANCAMENTO, NOTAS SJM, NOTAS CJM, NOTAS EJM, NOTAS PJM	CURSO, SEMESTRES, MODALIDADE ADMISSAO, COEFICIENTE, ANO GRADUACAO, TOTAL GRADUADOS
Estatísticas de Disciplinas	NOTAS SJM, NOTAS CJM, NOTAS EJM, NOTAS PJM	CODIGO DEPARTAMENTO, DESCRICAO, ANO, SEMESTRE, ESTUDANTES MATRICULADOS, ESTUDANTES REPROVADOS, ESTUDANTES APROVADOS, ESTUDANTE TRANCADOS, ESTUDANTES CANCELADOS, INDICES REPROVADOS
Estatísticas de Admissões	DADOS INGRESSANTES	ANO ADMISSAO, SEMESTRE ADMISSAO, MODALIDADE ADMISSAO, CURSO ADMISSAO, TOTAL ADMISSAO, TOTAL ADMISSAO GRADUADOS
Estatísticas de Evasões	DADOS INGRESSANTES	ANO ADMISSAO, SEMESTRE ADMISSAO, MODALIDADE ADMISSAO, CURSO ADMISSAO, TOTAL EVADIDOS
Notas	NOTAS SJM, NOTAS CJM, NOTAS EJM, NOTAS PJM	MATRICULA, CODIGO DISCIPLINA, CODIGO DEPARTAMENTO, DESCRICAO, ANO, SEMESTRE, RESULTADO, SITUACAO
Clusterização	DADOS INGRESSANTES, DADOS TRANCAMENTO, NOTAS SJM, NOTAS CJM, NOTAS EJM, NOTAS PJM	MATRICULA, CLUSTER, NOME, CURSO, SITUACAO, COR, SEXO, ANO ADMISSAO, SEMESTRE ADMISSAO, POLITICA AFIRMATIVA, ANO DIPLOMACAO, SEMESTRE DIPLOMACAO, TEMPO CURSADO

Fonte: Elaborado pelo autor.

3.3 Estrutura do Sistema

A subseção a seguir tem como objetivo detalhar a arquitetura e os componentes essenciais que compõem o sistema em desenvolvimento. Nesta parte, será abordada a organização das principais funcionalidades, os módulos que integram a solução, e a forma como eles se comunicam entre si para garantir o correto funcionamento do sistema.







3.3.1 Back-office

O *back-office* é responsável por todas as etapas do processo de ETL descritas nas seções: Extração (3.1.1), Transformação (3.1.2) e Carregamento (3.1.3). O usuário precisa somente anexar os arquivos de entrada em formato CSV. Após a entrada dos arquivos, o sistema realiza automaticamente todas as transformações necessárias e, em seguida, carrega os dados no banco de dados.

O *back-office* foi desenvolvido utilizando a linguagem de programação *Python*², amplamente reconhecida por sua rapidez no desenvolvimento, uma vasta gama de bibliotecas e *frameworks* e que possui alto desempenho para análises de dados e uma curva de aprendizado relativamente baixa (MUELLER; MASSARON, 2019). Oferece uma extensa coleção de bibliotecas voltadas para análise de dados, como *Pandas*³ para manipulação de dados, *NumPy*⁴ para cálculos numéricos avançados e *Scikit-learn*⁵ que oferece algoritmos e ferramentas para classificação, regressão, clusterização e pré-processamento de dados.

As principais tecnologias e bibliotecas utilizadas na construção do *back-office* estão descritas na Tabela 11.

Tabela 11 – Descrição das principais tecnologias utilizadas no *back-office*

Logo	Nome	Tipo	Descrição
	Python	Open Source	Linguagem de programação utilizada para o desenvolvimento.
	Pandas	Open Source	Biblioteca que disponibiliza ferramentas para análise e manipulação de dados.
	NumPy	Open Source	Biblioteca que disponibiliza ferramentas para cálculos numéricos avançados.
	Scikit-learn	Open Source	Biblioteca que disponibiliza ferramentas para classificação, regressão, clusterização e pré-processamento de dados.
	Psycopg ⁶	Open Source	Biblioteca que disponibiliza interface de conexão com o banco de dados.
	Tkinter ⁷	Open Source	Biblioteca que disponibiliza ferramentas para criação de interfaces de usuário.

Fonte: Elaborado pelo autor.

² Disponível em: <<https://www.python.org>>

³ Disponível em: <<https://pandas.pydata.org>>

⁴ Disponível em: <<https://numpy.org>>

⁵ Disponível em: <<https://scikit-learn.org>>

3.3.2 *Back-end*

O *back-end* é responsável por receber as requisições feitas pelo *front-end*, acessar e manipular o banco de dados, processar as informações e devolver os dados necessários para que a interface possa exibir gráficos e outras informações. Desenvolvido utilizando a linguagem de programação *TypeScript*⁸ e o *framework Express.js*⁹.

O *Express.js* auxilia na implementação da arquitetura de API *RESTful*, que oferece uma interface padronizada utilizando o protocolo *Hypertext Transfer Protocol (HTTP)*, facilitando a transferência de dados e a comunicação entre sistemas *web*, definindo regras para as requisições, que são conhecidas como métodos. Os métodos comumente usados, descritos a seguir, definem a finalidade da requisição (CLEMENTE, 2008):

- **GET**: Obtém uma representação do recurso.
- **DELETE**: Remove o recurso.
- **POST**: Cria um novo recurso com base na representação fornecida.
- **PUT**: Substitui completamente o estado do recurso com base na representação fornecida.
- **PATCH**: Atualiza parcialmente o recurso com base na representação parcial fornecida.

As principais tecnologias e bibliotecas utilizadas na construção do *back-end* estão descritas na Tabela 12.

⁵ Disponível em: <<https://www.drawio.com>>

⁶ Disponível em: <<https://www.psycopg.org>>

⁷ Disponível em: <<https://docs.python.org/pt-br/3/library/tkinter.html>>

⁸ Disponível em: <<https://www.typescriptlang.org>>

⁹ Disponível em: <<https://expressjs.com>>

¹⁰ Disponível em: <<https://nodejs.org>>







¹¹ Disponível em: <<https://www.prisma.io>>

¹² Disponível em: <<https://www.postgresql.org>>

¹³ Disponível em: <<https://jwt.io>>

¹⁴ Disponível em: <<https://www.npmjs.com/package/bcrypt>>

Tabela 12 – Descrição das principais tecnologias utilizadas no *back-end*

Logo	Nome	Tipo	Descrição
	TypeScript	<i>Open Source</i>	Linguagem de programação utilizada para o desenvolvimento.
	Node.js ¹⁰	<i>Open Source</i>	Ambiente de execução que permite criar servidores e sistemas web.
	Express.js	<i>Open Source</i>	Framework para o Node.js, mínimo e flexível que fornece um conjunto robusto de recursos para sistemas web.
	Prisma ORM ¹¹	<i>Open Source</i>	Biblioteca que simplifica a gestão de bancos de dados para consultas e migrações.
	Postgre SQL ¹²	<i>Open Source</i>	Banco de dados relacional que oferece alta performance e escalabilidade.
	Json Web Token ¹³	<i>Open Source</i>	Biblioteca que fornece um padrão seguro para transmissão de informações entre partes com tokens assinados digitalmente.
bcrypt	Bcrypt ¹⁴	<i>Open Source</i>	Biblioteca que gerencia hash seguro de senhas usando algoritmo de criptografia.

Fonte: Elaborado pelo autor.

3.3.3 *Front-end*

O *front-end* é a interface principal de interação entre o usuário e o sistema. Desenvolvido utilizando a linguagem de programação *TypeScript* e a biblioteca *React*¹⁵. A escolha dessas tecnologias é motivada por vários fatores: a facilidade de reutilização de componentes, o fato de serem *Open Source*, a presença de uma comunidade ativa e a ampla disponibilidade de bibliotecas com integração nativa. Além disso, oferecem suporte contínuo e ajudam a manter a conformidade com o conceito de *CA*, proporcionando flexibilidade para integração com outras bibliotecas e frameworks.


Um componente é uma unidade reutilizável e independente que define uma parte da interface do usuário, podendo operar isoladamente ou interagir com outros componentes. A componentização facilita a construção e manutenção de aplicações, pois erros não afetam os demais e alterações centralizadas se refletem em todas as instâncias.

¹⁵ Disponível em: <<https://react.dev>>

A performance da biblioteca está associada ao conceito de Virtual *Document Object Model* (DOM), as renderizações no DOM tradicional são caras e demoradas (AGGARWAL *et al.*, 2018). Portanto, quando um componente é renderizado, ele é armazenado de forma virtual em memória. Nas renderizações subsequentes, se não sofreu alterações, utiliza o componente previamente renderizado, atualizando apenas aqueles que sofreram modificações.

As principais tecnologias e bibliotecas utilizadas na construção do *front-end* estão descritas na Tabela 13.

Tabela 13 – Descrição das principais tecnologias utilizadas no *front-end*

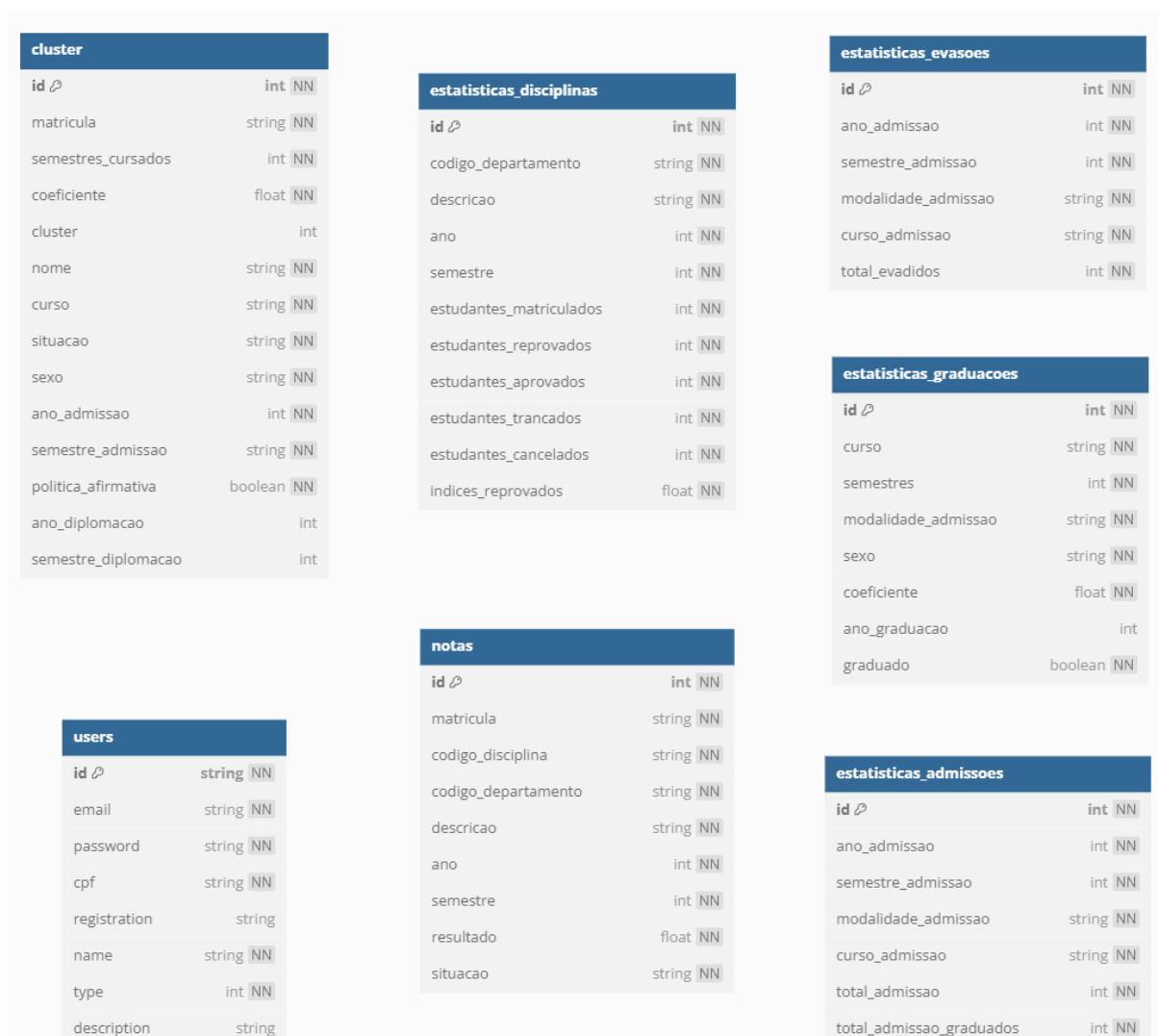
Logo	Nome	Tipo	Descrição
	TypeScript	<i>Open Source</i>	Linguagem de programação utilizada para o desenvolvimento.
	Node.js	<i>Open Source</i>	Ambiente de execução que permite criar servidores e sistemas web.
	React	<i>Open Source</i>	Biblioteca que permite criar interfaces de usuário interativas e reutilizáveis com uma abordagem baseada em componentes.
	Vite ¹⁶	<i>Open Source</i>	Ferramenta de construção que oferece um desenvolvimento rápido e otimizado para sistemas web.
	Material UI ¹⁷	<i>Open Source</i>	Biblioteca que fornece componentes de interface de usuário pré-estilizados e personalizáveis.
	Apex Charts ¹⁸	<i>Open Source</i>	Biblioteca que oferece componentes interativos e altamente configuráveis para visualização de dados em gráficos.
	Axios ¹⁹	<i>Open Source</i>	Biblioteca que simplifica a realização de requisições e o gerenciamento de respostas de rede.
	React Hook Form ²⁰	<i>Open Source</i>	Biblioteca que facilita a criação e validação de formulários.
	React Router Dom ²¹	<i>Open Source</i>	Biblioteca que gerencia o roteamento e navegação, permitindo a definição de rotas e a navegação entre páginas.
	React Flow ²²	<i>Open Source</i>	Biblioteca que permite criar e gerenciar diagramas e fluxos interativos com uma interface de arrastar e soltar.

Fonte: Elaborado pelo autor.

3.4 Modelagem do Banco de Dados

O banco de dados escolhido para armazenar os dados do sistema é do tipo relacional, o que normalmente sugere uma estrutura onde as tabelas estão interligadas por relações, chaves primárias e estrangeiras. No entanto, na modelagem atual, as tabelas não se relacionam entre si, sendo totalmente independentes uma da outra, trazendo maior flexibilidade para alterações individuais nas tabelas e abordagem sem interferência de outras tabelas. A Figura 4 mostra as tabelas finais.

Figura 4 – Modelagem do banco de dados



Fonte: Elaborado pelo autor no software Dbdiagram.io²³

16 Disponível em: <<https://vitejs.dev>>
 17 Disponível em: <<https://mui.com/material-ui>>
 18 Disponível em: <<https://apexcharts.com>>
 19 Disponível em: <<https://axios-http.com>>
 20 Disponível em: <<https://react-hook-form.com>>
 21 Disponível em: <<https://reactrouter.com>>
 22 Disponível em: <<https://reactflow.dev>>

3.5 Arquitetura do Sistema

A arquitetura do sistema, bem como a escolha das tecnologias e abordagens, foi projetada seguindo os conceitos da **CA**, com foco na reusabilidade e na independência de tecnologia, também a preferência por tecnologias *Open Source*, com comunidades ativas e engajadas em manter estável por um longo período. O *back-office* opera de forma independente do *back-end*, e o *front-end* é independente tanto do *back-end* quanto do *back-office*. Essa separação assegura a flexibilidade e facilita a manutenção e evolução do sistema.

Para a parte de *DevOps*, foram escolhidos o *Git*²⁴ para controle e versionamento do código, e a plataforma de hospedagem de código *GitHub*²⁵. Uma das principais vantagens dessa escolha é a integração nativa e simplificada com o *GitHub Actions*, uma ferramenta que se integra diretamente ao repositório remoto e facilita a implementação de **Integração Contínua (CI)** e **Entrega Contínua (CD)**.

O *back-end*, *front-end* e banco de dados são encapsulados em *containers Docker*²⁶. O uso do *Docker* permite a criação de ambientes isolados e reproduzíveis, o que simplifica o desenvolvimento, a implantação e a escalabilidade das aplicações. Cada componente do sistema é executado em seu próprio container, garantindo consistência entre diferentes ambientes e facilitando a gestão de dependências e configurações.

Além disso, o servidor *web Nginx*²⁷ é utilizado para otimizar o gerenciamento de tráfego e realizar o balanceamento de carga. Ele atua como um intermediário na comunicação entre o *back-end* e o *front-end*, recebendo requisições de forma segura via HTTPS e encaminhando-as para o *back-end* por meio de HTTP.

As Figuras 5 e 6 demonstram a arquitetura do sistema em alto nível. As áreas destacadas em cinza representam componentes fora dos *containers Docker*, enquanto as partes em azul indicam os elementos que estão conteinizados, e as seções em vermelho estão relacionadas às práticas de *DevOps* na arquitetura.

A Figura 5 demonstra as etapas do processo de **ETL** da arquitetura, o *back-office* é responsável por receber a entrada dos arquivos em formato **CSV** pelo auxiliar administrativo, transformar os dados e posteriormente carregar os resultados em um banco de dados relacional *PostgreSQL*.

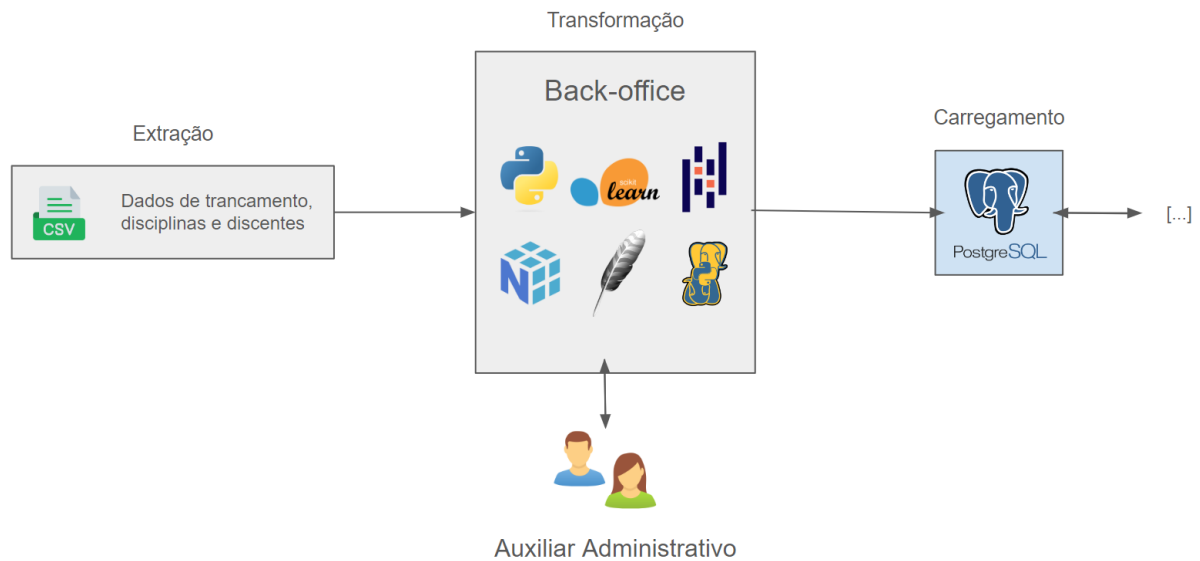
²⁴ Disponível em: <https://git-scm.com>

²⁵ Disponível em: <https://github.com>

²⁶ Disponível em: <https://www.docker.com>

²⁷ Disponível em: <https://nginx.org>

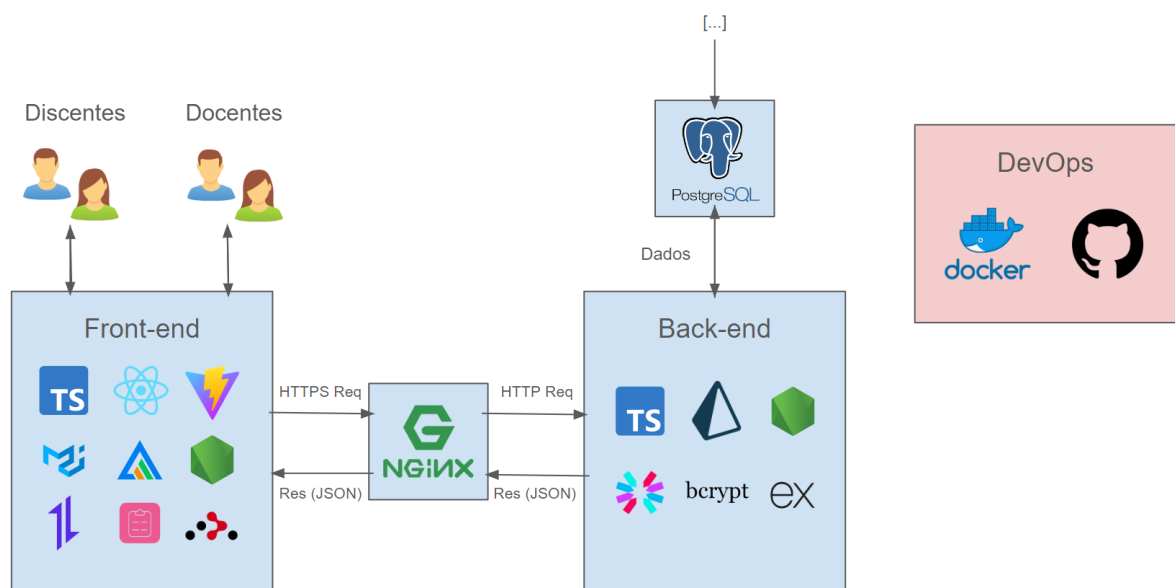
Figura 5 – Arquitetura do Sistema em Alto Nível - *back-office*



Fonte: Elaborado pelo autor.

A Figura 6 demonstra como ocorre a comunicação entre o *back-end* e o *front-end* do sistema, uma vez que os dados já estão disponíveis no banco de dados. O *front-end* atua como a interface que solicita ao *back-end* as informações requeridas pelos discentes e docentes. Essas requisições passam pelo servidor *web Nginx* antes de serem encaminhadas ao *back-end*, e as respostas, no formato *JavaScript Object Notation (JSON)*, são enviadas de volta para o *front-end*.

Figura 6 – Arquitetura do Sistema em Alto Nível - *back-end* e *front-end*



Fonte: Elaborado pelo autor.

3.6 Avaliação e Inspeção de Usabilidade do Sistema

Após a conclusão e implementação das principais funcionalidades do sistema, foi conduzida uma avaliação em um ambiente controlado, envolvendo usuários com as permissões apropriadas para testes. A avaliação consistiu em três etapas: preparação, execução da avaliação e elaboração do relatório consolidado.

Durante a etapa de preparação, o desenvolvedor (autor do presente trabalho) e os avaliadores (discentes do [ICEA](#)) se reuniram para preparar o ambiente e viabilizar a execução da inspeção. Foi decidido que os avaliadores conduziram a inspeção em duplas, totalizando cinco, com cada dupla utilizando um *notebook* com as especificações técnicas necessárias para executar o sistema. O sistema foi hospedado na máquina do desenvolvedor e disponibilizado na rede *wi-fi Eduroam* do [ICEA](#), com um endereço de acesso fornecido para os avaliadores. Foram criados usuários e senhas exclusivos para cada dupla, na categoria discente.

Após a etapa de preparação, iniciou-se a etapa de execução da avaliação, em que cada dupla realizou a inspeção heurística guiada pelas 10 heurísticas de [Nielsen e Molich \(1990\)](#). Todos os problemas de usabilidade identificados foram documentados em uma planilha, que incluiu a descrição do problema, local, criticidade, heurística associada e recomendação de solução. As heurísticas estão listadas a seguir:

1. **Visibilidade do status do sistema:** O sistema deve informar o usuário sobre o que está acontecendo através de *feedback* apropriado e em tempo hábil.
2. **Compatibilidade do sistema com o mundo real:** O sistema deve usar linguagem e conceitos familiares ao usuário, seguindo convenções do mundo real para facilitar a compreensão.
3. **Controle do usuário e liberdade:** O usuário deve ter a capacidade de desfazer e refazer ações facilmente, mantendo o controle sobre o sistema.
4. **Consistência e padrões:** Situações e ações semelhantes devem significar conceitos ou operações semelhantes.
5. **Prevenção de erros:** O sistema deve ser projetado para evitar a ocorrência de erros, ou seja, prevenir ações que possam levar a erros.
6. **Reconhecimento ao invés de lembrança:** O sistema deve minimizar a carga cognitiva do usuário, tornando as opções visíveis e não exigindo memorização.
7. **Flexibilidade e eficiência de uso:** O sistema deve ser flexível, permitindo atalhos para usuários experientes e oferecendo opções de personalização para aumentar a eficiência.
8. **Estética e design minimalista:** O *design* deve ser simples e direto, sem informações irrelevantes ou excessivas que possam distrair o usuário.

9. **Ajudar os usuários a reconhecer e corrigir erros:** As mensagens de erro devem ser claras e fornecer informações suficientes para que o usuário entenda e corrija o problema.
10. **Ajuda e documentação:** O sistema deve oferecer suporte acessível e útil, com documentação clara e direcionada para auxiliar o usuário quando necessário.

A terceira e última etapa trata da elaboração do relatório consolidado, promovendo a discussão entre os avaliadores de forma a identificar padrões recorrentes e elaborar um relatório unificado. Esse relatório consolidado foi elaborado de forma consensual, resumindo os problemas encontrados, conforme mostra a Tabela 14, onde temos o local do problema, a descrição do problema, a heurística utilizada, o nível de criticidade e a quantidade de duplas que identificaram o problema.

Tabela 14 – Resumo dos problemas encontrados com relatório consolidado

Local do Problema	Descrição do Problema	Heurística	Criticidade	Quantidade Avaliadores (Duplas)
Página inicial	Falta de descrição das funcionalidades	Reconhecimento ao invés de lembrança	Média	4
Página inicial	Algumas imagens não são compatíveis com a funcionalidade	Compatibilidade do sistema com o mundo real	Baixa	2
Página inicial	Logo não direciona para a página inicial	Compatibilidade do sistema com o mundo real	Baixa	2
Clusterização -> Gráficos	Falta legenda descritiva no gráfico de clusterização	Ajuda e Documentação	Baixa	5
Clusterização -> Filtros	Na aba filtros, apenas o valor indefinido é apresentado	Compatibilidade do sistema com o mundo real	Alta	1
Admissões -> Filtros	Não existe a opção de desmarcar / marcar todos	Flexibilidade e eficiência de uso	Média	2
Admissões -> Filtros	Filtro de agrupamento não foi compreendido	Ajuda e Documentação	Média	3
Graduações -> Geral	Não foi possível acessar a página	Prevenção de erros	Alta	3
Disciplinas -> Busca	É necessário digitar o nome completo da disciplina. Ausência de auto-complete	Prevenção de erros	Alta	1
Disciplinas -> Filtro	O agrupamento pelos anos não funciona	Compatibilidade do sistema com o mundo real	Alta	1
Admissões -> Ordenação	Não é possível escolher o critério de ordenação	Compatibilidade do sistema com o mundo real	Baixa	1
Grafos de Dependências	A funcionalidade apenas exibe a grade de um curso	Consistência e padrões	Alta	1
Referência da API	Sumário não leva para a informação esperada e funcionalidade não explicada	Consistência e padrões	Baixa	2
Rerelatório do Aluno	Ausência de título	Consistência e padrões	Baixa	1

Fonte: Elaborado pelo autor.

4 Resultados

Este capítulo tem como objetivo apresentar o resultado final de todo o desenvolvimento do sistema *web*. Logo, serão apresentados capturas de telas das principais funcionalidades presentes, bem como uma breve análise e síntese dos resultados apresentados.

4.1 Ajustes após Avaliação e Inspeção do Sistema

Com base na avaliação e inspeção de usabilidade pelos usuários do sistema detalhados na Seção 3.6, as ações tomadas para a versão final do sistema foram documentadas na Tabela 15, cobrindo a solução de todos os problemas encontrados.

Tabela 15 – Resumo dos problemas encontrados pelos usuários e ações tomadas

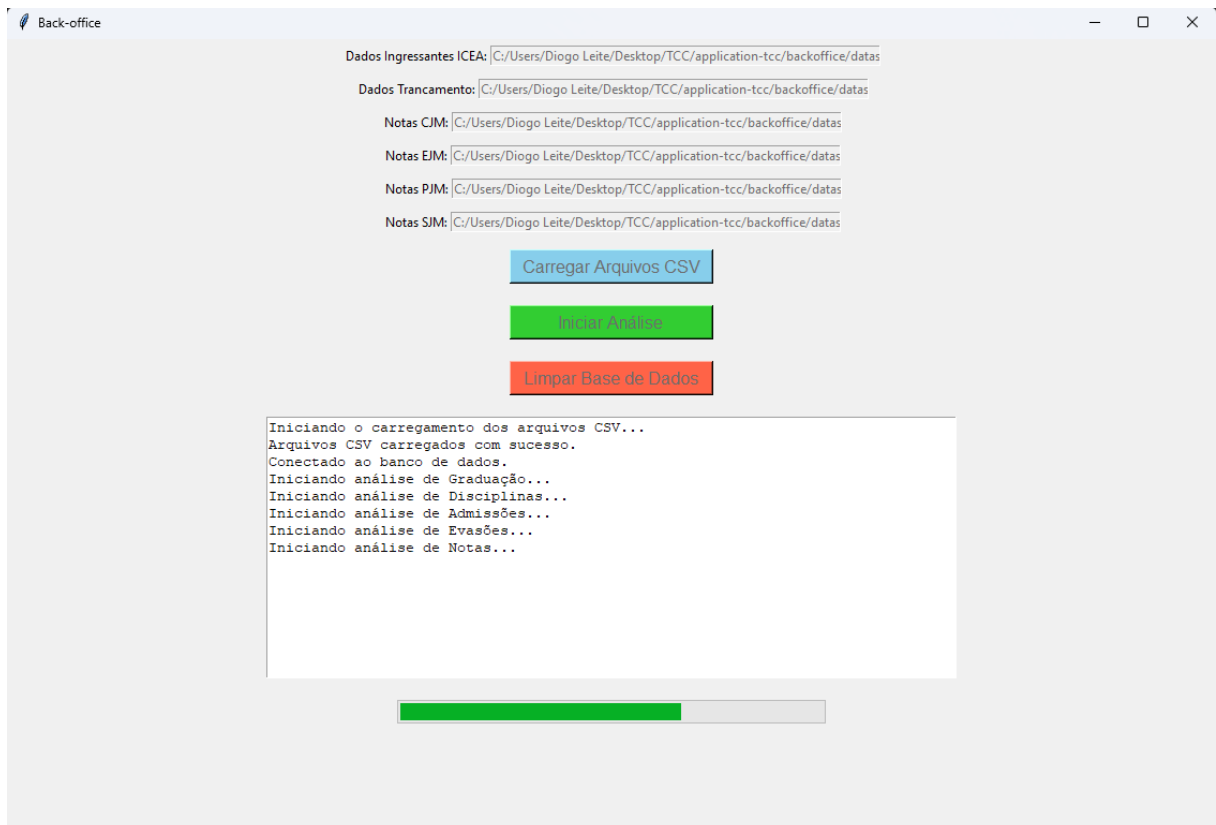
Local do Problema	Descrição do Problema	Ação
Página inicial	Falta de descrição das funcionalidades	Foram adicionados descrições em todas as funcionalidades, bem como a padronização da escrita.
Página inicial	Algumas imagens não são compatíveis com a funcionalidade	As imagens foram removidas para a padronização, pois nem todas as funções necessitam de ilustração.
Página inicial	Logo não direciona para a página inicial	Adicionado interação no logo, no qual encaminha o usuário para a página inicial.
Clusterização -> Gráficos	Falta legenda descritiva no gráfico de clusterização	Foram adicionadas descrições e legendas em todos os gráficos presente na clusterização.
Clusterização -> Filtros	Na aba filtros, apenas o valor indefinido é apresentado	Filtro indefinido removido e substituído por outro filtro válido.
Admissões -> Filtros	Não existe a opção de desmarcar / marcar todos	Os filtros vem desmarcado por padrão.
Admissões -> Filtros	Filtro de agrupamento não foi compreendido	Uma funcionalidade de "Ajuda" do filtro foi adicionado ao lado dos agrupamentos.
Graduações -> Geral	Não foi possível acessar a página	A página foi consertada e repaginada.
Disciplinas -> Busca	É necessário digitar o nome completo da disciplina. Ausência de auto-complete	A função de busca foi alterada para aceitar parte da descrição, não sendo mais necessário digitar o nome completo da disciplina.
Disciplinas -> Filtro	O agrupamento pelos anos não funciona	O agrupamento por ano foi corrigido.
Admissões -> Ordenação	Não é possível escolher o critério de ordenação	A função foi removida, pois estava depreciada.
Grafos de Dependências	A funcionalidade apenas exibe a grade de um curso	Todos os demais cursos disponíveis no campus foram adicionados.
Referência da API	Sumário não leva para a informação esperada e funcionalidade não explicada	O sumário foi consertado e todas as funcionalidades foram adicionadas guias de utilização.
Rerelatório do Aluno	Ausência de título	O título foi adicionado.

Fonte: Elaborado pelo autor.

4.2 Interface do *Back-office*

A Figura 7 apresenta a interface do *back-office*, na qual o usuário pode selecionar os arquivos a serem analisados, iniciar o processo de análise e, quando necessário, limpar a base de dados. Além disso, a interface inclui em seus componentes visuais uma barra de progresso que indica o estágio atual da ferramenta, além de fornecer registros de *logs*, permitindo a identificação de erros e o acompanhamento do *status* da análise de forma detalhada e precisa.

Figura 7 – Captura de tela do sistema Back-office



Fonte: Elaborado pelo autor.

4.3 Interface Login

A Figura 8 apresenta uma tela de *login*, na qual o usuário preenche o formulário com "cpf" e "senha" cadastrados, sendo possível guardar as credenciais de autenticação selecionando o botão "manter conectado". Após o preenchimento dos campos, basta selecionar a opção "entrar".

Caso as credenciais de acesso estejam corretas, uma mensagem de sucesso irá aparecer no canto inferior direito, como ilustrado na Figura 9, caso contrário, uma mensagem de erro irá aparecer na mesma posição, conforme a Figura 10.

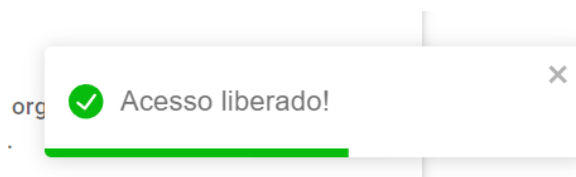
Figura 8 – Captura de tela da página de login



Logo do ICEA - Instituto de Ciências Exatas e Aplicadas. Título: ICEA - UFOP. Campos: CPF*, Senha* (com ícone de olho para alternar visibilidade). Opção: Manter conectado. Botão: ENTRAR.

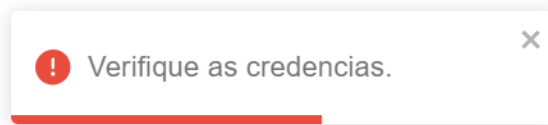
Fonte: Elaborado pelo autor.

Figura 9 – Captura de tela do sucesso de login



Fonte: Elaborado pelo autor.

Figura 10 – Captura de tela do erro de login



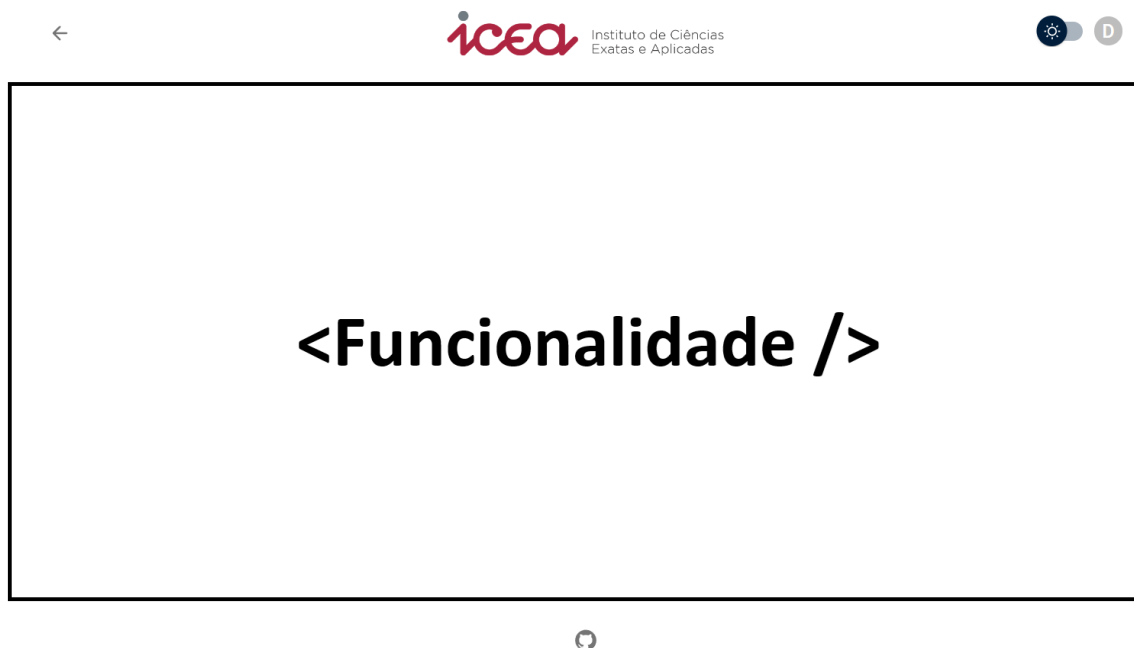
Fonte: Elaborado pelo autor.

4.4 Layout da Aplicação

Toda a aplicação é envolvida por um *container*, logo, a Figura 11 exemplifica como o conteúdo de cada funcionalidade é exibido e as funções padronizadas em todas as telas após a etapa de *login*. No canto superior direito se encontra a opção que ativa e desativa o modo noturno e o perfil, demonstrados conforme a Figura 12 no qual se encontra os botões para acessar "**minha conta**", "**usuários**" (disponível apenas para administradores) e "**sair**" para encerrar a seção.

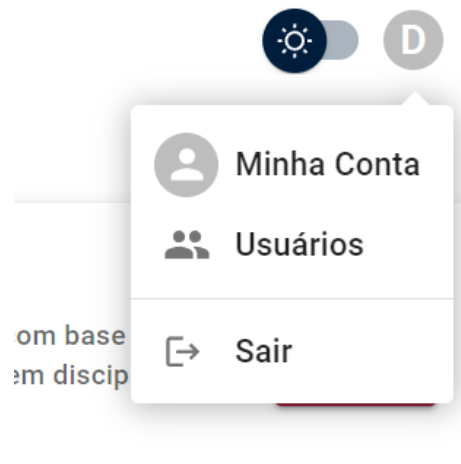
No canto superior esquerdo, se encontra a opção "**voltar**" ilustrada e representada por um ícone de seta, no qual retorna para a página anterior da navegação, e na região central o logo do [ICEA](#).

Figura 11 – Container da aplicação



Fonte: Elaborado pelo autor.

Figura 12 – Botão perfil



Fonte: Elaborado pelo autor.

4.5 Página Inicial

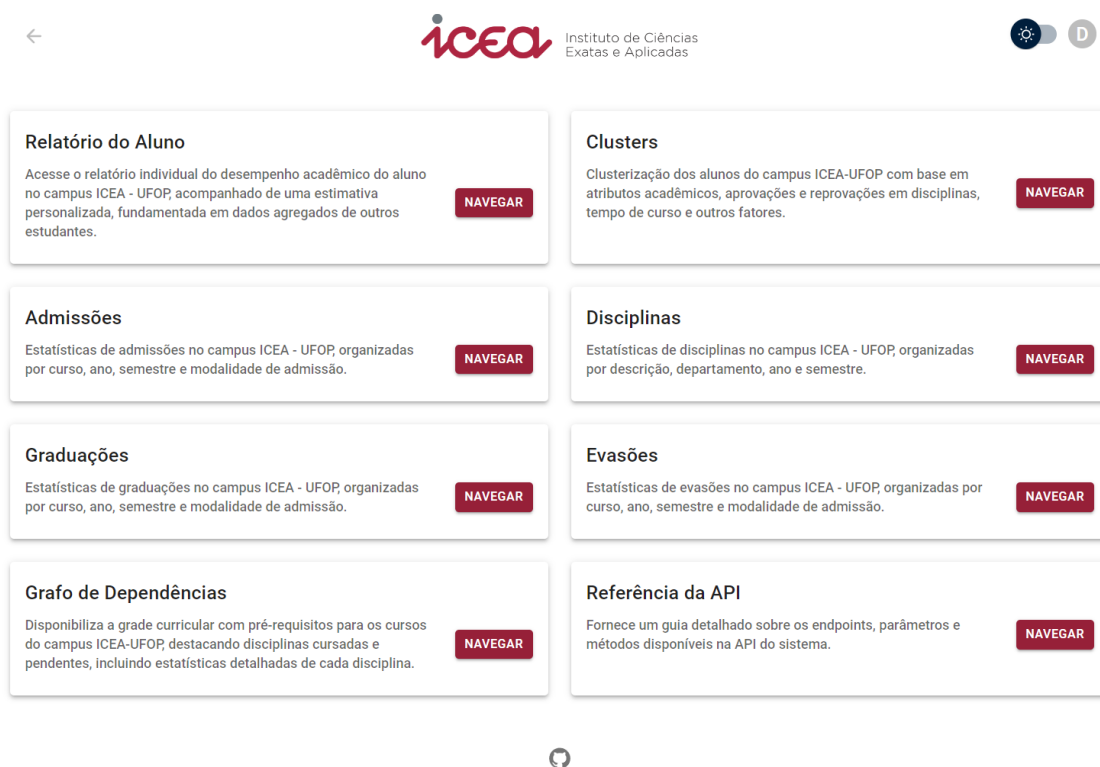
A primeira página na qual o usuário é redirecionado após o *login* é a página inicial, demonstrado conforme a Figura 13, em que estão listadas todas as funcionalidades do sistema.

Cada funcionalidade tem o seu título em evidência, uma breve descrição da funcionalidade e um botão "**navegar**", na qual direciona para a página de interesse.

4.5.1 Meu Perfil

A Figura 14 apresenta a página de meu perfil, na qual consiste em apresentar ao usuário como visualizar suas informações como: id, e-mail, nome, cpf e descrição. Ademais, é evidenciado o seu nível de acesso na parte inferior da interface.

Figura 13 – Captura de tela da página inicial



Fonte: Elaborado pelo autor.

4.6 Usuários

A Figura 15 demonstra a página de usuários, responsável por todo o controle de usuários do sistema. Analogamente, todos os usuários presentes do sistema são exibidos em uma tabela, na qual contém os nomes, e-mail, tipo de acesso, descrição e ações como editar e excluir, preenchendo a caixa de texto "**pesquisar**" para a tabela ser filtrada.

Figura 14 – Captura de tela da página meu perfil

← **ICEA** Instituto de Ciências Exatas e Aplicadas

Perfil

ID
934a7fb4-419a-4eff-a722-7ee34bd6b5d2

E-mail
diogo.lucas@aluno.ufop.edu.br

Nome
Diogo Leite Lucas

Matrícula
20.2.8072

CPF
111.111.111-11

Descrição
<https://github.com/diogoleite87/>

Usuário Administrador
Como usuário administrador, você possui privilégios especiais que permitem adicionar, excluir e editar usuários comuns. Isso significa que você tem controle total sobre a gestão de usuários na plataforma. Sinta-se à vontade para realizar operações como adicionar novos usuários, remover usuários existentes e editar as informações dos usuários conforme necessário para manter a integridade do sistema.

Fonte: Elaborado pelo autor.

4.6.1 Cadastrar Usuário

Selecionando a opção "**cadastrar**" na tela de usuários, um *modal* é aberto para o cadastro de um novo usuário, conforme apresentado na Figura 16. Assim, os campos "**nome**", "**e-mail**", "**matrícula**", "**cpf**", "**senha**" e "**descrição**" estão disponíveis para preenchimento, após o processo dessa funcionalidade, basta selecionar a opção "**cadastrar**".

Se o cadastro for efetuado, uma mensagem de sucesso irá ser exibida no canto inferior direito conforme a Figura 17, caso contrário, uma mensagem de erro será exibida na mesma posição, conforme a Figura 18.

4.6.2 Gerenciar Acesso do Usuário

Selecionando a opção "**editar**" na tela de usuários, um *modal* é aberto para a edição do nível de acesso do usuário, conforme mostra a Figura 19. Após escolher o acesso que deseja atribuir ao usuário, basta selecionar a opção "**atualizar**".

Figura 15 – Captura de tela da página de usuários

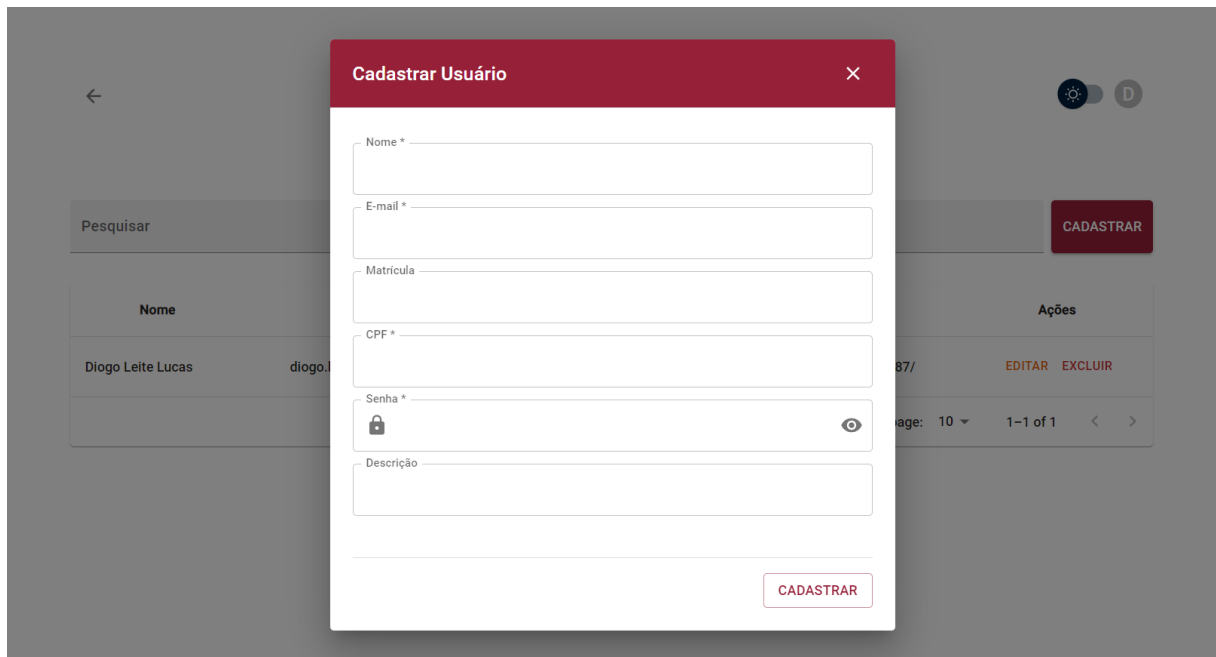
Nome	E-mail	Tipo	Descrição	Ações
Diogo Leite Lucas	diogo.lucas@aluno.ufop.edu.br	Administrador	https://github.com/diogoleite87/	EDITAR EXCLUIR
Usuário Teste	teste@gmail.com	Discente	Usuário de Teste	EDITAR EXCLUIR
Usuário Teste	teste1@gmail.com	Docente	Usuario de Teste	EDITAR EXCLUIR

Rows per page: 10 1-3 of 3 < >

Fonte: Elaborado pelo autor.

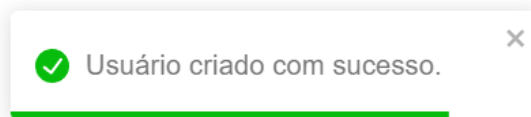
Se a atualização for efetuada, uma mensagem de sucesso irá ser exibida no canto inferior direito conforme a Figura 20, caso contrário, uma mensagem de erro será exibida na mesma posição, conforme a Figura 18.

Figura 16 – Captura de tela da página de usuários



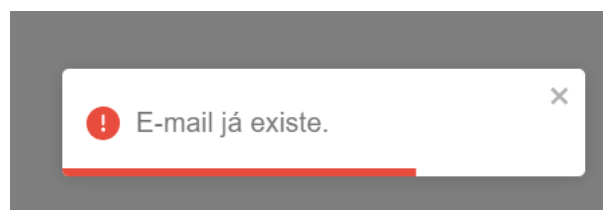
Fonte: Elaborado pelo autor.

Figura 17 – Captura de tela do sucesso em cadastrar usuário



Fonte: Elaborado pelo autor.

Figura 18 – Captura de tela do erro em cadastrar usuário

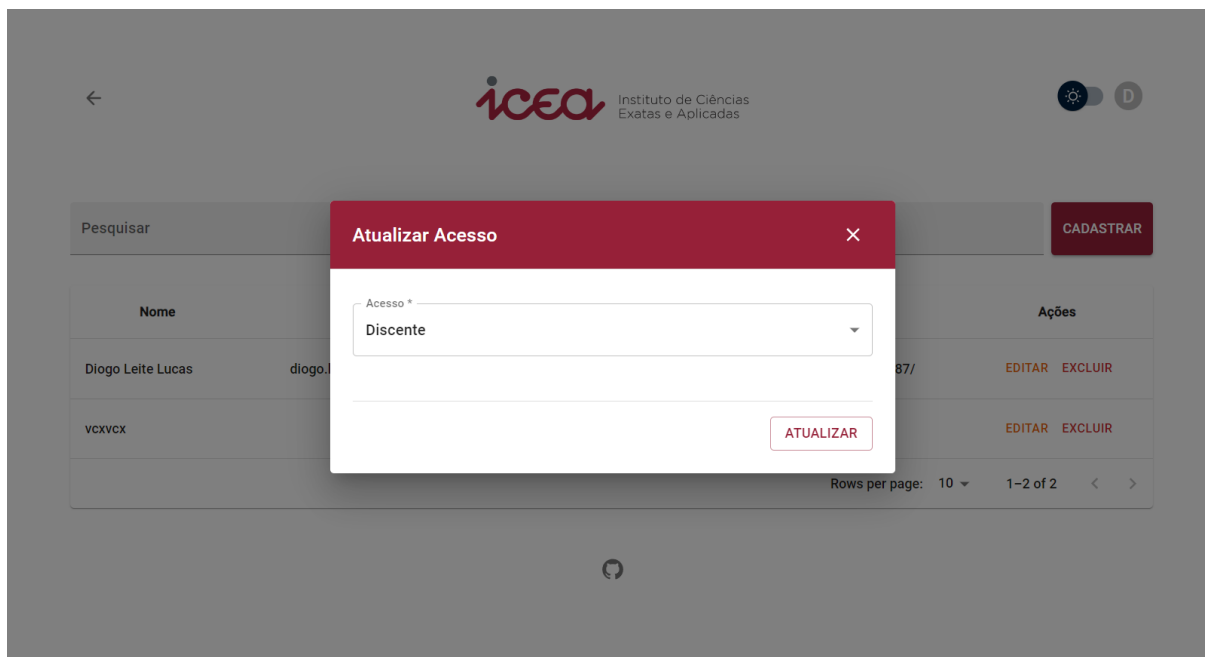


Fonte: Elaborado pelo autor.

4.6.3 Excluir Usuário

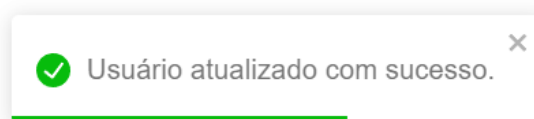
Selecionando a opção "**excluir**" na tela de usuários, um *modal* é aberto para a confirmação de exclusão do usuário, conforme mostra a Figura 22. Ao selecionar a opção "**excluir**" e a exclusão for efetuada, uma mensagem de sucesso irá ser exibida no canto inferior direito conforme a Figura 23, caso contrário, uma mensagem de erro será exibida na mesma posição, conforme a Figura 24.

Figura 19 – Captura de tela do gerenciamento de acesso do usuário



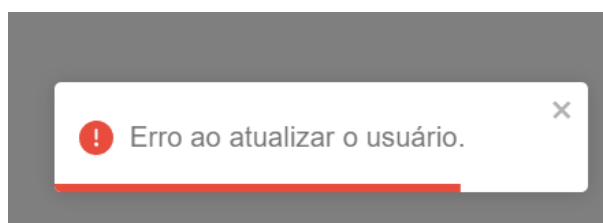
Fonte: Elaborado pelo autor.

Figura 20 – Captura de tela do sucesso em atualizar o acesso do usuário



Fonte: Elaborado pelo autor.

Figura 21 – Captura de tela do erro em atualizar o acesso do usuário

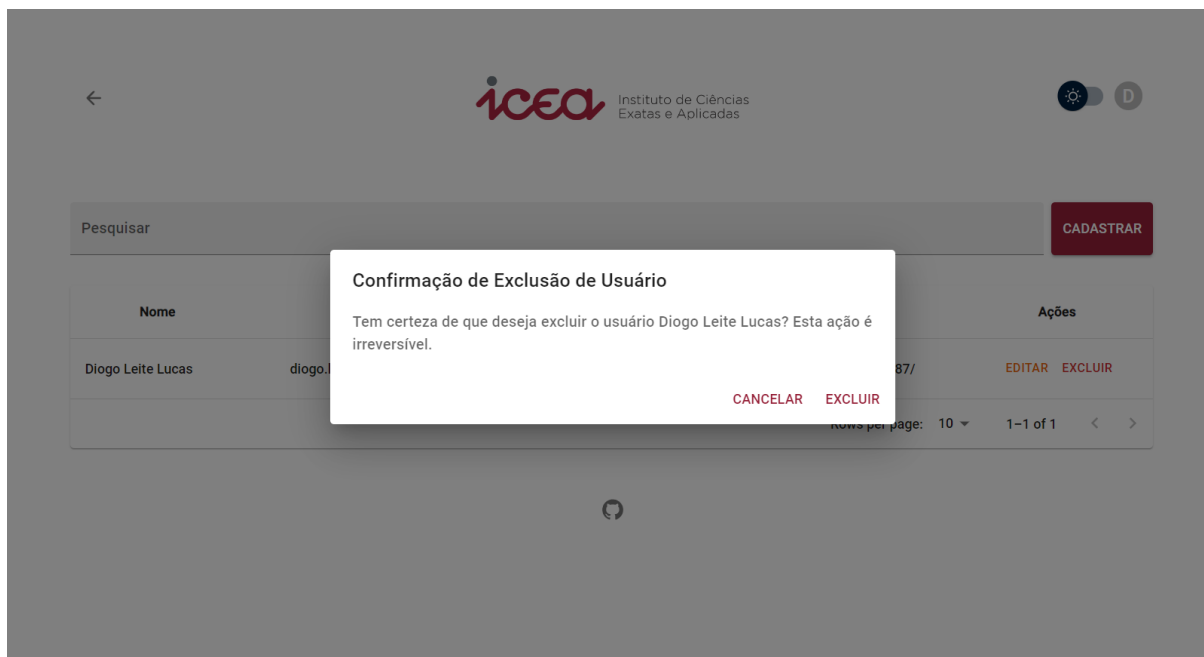


Fonte: Elaborado pelo autor.

4.7 Relatório do Aluno

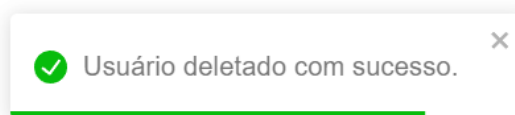
A página de relatório do aluno contém informações de um único aluno, no caso de usuários da categoria discente, aparecerá as informações referente ao seu usuário, conforme mostra a Figura 25. Para usuários da categoria docente, é possível pesquisar pela matrícula de discentes e obter o relatório, conforme mostra a Figura 26. Essa funcionalidade não está disponível para categoria discentes.

Figura 22 – Captura de tela da confirmação de exclusão do usuário



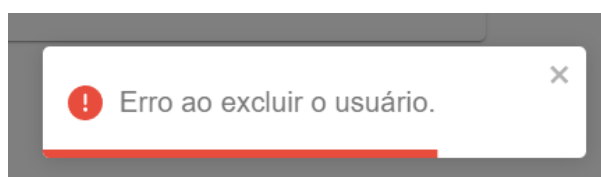
Fonte: Elaborado pelo autor.

Figura 23 – Captura de tela do sucesso em excluir o usuário



Fonte: Elaborado pelo autor.

Figura 24 – Captura de tela do erro em excluir o usuário



Fonte: Elaborado pelo autor.

A página contém métricas do aluno, como "Coeficiente", "Disciplinas Cursando", "Disciplinas Aprovadas", "Disciplinas Reprovadas", "Disciplinas Trancadas" e "Disciplinas Canceladas". Também estão disponíveis os gráficos de "Coeficiente Semestral", trazendo o coeficiente do aluno em cada período cursado, e o gráfico de "Disciplina Semestral", contendo o total de disciplinas, aprovadas, reprovadas, canceladas e trancadas, de cada período cursado.

Também está disponível o *cluster* atribuído para o aluno em questão durante o processo de clusterização, trazendo uma série de filtros para personalizar e ajudar a entender as características em comum com outros alunos, a seção da situação dos alunos em cada *cluster*, contendo o gráfico de calor e de barras, bem como o gráfico do tipo diagrama de caixa, para o coeficiente e o tempo de curso dos *clusters*.

Tanto para os docentes, que podem pesquisar informações sobre um aluno, quanto para os discentes, que têm acesso ao seu relatório individual, um sumário de fácil acesso com métricas e gráficos na página pode facilitar na identificação da situação do aluno em questão. Os gráficos de coeficiente semestral mostram o desempenho ao longo dos semestres, enquanto os gráficos de disciplinas detalham o número de aprovações, reprovações, cancelamentos e trancamentos ao longo do tempo.

Para os usuários docentes, a disposição dos *clusters* pode auxiliar na identificação de alunos que estão matriculados, mas que pertencem a um *cluster* com alto índice de evasão. Isso demonstra uma padronização que sugere uma tendência de evasão, com base nos demais alunos que possuem atributos acadêmicos semelhantes e que já evadiram da instituição. O mesmo pode ocorrer para os *clusters* relacionados ao tempo de curso, sinalizando a tendência de um aluno matriculado permanecer retido na instituição. A identificação de um aluno em um *cluster* com essas tendências permite um tratamento individualizado, ajudando a entender a situação e a tomar decisões para tentar reverter o quadro.

Para os usuários discentes, a sinalização do seu *cluster* atribuído, junto à disposição geral dos *clusters*, pode auxiliar na percepção do aluno sobre seu desempenho no curso e seu futuro na instituição. Assim, se julgar necessário, o aluno pode procurar o corpo docente do *campus*, ou colegiado do curso, em busca de apoio.

4.8 Clusterização

A página de *cluster* mostra os resultados e *clusters* obtidos para cada curso do *campus*, com base nos atributos acadêmicos dos alunos. A Figura 27 apresenta o mapa de calor e gráfico de barras com base na situação atual dos integrantes de cada *cluster*. A Figura 28 e 29 apresenta os diagramas de caixa do coeficiente e semestres cursado dos alunos em cada *cluster*, respectivamente. É necessário selecionar o curso na parte superior da página, também é possível personalizar a análise de acordo com a preferência do usuário, com os filtros disponíveis de sexo, situação e ações afirmativas. Uma breve explicação da página está disponível na parte inferior, cada gráfico possui um botão de "**ajuda**", no qual explica a visualização.

Usuários das categorias discentes e docentes têm acesso à página em questão. Dessa forma, os docentes podem visualizar informações relevantes sobre o desempenho dos alunos, identificar tendências e tomar decisões para aprimorar o acompanhamento e o apoio aos estudantes. Da mesma forma, a categoria discente pode realizar análises semelhantes, incentivando a participação ativa na tomada de decisões pelo colegiado do curso.

4.9 Admissões

A Figura 30 mostra a página de admissões, no qual está disponível informações dos gráficos de "Admissões por Período" no formato colunar e no formato radar, além do "Total de Admissões" no formato de pizza. É possível personalizar as análises de acordo com a preferência do usuário, tendo disponível os filtros de cursos, anos, semestres e modalidade de admissão. Também é necessário escolher um parâmetro de agrupamento dos dados que podem ser sendo do semestre, curso ou modalidade. Todas as funcionalidades são explicadas movendo o cursor para o ponto de interrogação ao lado da opção.

É possível também acessar os dados em formato de tabela, com os mesmos filtros, entretanto é possível selecionar mais de uma opção de agrupamento dos dados, sendo ano, semestre, modalidade e curso, conforme mostra a Figura 31. A tabela é paginada e os dados exibidos são alterados de acordo com a personalização dos agrupamento.

Usuários das categorias discentes e docentes têm acesso à página em questão. Dessa forma, os docentes podem analisar informações relevantes sobre o perfil dos novos alunos, identificar tendências de ingresso e tomar decisões embasadas para melhorar o processo de seleção e acolhimento. Da mesma forma, os discentes podem examinar esses dados, promovendo uma maior compreensão do perfil dos ingressantes e incentivando a participação ativa na discussão sobre melhorias no processo de admissão.

4.10 Graduações

A Figura 32 mostra a visualização de estatísticas gerais da página de graduações, no qual está disponível o gráfico "Quantidade de Aluno Graduado por Tempo" no campus ICEA, bem como métricas de "Quantidade Total de Alunos Graduados", "Média do Coeficiente de Alunos Graduados", "Quantidade de Semestres Médios para Graduar" e "Porcentagem de Alunos Graduados". É possível personalizar as análises de acordo com a preferência do usuário, tendo disponível os filtros de cursos, anos, semestres e modalidade de admissão. Também é necessário escolher um parâmetro de agrupamento dos dados, que podem ser sendo o semestre, curso ou modalidade. Todas as funcionalidades são explicadas movendo o cursor para o ponto de interrogação ao lado da opção.

A Figura 33 mostra a visualização de estatísticas por período de ingresso da página de graduações, similar a página de admissões, estão disponíveis informações os gráficos de "Graduados por Período" no formato colunar e no formato radar, também "Total de Graduações" no formato de pizza. Contendo os filtros de cursos, anos, semestres e modalidade de admissão, também é necessário escolher um parâmetro de agrupamento dos dados, sendo semestre, curso ou modalidade.

Usuários das categorias discentes e docentes têm acesso à página em questão. Dessa forma, os docentes podem analisar informações relevantes sobre o perfil dos graduandos, identificar tendências de diplomação e tomar decisões informadas para aprimorar os processos de ensino e apoio aos alunos. Da mesma forma, os discentes podem explorar esses dados, promovendo uma maior compreensão do desempenho de suas turmas e incentivando a participação ativa nas discussões sobre melhorias nos caminhos para a diplomação.

4.11 Evasões

Similar a página de admissões, a página de evasões apresenta os mesmos conjuntos de gráficos, sendo eles "Evasões por Período de Ingresso" no formato colunar e no formato radar, e "Total de Evasões" no formato de pizza, conforme mostra a Figura 34. Também estão disponíveis os filtros de cursos, anos, semestres, modalidades de admissão e agrupamento único de semestre, curso ou modalidade. A visualização em tabela também está disponível, contendo os mesmos filtros e podendo agrupar por ano, semestre, modalidade de admissão e curso.

Usuários das categorias discentes e docentes têm acesso à página em questão. Dessa forma, os docentes podem analisar os fatores que contribuem para a desistência dos alunos, identificando padrões e tendências que auxiliem na formulação de estratégias para apoio ao estudante. Da mesma forma, os discentes podem explorar esses dados, promovendo uma compreensão mais profunda dos desafios enfrentados e incentivando um diálogo construtivo sobre possíveis melhorias nos processos que favorecem a permanência e o sucesso acadêmico.

4.12 Disciplinas

A Figura 35 apresenta a página de disciplinas, contendo o mesmo padrão das páginas de admissões, graduações e evasões. Possuindo os filtros de nome da disciplina, departamentos, anos e semestres, os agrupamentos disponíveis são por departamento ou semestre. Um diferencial da página de disciplinas, é que existe a possibilidade de mudar a variável que deve ser representada no gráfico, sendo as opções aprovações, reprovações, trancamentos, cancelamentos e matriculados. Os mesmos gráficos estão disponíveis, bem como a visualização em tabela.

Usuários das categorias discentes e docentes têm acesso à página em questão. Dessa forma, os docentes podem analisar informações relevantes sobre o andamento das disciplinas, identificar taxas altas de reprovação e tomar decisões para aprimorar suas abordagens pedagógicas. Da mesma forma, os discentes podem explorar esses dados, promovendo uma reflexão sobre suas práticas de estudo e incentivando a busca por apoio, o que contribui para um ambiente de colaboração e comprometimento com o sucesso acadêmico.

4.13 Grafo de Dependências

A Figura 36 mostra a página de grafo de dependências, onde é possível visualizar as grades dos quatro cursos disponíveis no ICEA. Essa página serve como um guia onde o usuário da categoria discente pode visualizar sua grade curricular, evidenciando com cores as disciplinas que já foram concluídas (verde), cursando (azul), reprovadas (vermelho), canceladas (vermelho) e que faltam cursar (branco).

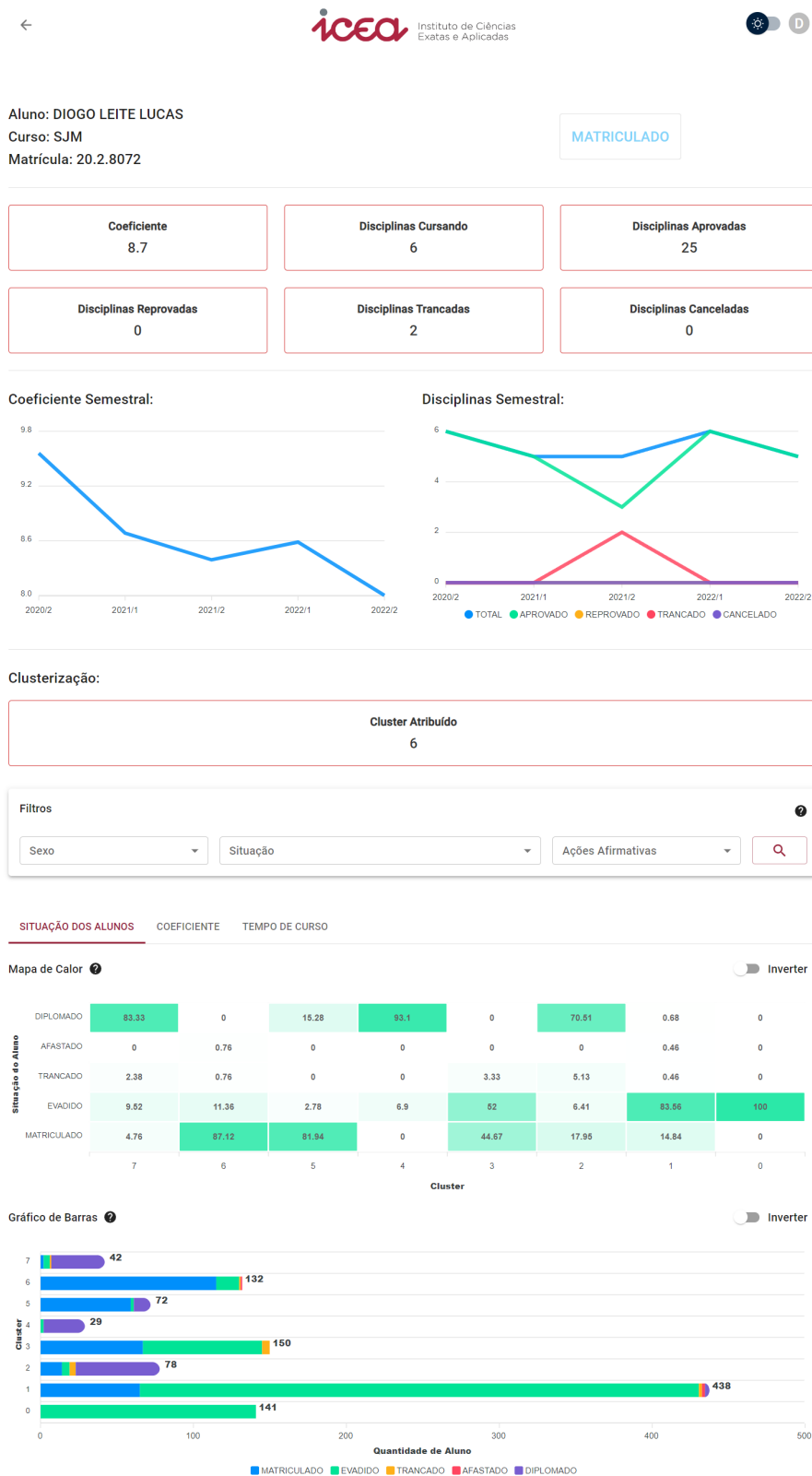
Com o intuito de auxiliar o usuário na tomada de decisão, as categorias discentes e docentes, selecionando qualquer disciplina disponível na grade, uma tela de estatísticas da disciplina é aberta contendo o gráfico temporal dos índices de aprovação, reprovação e trancamento da disciplina, conforme mostra a Figura 37. Essa informação pode auxiliar o usuário a escolher a melhor combinação de disciplina que deseja cursar nos próximos semestres, ou tomar a melhor decisão diante da sua situação atual, seja trancando disciplinas ou o período completo.

4.14 Referência da API

A página de referência da API serve como um guia de como utilizar os métodos e *endpoints* disponíveis, podendo ser utilizado para minerar dados, construir outras visualizações, análises e tratamentos. A Figura 38 mostra o índice da página.

A Figura 39 mostra um módulo de requisição para autenticar na API do sistema, cada método disponível tem seu próprio módulo e serve como um guia de utilização, contendo os parâmetros necessários, cabeçalho, corpo da requisição e a url. Também é possível visualizar um exemplo da estrutura e campos retornados de cada método, bem como o resultado após uma requisição bem sucedida.

Figura 25 – Captura de tela da página de relatório do aluno



Fonte: Elaborado pelo autor.

Figura 26 – Captura de tela da página de relatório do aluno para docentes

← **ICEA** Instituto de Ciências Exatas e Aplicadas

⚙️ T

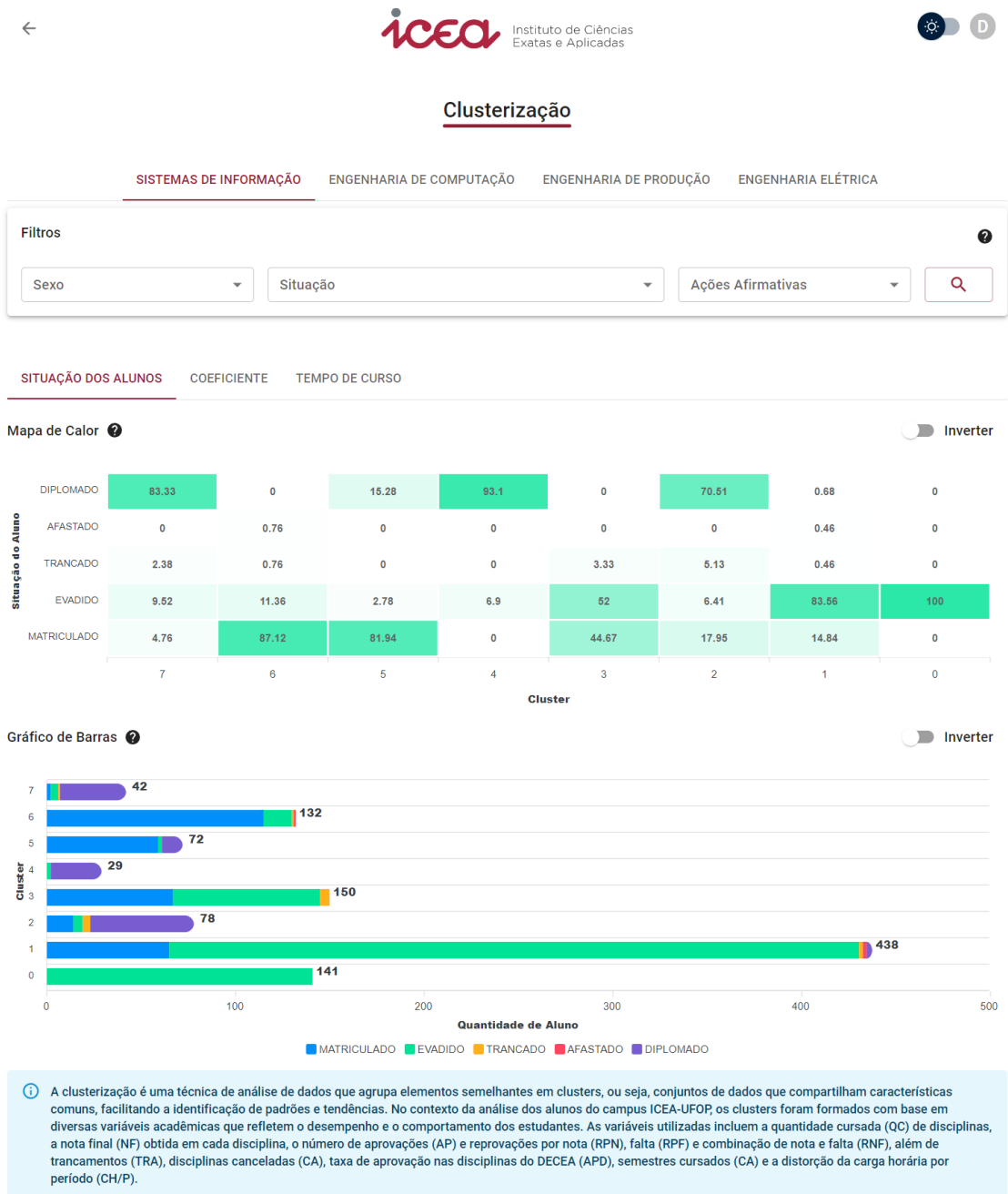
Matrícula 🔍

i **Insira uma Matrícula**
Para obter informações, insira a matrícula do aluno. Acesse o relatório individual do desempenho acadêmico no campus ICEA - UFOP, acompanhado de uma estimativa personalizada baseada em dados agregados de outros estudantes.

↻

Fonte: Elaborado pelo autor.

Figura 27 – Captura de tela da página de clusterização da seção de situação dos alunos



Fonte: Elaborado pelo autor.

Figura 28 – Captura de tela da página de clusterização da seção de coeficiente



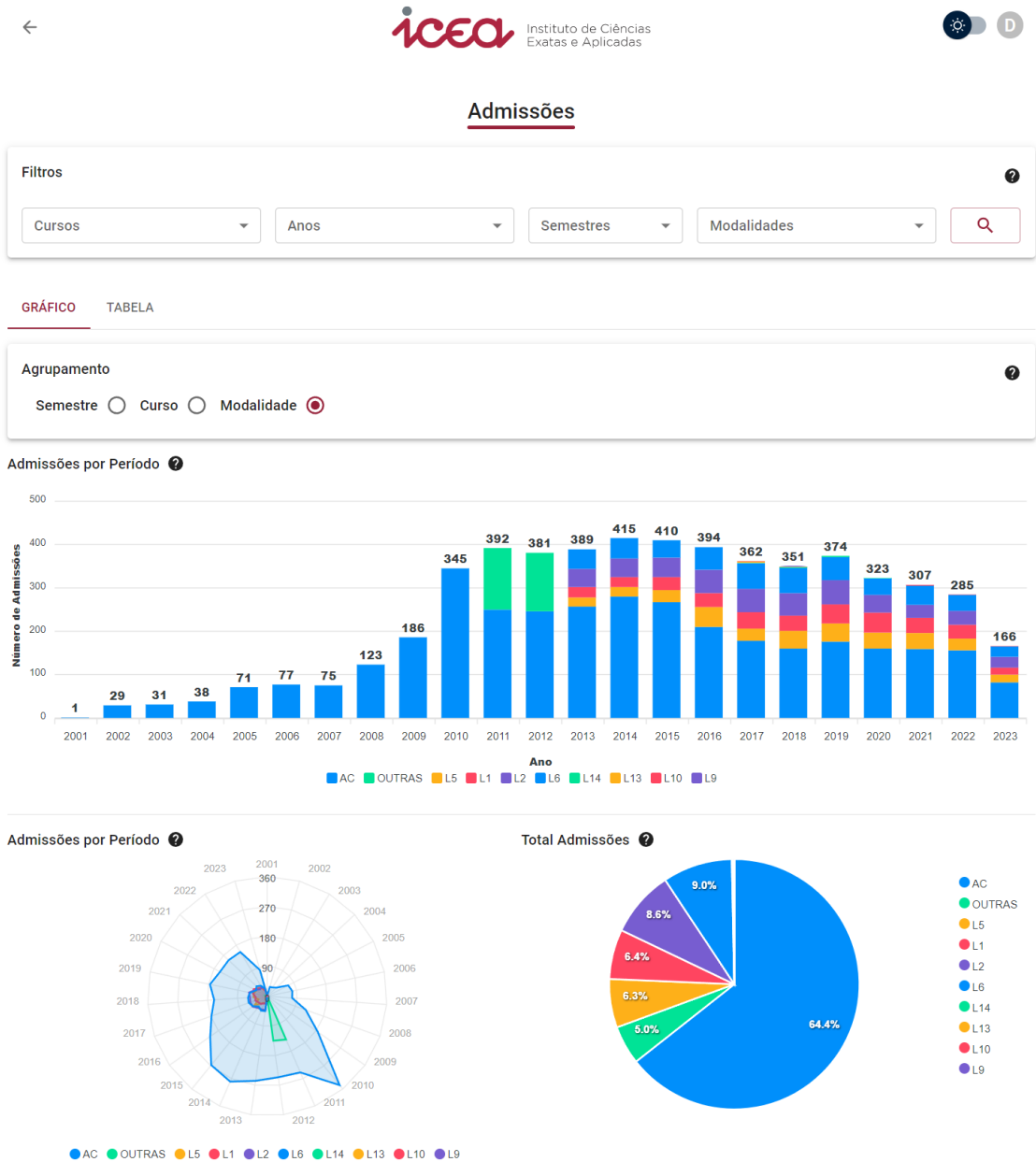
Fonte: Elaborado pelo autor.

Figura 29 – Captura de tela da página de clusterização da seção de tempo de curso



Fonte: Elaborado pelo autor.

Figura 30 – Captura de tela da página de admissões



Fonte: Elaborado pelo autor.

Figura 31 – Captura de tela da tabela da página de admissões

The screenshot displays the 'Admissões' (Admissions) page of the ICEA system. At the top, there is a navigation bar with the ICEA logo and the text 'Instituto de Ciências Exatas e Aplicadas'. Below the navigation bar, the title 'Admissões' is centered. A filter section labeled 'Filtros' contains dropdown menus for 'Cursos', 'Anos', 'Semestres', and 'Modalidades', along with a search icon. Below the filters, there are two tabs: 'GRÁFICO' and 'TABELA', with 'TABELA' being the active tab. Under the 'TABELA' tab, there is an 'Agrupamento' (Grouping) section with four checked checkboxes: 'Ano', 'Semestre', 'Modalidade', and 'Curso'. The main content is a table with the following data:

Ano	Semestre	Curso	Modalidade	Total Admissões
2015	1	EJM	AC	70
2014	2	EJM	AC	51
2010	1	CJM	AC	49
2010	1	EJM	AC	46
2010	2	EJM	AC	44
2010	1	PJM	AC	43
2010	2	PJM	AC	42
2017	1	CJM	AC	41
2016	2	CJM	AC	40
2010	2	SJM	AC	40

At the bottom of the table, there is a pagination control showing 'Número por página: 10' and a series of page numbers: 1, 2, 3, 4, 5, ..., 10, >. The number '1' is highlighted in a red circle, indicating the current page.

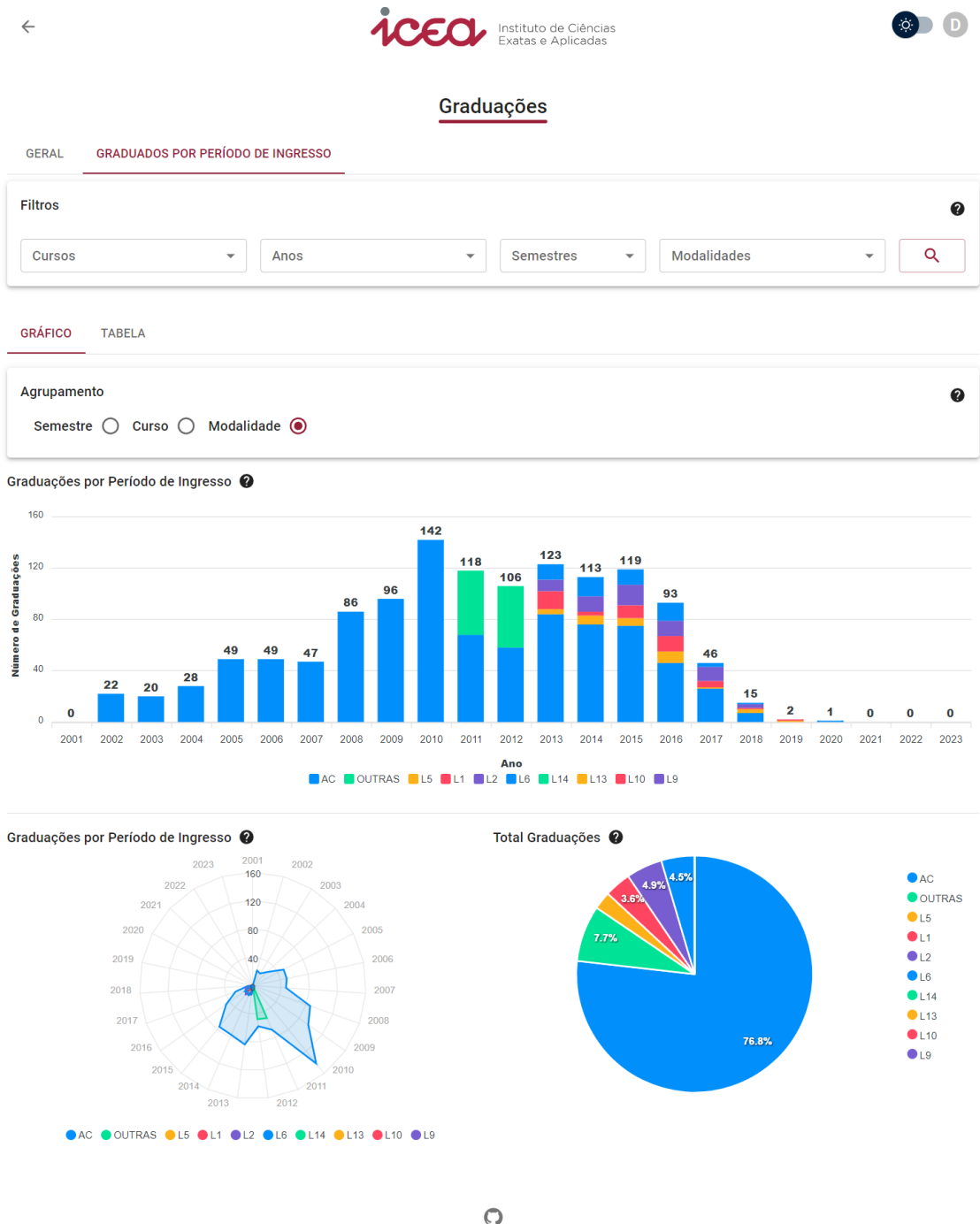
Fonte: Elaborado pelo autor.

Figura 32 – Captura de tela de estatísticas gerais página de graduações



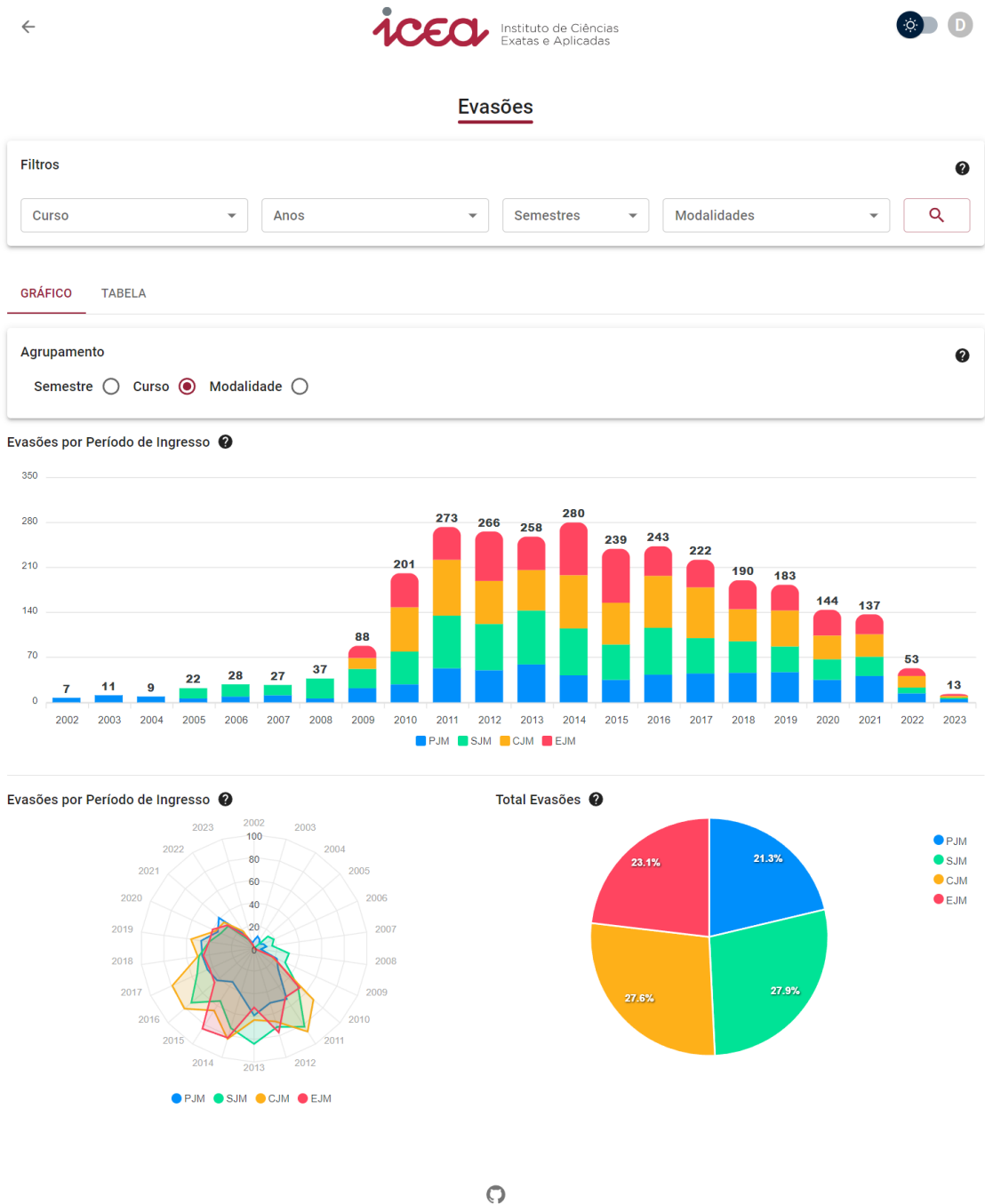
Fonte: Elaborado pelo autor.

Figura 33 – Captura de tela do graduados por período de ingresso da página de graduações



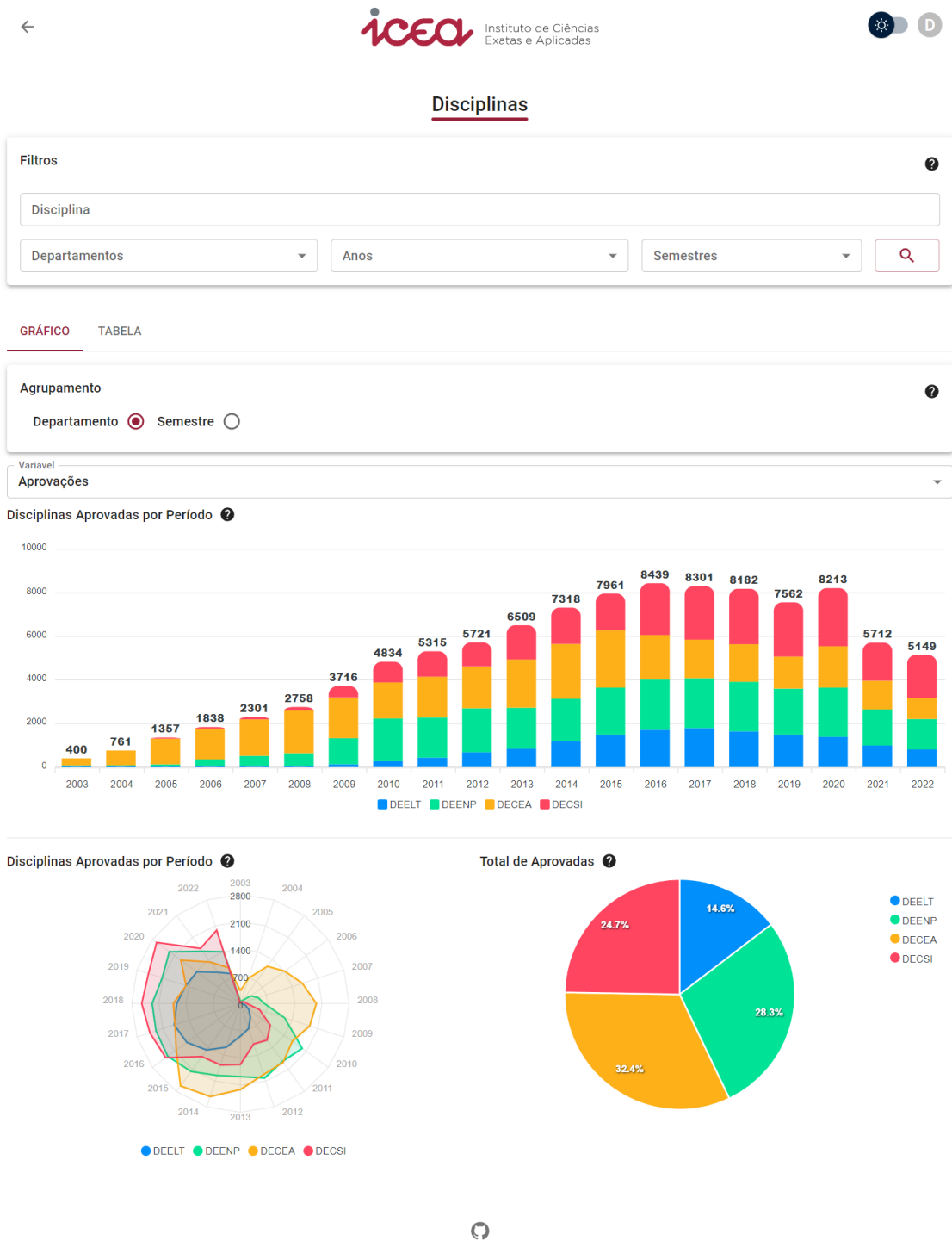
Fonte: Elaborado pelo autor.

Figura 34 – Captura de tela da página de evasões



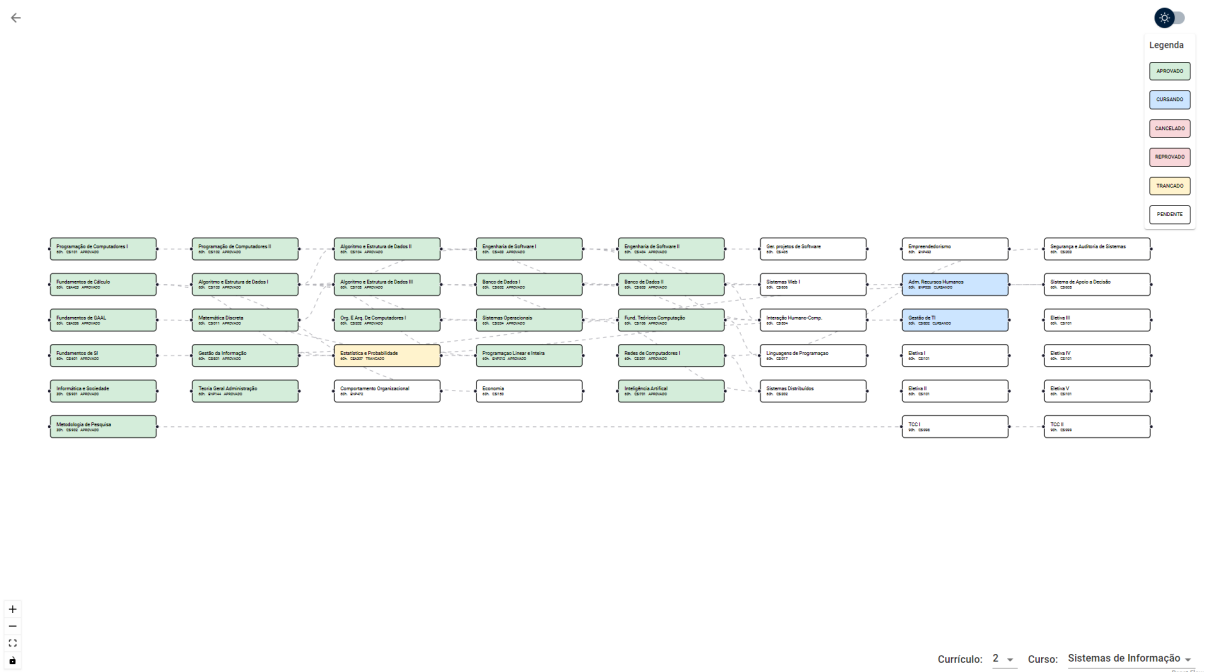
Fonte: Elaborado pelo autor.

Figura 35 – Captura de tela da página de disciplinas



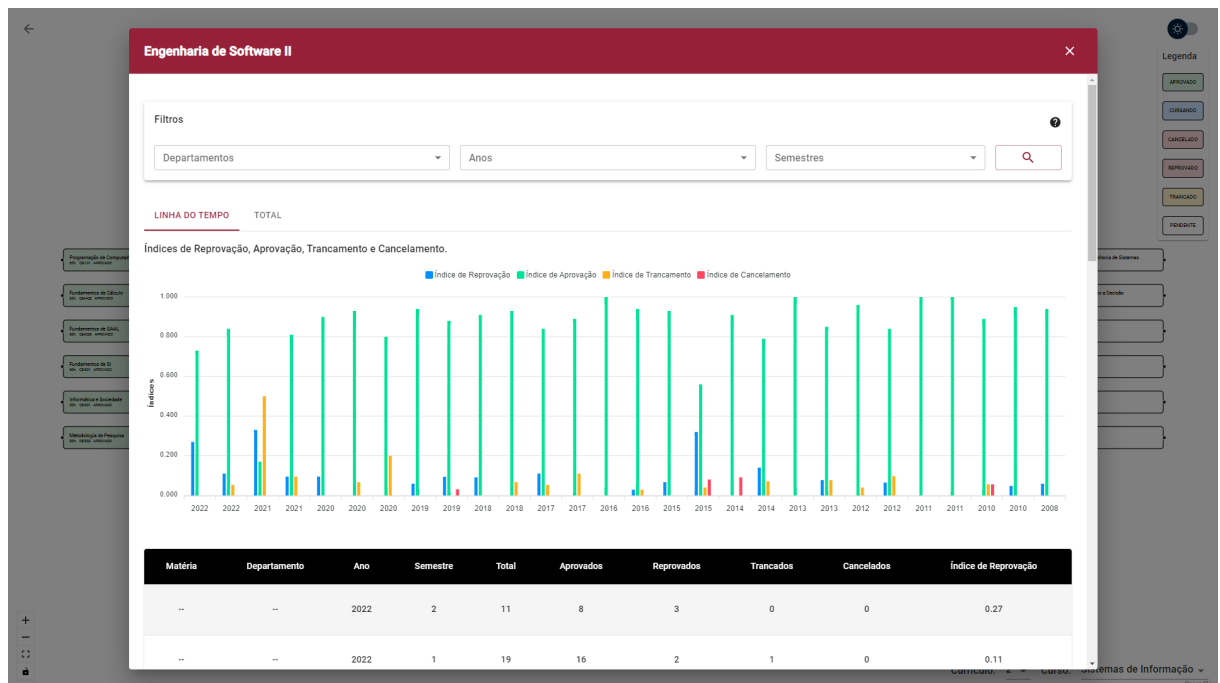
Fonte: Elaborado pelo autor.

Figura 36 – Captura de tela da grade de Sistemas de Informação na página grafo de dependências



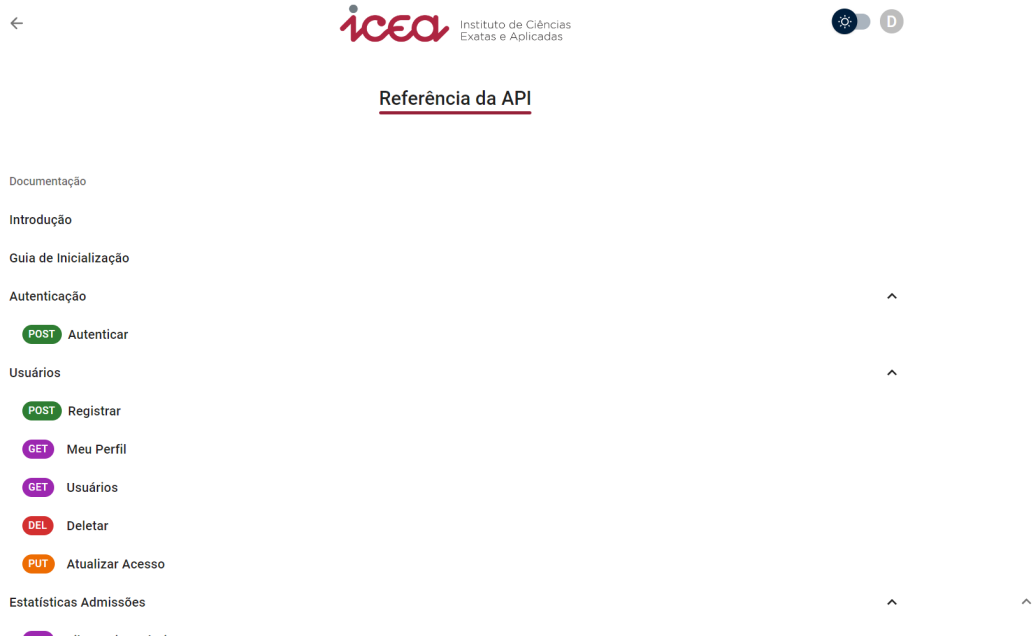
Fonte: Elaborado pelo autor.

Figura 37 – Captura de tela das estatísticas de disciplina na página grafo de dependências



Fonte: Elaborado pelo autor.

Figura 38 – Captura de tela do índice da página de referência da API



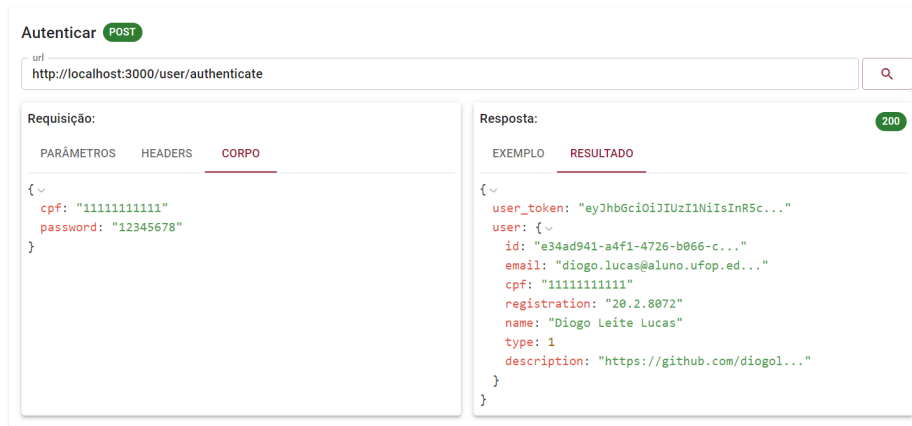
Fonte: Elaborado pelo autor.

Figura 39 – Captura de tela do módulo de requisição da página de referência da API

Antes de realizar qualquer requisição à API, é necessário obter suas credenciais de acesso. Normalmente, você precisará de uma chave de API ou token de autenticação. Solicite suas credenciais através do portal de desenvolvedores ou da equipe responsável.

Autenticação

Para interagir com as APIs, você precisará autenticar suas requisições usando um Bearer Token. O processo de autenticação envolve fazer uma requisição para o endpoint de autenticação com seu usuário e senha, que retornará um token que deve ser incluído nas chamadas subsequentes.



Usuários

A API de usuários oferece funcionalidades para gerenciar o controle de acesso, adicionar novos usuários e acessar informações detalhadas sobre os



Fonte: Elaborado pelo autor.

5 Conclusão

O presente trabalho consistiu no desenvolvimento de um sistema *web* de fácil acessibilidade e uso intuitivo, focado em disponibilizar informações relevantes sobre desempenho acadêmicos na instituição. O sistema facilita a visualização e interpretação dos dados por meio de gráficos interativos que melhoram a experiência do usuário com a aplicação, permitindo que a comunidade acadêmica tenha acesso a indicadores precisos para a tomada de decisões embasadas, promovendo, assim, uma orientação mais eficaz na identificação e combate ao abandono escolar e retenção dos alunos nos cursos.

Para projetos futuros, estão previstas importantes melhorias, como a automação do processo de atualização dos dados, garantindo que as informações estejam sempre atualizadas sem a necessidade de intervenções manuais. Além disso, pretende-se tornar o sistema responsivo, otimizando o seu uso em dispositivos móveis, bem como aprimorar a experiência do usuário por meio de um design mais acessível e inclusivo, beneficiando um público mais amplo. Também é necessário integrar as credenciais de acesso com a autenticação do sistema Minha Ufop.

Esses aperfeiçoamentos visam não apenas aumentar a eficiência do sistema, mas também promover o uso contínuo, integrado e sustentável da ferramenta pela comunidade acadêmica, potencializando seu impacto nas iniciativas educacionais da instituição.

Referências

- AGGARWAL, S. et al. Modern web-development using reactjs. **International Journal of Recent Research Aspects**, v. 5, n. 1, p. 133–137, 2018.
- CLEMENTE, Q. **Gestão de Frota de Veículos**. [S.l.]: "Lisboa IST.", 2008.
- COSTA, D. D. **EVASÃO NO ENSINO SUPERIOR DURANTE A PANDEMIA PELO COVID 19: UM ESTUDO DE CASO NA UNIVERSIDADE ESTADUAL DO RIO GRANDE DO SUL - UERGS**. 2022. UNIVERSIDADE ESTADUAL DO RIO GRANDE DO SUL. Disponível em: <https://repositorio.uergs.edu.br/xmlui/bitstream/handle/123456789/2379/_versao_final_formatada_diego.pdf?sequence=-1&isAllowed=y>.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge discovery and data mining: Towards a unifying framework. In: **Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)**. [S.l.]: s.n., 1996. v. 96, p. 82–88.
- FERREIRA, A. A. R. Caracterização do perfil de alunos do icea: uma análise guiada por dados e indicadores de sucesso. 2023.
- IEEE. **IEEE Recommended Practice for Architectural Description for Software-Intensive Systems**. 2000. Acesso em 07 de setembro de 2024. Disponível em: <<https://standards.ieee.org/ieee/1471/2187/>>.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 31, n. 3, p. 264–323, set. 1999. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/331499.331504>>.
- JOLLIFFE, I. T. **Principal component analysis for special types of data**. [S.l.]: Springer, 2002.
- KIMBALL, R.; ROSS, M. **The data warehouse toolkit: The complete guide to dimensional modeling**. [S.l.]: John Wiley & Sons, 2002.
- MARTIN, R. C. **Clean Architecture: A Craftsman's Guide to Software Structure and Design**. [S.l.]: Prentice Hall, 2017. (Robert C. Martin Series).
- MUELLER, J. P.; MASSARON, L. **Aprendizado de máquina para leigos**. [S.l.]: Alta Books Editora, 2019.
- MUKHERJEE, R.; KAR, P. A comparative review of data warehousing etl tools with new trends and industry insight. In: IEEE. **2017 IEEE 7th International Advance Computing Conference (IACC)**. [S.l.], 2017. p. 943–948.
- NIELSEN, J.; MOLICH, R. Heuristic evaluation of user interfaces. In: **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S.l.: s.n.], 1990. p. 249–256.
- PARANHOS, H. P. Desenvolvimento de um dashboard para análise e visualização dos dados educacionais dos discentes do instituto de ciências exatas e aplicadas da ufop. 2021.

Patrocínio, Wesley S. **Data Science é metodologia, e não tecnologia**. 2017. Acesso em 07 de setembro de 2024. Disponível em: <<https://medium.com/@wespatrocinio/data-science-%C3%A9-metodologia-e-n%C3%A3o-tecnologia-9c50fec73640>>.

PAULA, H. M.; ALMEIDA, L. F. F.; GORGES, V. d. O.; FERNANDES, C. H.; SANJULIÃO, L. K. A. F. A comparative review of data warehousing etl tools with new trends and industry insight. In: UTFPR. **Congresso Brasileiro de Engenharia de Produção**. [S.l.], 2020. p. 1–13.

RICHARDSON, L.; AMUNDSEN, M.; RUBY, S. **RESTful Web APIs: Services for a Changing World**. [S.l.]: "O'Reilly Media, Inc.", 2013.

SANTOS, G. **Sistema Web e Site, quais as diferenças entre eles?** 2018. Acesso em 06 de Setembro de 2024. Disponível em: <<http://cpejr.com.br/site/diferenca-sistema-web-site>>.

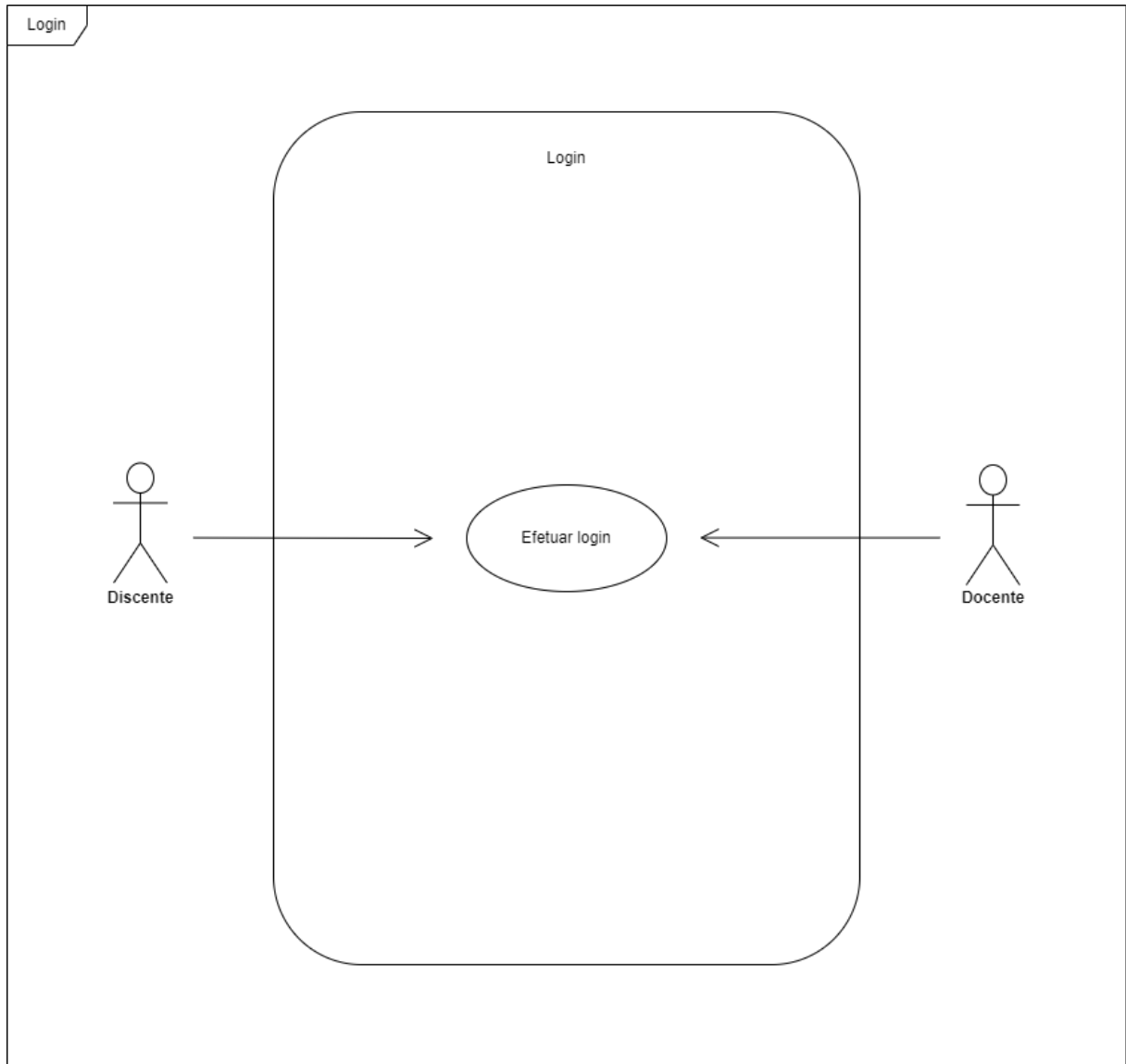
SEMESP. **Mapa de Ensino Superior**. 2024. Acesso em 30 de agosto de 2024. Disponível em: <<https://www.semesp.org.br/mapa/educacao-14/>>.

VALENTE, M. T. **Engenharia de Software Moderna: Princípios e Práticas para Desenvolvimento de Software com Produtividade**. [S.l.]: Editora: Independente, 2020.

APÊNDICE A – Diagramas de casos de uso

A.1 Login

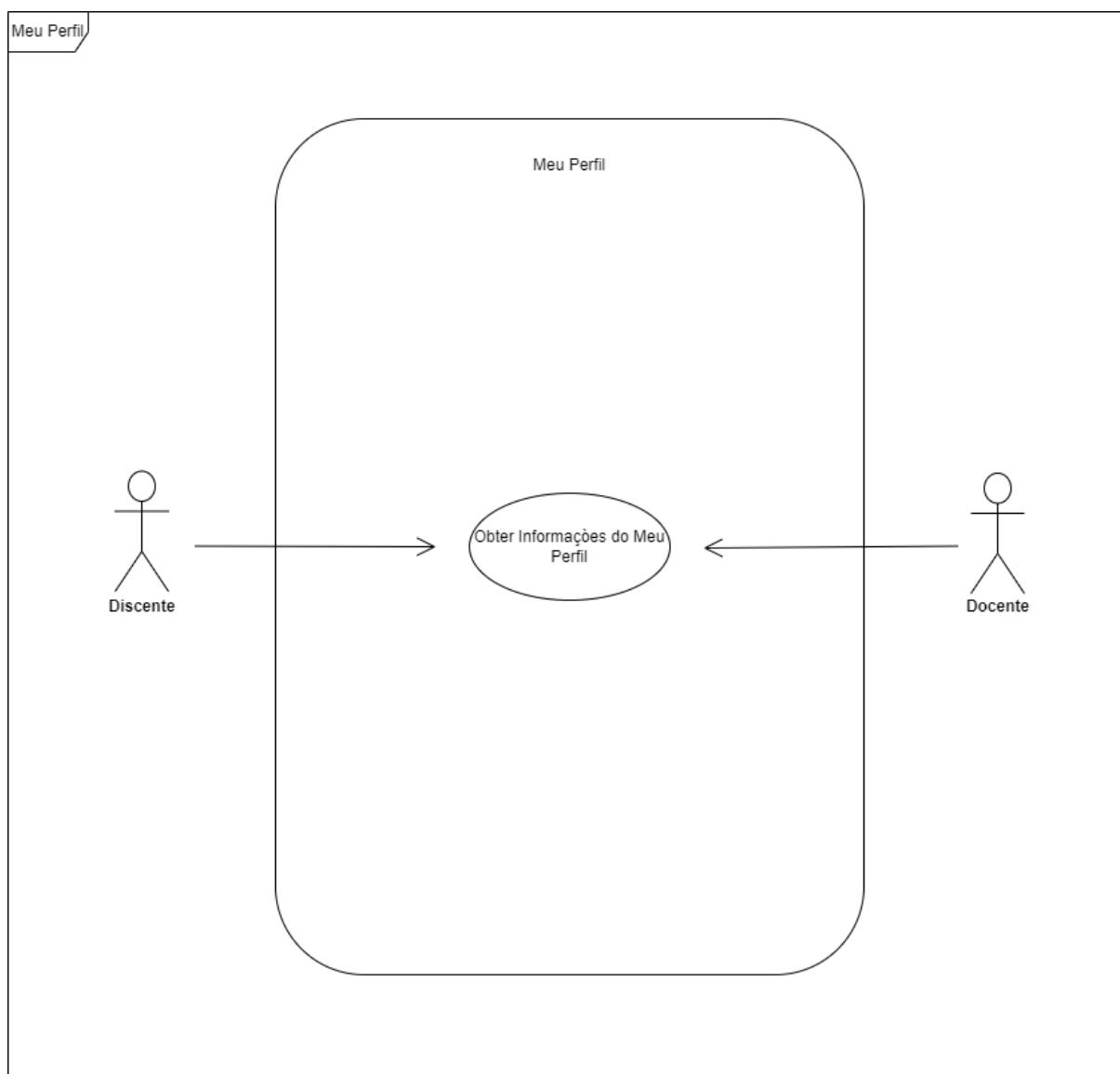
Figura 40 – Diagrama de casos de uso - Login



Fonte: Elaborado pelo autor no software Draw.io.

A.2 Meu Perfil

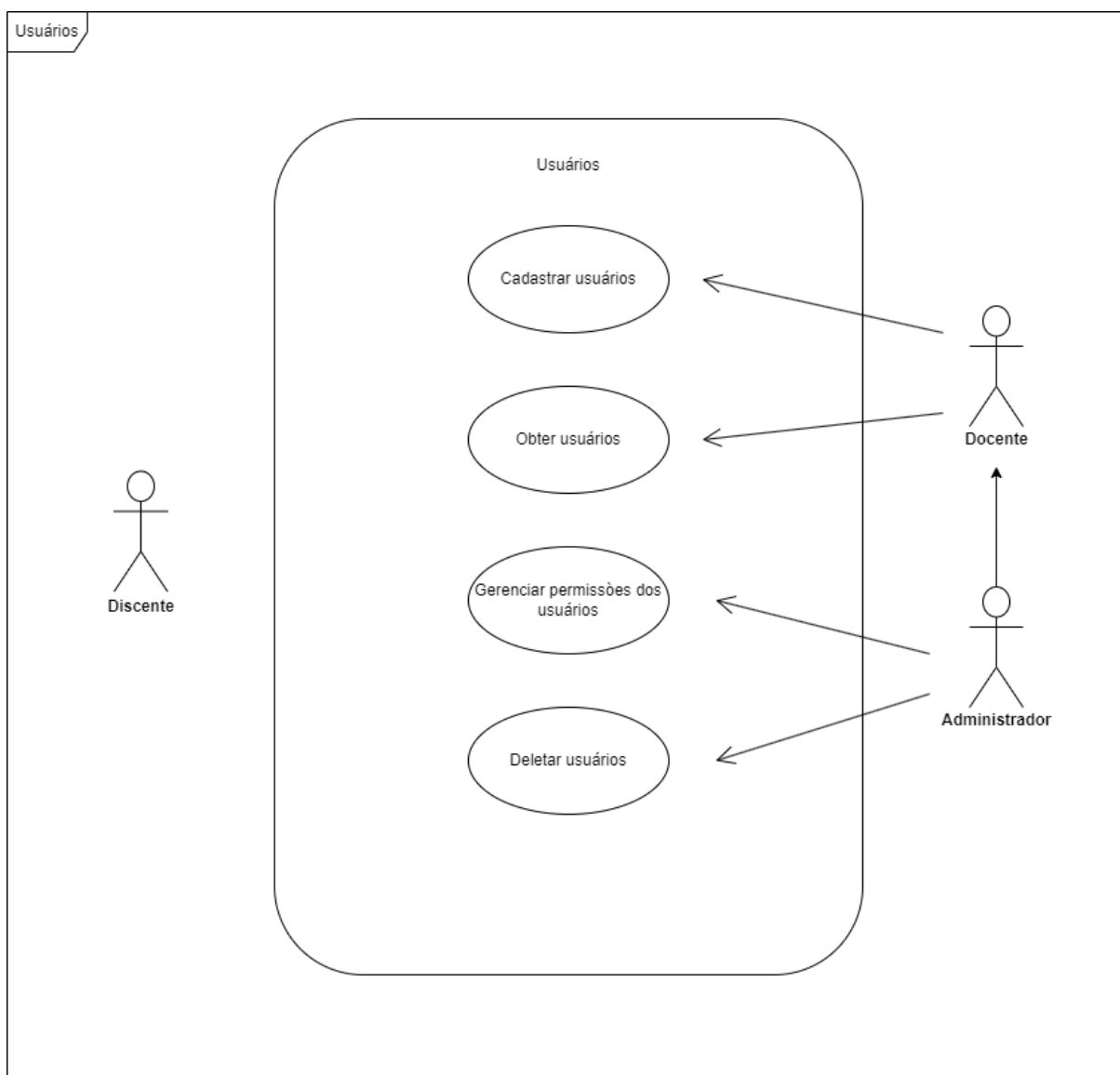
Figura 41 – Diagrama de casos de uso - Meu Perfil



Fonte: Elaborado pelo autor no software Draw.io.

A.3 Usuários

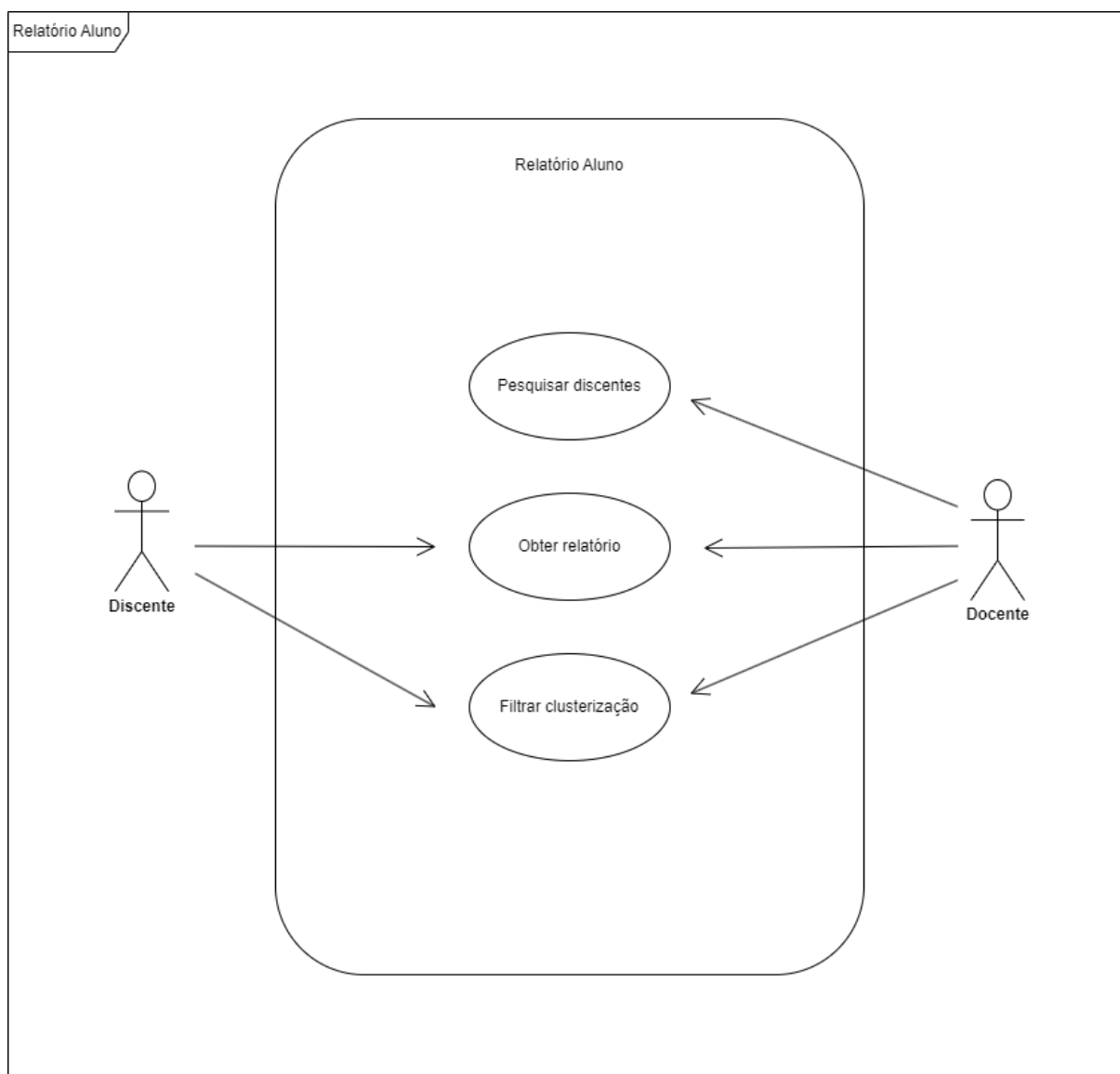
Figura 42 – Diagrama de casos de uso - Usuários



Fonte: Elaborado pelo autor no software Draw.io.

A.4 Relatório do Aluno

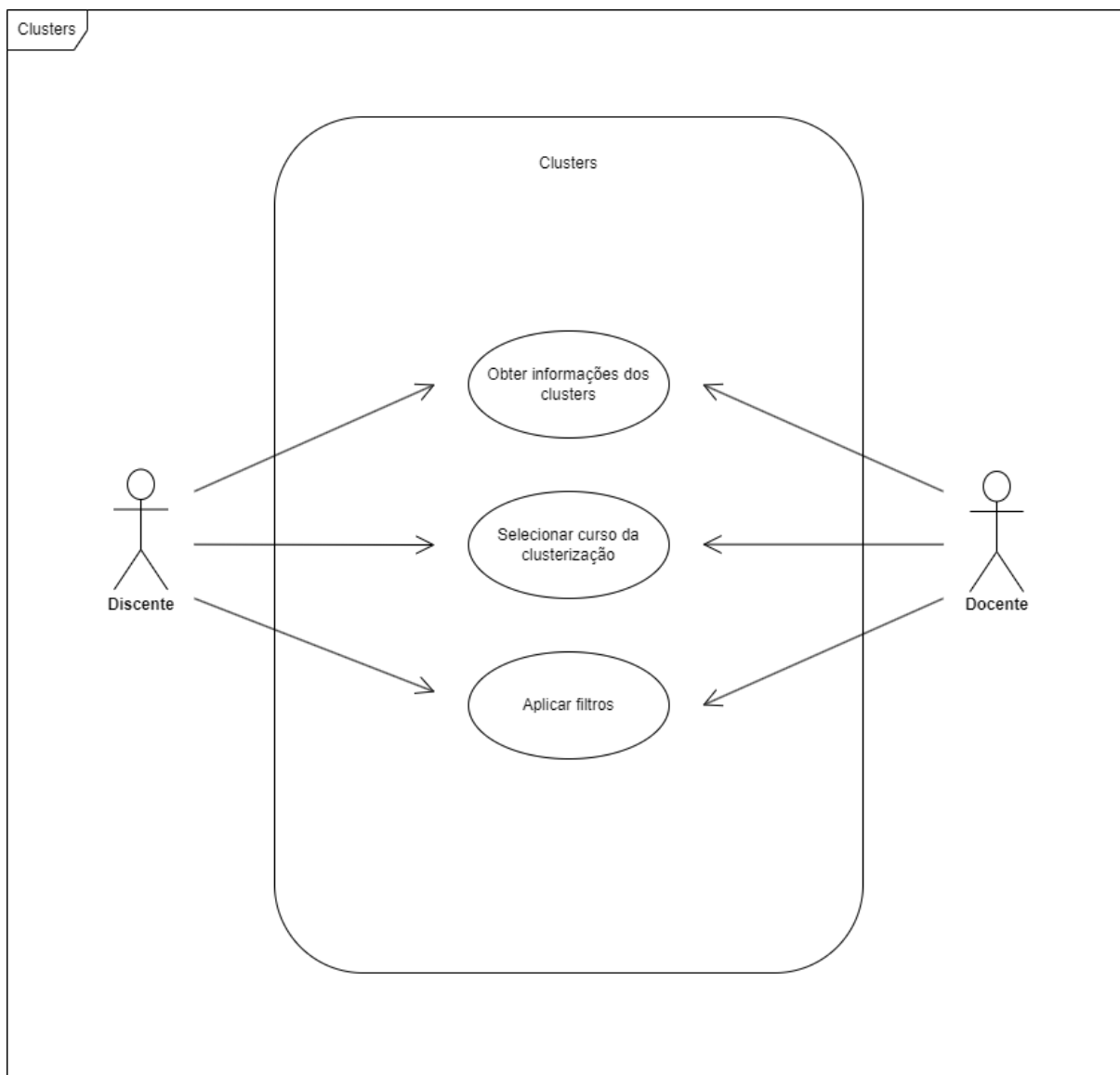
Figura 43 – Diagrama de casos de uso - Relatório do Aluno



Fonte: Elaborado pelo autor no software Draw.io.

A.5 Clusters

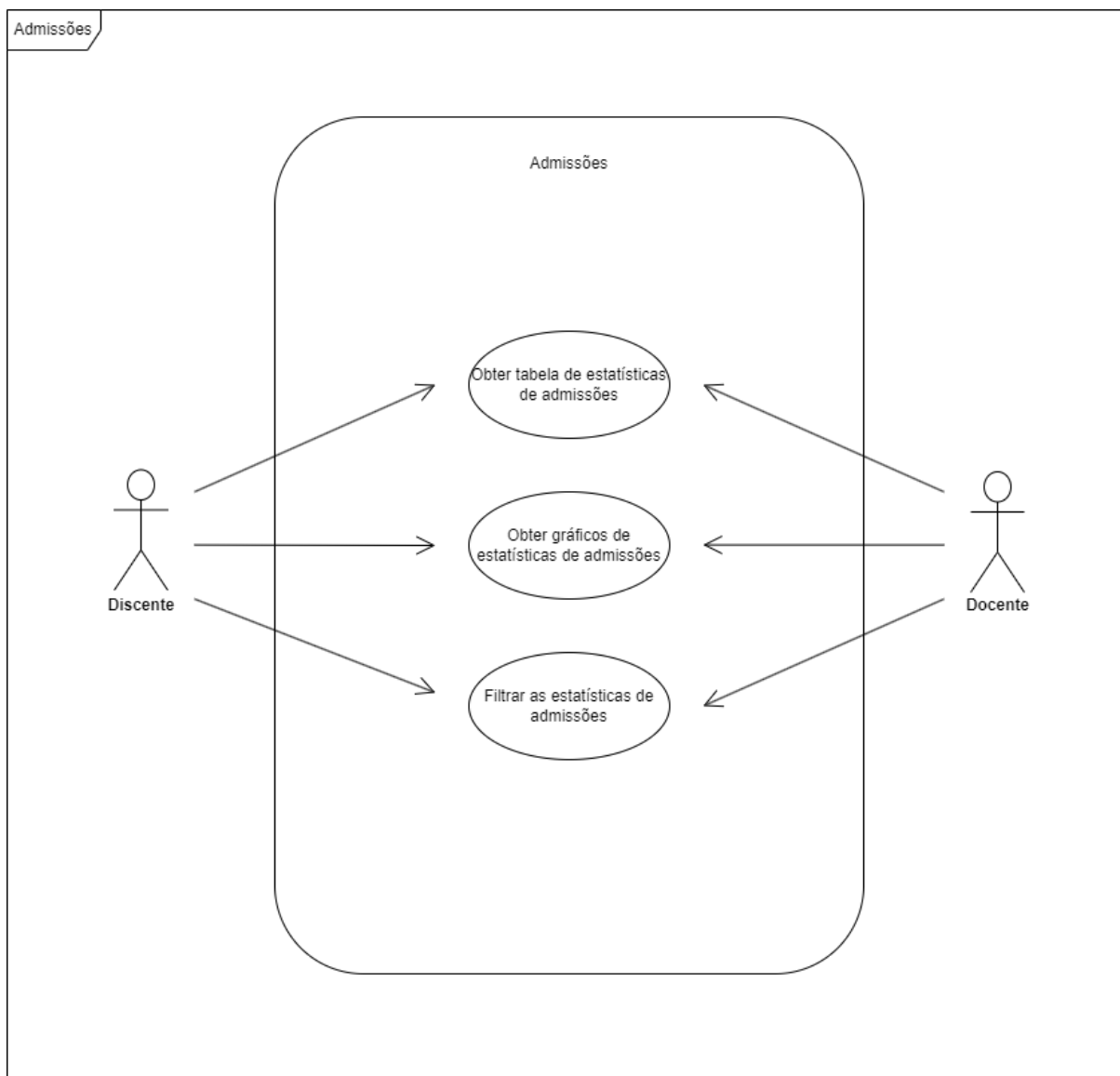
Figura 44 – Diagrama de casos de uso - Clusters



Fonte: Elaborado pelo autor no software Draw.io.

A.6 Admissões

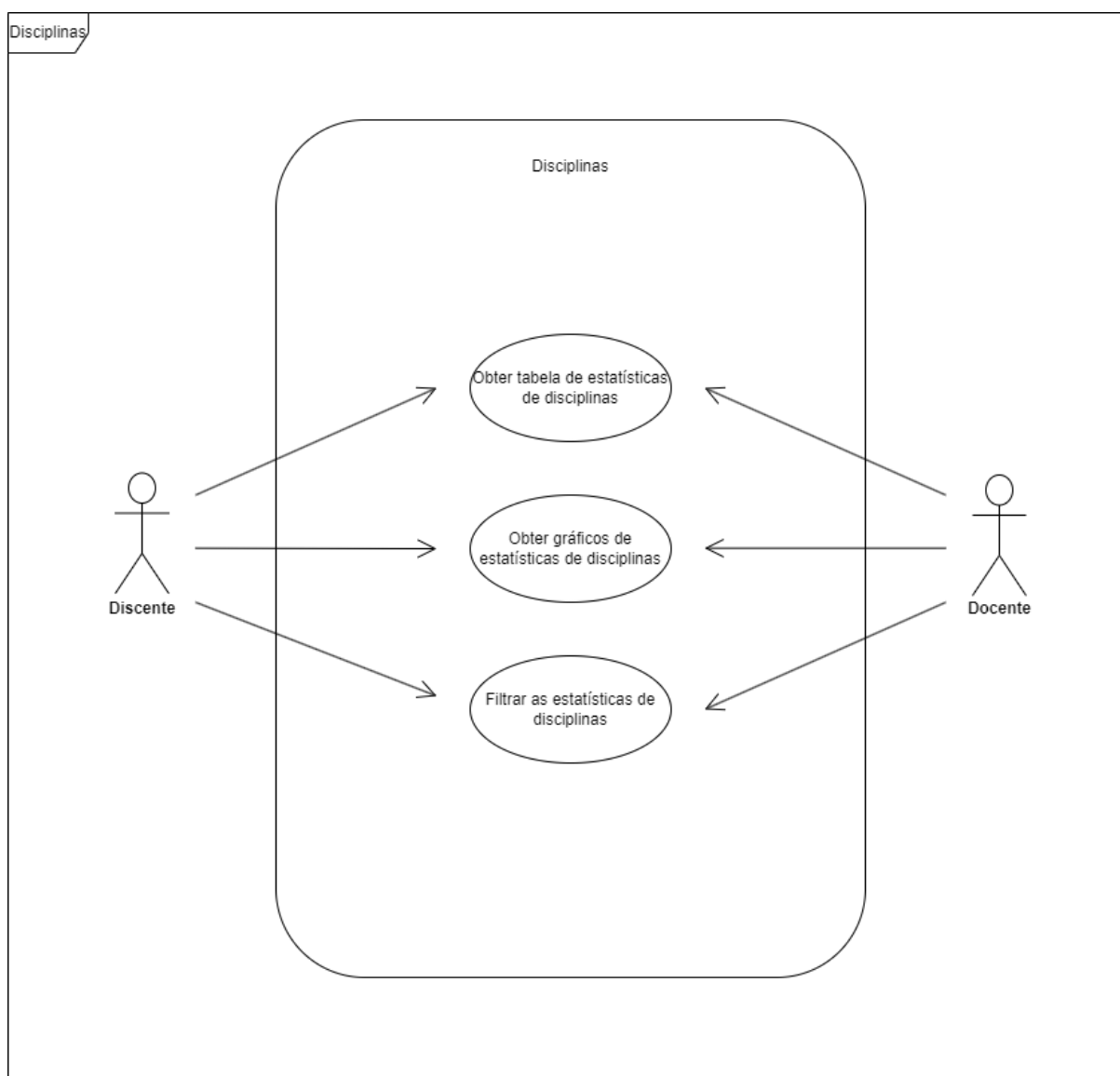
Figura 45 – Diagrama de casos de uso - Admissões



Fonte: Elaborado pelo autor no software Draw.io.

A.7 Disciplinas

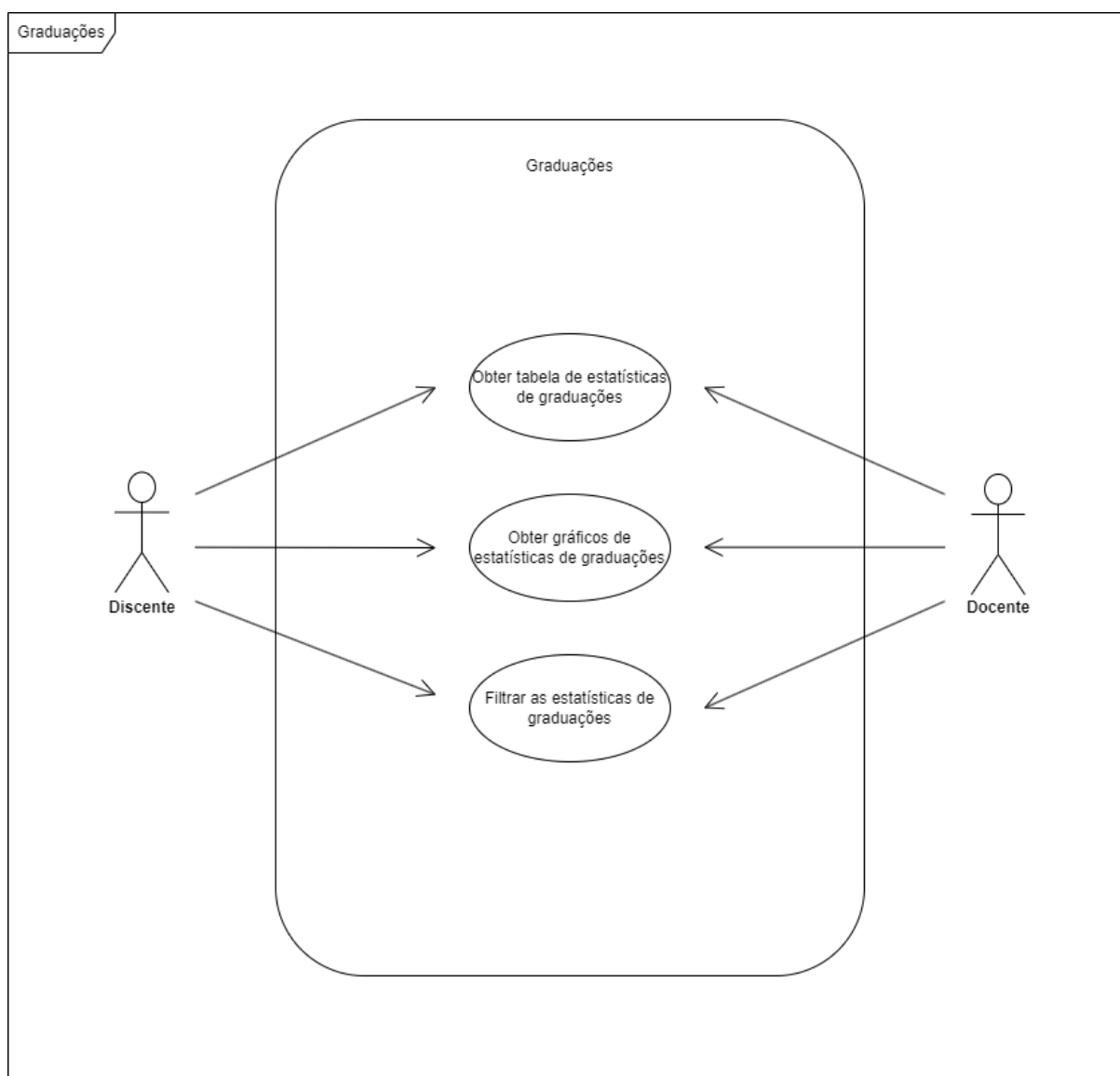
Figura 46 – Diagrama de casos de uso - Disciplinas



Fonte: Elaborado pelo autor no software Draw.io.

A.8 Graduações

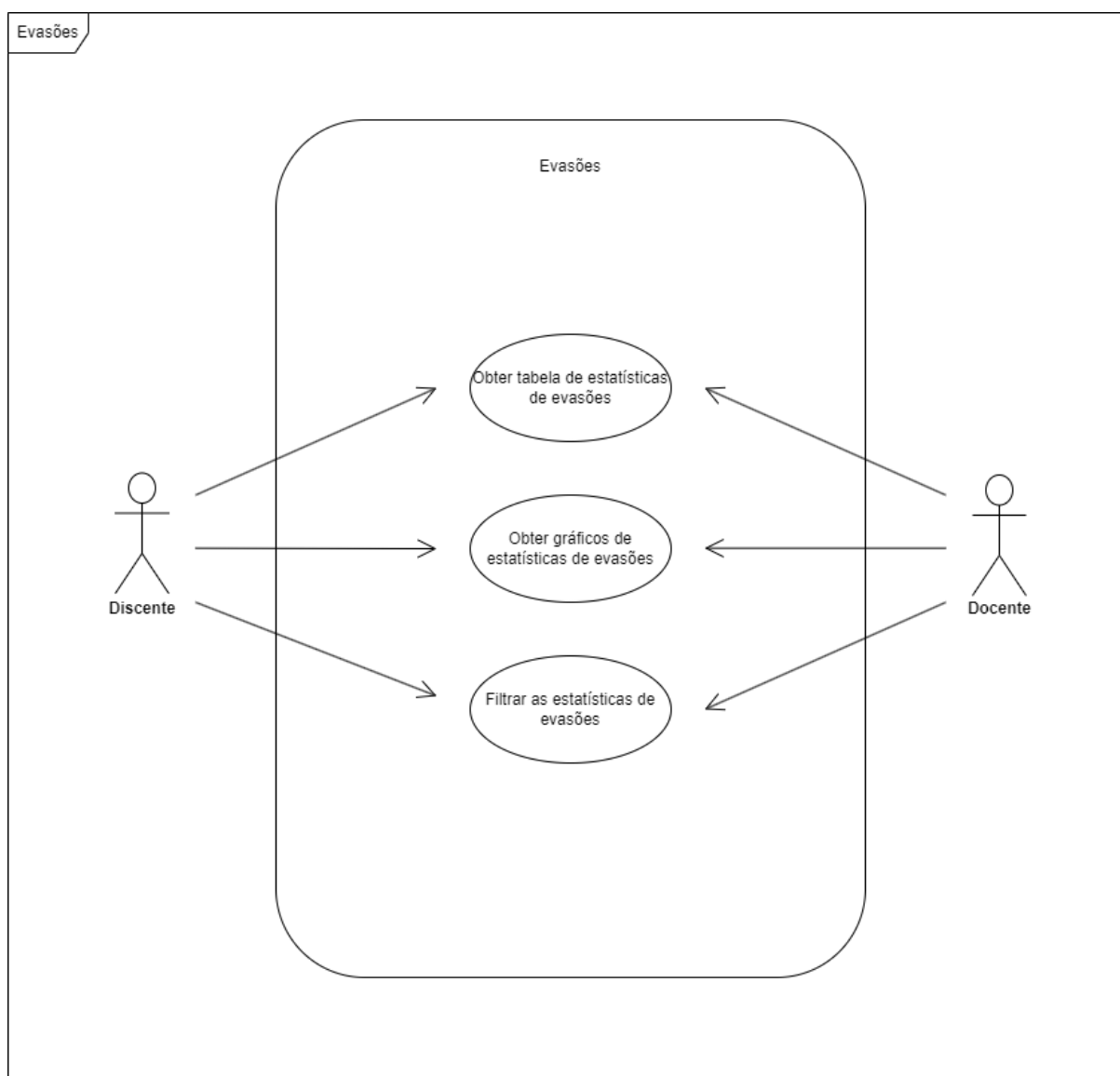
Figura 47 – Diagrama de casos de uso - Graduações



Fonte: Elaborado pelo autor no software Draw.io.

A.9 Evasões

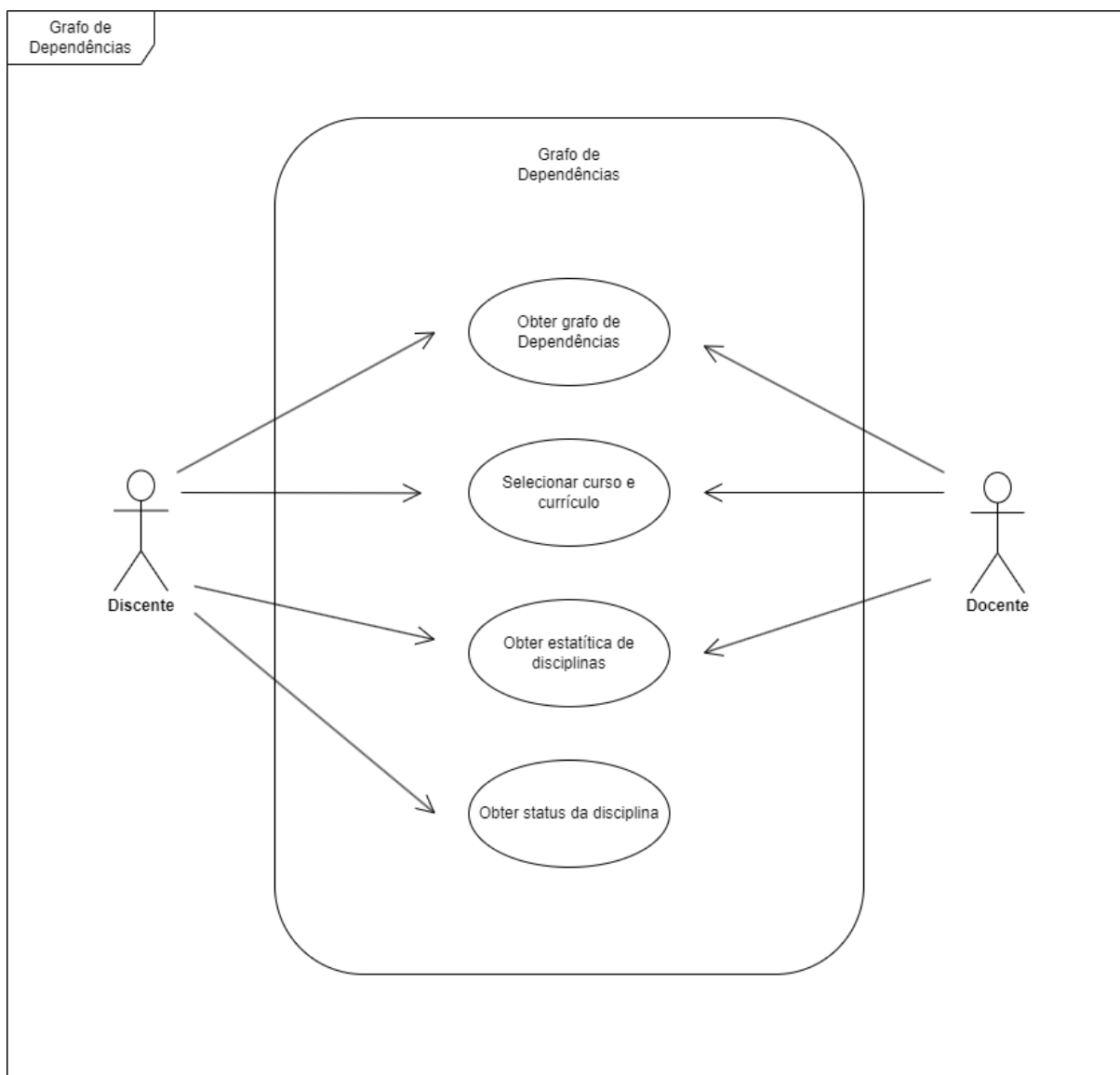
Figura 48 – Diagrama de casos de uso - Evasões



Fonte: Elaborado pelo autor no software Draw.io.

A.10 Grafo de Dependências

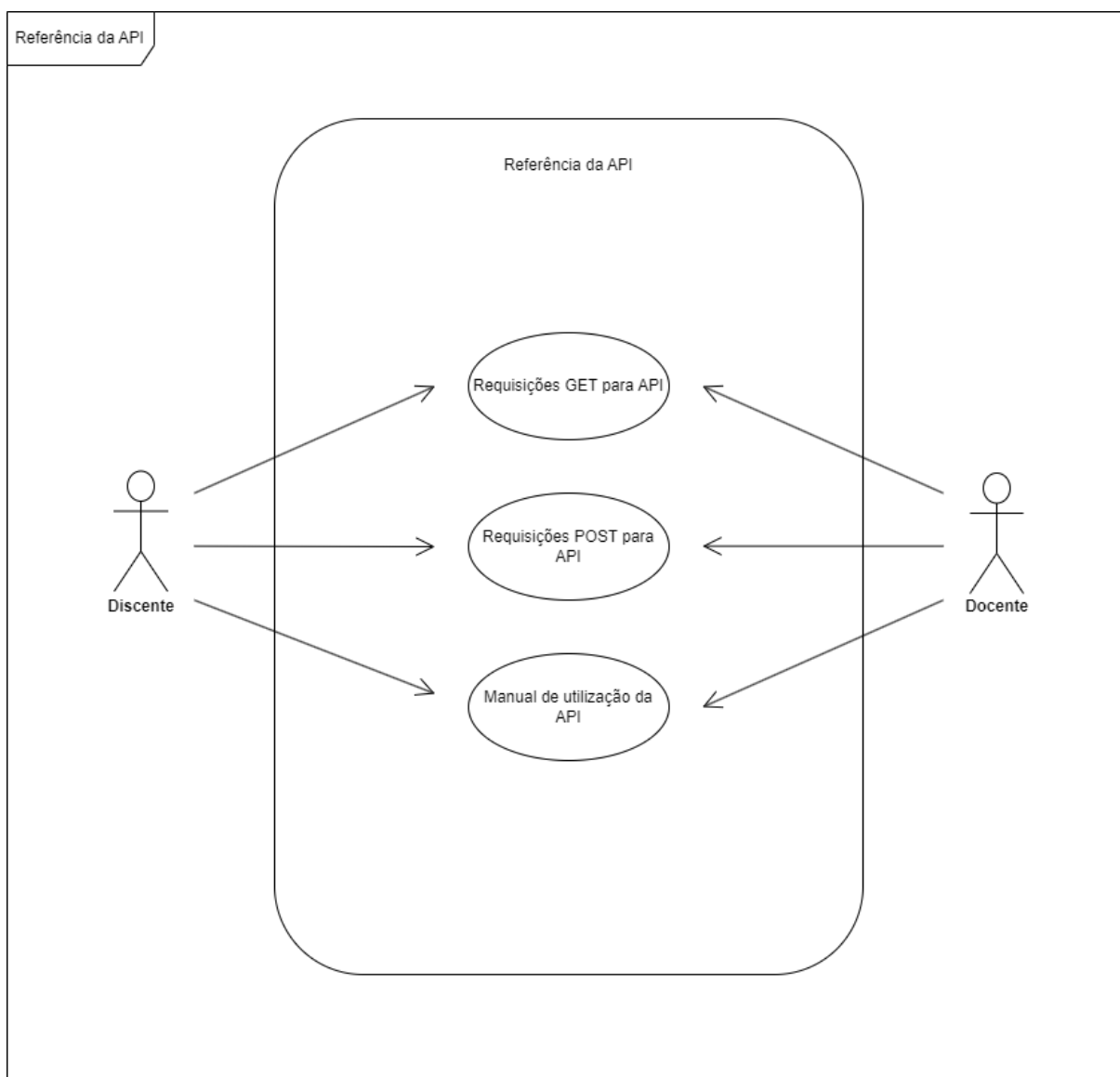
Figura 49 – Diagrama de casos de uso - Grafo de Dependências



Fonte: Elaborado pelo autor no software Draw.io.

A.11 Referência da API

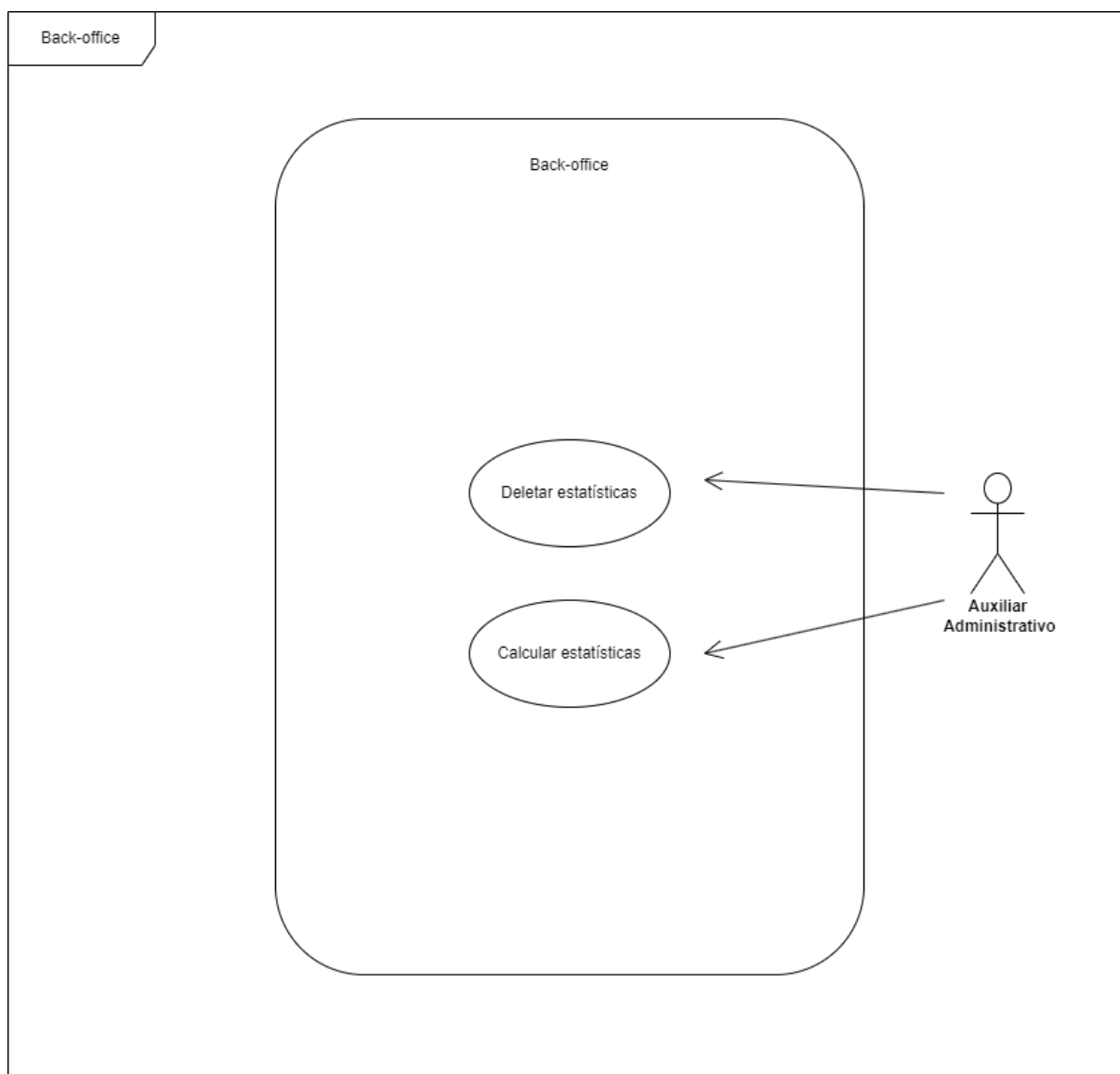
Figura 50 – Diagrama de casos de uso - Referência da API



Fonte: Elaborado pelo autor no software Draw.io.

A.12 Back-office

Figura 51 – Diagrama de casos de uso - Back-office



Fonte: Elaborado pelo autor no software Draw.io.