



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

COLETA E TRATAMENTO DE DADOS SOBRE A PRODUÇÃO TÉCNICA BRASILEIRA: UM ESTUDO BASEADO EM PATENTES

RAULIVAN RODRIGO DA SILVA

Orientador: Prof. Dr. Thiago Magela Rodrigues Dias
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientador: Prof. Dr. Washington Luís Ribeiro de Carvalho Segundo
Instituto Brasileiro de Informação em Ciência e Tecnologia, IBICT, Brasil

BELO HORIZONTE
MARÇO DE 2022

RAULIVAN RODRIGO DA SILVA

**COLETA E TRATAMENTO DE DADOS SOBRE A
PRODUÇÃO TÉCNICA BRASILEIRA: UM ESTUDO
BASEADO EM PATENTES**

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Prof. Dr. Thiago Magela Rodrigues Dias
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientador: Prof. Dr. Washington Luís Ribeiro de Carvalho Segundo
Instituto Brasileiro de Informação em Ciência e Tecnologia, IBICT, Brasil

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
MARÇO DE 2022

S586c Silva, Raulivan Rodrigo da
Coleta e tratamento de dados sobre a produção técnica brasileira: um estudo baseado em patentes / Raulivan Rodrigo da Silva. – 2022.
77 f. : il.

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientador: Thiago Magela Rodrigues Dias.
Coorientador: Washington Luís Ribeiro de Carvalho Segundo.
Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Banco de dados. 2. Mineração de dados (Computação). 3. Patentes. I. Dias, Thiago Magela Rodrigues. II. Segundo, Washington Luís Ribeiro de Carvalho. III. Centro Federal de Educação Tecnológica de Minas Gerais. IV. Título.

CDD: 005.74



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO CURSO DE MESTRADO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

**“COLETA E TRATAMENTO DE DADOS SOBRE A PRODUÇÃO
TÉCNICA BRASILEIRA: UM ESTUDO BASEADO EM PATENTES”**

Dissertação de Mestrado apresentada por **Raulivan Rodrigo da Silva**, em 17 de março de 2022, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Thiago Magela Rodrigues Dias
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Washington Luís Ribeiro de Carvalho Segundo
Instituto Brasileiro de Informação em Ciência e Tecnologia

Prof. Dr. Jesús Pascual Mena Chalco
Universidade Federal do ABC

Prof. Dr. Alisson Marques da Silva
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Profª. Drª. Elizabeth Fialho Wanner
Presidenta do Colegiado do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

Dedico este trabalho aos meus pais, minha esposa e aqueles(as) que estiveram ao meu lado em toda esta trajetória.

Agradecimentos

Agradeço primeiramente a Deus por ter me abençoado durante este estudo com saúde, força e sabedoria para conseguir chegar aqui.

Aos meus pais, minha irmã, esposa e filhos, que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava à realização deste estudo.

A todos meus professores que compartilharam seus conhecimentos, em especial ao meu orientador Dr. Thiago Magela Rodrigues Dias e meu co-orientador Dr. Washington Luís Ribeiro de Carvalho Segundo, por acreditar e incentivar o desenvolvimento deste estudo, apoiando e orientando com dedicação e amizade.

Ao Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG), pelo suporte e incentivo em minha formação acadêmica, em que proporcionou toda estrutura adequada para que eu pudesse executar as minhas atividades.

A todos que participaram, direta ou indiretamente do desenvolvimento deste estudo, contribuindo e enriquecendo o meu processo de aprendizado.

“A sabedoria oferece proteção, como o faz o dinheiro, mas a vantagem do conhecimento é esta: a sabedoria preserva a vida de quem a possui.”
(Eclesiastes 7:12)

Resumo

Este trabalho busca contribuir com a compreensão do cenário tecnológico nacional, tendo como principal objetivo coletar, agregar e realizar tratamentos em grandes conjuntos de dados de patentes, para que dessa forma seja possível traçar uma visão geral da produção técnica brasileira. Tal análise é fundamentada na coleta e tratamento de patentes depositadas no Instituto Nacional da Propriedade Industrial (INPI) e disponibilizadas no repositório internacional de patentes, Espacenet. Inicialmente, é apresentado um conjunto de estratégias que propiciou a coleta, tratamento e agregação dos dados contidos em documentos de patentes, processo este, que viabilizou a construção de uma base de dados local, flexível e ampla, composta por dados provenientes do INPI e Espacenet. Para agregar mais valor à base de dados, foram também coletados registros da base curricular da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A fim verificar a representatividade das patentes registradas nos currículos da plataforma e sua validação junto aos dados coletados na Espacenet. Os dados coletados foram analisados e os resultados são apresentados sob diversas perspectivas, por meio de técnicas de bibliometria, patentometria, mineração de textos e algoritmos de processamento de linguagem natural, foram levantados indicadores quantitativos que caracterizam a produção técnica brasileira, tais como, evolução temporal, áreas do conhecimento que mais depositam patentes de acordo com as classificações obtidas, maiores depositantes, instituições de ensino superior que mais depositam patentes, entre outras. Consequentemente, foi possível salientar que estudar os diversos aspectos da evolução tecnológica com base em informações oriundas de documentos de patentes, além de ser relevante, possibilita compreender tendências tecnológicas e identificar especialistas em determinadas áreas do conhecimento. Além disso, gera uma formulação de políticas e estratégias que potencialize um diferencial competitivo para empresas e organizações atuantes no ramo da inovação e tecnologia. Ademais, destaca-se o grande esforço despendido para análise de um grande volume de dados de patentes, que corrobora com a proposição de uma base de dados local.

Palavras-chave: Patentes, INPI, Plataforma Lattes, Patentometria, Bibliometria, Espacenet.

Abstract

This work seeks to contribute to the understanding of the national technological scenario, with the main objective of collecting, aggregating and performing treatments on large sets of patent data, so that it is possible to draw an overview of Brazilian technical production. Such analysis is based on the collection and treatment of patents deposited at the National Institute of Industrial Property (INPI) and made available in the international patent repository, Espacenet. Initially, a set of strategies is presented that allowed the collection, treatment and aggregation of data contained in patent documents, a process that enabled the construction of a local, flexible and broad database, composed of data from the INPI and Espacenet. . To add more value to the database, records were also collected from the curriculum base of the Lattes Platform of the National Council for Scientific and Technological Development (CNPq). In order to verify the representativeness of the patents registered in the platform's curricula and their validation with the data collected on Espacenet. The collected data were analyzed and the results are presented from different perspectives, through bibliometrics, patentometrics, text mining and natural language processing algorithms, quantitative indicators were raised that characterize the Brazilian technical production, such as temporal evolution , areas of knowledge that deposit the most patents according to the classifications obtained, the largest depositors, higher education institutions that deposit the most patents, among others. Consequently, it was possible to point out that studying the various aspects of technological evolution based on information from patent documents, in addition to being relevant, makes it possible to understand technological trends and identify specialists in certain areas of knowledge. In addition, it generates a formulation of policies and strategies that enhance a competitive advantage for companies and organizations operating in the field of innovation and technology. Furthermore, the great effort spent to analyze a large volume of patent data stands out, which corroborates the proposition of a local database.

Keywords: Patents, INPI, Lattes Platform, Patentometrics, Bibliometrics, Espacenet.

Lista de Figuras

Figura 1 – Ciclo de relacionamento entre criador e sociedade	7
Figura 2 – Grupos da Propriedade Intelectual	7
Figura 3 – Detalhamento de uma classificação de patente	12
Figura 4 – Linha do tempo da evolução dos estudos métricos.	16
Figura 5 – Diagrama de inter-relação entre os subcampos	17
Figura 6 – Visão geral da coleta de dados	26
Figura 7 – Formato de numeração antiga	31
Figura 8 – Formato atual de numeração	32
Figura 9 – Processo de pesquisa de patentes	35
Figura 10 – Visão geral do LattesDataXplorer	36
Figura 11 – Visão geral do processo de validação	37
Figura 12 – Representação da estrutura do arquivo	39
Figura 13 – Organização de arquivos de patente	40
Figura 14 – Diagrama de entidade e relacionamento	42
Figura 15 – Evolução temporal do depósito de patentes por ano	45
Figura 16 – Taxa percentual de crescimento anual de depósitos de patentes	46
Figura 17 – Patentes por classificação	47
Figura 18 – Patentes por seção e classe	51
Figura 19 – 10 maiores depositantes de patentes no Brasil	52
Figura 20 – Série temporal do depósito de patentes	53
Figura 21 – 10 maiores IES depositantes de patentes	54
Figura 22 – Distribuição do número de depósito de patentes de IES por estado.	55
Figura 23 – Nuvem de palavras	56
Figura 24 – Quantitativo de Patentes coletadas nos currículos da Plataforma Lattes	58
Figura 25 – Análise temporal do depósito das patentes informadas nos currículos da Plataforma Lattes	59
Figura 26 – Classificação das patentes registradas nos currículos da Plataforma Lattes.	60
Figura 27 – Formação dos proponentes de pantes	62
Figura 28 – Grande área de atuação dos proponentes de pantes	63
Figura 29 – Distribuição de proponentes de patentes por estado	64

Lista de Tabelas

Tabela 1 – Sumarização dos dados coletados no INPI	29
Tabela 3 – Natureza da Proteção	31
Tabela 4 – Exemplos de códigos de publicação	32
Tabela 5 – Listagem de pastas	40
Tabela 7 – Quantitativo das patentes identificadas	44
Tabela 9 – Indicador de classe da seção A	48
Tabela 10 – indicador de classe da seção b	49
Tabela 11 – indicador de classe da seção c	50
Tabela 12 – Palavras mais utilizadas	57
Tabela 13 – Classificação das patentes registradas nos currículos da Plataforma Lattes	61

Lista de Algoritmos

Algoritmo 1 – COLETA DE DADOS NO INPI	28
---	----

Lista de Abreviaturas e Siglas

CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CIP	Classificação Internacional de Patente
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPC	<i>Cooperative Patent Classification</i>
CPF	Cadastro de Pessoa Física
CSV	<i>comma-separated-values</i>
C&T	Ciência e Tecnologia
DBLP	<i>Digital Bibliography Library Project</i>
DOI	<i>Digital Object Identifier</i>
DOU	Diário Oficial da União
EPO	<i>European Patent Office</i>
EUA	<i>Estados Unidos da América</i>
Gb	<i>Gigabit</i>
HTML	<i>HyperText Markup Language</i>
ICT	Instituições de Ciência e Tecnologia
IES	Instituições de Ensino Superior
INPI	Instituto Nacional de Propriedade Industrial
IPC	<i>International Patent Classification</i>
JPO	<i>Japan Patent Office</i>
JSON	<i>JavaScript Object Notation</i>
MEC	Ministério da Educação
MDIC	Ministério do Desenvolvimento, Indústria e Comércio Exterior
MU	Patente de Modelo de Utilidade
NIT	Núcleo de Inovação Tecnológica

OMPI	Organização Mundial de Propriedade Intelectual
OPS	<i>Open Patent Services</i>
PDF	<i>Portable Document Format</i>
PI	Patente de Invenção
pePI	Pesquisa em Propriedade Industrial
SGBDOR	Sistema Gerenciador de Banco de Dados Objeto-Relacional
SIPO	<i>State Intellectual Property Office of China</i>
SQL	<i>Structured Query Language</i>
UFMG	Universidade Federal de Minas Gerais
URL	<i>Uniform Resource Locator</i>
UNICAMP	Universidade Estadual de Campinas
USP	Universidade de São Paulo
USPTO	<i>United States Patent and Trademark Office</i>
XML	<i>eXtensible Markup Language</i>
WEB	<i>World Wide Web</i>
WIPO	<i>World Intellectual Property Organization</i>
WWW	<i>World Wide Web</i>

Sumário

1 – Introdução	1
1.1 Descrição do problema	2
1.2 Motivação	4
1.3 Objetivos	4
1.4 Organização do trabalho	5
2 – Fundamentação Teórica	6
2.1 Propriedade Intelectual	6
2.2 Patentes	8
2.3 Repositórios para consulta de documentos de patentes	13
2.4 Cientometria	15
2.4.1 Informetria	17
2.4.2 Bibliometria	17
2.4.3 Webometria	18
2.4.4 Patentometria	18
3 – Trabalhos Relacionados	20
3.1 Patentes como fonte de dados de pesquisa	20
3.2 Diferenciais propostos	23
4 – Metodologia	25
4.1 Aquisição dos dados	25
4.1.1 Coleta de dados no INPI	26
4.1.2 Coleta de dados na Espacenet	30
4.1.3 Coleta de dados na Plataforma Lattes	35
4.2 Seleção, tratamento e organização dos dados	38
5 – Análise e Discussão dos Resultados	44
5.1 Caracterização geral	44
5.2 Repositórios internacionais como mecanismo de certificação	57
6 – Conclusão	66
6.1 Publicações	69
6.2 Trabalhos Futuros	69
Referências	71

1 Introdução

O século XXI tem sido solo fértil para a criação de estruturas tecnológicas e, mais do que nunca, a rapidez na evolução destas tecnologias tem sido visível. Diariamente novos dispositivos, aplicações, meios digitais permeiam o mercado, trazendo versões melhores de recursos e/ou funcionalidades que até então conhecíamos ou apresentavam novas soluções. Em consequência disso, as organizações desse mercado estão empenhadas em realizar um monitoramento constante de suas atividades e da viabilidade de seus produtos e serviços oferecidos, sendo necessário para implementar inovações que fidelizem ou aumentem a base de clientes (AMADEI; TORKOMIAN, 2009). Boa parte das invenções e inovações geradas possuem como origem pesquisas iniciadas em instituições de ensino públicas e privadas, que vem aumentando ao longo dos anos tornando-as um grande polo de inovação nacional. Este fator impulsionou a criação de leis e normas que visam proteger a propriedade intelectual gerada em universidades e também núcleos que visam auxiliar no processo de criação de patentes e viabilizando o relacionamento entre os setores acadêmicos e mercadológicos, fomentando para que instituições de ensino e pesquisadores se inclinem cada vez mais ao depósito de patentes (MUELLER; PERUCCHI, 2014).

Mediante ao exposto, acompanhar o desenvolvimento tecnológico, de um setor, ou um grupo de empresas tem se tornado um pré-requisito para empresas que desejam se destacar no meio tecnológico e de inovação. Dentre as diversas formas de acompanhar esse desenvolvimento, segundo Nascimento e Speziali (2020), o uso de informações tecnológicas contidas em documentos de patentes é uma excelente alternativa. A compreensão do estado da técnica da arte por meio de documentos de patentes, consequentemente apresenta um cenário mais assertivo a respeito de tendências tecnológicas, setores promissores, bem como, a possibilidade de novas tecnologias (NASCIMENTO; SPEZIALI, 2020). Sanz-Casado (2006) aponta que estudos e análise de documentos de patentes permitem identificar o conhecimento científico e convertê-lo em conhecimento tecnológico.

Assim como existem diversos repositórios disponíveis na internet que possibilitam a pesquisa de produções científicas publicadas, a saber, DBLP (*Digital Bibliography Library Project*)¹, ArnetMiner², Google Scholar³, Microsoft Academic Search⁴ e a Plataforma Lattes do CNPq⁵, sendo essa última um instrumento extremamente rico em dados para estudos sobre a produção científica e técnica brasileira. No contexto da produção técnica, as bases de dados de patentes oferecem acesso a milhares de documentos por meio da

¹ <<https://dblp.org>>

² <<https://www.aminer.org>>

³ <<https://scholar.google.com.br>>

⁴ <<https://academic.microsoft.com/home>>

⁵ <<http://lattes.cnpq.br>>

internet, como o pePI (Pesquisa em Propriedade Industrial)⁶ mantido pelo órgão brasileiro de gestão de patentes INPI (Instituto Nacional de Propriedade Intelectual)⁷. Cada país possui seu órgão responsável por gerenciar o depósito, conceder a patente e disponibilizar para consulta. Além disso, existem os repositórios internacionais, como a Espacenet⁸ de reconhecida relevância, tendo em vista a sua abrangência em armazenar e disponibilizar os dados registrados, viabilizando em um único repositório, a consulta de patentes de aproximadamente 70 países. Tais repositórios propiciam a pesquisa no estado da arte fundamentada em documentos de patentes, permitindo ao pesquisador obter acesso a manuais, fórmulas, desenhos técnicos, esquemas, a toda documentação necessária para reproduzir o invento em questão. Proporcionando a criação de novos produtos ou processos sendo estes, resultado do esforço do desenvolvimento de novas tecnologias ou combinações de tecnologias já existentes.

Logo, utilizando documentos de patentes, este trabalho tem como um dos intuitos definir estratégias para coleta de dados de patentes em repositórios nacionais e internacionais, de igual modo, em currículos da Plataforma Lattes que contenham informações de patentes, construindo assim, uma base de dados local, a fim de investigar fatores e indicadores que caracterizam o cenário tecnológico nacional. Além disso, com os dados da Plataforma Lattes extraídos, é possível validar por meio de cruzamentos entre os dados coletados e os dados patentários informados pelos pesquisadores em seus currículos, bem como avaliar a representatividade dos dados cadastrados nos currículos em repositórios internacionais, como a Espacenet.

1.1 Descrição do problema

De acordo com dados levantados pelo INPI (2021c), são publicados anualmente cerca de 28 mil registros de novas patentes, um rico arcabouço de informações tecnológicas em diversas áreas do conhecimento disponíveis para serem exploradas. Porém não é feito um levantamento de quais ou quantas foram advindas de projetos acadêmicos ou de membros da academia, qual é o percentual de contribuição que as universidades geram para o depósito de patentes, ou até mesmo quais são as áreas do conhecimento mais representativas, qual perfil de inventores mais contribuem para desenvolvimento tecnológico.

Conforme apontam Nascimento e Speziali (2020) o mapeamento de tecnologias utilizando informações contidas em documentos de patentes, é pouco explorado no Brasil, salientando que, por fazer parte de áreas estratégicas de muitas empresas, relatórios de mapeamentos tecnológicos mais completos não são disponibilizados de forma gratuita para consulta. Calzolaio et al. (2018) e Tanaka e Inui (2016) concordam que os dados de

⁶ <<https://busca.inpi.gov.br/pePI/jsp/patentes/PatenteSearchBasico.jsp>>

⁷ <<https://www.gov.br/inpi/pt-br>>

⁸ <<https://worldwide.espacenet.com>>

patentes contêm informações valiosas para análises técnicas, entretanto, esta é uma área pouco explorada também pelas universidades. A prospecção tecnológica é uma área de estudo recente principalmente no Brasil, que possui uma literatura limitada sobre o tema.

Publicações da Organização para a Cooperação e Desenvolvimento Econômico (OCED) e da revista Cadernos de Prospecção Tecnológica e do Fórum Nacional de Gestores de Inovação e Transferência de Tecnologia (FORTEC) evidenciam a importância de se utilizar documentos de patentes como fonte de informação tecnológica (SILVA; NASCIMENTO, 2020). Diversos trabalhos incluindo patentes têm citado técnicas para o gerenciamento dos dados que são publicados neste tipo de documento (IBM (2022); NCET (2019); HQIT (2017)).

No entanto, ter acesso aos dados de documentos de patentes não é considerada uma tarefa trivial. Apesar de existirem diversos repositórios que disponibilizam o acesso pago ou gratuito aos dados de patente, como o INPI e Espacenet, manipular grande volume de dados ocasiona limitações na recuperação dos dados. Essas limitações seriam: limite no volume de dados a ser manipulado em um único processo; limite de horários de acesso aos dados; limite de filtros e critérios de seleção dentre outros. Isso dificulta análises mais amplas, que abrangem um alto volume de dados, como por exemplo, avaliar o cenário de um país ou conjunto de instituições.

Nesse âmbito, a Plataforma Lattes tem recebido destaque especial no setor acadêmico nacional como um dos principais repositórios para registros de informações de publicações científicas e tecnológicas em diversas áreas de conhecimento. Os currículos que compõem a Plataforma Lattes também possuem uma área destinada a registrar informações referente a patentes, sendo estas de autoria única ou realizadas em colaboração com outros proponentes de empresas ou de instituições acadêmicas. Vale ressaltar que atualmente a Plataforma Lattes não possibilita a pesquisa em seus currículos a partir de registros de produções, entretanto permite a pesquisa a partir de nomes dos indivíduos mas, limitando o acesso para visualização unitária de cada currículo. Percebe-se que com a junção dos dados coletados em repositórios específicos de Patentes, aliado à informações registradas nos currículos da Plataforma Lattes, como formação acadêmica e áreas de atuação, viabilizará traçar um panorama inédito sobre como tem acontecido a produção técnica no país.

Nessa conjuntura, gerar inteligência de patentes, transformando dados em informações substanciais para análises técnicas, de fato é uma tarefa complexa com lacunas a serem exploradas.

1.2 Motivação

Os principais fatores que motivam o estudo para a compreensão do cenário tecnológico nacional é a oportunidade de compreender por meio de documentos de patentes a evolução tecnológica brasileira e seus diversos aspectos, tais como, os principais depositantes, as áreas do conhecimento que possuem maior representatividade, a evolução temporal, regiões do Brasil que mais produzem patentes. A oportunidade de traçar um retrato pouco explorado da produção técnica brasileira, demonstrar como a análise de informações patentárias pode auxiliar os núcleos de inovação tecnológica no processo de gestão da inovação tecnológica.

Consequentemente, contribuir e fomentar o desenvolvimento de pesquisas que tratem da temática, fornecendo uma base de dados composta por dados extraídos de patentes e dados de currículos da Plataforma Lattes, contribuindo assim com avanço da Ciência e Tecnologia.

1.3 Objetivos

Este trabalho tem como objetivo geral a coleta, organização e tratamentos de grandes conjuntos de dados de patentes. Para que dessa forma seja viabilizada diversas análises que utilizam-se de métricas bibliométricas e patentométricas para que seja possível construir retratos quantitativos e temporais focados em patentes, visando avaliá-las por áreas de conhecimento e instituições.

Para tanto, será necessário realizar os seguintes objetivos específicos:

- Processar e organizar o conjunto de dados de patentes inseridas nos currículos cadastrados na Plataforma Lattes, através da ferramenta LattesDataXplorer (DIAS, 2016);
- Recuperar o número do pedido, data de depósito, data de publicação, título, depositante, inventor, classificação ICP (*International Patent Classification*) das patentes registradas no INPI no período de janeiro de 1900 à dezembro de 2020;
- Recuperar e validar os dados coletados das patentes no repositório da Espacenet;
- Analisar os dados e definir melhor técnica de cruzamento das informações recuperadas;
- Organizar todos os conjuntos de dados coletados em um único repositório local.
- Elencar e aplicar as métricas nos dados de produção técnica (métodos quantitativos e temporais).

1.4 Organização do trabalho

O restante desta Dissertação está organizada da seguinte forma:

- No [Capítulo 2](#), é apresentada uma revisão bibliográfica que aborda os assuntos permeados por esta Dissertação. Este capítulo inclui os conceitos de patentes e suas características, sobre a bibliometria e suas principais métricas.
- No [Capítulo 3](#), são apresentados, discutidos e comparados alguns trabalhos relacionados aos assuntos abordados nesta Dissertação.
- No [Capítulo 4](#), é apresentado um arcabouço desenvolvido para identificação, coleta, tratamento e análise de dados do INPI, Espacenet e Plataforma Lattes, de igual modo, das extensões que foram desenvolvidos para o tratamento e análise dos dados.
- No [Capítulo 5](#), é apresentada uma caracterização geral da produção científica e tecnologia do território nacional, baseado nos dados coletados. De igual modo, é apresentado um panorama da representatividade da produção técnica registrada nos currículos da Plataforma Lattes.
- Por fim, no [Capítulo 6](#), são apresentadas as considerações finais destacando os principais resultados obtidos.

2 Fundamentação Teórica

Este capítulo apresenta um estudo bibliográfico referente aos assuntos abordados nesta Dissertação. Nas próximas seções, são apresentados: introdução e definição de propriedade intelectual ([Seção 2.1](#)), conceitos referente a documentos de patentes e seus processos ([Seção 2.2](#)), apresentação dos principais repositórios de consulta de documentos de patentes ([Seção 2.3](#)), e por fim, o contexto histórico e conceitos de cientometria ([Seção 2.4](#)).

2.1 Propriedade Intelectual

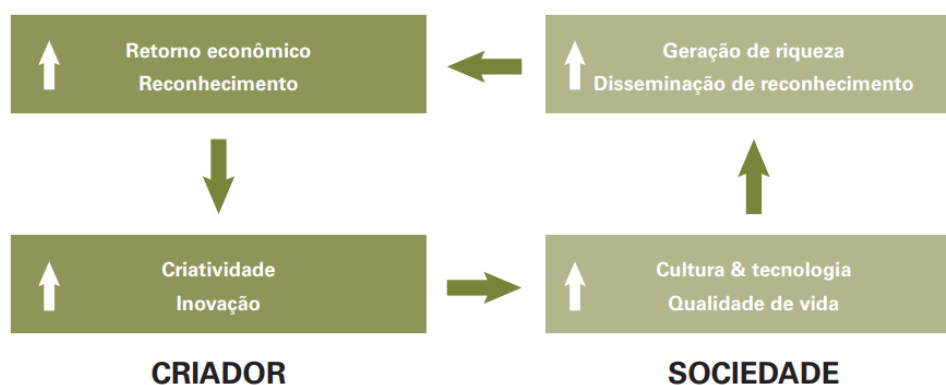
Para melhor compreensão do tópico Propriedade Intelectual se faz necessário conceituar separadamente as palavras “propriedade” e “intelectual”. De acordo com [Ferreira \(2010\)](#), “propriedade” é o “direito de usar, gozar e dispor de bens e de reavê-los do poder de quem quer que injustamente os possua” e “bens sobre os quais se exerce esse direito”; O adjetivo “intelectual” como “possuir dotes de espírito, de inteligência”. Dessa forma, é possível concluir que consiste em ter direito a tudo aquilo que pode ser considerado inventividade, conhecimento.

A Organização Mundial da Propriedade Intelectual (OMPI), em inglês *World Intellectual Property Organization* (WIPO), é uma entidade internacional criada em 1967 dentro do sistema das Nações Unidas, incluindo as Uniões de Paris e de Berna, que tem como objetivo a promoção da proteção da propriedade intelectual ao redor do mundo através da cooperação entre Estados ([BARBOSA, 2010](#)).

A Convenção da OMPI define como Propriedade intelectual, a soma dos direitos relativos às obras literárias, artísticas e científicas, às interpretações dos artistas intérpretes e às execuções dos artistas executantes, aos fonogramas e às emissões de radiodifusão, às invenções em todos os domínios da atividade humana, às descobertas científicas, aos desenhos e modelos industriais, às marcas industriais, comerciais e de serviço, bem como às firmas comerciais e denominações comerciais, à proteção contra a concorrência desleal e todos os outros direitos inerentes à atividade intelectual nos domínios industrial, científico, literário e artístico ([BARBOSA, 2010](#)).

Logo, no intuito de melhor compreender a relação existente entre algumas entidades envolvidas no processo de invenção, a [Figura 1](#) ilustra o ciclo em que estabelece as relações entre o inventor e a sociedade, em que o criador contribui para com a sociedade com um invento que melhora a qualidade de vida. Então, a sociedade por sua vez, reconhece e gera retorno financeiro ao inventor.

Figura 1 – Ciclo de relacionamento entre criador e sociedade

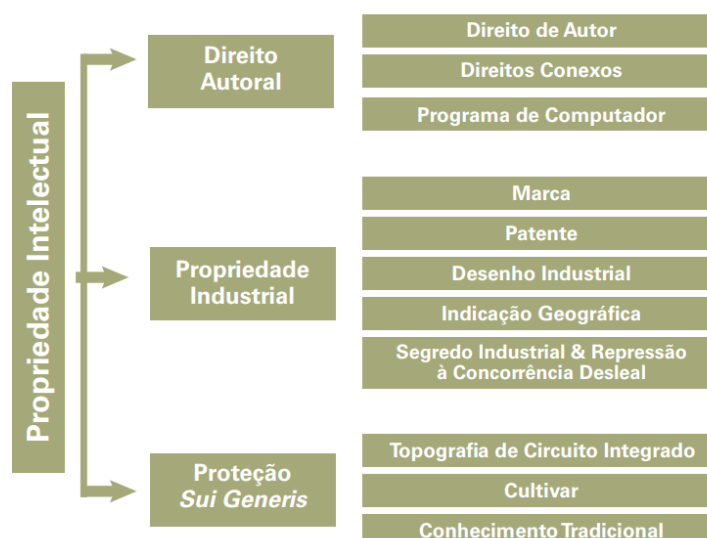


Fonte: Jungmann e Bonetti. (2010)

Conforme esboçado pela Figura 1 os inventores e depositantes, caracterizados como criadores da patente, geram produtos e/ou serviços que contribuem com a qualidade de vida da sociedade e consequentemente recebem o retorno financeiro e reconhecimento como fruto do invento depositado. Já a sociedade, consome este produto e/ou serviço, gerando riqueza e reconhecimento aos criadores. Esboçando de maneira sucinta o processo de Proteção Intelectual definido pela OMPI.

De acordo com Jungmann e Bonetti. (2010) a Propriedade Intelectual se divide em três grandes grupos, “direito autoral”, “propriedade industrial” e “proteção *sui generis*”. Neste contexto Quoniam, Kniess e Mazieri (2014) destacam que patentes, que é parte fundamental deste estudo, pertence ao grupo “propriedade industrial”, junto com marcas, desenhos industriais e indicações geográficas de origem, conforme é ilustrado na Figura 2.

Figura 2 – Grupos da Propriedade Intelectual



Fonte: Jungmann e Bonetti. (2010)

A propriedade intelectual é a área do Direito que, por meio de leis, estabelece um conjunto de direitos que garante aos inventores ou responsáveis o direito de obter por tempo determinado recompensa pela própria criação, seja ela industrial, científica, literária ou artística. No Brasil existem diversas leis que estabelecem as diretrizes do direito intelectual, dentre elas temos: Lei 9279/96¹, Lei de Propriedade Industrial, em vigor desde maio de 1997, substituiu o antigo Código da Propriedade Industrial; Lei 9456/97², Lei Cultivares, em vigor desde abril de 1997 institui a proteção para plantas; Lei 9609/98³, Lei de Programa de Computador, em vigor desde fevereiro de 1998; E por fim a lei 9610/98⁴, Lei de Direitos Autorais, em vigor desde junho de 1998, em substituição a Lei 5988/73 (INPI, 2021c).

Tais tópicos são importantes de serem fundamentados pois em geral estão intimamente relacionados às patentes, que são o objeto principal de estudo deste trabalho. Apesar do presente estudo não possuir ênfase nos termos direitos e propriedades, destaca-se que ambos são fundamentais no processo de produção técnica, pois a propriedade intelectual é o meio legal que garante que uma inovação será protegida, e que os responsáveis por ela, serão devidamente recompensados.

2.2 Patentes

Conforme apresentado anteriormente, patente é um dos ramos da proteção intelectual, considera-se patente um documento formal, que é expedido por uma repartição pública, pela qual se conferem e se reconhecem direitos de propriedade e uso exclusivo para uma invenção amplamente descrita (SEBRAE, 2021). Trata-se de um privilégio concedido aos idealizadores sejam eles pessoas física ou jurídica detentores do direito de invenção seja de produtos, processos de fabricação, ou até mesmo aperfeiçoamento de algum produto ou processo já existente. O Inventor fornece à sociedade um novo bem, ou seja, o conhecimento, divulgando as informações técnicas que nortearão para chegar ao objetivo. A sociedade por sua vez, oferece ao inventor exclusividade de exploração, ou seja, produção e comercialização do objeto de sua invenção, por um determinado intervalo de tempo.

De acordo com Mueller e Perucchi (2014) às inovações protegidas por patente tem potencial valor comercial, por tal motivo, existe legislação específica para garantir sua proteção. O valor de um documento de patente é territorial, ou seja, cada país tem a sua própria legislação e seu próprio escritório de exame e concessão de patentes. No Brasil entrou em vigor em 14 de maio de 1996, o Código de Propriedade Intelectual, lei 9279/96, que regula direitos e obrigações relativos à propriedade industrial (BRASIL, 1996). O órgão

¹ Disponível em <http://www.planalto.gov.br/ccivil_03/leis/l9279.htm> acessado em 16 de abril de 2021

² Disponível em <http://www.planalto.gov.br/ccivil_03/leis/l9456.htm> acessado em 16 de abril de 2021

³ Disponível em <http://www.planalto.gov.br/ccivil_03/leis/l9609.htm> acessado em 16 de abril de 2021

⁴ Disponível em <http://www.planalto.gov.br/ccivil_03/leis/l9610.htm> acessado em 16 de abril de 2021

encarregado pela sua aplicação é Instituto Nacional da Propriedade Industrial (INPI), criado em 1970, hoje uma autarquia federal vinculada ao Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC).

Desta forma, são estabelecidas regras específicas para o pedido de depósito de patente, incluindo seu processamento e publicação. Entretanto tais regras divergem em diferentes jurisdições. Em sua maioria, as jurisdições têm um sistema em vigor com um exame em que a matéria técnica reivindicada é examinada verificando se atende a determinadas condições de patenteabilidade, como novidade, atividade inventiva e aplicabilidade industrial. Na jurisdição brasileira, conforme é apontado pela Lei 9279/96, publicada por [Brasil \(1996\)](#), são estabelecidos os critérios de patenteabilidade, evidenciando que é patenteável a toda invenção que apresenta os requisitos de novidade, atividade inventiva e aplicação industrial. Não devem ser classificadas como descobertas, teorias científicas e métodos matemáticos; concepções puramente abstratas; esquemas, planos, princípios ou métodos comerciais, contábeis, financeiros, educativos, publicitários, de sorteio e de fiscalização; obras literárias, arquitetônicas, artísticas e científicas ou qualquer criação estética; programas de computador em si; apresentação de informações; regras de jogo; técnicas e métodos operatórios ou cirúrgicos, bem como métodos terapêuticos ou de diagnóstico, para aplicação no corpo humano ou animal; e todo ou parte de seres vivos naturais. Vale citar também os materiais biológicos encontrados na natureza, ou ainda que dela isolados, inclusive o genoma ou germoplasma de qualquer ser vivo natural e os processos biológicos naturais.

De acordo com o [INPI \(2021c\)](#) no Brasil as patentes podem ser divididas em três tipos, a saber:

- **Patente de Invenção (PI):** São os produtos ou processos que atendam aos requisitos de atividade inventiva, novidade e aplicação industrial. Sua validade é de 20 anos a partir da data do depósito.
- **Patente de Modelo de Utilidade (MU):** Objeto de uso prático, ou parte deste, suscetível de aplicação industrial, que apresente nova forma ou disposição, envolvendo ato inventivo, que resulte em melhoria funcional no seu uso ou em sua fabricação. Sua validade é de 15 anos a partir da data do depósito.
- **Certificado de Adição de Invenção:** Não é bem um tipo, porém um adendo a uma patente existente. Aperfeiçoamento ou desenvolvimento introduzido no objeto da invenção, mesmo que destituído de atividade inventiva, porém ainda dentro do mesmo conceito inventivo. O certificado será acessório à patente e com mesma data final de vigência desta.

Para realizar depósito de uma patente, o depositante tem que elaborar o "Pedido de Depósito". Documento este que consiste na elaboração de um relatório, em que o depositante deve descrever de maneira clara e com riqueza de detalhes o objeto ou o processo a ser patentado. Segundo [INPI \(2020b\)](#), o pedido de patente é composto pelos

seguintes documentos:

- **Relatório Descritivo:** É apresentado um relatório descritivo de maneira suficientemente clara para que um técnico no assunto possa reproduzir e indicar a melhor forma de execução. Destacando o problema no estado da técnica e a solução proposta pela invenção, especificando o setor técnico a qual se destina. Mostrando que de fato a solução apresentada é uma novidade frente às outras soluções existentes, ressaltando a evolução tecnológica e as vantagens em relação ao estado da técnica.
- **Quadro reivindicações:** São apresentadas as particularidades do pedido, apresentando de maneira precisa a matéria objeto da proteção.
- **Resumo:** No resumo é apresentada uma descrição sucinta da matéria exposta no pedido, podendo conter de 50 a 200 palavras, mantendo a clareza sobre o invento. Não excedendo o limite de 20 linhas, sendo vetada a menção ao mérito ou valor da invenção.
- **Desenhos:** Informações complementares, que possam auxiliar na compreensão da invenção, como gráficos, fluxogramas, diagramas, entre outras.

Após a elaboração e entrega do “Pedido de Patente”, caso o mesmo seja deferido e exista a confirmação de pagamento da retribuição correspondente, é expedido a respectiva carta-patente. De acordo com a [Brasil \(1996\)](#) para efetuar a concessão de uma patente é analisado se o pedido está de acordo com os Artigos 10 e 18, em que são apresentados condições na qual uma ideia não pode ser patenteada, a saber:

"Art. 10. Não se considera invenção nem modelo de utilidade: I - descobertas, teorias científicas e métodos matemáticos; II - concepções puramente abstratas; III - esquemas, planos, princípios ou métodos comerciais, contábeis, financeiros, educativos, publicitários, de sorteio e de fiscalização; IV - as obras literárias, arquitetônicas, artísticas e científicas ou qualquer criação estética; V - programas de computador em si; VI - apresentação de informações; VII - regras de jogo; VIII - técnicas e métodos operatórios ou cirúrgicos, bem como métodos terapêuticos ou de diagnóstico, para aplicação no corpo humano ou animal; e IX - o todo ou parte de seres vivos naturais e materiais biológicos encontrados na natureza, ou ainda que dela isolados, inclusive o genoma ou germoplasma de qualquer ser vivo natural e os processos biológicos naturais."

"Art. 18. Não são patenteáveis: I - o que for contrário à moral, aos bons costumes e à segurança, à ordem e à saúde públicas; II - as substâncias, matérias, misturas, elementos ou produtos de qualquer espécie, bem como a modificação de suas propriedades físico-químicas e os respectivos processos de obtenção ou modificação, quando resultantes de transformação do núcleo atômico; e III - o todo ou parte dos seres vivos, exceto os microorganismos transgênicos que atendam aos três requisitos de patenteabilidade - novidade, atividade inventiva e aplicação industrial - previstos no art. 8º e que não sejam mera descoberta. Parágrafo único. Para os fins desta Lei, microorganismos transgênicos são organismos, exceto o todo ou parte de plantas ou de animais, que expressem, mediante intervenção humana direta em sua composição genética, uma característica normalmente não alcançável pela espécie em condições naturais."

Após a análise de patenteabilidade, será feita a análise dos documentos de acordo com as exigências que cada área tecnológica, validando se a documentação é suficientemente detalhada para que um profissional da área consiga reproduzir seu resultado. Estando tudo em conformidade é concedido um documento oficial chamado de carta patente, contendo número, título e natureza, o nome do inventor, prazo de vigência, o relatório descritivo, as reivindicações e os desenhos, bem como os dados relativos à propriedade. A concessão é feita a quem a solicitou, não necessariamente ao inventor, formalizando assim a concessão dos direitos sobre a invenção ou modelo.

As patentes concedidas são classificadas de acordo com sua natureza seguindo o padrão do sistema de Classificação Internacional de Patente (CIP), em inglês *International Patent Classification* (IPC). O CIP é o sistema de classificação de patentes usado no Brasil e por mais de 90 outros países. Conforme Mueller e Perucchi (2014) vários países se mobilizaram para elaborar o sistema de classificação baseado no tratado internacional celebrado em 1954, a Convenção Europeia para a Classificação Internacional de Patentes de Invenções. Patrocinado pela OMPI foi celebrado em 1971 o Acordo de Estrasburgo Relativo à Classificação Internacional de Patentes, que entrou em vigor em 1975 (MUELLER; PERUCCHI, 2014). A versão mais recente, publicada em 2020 do Guia Internacional de patente, está disponível em inglês no site da WIPO ⁵, e em português no site do INPI ⁶.

De acordo com INPI (2021a) a CIP tem como objetivos permitir a busca e recuperação de documentos referentes às patentes; organizá-los para facilitar o acesso às informações tecnológicas; ajudar na investigação do estado da técnica em determinados campos da tecnologia; servir de base para a elaboração de estatísticas sobre propriedade industrial. Vale destacar que uma Patente pode receber mais de uma classificação, quantas forem necessárias, caso não exista uma seção específica que melhor classifique para o invento, ele será inserido na área que for mais apropriado.

A CIP é uma classificação hierárquica, que divide o conhecimento tecnológico em oito grandes áreas, denominadas seções. No nível mais alto, existem as seções, representadas pelas letras de A a H; cada seção é dividida em classes, cada classe em subclasses e cada subclasse em grupos, que podem conter subgrupos. As oito seções, primeiro nível hierárquico, são:

- A – Necessidades humanas;
- B – Operações de processamento, transportes;
- C – Química; Metalurgia;
- D – Têxteis; Papel;
- E – Construções fixas;

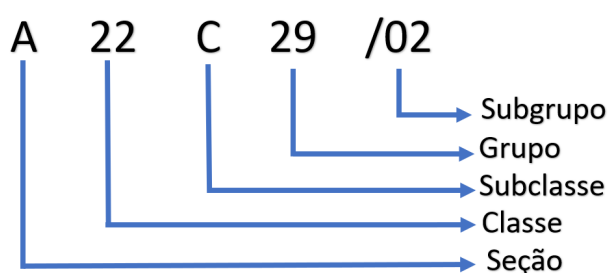
⁵ <<https://www.wipo.int/publications/en/details.jsp?id=4490&plang=EN>>

⁶ <<http://ipc.inpi.gov.br/classifications/ipc/ipcpub/?notion=scheme&version=20200101&symbol=none&menulang=pt&lang=pt&viewmode=f&fipcp=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>>

- F – Engenharia mecânica; Iluminação; Aquecimento; Armas; Explosões;
- G – Física;
- H – Eletricidade.

A classificação CIP recebida por uma patente é composta por uma combinação de letras e números, como a classificação A22C29/02 (WIPO, 2020). A Figura 3 apresenta o detalhamento cada elemento que compõem a classificação:

Figura 3 – Detalhamento de uma classificação de patente



Fonte: Elaboração do autor

- **Seção A**: Necessidades Humanas.
- **Classe 22**: Matança de animais; beneficiamento da carne; processamento de aves domésticas ou peixes.
- **Subclasse C**: Processamento de carnes, aves domésticas ou peixes.
- **Grupo 29**: Processamento de crustáceos, p. ex. ostras, lagostas.
- **Subgrupo 02**: Processamento de camarões, lagostas ou similares.

São também utilizados símbolos como pontos. Quanto mais pontos, mais específico será o detalhe a ser buscado.

O sistema CIP não é o único sistema de classificação de patentes. De acordo com [Carvalho e Santos \(2019\)](#) existe a CPC (*Cooperative Patent Classification*, em português, Classificação Cooperativa de Patentes). A CPC é um sistema de classificação criado pelo Escritório Europeu de Patentes (EPO) em parceria com o Escritório Norte Americano de Patentes (USPTO), baseado na CIP, porém com mais detalhes acerca da classificação. Enquanto a CIP possui em torno de 70 mil grupos, a CPC possui em torno de 200 mil grupos. A CPC é adotada pelo INPI, em conjunto com a CIP desde 2014.

É importante destacar que uma patente depositada no Brasil, só tem validade em território brasileiro, caso deseje proteger a invenção em outro país, é necessário realizar o depósito da patente no país desejado para proteger a mesma, constituindo assim uma família de patentes. De um modo geral uma família de patentes consiste em um conjunto de patentes depositadas em diferentes países, porém com o mesmo propósito, ou seja, proteger uma mesma invenção, o direito de prioridade do primeiro depósito é estendido aos demais nos diferentes países os quais foram depositados. De acordo com o [INPI \(2020a\)](#) família de

patentes, é um conjunto de patentes e/ou pedidos, sejam eles nacionais ou internacionais, que tenham relação direta com a reivindicação de propriedade ou compartilhamento do mesmo depósito internacional.

O conceito de família de patentes foi originado na Convenção da União de Paris (CUP), em 1883, em que foi criado sistema internacional de propriedade industrial, cujo o objetivo era compatibilizar os diversos sistemas jurídicos de diferentes países a respeito de Propriedade Industrial, em que foi estabelecido quatro princípios básicos, dentre eles, a Prioridade Unionista. Este princípio determina que caso existam sucessivos depósitos de uma mesma invenção em distintos países membros da União, dentro do período máximo de 12 meses, contando da data do primeiro depósito, a invenção estará protegida em todos esses países em que houveram os depósitos. Este conjunto de patentes referentes a uma mesma invenção em diferentes países da União é a família de patentes (WIPO, 2021). Várias atualizações foram introduzidas no texto inicial de 1883 através de 7 revisões, em Bruxelas em 14 de Dezembro de 1900, em Washington em 2 de Junho de 1911, na Haia em 6 de Novembro de 1925, em Londres em 2 de Junho de 1934, em Lisboa em 31 de Outubro de 1958, e em Estocolmo em 14 de Julho de 1967, e modificada em 2 de Outubro de 1979. Em 1992 o Brasil aderiu à União (INPI, 2021b). A OMPI disponibilizou em português o documento da Convenção de Paris para a Proteção da Propriedade Industrial, que pode ser acessado pelo endereço eletrônico <https://www.wipo.int/edocs/pubdocs/pt/wipo_pub_201.pdf> (WIPO, 2021).

A WIPO (2007) classifica a família de patentes em três grupos distintos:

- **Família de patentes domésticas:** uma família de patentes que consiste exclusivamente em diferentes publicações processuais para o mesmo aplicativo de origem.
- **Família de patente simples:** uma família de patente relacionada à mesma invenção, cada membro da qual tem como base de seu “direito de prioridade” e exatamente o mesmo aplicativo ou aplicativos de origem.
- **Família estendida de patentes:** uma família de patentes relacionada a uma ou mais invenções, cada membro de que tem, por base de seu direito de prioridade, pelo menos um pedido originário em comum em pelo menos um outro membro da família.

2.3 Repositórios para consulta de documentos de patentes

Conforme foi apresentado na Seção 2.2, para depositar um pedido de patente um dos requisitos é conhecer o estado da técnica, que por definição, é tudo aquilo que é acessível ao público antes da data de depósito do pedido da patente, podendo ser feito por meio de uma descrição textual, fazendo uso de qualquer tipo de dispositivo, sendo nacional ou internacional, inclusive sendo divulgada pelo próprio inventor, ou seja, tudo que já existe em relação ao pedido pleiteado no mundo. Os documentos de patente são

a principal fonte de informação tecnológica na atualidade, representando cerca de 70% do conhecimento tecnológico divulgado exclusivamente por esse tipo de documento. De acordo com a Organização Mundial da Propriedade Intelectual (OMPI), o número de patentes concedidas é de aproximadamente de 1,2 milhões por ano (INPI, 2020b). Todas as informações tecnológicas contidas nesses documentos estão disponíveis em bases de dados de livre acesso por meio da internet. Brandão (2016) destaca que dentre as diversas bases existentes mundialmente, se destaca sete bases de dados que disponibilizam a consulta de documentos de patentes gratuitamente pela internet. A saber:

- **PePI** (Pesquisa em Propriedade Industrial)⁷ também conhecido como Busca Web e, é mantido pelo órgão INPI, para busca de patentes depositadas no Brasil, bem como acesso a pedidos de patentes que não estão em sigilo, patentes já concedidas e patentes arquivadas. Por meio de acesso anônimo em que possibilita visualizar informações básicas de patentes, como título, resumo, autores, entre outras. Para visualizar documentos digitalizados é necessário realizar um cadastro, gratuitamente, para liberar acesso a tais informações;
- **Latipat**⁸ para buscas em documentos públicos de patentes da América Latina, como Brasil, Argentina, México, Cuba, Chile, Uruguai, Peru, entre outros, e ainda o único país europeu, Espanha;
- **Espacenet**⁹ mantido pelo Escritório Europeu de Patentes (EPO), é o mais conhecido e amplamente utilizado para busca de documentos de patentes. É uma base de dados contendo dados de mais de 120 milhões de documentos de patentes do mundo todo.
- **Patentscope**¹⁰ que é a ferramenta de buscas da OMPI, oferece acesso gratuito a documentos de mais de 40 países, como China, Estados Unidos, Austrália, Japão, México e Brasil. Incluindo os depósitos realizado via PCT (Tratado de Cooperação em matéria de Patentes);
- **USPTO** (*United States Patent and Trademark Office*)¹¹ para buscas em patentes depositadas nos Estados Unidos da América (EUA), é a página do escritório de registro de patentes norte americano;
- **SIPO** (*State Intellectual Property Office of China*)¹² para buscas em patentes depositadas na República Popular da China;
- **JPO** (*Japan Patent Office*)¹³ para as buscas das patentes depositadas no território japonês e, inclusive, no Google patents¹⁴.

Entretanto, é possível fazer uso de ferramentas pagas, que permitem fazer buscas

⁷ <<https://busca.inpi.gov.br/pePI/>>

⁸ <<https://lp.espacenet.com>>

⁹ <<https://worldwide.espacenet.com>>

¹⁰ <<https://patentscope.wipo.int/search/en/search.jsf>>

¹¹ <<https://www.uspto.gov>>

¹² <<https://english.cnipa.gov.cn>> ou <<http://english.sipo.gov.cn/>>

¹³ <<https://www.jpo.go.jp>>

¹⁴ <<https://patents.google.com>>

em diversos repositório em um único lugar de maneira transparente para o usuário, como, por exemplo a base *Derwent Innovation Index – DII46*, da *Thomson Reuters Scientific*, acessível pelas instituições participantes do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes/MEC) (BRANDÃO, 2016).

2.4 Cientometria

O catálogo das atividades de pesquisa científica de qualquer campo do conhecimento implica, em uma seleção criteriosa, diante do grande e crescente volume de informações registradas. Os formatos metodológicos de investigação nos diversos campos do conhecimento se destacam no processo de tratamento, organização, recuperação e transferência das informações. Cada ciência tem suas especificidades em relação à instrumentos de mensuração, por exemplo, a Biometria usada na Biologia, a Econometria usada na Economia, bem como a Cientometria que mede a ciência em um contexto geral (SILVA; BIANCHI, 2001a).

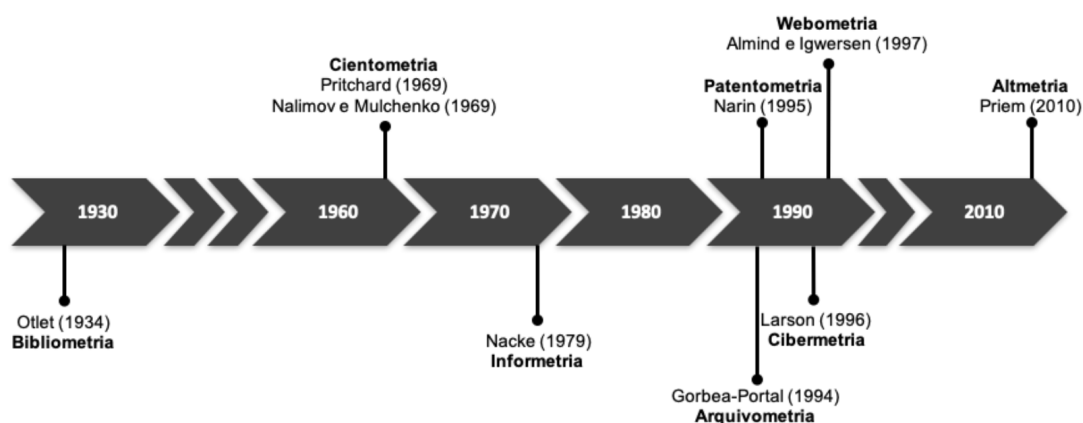
Os Estudos Métricos da Informação são uma das áreas de interesse da Ciência da Informação e tem como foco a identificação e avaliação da informação, seu alcance, influência e impacto. Outras novas técnicas vêm surgindo de maneira proporcional os novos meios de produção e circulação da informação. De acordo com o foco de interesse, da natureza da informação e do objeto de análise, os ramos dos estudos métricos podem ser classificados como Bibliométricos, Informétricos ou Infométricos, Cientométricos, Cibernométricos, Webométricos, Patentométricos e Arquivométricos (CURTY; DELBIANCO, 2020).

Cabrini e Gracio (2011) conceituam os estudos métricos como:

“Os “Estudos Métricos” compreendem o conjunto de estudos relacionados à avaliação da informação produzida, mais especialmente científica, em diferentes suportes, baseados em recursos quantitativos como ferramentas de análise. Fundamentados na sociologia da ciência, na ciência da informação, matemática, estatística e computação, são estudos de natureza teórico-conceitual, quando contribuem para o avanço do conhecimento da própria temática, propondo novos conceitos e indicadores, bem como reflexões e análises relativas à área. São, também, de natureza metodológica, quando se propõem a dar sustentação aos trabalhos de caráter teórico da área onde estão aplicados.”

De acordo com Curty e Delbianco (2020) a bibliometria (1934) é o subcampo pioneiro no ramo dos estudo métricos seguido pela cientometria (1969), a Figura 4 apresenta a uma linha do tempo com os marcos temporais aproximados do surgimento de cada campo dos estudos métricos.

Figura 4 – Linha do tempo da evolução dos estudos métricos.



Fonte: [Curty e Delbianco \(2020\)](#)

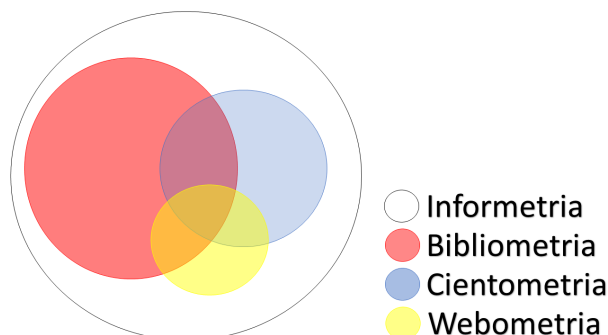
A cientometria consiste na análise quantitativa da ciência, mensuração do progresso científico e tecnológico, baseada em fontes secundárias, sem observação direta do processo de produção de conhecimento e sem avaliação direta dos resultados produzidos, inclui a bibliometria, que é estudo de citações e de produtividade científica ([SILVA; BIANCHI, 2001b](#)). Segundo [Parra, Coutinho e Pessano \(2019 apud VELHO, 1990\)](#) a cientometria consiste em desenvolver indicadores precisos por meio de técnicas matemáticas e estatísticas, para mensurar e analisar atividades científicas, servindo como parâmetros de avaliação. O termo cientometria é a tradução para o idioma Português do termo *scientometrics* em Inglês, onde teve suas primeiras menções, com o trabalho de Dobrov e Karennoi, em uma publicação do *All-Union Institut for Scientific and Thecnical Information* ([VANTI, 2002](#)). Entretanto é possível encontrar comumente na literatura o termo *cienciometria*, porém cientometria é uma tradução mais adequada ao termo *scientometrics*. O termo cientometria, teve sua origem na União Soviética e também na Europa Oriental, especialmente na Hungria por volta de 1969 alcançando maior notoriedade em 1977 com o início da publicação da revista *Scientometrics*, editada originalmente na Hungria e atualmente na Holanda ([VANTI, 2002](#)).

A cientometria aplica técnicas bibliométricas na ciência fazendo uso de métodos matemáticos e estatísticos para investigar as características da pesquisa científica. [Silva e Bianchi \(2001b apud TAUBES, 1993\)](#) destaca os principais indicadores bibliométricos utilizados por pesquisadores que estudam a Cientometria, a saber: total de pessoas com titulações acadêmicas ou científicas, o total patentes registradas por cientistas, o total de artigos científicos publicados, o somatório de cientistas que publicam artigos científicos, o total de referências bibliográficas citadas nos artigos científicos, o total de citações recebidas por artigo científico, o total de auxílios à pesquisa recebidos pelos cientistas e a quantidade de recursos destinados às atividades de pesquisa fomentadas pelas agências.

[Vanti \(2002\)](#) afirma que a cientometria é composta por outras áreas da Ciência da

Informação, como a Informetria, Bibliometria e Webometria, conforme é apresentado na Figura 5.

Figura 5 – Diagrama de inter-relação entre os subcampos



Fonte: Elaboração do autor, adaptado de (VANTI, 2005)

Além da já citada Cientometria, nas próximas subseções são conceituados os subcampos que compõem a informetria de acordo com Vanti (2005).

2.4.1 Informetria

Apesar de ser uma abordagem métrica mais ampla no que se refere à escopo, o termo Informetria foi utilizado pela primeira vez pelo o pesquisador alemão Otto Nacke em 1979 em seu trabalho, traduzido para o idioma espanhol, intitulado “*Informetría: un nuevo nombre para una nueva disciplina. Definición, estado de la ciencia y principios de desarrollo*” (CURTY; DELBIANCO, 2020). No estudo o autor apresenta a seguinte definição para o termo: “a aplicação de métodos matemáticos aos fatos e situações do campo da informação para descrever e analisar seus fenômenos, descobrir suas leis e servir de apoio para decisões” (CURTY; DELBIANCO, 2020 apud NACKE, 1983).

A informetria consiste nos estudos dos aspectos quantitativos da informação seja ela em qualquer formato, por isso é considerada a disciplina mais ampla no campo da ciência da informação pois inclui as áreas da bibliometria, cientometria, dentre outras. Não limitada a somente a informações científicas, mas informação referente a qualquer grupo social (MACIAS-CHAPULA, 1998 apud TAGUE-SUTCLIFFE, 1992).

2.4.2 Bibliometria

Usada pela primeira vez por Pritchard em 1969 em seu trabalho intitulado *Statistical bibliography or bibliometrics?* publicado no *Journal of Documentation*. A Bibliometria consiste nos estudos dos aspectos quantitativos da produção da informação registrada, fazendo uso de padrões e modelos matemáticos para mensurar a produção e disseminação,

gerando previsões e apoio para a tomada de decisão (MACIAS-CHAPULA, 1998 apud TAGUE-SUTCLIFFE, 1992).

2.4.3 Webometria

Embora considerado uma das métricas mais recentes dentro dos estudos da ciência da informação, a webometria, segundo Vanti (2005) utiliza técnicas quantitativas para medir a informação especificamente disponível na Web com base em abordagens bibliométricas e informétricas, consiste basicamente na análise de páginas da Web como se fossem documentos. O conceito de Webometria foi introduzido por Almind e Ingwersen (1997), como uma abordagem a informétrica voltada ao contexto da Web, com a conciliação de estratégias bibliométricas e informétricas aplicada a documentos disponibilizados na WWW (CURTY; DELBIANCO, 2020).

2.4.4 Patentometria

Apesar do termo não estar representado na Figura 5, o mesmo se apresenta relevante por compor recentemente o grupo da Bibliometria. Diversos trabalhos têm utilizado o termo “Patentometria” para descrever os estudos métricos de patentes. A patentometria faz parte do grupo que constitui a bibliometria, pois para a análise de documentos de patentes, são utilizadas adaptações dos indicadores aplicados à produção de outros tipos de documentos (FAGUNDES et al., 2014). Nascimento e Speziali (2020) definem patentometria como a ciência de analisar informações contidas em documentos de patente buscando identificar possíveis relacionamentos e tendências, por meio de estratégias, tais como:

- i. **Inteligência de patentes:** "Uso de informações de patentes para identificar os recursos técnicos de uma ICT (Instituições de Ciência e Tecnologia) ou NIT (Núcleo de Inovação Tecnológica) e o uso dessa inteligência para desenvolver um planejamento estratégico tecnológico"(NASCIMENTO; SPEZIALI, 2020);
- ii. **Mapeamento de patentes:** "Usa dados de patentes para criar representações gráficas que mostram um panorama tecnológico de uma área de conhecimento ou de uma invenção"(NASCIMENTO; SPEZIALI, 2020);
- iii. **Análise de citações de patentes:** "Auxilia na valoração de uma tecnologia e na identificação de parceiros e no licenciamento novas descobertas tecnológicas"(NASCIMENTO; SPEZIALI, 2020 apud TRIPPE, 2015).

O conceito de patentometria foi utilizado em 1995 por Narin (NARIN, 1995), contudo o termo foi estabelecido por Sánchez (1999) onde destaca que a Patentometria é uma abordagem que agrega o conhecimento certificado e a participação em inovação industrial.(CURTY; DELBIANCO, 2020). Já Aristodemou e Tietze (2018) definem patentometria como a ciência de analisar um grande volume de dados sobre propriedade intelectual,

visando descobrir relacionamentos, tendências e padrões. Fazendo uso de matemática, estatísticas, programação computacional e pesquisa operacional, para obter informações que possam auxiliar nas tomadas de decisão.

O conhecimento obtido com a patentometria pode ser aplicado a quase qualquer organização que se dedica à avaliação de tecnologia e seu impacto na sociedade. Governo e agências, bem como empresas privadas podem obter uma perspectiva valiosa sobre um determinado tipo de tecnologia ou desenvolvimento. (TRIPPE, 2015)

Além do exposto anteriormente Macias-Chapula (1998 apud TAGUE-SUTCLIFFE, 1992) destaca que tanto a bibliometria quanto a cienciometria são aplicadas em diversas áreas como, ciências sociais, documentação e biblioteconomia, política científica e indústria da informação entre outros. Fazendo uso de indicadores e métricas tais como:

- Aspectos estatísticos da linguagem e frequência de citação de frases, tanto em textos (linguagem natural), como em índices impressos e em formato eletrônico;
- Características da relação autor-produtividade medidas por meio do número de artigos ou outros meios; grau de colaboração;
- Características das publicações, sobretudo a distribuição em revistas de artigos relativos a uma disciplina;
- Análise de citação: distribuição entre autores, artigos, instituições, revistas, países; uso em avaliação; mapa de disciplinas baseado na co-citação;
- Uso da informação registrada: circulação em bibliotecas e uso de livros e revistas da própria instituição; uso de bases de dados;
- Obsolescência da literatura, avaliada pelo uso e pela citação;
- Crescimento de literaturas especializadas, bases de dados, bibliotecas; crescimento simultâneo de novos conceitos;
- Definição e medida da informação;
- Tipos e características dos níveis de desempenho da recuperação.

Dentre os diversos instrumentos de mensuração aqui apresentados, a patentometria recebe destaque no contexto deste trabalho, por aplicar métricas em informações oriundas de documentos de patentes, permitindo definir estratégias mais precisas no monitoramento e controle das atividades envolvidas na transferência de tecnologia.

3 Trabalhos Relacionados

Este capítulo apresenta a síntese de um conjunto de trabalhos referente aos assuntos permeados por esta Dissertação. Nas próximas seções, são apresentados: os principais trabalhos relacionados a temática patentes como fonte de dados de análise ([Seção 3.1](#)) e diferenciais propostos ([Seção 3.2](#)).

3.1 Patentes como fonte de dados de pesquisa

Estudos baseados nas informações obtidas através de dados de patentes, possuem diversas abordagens e apresentam caráter quantitativo, tecnológico, gestão da inovação, social, dentre outros. Por se tratar de um tema com uma ampla variedade de trabalhos publicados, um pequeno grupo foi selecionado, em que a principal temática é a patentometria. Os critérios definidos para a seleção desses estudos foram: publicações que estejam disponíveis na web, nos idiomas português ou inglês, publicados nos últimos 10 anos e que tenham como objetivo análise de qualquer natureza, de documentos de patentes. A pesquisa foi realizada no ano de 2021. Foram utilizados os seguintes repositórios: Arca (Fiocruz), Biblioteca Digital de Teses e Dissertações, Google Acadêmico, Microsoft Academic, Portal de Periódicos CAPES/MEC e SciELO.

O trabalho de [França \(2007\)](#) apresenta as patentes como fonte de informação tecnológica, em que é apresentado o contexto histórico da propriedade industrial, abordando conceitos importantes como marcas e patentes dando maior ênfase neste último. Apresenta um panorama geral sobre o documento de patente e suas particularidades, como o sistema de classificação e objeto que podem ser patenteados. Destacando a importância de considerar os dados contidos em documentos de patentes como fonte de informação primária a respeito do desenvolvimento tecnológico. Em seu trabalho o autor aponta a patente como uma rica fonte de informação tecnológica.

[Quoniam, Kniess e Mazieri \(2014\)](#) abordam em seu trabalho as dimensões da propriedade intelectual, conceituando e destacando patente como rica fonte de dados para pesquisas no campo das Ciências Humanas e Sociais, particularmente em Ciência da Informação e Comunicação, bem como a importância de utilizá-las na transferência de tecnologia entre países e organizações, uma vez que as informações contidas em documentos de patentes, em raros casos, são reaplicadas em outras fontes de informação. Em seu trabalho os autores não realizam de fato uma análise específica dados de patente, entretanto enfatizam a importância em utilizar as informações contidas em documentos de patentes e apresentam diversas estratégias de análise dos dados, apresentando direcionamentos para pesquisadores interessados no assunto.

Ribeiro e Garcia (2014) apresentam em seu trabalho um panorama da patentometria em periódicos internacionais da ciência da informação. Iniciam seu trabalho citando a bibliometria e a cientometria como as metodologias mais relevantes nos estudos da informetria, por abranger grande área de estudo. Complementando, destacam que a patentometria como uma ferramenta de inteligência competitiva em ciência e inovação, fornecendo indicadores como base substancial para a tomada de decisão. Uma vez que a patentometria é caracterizada por estabelecer indicadores que buscam identificar atividades de inovação e tecnologias, através das informações extraídas de documentos de patentes. Para estabelecer um panorama geral os autores realizaram um estudo no âmbito internacional, verificando o estado da arte das pesquisas em patentometria na área da ciência da informação. Por meio de pesquisas em periódicos com conceitos Qualis A1 a B2 compreendido entre os anos de 1998 a 2013, por termos referente a patentometria, tais como: *patent*, *industrial property*, *innovation patent* e *analysis patent*. Concluem o trabalho, afirmando que a análise de patentes constituem uma ferramenta importante de competitividade para pesquisas, que permite identificar e converter o conhecimento científico em conhecimento tecnológico.

Já Cattivelli (2016) apresenta em sua dissertação de mestrado, uma análise das patentes concedidas, das Universidades Públicas Brasileiras com o objetivo de identificar fatores e/ou características dessas que incentivam o depósito, e por consequência, a concessão de patentes. A coleta dos dados de patentes para análise foi realizada na base de dados do INPI, no qual foi possível coletar dados de patentes de 29 universidades, do total de 106 universidades pesquisadas, 65,2% corresponde a universidades da região sudeste do país. Com base na CIP (Classificação Internacional de Patentes) das patentes coletadas, foi possível observar que a área do conhecimento com maior quantidade de patentes concedidas são “Química e Metalurgia”, em que os pesquisadores que mais patenteiam, foram professores com formação em Química ou Engenharia Mecânica. A autora ainda destaca que existe uma forte presença do setor público, pois em sua maior parte, a cotitularidade é constituída por instituições de fomento e de pesquisa. Concluindo os autores destacam a importância dos Núcleos de Inovação Tecnológica (NITs) nas instituições públicas como atores que contribuem para implantar a cultura do patenteamento no país, que por consequência, contribuem de forma indispensável com a evolução tecnológica do país.

Mendes e Melo (2017) propõem em seu trabalho uma metodologia para avaliar a tecnologia, por meio de técnicas de cientometria e patentometria, servindo assim como ferramenta de apoio a tomada de decisão para as organizações no tangente a introdução de novas tecnologias. Por meio de um modelo composto por 9 etapas, (1) Definição de termos, em que serão definidas as palavras ou expressões, a quais deseja obter maiores informações ; (2) Validação dos termos por especialistas, em que os termos definidos são revisados e aprovados por especialistas na área afim; (3) Definição das bases a serem utilizadas, por meio de critérios previamente estabelecidos; (4) busca pelos termos nas bases definidas; (5) Extrair os registros encontrados, para formar uma base de dados local,

para viabilizar as análises; (6) Realizar a limpeza do dados, correção de possíveis erros de digitação, duplicidades e padronizações; (7) Gerar listas, gráficos e tabelas; (8) Analisar as informações geradas; (9) Conclusões e resultados. Os autores ainda argumentam que a proposta permite fornecer aos tomadores de decisão bases de dados científicas que servirão como fundamentação para tomada de decisões, como indícios de ofertas de conhecimento para o desenvolvimento da tecnologia, bem como sua demanda.

Os autores [Scartassini et al. \(2018\)](#), apresentam uma análise patentométrica sobre as patentes brasileiras com extensão de depósito via Tratado de Cooperação de Patentes (PCT). Um trabalho patentométrico quantitativo de natureza descritiva. Os autores coletaram os dados em março de 2017, dividido a metodologia em três etapas distintas, a primeira recuperando todas as patentes depositadas no Brasil indexadas na base Derwent Innovation Index (DII) entre o período de 2004 a 2016; na segunda etapa dentro do conjunto coletado, foi extraído as patentes com registro prioritário no Brasil; e por fim na terceira etapa foi extraído desse conjunto as patentes com depósito prioritário no Brasil e que realizaram pedido de proteção internacional na via PCT. Totalizando 190 patentes consideradas em seu estudo. No processo de análise foram levantados os maiores depositantes, análise temporal dos depósitos de patentes, o quantitativo das classificações recebidas pelas patentes. Concluindo que o depósito via PCT ainda não é uma prática comum para os depositantes brasileiros.

[Menezes \(2019\)](#) em sua dissertação de mestrado ao Programa de Pós-graduação Profissional em Gestão, Pesquisa e Desenvolvimento na Indústria Farmacêutica, do Instituto de Tecnologia em Fármacos na Fundação Oswaldo Cruz, realizou um estudo patentométrico em que se almejou-se mapear o cenário tecnológico do desenvolvimento de novas alternativas terapêuticas para o tratamento da tuberculose. Os dados para o estudo foram obtidos por meio de consultas no repositório Derwent World Patents Index através do Portal de Periódicos da Capes em setembro de 2018, dentro do período de 2008 a 2018, estabelecendo como critério de seleção as patentes cujo título ou resumo apresentem termos relacionados a tuberculose e possuam as classificações A61K-039/04 ou A61P-031/06. Obtendo um conjunto de 517 patentes consideradas no estudo. Mediante a análise dos indicadores quantitativos, o autor destaca um declínio no número de depósitos por ano, entretanto, foi possível observar a presença tanto de instituições públicas quanto privadas participando ativamente nos processos de depósito de patentes e estudos clínicos.

Já os autores [Junior e Kühl \(2020\)](#), realizaram um estudo bibliométrico por meio de pesquisa nos repositórios da SciELO e Capes no período de 2005 a 2019, complementado por um estudo patentométrico por meio da busca na base de patentes do Instituto Nacional da Propriedade Industrial (INPI) no período de 2005 a 2018. Tendo como principal objetivo mapear as inovações tecnológicas aplicáveis nas Ciências Contábeis. Na etapa metodológica foram estabelecidos os critérios de busca e seleção dos dados nos repositórios da

SciELO e Capes e INPI, considerando os termos “Contabilidade”, “Contabilidade *and* Inovação”, “Contabilidade *and* Tecnologia” e “Contabilidade *and* Tecnologia da Informação”, com caráter quantitativo. Destacando que as patentes identificadas por meio da busca realizada, obteve-se 10 resultados, em que todas as 10 patentes referem-se a Sistemas de Informação aplicáveis à contabilidade, tais como controle financeiro, controle de despesas e gastos, emissão de notas fiscais e o ensino de contabilidade. Os autores concluem considerando que os resultados do estudo bibliográfico estão em concordância com os resultados bibliométrico e patentométrico obtidos, destacando que, em maior parte, os dados levantados nos repositórios considerados no estudo, tratam-se de sistemas informatizados, associado a era da informatização e dos processos computadorizados nas Ciências Contábeis

Em [Silva e Nascimento \(2020\)](#) os autores apontam a patentometria como ferramenta indispensável para a proposição de tecnologias para indústria química, pois a prospecção baseada em métodos quantitativos e qualitativos, são de fato importantes para viabilizar a conceitualização de propostas plausíveis no estado da arte. Em sequência exploram alguns indicadores apresentando seu processo de construção de sua relevância no panorama tecnológico. Os autores concluem mencionando que a patentometria é uma área pouco explorada no Brasil apesar de seu valor competitivo.

Os autores [Souza et al. \(2021\)](#) abordam o mapeamento do depósito de patentes, marcas, desenhos industriais e registros de programas de computador registrados no INPI por discentes do Programa de Pós-Graduação em Ciência da Propriedade Intelectual da Universidade Federal de Sergipe, com o objetivo de avaliar sua representatividade. Os dados foram obtidos no repositório do Instituto Nacional da Propriedade Industrial (INPI) no período de 2014 a 2020. Mediante a análise dos indicadores quantitativos levantados os autores evidenciaram a existência de registro de propriedade intelectual em todas as categorias pesquisadas caracterizando assim a representatividade dos discentes.

3.2 Diferenciais propostos

Apesar de existirem diversos trabalhos utilizando patentes como fonte de dados para pesquisa, ainda existem lacunas a serem exploradas, novas abordagens e perspectivas de análise. A proposta deste trabalho é, primeiramente, propor uma estratégia de coleta e agregação de dados referente à produção técnica brasileira, o fazendo, através da coleta de dados de patentes depositadas no Brasil e disponibilizadas na Espacenet. Para abranger o maior número de patentes no estudo, o período de coleta terá como data inicial janeiro de 1900 objetivando pegar a patente mais antiga registrada e a data final setembro de 2021 conciliando com a data limite de coleta de dados para este estudo. Sendo este conjunto de dados composto por, identificação da patente, depositantes, inventores, classificações, famílias e propriedade de reivindicações. Estabelecendo assim um repositório de dados local

para pesquisa ampla sobre a produção técnica brasileira, que não é facilmente recuperada por pesquisas nos atuais repositórios disponíveis. Posteriormente, com todos os dados recuperados e estruturados, constituir indicadores para avaliar e mensurar a produção técnica brasileira, a fim de contribuir para a compreensão do cenário tecnológico do país.

4 Metodologia

Este trabalho trata-se de um estudo de caso, ou seja, um estudo de natureza empírica que investiga um determinado fenômeno, dentro de um contexto em que ainda há lacunas na literatura, conforme afirmam (SERRANO; JUNIOR, 2014).

Nas próximas seções, são apresentados: o processo de aquisição dos dados (Seção 4.1), e o processo de seleção, tratamento e visualização dos dados (Seção 4.2).

4.1 Aquisição dos dados

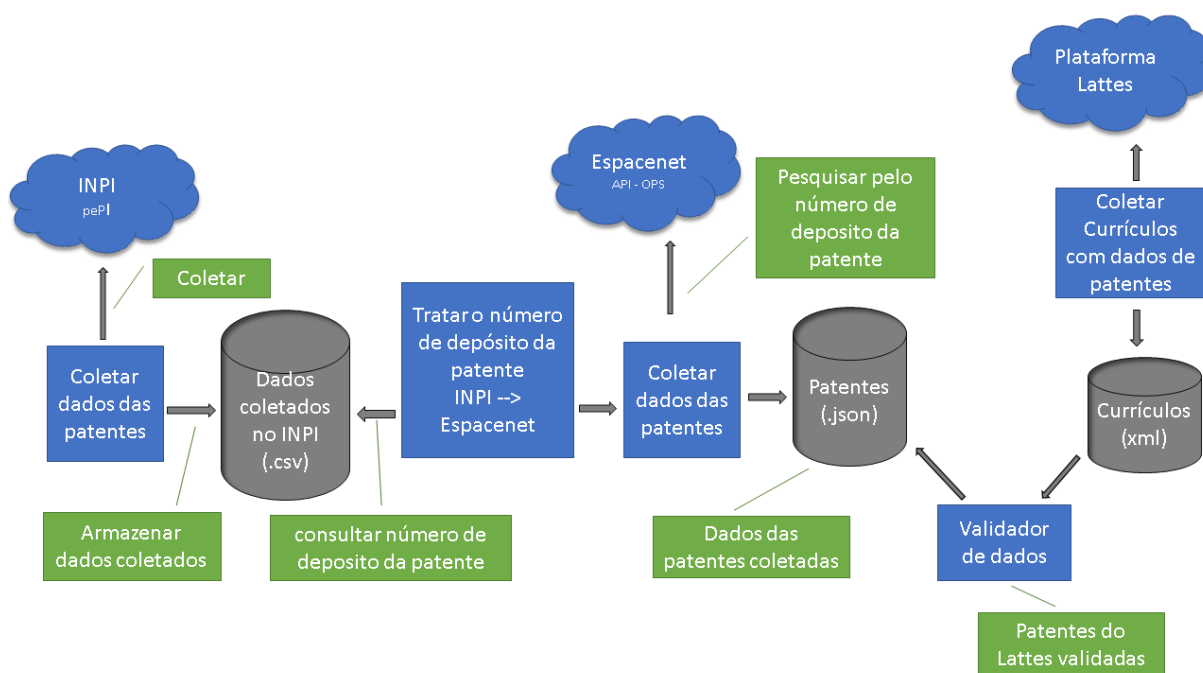
Para a construção do repositório local de dados que irá fundamentar esta dissertação, a priori foi realizada a coleta de informações referente a documentos de patentes depositadas no INPI no período de janeiro de 1900 à dezembro de 2020, este período foi definido objetivando compreender o maior número de patentes possível. Assim, de posse dos dados coletados e objetivando agregar ainda mais informações sobre o conjunto de patentes, foi realizada a consulta destes dados patentários no repositório da Espacenet, utilizando para isso o número de depósito das patentes coletadas no INPI. Esse conjunto de dados extraídos no INPI e validado na Espacenet é o conjunto de dados analisado neste estudo, estudo este de grande relevância, principalmente pelas técnicas implementadas, bem como, amplitude e consistência dos dados coletados.

Após a coleta dos dados das patentes e construção do repositório de dados local, no intuito de se propor uma estratégia de validação dos dados coletados com outras fontes de dados, utilizando o *framework* LattesDataXplorer desenvolvido por Dias (2016), realizou-se a coleta de dados contidos nos currículos cadastrados na Plataforma Lattes do CNPq, que possuem informações de registro e/ou participação no depósito de patentes. De posse desses dados, algoritmos foram desenvolvidos para equalizar toda a base de dados obtida dos currículos, e também, para remover informações de usuários sem produção técnica registrada ou sem número válido de registros das patentes. O objetivo desta estratégia é verificar a consistência dos dados de patentes registrados nos currículos cadastrados na Plataforma Lattes, propor uma estratégia de certificação de dados de patentes, bem como, viabilizar análises que possam considerar diversas informações dos proponentes que não estão nos registros das patentes mas, inseridas nos seus currículos, como por exemplo, dados de formação acadêmica, áreas de atuação e produção científica.

Inicialmente, o processo de aquisição dos dados de patentes foi dividido em três etapas, (1) inicialmente foi realizada a coleta dos dados no INPI e tratamento dos números de depósito de patentes, e posteriormente, (2) a validação e coleta de dados das patentes

extraídas no repositório da Espacenet, e por fim, (3) a coleta das patentes registradas nos currículos da Plataforma Lattes e sua validação com o conjunto coletado na Espacenet. A Figura 6 apresenta o esquema elaborado para o processo de coleta dos dados.

Figura 6 – Visão geral da coleta de dados



Fonte: Elaboração do autor

4.1.1 Coleta de dados no INPI

Para coletar os dados de patentes no INPI foi utilizada a ferramenta de pesquisa de patentes pePI (Pesquisa em Propriedade Industrial), também conhecida como BuscaWeb mantido pelo INPI, em que é possível realizar a consulta de documentos de patentes informando login e senha ou por acesso anônimo. O que difere as duas formas de identificação é que optando por informar o login e senha irá permitir acesso a mais serviços, como por exemplo, a disponibilização de documentos no formato PDF, porém para atingir o objetivo deste trabalho o acesso anônimo é o suficiente, por isso, o mesmo foi utilizado. Após acessar a ferramenta pelo método identificação anônima, foi selecionada a opção de “Patentes”, em que é apresentada uma página contendo opções de pesquisa. Logo, foi selecionada a opção “pesquisa avançada” para exibir mais critérios de pesquisa.

Ao informar no campo de pesquisa “(22) Data de Depósito” à data inicial “01/01/1900” e data final “31/12/2020” e selecionar a opção “pesquisar”, o sistema retorna uma página com a listagem de 862.726 patentes distribuídas em 8.627 páginas exibindo 100 registros por página. De forma ilustrativa, para exemplificar melhor a dimensão do volume dos dados, se executarmos a coleta manual de todas as patentes, isso demandaria um esforço humano

muito grande, considerando de forma hipotética que levasse aproximadamente 10 minutos para acessar os detalhes de cada patente e armazenar as informações de interesse em um processador de planilhas eletrônicas, o processo manual de coleta levaria em torno de 143.434 horas, dedicando 8 horas por dia, levaria cerca de 17.942 dias. Para mitigar este tempo e otimizar a coleta dos dados, foi proposto um algoritmo para viabilizar um processo computacional, composto por 5 etapas dispostas a seguir:

1. Realizar o login anônimo para recuperar as credenciais necessárias para realizar a pesquisa;
2. Acessar a pesquisa avançada, informando as credenciais obtidas na etapa anterior;
3. Na tela de pesquisa avançada, informar no campo “(22) Data de Depósito” a data inicial 01/01/1900 e a data final 31/12/2020 e disparar o evento de pesquisa;
4. Percorrer a toda a listagem de patentes apresentada na página de resultado:
 - a) Para cada patente, acessar a página de detalhes;
 - b) Analisar o conteúdo HTML (*HyperText Markup Language*) da página de detalhe e recuperar as informações: “Número do pedido”, “Data de depósito”, “Data de publicação”, “Título Depositante”, “Inventor” e “Classificação ICP”.
 - c) Armazenar as informações recuperadas em um arquivo CSV (*Comma-separated-values*);
 - d) Voltar a listagem de patentes;
5. Repetir a etapa 4 para todas as páginas de resultados da pesquisa.

Por meio de técnicas de *web scraping* e *web crawler*, toda essa estratégia foi codificada utilizando a linguagem de programação Python. Zhao (2017) define *web scraping*, que em tradução para português, raspagem web, como uma técnica de extração de dados em páginas disponíveis na WWW (*World Wide Web*) e armazená-los em arquivo ou em banco de dados para posterior análise, podendo ser realizado manualmente por um usuário ou automaticamente por um robô, um *web crawler*, um algoritmo usado para encontrar, ler e indexar páginas de um site. Mitchell (2019) explica que *web scraping* engloba um grande conjunto de técnicas de programação e diversas tecnologias, tais como, a análise de dados, *parsing* de idiomas naturais e segurança da informação, entre outras.

Durante os testes do algoritmo desenvolvido foi possível identificar uma limitação na abordagem estabelecida, devido ao grande volume de dados, por motivos de segurança da plataforma, as credenciais expiram depois de um determinado período. Para contornar essa limitação, foi utilizado períodos mensais para o filtro “Data de Depósito”, conseguindo assim, em apenas 0,5% do tempo, se comparado ao hipotético processo manual para coletar as informações das patentes. Armazenando os dados em arquivos CSV, um arquivo para cada ano, a coleta foi executada entre os meses de abril a dezembro de 2020. O pseudocódigo apresentado pelo Algoritmo 1 esboça o script desenvolvido.

Algoritmo 1: COLETA DE DADOS NO INPI

```
início
  Inicializa a variável anos com a lista anos dentro do intervalo 1900 à 2020;
  para cada  $ano_i \in anos$  faça
    Cria um arquivo CSV com o nome do ano corrente no loop;
     $meses = 12$ ;
    para cada  $mes_i \in meses$  faça
      Inicializa a variável dtInicio com primeiro dia do mês e ano corrente;
      Inicializa a variável dtFim com último dia do mês e ano corrente;
      Inicializa credenciais com o login anônimo, na busca INPI;
      Pesquisar patentes por data de depósito, informando credenciais,
        dtInicio e dtFim;
      Inicializa a variável paginas com todas as páginas de resultado;
      para cada  $pagina_i \in paginas$  faça
        para cada  $patente_i \in listaDePatentes$  faça
          Acessar página de detalhes da patente;
          Analisar o HTML e recuperar os dados de interesse;
          Armazenar os dados coletados no arquivo CSV;
        fim
      fim
    fim
  fim
fim
```

A [Tabela 1](#) apresenta a sumarização dos dados coletados no INPI, apresentando a quantidade coletada, e tempo gasto para a coleta.

Tabela 1 – Sumarização dos dados coletados no INPI

Ano de depósito	Quantidade patentes coletada	Tempo médio de coleta (segundos)	Tempo total de coleta (horas)
1900	10.117	3,00	8,43
1919	228	3,00	0,19
1974	835	3,00	0,69
1975	4.691	3,00	3,90
1976	7.625	3,00	6,35
1977	9.997	3,00	8,33
1978	9.442	3,00	7,87
1979	9.330	3,00	7,77
1980	10.968	3,00	9,14
1981	9.364	3,00	7,80
1982	10.499	3,00	8,75
1983	10.169	3,00	8,47
1984	9.654	3,00	8,04
1985	9.836	3,00	8,20
1986	10.295	3,00	8,58
1987	11.623	3,00	9,68
1988	12.218	3,00	10,18
1989	12.424	3,00	10,35
1990	12.755	3,00	10,63
1991	11.824	3,00	9,85
1992	10.593	3,00	8,83
1993	11.883	3,00	9,90
1994	11.646	3,00	9,70
1995	14.979	3,00	12,48
1996	17.326	3,00	14,44
1997	19.476	3,00	16,23
1998	19.450	3,00	16,21
1999	24.494	3,00	20,41
2000	21.138	3,00	17,61
2001	21.230	3,00	17,69
2002	20.039	3,00	16,70
2003	21.970	3,00	18,31
2004	23.474	3,00	19,56
2005	24.680	3,00	20,57
2006	26.001	3,00	21,67
2007	26.076	3,00	21,73
2008	27.459	3,00	22,88
2009	29.877	3,00	24,90
2010	32.448	3,00	27,04
2011	33.935	3,00	28,28
2012	33.812	3,00	28,18
2013	33.816	3,00	28,18
2014	31.883	3,00	26,57
2015	29.313	3,00	24,43
2016	28.452	4,50	35,56
2017	27.271	4,00	30,30
2018	25.451	4,35	30,75
2019	21.439	3,25	19,35
2020	9.221	2,18	5,58
Total	862.726	3,07	747,24

Conforme a [Tabela 1](#), é possível observar que anterior ao ano de 1974 há uma irregularidade no período de depósito de patentes, uma hipótese para tal, consiste que de acordo com [Brasil \(2021\)](#) o INPI foi criado pela Lei no 5.648, de 11 de dezembro de 1970, em que desde sua criação se estabelece uma regularidade no período de depósito de patentes. Destacando que o ano de 2020 contém somente as patentes publicadas até o mês de dezembro de 2020, período final para coleta dos dados no INPI.

4.1.2 Coleta de dados na Espacenet

Com a coleta de dados no INPI concluída, a próxima etapa deste trabalho foi identificar cada patente coletada no INPI, na Espacenet, e posteriormente coletar os dados que são disponibilizados neste repositório, dados estes que visam complementar e validar a coleta inicial, tendo em vista a quantidade de informações adicionais que estão presentes na Espacenet como informações sobre família das patentes e citações. Somente o conjunto de patentes que forem identificadas na Espacenet será considerado nas análises do presente estudo, devida sua completude e consistência dos dados.

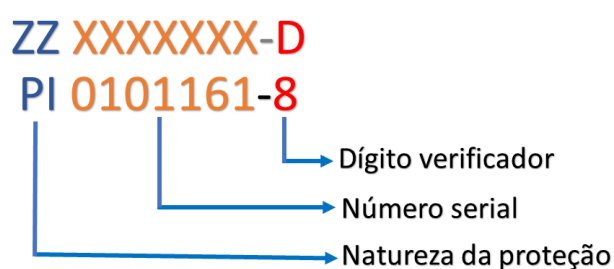
A Espacenet é um serviço de pesquisa amplo e consistente de cobertura mundial que oferece acesso gratuito a informações sobre invenções e desenvolvimentos técnicos desde os anos de 1782 até a atualidade. Sua interface de consulta é simples e intuitiva, tornando-a acessível mesmo para usuários inexperientes, contendo atualmente dados de mais de 120 milhões de documentos de patentes de todo o mundo ([ESPECENET, 2021](#)). A Plataforma oferece um recurso de pesquisa em que é possível informar o termo desejado e o mesmo é pesquisado em diversos campos da patente, podendo informar até 10 termos separados por espaço. Entretanto, o serviço foi projetado para ser usado por seres humanos, não permitindo realizar consultas automáticas ou recuperação em lotes, quando isso é necessário é recomendado o uso do OPS (*Open Patent Services*). OPS é um serviço da web que fornece acesso aos dados armazenados no banco de dados do EPO (*European Patent Office*) por meio de serviços web utilizando a arquitetura *RESTful*, fazendo uso dos padrões XML (*eXtensible Markup Language*) e JSON (*JavaScript Object Notation*) para formatar os dados de resposta às requisições, conforme a parametrização. Esta possibilidade consequentemente, torna-se viável o desenvolvimento de aplicativos e robôs de extração automática para baixar grandes volumes de dados.

A recuperação dos dados referente a cada patente é realizada utilizando os serviços disponíveis na OPS, utilizando o número de pedido de depósito da patente como critério de seleção. O número do pedido é importante para identificação da patente tanto no INPI quanto na Espacenet, pois cada patente possui seu próprio número único de depósito. A composição do número de pedido de depósito das patentes no INPI tem dois formatos distintos, o primeiro utilizado para patentes mais antigas e atualmente é adotado um segundo formato. Desde 02 de janeiro de 2012 para os novos pedidos de patente (de invenção e

modelo de utilidade), desenho industrial e indicação geográfica é atribuído o novo formato (UECE, 2021).

Já o formato atribuído às patentes depositadas até de 31/12/2011 é composto pelo seguinte formato **ZZ XXXXXXX-D**, em que **ZZ** é referente a natureza da proteção, **XXXXXXX** um número serial anual composto por 7 dígitos, e por fim, **D** que é o dígito verificador. A Figura 7 apresenta graficamente a composição do formato.

Figura 7 – Formato de numeração antiga



Fonte: Elaboração do autor

Já o novo formato estabelecido visa atender à política de integração internacional do INPI atendendo os padrões sugeridos internacionalmente pela OMPI. Esse novo formato possui a seguinte estrutura **BR ZZ AAAA XXXXXX D CP**, em que **BR** é a identificação do país, **ZZ** é a natureza da proteção (Tabela 3), **AAAA** ano de entrada no INPI, **XXXXXX** numeração que corresponde a ordem de depósito dos pedidos composto por 6 dígitos, **D** o dígito verificador e por fim **CP** que corresponde ao código de publicação, o status legal do pedido junto ao INPI (Tabela 4). A Figura 8 apresenta graficamente a composição do formato atual de numeração de pedidos de depósito de patentes.

Tabela 3 – Natureza da Proteção

Código	Tipo de Patente	Descrição
10	Patentes de Invenção	Pedido de patente de invenção depositado no INPI e via CUP (antigo PI)
11	Patentes de Invenção	Pedido de patente de invenção depositado via PCT
12	Patentes de Invenção	Pedidos divididos (antigo PI)
13	Patentes de Invenção	Certificado de adição (antigo C1, C2, etc)
14-19	Patentes de Invenção	Para atender necessidades da DIRPA
20	Modelo de Utilidade	Pedidos depositados por nacionais e via CUP (antigo MU)
21	Modelo de Utilidade	pedidos depositados via PCT (antigo MU PCT)
22	Modelo de Utilidade	Pedidos divididos (antigo MU)
23-29	Modelo de Utilidade	Para atender necessidades da DIRPA

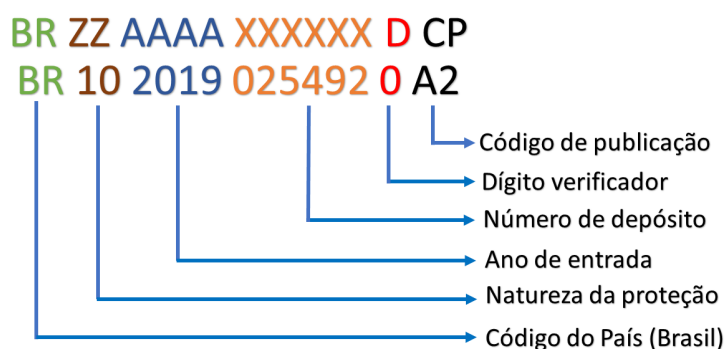
Fonte: adaptado de SAESP/DIRPA/INPI (2015).

Tabela 4 – Exemplos de códigos de publicação

Código	Descrição
A1	Publicação do pedido de patente
A2	Publicação do pedido sem o relatório de busca
A3	Publicação do pedido com o relatório de busca
B1	Publicação da patente concedida
B2	Republicação da patente, por estar ilegível

Fonte: adaptado de [SAESP/DIRPA/INPI \(2015\)](#).

Figura 8 – Formato atual de numeração



Fonte: Elaboração do autor

Após uma série de experimentos foi possível definir uma estratégia para tratar os números de pedido de depósito coletados no INPI, para consequentemente identificar a patente correspondente na Espacenet. Isso, objetivando, principalmente, a identificação das patentes mais antigas com o formato internacional atual adotado pela Espacenet. As regras definidas foram as seguintes:

1. Regra **R001**: Caso o número inicie com “BR”, será removido os espaços entre os caracteres e desconsiderados os 3 últimos caracteres. Exemplo:

BR 10 2019 025492 0 A2 → BR102019025492

2. Regra **R002**: Caso o número não inicie com “BR”, mas inicie com alguns dos seguintes prefixos: "C1", "DI", "MI", "MU" ou "PI", será adotada uma estratégia de força bruta:
 - a) Regra **R002-01**: A primeira tentativa é incluir o prefixo “BR”, desconsiderar todos os caracteres após o separador “-” e remover todos os espaços entre os caracteres. Exemplo:

PI 0000061-2 B1 → BRPI0000061.

Caso não apresente sucesso, iniciar o passo descrito na letra “b”.

- b) Regra **R002-02**: A segunda tentativa consiste em substituir os dois primeiros caracteres por “BR”, desconsiderar todos os caracteres após o separador “-” e remover todos os espaços entre os caracteres. Exemplo:

PI0300001-0 A2 → BR0300001.

Caso não apresente sucesso, iniciar o passo descrito na letra “c”.

- c) Regra **R002-03**: A terceira tentativa consiste em executar os mesmos passos definidos na regra R002-02, incluindo o caractere que estiver dentro do intervalo de “A-Z”, que houver após o separador. Exemplo:

PI0300001-0 A2 → BR0300001A.

Caso não apresente sucesso, iniciar o passo descrito na letra “d”.

- d) Regra **R002-04**: A quarta tentativa, consiste em executar os mesmos passos definidos na regra R002-03, incluindo todos os caracteres que existirem após o caractere dentro do intervalo de “A-Z”, que existir após o separador. Exemplo:

PI0300001-0 A2 → BR0300001A2.

Neste processo é considerando sucesso somente quando é localizado uma única patente na Espacenet. Não localizar a patente ou localizar mais de uma patente, será considerado como falha. Tomando como base o conjunto de regras, foi desenvolvido utilizando a linguagem de programação Python, um algoritmo que percorre todas as patentes coletadas no INPI, e aplica todas as regras definidas armazenando os resultados em arquivos CSV, um arquivo para cada ano de depósito, com a seguinte estrutura de campos:

- Número INPI: Número do pedido de depósito de patente da forma em que foi coletado no INPI;
- Número original: Número do pedido de depósito de patente da forma em que foi coletado no INPI, porém sem todos os espaços;
- Regra 001: Resultado do processamento da regra R001;
- Regra 002-01: Resultado do processamento da regra R002-01;
- Regra 002-02: Resultado do processamento da regra R002-02;
- Regra 002-03: Resultado do processamento da regra R002-03;
- Regra 002-04: Resultado do processamento da regra R002-04;
- Nome do arquivo: Nome que será atribuído aos dados coletados das patentes quando os mesmos forem localizados na Espacenet.

Após o tratamento dos números de pedido de depósito das patentes, foi desenvolvido um novo algoritmo utilizando também a linguagem de programação Python, que percorre

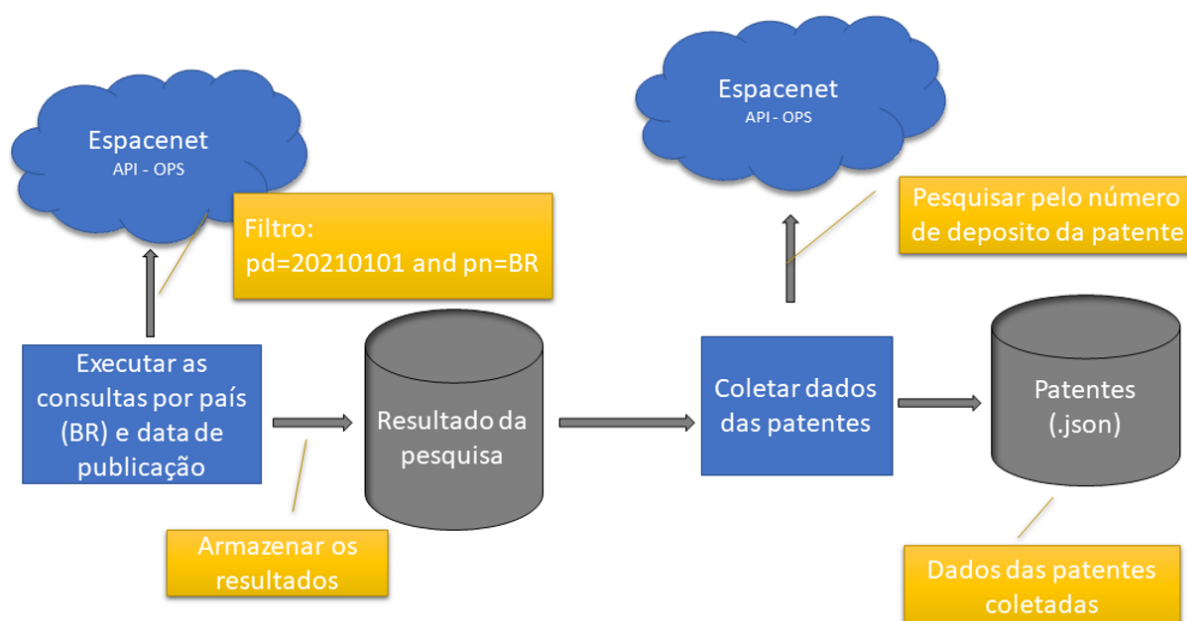
todos os arquivos CSV com os resultados do tratamento dos números de depósito de patentes e fazendo uso dos serviços disponíveis na OPS, realiza a consulta de cada patente, adotando como critério de busca os números de pedido de depósito de patente previamente tratados, armazenando cada patente localizada com sucesso no repositório da Espacenet, em um arquivo no formato .json. Após 241 horas de execução do algoritmo foi possível recuperar dados de 722.347 patentes identificadas com sucesso na Espacenet, cerca de 83% das 862.726 patentes coletadas no INPI. A coleta foi realizada entre os meses de julho de 2020 a fevereiro de 2021.

Apesar de considerar satisfatório o processo de identificação das patentes cadastradas no INPI na Espacenet, com objetivo de aumentar este quantitativo, novos estudos e testes foram realizados, chegando em uma abordagem complementar que permite, dando sequência a este projeto e para projetos futuros, traçar estratégias para manter a base de dados atualizada com frequência.

Dentre os serviços disponíveis na OPS, existe o serviço de pesquisa, no qual é possível informar filtros, tais como, data de publicação, país, título, nome do inventor e nome do depositante, dentre outros. Logo, seguindo o formato definido pela documentação da API (OPS, 2020), por exemplo, para consultar patentes brasileiras publicadas no período de janeiro de 2021 à dezembro de 2021, se faz necessário construir a consulta da seguinte forma: *search?q=pn=BR and pd="20210101 20210131"*, em que *search* é serviço de consulta de patente. O parâmetro *q* que recebe os critérios da consulta, já no parâmetro *pn* é informado o código do país no número da publicação e por fim, utilizando o conector "*and*" é informado o parâmetro *pd* em que é informado o período de publicação da patente, destacando a data inicial e final que devem ser informadas no formato AAAAMMDD (ano, mês, dia).

Fazendo uso do serviço de filtro, foi implementado um algoritmo, utilizando a linguagem de programação Python, que realiza requisições à API, informando a data de publicação e país de origem como critério de busca (*pd=20210101 20210101 and pn=BR*), efetuando uma requisição para cada dia, dentro do período de janeiro de 2020 a setembro de 2021, tendo em vista a possibilidade de recuperar as patentes recentes do ano de 2020 e as do ano de 2021 que não haviam sido consideradas anteriormente. Como resultado a API retorna uma lista contendo o número de depósito das patentes que atenderam aos filtros. De posse desta lista, novas requisições foram feitas na API, pesquisando a patente na Espacenet para obter os dados completos, conforme foi apresentado no início desta seção. A Figura 9 apresenta graficamente o processo de pesquisa aqui mencionado.

Figura 9 – Processo de pesquisa de patentes



Fonte: Elaboração do autor

Com a implementação desta estratégia torna-se viável manter a base de dados atualizada, realizando consultas por data de publicação, em que retornará todas as patentes publicadas no período especificado, geralmente após a data da última coleta, assim caso alguma patente publicada anteriormente, tenha sua data de publicação alterada, é sinal que seus dados foram alterados, permitindo assim identificá-las e recuperar os dados atualizados. Após a execução do algoritmo proposto foi possível recuperar 136.275 patentes, somando ao quantitativo já coletado totaliza 858.622 patentes identificadas na Espacenet. Entretanto é necessário salientar que neste conjunto coletado existem patentes em duplicidades, ou seja, ao executar o segundo processo de coleta, que se baseia na data de publicação, foram recuperadas patentes que tiveram atualizações, ocasionando as duplicidades. Após a identificação das duplicidades, foi obtido o quantitativo de 783.601 patentes distintas, o que representa 90,83% do conjunto de patentes recuperadas no INPI.

4.1.3 Coleta de dados na Plataforma Lattes

Dando sequência no processo de aquisição dos dados, a próxima etapa foi coletar currículos registrados na Plataforma Lattes que possuem informações de patentes, como o número do pedido de depósito ou o título da patente, visando validar tais informações com o conjunto coletado junto a Espacenet. Tal validação se faz importante pois todas as informações curriculares da Plataforma Lattes são incluídas pelo próprio indivíduo, sendo importante este processo de validação e posteriormente certificação dos dados. Os currículos da Plataforma Lattes estão disponíveis livremente na internet para consulta,

registrando toda a trajetória profissional, acadêmica e produções científicas de um indivíduo. O processo de coleta e seleção dos dados curriculares da Plataforma Lattes foi realizado por meio do *framework LattesDataXplorer* (DIAS, 2016). O *framework* possui um conjunto de técnicas e métodos responsáveis por coletar, selecionar, tratar e analisar os dados (Figura 10).

Figura 10 – Visão geral do LattesDataXplorer



Fonte: Dias (2016)

O *framework LattesDataXplorer* possui um rico arcabouço composto por diversos módulos, para extrair e selecionar os dados curriculares da Plataforma Lattes foram utilizados os módulos coleta e seleção. Todo o processo de coleta é composto por 3 etapas:

1. **Extração de URLs:** responsável por obter os códigos únicos de todos os currículos cadastrados, possibilitando o acesso individual a cada currículo, pois cada currículo na Plataforma Lattes possui um código único composto de oito dígitos.
2. **Extração de Identificadores e Data:** Acessar o cabeçalho de cada currículo para recuperar seu identificador único, a data de última atualização e o código do currículo. Armazenando os dados localmente em um arquivo de identificação que contém código, identificador, data da última atualização no CNPq e data da atualização do currículo.
3. **Extração dos currículos:** Extrair e armazenar os currículos. Caso a data de atualização na Plataforma Lattes seja divergente da data de atualização do currículo armazenado localmente, o currículo local é substituído pelo mais recente. Os currículos coletados possuem o formato XML, pois essa versão contém seções e campos bem delimitados.

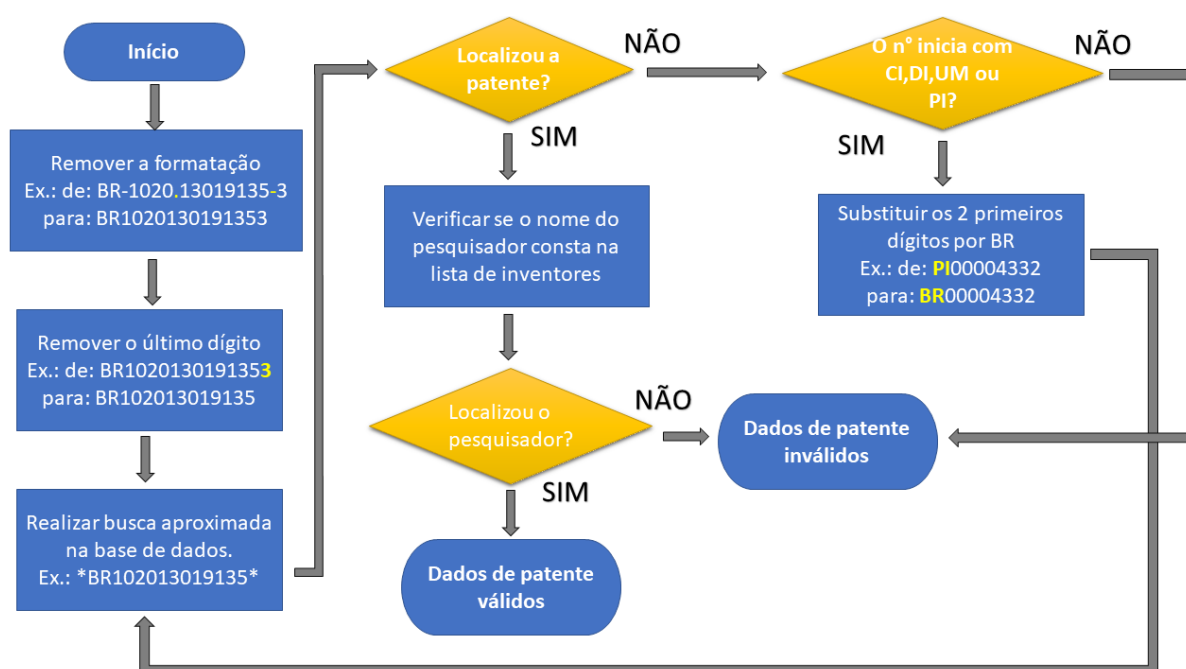
A coleta dos currículos foi realizada em julho de 2019, recuperando 72.256 registros com informações de patente, distribuídos em um total de 29.516 currículos, currículos estes,

selecionados do conjunto total de currículos da Plataforma Lattes por terem informações de patente. O extrator coleta os currículos e os armazena no formato XML em pastas, identificadas de 00 a 99. O nome da pasta selecionada para armazenar o arquivo e o nome do mesmo são definidas de acordo com seu número único de identificador de 16 dígitos, os dois primeiros números do identificador correspondem ao nome da pasta e os 14 dígitos restantes correspondem ao nome do arquivo salvo.

Um dos objetivos do estudo é a possibilidade de certificar os dados de patentes informados nos currículos cadastrados na Plataforma Lattes com informações registradas em repositórios de dados consistentes. Logo, se torna necessário validar as informações das patentes inseridas pelos pesquisadores em seus currículos para garantir a integridade e veracidade das informações. O processo de validação consiste em identificar a patente informada pelo pesquisador em seu currículo na Plataforma Lattes na base de dados de patentes recuperada e constituída neste trabalho, considerando todas as estratégias anteriormente apresentadas.

A Figura 11 apresenta a visão geral da estratégia adotada para realizar tal validação. A informação utilizada como critério de identificação é o número de depósito da patente.

Figura 11 – Visão geral do processo de validação



Fonte: Elaboração do autor

Nos currículos da Plataforma Lattes o número de depósito da patente, bem como as demais informações, são inseridas pelo próprio pesquisador, o que geralmente ocasiona uma falta de padrão no registro das informações. Portanto, se faz necessário implementar o “passo-1” da estratégia de validação proposta que consiste na remoção da formatação

do número de depósito, removendo todos os caracteres especiais como ponto, vírgula, símbolos, dentre outros. Em continuidade no tratamento do número de depósito, no “passo-2” é realizada a remoção do último dígito que compõem o número. Feito isto, o “passo-3” consiste em uma busca aproximada pelo o número de depósito já tratado, esta busca aproximada consiste em realizar a uma consulta SQL na base de dados utilizando como critério de seleção: %[número tratado]%, em que as porcentagens representam qualquer caractere, ou seja, qualquer que seja o valor antes e/ou depois do número tratado será considerado como resultado satisfatório da busca. Caso tenha localizado alguma patente que atenda aos critérios de busca no “passo-3”, é realizado o “passo-4”, em que é verificado se o nome do pesquisador consta na lista de inventores da patente, para isso, usa-se o nome do pesquisador conforme foi informado em seu currículo, a exemplo “Raulivan Rodrigo da Silva”, posteriormente o nome é dividido por espaços formando uma lista, com base no exemplo dado, a lista ficaria com 4 elementos “[Raulivan, Rodrigo, da, Silva]”, feito isto, verifica se todos os elementos da lista correspondem aos nome dos inventores, respeitando a mesma grafia e ordem de ocorrência, ignorando outros possíveis termos existentes no nome dos inventores. Caso tenha encontrado, finaliza-se o processo de validação considerando o número informado pelo pesquisado como válido, caso não tenha encontrado, finaliza-se o processo considerando o número informado pelo pesquisador como inválido. Ainda no “passo-3” existe um fluxo alternativo, em que, caso não tenha localizada uma patente é investigado se o número utilizado como critério de busca inicia com alguns dos prefixos “CI”, “DI”, “UM” ou “PI”, em caso negativo, finaliza o processo considerando o número informado inválido, mas em caso afirmativo, é feita a substituição do prefixo identificado por “BR”, voltando assim ao “passo-3” dando continuidade no fluxo já estabelecido.

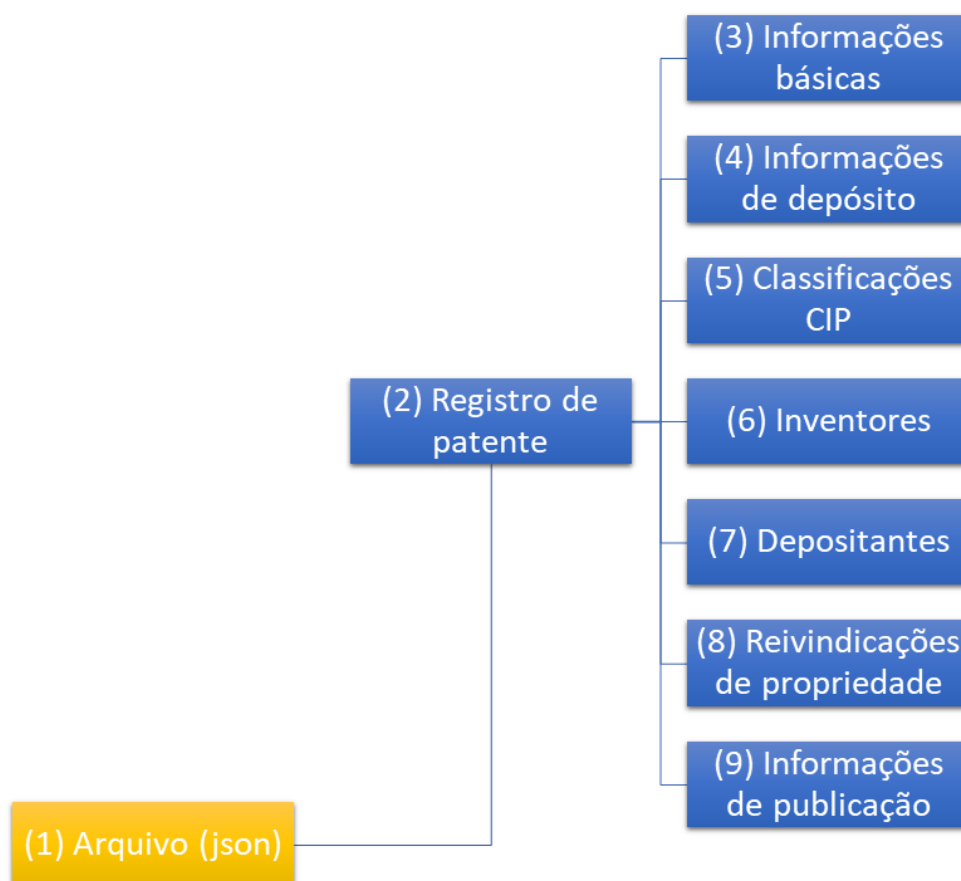
Para automatizar o processo de validação, foi implementado um algoritmo utilizando a linguagem de programação Python. Após a execução do algoritmo, do total de 72.256 registros com dados de patentes informado nos currículos coletados da Plataforma Lattes, 31.816 registros que contêm informações de patentes consideradas como válidas, ou seja, aproximadamente 44% deste conjunto puderam ser certificadas. Somente o conjunto de dados considerado válido será utilizado nas estatísticas deste estudo.

4.2 Seleção, tratamento e organização dos dados

As patentes coletadas na Espacenet, foram armazenadas em arquivos no formato .json, formato este que potencializa a manipulação dos dados. Cada arquivo possui informações de uma ou mais patentes, a [Figura 12](#) apresenta como é organizado o conteúdo de cada arquivo. O item (1) ilustra o arquivo gerado com os dados obtidos ao consultar um patente na Espacenet; já o item (2) representa um registro de patente, um arquivo pode conter um ou mais registros de patentes; no item (3) são as informações básicas, tais como título, resumo e código de identificação da família a qual a patente pertence; o item (4)

apresenta as informações de depósito da patente, como a data e número de depósito no formato original, tal como informado pelo INPI, e nos padrões EPODOC e DOCDB, estes dois últimos são regras de formatação aplicado no número de identificação da patente, DOCDB é usado na versão atual da Espacenet e OPS, e o EPODOC é usado na versão Classic da Espacenet e OPS; no item (5) são as classificações recebidas pela a patente; no item (6), a listagem dos nomes dos inventores; no item (7) contém a listagem dos nomes dos depositantes; no item (8) informações das reivindicações de propriedade; e por fim, no item (8) as informações de publicação da patente, como data de publicação e número de identificação da patente, assim como nos dados de depósito, os dados de publicação seguem os formatos original, EPODOC e DOCDB.

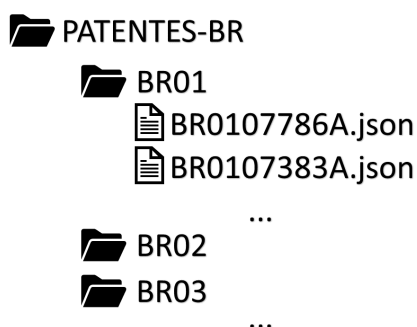
Figura 12 – Representação da estrutura do arquivo



Fonte: Elaboração do autor

Todos os arquivos foram armazenados em um diretório denominado "PATENTES-BR". O nome de cada arquivos é formado pelo número da patente o qual armazena os dados, por exemplo "BR0107786A.json", com base nesta informação, foram criados 69 diretórios para armazenar as patentes, cujo o nome do diretório é composto pelos 4 primeiros caracteres do nome do arquivo, de acordo com o exemplo citado, o nome do diretório a qual ele pertence é "BR01", um desenho esquemático é apresentado na [Figura 13](#).

Figura 13 – Organização de arquivos de patente



Fonte: Elaboração do autor

Essa estrutura de diretórios adotada, permite organizar melhor os arquivos, otimizando o armazenamento e recuperação de arquivos. A [Tabela 5](#) apresenta a listagem de pastas criadas e suas respectivas quantidade de arquivos armazenados. É possível observar que as pastas não seguem uma sequência exata, devido a ausência de pastas que ficariam vazias. Com essa estrutura, cada diretório ficou em média com 11.876,03 arquivos, totalizando 783.818 arquivos.

Tabela 5 – Listagem de pastas

Diretório	Quantidade	Diretório	Quantidade	Diretório	Quantidade
BR00	17.372	BR60	1.760	BR82	11.334
BR01	17.090	BR61	1.861	BR83	10.894
BR02	15.696	BR62	1.730	BR84	7.298
BR03	17.227	BR63	2.017	BR85	7.258
BR04	3	BR64	2.149	BR86	7.137
BR05	1	BR65	2.129	BR87	7.812
BR10	50.054	BR66	2.584	BR88	7.853
BR11	155.625	BR67	2.930	BR89	7.791
BR12	1.814	BR68	2.810	BR90	7.959
BR13	580	BR69	2.662	BR91	7.222
BR20	16.267	BR70	2.872	BR92	6.967
BR21	237	BR71	2.867	BR93	7.673
BR22	16	BR72	2.117	BR94	8.507
BR30	5.857	BR73	2.557	BR95	10.527
BR32	12	BR74	5.646	BR96	12.677
BR53	1	BR75	11.307	BR97	14.992
BR54	957	BR76	11.479	BR98	16.072
BR55	1.479	BR77	11.668	BR99	17.578
BR56	1.447	BR78	11.553	BRC1	1
BR57	1.639	BR79	12.221	BRDI	850
BR58	1.817	BR80	12.247	BRMU	24.129
BR59	1.802	BR81	12.322	BRPI	152.806

A estrutura de diretórios proposta, conforme já mencionado, otimiza o processo

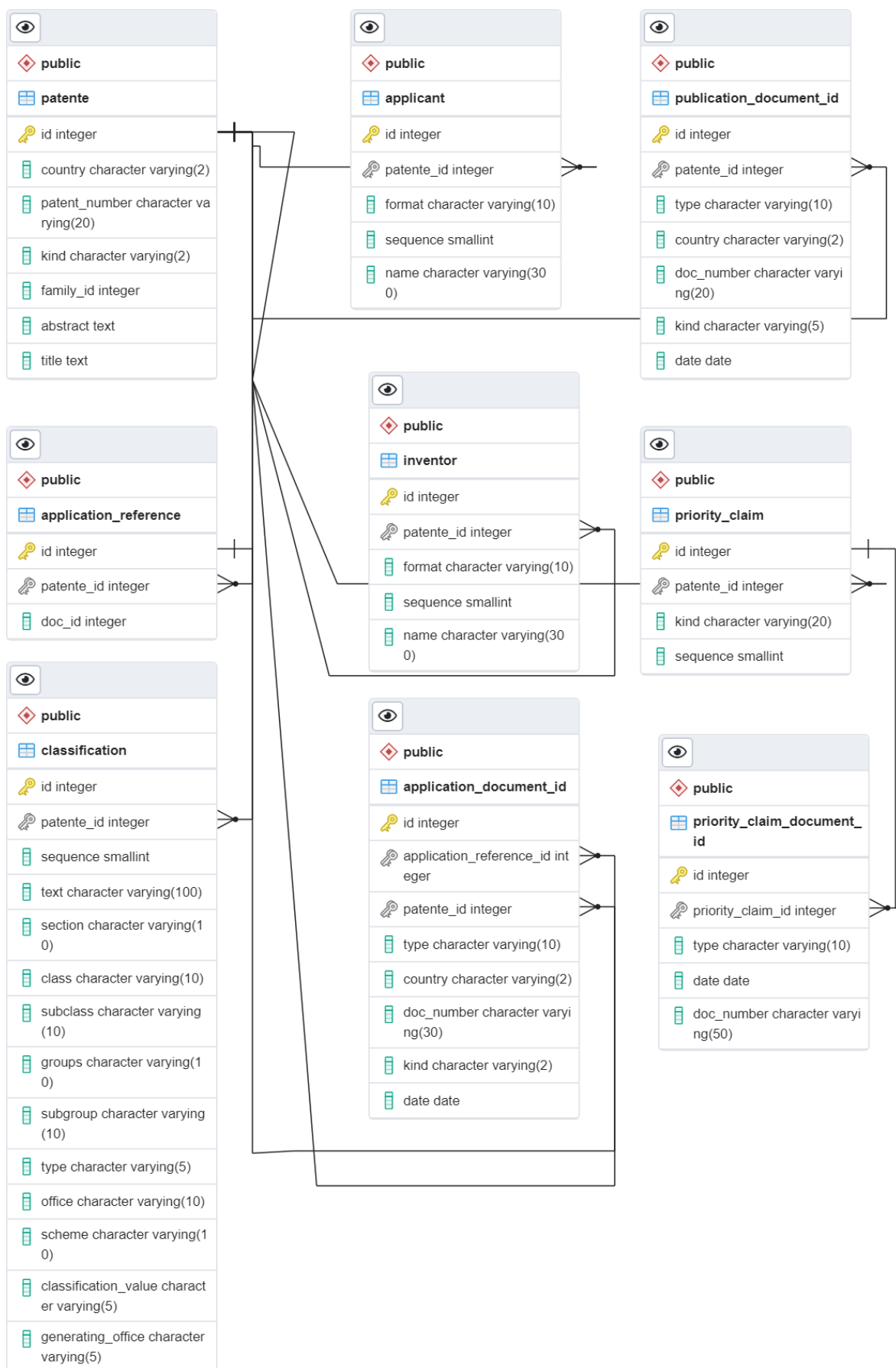
de recuperação e armazenamento, entretanto, para extrair informações que dependem de relacionamento e/ou cruzamento de dados, exige maior esforço computacional, na construção de algoritmos que coletem as informações necessárias dos arquivos. Como por exemplo, gerar um indicador de patentes por classificação, em que seria necessário percorrer todos os arquivos para coletar as informações de classificação e contabilizá-las. Com intuito de resolver este problema e viabilizar novas estratégias de processamento e cruzamento de informações, foi também adotado o uso de um banco de dados relacional, baseado no Modelo Entidade Relacionamento introduzido por Peter Pin-Shan Chen, em 1976 (DATE, 2003).

O banco de dados definido foi o PostgreSQL, um banco de dados que fornece acesso a pontos de relacionamento para os dados armazenados. Os Bancos de dados relacionais, como o PostgreSQL, são baseados no modelo relacional, uma maneira intuitiva e direta de representar dados em tabelas, em que cada linha na tabela é um registro com uma identificação exclusiva denominada chave. As colunas da tabela contém os atributos dos dados armazenados, viabilizando o estabelecimento das relações entre os pontos de dados (ELMASRI; NAVATHE, 2011).

O PostgreSQL é um Sistema Gerenciador de Banco de Dados Objeto-Relacional (SGBDOR) desenvolvido no Departamento de Ciência da Computação da Universidade da Califórnia em Berkeley (POSTGRESQL, 2005). Dentre suas diversas vantagens, sua escolha se deu pelo fato de ser de código aberto, sua compatibilidade com a linguagem SQL (*Structured Query Language*) e pela sua capacidade de lidar com grande volume de dados, sem a perda de performance.

O modelo de entidade e relacionamento definido é apresentado pela Figura 14 gerado pelo PostgreSQL, em que foram definidas 9 tabelas. Na tabela de “patente” é armazenado o cabeçalho das patentes, sendo esta a principal tabela em que é identificada a patente; na tabela “applicant” são armazenados dados dos depositantes das patentes; na tabela “application_document_id” é armazenado a data e número de depósito da patente em seus respectivos formatos; já na tabela “application_reference” contém informações do processo de depósito da patente; na tabela “classification” armazena as classificações das patentes; na tabela “inventor” são armazenados os nomes dos inventores; na tabela “priority_claim” e “priority_claim_document_id” são armazenadas informações referente as reivindicações de propriedade; e por fim a tabela “publication_document_id” armazena os formatos de identificação da patente no processo de publicação.

Figura 14 – Diagrama de entidade e relacionamento



Já os dados coletados da Plataforma Lattes, foram armazenados em um único arquivo do tipo .CSV, formato que define uma estrutura delimitada para cada informação, que facilita a recuperação dos dados. O arquivo é composto pelos seguintes campos: identificação do currículo na Plataforma Lattes, que será útil para obter maiores informações do pesquisador, como a grande área de atuação e instituição de atuação profissional; nome do pesquisador; título da patente informado pelo pesquisador; número de depósito da patente; e ano de depósito. Para complementar as informações, foi acrescentado aos dados da Plataforma Lattes as seguintes informações: o número da patente na Espacenet; formato de publicação do número na Espacenet (docdb, epodoc, original); título da patente na Espacenet; e por fim o ano de publicação da patente na Espacenet.

Logo, diante de todos os procedimentos metodológicos propostos, foi possível realizar uma análise quantitativa dos dados que serão apresentados no próximo capítulo. O que irá viabilizar a obtenção de um retrato até então inédito sobre o processo de produção tecnológica no Brasil. Bem como, de um primeiro esforço para a certificação de dados originados de bases locais, com a Plataforma Lattes com os dados recuperados de repositórios internacionais, como a Espacenet.

5 Análise e Discussão dos Resultados

Este capítulo apresenta uma análise dos dados de patentes obtidos no desenvolvimento deste trabalho, em especial os coletados na Espacenet e na Plataforma Lattes. Nas próximas seções, são apresentados: a caracterização geral dos dados ([Seção 5.1](#)) e a exposição de repositórios internacionais como mecanismo de certificação ([Seção 5.2](#)).

5.1 Caracterização geral

Os dados coletados no INPI e Espacenet no período compreendido entre janeiro de 1900 a setembro de 2021 totalizaram 858.622 registros de patentes, contendo 1.914.866 nomes de depositantes e 4.386.733 nomes de inventores, cerca de 25,6 Gb (Gigabyte) de dados. Em que o depositante é a pessoa física ou jurídica que tem os direitos de propriedade intelectual sobre a patente, já o inventor é o pesquisador que desenvolveu a patente, ressaltando que um mesmo indivíduo pode ser depositante e inventor. Deste montante, 783.601 são registros de patentes distintas, ou seja, 75.021 registros de patentes correspondem a publicações de atualizações de patentes publicadas anteriormente. O quantitativo de patentes distintas extraídas da Espacenet corresponde a cerca 90,83% do conjunto de patentes coletados no INPI, a [Tabela 7](#) apresenta o quantitativo das patentes coletadas e identificadas.

Tabela 7 – Quantitativo das patentes identificadas

Situação	Total de patentes	%
Patentes coletadas no INPI	862.726	100,00
Patentes identificadas e validadas na Espacenet	858.622	99,52
Patentes distintas identificadas e validadas Espacenet	783.601	90,83

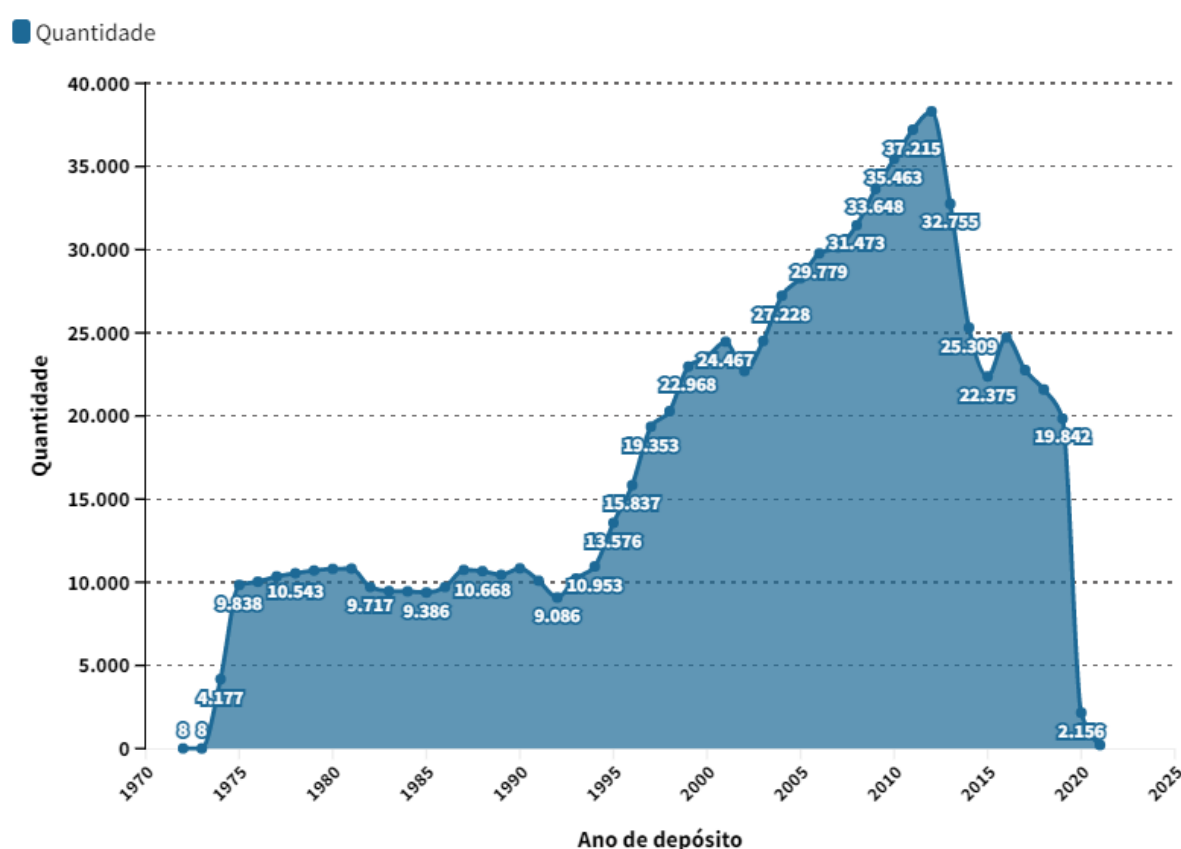
É importante destacar que 79.125 patentes não puderam ser identificadas na Espacenet, uma hipótese para tal, se dá que, apesar das estratégias adotadas para ajustar o número de depósito da patente do padrão usado no INPI anterior a 2012 para o padrão atual OMPI St13, não obteve sucesso em todos os casos, outra hipótese, consiste que as patentes não identificadas ainda não foram disponibilizadas no repositório da Espacenet ou se encontram em período de sigilo.

Como resultado, inicialmente, foi gerada uma base de dados com informações provenientes de patentes depositadas no INPI e disponibilizadas na Espacenet, de igual modo informações extraídas de currículos da Plataforma Lattes. Base esta que viabiliza a

análise de grande volumes de dados sobre produção técnica brasileira, se mostrando uma alternativa viável às limitações encontradas em repositórios online de consulta de dados patentes.

Para apresentar um panorama da evolução tecnológica nacional, foi realizada a análise dos depósitos anuais de patentes, apresentando dados entre 1972 a 2021, data de registro da patente mais antiga até o último ano de análise. A [Figura 15](#) apresenta a evolução temporal no número de depósito de patentes, ressaltando que seis patentes não contém a informação de ano de depósito.

Figura 15 – Evolução temporal do depósito de patentes por ano



Fonte: Elaboração do autor

Desde 1972 até 2021 no Brasil são realizados em média 17.172,32 depósitos de patentes por ano. O ano com maior número de patentes depositadas é o ano de 2012, contendo um total de 38.314 patentes depositadas. Destaca-se ainda que houve um crescimento contínuo no número de depósitos entre os anos de 1992 a 2012, podendo ser considerado um período de grande desenvolvimento científico e tecnológico. Posterior a este ano há uma queda considerável, uma hipótese para tal fato pode estar relacionado a possibilidade de existirem pedidos de patentes que ainda não foram disponibilizados para consulta, pois de acordo com [INPI \(2021c\)](#) depois de solicitar a patente, o pedido

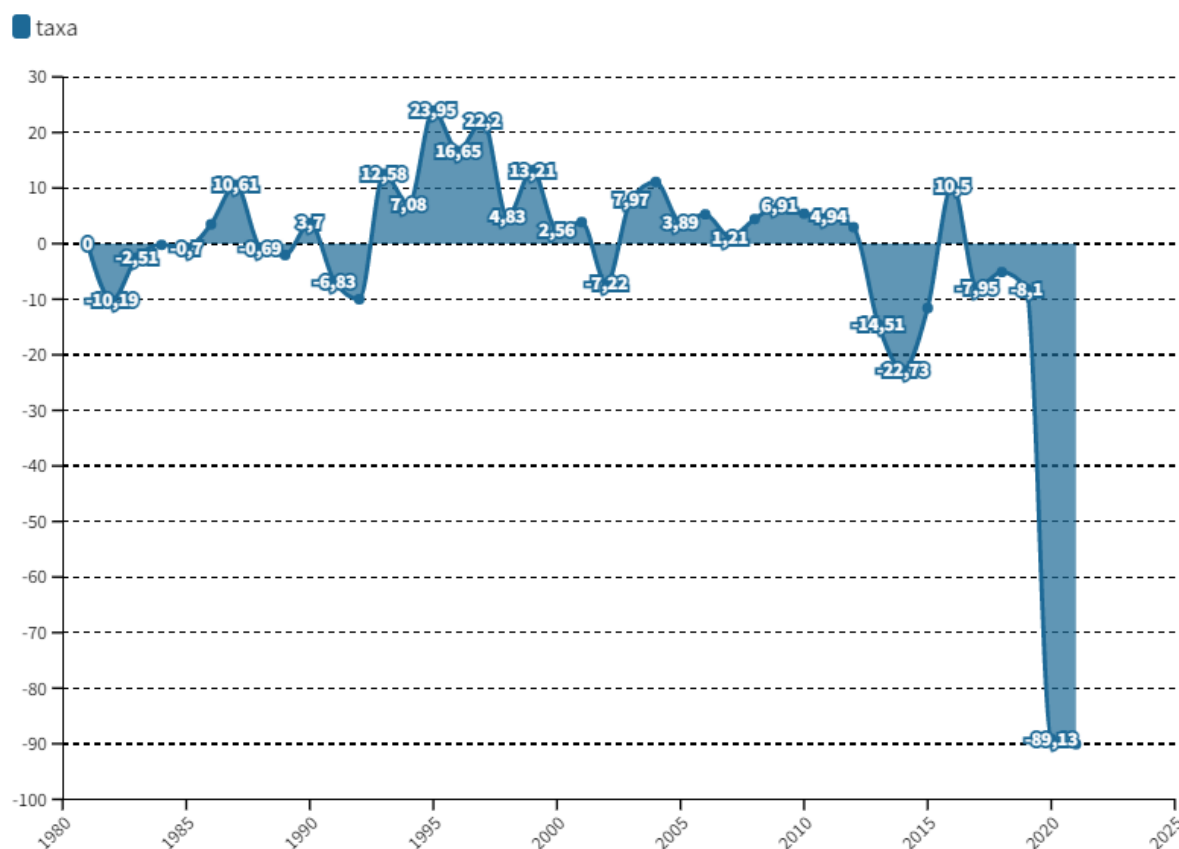
fica em sigilo por até 18 meses, e o depositante tem até 36 meses para solicitar ao INPI que examine esse pedido, levando em média 5,8 anos até sua publicação, em casos mais extremos até 14 anos.

O número de depósito de patentes é comumente utilizado para avaliar a produção tecnológica de empresas, universidades e áreas do conhecimento. No entanto, é relevante verificar as taxas de crescimento anuais para analisar a velocidade do desenvolvimento tecnológico. De acordo com Ramos (2012) o cálculo deste indicador será, portanto, equivalente à diferença entre o somatório atual de depósito de patentes e seu somatório no ano anterior, dividido pelo total de depósitos no ano anterior e depois multiplicado por 100. O resultado será expresso em porcentagem e indicará a evolução no número de depósito de patentes. De acordo com a Fórmula (1):

$$tx = \frac{\sum_{i=1}^{n_1} 1 - \sum_{i=1}^{n_2} 1}{\sum_{i=1}^{n_2} 1} . 100 \quad (1)$$

Sendo que n_1 é o total atual de depósitos, n_2 é total de depósitos no ano anterior. Realizando o cálculo da taxa de crescimento, conforme a Fórmula (1), para os últimos 40 anos, ou seja, para cada ano no intervalo de 1981 à 2021, temos o gráfico apresentado pela Figura 16.

Figura 16 – Taxa percentual de crescimento anual de depósitos de patentes

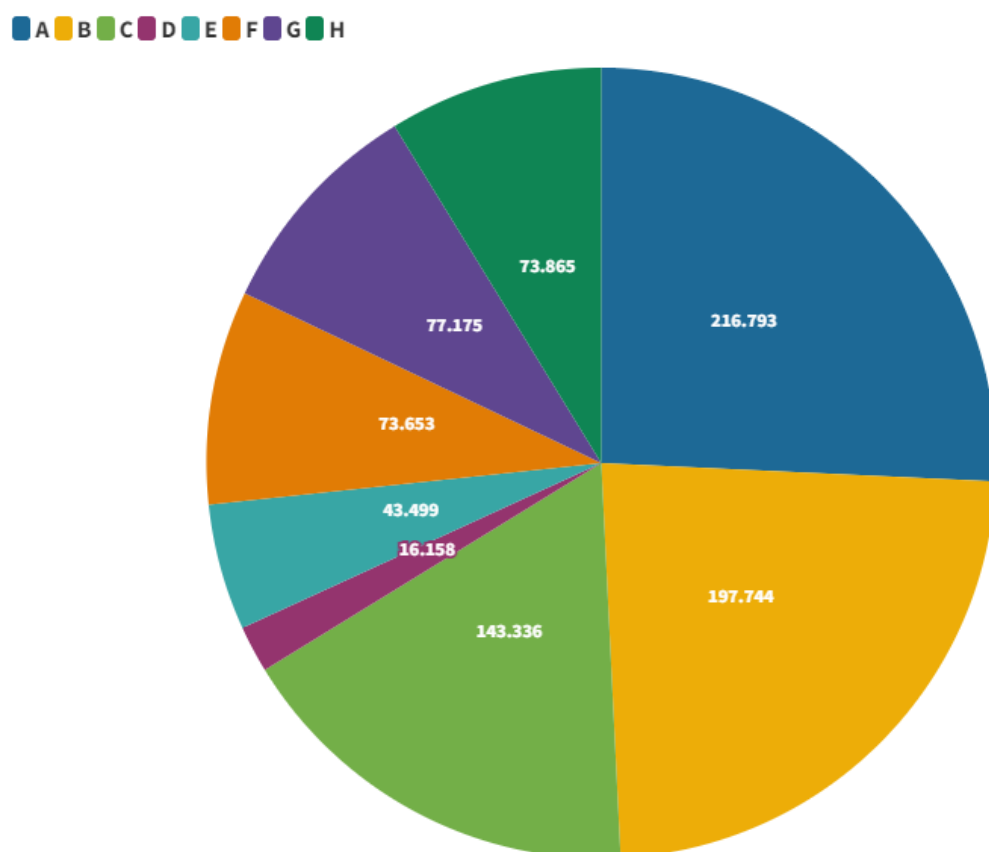


Fonte: Elaboração do autor

É possível observar que ocorrem oscilações na taxa chegando em determinados períodos a ficar com valores negativos, o que representa uma queda no número de depósitos. Os anos de 2020 e 2021 apresentaram valores de -89,13 e -90,07 respectivamente, especula-se que grande parte das patentes depositadas em 2020 e 2021 ainda estão no período de sigilo. Calculando a média desconsiderando os anos de 2020 e 2021, é possível caracterizar que o desenvolvimento tecnológico brasileiro cresce em média 0,02% ao ano.

Cada patente de acordo com sua natureza e finalidade recebe uma classificação de acordo com o sistema internacional de classificação de patentes, com base nestas classificações é possível compreender quais áreas do conhecimento têm gerado o maior número de patentes. Para compreender melhor este cenário foi compilado no gráfico, apresentado pela [Figura 17](#), as classificações recebidas pelas patentes brasileiras consideradas neste estudo.

Figura 17 – Patentes por classificação



A - Necessidades humanas; B - Operações de processamento, transportes; C - Química; Metalurgia; D – Têxteis, Papel; E – Construções fixas; F – Engenharia mecânica, Iluminação, Aquecimento, Armas, Explosões; G – Física; H – Eletricidade.

Fonte: Elaboração do autor

Analisando o gráfico da [Figura 17](#), em um contexto geral, as patentes brasileiras em sua maioria recebem as classificações: A - Necessidades humanas (25,74%); B - Operações

de processamento, transportes (23,47%); e C – Química, Metalurgia (17%). A classificação A-Necessidades humanas, incluem as áreas da agricultura, silvicultura, pecuária, caça, captura em armadilhas e pesca, que são fatores que impulsionam a economia do país. Tudo isso, indo ao encontro com o panorama da economia no brasileira que é fundamentada principalmente no setor agrícola, sendo um dos maiores exportadores de soja, carne de frango e suco de laranja. Além de ser o líder no ranking mundial de cultivo de cana-de-açúcar, o setor do agronegócio tem se apresentando um vetor crucial do crescimento econômico nacional de acordo com a Confederação da Agricultura e Pecuária do Brasil (CNA, 2021).

A classificação que uma patente recebe é composta por 5 elementos, a saber: Seção, Classe, Subclasse, Grupo e Subgrupo, aqui citados do mais abrangente ao mais restritivo. Para obter resultados mais específicos se faz necessário explorar os outros campos que compõem a classificação. Considerando o que foi exposto, foi realizada a análise das patentes por Seção e Classes, das classificações com maiores ocorrências. Para uma melhor visualização dos dados, foi considerado somente as 10 Classes com maior número de ocorrências.

A Tabela 9 apresenta o detalhamento das classificações agrupando as patentes por classes dentro da Seção A - Necessidades humanas, em que é possível observar que as três maiores ocorrências são: “CIÊNCIA MÉDICA OU VETERINÁRIA; HIGIENE” com 46,74%, a Classe “AGRICULTURA; SILVICULTURA; PECUÁRIA; CAÇA; CAPTURA EM ARMADILHAS; PESCA” com 32,71% do quantitativo, e com 9,25% a Classe “ALIMENTOS OU PRODUTOS ALIMENTÍCIOS; TRATAMENTO DOS MESMOS, NÃO ABRANGIDO POR OUTRAS CLASSES”.

Tabela 9 – Indicador de classe da seção A

Seção	Classe	Descrição	Quantidade
A	61	Ciência médica ou veterinária; higiene	101.339
A	1	Agricultura; silvicultura; pecuária; caça; captura em armadilhas; pesca	37.776
A	47	Móveis; artigos ou aparelhos domésticos; moinhos de café; moinhos de especiaria; aspiradores em geral	27.335
A	23	Alimentos ou produtos alimentícios; tratamento dos mesmos, não abrangido por outras classes	14.566
A	63	Esportes; jogos; recreação	9.111
A	45	Artigos portáteis ou de viagem	5.784
A	43	Calçados	3.321
A	41	Vestuário	3.232
A	24	Tabaco; charutos; cigarros; dispositivos simuladores de fumo; artigos para fumantes	3.190
A	21	Cozedura ao forno; equipamento para preparo ou processamento de massas; massas para cozedura ao forno	2.305

Neste cenário é possível apontar a importância da análise da produção técnica por diversos aspectos, a Classe 1 que contém a agricultura, está em segundo lugar no número de patentes, em primeiro lugar está a Classe 61 que contém as áreas da ciência médica,

veterinária e higiene. Com isso, é importante ressaltar que pela a análise quantitativa é possível gerar indicadores de maturidade do campo tecnológico, a quantidade cumulativa de depósitos de patentes ao longo dos anos, no entanto, esse indicador não avalia o valor da patente, ou seja, o retorno econômico e tecnológico que a patente traz.

Já a [Tabela 10](#) apresenta o detalhamento agrupando das patentes por classes dentro da Seção B - Operações de processamento, transportes, em que as três maiores ocorrências são: com 17,97% a Classe “TRANSPORTE; EMBALAGEM; ARMAZENAMENTO; MANIPULAÇÃO DE MATERIAL DELGADO OU FILAMENTAR”, com 15,72% a Classe “VEÍCULOS EM GERAL” e com 13,36% a Classe “PROCESSOS OU APARELHOS FÍSICOS OU QUÍMICOS EM GERAL”.

Tabela 10 – indicador de classe da seção b

Seção	Classe	Descrição	Quantidade
B	65	Transporte; embalagem; armazenamento; manipulação de material delgado ou filamentar	35.546
B	60	Veículos em geral	31.097
B	1	Processos ou aparelhos físicos ou químicos em geral	26.423
B	29	Processamento de matérias plásticas; processamento de substâncias em estado plástico em geral	12.551
B	62	Veículos terrestres para trafegar de outra maneira que não sobre trilhos	9.269
B	23	Máquinas-ferramenta; usinagem de metal não incluída em outro local	8.613
B	21	Trabalho mecânico de metais sem remoção essencial do material; punção de metais	6.297
B	5	Pulverização ou atomização em geral; aplicação de materiais fluentes a superfícies, em geral	6.286
B	32	Produtos em camadas	6.088
B	41	Impressão; máquinas para imprimir linhas; máquinas de escrever; carimbos	5.564

E na [Tabela 11](#) apresentando o agrupando dentro da Seção C-Química, Metalurgia, as três maiores ocorrências são: com 30,17% é a “QUÍMICA ORGÂNICA”, 19,74% é a Classe “COMPOSTOS MACROMOLECULARES ORGÂNICOS; SUA PREPARAÇÃO OU SEU PROCESSAMENTO QUÍMICO; COMPOSIÇÕES BASEADAS NOS MESMOS” e com 10,43% a Classe “BIOQUÍMICA; CERVEJA; ÁLCOOL; VINHO; VINAGRE; MICROBIOLOGIA; ENZIMOLOGIA; ENGENHARIA GENÉTICA OU DE MUTAÇÃO”.

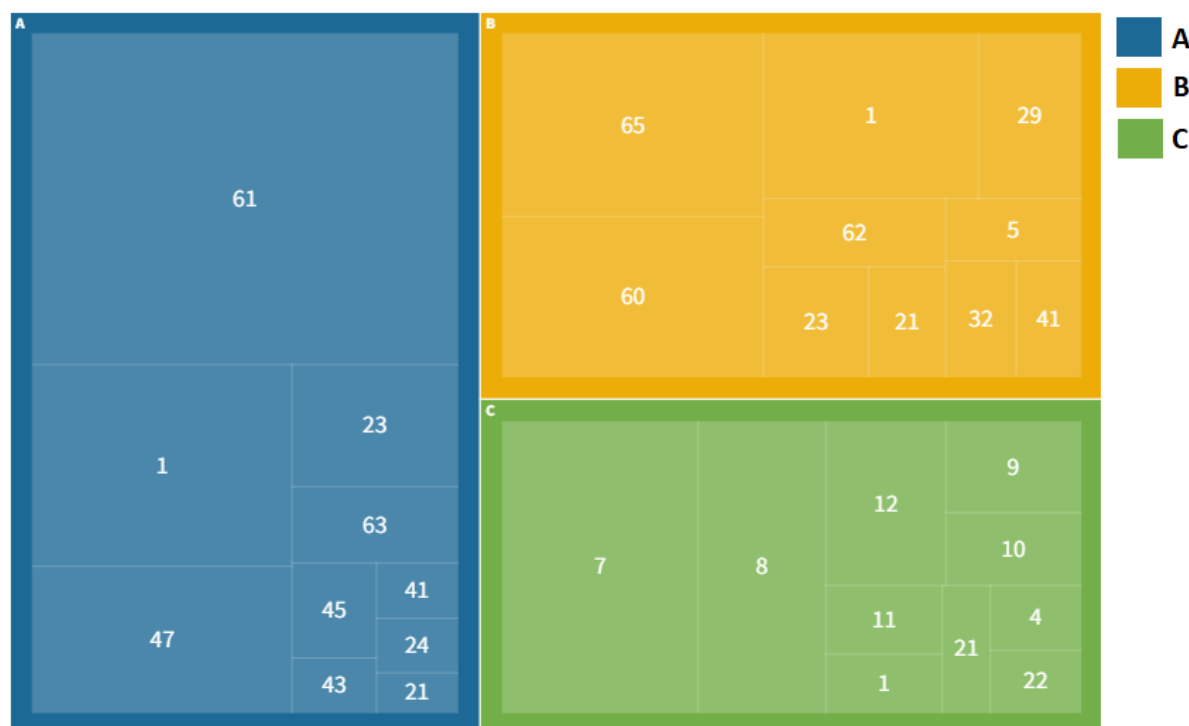
Tabela 11 – indicador de classe da seção c

Seção	Classe	Descrição	Quantidade
C	7	Química orgânica	43.246
C	8	Compostos macromoleculares orgânicos; sua preparação ou seu processamento químico; composições baseadas nos mesmos	28.299
C	12	Bioquímica; cerveja; álcool; vinho; vinagre; microbiologia; enzimologia; engenharia genética ou de mutação	14.956
C	9	Corantes; tintas; polímeros; resinas naturais; adesivos; composições não abrangidos em outros locais; aplicações de materiais não abrangidos em outros locais	9.348
C	10	Indústrias do petróleo, do gás ou do coque; gases técnicos contendo monóxido de carbono; combustíveis; lubrificantes; turfa	7.471
C	11	Óleos animais ou vegetais, gorduras, substâncias graxas ou ceras; ácidos graxos derivados dos mesmos; detergentes; velas	6.099
C	1	Química inorgânica	5.160
C	21	Metalurgia do ferro	4.663
C	4	Cimento; concreto; pedra artificial; cerâmica; refratários	4.459
C	22	Metalurgia; ligas ferrosas ou não-ferrosas; tratamento de ligas ou de metais não-ferrosos	4.267

Esta análise quantitativa permite inclusive avaliar a maturidade do campo tecnológico, avaliando a quantidade cumulativa de depósitos de patentes ao longo dos anos. Para ilustrar melhor, um produto ou serviço inovador pode ser identificado a partir de sua curva S de concessão de patentes que pode indicar se há uma tendência de aumento do depósito de patentes para proteger funcionalidades. A curva S divide o ciclo de vida tecnológico em quatro estágios: iniciação, crescimento, maturidade e saturação (MADEU; PELLANDA; PASSOS, 2021)

Buscando uma nova perspectiva de visualização dos dados apresentando a Figura 18 apresenta o gráfico com os dados apresentados nas tabelas citadas (Tabela 9, Tabela 10, Tabela 11).

Figura 18 – Patentes por seção e classe



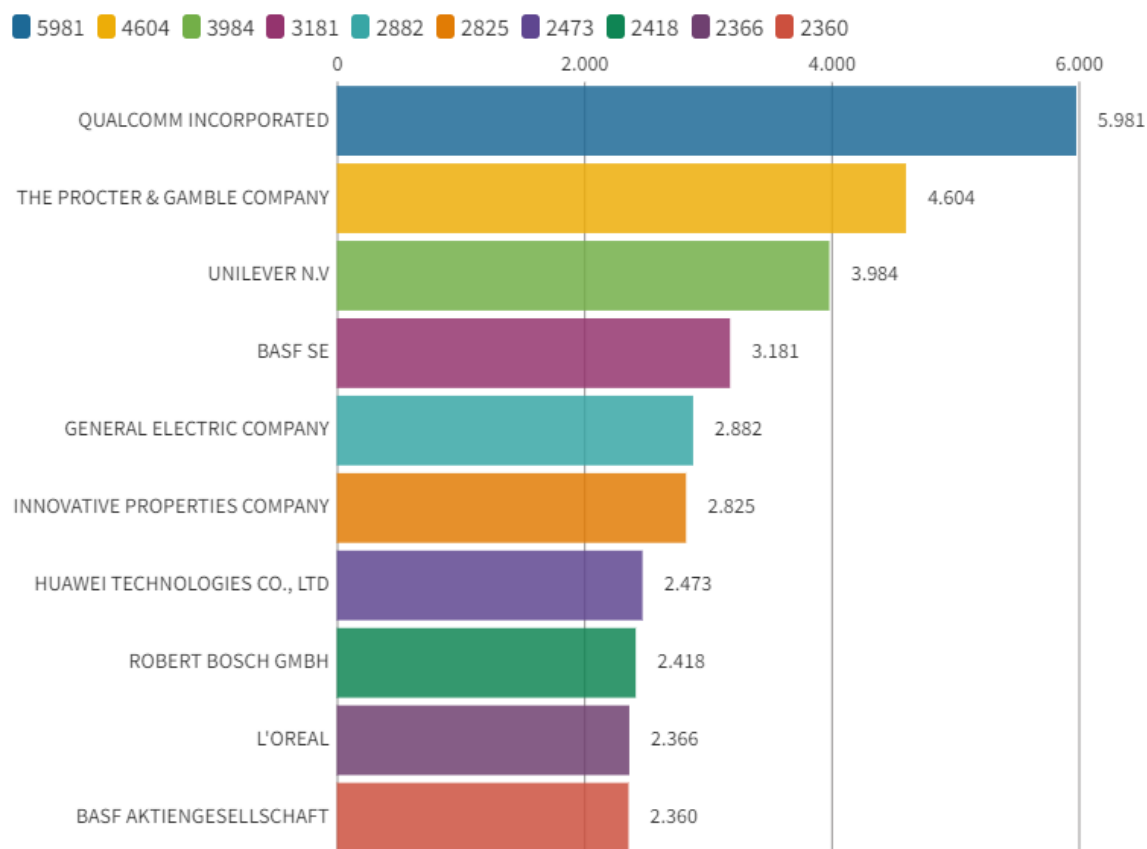
Fonte: Elaboração do autor

O estilo de visualização adotado pela Figura 18 permite identificar intuitivamente quais classes que compõem as seções são mais representativas. Os três grandes quadros que são apresentados na figura, correspondem a cada seção, e dentro de cada quadro, suas respectivas classes. Destacando que os valores numéricos apresentados na figura, não são valores quantitativos, mas sim, a representação das classes, pois a CIP¹ representa as seções utilizando as letras de A à H e as classes são representadas numericamente de 1 à 99.

No intuito de apresentar o resultado da análise que visa identificar os maiores depositantes de patentes no Brasil ao longo dos anos, utilizou-se dos nomes dos depositantes para a classificação de quem tem realizado pedidos. O gráfico apresentado na Figura 19 exibe os 10 maiores depositantes de patentes, em que os três com maiores números de depósito são: “QUALCOMM INCORPORATED”, empresa norte americana fundada em 1985 que produz os chipsets dos celulares, com 5.981 pedidos, seguido de “THE PROCTER GAMBLE COMPANY”, uma corporação norte americana de bens de consumo conhecida como PG, com 4.604 pedidos e “UNILEVER N.V”, multinacional britânica de bens de consumo, com 3.984 patentes depositadas.

¹ <<http://ipc.inpi.gov.br/classifications/ipc/ipcpub/?notion=scheme&version=20200101&symbol=none&menulang=pt&lang=pt&viewmode=f&fipcp=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>>

Figura 19 – 10 maiores depositantes de patentes no Brasil



Fonte: Elaboração do autor

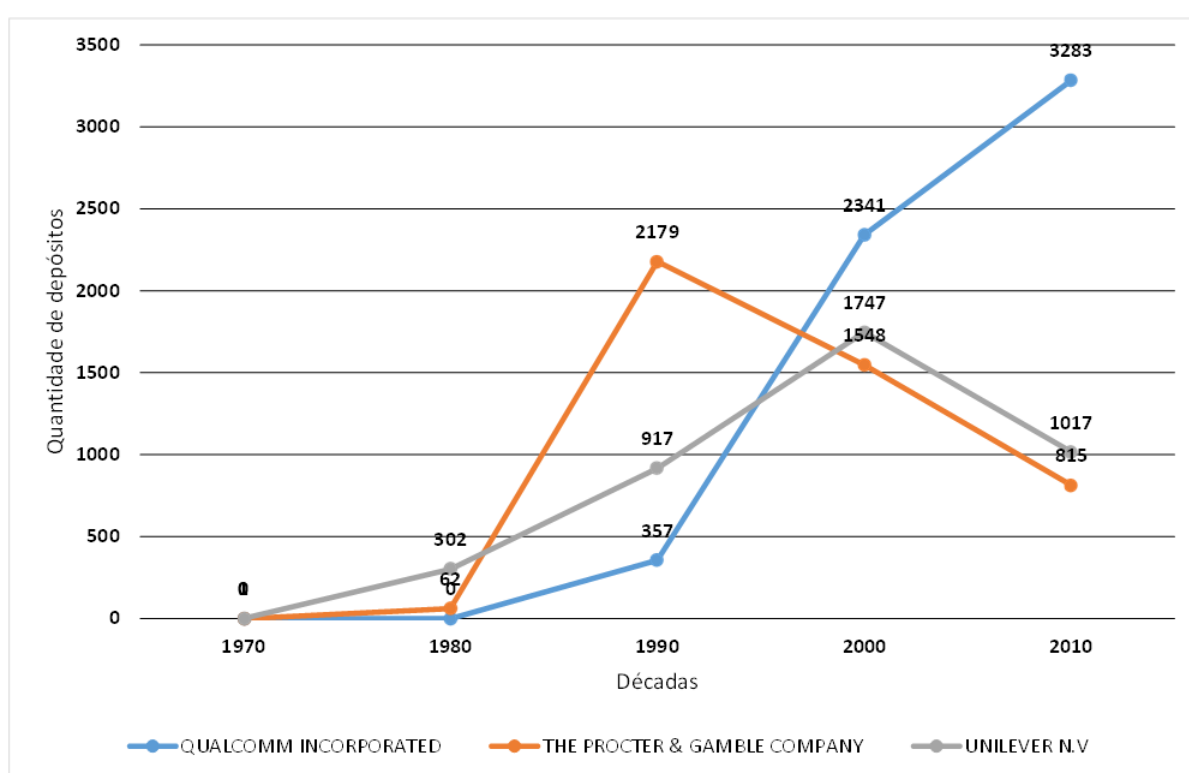
Esta análise permite observar que os maiores depositantes de patentes no Brasil, são empresas internacionais, a saber: Qualcomm Incorporated, The Procter Gamble Company, General Electric Company e 3M Innovative Properties Company que possuem sede nos Estados Unidos da América; a Unilever N.V que possui sede na Inglaterra; Basf SE e Robert Bosch GmbH possuem sede na Alemanha; Huawei Technologies CO com sede na China; e L'oreal com sede na França. Com intuito de também proteger suas invenções, depositaram as mesmas no Brasil. Importante aqui destacar que na ausência de um identificador único para os depositantes, seus nomes foram utilizados para fins de contabilização. Logo, problemas relacionados a ambiguação ou variação dos nomes não foram considerados.

Comparando ao ranking dos maiores depositantes de patentes em 2020 publicado pelo [INPI \(2022b\)](#), o maior depositante estrangeiro foi a empresa Huawei com 448 patentes, já o maior depositante nacional foi a Universidade Federal de Campina Grande na Paraíba com 96 patentes depositadas. Este cenário retrata que as instituições brasileiras precisam investir ainda mais no desenvolvimento tecnológico, destacando ainda que uma universidade pública foi a maior depositante no ano de 2020, ultrapassando empresas reconhecidas

como a Petrobras que depositou 79 patentes neste mesmo ano.

Para se obter uma visão histórica do depósito de patentes, a [Figura 20](#) apresenta a uma série temporal de produção tecnológica dos principais depositantes, se baseando na distribuição por décadas dos números de depósitos dos três maiores depositantes. É possível obser que “QUALCOMM INCORPORATED” começou a realizar seus depósitos na década de 90 e desde então tem aumentando o número de depósitos de patentes, já “THE PROCTER GAMBLE COMPANY” e “UNILEVER N.V” realizaram os primeiros depósitos da década de 70 e 80, com crescimento ao longo dos anos, mas que nos anos 2000 têm apresentado queda.

Figura 20 – Série temporal do depósito de patentes



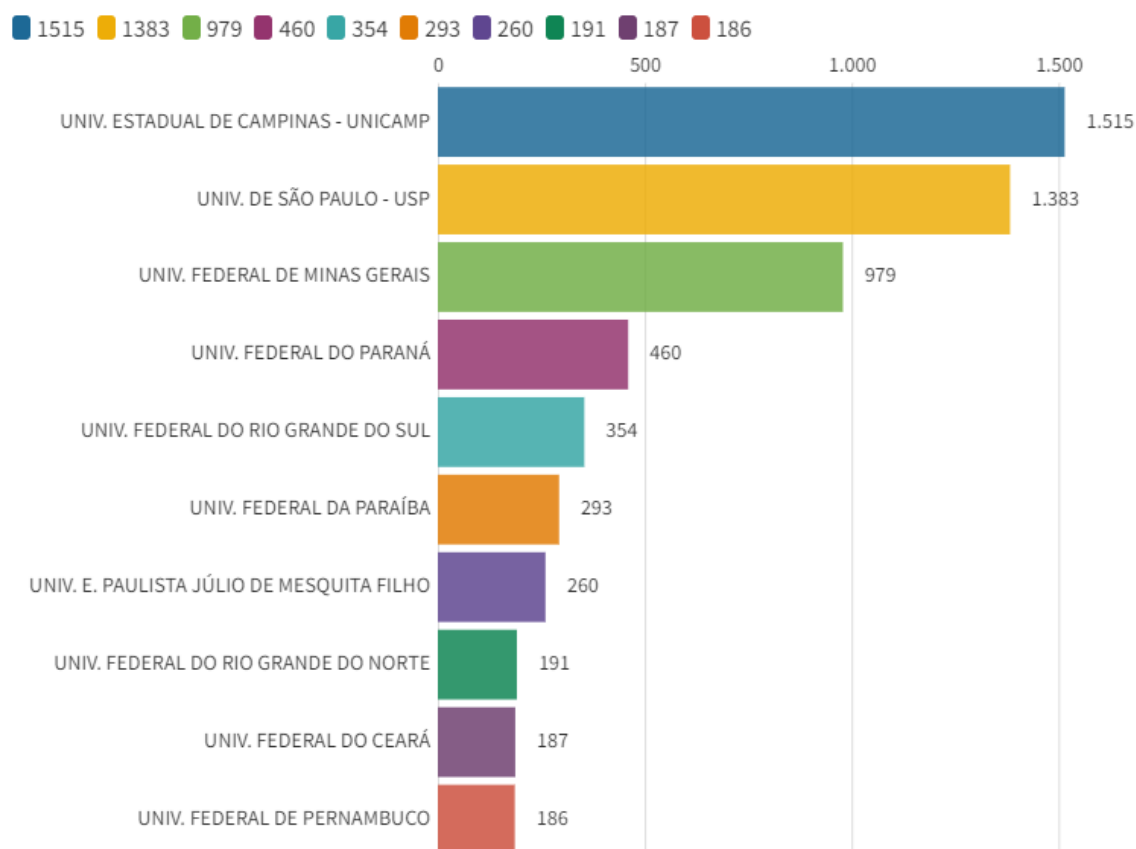
Fonte: Elaboração do autor

Uma hipótese do constante crescimento da empresa QUALCOMM INCORPORATED neste período observado, dar-se ao perfil da empresa, diretamente conectada ao ramo de inovações tecnológicas.

Com intuito de identificar as instituições de ensino superior (IES), que contribuem com o desenvolvimento tecnológico brasileiro, o gráfico apresentado na [Figura 21](#), exibe as 10 IES com maiores números de depósito de patentes, para seleção dos dados foi usado como critério de seleção depositantes que contenham os seguintes termos em seu nome: “UNIVERSIDADE” ou “FACULDADE” ou “ESCOLA” ou “CENTRO DE EDUCAÇÃO” ou “ENSINO SUPERIOR” ou “CENTRO UNIVERSITÁRIO” ou “INSTITUTO FEDERAL” ou

"CENTRO FEDERAL DE EDUCAÇÃO". Esses dados são promissores para identificar IES especialistas em determinada tecnologia e ou áreas do conhecimento.

Figura 21 – 10 maiores IES depositantes de patentes



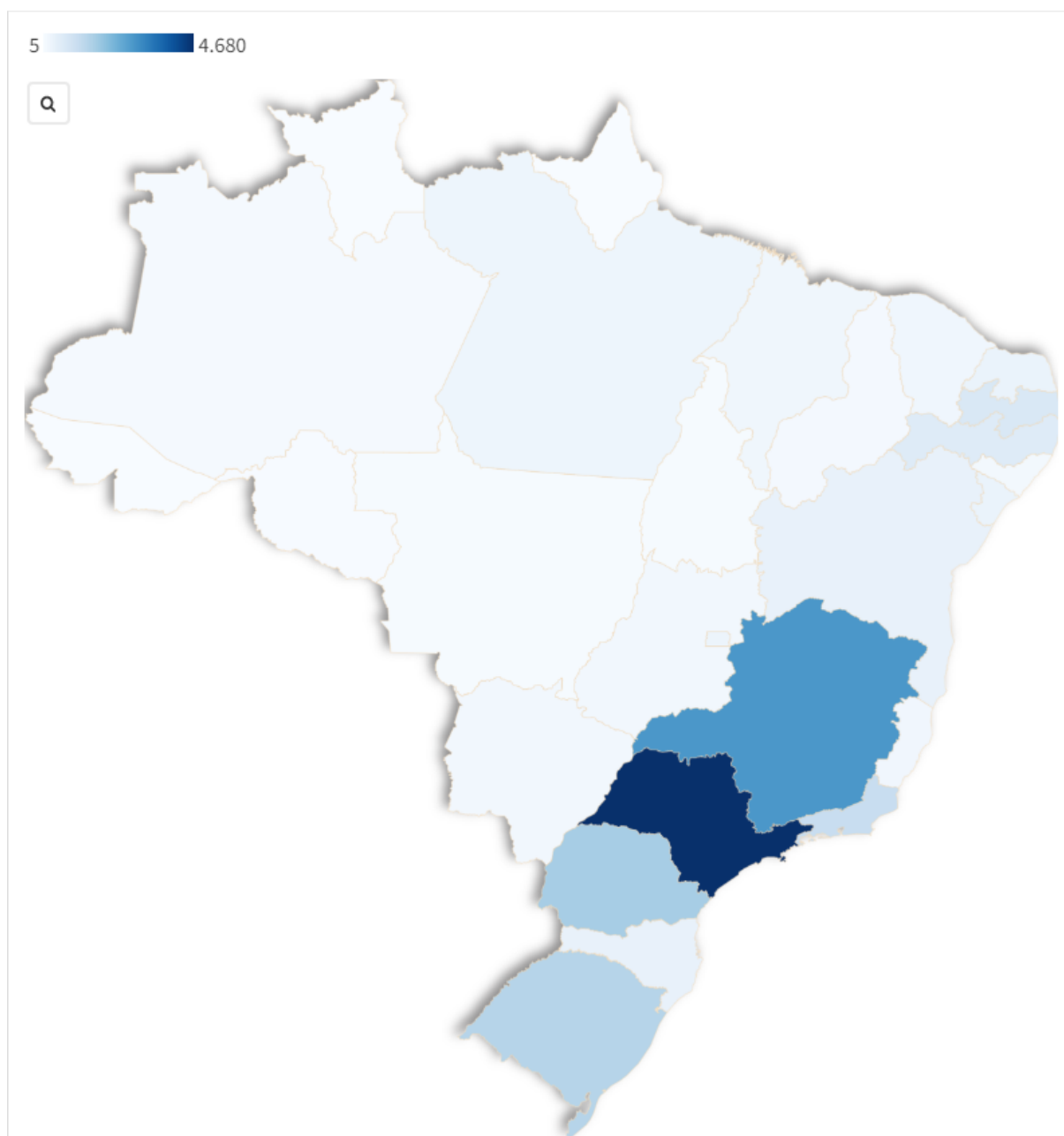
Fonte: Elaboração do autor

Parte da produção de patentes está surgindo em iniciativa das universidades, de acordo com o *ranking* feito pelo [INPI \(2022a\)](#), em 2018 o maior depositante foi a Universidade Federal da Paraíba com 94 depósitos, em 2019 novamente a Universidade Federal da Paraíba com 100 depósitos e em 2020 a Universidade Federal de Campina Grande também do estado da Paraíba, com 96 depósitos. No total acumulado do depósito de patentes considerando toda a série histórica, conforme [Figura 21](#), as universidades UNICAMP (Universidade Estadual de Campinas), USP (Universidade de São Paulo) e UFMG (Universidade Federal de Minas Gerais) se destacam como as IES produtoras de inovação.

Em um contexto geográfico, as maiores IES depositantes de patentes entre os anos de 2018 a 2020, conforme apresentado, são da região Nordeste do Brasil, já em um contexto geral, as IES com maior número de depósito são da região Sudeste do Brasil. Para compreender melhor esta situação e apontar a região brasileira em que as IES são mais propícias ao desenvolvimento tecnológico por meio do depósito de patentes, a [Figura 22](#) apresenta a distribuição do número de depósito de patentes por estados brasileiro

considerando toda a janela temporal.

Figura 22 – Distribuição do número de depósito de patentes de IES por estado.



Fonte: Elaboração do autor

As regiões sudeste e sul do Brasil concentram o maior número de depósito de patentes realizado por IES. Uma hipótese para este fenômeno, se dá ao fato de serem as regiões mais desenvolvidas economicamente e por abrigarem as maiores universidades públicas do país e consequentemente, concentrando um maior percentual de pesquisadores.

Utilizando como objeto de análise o título de todas as patentes coletadas, foi elaborada uma nuvem de palavras ([Figura 23](#)), com ênfase para as palavras mais utilizadas na

Tabela 12 – Palavras mais utilizadas

Ocorrências	Palavras
160.221	processo
155.653	método
99.013	sistema
89.467	dispositivo
58.424	composição
54.801	uso
53.109	aparelho
29.026	produção
28.439	disposição
28.411	tratamento

O título de uma patente deve definir de forma concisa, clara e precisa o escopo técnico da invenção, o título deve representar adequadamente as categorias de reivindicações e não conter expressões ou palavras irrelevantes ou desnecessárias. Dito isso, é possível verificar tendo em vista os principais termos utilizados que grande parte das patentes brasileiras envolvem criação ou melhoria de processos e métodos.

5.2 Repositórios internacionais como mecanismo de certificação

Esta seção tem como intuito apresentar os resultados obtidos quando se analisa os registros de patentes contidos nos currículos cadastrados na Plataforma Lattes como indicador do desenvolvimento tecnológico nacional. Atualmente a Plataforma Lattes é composta por mais de 7.4 milhões de currículos, que abrange indivíduos nos diversos níveis de formação acadêmica. Diversos trabalhos têm se utilizado de dados registrados nos currículos da Plataforma Lattes para diversas análises ([Mena-Chalco e Cesar-Junior \(2009\)](#); [Digiampietri \(2015\)](#); [Brito, Quoniam e Mena-Chalco \(2016\)](#); [Dias, Moreira e Dias \(2018\)](#)).

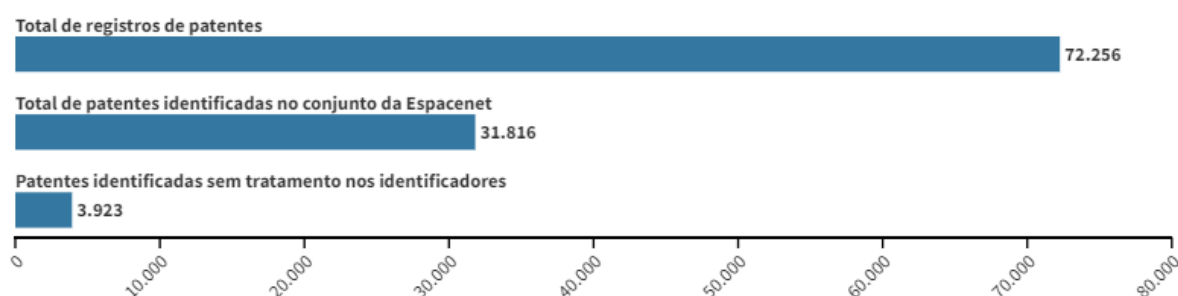
A certificação, bem como a integração de dados de repositórios locais com fontes alternativas e complementares como os dados da Espacenet ou outros repositórios internacionais se caracteriza como um excelente mecanismo para se viabilizar análises mais completas que não seriam facilmente obtidas. Um exemplo é considerar dados individuais dos indivíduos como os que podem ser extraídos dos currículos em conjunto de informações que não estão nos registros das patentes, mas que são recuperados com os dados da Espacenet.

No intuito de avaliar a representatividade e consistência dos registros de patentes informados nos currículos da Plataforma Lattes, foi realizada inicialmente uma busca pelos currículos que possuem registros de patentes. Do conjunto total de currículos, somente 29.516 possuem informações de patentes registradas, menos de 1% de toda a base de

dados curriculares da Plataforma Lattes. Conforme já mencionado os 29.516 currículos possuem juntos um total 72.256 registros com informações de patentes, contudo, não foram todos considerados, apenas 31.816 registros foram devidamente identificados na base de dados da Espacenet, totalizando 16.445 currículos. O restante não foi possível identificar na base de dados da Espacenet aplicando as estratégias definidas neste estudo.

A [Figura 24](#) apresenta um gráfico comparativo dos quantitativos obtidos referente ao processo de identificação, em que 72.256 é o somatório de registros dos currículos que contém número de depósito de patente informado. Deste total 31.816 registros foram identificados corretamente na Espacenet, conforme já apresentado. Destacando que apenas 3.923 dos registros identificados, puderam ser identificados sem a necessidade de tratamento no número de depósito da patente, ou seja, apenas 12% de todos os registros identificados nos currículos da Plataforma Lattes foram informados corretamente por seus respectivos pesquisadores em seus currículos.

Figura 24 – Quantitativo de Patentes coletadas nos currículos da Plataforma Lattes

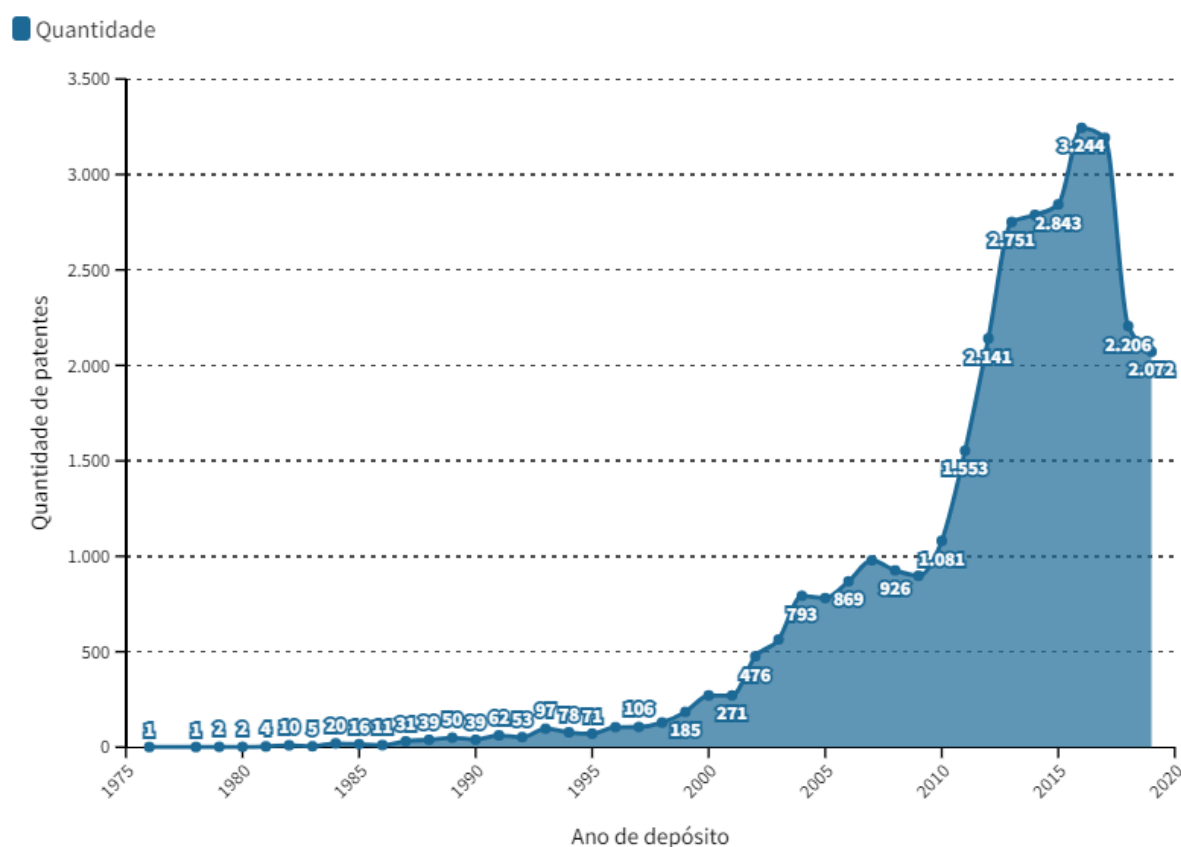


Fonte: Elaboração do autor

Esse resultado aponta a necessidade de se propor um certificador de registros de patentes em repositórios locais, como a Plataforma Lattes, no intuito de viabilizar a realização de análises confiáveis e consistentes. Tal fato, viabiliza estudos que considerem este tipo de produção com adoção de diversas técnicas bibliométricas e patentométricas. Além disso, com a certificação das patentes é possível aumentar a consistência de dados que são informados pelo próprio autor.

Em uma análise temporal das patentes informadas nos currículos da Plataforma Lattes, é possível observar, conforme apresentado na [Figura 25](#), que as patentes foram depositadas entre os anos de 1975 a 2019, tendo maior concentração entre os anos de 2002 e 2016. Ressalta-se que o conjunto de currículos utilizados na análise foram coletados em 2019 o que justifica a ausência de patentes nos últimos anos, tendo em vista que alguns currículos podem não terem sido recentemente atualizados, bem como a possibilidade de existirem patentes no período de sigilo que consequentemente não puderam ser cadastradas por seus proponentes.

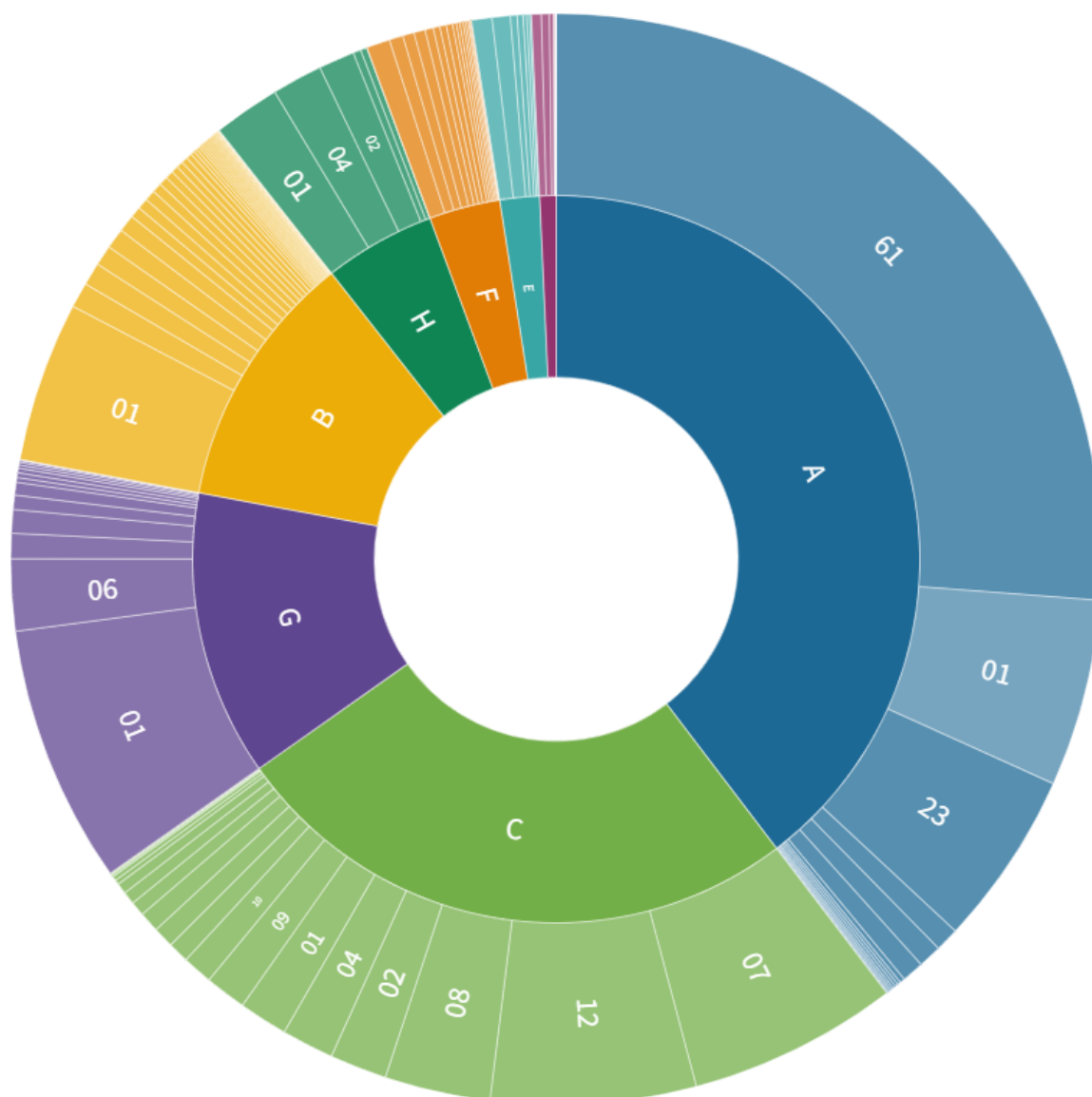
Figura 25 – Análise temporal do depósito das patentes informadas nos currículos da Plataforma Lattes



Fonte: Elaboração do autor

No intuito de se identificar baseado nas classificações das patentes, em quais áreas do conhecimento os pesquisadores têm dedicado seus esforços a [Figura 26](#) apresenta o quantitativo de patentes em relação às classificações obtidas pelas patentes no processo de depósito. É possível observar que a maior parte das patentes identificadas na Plataforma Lattes, possuem as classificações: “A-Necessidades humanas” com 12.479 patentes, “C-Química, Metalurgia” com 8.059 patentes e “G- Física” com 3.996 patentes. Alguns valores foram omitidos na representação para uma melhor visualização. Conforme já mencionado, a identificação das classificações recebidas pelas patentes, são valores numéricos, portanto, a [Tabela 13](#) apresenta a descrição das classificações apresentadas na [Figura 26](#).

Figura 26 – Classificação das patentes registradas nos currículos da Plataforma Lattes.



A - Necessidades humanas; B - Operações de processamento, transportes; C - Química; Metalurgia; D – Têxteis, Papel; E – Construções fixas; F – Engenharia mecânica, Iluminação, Aquecimento, Armas, Explosões; G – Física; H – Eletricidade.

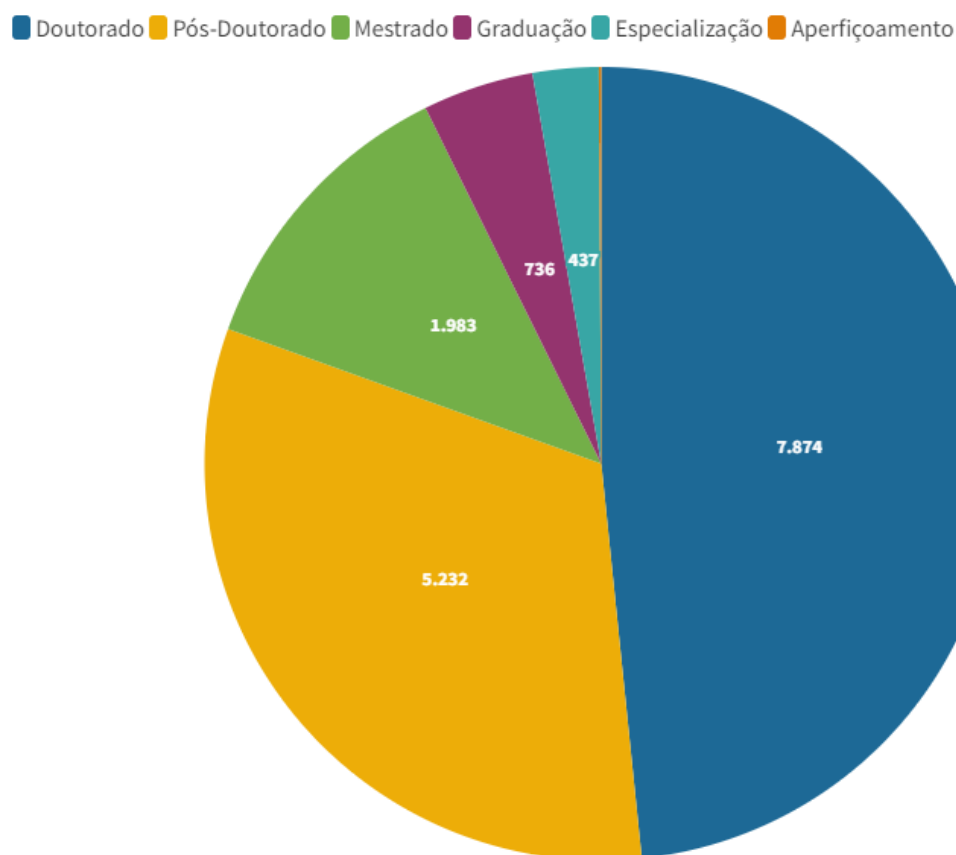
Fonte: Elaboração do autor

Tabela 13 – Classificação das patentes registradas nos currículos da Plataforma Lattes

Seção	Classe	Descrição	Quantidade
A	61	Ciência médica ou veterinária; higiene	8.249
A	01	Agricultura; silvicultura; pecuária; caça; captura em armadilhas; pesca	1.747
A	23	Alimentos ou produtos alimentícios; tratamento dos mesmos, não abrangido por outras classes	1.633
B	01	Processos ou aparelhos físicos ou químicos em geral	1.494
C	07	Química orgânica	1.961
C	12	Bioquímica; cerveja; álcool; vinho; vinagre; microbiologia; enzimologia; engenharia genética ou de mutação	1.912
C	08	Compostos macromoleculares orgânicos; sua preparação ou seu processamento químico; composições baseadas nos mesmos	987
C	02	Tratamento de água, de águas residuais, de esgotos ou de lamas e lodos	532
C	04	Cimento; concreto; pedra artificial; cerâmica; refratários	487
C	01	Química inorgânica	456
C	9	Corantes; tintas; polidores; resinas naturais; adesivos; composições não abrangidos em outros locais; aplicações de materiais não abrangidos em outros locais	386
C	10	Indústrias do petróleo, do gás ou do coque; gases técnicos contendo monóxido de carbono; combustíveis; lubrificantes; turfa	281
G	01	Medição; teste	2.410
G	06	Cômputo; cálculo ou contagem	669
H	01	Elementos elétricos básicos	621
H	04	Técnica de comunicação elétrica	472
H	02	Produção, conversão ou distribuição de energia elétrica	335

Uma análise relevante referente a produção técnica registrada nos currículos da Plataforma Lattes consiste em identificar a formação acadêmica dos proponentes. Este fator permite traçar o perfil dos pesquisadores que contribuem com o desenvolvimento tecnológico nacional, qual o nível de formação que mais favorece o crescimento do número de depósitos de patentes no Brasil. A [Figura 27](#) apresenta a distribuição dos pesquisadores considerando o maior nível de formação acadêmica informada, o critério de seleção adotado foi contabilizar todos os currículos que tiverem pelo menos uma patente válida na Espacenet e que possua informação academia registrada em seu currículo, considerando a de nível mais alto.

Figura 27 – Formação dos proponentes de patentes



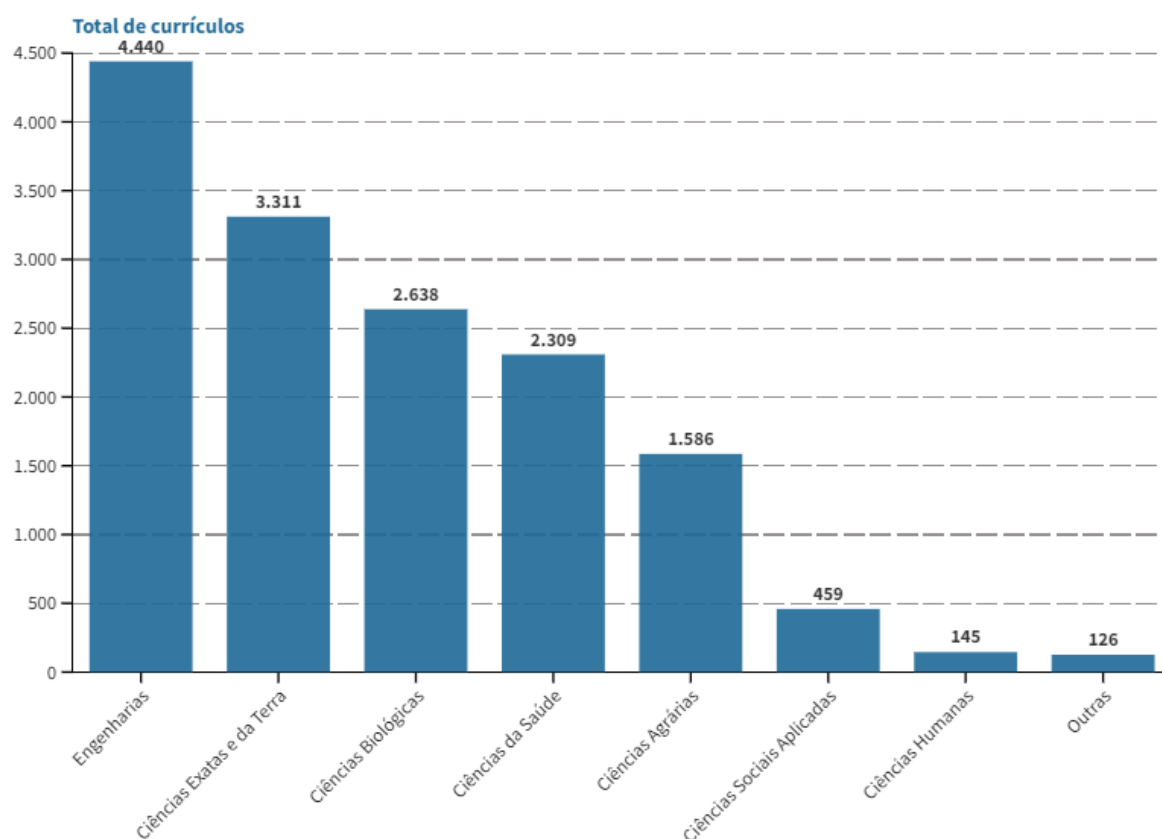
Fonte: Elaboração do autor

Na [Figura 27](#) os valores referente a aperfeiçoamento foram omitidos devido sua baixa representatividade e para melhor visualização dos dados. Avaliar o perfil dos depositantes de patente é importante para compreender o perfil destes proponentes, identificar especialistas em determinadas áreas do conhecimento, bem como assinalar possíveis parcerias para o desenvolvimento de novas tecnologias. De acordo com [Dias \(2016\)](#) os doutores são responsáveis pela maioria das produções científicas registradas nos currículos da Plataforma Lattes, e esse fenômeno se repete na produção técnica, doutores e pós-doutores juntos correspondem a 80,4% do total de depositantes das patentes analisadas. Apesar de pós-doutorado não ser um nível de formação acadêmica, os indivíduos foram classificados nesta categoria para melhor visualização e classificação.

Nos currículos da Plataforma Lattes os indivíduos podem informar em seus dados quais as suas principais áreas de atuação. Este fato permite averiguar qual a área do conhecimento mais representativa tendo em vista o perfil dos proponentes analisados, e conseqüentemente, avaliar qual área do conhecimento se caracteriza como propulsor do desenvolvimento tecnológico nacional. A [Figura 28](#) apresenta as grandes áreas do conhecimento e seus respectivos quantitativos de indivíduos. O critério de seleção adotado

foi contabilizar todos os currículos que tiverem pelo menos uma patente válida na Espacenet e que possua informação de área de atuação registrada em seu currículo. Como um currículo pode ter até seis registros de áreas de atuação, para esta análise a primeira área informada foi considerada.

Figura 28 – Grande área de atuação dos proponentes de patentes

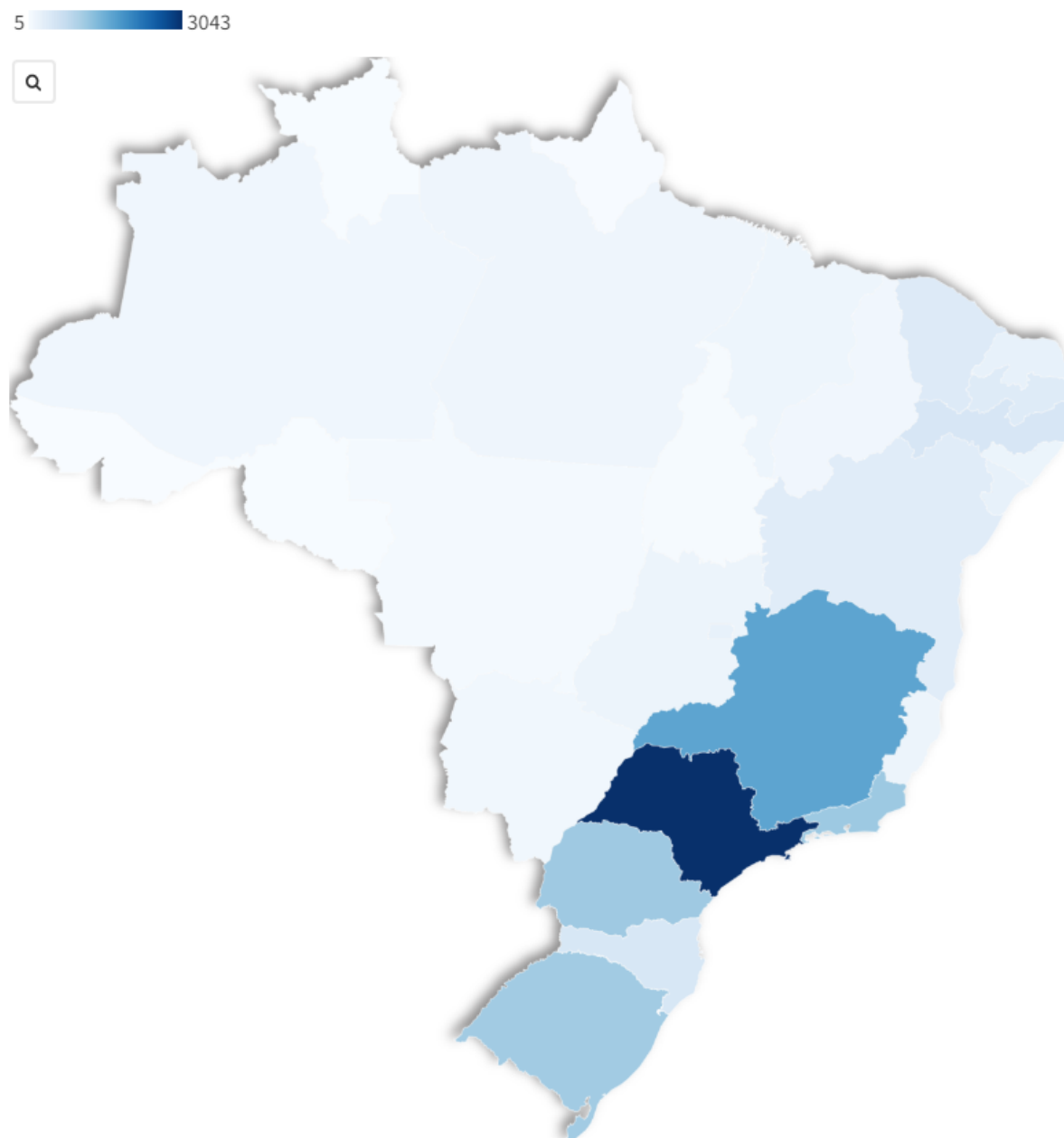


Fonte: Elaboração do autor

O desenvolvimento tecnológico está diretamente ligado às mais diversas áreas da do conhecimento, o que implica a constituição de equipes multidisciplinares de alto nível de conhecimento, no entanto, a área das engenharias se apresenta em destaque no desenvolvimento tecnológico no conjunto analisado.

Em complemento a análise sobre o perfil dos indivíduos que possuem patentes informadas em seus currículos cadastrados na Plataforma Lattes, a análise de seus endereços profissionais permite verificar como estão distribuídos estes proponentes de forma geográfica no país. A [Figura 29](#) apresenta a distribuição por estado onde estão atuando profissionalmente o conjunto de indivíduos analisados.

Figura 29 – Distribuição de proponentes de patentes por estado



Source: IGBE

Fonte: Elaboração do autor

Em concordância com o cenário nacional, as regiões mais expressivas são as regiões Sudeste e Sul, os estados de São Paulo, Minas Gerais e Rio de Janeiro correspondem a 47% do total de depositantes. Uma hipótese se dá ao fato que as maiores universidades depositantes de patentes se concentram também nestas regiões, e consequentemente, possuem um maior quantitativo de indivíduos atuando em pesquisa e desenvolvimento tecnológico nestas instituições.

Os indicadores gerados com base nos dados extraídos dos currículos cadastrados na

Plataforma Lattes destacam que a plataforma além de informações científicas, informações tecnológicas podem ser recuperadas e utilizadas como subsídios para diversas pesquisas.

Neste contexto, é importante destacar que vários registros não puderam ser analisados, devido a falta de vinculação com o repositório construído com os dados do INPI e Espacenet. Isso reforça a importância em se utilizar de estratégias para a vinculação e posterior certificação. Tal processo se torna importante tendo em vista que viabiliza de forma efetiva a possibilidade de validar os dados dos currículos com informações complementares validadas.

6 Conclusão

A inovação tecnológica e a competitividade dos países e empresas em um cenário globalizado voltado ao compartilhamento de informações e conhecimento, estão diretamente associadas aos direitos da Propriedade Intelectual. Empresas e Universidades sentem uma crescente necessidade em assegurar uma proteção jurídica para produtos e/ou tecnologias que desenvolvem, contribuindo para o contínuo e crescente número de depósito de patentes, constituindo de tal forma, um conjunto de dados com informações valiosas sobre o desenvolvimento tecnológico. Mediante a análise de documentos de patentes, pode-se investigar: quais foram as patentes mais depositadas por uma determinada organização em uma janela temporal ou em toda sua trajetória; identificar tecnologias emergentes; quais organizações ou indivíduos que estão patenteando em uma determinada área.

Mundialmente diversas organizações têm usado a análise de dados de patentes como um diferencial competitivo para se manter no mercado, pois permite identificar seus principais concorrentes e identificar os atores responsáveis por tecnologias de áreas de interesse da organização, fator importante para estabelecer parcerias. Ainda, para além das organizações, as Instituições de Ensino e Pesquisa têm assumido ao longo dos anos, um papel cada vez mais relevante no desenvolvimento tecnológico e econômico do país, agregando o empreendedorismo aos seus princípios de ensino, pesquisa e extensão. Transformando parte do conhecimento científico desenvolvido em laboratórios em publicações de patentes. Tal fato torna ainda mais relevante o estudo e acompanhamento tecnológico por meio de patentes.

Mediante ao exposto, as patentes se apresentam como fontes privilegiadas de informações precisas e atuais, no entanto, de difícil análise tendo em vista que os dados contidos neste tipo de documento nem sempre estão acessíveis de forma estruturada e consistentes em outras fontes de dados. Apesar de ser considerada uma metodologia relativamente recente no campo da ciência da informação, quando comparado a bibliometria e cientometria, a análise de patentes tem ganhado destaque, por meio de técnicas de patentometria, as informações contidas nestes documentos, são convertidas em indicadores da inovação tecnológica, possibilitando assim identificar comportamentos, tendências e consequentemente viabilizar o monitoramento de tecnologias e práticas emergentes.

Este trabalho abordou a perspectiva da ciência da informação, expondo os dados provenientes de patentes como uma fonte confiável e ampla no que se refere ao desenvolvimento tecnológico nacional. Os dados utilizados foram obtidos no repositório nacional de Pesquisa em Propriedade Industrial mantido pelo INPI, e no repositório internacional de patentes, a Espacenet, mantido pelo Escritório Europeu de Patentes. Adotando como

critério de seleção as patentes depositadas no INPI entre os anos de 1900 a 2021 e disponibilizadas na Espacenet. Para agregar valor ao estudo também foram coletados dados dos currículos cadastrados na Plataforma Lattes em que constasse algum registro de patente visando obter um retrato da representatividade dos dados cadastrados nos currículos bem como de sua consistência.

As bases de dados de patentes são ferramentas essenciais no processo de coleta de dados, são amplamente utilizadas na prospecção tecnológica para identificar oportunidades de exploração ou aprimoramento de tecnologias e produtos, além de serem utilizadas em atividades no meio acadêmico. No entanto, coletar e agrupar tais dados não é uma tarefa trivial, durante o desenvolvimento deste trabalho, algumas limitações foram encontradas. As opções de filtro disponibilizadas no acesso gratuito das bases de dados de patentes não permitem a elaboração de critérios mais específicos de seleção dos dados. Existe também a limitação quanto ao volume de dados trafegado, o que consequentemente inviabiliza a manipulação de grandes volumes de dados. Estes fatores provocaram a implementação de estratégias que viabilizassem a obtenção dos dados de forma a favorecer a construção de uma base de dados local que viabilizasse a manipulação e cruzamento dos dados de forma otimizada. Tal estratégia se tornou uma ótima alternativa às limitações encontradas.

A princípio, os dados coletados foram armazenados em arquivos de extensão JSON, um arquivo para cada patente. No entanto, essa forma de armazenamento ficou suscetível a redundâncias, o que demandaria muito esforço para manter a integridade dos dados. Então, foi adotado o uso do banco de dados PostgreSQL que auxiliou veemente na manipulação dos dados, desta forma as informações puderam ser recuperadas por meio de consultas SQL, o que otimizou a recuperação dos dados e consequentemente, diminuiu o custo computacional ao manipular grandes volumes de dados. O PostgreSQL armazena os dados de maneira estruturada, o que possibilitou a padronização do armazenamento das informações evitando a ocorrência de redundâncias.

A patentometria é uma ferramenta fundamental para a tomada de decisão, quando o assunto é inovação tecnológica, pois tem como principal resultado a geração de indicadores através das informações contidas nos documentos de patentes. A gestão por indicadores é um modelo gerencial estratégico que traz os dados para perto da tomada de decisões, traçando um panorama das atividades tecnológicas. Ao aplicar técnicas patentométricas nos dados coletados foi possível identificar alguns fatores que caracterizam a produção técnica brasileira, como: número total e evolução dos pedidos de patentes de forma temporal, quais empresas e instituições que lideram o número de depósito de patentes, quais áreas são as líderes de patenteamento, dentre outros.

Avaliar o número de depósito de patentes corresponde a um tópico relevante para compreensão do cenário tecnológico nacional, pois permite avaliar as taxas de crescimento do número de depósitos ao longo dos anos e também é usado para avaliar a produtividade

tecnológica do país. O Brasil apresenta aumento no número de depósitos a cada ano, entretanto ao mensurar a velocidade, este indicador aponta um lento crescimento tecnológico, quando comparado a países como a China e Estado Unidos (WIPO, 2022). Evidenciando uma carência de investimentos para o setor de ciência, tecnologia e inovação. Ao investigar os dez maiores depositantes de patentes no país, fica mais evidente essa necessidade do efetivo acompanhamento e investimento para o setor, visto que os maiores depositantes de patentes não são empresas brasileiras, são empresas multinacionais que buscam proteger suas invenções no Brasil.

As Instituições de Ensino Superior tem se apresentado como propulsores do desenvolvimento tecnológico nacional, consolidando a premissa da universidade como produtora de conhecimento patenteável. Este indicador revela as IES que possuem um ambiente propício à inovação, informação relevante para organizações que buscam parcerias com a universidade a fim de desenvolver novos produtos e serviços. Destacando que as universidades públicas federais e estaduais são os maiores depositantes ao se considerar somente as instituições de ensino e pesquisa do país, destacando a importância da educação pública de qualidade. Embora existam indicadores de depósito de patentes acadêmicas em todas as regiões do país, as regiões Sudeste e Sul se destacam das demais, as universidades UNICAMP, USP e UFMG são as universidades com maior quantitativo de patentes.

O indicador de classificação de patentes, aponta as áreas de patenteamento que mais se desenvolvem no país. Aproximadamente 25% das patentes depositadas no Brasil obtém a classificação “A-Necessidades humanas”, o que supõe uma relação com os fatores econômicos do país, como o setor agrícola (CNA, 2021).

Os indicadores sobre produção tecnológica extraídos de documentos de patentes refletem o panorama do cenário tecnológico nacional, que se caracteriza como uma excelente ferramenta que pode ser mais explorada no Brasil. Este trabalho evidenciou-se que no contexto geral a predominância dos depósitos das patentes é realizado por organizações cuja a sede destas não é no Brasil, o que implica na importância das multinacionais investirem no país. Já em relação aos depositantes nacionais, entre os maiores responsáveis estão as Universidades Públicas, desempenhando um papel importante no cenário econômico do país.

Dentro deste contexto das universidades como aspirantes protagonistas no depósito de patentes e que a maioria dos seus pesquisadores possuem currículos cadastrados na Plataforma Lattes, constatou-se que avaliar a representatividade das patentes cadastradas nos currículos em repositórios internacionais, como por exemplo nos dados extraídos da Espacenet, viabiliza a validação e consequentemente a certificação dos dados registrados nos currículos garantindo a confiabilidade das informações ali contidas. Diante do exposto, destaca-se que os pesquisadores informam em seus currículos registros de toda sua trajetória acadêmica e profissional, informações relevantes sobre a compreensão do cenário

tecnológico, o que viabiliza a identificação de atores chaves em determinadas áreas do conhecimento. Entretanto, apenas cerca de 1% de todos os currículos da base de dados da Plataforma Lattes, possuem informações sobre o depósito de patentes, base composta por mais de 7.4 milhões de currículos. Do conjunto de registros de patentes recuperados dos currículos, nem todos puderam ser validados na Espacenet devido à inconsistência nos dados registrados, notabilizando a necessidade da existência de mecanismos de validação e certificação dos dados patentários.

Em virtude dos dos resultados obtidos foi possível caracterizar a produção técnica brasileira, elencando sob diversos aspectos indicadores patentométricos. Foi possível validar a importância em adotar informações contidas em documentos de patentes como fonte de dados para análises da produção técnica. Foi possível ainda obter um retrato da produção técnica brasileira, apresentando suas principais características. Apesar de indicadores quantitativos não serem o suficiente para apresentarem o valor de uma patente, permite compreender todo o ecossistema de proteção de propriedade intelectual. Os indicadores contribuem de maneira significativa para o acompanhamento do desenvolvimento tecnológico, servindo como referência para prospecção de novas tecnologias. Por conseguinte é possível afirmar que a patentometria possui um grande valor científico e precisa ser mais explorada no cenário nacional.

6.1 Publicações

A seguir, são apresentadas as produções científicas desenvolvidas com base nos estudos realizados nesta dissertação. Foram produzidos e publicados os seguintes artigos:

- SILVA, R. R.; DIAS, T. M. R.; CARVALHO-SEGUNDO, W.. Uma estratégia para a identificação e extração de dados de patentes brasileiras. In: **18° Congresso Latino-americano de Software Livre e Tecnologias Abertas/Latin Science 2021**, Outubro 2021.
- SILVA, R. R.; DIAS, T. M. R.; CARVALHO-SEGUNDO, W.. Uma estratégia para recuperação de dados para análises sobre a produção técnica brasileira. In: **XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB**, Outubro 2021.

6.2 Trabalhos Futuros

Com o objetivo de dar continuidade aos trabalhos proposto nesta dissertação, a seguir são apresentadas algumas extensões para possíveis trabalhos futuros:

- **Desenvolvimento de Dashboard com os dados das patentes:** este estudo tem como objetivo desenvolver uma ferramenta Web, em que os pesquisadores poderão construir seus próprios indicadores com os dados provenientes da base de dados produzida por este trabalho, disponibilizando funcionalidades tais como, definição de quais dados serão considerados no indicador, e quais são os parâmetros de filtro de geração dos dados, e a definição do modo de visualização dos resultados, podendo definir como listagem ou gráficos de preferência, selecionando dentre as opções disponíveis. Possibilitando ainda, o consumo dos dados por API's.
- **Identificação de Especialistas:** tal estudo se beneficiaria da integração de dados coletados dos repositórios propostos nesta dissertação e utilizar-se de diversas informações para classificar os principais indivíduos responsáveis pela proposição de patentes. Vários desafios como desambiguação e deduplicação de dados devem ser enfrentados, e como resultado poderiam ser verificados indicadores sobre a representatividade de determinados indivíduos nas suas regiões de atuação profissional, áreas do conhecimento ou até mesmo temáticas de interesse.
- **Certificador de dados de patentes para a Plataforma Lattes:** devido a construção da base de dados proposta neste trabalho, esta viabiliza o desenvolvimento de um certificador de dados de patentes para os currículos da Plataforma Lattes, auxiliando assim, o pesquisador ao preencher estas informações em seu currículo. Tal funcionalidade poderá viabilizar que o pesquisador possa informar o número de depósito da patente que deseja registrar em seu currículo, retornando dessa forma via API informações como, título da patente, classificações, depositantes e inventores. Tal fato auxiliará o indivíduo no preenchimento de seus dados fornecendo informações precisas sobre o invento. Funcionalidade semelhante quando se preenche um artigo informando o DOI.
- **Análise de rede de colaborações:** devido ao fato de trabalhos estarem sendo realizados cada vez mais de forma colaborativa, uma análise baseada nas colaborações na proposição de patentes se torna relevante. O intuito do estudo é conseguir identificar as redes de colaboração e cooperação, bem como mensurar a eficiência dos laços constituídos. A exemplo do que ocorre com frequência no estudo sobre redes de colaboração científica, a análise sobre como ocorre a colaboração técnica poderia proporcionar um retrato inédito sobre como evoluiu a proposição de patentes de forma colaborativa, inclusive sobre a colaboração entre os depositantes (organizações/instituições de ensino).

Referências

- ALMIND, P. T. C.; INGWERSEN. Informetric analyses on the world wide web: methodological approaches to 'webometrics'. **Journal of Documentation**, v. 53, n. 4, p. 404–426, 1997. DOI <https://doi.org/10.1108/EUM0000000007205>. Citado na página 18.
- AMADEI, J. R. P.; TORKOMIAN, A. L. V. As patentes nas universidades: análise dos depósitos das universidades públicas paulistas. **Ciência da Informação**, v. 38, n. 02, p. 9–18, 2009. Citado na página 1.
- ARISTODEMOU, F. L.; TIETZE. The state-of-the-art on intellectual property analytics (ipa): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (ip) data. **World Patent Information**, v. 55, p. 37–51, 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0172219018300103>. Acesso em: 07 mai. de 2021. Citado na página 18.
- BARBOSA, D. B. **Uma Introdução À Propriedade Intelectual**. segunda. [S.l.]: Editora Lumen Juris, 2010. Disponível em: <http://s-inova.ucdb.br/wp-content/uploads/biblioteca/Uma%20Introducao%20a%20Propriedade%20Intelectual.pdf>. Acesso em: 13 abr. de 2021. Citado na página 6.
- BRANDÃO, F. G. **Democratização da informação a partir do uso de repositórios digitais institucionais : da comunicação científica às informações tecnológicas de patentes**. Dissertação (Mestrado) — Universidade Regional Integrada do Alto Uruguai e das Missões, set. 2016. Disponível em: <https://lume.ufrgs.br/handle/10183/179853>. Acesso em: 24 mar. de 2021. Citado 2 vezes nas páginas 14 e 15.
- BRASIL. Código de propriedade industrial - lei 9279/96. **Diário Oficial da União**, mai. 1996. Citado 3 vezes nas páginas 8, 9 e 10.
- BRASIL. **INPI chega aos 45 anos: conheça a história do Instituto**. Brasil, 2021. Disponível em: <https://www.gov.br/inpi/pt-br/assuntos/noticias/inpi-chega-aos-45-anos-conheca-a-historia-do-instituto>. Acesso em: 15 de dez. de 2021. Citado na página 30.
- BRITO, L. A. G. C. de; QUONIAM, J. P.; MENA-CHALCO. Exploração da plataforma lattes por assunto: proposta de metodologia. **Transinformação**, v. 28, n. 1, p. 77–86, abr. 2016. FapUNIFESP (SciELO) <http://dx.doi.org/10.1590/2318-08892016002800006>. Citado na página 57.
- CABRINI, E. F. T. de O. M. C.; GRACIO. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base scopus. **Perspectivas em Ciência da Informação**, v. 16, p. 16–28, out. 2011. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362011000400003&lng=pt&lng=pt. Acesso em: 08 mai. de 2021. Citado na página 15.
- CALZOLAIO, A. P. A. E. et al. Mapeamento dos registros de propriedade intelectual (patentes) na universidade federal do rio grande do sul. **Revista Brasileira de Gestão e Inovação**, v. 6, n. 1, p. 44–70, 2018. Disponível em: <http://www.ucs.br/etc/revistas/index.php/RBGI/article/view/5860>. Acesso em: 22 fev. de 2021. Citado na página 2.

CARVALHO, M. R. d. M. C. B. C. C. B.; SANTOS. A classificação internacional de patentes: Descrição e importância. **GEINTEC**, v. 09, n. 01, p. 4798–4808, jan. 2019. Citado na página 12.

CATIVELLI, A. S. **Patentes das Universidades Públicas Brasileiras: Análise das Concessões**. Dissertação (Dissertação do Mestrado Profissional em Gestão de Unidades de Informação) — Universidade do Estado de Santa Catarina, 2016. Citado na página 21.

CNA. **PIB do agronegócio avança no trimestre e acumula alta de 9,81% no primeiro semestre de 2021**. [S.l.], 2021. Disponível em <https://www.cnabrazil.org.br/assets/arquivos/boletins/dtec.pib_jun_2021.13set2021vff.pdf> Acesso em: 05 de jan. de 2022. Citado 2 vezes nas páginas 48 e 68.

CURTY, N. R. R. G.; DELBIANCO. As diferentes métricas dos estudos métricos da informação: evolução epistemológica, inter-relações e representações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 25, p. 01–21, 2020. Citado 4 vezes nas páginas 15, 16, 17 e 18.

DATE, C. J. **Introdução a Sistemas de Banco de Dados**. eighth. [S.l.]: Elsevier, 2003. Citado na página 41.

DIAS, T. H. J. T. M. R.; MOREIRA, P. M.; DIAS. Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do cnpq. **BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING**, n. 7, p. 64–75, 2018. Natal. Anais. Natal: Sbc. Citado na página 57.

DIAS, T. M. R. **Um Estudo da Produção Científica Brasileira a partir de Dados da Plataforma Lattes**. 181 p. Tese (Doutorado em Modelagem Matemática e Computacional) — Centro Federal de Educação Tecnológica de Minas Gerais, Setembro 2016. Citado 4 vezes nas páginas 4, 25, 36 e 62.

DIGIAMPIETRI, L. A. **Análise da Rede Social Acadêmica Brasileira**. Tese (LIVRE DOCÊNCIA - Informação e Tecnologia) — Escola de Artes Ciências e Humanidades da Universidade de São Paulo, 2015. São Paulo. Citado na página 57.

ELMASRI, S. B. R.; NAVATHE. **Sistema de Banco de Dados**. sexta. [S.l.]: São Paulo: Addison Wesley, 2011. Citado na página 41.

ESPECENET. **Espacenet patent search**. 2021. Disponível em: <<https://worldwide.espacenet.com/patent/>>. Acesso em: 13 de jan. de 2021. Citado na página 30.

FAGUNDES, P. A. d. A. M. C. et al. Perfil tecnológico da csu: Um estudo patentométrico. **Revista de Administração e Inovação (RAI)**, v. 11, n. 1, p. 276–294, mar. 2014. Citado na página 18.

FERREIRA, A. B. de H. **Dicionário Aurélio**. [S.l.]: Editora Positivo, 2010. Citado na página 6.

FRANÇA, R. O. Patente como fonte de informação tecnológica. **Perspectivas em Ciência da Informação**, v. 2, n. 2, 2007. Disponível em: <<https://periodicos.ufmg.br/index.php/pci/article/view/23241>>. Citado na página 20.

- HQIT, H. Q. I. T. C. L. **A patent database management analysis method**. 2017. Depositante: Hefei Qingzhuo Information Technology Co. Ltd. CN106407394A. Depósito: 19 set. 2016. Publicação: 15 fev. 2017. Citado na página 3.
- IBM, I. B. M. C. **Processing Large Machine Learning Datasets**. 2022. Depositante: International Business Machines Corporation. US2022075761A1. Depósito: 8 nov. 2020. Publicação: 10 fev. 2022. Citado na página 3.
- INPI. **Busca de patentes**. Brasil, 2020. Disponível em: <<https://www.gov.br/inpi/pt-br/assuntos/informacao/busca-de-patentes>>. Acesso em: 05 de jan. de 2020. Citado na página 12.
- INPI. **Guia Básico**. Brasil, 2020. Disponível em: <<https://www.gov.br/inpi/pt-br/servicos/patentes/guia-basico>>. Acesso em: 20 de abr. de 2020. Citado 2 vezes nas páginas 9 e 14.
- INPI. **Classificação de patentes**. Brasil, 2021. Disponível em: <<https://www.gov.br/inpi/pt-br/servicos/patentes/classificacao-de-patentes>>. Acesso em: 03 de mar. de 2020. Citado na página 11.
- INPI. **Convenção de Paris**. Brasil, 2021. Disponível em: <<https://www.gov.br/inpi/pt-br/backup/legislacao-1/cup.pdf>>. Acesso em: 18 de mai. de 2021. Citado na página 13.
- INPI. **Instituto Nacional da Propriedade Industrial**. Brasil, 2021. Disponível em: <<https://www.gov.br/inpi/pt-br>>. Acesso em: 20 de jan. de 2021. Citado 4 vezes nas páginas 2, 8, 9 e 45.
- INPI. **Estatísticas e Estudos Econômicos**. Brasil, 2022. Disponível em: <<https://www.gov.br/inpi/pt-br/central-de-conteudo/estatisticas/estatisticas>>. Acesso em: 08 de fev. de 2022. Citado na página 54.
- INPI. **Ranking dos maiores depositantes em 2020**. Brasil, 2022. Disponível em: <<https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/inpi-divulga-rankings-dos-maiores-depositantes-em-2020>>. Acesso em: 05 de fev. de 2022. Citado na página 52.
- JUNGSMANN, E. A. Diana de M.; BONETTI. **Inovação e Propriedade Intelectual: Guia para o Docente**. [S.l.], 2010. Citado na página 7.
- JUNIOR, M. R. A. C.; KÜHL. Análise das inovações tecnológicas aplicáveis nas ciências contábeis: Um olhar a partir da bibliometria e patentometria no período 2005-2019. **Brazilian Journal of Development**, v. 6, n. 12, p. 94465–94491, dez. 2020. DOI:10.34117/bjdv6n12-058. Citado na página 22.
- MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, n. 2, p. 134–140, mai. 1998. Disponível em: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200005#:~:text=Inicialmente%2C%20o%20trabalho%20apresenta%20as,as%20limita~oes%20dos%20indicadores%20científicos.>. Acesso em: 16 abr. de 2021. Citado 3 vezes nas páginas 17, 18 e 19.
- MADEU, P. C. F. C. B.; PELLANDA, A. C.; PASSOS. Prospecção tecnológica utilizando análise de patentes e o método ahp: estudo de caso dos materiais de carbono para o setor de defesa. **Revista Militar de Ciência e Tecnologia**, v. 1, n. 38, 2021. Disponível em: <http://rmct.ime.eb.br/arquivos/RMCT_1_tri_2021/RMCT_47419.pdf>. Citado na página 50.

MENA-CHALCO, R. M. J. P.; CESAR-JUNIOR. Scriptlattes: an open-source knowledge extraction system from the lattex platform. **Journal Of The Brazilian Computer Society**, v. 15, n. 4, p. 31–39, dez. 2009. Springer Nature America, Inc <<http://dx.doi.org/10.1007/bf03194511>>. Citado na página 57.

MENDES, D. R. A. d. M. L. S.; MELO. Avaliação tecnológica: Uma proposta metodológica. **ANPAD - Associação nacional de Pós-graduação e Pesquisa em Administração**, 2017. Disponível em: <<https://www.scielo.br/j/rac/a/Hb3HpXPBq4XZDFtLmC74RMh/?lang=pt#>>. Acesso em: 30 jun. de 2021. Citado na página 21.

MENEZES, M. S. **Estudo patentométrico relacionado à terapia de Tuberculose pulmonar**. Dissertação (Mestrado) — Fundação Oswaldo Cruz, Programa de Pós-graduação Profissional em Gestão, Pesquisa e Desenvolvimento na Indústria Farmacêutica, do Instituto de Tecnologia em Fármacos, 2019. Disponível em: <<https://www.arca.fiocruz.br/handle/icict/37031>>. Citado na página 22.

MITCHELL, R. **Web Scraping com Python: Coletando mais dados da web moderna**. segunda. [S.l.]: Novatec Editora., 2019. Citado na página 27.

MUELLER, V. S. P. M.; PERUCCHI. Universidades e a produção de patentes: tópicos de interesse para o estudioso da informação tecnológica. **Perspectivas em Ciência da Informação**, v. 19, n. 2, p. 15–36, abr. 2014. Citado 3 vezes nas páginas 1, 8 e 11.

NACKE, O. Informetría: un nuevo nombre para una nueva disciplina. definición, estado de la ciencia y principios de desarrollo. **Revista Española de Documentación Científica**, v. 6, n. 3, jan. 1983. Citado na página 17.

NARIN, F. Patents as indicators for the evaluation of industrial research output. **Scientometrics**, v. 34, n. 3, p. 489–496, jul. 1995. Citado na página 18.

NASCIMENTO, M. G. Raphael da S.; SPEZIALI. Patentometria: a utilização de dados contidos em patentes como mecanismo de análise da predominância tecnológica dos nits. **IV Encontro Internacional de Gestão, Desenvolvimento e Inovação**, nov. 2020. Citado 3 vezes nas páginas 1, 2 e 18.

NCET, N. C. E. T. **Management method for classifying patent files by using dates**. 2019. Depositante: Nanjing Chuqing Electronic Technology. CN110309384A. Depósito: 8 set. 2019. Publicação: 8 dez. 2019. Citado na página 3.

OPS. **Open Patent Services RESTful Web Services**. 1.3.16. ed. <https://www.epo.org/searching-for-patents/data/web-services/ops.html>, 2020. Citado na página 34.

PARRA, R. X. M. R.; COUTINHO, E. F. C.; PESSANO. Um breve olhar sobre a cienciometria: Origem, evolução, tendências e sua contribuição para o ensino de ciências. **Revista Contexto & Educação**, n. 107, abr. 2019. Disponível em: <<https://www.revistas.unijui.edu.br/index.php/contextoeducacao/article/view/7267>>. Acesso em: 30 mar. de 2021. Citado na página 16.

POSTGRESQL. **Documentação do PostgreSQL 8.0.0**. <http://pgdocptbr.sourceforge.net/pg80/>, 2005. Citado na página 41.

QUONIAM, C. T. L.; KNISS, M. R.; MAZIERI. A patente como objeto de pesquisa em ciências da informação e comunicação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 19, n. 39, p. 243–268, abr. 2014. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2014v19n39p243/26586>>. Acesso em: 13 abr. de 2021. Citado 2 vezes nas páginas 7 e 20.

RAMOS, R. C. **Elaboração de indicadores de patentes sobre nanotecnologia aplicada ao agronegócio**. Dissertação (Programa de Pós-Graduação em Ciência Tecnologia e Sociedade) — Universidade Federal de São Carlos, 2012. Citado na página 46.

RIBEIRO, S. P. M. J. C.; GARCIA. O estado da arte da patentometria em periódicos internacionais da ciência da informação. **Encontro Brasileiro de bibliometria e cientometria**, mai. 2014. Disponível em: <<https://www.brapci.inf.br/index.php/res/v/45400>>. Acesso em: 22 fev. de 2021. Citado na página 21.

SAESP/DIRPA/INPI, S. de Assuntos Especiais da Diretoria de P. **Manual Para o Depositante de Patentes**. [S.l.], 2015. Disponível em: <https://www.gov.br/inpi/pt-br/assuntos/arquivos-dirpa/ManualparaoDepositantedePatentes23setembro2015-versaoC_set_15.pdf>. Acesso em: 11 set. de 2021. Citado 2 vezes nas páginas 31 e 32.

SANZ-CASADO, E. **Los estudios métricos de la información y la evaluación del a actividad científica: conceptos básicos**. [S.l.]: São Paulo: ECA/USP, 2006. Material didático de curso “Os estudos métricos da informação”, ministrado no Programa de Pós-Graduação em Ciência da Informação da ECA/USP, novembro 2006. Citado na página 1.

SCARTASSINI, F. V. B. et al. Estudo patentométrico das patentes brasileiras na via pct. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (XIX ENANCIB)**, n. XIX ENANCIB, 2018. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/102164>>. Citado na página 22.

SEBRAE. **Definição de Patente**. Brasil, 2021. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/artigos/definicao-de-patente,230a634e2ca62410VgnVCM100000b272010aRCRD>>. Acesso em: 13 de jan. de 2021. Citado na página 8.

SERRANO, B. P.; JUNIOR, J. A. G. Redes de inovação: mapeamento de inventores de patentes em uma empresa do setor de cosméticos. **Revista GEPROS**, v. 09, n. 1, p. 101, jan 2014. Citado na página 25.

SILVA, M. d. L. P. José Aparecido da; BIANCHI. Cientometria: a métrica da ciência. **Paidéia (Ribeirão Preto)**, v. 11, n. 21, p. 5–10, 2001. Citado na página 15.

SILVA, M. d. L. P. José Aparecido da; BIANCHI. Cientometria: a métrica da ciência. **Paidéia (Ribeirão Preto)**, v. 11, n. 21, 2001. Disponível em: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-863X2001000200002>. Acesso em: 29 mar. de 2021. Citado na página 16.

SILVA, M. G. S. R. da; NASCIMENTO. Patentometria: uma ferramenta indispensável no estudo de desenvolvimento de tecnologias para a indústria química. **Quim. Nova**, v. 43, n. 10, p. 1538–1548, ago. 2020. DOI: <<http://dx.doi.org/10.21577/0100-4042.20170620>>. Acesso em: 08 dez. de 2021. Citado 2 vezes nas páginas 3 e 23.

SOUZA, A. A. A. S. N. et al. Mapeamento de depósitos de propriedade intelectual no inpi por discentes do programa de pós-graduação em ciência da propriedade intelectual da universidade federal de sergipe. **11TH INTERNATIONAL SYMPOSIUM ON TECHNOLOGICAL INNOVATION**, v. 11, n. 1, p. 1524–1533, apr 2021. D.O.I.: 10.51722/S2318-3403202100011344, ISSN: 2318-3403. Citado na página 23.

SÁNCHEZ, M. V. G. **Patentometría: herramienta para el análisis de oportunidades tecnológicas**. Tese (Doutorado em Gerência de Información Tecnológica) — Facultad de Economía - Universidad de La Habana, La Habana, dez. 1999. Citado na página 18.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. **Information Processing & Management**, v. 28, p. 1–3, 1992. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/030645739290087G>>. Acesso em: 16 abr. de 2021. Citado 3 vezes nas páginas 17, 18 e 19.

TANAKA, T. Y.; INUI. Preliminary study on why university researchers do not utilize patent information for their academic research in the field of science and engineering in japan. **Portland International Conference on Management of Engineering and Technology (PICMET)**, p. 1609–1618, 2016. Citado na página 2.

TAUBES, G. Measures for measure in science. **Science**, n. 260, p. 884–886, 1993. Citado na página 16.

TRIPPE, A. **Guidelines for Preparing Patent Landscape Reports**. [S.l.]: World Intellectual Property Organization (WIPO), 2015. Disponível em: <<https://www.wipo.int/publications/en/details.jsp?id=3938&plang=EN>>. Acesso em: 07 mai. de 2021. Citado 2 vezes nas páginas 18 e 19.

UECE. **INPI - Saiba mais sobre a nova numeração nos pedidos da DIRPA e da DICIG**. Universidade Federal do Ceará, 2021. Disponível em: <http://www.uece.br/nit/index.php?option=com_content&view=article&id=1654:inpi-saiba-mais-sobre-a-nova-numeracao-nos-pedidos-da-dirpa-e-da-dicig&catid=31:lista-de-noticias>. Acesso em: 11 de mai. de 2021. Citado na página 31.

VANTI, N. Os links e os estudos webométricos. **Ciência da Informação**, v. 34, n. 01, 2005. Disponível em: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652005000100009>. Acesso em: 09 abr. de 2021. Citado 2 vezes nas páginas 17 e 18.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, n. 02, p. 152–162, 2002. Disponível em: <https://www.scielo.br/scielo.php?pid=s0100-19652002000200016&script=sci_abstract&tlng=pt>. Acesso em: 09 abr. de 2021. Citado na página 16.

VELHO, L. Indicadores científicos: em busca de teoria. interciência. **Caracas**, v. 15, n. 3, p. 139–145, 1990. Citado na página 16.

WIPO. **WIPO Patent Drafting Manual**. <https://www.wipo.int/publications/en/details.jsp?id=297>, 2007. Citado na página 13.

WIPO. **Sistema de Classificação de patentes**. 2020. Disponível em: <<http://ipc.inpi.gov.br/classifications/ipc/ipcpub/?notion=scheme&version=20200101&symbol=none&menulang=pt&lang=pt&viewmode=f&fipcp=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>>. Acesso em: 15 de mar. de 2020. Citado na página 12.

WIPO. **Família de Patentes**. Brasil, 2021. Disponível em: <https://www.wipo.int/edocs/mdocs/aspac/en/wipo_ip_bkk_12/wipo_ip_bkk_12_www_238983.pdf>. Acesso em: 20 de abr. de 2021. Citado na página 13.

WIPO. **Innovation Perseveres: International Patent Filings via WIPO Continued to Grow in 2020 Despite COVID-19 Pandemic**. 2022. Disponível em: <https://www.wipo.int/pressroom/en/articles/2021/article_0002.html>. Acesso em: 11 de fev. de 2022. Citado na página 68.

ZHAO, B. Web scraping. **Springer International Publishing**, p. 1–3, mai. 2017. Disponível em: <https://www.researchgate.net/publication/317177787_Web_Scraping>. Acesso em: 07 mai. de 2021. Citado na página 27.