



Instituto Federal de Brasília
Campus Brasília
Tecnologia em Sistemas para Internet

Felipe Alves Vila Verde Gomes
Geovana Araujo Perazzo

**Análise Temática da Produção Acadêmica da Rede Federal:
Uma Abordagem Baseada em Web Scraping, Processamento de
Linguagem Natural e Visualização de Dados**

Brasília - DF
2026

Felipe Alves Vila Verde Gomes
Geovana Araujo Perazzo

**Análise Temática da Produção Acadêmica da Rede Federal:
Uma Abordagem Baseada em Web Scraping, Processamento de
Linguagem Natural e Visualização de Dados**

Trabalho de Conclusão de Curso apresentado
ao curso de Tecnologia em Sistemas para In-
ternet do Instituto Federal de Brasília do
Campus Brasília, como parte da exigência
para obtenção do título de tecnólogo.

Orientador: Prof. Msc. Gustavo Henrique
Dornelas de Deus
Instituto Federal de Brasília

G633 Gomes, Felipe Alves Vila Verde.

Análise temática da produção acadêmica da Rede Federal: uma abordagem baseada em Web Scraping, Processamento de Linguagem Natural e Visualização de Dados. / Felipe Alves Vila Verde Gomes; Geovana Araujo Perazzo. – Brasília, 2026.
81 f. : il. color.

Orientador: Gustavo Henrique Dornelas de Deus.

Trabalho de conclusão de curso (Graduação) – Instituto Federal de Educação, Ciência e Tecnologia de Brasília, Tecnologia em Sistemas para Internet, 2026.

1. Web scraping. 2. Visualização de dados. 3. Rede Federal. I. Perazzo, Geovana Araujo. II. Deus, Gustavo Henrique Dornelas de. (orient.). III. Título.

CDU 004.912:001.81

Elaborado com os dados fornecidos pelo autor.

Felipe Alves Vila Verde Gomes
Geovana Araujo Perazzo

**Análise Temática da Produção Acadêmica da Rede Federal:
Uma Abordagem Baseada em Web Scraping, Processamento de
Linguagem Natural e Visualização de Dados**

Trabalho de Conclusão de Curso apresentado ao curso de Tecnologia em Sistemas para Internet do Instituto Federal de Brasília do *Campus Brasília*, como parte da exigência para obtenção do título de tecnólogo.

Aprovado em 08 de junho de 2026

BANCA EXAMINADORA

Prof. MSc. Gustavo Henrique Dornelas de Deus
Instituto Federal de Brasília

Prof. MSc. Fernando Wagner Brito Hortencio Filho
Instituto Federal de Brasília

Prof. Dr. Fábio Henrique Monteiro Oliveira
Instituto Federal de Brasília

AGRADECIMENTOS

Felipe Alves Vila Verde Gomes

Agradeço, primeiramente, a todos os membros da minha família que durante toda a caminhada acadêmica me deram muito suporte e compreensão na hora de me dividir para a execução de tudo que me era proposto. Sou muito grato pela oportunidade de concluir esta primeira etapa da vida acadêmica, e que venham muitas outras daqui em diante.

A todos os professores que conheci ao longo dessa jornada, pelo apoio, incentivo, ensinamentos e compreensão durante os momentos de dedicação aos estudos e ao desenvolvimento deste trabalho.

Ao meu orientador, professor MSc. Gustavo Henrique Dornelas de Deus, pela orientação, pela confiança e pelo apoio no momento em que mais precisamos no último semestre. Agradeço pela paciência, incentivo e pelas contribuições fundamentais para a realização desta pesquisa.

Aos colegas e amigos, por compartilharem essa caminhada acadêmica, mesmo que por diferentes momentos e períodos dessa trajetória.

À Rede Federal de Educação Profissional, Científica e Tecnológica e às instituições participantes, pela disponibilização dos dados e repositórios utilizados nesta pesquisa.

Por fim, agradeço a todos que, de alguma forma, puderam contribuir para essa jornada e que, direta ou indiretamente, participaram da realização deste Trabalho de Conclusão de Curso.

Geovana Araujo Perazzo

Primeiramente, agradeço a Deus, por ter sido meu sustento em todos os momentos desta caminhada. Foi pela Sua graça, força e cuidado que consegui superar os desafios, encontrar motivação nos dias difíceis e chegar até aqui com esperança e gratidão.

À minha família, expressei minha mais profunda gratidão por todo amor, apoio, compreensão e incentivo ao longo dessa trajetória. Obrigada por acreditarem em mim mesmo quando eu duvidei das minhas próprias capacidades.

Aos meus amigos, agradeço pela companhia, pelo carinho, pelas palavras de apoio e pelos momentos compartilhados. Cada conversa, incentivo e gesto de amizade tornaram os desafios mais leves e os momentos felizes ainda mais especiais.

Ao meu orientador, professor MSc. Gustavo Henrique Dornelas de Deus, minha sincera gratidão pela orientação, paciência, dedicação e confiança durante o desenvolvimento deste trabalho. Obrigada pelos ensinamentos, pela disponibilidade e pelas contri-

buições fundamentais para minha formação acadêmica. Sua orientação foi essencial para a conclusão desta pesquisa.

Por fim, agradeço a todos que, de alguma forma, contribuíram direta ou indiretamente para a realização deste Trabalho de Conclusão de Curso e para minha trajetória até aqui. Esta conquista também carrega um pouco de cada pessoa que esteve ao meu lado ao longo do caminho.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.”

— Martin Luther King

RESUMO

A crescente disponibilidade de dados acadêmicos em repositórios institucionais e plataformas digitais tem ampliado as possibilidades de análise da produção científica por meio de técnicas computacionais. Nesse contexto, compreender as temáticas predominantes e as tendências de pesquisa em instituições de ensino torna-se relevante para apoiar processos de gestão acadêmica, planejamento científico e definição de políticas institucionais. Trabalhos recentes já demonstraram o potencial do uso de técnicas de Web Scraping e Processamento de Linguagem Natural para identificar padrões temáticos em grandes conjuntos de dados acadêmicos, como no caso da análise de Trabalhos de Conclusão de Curso (TCCs) da Rede Federal de Educação Profissional, Científica e Tecnológica. Diante desse cenário, este trabalho propõe o desenvolvimento de uma abordagem ampliada para análise temática da produção acadêmica da Rede Federal, incorporando não apenas TCCs, mas também artigos científicos, capítulos de livros e projetos institucionais de ensino, pesquisa e extensão. A proposta consiste na coleta automatizada de dados a partir de repositórios institucionais e bases acadêmicas, seguida de processos de extração, transformação e organização dos dados em um ambiente estruturado para análise. A partir desse conjunto de dados, serão aplicadas técnicas de mineração de texto e modelagem de tópicos, com o objetivo de identificar padrões temáticos, tendências de pesquisa e áreas de maior concentração de produção acadêmica. Além disso, pretende-se desenvolver mecanismos de visualização interativa que permitam explorar os resultados obtidos, facilitando a análise da evolução das temáticas ao longo do tempo e a identificação de relações entre diferentes tipos de produção acadêmica. Espera-se que os resultados contribuam para ampliar a compreensão sobre a dinâmica da produção científica na Rede Federal, além de fornecer subsídios para o planejamento institucional e para o desenvolvimento de novas pesquisas. Ao integrar diferentes tipos de produção acadêmica em uma mesma plataforma de análise, o estudo busca oferecer uma visão mais abrangente das áreas de conhecimento, linhas de pesquisa e tendências científicas presentes no contexto da educação profissional e tecnológica no Brasil.

Palavras-chave: Web scraping. Visualização de dados. Rede Federal.

ABSTRACT

The growing availability of academic data in institutional repositories and digital platforms has expanded the possibilities for analyzing scientific production through computational techniques. In this context, understanding predominant themes and research trends within educational institutions becomes relevant to support academic management, scientific planning, and institutional policy development. Recent studies have demonstrated the potential of Web Scraping and Natural Language Processing techniques to identify thematic patterns in large academic datasets, such as analyses of undergraduate theses within the Brazilian Federal Network of Professional, Scientific and Technological Education. In this context, this study proposes an expanded approach for thematic analysis of academic production within the Federal Network, incorporating not only undergraduate theses but also scientific articles, book chapters, and institutional projects related to teaching, research, and extension. The proposed methodology involves automated data collection from institutional repositories and academic databases, followed by extraction, transformation, and structuring processes to organize the data in an analytical environment. Based on this dataset, text mining and topic modeling techniques will be applied to identify thematic patterns, research trends, and areas with higher concentrations of academic production. Additionally, the study aims to develop interactive visualization mechanisms that allow users to explore the results and analyze the evolution of research themes over time, as well as the relationships among different types of academic outputs. The expected results aim to contribute to a broader understanding of the dynamics of scientific production within the Federal Network, providing insights that may support institutional planning and future research initiatives. By integrating multiple types of academic production into a single analytical framework, the study seeks to offer a comprehensive view of research areas, knowledge domains, and emerging scientific trends in the context of professional and technological education in Brazil.

Keywords: Web scraping. Data visualization. Federal Education Network.

Lista de Figuras

1	Interface principal do dashboard desenvolvido por Caixeta (2025)	5
2	Interface principal do sistema desenvolvido por Mota e Jesuino (2025) . . .	6
3	Página inicial do Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr).	8
4	Interface inicial do módulo “Visão Geral”	19
5	Evolução temporal das produções acadêmicas	20
6	Ranking geral por tipo de produção acadêmica	20
7	Análise de temáticas acadêmicas	21
8	Temas dominantes da produção acadêmica	22
9	Módulo de comparação entre servidores	23
10	Comparação visual entre servidores	23
11	Resumo comparativo e ranking entre servidores	24
12	Distribuição geográfica da produção acadêmica	25
13	Ranking de produções acadêmicas por estado	25
14	Visão geral do módulo analítico de TCCs	26
15	Instituições e cursos com maior número de TCCs	27
16	Análise de orientadores	28
17	Detalhamento analítico por orientador	28
18	Ranking completo de orientadores	29
19	Análise institucional dos TCCs	29
20	Análise detalhada das produções acadêmicas	30
21	Distribuição temática de TCCs	30
22	Análise temática dos TCCs	31
23	Análise detalhada por agrupamento temático	32
24	Cursos relacionados ao agrupamento temático selecionado	32
25	Mapa de calor entre temas e cursos relacionados aos TCCs	33
26	Busca avançada e similaridade de TCCs	33
27	Análise de similaridade entre Trabalhos de Conclusão de Curso	34
28	Visão geral do módulo analítico de artigos científicos	35
29	Visão geral do módulo analítico de artigos científicos	35
30	Análise temática e evolução temporal dos artigos científicos	36

31	Análise detalhada por tema em artigos científicos	37
32	Veículos de publicação e distribuição temática por instituição	37
33	Ranking de instituições por produção científica	38
34	Produção anual de artigos científicos por instituição	39
35	Detalhamento da produção científica por instituição	39
36	Análise de produtividade dos servidores em artigos científicos	40
37	Detalhamento da produção científica por servidor	41
38	Ranking completo de servidores por quantidade de artigos publicados . . .	41
39	Funcionalidade de busca de artigos científicos	42
40	Visão geral do módulo analítico de projetos acadêmicos	43
41	Distribuição dos projetos por natureza e instituição	43
42	Tabelas de instituições, financiadores e projetos acadêmicos	44
43	Análise temática dos projetos acadêmicos	44
44	Distribuição por natureza e listagem de projetos do tema selecionado . . .	45
45	Mapa de calor entre temas e instituições no módulo de projetos	46
46	Ranking de instituições por produção de projetos acadêmicos	46
47	Detalhamento de projetos por instituição	47
48	Top servidores por natureza de projeto — Parte 1	47
49	Top servidores por natureza de projeto — Parte 2	48
50	Ranking completo de servidores	48
51	Funcionalidade de busca de projetos acadêmicos	49
52	Projetos similares ao primeiro resultado encontrado	49
53	Diagrama Dimensional da Arquitetura Star Schema do Data Mart do sistema proposto.	62

Lista de Tabelas

1	Comparativo entre os sistemas analisados e o trabalho atual	9
2	Repositórios institucionais da Rede Federal utilizados na coleta de dados .	51
3	Distribuição de artigos por agrupamento temático	53
4	Distribuição de projetos por agrupamento temático	54
5	Distribuição de TCCs por agrupamento temático	55
6	Tópico de maior incidência por instituição	58
7	Estrutura da entidade <i>dim_instituicao</i>	63
8	Estrutura da entidade <i>dim_campus</i>	63
9	Estrutura da entidade <i>dim_curso</i>	63
10	Estrutura da entidade <i>dim_pessoa</i>	63
11	Estrutura da entidade <i>fato_tcc</i>	64
12	Estrutura da entidade <i>ponte_tcc_aluno</i>	64
13	Estrutura da entidade <i>ponte_tcc_orientador</i>	65
14	Estrutura da entidade <i>fato_artigo</i>	65
15	Estrutura da entidade <i>fato_projeto</i>	65

LISTA DE ABREVIATURAS

BI Business Intelligence. 6, 9

CSV Comma-Separated Values. 12, 18

ETL Extract, Transform, Load. 4, 5

LDA Latent Dirichlet Allocation. 4, 9

MER Modelo Entidade Relacionamento. 61

PLN Processamento de Linguagem Natural. 4, 11, 12, 15, 16, 21

RF Requisito Funcional. 16

RNF Requisito Não Funcional. 17

TCC Trabalho de Conclusão de Curso. 1, 9, 12, 15, 16, 19–22, 26, 29, 31

TCCs Trabalhos de Conclusão de Curso. 4, 9, 11

Sumário

1	Introdução	1
1.1	Tema	1
1.2	Problema	1
1.3	Objetivos	2
1.3.1	Objetivo geral	2
1.3.2	Objetivos específicos	2
2	Projetos e sistemas similares	4
2.1	Análise de temas de Trabalhos de Conclusão de Curso na Rede Federal: uma abordagem baseada em web scraping e visualização de dados	4
2.2	Visualização Interativa de Dados Docentes: Integração de Dados via Web Scraping e Business Intelligence	5
2.3	Portal Oasisbr	7
2.4	Comparativo entre projetos similares	8
3	Metodologia	11
3.1	Descrição geral da pesquisa	11
3.2	Instrumentos de pesquisa	12
3.3	Coleta e análise de dados	12
3.4	Fluxo metodológico da solução	13
3.4.1	Etapa 1 – Coleta automatizada de dados acadêmicos	13
3.4.2	Etapa 2 – Processo ETL e consolidação dos dados	13
3.4.3	Etapa 3 – Modelagem dimensional do Data Mart	13
3.4.4	Etapa 4 – Pré-processamento textual	14
3.4.5	Etapa 5 – Modelagem de tópicos e análise temática	14
3.4.6	Etapa 6 – Exportação e estrutura analítica	14
3.4.7	Etapa 7 – Desenvolvimento da interface analítica	14
3.5	Considerações sobre o Data Mart Acadêmico	15
4	Proposta	16
4.1	Especificação de requisitos	16
4.2	Requisitos funcionais	16
4.3	Requisitos não funcionais	17

4.4	Diagrama de casos de uso	17
4.5	Arquitetura da aplicação	17
4.5.1	Tecnologias utilizadas	18
5	Protótipo Funcional	19
5.1	Módulo de Visão Geral	19
5.2	Módulo de Temáticas	21
5.3	Módulo de Comparação entre Servidores	22
5.4	Módulo de Distribuição Geográfica	24
5.5	Módulo Analítico de TCCs	26
5.6	Módulo Analítico de Artigos Científicos	34
5.7	Módulo Analítico de Projetos	42
6	Análise dos resultados	51
6.1	Apresentação e Análise dos Resultados	51
6.1.1	Artigos Científicos	53
6.1.2	Projetos	54
6.1.3	Trabalhos de Conclusão de Curso (TCC)	54
6.2	Análise Comparativa dos Agrupamentos Temáticos	55
6.2.1	Legenda dos Principais Tópicos	56
7	Conclusão	57
	Referências	59
A	Apêndices	61
A.1	Modelagem Dimensional da Arquitetura Star Schema	62
A.2	Dicionário de Dados	63
A.2.1	Tabela: dim_instituicao	63
A.2.2	Tabela: dim_campus	63
A.2.3	Tabela: dim_curso	63
A.2.4	Tabela: dim_pessoa	63
A.2.5	Tabela: fato_tcc	64
A.2.6	Tabela: ponte_tcc_aluno	64
A.2.7	Tabela: ponte_tcc_orientador	65

A.2.8	Tabela: fato_artigo	65
A.2.9	Tabela: fato_projeto	65
B	Reconhecimento de Uso de Inteligência Artificial	66

1 Introdução

1.1 Tema

A Rede Federal de Educação Profissional, Científica e Tecnológica foi instituída pela Lei nº 11.892, de 29 de dezembro de 2008, responsável pela criação dos Institutos Federais de Educação, Ciência e Tecnologia, com atuação voltada ao ensino, pesquisa, extensão e desenvolvimento tecnológico em diferentes regiões do país (BRASIL, 2008). Nesse contexto, a produção acadêmica desenvolvida pelas instituições da Rede Federal representa uma importante fonte de informação científica e tecnológica, possibilitando análises relacionadas às áreas de pesquisa, tendências temáticas e características institucionais.

A proposta consiste em coletar dados acadêmicos provenientes de repositórios institucionais e bases digitais, organizar essas informações em uma base estruturada e aplicar técnicas de mineração de texto para identificar padrões temáticos e tendências de pesquisa presentes nos documentos analisados.

Além disso, o desenvolvimento da solução está alinhado aos princípios de estímulo à pesquisa científica, inovação e desenvolvimento tecnológico previstos no Marco Legal da Ciência, Tecnologia e Inovação, instituído pela Lei nº 13.243, de 11 de janeiro de 2016 (BRASIL, 2016). Adicionalmente, o tratamento e processamento dos dados observam os princípios relacionados ao uso responsável das informações digitais, conforme estabelecido pela Lei Geral de Proteção de Dados Pessoais – LGPD (BRASIL, 2018).

O tema deste trabalho consiste na análise temática da produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica por meio da utilização de técnicas de Web Scraping, Processamento de Linguagem Natural (PLN) e visualização de dados.

Por fim, este trabalho dá continuidade à abordagem proposta por Caixeta (2025), ampliando o escopo da análise ao incorporar diferentes tipos de produção acadêmica, como artigos científicos e projetos, além dos Trabalho de Conclusão de Cursos (TCCs), anteriormente explorados.

1.2 Problema

Embora a produção acadêmica das instituições da Rede Federal esteja amplamente disponível em ambientes digitais, grande parte dessas informações permanece dispersa em diferentes plataformas e repositórios institucionais. Essa fragmentação dificulta a realização de análises integradas, comprometendo a identificação de padrões, tendências e relações entre áreas do conhecimento.

Além disso, a análise manual de grandes volumes de documentos acadêmicos

mostra-se limitada, tanto pelo tempo necessário quanto pela complexidade envolvida na identificação de padrões temáticos e correlações entre diferentes produções.

Outro fator relevante é a ausência de ferramentas que integrem coleta automatizada, análise textual e visualização de dados em um único processo estruturado, o que dificulta a exploração eficiente dessas informações por pesquisadores e gestores acadêmicos.

Diante desse cenário, surge a seguinte questão de pesquisa:

Como técnicas de Web Scraping, Processamento de Linguagem Natural e visualização de dados podem ser utilizadas para identificar e analisar padrões temáticos na produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica?

1.3 Objetivos

A análise da produção acadêmica pode fornecer subsídios relevantes para o planejamento institucional, a formulação de políticas científicas e o acompanhamento da evolução das linhas de pesquisa em instituições de ensino. No entanto, muitas instituições ainda enfrentam dificuldades para explorar de forma sistemática os dados disponíveis em seus repositórios acadêmicos.

Nesse contexto, o uso de técnicas de coleta automatizada de dados e análise textual apresenta-se como uma alternativa promissora para ampliar a compreensão sobre a dinâmica da produção científica. A aplicação dessas técnicas permite não apenas organizar grandes volumes de dados, mas também identificar padrões e tendências que podem apoiar a tomada de decisões estratégicas.

Além disso, este trabalho diferencia-se por propor uma abordagem integrada, que combina múltiplas fontes de dados acadêmicos e diferentes técnicas analíticas em um único fluxo metodológico, permitindo não apenas a identificação de padrões temáticos, mas também uma visão comparativa entre diferentes tipos de produção acadêmica.

1.3.1 Objetivo geral

Desenvolver uma abordagem baseada em Web Scraping, Processamento de Linguagem Natural e visualização de dados para identificar e analisar padrões temáticos na produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica.

1.3.2 Objetivos específicos

- Identificar e coletar dados acadêmicos provenientes de repositórios institucionais e currículos acadêmicos.
- Desenvolver rotinas de Web Scraping para automatizar o processo de coleta de

informações acadêmicas.

- Estruturar e organizar os dados coletados em uma base adequada para análise textual.
- Aplicar técnicas de Processamento de Linguagem Natural e mineração de texto para identificar padrões temáticos nos documentos analisados.
- Desenvolver mecanismos de visualização de dados que permitam explorar e interpretar os resultados obtidos.
- Analisar as principais temáticas e tendências presentes na produção acadêmica da Rede Federal, considerando artigos científicos, Trabalhos de Conclusão de Curso e projetos de pesquisa, ensino, extensão e inovação.

2 Projetos e sistemas similares

Este capítulo apresenta um mapeamento de projetos e sistemas similares que servem como referência para a modelagem desta solução. São analisadas iniciativas focadas na coleta automatizada de dados, no processamento textual e na visualização de indicadores institucionais, destacando as metodologias e ferramentas empregadas em cenários correlatos de análise da produção acadêmica.

2.1 Análise de temas de Trabalhos de Conclusão de Curso na Rede Federal: uma abordagem baseada em web scraping e visualização de dados

Este trabalho tem como principal objetivo analisar as temáticas dos Trabalhos de Conclusão de Curso (TCCs) desenvolvidos nas instituições da Rede Federal de Educação Profissional, Científica e Tecnológica, utilizando técnicas de coleta automatizada de dados e análise computacional. A pesquisa parte da identificação de uma lacuna relacionada à ausência de uma visão consolidada sobre os temas abordados nos TCCs, uma vez que esses dados encontram-se dispersos em diferentes repositórios institucionais.

Para solucionar esse problema, foi desenvolvido um sistema de web scraping assíncrono, responsável por coletar metadados de TCCs (como título, autor, orientador, instituição, ano e resumo) em repositórios digitais de 30 instituições. Ao todo, foram coletados mais de 206 mil registros, dos quais cerca de 81 mil foram validados e estruturados após o processamento.

Os dados coletados passaram por um processo de Extract, Transform, Load (ETL) (Extração, Transformação e Carga), sendo organizados em um modelo dimensional do tipo Star Schema, permitindo melhor estruturação e análise das informações. Na etapa de análise, foram aplicadas técnicas de Processamento de Linguagem Natural (PLN) e mineração de texto, com destaque para o uso do algoritmo Latent Dirichlet Allocation (LDA), responsável por identificar e agrupar os TCCs em 10 tópicos temáticos distintos.

Os resultados permitiram identificar tendências relevantes na produção acadêmica, como o crescimento de determinadas áreas temáticas e a emergência de termos específicos, evidenciando mudanças no foco das pesquisas ao longo do tempo.

Por fim, o projeto culminou no desenvolvimento de um dashboard interativo utilizando Streamlit, que possibilita a exploração dinâmica dos dados por meio de filtros e visualizações, facilitando a análise por pesquisadores, gestores e estudantes. A pesquisa é classificada como aplicada, descritiva e exploratória, com abordagem predominantemente quantitativa, e demonstra que a integração entre web scraping, processamento de dados e visualização é eficaz para transformar dados acadêmicos dispersos em informações

estratégicas.

Figura 1: Interface principal do dashboard desenvolvido por Caixeta (2025)



Fonte: Caixeta (2025).

A Figura 1 apresenta a interface principal do sistema desenvolvido por Caixeta (2025), voltado à análise temática de Trabalhos de Conclusão de Curso da Rede Federal. A aplicação disponibiliza indicadores quantitativos, filtros interativos e visualizações gráficas destinadas à exploração dos agrupamentos temáticos identificados nos dados acadêmicos coletados.

2.2 Visualização Interativa de Dados Docentes: Integração de Dados via Web Scraping e Business Intelligence

O projeto propõe o desenvolvimento de um sistema web voltado à organização, tratamento e visualização de informações acadêmicas de docentes da Rede Federal de Educação Profissional, Científica e Tecnológica. A pesquisa parte do problema da dispersão, desatualização e falta de padronização dos dados docentes, que se encontram distribuídos em diferentes plataformas, como portais institucionais e a site do integra, dificultando análises e o uso estratégico dessas informações.

Para resolver essa questão, o trabalho propõe a construção de um sistema que utiliza técnicas de web scraping para coletar dados acadêmicos diretamente de fontes públicas, integrando essas informações em uma base estruturada e padronizada. Após a coleta, os dados passam por processos de tratamento, limpeza e normalização, garantindo consistência e permitindo comparações entre instituições e docentes. O sistema também aplica conceitos de ETL para organização eficiente das informações.

Um dos principais diferenciais do projeto é a utilização de técnicas de Business Intelligence (BI), com a criação de dashboards interativos que permitem a visualização de indicadores acadêmicos, como produção científica, áreas de atuação, projetos e orientações.

Além disso, o sistema oferece funcionalidades como busca por docentes, visualização de perfis acadêmicos, comparação entre instituições e professores, análise de produtividade científica e identificação de áreas de conhecimento e tendências.

Essas funcionalidades contribuem para facilitar a identificação de parcerias acadêmicas, apoiar a tomada de decisão institucional e promover maior transparência dos dados educacionais.

Metodologicamente, o trabalho é caracterizado como uma pesquisa aplicada, com abordagem prática voltada ao desenvolvimento de um sistema funcional, combinando revisão bibliográfica, análise de sistemas similares e implementação tecnológica.

Por fim, o projeto demonstra que a integração entre web scraping, BI e visualização de dados é eficaz para transformar dados acadêmicos dispersos em informações úteis, contribuindo para a gestão do conhecimento e a integração na Rede Federal.

Figura 2: Interface principal do sistema desenvolvido por Mota e Jesuíno (2025)



Fonte: Mota e Jesuíno (2025).

A Figura 2 apresenta a interface principal do sistema desenvolvido por Mota e Jesuíno (2025), voltado à visualização interativa de dados acadêmicos da Rede Federal. A aplicação disponibiliza funcionalidades relacionadas à busca de docentes, análise de produção científica, rankings acadêmicos e comparação entre professores e instituições,

utilizando dashboards e recursos gráficos para facilitar a interpretação das informações.

2.3 Portal Oasisbr

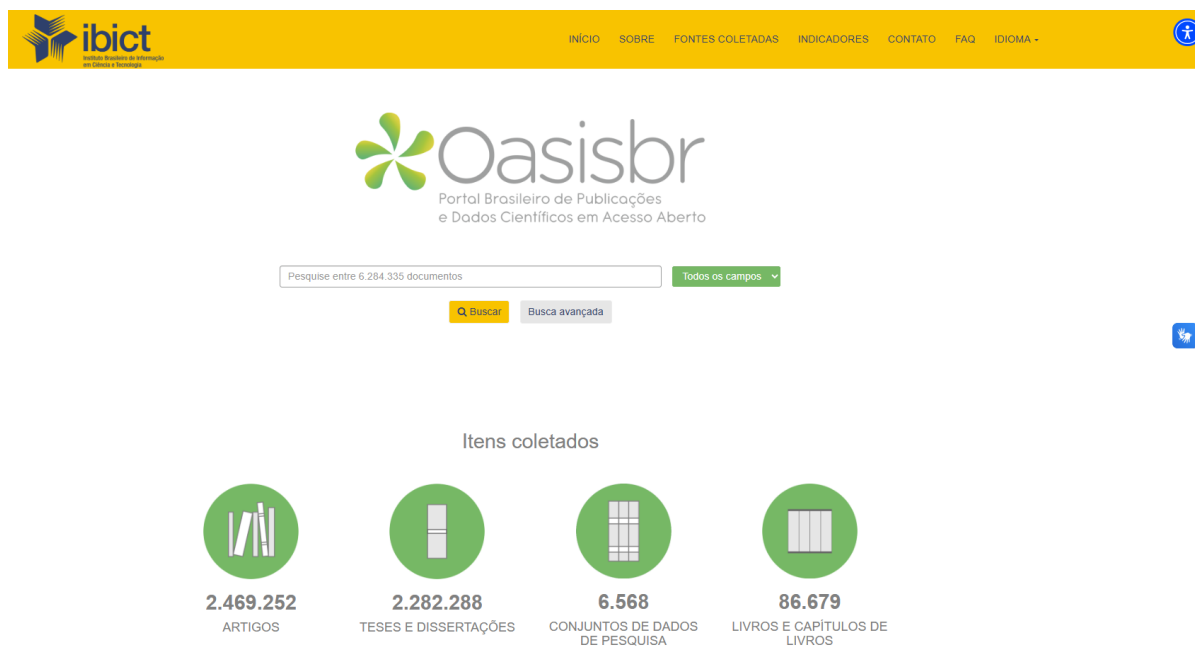
O Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr), desenvolvido e mantido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), reúne a produção científica disponibilizada por repositórios institucionais e outros provedores de dados em acesso aberto (IBICT, 2023). A plataforma constitui uma das principais iniciativas nacionais para integração e disseminação da produção científica em acesso aberto, reunindo metadados provenientes de repositórios institucionais, bibliotecas digitais e revistas científicas brasileiras, permitindo a recuperação unificada de diferentes tipos de documentos acadêmicos.

O Oasisbr disponibiliza mecanismos de busca por título, autor, assunto, instituição e outros metadados, facilitando o acesso à produção científica distribuída em diversas instituições de ensino e pesquisa. A plataforma reúne artigos científicos, teses, dissertações, trabalhos de conclusão de curso, livros, capítulos e outros documentos acadêmicos provenientes de centenas de repositórios brasileiros.

Embora apresente uma interface eficiente para recuperação de documentos e consulta aos metadados das publicações, o Oasisbr possui foco voltado principalmente à pesquisa bibliográfica. A plataforma não oferece recursos de Business Intelligence, dashboards interativos, análises estatísticas ou visualizações gráficas que permitam explorar tendências da produção científica ou realizar comparações entre instituições, autores e áreas do conhecimento.

Nesse contexto, o Oasisbr representa uma importante referência como sistema nacional de agregação e recuperação de informações científicas. Entretanto, diferencia-se da solução proposta neste trabalho por não disponibilizar funcionalidades analíticas capazes de apoiar gestores e pesquisadores na exploração visual dos dados acadêmicos da Rede Federal de Educação Profissional, Científica e Tecnológica.

Figura 3: Página inicial do Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr).



Fonte: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT, 2023).

A Figura 3 apresenta a interface inicial do Portal Oasisbr. É possível observar o mecanismo de busca unificada, que permite realizar pesquisas em diferentes campos do documento, bem como acessar a busca avançada para refinamento das consultas. A página também disponibiliza indicadores referentes ao volume de documentos indexados, como artigos científicos, teses e dissertações, conjuntos de dados de pesquisa e livros, evidenciando a abrangência do acervo disponibilizado pela plataforma.

Embora o Oasisbr seja voltado principalmente para a recuperação e disseminação da produção científica em acesso aberto, oferecendo mecanismos de pesquisa e acesso aos documentos, o sistema desenvolvido neste trabalho amplia essas funcionalidades ao incorporar técnicas de web scraping, processamento de linguagem natural, análise temática e dashboards interativos voltados especificamente para a produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica.

2.4 Comparativo entre projetos similares

Os trabalhos analisados apresentam propostas semelhantes ao utilizarem técnicas computacionais para a coleta, organização, processamento e análise de dados acadêmicos. Em todos os casos, observa-se o emprego de tecnologias voltadas ao tratamento e à recuperação de informações científicas, embora cada solução possua objetivos, escopos e funcionalidades distintas.

O trabalho desenvolvido por **Caixeta (2025)** tem como foco a análise temática de TCC, utilizando técnicas de mineração de texto e Processamento de Linguagem Natural, com destaque para o uso do algoritmo LDA. A proposta busca identificar padrões e tendências nos temas abordados na produção acadêmica da Rede Federal.

Por sua vez, o trabalho de **Mota e Jesuíno (2025)** direciona-se à análise de dados docentes, propondo o desenvolvimento de um sistema web interativo voltado à visualização de informações acadêmicas e indicadores institucionais, com ênfase na utilização de técnicas de BI para apoiar a análise gerencial dos dados.

Já o **Portal Oasisbr**, desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), atua como um agregador nacional de publicações científicas em acesso aberto, reunindo documentos provenientes de diversos repositórios institucionais brasileiros. A plataforma disponibiliza mecanismos de busca unificada e recuperação de informações, permitindo o acesso a artigos científicos, teses, dissertações, trabalhos de conclusão de curso e outros documentos acadêmicos. Entretanto, seu foco concentra-se na recuperação da informação, não oferecendo recursos voltados à análise temática, geração de indicadores ou visualizações analíticas dos dados recuperados.

O presente trabalho diferencia-se por reunir características observadas nas soluções analisadas e ampliá-las em uma única plataforma. Além da coleta automatizada de dados por meio de web scraping, o sistema integra técnicas de processamento de linguagem natural, análise temática e visualização de dados em dashboards interativos. Diferentemente dos trabalhos relacionados e do Portal Oasisbr, a solução proposta contempla simultaneamente múltiplos tipos de produção acadêmica (TCCs, artigos científicos publicados em periódicos e projetos acadêmicos), permitindo realizar análises comparativas entre instituições, servidores, temáticas e tendências da produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica.

Tabela 1: Comparativo entre os sistemas analisados e o trabalho atual

Critério	Caixeta	Mota e Jesuino	Oasisbr	Trabalho Atual
Objeto principal	TCCs	Dados docentes	Produção científica	TCCs, artigos e projetos
Web Scraping	Sim	Sim	Não	Sim
Busca textual	Não	Sim	Sim	Sim
Filtros dinâmicos	Sim	Sim	Sim	Sim
Dashboards interativos	Sim	Sim	Não	Sim
Análise temática	Sim	Não	Não	Sim
Business Intelligence	Parcial	Sim	Não	Sim
Comparação entre instituições	Não	Sim	Não	Sim
Visualizações gráficas	Sim	Sim	Limitadas	Sim
Integração de diferentes produções	Não	Não	Parcial	Sim
Foco na Rede Federal	Sim	Sim	Não	Sim

Fonte: Elaborado pelos autores (2026).

3 Metodologia

Este capítulo apresenta os procedimentos metodológicos utilizados no desenvolvimento deste trabalho, contemplando as etapas de coleta, tratamento, transformação, análise e visualização dos dados acadêmicos da Rede Federal de Educação Profissional, Científica e Tecnológica.

A metodologia adotada dá continuidade ao trabalho desenvolvido por Ana Luísa Caixeta (2025), que estabeleceu uma abordagem inicial para a coleta e análise de produções acadêmicas, com foco nos TCCs. A partir dessa base, o presente estudo amplia o escopo da investigação, incorporando outros tipos de produção acadêmica, como artigos científicos e projetos, além de refinar as técnicas de tratamento, análise e organização dos dados.

A estrutura metodológica foi definida considerando a necessidade de reunir informações acadêmicas dispersas em diferentes repositórios institucionais. Para isso, foram empregadas técnicas de *web scraping* para coleta automatizada dos dados, permitindo a extração estruturada de informações disponíveis em repositórios institucionais e bases digitais (MITCHELL, 2018). Em seguida, foram realizados processos de tratamento e organização dos dados, com o objetivo de garantir a consistência, padronização e confiabilidade das informações obtidas.

Na etapa de análise, foram utilizadas técnicas de PLN para identificação de padrões temáticos nos documentos analisados. Essa abordagem mantém alinhamento com a proposta inicial de Caixeta (2025), porém incorpora aprimoramentos nos critérios de agrupamento e na organização dos resultados. Por fim, os dados foram apresentados por meio de recursos de visualização, permitindo uma interpretação mais clara, estruturada e comparativa dos resultados obtidos.

3.1 Descrição geral da pesquisa

Quanto à finalidade, este trabalho pode ser classificado como uma pesquisa aplicada, pois tem como objetivo desenvolver uma solução prática voltada à análise da produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica, utilizando técnicas computacionais de mineração de dados e processamento textual, conforme definição apresentada por Gil (2008).

Em relação aos objetivos, a pesquisa caracteriza-se como descritiva e exploratória. O caráter descritivo está associado à organização, categorização e análise dos dados coletados, possibilitando a identificação de padrões temáticos e tendências nas produções acadêmicas. Já o aspecto exploratório justifica-se pela análise de um conjunto amplo e heterogêneo de dados, ainda pouco investigados de forma integrada, conforme discutido por Prodanov e Freitas (2013).

Quanto à abordagem, a pesquisa é predominantemente quantitativa, considerando o volume de dados analisados e a utilização de métricas de frequência e distribuição. No entanto, também apresenta aspectos qualitativos, especialmente na interpretação dos agrupamentos temáticos e na análise dos resultados obtidos por meio das técnicas de PLN, caracterizando uma abordagem mista, conforme proposto por Creswell (2010).

Em relação aos procedimentos técnicos, a pesquisa é classificada como documental e bibliográfica. Documental, por utilizar dados provenientes de repositórios institucionais e currículos acadêmicos; e bibliográfica, por se apoiar em estudos relacionados à coleta automatizada de dados, mineração de texto e visualização da informação, conforme Richardson (2017). Além disso, o desenvolvimento de um sistema computacional para coleta, tratamento, armazenamento e análise dos dados caracteriza o trabalho como uma pesquisa tecnológica aplicada.

3.2 Instrumentos de pesquisa

Os instrumentos utilizados nesta pesquisa consistem, predominantemente, em ferramentas computacionais desenvolvidas para as etapas de coleta, processamento e análise dos dados.

Na etapa de coleta, foram implementadas rotinas de *web scraping*, responsáveis por extrair automaticamente informações de repositórios acadêmicos e currículos Lattes, conforme abordagem também adotada por Caixeta (2025).

Para o tratamento dos dados, foram empregadas técnicas de limpeza, remoção de duplicidades, padronização textual e organização em formato estruturado Comma-Separated Values (CSV) e bancos de dados relacionais, possibilitando maior consistência e viabilidade para análise posterior.

Na etapa analítica, foram utilizadas bibliotecas e algoritmos de Processamento de Linguagem Natural (PLN), com o objetivo de identificar padrões semânticos e realizar o agrupamento temático das produções acadêmicas. Técnicas de PLN são amplamente utilizadas em tarefas de mineração textual e análise de documentos, permitindo identificar relações semânticas e padrões linguísticos em grandes volumes de dados textuais (MANNING; SCHÜTZE, 1999).

Por fim, ferramentas de visualização de dados foram utilizadas para representar os resultados de forma gráfica e tabular, facilitando a interpretação das informações, a identificação de tendências e a comparação entre os diferentes tipos de produção analisados.

3.3 Coleta e análise de dados

A coleta de dados foi realizada de forma automatizada por meio de técnicas de *web scraping*, considerando diferentes fontes institucionais da Rede Federal. Os dados coletados abrangem informações sobre Trabalhos de Conclusão de Curso, artigos científicos e projetos, ampliando o escopo metodológico em relação ao trabalho de Caixeta (2025), que se concentrou predominantemente em TCC.

Após a coleta, os dados passaram por um processo de tratamento, envolvendo a remoção de duplicidades, correção de inconsistências e padronização textual. Em seguida, as informações foram organizadas em uma base estruturada, permitindo sua utilização nas etapas subsequentes de análise.

A análise dos dados foi realizada por meio de técnicas quantitativas, como contagem de ocorrências por temática, distribuição de frequências e análise estatística descritiva, e qualitativas, por meio da interpretação dos agrupamentos gerados. As técnicas de Processamento de Linguagem Natural foram fundamentais para identificar padrões nos títulos e descrições das produções, possibilitando a classificação temática dos dados.

Essa abordagem segue a metodologia inicial proposta por Caixeta (2025), sendo, no entanto, ampliada e aprimorada neste trabalho, especialmente no que se refere à diversidade dos dados analisados e à consolidação dos agrupamentos temáticos.

Os resultados obtidos foram organizados em agrupamentos temáticos, posteriormente consolidados em tabelas, gráficos e visualizações interativas, permitindo uma análise comparativa entre os diferentes tipos de produção acadêmica (TCC, artigos e projetos). Essa organização possibilita identificar padrões, tendências e relações entre as áreas do conhecimento, em alinhamento com os objetivos propostos neste trabalho.

3.4 Fluxo metodológico da solução

A arquitetura metodológica da solução foi organizada em sete etapas principais, estruturadas em pipeline computacional automatizado.

-alterado-

3.4.1 Etapa 1 – Coleta automatizada de dados acadêmicos

A primeira etapa do processo consistiu na coleta automatizada de dados acadêmicos provenientes de repositórios institucionais da Rede Federal e currículos acadêmicos da Plataforma Lattes. Para essa finalidade, foram desenvolvidos scripts de web scraping utilizando a linguagem Python, responsáveis pela navigation automatizada, extração e consolidação das informações acadêmicas.

Os algoritmos implementados realizam: 1 — acesso automatizado às páginas institucionais; 2 — identificação de elementos HTML contendo metadados acadêmicos; 3 — extração de títulos, autores, orientadores, instituições, cursos, palavras-chave e descrições textuais; 4 — normalização preliminar dos dados coletados; 5 — armazenamento inicial em banco de dados intermediário denominado *integra.db*. Ao final desta etapa, o montante de registros brutos associados a TCCs, artigos científicos e projetos institucionais foi catalogado e direcionado para a esteira de tratamento.

3.4.2 Etapa 2 – Processo ETL e consolidação dos dados

Após a coleta, os dados passaram por um processo ETL (*Extract, Transform and Load*), responsável pela transformação, padronização e consolidação das informações acadêmicas, visando a criação do ambiente analítico dimensional.

Inicialmente, os dados extraídos do banco *integra.db* foram submetidos a validações institucionais para garantir a aderência estrita dos registros às diretrizes da Rede Federal de Educação Profissional, Científica e Tecnológica. Durante esta etapa, aplicou-se filtros de exclusão para expurgar registros que não pertenciam ao escopo delimitado ou que apresentavam falhas críticas de integridade referencial, como a quebra de chaves estrangeiras relacionadas às dimensões de campus e curso.

Os registros rejeitados foram salvos de forma isolada em arquivos de log específicos, garantindo a rastreabilidade e a auditoria completa da esteira de transformação. Após os procedimentos de filtragem, a base de dados consolidada e livre de inconsistências foi estruturada para compor o Data Mart analítico.

3.4.3 Etapa 3 – Modelagem dimensional do Data Mart

Com os dados consolidados, foi realizada a modelagem dimensional do ambiente analítico utilizando a abordagem *Star Schema*.

O modelo dimensional foi estruturado em tabelas fato e dimensões, permitindo consultas analíticas multidimensionais e integração entre diferentes tipos de produção acadêmica.

Foram desenvolvidas dimensões relacionadas a: instituições; campi; cursos; autores; orientadores; temporalidade; tópicos temáticos.

As tabelas fato foram responsáveis pelo armazenamento das produções acadêmicas, permitindo relacionamentos analíticos entre os diferentes elementos dimensionais. O banco analítico consolidado foi armazenado em *datamart.db*.

3.4.4 Etapa 4 – Pré-processamento textual

Na sequência, os dados textuais passaram por uma etapa de pré-processamento voltada à preparação do corpus textual para análise semântica e modelagem temática.

Os algoritmos implementados realizaram: 1 — remoção de caracteres especiais; 2 — normalização textual; 3 — conversão para caixa baixa; 4 — remoção de *stopwords*; 5 — remoção de registros vazios; 6 — identificação de idioma; 7 — remoção de duplicidades; 8 — tokenização textual; 9 — limpeza semântica e padronização textual dos documentos.

Durante esta etapa, os mecanismos automatizados de higienização linguística operaram de forma iterativa sobre cada natureza de documento. Registros redundantes que compartilhavam conteúdos idênticos e produções cujos textos principais foram identificados em idiomas inválidos ou fora do escopo nacional foram sumariamente expurgados. Essa filtragem foi mandatória para garantir a integridade semântica das matrizes textuais, gerando um corpus limpo e otimizado para a posterior etapa de modelagem de tópicos.

3.4.5 Etapa 5 – Modelagem de tópicos e análise temática

A etapa seguinte consistiu na aplicação de técnicas de Processamento de Linguagem Natural e modelagem de tópicos para identificação automática de agrupamentos temáticos.

Os algoritmos implementados utilizaram a técnica *Latent Dirichlet Allocation* (LDA), permitindo identificar padrões semânticos e agrupamentos temáticos latentes nos títulos e descrições das produções acadêmicas. O processo analítico foi executado separadamente para TCCs, artigos científicos e projetos institucionais.

Para definição da quantidade ideal de tópicos, foram realizados testes iterativos variando a quantidade de agrupamentos temáticos. Os algoritmos avaliaram diferentes configurações considerando métricas de coerência temática. Os melhores resultados identificados foram 11 tópicos para TCCs, 5 tópicos para artigos científicos e 5 tópicos para projetos institucionais. Além disso, foram gerados arquivos auxiliares contendo diagnósticos temáticos e métricas de validação dos modelos.

3.4.6 Etapa 6 – Exportação e estrutura analítica

Após o processamento analítico, os resultados foram exportados em arquivos no formato *Parquet*, visando maior eficiência de leitura e processamento no ambiente web, dividindo os dados entre tópicos, tccs, artigos e projetos. Essa estrutura permitiu otimizar consultas analíticas e reduzir o tempo de carregamento das visualizações interativas.

3.4.7 Etapa 7 – Desenvolvimento da interface analítica

A etapa final consistiu no desenvolvimento da interface web interativa da solução utilizando a plataforma Streamlit.

A aplicação foi estruturada para permitir: 1 - consulta e exploração dos dados acadêmicos; 2 - aplicação de filtros institucionais; 3 - análises comparativas entre produções; 4 - visualizações gráficas; 5 - dashboards temáticos; 6 - exploração de agrupamentos semânticos; 7 - análise temporal da produção acadêmica.

A interface foi desenvolvida visando acessibilidade, interatividade e facilidade de interpretação dos resultados obtidos pelos modelos analíticos implementados.

3.5 Considerações sobre o Data Mart Acadêmico

Com o objetivo de organizar e padronizar os dados coletados dos currículos Lattes e repositórios institucionais da Rede Federal, foi desenvolvido um modelo dimensional para armazenamento e análise das informações acadêmicas.

A estrutura proposta foi organizada em tabelas fato e dimensões utilizando a abordagem *Star Schema*, permitindo integração entre artigos científicos, projetos institucionais e (TCC).

O processo ETL foi responsável pela criação das dimensões relacionadas às instituições, campi, cursos e pessoas, além das tabelas fato responsáveis pelo armazenamento das produções acadêmicas.

Destaca-se ainda a dimensão temática associada aos tópicos identificados pelos modelos de PLN e *Latent Dirichlet Allocation* (LDA), permitindo associar cada produção acadêmica ao agrupamento temático correspondente.

Essa modelagem dimensional favorece análises multidimensionais relacionadas à produção acadêmica, científica e extensionista da Rede Federal, além de possibilitar análises comparativas entre instituições, áreas do conhecimento e tipos de produção acadêmica.

O Modelo Entidade-Relacionamento desenvolvido para o Data Mart acadêmico encontra-se apresentado no Apêndice A.1, permitindo visualizar a estrutura dimensional utilizada na organização dos dados analisados.

4 Proposta

Este capítulo apresenta a proposta de solução desenvolvida neste trabalho, com o objetivo de integrar, processar e analisar a produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica.

A proposta dá continuidade ao trabalho de Caixeta (2025), ampliando o escopo de análise ao incluir não apenas TCC, mas também artigos científicos e projetos. Dessa forma, busca-se oferecer uma visão mais abrangente e integrada da produção acadêmica, permitindo identificar padrões e tendências de forma mais consistente.

A solução proposta é baseada na integração de técnicas de *web scraping*, processamento de dados, PLN e visualização de informações, estruturadas em uma arquitetura modular.

4.1 Especificação de requisitos

A especificação de requisitos tem como finalidade definir as funcionalidades e características que o sistema deve atender, considerando os objetivos do trabalho.

O sistema proposto tem como principal finalidade coletar dados acadêmicos de fontes públicas, tratá-los, analisá-los e disponibilizá-los de forma estruturada e visualmente compreensível.

Os requisitos foram definidos com base nas necessidades identificadas ao longo da pesquisa, especialmente relacionadas à dificuldade de análise integrada da produção acadêmica da Rede Federal, conforme discutido na introdução.

4.2 Requisitos funcionais

O Requisito Funcional (RF), descreve as funcionalidades que o sistema deve oferecer ao usuário. Para a solução proposta, destacam-se:

- RF01 - Coletar automaticamente dados acadêmicos de repositórios institucionais e da Plataforma Lattes por meio de técnicas de *web scraping*;
- RF02 - Armazenar os dados coletados em formato estruturado, permitindo sua manipulação e análise;
- RF03 - Realizar o tratamento dos dados, incluindo limpeza, padronização e remoção de duplicidades;
- RF04 - Classificar os dados de acordo com o tipo de produção acadêmica (TCC, artigos e projetos);
- RF05 - Aplicar técnicas de Processamento de Linguagem Natural para identificar padrões temáticos;
- RF06 - Agrupar os dados em temáticas com base na análise textual;
- RF07 - Gerar quadros e visualizações que representem a distribuição temática dos dados;
- RF08 - Permitir a análise comparativa entre diferentes tipos de produção acadêmica;
- RF09 - Disponibilizar os resultados para consulta e exploração por meio de interface visual.

4.3 Requisitos não funcionais

O Requisito Não Funcional (RNF), descreve as características de qualidade do sistema, relacionadas ao seu desempenho, usabilidade e confiabilidade.

- RNF01 - O sistema deve ser capaz de processar grandes volumes de dados acadêmicos;
- RNF02 - A coleta de dados deve ser realizada de forma automatizada e eficiente;
- RNF03 - Os dados devem ser armazenados de forma padronizada e consistente;
- RNF04 - O sistema deve apresentar resultados de forma clara e intuitiva;
- RNF05 - A solução deve permitir escalabilidade para inclusão de novas fontes de dados;
- RNF06 - O sistema deve garantir reprodutibilidade das análises realizadas;
- RNF07 - O tempo de resposta para geração das análises deve ser adequado ao volume de dados processado.

4.4 Diagrama de casos de uso

O diagrama de casos de uso representa as principais interações entre o usuário e o sistema proposto.

De forma geral, o sistema possui como ator principal o usuário (pesquisador ou gestor acadêmico), que interage com as funcionalidades de coleta, análise e visualização dos dados.

Entre os principais casos de uso, destacam-se:

- Realizar coleta de dados acadêmicos;
- Processar e tratar dados;
- Executar análise temática;
- Visualizar resultados;
- Comparar produções acadêmicas;
- Explorar dados por meio de filtros.

O diagrama de casos de uso sintetiza essas interações, evidenciando o fluxo principal do sistema desde a coleta dos dados até a apresentação dos resultados.

4.5 Arquitetura da aplicação

A arquitetura da aplicação foi projetada de forma modular, visando facilitar a manutenção, escalabilidade e evolução do sistema.

A solução é composta pelos seguintes módulos principais:

- **Módulo de Coleta:** responsável pela extração automatizada de dados por meio de técnicas de *web scraping*;
- **Módulo de Tratamento:** realiza a limpeza, padronização e organização dos dados coletados;
- **Módulo de Análise:** aplica técnicas de Processamento de Linguagem Natural para identificação de padrões e agrupamentos temáticos;

- **Módulo de Visualização:** apresenta os resultados por meio de gráficos, tabelas e dashboards interativos.

Essa arquitetura segue o fluxo definido na metodologia, garantindo coerência entre as etapas do trabalho e permitindo a transformação de dados brutos em informações estruturadas.

Além disso, a modularização da solução permite a integração com novas fontes de dados e a evolução das técnicas de análise, tornando o sistema adaptável a diferentes contextos de aplicação.

4.5.1 Tecnologias utilizadas

A implementação da solução proposta utiliza tecnologias alinhadas às adotadas por Caixeta (2025), garantindo consistência metodológica e viabilidade técnica.

Para a etapa de coleta de dados, são empregadas linguagens e bibliotecas voltadas à automação de extração de informações, com destaque para o uso de *web scraping* em ambiente Python.

No tratamento e processamento dos dados, são utilizadas bibliotecas para manipulação de dados estruturados, permitindo a realização de etapas de limpeza, padronização e organização em formato CSV.

A análise textual é realizada por meio de bibliotecas de Processamento de Linguagem Natural, responsáveis pela extração de termos relevantes e identificação de padrões temáticos nos dados coletados.

Para a visualização dos resultados, é utilizada uma abordagem baseada em dashboards interativos, inspirada na solução desenvolvida por Caixeta (2025), permitindo a exploração dinâmica dos dados por meio de gráficos e filtros.

Essa combinação de tecnologias possibilita a construção de uma solução integrada, capaz de transformar dados acadêmicos dispersos em informações estruturadas e visualmente acessíveis.

5 Protótipo Funcional

A interface da plataforma foi estruturada em módulos analíticos independentes, permitindo a navegação entre diferentes áreas voltadas à exploração de TCCs, artigos científicos e projetos da Rede Federal de Educação Profissional, Científica e Tecnológica. Cada módulo foi desenvolvido com foco em uma camada específica de análise, integrando indicadores quantitativos, visualizações gráficas e mecanismos interativos de filtragem.

Além da organização visual, a aplicação mantém integração dinâmica entre filtros e parâmetros de consulta, permitindo que as análises sejam realizadas de forma contínua entre os diferentes módulos do sistema. Essa abordagem favorece maior interatividade, flexibilidade analítica e facilidade de interpretação dos dados acadêmicos processados pela plataforma.

As subseções a seguir apresentam os principais módulos analíticos implementados na solução.

5.1 Módulo de Visão Geral

O módulo de Visão Geral foi desenvolvido com o objetivo de apresentar uma síntese inicial da produção acadêmica analisada pela plataforma, reunindo indicadores quantitativos relacionados aos TCCs, artigos científicos e projetos da Rede Federal.

Nessa seção, o usuário pode consultar os totais consolidados de cada categoria de produção acadêmica, permitindo uma visão ampla do volume de dados disponível no sistema. Além disso, a interface disponibiliza gráficos e recursos visuais que auxiliam na exploração das informações, permitindo acompanhar tendências temporais, distribuições institucionais e padrões gerais identificados nos dados analisados.

A Figura 4 apresenta a interface principal do módulo de Visão Geral da plataforma.

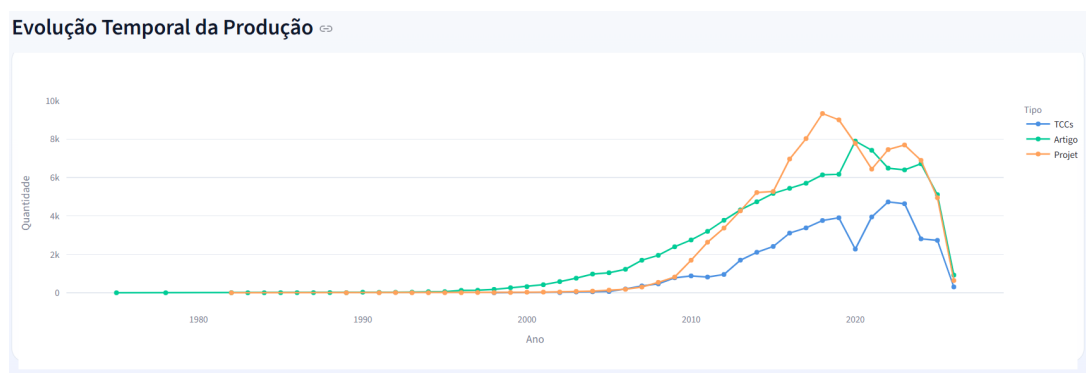
Figura 4: Interface inicial do módulo “Visão Geral”



Fonte: Elaborado pelos autores.

A Figura 5 apresenta a evolução temporal da produção acadêmica. O gráfico apresenta a distribuição anual da quantidade de TCCs, artigos científicos publicados em periódicos e projetos acadêmicos da Rede Federal, abrangendo o período de 1975 a 2026.

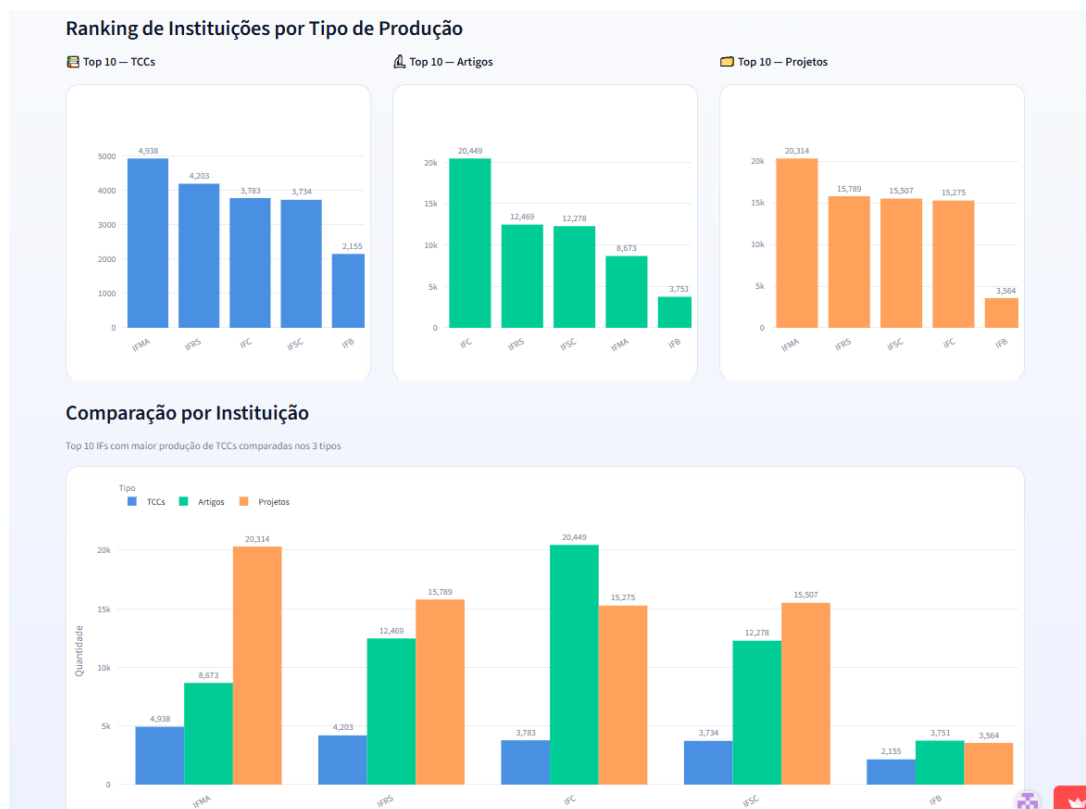
Figura 5: Evolução temporal das produções acadêmicas



Fonte: Elaborado pelos autores.

A Figura 6 apresenta o ranking das instituições por tipo de produção e comparação da produção acadêmica entre instituições da Rede Federal. Na parte superior da figura são apresentados os rankings das dez instituições com maior número de registros em cada categoria de produção acadêmica: TCCs, artigos científicos e projetos acadêmicos. Na parte inferior, é exibido um gráfico comparativo que reúne, para cada instituição do ranking, a quantidade de registros em cada tipo de produção, permitindo visualizar as diferenças de desempenho institucional entre as categorias analisadas. Para a categoria de artigos científicos, foram considerados exclusivamente os artigos publicados em periódicos científicos.

Figura 6: Ranking geral por tipo de produção acadêmica



Fonte: Elaborado pelos autores.

5.2 Módulo de Temáticas

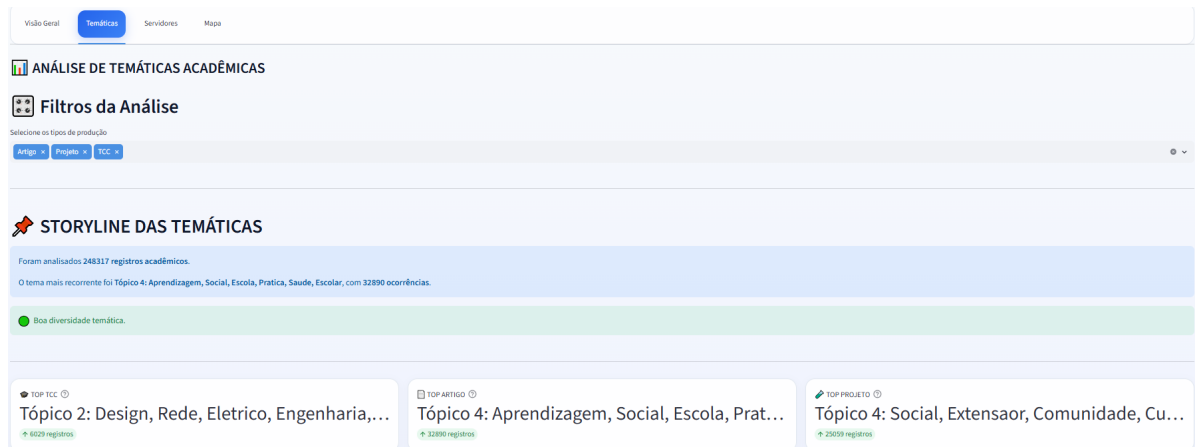
A plataforma também disponibiliza visualizações voltadas à análise dos agrupamentos temáticos identificados nos dados coletados. O módulo de Temáticas foi desenvolvido com o objetivo de explorar os principais tópicos recorrentes nas produções acadêmicas presentes na base analisada.

Nessa seção, o usuário pode aplicar filtros relacionados aos diferentes tipos de produção, como artigos científicos, projetos e TCCs, permitindo refinar a análise de acordo com o interesse da pesquisa. Além dos mecanismos de filtragem, a interface apresenta indicadores quantitativos e informações resumidas sobre os tópicos predominantes identificados por meio de técnicas de PLN e modelagem de tópicos.

A partir dessas visualizações, é possível identificar padrões temáticos recorrentes e compreender como determinados assuntos se distribuem entre os diferentes tipos de produção analisados.

A Figura 7 apresenta a interface inicial do módulo de análise temática, na qual o usuário pode selecionar, por meio dos filtros, um ou mais tipos de produção acadêmica (TCCs, artigos científicos e/ou projetos acadêmicos), permitindo que as análises sejam realizadas de forma individual ou combinada. Com base na seleção realizada, o sistema atualiza automaticamente as informações exibidas, apresentando o total de registros analisados, o tema mais recorrente no conjunto de dados e um resumo da distribuição temática. Os cartões Top TCC, Top Artigo e Top Projeto destacam, respectivamente, a temática com maior número de ocorrências entre os Trabalhos de Conclusão de Curso, os artigos científicos e os projetos acadêmicos, bem como a quantidade de registros associados a cada uma dessas temáticas.

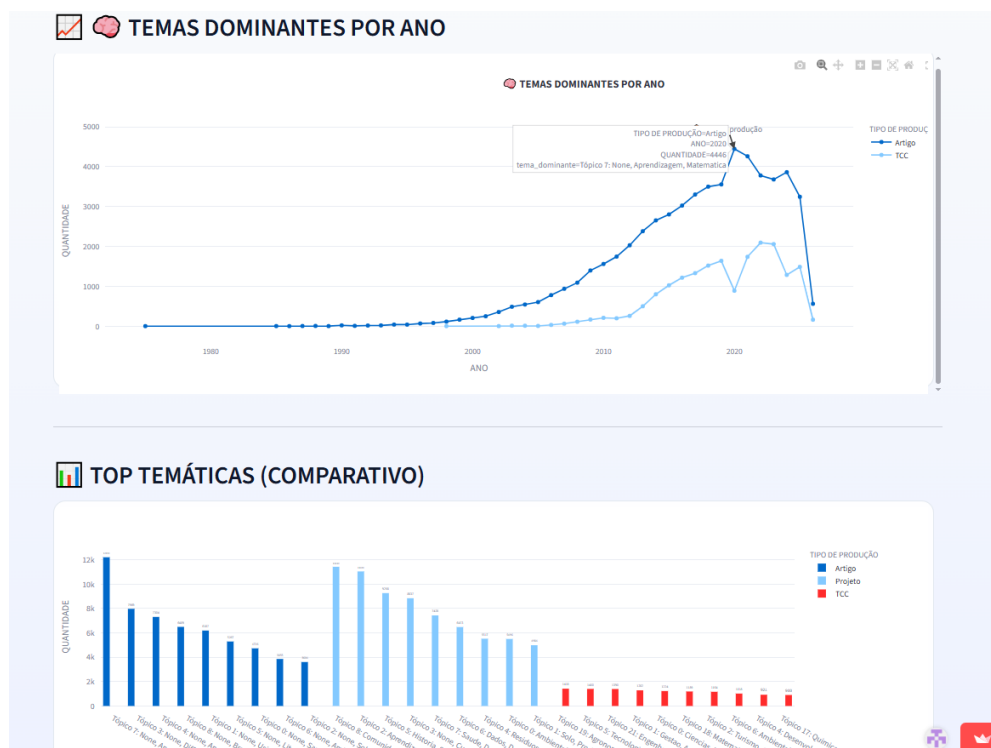
Figura 7: Análise de temáticas acadêmicas



Fonte: Elaborado pelos autores.

A Figura 8 apresenta evolução dos temas dominantes por ano e comparação das principais temáticas das produções acadêmicas. Na parte superior da figura é apresentado o gráfico Temas Dominantes por Ano, obtido a partir da seleção dos filtros para TCCs e artigos científicos, permitindo visualizar a evolução anual da quantidade de registros associados ao tema predominante em cada tipo de produção. Na parte inferior, o gráfico Top Temáticas (Comparativo) apresenta as temáticas com maior número de ocorrências em cada categoria de produção acadêmica (TCCs, artigos e projetos acadêmicos), possibilitando a comparação da frequência das principais temáticas entre os três tipos de produção.

Figura 8: Temas dominantes da produção acadêmica



Fonte: Elaborado pelos autores.

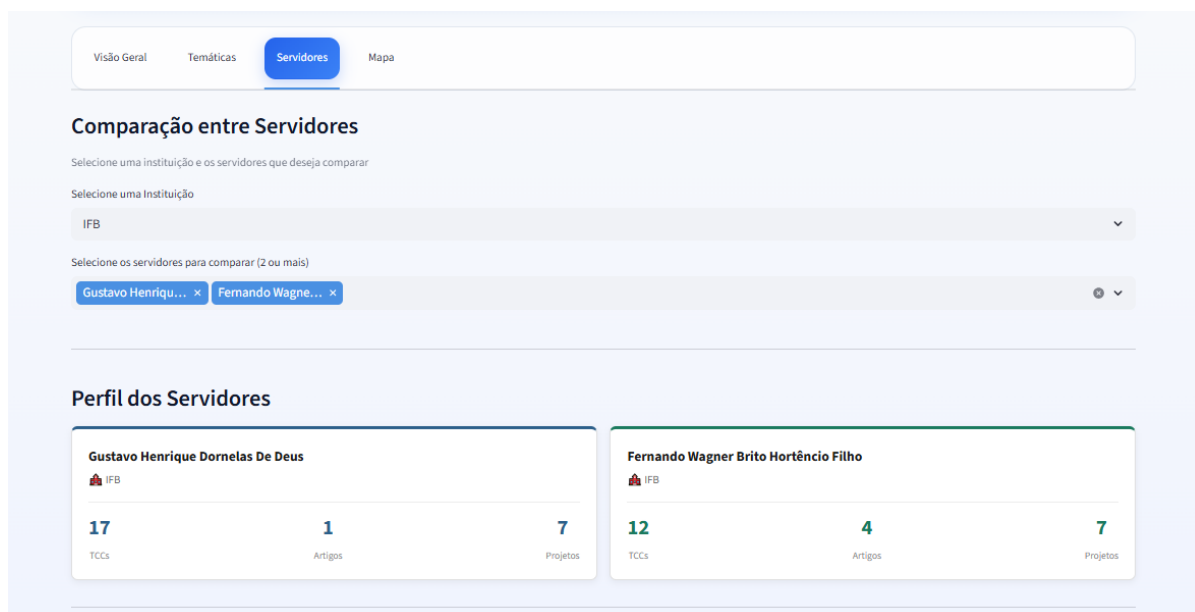
5.3 Módulo de Comparação entre Servidores

O módulo de Comparação entre Servidores foi desenvolvido para possibilitar análises comparativas da produção acadêmica vinculada aos servidores presentes na base de dados.

Nessa seção, o usuário pode selecionar instituições e comparar diferentes servidores com base em indicadores relacionados à quantidade de TCCs orientados, artigos publicados e projetos desenvolvidos. Além disso, a interface apresenta informações resumidas sobre os servidores selecionados, juntamente com gráficos comparativos que auxiliam na identificação de padrões de atuação acadêmica e diferenças entre os perfis analisados.

A Figura 9 apresenta a funcionalidade de comparação entre servidores do sistema, utilizando os filtros para seleção da instituição IFB e dos servidores Gustavo Henrique Dornelas de Deus e Fernando Wagner Brito Hortêncio Filho. A partir da seleção realizada, são exibidos os perfis individuais dos servidores, contendo a instituição de vínculo e a quantidade de produções acadêmicas registradas em cada categoria (TCCs, artigos científicos publicados em periódicos e projetos acadêmicos), permitindo a comparação direta entre seus respectivos perfis de produção.

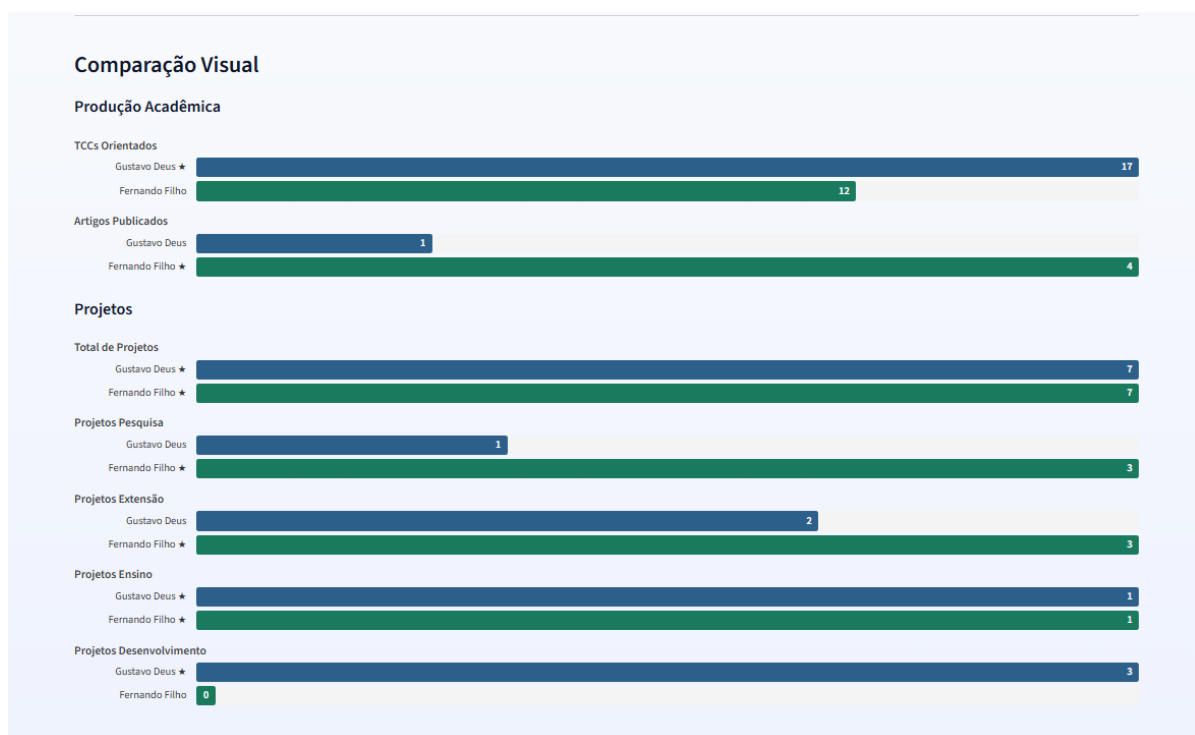
Figura 9: Módulo de comparação entre servidores



Fonte: Elaborado pelos autores.

A Figura 10 apresenta a visualização comparativa implementada no sistema.

Figura 10: Comparação visual entre servidores

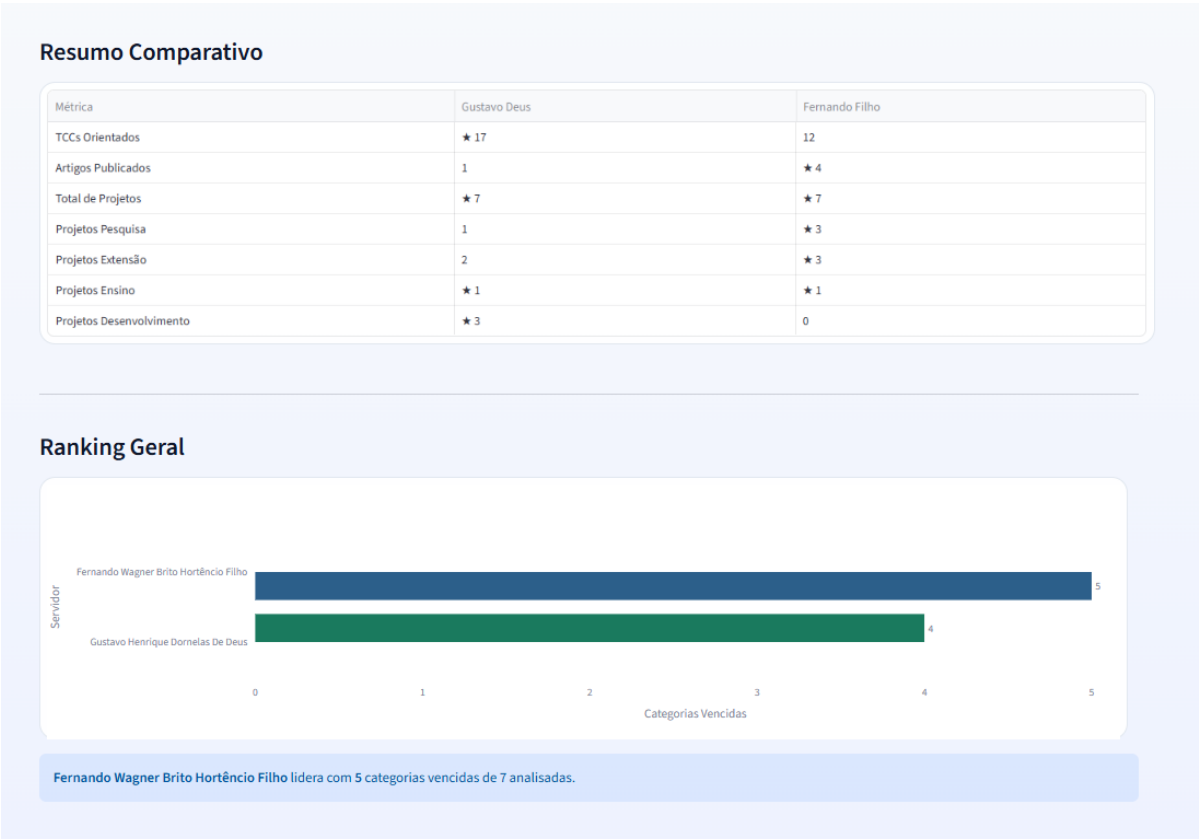


Fonte: Elaborado pelos autores.

Além da comparação inicial entre servidores, o sistema também disponibiliza recursos destinados à análise comparativa detalhada dos indicadores acadêmicos identificados para cada docente selecionado.

A Figura 11 apresenta uma visualização comparativa entre os agrupamentos temáticos identificados.

Figura 11: Resumo comparativo e ranking entre servidores



Fonte: Elaborado pelos autores.

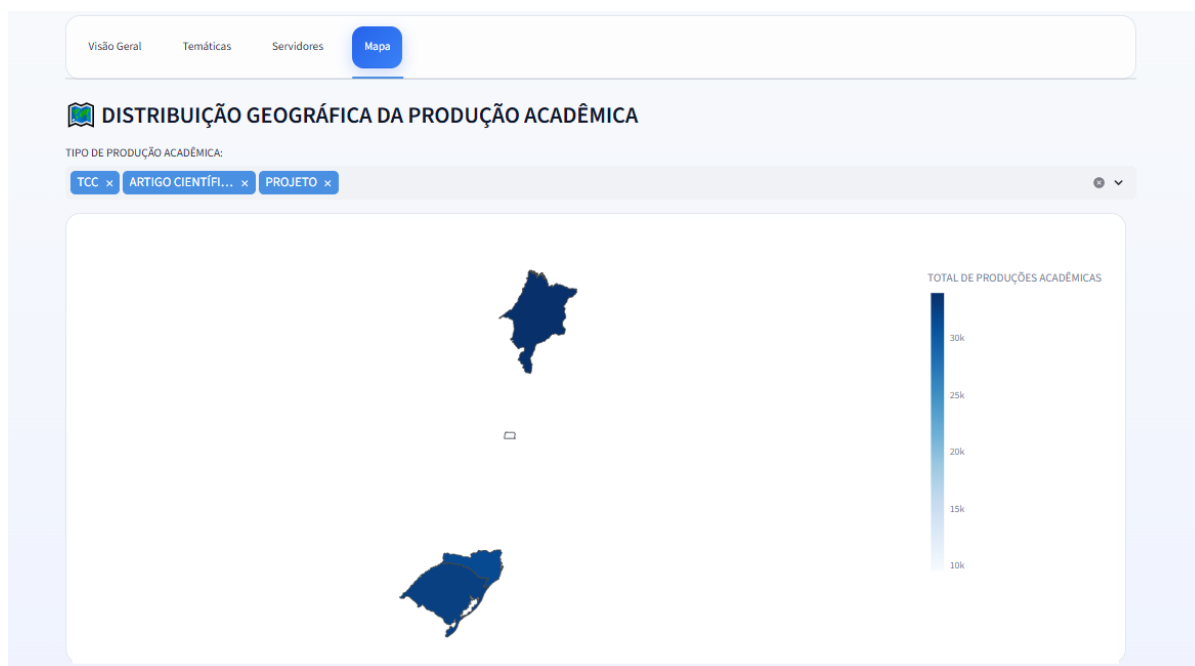
5.4 Módulo de Distribuição Geográfica

O módulo de distribuição geográfica foi desenvolvido para visualizar a concentração de produções acadêmicas nos diferentes estados da Rede Federal.

A interface permite selecionar diferentes tipos de produção acadêmica, possibilitando analisar a distribuição espacial dos registros presentes na base de dados. O mapa interativo utiliza recursos visuais para representar quantitativamente a produção acadêmica por unidade federativa, facilitando a identificação de regiões com maior concentração de registros.

A Figura 12 apresenta a distribuição geográfica das produções acadêmicas por instituição da Rede Federal, utilizando o filtro para exibição conjunta dos três tipos de produção: Trabalhos de Conclusão de Curso (TCCs), artigos científicos e projetos acadêmicos. No mapa, são destacados os estados onde estão localizadas as instituições analisadas (IFB, IFC, IFMA, IFRS e IFSC), e a intensidade da coloração representa o volume total de produções acadêmicas registradas em cada estado. Essa visualização permite identificar a distribuição espacial da produção e comparar a concentração de registros entre as diferentes regiões contempladas pelo estudo.

Figura 12: Distribuição geográfica da produção acadêmica

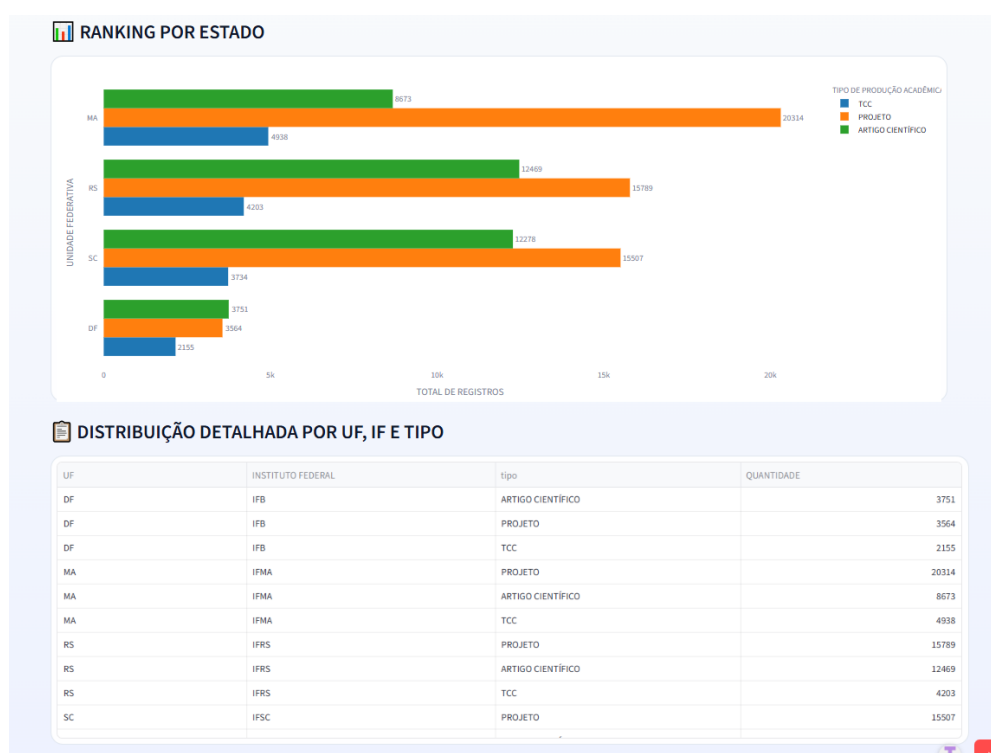


Fonte: Elaborado pelos autores.

Complementando a análise espacial, a plataforma apresenta rankings estaduais relacionados às produções acadêmicas analisadas.

A Figura 13 apresenta o ranking de produções acadêmicas por estado.

Figura 13: Ranking de produções acadêmicas por estado



Fonte: Elaborado pelos autores.

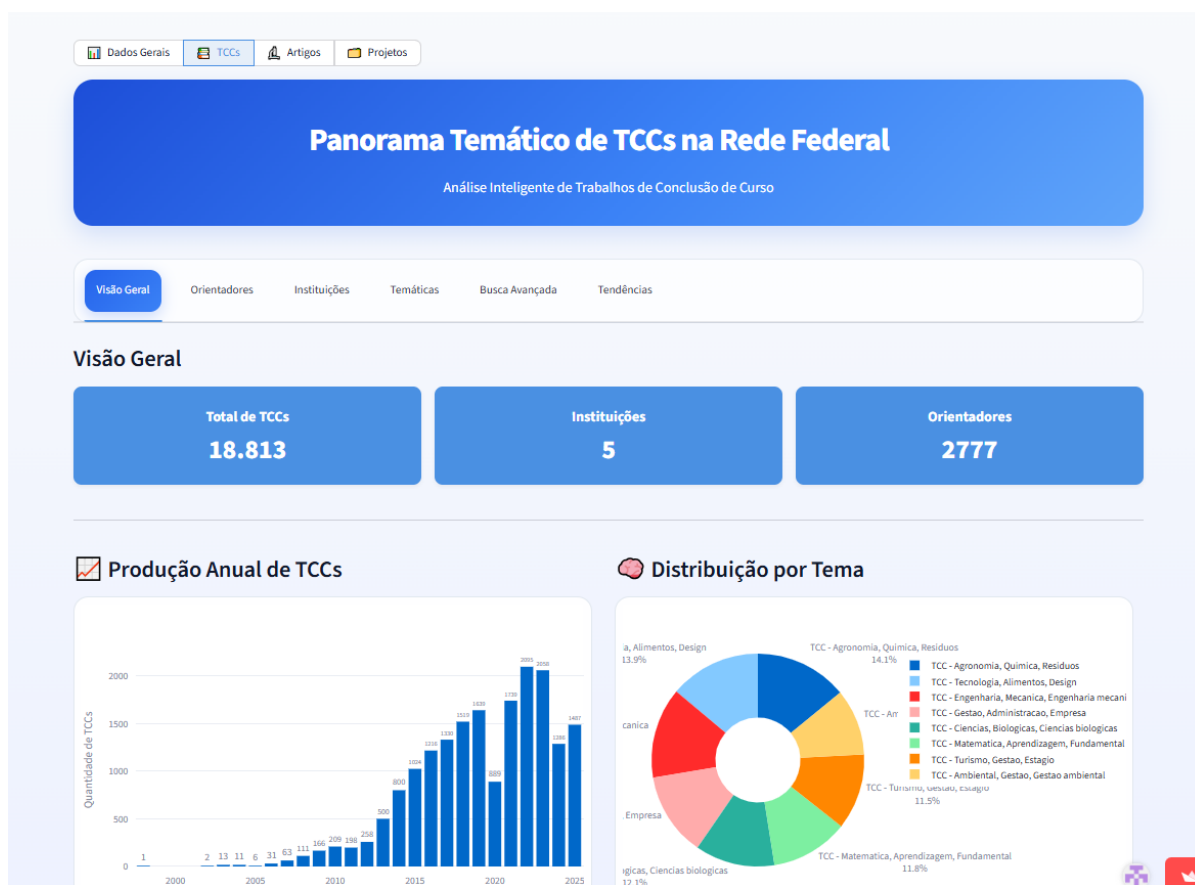
5.5 Módulo Analítico de TCCs

A aba específica de TCCs foi desenvolvida para disponibilizar indicadores analíticos detalhados relacionados às produções classificadas como Trabalhos de Conclusão de Curso.

A interface apresenta indicadores quantitativos gerais, como total de TCCs, quantidade de instituições analisadas e número de orientadores identificados na base de dados. Além disso, gráficos analíticos auxiliam na interpretação das informações, destacando tendências temporais, distribuição temática e padrões institucionais.

A Figura 14 apresenta a visão geral do módulo analítico de TCCs.

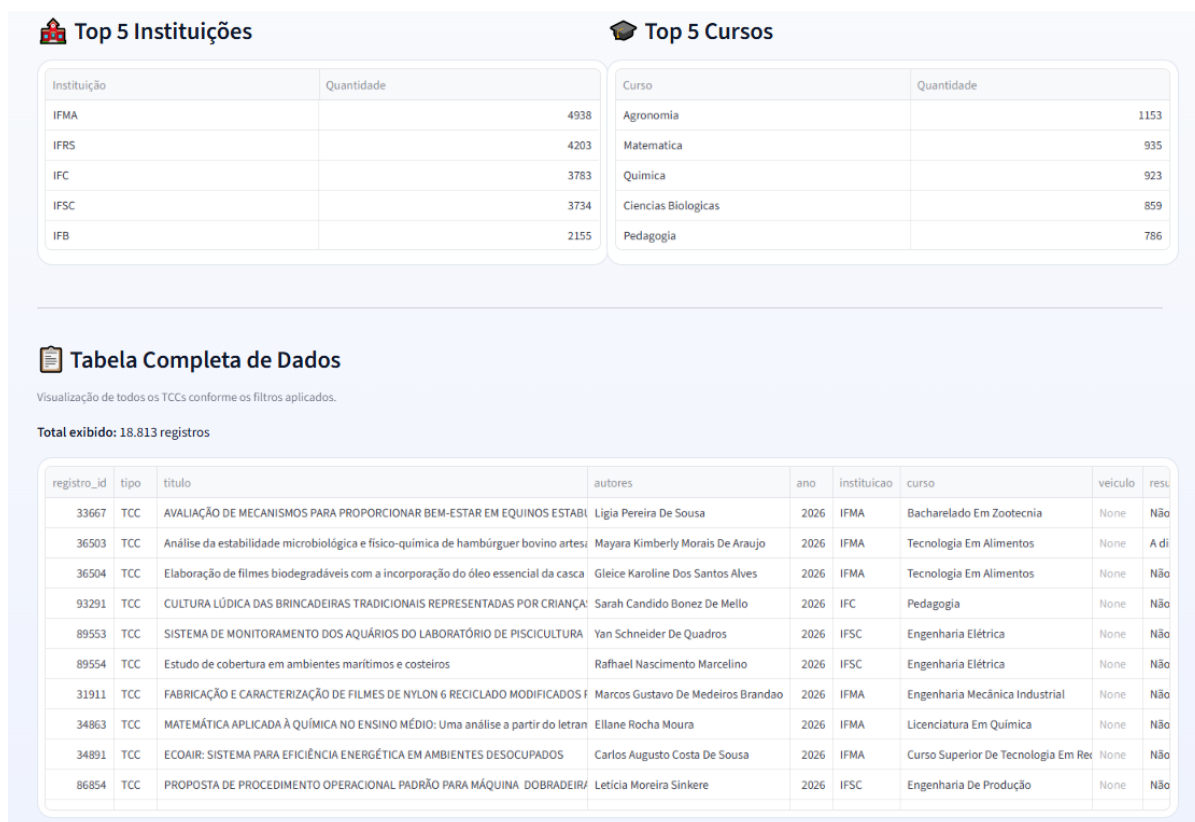
Figura 14: Visão geral do módulo analítico de TCCs



Fonte: Elaborado pelos autores.

A Figura 15 apresenta, sem aplicação de filtros, o Top 5 das instituições com maior quantidade de Trabalhos de Conclusão de Curso (TCCs) cadastrados, bem como o Top 5 dos cursos com maior número de trabalhos registrados. Além dos rankings, é exibida a tabela completa de dados, contendo todos os TCCs disponíveis na base, com informações como título, autores, ano de publicação, instituição, curso e demais atributos associados. Essa visualização possibilita identificar as instituições e os cursos com maior produção de TCCs, além de permitir a consulta detalhada dos registros que compõem a base analisada.

Figura 15: Instituições e cursos com maior número de TCCs



Fonte: Elaborado pelos autores.

Além das análises gerais, a plataforma também disponibiliza módulos voltados à análise dos orientadores vinculados às produções acadêmicas presentes na base de dados.

Nessa seção, são apresentados indicadores quantitativos relacionados ao número de orientações registradas, médias de orientações por docente e distribuição dos orientadores com maior volume de produções identificadas no sistema.

A Figura 16 apresenta os indicadores relacionados aos orientadores identificados no sistema.

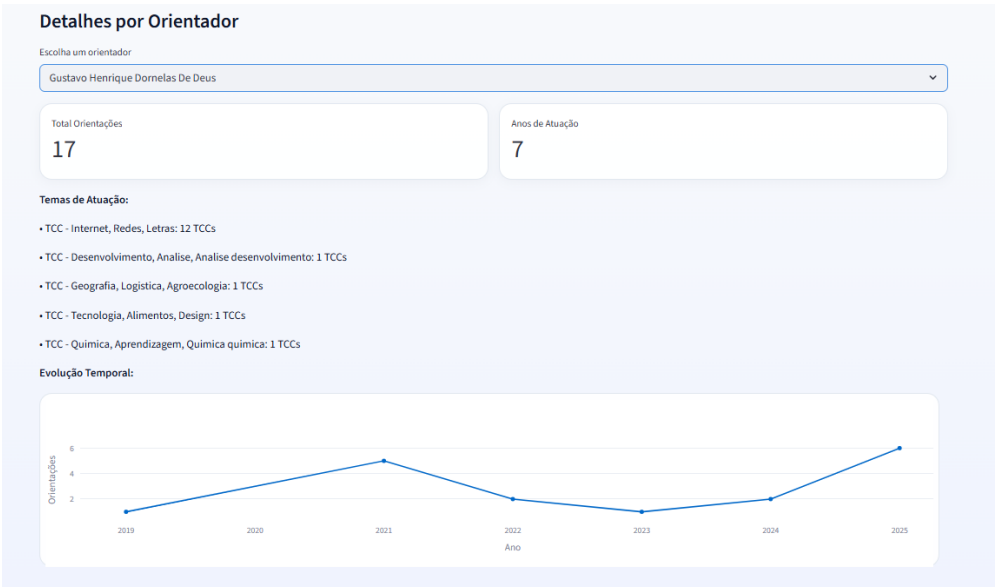
Figura 16: Análise de orientadores



Fonte: Elaborado pelos autores.

A Figura 17 apresenta os detalhes analíticos por orientador.

Figura 17: Detalhamento analítico por orientador



Fonte: Elaborado pelos autores.

A Figura 18 apresenta um ranking consolidado dos orientadores identificados na plataforma, relacionando a quantidade de orientações realizadas e os principais temas associados às produções

acadêmicas. A funcionalidade permite identificar áreas temáticas predominantes e docentes com maior participação na orientação de Trabalhos de Conclusão de Curso.

Figura 18: Ranking completo de orientadores

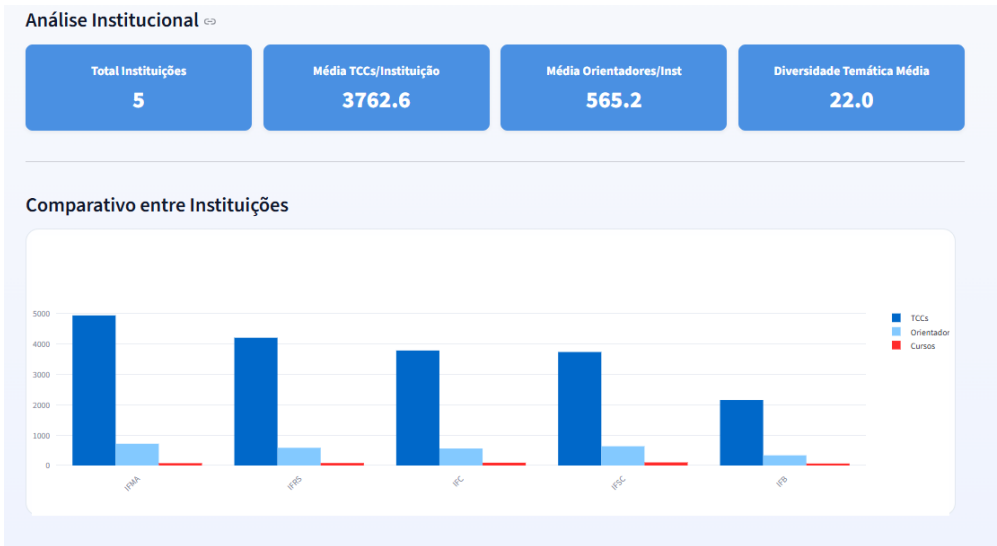
Ranking Completo de Orientadores		
Orientador	Orientações	Tema Principal
Ronaldo Ribeiro Corrêa	103	TCC - Engenharia, Mecânica, Engenharia mecânica
Simão Alberto	61	TCC - Ciências, Biológicas, Ciências biológicas
Fernanda Zorzi	60	TCC - Matemática, Aprendizagem, Fundamental
Yrla Nívea Oliveira Magalhães	59	TCC - Ciências, Biológicas, Ciências biológicas
Alexandra Goede De Souza	57	TCC - Engenharia, Arquitetura, Agronomia
Carina Faccio	54	TCC - Tecnologia, Alimentos, Design
Marcelo Moizinho Oliveira	53	TCC - Química, Aprendizagem, Química química
Fábio Henrique Silva Sales	51	TCC - Física, Física física, Deficiência
Josilene Lima Serra Pereira	50	TCC - Tecnologia, Alimentos, Design
Rubípara Cavalcante Fernandes	48	TCC - Elétrica, Energia, Engenharia

Fonte: Elaborado pelos autores.

Destaca-se, entre os resultados apresentados, o orientador Ronaldo Ribeiro Corrêa, identificado como o docente com maior número de orientações registradas na base analisada, totalizando 103 orientações. Observa-se também predominância de produções relacionadas à área de Engenharia Mecânica, evidenciando forte atuação temática nesse campo do conhecimento.

A Figura 19 apresenta os indicadores institucionais relacionados aos TCCs analisados.

Figura 19: Análise institucional dos TCCs

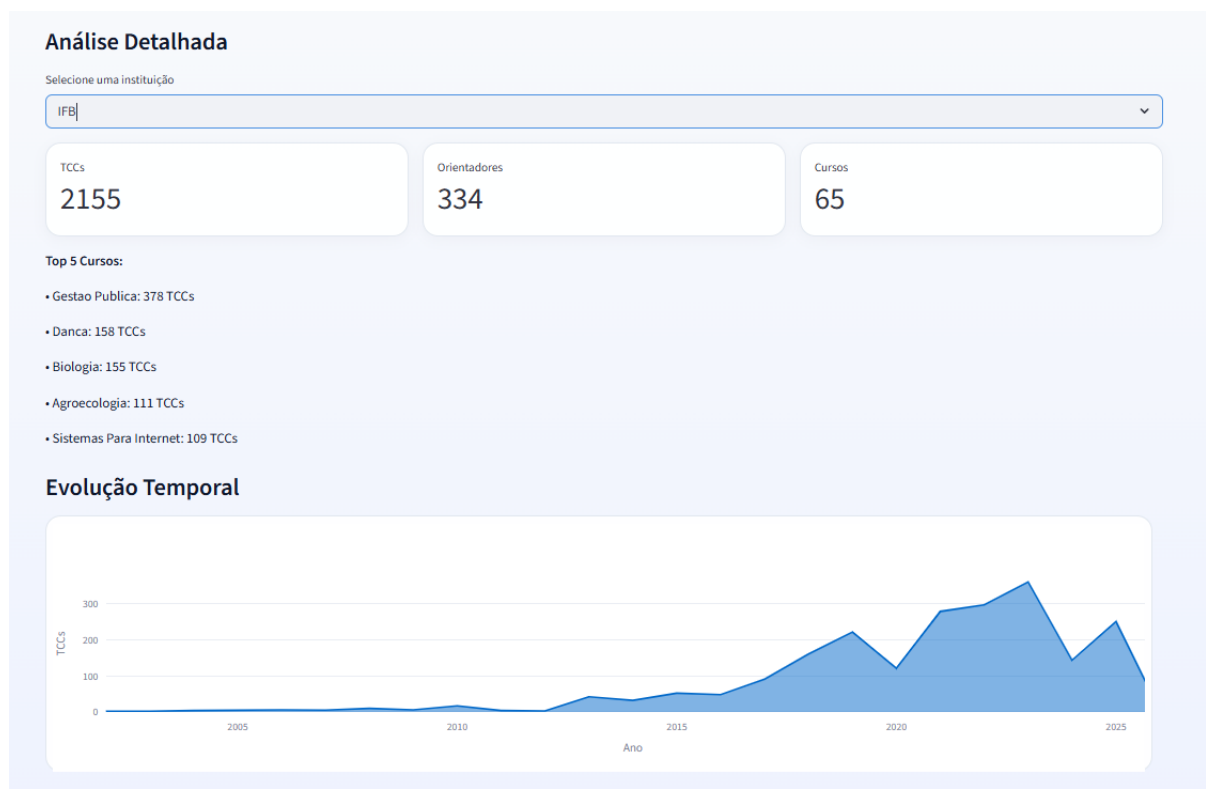


Fonte: Elaborado pelos autores.

A Figura 20 apresenta a análise detalhada da produção acadêmica após a seleção do Instituto Federal de Brasília (IFB) no filtro de instituições. O painel exibe indicadores referentes ao total de TCCs cadastrados, quantidade de orientadores e número de cursos da instituição. Também é apresentado o ranking dos cinco cursos com maior quantidade de Trabalhos de Conclusão de Curso, possibilitando

identificar as áreas de maior produção. Complementando a análise, o gráfico de evolução temporal ilustra a distribuição anual dos TCCs do IFB, permitindo observar o crescimento e as variações da produção acadêmica ao longo dos anos.

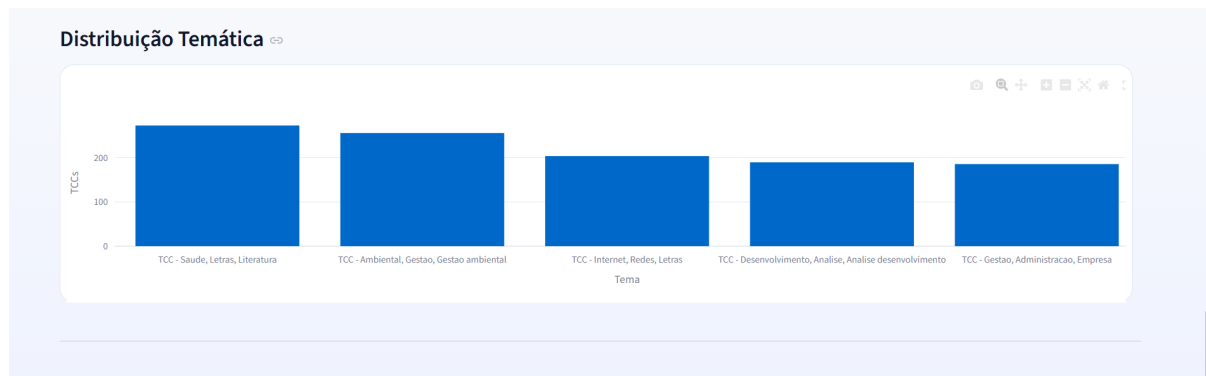
Figura 20: Análise detalhada das produções acadêmicas



Fonte: Elaborado pelos autores.

A plataforma permite realizar análises temáticas específicas por instituição, possibilitando identificar os principais agrupamentos presentes nas produções acadêmicas de cada unidade da Rede Federal. A Figura 21 apresenta a distribuição temática dos TCCs vinculados ao Instituto Federal de Brasília (IFB).

Figura 21: Distribuição temática de TCCs



Fonte: Elaborado pelos autores.

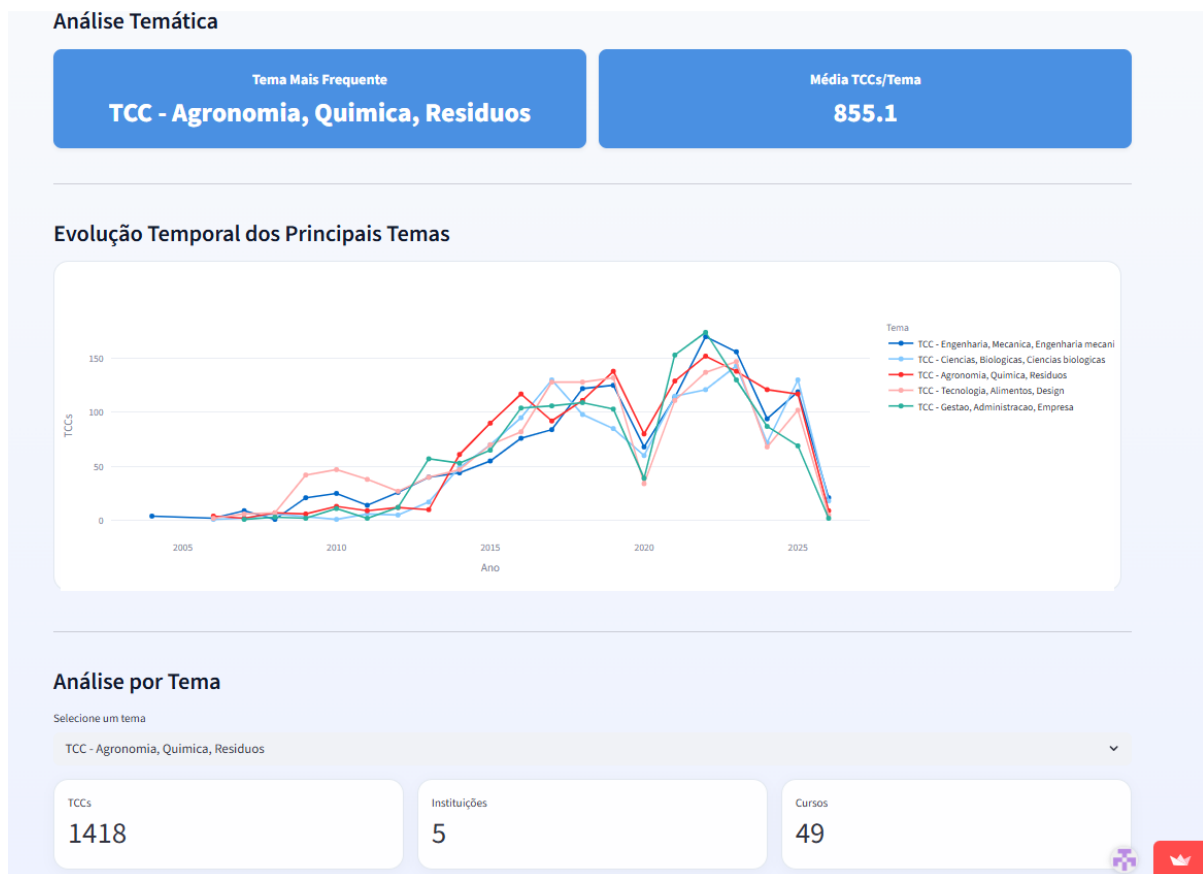
Observa-se predominância de agrupamentos relacionados às áreas de saúde, letras e literatura, seguidos por temas associados à gestão ambiental, redes e desenvolvimento. Esses resultados demonstram a

diversidade temática presente nos TCCs do IFB, evidenciando a presença de produções voltadas tanto às áreas tecnológicas quanto às ciências humanas e sociais aplicadas.

Por fim, a plataforma disponibiliza visualizações relacionadas aos agrupamentos temáticos identificados pelos modelos de Processamento de Linguagem Natural.

A Figura 22 apresenta a análise temática específica para os TCCs.

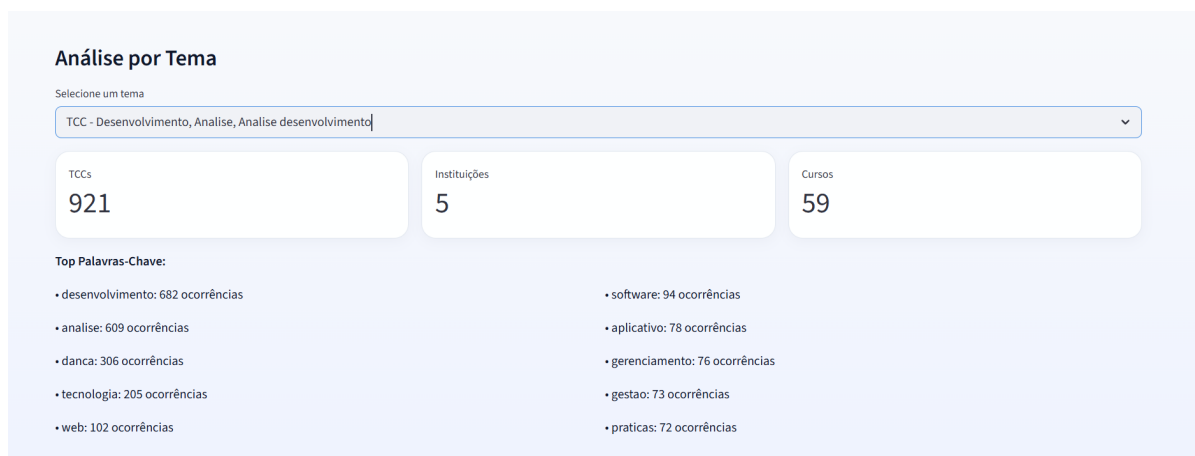
Figura 22: Análise temática dos TCCs



Fonte: Elaborado pelos autores.

O sistema também disponibiliza uma funcionalidade de análise detalhada por agrupamento temático, permitindo ao usuário selecionar um tema específico e visualizar indicadores quantitativos associados às produções acadêmicas correspondentes. A Figura 23 apresenta um exemplo dessa funcionalidade aplicada ao agrupamento temático relacionado a desenvolvimento e análise.

Figura 23: Análise detalhada por agrupamento temático

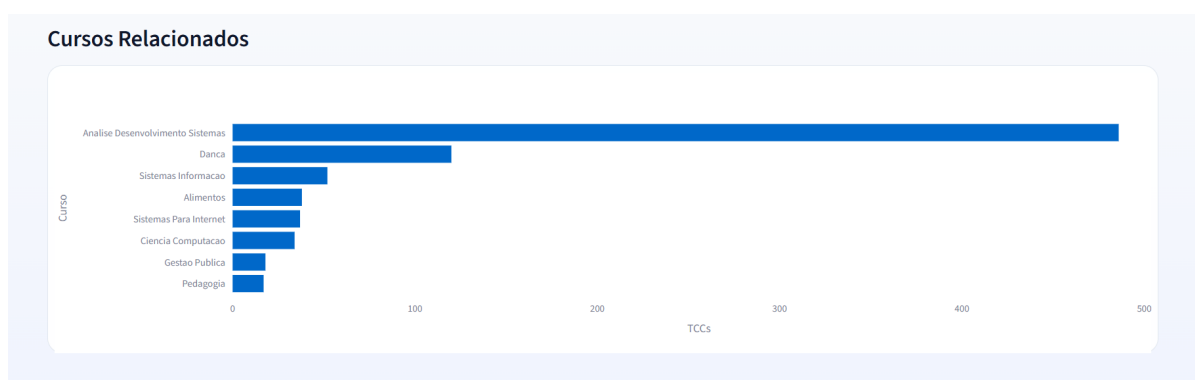


Fonte: Elaborado pelos autores.

Ao selecionar o agrupamento temático “Desenvolvimento, Análise”, o sistema apresenta informações relacionadas à quantidade de TCCs identificados, instituições participantes e cursos associados ao tema. Além disso, são exibidas palavras-chave com maior frequência de ocorrência, como “desenvolvimento”, “análise”, “software” e “tecnologia”, permitindo compreender os principais conceitos vinculados ao agrupamento analisado.

A plataforma também apresenta a distribuição dos cursos mais relacionados ao agrupamento temático selecionado, permitindo identificar quais áreas de formação possuem maior concentração de produções acadêmicas associadas ao tema analisado. A Figura 24 apresenta os cursos relacionados ao agrupamento “Desenvolvimento, Análise”.

Figura 24: Cursos relacionados ao agrupamento temático selecionado



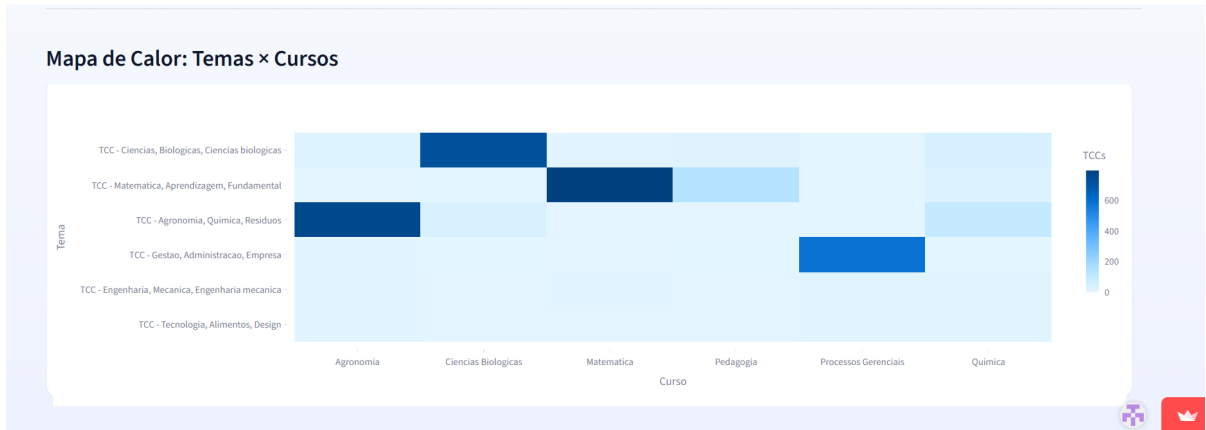
Fonte: Elaborado pelos autores.

Observa-se que o curso de Análise e Desenvolvimento de Sistemas apresenta a maior quantidade de TCCs associados ao agrupamento temático selecionado, indicando forte relação entre os temas de desenvolvimento e análise com áreas voltadas à computação e tecnologia. Também se destacam cursos como Dança, Sistemas de Informação e Ciência da Computação, evidenciando a diversidade de áreas contempladas pela análise temática da plataforma.

A plataforma também disponibiliza visualizações analíticas baseadas em mapas de calor, permitindo identificar relações entre agrupamentos temáticos e cursos associados aos Trabalhos de Conclusão de

Curso. A Figura 25 apresenta a intensidade da relação entre temas identificados e cursos da Rede Federal.

Figura 25: Mapa de calor entre temas e cursos relacionados aos TCCs



Fonte: Elaborado pelos autores.

Observa-se que determinados cursos apresentam maior concentração em temas específicos, evidenciada pelas regiões com coloração mais intensa no mapa de calor. Destacam-se, por exemplo, as associações entre o curso de Agronomia e o tema relacionado a “Agronomia, Química e Resíduos”, além da forte relação entre Matemática e o agrupamento temático “Matemática, Aprendizagem, Fundamental”. Esses resultados demonstram a capacidade da plataforma em identificar padrões temáticos predominantes entre diferentes áreas de formação.

Além das visualizações analíticas, a plataforma disponibiliza mecanismos de busca avançada voltados à recuperação de Trabalhos de Conclusão de Curso. A funcionalidade permite realizar pesquisas utilizando termos relacionados a títulos, autores, orientadores ou descrições dos trabalhos, retornando resultados semanticamente relacionados ao conteúdo pesquisado. A Figura 26 apresenta a interface de busca avançada implementada no módulo analítico de TCCs.

Figura 26: Busca avançada e similaridade de TCCs

Fonte: Elaborado pelos autores.

Na busca apresentada, foi utilizado o termo relacionado ao orientador “Gustavo Henrique Dornela de Deus”, resultando na recuperação de 17 trabalhos associados ao contexto pesquisado. A interface exibe títulos relacionados ao termo informado, permitindo identificar rapidamente produções acadêmicas semelhantes ou vinculadas ao mesmo orientador. Essa funcionalidade amplia a capacidade exploratória da plataforma, facilitando consultas específicas e análises comparativas entre produções acadêmicas.

A plataforma também disponibiliza funcionalidades voltadas à análise de similaridade entre TCCs, permitindo identificar produções acadêmicas semanticamente relacionadas. A funcionalidade utiliza informações textuais dos TCCs para localizar trabalhos com maior proximidade temática, auxiliando na exploração de conteúdos correlatos e na identificação de padrões de pesquisa. A Figura 27 apresenta a interface de análise de similaridade implementada no sistema.

Figura 27: Análise de similaridade entre Trabalhos de Conclusão de Curso

Análise de Similaridade entre TCCs

Selecione um TCC para encontrar trabalhos similares

Automação de um sistema de irrigação para agricultura familiar

Quantidade: 5

Buscar TCCs Similares

TCC de Referência:

Instituição: IFB

Título: Automação de um sistema de irrigação para agricultura familiar

Curso: Automação Industrial

Autores: Thauany Soares Da Silva

Tema: TCC - Automação, Engenharia, Industrial

Ano: 2023

Top 5 TCCs Mais Similares:

- #1 - Desenvolvimento de um protótipo para automação e supervisão remota de bombas d'água (73.9% similar)
- #2 - Caixa de Banho Térmico (73.7% similar)
- #3 - Sistema de Controle de Pressurização de Água (73.7% similar)
- #4 - Climatização de Adega (73.7% similar)
- #5 - Sistema de Controle para Fresadora CNC (73.7% similar)

Fonte: Elaborado pelos autores.

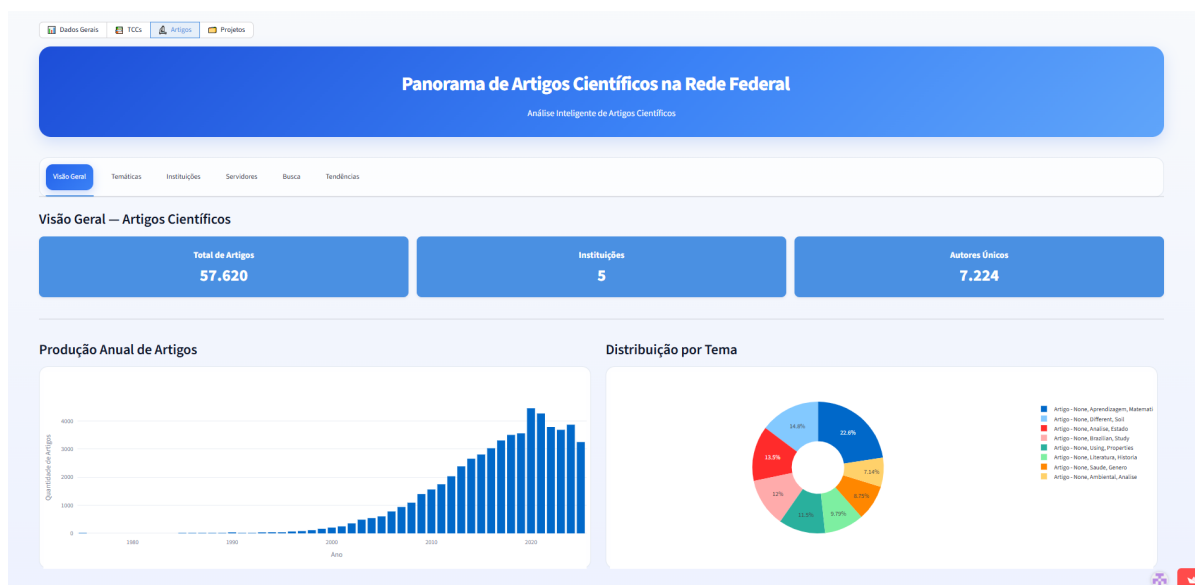
Na análise apresentada, foi selecionado o TCC “Automação de um sistema de irrigação para agricultura familiar”, pertencente ao Instituto Federal de Brasília (IFB). A plataforma identificou trabalhos semanticamente relacionados ao tema selecionado, destacando produções voltadas à automação industrial, controle de sistemas e supervisão remota. Os resultados são apresentados com percentuais de similaridade, permitindo avaliar o grau de proximidade entre os trabalhos encontrados e o TCC de referência.

5.6 Módulo Analítico de Artigos Científicos

O módulo analítico de artigos científicos foi desenvolvido para permitir a exploração quantitativa e temática das produções científicas identificadas na base de dados da Rede Federal. A interface integra indicadores estatísticos, gráficos temporais e visualizações temáticas, possibilitando analisar padrões de publicação, crescimento da produção acadêmica e distribuição dos principais temas presentes nos artigos coletados.

A Figura 28 apresenta a interface inicial do módulo analítico de artigos científicos da plataforma.

Figura 28: Visão geral do módulo analítico de artigos científicos



Fonte: Elaborado pelos autores.

Observa-se que o módulo apresenta indicadores relacionados ao total de artigos científicos coletados, quantidade de instituições analisadas e número de autores identificados na base de dados. Além disso, os gráficos permitem visualizar a evolução anual das publicações científicas e a distribuição dos principais agrupamentos temáticos identificados nos artigos analisados, possibilitando a identificação de tendências e áreas de maior incidência na produção científica da Rede Federal.

A Figura 29 apresenta informações complementares relacionadas às instituições com maior quantidade de artigos científicos identificados, aos principais veículos de publicação e à tabela detalhada das produções analisadas na plataforma.

Figura 29: Visão geral do módulo analítico de artigos científicos

Top 5 Instituições				Top 10 Veículos de Publicação			
Instituição	Artigos			Veículo	Artigos		
IFC		20449		Brazilian Journal of Development		615	
IFRS		12469		RESEARCH, SOCIETY AND DEVELOPMENT		521	
IFSC		12278		Higiene Alimentar		273	
IFMA		8673		OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA		236	
IFB		3751		Ciência Rural (UFPA, Impresso)		204	
				RENOTE. REVISTA NOVAS TECNOLOGIAS NA EDUCAÇÃO		189	
				CADERNO PEDAGÓGICO (LAEAD). ONLINE		178	
				CONTRIBUCIONES A LAS CIENCIAS SOCIALES		160	
				Revista Elio		154	
				Year: Revista de Educação, Ciência e Tecnologia		148	

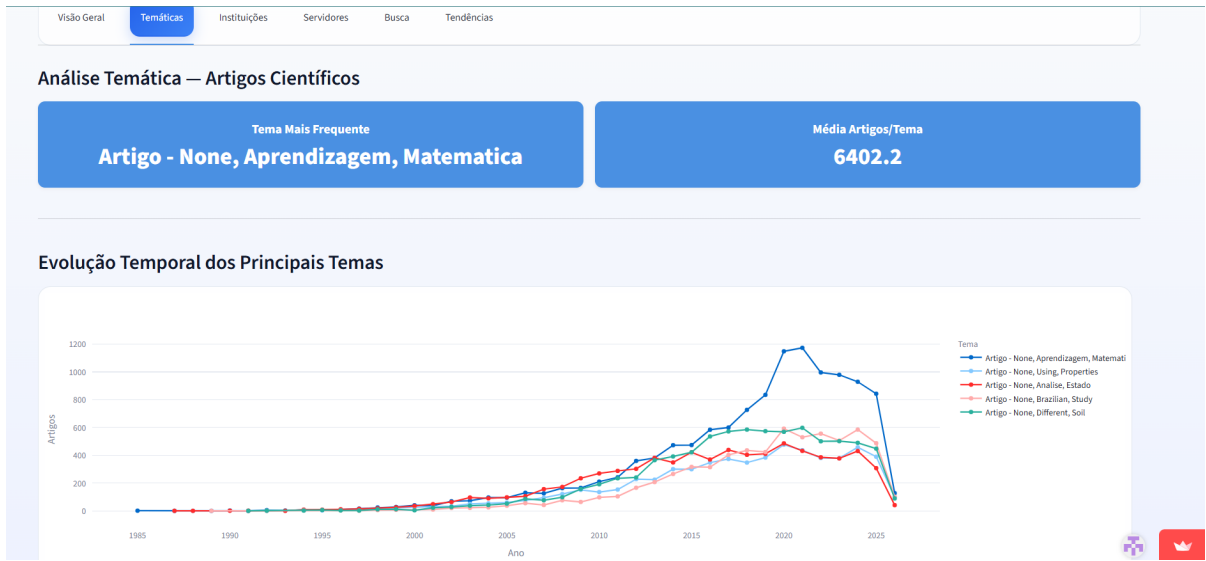
Tabela de Artigos						
Total exibido: 57.620 registros						
Titulo	autores	ano	instituição	veículo	nome_topico	
Spectroscopy-driven authentication of agro-industrial products: A comprehensive review of NMR, FT-IR, NIR	Cacilene Peinhoff Mega	2026	IFC	FOOD CHEMISTRY	Artigo - Tópico 1: None, Using, Properties	
Defoliation management and boron application as a strategy to reduce poor fruit set and improve the quality	Rogério De Oliveira Azeite	2026	IFSC	ACTA SCIENTIARUM-AGRONOMY	Artigo - Tópico 1: None, Different, Soil	
Zinc Bisulfate-Based Nanocomposites: Promising Antimicrobial Properties for Biomedical Applications	Mayara Coelho Sá	2026	IFMA	JOURNAL OF THE BRAZILIAN CHEMICAL SOCIETY	Artigo - Tópico 1: None, Using, Properties	
GOVERNANÇA DE DADOS E CONFORMIDADE À LGPD NO SETOR PREVIDENCIÁRIO: UM ESTUDO DE CASO SOB	Rodrigo Da Silva	2026	IFC	FOCO (FACULDADE NOVO MILÊNIO)	Artigo - Tópico 4: None, Análise, Estado	
Spectroscopy-driven authentication of agro-industrial products: a comprehensive review of NMR, FT-IR, NIR	Volmir Kist	2026	IFC	Food Chemistry	Artigo - Tópico 1: None, Using, Properties	
Avaliação de diferentes métodos de ponderação na formulação de indicadores e índices para gerenciamento	Camilla Angélica Baum	2026	IFRS	REVISTA DAE	Artigo - Tópico 4: None, Análise, Estado	
Pagamento por serviços ambientais como ferramenta para otimização da produção hidrelétrica	Camilla Angélica Baum	2026	IFRS	Revista Tamoios	Artigo - Tópico 4: None, Análise, Estado	
Letramento digital e tabuada: uma proposta pedagógica de integração de tecnologias interativas para ensino	Diego Ricardo Krohl	2026	IFC	https://doi.org/10.47238/2594-7036.2026.v30.1395	Artigo - Tópico 7: None, Aprendizagem, Matemática	
Uma autobiografia do tomar-se: eu-menina, eu-professora da infância	Luciana Carlos Geronzi	2026	IFC	DIALOGIA	Artigo - Tópico 5: None, Literatura, História	
Anti-scratching activity of in acute and dry skin models	Rosane Schenkel De Aquino	2026	IFSC	NATURAL PRODUCTS RESEARCH	Artigo - Tópico 3: None, Different, Soil	

Fonte: Elaborado pelos autores.

Observa-se que o Instituto Federal Catarinense (IFC) apresenta a maior quantidade de artigos científicos registrados entre as instituições analisadas, seguido pelo Instituto Federal do Rio Grande do Sul (IFRS) e pelo Instituto Federal de Santa Catarina (IFSC). Em relação aos veículos de publicação, destacam-se periódicos como o *Brazilian Journal of Development* e a revista *Research, Society and Development*, evidenciando diversidade temática e ampla disseminação científica. A tabela detalhada permite visualizar informações específicas sobre os artigos coletados, incluindo autores, instituição de origem, ano de publicação, veículo científico e agrupamento temático associado.

A Figura 30 apresenta a aba de temáticas do módulo analítico de artigos científicos, responsável por demonstrar os principais agrupamentos temáticos identificados na produção científica analisada.

Figura 30: Análise temática e evolução temporal dos artigos científicos

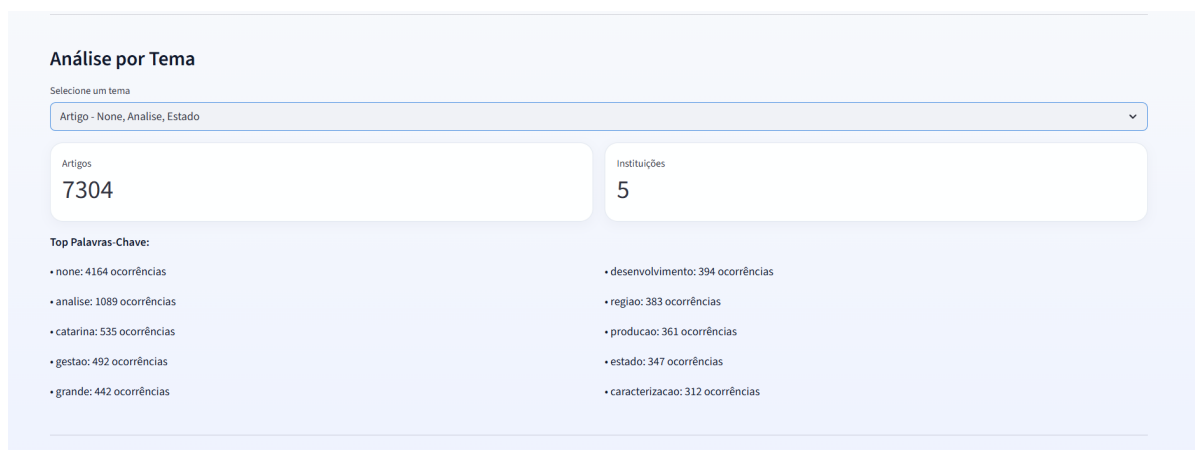


Fonte: Elaborado pelos autores.

Na interface apresentada, observam-se indicadores relacionados ao tema mais frequente identificado entre os artigos científicos, além da média de artigos por agrupamento temático. O gráfico de evolução temporal demonstra o crescimento gradual das produções ao longo dos anos, evidenciando maior incidência do agrupamento relacionado a aprendizagem e matemática a partir de 2010. Também é possível perceber que os demais temas acompanham tendência semelhante de crescimento, indicando expansão e diversificação da produção científica nas instituições analisadas.

A Figura 31 apresenta a funcionalidade de análise por tema disponível no módulo de artigos científicos, permitindo explorar informações detalhadas sobre agrupamentos temáticos específicos identificados na base de dados.

Figura 31: Análise detalhada por tema em artigos científicos

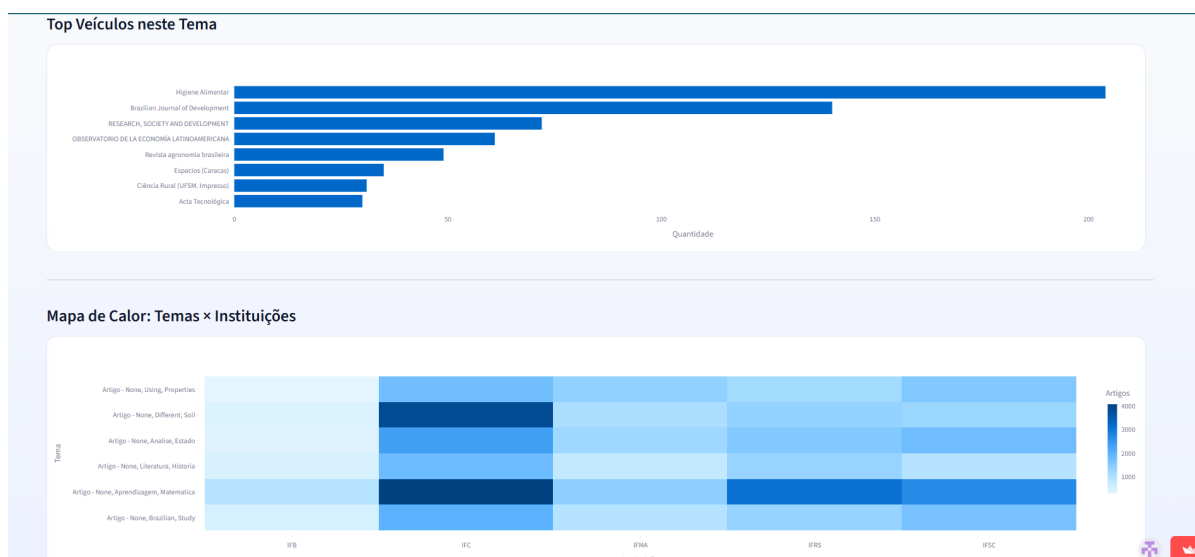


Fonte: Elaborado pelos autores.

Na imagem apresentada, foi selecionado o agrupamento temático “Artigo - None, Analise, Estado”, possibilitando visualizar métricas relacionadas ao tema escolhido. O sistema informa a quantidade total de artigos associados ao agrupamento, além do número de instituições em que o tema foi identificado. Também são exibidas as palavras-chave mais frequentes relacionadas ao agrupamento, com destaque para os termos “analise”, “gestao”, “desenvolvimento” e “estado”, evidenciando características predominantes das produções científicas vinculadas ao tema selecionado.

A Figura 32 apresenta visualizações complementares relacionadas aos agrupamentos temáticos dos artigos científicos, destacando os principais veículos de publicação associados aos temas identificados e a distribuição temática entre as instituições analisadas.

Figura 32: Veículos de publicação e distribuição temática por instituição



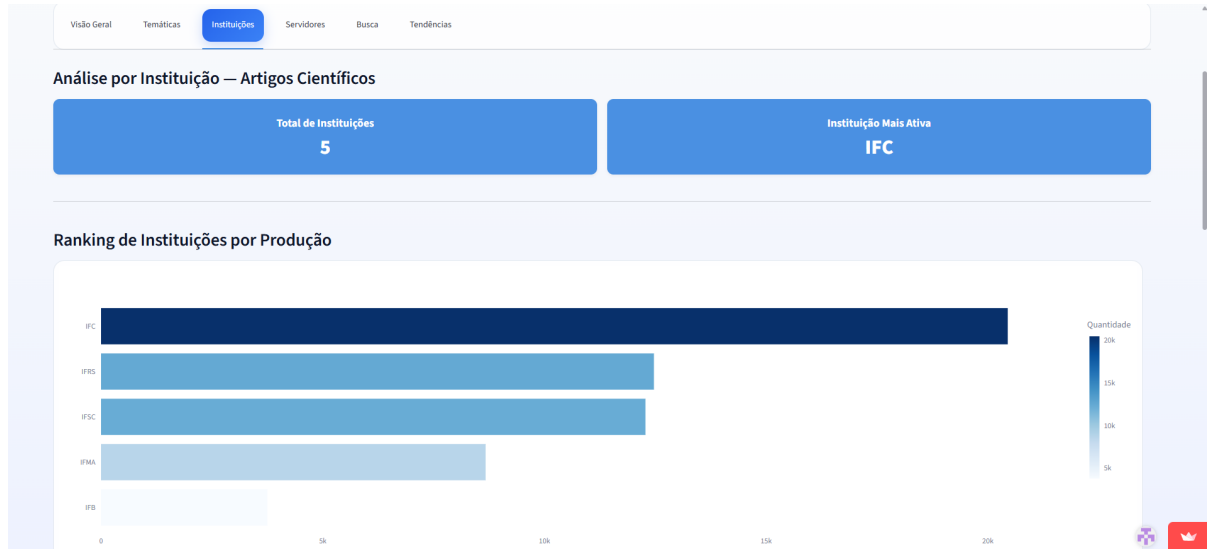
Fonte: Elaborado pelos autores.

No gráfico superior, observa-se a relação dos principais veículos de publicação vinculados ao tema selecionado, com destaque para o periódico “Higiene Alimentar”, que apresenta a maior quantidade de artigos associados ao agrupamento temático analisado. Já o mapa de calor permite visualizar a

intensidade da distribuição dos temas entre as instituições participantes, evidenciando maior concentração de produções no IFC e no IFRS em determinados agrupamentos temáticos. A variação de tonalidade facilita a identificação das instituições com maior incidência de artigos em cada tema analisado.

A Figura 33 apresenta a aba de análise por instituições do módulo de artigos científicos, destinada à visualização comparativa da produção acadêmica entre as instituições participantes da base analisada.

Figura 33: Ranking de instituições por produção científica

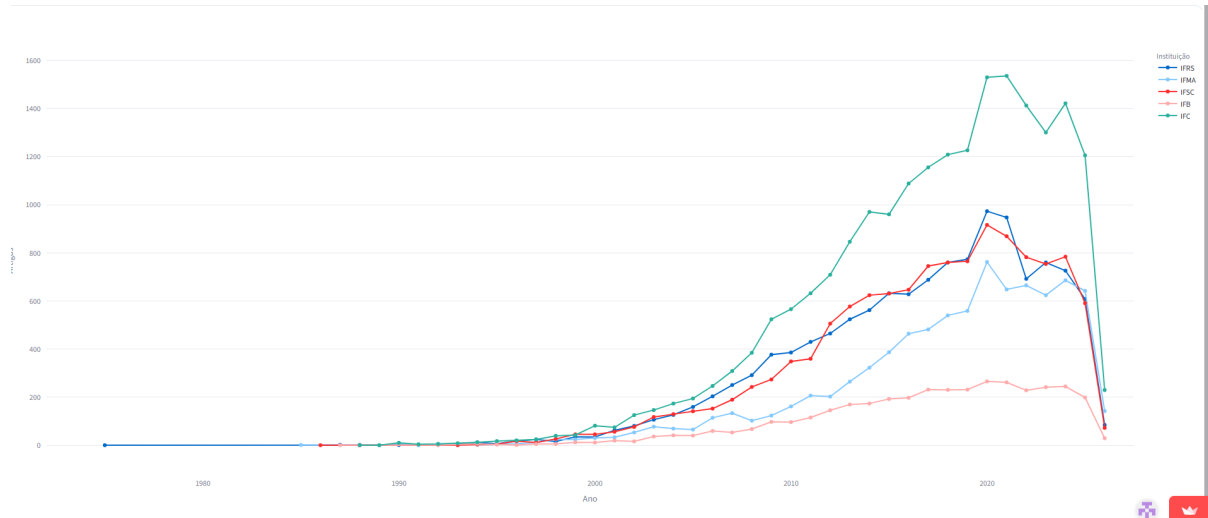


Fonte: Elaborado pelos autores.

Na interface apresentada, são exibidos indicadores relacionados à quantidade total de instituições analisadas e à instituição com maior volume de produção científica identificada na base de dados. O gráfico de ranking evidencia a distribuição da quantidade de artigos por instituição, destacando o IFC como a instituição mais ativa, apresentando volume significativamente superior em relação às demais. Também é possível observar que IFRS e IFSC ocupam posições próximas no ranking, demonstrando forte participação na produção científica da Rede Federal.

A Figura 34 apresenta a evolução anual da produção científica das instituições analisadas, permitindo comparar o crescimento das publicações ao longo do tempo.

Figura 34: Produção anual de artigos científicos por instituição

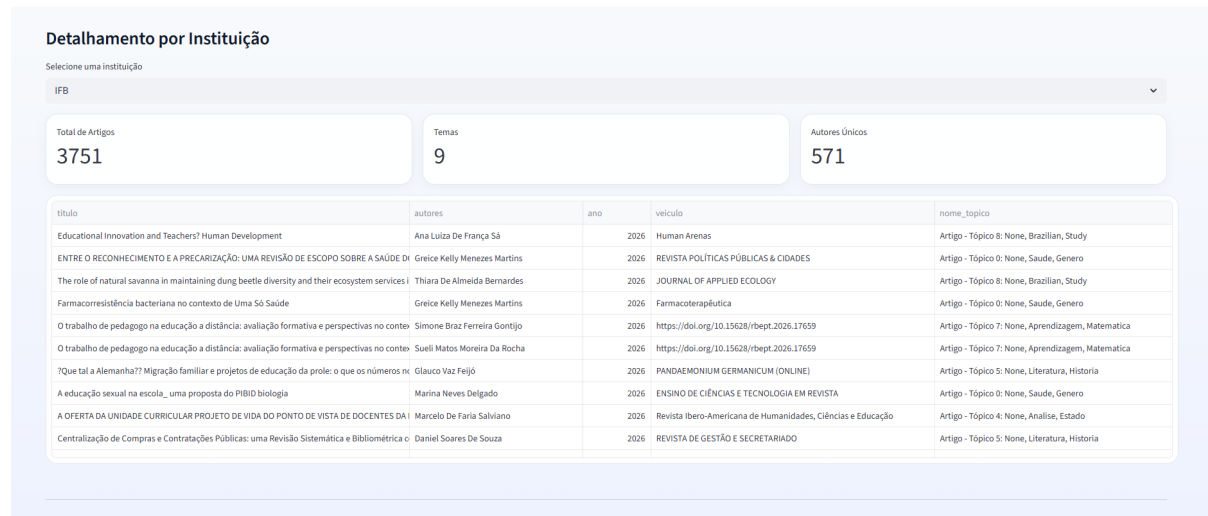


Fonte: Elaborado pelos autores.

O gráfico evidencia a evolução temporal da produção científica das instituições participantes, demonstrando crescimento progressivo no número de artigos publicados ao longo dos anos. Observa-se que o IFC apresenta a maior quantidade de publicações durante grande parte do período analisado, alcançando os valores mais elevados da série histórica. IFRS e IFSC também demonstram crescimento significativo na produção científica, enquanto IFMA e IFB apresentam quantitativos menores, porém mantendo evolução contínua ao longo do tempo. A visualização permite identificar tendências de expansão da produção acadêmica na Rede Federal e comparar o desempenho institucional em diferentes períodos.

A Figura 35 apresenta a funcionalidade de detalhamento por instituição disponível no módulo de artigos científicos, permitindo visualizar informações específicas relacionadas à produção acadêmica de uma instituição selecionada.

Figura 35: Detalhamento da produção científica por instituição



Fonte: Elaborado pelos autores.

Na interface apresentada, foi selecionado o IFB para análise detalhada da produção científica

institucional. O sistema exibe indicadores relacionados à quantidade total de artigos publicados, número de temas identificados e total de autores únicos associados à instituição. Além disso, é apresentada uma tabela contendo informações detalhadas dos artigos, incluindo título, autores, ano de publicação, veículo de divulgação científica e agrupamento temático correspondente. Essa funcionalidade possibilita explorar de forma organizada a produção acadêmica institucional, facilitando consultas e análises específicas sobre os trabalhos publicados.

A Figura 36 apresenta a aba de análise de servidores do módulo de artigos científicos, destinada à visualização de indicadores relacionados à produtividade acadêmica dos autores identificados na base de dados.

Figura 36: Análise de produtividade dos servidores em artigos científicos

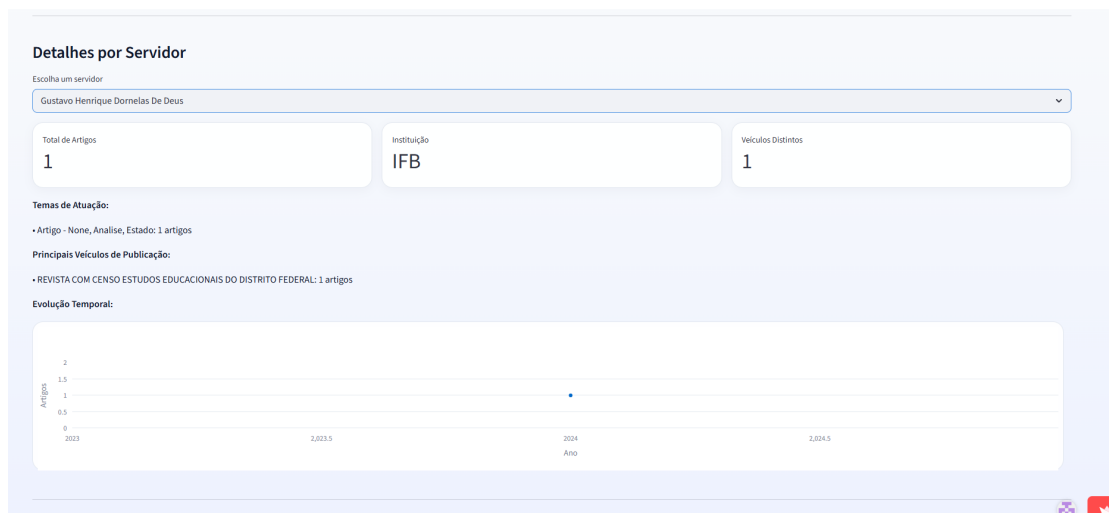


Fonte: Elaborado pelos autores.

Na interface apresentada, são exibidos indicadores relacionados à quantidade total de servidores identificados na base analisada, à média de artigos publicados por servidor e ao maior quantitativo de produções registrado entre os autores. O gráfico de ranking evidencia os 15 servidores com maior número de artigos publicados, destacando Leandro Luiz Marcuzzo como o pesquisador com maior volume de produção científica no conjunto analisado. A visualização permite identificar padrões de produtividade acadêmica e comparar o desempenho dos servidores em relação à quantidade de publicações realizadas.

A Figura 37 apresenta a funcionalidade de detalhamento por servidor, permitindo analisar individualmente a produção acadêmica associada a um autor específico.

Figura 37: Detalhamento da produção científica por servidor



Fonte: Elaborado pelos autores.

Na interface apresentada, foi selecionado o servidor Gustavo Henrique Dornelas De Deus para análise individual da produção científica. O sistema exibe informações relacionadas à quantidade total de artigos publicados, instituição vinculada e número de veículos distintos de publicação identificados. Além disso, são apresentados os principais temas de atuação e os veículos científicos associados às produções do servidor. O gráfico de evolução temporal permite visualizar a distribuição das publicações ao longo dos anos, contribuindo para análises específicas sobre a atuação acadêmica individual dos autores cadastrados na base.

A Figura 38 apresenta o ranking completo de servidores com maior número de artigos científicos publicados.

Figura 38: Ranking completo de servidores por quantidade de artigos publicados

Servidor	Artigos	Instituição	Tema Principal
Leandro Luiz Marcuzzo	318	IFC	Artigo - None, Solo, Língua
Ana Sara Castaman	270	IFC	Artigo - None, Aprendizagem, Matemática
Roberta Pasquali	217	IFC	Artigo - None, Aprendizagem, Matemática
Douglas Andre Würz	198	IFSC	Artigo - None, Analise, Estado
Vanessa Peripolli	164	IFC	Artigo - None, Different, Soil
Ricardo Evandro Mendes	154	IFC	Artigo - None, Different, Soil
Cássio Aurelio Suski	154	IFC	Artigo - None, Analise, Estado
Michel Goudart Da Silva	153	IFC	Artigo - None, Literatura, História
Antonio Mendes De Oliveira Neto	140	IFC	Artigo - None, Solo, Língua
Renata Assis Casagrande	138	IFC	Artigo - None, Different, Soil

Fonte: Elaborado pelos autores.

A tabela exibe os servidores com maior volume de publicações científicas, apresentando informações relacionadas à quantidade de artigos publicados, instituição de vínculo e principal temática associada às produções acadêmicas. Observa-se que o servidor Leandro Luiz Marcuzzo, vinculado ao IFC, ocupa a primeira posição do ranking, com 318 artigos publicados. A funcionalidade permite identificar pesquisadores com maior destaque em produção científica, além de evidenciar as principais áreas temáticas associadas aos autores mais produtivos da base analisada.

A Figura 39 apresenta a funcionalidade de busca de artigos científicos.

Figura 39: Funcionalidade de busca de artigos científicos

título	autores	ano	instituição	veículo	nome_topico
Análise de Documentos de Patente Utilizando um Sistema Multimodal: Proposta para Integ	Thiago Domingos Marques	2026	IFC	REVISTA MULTIDISCIPLINAR DO NORDESTE MINEIRO	Artigo - Tópico 7: None, Aprendizagem, Matematica
GOVERNANÇA DE DADOS E CONFORMIDADE À LGPD NO SETOR PREVIDENCIÁRIO: UM ESTUD	Rodrigo Da Silva	2026	IFSC	FOCO (FACULDADE NOVO MILÊNIO)	Artigo - Tópico 4: None, Analise, Estado
Quando dados viram decisão: o desafio real da gestão pública em Educação	Luis Carlos De Lima	2026	IFC	Revista Coleta Científica	Artigo - Tópico 7: None, Aprendizagem, Matematica
Compliance e Sanções da Lei Geral de Proteção de Dados	Ailton Zancanaro	2026	IFC	Revista Eletrônica de Iniciação Científica em Computação	Artigo - Tópico 4: None, Analise, Estado
GOVERNANÇA DE DADOS E CONFORMIDADE À LGPD NO SETOR PREVIDENCIÁRIO: UM ESTUD	Rodrigo Da Silva	2026	IFC	FOCO (FACULDADE NOVO MILÊNIO)	Artigo - Tópico 4: None, Analise, Estado
Compliance e Sanções da Lei Geral de Proteção de Dados	Carlos Alexandre Gouveia Da Silva	2026	IFC	Revista Eletrônica de Iniciação Científica em Computação	Artigo - Tópico 4: None, Analise, Estado
Compliance e Sanções da Lei Geral de Proteção de Dados	Regina Palva Melo Marin	2026	IFC	REVISTA ELETRÔNICA DE INICIAÇÃO CIENTÍFICA	Artigo - Tópico 4: None, Analise, Estado
Uma análise da produção científica brasileira em torno do tema da Biblioterapia indexada n	Monique Izoton	2025	IFRS	Revista Bibliomar	Artigo - Tópico 5: None, Literatura, História
Análise de dados estatísticos com auxílio do software GeoGebra: vivência em um curso de D	Raimundo Santos De Castro	2025	IFMA	CADERNO PEDAGÓGICO (LAJEADO, ONLINE)	Artigo - Tópico 7: None, Aprendizagem, Matematica
Gestão integrada de riscos em ambientes universitários: cibersegurança, informação e priv	Valdeir Reis Quietinho Leithardt	2025	IFC	Brazilian Journal of Development	Artigo - Tópico 7: None, Aprendizagem, Matematica

Fonte: Elaborado pelos autores.

A funcionalidade permite realizar pesquisas por palavras-chave presentes no título, autores ou resumo dos artigos científicos cadastrados na plataforma. No exemplo apresentado, foi realizada uma busca utilizando o termo “dados” no campo de título, retornando 302 artigos relacionados ao tema pesquisado. Os resultados são exibidos em formato tabular, contendo informações como título do artigo, autores, ano de publicação, instituição, veículo de publicação e temática associada. Esse recurso possibilita consultas rápidas e direcionadas, facilitando a localização de produções científicas específicas dentro da base analisada.

5.7 Módulo Analítico de Projetos

A aba específica de projetos foi desenvolvida para disponibilizar indicadores analíticos detalhados relacionados às produções classificadas como projetos acadêmicos e institucionais. O módulo permite explorar informações sobre produção anual, instituições, servidores, temáticas e tendências associadas aos projetos cadastrados na plataforma, possibilitando análises comparativas e identificação de padrões relevantes no contexto da Rede Federal.

A Figura 40 apresenta a interface da aba Visão Geral do módulo analítico de projetos acadêmicos da plataforma.

Figura 40: Visão geral do módulo analítico de projetos acadêmicos

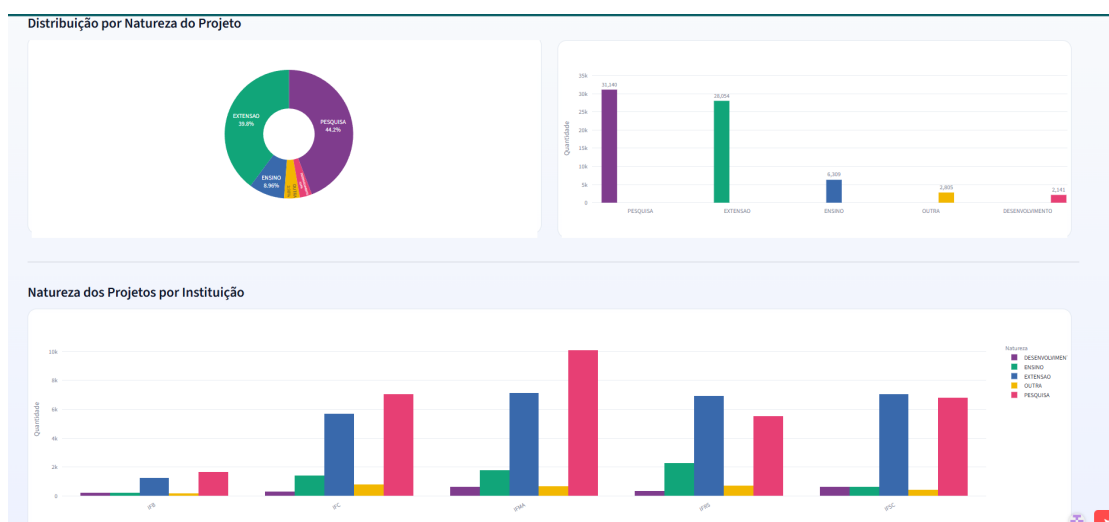


Fonte: Elaborado pelos autores.

A aba de visão geral concentra indicadores iniciais relacionados aos projetos acadêmicos cadastrados na plataforma, permitindo uma análise rápida do volume de produções existentes na base de dados. Entre os indicadores apresentados estão o total de projetos registrados, a quantidade de instituições participantes e o número de autores únicos vinculados às produções analisadas.

A Figura 41 apresenta a distribuição dos projetos acadêmicos conforme sua natureza, além da relação entre as categorias de projetos e as instituições participantes da base analisada.

Figura 41: Distribuição dos projetos por natureza e instituição



Fonte: Elaborado pelos autores.

Os gráficos apresentados permitem analisar a composição dos projetos acadêmicos segundo suas naturezas institucionais, destacando categorias como pesquisa, extensão, ensino, desenvolvimento e outras classificações. Observa-se que os projetos de pesquisa representam a maior parcela da base de dados, seguidos pelos projetos de extensão. Além disso, o gráfico inferior evidencia a distribuição dessas categorias entre as instituições da Rede Federal, possibilitando identificar diferenças no perfil de atuação de cada instituição em relação aos tipos de projetos desenvolvidos.

A Figura 42 apresenta tabelas com informações complementares relacionadas aos projetos acadêmicos cadastrados na plataforma.

Figura 42: Tabelas de instituições, financiadores e projetos acadêmicos

Top 5 Instituições			Top 10 Financiadores		
Instituição	Projetos		Financiador	Projetos	
IFMA	20314		None	58642	
IFRS	15789		Instituto Federal de Santa Catarina	2138	
IFSC	15507		Conselho Nacional de Desenvolvimento Científico e Tecnológico	1654	
IFC	15275		Fundação de Amparo à Pesquisa ao Desenvolvimento Científico e Tecnológico - MA	1126	
IFB	3564		Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul	1094	
			Instituto Federal do Maranhão	907	
			Instituto Federal Catarinense	708	
			Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina	403	
			Instituto Federal de Brasília	329	
			FUNDACAO DE AMPARO A PESQUISA DO ESTADO DO RIO GRANDE DO SUL	272	

Tabela de Projetos						
Total exibido: 70.449 registros						
título	autores	periodo	instituição	natureza	nome_topico	financiadores
Pogtherm: Aquecimento de Solar	Lucas Fernandes De Lima Lira		IFB	PESQUISA	Projeto - Tópico 7: Saúde, Desenvolvimento, Póde	None
Sonho de Banheiro	Lucas Fernandes De Lima Lira		IFB	PESQUISA	Projeto - Tópico 5: História, Social, Leitura	None
ÁGUA NOSSA DE CADA DIA REAPROVEITAMENTO DE ÁGUAS RESIDUAIS PARA USO DOMÉSTICO	Lucas Fernandes De Lima Lira		IFB	EXTENSÃO	Projeto - Tópico 4: Resíduos, Produção, Produtos	None
desenvolver procedimento de visita predial com vistas à prevenção de acidentes e diretrizes para re	Lucas Fernandes De Lima Lira		IFB	EXTENSÃO	Projeto - Tópico 0: Ambiental, Produção, Desenvolvimento	None
Elaboração de Laudos e capacitação dos servidores, equipe técnica de inspeção, para a elaboração de	Lucas Fernandes De Lima Lira		IFB	EXTENSÃO	Projeto - Tópico 0: Ambiental, Produção, Desenvolvimento	None
Elaboração de laudos periciais de engenharia sobre as condições estruturais e de solidez dos imóveis	Lucas Fernandes De Lima Lira		IFB	EXTENSÃO	Projeto - Tópico 7: Saúde, Desenvolvimento, Póde	None
CompactAR - Amassador de latas para catadores	Gabriel Queiroz Negreiro		IFB	DESENVOLVIMENTO	Projeto - Tópico 5: História, Social, Leitura	Fábrica de Ideias Inovadoras
CompactAR V2	Gabriel Queiroz Negreiro		IFB	DESENVOLVIMENTO	Projeto - Tópico 3: Nome, Curso, Programa	Instituto Federal de Brasília
Brinkando	Gabriel Queiroz Negreiro		IFB	DESENVOLVIMENTO	Projeto - Tópico 3: Nome, Curso, Programa	Instituto Federal de Brasília

Fonte: Elaborado pelos autores.

As tabelas apresentadas permitem visualizar informações complementares relacionadas aos projetos acadêmicos cadastrados na plataforma. A primeira tabela apresenta as instituições com maior quantidade de projetos registrados, destacando o IFMA como a instituição com maior volume de produções. Em seguida, é exibida a relação dos principais financiadores associados aos projetos acadêmicos, evidenciando a participação de instituições de fomento e órgãos vinculados à Rede Federal. Por fim, a tabela de projetos apresenta os registros detalhados das produções cadastradas, incluindo informações como título, autores, instituição, natureza do projeto, temática associada e financiadores vinculados.

A Figura 43 apresenta a aba de análise temática do módulo de projetos acadêmicos.

Figura 43: Análise temática dos projetos acadêmicos

Visão Geral

Temáticas

Instituições

Servidores

Busca

Tendências

Análise Temática — Projetos Acadêmicos

Tema Mais Frequente

Projeto - Comunidade, Atividades, Extensao

Média Projetos/Tema

7827.7

Análise por Tema

Selecione um tema

Projeto - Comunidade, Atividades, Extensao

Projetos

11418

Instituições

5

Natureza Principal

EXTENSÃO

Com Período

0

Top Palavras-Chave:

• comunidade: 8985 ocorrências

• extensao: 6133 ocorrências

• atividades: 6024 ocorrências

• estudantes: 4397 ocorrências

• evento: 4315 ocorrências

• curso: 3845 ocorrências

• ifsc: 3791 ocorrências

• desenvolvimento: 3732 ocorrências

• ifrs: 3117 ocorrências

• tecnologia: 3106 ocorrências

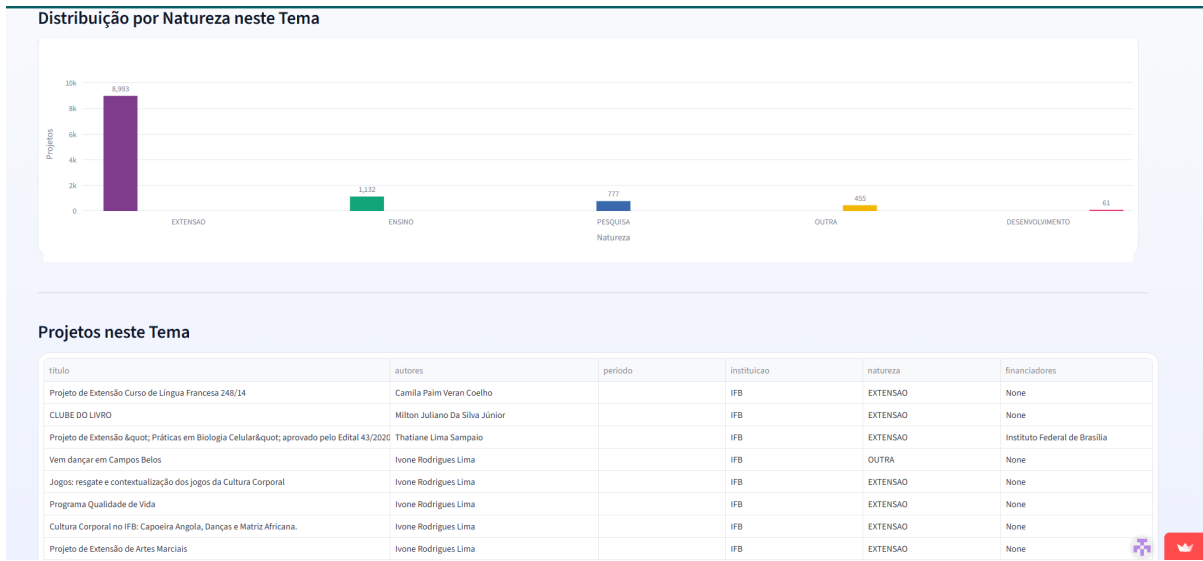
Fonte: Elaborado pelos autores.

A funcionalidade de análise temática permite explorar os principais temas identificados nos projetos

acadêmicos cadastrados na plataforma. Na interface apresentada, observa-se o tema mais frequente da base de dados, além da média de projetos associados às temáticas mapeadas. A seção de análise por tema possibilita selecionar um agrupamento temático específico, retornando indicadores relacionados à quantidade de projetos, instituições participantes, natureza principal do projeto e principais palavras-chave associadas ao tema selecionado. No exemplo apresentado, o agrupamento “Projeto - Comunidade, Atividades, Extensão” destaca-se como o tema mais frequente da base analisada, evidenciando forte associação com atividades extensionistas e ações voltadas à comunidade.

A Figura 44 apresenta a distribuição das naturezas dos projetos associados ao tema selecionado, bem como a listagem detalhada dos projetos relacionados.

Figura 44: Distribuição por natureza e listagem de projetos do tema selecionado

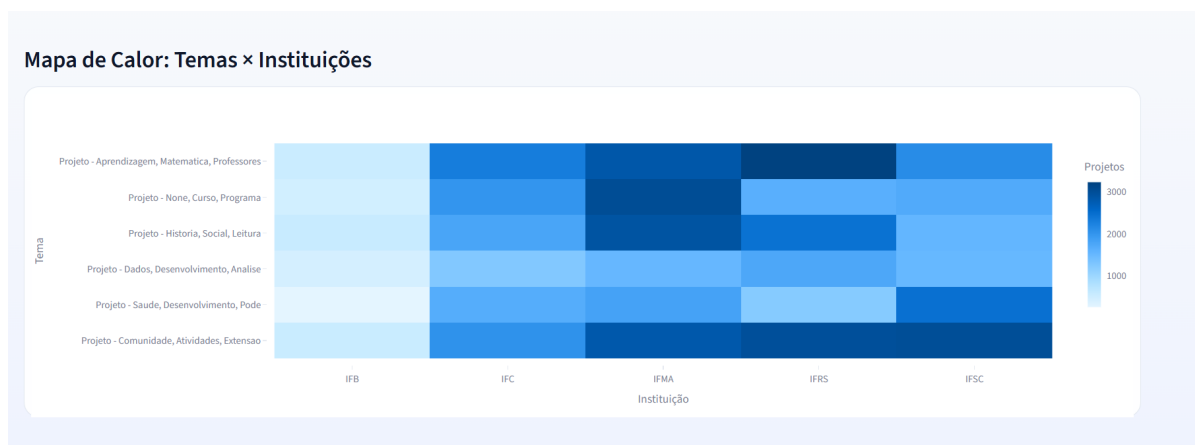


Fonte: Elaborado pelos autores.

A interface permite analisar a distribuição dos projetos de acordo com sua natureza dentro do tema selecionado. O gráfico evidencia a predominância de projetos classificados como extensão, seguidos por iniciativas de ensino, pesquisa, outras categorias e desenvolvimento. Complementando a análise, a tabela inferior apresenta os projetos vinculados ao tema, exibindo informações como título, autores, instituição, natureza e financiadores. Essa abordagem possibilita identificar tanto o perfil quantitativo das categorias de projetos quanto os registros específicos associados à temática analisada.

A Figura 45 apresenta um mapa de calor relacionando os principais temas identificados nos projetos acadêmicos.

Figura 45: Mapa de calor entre temas e instituições no módulo de projetos



Fonte: Elaborado pelos autores.

O mapa de calor permite visualizar a distribuição dos projetos entre os diferentes temas e instituições analisadas. Tons mais escuros indicam maior quantidade de projetos associados à combinação entre determinado tema e instituição. Observa-se destaque para os temas relacionados à comunidade, extensão, aprendizagem e desenvolvimento, principalmente nas instituições IFMA, IFRS e IFSC, evidenciando maior concentração de projetos nessas áreas. A visualização facilita a identificação de padrões de atuação institucional e das temáticas com maior representatividade dentro da base analisada. A Figura 46 apresenta a análise institucional do módulo de projetos acadêmicos, destacando o ranking das instituições de acordo com a quantidade de projetos registrados na base analisada.

Figura 46: Ranking de instituições por produção de projetos acadêmicos

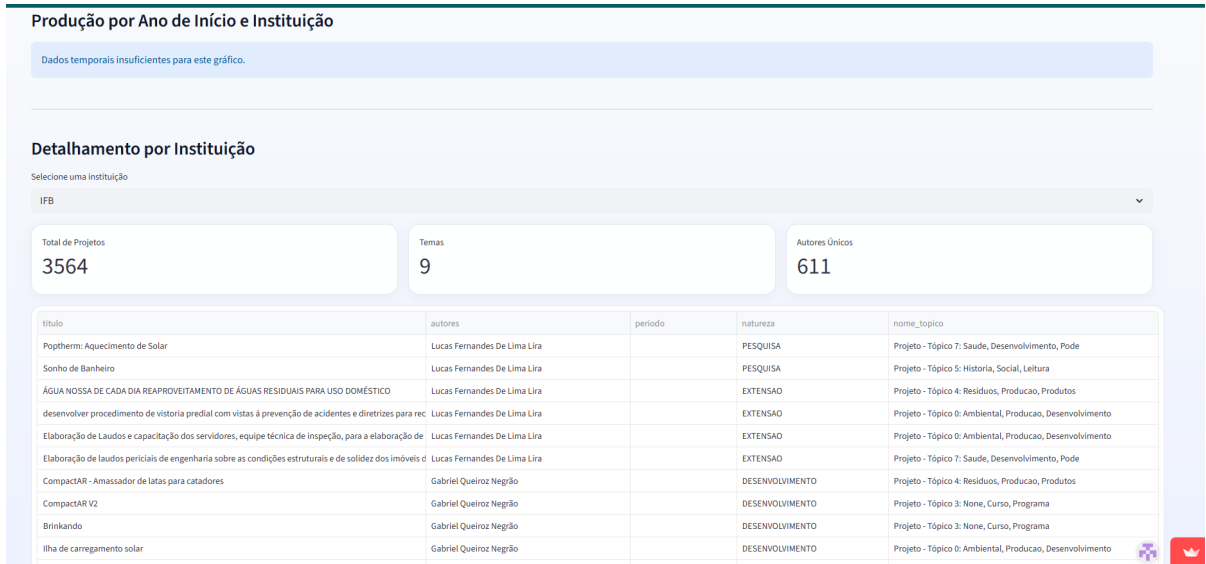


Fonte: Elaborado pelos autores.

Observa-se que o IFMA apresenta a maior quantidade de projetos cadastrados, sendo identificado como a instituição mais ativa dentro do conjunto analisado. Em seguida, destacam-se IFRS, IFSC e IFC, que apresentam quantitativos próximos entre si, evidenciando uma distribuição relativamente equilibrada entre essas instituições.

A Figura 47 apresenta o detalhamento dos projetos associados a uma instituição específica, permitindo visualizar indicadores quantitativos e exemplos de registros vinculados à instituição selecionada.

Figura 47: Detalhamento de projetos por instituição

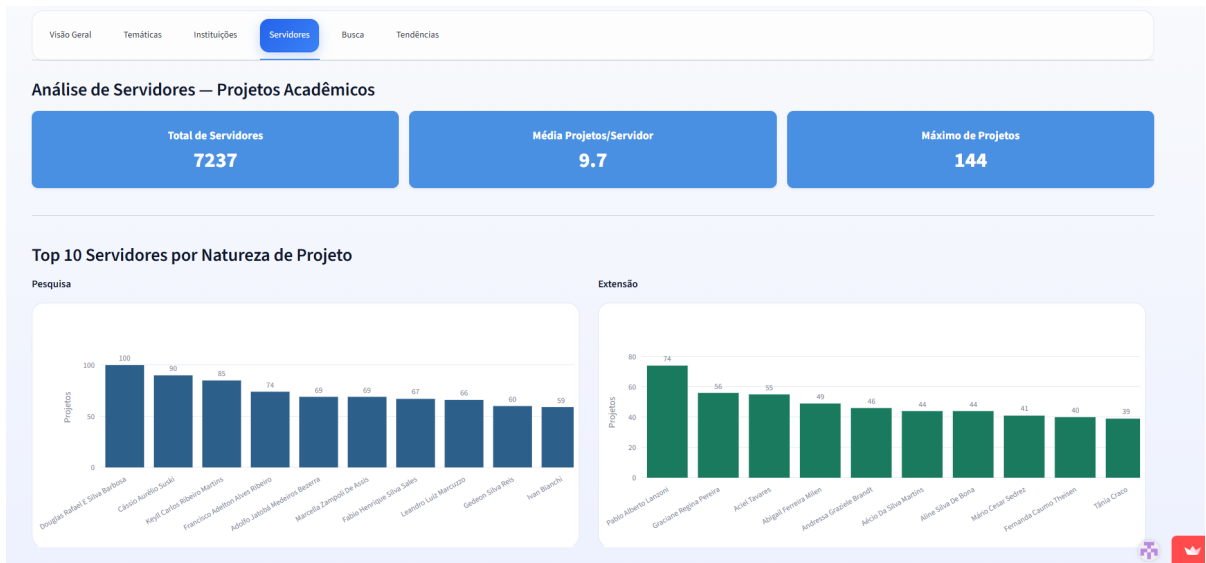


Fonte: Elaborado pelos autores.

Na visualização apresentada, observa-se o detalhamento da instituição IFB, contendo informações como quantidade total de projetos, número de temas identificados e total de autores únicos associados. A tabela inferior exibe exemplos de projetos vinculados à instituição, incluindo título, autores, natureza do projeto e temática correspondente.

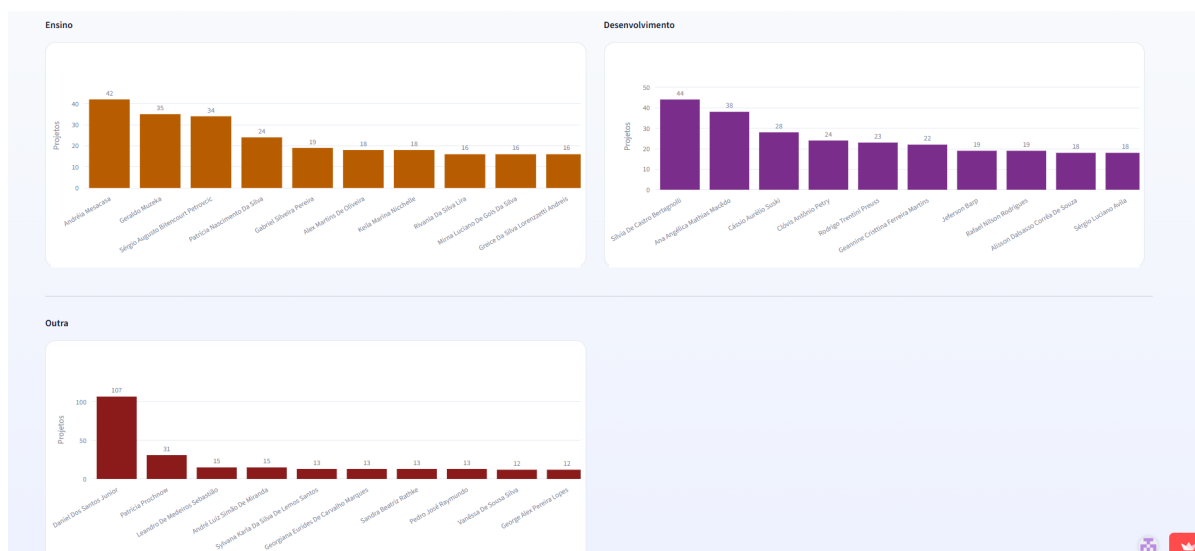
As Figuras 48 e 49 apresentam a análise dos servidores com maior participação em projetos acadêmicos, segmentados de acordo com a natureza dos projetos desenvolvidos.

Figura 48: Top servidores por natureza de projeto — Parte 1



Fonte: Elaborado pelos autores.

Figura 49: Top servidores por natureza de projeto — Parte 2



Fonte: Elaborado pelos autores.

Os gráficos evidenciam os servidores com maior quantidade de projetos em diferentes categorias, incluindo pesquisa, extensão, ensino, desenvolvimento e outras naturezas de projeto. Observa-se que determinados servidores se destacam significativamente em áreas específicas, indicando forte atuação em determinadas modalidades acadêmicas. Na categoria “Pesquisa”, por exemplo, alguns servidores apresentam quantitativos superiores aos demais, enquanto em “Extensão” e “Ensino” há uma distribuição mais equilibrada entre os participantes. A segmentação por natureza permite identificar perfis de atuação distintos entre os servidores da Rede Federal, contribuindo para análises comparativas sobre produção acadêmica e participação institucional em diferentes tipos de projetos.

A Figura 50 apresenta o ranking completo dos servidores com maior participação em projetos acadêmicos, incluindo a distribuição das produções por natureza de projeto.

Figura 50: Ranking completo de servidores

Servidor	Total	Instituição	Primeiro Ano	Último Ano	Desenvolvimento	Ensino	Extensão	Outra	Pesquisa
Cássio Aurélio Sucki	144	IFC	None	None	28	4	22	0	90
Keyll Carlos Ribeiro Martins	125	IFMA	None	None	8	1	24	7	85
Daniel Dos Santos Junior	114	IFC	None	None	0	1	6	107	0
Douglas Rafael E Silva Barbosa	108	IFMA	None	None	5	2	1	0	100
Fabio Henrique Silva Sales	103	IFMA	None	None	6	2	16	12	67
Ana Angélica Mathias Macêdo	100	IFMA	None	None	38	5	4	0	53
Iury De Almeida Accordi	100	IFRS	None	None	0	14	38	0	48
Francisco Adelfon Alves Ribeiro	99	IFMA	None	None	5	7	13	0	74
Marcella Zampoli De Assis	98	IFC	None	None	8	7	13	1	69
Graciane Regina Pereira	92	IFC	None	None	0	0	56	0	36

Fonte: Elaborado pelos autores.

A tabela permite visualizar detalhadamente os servidores com maior número de projetos registrados, apresentando informações como instituição de vínculo, quantidade total de projetos e distribuição entre as categorias de desenvolvimento, ensino, extensão, pesquisa e outras naturezas. Observa-se que

diferentes servidores apresentam perfis distintos de atuação acadêmica, com alguns concentrando suas produções principalmente em pesquisa, enquanto outros se destacam em extensão ou desenvolvimento. A Figura 51 apresenta a funcionalidade de busca de projetos acadêmicos, permitindo localizar registros a partir de palavras-chave presentes no título, autores ou resumo dos projetos cadastrados.

Figura 51: Funcionalidade de busca de projetos acadêmicos

Visão GeralTemáticasInstituiçõesServidoresBuscaTendências

Busca de Projetos Acadêmicos

Buscar por título, autor ou resumo

sistemas

Campo

Título

849 projeto(s) encontrado(s)

título	autores	periodo	instituicao	natureza	nome_topico
Como as ações antrópicas vêm interferindo nos ecossistemas aquáticos brasileiros?	Fernanda Keley Silva Pereira Navarro		IFB	PESQUISA	Projeto - Tópico 1: Solo, Producao, Especies
Sistemas Integrados de produção para a formação e recuperação de pastagens via consorci	Fabiano Gama De Sousa		IFB	PESQUISA	Projeto - Tópico 1: Solo, Producao, Especies
Sistemas integrados de produção de sorgo e capim sob sistema de plantio direto para form	Fabiano Gama De Sousa		IFB	PESQUISA	Projeto - Tópico 1: Solo, Producao, Especies
Projeto de Extensão para Constituição e Implantação da Empresa Júnior CODAMAS do Cur	Cristiane Jorge De Lima Bonfim		IFB	EXTENSAO	Projeto - Tópico 6: Dados, Desenvolvimento, Analise
Projeto de Extensão para Institucionalizar a Empresa Júnior (EJ) CODAMAS do Curso Super	Cristiane Jorge De Lima Bonfim		IFB	EXTENSAO	Projeto - Tópico 6: Dados, Desenvolvimento, Analise
Agrobiodiversidade dos sistemas produtivos no município de Coari-AM	Bruna Aparecida Madureira De Souza		IFB	PESQUISA	Projeto - Tópico 1: Solo, Producao, Especies
Concepção e Desenvolvimento de Sistemas de Gerenciamento de Informação Pessoal	Heitor José Dos Santos Barros		IFB	PESQUISA	Projeto - Tópico 6: Dados, Desenvolvimento, Analise
Concepção e Desenvolvimento de Sistemas Educacionais Adaptativos Baseados na Web Se	Heitor José Dos Santos Barros		IFB	PESQUISA	Projeto - Tópico 6: Dados, Desenvolvimento, Analise
Laboratório de Processamento e Reconhecimento Inteligente de Sistemas e Métodos Analít	Michele Calefe		IFB	PESQUISA	Projeto - Tópico 6: Dados, Desenvolvimento, Analise
ESTUDO E DESENVOLVIMENTO TECNOLÓGICO DE SISTEMAS GERADORES DE PLASMÓIDE PC	Giuliano Gonçalves De Souza		IFB	PESQUISA	Projeto - Tópico 7: Saude, Desenvolvimento, Pode

Fonte: Elaborado pelos autores.

Na interface apresentada, foi realizada uma busca utilizando o termo “sistemas”, resultando na recuperação de diversos projetos relacionados ao tema pesquisado. A tabela exibe informações relevantes como título do projeto, autores, instituição vinculada, natureza do projeto e temática associada. O recurso de busca permite explorar rapidamente os registros armazenados na plataforma, facilitando a localização de projetos específicos e contribuindo para análises temáticas e institucionais mais detalhadas.

A Figura 52 apresenta a funcionalidade de recomendação de projetos similares, baseada na análise de similaridade entre os registros cadastrados na plataforma.

Figura 52: Projetos similares ao primeiro resultado encontrado

Projetos Similares ao Primeiro Resultado

título	autores	periodo	instituicao	similaridade
VEM, ENADE!	Hosana Maria Da Silva Sousa		IFMA	0.8435
Ação de Ensino: Vem chegando o verão! ? Canga de praia com pintura em aquarela silk	Andréia Mesacasa		IFRS	0.795
Ação de Ensino: Vem chegando o verão! ? Bolsa multifuncional com tingimento em urucum e patch em aq	Andréia Mesacasa		IFRS	0.795
Ação de Ensino: Vem chegando o verão! ? Turbante com pintura em aquarela silk e sal grosso	Andréia Mesacasa		IFRS	0.795
V Encontro de Libras e II Seminário de Educação Inclusiva do Litoral Norte: vem pra live	Aline Dubal Machado		IFRS	0.7078

Fonte: Elaborado pelos autores.

A funcionalidade permite identificar projetos com características semelhantes ao registro selecionado, utilizando métricas de similaridade aplicadas aos dados textuais dos projetos. A tabela apresenta informações como título, autores, instituição vinculada e índice de similaridade calculado para cada

projeto relacionado. Esse recurso auxilia na descoberta de projetos correlatos, favorecendo análises comparativas, identificação de temas próximos e exploração de conteúdos acadêmicos relacionados dentro da plataforma.

6 Análise dos resultados

6.1 Apresentação e Análise dos Resultados

A partir da execução do pipeline automatizado de engenharia de dados desenvolvido nesta pesquisa, a esteira de extração e tratamento consolidou volumes expressivos de registros acadêmicos. Inicialmente, o processo de coleta automatizada por *web scraping* obteve uma volumetria bruta de 228.295 registros de Trabalhos de Conclusão de Curso (TCC), 296.318 artigos científicos e 295.162 projetos institucionais de ensino, pesquisa e extensão.

Ao submeter essa massa de dados à esteira de ETL, o mapeamento rigoroso de escopo resultou no descarte de 123.332 registros de TCCs por não pertencerem estritamente à Rede Federal de Educação Profissional, Científica e Tecnológica, além da remoção de 7.598 registros discentes que violavam as restrições de integridade referencial devido a falhas de consistência nas dimensões de campus e curso.

Na sequência do fluxo, a etapa de pré-processamento textual aplicou filtros linguísticos e de redundância sobre o corpus, expurgando por duplicidade um total de 3.081 TCCs, 27.832 artigos e 46.414 projetos. Adicionalmente, as rotinas de identificação de idioma removeram documentos inconsistentes ou redigidos em línguas fora do escopo analítico, totalizando a exclusão de 2.409 TCCs, 12.502 artigos e 4.930 projetos. Após a conclusão de toda a higienização, a base de dados analítica final integrada no arquivo *datamart.db* foi consolidada com 91.875 TCCs, 255.970 artigos científicos e 243.810 projetos institucionais válidos.

Destaca-se que parte significativa dos dados relacionados aos TCCs foi obtida a partir de repositórios institucionais baseados em plataformas padronizadas, muitas delas associadas a estruturas conhecidas como *.integra*. Esses repositórios apresentam organização relativamente homogênea, o que favorece a aplicação de técnicas de *web scraping* e contribui para a consistência dos dados coletados.

A Tabela 2 apresenta exemplos dessas fontes, evidenciando a diversidade de instituições contempladas na coleta de dados e reforçando a abrangência da base analisada.

Tabela 2: Repositórios institucionais da Rede Federal utilizados na coleta de dados

Instituição	URL do Repositório
Instituto Federal de Brasília (IFB)	https://repositorio.ifb.edu.br
Instituto Federal de Goiás (IFG)	https://repositorio.ifg.edu.br
Instituto Federal Goiano (IF Goiano)	https://repositorio.ifgoiano.edu.br
Instituto Federal do Tocantins (IFTO)	https://repositorio.ifto.edu.br
Instituto Federal do Pará (IFPA)	https://repositorio.ifpa.edu.br
Instituto Federal do Amazonas (IFAM)	https://repositorio.ifam.edu.br
Instituto Federal de Rondônia (IFRO)	https://repositorio.ifro.edu.br
Instituto Federal do Acre (IFAC)	https://repositorio.ifac.edu.br

Instituição	URL do Repositório
Instituto Federal de Roraima (IFRR)	https://repositorio.ifrr.edu.br
Instituto Federal do Amapá (IFAP)	https://repositorio.ifap.edu.br
Instituto Federal do Maranhão (IFMA)	https://repositorio.ifma.edu.br
Instituto Federal do Piauí (IFPI)	https://repositorio.ifpi.edu.br
Instituto Federal do Ceará (IFCE)	https://repositorio.ifce.edu.br
Instituto Federal do Rio Grande do Norte (IFRN)	https://repositorio.ifrn.edu.br
Instituto Federal da Paraíba (IFPB)	https://repositorio.ifpb.edu.br
Instituto Federal de Pernambuco (IFPE)	https://repositorio.ifpe.edu.br
Instituto Federal de Alagoas (IFAL)	https://repositorio.ifal.edu.br
Instituto Federal de Sergipe (IFS)	https://repositorio.ifs.edu.br
Instituto Federal da Bahia (IFBA)	https://repositorio.ifba.edu.br
Instituto Federal Baiano (IF Baiano)	https://repositorio.ifbaiano.edu.br
Instituto Federal de Minas Gerais (IFMG)	https://repositorio.ifmg.edu.br
Instituto Federal do Espírito Santo (IFES)	https://repositorio.ifes.edu.br
Instituto Federal do Rio de Janeiro (IFRJ)	https://repositorio.ifrj.edu.br
Instituto Federal Fluminense (IFF)	https://repositorio.iff.edu.br
Instituto Federal de São Paulo (IFSP)	https://repositorio.ifsp.edu.br
Instituto Federal do Paraná (IFPR)	https://repositorio.ifpr.edu.br
Instituto Federal de Santa Catarina (IFSC)	https://repositorio.ifsc.edu.br
Instituto Federal Catarinense (IFC)	https://repositorio.ifc.edu.br
Instituto Federal do Rio Grande do Sul (IFRS)	https://repositorio.ifrs.edu.br
Instituto Federal Sul-rio-grandense (IFSUL)	https://repositorio.ifsul.edu.br

Fonte: Elaborado pelos autores (2026).

A utilização dessas fontes contribui diretamente para a confiabilidade dos resultados apresentados, uma

vez que os dados analisados são provenientes de bases institucionais oficiais. Além disso, a padronização estrutural desses repositórios favorece a extração consistente de informações, reduzindo ruídos no processo de coleta.

Em consonância com a metodologia proposta, foram aplicadas técnicas de Processamento de Linguagem Natural para identificação de padrões nos dados textuais, possibilitando a formação de agrupamentos temáticos representativos em cada conjunto analisado.

Essa abordagem permite não apenas identificar tendências específicas em cada tipo de produção acadêmica, mas também realizar uma análise comparativa entre elas, evidenciando relações entre ensino, pesquisa e extensão no contexto da Rede Federal.

Ressalta-se que a identificação dos agrupamentos temáticos foi realizada com base nas técnicas de aprendizado de máquina propostas por Caixeta (2025), especialmente no uso de modelagem de tópicos para análise textual. Diferentemente do estudo original, no entanto, tais algoritmos foram aplicados sobre um conjunto de dados ampliado e com maior diversidade de atributos, o que contribuiu para o refinamento e maior robustez dos agrupamentos temáticos obtidos.

6.1.1 Artigos Científicos

A análise dos artigos científicos evidencia uma forte concentração em temáticas relacionadas à análise experimental, propriedades de materiais e estudos baseados em métodos científicos. Observa-se também a presença de áreas interdisciplinares, como educação, meio ambiente e saúde, indicando diversidade temática e maturidade na produção científica.

Para a construção dos agrupamentos temáticos, foram considerados como atributos textuais principais o **título** e as **palavras-chave** dos artigos. Essa escolha se justifica pelo fato de que esses elementos sintetizam os principais conceitos e objetivos das produções científicas, permitindo maior precisão na identificação de padrões semânticos.

Destacam-se agrupamentos com termos como “using”, “based” e “study”, que indicam trabalhos orientados à aplicação de métodos e experimentação. Além disso, temas relacionados à pandemia de COVID-19 e à gestão ambiental reforçam a aderência da produção científica a demandas contemporâneas.

A distribuição completa dos agrupamentos temáticos identificados pode ser observada na Tabela 3.

Tabela 3: Distribuição de artigos por agrupamento temático

Nº do Tópico	Temática	Quantidade
1	Using, Based, Study	51.314
2	Properties, Effect, Different	49.843
3	Matemática, Aprendizagem, Professores	33.779
4	Ambiental, Gestão, Análise	32.013
5	História, Literatura, Social	30.135
6	Análise, Production, Desempenho	29.885
7	Análise, Química, Tecnologias	26.751
8	Solo, Produção, Sementes	20.144
9	Saúde, Covid, Pandemia	20.134

Fonte: Elaborado pelos autores (2026).

6.1.2 Projetos

Os projetos apresentam uma forte orientação extensionista, com predominância de temas relacionados à interação com a comunidade, ações educacionais e desenvolvimento social. Termos como “comunidade”, “extensão” e “curso” aparecem com frequência, evidenciando o papel institucional da Rede Federal na promoção de impacto social.

Para a identificação dos agrupamentos temáticos, foram utilizados como base textual o **título** e a **descrição** dos projetos. Esses campos permitem capturar tanto o objetivo geral quanto o contexto de aplicação das iniciativas, sendo especialmente relevantes em projetos de ensino, pesquisa e extensão.

Além disso, observa-se a presença de projetos voltados à inovação, sustentabilidade e desenvolvimento tecnológico, indicando integração entre ensino, pesquisa e aplicação prática.

A distribuição dos agrupamentos temáticos dos projetos está apresentada na Tabela 4.

Tabela 4: Distribuição de projetos por agrupamento temático

Nº	Temática	Quantidade
1	Dados, Análise, Energia	38.674
2	Desenvolvimento, Aprendizagem, Matemática	37.436
3	Comunidade, Extensão, Curso	33.624
4	Comunidade, Cultura, Cultural	31.806
5	Curso, Professores, Estudantes	28.831
6	Saúde, Leitura, Língua	28.113
7	Produção, Desenvolvimento, Produtos	25.367
8	Produção, Resíduos, Solo	23.423
9	Edital, Programa, Extensão	22.103

Fonte: Elaborado pelos autores (2026).

6.1.3 Trabalhos de Conclusão de Curso (TCC)

Os Trabalhos de Conclusão de Curso apresentam maior dispersão temática, refletindo a diversidade dos cursos ofertados na Rede Federal. Há predominância de áreas aplicadas, especialmente engenharias, agronomia e tecnologia, além de temas relacionados à educação.

Diferentemente do trabalho de Caixeta (2025), que considerava principalmente título e resumo dos TCCs, nesta pesquisa foram utilizados três atributos textuais para a identificação das temáticas: o **título do curso**, o **título do TCC** e o **resumo**. A inclusão do título do curso permitiu enriquecer o contexto semântico dos dados, contribuindo para uma classificação temática mais precisa e alinhada às áreas de formação.

Observa-se que os TCCs possuem caráter mais prático, voltado ao desenvolvimento e implementação de soluções, em consonância com o perfil formativo desses cursos. Essa característica é semelhante à identificada no trabalho de Ana Luísa, porém com maior robustez analítica devido à ampliação dos atributos considerados.

A distribuição detalhada dos agrupamentos temáticos pode ser observada na Tabela 5.

Tabela 5: Distribuição de TCCs por agrupamento temático

Nº	Temática	Quantidade
1	Desenvolvimento, Análise	9.816
2	Agronomia, Produção, Engenharia	7.121
3	Física, Alimentos, Tecnologia	6.521
4	Matemática, Moda	5.308
5	Engenharia Civil, Automação	5.215
6	Gestão, Tecnologia, Turismo	4.980
7	Pedagogia, Aprendizagem, Escola	4.841
8	Engenharia Mecânica, Produção	4.776
9	Ciências Biológicas	4.604
10	Química, Arquitetura, Urbanismo	4.578
11	Ambiental, Gestão Ambiental	4.435
12	Engenharia Elétrica, Energia	4.175

Fonte: Elaborado pelos autores (2026).

6.2 Análise Comparativa dos Agrupamentos Temáticos

A análise temática dos dados coletados permitiu identificar padrões predominantes na produção acadêmica da Rede Federal de Educação Profissional, Científica e Tecnológica, considerando artigos científicos, projetos institucionais e Trabalhos de Conclusão de Curso (TCCs).

Os resultados demonstram que a produção acadêmica da Rede Federal apresenta forte diversidade temática, contemplando áreas tecnológicas, educacionais, ambientais, sociais e de engenharia. Além disso, observou-se predominância de conteúdos associados ao desenvolvimento tecnológico, análise de dados, processos produtivos e aplicações práticas.

Nos artigos científicos, verificou-se maior incidência de agrupamentos relacionados a pesquisas experimentais e tecnológicas, especialmente representadas pelos tópicos “Using, Based, Study” e “Properties, Effect, Different”. Esse comportamento sugere forte presença de publicações voltadas à internacionalização científica, considerando o uso recorrente de terminologias em língua inglesa e características comuns de artigos indexados em periódicos internacionais.

Também foram identificados agrupamentos ligados às áreas de ensino, aprendizagem e formação docente, evidenciando que a Rede Federal mantém forte participação em pesquisas educacionais e em estudos relacionados às licenciaturas e às ciências humanas.

Nos projetos institucionais, observou-se predominância de iniciativas voltadas ao desenvolvimento tecnológico, extensão comunitária e sustentabilidade. Os tópicos associados a dados, energia, matemática, comunidade e cultura demonstram que os projetos possuem forte caráter aplicado e socialmente orientado, reforçando o papel dos Institutos Federais no desenvolvimento regional e na integração entre ensino, pesquisa e extensão.

Em relação aos Trabalhos de Conclusão de Curso, verificou-se concentração significativa em áreas relacionadas ao desenvolvimento de software, engenharias, automação e produção industrial. Esse comportamento indica que a produção discente da Rede Federal possui perfil fortemente aplicado, direcionado à resolução de problemas práticos e ao desenvolvimento de soluções tecnológicas.

Além disso, a diversidade temática observada nos TCCs demonstra a natureza multidisciplinar dos

Institutos Federais, abrangendo áreas como pedagogia, turismo, química, arquitetura, ciências biológicas e gestão ambiental.

6.2.1 Legenda dos Principais Tópicos

- **TCCs:** Tópico 3 (Desenvolvimento de Software/Web); Tópico 6 (Ciências da Natureza e Aprendizagem); Tópico 9 (Engenharias e Construção).
- **Projetos:** Tópico 0 (Produção, Água e Solo); Tópico 1 (Desenvolvimento, Energia e Dados); Tópico 2 (Ensino e Matemática); Tópico 3 (Extensão e Comunidade); Tópico 4 (Social, Cultura e Mulher).
- **Artigos:** Tópico 1 (Using/Based/Study); Tópico 2 (Properties/Effect); Tópico 3 (Matemática e Professores); Tópico 4 (Ambiental e Gestão); Tópico 5 (História e Social); Tópico 8 (Solo, Sementes e Produção).

7 Conclusão

A análise comparativa dos três conjuntos evidencia a complementaridade entre ensino, pesquisa e extensão, pilares fundamentais da Rede Federal.

Os TCCs concentram-se no desenvolvimento de soluções aplicadas, refletindo o caráter formativo e prático dos cursos. Os artigos científicos apresentam maior rigor metodológico e aprofundamento investigativo, enquanto os projetos destacam-se pela forte orientação social e extensionista.

A presença de temáticas transversais, como educação, tecnologia e meio ambiente, em todas as categorias analisadas, indica alinhamento com demandas contemporâneas e reforça a integração entre diferentes áreas do conhecimento.

Comparativamente ao trabalho de Caixeta (2025), observa-se um avanço na abrangência e na diversidade dos dados analisados, bem como na consolidação dos agrupamentos temáticos, decorrente da ampliação das fontes de dados e da integração de diferentes abordagens analíticas.

Dessa forma, os resultados demonstram que a combinação de técnicas de coleta automatizada, processamento de linguagem natural e visualização de dados constitui uma estratégia eficaz para análise integrada da produção acadêmica.

Entretanto, a pesquisa apresenta algumas limitações. Entre elas, destacam-se a dependência da qualidade e padronização dos metadados disponibilizados nos repositórios analisados, a possibilidade de perda de informações relevantes durante as etapas de pré-processamento textual e as limitações inerentes aos modelos de agrupamento temático, que podem gerar sobreposições ou simplificações das áreas do conhecimento. Além disso, parte dos documentos analisados apresentava inconsistências textuais, ausência de resumos ou divergências terminológicas, o que pode impactar a precisão dos agrupamentos identificados.

Como contribuições, o estudo disponibiliza uma abordagem metodológica replicável para mapeamento e análise da produção acadêmica da Rede Federal, integrando técnicas de web scraping, processamento de linguagem natural, modelagem de tópicos e visualização de dados. Adicionalmente, a pesquisa contribui para ampliação da transparência e da compreensão sobre as dinâmicas temáticas institucionais, podendo subsidiar gestores, pesquisadores e formuladores de políticas educacionais na identificação de tendências, lacunas e potencialidades da produção científica e extensionista da Rede Federal.

Tabela 6: Tópico de maior incidência por instituição

Instituição	TCC	Projeto	Artigo
IFAC	Tópico 3 / 42%	Tópico 3 / 38%	Tópico 4 / 31%
IFAL	Tópico 3 / 45%	Tópico 3 / 41%	Tópico 3 / 35%
IFAM	Tópico 3 / 47%	Tópico 4 / 33%	Tópico 4 / 29%
IFB	Tópico 3 / 49%	Tópico 3 / 37%	Tópico 5 / 30%
IFBA	Tópico 3 / 43%	Tópico 1 / 39%	Tópico 1 / 28%
IFCE	Tópico 3 / 52%	Tópico 1 / 46%	Tópico 1 / 34%
IFES	Tópico 3 / 40%	Tópico 1 / 35%	Tópico 2 / 27%
IFFARROUPILHA	Tópico 3 / 38%	Tópico 0 / 34%	Tópico 8 / 25%
IFFLUMINENSE	Tópico 3 / 44%	Tópico 1 / 42%	Tópico 2 / 31%
IFG	Tópico 3 / 39%	Tópico 2 / 33%	Tópico 3 / 30%
IFGOIANO	Tópico 3 / 35%	Tópico 0 / 45%	Tópico 8 / 38%
IFMA	Tópico 3 / 41%	Tópico 1 / 36%	Tópico 2 / 29%
IFMG	Tópico 3 / 48%	Tópico 1 / 40%	Tópico 1 / 32%
IFMT	Tópico 3 / 37%	Tópico 0 / 39%	Tópico 4 / 33%
IFMS	Tópico 3 / 51%	Tópico 3 / 38%	Tópico 1 / 28%
IFPA	Tópico 3 / 36%	Tópico 0 / 35%	Tópico 4 / 31%
IFPB	Tópico 3 / 54%	Tópico 1 / 44%	Tópico 1 / 36%
IFPE	Tópico 3 / 46%	Tópico 1 / 38%	Tópico 6 / 30%
IFPI	Tópico 3 / 45%	Tópico 2 / 36%	Tópico 3 / 32%
IFPR	Tópico 3 / 40%	Tópico 2 / 34%	Tópico 3 / 29%
IFRJ	Tópico 6 / 32%	Tópico 4 / 39%	Tópico 5 / 35%
IFRN	Tópico 3 / 46%	Tópico 3 / 40%	Tópico 1 / 31%
IFRO	Tópico 3 / 39%	Tópico 0 / 37%	Tópico 8 / 30%
IFRR	Tópico 3 / 37%	Tópico 3 / 42%	Tópico 3 / 34%
IFRS	Tópico 3 / 49%	Tópico 1 / 38%	Tópico 1 / 29%
IFSC	Tópico 3 / 43%	Tópico 1 / 37%	Tópico 2 / 30%
IFSP	Tópico 3 / 50%	Tópico 1 / 43%	Tópico 1 / 37%
IFSUDESTEMG	Tópico 3 / 41%	Tópico 3 / 35%	Tópico 6 / 28%
IFSUL	Tópico 3 / 47%	Tópico 1 / 39%	Tópico 1 / 32%
IFTO	Tópico 3 / 44%	Tópico 3 / 38%	Tópico 3 / 29%
IFSertãoPE	Tópico 3 / 38%	Tópico 0 / 41%	Tópico 8 / 34%

Fonte: Elaborado pelos autores (2026).

Referências

- BRASIL. Lei nº 11.892, de 29 de dezembro de 2008. Institui a Rede Federal de Educação Profissional, Científica e Tecnológica, cria os Institutos Federais de Educação, Ciência e Tecnologia, e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, DF, 30 dez. 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111892.htm. Acesso em: 21 maio 2026.
- BRASIL. Lei nº 13.243, de 11 de janeiro de 2016. Dispõe sobre estímulos ao desenvolvimento científico, à pesquisa, à capacitação científica e tecnológica e à inovação. **Diário Oficial da União**: seção 1, Brasília, DF, 12 jan. 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/lei/113243.htm. Acesso em: 21 maio 2026.
- BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 21 maio 2026.
- CAIXETA, Ana Luísa. **Análise temática de Trabalhos de Conclusão de Curso na Rede Federal**: uma abordagem baseada em web scraping e visualização de dados. 2025. Trabalho de Conclusão de Curso (Tecnólogo em Sistemas para Internet) – Instituto Federal de Brasília, Brasília, 2025.
- CRESWELL, John W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre: Artmed, 2010.
- GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Oasisbr: Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto**. Brasília, DF: IBICT, [s.d.]. Disponível em: <https://oasisbr.ibict.br/>. Acesso em: 18 jun. 2026.
- MANNING, Christopher D.; SCHÜTZE, Hinrich. **Foundations of Statistical Natural Language Processing**. Cambridge: MIT Press, 1999.
- MITCHELL, Ryan. **Web Scraping with Python**: Collecting More Data from the Modern Web. 2. ed. Sebastopol: O'Reilly Media, 2018.
- MOTA, Ana Caroline Ribeiro; JESUÍNO, Fillipy Antony Gabriel. **Visualização interativa de dados docentes**: integração de dados via web scraping e Business Intelligence. 2025. Trabalho de Conclusão de Curso (Tecnólogo em Sistemas para Internet) – Instituto Federal de Brasília, Brasília, 2025.
- PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

RICHARDSON, Roberto Jarry. **Pesquisa social:** métodos e técnicas. 3. ed. São Paulo: Atlas, 2017.

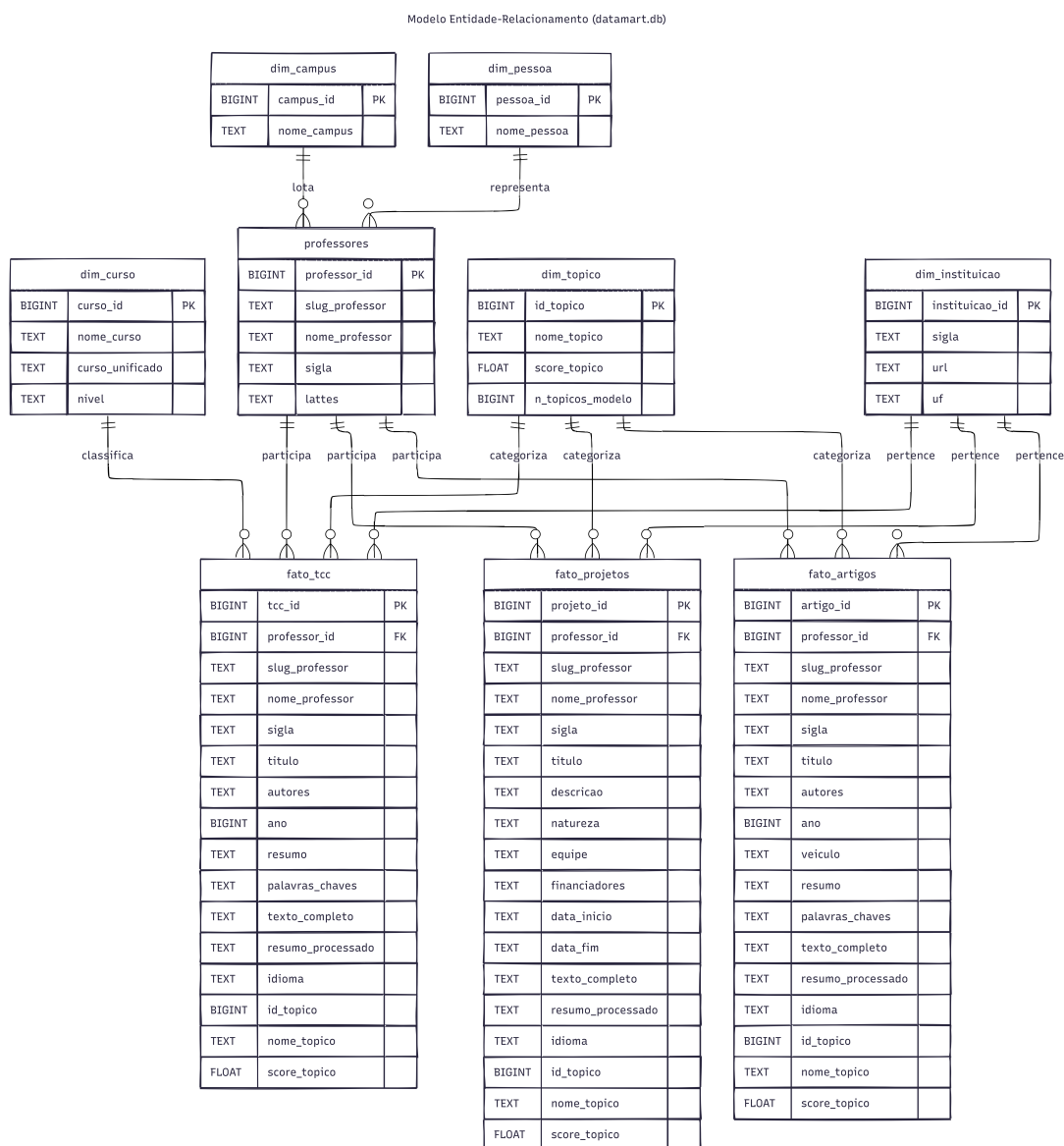
A Apêndices

Os apêndices apresentam materiais complementares utilizados no desenvolvimento do sistema proposto neste trabalho. São apresentados o Modelo Entidade Relacionamento (MER), do banco de dados utilizado na aplicação e o dicionário de dados correspondente às estruturas de armazenamento das informações coletadas.

A.1 Modelagem Dimensional da Arquitetura Star Schema

Esta seção apresenta a estrutura do Esquema Dimensional (*Star Schema*), utilizado para modelar o banco de dados analítico (*Data Mart*) da aplicação. O modelo representa a organização das Tabelas Fato e Tabelas Dimensão envolvidas no armazenamento, cruzamento e agregação das informações acadêmicas coletadas, bem como as restrições de integridade e os relacionamentos existentes entre elas para viabilizar consultas de alta performance.

Figura 53: Diagrama Dimensional da Arquitetura Star Schema do Data Mart do sistema proposto.



Fonte: Elaborado pelos autores (2026).

A.2 Dicionário de Dados

O dicionário de dados descreve as estruturas utilizadas na camada de transformação do projeto, organizada em modelo estrela (*star schema*). Nessa etapa, os dados extraídos do portal Integra são reorganizados em dimensões, tabelas fato e tabelas ponte, com o objetivo de apoiar consultas analíticas e a etapa de análise temática.

Para cada entidade do banco de dados transformado, são apresentados os atributos, seus respectivos tipos de dados e descrições, de forma a facilitar a compreensão da modelagem adotada.

A.2.1 Tabela: dim_instituicao

Tabela 7: Estrutura da entidade *dim_instituicao*

Campo	Tipo	Descrição
instituicao_id	BIGINT	Identificador único da instituição
sigla	TEXT	Sigla da instituição
nome_completo	TEXT	Nome completo da instituição
uf	TEXT	Unidade federativa da instituição
url	TEXT	Endereço eletrônico associado à instituição

Fonte: Elaborado pelos autores (2026).

A.2.2 Tabela: dim_campus

Tabela 8: Estrutura da entidade *dim_campus*

Campo	Tipo	Descrição
nome_campus	TEXT	Nome do campus
campus_id	BIGINT	Identificador único do campus

Fonte: Elaborado pelos autores (2026).

A.2.3 Tabela: dim_curso

Tabela 9: Estrutura da entidade *dim_curso*

Campo	Tipo	Descrição
nome_curso	TEXT	Nome do curso
nivel	TEXT	Nível do curso
curso_id	BIGINT	Identificador único do curso

Fonte: Elaborado pelos autores (2026).

A.2.4 Tabela: dim_pessoa

Tabela 10: Estrutura da entidade *dim_pessoa*

Campo	Tipo	Descrição
nome_pessoa	TEXT	Nome da pessoa
pessoa_id	BIGINT	Identificador único da pessoa

Fonte: Elaborado pelos autores (2026).

A.2.5 Tabela: *fato_tcc*

Tabela 11: Estrutura da entidade *fato_tcc*

Campo	Tipo	Descrição
tcc_id	BIGINT	Identificador único do TCC
titulo	TEXT	Título do trabalho
resumo	TEXT	Resumo ou descrição do TCC
palavras_chaves	TEXT	Palavras-chave associadas ao TCC
ano	TEXT	Ano do trabalho
curso_id	FLOAT	Referência ao curso associado
instituicao_id	BIGINT	Referência à instituição associada
campus_id	FLOAT	Referência ao campus associado

Fonte: Elaborado pelos autores (2026).

A.2.6 Tabela: *ponte_tcc_aluno*

Tabela 12: Estrutura da entidade *ponte_tcc_aluno*

Campo	Tipo	Descrição
tcc_id	BIGINT	Referência ao TCC
aluno_id	BIGINT	Referência ao aluno na dimensão pessoa

Fonte: Elaborado pelos autores (2026).

A.2.7 Tabela: *ponte_tcc_orientador*

Tabela 13: Estrutura da entidade *ponte_tcc_orientador*

Campo	Tipo	Descrição
tcc_id	BIGINT	Referência ao TCC
orientador_id	BIGINT	Referência ao orientador na dimensão pessoa

Fonte: Elaborado pelos autores (2026).

A.2.8 Tabela: *fato_artigo*

Tabela 14: Estrutura da entidade *fato_artigo*

Campo	Tipo	Descrição
artigo_id	BIGINT	Identificador único do artigo
slug_professor	TEXT	Identificador textual do professor
nome_professor	TEXT	Nome do professor vinculado ao artigo
sigla	TEXT	Sigla da instituição de origem
ano	TEXT	Ano do artigo
titulo	TEXT	Título do artigo
journal	TEXT	Nome do periódico ou revista
doi	TEXT	Identificador DOI do artigo
palavras_chaves	TEXT	Palavras-chave associadas ao artigo
resumo	TEXT	Resumo do artigo, quando disponível

Fonte: Elaborado pelos autores (2026).

A.2.9 Tabela: *fato_projeto*

Tabela 15: Estrutura da entidade *fato_projeto*

Campo	Tipo	Descrição
projeto_id	BIGINT	Identificador único do projeto
slug_professor	TEXT	Identificador textual do professor
nome_professor	TEXT	Nome do professor vinculado ao projeto
sigla	TEXT	Sigla da instituição de origem
titulo	TEXT	Título do projeto
descricao	TEXT	Descrição do projeto
natureza	TEXT	Natureza do projeto
equipe	TEXT	Integrantes da equipe do projeto
financiadores	TEXT	Instituições financiadoras do projeto

Fonte: Elaborado pelos autores (2026).

B Reconhecimento de Uso de Inteligência Artificial

Ferramentas de Inteligência Artificial generativa foram utilizadas como apoio complementar no desenvolvimento deste trabalho, especialmente em atividades relacionadas à revisão textual, reorganização estrutural de seções, refinamento gramatical e apoio à escrita acadêmica.

O uso dessas ferramentas ocorreu sob supervisão dos autores, que permaneceram responsáveis pela validação, adaptação, verificação e adequação de todo o conteúdo produzido ao contexto da pesquisa desenvolvida.

As ferramentas utilizadas não substituíram a elaboração intelectual dos autores, sendo empregadas exclusivamente como suporte auxiliar durante o processo de desenvolvimento do trabalho acadêmico.