

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DE SANTA CATARINA
CAMPUS SÃO JOSÉ

SCHAIANA SONAGLIO

**ANÁLISE DA ADERÊNCIA DAS DISCIPLINAS OPTATIVAS DO
CURSO DE ENGENHARIA DE TELECOMUNICAÇÕES DO IFSC –
CÂMPUS SÃO JOSÉ FRENTE ÀS DEMANDAS DO MERCADO
DE TRABALHO EM SANTA CATARINA**

SÃO JOSÉ

2025

Schaiana Sonaglio

Análise da aderência das disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC – Câmpus São José frente às demandas do mercado de trabalho em Santa Catarina

Monografia apresentada ao Curso de Engenharia de Telecomunicações do Instituto Federal de Santa Catarina, para a obtenção do título de bacharel em Engenharia de Telecomunicações.

Área de concentração: Telecomunicações

Orientador: Prof. Ederson Torresini, Me.

São José

2025

Schaiana Sonaglio

Análise da aderência das disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC – Câmpus São José frente às demandas do mercado de trabalho em Santa Catarina

Monografia apresentada ao Curso de Engenharia de Telecomunicações do Instituto Federal de Santa Catarina, para a obtenção do título de bacharel em Engenharia de Telecomunicações.

São José, 25 de julho de 2025.

Prof. Ederson Torresini, Me.
Instituto Federal de Santa Catarina

Prof. Cleber Jorge Amaral, Dr.
Instituto Federal de Santa Catarina

Prof. Clayrton Monteiro Henrique, Dr.
Instituto Federal de Santa Catarina

AGRADECIMENTOS

Ao meu esposo, pelo apoio constante, pelo aconchego e incontáveis cafés da manhã trazidos na cama.

Ao meu orientador, que, além de ser uma pessoa leve e divertida, muitas vezes teve mais motivação do que eu para que este trabalho fosse concluído.

À minha psicóloga, que me ajuda a enxergar a vida — e a mim mesma — com mais suavidade e compaixão.

Às minhas irmãs, que passaram por esta instituição antes de mim e abriram caminho para que a minha jornada fosse mais serena.

Aos meus pais, que se sacrificaram tanto ao deixarem o interior e recomeçarem, para que pudéssemos ter o que eles nunca tiveram.

E a mim mesma, por não desistir de nada que realmente importa para mim.

*Não é o crítico que importa, nem aquele que aponta onde o outro tropeçou.
O crédito pertence àquele que está realmente na arena,
com o rosto marcado de poeira, suor e sangue;
que luta bravamente, erra e recomeça, mas nunca se omite.
Que, se fracassar, fracassa ousando grandemente.*

T. Roosevelt

RESUMO

A formação acadêmica em Engenharia de Telecomunicações deve considerar, além dos conteúdos obrigatórios, o papel das disciplinas optativas na construção das competências profissionais. Inseridas como parte flexível do currículo, essas disciplinas têm o potencial de incorporar com mais agilidade as transformações tecnológicas e setoriais. Diante disso, investiga-se neste trabalho em que medida as competências desenvolvidas nas disciplinas optativas do curso de Engenharia de Telecomunicações do Instituto Federal de Santa Catarina (IFSC) – Câmpus São José se alinham às demandas do mercado de trabalho no estado de Santa Catarina. A metodologia consiste na extração de palavras-chave a partir dos objetivos das disciplinas optativas, normalizadas e validadas com base em sua recorrência em descrições de vagas. Em seguida, utiliza-se *web scraping* e mineração de texto para identificar essas palavras-chave em anúncios de emprego. Os resultados permitirão avaliar a aderência entre currículo e mercado, subsidiando tanto o aprimoramento das disciplinas quanto a tomada de decisão dos estudantes na escolha de suas trajetórias formativas.

Palavras-chave: Engenharia de Telecomunicações. Competências. Disciplinas optativas. Mercado de trabalho. Análise curricular.

ABSTRACT

The undergraduate program in Telecommunications Engineering must consider not only the mandatory content but also the strategic role of elective courses in developing professional competencies. As a more adaptable part of the curriculum, these courses are capable of incorporating technological and sectoral changes more rapidly. In this context, this study investigates the extent to which the competencies developed in the elective courses of the Telecommunications Engineering program at IFSC – São José Campus align with the demands of the labor market in the state of Santa Catarina. The methodology involves extracting keywords from the learning objectives of elective courses, followed by a process of semantic normalization and recurrence validation based on their frequency in job postings. Then, web scraping and text mining techniques are used to identify these keywords in job advertisements. The results aim to assess the alignment between curriculum and market, supporting both curricular improvement and student decision-making in selecting relevant elective courses.

Keywords: Telecommunications Engineering. Competencies. Elective courses. Labor market. Curriculum analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ingressantes (vagas novas) em cursos de engenharia (em milhares) . . .	14
Figura 2 – Concluintes / Ingressantes em cursos de engenharia (em milhares) . . .	15
Figura 3 – Ocorrência de palavras-chave por disciplina optativa	41
Figura 4 – Top 50 vagas com maior número de palavras-chave mapeadas (optativas)	42
Figura 5 – Ocorrência de palavras-chave	43

LISTA DE TABELAS

Tabela 1 – Dados da disciplina Banco de Dados (BCD)	28
Tabela 2 – Verbos extraídos dos objetivos da disciplina Banco de Dados e seus níveis na Taxonomia de Bloom	29
Tabela 3 – Competências da disciplina BCD o modelo Competências, Habilidades e Atitudes (CHA)	31

LISTA DE CÓDIGOS

Código 3.1 – Função para classificar frases segundo a metodologia CHA	30
Código 3.2 – Mensagem de erro - Robots.txt - Indeed	35
Código 3.3 – Erro HTTP retornado pelo Cloudflare	35
Código 3.4 – Autenticação na API da Apify	37
Código 3.5 – Execução do ator para cada termo	37
Código 3.6 – Gravação dos resultados localmente	37
Código 3.7 – Iteração sobre os arquivos JSON	38
Código 3.8 – Montagem da estrutura consolidada	38
Código 3.9 – Contagem de termos nas vagas	39
Código 3.10–Exemplo de relatório de frequência de termos	39
Código A.1–Mensagem de erro do robots.txt	53
Código B.1–Erro HTTP retornado pelo Cloudflare	54
Código C.1–Código-fonte do script <code>apify_indeed_scraper.py</code>	55
Código D.1–Código-fonte do script <code>consolidate_results.py</code>	60
Código E.1–Código-fonte do script <code>term_frequency_analysis.py</code>	61
Código F.1–Código-fonte do script <code>term_frequency_report_ascii.json</code>	68
Código G.1–Código-fonte do script <code>term_frequency_keywords.py</code>	93
Código H.1–Código-fonte do script <code>term_frequency_courses.py</code>	94

LISTA DE ABREVIATURAS E SIGLAS

API *Application Programming Interface.*

APUFSC Associação dos Professores da Universidade Federal de Santa Catarina.

BCD Banco de Dados.

CAPTCHA *Completely Automated Public Turing test to tell Computers and Humans Apart.*

CHA Competências, Habilidades e Atitudes.

CNPJ Cadastro Nacional da Pessoa Jurídica.

DBA Database Administrator.

DBDD *Database-driven Development.*

DBMS *Database Management System.*

DCN Diretrizes Curriculares Nacionais.

ESCO *European Skills, Competences, Qualifications and Occupations.*

IFSC Instituto Federal de Santa Catarina.

LGPD Lei Geral de Proteção de Dados Pessoais.

PPC Projeto Pedagógico do Curso.

RDBMS *Relational Database Management System.*

SGBD Sistema Gerenciador de Banco de Dados.

SQL *Structured Query Language.*

TF *Term Frequency.*

UC Unidade Curricular.

UCs Unidades Curriculares.

UFSC Universidade Federal de Santa Catarina.

SUMÁRIO

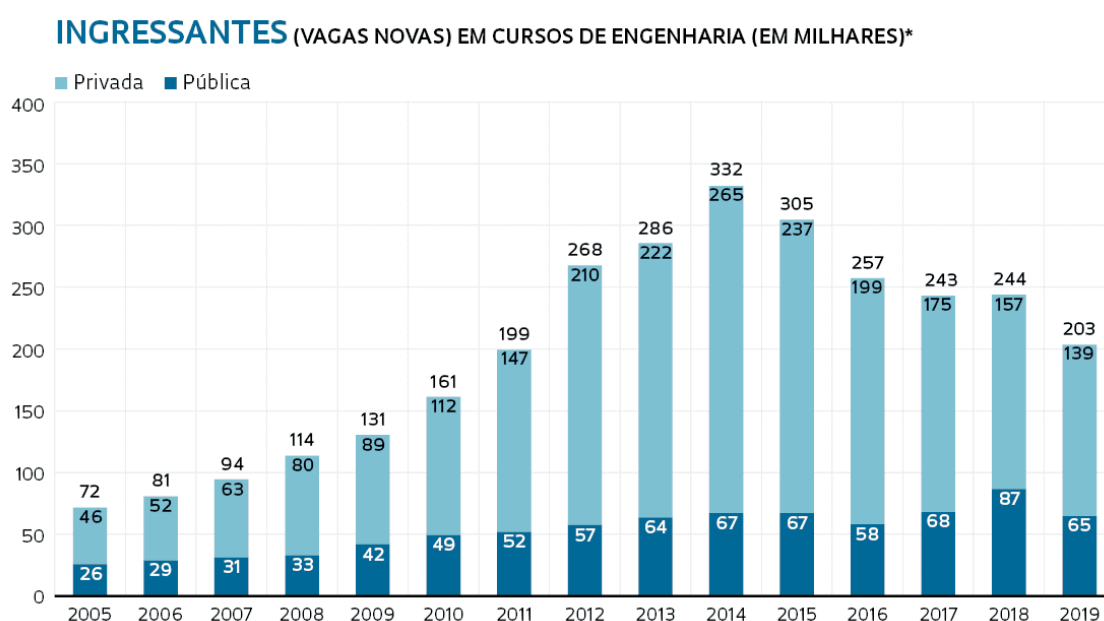
1	INTRODUÇÃO	14
1.1	OBJETIVOS	17
1.1.1	Objetivo geral	17
1.1.2	Objetivos específicos	17
2	REVISÃO BIBLIOGRÁFICA	18
2.1	O QUE SÃO OS INSTITUTOS FEDERAIS E PORQUE ELES EXISTEM	18
2.2	ENSINO POR COMPETÊNCIA	19
2.2.1	Modelo Competências, Habilidades e Atitudes (CHA): Conhecimentos, Habilidades e Atitudes	21
2.2.2	Taxonomias educacionais	22
2.2.3	Classificação heurística de competências	22
2.2.4	Geração de Palavras-chave	23
2.2.4.1	<i>Normalização Semântica</i>	23
2.2.4.2	<i>Validação por Recorrência</i>	24
2.3	<i>WEB SCRAPING</i>	24
2.3.1	<i>Web Scraping</i>: implicações legais e limitações técnicas	25
2.3.2	Mineração de Texto	26
3	DESENVOLVIMENTO	27
3.1	METODOLOGIA	27
3.1.1	Levantamento dos objetivos das disciplinas	28
3.1.2	Definição de verbos de acordo com a Taxonomia de Bloom e Krathwohl	28
3.1.3	Aplicação da metodologia CHA para definição de competências	29
3.1.4	Definição das palavras-chave	30
3.1.4.1	<i>Normalização semântica</i>	31
3.1.4.2	<i>Validação por recorrência</i>	33
3.2	COLETA E PROCESSAMENTO DE DADOS	34
3.2.1	Escolha da plataforma de empregos	34
3.2.2	Utilização da API da Indeed	34
3.2.3	<i>Uso de Web Scraping</i>	34
3.2.4	Desenvolvimento de um <i>script</i> para coleta de dados	35
3.2.5	Uso de um <i>script open source</i>	36
3.2.6	Ferramenta <i>Indeed Scraper - Apify</i>	36
3.2.7	Consolidação e processamento dos resultados	37

3.2.7.1	<i>Consolidação de arquivos</i>	38
3.2.7.2	<i>Análise de frequência de termos</i>	38
3.3	ANÁLISE DOS RESULTADOS	40
4	CONCLUSÕES	44
	Referências	46
	APÊNDICE A – MENSAGEM DE ERRO - ROBOTS.TXT - IN- DEED	53
	APÊNDICE B – MENSAGEM DE ERRO - CAPTCHA - INDEED	54
	APÊNDICE C – SCRIPT PARA COLETA DE DADOS VIA API DA APIFY	55
	APÊNDICE D – SCRIPT PARA CONSOLIDAÇÃO DOS ARQUI- VOS DE VAGAS	60
	APÊNDICE E – SCRIPT PARA ANÁLISE DE FREQUÊNCIA DE TERMOS	61
	APÊNDICE F – RESULTADO CONSOLIDADO - ANÁLISE DE FREQUÊNCIA DE PALAVRAS-CHAVE	68
	APÊNDICE G – SCRIPT PARA ILUSTRAR RESULTADOS: OCOR- RÊNCIA DE PALAVRAS-CHAVE	93
	APÊNDICE H – SCRIPT PARA ILUSTRAR RESULTADOS: OCOR- RÊNCIA DE PALAVRAS-CHAVE POR DISCI- PLINA	94

1 INTRODUÇÃO

Nos últimos anos, o Brasil registrou uma redução significativa no número de ingressantes e concluintes nos cursos de engenharia. Segundo a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2021), o número de novos ingressos caiu 17% entre 2018 e 2019, e 39% em comparação ao pico de 2014, conforme ilustrado na Figura 1.

Figura 1 – Ingressantes (vagas novas) em cursos de engenharia (em milhares)



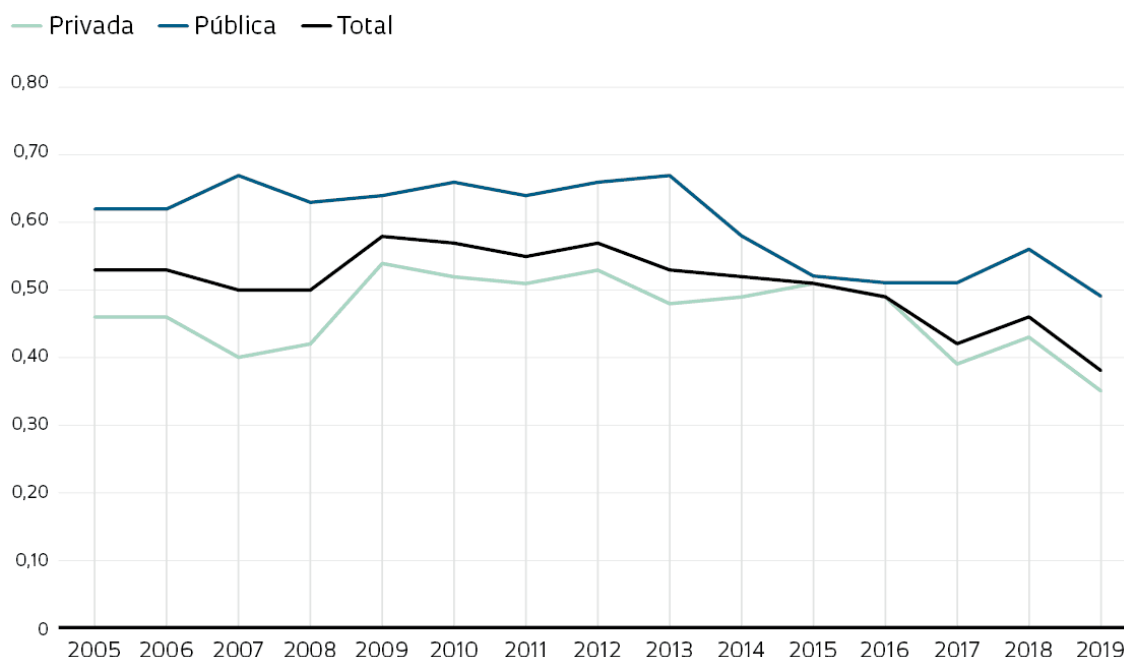
Fonte: FAPESP (2021)

Ao mesmo tempo, o índice de conclusão desses cursos também apresentou queda expressiva. Entre 2015 e 2019, o número de concluintes caiu 39%, segundo a mesma fonte. A Figura 2 mostra a relação entre concluintes e ingressantes, com defasagem de cinco anos, que é a duração média dos cursos de engenharia.

Esse cenário reflete uma série de desafios enfrentados pela formação em engenharia no Brasil. Embora a educação formal ainda seja considerada um requisito importante para o ingresso no mercado de trabalho, a realidade tem demonstrado que, por si só, ela não garante a inserção nem o sucesso profissional. As dificuldades de empregabilidade e o descompasso entre a formação acadêmica de engenharia e as exigências do mercado contribuem diretamente para o cenário de evasão e desinteresse dos estudantes (PARIZOTO, 2018).

Em escala nacional, segundo Borges e Almeida (2013), embora a formação técnica dos engenheiros continue sendo rigorosa, o mercado tem demonstrado baixa capa-

Figura 2 – Concluintes / Ingressantes em cursos de engenharia (em milhares)

CONCLUINTES/INGRESSANTES (DEFASAGEM DE CINCO ANOS)

Fonte: FAPESP (2021)

cidade de absorver esses profissionais em funções compatíveis com sua formação. Como consequência, muitos graduados acabam migrando para áreas administrativas ou tecnológicas, enquanto atividades técnicas típicas da engenharia vêm sendo progressivamente terceirizadas ou automatizadas. Esse processo de subutilização contribui para a perda de atratividade dos cursos, especialmente entre jovens que buscam formações mais conectadas às dinâmicas da economia contemporânea.

Esses desafios também se manifestam de forma evidente no estado de Santa Catarina. Um levantamento da Associação dos Professores da Universidade Federal de Santa Catarina (APUFSC) - Sindical revelou que a Universidade Federal de Santa Catarina (UFSC) teve uma queda de 34,8% na procura por seus cursos de graduação entre 2017 e 2025, com destaque para as Engenharias Civil e Mecânica, cujas relações candidato/-vaga caíram 77,9% e 68,6%, respectivamente, no período analisado (APUFSC-SINDICAL, 2025).

Além disso, uma reportagem publicada pela APUFSC em parceria com o jornal O Globo aponta que os cursos de engenharia no estado enfrentam dificuldades crescentes para atrair e reter estudantes, em um cenário agravado por um mercado de trabalho que demonstra baixa capacidade de absorver engenheiros recém-formados em funções compatíveis com sua qualificação (APUFSC-SINDICAL, 2025).

Dados específicos do IFSC – Câmpus Florianópolis, que oferta cursos de Engenharia Elétrica, Eletrônica, Mecatrônica e Civil desde 2013, indicam índices elevados de

evasão, com média de aproximadamente 26% — superior à média nacional de 22% para todos os cursos de graduação e de 23% para os cursos de Engenharia em geral (SILVA FILHO et al., 2007; ALVES NETO; CUSTÓDIO, 2023). A evasão nos cursos de Engenharia também pode ser explicada por fatores ligados à identidade profissional dos estudantes, à forma como se relacionam com o conhecimento e à integração social no ambiente acadêmico. Quando o aluno não se reconhece como futuro engenheiro, não percebe sentido nos conteúdos, torna-se mais difícil criar um vínculo com os estudos — o que contribui diretamente para a desmotivação e o abandono precoce do curso (ALVES NETO; CUSTÓDIO, 2023).

Essa mudança de interesse dos estudantes reforça a necessidade de refletir sobre a efetividade da formação acadêmica nos cursos de engenharia, principalmente em relação à adequação das habilidades e competências desenvolvidas em sala de aula às exigências reais do mercado de trabalho, que se mostra cada vez mais dinâmico.

Nesse contexto, destaca-se o papel dos Institutos Federais — como o IFSC – Câmpus São José — cuja missão, segundo Pacheco (2010), é oferecer uma formação técnica de qualidade, contextualizada e integrada às necessidades regionais. A proposta pedagógica dessas instituições busca equilibrar fundamentos científicos, competências profissionais e articulação com a realidade social e produtiva local.

As Diretrizes Curriculares Nacionais (DCN) para os cursos de graduação preveem a possibilidade de flexibilização curricular por meio da inclusão de componentes optativos, com o objetivo de permitir maior diversidade de itinerários formativos e adaptação às transformações sociais, científicas e tecnológicas (BRASIL, 2004). Nesse contexto, as disciplinas optativas desempenham um papel estratégico ao viabilizar a atualização contínua do currículo, possibilitando uma resposta mais ágil às demandas específicas do setor produtivo e às particularidades regionais.

Diante desse compromisso institucional e normativo com a formação adaptável ao mercado, este trabalho se concentra nas disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC - Câmpus São José. Pretende-se analisar em que medida suas competências estão alinhadas às exigências profissionais no estado de Santa Catarina, identificando convergências, lacunas e oportunidades de aprimoramento curricular. Além disso, busca-se oferecer aos estudantes subsídios mais objetivos para a escolha das disciplinas optativas mais coerentes com seus interesses e com as exigências atuais do mercado de trabalho.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Analisar a aderência das competências desenvolvidas nas disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC – Câmpus São José às exigências do mercado de trabalho em Santa Catarina.

1.1.2 Objetivos específicos

- Identificar as competências previstas nas disciplinas obrigatórias e optativas, conforme o Projeto Pedagógico do Curso (PPC);
- Construir um conjunto de palavras-chave representativas dessas competências, a partir de processos de normalização e validação terminológica;
- Verificar a presença dessas palavras-chave em descrições de vagas da área tecnológica publicadas no estado de Santa Catarina;
- Avaliar, com base nos dados coletados, a aderência das disciplinas optativas aos requisitos profissionais observados no mercado.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados os principais conceitos e referenciais teóricos que sustentam este trabalho, abordando temas relacionados à educação profissional e tecnológica, formação por competências, tecnologias e metodologias de análise curricular baseadas em dados, como mineração de texto, análise de frequência de termos e uso de *web scraping* para coleta automatizada de informações. A integração desses campos teóricos fornece a base para compreender o contexto da formação em Engenharia de Telecomunicações e justificar a metodologia adotada na pesquisa.

2.1 O QUE SÃO OS INSTITUTOS FEDERAIS E PORQUE ELES EXISTEM

O Decreto nº 7.566, de 23 de setembro de 1909, estabeleceu a criação de Escolas de Aprendizes Artífices nas capitais estaduais do Brasil, com a intenção de preparar os alunos com habilidades práticas diretamente aplicáveis ao mercado de trabalho (BRASIL, 1909).

Como destaca Almeida (2016), o foco inicial desses centros de formação era proporcionar aos estudantes uma educação técnica voltada para as necessidades da indústria local, promovendo uma interação direta entre o aprendizado e a produção industrial. Além disso, as escolas também incentivavam a participação dos alunos na criação e comercialização de artefatos, como parte de uma abordagem prática de ensino.

Ao longo das décadas, essas instituições passaram por diversas transformações, evoluindo para a configuração atual dos Institutos Federais. Durante esse processo, houve uma ampliação e diversificação dos cursos, que passaram a englobar não apenas a formação técnica, mas também o ensino médio e superior. O objetivo principal sempre foi manter a conexão com as necessidades econômicas locais, formando profissionais capacitados para contribuir com o desenvolvimento regional.

A criação do curso de Engenharia de Telecomunicações no IFSC - Câmpus São José segue esse compromisso institucional com a adaptação às necessidades do mercado local. Assim como as Escolas de Aprendizes Artífices foram estabelecidas para capacitar os jovens para áreas específicas e emergentes, o curso de Engenharia de Telecomunicações foi criado para suprir a carência de profissionais qualificados na área de telecomunicações.

O IFSC (2022) menciona no PPC do curso de Engenharia de Telecomunicações - IFSC - Câmpus São José:

Analizando as necessidades que o mercado local da Grande Florianópolis necessita, onde se localiza o IFSC Campus São José, sabendo da escassez de cursos de Engenharia de Telecomunicações no Sul do Brasil e que o quadro de engenheiros que atua nessa área são oriundos dos cursos de engenharia elétrica, eletrônica, ou mesmo computação, para os quais a formação mais específica acaba sendo adquirida nas próprias empresas do setor, o IFSC vem propor a criação deste curso que contribuirá para atender esta demanda de profissionais na região.

Nesse contexto institucional, as disciplinas optativas assumem papel estratégico na atualização curricular. Como previsto nas DCN para os cursos de graduação, a inclusão de componentes optativos permite maior flexibilidade formativa, favorecendo a diversificação de trajetórias acadêmicas e a adequação às transformações sociais, científicas e tecnológicas (BRASIL, 2004).

No caso específico do curso de Engenharia de Telecomunicações do IFSC – Câmpus São José, o PPC prevê não apenas a presença de disciplinas optativas na matriz curricular, mas também apresenta sugestões de temáticas para sua oferta. Essa previsão visa promover maior alinhamento entre a formação acadêmica e as demandas emergentes do mercado de trabalho e da sociedade.

Dessa forma, torna-se relevante investigar em que medida o currículo, especialmente por meio das disciplinas optativas, contribui para o desenvolvimento de competências condizentes com as exigências atuais do setor produtivo. Uma das abordagens que orienta essa adaptação curricular é a formação por competências, que será explorada na seção 2.2.

2.2 ENSINO POR COMPETÊNCIA

A ideia de competência tem origem na linguística, onde se introduziu o conceito de “competência linguística” como o conhecimento interno que uma pessoa tem sobre as regras da sua língua mesmo sem nunca ter estudado gramática formalmente (CHOMSKY, 1965). Essa concepção foi posteriormente reinterpretada no campo da educação, assumindo um novo significado: o de um conjunto de habilidades e conhecimentos mensuráveis, voltados ao desempenho prático em situações reais. Segundo Butova (2015), essa transição conceitual esteve ligada às transformações no mundo do trabalho e à busca por modelos de ensino mais eficazes e adaptados às exigências das sociedades pós-industriais.

No Brasil, essa abordagem passou a ganhar força a partir da década de 1990 com a reforma da Lei de Diretrizes e Bases da Educação Nacional (Lei nº 9.394/1996), que incorporou os princípios da educação por competências, voltando-se à formação integral do educando, ao desenvolvimento de capacidades práticas e à articulação entre educação e mercado de trabalho (BRASIL, 1996).

O foco na competência surge como uma crítica ao modelo educacional centrado na transmissão de conteúdos e na avaliação da memorização. Como aponta Perrenoud (1999), o ensino tradicional prepara estudantes para responder a perguntas conhecidas, mas não os capacita a resolver situações novas e complexas. A proposta da formação por competências é preparar o aluno para “agir com pertinência” em contextos reais, articulando saberes, saber-fazer e saber-ser.

Nesse modelo, a competência não é uma habilidade isolada ou uma tarefa mecânica, mas uma construção complexa que envolve a mobilização de múltiplos recursos (cognitivos, técnicos, sociais e afetivos) diante de um desafio. Tardif (2002) complementa que as competências são situadas: elas só podem ser avaliadas em situações autênticas de prática, e se desenvolvem gradualmente com a experiência.

O ensino por competências traz profundas implicações para a organização curricular. Ele exige a ruptura com a fragmentação disciplinar, a promoção de atividades integradoras, a definição de perfis de egresso baseados em desempenhos observáveis e a articulação entre conhecimentos acadêmicos e saberes do mundo do trabalho, como afirmam Ramos e Paranhos (2018):

A competência, enquanto princípio de organização curricular, implica a articulação de saberes disciplinares e não disciplinares, de saberes escolares e extraescolares, de saberes científicos e do senso comum, de saberes teóricos e práticos, de saberes técnicos e sociais. Envolve, portanto, a superação da fragmentação do conhecimento e da organização curricular baseada em disciplinas estanques, em favor de uma abordagem integrada e situada no contexto do mundo do trabalho.

No contexto do IFSC, a formação por competências está presente em cursos técnicos e superiores. O curso de Engenharia de Telecomunicações, por sua vez, propõe um conjunto de competências gerais e específicas no seu PPC, com, por exemplo, Projetos Integradores como estratégia para promover a interdisciplinaridade e a contextualização da aprendizagem.

A adoção do ensino por competência nas instituições de educação profissional requer modelos que orientem a definição e o desenvolvimento das competências a serem trabalhadas ao longo do processo formativo. Nesse contexto, destaca-se o modelo CHA, amplamente utilizado como referência para descrever os atributos esperados de um profissional em diferentes contextos ocupacionais e educacionais, conforme destaca Dutra (2004).

2.2.1 Modelo Competências, Habilidades e Atitudes (CHA): Conhecimentos, Habilidades e Atitudes

O modelo CHA é uma das estruturas mais utilizadas para a definição e análise de competências no campo da educação e da gestão de pessoas.

Esse modelo representa as três dimensões complementares de uma competência:

- **Conhecimentos (C):** referem-se ao saber teórico, técnico ou científico que o indivíduo detém sobre determinado assunto. São adquiridos por meio de estudos formais, leitura, cursos e experiências de aprendizagem estruturadas.
- **Habilidades (H):** dizem respeito ao saber-fazer, ou seja, à capacidade prática de aplicar os conhecimentos em situações reais. Estão ligadas a capacidades cognitivas, motoras, interpessoais e operacionais.
- **Atitudes (A):** compreendem o saber-ser, ou a disposição pessoal e social para agir com responsabilidade, iniciativa, ética e colaboração. Representam as motivações e valores que influenciam o comportamento do indivíduo.

Segundo Dutra (2004), o modelo CHA tem sido amplamente utilizado em projetos pedagógicos, diretrizes curriculares, programas de formação continuada e estruturas de avaliação de desempenho. Um exemplo de uso é a *European Skills, Competences, Qualifications and Occupations* (ESCO), que adota a estrutura para organizar e descrever competências profissionais em diversos setores da economia.

A aplicação da metodologia CHA permite a análise integrada das competências, favorecendo a identificação de lacunas formativas e o alinhamento entre a formação educacional e as exigências do mercado de trabalho.

No contexto do curso de Engenharia de Telecomunicações, a presença de disciplinas optativas representa uma oportunidade estratégica para adaptar a formação dos estudantes às exigências específicas do cenário profissional. Quando essas unidades curriculares são selecionadas com base em evidências extraídas de plataformas de recrutamento e análise de mercado, a formação se torna mais direcionada e condizente com as competências efetivamente demandadas, ampliando a chance de empregabilidade.

Contudo, para que a classificação de competências segundo o modelo CHA seja realizada de forma criteriosa e fundamentada, é necessário recorrer a estruturas educacionais consolidadas que auxiliem na identificação das ações cognitivas e atitudinais envolvidas. Nesse sentido, as taxonomias educacionais propostas por Anderson e Krathwohl (2001) oferecem subsídios teóricos relevantes para a análise dos verbos e objetivos de aprendizagem.

2.2.2 Taxonomias educacionais

A definição de objetivos educacionais passou a contar com maior sistematização a partir da publicação da Taxonomia de Bloom, em 1956. Essa taxonomia organiza o domínio cognitivo da aprendizagem em uma hierarquia composta por seis níveis: lembrar, compreender, aplicar, analisar, avaliar e criar. Cada nível representa um grau crescente de complexidade cognitiva e serve como base para a elaboração de objetivos de ensino e instrumentos de avaliação coerentes (BLOOM et al., 1956).

Posteriormente, uma versão revisada foi apresentada por Anderson e Krathwohl (2001), na qual houve mudanças na terminologia dos níveis, priorizando verbos no lugar de substantivos, além de uma reorganização da ordem hierárquica. Essa revisão visa maior aplicabilidade prática, sobretudo no ensino superior, e tem sido amplamente utilizada em propostas pedagógicas baseadas em competências (FERRAZ; BELHOT, 2010).

Além do domínio cognitivo, Krathwohl, Bloom e Masia (1964) desenvolveram a Taxonomia Afetiva, voltada à estruturação de objetivos relacionados a valores, atitudes, ética e disposição pessoal do estudante. Essa taxonomia é organizada em cinco níveis progressivos: receber, responder, valorizar, organizar e caracterizar por um valor ou valor complexo.

Dessa forma, a combinação da taxonomia cognitiva de Bloom et al. (1956) e da taxonomia afetiva de Krathwohl, Bloom e Masia (1964) oferece uma base conceitual para a análise e classificação de objetivos educacionais em competências estruturadas segundo o modelo CHA. Ao priorizar o uso de verbos operacionais, a conversão dos objetivos das disciplinas em competências observáveis e mensuráveis é facilitada e pode ser utilizada em propostas pedagógicas baseadas em competências.

Para que essas classificações possam ser utilizadas de maneira sistemática e replicável ao longo do trabalho, foi necessário desenvolver uma estratégia automatizada capaz de aplicar as categorias do modelo CHA a um conjunto amplo de objetivos educacionais. É nesse contexto que se insere o uso de uma função heurística, explicada na seção subseção 2.2.3.

2.2.3 Classificação heurística de competências

Para viabilizar a classificação automatizada das competências segundo o modelo CHA, foi adotado o conceito de função heurística, amplamente utilizado em ciência da computação, inteligência artificial e áreas correlatas. Segundo Russell e Norvig (2010), heurísticas são estratégias práticas que orientam a tomada de decisão com base em aproximações, permitindo encontrar soluções viáveis em contextos onde métodos exatos seriam inviáveis ou ineficientes.

Mais especificamente, uma função heurística consiste em uma função de avaliação

que estima o valor ou a relevância de uma alternativa com base em critérios predefinidos, guiando algoritmos de busca, classificação ou inferência (SHARMA; KUMAR; GROVER, 2020). Trata-se, portanto, de uma abordagem baseada em regras para resolver problemas complexos com rapidez e custo computacional reduzido.

Neste trabalho, a classificação heurística foi aplicada à análise automatizada de objetivos educacionais. Implementou-se uma função heurística capaz de classificar frases com base em verbos previamente associados às dimensões do modelo CHA. Essa estratégia permitiu aplicar os referenciais educacionais de forma sistemática e replicável, integrando-os ao processo de análise curricular com apoio computacional.

Como etapa complementar, foi realizada a geração de palavras-chave a partir das competências identificadas. Essa transformação busca aproximar a linguagem acadêmica dos termos comumente utilizados em descrições de vagas de emprego, aperfeiçoando a coleta de informações através de técnicas como *web scraping* e a análise automatizada do mercado de trabalho por meio de técnicas de mineração de texto.

2.2.4 Geração de Palavras-chave

A conversão das competências extraídas dos objetivos das disciplinas em palavras-chave tem como propósito aproximar a linguagem acadêmica — baseada em estruturas conceituais e objetivos educacionais — do vocabulário utilizado em descrições de vagas de emprego. Segundo Pejić Bach et al. (2020), essa abordagem facilita o alinhamento entre perfis formativos e demandas profissionais, permitindo uma comparação mais direta e objetiva entre o conteúdo curricular e os requisitos do mercado de trabalho.

Conforme afirmado por Firoozeh et al. (2019), a eficácia de métodos de extração de palavras-chave está diretamente ligada à capacidade de adaptação ao contexto específico de domínio, garantindo aplicabilidade no domínio alvo. Sabendo disso, algumas técnicas podem ser aplicadas para a extração de palavras-chave para o contexto específico de domínio, como competências advindas de disciplinas de um curso, para o domínio alvo, descrições de vagas de emprego, como:

2.2.4.1 Normalização Semântica

A normalização semântica consiste na unificação de diferentes formas linguísticas que compartilham o mesmo significado essencial. Esse processo pode envolver a consolidação de variações gráficas - como “multi thread” e “multithread”- ou a tradução entre idiomas - como “gerenciamento de memória” e “memory management”. Essa prática é comum na análise de conteúdo (BARDIN, 2016) e também é reconhecida na literatura de extração de palavras-chave como uma etapa fundamental para evitar redundância semântica e garantir consistência terminológica (PAPAGIANNOPPOULOU; TSOUMAKAS, 2019).

2.2.4.2 Validação por Recorrência

Após a normalização, a seleção da forma mais recorrente no contexto alvo é uma etapa essencial para selecionar as palavras-chave. Essa escolha, no contexto do mercado de trabalho, considera a frequência de uso dos termos em descrições de vagas e perfis ocupacionais. Segundo Khaouja, Kassou e Ghogho (2021), a análise de termos recorrentes em anúncios de emprego permite identificar quais competências e habilidades são efetivamente valorizadas, auxiliando no alinhamento entre currículos acadêmicos e exigências ocupacionais reais.

Após a aplicação das técnicas de Normalização Semântica e Validação por Recorrência para geração de palavras-chave, é possível avançar para a etapa de coleta das descrições de vagas de emprego. Para isso, são necessárias estratégias automatizadas, como *web scraping* e mineração de texto, que possibilitem a extração desses dados em larga escala.

2.3 WEB SCRAPING

O *web scraping*, ou raspagem de dados, consiste em uma técnica de extração automatizada de informações disponíveis em páginas da internet, realizada por meio de *scripts* ou ferramentas específicas. Essa prática tem aplicações em diversas áreas, como pesquisas acadêmicas, análise de mercado e monitoramento de conteúdos públicos. De acordo com Ferrara et al. (2014), a extração automatizada de dados provenientes da internet é reconhecida como uma ferramenta metodológica relevante em pesquisas científicas, especialmente naquelas voltadas à análise de sistemas complexos e ao estudo de comportamentos humanos em ambientes digitais.

Em contextos acadêmicos, a técnica de raspagem automatizada de dados é amplamente utilizada quando *Application Programming Interface* (API)s não disponibilizam informações estruturadas suficientes, sendo empregada para a coleta sistemática de dados públicos em larga escala (BRENNING; HENN, 2023). Segundo Struck (2025), a automação proporcionada por esses processos reduz a ocorrência de erros manuais e otimiza o tempo dedicado à coleta, tornando-se uma alternativa eficaz aos métodos tradicionais.

Ferramentas como o pacote **RCrawler**, desenvolvido por Khalil e Fakir (2017), ilustram o uso da raspagem automatizada no contexto da pesquisa científica, possibilitando a obtenção sistemática de dados públicos em larga escala.

Considerando sua ampla aplicação em ambientes digitais, é fundamental discutir os limites legais e as barreiras técnicas associados ao uso do *web scraping*. A utilização dessa técnica, embora metodologicamente válida, exige atenção quanto à proteção de dados e ao respeito às políticas de acesso das plataformas. A seção 2.3.1 aborda essas implicações, destacando os pontos previstos na Lei Geral de Proteção de Dados Pessoais

(LGPD), bem como alguns mecanismos técnicos que restringem acessos automatizados.

2.3.1 **Web Scraping: implicações legais e limitações técnicas**

No contexto da LGPD - Lei nº 13.709/2018, a proteção legal recai exclusivamente sobre dados pessoais, definidos como informações relacionadas a pessoa natural identificada ou identificável (BRASIL, 2018).

Dessa forma, dados institucionais ou empresariais, como razão social, Cadastro Nacional da Pessoa Jurídica (CNPJ), setor de atuação ou nome da empresa, não são considerados dados pessoais e, portanto, não se enquadram nas restrições diretas impostas pela LGPD.

A coleta automatizada dessas informações por meio de técnicas de *web scraping*, como em plataformas de anúncios de emprego, é legalmente permitida, desde que não envolva o tratamento de dados pessoais de indivíduos, como nomes de recrutadores ou contatos pessoais, os quais exigem base legal específica para tratamento.

Mesmo que a extração de dados das plataformas de anúncio de emprego seja legalmente permitida, existem mecanismos implementados para restringir o acesso automatizado e evitar a extração massiva de dados. Entre as barreiras técnicas mais comuns estão sistemas de detecção de robôs, como o uso do arquivo `robots.txt`, e uso de *Completely Automated Public Turing test to tell Computers and Humans Apart* (CAPTCHA) (VENUGOPALAN et al., 2024).

O arquivo `robots.txt` é um recurso previsto pelo protocolo *Robots Exclusion Protocol*, que orienta agentes automatizados sobre quais páginas do site podem ou não ser acessadas. Embora seu cumprimento não seja obrigatório nem possua força legal, é amplamente considerado uma norma ética no desenvolvimento de aplicações que realizam raspagem de dados (MUNZERT et al., 2015).

Outro mecanismo comum é o uso de CAPTCHA, que consiste em testes projetados para distinguir usuários humanos de *scripts* automatizados, como desafios visuais, reconhecimento de imagens ou digitação de códigos. Tais ferramentas desempenham um papel central na prevenção de acessos automatizados excessivos, dificultando ou até impedindo a coleta de dados por métodos de *web scraping* (AL-FANNAH, 2017).

Além de barreiras como `robots.txt` e CAPTCHAs, muitas plataformas utilizam *proxies* reversos para gerenciar e filtrar o tráfego de entrada. Esses sistemas interceptam as requisições, permitindo a aplicação de políticas automáticas de bloqueio, limitação de requisições - *rate limit* - e desafios adicionais, que estratégias eficazes para conter acessos automatizados excessivos (MOHAN; RENGARAJAN, 2024; NASIRIANI; ZOLA; MALAIYA, 2017).

O *Cloudflare* é um exemplo amplamente utilizado dessa tecnologia. Conforme ar-

quitetado pela própria empresa, seus *proxies* reversos inspecionam solicitações na borda da rede, aplicando mecanismos de segurança antes que o tráfego alcance os servidores de destino (CLOUDFLARE, 2025). Essa proteção garante maior controle sobre comportamentos automatizados, protegendo a estabilidade dos sistemas web frente a acessos não autorizados.

Após a coleta de dados automatizada por meio de *web scraping*, os dados extraídos podem então ser submetidos a técnicas de mineração de texto, que visam transformar o conteúdo textual bruto, como descrições de vagas, em informações estruturadas e analisáveis, permitindo identificar padrões de demanda e avaliar a aderência entre as competências do curso e as exigências do mercado de trabalho.

2.3.2 Mineração de Texto

A mineração de texto, também denominada *text mining*, refere-se ao processo de extração de informações relevantes a partir de dados textuais não estruturados, permitindo a identificação de padrões e frequências de termos (MACHADO; ARANHA; PASSOS, 2010; FERREIRA; CORREA, 2020).

Uma das técnicas empregadas nesse processo é a análise de frequência simples de termos (*Term Frequency* (TF)), que contabiliza o número absoluto de ocorrências de termos pré-definidos em um corpus textual. Essa abordagem é amplamente utilizada em tarefas de mineração de texto e recuperação da informação, por possibilitar a mensuração direta da relevância local de um termo dentro de um conjunto de documentos (SILVA, J.; SOUZA, M., 2007).

No contexto deste trabalho, os dados coletados por meio de *web scraping* foram analisados com base na métrica de TF, permitindo identificar quais termos aparecem com maior frequência nas descrições de vagas. Essa estratégia fornece uma medida objetiva e interpretável da demanda por determinadas competências no mercado de trabalho.

3 DESENVOLVIMENTO

Diante do contexto passado, este capítulo descreve a construção do processo analítico adotado. A proposta parte da organização dos dados curriculares do curso de Engenharia de Telecomunicações do IFSC - Câmpus São José e avança até a coleta e análise de informações extraídas de uma plataforma de vagas de emprego, para, por fim, verificar a aderência das competências desenvolvidas nas Unidades Curriculares (UCs) optativas do curso e aquelas exigidas pelo mercado de trabalho.

Para facilitar a compreensão do processo, este capítulo foi dividido em três seções:

3.1 – Metodologia: apresenta o processo de identificação das competências das UCs com a aplicação da metodologia CHA combinada com estratégias para transformar essas competências em palavras-chave.

3.2 – Coleta e Processamento de Dados: detalha a construção do *script* utilizado para coletar e armazenar descrições de vagas com base nas palavras-chave geradas, bem como o processo de cruzamento entre os dados coletados e as disciplinas optativas do curso.

3.3 – Análise dos resultados: discute os dados obtidos com a aplicação prática da metodologia e apresenta os principais padrões e tendências observados.

3.1 METODOLOGIA

A metodologia para identificação das competências e habilidades desenvolvidas durante o curso se dividiu nas seguintes seções:

3.1.1 - Levantamento dos objetivos das UCs de todos os eixos formativos com exceção do eixo básico, descritos no PPC;

3.1.2 - Definição de verbos de acordo com a Taxonomia de Bloom et al. (1956) e Krathwohl, Bloom e Masia (1964);

3.1.3 - Aplicação da metodologia CHA para identificar as competências associadas às UCs;

3.1.4 - Definição das palavras-chaves associadas às UCs.

Para mostrar o método aplicado, utilizaremos uma das UCs como exemplo. A Unidade Curricular (UC) escolhida, de forma arbitrária, foi BCD.

3.1.1 Levantamento dos objetivos das disciplinas

Primeiramente, os objetivos de todas as UCs, com exceção das matérias do eixo formativo básico, foram reunidos de forma manual em uma planilha eletrônica disponibilizada no *GitHub* (SONAGLIO, 2025b). Um exemplo pode ser visto na Tabela 1. As UCs do eixo formativo básico foram excluídas da análise porque as competências desenvolvidas nelas, por serem a base para as demais UCs, não seriam facilmente encontradas em vagas de emprego. Um exemplo são as disciplinas como Cálculo I, II, Álgebra Linear, que possuem objetivos como: “Compreender os conceitos e calcular integrais.”, “Resolver equações diferenciais lineares de primeira e segunda ordem.”, “Revisar as operações com matrizes, as regras de cálculo das matrizes inversas e de determinantes e os métodos de resolução de sistemas lineares”.

Tabela 1 – Dados da disciplina BCD

Disciplina	Banco de Dados
Objetivos	Projetar e modelar banco de dados relacional. Usar ferramentas livres de Sistema Gerenciador de Banco de Dados (SGBD) para criar e manter banco de dados. Desenvolver aplicações que façam uso de banco de dados. Conhecer conceitos fundamentais sobre organização da informação. Modelar banco de dados relacionais. Conhecer e usar sistemas de banco de dados relacionais. Desenvolver aplicações <i>desktop</i> e <i>web</i> que façam uso de banco de dados. Conhecer conceitos sobre bancos de dados não relacionais.

Fonte: Elaborado pela autora e baseado no PPC do curso (IFSC, 2022).

3.1.2 Definição de verbos de acordo com a Taxonomia de Bloom e Krathwohl

Após a coleta dos objetivos das UCs, utilizou-se a Taxonomia Cognitiva de Anderson e Krathwohl (2001) — uma revisão da taxonomia de Bloom et al. (1956) — com o objetivo de identificar os verbos centrais que expressam ações cognitivas nos objetivos educacionais.

Essa etapa é fundamental para a aplicação da metodologia CHA, pois fornece um critério sistemático para associar os objetivos das disciplinas a ações observáveis, que podem então ser classificadas como relativas a conhecimentos, habilidades ou atitudes. Em outras palavras, os verbos identificados por meio da taxonomia funcionam como ponte entre a formulação dos objetivos curriculares e a estrutura analítica da abordagem CHA.

O processo foi realizado de forma integrada, e não individualmente por disciplina, mas, para exemplificar esta parte da metodologia, continuaremos com o uso da UC de Banco de Dados. As seguintes etapas foram aplicadas:

1. **Identificação e normalização dos verbos selecionados:** Realizou-se a leitura dos objetivos das disciplinas presentes no PPC, com o intuito de identificar os verbos que estruturam os enunciados e que são mais utilizados. Na disciplina de BCD, os seguintes verbos foram identificados e normalizados em sua forma infinitiva:

- Projetar;
- Modelar;
- Usar;
- Criar;
- Desenvolver;
- Conhecer.

2. **Associação com a Taxonomia de Bloom e Krathwohl:** Os verbos então foram comparados com os níveis da Taxonomia de Bloom et al. (1956) e Anderson e Krathwohl (2001), a fim de relacioná-los às dimensões cognitivas proposta por eles. A Tabela 2 apresenta o resultado dessa associação.

Tabela 2 – Verbos extraídos dos objetivos da disciplina Banco de Dados e seus níveis na Taxonomia de Bloom

Verbo	Nível da Taxonomia de Bloom
Projetar	Criar
Modelar	Aplicar
Usar	Aplicar
Criar	Criar
Desenvolver	Criar
Conhecer	Lembrar

Fonte: Elaborado pela autora.

3.1.3 Aplicação da metodologia CHA para definição de competências

Após a identificação e classificação dos verbos com base na Taxonomia de Bloom et al. (1956) e Anderson e Krathwohl (2001), foi aplicada a metodologia CHA para identificar as competências associadas às UCs. A taxonomia cognitiva permitiu interpretar a intencionalidade pedagógica de cada verbo — ou seja, o tipo de operação esperada do estudante. Com isso, tornou-se possível associar esses verbos às três dimensões fundamentais do CHA mostradas abaixo:

- **Conhecimentos (C):** verbos que indicam aquisição e domínio teórico de conteúdos.
- **Habilidades (H):** verbos relacionados à aplicação prática de técnicas e procedimentos.
- **Atitudes (A):** verbos que expressam valores, posturas, colaboração e ética profissional.

A associação dos verbos aos componentes da metodologia CHA foi realizada de forma automatizada, por meio de um *script* desenvolvido especificamente para essa finalidade, disponibilizado no *GitHub* (SONAGLIO, 2025f). Com os verbos previamente agrupados por categoria, foi implementada uma função heurística simples, em linguagem *Python*, capaz de classificar automaticamente cada frase com base no verbo identificado. A Código 3.1 apresenta a função central responsável por essa tarefa, que identifica a dimensão correspondente (Conhecimentos, Habilidades ou Atitudes) com base em verbos-chave extraídos dos objetivos das disciplinas.

Código 3.1 – Função para classificar frases segundo a metodologia CHA

```

1 def classificar_cha(frase):
2     frase = frase.lower()
3     if any(verbo in frase for verbo in ["compreender", "lembrar", "conhecer", "
4         entender", "estudar", "analisar", "identificar"]):
5         return "C"
6     elif any(verbo in frase for verbo in ["aplicar", "criar", "usar", "utilizar", "
7         projetar", "resolver", "desenvolver", "implementar", "programar", "calcular", "
8         montar"]):
9         return "H"
10    elif any(verbo in frase for verbo in ["demonstrar", "atuar", "posicionar", "
        cooperar", "valorizar", "respeitar", "trabalhar"]):
            return "A"
    else:
        return "?"

```

Os resultados da execução do *script* foram revisados manualmente para garantir a consistência semântica entre os objetivos originais e a categoria CHA atribuída. O resultado da metodologia CHA aplicada na UC BCD é apresentada na Tabela 3. Os demais resultados estão disponibilizados no *GitHub* (SONAGLIO, 2025e).

3.1.4 Definição das palavras-chave

Após a identificação das competências por meio da metodologia CHA, foi feita a conversão dessas competências em palavras-chave. O objetivo foi aproximar a linguagem acadêmica — frequentemente baseada em descrições conceituais — do vocabulário

Tabela 3 – Competências da disciplina BCD o modelo CHA

Dimensão	Descrição da competência
Conhecimentos (C)	Modelagem relacional; fundamentos de bancos de dados; organização da informação.
Habilidades (H)	Projetar, criar e manter bancos de dados; desenvolver aplicações integradas.
Atitudes (A)	Interesse por boas práticas de armazenamento e manipulação de dados; atenção à integridade dos dados.

Fonte: Elaborado pela autora.

utilizado nas descrições reais de vagas de emprego, facilitando a posterior aplicação de técnicas de mineração de texto em plataformas de busca de oportunidades profissionais.

Para associar as competências identificadas às palavras-chave, adotou-se um critério combinado de normalização semântica e validação por recorrência.

Inicialmente, as competências foram analisadas com base em seu núcleo conceitual — ou seja, a ação principal, o campo de conhecimento ou a atitude descrita. Em seguida, realizaram-se buscas manuais com esses termos e suas variações em plataformas de emprego, com o intuito de verificar quais expressões são mais utilizadas no mercado de trabalho. As palavras-chave finais foram selecionadas com base em dois critérios:

- Fidelidade ao conteúdo original da competência
- Frequência significativa nas descrições de vagas, indicando maior potencial de empregabilidade e aderência à linguagem profissional.

3.1.4.1 Normalização semântica

A partir das competências extraídas da disciplina BCD ilustrada na Tabela 3, foram identificados os conceitos centrais:

- Modelagem de banco de dados relacionais;
- SGBD;
- Linguagem de manipulação e consulta de dados (*Structured Query Language* (SQL));
- Normalização e álgebra relacional;
- Desenvolvimento de aplicações com banco de dados;
- Bancos de dados não relacionais;
- Administração e otimização de banco de dados.

Com base nesses conceitos, realizou-se a etapa de normalização semântica, que consistiu na identificação de variações linguísticas, sinônimos e traduções frequentemente utilizadas no vocabulário técnico da área de tecnologia da informação. A normalização buscou alinhar os termos acadêmicos às formas mais utilizadas no mercado de trabalho, especialmente aquelas presentes em descrições de vagas.

As seguintes padronizações foram adotadas:

- **Modelagem relacional** foi normalizada para Modelagem de Dados e sua correspondente em língua inglesa - *Data Modeling* - e *Design* de Banco de Dados - com a correspondente em inglês - *Database Design* - termos amplamente utilizados em processos de estruturação de dados e arquitetura lógica.
- **SGBD** foi representado por siglas e tecnologias específicas como *Database Management System* (DBMS) e *Relational Database Management System* (RDBMS), *PostgreSQL*, *MySQL* e *SQLite*, que refletem as ferramentas do mercado.
- **Linguagem de manipulação de dados** foi padronizada como SQL, termo central em quase todas as descrições de vagas relacionadas a banco de dados.
- **Normalização e álgebra relacional** foram mantidas como Normalização e sua correspondente em inglês - *normalization* - e Álgebra Relacional e sua correspondente - *relational algebra*, por serem formas reconhecidas tanto no meio acadêmico quanto profissional.
- **Desenvolvimento de aplicações com banco de dados** foi representado por expressões como Desenvolvimento com Banco de Dados e a sua correspondência em língua inglesa - *Database-driven Development* (DBDD) - e Aplicações *Desktop* e *Web* com Banco de Dados e sua correspondência em língua inglesa - *Desktop and Web Applications with Database Integration*, que descrevem as habilidades práticas exigidas de desenvolvedores.
- **Bancos de dados não relacionais** foi associado a *NoSQL* e *MongoDB*, termos adotados em contextos de dados não estruturados e aplicações escaláveis.
- **Administração e desempenho de banco de dados** foi representado por Administração de banco de dados e a sua correspondência em inglês - *Database Administration* - e Otimização de *Queries* - *Query Optimization* - refletindo exigências frequentes para funções de Database Administrator (DBA) e *performance tuning*.

3.1.4.2 Validação por recorrência

Após a normalização semântica, realizou-se a validação por recorrência para confirmar quais das variações apresentavam maior aderência ao vocabulário do mercado de trabalho.

Essa etapa consistiu em buscas exploratórias em plataformas de emprego, como Indeed e *LinkedIn*, utilizando cada termo padronizado para verificar sua frequência de ocorrência em descrições de vagas da área de tecnologia. Aquelas expressões que apresentaram maior incidência foram mantidas como palavras-chave finais, assegurando a relevância prática das competências traduzidas.

Os resultados observados para BCD foram:

- *SQL*, *PostgreSQL* e *MySQL* apresentaram alta recorrência em vagas de desenvolvimento *backend* e análise de dados, sendo habilidades básicas exigidas em grande parte das descrições.
- Modelagem de dados e *design de banco de dados* apareceram com frequência em cargos voltados à arquitetura de sistemas, destacando a importância de compreender estrutura e organização da informação.
- Normalização e álgebra relacional surgiram em vagas que exigem conhecimentos sólidos em fundamentos teóricos e boas práticas de modelagem relacional.
- Bancos de dados *NoSQL* e *MongoDB* demonstraram presença significativa em vagas relacionadas a grandes volumes de dados, aplicações escaláveis e soluções modernas de armazenamento.
- Administração de banco de dados e otimização de *queries* foram encontrados com frequência em vagas para DBA.
- Desenvolvimento com banco de dados e aplicações *desktop* e *web* com banco de dados apareceram em descrições de cargos voltados à integração entre sistemas e persistência de dados, sobretudo em ambientes corporativos e sistemas de informação.

Os resultados para as demais disciplinas foram disponibilizados no *GitHub* (SONAGLIO, 2025d).

Com a definição das palavras-chave concluída, a próxima seção, 3.2, descreve como foi realizado o processo de coleta e processamento automatizado das descrições de vagas de emprego.

3.2 COLETA E PROCESSAMENTO DE DADOS

Com as palavras-chave definidas, foi implementado um processo automatizado para a coleta de vagas de emprego, de modo a viabilizar o tratamento de um grande volume de dados provenientes de múltiplas combinações de busca. Para essa finalidade, foi selecionada uma plataforma de ampla cobertura e relevância no mercado, a qual permitisse acesso programático aos dados de interesse, seja por meio de API pública ou por extração via *web scraping*.

Esta seção foi estruturada da seguinte forma:

3.2.1 - Escolha da plataforma de empregos para buscas das vagas;

3.2.2 - Primeira abordagem para coleta de dados: uso da API pública Indeed;

3.2.3 - Segunda abordagem para coleta de dados: *Web Scraping*.

3.2.1 Escolha da plataforma de empregos

A plataforma escolhida como fonte primária para a coleta de dados foi a plataforma Indeed. A escolha se fundamenta num fator que assegura a relevância e a eficácia da pesquisa: o Indeed é amplamente reconhecido como um dos maiores agregadores de vagas de emprego em escala global, abrangendo uma ampla variedade de setores e regiões. Essa abrangência proporciona uma amostra diversificada e representativa das exigências do mercado, aspecto essencial para análises precisas e generalizáveis.

3.2.2 Utilização da API da Indeed

Após a escolha da plataforma, a primeira abordagem foi utilizar a API pública do Indeed para coletar os dados das vagas de emprego para posterior análise. Entretanto, essa abordagem foi descartada, uma vez que a documentação da API da Indeed não disponibiliza consulta de vagas com exceção das de criação própria (INDEED, 2025), impossibilitando a coleta de dados por esse meio.

3.2.3 Uso de *Web Scraping*

Como a API pública do Indeed não disponibiliza consulta de vagas nos critérios desejados, ou seja, vagas de qualquer empregador e não de criação própria, fez-se necessária uma abordagem alternativa onde fosse possível obter os dados das vagas da plataforma. Essa segunda abordagem foi a técnica de *web scraping*.

Neste contexto, objetivo era implementar um *script* para coletar descrições textuais de vagas de emprego publicadas e limitar a coleta a informações públicas e não identificáveis, como títulos e descrições de cargos, sem envolver quaisquer dados pessoais

de candidatos ou empregadores, mantendo a conformidade com a LGPD. A atividade teve finalidade estritamente acadêmica, sem qualquer uso comercial.

A implementação do *script* para *web scraping* concentrou-se na identificação de vagas relacionadas às palavras-chave previamente definidas na Seção 3.1.4.

3.2.4 Desenvolvimento de um *script* para coleta de dados

A implementação do *script* para *web scraper*, disponibilizado no *GitHub* (SONA-GLIO, 2025a), foi focada em realizar buscas na plataforma e armazenar localmente as descrições das vagas encontradas para posterior análise.

A primeira versão implementada utilizou técnicas básicas, sem a aplicação de medidas específicas para evitar a detecção por mecanismos *anti-bot* ou desafios de verificação, como CAPTCHA. Nessa fase inicial, adotou-se um identificador de agente de usuário (*user-agent*) padrão, respeitando as diretrizes estabelecidas no arquivo `robots.txt` da plataforma.

Ao executar o *script*, descobriu-se que o arquivo `robots.txt` proíbe a realização de *scraping* em qualquer parte do site. A mensagem do código A de erro foi retornada.

Código 3.2 – Mensagem de erro - Robots.txt - Indeed

```
1 Access to search URL br.indeed.com/jobs?q=Scrum&l=Santa%20Catarina&from=searchOnDesktopSerp is disallowed by robots.txt.
```

As tentativas foram bloqueadas pelo sistema de proteção *Cloudflare*, que empregam técnicas avançadas de detecção de tráfego automatizado.

Diante da tentativa falha, foram introduzidos atrasos randômicos entre as requisições, a fim de simular o comportamento humano. Ainda assim a adaptação não foi suficiente, dado que o site apresentou desafios de CAPTCHA, exigindo novas estratégias para prosseguir com a automação. Ao tentar resolver os desafios de CAPTCHA, a mensagem do código B de erro foi retornada.

Código 3.3 – Erro HTTP retornado pelo Cloudflare

```
1 HTTPError: 403 Client Error: Forbidden for url: br.indeed.com/jobs?q=Scrum&l=Santa%20Catarina&from=searchOnDesktopSerp
2 Detail: Cloudflare verification failed. Missing or invalid cf_clearance cookie
```

Para lidar com os CAPTCHA, a abordagem foi modificada: passou-se a detectar a presença do desafio da *Cloudflare* e salvar a página correspondente, permitindo a resolução manual do CAPTCHA e, então, a continuidade do *scraping*. Contudo, mesmo após a resolução manual do CAPTCHA, o site ainda identificava as requisições como provenientes

de *scripts* e bloqueava o acesso, mostrando que a política de proteção do Indeed cumpriu o seu papel, sendo necessário optar por outra alternativa, descrita na subseção 3.2.5.

3.2.5 Uso de um *script open source*

Diante das tentativas falhas mencionadas anteriormente, uma alternativa viável foi procurar por uma ferramenta de *scraping open source*. A ferramenta utilizada foi a *Indeed Job Scraper* (GANA EBENEZER, 2023).

No entanto, essa ferramenta também apresentou limitações, sendo incapaz de contornar os mecanismos de proteção da plataforma pelos mesmo desafios citados na subseção 3.2.4. A alternativa então foi buscar alguma ferramenta comercial que focasse em *web scraping*.

3.2.6 Ferramenta *Indeed Scraper - Apify*

Como as tentativas de coleta automatizada de dados da plataforma Indeed enfrentaram diversas barreiras técnicas, como a presença de mecanismos de proteção contra *scraping*, optou-se então por utilizar uma ferramenta paga, a API da Apify (2025), mais especificamente o ator *misceres/indeed-scraper*, para realizar a coleta automatizada de dados de vagas.

A *Apify* é uma plataforma para execução remota e automatizada de tarefas de extração de dados, baseada em unidades reutilizáveis chamadas de atores - *actors*. Diferentemente de bibliotecas locais, a ferramenta opera em nuvem e permite o controle total via API, o que facilita a automação de coletas recorrentes e a persistência de resultados estruturados em diferentes formatos.

Para interagir com a API, foi desenvolvido o *script* disponibilizado no *GitHub* (SONAGLIO, 2025c). Esse *script* utiliza a *Apify API Client for Python*, que é a biblioteca oficial da *Apify* para execução de atores. A estrutura geral do *script* consiste em:

- Autenticação via *token* de acesso;
- Definição de uma lista com termos de busca - palavras-chaves de todas as disciplinas;
- Execução de um ator da *Apify* para cada termo;
- Salvamento local dos resultados obtidos.

Inicialmente, o *script* se autentica na API utilizando o *token* pessoal de acesso, conforme o código 3.4:

Código 3.4 – Autenticação na API da Apify

```

1 from apify_client import ApifyClient
2
3 API_TOKEN = "apify_api_XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
4 client = ApifyClient(API_TOKEN)

```

Em seguida, é definida uma lista com mais de 100 palavras-chaves associadas aos conteúdos da formação. Para cada item da lista, é executado o ator remoto na plataforma *Apify*, com a delimitação de região e quantidade de resultados, como mostra o trecho de código 3.5:

Código 3.5 – Execução do ator para cada termo

```

1 for i, term in enumerate(search_terms, 1):
2     run_input = {
3         "country": "BR",
4         "location": "Santa Catarina",
5         "position": term,
6         "maxItems": 50
7     }
8     run = client.actor("misceres~indeed-scraper").call(run_input=run_input)
9     client.run(run["id"]).wait_for_finish()

```

Por fim, os dados retornados pela execução, ou seja, as vagas retornadas para cada palavra-chave, são salvos localmente em arquivos no formato `.json`, com nomes derivados dos termos buscados, como mostra o trecho de código 3.6):

Código 3.6 – Gravação dos resultados localmente

```

1 filename = f"results_{i:02d}_{sanitize_filename(term[:40])}.json"
2 with open(filename, "w", encoding="utf-8") as f:
3     json.dump(dataset_items, f, ensure_ascii=False, indent=2)

```

Os dados brutos coletados por meio do *script* descrito na Seção 3.2.6 foram organizados em arquivos no formato `.json` e disponibilizados no *GitHub* (SONAGLIO, 2025c). Esse material contém os resultados para cada termo de busca empregado na coleta de vagas.

3.2.7 Consolidação e processamento dos resultados

Após a raspagem automatizada (técnica de *web scraping*) das vagas por meio da plataforma *Apify*, foi necessário realizar o tratamento e análise dos dados brutos salvos em arquivos no formato `.json`. Para isso, dois *scripts* em *Python* foram desenvolvidos: o primeiro para consolidação dos resultados em formato tabular e o segundo para análise de frequência de termos técnicos.

3.2.7.1 Consolidação de arquivos

O *script* listado no Apêndice D, percorre todos os arquivos `results_*.json` gerados durante a coleta de dados e os agrupa em um único arquivo no formato `.csv`, facilitando a inspeção tabular.

O processo se inicia com a identificação dos arquivos no diretório atual, como mostra o trecho de código 3.7:

Código 3.7 – Iteração sobre os arquivos JSON

```
1 for filename in os.listdir(folder_path):
2     if filename.startswith("results_") and filename.endswith(".json"):
3         with open(os.path.join(folder_path, filename), encoding="utf-8") as f:
4             data = json.load(f)
```

Em seguida, os campos principais de cada vaga (termo de busca, título, empresa, localização, descrição e link) são normalizados e armazenados em um `DataFrame`, conforme o trecho de código 3.8:

Código 3.8 – Montagem da estrutura consolidada

```
1 for job in data:
2     all_items.append({
3         "Termo de busca": term,
4         "Título": job.get("title"),
5         "Empresa": job.get("company"),
6         "Local": job.get("location"),
7         "Descrição": job.get("description"),
8         "Link": job.get("url")
9     })
```

Por fim, os dados consolidados são gravados no arquivo `vagas_unificadas.csv`.

3.2.7.2 Análise de frequência de termos

O segundo *script*, listado no Apêndice E, realiza a contagem da frequência de palavras-chave nas descrições das vagas. O objetivo é identificar quais competências aparecem com maior recorrência, a partir da lista de palavras-chave definida.

O *script* percorre os arquivos JSON e, para cada vaga, concatena os campos de texto e aplica expressões regulares que respeitam as bordas de palavra (`\b`). O trecho de código 3.9 mostra a lógica central de contagem:

Código 3.9 – Contagem de termos nas vagas

```

1 for job in jobs:
2     texto = " ".join([
3         job.get("title", ""),
4         job.get("description", ""),
5         job.get("skills", ""),
6         job.get("requirements", "")
7     ]).lower()
8
9     for term in search_terms:
10        if re.search(r'\b' + re.escape(term.lower()) + r'\b', texto):
11            counts[term]["matches"] += 1
12            counts[term]["jobs"].add(job.get("id") or filename)

```

Ao final da execução, é gerado o relatório `term_frequency_report.json`, contendo as estatísticas estruturadas para cada termo (número de ocorrências e número de vagas únicas em que apareceu). Um trecho deste resultado pode ser visto no trecho de código F.

Código 3.10 – Exemplo de relatório de frequência de termos

```

1 [
2     {
3         "term": "DWDM",
4         "match_count": 3,
5         "unique_jobs": 1
6     },
7     {
8         "term": "GPON",
9         "match_count": 18,
10        "unique_jobs": 6
11    },
12    {
13        "term": "Micro-ondas",
14        "match_count": 6,
15        "unique_jobs": 5
16    },
17    {
18        "term": "Internet das Coisas",
19        "match_count": 2,
20        "unique_jobs": 1
21    },
22    {
23        "term": "IoT",
24        "match_count": 50,
25        "unique_jobs": 16
26    },

```



```

27 {
28   "term": "SQL",
29   "match_count": 319,
30   "unique_jobs": 150
31 }
32 ]

```

Esses dois *scripts* completam o pós-processamento, permitindo que os dados brutos da coleta sejam utilizados de forma estruturada e quantitativa para análise.

3.3 ANÁLISE DOS RESULTADOS

Os dados consolidados no Apêndice F permitem avaliar o grau de aderência entre as competências desenvolvidas pelas disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC e as exigências observadas no mercado catarinense. Nesta seção, os resultados são analisados em relação ao desafio identificado na Introdução: a dificuldade de inserção dos engenheiros no mercado de trabalho devido à possível defasagem entre as competências desenvolvidas no curso e as demandas profissionais em Santa Catarina.

A Figura 3 demonstra que Engenharia de Software e Banco de Dados concentram, respectivamente, 45,4% e 26,2% de todas as ocorrências de palavras-chave mapeadas. Essa predominância sugere que as competências abordadas nessas disciplinas possuem maior aderência às tecnologias e práticas atualmente exigidas em anúncios de vagas de emprego.

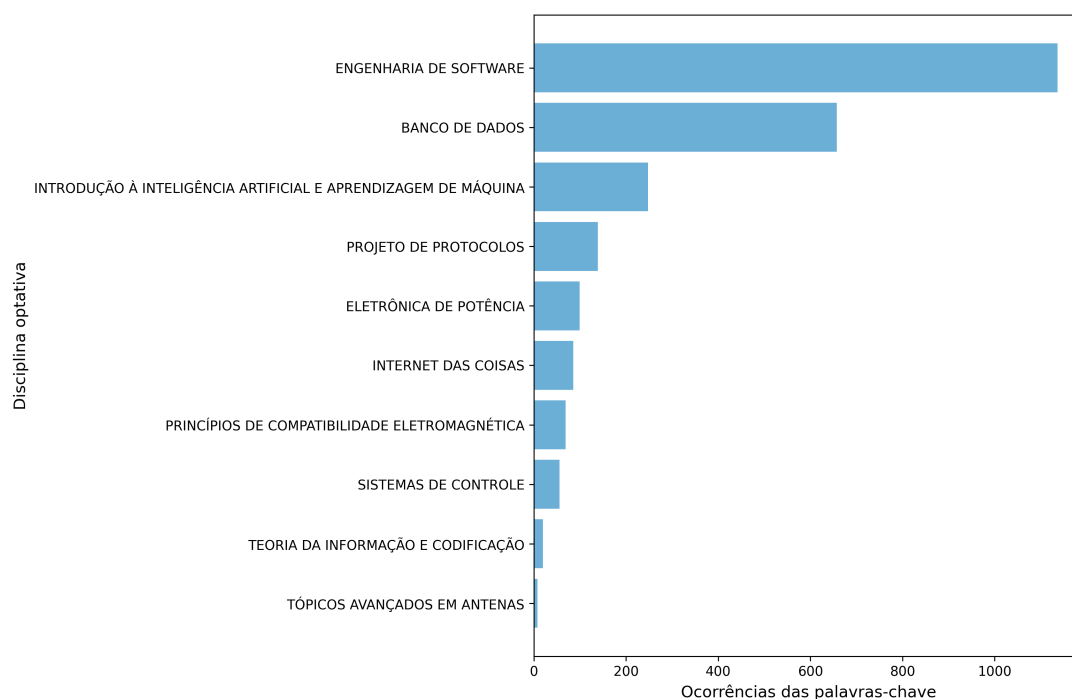
Em um segundo grupo aparecem as disciplinas Introdução à Inteligência Artificial e Aprendizagem de Máquina (9,8%) e Projeto de Protocolos (5,5%). A presença recorrente dessas unidades curriculares aponta para uma demanda concreta por competências relacionadas à ciência de dados, aprendizado de máquina e à aplicação prática de projetos com metodologias ativas — aspectos valorizados por empresas de tecnologia em ambientes inovadores.

Por outro lado, disciplinas como Teoria da Informação e Codificação e Tópicos Avançados em Antenas apresentaram porcentagens significativamente menores. A Figura 3 mostra que, somadas, essas três disciplinas representam menos de 0,9% das palavras-chave identificadas, evidenciando sua baixa representatividade no conjunto de competências demandadas pelas vagas analisadas.

A Figura 4 mostra a distribuição das 50 vagas com maior número de palavras-chave mapeadas a partir das disciplinas optativas. Esse gráfico permite observar como determinadas vagas concentram múltiplas competências alinhadas ao currículo optativo do curso.

Observa-se que cargos voltados a desenvolvimento de software e ciência de dados —

Figura 3 – Ocorrência de palavras-chave por disciplina optativa



Fonte: Elaborado pela autora.

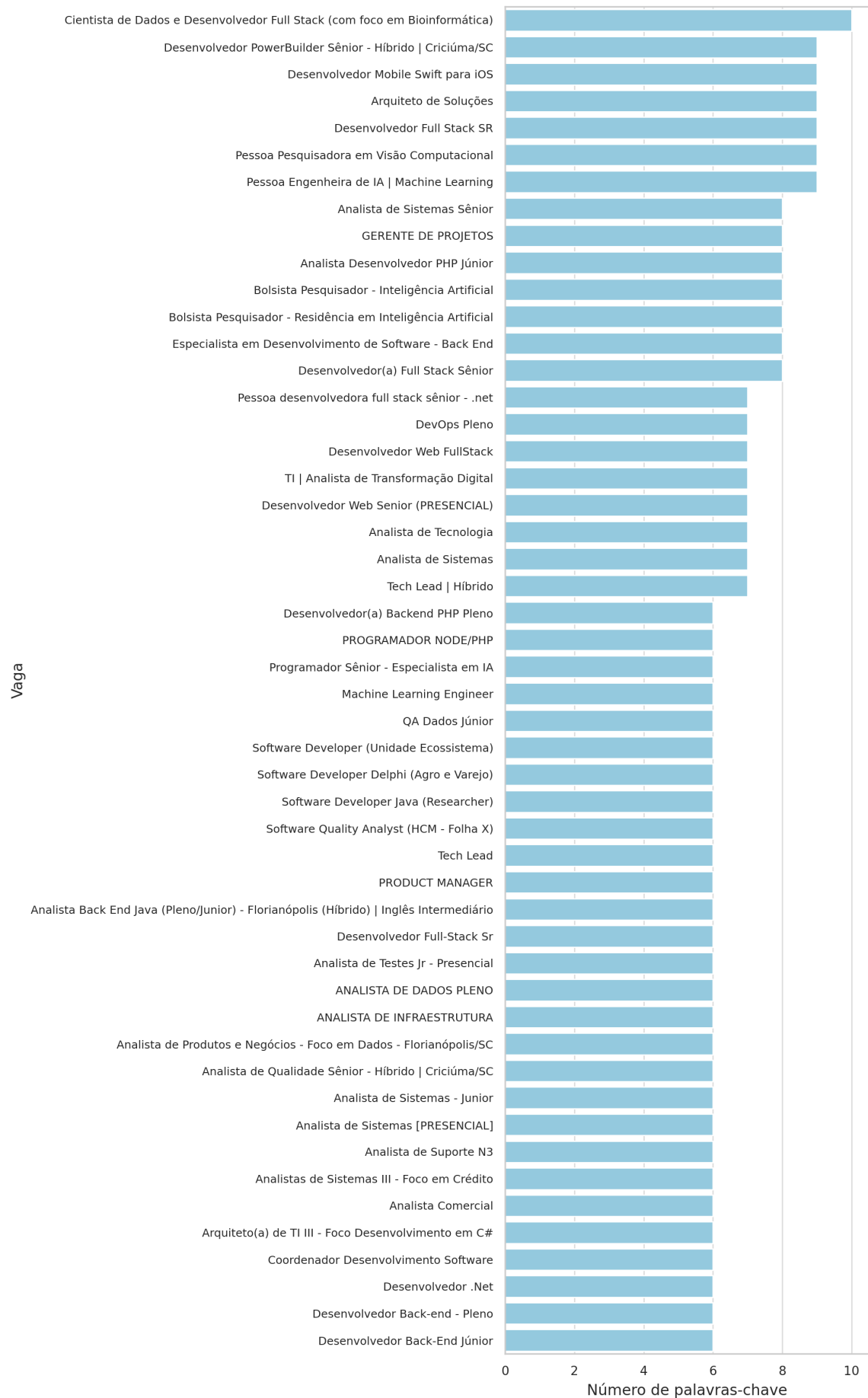
como Cientista de Dados e Desenvolvedor de Software — lideram em número de palavras-chave identificadas. Isso evidencia a forte presença de competências ligadas a engenharia de software, bancos de dados e metodologias ágeis, o que corrobora a análise anterior e indica que a formação optativa se alinha com as exigências do setor produtivo digital.

Em contrapartida, disciplinas como Tópicos Avançados em Antenas e Teoria da Informação e Codificação apresentaram baixa ou nenhuma incidência no ranking. Essa ausência reforça a ideia de que essas competências têm tido menor demanda no mercado catarinense, ou que estão sendo absorvidas em menor escala por empresas locais.

Essa distribuição também mostra a diversidade das vagas em que os egressos do curso podem atuar, com competências transferíveis sendo valorizadas em diferentes contextos — de pesquisa a ambientes corporativos. Conforme demonstrado na Figura 5, a ampla presença de termos como *SQL*, Metodologias Ágeis, *Kanban*, *Scrum*, *PostgreSQL* indica uma ênfase no domínio de ferramentas e práticas de desenvolvimento de software, o que reforça a importância de mantê-las no projeto pedagógico.

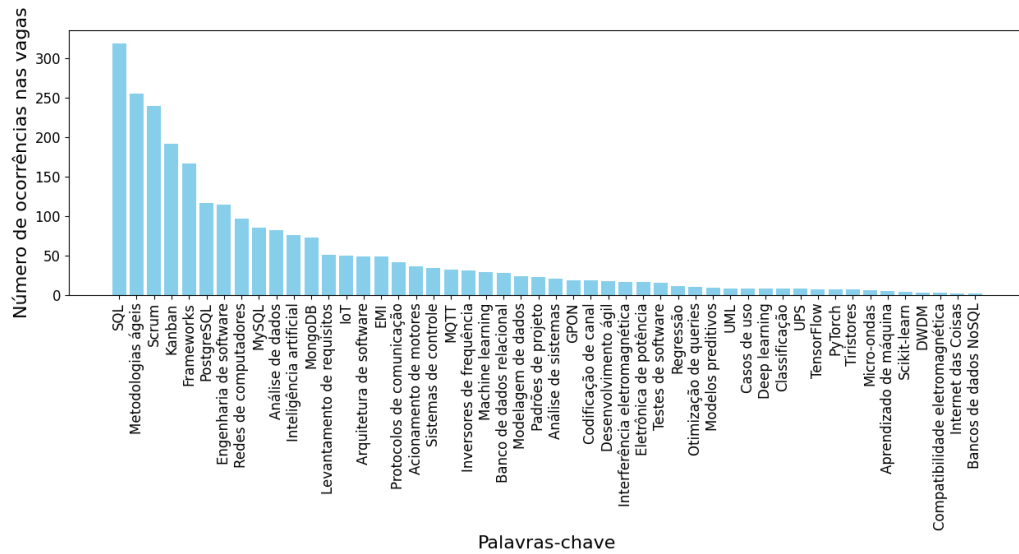
Os resultados da análise apontam a importância de que a política de flexibilização curricular prevista pelas DCN seja implementada de maneira dinâmica e sensível às transformações do setor produtivo. Nesse sentido, a realização periódica de análises empíricas, como a conduzida neste estudo, pode subsidiar a atualização das disciplinas optativas e fortalecer a aderência do currículo às demandas profissionais emergentes.

Figura 4 – Top 50 vagas com maior número de palavras-chave mapeadas (optativas)



Fonte: Elaborado pela autora.

Figura 5 – Ocorrência de palavras-chave



Fonte: Elaborado pela autora.

Além das implicações curriculares discutidas, a partir de análises como a conduzida neste estudo, seria possível oferecer aos estudantes recomendações mais precisas sobre quais disciplinas optativas melhor se alinham aos seus interesses e às demandas profissionais predominantes no mercado regional. Essa abordagem orientada por evidências pode contribuir para decisões mais informadas ao longo do percurso formativo, promovendo maior engajamento, adaptabilidade e inserção no mercado de trabalho.

4 CONCLUSÕES

Este trabalho teve como objetivo principal analisar a aderência das competências desenvolvidas nas disciplinas optativas do curso de Engenharia de Telecomunicações do IFSC – Câmpus São José frente às exigências do mercado de trabalho no estado de Santa Catarina. A motivação para essa investigação partiu da constatação de que os cursos de engenharia no Brasil vêm enfrentando queda na atratividade e na taxa de conclusão, em parte devido ao descompasso entre a formação acadêmica e as demandas profissionais.

A metodologia adotada envolveu a identificação das competências previstas no PPC do curso através da metodologia CHA e a conversão destas em palavras-chave por meio de processos de normalização terminológica. Com as palavras-chave determinadas, foi feita a análise da ocorrência desses termos em descrições de vagas da plataforma Indeed, focando no estado de Santa Catarina. Esse procedimento permitiu verificar, de forma empírica, o grau de correspondência entre o conteúdo ofertado pelas disciplinas optativas e os requisitos mais recorrentes no mercado catarinense.

Os resultados indicaram que disciplinas como Engenharia de Software, Banco de Dados e Introdução à Inteligência Artificial e Aprendizagem de Máquina apresentam elevada aderência às competências demandadas, evidenciando pontos de convergência entre a formação ofertada e o perfil profissional esperado. Em contrapartida, outras disciplinas, como Teoria da Informação e Codificação e Tópicos Avançados em Antenas, apresentaram baixa incidência de termos associados às vagas analisadas, sugerindo possíveis lacunas a serem consideradas em futuros processos de revisão e atualização curricular.

O estudo também demonstrou o potencial de abordagens orientadas por dados como instrumento de apoio à gestão acadêmica e à orientação discente. A análise baseada na frequência e distribuição de palavras-chave permite subsidiar decisões sobre o planejamento curricular e fornecer aos estudantes informações objetivas para a escolha de componentes optativos mais alinhados às exigências do setor produtivo.

Como trabalhos futuros, abrem-se possibilidades para aprofundar e expandir a metodologia adotada neste estudo. Uma possibilidade seria a construção de uma taxonomia das vagas de emprego na área de tecnologia em Santa Catarina, com o objetivo de identificar com maior precisão quais competências são prioritariamente exigidas de engenheiros recém-formados. Esse mapeamento permitiria medir, de forma sistemática, o percentual de aderência do curso às exigências reais do mercado.

Outra aplicação relevante seria empregar a metodologia desenvolvida como ferramenta de apoio à proposição de novas disciplinas optativas. A análise de palavras-chave, aliada aos *scripts* desenvolvidos, pode orientar a criação de componentes curriculares de

maneira orientada por dados — promovendo uma atualização mais ágil e embasada nas tendências do mercado de trabalho.

De forma semelhante, a abordagem pode ser estendida à revisão de disciplinas já existentes, tanto na base obrigatória quanto no conjunto de optativas. O cruzamento entre conteúdos curriculares e demandas do mercado viabiliza decisões pedagógicas mais informadas, com potencial para aumentar a relevância prática do curso.

Por fim, recomenda-se ampliar o escopo da análise para outras plataformas de emprego além da utilizada neste estudo, como o LinkedIn. Alternativamente, plataformas agregadoras como a TheirStack — que consolida dados de mais de 16 mil fontes, incluindo o próprio LinkedIn, Indeed e Glassdoor (THEIRSTACK, 2025) — podem ser incorporadas à metodologia. Essa ampliação permitiria uma coleta de dados mais abrangente, automatizada e atualizada, fortalecendo a robustez da análise e permitindo a detecção de padrões emergentes de competências e tecnologias exigidas no mercado.

REFERÊNCIAS

- ALLAHYARI, Mehdi et al. A brief survey of text mining: classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*, 2017.
- ALMEIDA, Alcides Vieira de. *100 anos de Escolas de Aprendizes Artífices*. 1. ed. Florianópolis: Publicações do IF-SC, 2016.
- ALVES NETO, José de Pinho; CUSTÓDIO, José Francisco. Evasão nos cursos de Engenharia do IFSC/FLN e a relação com o saber. *REAMEC – Revista de Ensino de Engenharia, Ambiente e Cidade*, v. 11, n. 1, e23068, 2023. DOI: 10.26571/reamec.v11i1.16487.
- ANDERSON, Lorin W.; KRATHWOHL, David R. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.
- APIFY. Web Scraping, Data Extraction, and Automation Platform. Acesso em: 12 maio 2025. 2025. Disponível em: <<https://apify.com/>>.
- APUFSC-SINDICAL. Procura por cursos de graduação na UFSC cai 34,8%. Acesso em: 27 jul. 2025. 2025. Disponível em: <<https://www.apufsc.org.br/2025/05/26/engenharia-tem-queda-de-alunos-nos-ultimos-oito-anos-e-faz-mec-planejar-reformulacao/>>. Acesso em: 27 jul. 2025.
- APUFSC-SINDICAL, O Globo e. Engenharia tem queda de alunos nos últimos oito anos e faz MEC planejar reformulação. 2025. Reportagem online. Acesso em: 27 jul. 2025.
- BARDIN, Laurence. *Análise de conteúdo*. São Paulo: Edições 70, 2016.
- BASAK, Tapati; RAZZAQUE, Muhammad Abdur. Text mining in unstructured text: techniques, methods and analysis. *World Scientific News*, v. 174, 2022. Acesso em: 5 jul. 2025. Disponível em: <<https://www.researchgate.net/publication/375583862>>.
- BERTONI, Ricardo. Desenvolvimento e uso de web crawlers para extração de dados na internet. 2017. <https://www.acme.edu.br>. Acesso em: 11 mar. 2025.
- BLOOM, Benjamin S. et al. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay Company, 1956.
- BORGES, Mario Neto; ALMEIDA, Nival Nunes de. Perspectivas parReflexões sobre o ensino da Engenharia Nacional – Desafios e Oportunidadesno Brasil: formação, identidade e mercado de trabalho. 2013. <https://ptdocz.com/doc/330551/baixar-este-arquivo-pdf>. Acesso em: 24 mar. 2025.

BRASIL. Decreto n. 7.566, de 23 de setembro de 1909. 1909. Diário Oficial da União. Disponível em: <<https://www2.camara.leg.br/legin/fed/decret/1900-1909/decreto-7566-23-setembro-1909-525411-publicacaooriginal-1-pe.html>>. Acesso em: 11 mar. 2025.

_____. Diretrizes Curriculares Nacionais dos Cursos de Graduação: Resolução CNE/CES nº 54, de 28 de dezembro de 2004. 2004. Ministério da Educação, Brasília. Disponível em: <https://portal.mec.gov.br/cne/arquivos/pdf/pces0054_04.pdf>. Acesso em: 11 mar. 2025.

_____. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). 2018. Diário Oficial da União: seção 1, Brasília, DF, ano 156, n. 157, p. 1–9, 15 ago. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 18 mai. 2025.

_____. Lei nº 9.394, de 20 de dezembro de 1996. 1996. Diário Oficial da União: seção 1, Brasília, DF. Estabelece as diretrizes e bases da educação nacional. Publicado em: 23 dez. 1996. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l9394.htm>. Acesso em: 25 mai. 2025.

BRENNING, Alexander; HENN, Sebastian. Web scraping: a promising tool for geographic data acquisition. *arXiv preprint arXiv:2305.19893*, 2023. Disponível em: <<https://arxiv.org/abs/2305.19893>>.

BUTOVA, Yelena. The history of development of competency-based education. *European Scientific Journal*, v. 11, n. 10, p. 250–255, 2015. Disponível em: <<https://eujournal.org/index.php/esj/article/view/5728>>. Acesso em: 25 mai. 2025.

CHIARELLO, Francesco et al. Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change*, v. 173, p. 121177, 2021. DOI: 10.1016/j.techfore.2021.121177. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0040162521006107>>.

CHOMSKY, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.

CLOUDFLARE. Cloudflare Security Architecture. 2025. Acesso em: 5 jul. 2025. Disponível em: <<https://developers.cloudflare.com/reference-architecture/architectures/security/>>.

DUTRA, Joel Souza. *Gestão por competências: um modelo avançado para o gerenciamento de pessoas*. São Paulo: Gente, 2004.

AL-FANNAH, Nasser Mohammed. Making Defeating CAPTCHAs Harder for Bots. *arXiv preprint arXiv:1704.02803*, 2017. Disponível em: <<https://arxiv.org/abs/1704.02803>>.

FAPESP. Cursos de engenharia: queda nos números de ingressantes e de concluintes. *Revista Pesquisa FAPESP*, n. 299, p. 11, 2021. Acesso em: 21 mar. 2025. Disponível em: <<https://revistapesquisa.fapesp.br/cursos-de-engenharia-queda-nos-numeros-de-ingressantes-e-de-concluintes/>>.

FELDMAN, Ronen; SANGER, James. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press, 2007. ISBN 9780521836579. Disponível em: <<https://www.cambridge.org/core/books/text-mining-handbook/0634B1DF14259CB43FCCF28972AE4382>>.

FERRARA, Emilio et al. Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, Elsevier, v. 70, p. 301–323, 2014. DOI: 10.1016/j.knosys.2014.07.007. Disponível em: <<https://doi.org/10.1016/j.knosys.2014.07.007>>.

FERRAZ, Ana Paula Cabral Seixas Costa; BELHOT, Renato Vidal. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. *Gestão & Produção*, Universidade Federal de São Carlos, v. 17, n. 2, p. 421–431, 2010. Disponível em: <<https://www.scielo.br/j/gp/a/bRkFgcJqbGCDp3HjQqFdqBm/?lang=pt>>.

FERREIRA, M. H. W.; CORREA, R. F. Mineração de textos científicos: análise de artigos brasileiros. *Em Questão*, UFPR, v. 27, n. 1, p. 237–262, 2020. DOI: 10.19132/1808-5245271.237-262.

FERREIRA-MELLO, Rubens; BASTOS, Diego; BARROS, Marcelo. Mining Educational Texts: A Systematic Mapping of Text Mining in Educational Contexts. *Journal of Educational Data Science*, v. 9, n. 1, p. 1–20, 2024. DOI: 10.1000/eds2024.123456.

FIROOZEH, Nazanin et al. Keyword extraction: Issues and methods. *Natural Language Engineering*, v. 26, n. 3, 2019. Disponível em: <<https://doi.org/10.1017/S1351324919000457>>.

FLEURY, Maria Tereza Leme; FLEURY, Afonso Carlos Corrêa. Construindo o conceito de competência. *Revista de Administração Contemporânea*, v. 5, Esp., p. 183–196, 2001. Acesso em: 24 mar. 2025. Disponível em: <<https://www.scielo.br/j/rac/a/X6JmndKNK8VtZz7GzpFQshH/?lang=pt>>.

FREITAS, Luiz Carlos de. *A pedagogia das competências: autonomia ou adaptação?* São Paulo: Cortez, 2003.

GAN EBENEZER. IndeedJobScraper. Acesso em: 12 maio 2025. 2023. Disponível em: <<https://github.com/Eben001/IndeedJobScraper>>.

GROOT, Maurits de; SCHUTTE, Jelle; GRAUS, David. Job Posting-Enriched Knowledge Graph for Skills-based Matching. *arXiv preprint arXiv:2109.02584*, 2021.

IBM. What Is Text Mining? 2025. Acesso em: 5 jul. 2025. Disponível em: <<https://www.ibm.com/topics/text-mining>>.

IFSC. Projeto Pedagógico do Curso de Engenharia de Telecomunicações. Dez. 2022. Versão 2.2, aprovado em 6 de dezembro de 2022. Disponível em: <<https://www.ifsc.edu.br>>.

INDEED. Indeed Docs. Acesso em: 09 maio 2025. 2025. Disponível em: <<https://docs.indeed.com/getstarted/>>.

JOGARAO, Malla; HEMALATHA, T.; NAIDU, S. T. Leveraging HR Analytics for Data-Driven Decision Making: A Comprehensive Review. *International Journal of Scholarly Research and Reviews*, v. 11, n. 10, p. 1774–1790, 2022. Disponível em: <https://www.researchgate.net/publication/372216539_Leveraging_HR_Analytics_for_Data-Driven_Decision-Making_A_Comprehensive_Review>. Acesso em: 25 mai. 2025.

JURAFSKY, Daniel; MARTIN, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3. ed.: Prentice Hall, 2021. Acesso em: 21 mar. 2025. Disponível em: <<https://web.stanford.edu/~jurafsky/slp3/>>.

_____. _____. 3. ed. New York: Cambridge University Press, 2025. Disponível em: <<https://web.stanford.edu/~jurafsky/slp3/>>.

KELLEHER, John D.; TIERNEY, Brendan. *Data Science*. Cambridge, MA: MIT Press, 2018. ISBN 9780262535434.

KHALIL, Salim; FAKIR, Mohamed. RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, Elsevier, v. 6, p. 98–106, 2017. DOI: 10.1016/j.softx.2017.04.004. Disponível em: <<https://doi.org/10.1016/j.softx.2017.04.004>>.

KHAOUJA, Imane; KASSOU, Ismail; GHOGHO, Mounir. A Survey on Skill Identification from Online Job Ads. *IEEE Access*, v. 9, p. 118134–118153, 2021. DOI: 10.1109/ACCESS.2021.3106120. Disponível em: <<https://doi.org/10.1109/ACCESS.2021.3106120>>.

KRATHWOHL, David R.; BLOOM, Benjamin S.; MASIA, Bertram B. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Company, 1964.

LAL, S.; AL., et. Evaluation of the trends in jobs and skill-sets using data analytics. *Journal of Big Data*, v. 9, n. 1, p. 1–17, 2022. DOI: 10.1186/s40537-022-00576-5.

LIGHTCAST. Academic Research Using Lightcast Data. 2025. <https://lightcast.io/resources/new-academic-research>. Acesso em: 25 mai. 2025.

MACHADO, A.; ARANHA, C.; PASSOS, E. Mineração de textos: conceito, processo e aplicações. *Repositório da UDESC*, 2010. Definição de mineração de textos como processo de extração de padrões por frequência e temática.

MOHAN, Kavya; RENGARAJAN, A. Reverse Proxy Technology: A Comprehensive Review. *International Journal of Innovative Research in Computer and Communication Engineering*, 2024. Acesso em: 5 jul. 2025. Disponível em: <https://www.researchgate.net/publication/379313742_Reverse_Proxy_Technology>.

MUNZERT, Simon et al. *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. Chichester: John Wiley & Sons, 2015.

NASIRIANI, Neda; ZOLA, Enrico; MALAIYA, Yashwant K. Moving-target Defense against Botnet Reconnaissance and an Adversarial Coupon-Collection Model. *arXiv preprint arXiv:1712.01102*, 2017. Acesso em: 5 jul. 2025. Disponível em: <<https://arxiv.org/abs/1712.01102>>.

OECD. The Future of Education and Skills: Education 2030. 2020. Acesso em: 21 mar. 2025. Disponível em: <<https://www.oecd.org/education/2030-project/>>.

PACHECO, Eliezer. *Os Institutos Federais: uma revolução na educação profissional e tecnológica*. Natal: Editora do IFRN, 2010. Acesso em: 24 mar. 2025. Disponível em: <<https://ptdocz.com/doc/330551/baixar-este-arquivo-pdf>>.

PAPAGIANNOPPOULOU, Evangelia; TSOUMAKAS, Grigorios. A review of keyphrase extraction. 2019. Disponível em: <<https://arxiv.org/abs/1905.05044>>.

PARIZOTO, Gabriel Henrique Silveira. Considerações sobre a formação do engenheiro nas escolas brasileiras. 2018. Trabalho de Conclusão de Curso (Graduação em Engenharia Eletrônica) – Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco. Acesso em: 19 jul. 2025. Disponível em: <<https://riut.utfpr.edu.br/jspui/bitstream/1/24875/1/formacaoengenheiroescolasbrasileiras.pdf>>.

PEJIĆ BACH, Mirjana et al. Text mining of Industry 4.0 job advertisements. *International Journal of Information Management*, v. 50, p. 416–431, 2020. DOI: 10.1016/j.ijinfomgt.2019.07.014. Disponível em: <<https://doi.org/10.1016/j.ijinfomgt.2019.07.014>>.

PERRENOUD, Philippe. *Construir as Competências desde a Escola*. Artmed, 2000.

_____. *Dez novas competências para ensinar*. Porto Alegre: Artmed, 1999.

PESTANA, Márcio. Os princípios no tratamento de dados na LGPD (Lei Geral de Proteção de Dados Pessoais). 2023. Consultor Jurídico. Acesso em: 18 maio 2025. Disponível em: <<https://www.conjur.com.br/wp-content/uploads/2023/09/artigo-marcio-pestana-lgpd.pdf>>.

RAMOS, Marise. *A pedagogia das competências: autonomia ou adaptação?* Cortez Editora, 2001.

RAMOS, Marise; PARANHOS, Michelle. A noção de competência enquanto princípio de organização curricular. *Revista Brasileira de Educação*, v. 23, n. 74, p. 1–23, 2018. DOI: 10.1590/S1413-24782018230074. Disponível em: <<https://www.scielo.br/j/rbedu/a/DzH3VMkvXTVv6FCR7TLjPzG/>>.

RAMOS, Marise Nogueira. A pedagogia das competências: autonomia ou adaptação? In: FRIGOTTO, Gaudêncio; CIAVATTA, Maria; RAMOS, Marise (Ed.). *Ensino médio integrado: concepção e contradições*. São Paulo: Cortez, 2001. P. 11–35.

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. Upper Saddle River, NJ: Prentice Hall, 2010.

- SAVIANI, Dermeval. *Escola e Democracia*. Cortez Editora, 2008.
- SHARMA, Anju; KUMAR, Yogesh; GROVER, Pardeep. A review of heuristic optimization algorithms. *International Journal of Computer Applications*, v. 975, n. 8887, 2020.
- SILVA, João; SOUZA, Maria. *Mineração de Textos – Frequência de Termo (Term Frequency)*. 2007. Também denominada “term frequency”, mede quantidade de vezes que um termo aparece em um documento.
- SILVA, Rafael de Oliveira; SOUZA, Fernanda. Mineração de Texto: Uma Revisão Sistemática da Literatura. *Revista de Sistemas e Computação*, v. 10, n. 1, p. 50–60, 2020. ISSN 2238-1031. Disponível em: <<https://revistas.unifacs.br/index.php/rsc/article/view/7015>>.
- SILVA FILHO, Ruy Leme da et al. Evasão no Ensino Superior. *Cadernos de Pesquisa*, v. 37, n. 132, p. 641–659, 2007. DOI: 10.1590/S0100-15742007000300007.
- SONAGLIO, Schaiana. Módulo Web Scraper. 2025. https://github.com/schaisonaglio/web_scraper. Acesso em: 12 jul. 2025.
- _____. Planilha de objetivos das UCs. 2025. https://github.com/schaisonaglio/indeed-scraper-tcc/blob/main/Objetivos_UCs.xlsx. Acesso em: 12 jul. 2025.
- _____. Repositório Indeed Scraper. 2025. <https://github.com/schaisonaglio/indeed-scraper-tcc>. Acesso em: 12 jul. 2025.
- _____. Resultados após normalização (CHA_apos_normalizacao.xlsx). 2025. https://github.com/schaisonaglio/indeed-scraper-tcc/blob/main/CHA_apos_normalizacao.xlsx. Acesso em: 12 jul. 2025.
- _____. Resultados CHA (planilha CHA.xlsx). 2025. <https://github.com/schaisonaglio/indeed-scraper-tcc/blob/main/CHA.xlsx>. Acesso em: 12 jul. 2025.
- _____. Script de análise CHA (metodologia_cha.py). 2025. https://github.com/schaisonaglio/indeed-scraper-tcc/blob/main/metodologia_cha.py. Acesso em: 12 jul. 2025.
- SOUZA, Vanderli Corrêa de; BORGES, Mario Neto; ALMEIDA, Nival Nunes de. Reflexões sobre o ensino da engenharia no Brasil: formação, identidade e mercado de trabalho. Acesso em: 24 mar. 2025. Disponível em: <<https://ptdocz.com/doc/330551/baixar-este-arquivo-pdf>>.
- STRUCK, Joshua. Introduction to Web Scraping with R. 2025. Social Science Computing Cooperative, University of Wisconsin-Madison. Acesso em: 18 maio 2025. Disponível em: <<https://sscc.wisc.edu/sscc/pubs/webscraping-r/>>.
- TARDIF, Maurice. *Saberes docentes e formação profissional*. Petrópolis: Vozes, 2002.
- THEIRSTACK. TheirStack Job Postings API. 2025. <https://publicapis.dev/resource/theirstack-s-job-postings/42dca0bb-41d7-42ed-a521-0e4bca8f688a>. Acesso em: 19 jul. 2025.

VENUGOPALAN, Hari et al. FP-Inconsistent: Detecting Evasive Bots using Browser Fingerprint Inconsistencies. *arXiv preprint arXiv:2406.07647*, 2024. Preprint. Disponível em: <<https://arxiv.org/abs/2406.07647>>.

WEBBER, Karen L. *Data Analytics and the Imperatives for Data-Informed Decision Making in Higher Education*. 2019. Disponível em: <https://ihe.uga.edu/sites/default/files/inline-files/Webber_2019004_paper_2.pdf>. Acesso em: 25 mai. 2025.

WINCH, Christopher. From labour market to educational system: the case of ESCO. *BWP – Vocational Training in Research and Practice*, 2025. Disponível em: https://www.bwp-zeitschrift.de/en/bwp_205647.php. Acesso em: 25 mai. 2025.

APÊNDICE A – MENSAGEM DE ERRO - ROBOTS.TXT - INDEED

Código A.1 – Mensagem de erro do robots.txt

```
1 Access to search URL br.indeed.com/jobs?q=Scrum&l=Santa%20Catarina&from=
  searchOnDesktopSerp is disallowed by robots.txt.
```

APÊNDICE B – MENSAGEM DE ERRO - CAPTCHA - INDEED

Código B.1 – Erro HTTP retornado pelo Cloudflare

```
1 HTTPError: 403 Client Error: Forbidden for url: br.indeed.com/jobs?q=Scrum&l=Santa  
  %20Catarina&from=searchOnDesktopSerp  
2 Detail: Cloudflare verification failed. Missing or invalid cf_clearance cookie
```

APÊNDICE C – SCRIPT PARA COLETA DE DADOS VIA API DA APIFY

Código C.1 – Código-fonte do script `apify_indeed_scraper.py`

```
1 import time
2 import json
3 import re
4 from apify_client import ApifyClient
5
6 API_TOKEN = "apify_api_XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
7
8 # Inicializar Apify client
9 client = ApifyClient(API_TOKEN)
10
11 # Limpar nomes
12 def sanitize_filename(text):
13     return re.sub(r'[\\/*?:"<>|;\|]', "_", text)
14
15 # Termos procurados
16 search_terms = [
17     "Design Thinking", # Design centrado no usuário
18     "Scrum", # Metodologia ágil
19     "Kanban", # Gestão visual de fluxo de trabalho
20     "Agile", # Desenvolvimento ágil
21     "Gestão de produtos",
22     "Gestão de projetos",
23     "Programação C/C++",
24     "Linguagem C",
25     "Desenvolvimento C",
26     "Depuração de código",
27     "Álgebra Booleana",
28     "Aritmética binária",
29     "Circuitos digitais",
30     "Algoritmos",
31     "Estrutura de dados",
32     "Circuitos elétricos",
33     "Corrente contínua",
34     "Leis de Kirchhoff",
35     "Análise de circuitos",
36     "Teoremas de Superposição",
37     "Thévenin e Norton",
38     "Ferramentas SPICE",
39     "Simulações Spice",
40     "LT-Spice",
41     "Tina-TI",
```



```
42 "Pspice",
43 "Circuitos analógicos",
44 "HDL", # Hardware Description Language
45 "VHDL", # Linguagem de descrição de hardware
46 "FPGA", # Circuitos programáveis
47 "CMOS", # Tecnologia de semicondutores
48 "TTL", # Transistor-Transistor Logic
49 "Simulação de hardware digital",
50 "Simulação e validação de hardware",
51 "Programação orientada a objetos",
52 "POO",
53 "Java",
54 "UML", # Unified Modeling Language
55 "RUP", # Rational Unified Process
56 "Git", # Controle de versão
57 "JUnit", # Testes unitários em Java
58 "Testes unitários",
59 "Unit tests",
60 "TDD", # Test-Driven Development
61 "Assembly", # Linguagem de baixo nível
62 "Matlab",
63 "Fourier", # Transformadas de Fourier
64 "Transformada de Laplace",
65 "Transformada Z",
66 "Corrente alternada",
67 "Fator de potência",
68 "Sistemas trifásicos",
69 "Potência ativa, reativa e aparente",
70 "Transformadores",
71 "Circuitos magnéticos e de indutância",
72 "Campos eletromagnéticos",
73 "Montagem de circuitos elétricos",
74 "Multímetro",
75 "Osciloscópio",
76 "Geradores de sinal",
77 "Processamento de sinais",
78 "Processamento digital de sinais",
79 "Transformada Discreta de Fourier",
80 "Filtros digitais",
81 "IIR", # Infinite Impulse Response
82 "FIR", # Finite Impulse Response
83 "Simulink",
84 "Matlab/Simulink",
85 "Redes de computadores",
86 "Redes IP",
87 "TCP/IP",
88 "UDP",
```

```
89     "TCP",
90     "WAN",
91     "MAN",
92     "LAN",
93     "HTTP",
94     "FTP",
95     "SMTP",
96     "Multithread",
97     "Multi thread",
98     "Multi-thread",
99     "Linux",
100    "Memory management",
101    "Gerenciamento de memória",
102    "Heap",
103    "Stack",
104    "File system",
105    "Amplificadores operacionais",
106    "Transistores",
107    "MOSFET",
108    "Eletrônica de potência",
109    "Sistemas de potência",
110    "Cabo coaxial",
111    "Fibra óptica",
112    "Atenuação e dispersão em meios guiados",
113    "Cabos trançados",
114    "Par trançado",
115    "Processos estocásticos",
116    "Métodos estocásticos",
117    "Estocásticos",
118    "Probabilidade e estatística",
119    "Variáveis aleatórias",
120    "Sistemas distribuídos",
121    "REST", # API REST
122    "SOAP", # Web Services
123    "DNS", # Sistema de nomes de domínio
124    "API",
125    "Fault tolerance",
126    "Tolerância a falhas",
127    "Segurança em desenvolvimento de software",
128    "Segurança da informação",
129    "Antenas",
130    "Radiofrequência",
131    "Antenas dipolo",
132    "Antenas yagi",
133    "Antenas parabólica",
134    "Amplificadores de potência",
135    "Amplificadores de pequeno sinal",
```

```
136     "Simulação de circuitos AC",
137     "Microprocessadores",
138     "Microcontroladores",
139     "GPIO", # General Purpose Input/Output
140     "ADC", # Conversor analógico-digital
141     "UART",
142     "Firmware",
143     "JTAG", # Interface de depuração
144     "Ethernet",
145     "Wi-Fi",
146     "xDSL",
147     "FTTH",
148     "FTTX",
149     "FTTN",
150     "FTTC",
151     "IoT", # Internet das Coisas
152     "Internet das Coisas",
153     "BGP", # Border Gateway Protocol
154     "Modulação",
155     "AM", # Modulação em amplitude
156     "FM", # Modulação em frequência
157     "Modulação em banda base",
158     "Codificação de linha",
159     "Digitalização de sinal analógico",
160     "Largura de banda",
161     "ASK", # Amplitude Shift Keying
162     "FSK", # Frequency Shift Keying
163     "PSK", # Phase Shift Keying
164     "QAM", # Quadrature Amplitude Modulation
165     "Sincronismo",
166     "Analisadores de espectro",
167     "SNR", # Signal-to-Noise Ratio
168     "BER", # Bit Error Rate
169     "Bit error rate",
170     "Rádio Definido por Software",
171     "Amplificadores de RF",
172     "LNA de baixo ruído",
173     "Filtros sintonizados de RF",
174     "PLLs", # Phase-Locked Loops
175     "Figura de ruído",
176     "Casamento de impedância",
177     "Redes móveis",
178     "Multiple Input Multiple Output",
179     "FDMA",
180     "TDMA",
181     "CDMA",
182     "OFDMA",
```

```

183     "4G",
184     "5G",
185     "6G",
186     "LPWAN",
187     "LTE",
188     "Sistemas Embarcados",
189     "RTOS", # Real-Time Operating System
190     "Sistemas operacionais de tempo real",
191     "CAN", # Barramento de comunicação
192     "Integração de software e hardware",
193     "Desenvolvimento de Hardware/Software Embarcado"
194 ]
195
196
197 # 3. Rodar ator Apify para cada termo
198 for i, term in enumerate(search_terms, 1):
199     print(f"\n({i}/{len(search_terms)}) Buscando: {term[:60]}...")
200
201     run_input = {
202         "country": "BR",
203         "followApplyRedirects": True,
204         "location": "Santa Catarina",
205         "maxItems": 50,
206         "parseCompanyDetails": False,
207         "position": term,
208         "saveOnlyUniqueItems": True
209     }
210
211     try:
212
213         run = client.actor("misceres-indeed-scraper").call(run_input=run_input)
214
215         client.run(run["id"]).wait_for_finish()
216
217         dataset_items = client.dataset(run["defaultDatasetId"]).list_items().items
218
219         filename = f"results_{i:02d}_{sanitize_filename(term[:40])}.json"
220         with open(filename, "w", encoding="utf-8") as f:
221             json.dump(dataset_items, f, ensure_ascii=False, indent=2)
222
223         print(f"{len(dataset_items)} Resultados salvos em: {filename}")
224         time.sleep(5)
225
226     except Exception as e:
227         print(f"Erro ao processar: {term}\n{e}")

```

APÊNDICE D – SCRIPT PARA CONSOLIDAÇÃO DOS ARQUIVOS DE VAGAS

Código D.1 – Código-fonte do script consolidate_results.py

```
1 import os
2 import json
3 import pandas as pd
4
5
6 folder_path = "./"
7
8 # Lista para armazenar os dados
9 all_items = []
10
11 # Percorrer todos os arquivos JSON
12 for filename in os.listdir(folder_path):
13     if filename.endswith(".json") and filename.startswith("results_"):
14         with open(os.path.join(folder_path, filename), "r", encoding="utf-8") as
15             file:
16                 data = json.load(file)
17                 search_term = filename.split("_", 2)[2].replace(".json", "").replace("_", " ")
18                 for item in data:
19                     all_items.append({
20                         "Termo de busca": search_term,
21                         "Título da vaga": item.get("title"),
22                         "Empresa": item.get("company"),
23                         "Localização": item.get("location"),
24                         "Descrição": item.get("description"),
25                         "Link": item.get("url")
26                     })
27
28 df = pd.DataFrame(all_items)
29
30
31 output_file = "vagas_unificadas.csv"
32 df.to_csv(output_file, index=False, encoding="utf-8-sig")
33
34 print(f"Arquivo CSV gerado: {output_file} ({len(df)} linhas)")
```

APÊNDICE E – SCRIPT PARA ANÁLISE DE FREQUÊNCIA DE TERMOS

Código E.1 – Código-fonte do script `term_frequency_analysis.py`

```

1 import os
2 import json
3 import re
4 from collections import defaultdict
5
6 # Termos de busca
7 search_terms = [
8     "Radiodifusão", # Transmissão de rádio e TV
9     "Rádio enlace", # Comunicação ponto a ponto e multiponto
10    "Ponto a ponto",
11    "Ponto multiponto",
12    "Multiponto",
13    "DWDM", # Dense Wavelength Division Multiplexing
14    "CWDM", # Coarse Wavelength Division Multiplexing
15    "GPON", # Gigabit-capable Passive Optical Network
16    "Regulamentação ANATEL", # Normas brasileiras de telecomunicações
17    "Antenas de alta frequência",
18    "Projeto de antenas",
19    "Micro-ondas",
20    "Miniaturização de antenas",
21    "Antenas microstrip",
22    "Smart antenas",
23    "Antenas fractais",
24    "Medições de antenas",
25    "Sistemas de comunicações RF",
26    "Comunicações móveis",
27    "Engenharia de radiofrequência",
28    "RF design",
29    "Microwave engineering",
30    "High-frequency measurements",
31    "Sistemas de radar",
32    "Radar CW",
33    "Radar pulsado",
34    "Phased array SAR", # Synthetic Aperture Radar com phased array
35    "Sinais de RF CW",
36    "Sinais de RF pulsado",
37    "Radar secundário",
38    "Radar naval",
39    "Radar aéreo",
40    "Radar satelital",
41    "Radar ambiental",

```

```
42 "Radar militar",
43 "Radar geográfico",
44 "Radar para policiamento",
45 "Análise de sinais de radar",
46 "Especificação de sistemas de radar",
47 "Comunicações via satélite",
48 "Sistemas satelitais",
49 "Segmento espacial",
50 "Segmento terrestre",
51 "Link budget",
52 "Enlace satelital",
53 "Cubesats",
54 "Nanosatélites",
55 "Satélites de relay",
56 "Payload de satélites",
57 "Bus de satélites",
58 "Aplicações satelitais",
59 "Engenharia de micro-ondas",
60 "Parâmetros de espalhamento",
61 "Carta de Smith",
62 "Transformação de impedância",
63 "Circuitos em micro-ondas",
64 "Ressonadores",
65 "Acopladores direcionais",
66 "Filtros de micro-ondas",
67 "Estruturas periódicas",
68 "Amplificadores de micro-ondas",
69 "Osciladores de micro-ondas",
70 "Simulação de RF",
71 "Ferramentas de simulação (HFSS, CST, ADS)",
72 "Análise de circuitos de alta frequência",
73 "Desempenho de redes",
74 "Teoria de filas",
75 "Cadeias de Markov",
76 "Simulação a eventos discretos",
77 "Simulação por eventos discretos",
78 "Geração de números randômicos",
79 "Medição ativa e passiva",
80 "Simuladores de redes",
81 "Tuning de parâmetros de rede",
82 "Modelagem analítica",
83 "NS-3", # Network Simulator 3
84 "OMNeT++", # Simulador de redes
85 "Planejamento de capacidade de redes",
86 "Análise preditiva de redes",
87 "Internet das Coisas",
88 "IoT",
```

```
89  "Arquitetura de dispositivos",
90  "Conectividade IoT",
91  "MQTT",
92  "CoAP",
93  "Hardware embarcado",
94  "Sensores e atuadores",
95  "Redes de sensores sem fio",
96  "AWS IoT",
97  "Azure IoT",
98  "Desenvolvimento de aplicações IoT",
99  "Teoria dos grafos",
100 "Percurso euleriano",
101 "Percurso hamiltoniano",
102 "Árvores geradoras mínimas",
103 "Árvore de extensão mínima",
104 "AGM", # Algoritmo de Kruskal/Prim para árvores mínimas
105 "Coloração de grafos",
106 "Redes de fluxo",
107 "Fluxo máximo",
108 "Fluxo com custo",
109 "Banco de dados relacional",
110 "RDBMS", # Relational Database Management System
111 "Modelagem de dados",
112 "Álgebra relacional",
113 "SQL",
114 "Normalização",
115 "SGBD", # Sistema de Gerenciamento de Banco de Dados
116 "PostgreSQL",
117 "MySQL",
118 "SQLite",
119 "Desenvolvimento com banco de dados",
120 "Aplicações desktop e web com banco de dados",
121 "Bancos de dados NoSQL",
122 "MongoDB",
123 "Administração de banco de dados",
124 "Design de banco de dados",
125 "Otimização de queries",
126 "Engenharia de software",
127 "Análise orientada a objetos",
128 "Projeto orientado a objetos",
129 "UML",
130 "Modelagem de software",
131 "Levantamento de requisitos",
132 "Casos de uso",
133 "Arquitetura de software",
134 "Padrões de projeto",
135 "Frameworks",
```



```
136 "Desenvolvimento ágil",
137 "Metodologias ágeis",
138 "Scrum",
139 "Kanban",
140 "Testes de software",
141 "Gerenciamento de projetos de software",
142 "Inteligência artificial",
143 "Aprendizado de máquina",
144 "Machine learning",
145 "Deep learning",
146 "Preparação de dados",
147 "Análise de dados",
148 "Modelos preditivos",
149 "TensorFlow",
150 "PyTorch",
151 "Scikit-learn",
152 "Avaliação de modelos",
153 "Agentes inteligentes",
154 "Classificação",
155 "Regressão",
156 "Redes neurais artificiais",
157 "Algoritmos supervisionados e não supervisionados",
158 "Projeto de protocolos",
159 "Especificação de protocolos",
160 "Comunicação de rede",
161 "Máquinas de estados",
162 "FSM", # Finite State Machine
163 "Modelagem de protocolos",
164 "Codificação de mensagens",
165 "Programação assíncrona",
166 "Model checking",
167 "Verificação formal",
168 "Protocolos de comunicação",
169 "Implementação de protocolos",
170 "Propriedades de protocolos",
171 "Redes de computadores",
172 "Protocolos personalizados",
173 "Técnicas de modelagem de protocolos",
174 "Sistemas de controle",
175 "Controle clássico",
176 "Controle digital",
177 "Análise de sistemas",
178 "Resposta transitória",
179 "Resposta em regime permanente",
180 "Lugar das raízes",
181 "Resposta em frequência",
182 "Controladores PID",
```

```
183 "Espaço de estados",
184 "Projeto de controle",
185 "Simulação de sistemas de controle",
186 "Matlab",
187 "Simulink",
188 "Compatibilidade eletromagnética",
189 "CEM", # Compatibilidade Eletromagnética
190 "Interferência eletromagnética",
191 "EMI",
192 "Emissões irradiadas",
193 "Emissões conduzidas",
194 "Suscetibilidade eletromagnética",
195 "Blindagem eletromagnética",
196 "Diafonia",
197 "Descargas eletrostáticas",
198 "Normas EMC",
199 "Medições eletromagnéticas",
200 "Espectro eletromagnético",
201 "Princípios eletromagnéticos",
202 "Eletrônica de potência",
203 "Conversores estáticos",
204 "Retificadores monofásicos e trifásicos",
205 "Variadores de tensão",
206 "CA-CC",
207 "CA-CA",
208 "CC-CC",
209 "CC-CA",
210 "MOSFET",
211 "IGBT",
212 "Tiristores",
213 "Fontes chaveadas",
214 "Filtros ativos",
215 "UPS", # Uninterruptible Power Supply
216 "Correção de fator de potência",
217 "Acionamento de motores",
218 "Inversores de frequência",
219 "Simulação de conversores",
220 "Reatores eletrônicos",
221 "Teoria da informação",
222 "Codificação de fonte",
223 "Codificação de canal",
224 "Códigos corretores de erro",
225 "Shannon",
226 "Capacidade do canal",
227 "Entropia",
228 "Compressão de dados",
229 "Detecção e correção de erros",
```

```

230     "Códigos de Hamming",
231     "Códigos convolucionais",
232     "Códigos LDPC" # Low-Density Parity-Check
233 ]
234
235
236 # Diretório contendo os arquivos JSON gerados
237 data_dir = "."
238
239 # Dicionário para armazenar as contagens
240 counts = {term: {"matches": 0, "jobs": set()} for term in search_terms}
241
242 # Percorra todos os arquivos JSON
243 for filename in os.listdir(data_dir):
244     if filename.endswith(".json") and filename.startswith("results_"):
245         with open(os.path.join(data_dir, filename), encoding="utf-8") as f:
246             try:
247                 jobs = json.load(f)
248             except json.JSONDecodeError:
249                 print(f"Aviso: falha ao ler {filename}. Pulando.")
250                 continue
251
252     # Para cada vaga, combine todos os campos relevantes em um único texto
253     for job in jobs:
254         text = " ".join([
255             str(job.get("title", "")),
256             str(job.get("description", "")),
257             str(job.get("skills", "")),
258             str(job.get("requirements", "")),
259         ]).lower()
260
261     # Verifique presença de cada termo
262     for term in search_terms:
263         if re.search(r'\b' + re.escape(term.lower()) + r'\b', text):
264             counts[term]["matches"] += 1
265             counts[term]["jobs"].add(job.get("id") or job.get("link") or
filename)
266
267 # Exporte resultados consolidados
268 report = []
269 for term, info in counts.items():
270     report.append({
271         "term": term,
272         "match_count": info["matches"],
273         "unique_jobs": len(info["jobs"])
274     })
275

```

```
276 # Salve o relatório final
277 with open("term_frequency_report.json", "w", encoding="utf-8") as f:
278     json.dump(report, f, ensure_ascii=False, indent=2)
279
280 print("Análise concluída. Verifique 'term_frequency_report.json'.")
```

APÊNDICE F – RESULTADO CONSOLIDADO - ANÁLISE DE FREQUÊNCIA DE PALAVRAS-CHAVE

Código F.1 – Código-fonte do script term_frequency_report_ascii.json

```
1 [
2   {
3     "term": "Radiodifusao",
4     "match_count": 0,
5     "unique_jobs": 0
6   },
7   {
8     "term": "Radio enlace",
9     "match_count": 0,
10    "unique_jobs": 0
11  },
12  {
13    "term": "Ponto a ponto",
14    "match_count": 0,
15    "unique_jobs": 0
16  },
17  {
18    "term": "Ponto multiponto",
19    "match_count": 0,
20    "unique_jobs": 0
21  },
22  {
23    "term": "Multiponto",
24    "match_count": 0,
25    "unique_jobs": 0
26  },
27  {
28    "term": "DWDM",
29    "match_count": 3,
30    "unique_jobs": 1
31  },
32  {
33    "term": "CWDM",
34    "match_count": 0,
35    "unique_jobs": 0
36  },
37  {
38    "term": "GPON",
39    "match_count": 18,
40    "unique_jobs": 6
```

```
41 },
42 {
43     "term": "Regulamentacao ANATEL",
44     "match_count": 0,
45     "unique_jobs": 0
46 },
47 {
48     "term": "Antenas de alta frequencia",
49     "match_count": 0,
50     "unique_jobs": 0
51 },
52 {
53     "term": "Projeto de antenas",
54     "match_count": 0,
55     "unique_jobs": 0
56 },
57 {
58     "term": "Micro-ondas",
59     "match_count": 6,
60     "unique_jobs": 5
61 },
62 {
63     "term": "Miniaturizacao de antenas",
64     "match_count": 0,
65     "unique_jobs": 0
66 },
67 {
68     "term": "Antenas microstrip",
69     "match_count": 0,
70     "unique_jobs": 0
71 },
72 {
73     "term": "Smart antennas",
74     "match_count": 0,
75     "unique_jobs": 0
76 },
77 {
78     "term": "Antenas fractais",
79     "match_count": 0,
80     "unique_jobs": 0
81 },
82 {
83     "term": "Medicoes de antenas",
84     "match_count": 0,
85     "unique_jobs": 0
86 },
87 {
```

```

88     "term": "Sistemas de comunicacoes RF",
89     "match_count": 0,
90     "unique_jobs": 0
91 },
92 {
93     "term": "Comunicacoes moveis",
94     "match_count": 0,
95     "unique_jobs": 0
96 },
97 {
98     "term": "Engenharia de radiofrequencia",
99     "match_count": 0,
100    "unique_jobs": 0
101 },
102 {
103     "term": "RF design",
104     "match_count": 0,
105     "unique_jobs": 0
106 },
107 {
108     "term": "Microwave engineering",
109     "match_count": 0,
110     "unique_jobs": 0
111 },
112 {
113     "term": "High-frequency measurements",
114     "match_count": 0,
115     "unique_jobs": 0
116 },
117 {
118     "term": "Sistemas de radar",
119     "match_count": 0,
120     "unique_jobs": 0
121 },
122 {
123     "term": "Radar CW",
124     "match_count": 0,
125     "unique_jobs": 0
126 },
127 {
128     "term": "Radar pulsado",
129     "match_count": 0,
130     "unique_jobs": 0
131 },
132 {
133     "term": "Phased array SAR",
134     "match_count": 0,

```

```
135     "unique_jobs": 0
136 },
137 {
138     "term": "Sinais de RF CW",
139     "match_count": 0,
140     "unique_jobs": 0
141 },
142 {
143     "term": "Sinais de RF pulsado",
144     "match_count": 0,
145     "unique_jobs": 0
146 },
147 {
148     "term": "Radar secundario",
149     "match_count": 0,
150     "unique_jobs": 0
151 },
152 {
153     "term": "Radar naval",
154     "match_count": 0,
155     "unique_jobs": 0
156 },
157 {
158     "term": "Radar aereo",
159     "match_count": 0,
160     "unique_jobs": 0
161 },
162 {
163     "term": "Radar satelital",
164     "match_count": 0,
165     "unique_jobs": 0
166 },
167 {
168     "term": "Radar ambiental",
169     "match_count": 0,
170     "unique_jobs": 0
171 },
172 {
173     "term": "Radar militar",
174     "match_count": 0,
175     "unique_jobs": 0
176 },
177 {
178     "term": "Radar geografico",
179     "match_count": 0,
180     "unique_jobs": 0
181 },
```



```
182 {
183   "term": "Radar para policiamento",
184   "match_count": 0,
185   "unique_jobs": 0
186 },
187 {
188   "term": "Analise de sinais de radar",
189   "match_count": 0,
190   "unique_jobs": 0
191 },
192 {
193   "term": "Especificacao de sistemas de radar",
194   "match_count": 0,
195   "unique_jobs": 0
196 },
197 {
198   "term": "Comunicacoes via satelite",
199   "match_count": 0,
200   "unique_jobs": 0
201 },
202 {
203   "term": "Sistemas satelitais",
204   "match_count": 0,
205   "unique_jobs": 0
206 },
207 {
208   "term": "Segmento espacial",
209   "match_count": 0,
210   "unique_jobs": 0
211 },
212 {
213   "term": "Segmento terrestre",
214   "match_count": 0,
215   "unique_jobs": 0
216 },
217 {
218   "term": "Link budget",
219   "match_count": 0,
220   "unique_jobs": 0
221 },
222 {
223   "term": "Enlace satelital",
224   "match_count": 0,
225   "unique_jobs": 0
226 },
227 {
228   "term": "Cubesats",
```

```
229     "match_count": 0,
230     "unique_jobs": 0
231 },
232 {
233     "term": "Nanosatelites",
234     "match_count": 0,
235     "unique_jobs": 0
236 },
237 {
238     "term": "Satelites de relay",
239     "match_count": 0,
240     "unique_jobs": 0
241 },
242 {
243     "term": "Payload de satelites",
244     "match_count": 0,
245     "unique_jobs": 0
246 },
247 {
248     "term": "Bus de satelites",
249     "match_count": 0,
250     "unique_jobs": 0
251 },
252 {
253     "term": "Aplicacoes satelitais",
254     "match_count": 0,
255     "unique_jobs": 0
256 },
257 {
258     "term": "Engenharia de micro-ondas",
259     "match_count": 0,
260     "unique_jobs": 0
261 },
262 {
263     "term": "Parametros de espalhamento",
264     "match_count": 0,
265     "unique_jobs": 0
266 },
267 {
268     "term": "Carta de Smith",
269     "match_count": 0,
270     "unique_jobs": 0
271 },
272 {
273     "term": "Transformacao de impedancia",
274     "match_count": 0,
275     "unique_jobs": 0
```

```
276 },
277 {
278     "term": "Circuitos em micro-ondas",
279     "match_count": 0,
280     "unique_jobs": 0
281 },
282 {
283     "term": "Ressonadores",
284     "match_count": 0,
285     "unique_jobs": 0
286 },
287 {
288     "term": "Acopladores direcionais",
289     "match_count": 0,
290     "unique_jobs": 0
291 },
292 {
293     "term": "Filtros de micro-ondas",
294     "match_count": 0,
295     "unique_jobs": 0
296 },
297 {
298     "term": "Estruturas periodicas",
299     "match_count": 0,
300     "unique_jobs": 0
301 },
302 {
303     "term": "Amplificadores de micro-ondas",
304     "match_count": 0,
305     "unique_jobs": 0
306 },
307 {
308     "term": "Osciladores de micro-ondas",
309     "match_count": 0,
310     "unique_jobs": 0
311 },
312 {
313     "term": "Simulacao de RF",
314     "match_count": 0,
315     "unique_jobs": 0
316 },
317 {
318     "term": "Ferramentas de simulacao (HFSS, CST, ADS)",
319     "match_count": 0,
320     "unique_jobs": 0
321 },
322 {
```

```
323     "term": "Analise de circuitos de alta frequencia",
324     "match_count": 0,
325     "unique_jobs": 0
326 },
327 {
328     "term": "Desempenho de redes",
329     "match_count": 0,
330     "unique_jobs": 0
331 },
332 {
333     "term": "Teoria de filas",
334     "match_count": 0,
335     "unique_jobs": 0
336 },
337 {
338     "term": "Cadeias de Markov",
339     "match_count": 0,
340     "unique_jobs": 0
341 },
342 {
343     "term": "Simulacao a eventos discretos",
344     "match_count": 0,
345     "unique_jobs": 0
346 },
347 {
348     "term": "Simulacao por eventos discretos",
349     "match_count": 0,
350     "unique_jobs": 0
351 },
352 {
353     "term": "Geracao de numeros randomicos",
354     "match_count": 0,
355     "unique_jobs": 0
356 },
357 {
358     "term": "Medicao ativa e passiva",
359     "match_count": 0,
360     "unique_jobs": 0
361 },
362 {
363     "term": "Simuladores de redes",
364     "match_count": 0,
365     "unique_jobs": 0
366 },
367 {
368     "term": "Tuning de parametros de rede",
369     "match_count": 0,
```

```
370     "unique_jobs": 0
371 },
372 {
373     "term": "Modelagem analitica",
374     "match_count": 0,
375     "unique_jobs": 0
376 },
377 {
378     "term": "NS-3",
379     "match_count": 0,
380     "unique_jobs": 0
381 },
382 {
383     "term": "OMNeT++",
384     "match_count": 0,
385     "unique_jobs": 0
386 },
387 {
388     "term": "Planejamento de capacidade de redes",
389     "match_count": 0,
390     "unique_jobs": 0
391 },
392 {
393     "term": "Analise preditiva de redes",
394     "match_count": 0,
395     "unique_jobs": 0
396 },
397 {
398     "term": "Internet das Coisas",
399     "match_count": 2,
400     "unique_jobs": 1
401 },
402 {
403     "term": "IoT",
404     "match_count": 50,
405     "unique_jobs": 16
406 },
407 {
408     "term": "Arquitetura de dispositivos",
409     "match_count": 0,
410     "unique_jobs": 0
411 },
412 {
413     "term": "Conectividade IoT",
414     "match_count": 0,
415     "unique_jobs": 0
416 },
```

```
417 {
418   "term": "MQTT",
419   "match_count": 32,
420   "unique_jobs": 6
421 },
422 {
423   "term": "CoAP",
424   "match_count": 0,
425   "unique_jobs": 0
426 },
427 {
428   "term": "Hardware embarcado",
429   "match_count": 0,
430   "unique_jobs": 0
431 },
432 {
433   "term": "Sensores e atuadores",
434   "match_count": 0,
435   "unique_jobs": 0
436 },
437 {
438   "term": "Redes de sensores sem fio",
439   "match_count": 0,
440   "unique_jobs": 0
441 },
442 {
443   "term": "AWS IoT",
444   "match_count": 0,
445   "unique_jobs": 0
446 },
447 {
448   "term": "Azure IoT",
449   "match_count": 0,
450   "unique_jobs": 0
451 },
452 {
453   "term": "Desenvolvimento de aplicacoes IoT",
454   "match_count": 0,
455   "unique_jobs": 0
456 },
457 {
458   "term": "Teoria dos grafos",
459   "match_count": 0,
460   "unique_jobs": 0
461 },
462 {
463   "term": "Percorso euleriano",
```

```
464     "match_count": 0,
465     "unique_jobs": 0
466 },
467 {
468     "term": "Percurso hamiltoniano",
469     "match_count": 0,
470     "unique_jobs": 0
471 },
472 {
473     "term": "Arvores geradoras minimas",
474     "match_count": 0,
475     "unique_jobs": 0
476 },
477 {
478     "term": "Arvore de extensao minima",
479     "match_count": 0,
480     "unique_jobs": 0
481 },
482 {
483     "term": "AGM",
484     "match_count": 0,
485     "unique_jobs": 0
486 },
487 {
488     "term": "Coloracao de grafos",
489     "match_count": 0,
490     "unique_jobs": 0
491 },
492 {
493     "term": "Redes de fluxo",
494     "match_count": 0,
495     "unique_jobs": 0
496 },
497 {
498     "term": "Fluxo maximo",
499     "match_count": 0,
500     "unique_jobs": 0
501 },
502 {
503     "term": "Fluxo com custo",
504     "match_count": 0,
505     "unique_jobs": 0
506 },
507 {
508     "term": "Banco de dados relacional",
509     "match_count": 28,
510     "unique_jobs": 13
```

```
511 },
512 {
513     "term": "RDBMS",
514     "match_count": 0,
515     "unique_jobs": 0
516 },
517 {
518     "term": "Modelagem de dados",
519     "match_count": 24,
520     "unique_jobs": 18
521 },
522 {
523     "term": "Algebra relacional",
524     "match_count": 0,
525     "unique_jobs": 0
526 },
527 {
528     "term": "SQL",
529     "match_count": 319,
530     "unique_jobs": 150
531 },
532 {
533     "term": "Normalizacao",
534     "match_count": 0,
535     "unique_jobs": 0
536 },
537 {
538     "term": "SGBD",
539     "match_count": 0,
540     "unique_jobs": 0
541 },
542 {
543     "term": "PostgreSQL",
544     "match_count": 116,
545     "unique_jobs": 55
546 },
547 {
548     "term": "MySQL",
549     "match_count": 85,
550     "unique_jobs": 45
551 },
552 {
553     "term": "SQLite",
554     "match_count": 0,
555     "unique_jobs": 0
556 },
557 {
```



```
558     "term": "Desenvolvimento com banco de dados",
559     "match_count": 0,
560     "unique_jobs": 0
561 },
562 {
563     "term": "Aplicacoes desktop e web com banco de dados",
564     "match_count": 0,
565     "unique_jobs": 0
566 },
567 {
568     "term": "Bancos de dados NoSQL",
569     "match_count": 2,
570     "unique_jobs": 1
571 },
572 {
573     "term": "MongoDB",
574     "match_count": 72,
575     "unique_jobs": 27
576 },
577 {
578     "term": "Administracao de banco de dados",
579     "match_count": 1,
580     "unique_jobs": 1
581 },
582 {
583     "term": "Design de banco de dados",
584     "match_count": 0,
585     "unique_jobs": 0
586 },
587 {
588     "term": "Otimizacao de queries",
589     "match_count": 10,
590     "unique_jobs": 6
591 },
592 {
593     "term": "Engenharia de software",
594     "match_count": 114,
595     "unique_jobs": 46
596 },
597 {
598     "term": "Analise orientada a objetos",
599     "match_count": 0,
600     "unique_jobs": 0
601 },
602 {
603     "term": "Projeto orientado a objetos",
604     "match_count": 0,
```

```
605     "unique_jobs": 0
606   },
607   {
608     "term": "UML",
609     "match_count": 8,
610     "unique_jobs": 4
611   },
612   {
613     "term": "Modelagem de software",
614     "match_count": 2,
615     "unique_jobs": 1
616   },
617   {
618     "term": "Levantamento de requisitos",
619     "match_count": 51,
620     "unique_jobs": 21
621   },
622   {
623     "term": "Casos de uso",
624     "match_count": 8,
625     "unique_jobs": 5
626   },
627   {
628     "term": "Arquitetura de software",
629     "match_count": 49,
630     "unique_jobs": 22
631   },
632   {
633     "term": "Padroes de projeto",
634     "match_count": 22,
635     "unique_jobs": 8
636   },
637   {
638     "term": "Frameworks",
639     "match_count": 166,
640     "unique_jobs": 71
641   },
642   {
643     "term": "Desenvolvimento agil",
644     "match_count": 17,
645     "unique_jobs": 8
646   },
647   {
648     "term": "Metodologias ageis",
649     "match_count": 255,
650     "unique_jobs": 114
651   },
```

```
652 {
653   "term": "Scrum",
654   "match_count": 239,
655   "unique_jobs": 103
656 },
657 {
658   "term": "Kanban",
659   "match_count": 191,
660   "unique_jobs": 82
661 },
662 {
663   "term": "Testes de software",
664   "match_count": 15,
665   "unique_jobs": 10
666 },
667 {
668   "term": "Gerenciamento de projetos de software",
669   "match_count": 0,
670   "unique_jobs": 0
671 },
672 {
673   "term": "Inteligencia artificial",
674   "match_count": 76,
675   "unique_jobs": 31
676 },
677 {
678   "term": "Aprendizado de maquina",
679   "match_count": 5,
680   "unique_jobs": 4
681 },
682 {
683   "term": "Machine learning",
684   "match_count": 29,
685   "unique_jobs": 20
686 },
687 {
688   "term": "Deep learning",
689   "match_count": 8,
690   "unique_jobs": 6
691 },
692 {
693   "term": "Preparacao de dados",
694   "match_count": 1,
695   "unique_jobs": 1
696 },
697 {
698   "term": "Analise de dados",
```

```
699     "match_count": 82,
700     "unique_jobs": 63
701 },
702 {
703     "term": "Modelos preditivos",
704     "match_count": 9,
705     "unique_jobs": 6
706 },
707 {
708     "term": "TensorFlow",
709     "match_count": 7,
710     "unique_jobs": 5
711 },
712 {
713     "term": "PyTorch",
714     "match_count": 7,
715     "unique_jobs": 5
716 },
717 {
718     "term": "Scikit-learn",
719     "match_count": 4,
720     "unique_jobs": 3
721 },
722 {
723     "term": "Avaliacao de modelos",
724     "match_count": 0,
725     "unique_jobs": 0
726 },
727 {
728     "term": "Agentes inteligentes",
729     "match_count": 0,
730     "unique_jobs": 0
731 },
732 {
733     "term": "Classificacao",
734     "match_count": 8,
735     "unique_jobs": 5
736 },
737 {
738     "term": "Regressao",
739     "match_count": 11,
740     "unique_jobs": 4
741 },
742 {
743     "term": "Redes neurais artificiais",
744     "match_count": 0,
745     "unique_jobs": 0
```

```
746 },
747 {
748     "term": "Algoritmos supervisionados e nao supervisionados",
749     "match_count": 0,
750     "unique_jobs": 0
751 },
752 {
753     "term": "Projeto de protocolos",
754     "match_count": 0,
755     "unique_jobs": 0
756 },
757 {
758     "term": "Especificacao de protocolos",
759     "match_count": 0,
760     "unique_jobs": 0
761 },
762 {
763     "term": "Comunicacao de rede",
764     "match_count": 0,
765     "unique_jobs": 0
766 },
767 {
768     "term": "Maquinas de estados",
769     "match_count": 0,
770     "unique_jobs": 0
771 },
772 {
773     "term": "FSM",
774     "match_count": 0,
775     "unique_jobs": 0
776 },
777 {
778     "term": "Modelagem de protocolos",
779     "match_count": 0,
780     "unique_jobs": 0
781 },
782 {
783     "term": "Codificacao de mensagens",
784     "match_count": 0,
785     "unique_jobs": 0
786 },
787 {
788     "term": "Programacao assincrona",
789     "match_count": 0,
790     "unique_jobs": 0
791 },
792 {
```

```
793     "term": "Model checking",
794     "match_count": 0,
795     "unique_jobs": 0
796 },
797 {
798     "term": "Verificacao formal",
799     "match_count": 0,
800     "unique_jobs": 0
801 },
802 {
803     "term": "Protocolos de comunicacao",
804     "match_count": 41,
805     "unique_jobs": 12
806 },
807 {
808     "term": "Implementacao de protocolos",
809     "match_count": 0,
810     "unique_jobs": 0
811 },
812 {
813     "term": "Propriedades de protocolos",
814     "match_count": 0,
815     "unique_jobs": 0
816 },
817 {
818     "term": "Redes de computadores",
819     "match_count": 96,
820     "unique_jobs": 32
821 },
822 {
823     "term": "Protocolos personalizados",
824     "match_count": 0,
825     "unique_jobs": 0
826 },
827 {
828     "term": "Tecnicas de modelagem de protocolos",
829     "match_count": 0,
830     "unique_jobs": 0
831 },
832 {
833     "term": "Sistemas de controle",
834     "match_count": 34,
835     "unique_jobs": 16
836 },
837 {
838     "term": "Controle classico",
839     "match_count": 0,
```

```
840     "unique_jobs": 0
841 },
842 {
843     "term": "Controle digital",
844     "match_count": 0,
845     "unique_jobs": 0
846 },
847 {
848     "term": "Analise de sistemas",
849     "match_count": 20,
850     "unique_jobs": 10
851 },
852 {
853     "term": "Resposta transitoria",
854     "match_count": 0,
855     "unique_jobs": 0
856 },
857 {
858     "term": "Resposta em regime permanente",
859     "match_count": 0,
860     "unique_jobs": 0
861 },
862 {
863     "term": "Lugar das raizes",
864     "match_count": 0,
865     "unique_jobs": 0
866 },
867 {
868     "term": "Resposta em frequencia",
869     "match_count": 0,
870     "unique_jobs": 0
871 },
872 {
873     "term": "Controladores PID",
874     "match_count": 0,
875     "unique_jobs": 0
876 },
877 {
878     "term": "Espaco de estados",
879     "match_count": 0,
880     "unique_jobs": 0
881 },
882 {
883     "term": "Projeto de controle",
884     "match_count": 0,
885     "unique_jobs": 0
886 },
```

```
887 {
888   "term": "Simulacao de sistemas de controle",
889   "match_count": 0,
890   "unique_jobs": 0
891 },
892 {
893   "term": "Matlab",
894   "match_count": 0,
895   "unique_jobs": 0
896 },
897 {
898   "term": "Simulink",
899   "match_count": 0,
900   "unique_jobs": 0
901 },
902 {
903   "term": "Compatibilidade eletromagnetica",
904   "match_count": 3,
905   "unique_jobs": 1
906 },
907 {
908   "term": "CEM",
909   "match_count": 2,
910   "unique_jobs": 2
911 },
912 {
913   "term": "Interferencia eletromagnetica",
914   "match_count": 16,
915   "unique_jobs": 2
916 },
917 {
918   "term": "EMI",
919   "match_count": 48,
920   "unique_jobs": 4
921 },
922 {
923   "term": "Emissoes irradiadas",
924   "match_count": 0,
925   "unique_jobs": 0
926 },
927 {
928   "term": "Emissoes conduzidas",
929   "match_count": 0,
930   "unique_jobs": 0
931 },
932 {
933   "term": "Suscetibilidade eletromagnetica",
```



```
934     "match_count": 0,
935     "unique_jobs": 0
936 },
937 {
938     "term": "Blindagem eletromagnetica",
939     "match_count": 0,
940     "unique_jobs": 0
941 },
942 {
943     "term": "Diafonia",
944     "match_count": 0,
945     "unique_jobs": 0
946 },
947 {
948     "term": "Descargas eletrostaticas",
949     "match_count": 0,
950     "unique_jobs": 0
951 },
952 {
953     "term": "Normas EMC",
954     "match_count": 0,
955     "unique_jobs": 0
956 },
957 {
958     "term": "Medicoes eletromagneticas",
959     "match_count": 0,
960     "unique_jobs": 0
961 },
962 {
963     "term": "Espectro eletromagnetico",
964     "match_count": 0,
965     "unique_jobs": 0
966 },
967 {
968     "term": "Principios eletromagneticos",
969     "match_count": 0,
970     "unique_jobs": 0
971 },
972 {
973     "term": "Eletronica de potencia",
974     "match_count": 16,
975     "unique_jobs": 3
976 },
977 {
978     "term": "Conversores estaticos",
979     "match_count": 0,
980     "unique_jobs": 0
```

```
981 },
982 {
983     "term": "Retificadores monofasicos e trifasicos",
984     "match_count": 0,
985     "unique_jobs": 0
986 },
987 {
988     "term": "Variadores de tensao",
989     "match_count": 0,
990     "unique_jobs": 0
991 },
992 {
993     "term": "CA-CC",
994     "match_count": 0,
995     "unique_jobs": 0
996 },
997 {
998     "term": "CA-CA",
999     "match_count": 0,
1000     "unique_jobs": 0
1001 },
1002 {
1003     "term": "CC-CC",
1004     "match_count": 0,
1005     "unique_jobs": 0
1006 },
1007 {
1008     "term": "CC-CA",
1009     "match_count": 0,
1010     "unique_jobs": 0
1011 },
1012 {
1013     "term": "MOSFET",
1014     "match_count": 0,
1015     "unique_jobs": 0
1016 },
1017 {
1018     "term": "IGBT",
1019     "match_count": 0,
1020     "unique_jobs": 0
1021 },
1022 {
1023     "term": "Tiristores",
1024     "match_count": 7,
1025     "unique_jobs": 1
1026 },
1027 {
```

```
1028     "term": "Fontes chaveadas",
1029     "match_count": 0,
1030     "unique_jobs": 0
1031 },
1032 {
1033     "term": "Filtros ativos",
1034     "match_count": 0,
1035     "unique_jobs": 0
1036 },
1037 {
1038     "term": "UPS",
1039     "match_count": 8,
1040     "unique_jobs": 7
1041 },
1042 {
1043     "term": "Correcao de fator de potencia",
1044     "match_count": 0,
1045     "unique_jobs": 0
1046 },
1047 {
1048     "term": "Acionamento de motores",
1049     "match_count": 36,
1050     "unique_jobs": 3
1051 },
1052 {
1053     "term": "Inversores de frequencia",
1054     "match_count": 31,
1055     "unique_jobs": 14
1056 },
1057 {
1058     "term": "Simulacao de conversores",
1059     "match_count": 0,
1060     "unique_jobs": 0
1061 },
1062 {
1063     "term": "Reatores eletronicos",
1064     "match_count": 0,
1065     "unique_jobs": 0
1066 },
1067 {
1068     "term": "Teoria da informacao",
1069     "match_count": 0,
1070     "unique_jobs": 0
1071 },
1072 {
1073     "term": "Codificacao de fonte",
1074     "match_count": 0,
```

```
1075     "unique_jobs": 0
1076 },
1077 {
1078     "term": "Codificacao de canal",
1079     "match_count": 18,
1080     "unique_jobs": 1
1081 },
1082 {
1083     "term": "Codigos corretores de erro",
1084     "match_count": 0,
1085     "unique_jobs": 0
1086 },
1087 {
1088     "term": "Shannon",
1089     "match_count": 0,
1090     "unique_jobs": 0
1091 },
1092 {
1093     "term": "Capacidade do canal",
1094     "match_count": 0,
1095     "unique_jobs": 0
1096 },
1097 {
1098     "term": "Entropia",
1099     "match_count": 0,
1100     "unique_jobs": 0
1101 },
1102 {
1103     "term": "Compressao de dados",
1104     "match_count": 0,
1105     "unique_jobs": 0
1106 },
1107 {
1108     "term": "Deteccao e correcao de erros",
1109     "match_count": 0,
1110     "unique_jobs": 0
1111 },
1112 {
1113     "term": "Codigos de Hamming",
1114     "match_count": 0,
1115     "unique_jobs": 0
1116 },
1117 {
1118     "term": "Codigos convolucionais",
1119     "match_count": 0,
1120     "unique_jobs": 0
1121 },
```

```
1122 {  
1123   "term": "Codigos LDPC",  
1124   "match_count": 0,  
1125   "unique_jobs": 0  
1126 }  
1127 ]
```

APÊNDICE G – SCRIPT PARA ILUSTRAR RESULTADOS: OCORRÊNCIA DE PALAVRAS-CHAVE

Código G.1 – Código-fonte do script `term_frequency_keywords.py`

```
1 import json
2 import matplotlib.pyplot as plt
3
4 with open("term_frequency_report.json", encoding="utf-8") as f:
5     report = json.load(f)
6
7 # 2. Ordena
8 report_sorted = sorted(report, key=lambda x: x["match_count"], reverse=True)
9
10 # 3. Seleciona os 50 termos mais frequentes
11 top_n = 50
12 top_terms = report_sorted[:top_n]
13 terms = [item["term"] for item in top_terms]
14 counts = [item["match_count"] for item in top_terms]
15 jobs = [item["unique_jobs"] for item in top_terms]
16
17 # 4. Calcular o total de descrições de vagas
18 total_jobs = sum(jobs)
19
20 # 5. Gera gráfico
21 plt.figure(figsize=(14, 8))
22 plt.bar(terms, counts, color='skyblue')
23 plt.xlabel("Palavras-chave", fontsize=16)
24 plt.ylabel("Número de ocorrências nas vagas", fontsize=16)
25 plt.xticks(rotation=90, fontsize=12)
26 plt.yticks(fontsize=12)
27 plt.tight_layout(pad=6.0)
28 plt.savefig("top50_frequencia_termos.png")
```

APÊNDICE H – SCRIPT PARA ILUSTRAR RESULTADOS: OCORRÊNCIA DE PALAVRAS-CHAVE POR DISCIPLINA

Código H.1 – Código-fonte do script `term_frequency_courses.py`

```

1 import json
2 import pandas as pd
3 import matplotlib.pyplot as plt
4
5 # CONFIGURACAO
6 json_path      = "term_frequency_report.json"
7 excel_path     = "CHA_apos_normalizacao_opt.xlsx"
8 sheet_name     = 0      # Primeira aba (ou pelo nome, se preferir)
9 top_n_terms    = 50     # termos de maior frequencia considerados
10 top_n_disciplinas = 20  # num de barras exibidas no gráfico
11
12 # Passo 1 - Carrega o relatorio de frequencia de termos
13 with open(json_path, encoding="utf-8") as f:
14     report = json.load(f)
15
16 freq_df = pd.DataFrame(report)
17 freq_df["term"] = freq_df["term"].str.lower()
18
19 if top_n_terms:
20     freq_df = freq_df.nlargest(top_n_terms, "match_count")
21
22 # Passo 2 - Carrega a planilha
23 cha_df = pd.read_excel(excel_path, sheet_name=sheet_name)
24
25 records = []
26 for _, row in cha_df.iterrows():
27     disciplina = row["Disciplina"]
28     raw_kw     = row["Palavras-chave"]
29     if pd.notna(raw_kw):
30         for kw in map(str.strip, raw_kw.lower().split(", ")):
31             if kw:
32                 records.append({"keyword": kw, "disciplina": disciplina})
33
34 keywords_df = pd.DataFrame(records)
35
36 # Passo 3 - Junta frequencias e palavras-chave
37 merged = pd.merge(freq_df,
38                   keywords_df,
39                   left_on="term",
40                   right_on="keyword",

```

```
41         how="inner") # apenas termos que aparecem na planilha
42
43 # Passo 4 - Soma ocorrencias por disciplina
44 disc_summary = (merged.groupby("disciplina")["match_count"]
45                 .sum()
46                 .reset_index()
47                 .sort_values("match_count", ascending=False))
48
49 # Passo 5 - Gera o grafico
50 if top_n_disciplinas:
51     disc_summary = disc_summary.head(top_n_disciplinas)
52
53 plt.figure(figsize=(12, 8))
54 plt.barh(disc_summary["disciplina"],
55          disc_summary["match_count"],
56          color="#6baed6",
57          edgecolor="#6baed6")
58 plt.gca().invert_yaxis()
59 plt.xlabel("Ocorrências das palavras-chave", fontsize=12)
60 plt.ylabel("Disciplina optativa", fontsize=12)
61 plt.tight_layout(pad=3)
62 plt.savefig("ocorrencias_por_disciplina_optativas.png", dpi=300)
63 plt.close()
64
65 print("Gráfico salvo em 'ocorrencias_por_disciplina_optativas.png'")
```