

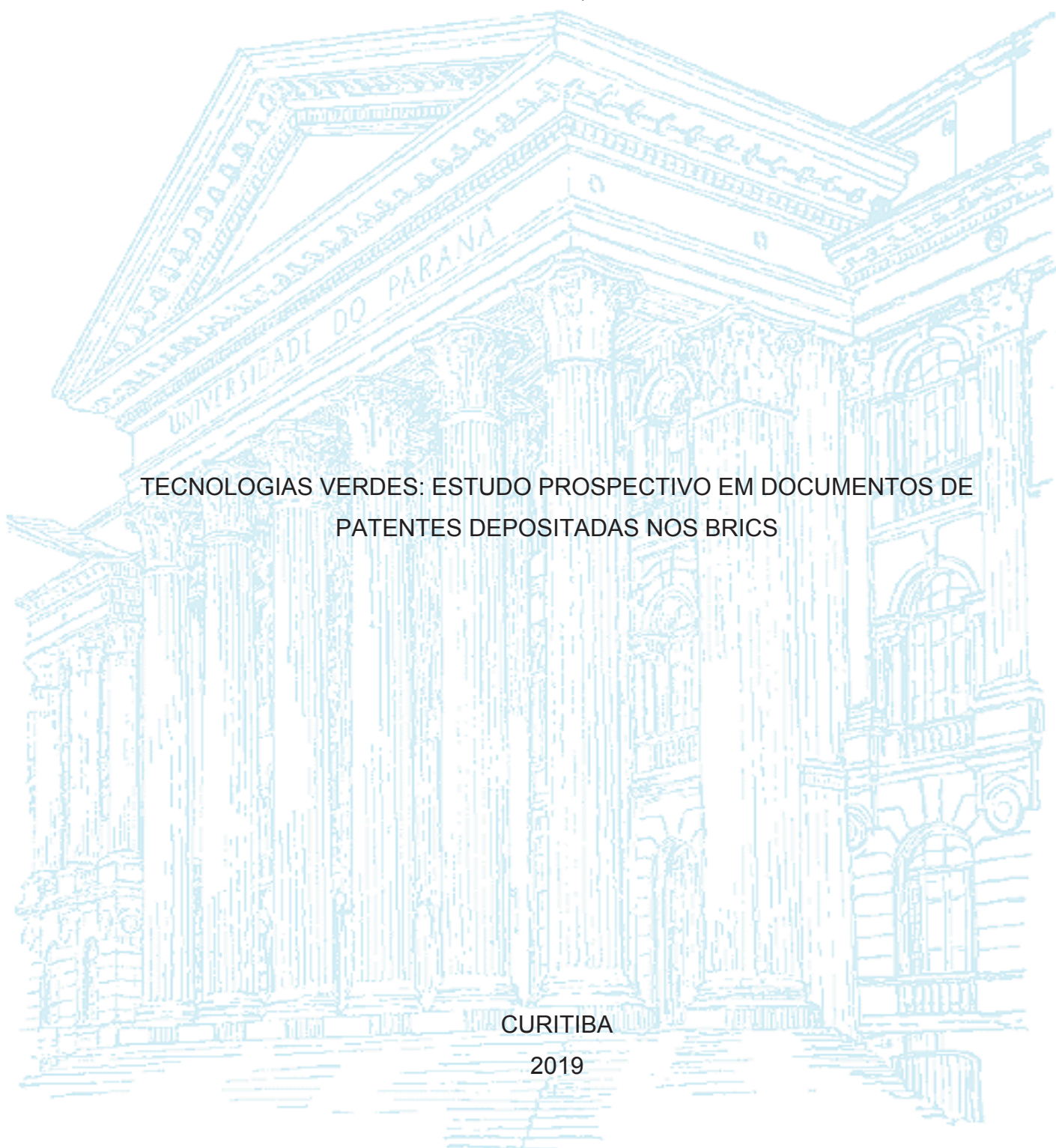
UNIVERSIDADE FEDERAL DO PARANÁ

ANDRÉ IRAZOQUI DE LIMA

TECNOLOGIAS VERDES: ESTUDO PROSPECTIVO EM DOCUMENTOS DE
PATENTES DEPOSITADAS NOS BRICS

CURITIBA

2019



ANDRÉ IRAZOQUI DE LIMA

TECNOLOGIAS VERDES: ESTUDO PROSPECTIVO EM DOCUMENTOS DE
PATENTES DEPOSITADAS NOS BRICS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia de Produção, na área de Tecnologia e Inovação em Projeto, Produtos e Processos, Setor de Tecnologia, da Universidade Federal do Paraná.

Orientadora: Prof^a. Dr^a. Izabel Cristina Zattar

CURITIBA

2019

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

L732t Lima, André Irazoqui de
Tecnologias verdes: estudo prospectivo em documentos de patentes depositadas nos BRICS [recurso eletrônico] / André Irazoqui de Lima – Curitiba, 2019.

Dissertação - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-graduação em Engenharia de produção.

Orientadora: Izabel Cristina Zattar

1. Patentes. 2. Patentes verdes (Setor automotivo). 3. Patentometria. I. Universidade Federal do Paraná. II. Zattar, Izabel Cristina. III. Título.

CDD: 608.71

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585



MINISTÉRIO DA EDUCAÇÃO
 SETOR SETOR DE TECNOLOGIA
 UNIVERSIDADE FEDERAL DO PARANÁ
 PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
 PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE
 PRODUÇÃO - 40001016070P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ANDRÉ IRAZOQUI DE LIMA** intitulada: **Tecnologias Verdes: Estudo Prospectivo em Documentos de Patentes Depositados nos BRICS**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 20 de Fevereiro de 2019.

IZABEL CRISTINA ZATTAR

Presidente da Banca Examinadora (UFPR)

DOUGLAS ALVES SANTOS

Avaliador Externo (INPI)

MARCOS AUGUSTO MENDES MARQUES

Avaliador Interno (UFPR)

MARIANA KLEINA

Avaliador Interno (UFPR)

AGRADECIMENTOS

Agradeço primeiramente a Deus por me iluminar e dar força para vencer este desafio;

Aos meus pais Eunice Marques Irazoqui de Lima e Bernardino Pereira de Lima e minha irmã Ana Teresa Irazoqui de Lima por estarem sempre ao meu lado;

À minha noiva, amiga, companheira e futura esposa Aline Brum Argenta por me apoiar e incentivar nas horas boas e ruins;

À Professora Izabel Cristina Zattar, pela orientação dada, pela experiência compartilhada e pela grande contribuição à minha formação;

A todos os professores do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Paraná. Em especial aos professores Dr. Marcos Augusto Mendes Marques, Dr^a. Mariana Kleina e Dr. Gustavo Valentim Loch pelas contribuições durante a realização desta pesquisa;

Ao Dr. Douglas Alves Santos pelo incentivo, disposição e contribuições para o desenvolvimento desta pesquisa;

À Rafaella Laureano por todo auxílio durante a realização do mestrado;

Aos meus primos, demais familiares e amigos pelo apoio;

Aos meus colegas Andreia, Anis, Carlos Brasil, Carlos Flesch, Daniel Ayub, Daniel Oliveira, Danielle, Gustavo, Gregori, Henrique, Iris, Jefferson, João Vitor, Lívia, Majed, Matheus, Natália, Nicolas, Paulo, Ricardo, Rodrigo, Tarley, Thiago, Thiago e Vitor por todo aprendizado, amizade e dedicação ao longo deste período;

E a todos os professores que tive ao longo da vida, sem a contribuição deles nada seria possível.

RESUMO

Com a crescente preocupação relacionada aos impactos ambientais, causados pelo crescimento populacional, juntamente com o aumento da atividade industrial ao redor do mundo, países como Brasil e China tem criado políticas de incentivo a produção de veículos elétricos e híbridos. Estas políticas, além de buscarem a mitigação dos danos causados ao meio ambiente, também irão causar fortes impactos no setor energético, alterando sua demanda a médio e longo prazo. Desta forma, por uma questão de sobrevivência no mercado e adequação a regulações e legislações vigentes, as empresas estão constantemente promovendo inovações tanto em seus produtos quanto em seus processos. No âmbito da propriedade intelectual, diversos países criaram os chamados programas de Patentes Verdes, os quais buscam acelerar o exame dos pedidos de patentes das tecnologias consideradas verdes (ambientalmente amigáveis), dentre estes países está o Brasil. Neste contexto, este trabalho tem como objetivo comparar os movimentos de inovação tecnológica nos BRICS, tomando como base a categoria de transportes do programa de Patentes Verdes brasileiro. Para isto, foi utilizada a análise patentométrica em documentos oriundos da base *Derwent Innovations Index*. Como principais resultados, observou-se que a China possui um volume de depósitos muito superior aos demais países do bloco, por este motivo foi retirada das análises estatísticas. Brasil, Rússia e Índia se alternam nas posições subsequentes, ficando a África do Sul com os menores indicadores em todas subcategorias. O número de depósitos efetuados nos códigos IPC relacionados aos veículos elétricos e híbridos são semelhantes. Porém, vale salientar a alta intensidade das tecnologias depositadas para estações de recarga de veículos elétricos. Em relação aos principais depositantes, Toyota, Ford, Nissan e Qualcomm se destacam.

Palavras-chave: Patentes. Patentes Verdes. Transportes. Setor automotivo. Patentometria. Mineração de texto. Programação em R.

ABSTRACT

With increasing concern about the environmental impacts caused by population growth coupled with increased industrial activity around the world, countries like Brazil and China have created policies to encourage the production of electric and hybrid vehicles. These policies, in addition to seeking to mitigate the damages caused to the environment, will also cause strong impacts in the energy sector, changing their demand in the medium and long term. Thus, due to market survival and compliance with current regulations and legislation, companies are constantly promoting innovations both in their products and in their processes. In the field of intellectual property, several countries have created so-called Green Patents programs, which seek to accelerate the examination of patent applications for technologies considered green (environmentally friendly), among these countries is Brazil. In this context, this work aims to compare the movements of technological innovation in BRICS, based on the transport category of the Brazilian Green Patents program. For this, the patentometric analysis was used in documents originating from the Derwent Innovations Index base. As main results, it was observed that China has a volume of deposits much higher than the other countries of the bloc, for this reason was removed from the statistical analyzes. Brazil, Russia and India alternate in the subsequent positions, leaving South Africa with the lowest indicators in all subcategories. The number of deposits made in the IPC codes related to electric and hybrid vehicles are similar. However, it is worth noting the high intensity of the technologies deposited for charging stations of electric vehicles. In relation to the main depositors, Toyota, Ford, Nissan and Qualcomm stand out.

Keywords: Patents. Green Patents. Transport. Automotive industry. Patentometry. Text Mining. Programming in R.

LISTA DE FIGURAS

FIGURA 1 - ÁREAS DA PROPRIEDADE INTELECTUAL	21
FIGURA 2 - VOLUME DE DEPÓSITOS DE PATENTES	24
FIGURA 3 - ORIGEM DOS PAÍSES DEPOSITANTES (em valores percentuais).....	25
FIGURA 4 - ORIGEM DOS DEPOSITANTES NACIONAIS	26
FIGURA 5 – CATEGORIAS DO PROGRAMA DE PATENTES VERDES.....	28
FIGURA 6 - ETAPAS DA PATENTOMETRIA	31
FIGURA 7 - FLUXO DO KDD.....	33
FIGURA 8 - COMPARAÇÃO DAS ESTRATÉGIAS DE BUSCA.....	36
FIGURA 9 - ABORDAGEM DE HOFMANN et al. 2019.....	37
FIGURA 10 - PASSOS MINERAÇÃO DE TEXTO	37
FIGURA 11 - PROTOCOLO DA PESQUISA.....	41
FIGURA 12 – PERCENTUAL DE PARTICIPAÇÃO POR SUBCATEGORIA	45
FIGURA 13 – EXEMPLO CLASSIFICAÇÃO CÓDIGO IPC.....	47
FIGURA 14 - PACOTES UTILIZADOS NA MINERAÇÃO DE TEXTO	49
FIGURA 15 - IMPORTAR DADOS PARA O R.....	50
FIGURA 16 - LIMPEZA DOS DADOS.....	50
FIGURA 17 - LISTA DAS <i>STOP WORDS</i> DO PACOTE 'TM'	51
FIGURA 18 - ENTRADA E SAÍDA DA ETAPA DE ORGANIZAÇÃO DOS DADOS..	51
FIGURA 19 - CÓDIGO, GRÁFICO EM BARRAS E NUVEM DE PALAVRAS.....	52
FIGURA 20 - MATRIZ DOS TERMOS DO DOCUMENTO.....	53
FIGURA 21 - FILTRO REMOVEPARSETERMS	54
FIGURA 22 - CONSULTA FRACIONADA UTILIZANDO FUNÇÃO ' <i>INSPECT</i> '	54
FIGURA 23 - CONSULTA DE TERMOS ESPECÍFICOS.....	55
FIGURA 24 - CÁLCULO DA MATRIZ DE FREQUÊNCIA RELATIVA.....	56
FIGURA 25 - TESTE DE NORMALIDADE BRICS	58
FIGURA 26 - TESTE DE NORMALIDADE BRASIL, RÚSSIA, ÍNDIA E ÁFRICA DO SUL	59
FIGURA 27 - TESTE DE BARTLETT	60
FIGURA 28 – ANOVA	60
FIGURA 29 - RESULTADO ANOVA	60
FIGURA 30 - BOXPLOT TESTE DE TUKEY	61

FIGURA 31 - LISTA DE CÓDIGOS IPC FORD NA RÚSSIA.....	68
FIGURA 32 - TERMOS MAIS FREQUENTES BRASIL	69
FIGURA 33 - TERMOS MAIS FREQUENTES PATENTES TETRÁDICAS.....	69
FIGURA 34 - NUVEM DE PALAVRAS	70
FIGURA 35 - MATRIZ FREQUENCIA ABSOLUTA DOS TERMOS	71
FIGURA 36 - MATRIZ FREQUENCIA RELATIVA DOS TERMOS.....	71
FIGURA 37 - NUVEM DE PALAVRAS VEÍCULOS EM GERAL	72
FIGURA 38 - EXEMPLO ESTAÇÃO DE RECARGA.....	73
FIGURA 39 - ESTAÇÃO DE RECARGA NISSAN.....	73
FIGURA 40 - ESTAÇÃO RECARGA TOYOTA	74
FIGURA 41 - SISTEMA PARA ALINHAMENTO DO VEÍCULO	75
FIGURA 42 - TRIANGULO DO LÍTIO	76

LISTA DE QUADROS

QUADRO 1 - ENQUADRAMENTO METODOLÓGICO.....	39
QUADRO 2 - OBJETIVOS ESPECÍFICOS E FERRAMENTAS.....	40
QUADRO 3 - IPC'S DE MAIOR FREQUÊNCIA NA AMOSTRA.....	63
QUADRO 4 - PERCENTUAL DOS DEPÓSITOS POR CÓDIGO IPC.....	63
QUADRO 5 - NÚMERO TOTAL DE PALAVRAS POR ANO.....	70

LISTA DE TABELAS

TABELA 2 - ESTRATÉGIA DE BUSCA	47
TABELA 3 - PATENTES DEPOSITADAS (VEÍCULOS EM GERAL)	58
TABELA 4 - FAMÍLIAS DE PATENTES	62
TABELA 5 - DEPÓSITOS PARA ESTAÇÕES DE RECARGA DE VEÍCULOS ELÉTRICOS	64
TABELA 6 - PROPULSÃO ELÉTRICA COM FONTE DE ENERGIA NO EXTERIOR DO VEÍCULO ATRAVÉS DE CÉLULAS COMBUSTÍVEL	65
TABELA 7 - COMÉRCIO INTERNACIONAL DE LÍTIO	77

LISTA DE SIGLAS

BRIC	– Conjunto de países formado por Brasil, Rússia, Índia e China
BRICS do Sul	– Conjunto de países formado por Brasil, Rússia, Índia, China e África
CO ₂	– Dióxido de Carbono
INPI	– Instituto Nacional da Propriedade Industrial
IPC	– <i>International Patent Classification</i>
IPCC	– <i>Intergovernmental Panel on Climate Change</i>
LEV	– <i>Low Emission Vehicle</i>
NO _x	– Óxido de Nitrogênio
OMPI	– Organização Mundial da Propriedade Intelectual
ONU	– Organização das Nações Unidas
P&D	– Pesquisa e Desenvolvimento
PCT	– <i>Patent Cooperation Treaty</i>
PI	– Propriedade Intelectual
PIB	– Produto Interno Bruto
RSL	– Revisão Sistemática da Literatura

SUMÁRIO

1	INTRODUÇÃO	15
1.1	JUSTIFICATIVA.....	17
1.2	OBJETIVOS	18
1.2.1	Objetivos Específicos	18
1.3	DELIMITAÇÃO DO TRABALHO.....	19
2	REVISÃO DE LITERATURA.....	20
2.1	PROPRIEDADE INTELECTUAL	20
2.2	PATENTES.....	22
2.3	PATENTES VERDES	26
2.4	PATENTOMETRIA	29
2.5	DATA MINING	32
2.6	TRABALHOS CORRELATOS	34
3	METODOLOGIA.....	39
3.1	CLASSIFICAÇÃO DA PESQUISA.....	39
3.2	PROTOCOLO DE PESQUISA	40
4	DESENVOLVIMENTO.....	44
4.1	DESENVOLVIMENTO DO ALGORITMO EM R	44
4.2	ESCOLHA DAS CATEGORIAS E BASE PATENTÁRIA	45
4.3	DEFINIÇÃO DA ESTRATÉGIA DE BUSCA	46
4.4	SELEÇÃO DOS INDICADORES	48
4.5	EXTRAÇÃO DOS DADOS DA BASE PATENTÁRIA.....	48
4.6	LIMPEZA, CLASSIFICAÇÃO E ORGANIZAÇÃO DOS DADOS.....	49
5	APRESENTAÇÃO DOS RESULTADOS	57
5.1	INDICADORES DE ATIVIDADE	57
5.1.1	Volume de depósitos	57
5.1.2	Códigos IPC	62
5.1.3	Principais depositantes.....	66
5.2	INDICADORES RELACIONAIS DE SEGUNDA GERAÇÃO	68
6	CONCLUSÃO.....	78
	REFERÊNCIAS.....	81
	APÊNDICE I - ALGORITMO.....	88
	APÊNDICE II – INDICADORES DE ATIVIDADE.....	97

ANEXO I – RESOLUÇÃO Nº 175/2016	100
ANEXO II – SOFTWARES <i>OPEN SOURCE</i> PARA MINERAÇÃO DE TEXTO	106
ANEXO III – CÓDIGOS IPC CATEGORIA TRANSPORTE.....	107
ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES.....	108
ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES.....	109
ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES.....	110
ANEXO V – PRINCIPAIS DEPOSITANTES	111

1 INTRODUÇÃO

O desenvolvimento sustentável e a disseminação da utilização de tecnologias verdes vêm ganhando força em todos os setores da economia. A velocidade do avanço tecnológico tem causado profundas transformações na sociedade, proporcionando um aumento na expectativa de vida, além de melhoras substanciais nos indicadores de desenvolvimento tanto econômicos quanto sociais. Entretanto, para Raiser et al. (2017), o crescimento desenfreado trouxe reflexos para o meio ambiente, pois quanto maior o consumo da sociedade maior será o impacto gerado na natureza.

Atualmente há uma crescente preocupação com os impactos ambientais causados pela manufatura. Para Cheng (2016), a rápida industrialização e urbanização aumentam acentuadamente o consumo de energia e produtos, resultando em altos índices de poluição. Ali et al. (2016) ressaltam que nas últimas décadas houve um aumento no interesse de companhias e indústrias ao redor do mundo em desenvolver iniciativas sustentáveis.

Durante a Conferência das Nações Unidas sobre Desenvolvimento Sustentável realizada na cidade do Rio de Janeiro em 2012, também conhecida como Rio+20, diversas nações renovaram seus compromissos com o desenvolvimento sustentável e a erradicação da pobreza (ONU, 2017a). Estes eventos e ações governamentais, que contam com a participação das principais potências mundiais, reforçam a importância de se estimular a criação de tecnologias verdes de forma a não agredir o meio ambiente.

Dentre estas iniciativas há a preocupação com o desenvolvimento de produtos e processos que ao longo de seu ciclo de vida não tragam prejuízos ao meio ambiente. Neste sentido, ganham cada vez mais espaço as tecnologias verdes, também denominadas de ambientalmente amigáveis. Segundo Morais (2014), estas elevam a importância das questões ambientais no direcionamento do desenvolvimento tecnológico, onde estão concentradas as atividades de P&D (Pesquisa & Desenvolvimento).

No âmbito governamental, alguns órgãos desenvolveram mecanismos para mitigar os danos causados à natureza. Estas ações vão desde o estímulo à inovação em determinados campos tecnológicos até a coibição da atividade em outros.

Em meio a as ações de estímulo a inovação, Zeng et al. (2017) analisaram como os BRICS (acrônimo para Brasil, Rússia, Índia, China e África do Sul) buscam desenvolver as tecnologias de energia renovável. Os autores identificaram diversos modelos de financiamento (públicos, privados, internacionais), porém o acesso de pequenas e médias empresas a eles é bastante escasso. O estudo mostra que, apesar dos esforços, as políticas governamentais têm se mostrado imperfeitas no alcance dos objetivos.

Além da criação de linhas de financiamento, um dos mecanismos criados pelo governo brasileiro, através do Instituto Nacional da Propriedade Industrial (INPI), foi o programa de Patentes Verdes. O programa brasileiro, criado em 2012, busca acelerar o exame dos pedidos de patentes visando fomentar a criação de produtos e processos que mitiguem os danos causados ao meio ambiente. Este programa é dividido em cinco categorias: energia alternativa, conservação de energia, transportes, gerenciamento de resíduos e agricultura sustentável. (INPI, 2018).

Um dos setores que passaram por profundas transformações e está sendo contemplado pelo programa de Patentes Verdes é o automobilístico. Representando cerca de 4% do PIB nacional, o setor vem recebendo incentivos para produção e venda de modelos elétricos e híbridos. (BRASIL, 2018). Tal mudança no panorama também afetará o consumo de energia. Portanto identificar e compreender os movimentos de inovação nas tecnologias ambientalmente amigáveis se tornará fator chave nas tomadas de decisões de diversas organizações.

Uma das formas de se identificar estes movimentos de inovação é através da análise de patentes. Segundo Lee et al. (2015), o resultado desta análise pode ser representado de diversas formas, através de gráficos, tabelas e redes. Para os autores, recentemente houve um crescimento no interesse da aplicação de técnicas para mineração de dados integradas à estas análises.

Devido ao crescente número de documentos digitais disponíveis a cada ano, se faz necessária a aplicação de técnicas de mineração de dados e textos. Logo, pesquisadores e profissionais devem utilizar ferramentas (*softwares*) disponíveis para analisar o conteúdo de documentos em larga escala. (KAYSER e BLIND, 2017).

Tendo em vista o contexto apresentado, a questão de pesquisa que o presente estudo buscou responder através de um estudo aplicado pode ser definida

como: A aplicação da mineração de texto em documentos de patentes, com a utilização de um algoritmo desenvolvido em *software* de código aberto, auxilia na identificação das tecnologias depositadas?

1.1 JUSTIFICATIVA

A presente pesquisa tem como foco comparar os depósitos de patentes no Brasil e nos demais países pertencentes aos BRICS. Nos últimos anos este bloco têm criado iniciativas para o desenvolvimento das tecnologias ambientalmente amigáveis. Além de possuírem indicadores socioeconômicos semelhantes, estas nações desempenham um papel fundamental no desenvolvimento de suas respectivas regiões. Portanto, alterações nos seus mercados internos podem repercutir nos países que fazem fronteira/negócios com elas.

Recentemente, diversos países como Brasil e China, adotaram políticas para incentivar o número de veículos elétricos em sua frota. Estas políticas objetivam reduzir a emissão de poluentes na atmosfera melhorando a qualidade do ar, especialmente em áreas urbanas. Assumindo que veículos elétricos substituíssem todos outros, a estimativa de redução de óxidos de nitrogênio (NO_x) e dióxido de carbono (CO_2) seria de aproximadamente 25,7% e 14,4%, respectivamente. Assim, os impactos gerados, principalmente em zonas com alto fluxo de veículos, seriam mitigados. (FERRERO et al., 2016).

Cada vez mais as tecnologias de mitigação da mudança climática cobrem uma gama maior de opções, indo desde o aumento da eficiência energética dos processos de produção atuais até o uso de dióxido de carbono capturado para fertilizar algas na produção de biomassa. Dentro destas opções, no entanto, quatro grupos distintos podem ser diferenciados: melhorias de eficiência energética, armazenamento de captura de carbono, energia nuclear e energia renovável. (KRIEGLER et al., 2014). Além desses grupos, a transformação do uso de energia, como o desenvolvimento de automóveis elétricos e híbridos, de outros meios de transporte, também pode contribuir para um portfólio de tecnologia de mitigação geral. (IPCC, 2007). Neste sentido, além da contribuição ao meio ambiente, o setor automobilístico possui forte relevância na economia brasileira e, antecipar suas mudanças pode ser crucial em toda a cadeia de suprimentos.

Uma das formas de se antecipar tendências é através da análise de patentes. Todavia, o elevado número de depósitos em determinadas áreas (Ex: automobilística) dificulta o processo de leitura individual dos documentos, tornando parte do desafio a busca por *softwares* que devem ser capazes de realizar a leitura e organizar os dados. Além disto, devido aos inúmeros campos e focos de cada pesquisa, em alguns casos se faz necessário utilizar mais de uma ferramenta. Portanto, ferramentas elaboradas em programação cujo código é aberto possibilita ao usuário fazer as alterações pertinentes, tornando-o assim capaz de atender aos anseios da pesquisa em sua plenitude.

Sendo assim, a presente pesquisa visa contribuir com o conhecimento científico em três áreas, são elas: acadêmica, engenharia de produção e indústria.

Para a academia, anualmente são investidos inúmeros recursos, tanto financeiros quanto intelectuais, no desenvolvimento de tecnologias que já nascem obsoletas. Identificar o que as principais instituições e empresas estão desenvolvendo contribuirá com pesquisas futuras auxiliando no processo de inovação tecnológica do país. Assim, a disponibilização de um algoritmo capaz de minerar dados e textos, em documentos de patentes, auxiliará pesquisadores em suas jornadas.

No âmbito da engenharia de produção e da indústria, a pesquisa se mostra relevante, pois o engenheiro de produção é um profissional que lida diretamente com questões estratégicas dentro das empresas. Antecipar as tendências tecnológicas pode trazer um diferencial competitivo na busca pró-ativa por matérias-primas, desenvolvimento de fornecedores, além de outros fatores como posicionamento da marca no mercado.

1.2 OBJETIVOS

Desenvolver um algoritmo, em linguagem de programação *open source*, capaz de minerar dados e textos em documentos de patentes, a fim de automatizar o processo de análise neste tipo de documento.

1.2.1 Objetivos Específicos

Para alcançar o objetivo geral deste trabalho foram elencados os objetivos específicos a seguir:

- a) Identificar as linguagens e *softwares open source* utilizados atualmente no processo de mineração de texto;
- b) Desenvolver algoritmo para mineração de texto a fim de comparar o conteúdo dos documentos depositados no Brasil e nas patentes tetrádicas.
- c) Recuperar informações tecnológicas, entre elas o volume de depósitos, empresas líderes, códigos IPC e frequência dos termos, através de estudo prospectivo em base patentária na categoria de transportes;
- d) Traçar a tendência tecnológica das publicações na área de transportes, contribuindo com o mapeamento das tecnologias desenvolvidas;
- e) Identificar as potenciais tecnologias verdes que estarão disponíveis na área de transportes.

1.3 DELIMITAÇÃO DO TRABALHO

Esta pesquisa foi desenvolvida com base em dados secundários, oriundos da base patentária *Derwent Innovations Index*®. Quanto a delimitação geográfica, os países analisados pertencem ao bloco econômico dos BRICS, composto por Brasil, Rússia, Índia, China e África do Sul. A comparação da análise de conteúdo das patentes se limitou às patentes depositadas nos escritórios norte americano, europeu, japonês e chinês (tetrádicas). Esta pesquisa analisou os dados da subcategoria veículos em geral, pois é a mais relevante, com 60% do total de depósitos, dentre as cinco subcategorias existentes dentro da categoria de transportes. Os indicadores relacionais de primeira geração e das famílias de patentes da patentometria não foram gerados, pois não fazem parte do escopo da pesquisa.

Embora sejam encontrados na literatura diversos sinônimos para as tecnologias ambientalmente amigáveis, neste trabalho será adotado o termo tecnologias verdes.

2 REVISÃO DE LITERATURA

Neste capítulo serão abordados os conceitos relevantes para contextualização da presente pesquisa. Na primeira parte será feita a apresentação dos principais aspectos da propriedade intelectual, logo após serão descritos os fundamentos relacionados às patentes. Na sequência será apresentado o programa de Patentes Verdes brasileiro, que serviu como base para indicar os códigos IPC utilizados neste trabalho. Por fim, será feita uma breve descrição do bloco político-econômico BRICS.

2.1 PROPRIEDADE INTELECTUAL

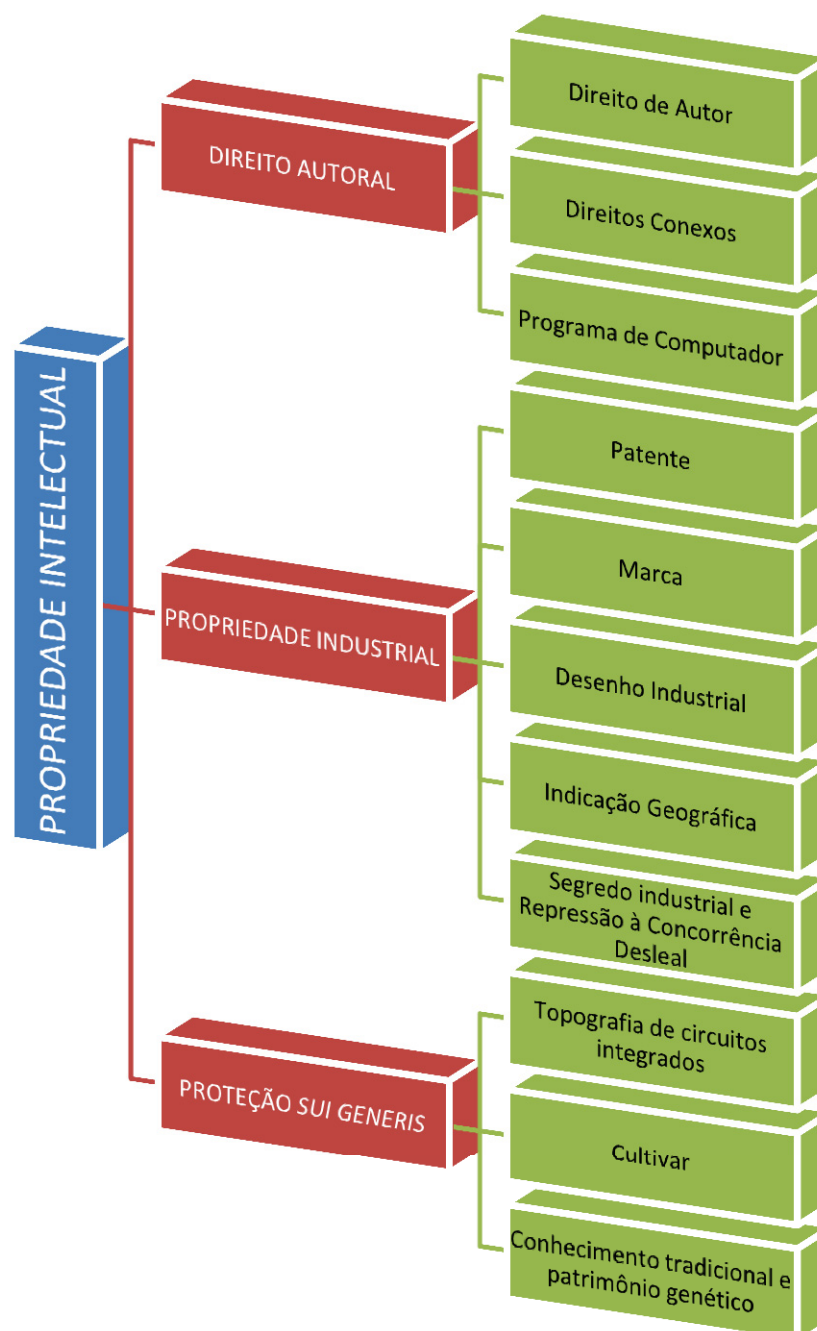
A criatividade humana se manifesta de diversas formas, sejam produções culturais, artísticas, científicas ou qualquer outra forma de expressão. (SILVA, 2016). No campo jurídico, a Constituição Federal protege a Propriedade Intelectual em seu Art. 5º, onde assegura:

“aos autores de inventos industriais privilégio temporário para sua utilização, bem como proteção às criações industriais, à propriedade das marcas, aos nomes de empresas e a outros signos distintivos, tendo em vista o interesse social e o desenvolvimento tecnológico e econômico do País”. (BRASIL. Constituição, 1988).

Severi (2013) define Propriedade Intelectual (PI) como um ramo do direito onde são abrigados os direitos autorais, as propriedades industriais, os programas de computadores, além de outras operações chamadas *sui generis* (cultivares, conhecimentos tradicionais e topografias de circuitos fechados). Portanto, a PI dá aos inventores o direito legal para que possam utilizar de forma comercial as suas criações. O autor ainda ressalta que estes direitos estão baseados nas leis de direitos autorais (Lei nº. 9.610/1998), na lei do *Software* (Lei nº. 9.609/1998) e na lei de Propriedade Industrial (Lei nº. 9.279/1996).

A propriedade intelectual é dividida em três grandes áreas, sendo elas o Direito Autoral, a Propriedade Industrial e a Proteção *sui generis* (SILVA, 2016). A Figura 1 ilustra as áreas e suas respectivas divisões.

FIGURA 1 - ÁREAS DA PROPRIEDADE INTELECTUAL



FONTE: adaptado de Vanin (2017).

No Brasil, a lei que trata do direito autoral é a de n°. 9.610¹ tendo sua publicação feita em 19 de Fevereiro de 1998. Esta lei altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências (BRASIL, 1998).

¹ BRASIL, Lei n° 9.610, de 19 de Fevereiro de 1998. Disponível em http://www.planalto.gov.br/ccivil_03/leis/L9610.htm. Acesso em: 05 nov 2017.

Conhecida também como Lei de Direito Autoral, ela protege o criador de obras artísticas, literárias e científicas. Os chamados direitos conexos visam proteger às empresas de radiodifusão, os intérpretes e os produtores fonográficos. Vale salientar que o direito autoral também garante proteção aos desenvolvedores de programa de computador, porém, para estes também há um dispositivo legal traduzido na Lei nº. 9.609/1998 denominada de Lei do *Software*.

No território nacional, a lei que rege os direitos e obrigações relativos à propriedade industrial é a de número 9.279/1996. Em seu artigo segundo ela deixa claro que o desenvolvimento tecnológico e econômico do país se efetua mediante a concessão de: patentes de invenção e de modelos de utilidade; de registro de marca e; de registro de desenho industrial, além da repressão às falsas indicações geográficas e repressão à concorrência desleal.

O ramo da proteção *sui generis* envolve o conhecimento tradicional que é composto pelos saberes empíricos, práticas, costumes e crenças passados de pais para filhos. Também faz parte desta área o cultivar que é a nomenclatura dada às novas variedades de plantas não encontradas na natureza. Por fim, a topografia de circuito integrado também está classificada dentro da proteção *sui generis*. (JUNGMANN, 2010).

Há diversas organizações que tratam dos temas relacionados a propriedade intelectual no mundo, dentre elas estão: a Organização Mundial da Propriedade Intelectual (OMPI), *Agreement of Trade-Related Aspects of Intellectual Property Rights* (TRIPS) e o *Patent Cooperation Treaty* (PCT). (CARLI, 2015).

No Brasil, o órgão responsável pelo sistema de propriedade intelectual é o Instituto Nacional da Propriedade Industrial (INPI). Este órgão é uma autarquia federal vinculada ao Ministério da Indústria, Comércio Exterior e Serviços cujo objetivo é aperfeiçoar, disseminar e gerir o sistema brasileiro de concessão e garantia de direitos de propriedade intelectual para a indústria. (INPI, 2018a).

2.2 PATENTES

A lei de propriedade industrial (LPI) (Lei nº 9.279/1996) classifica as patentes em dois tipos: a Patente de Invenção cuja validade é de 20 anos e a patente de Modelo de Utilidade válida por 15 anos. Silva et al. (2016) salienta que estes prazos têm início a partir da data do depósito. As patentes de invenção, que

podem ser tanto um produto quanto um processo, devem trazer uma nova solução para um problema existente, atendendo assim aos requisitos de novidade, atividade inventiva e aplicação industrial. Já um modelo de utilidade, refere-se a um objeto de uso prático, ele deve apresentar um novo design ou disposição, ato inventivo, resultando em uma melhoria funcional do seu uso ou em sua fabricação. (BRASIL, 1996).

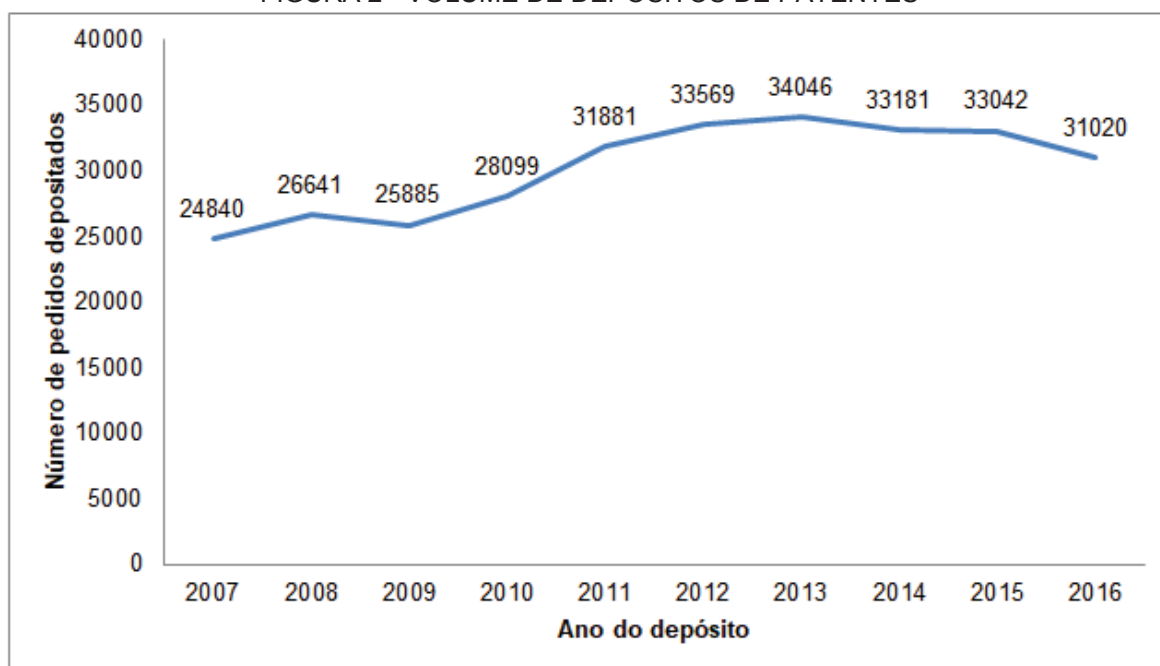
Para o INPI (2017a) a concessão de patentes promove novas tecnologias e inovações, disponibilizando-as com segurança, evitando que as mesmas permaneçam em segredo. Já a OMPI (2017) ressalta que o documento de patente garante ao inventor o direito de explorar a sua invenção, por si mesmo ou por terceiros. Entretanto, Russo et al. (2012) alerta que embora a patente traga garantias ao inventor ela ainda é um direito territorial, sendo sua abrangência limitada ao território do país cuja proteção foi concedida.

Vale destacar que embora o documento de patente tenha validade apenas no território em que a mesma foi concedida, o Brasil é signatário de alguns tratados internacionais sobre patentes. Dentre eles podemos destacar o Tratado de Cooperação em Matéria de Patentes (*Patent Cooperation Treaty – PCT*) desde abril de 1978. Este tratado multilateral permite que o requerente solicite a proteção patentária de sua invenção, simultaneamente, em um grande número de países através de um único pedido internacional de patente. (INPI, 2017b).

Ainda no âmbito dos tratados internacionais, o Brasil também é signatário do Acordo de Strasbourg desde outubro de 1975. Este acordo estabeleceu a Classificação Internacional de Patentes, também conhecida pela sua sigla em inglês IPC (*International Patent Classification*). O IPC prevê através de um sistema hierárquico de símbolos a classificação da patente de acordo com as diferentes áreas tecnológicas a que pertencem. (INPI, 2017c).

Em seu relatório de indicadores de propriedade intelectual de 2017, o INPI ressalta os 31020 pedidos de patentes recebidos no ano anterior. Dentre estes pedidos, cerca de 90% foram patentes de invenção. Embora o volume de pedidos seja expressivo, ele vem caindo desde 2014 conforme ilustra a Figura 2.

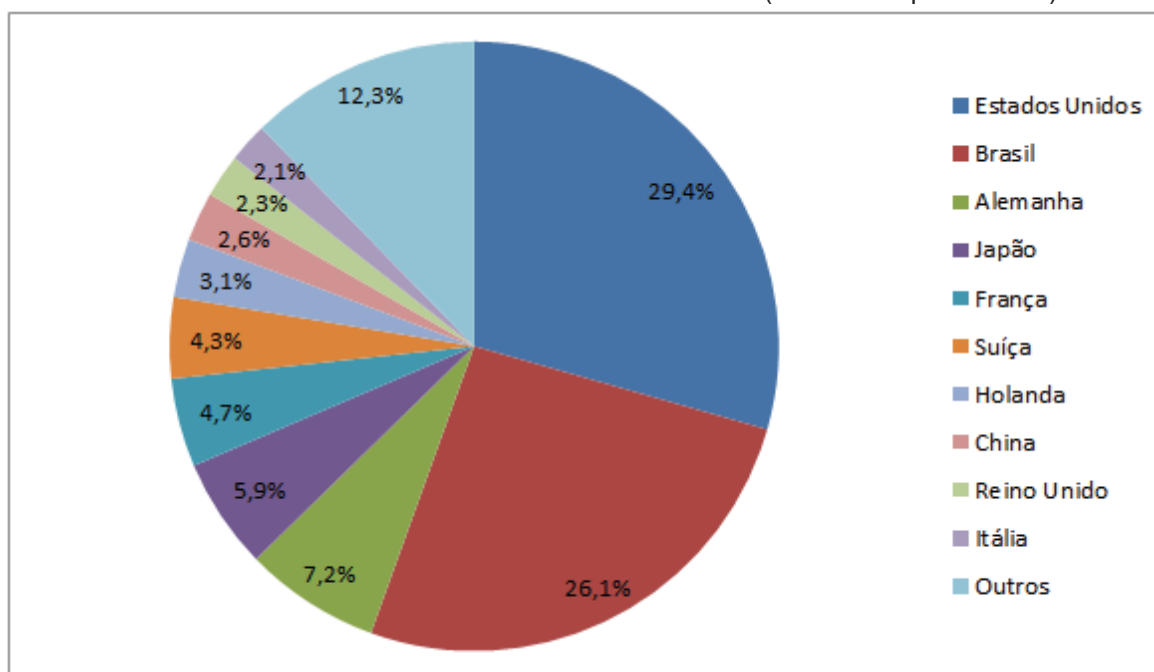
FIGURA 2 - VOLUME DE DEPÓSITOS DE PATENTES



FONTE: INPI, Assessoria de Assuntos econômicos, BADEPI v4.0 (2017).

Quanto à origem dos países depositantes em 2016, os residentes dos Estados Unidos da América apresentaram um número maior que os próprios depositantes nacionais. Vale destacar também a participação da Alemanha, do Japão, da França e da Suíça que juntos representam 22,1% dos depósitos, segundo mostra a Figura 3.

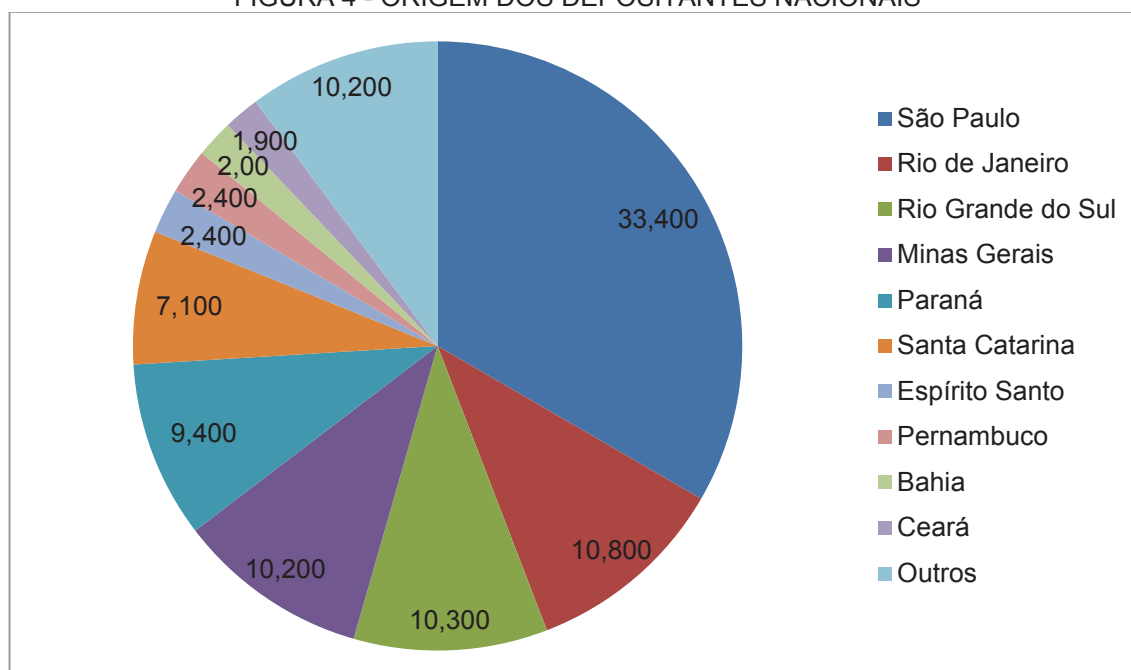
FIGURA 3 - ORIGEM DOS PAÍSES DEPOSITANTES (em valores percentuais)



FONTE: INPI, Assessoria de Assuntos econômicos, BADEPI v4.0 (2017).

Em relação aos depósitos nacionais, os estados das regiões Sudeste e Sul são os responsáveis pelos maiores volumes de depósitos. O estado de São Paulo, com sua economia e indústria bastante desenvolvidas, se destaca perante aos demais sendo responsável por 33% dos depósitos. Os três estados da região Sul também possuem forte relevância nacional sendo responsáveis por aproximadamente 26% dos pedidos, como mostra a Figura 4.

FIGURA 4 - ORIGEM DOS DEPOSITANTES NACIONAIS



FONTE: INPI, Assessoria de Assuntos econômicos, BADEPI v4.0 (2017).

Embora tenha um volume considerável de depósitos de pedidos de patente anualmente, o Brasil possuía um *backlog* de aproximadamente 240 mil pedidos pendentes no ano de 2015. (INPI, 2017d). Segundo Garcez Júnior e Moreira (2017), o tempo para concessão de um pedido de patente no Brasil saltou 4 anos em uma década, chegando no patamar de 10,8 anos em 2013.

Entretanto, com o objetivo de acelerar o prazo para concessão, foram criados alguns programas prioritários, dentre eles se destaca o programa de patentes verdes, o qual será apresentado na seção 2.3.

2.3 PATENTES VERDES

O desenvolvimento sustentável e a conscientização sobre as questões ambientais vêm ganhando maior importância nos últimos anos. Segundo relatório publicado pela Comissão Mundial sobre o Meio Ambiente e Desenvolvimento das Nações Unidas (ONU, 1987), o termo desenvolvimento sustentável significa "aquele que atende às necessidades do presente sem comprometer a capacidade das gerações futuras de atenderem às suas próprias necessidades".

Neste contexto, a preocupação com desenvolvimento de tecnologias verdes ganha cada vez mais espaço, pois está alinhada aos preceitos das questões

ambientais. Em busca do desenvolvimento sustentável, muitas organizações já focam em desenvolver tecnologias limpas (ou verdes). Durante a Conferência das Nações Unidas sobre o Meio Ambiente no ano de 1992, foi reconhecido o conceito de desenvolvimento sustentável. No documento chamado de Agenda 21, as tecnologias ambientalmente saudáveis foram assim determinadas (ONU, 1995):

As tecnologias ambientalmente saudáveis protegem o meio ambiente, são menos poluentes, usam todos os recursos de forma mais sustentável, reciclam mais seus resíduos e produtos e tratam os dejetos residuais de uma maneira mais aceitável do que as tecnologias que vieram substituir. As tecnologias ambientalmente saudáveis, no contexto da poluição, são "tecnologias de processos e produtos" que geram poucos ou nenhum resíduo, para a prevenção da poluição. Também compreendem tecnologias de "etapa final" para o tratamento da poluição depois que esta foi produzida.

Segundo essa definição, as tecnologias verdes tendem a mitigar o desperdício reduzindo a quantidade de resíduos lançados no meio ambiente. Com o intuito de fomentar o desenvolvimento dessas tecnologias, a OMPI lançou, em 2010, uma ferramenta online (*IPC Green Inventory*) com a relação do inventário das tecnologias consideradas verdes. Esse inventário foi desenvolvido com o objetivo de facilitar a busca por informações relacionadas às tecnologias verdes, que também são chamadas de Tecnologias Ambientalmente Amigáveis. (MARTINS et al. 2016).

Com o objetivo de incentivar a criação destas tecnologias verdes, diversos países criaram mecanismos para beneficiar estas invenções. Através dos escritórios de propriedade intelectual, entidades governamentais criaram os chamados programas de Patentes Verdes. Estes programas visam dar prioridade no exame de concessão para todas tecnologias que se enquadrem nas categorias ambientalmente amigáveis.

No ano de 2009, o Reino Unido inaugurou o primeiro programa de patentes verdes, o programa britânico nasceu com a expectativa de conceder em nove meses as patentes que nele se enquadram, sendo a duração normal para exame de patentes de aproximadamente dois a três anos. No mesmo ano outros países também lançaram seus programas, sendo eles: Austrália, Coréia do Sul, Japão, Estados Unidos e Israel. Em 2011 foi a vez do escritório canadense ter seu próprio programa. O Brasil foi o primeiro país dos BRICS a ter programa similar, a publicação da Resolução nº 283 em Abril de 2012 deu início ao programa de Patentes Verdes nacional. Ainda em 2012, a China lançou seu próprio programa. Já

em 2013 a Letônia implantou iniciativa similar. (DECHEZLEPRÊTRE, 2013; REIS et al. 2016).

Passados quatro anos do início do programa-piloto brasileiro de patentes verdes, desde o dia 6 de Dezembro de 2016 o mesmo passou a ser um serviço permanente do INPI. A resolução nº 175/2016 traz a listagem das tecnologias que o programa abrange (ANEXO I). Esta lista é dividida em cinco grupos distintos: energia alternativa, transporte, conservação de energia, gerenciamento de resíduos e agricultura, conforme mostra a Figura 5.

FIGURA 5 – CATEGORIAS DO PROGRAMA DE PATENTES VERDES



FONTE: O autor (2018).

A primeira das categorias busca fomentar a criação de energias alternativas. Dentro desta categoria se encontram as tecnologias voltadas aos biocombustíveis, células-combustível, aproveitamento de energia a partir de resíduos e energias renováveis (hidráulica, térmica, eólica, solar e geotérmica), além de outros tipos também utilizados. (BRASIL, 2016).

Na sequência, a categoria de transportes visa estimular o desenvolvimento de veículos ambientalmente amigáveis. Além dos veículos elétricos e híbridos, esta categoria também engloba as tecnologias voltadas a estações de carregamento para

veículos elétricos, veículos cujas fontes de alimentação são extraídas da natureza, e redução nas perdas de energia oriundas nos sistemas de frenagem e transmissão. Também faz parte da categoria de transportes carrocerias que possuem baixo arrasto aerodinâmico. (BRASIL, 2016).

Já a categoria relacionada à conservação de energia atende às tecnologias que visam armazenar e otimizar a energia. Esta categoria não só estimula a redução do consumo energético, principalmente através dos sistemas de iluminação de baixo consumo, como também engloba tecnologias relativas ao isolamento térmico de edificações e a recuperação de energia mecânica, tais como balanço, rolamento e arfagem. (BRASIL, 2016).

As tecnologias voltadas ao gerenciamento de resíduos estão listadas no item 4 do anexo I da resolução 175/2016. Nesta categoria são elencadas tecnologias para eliminação e tratamento de resíduos. Dentro do tratamento de resíduos são listadas tecnologias para produção de bens com base em aproveitamento de restos ou refugos, tecnologias voltadas à qualidade do ar, ao controle da poluição da água e meios para prevenção de contaminação radioativa em caso de vazamento de reatores. (BRASIL, 2016).

Por fim, a quinta e última categoria se refere às tecnologias de agricultura sustentável. Este item busca atender novas técnicas de reflorestamento, técnicas alternativas de irrigação, uso de pesticidas alternativos e melhoria do solo.

A resolução 175/2016, que disciplina o exame prioritário de Patente Verde, dispõe em seu artigo 2º que Patente Verde “*é o pedido de patente considerado apto ao exame prioritário, conforme a listagem presente no ANEXO I deste documento*”. O pedido é válido em todo território nacional (Art. 4º) e deve conter no máximo quinze reivindicações, das quais três delas independentes (Art. 5º). As demais disposições a cerca do pedido podem ser consultadas na resolução citada anteriormente.

2.4 PATENTOMETRIA

Morais (2014) destaca que nos estudos métricos da informação a bibliometria e a cientometria são as metodologias mais relevantes e, recentemente, tanto a webometria quanto a patentometria têm se mostrado poderosas ferramentas de informação. Desta forma, os estudos métricos podem ser considerados como

ferramenta de inteligência competitiva auxiliando empresas e pesquisadores nas tomadas de decisões estratégicas.

Conforme salientado anteriormente, dentre os estudos métricos a webometria e a patentometria são os termos mais recentes. A primeira é voltada aos recursos cuja fonte está disponibilizada na web, já a segunda tem como base a análise em documentos de patentes medindo o grau de tecnologia e inovação de um país ou instituição em determinado segmento da indústria. (GUZMÁN SÁNCHEZ, 1999).

A patentometria, para alguns autores, pode ser considerada uma técnica oriunda dos métodos analíticos que pertencem à bibliometria. Os indicadores utilizados na análise de documentos de patentes sofrem adaptações de outros já utilizados em outros tipos de documentos. (FAGUNDES et al., 2014).

Ainda que o termo patentometria seja bastante novo, outras formas de se referir ao estudo das patentes vêm sendo utilizada por pesquisadores, tais como: análise de patentes, estatística de patentes ou até mesmo bibliometria de patentes. (GRANT, HOF e GOLD, 2014).

Entretanto, algumas características presentes nos documentos de patentes são de suma importância neste tipo de pesquisa. Em primeiro lugar a padronização das informações nos documentos, seguido pela organização e classificação das patentes facilitando assim pesquisar sobre determinados nichos de mercado possibilitando estudar as publicações de países, empresas e instituições e, por último, mas não menos importante, a disponibilidade de dados em bases automatizadas. (GUZMAN SÁNCHEZ, 1999).

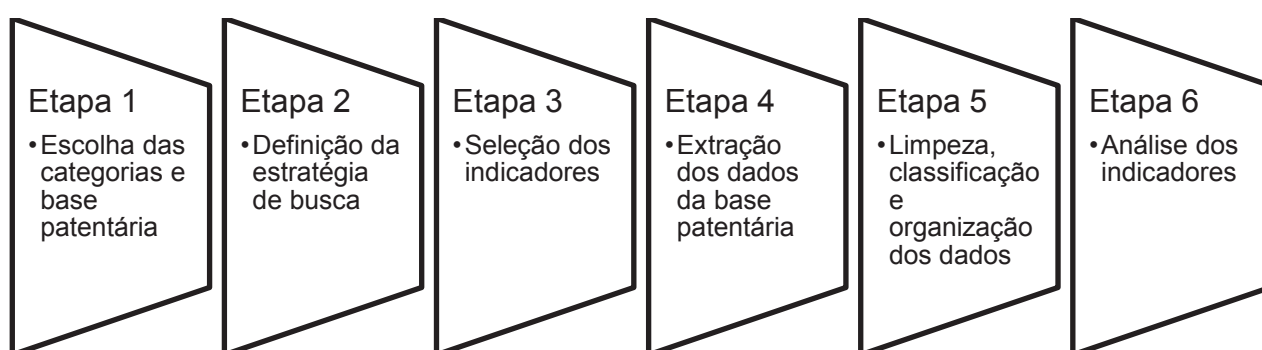
Guzman Sánchez (1999) divide os indicadores patentométricos em quatro grupos distintos: indicadores relacionais de primeira geração, indicadores relacionais de segunda geração, indicadores de atividades e família de patentes.

Segundo Fagundes et al. (2014), os indicadores relacionais de primeira geração buscam identificar inovações e frentes tecnológicas além de analisar a integração de determinada comunidade. Estes indicadores também têm como função estruturar determinados aspectos, tais como: redes de citações e cocitação. Já os indicadores relacionais de segunda geração, foram criados para analisar o conteúdo do documento, considerando as informações presentes no título, no resumo ou no próprio texto.

De acordo com o mesmo autor, os indicadores de atividade têm como objetivo medir as forças e fraquezas da tecnologia, além de identificar as empresas e países líderes em determinado segmento, três aspectos são analisados, sendo eles: distribuição das publicações, produtividade dos agentes de inovação e o número de citações. Por último, as famílias de patentes, são consideradas uma poderosa fonte de indicadores.

A revisão da bibliometria não deixa claro quais as etapas que devem compor uma patentometria. Os passos seguidos apresentam variações de autor para autor. Porém, como nos demais estudos métricos há uma convergência no método adotado pelos autores. (ARDITO et al., 2018; FAGUNDES et al. 2014; GONÇALVES et al., 2018; GÚZMAN-SÁNCHEZ, 1999; MORAIS, 2014). A Figura 6 sintetiza estas etapas.

FIGURA 6 - ETAPAS DA PATENTOMETRIA



FONTE: O autor (2018).

Tais etapas podem ser definidas como:

- Etapa 1 – Escolha das categorias e base patentária: Faz parte do planejamento da pesquisa. Sua escolha depende dos objetivos do pesquisador.
- Etapa 2 – Definição da estratégia de busca: Assim como na etapa 1, também faz parte do planejamento. As estratégias de busca variam de autor para autor, porém é comum utilizar os códigos IPC (ARDITO et al., 2018) ou termos de busca. (GONÇALVES et al., 2018).
- Etapa 3 – Seleção dos indicadores: A seleção dos indicadores está relacionada aos objetivos da pesquisa. O pesquisador deve selecionar o(s) grupo(s) que possui as informações mais relevantes para a pesquisa.

Conforme salientado por Fagundes et al. (2014), a análise das famílias de patentes costuma ser essencial na patentometria.

- Etapa 4 – Extração dos dados da base patentária: Etapa consiste em extrair os dados da base automatizada.
- Etapa 5 – Limpeza, classificação e organização dos dados: Consiste em eliminar dados desnecessários e/ou duplicados, classificar e organizar o restante das informações para atender os requisitos da última etapa.
- Etapa 6 – Análise dos indicadores: São analisados os indicadores planejados na Etapa 3 e gerados nas Etapas 4 e 5.

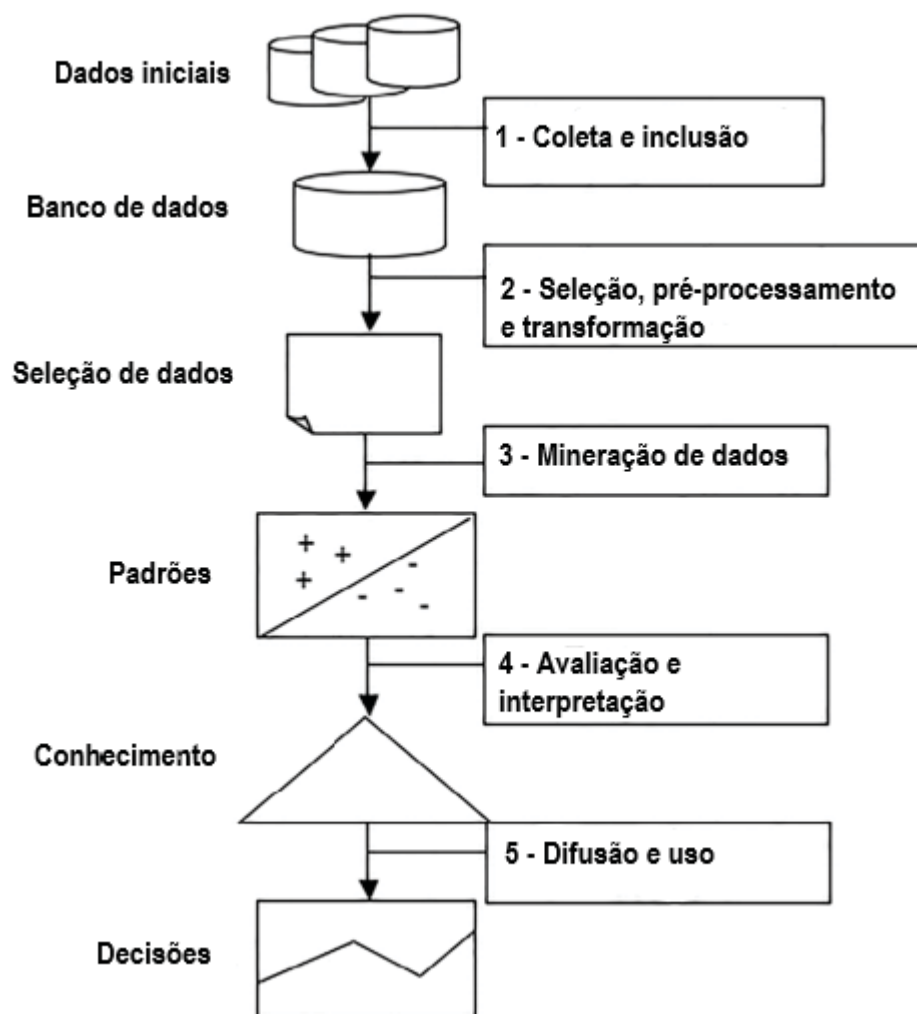
2.5 DATA MINING

A mineração de dados tem sido um dos temas de pesquisa mais modernos do nosso século. Ela tem feito contribuições importantes para o avanço científico e tecnológico em um vasto número de campos. (PHRIDVIRAJ e GURURAO, 2014). Segundo Alves (2019), as técnicas de mineração de dados buscam identificar e extrair padrões relevantes em um conjunto de dados. Conforme salientam Lara et al. (2014), as técnicas de mineração de dados oferecem uma ampla gama de soluções para diferentes problemas tais como: classificação, pesquisa de regras de associação, agrupamento ou regressão.

Para Fayyad et al. (1996), há uma necessidade de se desenvolver ferramentas capazes de processar o conteúdo do rápido crescimento de dados digitais. Segundo os autores, estas ferramentas fazem parte do chamado *KDD* (*Knowledge Discovery in Databases*). Este campo do conhecimento tem como principal objetivo o desenvolvimento de técnicas e métodos para compreender os dados.

O fluxo do processo do KDD, Figura 7, pode ser dividido em cinco etapas, partindo da coleta dos dados até sua difusão e uso. (FERNÁNDEZ-ARTEAGA et al., 2016;). O KDD estende-se desde o entendimento e preparação dos dados até a interpretação e uso dos resultados do processamento dos dados. Portanto, a mineração de dados é um estágio do processo do KDD durante o qual um conjunto de técnicas e ferramentas é aplicado para estudar os dados e extrair qualquer informação útil oculta. (LARA et al., 2014).

FIGURA 7 - FLUXO DO KDD



Fonte: Adaptado de Fernández-Arteaga et al. (2016).

No entanto, dados estruturalmente mais complexos estão aparecendo e, desta forma, é necessário adotar novas formas de abordagem. Conforme descrevem Silva et al. (2016), os dados são considerados a matéria-prima para que a mineração de dados ocorra. Segundo os autores, a forma com a qual esta matéria-prima se manifesta pode ser dividida em dois grandes grupos: estruturada e não estruturada.

Dados estruturados são encontrados, geralmente, em sistemas informatizados convencionais. Suas bases organizam os dados em estruturas tabulares onde as linhas armazenam uma ocorrência e as colunas representam características que a descrevem. Já os dados não estruturados são disponibilizados

sem uma estrutura definida, sua apresentação pode vir em forma de texto, imagem, vídeos e sons. (SILVA et al., 2016).

Portanto, conforme surgem cada vez mais dados estruturalmente mais complexos, novas abordagens de descoberta de conhecimento são necessárias. Em alguns domínios as informações sobre indivíduos (ou eventos) podem ser compostas por um amplo conjunto de dados inter-relacionados ao invés de ser representados por uma única tabela de atributos simples. (LARA et al, 2014).

Um dos ramos da mineração de dados que vem crescendo nos últimos anos é a mineração de texto, suas tecnologias podem combinar recursos de conhecimento, análise linguística e aprendizado de máquina. Além disso, as ferramentas de mineração de texto extraem não somente termos do texto, mas também as relações entre eles. Ferramentas e métodos de mineração de texto têm sido criados e usados para análise de publicações em diversas áreas. (CHAIX et al., 2018).

Na próxima seção serão apresentados os trabalhos correlatos a cerca da mineração de texto.

2.6 TRABALHOS CORRELATOS

A mineração de texto, assim como a análise de redes, tem sido utilizada para identificar os rumos em diversos campos de pesquisa. (DING et al., 2018). Em estudo recente, Fleuren e Alkema (2015) analisaram a aplicação da mineração de texto no domínio biomédico. Segundo os autores, a mineração de texto é utilizada em pesquisas bibliográficas dedicadas, respondendo a perguntas específicas. Kumar e Ravi (2016) destacam que as técnicas de mineração de texto são capazes de extrair conhecimento dos arquivos de texto. Entretanto, antes de efetuar a mineração é preciso converter os dados de texto em um formato estruturado.

Neste caso, o processo de mineração em documentos textuais pode ser dividido em dois passos: pré-processamento e extração do conhecimento. No primeiro, é feita a conversão dos dados não estruturados em uma matriz de termos do documento através do processo de tokenização, que são operações utilizadas por mecanismos para a análise morfológica do texto. (IBM KNOWLEDGE CENTER, 2019). Também no pré-processamento, são removidas as chamadas *stop words*

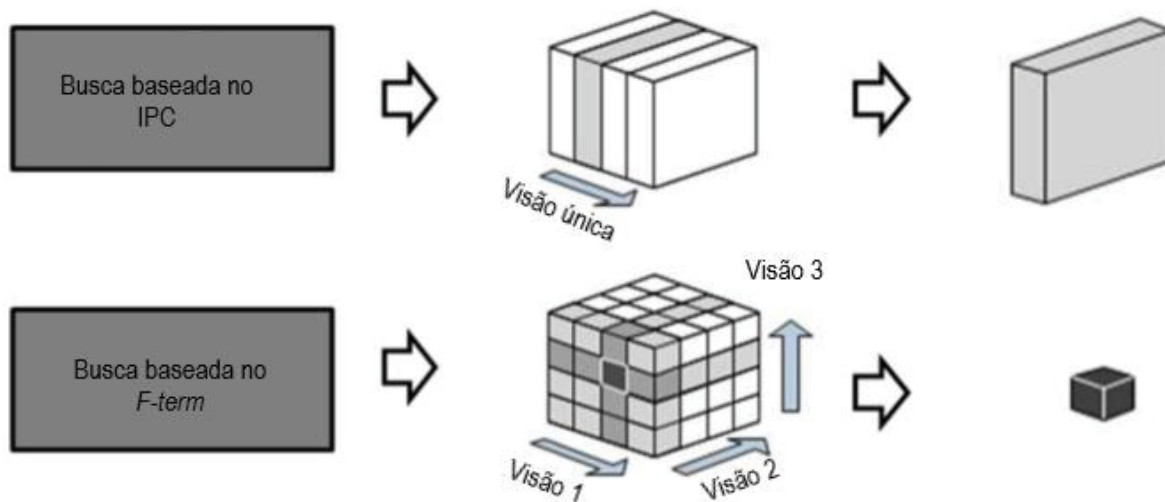
(palavras de ligação), além de outros termos desejados pelos pesquisadores. (KUMAR e RAVI, 2016).

Salton e Buckley (1988) apresentaram três diferentes formas de se montar uma matriz de termos: Termos Binários (BT), Frequência dos Termos (TF) e Frequência dos Termos e Frequência Inversa dos Documentos (TF-IDF). A análise dos termos binários considera apenas se determinado termo está ou não presente em um documento. Já a frequência dos termos informa quantas vezes um termo x aparece em um documento y . Outra abordagem, segundo Silge e Robinson (2017), é examinar a frequência inversa dos documentos de um termo, desta forma são considerados os pesos que cada termo tem em cada um dos documentos analisados.

Recentemente pesquisadores sul coreanos utilizaram a análise do *F-term*, aliado a técnicas de mineração de texto, para descobrir novas oportunidades tecnológicas baseado em análise patentária. O *F-term* é um sistema de classificação utilizado na lei de patentes japonesa. Embora útil, pois classifica os documentos de patentes de acordo com suas características técnicas, os autores lembram que para sua utilização é necessário que as patentes sejam arquivadas no escritório japonês. (SONG et al., 2017).

A estratégia de busca de Song et al. (2017) buscou realizar um cruzamento entre diferentes áreas representadas pelos códigos do *F-term*, Figura 8. O resultado desta abordagem é a combinação de ideias de diferentes campos, isso permitiu uma análise com maior foco em tecnologias específicas, possibilitando a criação de novas ideias através da fusão das tecnologias.

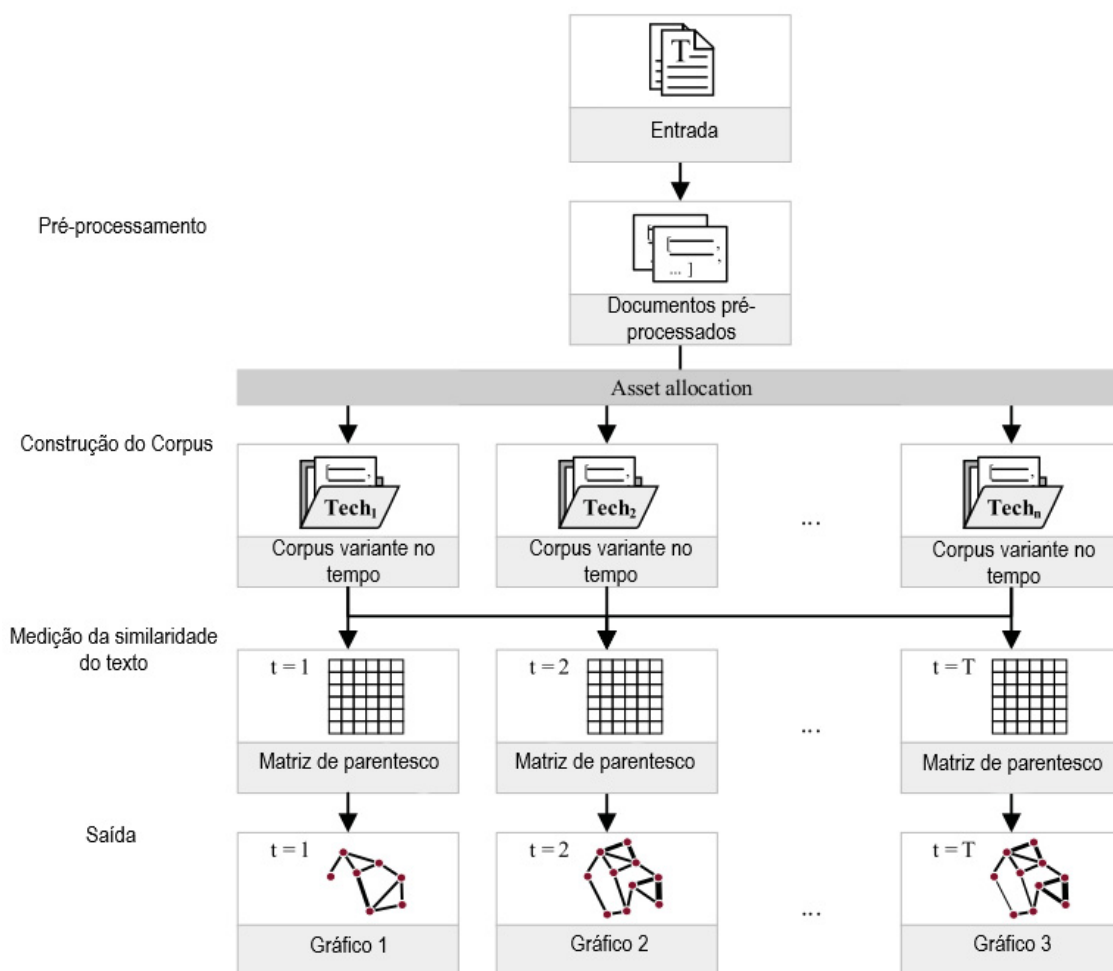
FIGURA 8 - COMPARAÇÃO DAS ESTRATÉGIAS DE BUSCA.



Fonte: Adaptado de Song et al. (2017).

A abordagem de Song et al. (2017) utilizou a mineração de texto no resultado da interseção da busca baseada no *F-term*. Já Hofmann et al. (2019) utilizaram a metodologia de busca pelos códigos IPC. Os autores desenvolveram um método, processando dados textuais de diferentes bases, para criar uma rede de relacionamento de intertecnologia analisável e legível, sua metodologia envolveu a criação de um corpus para medir a similaridade dos textos presentes em patentes e documentos científicos, Figura 9.

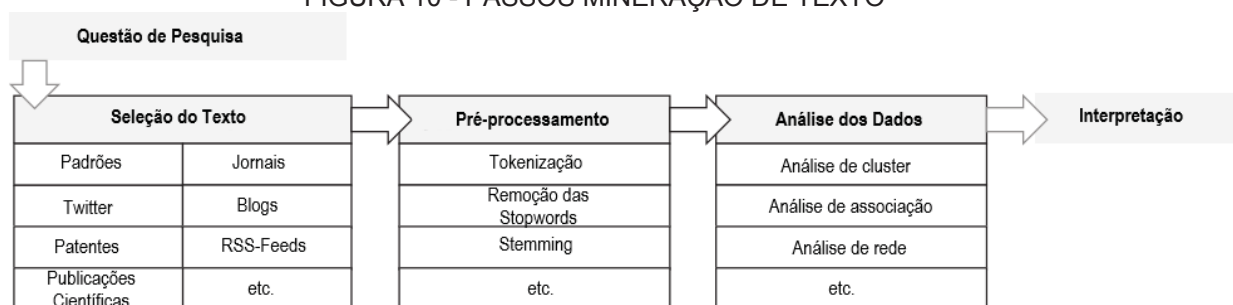
FIGURA 9 - ABORDAGEM DE HOFMANN et al. 2019



Fonte: Adaptado de Hofmann et al. (2019).

Independentemente da abordagem adotada por diferentes pesquisadores (DEMETER et al., 2018; FATTORI et al., 2003; SONG et al., 2017), a etapa de mineração de texto segue um padrão de passos conforme apresentam Kayser e Blind (2017), Figura 10.

FIGURA 10 - PASSOS MINERAÇÃO DE TEXTO



Fonte: Adaptado de Kayser e Blind (2017).

Dado o imenso volume de arquivos digitais disponíveis hoje em dia, a utilização de ferramentas capazes de processar se torna essencial. Sendo assim, foram comparados 12 softwares *open source*, capazes de executar a mineração de texto, disponíveis no mercado, Anexo II. Segundo Kaur e Chopra (2016), existem inúmeras técnicas para mineração de texto, dentre as mais populares podemos citar o processamento de linguagem natural, aprendizado de máquina, métodos estatísticos, inteligência artificial, técnicas de classificação, aprendizado linguístico e análise semântica.

Algumas características estão presentes nas variadas ferramentas para mineração de texto, tais como: análise de texto, processamento de texto, classificação/categorização, análise de sentimento, descoberta de conhecimento. Apesar da mineração de dados ser um campo já estabilizado, muitas das suas ferramentas não estão aptas a lidar com fontes de dados não estruturados, as quais requerem estas características citadas. Portanto, ferramentas para mineração de texto se tornam necessárias a todos pesquisadores e profissionais que trabalham com texto. (KAUR e CHOPRA, 2016).

3 METODOLOGIA

Neste capítulo, primeiramente, será apresentada a caracterização da pesquisa quanto à sua natureza, abordagem, objetivo e ao procedimento de pesquisa a ser utilizado. Na sequência, será apresentado o protocolo de pesquisa com as etapas propostas para que o objetivo geral seja alcançado.

3.1 CLASSIFICAÇÃO DA PESQUISA

A classificação de uma pesquisa pode variar de distintas formas. Segundo Gil (2010), elas vão desde questões relacionadas à sua natureza (básica ou aplicada), a forma de abordagem (qualitativa, quantitativa ou ambas), aos objetivos (exploratória, descritiva ou explicativa) e quanto aos seus procedimentos. Desta forma, o enquadramento metodológico que proporciona maior familiaridade com o tema proposto está apresentado no Quadro 1.

QUADRO 1 - ENQUADRAMENTO METODOLÓGICO

Classificação	Enquadramento
Natureza	Básica
Abordagem	Quali-quantitativa
Objetivos	Descritiva
Procedimento	Bibliográfica

FONTE: O autor (2018).

A natureza desta pesquisa pode ser classificada como básica, pois objetiva gerar conhecimentos novos úteis para o avanço da ciência sem aplicação prática prevista. Sua abordagem será quali-quantitativa, pois buscará a partir da análise de documentação obter dados numéricos e realizar um tratamento estatístico embasando assim as interpretações posteriores. Quanto ao seu objetivo, ela será descritiva, tendo a função de descrever as características relevantes dos BRICS. Por fim, o procedimento adotado será bibliográfico, uma vez que fará uso de dados presentes em documentos de patentes. (MIGUEL, 2010).

Para atingir os objetivos específicos propostos na seção 1.2.1 deste trabalho, foram utilizadas as ferramentas descritas no Quadro 2.

QUADRO 2 - OBJETIVOS ESPECÍFICOS E FERRAMENTAS

Objetivo específico	Ferramenta
Identificar as linguagens e softwares <i>open source</i> utilizados atualmente no processo de mineração de texto;	Revisão da literatura
Desenvolver algoritmo para mineração de texto a fim de comparar o conteúdo dos documentos (título e resumo) depositados no Brasil e nas patentes tetrádicas	Programação em R
Recuperar informações tecnológicas, entre elas o colume de depósitos, empresas líderes, códigos IPC e frequência dos termos, através de estudo prospectivo em base patentária na categoria de transportes	Patentometria e Mineração de Texto
Traçar a tendência tecnológica das publicações na área de transportes, contribuindo com o mapeamento das tecnologias desenvolvidas	Patentometria e Mineração de Texto
Traçar a tendência tecnológica das publicações na área de transportes, contribuindo com o mapeamento das tecnologias desenvolvidas	Patentometria e Mineração de Texto
Identificar as potenciais tecnologias verdes que estarão disponíveis na área de transportes	Patentometria e Mineração de Texto

FONTE: O autor (2018).

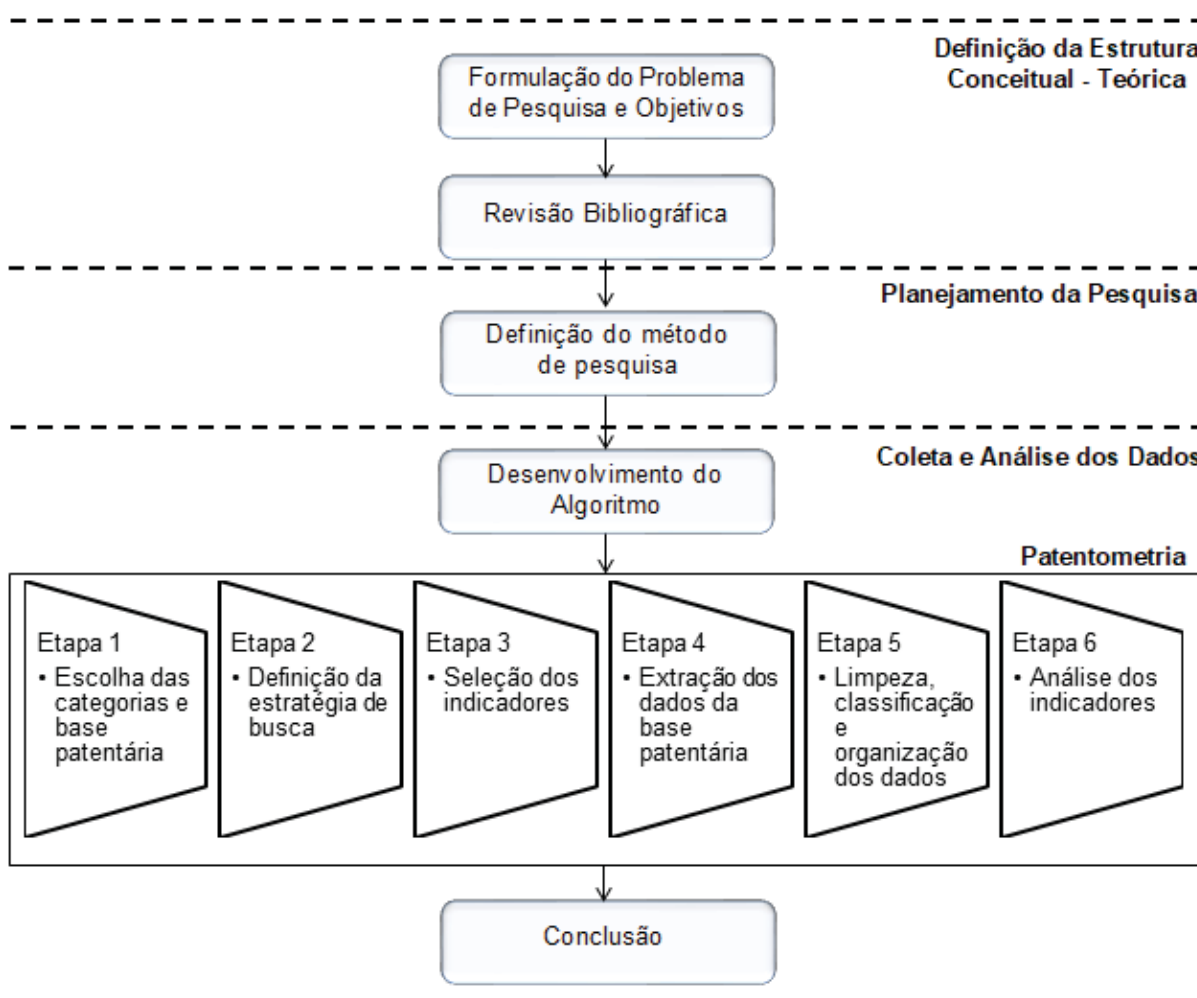
Através da análise das patentes (patentometria) é possível monitorar as tendências tecnológicas e identificar os principais desenvolvedores em determinado campo. (CHANG, WU e LEU, 2010). Para fazer esta análise foi selecionada a base *Derwent Innovations Index*®. Esta poderosa ferramenta combina informações das bases de dados *Derwent Patents Index*, *Derwent Patent Citation Index* e *Derwent Chemistry Resource*, sendo atualizada semanalmente e contendo mais de 11 milhões de invenções. (DERWENT, 2018).

3.2 PROTOCOLO DE PESQUISA

Tendo em vista o objetivo geral descrito no Capítulo 1 “Elaborar um estudo comparativo em base patentária, identificando as tecnologias verdes da categoria transportes depositadas nos BRICS para obtenção de informações que auxiliem empresas e pesquisadores nas tomadas de decisões”, foi elaborado o protocolo de pesquisa apresentado na Figura 11. Para Mello et al. (2012), a elaboração de um protocolo contribui com o aumento da confiabilidade da pesquisa. Este protocolo está dividido em três macro fases: Definição da estrutura conceitual-teórica, planejamento da pesquisa e coleta e análise dos dados (Patentometria). Sendo assim, na sequência são descritas as etapas do processo.

- a) **Formulação do problema de pesquisa e objetivos:** É a primeira etapa da pesquisa, a qual consistiu na formulação do problema, dos objetivos e na delimitação do tema, capítulo 1.
- b) **Revisão da bibliografia:** A revisão da bibliografia, assim como a etapa anterior, também faz parte da macro fase de definição da estrutura conceitual-teórica, foi realizada com base em livros, artigos, páginas eletrônicas, dissertações, teses e relatórios acerca de temas correlatos, como Propriedade Intelectual, Patentes, Patentes Verdes, Patentometria, *Data Mining* e Trabalhos Correlatos na qual se encontram no Capítulo 2 do presente trabalho.

FIGURA 11 - PROTOCOLO DA PESQUISA



FONTE: O autor (2018).

- c) **Definição do método de pesquisa:** Refere-se à identificação e seleção do método utilizado na pesquisa. Dentre os métodos que poderiam ser utilizados

nesta pesquisa estão: a bibliometria, a webometria e a patentometria. Optou-se para a condução deste estudo pelo uso da patentometria, pois segundo Pereira (2008), dentre outros estudos métricos, a patentometria é a que mais aproxima a academia com a indústria.

- d) **Desenvolvimento do algoritmo:** Como ferramenta auxiliar no estudo do conteúdo das patentes, foi desenvolvido um algoritmo em linguagem de programação R para realizar a mineração de texto no título e resumo dos documentos extraídos. A linguagem de programação R se mostrou a melhor opção por ser um código *Open Source* e já possuir pacotes voltados à manipulação de *strings* e mineração de texto. Os pacotes utilizados foram: ‘tm’ (FEINERER e HORNIK, 2018), ‘stringr’ (WICKHAM, 2018), ‘tidyverse’ (WICKHAM, 2017), ‘wordcloud’ (FELLOWS, 2018), ‘pdftools’ (OOMS, 2018), ‘ggplot’, ‘ggplot2’ (WICKHAM, 2016), ‘dplyr’ (WICKHAM et al., 2018), ‘tidytext’ (SILGE E ROBINSON, 2016), ‘tidyr’ (WICKHAM e HENRY, 2018), ‘ggraph’ (PEDERSEN, 2018), ‘scales’ (WICKHAM, 2018b), ‘igraph’ (CSARDI e NEPUSZ, 2006), ‘nortest’ (GROSS e LIGGES, 2015), ‘quanteda’ (BENOIT et al., 2018), ‘tokenizers’ (MULLEN et al., 2018) e ‘wordnet’ (FEINERER e HORNIK, 2017; WALLACE, 2007; FELLBAUM, 1998). O algoritmo pode ser consultado no Apêndice I deste documento e sua utilização está descrita com maiores detalhes na seção 4.2.2.
- e) **Escolher as categorias e base patentária:** Consistiu na definição da categoria a ser analisada no presente estudo. Tomando como base o programa de Patentes Verdes brasileiro, foi escolhida a categoria de transportes. Já a definição da base patentária contemplou a identificação e seleção da base de dados utilizada na pesquisa. Segundo o manual estatístico da Organização para a Cooperação e Desenvolvimento Econômico (OECD, 2009), os bancos de dados de patentes foram criados a muito tempo e, na última década expandiram o leque de informações disponíveis. Conforme salientado anteriormente, a base *Derwent Innovations Index*® serviu como fonte dos dados.
- f) **Definir a estratégia de busca:** Os códigos de busca adotados nesta pesquisa são os indicados no Inventário Verde da Organização Mundial da Propriedade Intelectual (OMPI), da sigla em inglês *IPC Green Inventory* (Anexo III). A extração dos registros foi realizada com um recorte anual, portanto foi gerado

um arquivo para cada família de patentes por ano. Esta estratégia permitiu identificar a variação da frequência dos termos ao longo do período analisado.

- g) **Seleção dos indicadores:** Os indicadores selecionados para análise foram os de atividade e relacionais de segunda geração uma vez que atendem as necessidades da pesquisa. A partir da base patentária foram extraídos os indicadores de atividade, já os relacionais de segunda geração foram gerados através do algoritmo desenvolvido.
- h) **Extração dos dados da base patentária:** Etapa em que os registros são exportados da base *Derwent Innovations Index*® e transformados em arquivo de texto.
- i) **Limpeza, classificação e organização dos dados:** Utilizou-se da programação em *R* através de ferramentas e pacotes que buscam expressões regulares. Também foram geradas tabelas, listas e gráficos oriundos tanto da base patentária quanto do algoritmo.
- j) **Análise dos indicadores:** Para análise dos indicadores, também foi utilizada a programação em *R*. Esta etapa buscou identificar a atividade de inovação e tecnologia nos BRICS contidas nos documentos de patentes. Ela possibilitou conhecer a atividade tecnológica e refletir as tendências de mudanças técnicas ao longo do tempo.
- k) **Conclusão.**

4 DESENVOLVIMENTO

Para atingir o objetivo proposto, de comparar as tecnologias depositadas no Brasil e nos demais países pertencentes aos BRICS, este capítulo apresenta o desenvolvimento do algoritmo em linguagem R, a escolha das categorias e bases patentárias, como foi definida a estratégia de busca, os indicadores selecionados para análise, a extração dos dados da base patentária e por fim a limpeza, classificação e organização dos dados.

4.1 DESENVOLVIMENTO DO ALGORITMO EM R

Como ferramenta auxiliar na pesquisa das patentes, foi desenvolvido um código em R para minerar o texto presente no título e no resumo dos documentos depositados. A criação deste algoritmo foi necessária, pois os *softwares* disponíveis como AntConc, Aika e Sobek não atendiam as demandas da pesquisa. Dentre os programas pesquisados, embora tenham funções úteis para a análise, não havia flexibilidade para excluir palavras indesejadas, criar gráficos de barras e/ou nuvem de palavras e funções para cálculo da frequência relativa. Como o R é uma linguagem *open source*, também trouxe uma flexibilidade maior para a análise do conteúdo. Conforme descrito no item 1.3 do presente documento, a análise do conteúdo foi realizada nos depósitos efetuados no escritório brasileiro em comparação com as patentes tetrádicas (documentos depositados simultaneamente nos escritórios norte americano, japonês, europeu e chinês). As patentes tetrádicas possuem alto valor comercial e, conseqüentemente, são consideradas de melhor qualidade.

Para realizar a mineração de texto, foi necessário instalar pacotes que não fazem parte do **R-base**. Estes *packages* possuem funcionalidades específicas e também são chamados de *contributed packages*, são desenvolvidos pela comunidade e disponibilizados de forma oficial na página do R.

O algoritmo desenvolvido visa facilitar e automatizar a análise do conteúdo dos documentos. Vale salientar que não foi desenvolvido nenhum pacote ou função nova. O Apêndice I apresenta a lista de códigos utilizados. Após o desenvolvimento do algoritmo, a próxima etapa da pesquisa é a escolha das categorias e base patentária.

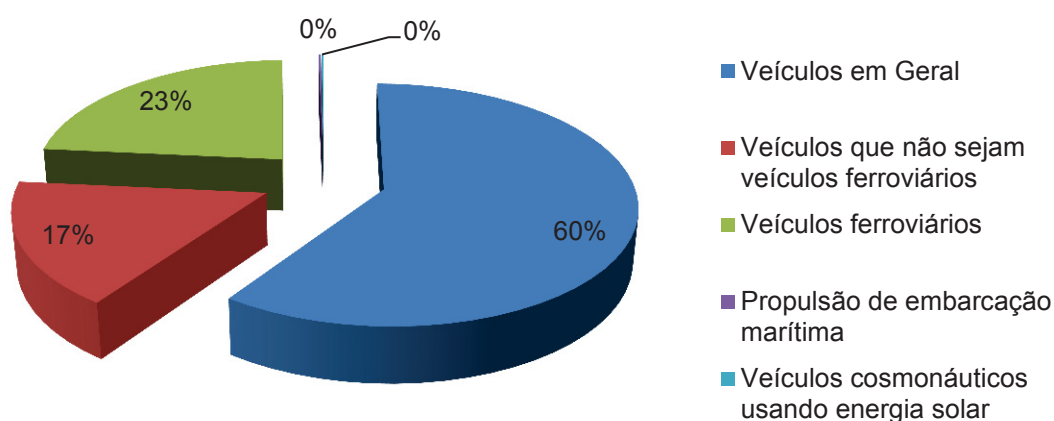
4.2 ESCOLHA DAS CATEGORIAS E BASE PATENTÁRIA

O inventário verde do IPC possui sete categorias, sendo elas: produção de energia alternativa, transporte, conservação de energia, gerenciamento de resíduos, agricultura, aspectos administrativos, regulamentares ou de desenho e geração de energia nuclear. (WIPO, 2018). Já o programa de patentes verdes brasileiro é subdividido nas categorias: energia alternativa, transportes, conservação de energia, gerenciamento de resíduos e agricultura sustentável. Dentre as cinco opções do programa brasileiro, que serviu como base para a presente pesquisa, a categoria de transportes foi selecionada devido às mudanças que vem ocorrendo a nível mundial e já explanadas nas seção 1.1 deste trabalho.

Tomando então como base esta categoria de transportes, o inventário verde do IPC (ANEXO III) a divide em cinco subcategorias. Cada uma delas possui um código IPC (ou *cluster* de códigos) que se refere a determinado tipo de tecnologia. O Anexo IV apresenta as descrições das principais seções, classes e subclasses da categoria.

O percentual de participação de cada subcategoria dentro do espaço amostral é apresentado na Figura 12. Veículos em geral, Veículos que não sejam veículos ferroviários (tradução livre) e Veículos ferroviários representam 99,66% do total. Já as subcategorias Propulsão de embarcação marítima e Veículos cosmonáuticos usando energia solar representam 0,34% da amostra.

FIGURA 12 – PERCENTUAL DE PARTICIPAÇÃO POR SUBCATEGORIA



FONTE: O autor (2018).

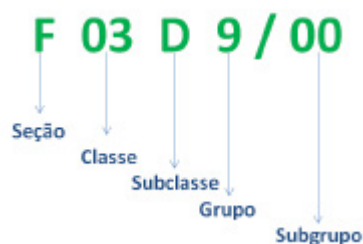
No momento, existem diversas bases disponíveis para pesquisa, a utilização de cada uma depende dos objetivos de cada pesquisador. Conforme Silva (2016), dentre alguns exemplos de bases podemos citar a do INPI, a ESPACENET, LATIPAT, PATENTSCOPE, USPTO e a *Derwent Innovations Index*®. Todas possuem mecanismos de busca, facilitando assim o acesso às informações. Na condução deste trabalho, optou-se pela base *Derwent Innovations Index*® por atender as demandas geradas mediante análise dos objetivos da pesquisa. A estratégia de busca utilizada será explicada com maior detalhe na seção seguinte.

4.3 DEFINIÇÃO DA ESTRATÉGIA DE BUSCA

O primeiro passo para definição da estratégia foi a escolha dos países, com isto, foi possível identificar o padrão no número das patentes depositadas nos respectivos escritórios nacionais. Com o período de anos definido (2012 – 2016), para a coleta dos arquivos referentes à categoria de transportes, era possível adotar diferentes abordagens. Dentre estas abordagens, poderiam ser pesquisados por tópico, códigos de classe e códigos manuais da própria base e a classificação internacional de patentes (IPC). Os códigos IPC foram criados para facilitar a busca de patentes de acordo com seu campo tecnológico, atualmente possuem a adesão em mais de 100 países e, em alguns casos, são a única forma de classificação dos documentos. (OECD, 2009). Portanto, a classificação internacional de patentes se mostra a alternativa mais adequada para a pesquisa, uma vez que não é exclusiva da base utilizada e seus códigos podem ser pesquisados nas demais bases existentes.

A classificação documentária das patentes é composta por letras, números e símbolos os quais facilitam a identificação da subdivisão da Classificação Internacional de Patentes. Para exemplificar, a Figura 13 mostra a classificação referente a uma “Turbina eólica de energia para veículos elétricos e híbridos”. (MORAIS, 2014).

FIGURA 13 – EXEMPLO CLASSIFICAÇÃO CÓDIGO IPC



FONTE: MORAIS (2014).

- F – Engenharia mecânica;
- F03 – Máquinas ou motores para líquidos; motores movidos a vento, molas, pesos ou outros; produção de força mecânica ou de empuxo propulsivo por reação, não incluída em outro local;
- F03D – Motores movidos a vento;
- F03D 9/00 – Adaptações de motores a vento para uso especial; Combinações de motores a vento com aparelhos por eles acionados.

Definidas as expressões de busca, a Tabela 2 apresenta de forma resumida a concatenação entre elas.

TABELA 1 - ESTRATÉGIA DE BUSCA

Expressão de busca (Número da patente)	Operador booleano	Expressão de busca (Ano)	Operador booleano	Expressão de busca (Código IPC)
BR*		2012		
RU*		2013		
IN*	AND	2014	AND	ANEXO III
CN*		2015		
ZA*		2016		

FONTE: O autor (2018).

Vale ressaltar, que determinados *clusters* de códigos extrapolavam os limites permitidos de operadores booleanos na base, portanto foi necessário utilizar os chamados caracteres curinga. Tendo isto em vista, a etapa seguinte da patentometria foi a seleção dos indicadores.

4.4 SELEÇÃO DOS INDICADORES

Para a OECD (2009), há uma grande variedade de estudos e indicadores baseados em dados de patentes, seus formatos de publicação vão desde relatórios de políticas até pesquisas acadêmicas. Dentre os tópicos de investigação elencados pela organização estão: desempenho tecnológico, tecnologias emergentes, geografia das invenções, o papel das universidades no desenvolvimento tecnológico, globalização das atividades de P&D e as estratégias de patenteamento por empresas.

Conforme descrito na seção 2.6, a patentometria possui quatro tipos de indicadores: indicadores relacionais de primeira geração, indicadores relacionais de segunda geração, famílias de patentes e indicadores de atividade. Para atender aos objetivos da presente pesquisa, os indicadores selecionados foram os de atividade e os relacionais de segunda geração. Os indicadores de atividade medem as forças e fraquezas da tecnologia, além disso, foram identificadas as empresas líderes, a distribuição das publicações (nos BRICS) e a produtividade dos agentes. Já os indicadores relacionais de segunda geração, através do auxílio do algoritmo em R, buscaram informações presentes no título e no resumo dos documentos para análise do conteúdo.

4.5 EXTRAÇÃO DOS DADOS DA BASE PATENTÁRIA

A extração dos dados da base patentária foi executada em duas etapas. A primeira delas foi responsável pela extração dos indicadores de atividade, onde foram exportados os volumes de depósitos, a lista com empresas líderes e produtividade dos agentes. Após a execução da estratégia de busca, a base patentária possui uma ferramenta chamada 'Analisar resultados', onde é possível extrair as informações referentes às áreas de conhecimento, nome de depositante (empresas), código do depositante, inventor, código internacional de classificação de patentes (IPC), código de classe e código manual no *Derwent*. Estes últimos servem como fonte auxiliar de dados para interpretação de alguns resultados.

Já a segunda etapa da extração, foi responsável pela obtenção do conteúdo presente no título e no resumo dos documentos. Os arquivos extraídos da base são em formato 'txt' e os dados não são estruturados. Portanto, se torna necessária a

etapa de limpeza, classificação e organização dos dados na patentometria, justamente para transformar estes dados em informações que sejam úteis.

4.6 LIMPEZA, CLASSIFICAÇÃO E ORGANIZAÇÃO DOS DADOS

Após a extração dos dados, a etapa de limpeza, classificação e organização é iniciada. Os indicadores de atividade, assim como os de segunda geração, são exportados em arquivos no formato 'txt'. Entretanto, a formatação dos dados é estruturada, desta forma é preciso apenas formatar os dados para apresentação na forma de quadros e tabelas. Já os indicadores relacionais de segunda geração são limpos, classificados e organizados com a utilização do algoritmo em R e, seu primeiro passo é a instalação dos pacotes descritos na Figura 14.

FIGURA 14 - PACOTES UTILIZADOS NA MINERAÇÃO DE TEXTO

```
#Instalação de pacotes
install.packages('tm')
install.packages('stringr')
install.packages('tidyverse')
install.packages('wordcloud')
install.packages('pdftools')
install.packages('ggplot')
install.packages('ggplot2')
install.packages('dplyr')
install.packages('tidytext')
install.packages('tidyr')
install.packages('ggraph')
install.packages('scales')
install.packages('igraph')
install.packages('nortest')
install.packages('quanteda')
install.packages('tokenizers')
install.packages('wordnet')
install.packages("Rgraphviz")
```

FONTE: O autor (2018).

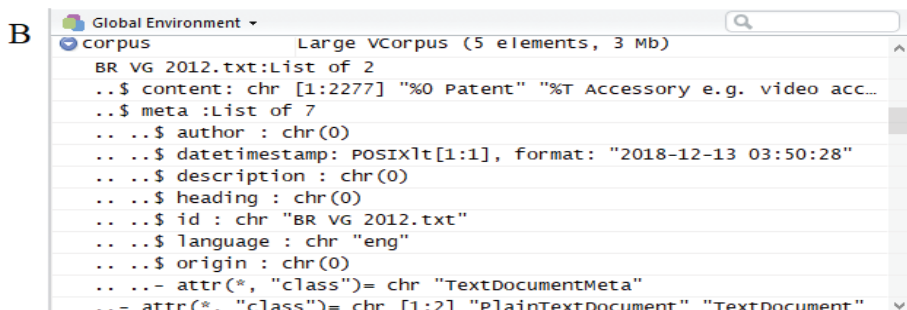
Após a instalação dos pacotes foi utilizado o comando Library para carregar as funções disponibilizadas por eles. Na sequência, devem ser importados os dados extraídos da base escolhida, por padrão o parâmetro de leitura (reader=readPlain) está configurado para arquivos no formato ".txt", porém pode ser alterado para arquivos nos formatos PDF (readPDF), RCV1 (readRCV1), *Reuters-21578* (readReut21578XML), XML (readXML) e *POS-tagged Word Text* (readTagged).

Com os pacotes instalados e carregados, a segunda etapa do algoritmo é importar os documentos presentes no diretório para o R. O código utilizado está descrito na Figura 15 (A). Os dados ficam armazenados na variável **corpus** que pode ser conferida na pilha, Figura 15 (B).

FIGURA 15 - IMPORTAR DADOS PARA O R

A

```
#Importar CORPUS
corpus <- vCorpus(DirSource("C:/Users/Pichau/OneDrive/UFPR/Dissertação/Corpus", encoding="UTF-8"),
  readerControl=list(reader=readPlain,language="eng"))
```



FONTE: O autor (2018).

A terceira etapa do algoritmo consiste em limpar o conteúdo não desejável. Este bloco de funções (Figura 16) deve ser utilizado de acordo com o que deseja ser analisado.

FIGURA 16 - LIMPEZA DOS DADOS

```
60 #LIMPEZA DOS DADOS
61 corpus <- tm_map(corpus, content_transformer(tolower))
62 corpus <- tm_map(corpus, remove_words, stopwords("english"))
63 corpus <- tm_map(corpus, remove_words, c("the", "novelty", "advantage", "can", "drawing", "drawings", "isididw",
64 "usa", "description", "woa", "didw", "for", "patent", "aua", "inventor", "use"))
65 corpus <- tm_map(corpus, strip_whitespace)
66 corpus <- tm_map(corpus, remove_punctuation)
67 corpus <- tm_map(corpus, remove_numbers)
68 corpus <- tm_map(corpus, stem_document)
```

FONTE: O autor (2018).

Cada função do bloco elimina/altera determinados tipos de caracteres/palavras, a finalidade de cada uma delas está descrita abaixo.

- Linha 61 – Transforma todas as letras em minúsculas;
- Linha 62 – Remove as chamadas “palavras de paradas” (*stop words*). Como o conteúdo importado está no idioma inglês, é necessário especificar na função. A lista das *stop words* cadastradas no pacote ‘tm’ pode ser consultada na Figura 17;
- Linhas 63 e 64 – Exclui palavras adicionais que o usuário deseja;
- Linha 65 – Remove os espaços em branco em excesso no texto;
- Linha 66 – Remove as pontuações;
- Linha 67 – Remove os caracteres numéricos;
- Linha 68 – Mantém apenas o radical das palavras. Exemplo: **casa**, **casas**, **casinha**.

FIGURA 17 - LISTA DAS STOP WORDS DO PACOTE 'TM'

```

> stopwords("en")
[1] "i" "me" "my" "myself" "we" "our" "ours" "ourselves" "you" "your" "yours"
[12] "yourself" "yourselves" "he" "him" "his" "himself" "she" "her" "hers" "herself" "it"
[23] "its" "itself" "they" "them" "their" "theirs" "what" "which" "who" "whom"
[34] "this" "that" "these" "those" "am" "is" "are" "was" "were" "be" "been"
[45] "being" "have" "has" "had" "having" "do" "does" "did" "doing" "would" "should"
[56] "could" "ought" "i'm" "you're" "he's" "she's" "it's" "we're" "they're" "i've"
[67] "we've" "they've" "i'd" "you'd" "he'd" "she'd" "we'd" "they'd" "i'll" "you'll"
[78] "she'll" "we'll" "they'll" "isn't" "aren't" "wasn't" "weren't" "hasn't" "haven't" "hadn't" "doesn't"
[89] "don't" "didn't" "won't" "wouldn't" "shan't" "shouldn't" "can't" "cannot" "couldn't" "mustn't" "let's"
[100] "that's" "who's" "what's" "here's" "there's" "when's" "where's" "why's" "how's" "a" "an"
[111] "the" "and" "but" "if" "or" "because" "as" "until" "while" "of" "at"
[122] "by" "for" "with" "about" "against" "between" "into" "through" "during" "before" "after"
[133] "above" "below" "to" "from" "up" "down" "in" "out" "on" "off" "over"
[144] "under" "again" "further" "then" "once" "here" "there" "when" "where" "why" "how"
[155] "all" "any" "both" "each" "few" "more" "most" "other" "some" "such" "no"
[166] "nor" "not" "only" "own" "same" "so" "than" "too" "very" "will"

```

FONTE: O autor (2018).

Cabe salientar que é imprescindível a transformação dos caracteres em letras minúsculas (linha 61), caso contrário o algoritmo pode considerar que *strings* escritas com alternância da fonte (maiúsculas/minúsculas) são diferentes.

Com os dados limpos, o próximo passo é a organização dos mesmos. Para tal, os termos são organizados em uma matriz através da função *TermDocumentMatrix*. As demais funções utilizadas servem para classificar os termos em ordem decrescente em um *Data Frame*, Figura 18. Para somar a quantidade de palavras, é utilizada a função 'sum(matriz\$freq)'.

FIGURA 18 - ENTRADA E SAÍDA DA ETAPA DE ORGANIZAÇÃO DOS DADOS

Entrada	Saída
<pre> #TermDocumentMatrix & DocumentTermMatrix (ORGANIZAR DADOS) freq <- TermDocumentMatrix(corpus) matriz <- as.matrix(freq) matriz <- sort(rowSums(matriz),decreasing=T) matriz <- data.frame(word=names(matriz),freq=matriz) #matriz head(matriz, n=10) #tail(matriz) </pre>	<pre> > head(matriz, n=10) power power 5758 vehicle vehicle 4789 system system 3639 device device 3403 electric electric 3088 battery battery 2701 unit unit 2594 charging charging 2113 gear gear 2112 control control 1957 </pre>

FONTE: O autor (2018).

Como resultado, é gerada a lista com a frequência absoluta dos termos. A partir desta lista, são gerados os gráficos de barras (A) e a nuvem de palavras (B), Figura 19. Nuvens de palavras são visualizações que exibem as palavras que se repetem frequentemente em um texto. Para Fleuren e Alkema (2015), elas são particularmente úteis quando não se tem ideia preconcebida dos conceitos que devem ocorrer no conteúdo dos documentos analisados. Nesta expressão gráfica, as palavras com maiores frequência são plotadas em fontes maiores do que as palavras de menor frequência.

FIGURA 20 - MATRIZ DOS TERMOS DO DOCUMENTO

Entrada	<pre>#ANALISAR FREQUÊNCIAS dtm <- DocumentTermMatrix(corpus) inspect(dtm) dtm nDocs(dtm) nTerms(dtm)</pre>
Saída	<pre>> dtm <- DocumentTermMatrix(corpus) > inspect(dtm) <<DocumentTermMatrix (documents: 5, terms: 8398)>> Non-/sparse entries: 18871/23119 Sparsity : 55% Maximal term length: 36 weighting : term frequency (tf) sample : Terms Docs BR VG 2012.txt battery charging control device electric gear power system unit vehicle 642 439 313 642 745 324 1173 715 557 860 BR VG 2013.txt 370 348 346 666 549 463 769 645 443 984 BR VG 2014.txt 709 350 460 580 713 367 1234 751 545 1060 BR VG 2015.txt 480 530 373 814 545 467 1310 805 599 849 BR VG 2016.txt 500 446 465 701 536 491 1272 723 450 1036 > dtm <<DocumentTermMatrix (documents: 5, terms: 8398)>> Non-/sparse entries: 18871/23119 Sparsity : 55% Maximal term length: 36 weighting : term frequency (tf) > nDocs(dtm) [1] 5 > nTerms(dtm) [1] 8398</pre>

FONTE: O autor (2018).

Ao inspecionar a variável **dtm** outras informações são geradas, como o número de documentos analisados (5) e a quantidade de termos (8398). Além disso, ele informa a *sparsity* (esparcidade) que é o número de posições na matriz cujo valor é nulo. Neste caso, 55% (23119) das posições possuem valor 0. Para remover esta esparcidade pode ser utilizada a função *'removeSparseTerms'* que serve como um filtro adicional na análise, Figura 21.

FIGURA 21 - FILTRO REMOVEPARSETERMS

Entrada	<pre>fgh <- dtm removeSparseTerms(fgh, 0.8) removeSparseTerms(fgh, 0.6) removeSparseTerms(fgh, 0.4) removeSparseTerms(fgh, 0.2)</pre>
Saída	<pre>> removeSparseTerms(fgh, 0.8) <<DocumentTermMatrix (documents: 5, terms: 8398)>> Non-/sparse entries: 18871/23119 Sparsity : 55% Maximal term length: 36 weighting : term frequency (tf) > removeSparseTerms(fgh, 0.6) <<DocumentTermMatrix (documents: 5, terms: 2854)>> Non-/sparse entries: 11995/2275 Sparsity : 16% Maximal term length: 26 weighting : term frequency (tf) > removeSparseTerms(fgh, 0.4) <<DocumentTermMatrix (documents: 5, terms: 2045)>> Non-/sparse entries: 9568/657 Sparsity : 6% Maximal term length: 18 weighting : term frequency (tf) > removeSparseTerms(fgh, 0.2) <<DocumentTermMatrix (documents: 5, terms: 1388)>> Non-/sparse entries: 6940/0 Sparsity : 0% Maximal term length: 16 weighting : term frequency (tf)</pre>

FONTE: O autor (2018).

O *'removeSparseTerms'* se mostra bastante útil para eliminar os termos que aparecem em poucos documentos, principalmente quando o volume de arquivos analisados for substancialmente alto. Já a função *'inspect'*, também permite a consulta dos termos em intervalos específicos (Figura 22).

FIGURA 22 - CONSULTA FRACIONADA UTILIZANDO FUNÇÃO 'INSPECT'

```
> inspect(dtm[1:5, 100:105])
<<DocumentTermMatrix (documents: 5, terms: 6)>>
Non-/sparse entries: 24/6
Sparsity           : 20%
Maximal term length: 9
weighting          : term frequency (tf)
sample            :
Terms
Docs  acting action actions activate activated activates
BR VG 2012.txt  1    10      2      0      14      0
BR VG 2013.txt  3    10      0      11     15      3
BR VG 2014.txt  5     6      0      1     10      4
BR VG 2015.txt  7     1      0      3     18      2
BR VG 2016.txt  1     6      0      3     8       3
```

FONTE: O autor (2018).

Dando sequência na análise, a frequência absoluta de alguns termos pode indicar a tendência das tecnologias desenvolvidas. Termos específicos podem ser consultados também com a função ‘inspect’, Figura 23.

FIGURA 23 - CONSULTA DE TERMOS ESPECÍFICOS

```
> inspect(DocumentTermMatrix(corpus, list(dictionary= c("wireless","battery","hybrid"))))
<<DocumentTermMatrix (documents: 5, terms: 3)>>
Non-/sparse entries: 15/0
Sparsity           : 0%
Maximal term length: 8
Weighting          : term frequency (tf)
Sample            :
      Terms
Docs   battery hybrid wireless
BR VG 2012.txt    642   162     62
BR VG 2013.txt    370   205     69
BR VG 2014.txt    709   269    111
BR VG 2015.txt    480   230    202
BR VG 2016.txt    500   256    158
```

FONTE: O autor (2018).

Para calcular a frequência relativa dos termos foi utilizada a Equação 1, onde é dividida a frequência absoluta do termo ‘n’ pela quantidade total de palavras da amostra.

$$Freq_{rel} = \frac{Freq_{abs} \text{ do termo } n}{Quantidade \text{ total de palavras}} \quad (1)$$

Para o cálculo da frequência relativa é necessário saber a quantidade total de palavras. Para isto, foi adicionada a função de soma na variável **dtm** e os resultados são inseridos dentro do vetor **palavras**. Com a quantidade de palavras total e a matriz dos termos é possível gerar a matriz de frequência relativa e, assim como nos exemplos anteriores também há a possibilidade de consultar termos de forma individual, Figura 24.

FIGURA 24 - CÁLCULO DA MATRIZ DE FREQUÊNCIA RELATIVA

```

> TotalPalavras2012 <- sum(dtm[1:1, 1:8398])
> TotalPalavras2012
[1] 43359
> TotalPalavras2013 <- sum(dtm[2:2, 1:8398])
> TotalPalavras2013
[1] 42354
> TotalPalavras2014 <- sum(dtm[3:3, 1:8398])
> TotalPalavras2014
[1] 45640
> TotalPalavras2015 <- sum(dtm[4:4, 1:8398])
> TotalPalavras2015
[1] 50823
> TotalPalavras2016 <- sum(dtm[5:5, 1:8398])
> TotalPalavras2016
[1] 49263
> TotalPalavras <- TotalPalavras2012 + TotalPalavras2013 + TotalPalavras2014 + TotalPalavras2015 + TotalPalavras2016
> TotalPalavras
[1] 231439
>
> palavras <- c(TotalPalavras2012, TotalPalavras2013, TotalPalavras2014, TotalPalavras2015, TotalPalavras2016)
>
> #freqrel <- inspect(DocumentTermMatrix(corpus, list(dictionary= c("wireless")))) #consulta de termos individuais
> freqrel <- inspect(dtm)
<<DocumentTermMatrix (documents: 5, terms: 8398)>>
Non-/sparse entries: 18871/23119
Sparsity : 55%
Maximal term length: 36
weighting : term frequency (tf)
Sample :
      Terms
Docs  battery charging control device electric gear power system unit vehicle
BR VG 2012.txt 642 439 313 642 745 324 1173 715 557 860
BR VG 2013.txt 370 348 346 666 549 463 769 645 443 984
BR VG 2014.txt 709 350 460 580 713 367 1234 751 545 1060
BR VG 2015.txt 480 530 373 814 545 467 1310 805 599 849
BR VG 2016.txt 500 446 465 701 536 491 1272 723 450 1036
>
> freqrel <- freqrel/palavras
> freqrel
      Terms
Docs  battery charging control device electric gear power system unit vehicle
BR VG 2012.txt 0.014806615 0.010124772 0.007218801 0.01480661 0.01718213 0.007472497 0.02705321 0.01649023 0.012846237 0.01983441
BR VG 2013.txt 0.008735893 0.008216461 0.008169240 0.01572461 0.01296218 0.010931671 0.01815649 0.01522879 0.010459461 0.02323275
BR VG 2014.txt 0.015534619 0.007668712 0.010078878 0.01270815 0.01562226 0.008041192 0.02703769 0.01645486 0.011941280 0.02322524
BR VG 2015.txt 0.009444543 0.010428349 0.007339197 0.01601637 0.01072349 0.009188753 0.02577573 0.01583929 0.011786002 0.01670504
BR VG 2016.txt 0.010149605 0.009053448 0.009439133 0.01422975 0.01088038 0.009966912 0.02582060 0.01467633 0.009134645 0.02102998

```

FONTE: O autor (2018).

Assim sendo, com os dados já estruturados, a etapa de análise dos indicadores pode ser iniciada. O desenvolvimento do algoritmo foi de suma importância para a condução da pesquisa, uma vez que a linguagem de programação em R trouxe flexibilidade nas suas funcionalidades além de já possuir pacotes específicos voltado ao tratamento e manipulação de *strings*.

5 APRESENTAÇÃO DOS RESULTADOS

Este capítulo visa analisar o conteúdo dos depósitos em quatro aspectos, sendo o primeiro constituído pela análise dos volumes de depósitos nos BRICS. No segundo e terceiro passos, foram elencados os principais códigos IPC (com auxílio dos códigos manuais da base *Derwent Innovations Index*®) e os principais depositantes em cada país. Por fim, foi utilizado o algoritmo, apresentado na seção 4.1, para gerar os indicadores relacionais de segunda geração no Brasil, indicadores estes, que foram comparados aos indicadores das patentes tetrádicas no mesmo período.

5.1 INDICADORES DE ATIVIDADE

Conforme descrito na seção 2.6, estes indicadores tem como objetivo medir as forças e fraquezas da tecnologia. Sua utilização permite identificar os principais *players* e dão um indicativo dos focos em P&D no setor.

5.1.1 Volume de depósitos

A subcategoria “Veículos em Geral” engloba tecnologias e códigos referentes a sistemas de controle, embreagens, motores sem escova, sistemas de freio regenerativo, propulsão elétrica advindas de forças da natureza, propulsão elétrica com fonte externa ao veículo, propulsão elétrica através de células combustíveis, motores de combustão operando com combustíveis gasosos (exemplo: veículos movidos a hidrogênio) e estações de recarga para veículos elétricos.

Foram feitas buscas na base patentária *Derwent Innovations Index*®, conforme esquema já apresentado na Tabela 2, e os resultados obtidos se encontram na Tabela 3.

TABELA 2 - PATENTES DEPOSITADAS (VEÍCULOS EM GERAL)

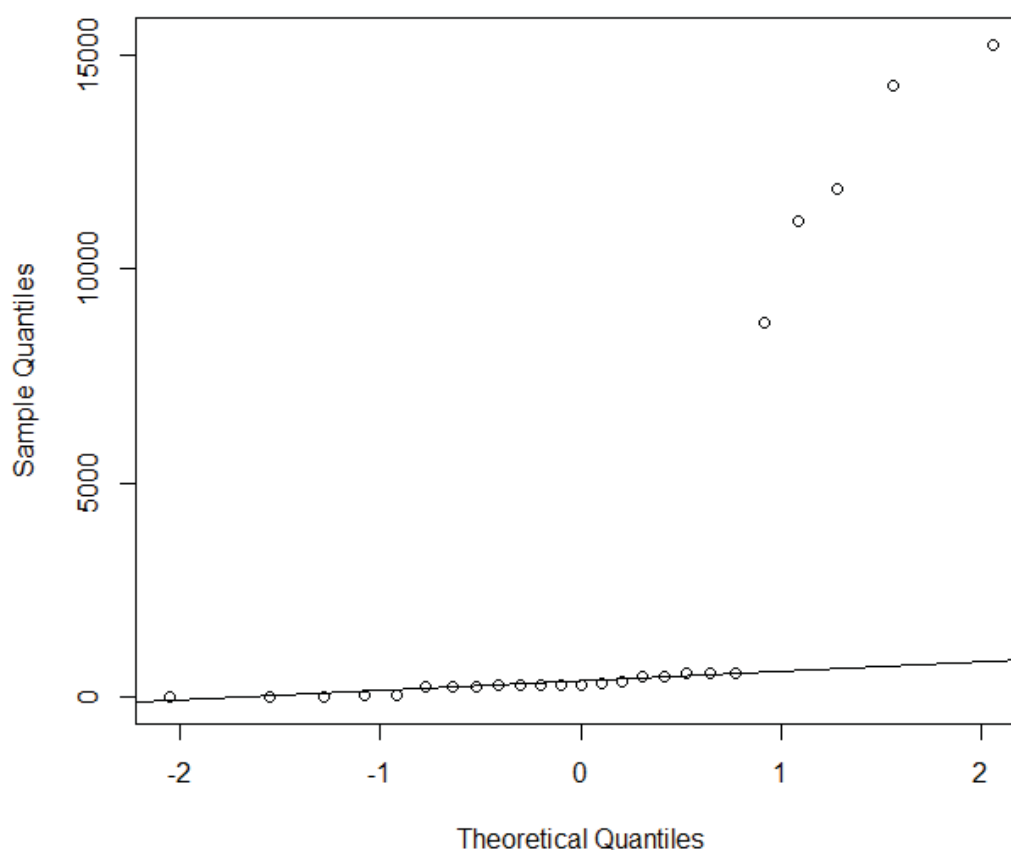
ANO	Brasil	Rússia	Índia	China	África do Sul
2012	253	310	554	8735	30
2013	237	352	470	11123	39
2014	253	298	480	11884	37
2015	292	300	552	14283	24
2016	296	282	570	15252	6
TOTAL	1331	1542	2626	61277	136

FONTE: O autor (2018).

Após o levantamento dos dados, foi realizada a análise estatística. Primeiramente foi feita a análise de normalidade com os dados dos cinco países pertencentes ao bloco, Figura 25. O resultado mostra que, através do método de Kolmogorov-Smirnov, temos um p-valor menor do que 1%, portanto podemos afirmar que os dados não são normalmente distribuídos. Foram excluídos os dados do escritório chinês, pois o volume de depósitos/patentes neste país é muito superior aos dos demais escritórios do bloco.

FIGURA 25 - TESTE DE NORMALIDADE BRICS

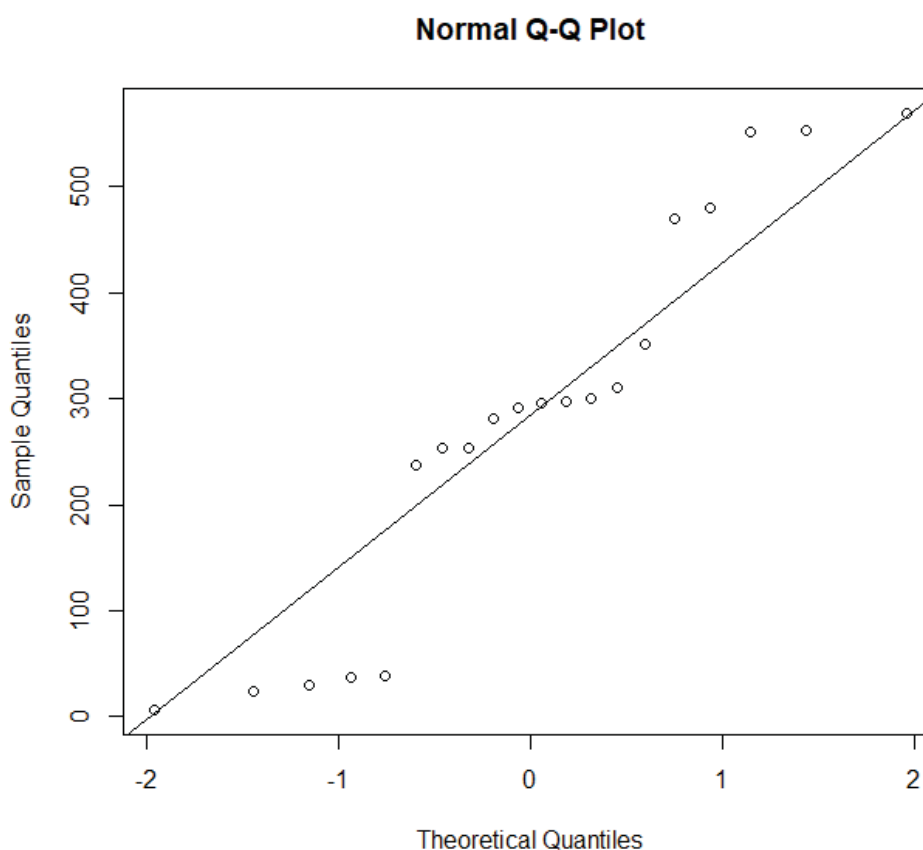
Normal Q-Q Plot



FONTE: O autor (2019).

O segundo teste de normalidade considerando apenas os dados de Brasil, Rússia, Índia e África do Sul, utilizando o mesmo método anterior, resultou em dados normalmente distribuídos, Figura 26.

FIGURA 26 - TESTE DE NORMALIDADE BRASIL, RÚSSIA, ÍNDIA E ÁFRICA DO SUL



FONTE: O autor (2019).

Sendo os dados normalmente distribuídos, o segundo pressuposto para a análise de variância (ANOVA) é de que deve haver homocedasticidade. Neste caso foi utilizado o método de Bartlett com um intervalo de confiança de 95%. O resultado obtido foi um p-valor de 0,1627, Figura 23. Sendo assim, podemos concluir que não há heterocedasticidade nos dados analisados permitindo que seja feita a análise de variância.

FIGURA 27 - TESTE DE BARTLETT

```
> bartlett.test(response ~ factor, data1)

Bartlett test of homogeneity of variances

data: response by factor
Bartlett's K-squared = 5.128, df = 3, p-value = 0.1627
```

FONTE: O autor (2019).

O teste da ANOVA indicou através do p-valor que há diferença entre os países. O valor encontrado foi muito próximo ao zero conforme mostra o resultado na Figura 28. A primeira linha do código indicado serve para selecionar o arquivo de texto (.txt) onde se encontram os dados da pesquisa, na sequência é executado o teste com o comando para mostrar a tabela da ANOVA.

FIGURA 28 – ANOVA

```
> dados <- read.table(file.choose(), header=TRUE)
> dados.anova <- aov(dados$Patentes ~ dados$País)
> summary(dados.anova)

      Df Sum Sq Mean Sq F value    Pr(>F)
dados$País    3  625078   208359    224 2.82e-13 ***
Residuals   16  14882     930

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FONTE: O autor (2019).

Pelo método de Tukey observamos a formação de três grupos distintos, sendo o primeiro deles composto apenas pela Índia, o segundo por Brasil e Rússia e um terceiro grupo representado pela África do Sul, Figura 29. O intervalo de confiança adotado foi de 95%.

FIGURA 29 - RESULTADO ANOVA

```
> TukeyHSD(dados.anova)
Tukey multiple comparisons of means
95% family-wise confidence level

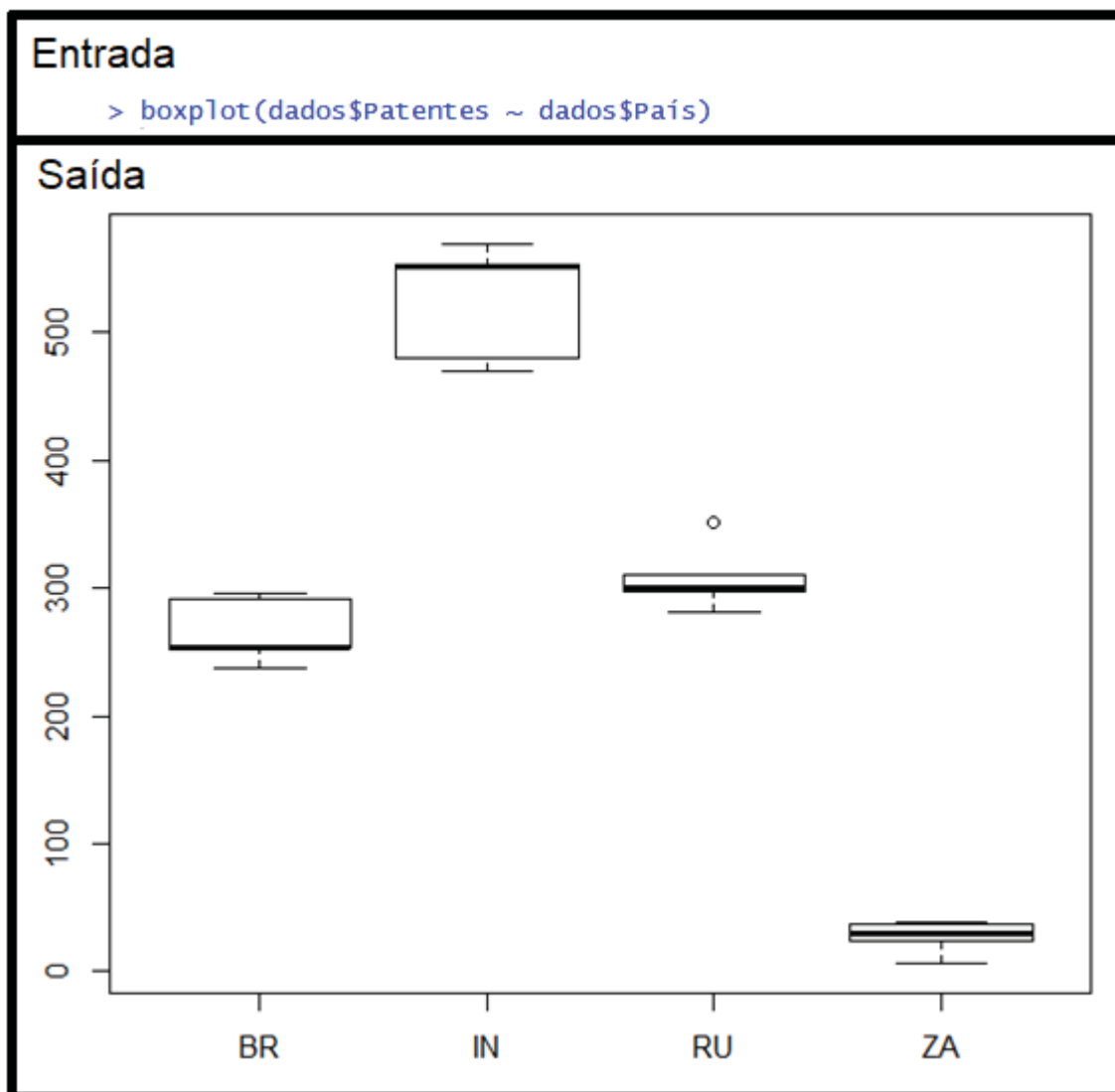
Fit: aov(formula = dados$Patentes ~ dados$País)

$`dados$País`
      diff          lwr          upr      p adj
IN-BR  259.0    203.81568    314.18432 0.0000000
RU-BR   42.2    -12.98432     97.38432 0.1688327
ZA-BR -239.0   -294.18432   -183.81568 0.0000000
RU-IN -216.8   -271.98432   -161.61568 0.0000000
ZA-IN -498.0   -553.18432   -442.81568 0.0000000
ZA-RU -281.2   -336.38432   -226.01568 0.0000000
```

FONTE: O autor (2019).

Após rodar o teste de Tukey, foi gerado o gráfico do tipo boxplot com a mesma entrada de dados, Figura 30. Através da análise gráfica podemos confirmar o dado do 'p adj' apresentado no teste.

FIGURA 30 - BOXPLOT TESTE DE TUKEY



FONTE: O autor (2019).

Assim, podemos afirmar que Brasil e Rússia possuem características muito semelhantes, sendo a Índia se destacando dos demais e a África do Sul ficando com indicadores abaixo de todos os países do bloco. A China possui indicadores tão elevados que precisou ser retirada dos cálculos devido às distorções que ocasionava.

Apesar de algumas semelhanças, ao analisarmos as famílias de patentes (documentos depositados em mais de um país, neste caso o comparativo entre os

BRICS), podemos perceber alguns padrões, Tabela 4. Aqui foram adicionados os dados dos documentos depositados concomitantemente nos escritórios dos Estados Unidos, Japão, Europeu e Chinês (patentes tetrádicas), que no período (2012 – 2016) alcançaram a marca de 5569 registros. Dos 1331 documentos encontrados no Brasil, aproximadamente 83% deles também estão depositados na China. Fazendo um comparativo com as patentes tetrádicas, este número chega a 55%.

TABELA 3 - FAMÍLIAS DE PATENTES

ANO	Brasil	Rússia	Índia	China	África do Sul	Tetrádicas
Brasil		488	793	1104	69	738
Rússia	488		518	984	43	588
Índia	793	518		2217	76	1264
China	1104	984	2217		111	5569
África do Sul	69	43	76	111		76
Tetrádicas	738	588	1264	5569	76	

FONTE: O autor (2019).

Portanto, existe uma gama de tecnologias com alto valor comercial que não estão resguardadas judicialmente em território nacional. Este fenômeno pode ser explicado pelas diferentes legislações vigentes, custos operacionais, estratégia das empresas e nível de concorrência no mercado local. O mercado indiano, assim como o brasileiro, também possui boa parte de suas patentes (84%) depositadas na China.

5.1.2 Códigos IPC

A análise do Quadro 3 indica grande presença do código IPC H02J-007/00 nas patentes depositadas nos BRICS, ele representa as disposições de circuitos para carregar ou despolarizar baterias ou para alimentar o carregamento de baterias. (INPI, 2018). Na sequência deste IPC temos o código B60L-011/18 que representa a propulsão elétrica com fonte de energia no interior do veículo. Estes códigos, conforme confirmado por Assis (2018), fazem parte do *cluster* de códigos relacionados aos veículos elétricos puros.

QUADRO 3 - IPC'S DE MAIOR FREQUÊNCIA NA AMOSTRA

Rank	BRASIL		RÚSSIA		ÍNDIA		CHINA		ÁFRICA DO SUL	
1º	H02J-007/00	42,8%	H02J-007/00	32,2%	H02J-007/00	49,7%	H02J-007/00	64,2%	H02J-007/00	40,4%
2º	B60L-011/18	28,1%	B60L-011/18	21,9%	B60L-011/18	28,1%	B60L-011/18	17,3%	H02J-000/00	22,8%
3º	B60W-020/00	16,7%	B60W-020/00	18,1%	H02J-007/02	18,1%	B60W-020/00	6,2%	B60L-011/18	17,6%
4º	H02J-007/02	14,5%	B60W-010/08	12,6%	B60W-020/00	13,6%	H02J-007/02	5,9%	B60L-000/00	15,4%
5º	B60W-010/08	13,2%	B60W-010/06	12,1%	H01M-010/44	10,4%	H01M-010/44	5,5%	B60K-000/00	13,2%
6º	B60W-010/06	11,2%	H02J-007/02	8,4%	B60W-010/08	9,9%	H02J-017/00	4,5%	F16H-000/00	11,8%
7º	B60L-003/00	9,1%	B60K-006/48	7,4%	B60L-003/00	9,9%	B60W-010/08	4,4%	H02J-007/02	11,0%
8º	H02J-017/00	8,6%	F02M-021/02	6,9%	H02J-017/00	9,0%	B60W-010/06	3,9%	B60W-020/00	10,3%
9º	H01M-010/44	8,2%	H01M-010/44	6,9%	H01M-010/48	8,3%	H01M-010/48	3,5%	H02J-003/38	9,6%
10º	H02J-005/00	7,4%	B60L-003/00	6,3%	H02J-050/12	8,1%	B60L-003/00	3,5%	B60L-009/00	8,1%

FONTE: O autor (2018).

Os IPCs da subclasse B60W se referem às tecnologias para veículos híbridos, o grupo de maior destaque na lista é o B60W-020/00 que abrange os sistemas de controle adaptados a veículos desta natureza. Outros destaques dentro desta subclasse são os códigos B60W-010/08 e B60W-010/06 que contém tecnologias de controle conjugado para subunidade de veículos de tipo ou função diferente, sendo incluídos no subgrupo 08 o controle de unidades com propulsão elétrica e no subgrupo 06 o controle de motores a combustão.

Para efeitos de comparação, foi gerada a lista dos códigos manuais da *Derwent Innovations Index*®, estes códigos são atribuídos às patentes por indexadores da base. Eles são comumente usados para indicar não só as aplicações como também os aspectos técnicos de inovação das invenções. (CLARIVATE ANALYTICS, 2018). O Quadro 4 lista os códigos das principais invenções por país.

QUADRO 4 - PERCENTUAL DOS DEPÓSITOS POR CÓDIGO IPC

Rank	BRASIL		RÚSSIA		ÍNDIA		CHINA		ÁFRICA DO SUL	
1º	X21-A01F	15,9%	X21-A01F	15,0%	X21-A01F	19,1%	X16-G	44,7%	X16-G	11,0%
2º	T01-J07D1	13,1%	X16-G	13,1%	X16-G	15,3%	X21-A01F	19,7%	X21-A01F	9,6%
3º	X21-A01D	13,1%	X21-A01D	10,8%	X21-A01D	13,9%	X16-B01	13,0%	U24-H	5,9%
4º	X16-G	12,6%	X22-P04A	10,2%	T01-J07D1	10,5%	T01-J07D1	10,2%	T01-J07D1	4,4%
5º	X21-A02A	9,9%	X21-A02A	9,0%	X22-P04A	10,2%	X21-B01A1	10,2%	W01-C01D3C	4,4%
6º	X22-P04A	9,8%	T01-J07D1	8,4%	X16-B01	8,6%	X21-A01D	6,9%	X21-B03	4,4%
7º	X21-A01D1	9,2%	X21-A01D1	8,3%	X21-A01D1	7,7%	W01-C01D3C	6,7%	T01-F05B2	3,7%
8º	T01-S03	8,9%	X16-B01	6,9%	X21-B01A1	7,5%	T01-C07C4A	5,9%	X22-A20C	3,7%
9º	X21-B01A1	7,0%	X22-G01	5,3%	X16-G01	6,7%	X16-F06A	5,2%	X22-G01	3,7%
10º	X16-B01	5,9%	X21-B01A1	5,3%	X21-A02A	6,4%	X16-B01F1	5,0%	X22-G03	3,7%

FONTE: O autor (2018).

Assim como nos códigos IPC, os indexadores da base também apresentam um padrão nas publicações. O indexador X21-A01F é referente a propulsão e frenagem de veículos elétricos e possui maior volume que o indexador dos veículos híbridos (X21-A01D). Brasil, Rússia, Índia e China possuem proporções semelhantes, já a África do Sul apresenta o indexador de veículos híbridos apenas na 15ª posição (2,9%).

Entretanto, o grande *gap* existente entre os depósitos na China e os demais países dos BRICS, pode ser explicado quando observamos as tecnologias voltadas às “Estações de recarga de veículos elétricos”, Tabela 5. O alto volume de depósitos, aliado a força da movimentação em atividade tecnológica neste grupo, é um forte indicativo de que as estações de recarga devem ser protagonistas na era dos veículos elétricos.

TABELA 4 - DEPÓSITOS PARA ESTAÇÕES DE RECARGA DE VEÍCULOS ELÉTRICOS

ANO	Brasil	Rússia	Índia	China	África do Sul
2012	130	117	257	5322	13
2013	87	108	185	7036	9
2014	104	110	244	7833	19
2015	136	93	314	9487	12
2016	112	67	304	9644	2
TOTAL	569	495	1304	39322	55

FONTE: O autor (2018).

Enquanto na China o desenvolvimento e busca por proteção deste tipo de tecnologia quase dobrou no intervalo de 5 anos, Brasil, Rússia, Índia e África do Sul tiveram pequenas flutuações nos depósitos. Estes indicadores reafirmam a força do mercado chinês e apontam o foco dos esforços de P&D das empresas do setor.

Os códigos manuais do *Derwent* demonstraram mais força para as tecnologias de veículos elétricos e, aliados aos indicadores de atividade para as estações de recarga, demonstram uma maior tendência de desenvolvimento para veículos deste tipo. Todavia, é importante analisar a dispersão das patentes nos demais *clusters* de IPC (ANEXO III).

Dentre os demais *clusters*, podemos citar o de propulsão elétrica com fonte de energia no exterior do veículo através de células de combustível, Tabela 6.

TABELA 5 - PROPULSÃO ELÉTRICA COM FONTE DE ENERGIA NO EXTERIOR DO VEÍCULO ATRAVÉS DE CÉLULAS COMBUSTÍVEL

ANO	Brasil	Rússia	Índia	China	África do Sul
2012	83	72	178	1475	4
2013	54	67	139	1799	11
2014	92	51	145	1738	5
2015	75	61	143	2129	3
2016	70	85	132	3440	1
TOTAL	374	336	737	10581	24

FONTE: O autor (2018).

Este grupo de patentes utiliza, por exemplo, células de combustível com reagentes de hidrogênio e oxigênio. Segundo Kaya e Hames (2019), células de combustível de hidrogênio possuem alta eficiência energética e baixa emissão, contribuindo assim para um desenvolvimento tecnológico mais limpo. Por outro lado, estas células apresentam alguns problemas, como a baixa densidade de potência e resposta lenta. Uma das soluções apresentadas é a utilização conjunta com outros tipos de armazenamento de energia, tais como baterias e supercapacitores.

Demais tabelas com os grupos de códigos descritos no ANEXO III podem ser consultadas no Apêndice II deste trabalho.

Outra função inserida no código é a busca por expressões regulares, também conhecida por *Regex Matches*. Este bloco de código é bastante útil, quando se encontra dentro de um documento de texto dados semi-estruturados. No exemplo aplicado foram extraídas as datas de depósito dos documentos pesquisados na base, Figura 31.

FIGURA 31 - BUSCA POR EXPRESSÕES REGULARES

```
#EXTRAÇÃO DAS DATAS DE DEPÓSITO
s <- corpus
ipc <- stringr::str_extract_all(s, "\\s\\d\\d\\d\\s*\\-\\d\\d\\d\\d\\d\\d\\d\\d\\d\\d\\s*")
ipc

ipc <- stringr::str_extract_all(s, "b62k-019/40")
ipc <- ipc[!is.na(ipc)]
ipc

bra <- stringr::str_extract_all(s, "BR\\s+\\s* BR\\s+\\s* \\d\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\d\\d\\s*")
bra <- bra[!is.na(bra)]
bra
rus <- stringr::str_extract_all(s, "RU\\s+\\s* RU\\s+\\s* \\d\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\d\\d\\s*")
rus <- rus[!is.na(rus)]
rus
ind <- stringr::str_extract_all(s, "IN\\s+\\s* IN\\s+\\s* \\d\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\d\\d\\s*")
ind <- ind[!is.na(ind)]
ind
chn <- stringr::str_extract_all(s, "CN\\s+\\s* CN\\s+\\s* \\d\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\d\\d\\s*")
chn <- chn[!is.na(chn)]
chn
afr <- stringr::str_extract_all(s, "ZA\\s+\\s* ZA\\s+\\s* \\d\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\d\\d\\s*")
afr <- afr[!is.na(afr)]
afr
```

FONTE: O autor (2018).

A busca pelas expressões regulares em R funciona de forma semelhante a outras linguagens (python, C, C++, etc) e sistemas (Windows, Linux, etc). Após a importação do corpus e inserção do mesmo na variável 's', são utilizados os códigos relacionados a cada tipo de caractere. Este bloco de funções tem o objetivo de extrair as datas de depósito dos documentos que estão nos escritórios dos BRICS. As duas letras iniciais indicadas no parâmetro (BR, RU, IN, CN e ZA) se referem ao início do número da patente. Como o arquivo extraído apresenta duas vezes o número da patente, foi necessário prever na busca pela expressão regular esta característica dos documentos. Na sequência do código (`\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*`) foi inserida a formatação da data (Exemplo: 19 jun 2015).

5.1.3 Principais depositantes

Para compreender as diferenças na intensidade de movimentação tecnológica no setor, é necessário identificar quais são os principais *players*. O Anexo V apresenta a lista dos 25 maiores depositantes por país entre os anos de 2012 e 2016. Embora haja diferença no “ranqueamento”, algumas empresas se fazem presentes em mais de uma lista.

Fica clara a forte presença da Toyota ocupando as primeiras posições no Brasil, Rússia, Índia e China. A montadora de origem japonesa possui expressiva participação no depósito de patentes em território nacional, chegando a aproximadamente 7,5% do total no período. Scania e Nissan também ocupam posto de destaque com 6,5% e 6,0% respectivamente.

O mercado russo é fortemente alimentado por tecnologias da Ford, Nissan e Toyota. Já a amostra extraída da Índia apresenta uma empresa de tecnologia no primeiro posto, a Qualcomm é uma reconhecida empresa desenvolvedora de *chipsets* para automatização não só veicular como também de *smartphones*. O maior mercado automotivo do mundo é o que possui maior dispersão no número de depositantes, os 10 principais depositantes no escritório chinês acumulam 15% do total de depósitos, enquanto nos demais países esta soma atinge no mínimo 38%. A África do Sul, por possuir um baixo volume de documentos analisados, tem características de depositantes completamente distintas do restante dos BRICS. A

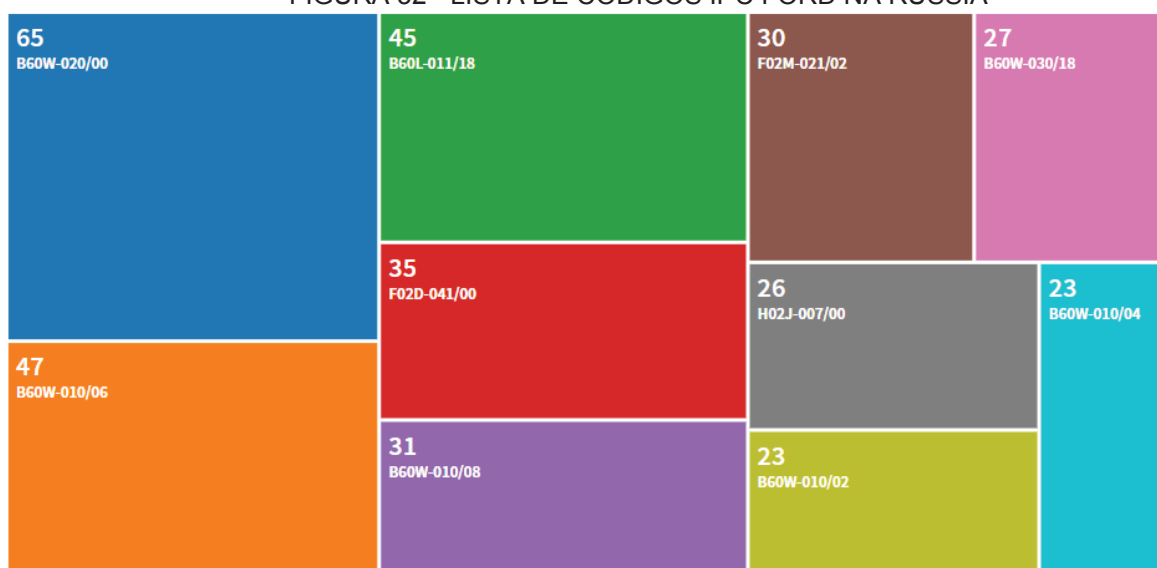
lista do país africano carece das principais montadoras mundiais nas primeiras posições.

Dentre os principais depositantes, a Toyota tem desenvolvido tecnologias de propulsão e frenagem tanto para veículos elétricos (31,31%) quanto para híbridos (29,29%) no Brasil. Na Rússia, a proporção fica 34,3% para veículos híbridos e 27,0% para veículos elétricos. Os depósitos da Toyota na Índia são semelhantes aos do Brasil (33,8% híbridos e 32,4% elétricos). Já no maior mercado mundial, o chinês, a montadora japonesa possui uma distribuição dos depósitos em 36,4% para veículos híbridos e 32,6% voltados aos veículos elétricos. Entretanto, dos 1584 registros encontrados na China, 75 estão depositados concomitantemente no Brasil, na Rússia e na Índia, demonstrando que muitas de suas tecnologias não ficam resguardadas judicialmente no restante dos BRICS.

Outro dos principais depositantes, a Nissan, está presente nas listas de Brasil, Rússia, Índia e China. Diferentemente da Toyota, a Nissan possui maior parte de seus depósitos voltados às tecnologias para veículos movidos a propulsão elétrica. No Brasil, a proporção dos depósitos é de 58,8% e 33,8% para veículos elétricos e híbridos, respectivamente. Enquanto na Rússia, fica em 52,2% com propulsão elétrica e 26,8% para híbridos. Já na Índia, a estratégia da empresa também está fortemente voltada ao desenvolvimento de veículos elétricos, chegando a 57,1% e 34,1% dos depósitos relacionados às tecnologias híbridas. Por fim, na China, apesar da maior dispersão, a estratégia se mantém a mesma com 44,2% e 25,2% para elétricos e híbridos.

A Ford, que ocupa o primeiro posto na lista russa, possui claramente uma estratégia mais voltada às tecnologias de veículos híbridos, Figura 32. Dentro dos 166 registros encontrados, após filtrar apenas os resultados da montadora, o maior destaque fica para o IPC B60W-020/00 (sistemas de controle adaptados a veículos híbridos), seguido do código B60W-010/06 (controle conjugado para subunidade de veículos de tipo ou função diferente incluindo o controle de motores a combustão) e B60L-011/18 (propulsão elétrica com fonte de energia no interior do veículo). Na China, a montadora norte-americana, também possui a maior parte de seus depósitos direcionados às tecnologias híbridas.

FIGURA 32 - LISTA DE CÓDIGOS IPC FORD NA RÚSSIA



FONTE: DERWENT (2019).

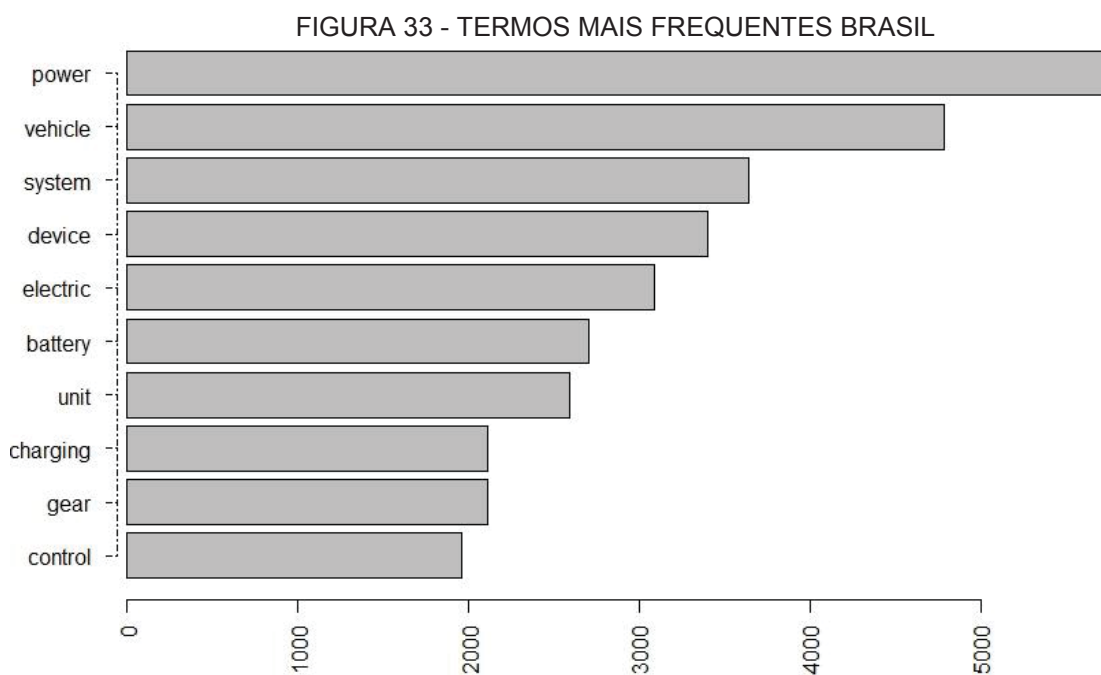
Um dos destaques das listas brasileira e russa é a Scania, a empresa sueca fundada em 1891 tem desenvolvido tecnologias para veículos híbridos. Foram encontrados 87 registros no Brasil e 50 na Rússia, destes, 48 estão depositados em ambos os países. Portanto, existe um indicativo de que a estratégia da empresa é utilizar a mesma tecnologia nos dois países.

5.2 INDICADORES RELACIONAIS DE SEGUNDA GERAÇÃO

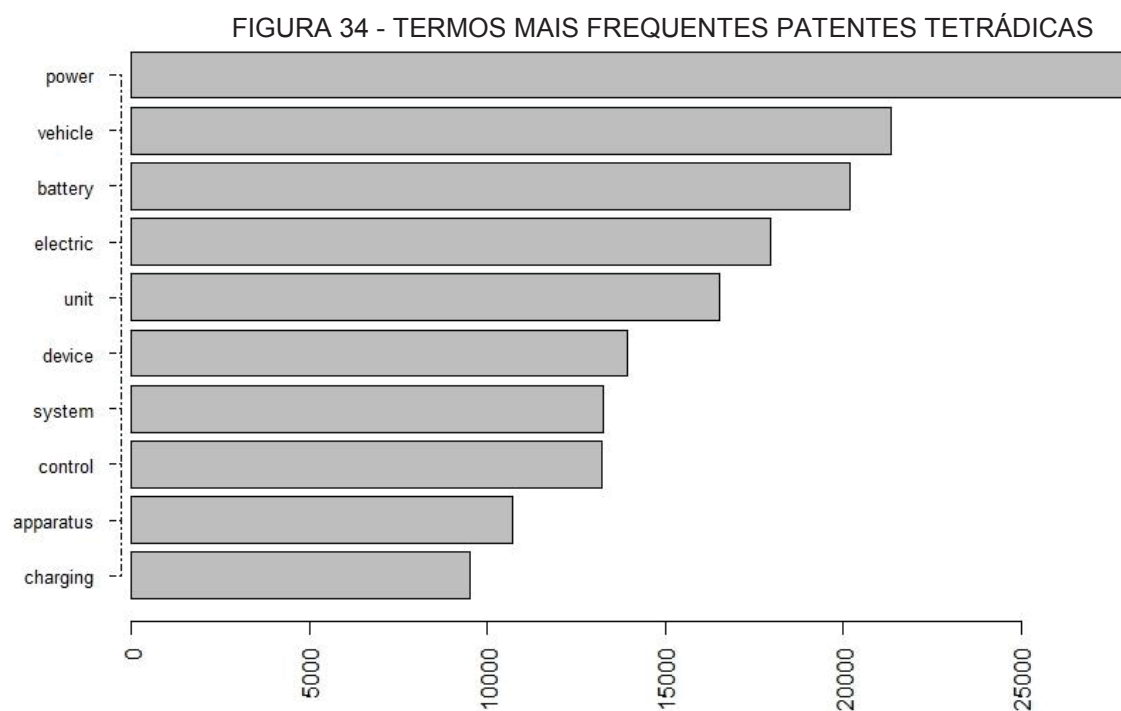
Dando seguimento à execução da pesquisa, este tópico visa analisar o conteúdo dos documentos extraídos da base patentária. Para tanto, como ferramenta auxiliar no processo, foi utilizado o código desenvolvido em linguagem R e os resultados já obtidos na seção 4.6. Esta análise é feita através das frequências absoluta e relativa dos termos.

Conforme definido na estratégia da pesquisa, foi feito um recorte anual na extração dos dados. Este método possibilitou identificar as variações dos depósitos e termos ao longo dos cinco anos analisados. Desta forma, foram gerados dez arquivos, cada um contendo o conteúdo do título e resumo dos documentos depositados no Brasil (cinco arquivos representando os anos de 2012 até 2016) e das patentes tetrádicas (cinco arquivos representando os anos de 2012 até 2016).

Portanto, os procedimentos adotados na seção 4.6 foram executados em cada arquivo gerado. De posse destes dados tratados, iniciou-se a análise da frequência dos termos. A lista dos dez termos mais frequentes no Brasil é apresentada na Figura 33 e a lista dos termos mais frequentes nas patentes tetrádicas está na Figura 34.



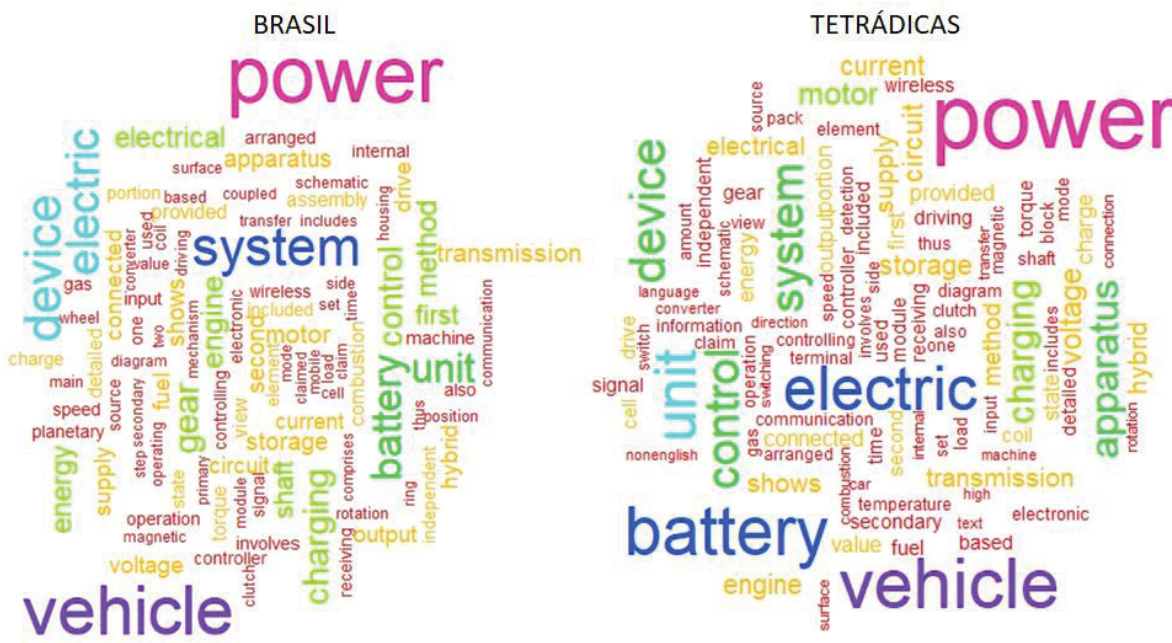
FONTE: O autor (2019).



FONTE: O autor (2019).

Como forma de representar graficamente a relação dos 100 termos mais frequentes em cada amostra, foi gerado o gráfico da nuvem de palavras com os dados acumulados de 2012 até 2016, Figura 35.

FIGURA 35 - NUVEM DE PALAVRAS



FONTE: O autor (2019).

A quantidade total de termos encontrados, nos documentos brasileiros, foi 8398 e o número absoluto de palavras 231439. Já nas patentes tetrádicas o número de termos foi 14939 e o número total de palavras 954811. O volume de documentos analisados nas patentes tetrádicas (5569) é bastante superior aos documentos brasileiros (1331), isto explica a grande diferença no número de termos e palavras. O Quadro 5 apresenta a quantidade absoluta de palavras ao longo dos anos.

QUADRO 5 - NÚMERO TOTAL DE PALAVRAS POR ANO

ANO	Brasil	Tetrádicas
2012	43.359	213.126
2013	42.354	223.070
2014	45.640	194.012
2015	50.823	183.880
2016	49.263	140.723

TOTAL **231.439** **954.811**

FONTE: O autor (2019).

O resultado desta etapa são duas matrizes, a que possui os documentos brasileiros possui 5 linhas e 8398 colunas, já a matriz referente às patentes tetrádicas tem 5 linhas e 14939 colunas, Figura 36.

FIGURA 36 - MATRIZ FREQUENCIA ABSOLUTA DOS TERMOS

BRASIL

Docs	Terms									
	battery	charging	control	device	electric	gear	power	system	unit	vehicle
BR VG 2012.txt	642	439	313	642	745	324	1173	715	557	860
BR VG 2013.txt	370	348	346	666	549	463	769	645	443	984
BR VG 2014.txt	709	350	460	580	713	367	1234	751	545	1060
BR VG 2015.txt	480	530	373	814	545	467	1310	805	599	849
BR VG 2016.txt	500	446	465	701	536	491	1272	723	450	1036

TETRÁDICAS

Docs	Terms									
	apparatus	battery	charging	control	device	electric	power	system	unit	vehicle
QD VG 2012.txt	2666	4804	2101	3246	2975	4408	5989	2705	3851	5138
QD VG 2013.txt	2471	5072	2202	3203	3380	4667	6600	3141	3823	5931
QD VG 2014.txt	2082	4073	2014	2607	2612	3526	5675	2753	3344	4039
QD VG 2015.txt	1947	3447	1830	2352	2776	3182	5685	2567	3368	3720
QD VG 2016.txt	1560	2791	1355	1801	2197	2190	4219	2099	2133	2506

FONTE: O autor (2019).

Contudo, a frequência absoluta pode mascarar algumas situações, como por exemplo, se o número total de documentos cresce, há uma tendência de também crescer a frequência absoluta dos termos. Assim, com estes dados disponíveis, criamos a matriz com a frequência relativa dos termos. Esta, tem como principal objetivo mostrar a força dos termos dentro do universo total de palavras e, tal como na primeira matriz, possui os mesmos números de linhas e colunas, Figura 37.

FIGURA 37 - MATRIZ FREQUENCIA RELATIVA DOS TERMOS

BRASIL

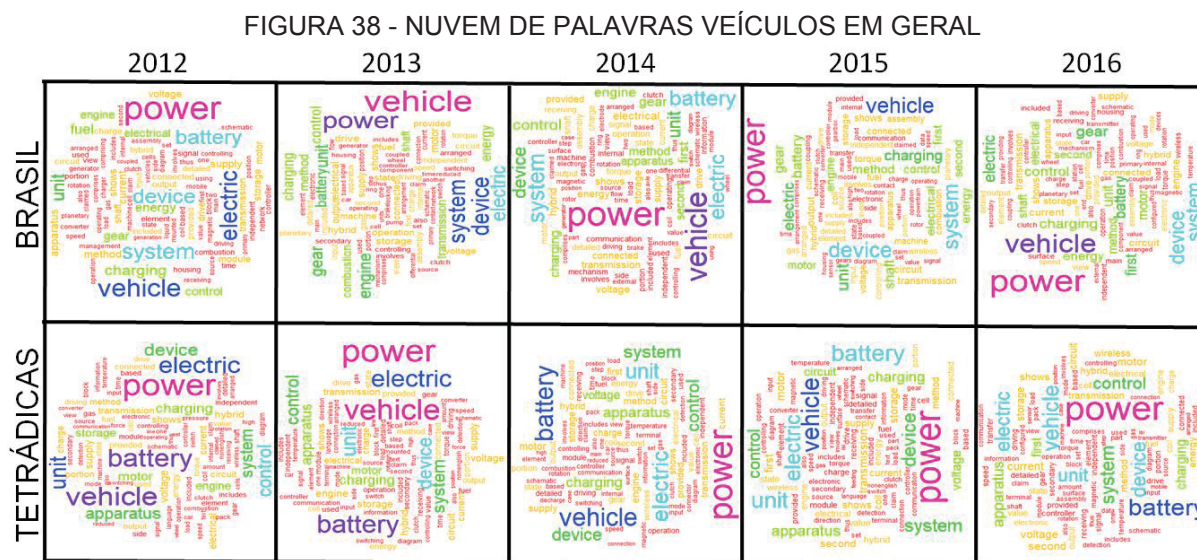
Docs	Terms									
	battery	charging	control	device	electric	gear	power	system	unit	vehicle
BR VG 2012.txt	0.014806615	0.010124772	0.007218801	0.01480661	0.01718213	0.007472497	0.02705321	0.01649023	0.012846237	0.01983441
BR VG 2013.txt	0.008735893	0.008216461	0.008169240	0.01572461	0.01296218	0.010931671	0.01815649	0.01522879	0.010459461	0.02323275
BR VG 2014.txt	0.015534619	0.007668712	0.010078878	0.01270815	0.01562226	0.008041192	0.02703769	0.01645486	0.011941280	0.02322524
BR VG 2015.txt	0.009444543	0.010428349	0.007339197	0.01601637	0.01072349	0.009188753	0.02577573	0.01583929	0.011786002	0.01670504
BR VG 2016.txt	0.010149605	0.009053448	0.009439133	0.01422975	0.01088038	0.009966912	0.02582060	0.01467633	0.009134645	0.02102998

TETRÁDICAS

Docs	Terms									
	apparatus	battery	charging	control	device	electric	power	system	unit	vehicle
QD VG 2012.txt	0.01250903	0.02254066	0.009858018	0.01523043	0.01395888	0.02068260	0.02810075	0.01269202	0.01806912	0.02410780
QD VG 2013.txt	0.01107724	0.02273726	0.009871341	0.01435872	0.01515219	0.02092168	0.02958713	0.01408078	0.01713812	0.02658807
QD VG 2014.txt	0.01073129	0.02099355	0.010380801	0.01343731	0.01346308	0.01817413	0.02925077	0.01418984	0.01723605	0.02081830
QD VG 2015.txt	0.01058843	0.01874592	0.009952143	0.01279095	0.01509680	0.01730476	0.03091690	0.01396019	0.01831629	0.02023059
QD VG 2016.txt	0.01108561	0.01983329	0.009628845	0.01279819	0.01561223	0.01556249	0.02998088	0.01491583	0.01515744	0.01780803

FONTE: O autor (2019).

Ao longo do tempo, a lista dos termos mais frequentes nas patentes depositadas no Brasil em comparação com as tetrádicas é muito semelhante, Figura 38. As flutuações no período analisado são sutis e há uma pequena variação na posição dos termos.

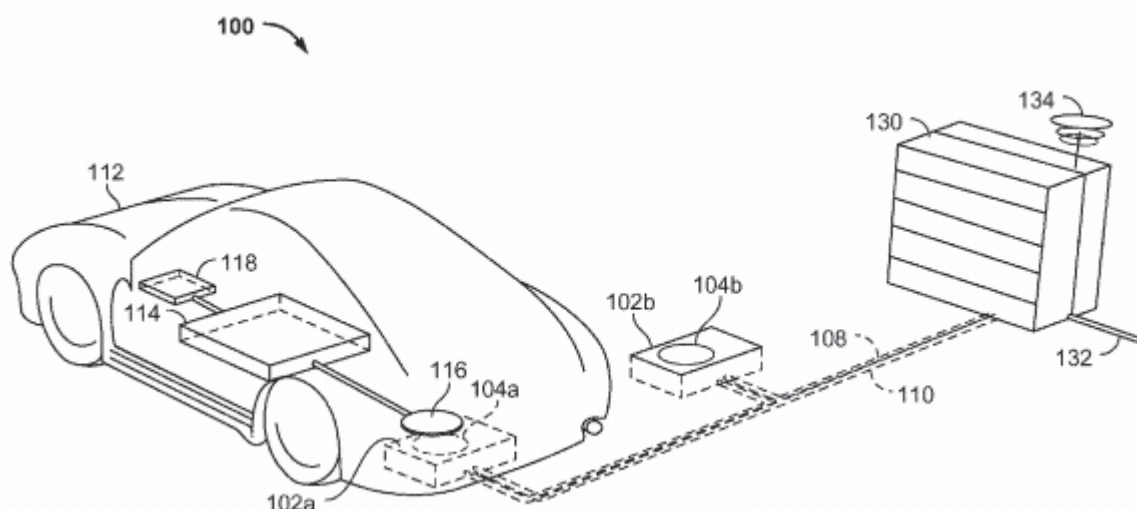


FONTE: O autor (2018).

Apesar da similaridade dos termos em ambos os conjuntos de dados, podemos perceber um primeiro indicativo de que a frequência do termo “battery” vem caindo tanto no Brasil quanto nas tetrádicas. No entanto, a frequência relativa nos últimos anos da amostra chega a ser o dobro nas tetrádicas em comparação com as brasileiras. Isto mostra que, mesmo perdendo força, as baterias ainda estão no foco dos desenvolvedores. Assim como visto na análise dos indicadores de atividade, mesmo os veículos elétricos alimentados por células de combustível, ainda é necessária a utilização das baterias como forma complementar no fornecimento de energia.

Uma análise mais profunda na matriz de frequência relativa apresentou um crescimento expressivo no termo *wireless*. Este termo quando relacionado a patentes de bateria e recarga, apresentou tecnologias como as da Qualcomm (2015) que permite recarregar a bateria através de um receptor sem fio, Figura 39. Neste exemplo, podemos ver que parte das tecnologias já utilizadas em *smartphones* está sendo migrada para o setor automotivo.

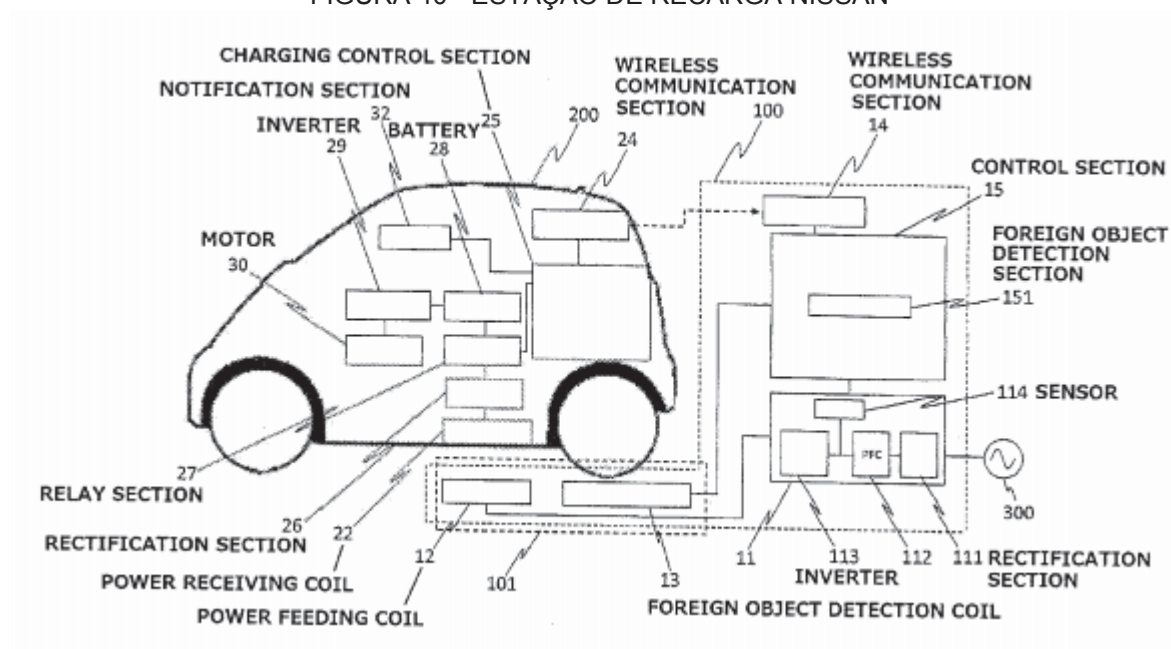
FIGURA 39 - EXEMPLO ESTAÇÃO DE RECARGA



FONTE: Qualcomm Incorporated (2015).

Outra empresa que também têm desenvolvido tecnologia similar a esta da Qualcomm é a Nissan. Conforme exposto, a montadora possui a maior parte de seus depósitos voltados para veículos elétricos, a Figura 40 ilustra uma de suas estações de recarga.

FIGURA 40 - ESTAÇÃO DE RECARGA NISSAN

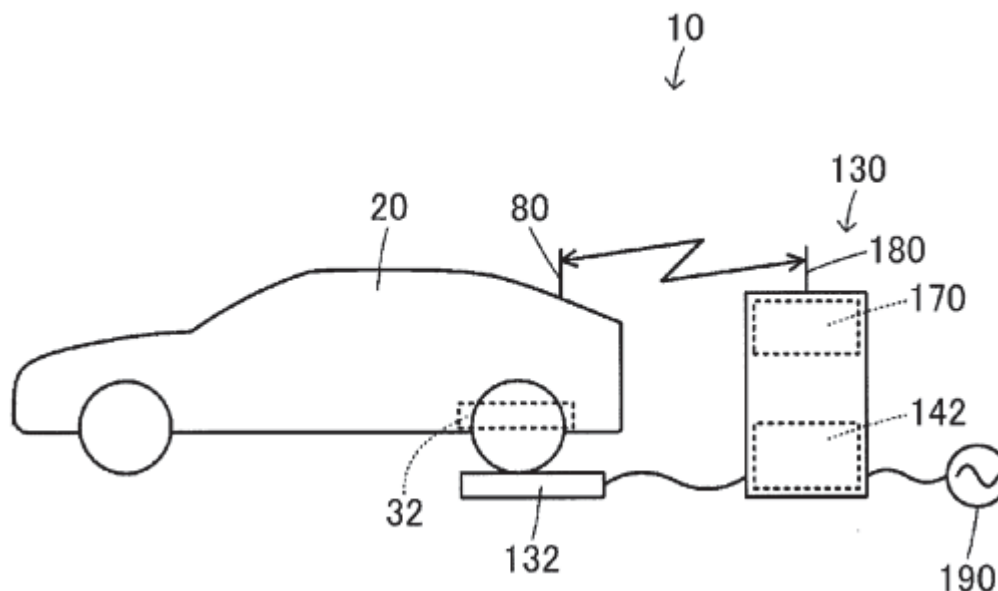


FONTE: Nissan Motor Co Ltd (2013).

Confirmando esta tendência, a Toyota também desenvolveu tecnologia para estação de recarga sem fio, Figura 41. A principal montadora da lista tem criado,

além de tudo, dispositivos de segurança e que aumentam a eficiência dos veículos elétricos e híbridos.

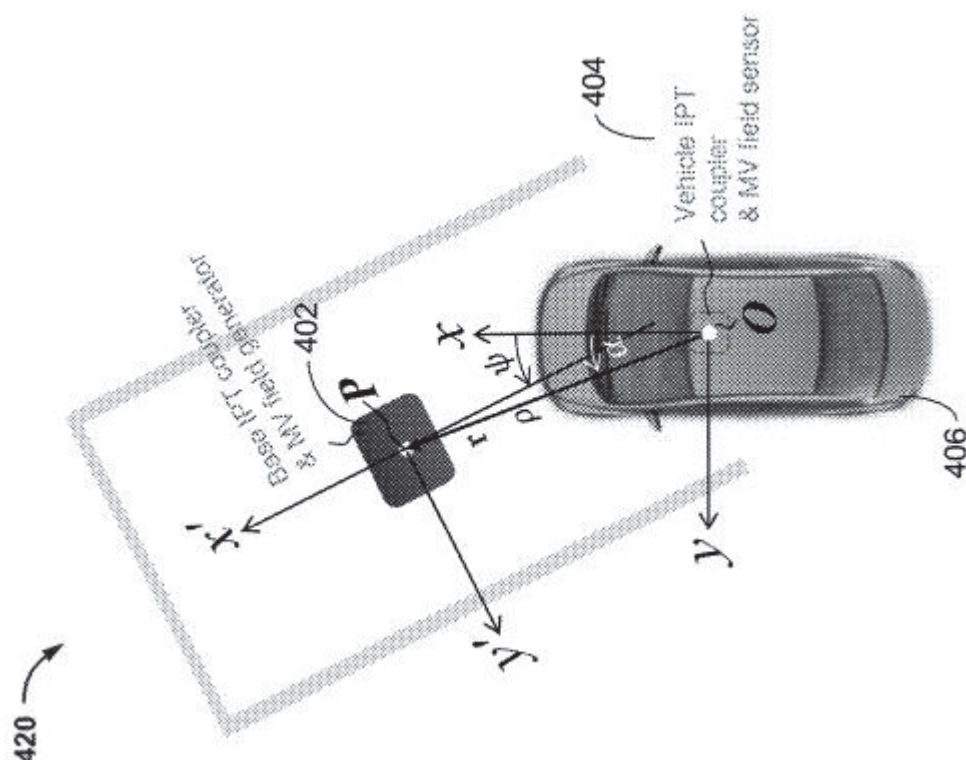
FIGURA 41 - ESTAÇÃO RECARGA TOYOTA



FONTE: Toyota Jidosha Kabushiki Kaisha (2016).

Dentre os termos mais frequentes encontramos a palavra *'apparatus'*, quando relacionado aos termos anteriores da pesquisa surgem tecnologias como as da também Qualcomm, Figura 42. Este sistema é capaz de determinar a posição relativa do veículo, seu objetivo é evitar tentativas mal sucedidas dos veículos elétricos nas estações de carregamento sem fio. Tecnologias como esta visam melhorar a experiência e conveniência dos usuários.

FIGURA 42 - SISTEMA PARA ALINHAMENTO DO VEÍCULO



FONTE: Qualcomm Incorporated (2016).

Ademais, outro termo fortemente relacionado a estas patentes é o Lítio. Apesar de mostrar queda na frequência relativa, ainda se encontra entre as palavras mais frequentes no código H01M-010/44 (Métodos para carregar ou descarregar). Este elemento está presente nas baterias de veículos tanto elétricos quanto híbridos.

Narins (2017) ressalta que a grande parte (70%) da mineração do lítio no mundo ocorre em quatro países, Argentina, Bolívia, Chile e China. O autor afirma que atualmente, a extração/produção economicamente mais viável do mineral ocorre no triângulo entre Bolívia, Chile e Argentina (Figura 43) e na região Sudoeste da China.

FIGURA 43 - TRIANGULO DO LÍTIO



FONTE: Narins (2017).

Segundo Sun et al. (2017) a produção e consumo do lítio ocorrem principalmente em alguns países chave. Os autores também destacam o Chile como um dos principais mineradores a nível mundial, sendo responsável por 36% da mineração do recurso. Para eles o mercado global de veículos elétricos sofreu aumento significativo a partir de 2014, com isto o mercado do lítio teve um forte impacto desde então, sendo a proporção do comércio internacional apresentada na Tabela 7.

TABELA 6 - COMÉRCIO INTERNACIONAL DE LÍTIO.

País	Código	Proporção comércio internacional
União Europeia	EU	15,51%
China	CHN	14,43%
Estados Unidos	USA	13,53%
Japão	JPN	5,04%
Coréia do Sul	KOR	3,69%
Canadá	CAN	3,14%
México	MEX	2,67%
Rússia	RUS	2,63%
Índia	IND	2,61%
Arábia Saudita	SAL	1,71%
Austrália	AUS	1,57%
Brasil	BRA	1,52%
Turquia	TUR	1,34%
Indonésia	IDN	1,19%
África do Sul	ZAF	0,54%
Chile	CHL	0,50%
Argentina	ARG	0,45%
TOTAL		72,07%

FONTE: Adaptado de SUN et al (2017).

Conforme a pesquisa de Sun et al. (2017), dentre os quinze principais países no comércio internacional de lítio, os cinco membros dos BRICS estão presentes. O grupo representa cerca de 21,83% do mercado consumidor. Vale ainda salientar que o Brasil está próximo (geograficamente) de duas das principais origens do mineral (Chile e Argentina).

6 CONCLUSÃO

Técnicas de mineração de dados vem sendo amplamente utilizadas em diversas áreas do conhecimento. Uma de suas vertentes é a mineração de dados textuais, sendo seu principal objetivo a busca por padrões em um conjunto de dados não estruturados. Com a crescente globalização, cada vez mais temos acesso à conteúdos digitais e seu alto volume de dados por vezes torna inviável a leitura individual de cada documento.

Isto posto, a presente pesquisa teve como objetivo desenvolver um algoritmo, em linguagem de programação *open source*, capaz de minerar dados e texto em documentos de patentes. Como plano de fundo, o algoritmo foi aplicado em documentos extraídos da ferramenta *Derwent Innovations Index* tomando como base a categoria de transportes do programa de patentes verdes brasileiro.

O setor automotivo possui grande relevância no mercado mundial, portanto as evoluções tecnológicas dele acabam afetando uma imensa cadeia de fornecedores. No momento em que países sinalizam que devem alterar a composição de suas frotas, a forma com que os *players* interagem entre si tanto no desenvolvimento das tecnologias quanto no fornecimento de serviços e matérias-primas sofrerá profundas mudanças. Esta migração dos veículos movidos à combustão para modelos elétricos ou híbridos faz com que mude não apenas a tecnologia embarcada nos veículos como também a fonte de energia utilizada.

O desenvolvimento do algoritmo em linguagem de programação R possibilitou uma análise mais profunda nos documentos. Atualmente existem diversas ferramentas para análise de dados estruturados, entretanto a complexidade do estudo em documentos cujos dados não estão organizados de forma estruturada, tais como: arquivos de áudio, vídeo, imagem e texto, é muito maior. Uma das principais vantagens da linguagem de programação R é a existência dos chamados *contributed packages*.

Estes pacotes disponibilizados pela comunidade facilitam o processo de programação e acabam por deixar o algoritmo desenvolvido de maneira mais didática, contribuindo com a disseminação do conhecimento. Hoje em dia há inúmeros pacotes voltados à manipulação de *strings* no R. Concomitante a isto, por se tratar de uma ferramenta estatística/gráfica foi possível adicionar blocos de

funções estatísticas (teste de normalidade, análise de variância e ANOVA) tornando a ferramenta mais completa.

Por se tratar de uma linguagem *open source*, a flexibilidade para inserção de novas palavras nas *stop words*, exclusão de termos, manipulação de caracteres e geração e gráficos personalizados dão ao R uma importante vantagem frente à outras linguagens/*softwares*. Sua aplicação independe do idioma dos arquivos analisados e vão além dos documentos de patentes, podendo o algoritmo ser utilizado para análise de documentos científicos e técnicos. Desta forma, os resultados alcançados comprovaram a eficiência da aplicação de métodos de mineração de dados e texto em documentos de patentes conforme apresentado nos parágrafos seguintes.

Apesar do forte apelo para redução das emissões de gases poluentes, o volume de depósitos no Brasil apresentou uma variação tímida dentro do período pesquisado. Na mesma direção, Rússia, Índia e África do Sul também não demonstram crescimento nos seus indicadores. Entretanto, no escritório chinês o volume de depósitos vem crescendo ano após ano de maneira sólida.

Os indicadores de atividade mostraram que os esforços do governo chinês tem surtido efeito nas tecnologias depositadas. O número de patentes depositadas, voltadas às tecnologias ambientalmente amigáveis, no escritório do país quase dobrou nos cinco anos analisados. Por ser detentora do maior mercado de veículos no mundo, naturalmente há uma preocupação maior por parte das empresas em se resguardarem judicialmente no país asiático.

Ainda os indicadores de atividade, os códigos IPC traduzem os esforços de P&D no desenvolvimento de tecnologias para carregar baterias. Nos cinco países do bloco, o código que aparece com maior frequência é o mesmo (H02J-007/00). Esta tendência demonstrou que além do desenvolvimento de novas baterias, um importante gargalo são as estações de recarga para as mesmas. Com o auxílio dos códigos manuais de classe da base *Derwent*, foi possível identificar um volume de tecnologias semelhantes para veículos elétricos e híbridos no Brasil. Rússia e Índia, possuem indicadores similares aos brasileiros, porém há uma diferença maior nos depósitos entre elétricos e híbridos. Já a China demonstrou um foco claramente voltado às tecnologias para veículos elétricos, seus indicadores apresentaram uma diferença de quase 200%.

Dentre as empresas que mais se destacam nos BRICS, confirmou-se que a Toyota lidera as iniciativas no setor. A montadora japonesa divide seu foco no desenvolvimento de veículos híbridos e elétricos e está entre os principais depositantes nos BRICS. Outra montadora que se destaca, de maneira mais forte nos mercados russo e chinês é a Ford, a empresa norte americana possui a maior parte de seus indicadores relacionados aos veículos híbridos. Em contra partida à Ford, a japonesa Nissan apresenta indicadores mais voltados aos veículos elétricos. Diferentemente de Brasil, Rússia e China, a Índia possui como principal depositante uma empresa de tecnologia (Qualcomm) e não uma montadora.

Assim, o desenvolvimento do algoritmo permitiu identificar os principais termos contidos nos documentos. A automatização neste processo de análise, em tempos onde há uma grande quantidade de informações disponíveis, se torna uma importante ferramenta nesta e em pesquisas futuras. Através da análise dos termos, foi possível identificar elementos chave dos documentos.

REFERÊNCIAS

ALI, F.; BOKS, C.; BEY, N. Design for Sustainability and Project Management Literature - A Review. **Procedia CIRP**, v. 48, p. 28–33, 2016.

ARDITO, L.; ADDA, D. D.; PETRUZZELLI, A. Mapping innovation dynamics in the Internet of Things domain: Evidence from patent analysis. **Technological Forecasting & Social Change**, v. 136, n. April, p. 317–330, 2018. Disponível em: <<http://dx.doi.org/10.1016/j.techfore.2017.04.022>>.

ASSIS, F. S. DE. **Análise da estrutura relacional das patentes: o caso domínio tecnológico no seguimento dos veículos elétricos e híbridos**, 2018. Universidade Federal de São Carlos.

BARBIERI, N. Investigating the impacts of technological position and European environmental regulation on green automotive patent activity. **Ecological Economics**, v. 117, p. 140–152, 2015. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.ecolecon.2015.06.017>>.

BENOIT, K.; WATANABE, K.; WANG, H.; et al. quanteda: An R package for the quantitative analysis of textual data. , 2018. Journal of Open Source Software. Disponível em: <<https://quanteda.io>>.

BERGEK, A.; BERGGREN, C. The impact of environmental policy instruments on innovation: A review of energy and automotive industry studies. **Ecological Economics**, v. 106, p. 112–123, 2014. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.ecolecon.2014.07.016>>.

BORGSTEDT, P.; NEYER, B.; SCHEWE, G. Paving the Road to Electric Vehicles – A Patent Analysis of the Automotive Supply Industry. **Journal of Cleaner Production**, v. 167, p. 75–87, 2017. Elsevier Ltd. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0959652617318905>>.

BRASIL. Estímulo a compra de veículos elétricos recebe apoio da comissão de meio ambiente. Disponível em: <<https://www12.senado.leg.br/noticias/materias/2016/05/10/estimulo-a-compra-de-veiculos-eletricos-recebe-apoio-da-comissao-de-meio-ambiente>> Acesso em: 06/01/2018.

BRASIL. Lei nº 9.279, de 14 de maio de 1996. Regula direitos e obrigações relativos à propriedade industrial. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l9279.htm> Acesso em: 07/01/2018.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l9610.htm> Acesso em: 07/01/2018.

BRASIL. Lei nº 9.609, de 19 de fevereiro de 1998. Dispõe sobre a proteção da propriedade intelectual de programa de computador, sua comercialização no País, e

dá outras providências. Disponível em: <
http://www.planalto.gov.br/ccivil_03/leis/l9609.htm> Acesso em: 07/01/2018.

BRASIL. Resolução 175/2016, de 05 de novembro de 2016. Disciplina o exame prioritário de “Patente Verde”. Disponível em: <
http://www.inpi.gov.br/menu-servicos/arquivos-dirpa/Resoluon1752016_Patentesverdes_21112016julio_docx.pdf> Acesso em: 07/01/2018.

CARLI, E. de. **Caracterização e análise da produção de patentes depositadas por instituições científicas e tecnológicas do Brasil de 2004 a 2013**. 137f. Dissertação (mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Administração, Curitiba, 2015.

CECERE, G.; CORROCHER, N.; GOSSART, C.; OZMAN, M. Technological pervasiveness and variety of innovators in Green ICT: A patent-based analysis. **Research Policy**, v. 43, n. 10, p. 1827–1839, 2014. Elsevier B.V. Disponível em: <
<http://dx.doi.org/10.1016/j.respol.2014.06.004>>.

CHANG, P.; WU, C.; LEU, H.; Using patent analysis to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. **Scientometrics**, v. 82, n. 1, p. 5-19, 2010.

CHENG, Z. The spatial correlation and interaction between manufacturing agglomeration and environmental pollution. **Ecological Indicators**, v. 61, p. 1024–1032, 2016. Elsevier Ltd. Disponível em: <
<http://dx.doi.org/10.1016/j.ecolind.2015.10.060>>.

CHRISTENSEN, T. B. Modularised eco-innovation in the auto industry. **Journal of Cleaner Production**, v. 19, n. 2–3, p. 212–220, 2011. Elsevier Ltd. Disponível em: <
<http://dx.doi.org/10.1016/j.jclepro.2010.09.015>>.

CLARIVATE ANALYTICS. Códigos manuais Derwent. Disponível em: <
https://images.webofknowledge.com/WOKRS523R4/help/pt_BR/DII/hs_derwent_m_anual_code.html>. Acesso em: 7/7/2018.

CLERIDES, S.; ZACHARIADIS, T. The effect of standards and fuel prices on automobile fuel economy: An international analysis. **Energy Economics**, v. 30, n. 5, p. 2657–2672, 2008.

CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. , 2006. InterJournal. Disponível em: <
<http://igraph.org>>.

DECHEZLEPRÊTRE, A. Fast-Tracking “Green” Patent Applications: An Empirical Analysis. **Centre for Economic Performance**, n. 1197, p. 27, 2013.

DERWENT. **Derwent Innovations Index**. Web of Science. 2018.

FAGUNDES, M. C.; FERREIRA, M. A.; MOTTA, G. S.; LEAL JÚNIOR, I. C. Patentometric analysis of the process Technologies of Companhia Siderúrgica

Nacional (CSN). **Revista GEINTEC Gestão, Inovação e Tecnologias**. v. 6, p. 3226-3240. 2014.

FEINERER, I.; HORNIK, K. wordnet: wordNet Interface. , 2017. Disponível em: <<https://cran.r-project.org/package=wordnet>>.

FEINERER, I.; HORNIK, K. tm: Text Mining Package. R package version 0.7-5. , 2018. Disponível em: <<https://cran.r-project.org/package=tm>>.

FELLBAUM, C. WordNet: An Electronic Lexical Database. , 1998. Bradford Books.

FELLOWS, I. wordcloud: Word Clouds. , 2018. Disponível em: <<https://cran.r-project.org/package=wordcloud>>.

FERRERO, E.; ALESSANDRINI, S.; BALANZINO, A. Impact of the electric vehicles on the air pollution from a highway. **Applied Energy**, v. 169, n. x, p. 450–459, 2016. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.apenergy.2016.01.098>>.

FIALHO, B. G. **Impactos dos investimentos em P & D nos depósitos de patentes nos países do BRICS**. 68f. Dissertação (mestrado) – Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Produção, Santa Maria, 2015.

FLEUREN, W. W. M.; ALKEMA, W. Application of text mining in the biomedical domain. **Methods**, v. 74, p. 97–106, 2015.

GACKSTATTER, S.; KOTZEMIR, M.; MEISSNER, D. Building an innovation-driven economy – The case of BRIC and GCC countries. **Foresight**. v. 16, 2014.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. Ed. São Paulo: Atlas, 2010.

GONÇALVES, P.; SAMPAIO, V.; ORESTES, M.; et al. Photovoltaic technologies : Mapping from patent analysis. **Renewable and Sustainable Energy Reviews**, v. 93, n. April, p. 215–224, 2018. Disponível em: <<https://doi.org/10.1016/j.rser.2018.05.033>>.

GROSS, J.; LIGGES, U. nortest: Tests for Normality. , 2015. Disponível em: <<https://cran.r-project.org/package=nortest>>.

GRANT, E.; HOF, M. V. D.; GOLD, E. R. Patent landscape analysis: A methodology in need of harmonized standards of disclosure. **World Patent Information**. 2014.

GUZMÁN SÁNCHEZ, M. V. **Patentometria: herramienta para el análisis de oportunidades tecnológicas**. 130f. Tese (Doutorado em Gerência de Informação Tecnológica) – Facultad de Economía, Universidade de La Habana, Cuba, 1999.

INPI. Instituto Nacional da Propriedade Industrial. Disponível em: <<http://www.inpi.gov.br/noticias/Patentes%20Verdes>> Acesso em: 20/02/2018.

INPI. Instituto Nacional da Propriedade Industrial. Disponível em: <

<http://www.inpi.gov.br/Portal/sobre/estrutura>> Acesso em: 20/02/2018a.

INPI. Instituto Nacional da Propriedade Industrial. Disponível em: <<http://www.inpi.gov.br/pedidos-em-etapas/Portal/aceso-a-informacao/aceso-a-informacao>> Acesso em: 19/11/2017a.

INPI. Instituto Nacional da Propriedade Industrial. Disponível em: <<http://www.inpi.gov.br/menu-servicos/patente/Como-protoger-patente-no-externo/Protoger-patente-no-externo>> Acesso em: 19/11/2017b.

INPI. Instituto Nacional da Propriedade Industrial. Disponível em: <<http://www.inpi.gov.br/menu-servicos/patente/classificacao-de-patentes>> Acesso em: 19/11/2017c.

INPI. Relatório Dados da produtividade até setembro de 2016 e cenário operacional do INPI. Disponível em: <http://www.inpi.gov.br/noticias/em-reuniao-da-mei-presidente-do-inpi-apresenta-dados-de-produtividade-e-cenario-operacional/INPI_Relatorio_p_CNOIMEI_30set2016.pdf> Acesso em: 20/12/2017d.

IPCC. Towards New scenarios for Analysis of Emissions, Climate Change, Impacts, and Response strategies. **IPCC Expert meeting report**. 2007.

ITAMARATY. Política externa: mecanismos inter-regionais. Disponível em: <<http://www.itamaraty.gov.br/pt-BR/politica-externa/mecanismos-inter-regionais/3672-brics>> Acesso em: 10/12/2017.

JUNGMANN, D. M. A caminho da inovação: proteção e negócios com bens de propriedade intelectual: guia para o empresário, Brasília: IEL, 2010. Disponível em: <http://www.inpi.gov.br/sobre/arquivos/guia_empresa_iel-senai-e-inpi.pdf> Acesso em: 07/01/2018.

GARCEZ JÚNIOR, S. S. G.; MOREIRA, J. DE J. DA S. O backlog de patentes no Brasil: o direito à razoável duração do procedimento administrativo. **Revista Direito GV**, p. 171–203, 2017.

KRIEGLER, E.; WEYANT, J.; BLANFORD, G.; KREY, V.; CLARKE, L.; EDMONDS, J.; FAWCETT, A.; LUDERER, G.; RIAHI, K.; RICHELIS, R.; ROSE, S.; TAVONI, M.; VAN VUUREN, D. The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. **Climatic Change**. v. 123, p. 353-367, 2014.

KAYA, K.; HAMES, Y. Two new control strategies: For hydrogen fuel saving and extend the life cycle in the hydrogen fuel cell vehicles. **International Journal of Hydrogen Energy**, p. 14, 2019. Elsevier Ltd. Disponível em: <<https://doi.org/10.1016/j.ijhydene.2018.12.111>>.

MARTINS, B.; CHIU, B.; SOUZA, C. DE. Evolução do programa de patentes verdes no Brasil: definição das tecnologias verdes depositadas no INPI. **Congresso Nacional de Excelência em Gestão**, 2016.

MELLO, C. H. P.; TURRIONI, J. B.; XAVIER, A. F.; CAMPOS, D. F. Pesquisa-ação na engenharia de produção: proposta de estruturação para sua condução. **Production**, v. 22, n. 1, p. 1–13, 2012. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132012000100001&lng=pt&tlng=pt>.

MIGUEL, P. A. C. (org.). **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Rio de Janeiro: Elsevier, 2010.

MORAIS, S. M. P. de. **Prospecção tecnológica em documentos de patentes verdes**. 2014. 108f. Dissertação (mestrado) – Universidade Federal da Paraíba, Centro de Ciências Sociais Aplicadas, Departamento de Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação, João Pessoa, 2014.

MULLEN, L. A.; BENOIT, K.; KEYES, O.; SELIVANOV, D.; ARNOLD, J. Fast, Consistent Tokenization of Natural Language Text. , 2018. Disponível em: <<https://doi.org/10.21105/joss.00655>>.

NARINS, T. P. The battery business: Lithium availability and the growth of the global electric car industry. **Extractive Industries and Society**, v. 4, n. 2, p. 321–328, 2017

NASCIMENTO, F.; AZEVEDO, A. F. Z. DE. O perfil exportador brasileiro para o BRICS no período de 2000 a 2010. 90f. Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Economia, São Leopoldo, 2013.

NISSAN MOTOR CO LTD. Maikawa, K.; Imazu, T. **Contactless electricity supply**. US 20130169062-A1, 4 Jul. 2013.

O'NEILL, J. Building Better Global Economic BRICs. **Global economics Paper**, [S.1], n. 66, p. 01-23, 30 Nov. 2001.

OECD. **OECD Patent Statistics Manual**. 2009. Disponível em: <<http://www.oecd.org/sti/inno/oecdpatentstatisticsmanual.htm>> Acesso em: 17/10/2018.

OLTRA, V.; SAINT JEAN, M. Variety of technological trajectories in low emission vehicles (LEVs): A patent data analysis. **Journal of Cleaner Production**, v. 17, n. 2, p. 201–213, 2009.

OOMS, J. PDFTools. , 2018. Disponível em: <<https://cran.r-project.org/package=pdfutils>>.

OMPI. Organização Mundial da Propriedade Intelectual. Disponível em: <<http://www.wipo.int/portal/en/>> Acesso em: 19/11/2017.

ONU. Conferência das Nações Unidas sobre o Desenvolvimento Sustentável. Disponível em: <http://www.rio20.gov.br/sobre_a_rio_mais_20.html> Acesso em: 27/11/2017a.

ONU. Conferência das Nações Unidas sobre o Desenvolvimento Sustentável. Disponível em: < <http://www.onu.org.br/rio20/img/2012/01/agenda21.pdf>> Acesso em: 27/11/2017.

ONU. Conferência das Nações Unidas sobre o Desenvolvimento Sustentável. Brundtland Report. 1987.

PEDERSEN, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. , 2018. Disponível em: <<https://cran.r-project.org/package=ggraph>>.

PEREIRA, C. A. **O fluxo e as dimensões socioespacial e socioinstitucional do conhecimento em Ciência, Tecnologia & Inovação: um estudo patentométrico da produção tecnológica da Unicamp**. Dissertação (Mestrado em Ciência da Informação) – Pontifícia Universidade Católica de Campinas, Campinas, 2008.

PILKINGTON, A. The fit and misfit of technological capability: Responses to vehicle emission regulation in the US. **Technology Analysis and Strategic Management**, 10, pp. 211-224. 1998.

PRADO, M. A. R. DO; OLIVEIRA, E. F. T. DE. MARCOS APARECIDO RODRIGUES DO PRADO OS BRICS NA PERSPECTIVA GEOPOLÍTICA: uma análise cientométrica do período de 2001 a 2010, p. 5–181, 2014.

QUALCOMM INCORPORATED. Kawashima, K. **Device Alignment in inductive power transfer systems**. US 20150094887-A1, 2 Apr. 2015.

QUALCOMM INCORPORATED. WIDMER, H. P.; DAETWYLER, A.; SIEBER, L. **Systems, methods and apparatuses for guidance and alignment in electric vehicles wireless inductive charging systems**. US 20160380488-A1, 29 Dec. 2016.

RAISER, K.; NAIMS, H.; BRUHN, T. Corporatization of the climate? Innovation, intellectual property rights, and patents for climate change mitigation. **Energy Research and Social Science**, v. 27, p. 1–8, 2017. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.erss.2017.01.020>>.

REIS, P. C.; OSAWA, C. C.; MARTINEZ, M. E. M.; MOREIRA, J. C. C. B. R.; SANTOS, D. A. Programa das Patentes Verdes no Brasil : Aliança Verde entre o Desenvolvimento Tecnológico, Crescimento Econômico e a Degradação Ambiental, p. 1–17, 2016.

RENNINGS, K. Redefining innovation - Eco-innovation research and the contribution from ecological economics. **Ecological Economics**, v. 32, n. 2, p. 319–332, 2000.

RUSSO, S. L.; SILVA, G. F.; NUNES, M. A. **Capacitação em Inovação Tecnológica para Empresários**. São Cristóvão: Editora UFS, 2012.

SANTOS, R. X. dos; MATHIAS, S. K.; CARVALHO, C. E. F. de. BRICS e as contingências do discurso. 88f. Dissertação (mestrado) – Universidade Estadual Júlio de Mesquita Filho (UNESP), Programa de Pós-Graduação em Ciência da

Informação, São Paulo, 2017.

SAVOI, F.; UNIVERSIDADE, D. A.; SÃO, F. DE. Domínio tecnológico no setor automotivo: Análise da estrutura relacional das patentes Technological domain in the automotive sector : Analysis of the relational structure of patents. 2017.

SILGE, J.; ROBINSON, D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. , 2016. Disponível em: <<http://dx.doi.org/10.21105/joss.00037>>.

SEVERI, F. Introdução à propriedade intelectual. In: **Gestão da Inovação e Empreendedorismo**, 2013.

SILVA, M. S. DA. ANÁLISE DOS DOCUMENTOS DE PATENTES CORRELACIONADOS A TECNOLOGIAS VERDES DEPOSITADOS POR UNIVERSIDADES PÚBLICAS BRASILEIRAS, 2016.

SUN, X.; HAO, H.; ZHAO, F.; LIU, Z. Tracing global lithium flow: A trade-linked material flow analysis. **Resources, Conservation and Recycling**, v. 124, n. January, p. 50–61, 2017.

TOYOTA JIDOSHA KABUSHIKI KAISHA. MISAWA, K. **Electric power transmission device**. US 20160156198-A1, 2 Jun. 2016.

WALLACE, M. Jawbone Java WordNet API. , 2007. Disponível em: <<http://mfwallace.googlepages.com/jawbone>>.

WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. , 2016. Disponível em: <<http://ggplot2.org>>.

WICKHAM, H. Tidyverse. , 2017. Disponível em: <<https://cran.r-project.org/package=tidyverse>>.

WICKHAM, H. stringr: Simple, Consistent Wrappers for Common String Operations. , 2018. Disponível em: <<https://cran.r-project.org/package=stringr>>.

WICKHAM, H. scales: Scale Functions for Visualization. , 2018b. Disponível em: <<https://cran.r-project.org/package=scales>>.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. dplyr: A Grammar of Data Manipulation. , 2018. Disponível em: <<https://cran.r-project.org/package=dplyr>>.

WICKHAM, H.; HENRY, L. tidyr: Easily Tidy Data with “spread()” and “gather()” Functions. , 2018. Disponível em: <<https://cran.r-project.org/package=tidyr>>.

WIPO. IPC Green Inventory. Disponível em: <https://www.wipo.int/classifications/ipc/en/green_inventory/>. Acesso em: 7/7/2018.

ZENG, S.; LIU, Y.; LIU, C.; NAN, X. A review of renewable energy investment in the BRICS countries: History, models, problems and solutions. **Renewable and Sustainable Energy Reviews**, v. 74, p. 860–872, 2017.

APÊNDICE I - ALGORITMO

```
#Instalação de pacotes
install.packages('tm')
install.packages('stringr')
install.packages('tidyverse')
install.packages('wordcloud')
install.packages('pdftools')
install.packages('ggplot')
install.packages('ggplot2')
install.packages('dplyr')
install.packages('tidytext')
install.packages('tidyr')
install.packages('gggraph')
install.packages('scales')
install.packages('igraph')
install.packages('nortest')
install.packages('quanteda')
install.packages('tokenizers')
install.packages('wordnet')
install.packages("Rgraphviz")
install.packages('car')

#Carregar pacotes
library("tm")
library("stringr")
library("tidyverse")
library("wordcloud")
library("pdftools")
library("ggplot")
```



```
library("ggplot2")
library("dplyr")
library("tidytext")
library("tidyr")
library("ggraph")
library("scales")
library("igraph")
library("nortest")
library("quanteda")
library("tokenizers")
library("wordnet")
library("Rgraphviz")
library("car")

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("Rgraphviz", version = "3.8")

#Importação e Leitura dos dados
getSources()
getReaders()

inspect(corpus)
inspect(corpus[1:1])
meta(corpus[[1]])
inspect(corpus2[[1]])
class(corpus)

#Importar CORPUS
corpus <-
```

```

VCorpus (DirSource ("C:/Users/Pichau/OneDrive/UFPR/Dissertação/Corpus", encoding="UTF-8"),
readerControl=list (reader=readPlain,language="eng"))

#corpus <-
VCorpus (DirSource ("C:/Users/André/OneDrive/UFPR/Dissertação/Corpus", encoding="UTF-8"),readerControl=list (reader=readPlain,language="eng"))

#LIMPEZA DOS DADOS

corpus <- tm_map (corpus, content_transformer (tolower))

corpus <- tm_map (corpus,removeWords, stopwords ("english"))

corpus <- tm_map (corpus,removeWords, c ("the", "novelty",
"advantage", "can", "drawing", "drawings", "isidiidw",
"usa","description", "woa",
"diidw","for","patent","aua","inventor", "use"))

corpus <- tm_map (corpus, stripWhitespace)

corpus <- tm_map (corpus, removePunctuation)

corpus <- tm_map (corpus, removeNumbers)

corpus <- tm_map (corpus, stemDocument)

#TermDocumentMatrix & DocumentTermMatrix (ORGANIZAR DADOS)

freq <- TermDocumentMatrix (corpus)

matriz <- as.matrix (freq)

matriz <- sort (rowSums (matriz),decreasing=T)

matriz <- data.frame (word=names (matriz),freq=matriz)

matriz

head (matriz, n=50)

tail (matriz)

soma <- sum (matriz$freq)

soma

```

```
#GERAR GRÁFICO EM BARRAS

barplot(matriz$freq[10:1], names.arg=matriz$word[10:1],
horiz=T, axis.lty=10, las=2, cex.names = 0.8) #horizontal

barplot(matriz$freq[1:10], names.arg=matriz$word[1:10],
horiz=F, axis.lty=10, las=2, cex.names = 0.7) #vertical

#GERAR NUVEM DE PALAVRAS

wordcloud(corpus, max.words=100,
random.order=T, colors=rainbow(8), rot.per=0.5, use.r.layout=T)

#ANALISAR FREQUÊNCIAS

dtm <- DocumentTermMatrix(corpus)

inspect(dtm)

dtm

nDocs(dtm)

nTerms(dtm)

termos <- nTerms(dtm)

inspect(dtm[1:5, 4340:4350])

inspect(dtm[1:5, 1:10])

fgh <- dtm

removeSparseTerms(fgh, 0.8)

removeSparseTerms(fgh, 0.6)

removeSparseTerms(fgh, 0.4)

removeSparseTerms(fgh, 0.2)

findFreqTerms(dtm, 200)

findAssocs(dtm, "battery", 0.95)
```

```
inspect(removeSparseTerms(dtm, 0.4))

lista <- c("wireless","battery","hybrid", "hydrogen", "cell",
"charge", "charging")

inspect(DocumentTermMatrix(corpus, list(dictionary= lista)))

TotalPalavras2012 <- sum(dtm[1:1, 1:termos])
TotalPalavras2012
TotalPalavras2013 <- sum(dtm[2:2, 1:termos])
TotalPalavras2013
TotalPalavras2014 <- sum(dtm[3:3, 1:termos])
TotalPalavras2014
TotalPalavras2015 <- sum(dtm[4:4, 1:termos])
TotalPalavras2015
TotalPalavras2016 <- sum(dtm[5:5, 1:termos])
TotalPalavras2016
TotalPalavras <- TotalPalavras2012 + TotalPalavras2013 +
TotalPalavras2014 + TotalPalavras2015 + TotalPalavras2016
TotalPalavras

palavras <- c(TotalPalavras2012, TotalPalavras2013,
TotalPalavras2014, TotalPalavras2015, TotalPalavras2016)

freqrel <- inspect(DocumentTermMatrix(corpus, list(dictionary=
lista))) #consulta de termos individuais

#freqrel <- inspect(dtm)

freqrel <- freqrel/palavras
freqrel

#ANÁLISE ESTATÍSTICA
```

```
A <- c(253, 237, 253, 292, 296)
B <- c(310, 352, 298, 300, 282)
C <- c(554, 470, 480, 552, 570)
D <- c(8735, 11123, 11884, 14283, 15252)
E <- c(30, 39, 37, 24, 6)
```

```
BR <- A
```

```
RU <- B
```

```
IN <- C
```

```
CN <- D
```

```
ZA <- E
```

```
#MÉDIAS
```

```
mean(BR)
```

```
mean(RU)
```

```
mean(IN)
```

```
mean(CN)
```

```
mean(ZA)
```

```
#MEDIDAS DE DISPERSÃO
```

```
#MÁXIMOS
```

```
max(BR)
```

```
max(RU)
```

```
max(IN)
```

```
max(CN)
```

```
max(ZA)
```

```
#MÍNIMOS
```

```
min(BR)
```

```
min(RU)
```

min (IN)

min (CN)

min (ZA)

#AMPLITUDE

range (BR)

max (BR) -min (BR)

range (RU)

max (RU) -min (RU)

range (IN)

max (IN) -min (IN)

range (CN)

max (CN) -min (CN)

range (ZA)

max (ZA) -min (ZA)

#VARIÂNCIA

var (BR)

var (RU)

var (IN)

var (CN)

var (ZA)

#DESVIO PADRÃO

sd (BR)

sd (RU)

sd (IN)

sd (CN)

sd (ZA)

```
#TESTE DE NORMALIDADE

response <- c(BR, RU, IN, CN, ZA)

factor <- c(rep("BR", length(BR)), rep("RU", length(RU)),
rep("IN", length(IN)), rep("CN", length(CN)), rep("ZA",
length(ZA)))

response <- c(BR, RU, IN, ZA)

factor <- c(rep("BR", length(BR)), rep("RU", length(RU)),
rep("IN", length(IN)), rep("ZA", length(ZA)))

data1 <- data.frame(response, factor)

plot(response ~ factor, data1)

tapply(data1$response, data1$factor, var)

qqnorm(data1$response)

qqline(data1$response)

data1

ks.test(response, pnorm, mean(response), sd(response),
alternative="greater")

help(ks.test)

#TESTE DE BARTLETT

bartlett.test(response ~ factor, data1)

#ANOVA

dados <- read.table(file.choose(), header=TRUE)

dados.anova <- aov(dados$Patentes ~ dados$País)

summary(dados.anova)

TukeyHSD(dados.anova)

boxplot(dados$Patentes ~ dados$País)

#EXTRAÇÃO DAS DATAS DE DEPÓSITO
```

```

s <- corpus

ipc <- stringr::str_extract_all(s, "\\S\\d\\d\\S*\\-\\d\\d\\d\\d\\/\\d\\d*")

ipc

ipc <- stringr::str_extract_all(s, "b62k-019/40")

ipc <- ipc[!is.na(ipc)]

ipc

bra <- stringr::str_extract_all(s, "BR\\S+\\s* BR\\S+\\s*\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*")

bra <- bra[!is.na(bra)]

bra

rus <- stringr::str_extract_all(s, "RU\\S+\\s* RU\\S+\\s*\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*")

rus <- rus[!is.na(rus)]

rus

ind <- stringr::str_extract_all(s, "IN\\S+\\s* IN\\S+\\s*\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*")

ind <- ind[!is.na(ind)]

ind

chn <- stringr::str_extract_all(s, "CN\\S+\\s* CN\\S+\\s*\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*")

chn <- chn[!is.na(chn)]

chn

afr <- stringr::str_extract_all(s, "ZA\\S+\\s* ZA\\S+\\s*\\d\\d\\s\\w\\w\\w\\s\\d\\d\\d\\d\\s*")

afr <- afr[!is.na(afr)]

afr

```


APÊNDICE II – INDICADORES DE ATIVIDADE

Hybrid vehicles, e.g. Hybrid Electric Vehicles (HEVs)					
	Brasil	Rússia	Índia	China	África do Sul
2012	8	11	20	68	0
2013	4	14	21	93	1
2014	7	17	11	67	0
2015	6	11	19	78	0
2016	16	17	26	83	0
TOTAL	41	70	97	389	1

Control systems					
	Brasil	Rússia	Índia	China	África do Sul
2012	32	48	102	717	4
2013	46	74	78	839	7
2014	49	52	68	762	2
2015	38	54	49	842	1
2016	57	51	60	629	0
TOTAL	222	279	357	3789	14

Gearings therefor					
	Brasil	Rússia	Índia	China	África do Sul
2012	42	67	72	1051	4
2013	45	76	80	1319	7
2014	50	67	60	1446	5
2015	81	82	74	1575	4
2016	68	77	74	1440	1
TOTAL	286	369	360	6831	21

Brushless motors					
	Brasil	Rússia	Índia	China	África do Sul
2012	3	3	8	105	1
2013	1	2	5	129	0
2014	3	3	10	100	0
2015	4	0	7	101	1
2016	9	3	11	121	0
TOTAL	20	11	41	556	2

Electromagnetic clutches					
	Brasil	Rússia	Índia	China	África do Sul
2012	6	4	6	48	0
2013	4	1	5	44	2
2014	5	1	5	52	1
2015	2	5	5	74	1
2016	4	3	7	215	0
TOTAL	21	14	28	433	4

Regenerative braking systems					
	Brasil	Rússia	Índia	China	África do Sul
2012	10	12	19	230	0
2013	10	15	31	277	2
2014	21	6	24	266	4
2015	12	12	22	313	1
2016	14	19	15	275	1
TOTAL	67	64	111	1361	8

Electric propulsion with power supply from force of nature, e.g. sun, wind					
	Brasil	Rússia	Índia	China	África do Sul
2012	3	5	3	273	0
2013	4	14	6	289	0
2014	4	3	9	250	1
2015	2	3	5	369	0
2016	1	0	4	311	0
TOTAL	14	25	27	1492	1

Electric propulsion with power supply external to vehicle					
	Brasil	Rússia	Índia	China	África do Sul
2012	12	20	33	141	0
2013	13	17	22	178	6
2014	12	16	14	160	3
2015	6	13	6	147	2
2016	7	7	6	117	0
TOTAL	50	73	81	743	11

Combustion engines operating on gaseous fuels, e.g. hydrogen					
	Brasil	Rússia	Índia	China	África do Sul
2012	27	28	47	361	7
2013	28	42	47	329	1
2014	20	42	48	364	1
2015	26	45	37	482	3
2016	15	23	42	494	0
TOTAL	116	180	221	2030	12

Power supply from force of nature, e.g. sun, wind					
	Brasil	Rússia	Índia	China	África do Sul
2012	4	1	6	84	2
2013	2	5	4	87	0
2014	0	1	3	56	0
2015	1	1	1	74	0
2016	2	2	3	76	1
TOTAL	9	10	17	377	3

ANEXO I – RESOLUÇÃO Nº 175/2016**MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL**

RESOLUÇÃO Nº 175, DE 05 DE NOVEMBRO DE 2016.

Assunto: Disciplina o exame prioritário de pedidos de “Patente Verde”.

O PRESIDENTE e o DIRETOR DE PATENTES DO INSTITUTO NACIONAL **DA PROPRIEDADE INDUSTRIAL** – INPI, no uso das atribuições legais e regimentais previstas no Decreto nº 8.854, de 22 de setembro de 2016,

CONSIDERANDO o disposto na Lei nº 9.279, de 14 de maio de 1996, e na Instrução Normativa PR nº 30, de 04 de dezembro de 2013, do INPI;

CONSIDERANDO o disposto na Lei nº 12.187, de 29 de dezembro de 2009, que institui a Política Nacional sobre a Mudança no Clima, e o Decreto nº 7.390, de 9 de dezembro de 2010, que regulamenta os artigos 6º, 11 e 12, da Lei nº 12.187, de 29 de dezembro de 2009;

RESOLVEM

Art. 1º Esta Resolução disciplina o exame prioritário de pedidos de “Patente Verde”. Art. 2º “Patente Verde” é o pedido de patente considerado apto ao exame prioritário, conforme listagem apresentada no Anexo I desta Resolução.

Parágrafo único. Aplica-se ao pedido o disposto nos artigos 10 e 18, da Lei nº 9.279, de 14 de maio de 1996.

Art. 3º A Diretoria de Patentes - DIRPA selecionará, analisará e decidirá acerca dos pedidos prioritários de “Patente Verde”.

Parágrafo único. O INPI publicará despacho na Revista Eletrônica da Propriedade Industrial, RPI, declarando ser o pedido “apto” ou “não apto” ao exame prioritário de “Patente Verde”.

Art. 4º O pedido prioritário “Patente Verde” é considerado um pedido nacional.

§1º São pedidos nacionais aqueles que são aceitos no exame formal preliminar ou no exame de admissibilidade para a entrada na fase nacional.

Art. 5º O pedido submetido ao exame prioritário conterà no máximo 15 (quinze) reivindicações, das quais até 3 (três) independentes.

Art. 6º A participação do pedido como prioritário de “Patente Verde” observará o seguinte:

I - ser publicado ou ter o requerimento de publicação antecipado, conforme o disposto no artigo 30 *caput* e parágrafo 1º, da Lei nº 9.279, de 14 de maio de 1996;

II - solicitar o exame do pedido de patente, conforme o disposto no artigo 33, da Lei nº 9.279, de 14 de maio de 1996;

III - a solicitação de exame prioritário será realizada pelo depositante ou representante legal constituído;

IV - o pedido não pode ter sido submetido a exame técnico regular, com a consequente publicação na Revista Eletrônica da Propriedade Industrial;

V - não se referirá a:

a) pedido de patente cujo exame esteja suspenso para o cumprimento de exigência técnica anteriormente formulada pela DIRPA;

b) pedido de patente, publicado na Revista Eletrônica da Propriedade Industrial, em razão de submissão prévia aos outros exames prioritários.

VI - pagamento das retribuições anuais, conforme o disposto no artigo 84, da Lei nº 9.279, de 14 de maio de 1996.

Art. 7º A concessão da “Patente Verde” sobre produto acabado ou sobre material reprodutivo obtido a partir de acesso a patrimônio genético ou a conhecimento tradicional associado fica condicionada ao cadastramento ou autorização, nos termos da Lei 13.123, de 20 de maio de 2015.

Art. 8º Caso haja necessidade de adequar ou alterar o quadro reivindicatório para atender ao disposto no artigo 5º desta Resolução, o depositante ou seu representante legal cumprirá o disposto no art. 32, da Lei nº 9.279, de 14 de maio de 1996.

Art. 9º. Os atos de que trata esta Resolução, quando não praticados pelo depositante, observarão o disposto no parágrafo 1º, do artigo 216, da Lei nº 9.279, de 14 de maio de 1996.

Art. 10. Esta Resolução entra em vigor na data de sua publicação na Revista Eletrônica da Propriedade Industrial.

LUIZ OTÁVIO PIMENTEL

Presidente

JULIO CÉSAR CASTELO BRANCO REIS MOREIRA

Diretor de Patentes

Anexo I (resolução 175/2016)

Listagem das “tecnologias verdes”, baseada no inventário publicado pela Organização Mundial da Propriedade Intelectual

1. Energias alternativas

- Biocombustíveis
 - Combustíveis sólidos
 - Combustíveis líquidos (óleos vegetais, biodiesel, bioetanol)
 - Biogás
 - Biocombustíveis de organismos geneticamente modificados
- Ciclo combinado de gaseificação integrada (IGCC)
- Células-combustível
- Pirólise ou gaseificação de biomassa
- Aproveitamento de energia a partir de resíduos humanos
 - A partir de resíduos agrícolas
 - Gaseificação
 - Resíduos químicos
 - Resíduos industriais
 - Utilizando os gases de saída de alto-fornos
 - Licores de polpa
 - Digestão anaeróbica de resíduos industriais
 - Resíduos industriais de madeira
 - Resíduos hospitalares
 - Gás de aterros
 - Separação dos componentes
 - Resíduos domiciliares e urbanos
- Energia hidráulica
 - Usinas hidrelétricas (PCH e MCH)
 - Energia das ondas ou marés
 - Meios de regulagem, controle ou segurança de máquinas ou motores acionados por líquidos
 - Propulsão pela utilização de energia derivada do movimento da água circundante
- Conversão da energia térmica dos oceanos (OTEC)
- Energia eólica
- Energia Solar
 - Energia solar fotovoltaica (PV)
 - Energia solar térmica
 - Sistemas solares híbridos (térmico-fotovoltaicos)
 - Propulsão de veículos usando energia solar
 - Produção de energia mecânica a partir da energia solar
 - Aspectos de cobertura de telhados com dispositivos de coleta de energia solar
 - Geração de vapor usando energia solar
 - Sistemas de refrigeração ou bombas de calor usando energia solar
 - Secagem de materiais ou objetos utilizando energia solar
 - Dispositivos para a concentração da irradiação solar
 - Coletores de calor solar com o fluido de trabalho conduzido através do coletor

- Energia geotérmica
- Outros tipos de produção ou utilização de calor não derivado de combustão
- Utilização de calor residual
- Dispositivos para a produção de energia mecânica a partir de energia muscular

2. Transportes

- Veículos híbridos
- Veículos elétricos
- Estações de carregamento para veículos elétricos
- Veículos alimentados por energia extraída das forças da natureza (sol, vento, ondas, etc.)
- Veículos alimentados por fonte de potência externa (energia elétrica, etc.)
 - Veículos alimentados por células combustíveis
 - Veículos alimentados por hidrogênio
 - Veículos com propulsão muscular
- Veículos com freios regenerativos
- Veículos cuja carroceria possui baixo arrasto aerodinâmico
- Veículos com embreagem eletromagnética (menor perda na transmissão)

3. Conservação de energia

- Armazenagem de energia elétrica
- Circuitos de alimentação de energia elétrica
- Medição do consumo de eletricidade
- Armazenamento de energia térmica
- Iluminação de baixo consumo energético
- Isolamento térmico de edificações
- Recuperação de energia mecânica (ex: balanço, rolamento, arfagem)

4. Gerenciamento de resíduos

- Eliminação de resíduos
- Tratamento de resíduos
 - Destruição de resíduos por combustão
 - Reutilização de materiais usados
 - Utilização de restos ou refugos de borracha na fabricação de calçados
 - Manufatura de artigos de sucata ou de refugo de partículas metálicas
 - Produção de cimento hidráulico a partir de resíduos
 - Utilização de resíduos como material de enchimento para argamassas ou concreto
 - Utilização de resíduos para a produção de fertilizantes
 - Recuperação ou aproveitamento de resíduos
- Controle de poluição

- Sequestro e armazenamento de carbono
- Gestão da qualidade do ar
 - Tratamento de gases residuais
 - Separação de partículas dispersas em gases ou vapores
 - Aplicação de aditivos em combustíveis ou nas chamas para redução de fumaça e facilitar a remoção de fuligem
 - Disposição dos dispositivos para tratamento de fumaça ou de emanações aparelhos combustores
 - Materiais para captação ou absorção de poeira
 - Alarmes de poluição
- Controle da poluição da água
 - Tratamento de águas residuais ou esgoto
 - Materiais para tratamento de líquidos poluentes
 - Remoção de poluentes de águas a céu aberto
 - Instalações de encanamentos para águas residuais
 - Gerenciamento de esgotos
- Meios para prevenir contaminação radioativa em caso de vazamento no reator

5. Agricultura sustentável

- Técnicas de reflorestamento
- Técnicas alternativas de irrigação
- Pesticidas alternativos
- Melhoria do solo (ex: fertilizantes orgânicos derivados de resíduos)

ANEXO II – SOFTWARES OPEN SOURCE PARA MINERAÇÃO DE TEXTO

Ferramenta	Técnicas	Características/Usos	Website
Visualtext	Sistemas de processamento de linguagem natural	Sistemas de extração de informações e analisadores de texto.	www.textanalysis.com
Aika	Aprendizado de máquina, redes neurais artificiais, mineração de padrão frequente e indução de gramática.	Silabificação	www.aika-software.org
Data Science Toolkit	Algoritmos avançados	Análise de sentimento. Identificação de idioma, classificação de tópicos	www.datasciencetoolkit.org
Datumbbox	Aprendizado de máquina e extração de palavras-chave.	Análise de texto, otimização de mecanismo de busca de extração de palavra-chave, monitoramento de mídia social, análise de sentimento	www.datumbbox.com
GATE	Processamento de linguagem natural	Processamento de texto	gate.ac.uk
Lingpipe	Linguagem computacional	Processamento e classificação de texto.	aliasi.com/lingpipe
Open Calais	Mecanismo de marcação aprimorado	Processamento de texto	www.opencalais.com
Rapidminer Text Mining	Aprendizado de máquina	Mineração de dados e texto, análise e processamento de texto	docs.rapidminer.com
Reverb	NLP, algoritmos WEKA	Extração de informação	reverb.cs.washington.edu
S-EM (Spy-EM)	Naive Bayes e algoritmos EM	Aprendizado de texto ou sistema de classificação	www.cs.uic.edu/~liub/S-EM/S-EM-download
TXM	Funções estatísticas baseadas em pacotes do R	Análise de texto	sourceforge.net/projects/txm/
Programação em R	Técnicas estatísticas e gráficas	transformação de dados e análise de texto	www.r-project.org

Fonte: Adaptado de Kaur e Chopra (2016).

ANEXO III – CÓDIGOS IPC CATEGORIA TRANSPORTE

▢ TRANSPORTATION	
▢ . Vehicles in general	
▢ . . Hybrid vehicles, e.g. Hybrid Electric Vehicles (HEVs)	B60K 6/00, 6/20
. . . Control systems	B60W 20/00
. . . Gearings therefor	F16H 3/00-3/78, 48/00-48/30
. . Brushless motors	H02K 29/08
. . Electromagnetic clutches	H02K 49/10
. . Regenerative braking systems	B60L 7/10-7/22
. . Electric propulsion with power supply from force of nature, e.g. sun, wind	B60L 8/00
▢ . . Electric propulsion with power supply external to vehicle	B60L 9/00
. . . With power supply from fuel cells, e.g. for hydrogen vehicles	B60L 11/18
. . Combustion engines operating on gaseous fuels, e.g. hydrogen	F02B 43/00 F02M 21/02, 27/02
. . Power supply from force of nature, e.g. sun, wind	B60K 16/00
. . Charging stations for electric vehicles	H02J 7/00

Fonte: http://www.wipo.int/classifications/ipc/en/green_inventory/

ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES

(continua)

SEÇÃO	CLASSE	SUBCLASSE
B - Operações de processamento; transporte	60 - Veículos em Geral	K - Disposições ou montagem de unidades de propulsão ou de transmissões em veículos; disposição ou montagem de várias máquinas motrizes diferentes em veículos; acionamentos auxiliares para veículos; instrumental ou painéis de instrumentos para veículos; dispositivos correlatos a resfriamento, tomada de ar, exaustão de gases ou alimentação de combustível de unidades de propulsão em veículos
		L - Propulsão de veículos de propulsão elétrica; suprimento de energia elétrica para equipamentos auxiliares dos veículos de propulsão elétrica, sistemas de freios eletrodinâmicos para veículos, em geral; suspensão magnética ou levitação para veículos; monitoração de variáveis operacionais de veículos de propulsão elétrica; dispositivos de segurança elétrica de veículos de propulsão elétrica.
		W - Controle conjugado para sub unidade de veículos de tipo ou função diferente; sistemas de controle especialmente adaptados para veículos híbridos; sistemas de controle de veículos terrestres não relacionados ao controle de uma sub unidade particular.
61 - Ferrovias		D - Detalhes das caixas ou tipos de veículos ferroviários

ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES

(continuação)

<p>B - Operações de processamento; transporte</p>	<p>62 - Veículos terrestres para trafegar de outra maneira que não sobre trilhos</p> <p>D - VEÍCULOS MOTORIZADOS; REBOQUES (direção ou direcionamento em uma trilha desejada, de máquinas ou implementos de agricultura A01B 69/00; rodas, rodas pivotadas, eixos, dispositivos para aumento de adesão das rodas B60B; pneus de veículo, enchimento, troca ou reparo dos pneus B60C; conexões entre veículos de um trem ou similar B60D; veículos para uso sobre trilhos e rodovia, veículos anfíbios ou conversíveis B60F; arranjos de suspensão B60G; dispositivos de aquecimento, de resfriamento, de ventilação ou outros dispositivos para tratamento do ar B60H; janelas, para-brisas, tetos que não sejam fixos, portas ou dispositivos similares, coberturas de proteção para veículos fora de uso B60J; disposições da instalação de propulsão, acionamentos auxiliares, transmissões, controles, instrumentação ou painéis B60K; equipamento elétrico ou propulsão dos veículos movidos a eletricidade B60L; alimentação de energia para os veículos movidos a eletricidade B60M; acomodações de passageiros, não incluídas em outro local B60N; adaptações para transporte ou carregamento de objetos ou cargas especiais B60P; disposições de iluminação ou sinalização, as montagens ou fixação dos mesmos ou dos circuitos para os mesmos, para veículos em geral B60Q; veículos, acessórios de veículos ou peças de veículos não incluídos em outro local B60R; manutenção, limpeza, reparos, sustentação, suspensão ou manobra, não incluídos em outro local B60S; disposições de freios, sistemas de controle de freios ou partes dos mesmos B60T; veículos a colchão de ar B60V; motocicletas, acessórios para as mesmas B62J, B62K; testes de veículos G01M).</p>
	<p>K - Bicicletas; quadros de bicicletas; dispositivos de direção de bicicletas; controles terminais, operados pelo ciclista, especialmente adaptados a bicicletas; suspensões de eixos de bicicletas; carrinhos laterais, carrinhos dianteiros ou similares, para bicicletas</p>

ANEXO IV – DESCRIÇÃO DAS SEÇÕES, CLASSES E SUBCLASSES

(conclusão)

<p>F - Engenharia mecânica; iluminação; aquecimento; armas; explosão</p>	<p>02 - Motores de combustão; instalações de motores a gás quente ou de produtos de combustão</p>	<p>M - Propulsão pelo condutor de veículos com rodas ou de trenós; propulsão mecânica de trenós ou bicicletas; transmissões especialmente adaptadas para tais veículos (disposições ou montagem de transmissões em veículos em geral b60k; elementos de transmissão per se f16)</p> <p>B - MOTORES DE COMBUSTÃO INTERNA DE PISTÕES; MOTORES DE COMBUSTÃO EM GERAL (válvulas de funcionamento cíclico para os mesmos F01L; lubrificação dos motores de combustão interna F01M; silenciadores de fluxo de gás ou aparelhos de exaustão para os mesmos F01N; resfriamento dos motores de combustão interna F01P; turbinas de combustão interna F02C; instalações em que os motores utilizam produtos de combustão F02C, F02G)</p> <p>M - ALIMENTAÇÃO DE MOTORES DE COMBUSTÃO EM GERAL COM MISTURAS COMBUSTÍVEIS OU SEUS COMPONENTES (carregamento desses motores F02B)</p>
<p>16 - Elementos ou unidades de engenharia; medidas gerais para assegurar e manter o funcionamento efetivo de máquinas ou instalações; isolamento térmico em geral</p>	<p>H - SISTEMAS DE ENGRENAGENS</p>	

FONTE: Adaptado de INPI (2018).

ANEXO V – PRINCIPAIS DEPOSITANTES

Rank	BRASIL	VG RÚSSIA	VG ÍNDIA	VG CHINA	VG ÁFRICA DO SUL
1º	TOYOTA JIDOSHA KK	FORD GLOBAL TECHNOLOGIES LLC	QUALCOMM INC	TOYOTA JIDOSHA KK	GENERAL ELECTRIC CO
2º	TOYOTA MOTOR CORP	FORD GLOBAL TECHNOLOGY CO LTD	TOYOTA JIDOSHA KK	TOYOTA MOTOR CORP	CATERPILLAR INC
3º	SCANIA CV AB	NISSAN MOTOR CO LTD	TOYOTA MOTOR CORP	FORD GLOBAL TECHNOLOGIES LLC	EATON CORP
4º	NISSAN MOTOR CO LTD	TOYOTA JIDOSHA KK	HONDA MOTOR CO LTD	FORD GLOBAL TECHNOLOGY CO LTD	FLUIDIC INC
5º	GENERAL ELECTRIC CO	TOYOTA MOTOR CORP	SUZUKI MOTOR CORP	HYUNDAI MOTOR CO LTD	JAGUAR LAND ROVER LTD
6º	QUALCOMM INC	SCANIA CV AB	SUZUKI KK	STATE GRID CORP CHINA	JAGUAR CARS LTD
7º	BOMBARDIER TRANSPORTATION GMBH	SIEMENS AG	NISSAN MOTOR CO LTD	GM GLOBAL TECHNOLOGY OPERATIONS INC	JOY MM DELAWARE INC
8º	HONDA MOTOR CO LTD	MAN TRUCK BUS AG	BOSCH GMBH ROBERT	BOSCH GMBH ROBERT	PHILIP MORRIS PROD SA
9º	BOMBARDIER TRANSPORT GMBH	BOMBARDIER TRANSPORTATION GMBH	GENERAL ELECTRIC CO	GM GLOBAL TECHNOLOGIES OPERATIONS INC	SANDVIK MINING CONSTR OY
10º	MAN TRUCK BUS AG	MAKITA CORP	GUANGDONG OPPO MOBILE TELECOM CORP LTD	KUNSHAN HOLY LIGHT NEW ENERGY TECHNOLOGY	SANDVIK TAMROCK OY
11º	BOMBARDIER PRIMOVE GMBH	BOMBARDIER TRANSPORT GMBH	MITSUBISHI ELECTRIC CORP	NISSAN MOTOR CO LTD	INFLEUNT CORP
12º	BERGQUIST M	RENAULT SAS	HONDA GIKEN KOGYO KK	HONDA MOTOR CO LTD	KONINK PHILIPS BV
13º	DEERE CO	KONINK PHILIPS NV	GUANGDONG OPPO MOBILE COMMUNICATION CO L	BYD CO LTD	KONINK PHILIPS NV
14º	RENAULT SAS	BERGQUIST M	GM GLOBAL TECHNOLOGY OPERATIONS INC	GEN MOTORS GLOBAL OPERATION TECHNOLOGY	MCMILLAN P J
15º	VOLVO TRUCK CORP	BOMBARDIER PRIMOVE GMBH	SONY CORP	SAMSUNG SDI CO LTD	PHILIP MORRIS INC
16º	EATON CORP	RENAULT	SAMSUNG ELECTRONICS CO LTD	KIA MOTORS CORP	DAIGLE J
17º	KONINK PHILIPS NV	ENERGY STORAGE CONTROL SYSTEMS LLC	INTEL CORP	SONY CORP	DARNELL P
18º	BOSCH GMBH ROBERT	KONINK PHILIPS ELECTRONICS NV	ALLISON TRANSMISSION INC	ZF FRIEDRICHSHAFEN AG	DEERE CO
19º	SAAB SCANIA AB	SAAB SCANIA AB	KONINK PHILIPS NV	TOYOTA MOTOR CO LTD	EDLER A N
20º	B/OERKMAN M	B/OERKMAN M	BOMBARDIER TRANSPORTATION GMBH	AISIN AW CO LTD	FRIESEN C A
21º	ETHICON LLC	EVTODEEV YU L	ETHICON ENDO SURGERY INC	PANASONIC INTELLECTUAL PROPERTY MANAGEME	HAYES J
22º	INTEL CORP	KAMAZ STOCK CO	SIEMENS AG	KIA MOTOR CO LTD	HEMES E
23º	PETTERSSON N	LINDSTROEM J	DENSO CORP	HON HAI PRECISION IND CO LTD	KONINK PHILIPS ELECTRONICS NV
24º	RENAULT	PETTERSSON N	BOMBARDIER TRANSPORT GMBH	BEIQI FOTON MOTOR CO LTD	KOUHIA S
25º	LINDSTROEM J	HONDA MOTOR CO LTD	SCANIA CV AB	LG CHEM LTD	KOUVO M