

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA EM
SISTEMAS PRODUTIVOS

EMERSON DAMO

MAPEAMENTO DO DESENVOLVIMENTO TECNOLÓGICO NA INDÚSTRIA POR
MEIO DE DOCUMENTOS DE PATENTES:
ANÁLISE DA INOVAÇÃO NA INDÚSTRIA AUTOMOBILÍSTICA

São Paulo

Abril/2021

EMERSON DAMO

MAPEAMENTO DO DESENVOLVIMENTO TECNOLÓGICO NA INDÚSTRIA POR
MEIO DE DOCUMENTOS DE PATENTES:
ANÁLISE DA INOVAÇÃO NA INDÚSTRIA AUTOMOBILÍSTICA

Dissertação apresentada como exigência parcial para a obtenção do título de Mestre em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação do Prof. Dr. António César Galhardi.

São Paulo

Abril/2021

Damo, Emerson

D163m Mapeamento do desenvolvimento tecnológico na indústria por meio de documentos de patentes: análise da inovação na indústria automobilística / Emerson Damo. – São Paulo: CPS, 2021.
106 f. : il.

Orientador: Prof. Dr. Antônio Cesar Galhardi

Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos) – Centro Estadual de Educação Tecnológica Paula Souza, 2021.

1. Prospeção tecnológica. 2. Patentometria. 3. Inovação. 4. Web crawler. 5. Indústria automobilística. I. Galhardi, Antônio Cesar. II. Centro Estadual de Educação Tecnológica Paula Souza. III. Título.

EMERSON DAMO

MAPEAMENTO DO DESENVOLVIMENTO TECNOLÓGICO NA INDÚSTRIA POR
MEIO DE DOCUMENTOS DE PATENTES:
ANÁLISE DA INOVAÇÃO NA INDÚSTRIA AUTOMOBILÍSTICA

Prof. Dr. António César Galhardi

Profa. Dra. Marcia Regina Neves Guimarães

Prof. Dr. Fabrício José Piacente

São Paulo, 07 de abril de 2021

Para os meus queridos pais Roberto (in
Memorian) e Alice pelo constante incentivo
aos meus estudos, a minha esposa Carminha e
aos meus filhos pelo apoio, paciência e amor
incondicional a mim devotados em todos os
momentos, dedico.

AGRADECIMENTOS

Aos meus pais Roberto e Alice, pela educação e amor com que me criaram. O insistente incentivo aos meus estudos me auxiliou a chegar até aqui.

À minha família: Carminha, Rodrigo e Pedro José, pela compreensão e amor que devotaram a mim, respeitando meus valores e ideais.

Ao meu orientador Prof. Dr. Antônio César Galhardi, pela disponibilidade e dedicação nas orientações, e pelos relevantes ensinamentos no processo de aprendizagem em pesquisa.

Aos professores membros da banca avaliadora desta dissertação, Profa. Dra. Marcia Regina Neves Guimarães e Prof. Dr. Fabrício José Piacente, pelas ricas contribuições compartilhadas.

Aos professores do programa de Mestrado Profissional do Centro Paula Souza, por todo conhecimento compartilhado, e aos colegas de turma pela amizade e companheirismo.

Aos colegas de turma Leonardo e Rodrigo pela ajuda nos momentos de dificuldades.

Ao estimado amigo Edney, pelo incentivo a continuar os meus estudos e a ingressar no programa de Mestrado Profissional do Centro Paula Souza.

Aos profissionais da Volkswagen do Brasil, pelo apoio e contribuições, em especial aos executivos do Departamento de Design da América do Sul, José Carlos Pavone e Fernanda Mendes.

A todos aqueles que de alguma forma contribuíram com esta pesquisa, o meu muito obrigado!

*Es ist nicht genug zu wissen, man muß auch
anwenden. Es ist nicht genug zu wollen, man
muß auch tun.*

*Knowing is not enough, we must apply.
Willing is not enough; we must do.*

Não basta saber, é preciso aplicar. Não basta
querer, é preciso também fazer.

Johann Wolfgang von Goethe

RESUMO

DAMO, E. **Mapeamento do Desenvolvimento Tecnológico na Indústria por meio de documentos de patentes**: análise da inovação na indústria automobilística. 106 f. Dissertação (Mestrado Profissional em Gestão e Desenvolvimento da Educação Profissional). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2020.

O presente trabalho tem por objetivo criar um software de prospecção tecnológica que permita, por meio do uso de palavras chaves, promover a coleta das informações contidas nos documentos de patentes disponíveis na plataforma *Derwent Innovation Index (DII)* na base de dados *Web of Science (WoS)*. A pesquisa se justifica pela contribuição que traz ao estudo do comportamento e das características da inovação tecnológica, com base em variáveis obtidas em documentos de patentes, um *gap* de pesquisa em uma área temática pouco explorada pela comunidade científica. A metodologia utilizada foi uma pesquisa empírica, de natureza aplicada com objetivo exploratório e abordagem qualitativa. Desenvolveu-se um programa em *Python* e, por meio da técnica de *web crawler*, obteve-se dados de documentos de patentes. As tecnologias pesquisadas basearam-se em tecnologias automotivas, consideradas relevantes em função da capacidade de inovação da indústria automotiva: *Vehicular Safety*; *Infotainment Systems*; *Smart Cities*; *Smart Mobility*; *Advanced Driver Assistance Systems*; e *Autonomous Vehicles*. Os resultados deste estudo permitiram a criação de um software de prospecção tecnológica que, por meio de palavras-chave, promove a coleta das informações contidas nos documentos de patentes disponíveis na plataforma *DII* na base de dados *WoS*. O software gera diversos bancos de dados relacionais, o que permite o uso estruturado destas informações para estudo de um determinado campo tecnológico, por meio da criação de *dashboards* e gráficos com os softwares *Excel* e *PowerBI*. Por fim utilizou-se os conceitos de busca de anterioridade, por meio de citações de patentes, e de redes de conhecimento, para elaborar grafos, o que possibilitou o estudo do fluxo de conhecimento para a tecnologia *Smart Mobility* e a verificação da ocorrência, nesta tecnologia, de *spillover* (quando uma tecnologia desenvolvida com um propósito, “transborda” para outra área e ou outra tecnologia).

Palavras-chave: Prospecção tecnológica; Patentometria; Inovação; *Web crawler*; Indústria automotiva.

ABSTRACT

DAMO, E. **Mapeamento do Desenvolvimento Tecnológico na Indústria por meio de documentos de patentes:** análise da inovação na indústria automobilística. 106 f. Dissertação (Mestrado Profissional em Gestão e Desenvolvimento da Educação Profissional). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2020.

This work aims to create a technological prospecting software that allows, using keywords, to promote the collection of information contained in patent documents available on the Derwent platform in the Web of Science database. The research is justified by the contribution it brings to the study of the behavior and characteristics of technological innovation, based on variables obtained in patent documents, a research gap in a thematic area little explored by the scientific community. The methodology used was an empirical research, of an applied nature with an exploratory objective and a qualitative approach. A Python program was developed and, using the web crawler technique, data from patent documents. The researched technologies were based on automotive technologies, considered relevant due to the innovative capacity of the automotive industry: Vehicular Safety; Infotainment Systems; Smart Cities; Smart Mobility; Advanced Driver Assistance Systems; and Autonomous Vehicles. The results of this study allowed the creation of a technological prospecting software that, through keywords, promotes the collection of information contained in the patent documents available on the DII platform in the WoS database. The software generates several relational databases, which allows the structured use of this information to study a certain technological field, through the creation of dashboards and graphics with Excel and PowerBI software. Finally, the concepts of search for anteriority were used, through patent citations, and knowledge network, to create graphs, which enabled the study of the flow of knowledge for the Smart Mobility technology and the verification of the occurrence of spillover in this technology.

Keywords: Technological Prospecting; Patinformatics; Innovation; Web Crawler; Automotive Industry;

LISTA DE QUADROS

Quadro 1 - Tecnologia, String de busca, data de acesso e quantidade de patentes encontradas na base WoS	65
Quadro 2 - Campos criados e descritivo do bd1_dados_basicos.....	65
Quadro 3 - Campos criados e descritivo do bd2_dados_ipc	66
Quadro 4 - Campos criados e descritivo do bd3_depositantes.....	66
Quadro 5 - Campos criados e descritivo do bd4_paises.....	67
Quadro 6 - Campos criados e descritivo do bd5_inventores	67
Quadro 7 - Campos criados e descritivo do bd6_tabela_IPC	68
Quadro 8 - Campos criados e descritivo do bd7_codigo_paises	68
Quadro 9 - Campos criados e descritivo do bd8_controle_acesso_base_dados.....	68
Quadro 10 - Campos de entrada (<i>dataframe</i> df1) e de saída (<i>dataframe</i> df2) para o primeiro nível do fluxo de conhecimento.....	73
Quadro 11 - Campos de entrada (<i>dataframe</i> df1) e de saída (<i>dataframe</i> df2) para o primeiro nível do fluxo de conhecimento.....	74

LISTA DE TABELAS

Tabela 1 – Top 10 países maiores solicitantes de patentes no mundo em 2019	20
Tabela 2 - Investimentos mundiais em P&D por setor industrial em 2019.....	40
Tabela 3 - Dez maiores empresas em investimentos em P&D no setor da indústria automobilística no mundo em 2019.....	41

LISTA DE FIGURAS

Figura 1 - Curva de pedidos de patentes no mundo por ano	20
Figura 2 - Concentração de pedidos de patentes por continente em 2019	21
Figura 3 - Visão geral da estrutura do trabalho	24
Figura 4 - Curva da Difusão da Inovação.....	29
Figura 5 - Investimentos em P&D no setor automotivo por ano e por região.....	41
Figura 6 - Divisão das áreas de <i>Smart Mobility</i> dentro do conceito de <i>Smart City</i>	46
Figura 7 - Estrutura geral de um <i>web crawler</i>	47
Figura 8 - Fluxo de dados e componentes do Scrapy	49
Figura 9 - Estrutura de diretórios e arquivos gerados pelo <i>Scrapy</i> após o comando <i>startproject</i>	50
Figura 10 - Estrutura de diretórios e arquivos gerados pelo <i>Scrapy</i> após o comando <i>genspider</i>	51
Figura 11 - Programa <i>wos.py</i> gerado automaticamente pelo <i>Scrapy</i> após o comando <i>genspider</i>	52
Figura 12 - Linhas de código do programa <i>wos.py</i> com a importação das bibliotecas	53
Figura 13 - Linhas de código do programa <i>wos.py</i> (class <i>Tela1_WOS</i> e class <i>Tela2_WOS</i>)	53
Figura 14 - Linhas de código do programa <i>wos.py</i> , com as etapas de criação da tela informativa do sistema	54
Figura 15 - Linhas de código do programa <i>wos.py</i> - etapas de criação dos arquivos tipo CSV	54
Figura 16 - Linhas de código do programa <i>wos.py</i> - etapas de interação com o <i>Chrome</i>	55
Figura 17 - Linhas de código do programa <i>wos.py</i> - etapas de obtenção das <i>URLs</i> iniciais	56
Figura 18 - Linhas de código do programa <i>wos.py</i> relativas ao <i>Scrapy Engine</i>	57
Figura 19 - Tela 1 do <i>Spider wos</i>	58
Figura 20 - Tela 2 do <i>Spider wos</i>	59
Figura 21 - Página web “login.webofknowledge.com/” com as ações autônomas de 1 a 5.....	60
Figura 22 - Página web “login.webofknowledge.com/” com as ações autônomas de 6 a 8	60
Figura 23 - Página web “login.webofknowledge.com/” com a ação autônoma 9.....	61
Figura 24 - Tela 3 do <i>Spider wos</i>	61
Figura 25 - Tela do <i>Windows PowerShell</i> no início do <i>spider engine</i>	62
Figura 26 - Tela do <i>Windows PowerShell</i> no final do <i>spider engine</i>	63

Figura 27 - Estrutura de diretórios e arquivos gerados após o processo <i>spider engine</i>	63
Figura 28 - Funcionamento do programa <i>web crawler</i> <i>wos.py</i>	64
Figura 29 - Estrutura do banco de dados relacional no software <i>Power BI</i>	69
Figura 30 – Tela do Software Anaconda Navigator.	70
Figura 31 - Tela para a criação de um <i>notebook</i> no <i>software Jupyter</i>	71
Figura 32 - Notebook criado no <i>software Jupyter</i>	71
Figura 33 - Código do programa <i>fluxo_spider.ipynb</i> relativas a abertura das bibliotecas <i>Python</i> e as configurações ao <i>software Jupyter</i> para funcionar como <i>web crawler</i>	72
Figura 34 - Código do programa <i>fluxo_spider.ipynb</i> relativo a criação do <i>dataframe</i> de entrada <i>df1</i> e do <i>dataframe</i> de saída <i>df2</i>	73
Figura 35 – Linha de código adicionada ao código original do programa <i>web crawler</i> <i>wos.py</i>	75
Figura 36 – Linhas de código do programa <i>fluxo_spider.ipynb</i> para interação com o navegador <i>Google Chrome</i>	75
Figura 37 - Linhas de 1 a 55 do código do programa <i>fluxo_spider.ipynb</i> para a execução do <i>Scrapy Engine</i>	76
Figura 38 - Funcionamento do programa <i>web crawler</i> <i>fluxo_spider.ipynb</i>	77
Figura 39 – Linhas de código do programa <i>gerar_grafo.ipynb</i> para a construção do mapa de fluxo de conhecimento.....	79
Figura 40 - Funcionamento do programa <i>gerar_grafo.ipynb</i>	80
Figura 41 - <i>Dashboard</i> Banco de Dados <i>bd1_dados_basicos</i>	81
Figura 42 - <i>Dashboard</i> Banco de Dados <i>bd2_dados_ipc</i>	81
Figura 43 - <i>Dashboard</i> Banco de Dados <i>bd3_depositantes</i>	82
Figura 44 - <i>Dashboard</i> Banco de Dados <i>bd4_paises</i>	82
Figura 45 - <i>Dashboard</i> Banco de Dados <i>bd5_inventores</i>	83
Figura 46 - <i>Dashboard</i> Banco de Dados <i>bd6_tabela_IPC</i>	83
Figura 47 - <i>Dashboard</i> Banco de Dados <i>bd7_codigo_paises</i>	84
Figura 48 - <i>Dashboard</i> Banco de Dados <i>bd8_controle_acesso</i>	84
Figura 49 - <i>Dashboard</i> Banco de Dados <i>bd9_calendario</i>	85
Figura 60 - <i>Dashboard</i> número de patentes pelo ciclo de vida por tecnologia	85
Figura 51 - <i>Dashboard</i> número de citações das patentes pelo ciclo de vida por tecnologia....	86
Figura 52 - <i>Dashboard</i> número de países com patentes registradas por tecnologia	86
Figura 53 - <i>Dashboard</i> número de campos tecnológicos (IPC) por tecnologia	87
Figura 54 - <i>Dashboard</i> número de proponentes por tecnologia.....	87

Figura 55 – Quantidade de Patentes por Ano de Publicação <i>Autonomous Vehicles</i>	88
Figura 56 – Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Autonomous Vehicles</i>	89
Figura 57 – Quantidade de Patentes por Ano de Publicação <i>Advanced Driver Assistance Systems (ADAS)</i>	89
Figura 58 - Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Advanced Driver Assistance Systems (ADAS)</i>	90
Figura 59 – Quantidade de Patentes por Ano de Publicação <i>Smart City</i>	90
Figura 60 – Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Smart City</i> ...	91
Figura 61 – Quantidade de Patentes por Ano de Publicação <i>Smart Mobility</i>	91
Figura 62 – Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Smart Mobility</i>	92
Figura 63 – Quantidade de Patentes por Ano de Publicação <i>Vehicular Safety</i>	92
Figura 64 – Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Vehicular Safety</i>	93
Figura 65 – Quantidade de Patentes por Ano de Publicação <i>Infotainment System</i>	93
Figura 66 – Proporção de Patentes por Ano de Publicação por Setor Industrial <i>Infotainment System</i>	94
Figura 67 - Fluxo de conhecimento número Derwent 2002-131700 tecnologia Smart Mobility	96
Figura 68 - Fluxo de conhecimento das patentes da tecnologia <i>Smart Mobility</i> (fase inicial).97	
Figura 69 - Fluxo de conhecimento das patentes da tecnologia <i>Smart Mobility</i> (fase final) ...	98
Figura 70 – Legenda Mapa do Fluxo de Conhecimento	99

LISTA DE SIGLAS

ABIPÉÇAS	Associação Brasileira da Indústria de Autopeças
ABPI	Associação Brasileira da Propriedade Intelectual
ABS	<i>Anti-Block System</i>
ACC	<i>Adaptative Cruiser Controllers</i>
ADAS	Sistemas de Assistência Avançada ao Motorista
AIS	Informática e Entretenimento Automotivo
ANFAVEA	Associação Nacional dos Fabricantes de Veículos Automotores
AV	Veículos Autônomos
C	Certificado de Adição de Invenção
COFINS	Contribuição para o Financiamento da Seguridade Social
CSV	<i>Comma Separated Values</i>
DII	<i>Derwent Innovations Index</i>
ESPACENET	Base de Patentes do Escritório Europeu de Patentes
Eurostat	Gabinete de Estatísticas da União Europeia
HMI	<i>Human Machine Interface</i>
HTML	Linguagem de Marcação de Hipertexto
ICMS	Imposto sobre Circulação de Mercadorias e Serviços
INPI	Instituto Nacional da Propriedade Industrial
IP	Propriedade Intelectual
IPC	Classificação Internacional de Patentes
IPI	Imposto sobre Produtos Industrializados
IPVA	Imposto sobre a Propriedade de Veículos Automotores
IRI	<i>Economics of Industrial Research and Innovation</i>
ISO	<i>International Organization for Standardization</i>
LATIPAT	Pesquisa de Patentes da Espanha e América Latina
LED	Diodo Emissor de Luz
MU	Patente de Modelo de Utilidade
NIT	Núcleos de Inovação Tecnológica
OECD	Organização para a Cooperação e Desenvolvimento Econômico
OICA	Organização Internacional dos Fabricantes de Automóveis

P&D	Pesquisa e Desenvolvimento
PI	Patente de Invenção
PIS	Programa de Integração Social
SINDIPEÇAS	Sindicato Nacional da Indústria de Componentes para Veículos Automotores
TI	Tecnologias da Informação
TRIPS	Acordo sobre Aspectos do Direito de Propriedade Intelectual Relacionados ao Comércio
TRL	Níveis de Maturidade Tecnológica
URL	<i>Uniform Resource Locator</i>
USPTO	<i>United States Patent and Trademark Office</i>
VS	Segurança Veicular
WHO	Organização Mundial de Saúde
WIPO	Organização Mundial da Propriedade Intelectual
WoS	<i>Web of Science</i>
WTO	Organização Mundial do Comércio

SUMÁRIO

INTRODUÇÃO	19
1 FUNDAMENTAÇÃO TEÓRICA.....	25
1.1 INOVAÇÃO	25
1.2 TEORIA DA DIFUSÃO DA INOVAÇÃO	28
1.3 TRANSBORDAMENTO DE CONHECIMENTO (<i>SPILLOVERS</i>)	30
1.4 PROPRIEDADE INTELECTUAL E PATENTES	31
1.5 PROSPECÇÃO TECNOLÓGICA E PATENTOMETRIA	34
1.6 TECNOLOGIAS AUTOMOTIVAS	38
1.6.1 <i>Veículos Autônomos</i>	42
1.6.2 <i>Sistemas de Assistência Avançada ao Motorista</i>	43
1.6.3 <i>Tecnologias de Segurança Veicular</i>	43
1.6.4 <i>Informática e Entretenimento Automotivo</i>	44
1.6.5 <i>Cidades Inteligentes (Smart Cities)</i>	44
1.6.6 <i>Mobilidade Inteligente (Smart Mobility)</i>	45
2 METODOLOGIA.....	46
2.1 CÓDIGO DE PROGRAMAÇÃO <i>WEB CRAWLER WOS.PY</i>	50
2.1.1 <i>Abertura das bibliotecas Python</i>	52
2.1.2 <i>Programação das Telas de Interação com o Usuário</i>	53
2.1.3 <i>Tela de Mensagem Informativa do Programa</i>	54
2.1.4 <i>Criação dos Arquivos no Formato CSV</i>	54
2.1.5 <i>Interação com o Navegador Google Chrome</i>	55
2.1.6 <i>Obtenção das URLs Iniciais (Requests)</i>	56
2.1.7 <i>Processamento do Web Crawler (Scrapy Engine)</i>	57
2.2 EXECUÇÃO DO <i>WEB CRAWLER WOS.PY</i>	58
2.3 ESTRUTURA DO BANCO DE DADOS.....	64
2.4 CRIAÇÃO DO BANCO DE DADOS DO FLUXO DE CONHECIMENTO	70
2.4.1 <i>Código de Programação fluxo_spider.ipynb – Abertura das bibliotecas Python</i>	72
2.4.2 <i>Código de Programação fluxo_spider.ipynb – Criação dos dataframes</i>	73
2.4.3 <i>Código de Programação fluxo_spider.ipynb – interação com o navegador</i>	74
2.4.4 <i>Código de Programação fluxo_spider.ipynb – processamento da web crawler (Scrape Engine) e geração do arquivo bd9_fluxo_conhecimento.CSV</i>	75

2.5 CRIAÇÃO DO MAPA DO FLUXO DO CONHECIMENTO	77
3 RESULTADOS E DISCUSSÕES	80
4. CONSIDERAÇÕES FINAIS E SUGESTÕES PARA TRABALHOS FUTUROS	100

INTRODUÇÃO

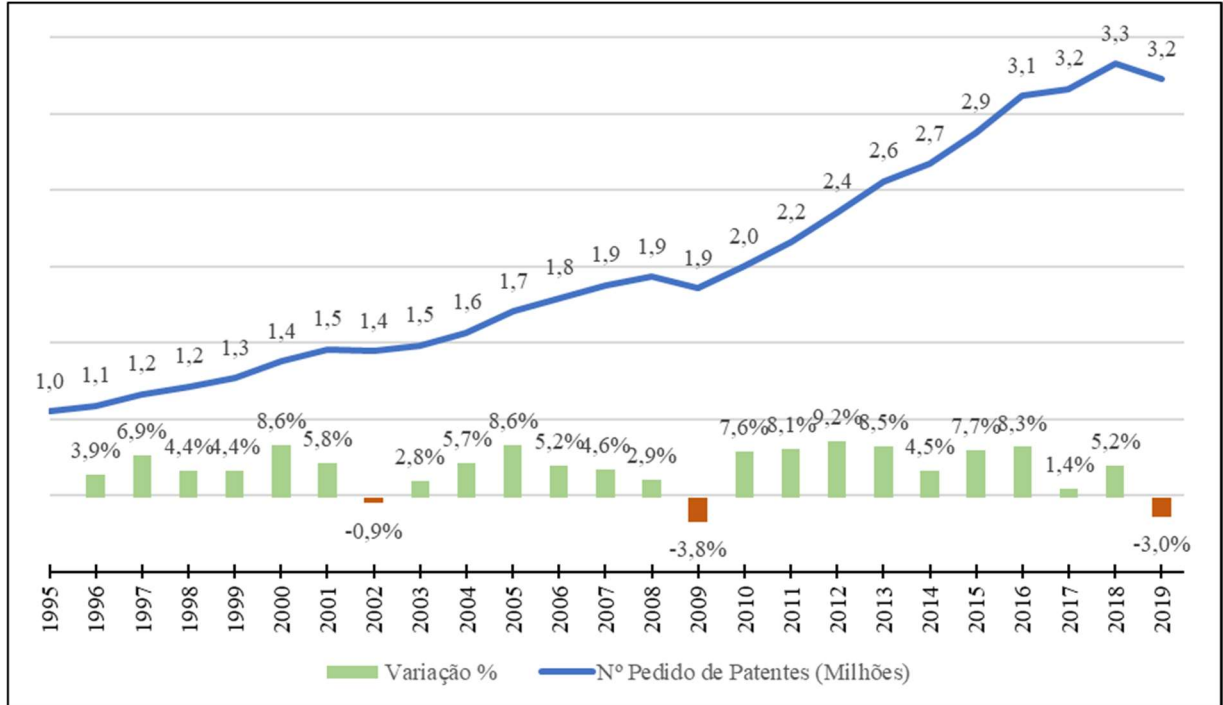
O fenômeno da globalização facilita a expansão do progresso tecnológico ao excluir fronteiras e ampliar conhecimentos, o que resulta em uma economia globalizada, onde a virtualidade está cada vez mais presente e as transformações são velozes, trazendo a necessidade de as organizações buscarem incessantemente novas tecnologias e inovações. Este cenário de crescente oferta e demanda por produtos e serviços incentiva que empreendedores, inventores e outros agentes procurem criar, proteger, divulgar e comercializar suas criações, objetivando retorno econômico, pessoal e para a sociedade. Existem várias formas de proteção das criações intelectuais, sendo que os documentos de patente se destacam por estarem muito concatenados com o desenvolvimento tecnológico (PARANHOS; RIBEIRO, 2018).

De acordo com a Organização Mundial da Propriedade Intelectual (WIPO, do inglês *World Intellectual Property Organization*), em 2019 foram apresentados no mundo 3,2 milhões de pedidos de patente, o que representa uma queda de 3% em relação a 2018. Um declínio substancial nos pedidos de patentes na China foi o principal impulsionador da redução no total global, visto que o escritório de patentes da China recebeu cerca de 141.300 pedidos a menos em 2019 se comparado com 2018. Outros países também apresentaram forte queda nos pedidos de patentes: Japão (-5.598); Federação Russa (-2.446); e o Reino Unido (-1.691). O estudo da curva de pedidos de patentes no mundo por ano, a longo prazo, mostra tendência de crescimento desde 1995, exceto pelos anos de 2002, 2009 e 2019, que diminuiram respectivamente 0,9%, 3,8% e 3% (WIPO, 2020b). A Figura 1 mostra a curva de pedidos de patentes no mundo por ano no período de 1995 a 2019.

No Brasil a legislação que regula a propriedade intelectual é a lei 9.279, de 14 de maio de 1996. O órgão responsável pela concessão dos direitos de propriedade intelectual no Brasil é uma autarquia federal, o Instituto Nacional da Propriedade Industrial (INPI). Em 2019 foram apresentados no Brasil 25.396 pedidos de patentes, um aumento de 2,2% em relação a 2018. Esta quantidade de pedidos de patentes representou 0,84% do total mundial e colocou o país como o 10º maior solicitante de pedidos de patentes no mundo. China e Estados Unidos representaram os maiores solicitantes de patentes no mundo, com quase 67% do total. Em função da presença da China, Japão e Coreia do Sul o continente asiático concentrou 65% dos pedidos de patentes no mundo em 2019 (WIPO, 2020a). A Tabela 1 mostra os dez maiores

países solicitantes de patentes no mundo em 2019. A Figura 2 mostra a concentração de pedidos de patentes por continente em 2019.

Figura 1 - Curva de pedidos de patentes no mundo por ano



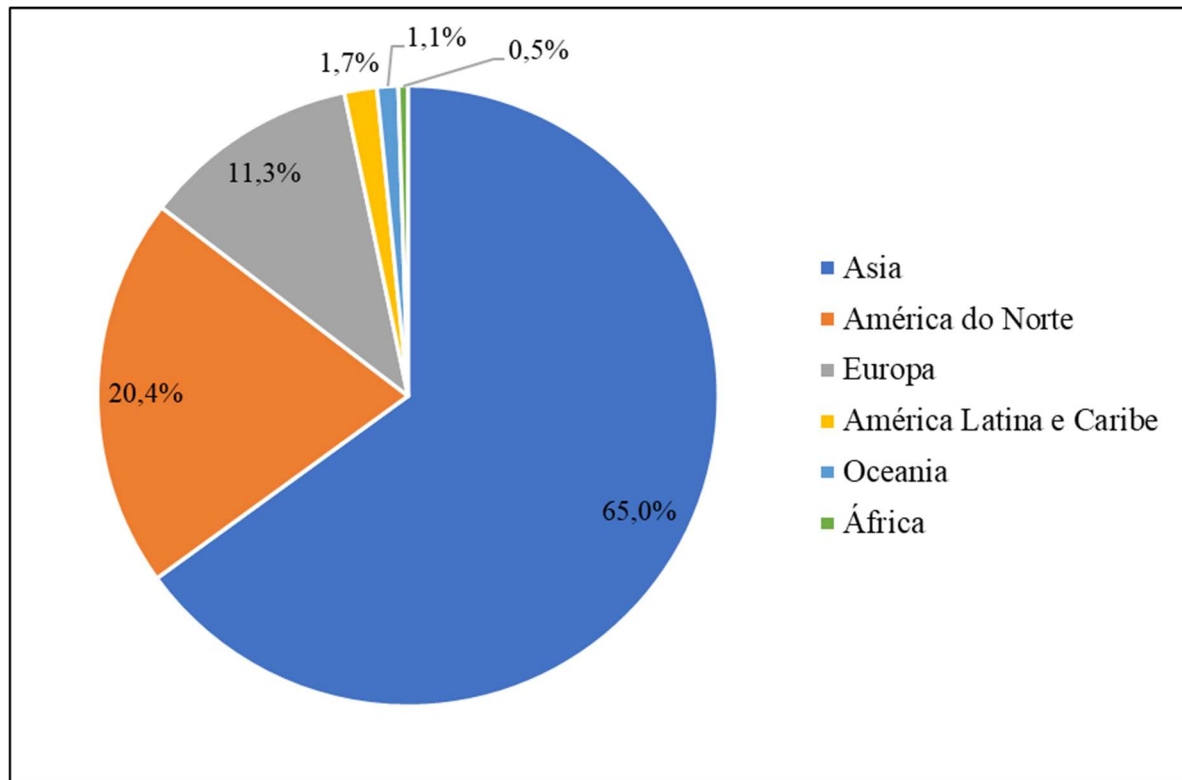
Fonte: adaptado de (WIPO, 2020a)

Tabela 1 – Top 10 países maiores solicitantes de patentes no mundo em 2019

País	2018	2019	Variação %	% Mundo
China	1.542.002	1.400.661	-9,17%	46,19%
Estados Unidos	597.141	621.453	4,07%	20,50%
Japão	313.567	307.969	-1,79%	10,16%
Coréia do Sul	209.992	218.975	4,28%	7,22%
Alemanha	67.898	67.434	-0,68%	2,22%
Índia	50.055	53.627	7,14%	1,77%
Canadá	36.161	36.488	0,90%	1,20%
Rússia	37.957	35.511	-6,44%	1,17%
Austrália	29.957	29.758	-0,66%	0,98%
Brasil	24.857	25.396	2,17%	0,84%

Fonte: adaptado de (WIPO, 2020a)

Figura 2 - Concentração de pedidos de patentes por continente em 2019



Fonte: adaptado de (WIPO, 2020b)

Drucker (1998) afirma que “a inovação pode ser gerida, desde que se saiba para onde e como olhar”, sendo que “a maioria das inovações e, especialmente, as de maior sucesso, resultam de uma procura consciente e intencional de oportunidades de inovação”.

Neste sentido as técnicas de prospecção tecnológica são fundamentais não apenas para verificar o estado da arte de um campo tecnológico, mas também para indicar territórios ou mercados potenciais, permitindo o mapeamento de eventuais parceiros ou concorrentes, e garantindo a exclusividade necessária à introdução de uma inovação no mercado (FORTEC, 2019).

Os métodos de prospecção tecnológica diferem nos tipos de abordagens e nas habilidades requisitadas (PARANHOS; RIBEIRO, 2018).

Existem diversas técnicas de prospecção tecnológica que podem ser usadas para a geração de informações, aplicadas isoladamente ou em conjunto com outras em estudos de inteligência competitiva e *foresight*: a) *brainstorming*; b) conferências e *workshops*; c) entrevistas; d) *genius forecasting*; e) mapeamento de arcabouço legal e de jurisprudência associada; f) mapeamento de dados empresariais; g) mapeamento patentário; h) *nominal group process*; i) opinião de especialistas; j) painéis de especialistas; k) painéis sociais; l) pesquisa

delphi; m) questionários e pesquisas de levantamento; n) Revisão bibliográfica; e o) votação (ANTUNES *et al.*, 2018).

A busca de anterioridade de patentes é provavelmente a forma mais simples de prospecção tecnológica e consiste em uma revisão minuciosa sobre o estado da técnica ao buscar patentes correlacionadas à invenção e trabalhos científicos. O mapeamento patentário, com o uso de citações de documentos de patente, permitem relacionar fluxos de conhecimento sobre um determinado campo tecnológico, a partir de buscas por palavras-chave ou pela Classificação Internacional de Patentes (IPC, do inglês *International Patents Classification*), sendo esse tipo de busca especialmente eficiente se houver uma família de patentes grande (QUINTELLA *et al.*, 2018).

Desde a invenção do primeiro automóvel em 1886, por Karl Benz (VON FERSEN, 2013), e a instauração da primeira linha de produção em massa no final de 1913, por Henry Ford (VON FERSEN, 2013), a indústria automobilística tem como uma de suas características mais forte a busca constante pela inovação dos veículos automotores e dos processos que são utilizados para produzi-los.

Como premissa deste trabalho ressalta-se o uso das informações contidas nos documentos de patentes, disponíveis na plataforma *Derwent* na base de dados *Web of Science*, a fim de evidenciar o comportamento de uma determinada tecnologia automotiva em função do tempo. Este estudo também buscou criar mecanismos para determinar a importância e duração de uma determinada tecnologia, bem como o transbordamento desta tecnologia do setor automotivo para outros setores da indústria.

A atuação do autor como Engenheiro do Produto, especificamente nas áreas de construção de protótipos, veículos especiais, modelos de controle e design, e gerenciamento de projetos, na Volkswagen do Brasil Indústria de Veículos Automotores Ltda., situada na cidade de São Bernardo do Campo, Estado de São Paulo, foi a principal motivação que o levou à realização desta pesquisa.

A pesquisa se justifica pela contribuição que traz ao estudo do comportamento e das características da inovação tecnológica, com base em variáveis obtidas em documentos de patentes, um gap de pesquisa em uma área temática pouco explorada pela comunidade científica.

Os resultados deste estudo permitiram a criação de um *software* de prospecção tecnológica que, por meio de palavras-chave, promovem a coleta das informações contidas nos

documentos de patentes disponíveis na plataforma *Derwent* na base de dados *Web of Science*. O *software* gera diversos bancos de dados relacionais, o que permite o uso estruturado destas informações para estudo de um determinado campo tecnológico.

Baseado no contexto aqui apresentado o Objetivo Geral desta pesquisa é desenvolver um aplicativo de software para *web crawler*, em linguagem *Python*, que possibilite a prospecção tecnológica, a partir de palavras-chave, de informações nos documentos de patentes da base *WoS*.

Como objetivos específicos desta pesquisa elenca-se:

1. Obter e organizar informações obtidas nos documentos de patentes da base *WoS* em banco de dados relacionais, mais especificamente nas áreas de segurança veicular, informações e entretenimento veicular, mobilidade urbana, cidades inteligentes, veículos conectados e automatizados, reconhecidas como sendo tecnologias relevantes e emergentes na indústria automobilística.
2. Criar, por meio das informações de documentos de patentes organizadas em bancos de dados relacionais, gráficos e ferramentas que possibilitem o estudo e a interpretação dos dados das tecnologias pesquisadas.
3. Desenvolver um software em linguagem *Python* para, a partir dos bancos de dados gerados, estruturar redes de pesquisa que possibilitem o mapeamento do fluxo de conhecimento das tecnologias pesquisadas.
4. Verificar a ocorrência de *spillover* (quando uma tecnologia desenvolvida com um propósito, “transborda” para outra área e ou outra tecnologia).

Além desta introdução, este trabalho contém quatro seções: a primeira seção apresenta a fundamentação teórica com conceitos de inovação, transbordamento de conhecimento, teorias da difusão da inovação, propriedade intelectual e patentes, prospecção tecnológica e patentometria, e tecnologias automotivas; a segunda seção trata da metodologia utilizada neste estudo, com a apresentação dos códigos de programação em *Python*, execução do aplicativo *web crawler*, estrutura do banco de dados e criação do mapa do fluxo de conhecimento; os resultados e discussões é a temática desenvolvida na terceira seção; finalmente, na quarta seção, apresentam-se as considerações finais e as sugestões de trabalhos futuros. A Figura 3 apresenta a estrutura desta dissertação e realça os principais tópicos e subtópicos abordados em cada um dos capítulos.

Figura 3 - Visão geral da estrutura do trabalho

Introdução
<ul style="list-style-type: none"> • Objetivo Geral • Objetivos Específicos
Fundamentação Teórica
<ul style="list-style-type: none"> • Inovação • Transbordamento de Conhecimento (Spillovers) • Teorias da Difusão da Inovação e Aceitação da Tecnologia • Propriedade Intelectual e Patentes • Prospecção Tecnológica e Patentometria • Tecnologias Automotivas
Metodologia
<ul style="list-style-type: none"> • Código de Programação <i>web crawler</i> <i>wos.py</i> • Execução do <i>Web Crawler</i> <i>wos.py</i> • Estrutura do Banco de Dados • Criação do Mapa do Fluxo do Conhecimento
Resultados e Discussões
Considerações Finais e Sugestões de Trabalhos Futuros
Referências

1 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica desta pesquisa está balizada em autores de diferentes áreas do conhecimento, com destaque para as áreas de: Inovação; Prospecção Tecnológica; Propriedade Intelectual; Patentes; e Engenharia Automobilística.

O referencial utilizado apresenta conceitos relacionados a: Inovação; Transbordamento de Conhecimento (*Spillover*); Teorias da Difusão da Inovação e Aceitação da Tecnologia; Documentos de Patentes; Prospecção Tecnológica e Patentometria; e Tecnologias Automotivas.

1.1 INOVAÇÃO

Schumpeter (1961) afirma que a inovação é um elemento básico para o desenvolvimento econômico. Para o autor a inovação se processa na forma de: a) introdução no mercado de um novo ou aperfeiçoado produto; b) adoção de um novo método de produção; c) abertura de um novo mercado; d) a conquista de uma nova fonte de suprimento de matérias-primas; e e) execução de um novo sistema de organização empresarial.

Tigre (1998) afirma que, do ponto de vista financeiro, a primeira aplicação comercial de uma inovação pode não representar impactos significativos, sendo, portanto, mais importante a velocidade e abrangência da difusão destas inovações na economia. Para o autor a difusão de inovações depende de condições favoráveis, como inovações complementares, infraestrutura apropriada, aceitação por parte de empresários e consumidores, mudanças na legislação e aprendizado no uso de novas tecnologias. Para Tigre (1998), embora a inovação permita o crescimento das empresas, com a criação de novos mercados e o exercício temporário de poder monopolístico, somente a ampla difusão desta inovação tem impacto macroeconômico.

Para Tigre (2006) a estratégia tecnológica defensiva das empresas englobam esforços de inovação sistêmicos e desenvolvimento de produtos constante. O autor observa que, em sua

estratégia tecnológica, algumas empresas preferem não correr riscos de serem as primeiras a difundir uma inovação, optando por esperar que os seus concorrentes o façam, aprendendo com os erros deles para que, quando entrar neste novo mercado, o faça com produtos melhores.

Schmoch (2007) estuda o comportamento do desenvolvimento de campos tecnológicos em função do tempo, por meio de variáveis de patentes, com base nos modelos lineares de inovação *science-push* e *market-pull*, e em modelos não lineares. Para Schmoch (2007) as patentes refletem as atividades inventivas nos campos tecnológicos, sendo que cerca de 60% representam inovações. Desta forma considera-se que o número de patentes representa o nível de atividades tecnológicas em um campo específico, sendo possível, por meio das estatísticas de patentes, rastrear estas atividades tecnológicas em função do tempo. O autor define o modelo *science-push* como sendo o caracterizado por uma sequência de etapas, começando pela ‘pesquisa básica’ pura (ou pesquisa orientada pela curiosidade), seguida por ‘pesquisa aplicada’, ‘desenvolvimento experimental’, e, por fim, a ‘inovação’ em si. No modelo *market-pull* (ou *demand-pull*) a etapa inicial ‘pesquisa básica’ é substituída por ‘descoberta tecnológica’ ou ‘demanda de mercado’, seguindo as demais etapas de forma análoga ao modelo *science-push*. Schmoch (2007) observa que a percepção dos processos de inovação foi radicalmente alterada por abordagens de inovação não lineares. Kline (1985), *apud* Schmoch (2007), propõe um modelo não linear de inovação, conhecido como *chain-linked*, focado em diferentes estágios de desenvolvimento e produção, onde cada uma destas etapas estão ligadas umas às outras por meio de *loops* de *feedback*, sendo que esses loops implicam, a cada passo do desenvolvimento, na criação de conhecimento útil para etapas anteriores. Schmoch (2007) conclui que a investigação a longo prazo do desenvolvimento de tecnologia revela mais frequentemente ciclos bimodais do que simples curvas em forma de sino, principalmente quando aplicada a campos tecnológicos complexos e baseados em ciência. O autor entende que o primeiro ciclo pode ser associado ao modelo de inovação *science-push* e o segundo ciclo ao modelo *market-pull*, porém não com sequenciamento linear.

Carvalho (2008) afirma que a inovação na indústria automobilística é resultado dos processos de Pesquisa e Desenvolvimento (P&D). O autor explana várias inovações importantes originadas na indústria automobilística: a) nos anos de 1970 a emergência das montadoras japonesas no cenário internacional introduziu métodos de organização e de gestão de produção criados e desenvolvidos pela Toyota; b) a década de 80, com os avanços da microeletrônica e dos sistemas de informação, foi marcada pela difusão do toyotismo, com a introdução do conceito de produção enxuta (*lean manufacturing*); c) os anos de 1990 destacou-

-se o deslocamento do foco competitivo para o desenvolvimento de produtos e para o avanço do processo de globalização; e d) a partir de meados dos anos 90 observou-se um crescente empenho das empresas automobilísticas no desenvolvimento das chamadas tecnologias automotivas avançadas. Para o autor a indústria automobilística incrementou a utilização de sistemas e de componentes eletrônicos, sendo que praticamente todas as principais funções dos veículos automotores - como aceleração, frenagem, controle de tração, de estabilidade, injeção e combustão de combustível, segurança veicular ativa e passiva, navegação, comunicação entre veículos e o ambiente, navegação e entretenimento - já são controladas e/ou viabilizadas pela eletrônica embarcada.

Segundo Ramos e Zilber (2015) investimentos em P&D favorecem a cultura da inovação, pois expõe as empresas aos benefícios advindos no processo de assimilação de conteúdo, treinamento técnico e operacional necessário, assim como na troca de conhecimentos tácitos com os fornecedores das novas tecnologias. Para os autores as empresas que atuam em mercados competitivos necessitam investir continuamente na aquisição e no desenvolvimento de capacidades inovadoras, pois são estas aptidões que garantem a sobrevivência e impulsionam os negócios no mercado.

A capacidade de uma empresa em ser inovadora está relacionada com a sua capacidade em investir em P&D, disponibilizando uma base de fomentação e viabilização de projetos inovadores, o que resulta em benefícios econômicos duradouros, como a elevação nas vendas e nos lucros. As empresas que atuam em mercados competitivos necessitam investir continuamente na aquisição de P&D e no desenvolvimento de capacidades inovadoras (DA SILVA MORAES *et al.*, 2018).

A Organização para a Cooperação e Desenvolvimento Econômico e o Gabinete de Estatísticas da União Europeia (OECD/EUROSTAT, 2018), na última edição do Manual de Oslo, define a inovação empresarial como um produto ou processo de negócios novo ou aprimorado, que difere significativamente dos produtos ou processos de negócios anteriores da empresa, e que tenha sido introduzido no mercado ou colocado em uso pela empresa. Um produto pode ser definido como um bem ou serviço (ou uma combinação de ambos). Os processos de negócios incluem todas as atividades essenciais da empresa para a produção dos produtos, incluindo todas as suas atividades auxiliares ou de apoio para este fim.

Segundo a OECD/Eurostat (2018) as atividades de inovação empresarial abrangem todas as atividades de desenvolvimento, financeiras e comerciais realizadas por uma empresa e que se destinam a resultar em uma inovação. Pode-se listar como características das atividades

de inovação empresarial: a) as empresas podem realizar atividades de inovação internamente ou por meio de organizações externas; b) as atividades de inovação podem ser adiadas ou abandonadas devido a vários motivos; c) as atividades de inovação podem criar conhecimento ou informação que não são usados para introduzir uma inovação; e d) as empresas podem usar os resultados de suas atividades de inovação, incluindo inovações, novos conhecimentos e novas informações, para seu próprio benefício, reter os resultados para uso interno ou transferindo, vendendo ou licenciando estes resultados para outras empresas ou organizações.

A OECD/Eurostat (2018) identifica oito grandes tipos de atividades que as empresas podem realizar na busca da inovação: a) atividades de P&D experimental; b) engenharia, design e outras atividades de trabalho criativo; c) atividades de marketing; c) atividades relacionadas à propriedade intelectual (IP, do inglês *intellectual property*); d) atividades de treinamento de funcionários; e) atividades de desenvolvimento de software e banco de dados; f) atividades relacionadas com a aquisição ou arrendamento de ativos tangíveis; e g) atividades de gestão da inovação.

As atividades relacionadas à propriedade intelectual incluem as atividades da empresa relacionadas à proteção de conhecimentos vindos por meio de P&D, desenvolvimento de software e engenharia, design e outros trabalhos criativos. As atividades de IP reúnem todo o trabalho administrativo e jurídico para se inscrever, registrar, documentar, gerenciar, comercializar, licenciar e fazer cumprir os direitos de IP da própria empresa, a aquisição definitiva ou por meio de licenciamento de direitos de IP de outras organizações, e as atividades de venda de IP a terceiros (OECD/EUROSTAT, 2018).

1.2 TEORIA DA DIFUSÃO DA INOVAÇÃO

Rogers (2010) define a difusão da inovação como um processo em resposta ao aprendizado sobre uma nova tecnologia, comunicada por meio de determinado canal, ao longo do tempo, entre os membros de um sistema social.

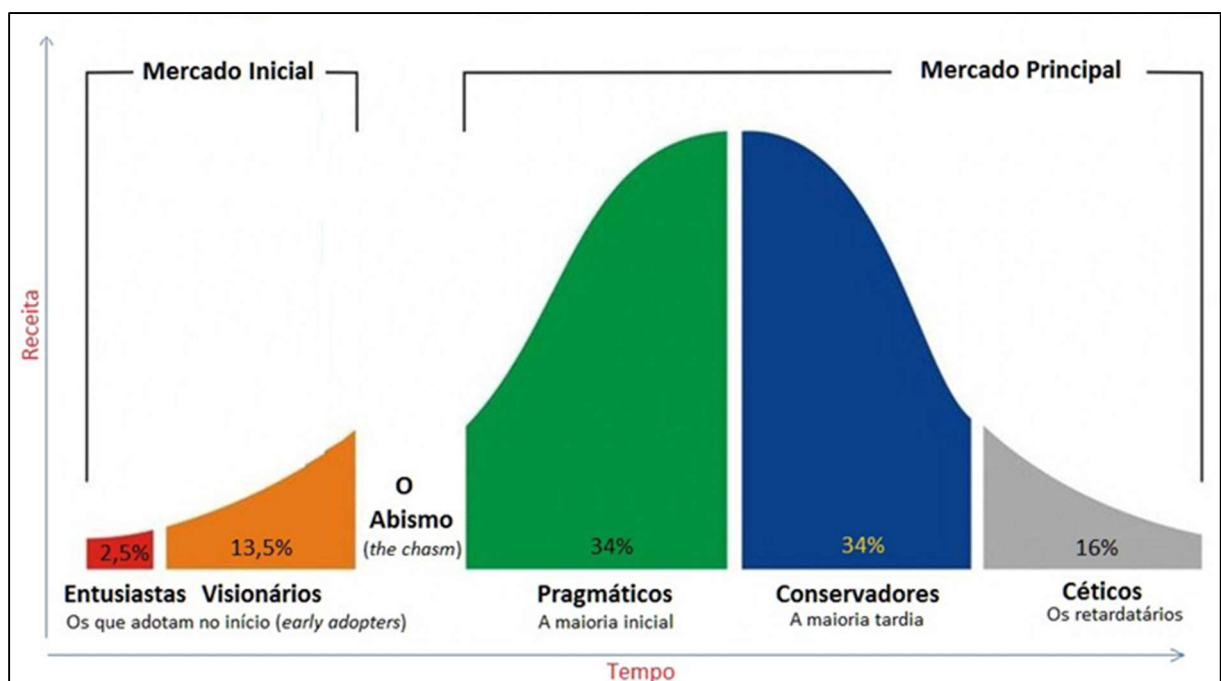
Rogers (2010) e Moore (2014) identificam os cinco comportamentos de adoção de tecnologia: a) Entusiastas (*Innovators*, no termo original em inglês) são os primeiros a adotar inovações, e, como são entusiastas de novas tecnologia, gostam de aprender sobre tecnologia e testam novos produtos assim que disponíveis, não sendo influenciados pelas regras sociais; b)

Visionários (*Early adopters*) estão cientes das vantagens competitivas proporcionadas por tecnologias avançadas e, assim que percebem que as vantagens superam as desvantagens, adotam a nova tecnologia; c) Pragmáticos (*Early majority*) focam em inovações que já provaram sua eficiência e seguem a estratégia dos Visionários; d) Conservadores (*Late majority*), são preocupados em reduzir riscos e custos e adotam novas tecnologias em função da influência social para fazê-lo, ou para evitar ser deixado para trás pelos seus concorrentes; e) Por último, os Céticos (*Laggards*), tem uma abordagem conservadora, sendo resistentes a mudanças e menos suscetíveis à pressão da sociedade para adotar uma nova tecnologia, sendo que, em alguns casos, não a adotam.

Moore (2014) define O Abismo (*The Chasm*) como uma quebra no crescimento da curva de adoção de uma inovação, que ocorre quando as empresas falham ao não adaptarem sua oferta de valor aos requisitos dos pragmáticos. É neste período que muitas inovações de alta tecnologia falham, não atingindo o mercado consumidor principal, e são descontinuadas. A Figura 4 mostra a curva da adoção da inovação.

Para Dearing e Cox (2018) a difusão da inovação, ou a falta dela, pode ser explicada por três conjuntos gerais de variáveis: a) os prós e contras da inovação ou seus atributos; b) as características dos adotantes da nova tecnologia, especialmente as percepções dos formadores de opinião; e c) o contexto social e político, isto é, como os proponentes e opositores enquadram o significado da inovação e o momento da sua introdução.

Figura 4 - Curva da Difusão da Inovação



Fonte: adaptado de Moore (2014)

1.3 TRANSBORDAMENTO DE CONHECIMENTO (*SPILLOVERS*)

Xiang *et al.* (2013) propõe um método para a construção de redes de *spillovers* de conhecimento utilizando as citações de patentes em conjunto com as informações de inventores. O autor afirma que existem dois tipos de conhecimento em uma organização: a) o conhecimento explícito, que é o que pode ser codificado e registrado em documentos, e facilmente difundido por vários meios; e b) o conhecimento tácito, interno à organização, altamente específico ao contexto, e não pode ser facilmente transferido exceto por meio de comunicação interpessoal intensiva e cooperação. O autor conclui que estudos de *spillovers* baseados em citações de patentes podem levar à compreensão da difusão de tecnologia, porém é um indicador falho ao refletir a difusão de conhecimento tácito, em função de sua limitação em identificar os *spillovers* que ocorrem por meio de contatos pessoais

Purificação (2015) estuda a análise de transbordamento de conhecimento por meio de redes sociais. A autora define redes sociais como uma estrutura constituída por uma série de agentes que estão conectados por uma ou mais relações de interdependência. No estudo criou-se redes sociais de transbordamento utilizando-se como agentes destas redes as variáveis de documentos de patentes: a) citação de patentes; b) inventores; e c) depositantes. A metodologia do trabalho aplica a teoria dos grafos como ferramenta matemática na análise das relações existentes entre os agentes nas redes de transbordamento criadas. Purificação (2015) define *cluster* industriais como um local geográfico, com concentração de firmas de alta tecnologia e centros de pesquisas científicas, como universidades ou laboratórios, cujo ambiente, em função destas características, é próprio ao transbordamento de conhecimento e a fomentação da inovação tecnológica.

Vilela Júnior (2015) analisa a importância da ocorrência de *spillovers* em *cluster* industriais para a criação de conhecimento, identifica os elementos associados que influenciam nesta ocorrência e a forma como essa influência acontece. O autor define *spillover* como diferentes conhecimentos criados em determinada organização ou em atividades conjuntas e extravasados de onde tenham sido originados. Vilela Júnior (2015) afirma que a criação de conhecimento nos *clusters* industriais é possível em função da propagação do conhecimento em um local e a difusão deste conhecimento por meio de *spillovers*. Segundo o autor os elementos que possibilitam a difusão do conhecimento em clusters industriais são: a) as relações informais existentes; b) a rotatividade de pessoal entre as indústrias; e c) atividades colaborativas entre os membros do *cluster* industrial. A convergência de interesses e de setores

de atuação entre os membros de um *cluster* industrial facilita a ocorrência de *spillovers*, enquanto a existência de diferenças tecnológicas dificulta esta ocorrência. O autor conclui que os *spillovers* favorecem a criação de conhecimento nos *clusters* industriais e não apenas em seus membros, sendo necessário que os *spillovers* superem o nível interorganizacional e sejam amplamente difundidas.

Jardim (2019) estuda a transferência de conhecimento das empresas brasileiras internacionalizadas por meio das variáveis de patentes. A autora lista quatro elementos que proporcionam os efeitos positivos dos *spillovers* para estas empresas: a) o enraizamento da empresa na localidade onde atua, sendo que este enraizamento é maior quanto maiores as relações entre produtores, fornecedores e consumidores, mediante o compartilhamento de conhecimento; b) a ocorrência de *spillovers* entre as subsidiárias internacionalizadas e as empresas locais, sendo que a empresa de menor produtividade tenderá a obter mais *spillovers* do que a de maior produtividade; c) quanto menor for o grau de rivalidade de conhecimento entre os agentes, maiores serão as possibilidades de apropriação de *spillovers* por uma empresa, sendo que o grau de rivalidade determina o quanto que as organizações do país hospedeiro poderiam apropriar-se de *spillovers* procedentes de outras organizações; e d) quanto maior o nível de capacitação e absorção dos agentes, maior a captura de *spillovers*. Segundo Jardim (2019), a metodologia mais utilizada para identificar a captura de *spillovers* de conhecimento e suas características utiliza citações de patentes. Para a autora as patentes representam um repositório de informações sobre o desenvolvimento de uma tecnologia, referenciando-se o conhecimento prévio existente que caracteriza o escopo tecnológico de uma invenção por meio das citações de patentes. Quando uma patente cita outra patente, gera-se um *spillover* com origem na patente citada, e, como as patentes são geradas por equipes de inventores, associa-se a ocorrência de *spillovers* por meio do fluxo de conhecimento entre as equipes envolvidas nas patentes citadas e citantes. Jardim (2019) observa que, uma limitação ao uso das citações de patentes como meio para o mapeamento da ocorrência de *spillovers* de conhecimento, é o fato destas citações poderem ser determinadas também pelo examinador da patente, e não apenas por seus inventores ou depositantes.

1.4 PROPRIEDADE INTELECTUAL E PATENTES

A convenção da WIPO define como propriedade intelectual “a soma dos direitos relativos às obras literárias, artísticas e científicas, às interpretações dos artistas intérpretes e às

execuções dos artistas executantes, aos fonogramas e às emissões de radiodifusão, às invenções em todos os domínios da atividade humana, às descobertas científicas, aos desenhos e modelos industriais, às marcas industriais, comerciais e de serviço, bem como às firmas comerciais e denominações comerciais, à proteção contra a concorrência desleal e todos os outros direitos inerentes à atividade intelectual nos domínios industrial, científico, literário e artístico” (WIPO, 1967).

De acordo com o *United States Patent and Trademark Office* (USPTO), órgão que regulariza a concessão de patentes e o registro de marcas nos Estados Unidos, a patente de uma invenção é a concessão de um direito de propriedade ao inventor, com prazo de 20 anos. O direito conferido pela concessão de patente é, na linguagem da própria concessão, “o direito de excluir terceiros de fazer, usar, oferecer para venda ou vender” ou “importar” a invenção nos Estados Unidos (USPTO, 2015).

Segundo Garcez Júnior e Moreira (2017) a função do sistema de propriedade intelectual de um país é garantir a propriedade ou a exclusividade resultante da atividade intelectual nos campos industrial, científico, artístico e literário. Para o autor as patentes apresentam-se como um importante mecanismo de troca, onde a sociedade oferece o privilégio da proteção contra a concorrência, no curto prazo, em troca dos frutos da inovação a longo prazo.

Braga *et al.* (2018) afirma que, quando um pedido de patente é depositado, o requerente, entre outros requisitos, deve divulgar suficientemente a invenção reivindicada de modo que um especialista técnico possa reproduzi-la, sendo que esta suficiência de divulgação desempenha um papel importante no chamado “*quid pro quo*”. Portanto considera-se uma patente não apenas como uma “recompensa”, mas também como um contrato entre os inovadores e a sociedade, no qual um direito de propriedade temporário é concedido em troca da divulgação da inovação.

A Associação Brasileira da Propriedade Intelectual (ABPI), em seu site, afirma que a propriedade intelectual garante a proteção jurídica de uma invenção, por determinado período, em nome do seu inventor ou titular do direito sobre a invenção, protegendo não apenas a atividade criativa em si, mas também os investimentos que são feitos para levar estas invenções ao mercado. No mundo inteiro os detentores de direitos de propriedade intelectual são protegidos por leis específicas contra o uso não autorizado de seus trabalhos, produtos, processos, marcas e serviços. Este direito temporário de exploração comercial exclusiva de uma propriedade intelectual estimula a inovação e o empreendedorismo (ABPI, 2020).

Ainda segundo a ABPI (2020), a Organização Mundial do Comércio (WTO, do inglês *World Trade Organization*) criou o Acordo sobre Aspectos do Direito de Propriedade Intelectual Relacionados ao Comércio (TRIPS, do inglês *Trade Related Aspects of Intellectual Property Rights*), do qual o Brasil tornou-se signatário desde 1994. O TRIPS estabelece um padrão de proteção mínima à propriedade intelectual, e os países que o assinaram obrigaram-se a revisar suas leis nacionais de modo a adaptá-las a esse padrão. No Brasil a legislação que regula a propriedade intelectual é a lei 9.279, de 14 de maio de 1996. O órgão responsável pela concessão dos direitos de propriedade intelectual no Brasil é uma autarquia federal, o Instituto Nacional da Propriedade Industrial (INPI).

A propriedade intelectual pode ser dividida em: a) a propriedade industrial, que se refere a marcas, patentes, desenhos industriais e indicações geográficas utilizadas no meio empresarial; b) o direito autoral, que compreende os trabalhos artísticos, como livros, pinturas, músicas, filmes, poesias *etc.*; c) a proteção *sui generis*, que inclui obras como a topografia de circuito integrado, o cultivar e o conhecimento tradicional (ABPI, 2020).

Em seu *site* o INPI define patente como “um título de propriedade temporária sobre uma invenção ou modelo de utilidade, outorgado pelo Estado aos inventores ou autores ou outras pessoas físicas ou jurídicas detentoras de direitos sobre a criação. Com este direito, o inventor ou o detentor da patente tem o direito de impedir terceiros, sem o seu consentimento, de produzir, usar, colocar à venda, vender ou importar produto objeto de sua patente e/ ou processo ou produto obtido diretamente por processo por ele patenteado. Em contrapartida, o inventor se obriga a revelar detalhadamente todo o conteúdo técnico da matéria protegida pela patente”. Ainda em seu *site* o INPI define os tipos de patentes e o tempo de duração: a) Patente de Invenção (PI), concedida a “produtos ou processos que atendam aos requisitos de atividade inventiva, novidade e aplicação industrial”, cuja validade é de “20 anos a partir da data do depósito”; b) Patente de Modelo de Utilidade (MU), concedida a “objeto de uso prático, ou parte deste, suscetível de aplicação industrial, que apresente nova forma ou disposição, envolvendo ato inventivo, que resulte em melhoria funcional no seu uso ou em sua fabricação”, cuja validade é de “15 anos a partir da data do depósito”; c) Certificado de Adição de Invenção (C), concedido a “aperfeiçoamento ou desenvolvimento introduzido no objeto da invenção, mesmo que destituído de atividade inventiva, porém ainda dentro do mesmo conceito inventivo”, cujo “o certificado será acessório à patente e com mesma data final de vigência desta” (INPI, 2020).

1.5 PROSPECÇÃO TECNOLÓGICA E PATENTOMETRIA

Trippe (2003) define patentometria (ou *Patinformatics*, no termo em inglês) como a ciência de analisar informações de patentes com o objetivo de descobrir relacionamentos e tendências tecnológicas. O autor descreve a patentometria como uma ciência de nível macro, isto é, uma análise que lida com grandes quantidades de informações de patentes. Trippe (2003) descreve como ferramentas da patentometria: a) inteligência de patentes, que consiste no uso de informações de patentes para identificar as capacidades técnicas de uma organização e o uso dessa inteligência para desenvolver um planejamento estratégico; b) mapeamento de patentes, que usa as informações contidas nos documentos de patentes para representar gráfica ou fisicamente uma visão do estado da arte de um campo tecnológico ou de uma invenção; e c) análise de citação de patentes para determinar a importância de uma patente, ou auxiliar na identificação de potenciais parceiros para o licenciamento de novos produtos.

Segundo Mayerhoff (2008) os estudos de prospecção tecnológica constituem uma ferramenta básica para a fundamentação nos processos de tomada de decisão e podem ser definidos como “um meio sistemático de mapear desenvolvimentos científicos e tecnológicos futuros capazes de influenciar de forma significativa uma indústria, a economia ou a sociedade como um todo”. A terminologia comumente utilizada nos estudos de prospecção tecnológica inclui as expressões *Future Research*, *Future Studies*, *Prospective Studies*, Prospectiva Estratégica, *Futuribles*, *Forecasting*, *Foresight*, entre outros. Para o autor os estudos de prospecção tecnológica encontram no sistema de propriedade intelectual, mais propriamente no registro de patentes, um recurso valioso, visto que as bases de dados indexadoras de patentes crescem significativamente em função da crescente importância das patentes na economia mundial.

Quintella *et al.* (2011) apresentam exemplos de prospecção tecnológica de patentes por meio da busca de anterioridade. Os autores afirmam que, por meio da busca de anterioridade, é possível verificar se uma tecnologia já foi desenvolvida anteriormente, se já foi apropriada, ou se já está disponível no mercado. Segundo os autores, por meio da prospecção tecnológica pode-se: a) verificar as tecnologias existentes; b) identificar o estágio de maturidade das tecnologias em pesquisa; c) a inserção destas tecnologias na sociedade; d) aspectos de tecnologias concorrentes e lacunas de mercado a serem preenchidas; e) inventores que pesquisam o mesmo

tema; f) países de origem das patentes; g) países onde ocorreram os depósitos de patentes; h) principais empresas depositantes; e i) classificação dos depositantes das patentes.

Trippe (2015) observa que os documentos de patentes disponibilizam uma grande quantidade de informações, de forma estruturada, divididas em uma série de seções que são encontradas em quase todos os documentos, independentemente do país de origem da patente. Essas seções de um documento de patente são representadas por uma página inicial com dados bibliográficos, uma descrição e uma seção de reivindicação. As seções são divididas em subseções que fornecem informações específicas sobre o documento de patente. Essas subseções são segmentadas em campos individuais quando os documentos de patente são processados para entrega eletrônica ou na geração de banco de dados. Estes campos são a base para o processo de patentometria para a geração de relatórios. Trippe (2015) lista os campos das subseções de um documento de patentes: a) requerente, que é a entidade ou pessoa que apresenta o pedido de concessão de patente; b) inventor, que é a pessoa autora da invenção; c) dados prioritários, que identifica os pedidos de patentes anteriores e é composto pelo número do pedido, data do depósito e país ou organização onde o respectivo pedido anterior foi depositado; d) classificação, que significa um código específico que subdivide a tecnologia em unidades distintas; e) citações, que indica a *prior-art* da invenção, isto é, as invenções ou tecnologias que precederam a invenção objeto do documento de patente; e f) descrição da patente, que geralmente especifica o campo técnico que uma invenção se refere, inclui um breve resumo dos antecedentes técnicos da invenção e descreve as características essenciais da invenção com referência a quaisquer desenhos anexos; Para Trippe (2015), qualquer discussão sobre citações de patentes começa com um documento de patente raiz. As referências que o documento raiz cita são definidas como citações reversas, uma vez que são referências que foram publicadas antes do documento raiz. Por outro lado, avançando no tempo a partir do documento raiz, qualquer documento mais recente que faça referência ao documento raiz é referido como uma citação direta. O autor afirma que as citações representam uma relação potencial entre duas invenções, fornecendo um meio de identificar documentos seminais que têm um alto impacto no desenvolvimento de uma tecnologia. Para o autor patentes que são citadas por outras frequentemente possivelmente estão relacionadas a inovações importantes.

Antunes *et al.* (2018) procedem uma revisão bibliográfica e documental com conceitos e técnicas de inteligência competitiva, de *foresight* e de outros temas relacionados, como prospecção tecnológica. Os autores observam que, para se manterem competitivas, as empresas precisam se diferenciar dos seus concorrentes no mercado, por meio da busca da inovação e da

antecipação de tendências e sinais de mudanças. A partir da aplicação de métodos de inteligência competitiva e de *technology foresight* as empresas podem fazer monitoramento do ambiente externo às empresas ou organizações, com foco principalmente mercadológico e tecnológico. Os autores afirmam que a prospecção tecnológica usa diversas atividades e métodos de captação, tratamento e análise de informações para dar subsídio a processos de tomada de decisão. Antunes *et al.* (2018) definem inteligência competitiva como um processo no qual avalia-se a evolução da indústria e o comportamento dos concorrentes, objetivando a manutenção ou o desenvolvimento de uma vantagem competitiva. *Foresight*, segundo os autores, está relacionado à uma visão de futuro, ou seja, envolve um conjunto de abordagens ligadas a considerações de longo prazo para tomada de decisão, sendo considerado como um processo contínuo de comunicação e aprendizado, com o objetivo de ampliar o questionamento e a conscientização sobre o futuro da organização e seu ambiente. Para Antunes *et al.* (2018) o conceito de *technology foresight* pode ser descrito como um processo analítico do futuro, a longo prazo, da ciência, tecnologia, economia e sociedade, com o objetivo de identificar áreas estratégicas de pesquisa e tecnologias emergentes. Segundo Antunes *et al.* (2018) o mapeamento e análise de patentes geram insumos importantes para os estudos de inteligência competitiva e de *foresight*, mostrando tendências e atores relevantes. Os autores identificam o mapeamento patentário como uma das técnicas disponíveis de prospecção tecnológica e afirma ser possível identificar por esta técnica os inventores, os titulares, os campos tecnológicos e as citações a patentes e a artigos anteriores, desta forma identificando os principais provedores de tecnologias. As análises quantitativas usam métodos estatísticos para observar o número de documentos depositados de patentes, enquanto as análises qualitativas podem focar mais no conteúdo das patentes. Antunes *et al.* (2018) afirmam que o mapeamento patentário é realizado em base de dados diversas e lista as etapas para aplicação da técnica: a) definição das melhores bases de dados; b) definição de escopo da busca patentária; c) realização do download dos documentos selecionados pelo escopo; d) remoção das duplicidades, redundâncias e documentos espúrios; e e) preparo das planilhas ou banco de dados para análises estatísticas e análises qualitativas.

Quintella *et al.* (2018) abordam a busca da anterioridade em base de patentes e discutem as motivações e os procedimentos para a realização desta busca, bem como as etapas necessárias para determinar preliminarmente os Níveis de Maturidade Tecnológica (TRL, do inglês *Technology Readiness Level*) de uma tecnologia. Os autores afirmam que a busca de anterioridade é provavelmente a forma mais simples de prospecção tecnológica, e objetiva

investigar se a tecnologia a ser patenteada já existe, bem como identificar tecnologias semelhantes ou afins que já tenham sido divulgadas ou já estejam em uso no mercado. De acordo com Quintella *et al.* (2018) a busca da anterioridade consiste em uma revisão minuciosa sobre o estado da técnica ao buscar patentes correlacionadas à invenção e trabalhos científicos sobre a temática. Os autores citam como fontes de informações científicas, tecnológicas e de mercado nas buscas de anterioridade as bases de dados indexadores de documentos de patentes: a) Base de Patentes do Instituto Nacional da Propriedade Industrial (INPI); b) Base de Patentes do Escritório Americano de Patentes e Marcas (USPTO); c) Base de Patentes do Escritório Europeu de Patentes (ESPACENET); d) Base de Patentes do LATIPAT; e) PATENTSCOPE; f) *Derwent Innovations Index* (DII); e g) *Google Patents*. Para Quintella *et al.* (2018) a utilização de bases indexadoras de patentes como fonte nas buscas de anterioridades é recomendável, visto que contempla aspectos de apropriação, engloba todos os campos tecnológicos, possui indexação regular e sistemática, e padronização internacional dos campos bibliográficos dos documentos de patentes. Os autores mencionam como método de busca de anterioridade o uso das citações em documentos de patente como meio para relacionar fluxos de conhecimento sobre um determinado campo tecnológico, a partir de buscas por palavras-chave ou pelo IPC, sendo esse tipo de busca especialmente eficiente se houver uma família de patentes grande.

Nascimento (2020) propõe a utilização de dados contidos em patentes como suporte em Núcleos de Inovação Tecnológica (NIT), por meio da patentometria, nas etapas relacionadas com: a) a avaliação mercadológica e estratégica, isto é, como fonte de informação tecnológica nas etapas de comunicação da invenção, parceria e vitrine tecnológica; b) na busca da anterioridade de uma tecnologia como auxílio no conhecimento do estado da arte, além do monitoramento de tecnologias existentes, emergentes e obsoletas; c) na captação de parceiros objetivando codesenvolvimento e prospecção de possíveis interessados em tecnologias, elaborando panoramas e perfis tecnológicos da invenção, mercado e academia; e d) na alocação de recursos e esforços como suporte ao conhecimento interno das tecnologias predominantes, políticas de inovação, incentivo ao desenvolvimento de novas tecnologias e programas de incentivo tecnológico. Para o autor a utilização da patentometria, quando incorporada à gestão de propriedade intelectual, suporta definições estratégicas e decisões mais precisas, uma vez que permite um melhor monitoramento e controle do encadeamento das atividades envolvidas na transferência de tecnologia, além de tornar a gestão da propriedade intelectual mais clara e pautada em cenários tecnológicos.

1.6 TECNOLOGIAS AUTOMOTIVAS

De acordo com a Organização Internacional dos Fabricantes de Automóveis (OICA, do francês *Organisation Internationale des Constructeurs d'Automobiles*), a produção mundial de veículos em 2020 foi de 77,6 milhões de unidades, uma queda de 15,4% em relação ao total produzido em 2019.

Segundo a OICA, em 2020 foram produzidos no Brasil 2,01 milhões de veículos, o que o classifica como 9º maior produtor mundial de veículos, atrás da China, Estados Unidos, Japão, Alemanha, Coreia do Sul, Índia, México e Espanha. No mesmo ano o Brasil comercializou 2,06 milhões de veículos, apresentando-se como o 7º maior mercado automobilístico do mundo, atrás da China, Estados Unidos, Japão, Alemanha, Índia e França (OICA, 2021).

Segundo a Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA) em 2018 o setor representou 3% do produto interno bruto (PIB) brasileiro e 18% da indústria de transformação. Em 2019 o faturamento bruto foi de US\$ 59,2 bilhões. Neste mesmo ano foram gerados R\$ 79,1 bilhões em tributos diretos (IPI, PIS/COFINS, ICMS, IPVA). A indústria automobilística emprega, direta ou indiretamente, 1,3 milhão de pessoas. A balança comercial em 2019 foi de US\$ -684 milhões. As exportações representaram US\$ 11,9 bilhões e as importações US\$ 12,6 bilhões (ANFAVEA, 2021).

A história da indústria automobilística brasileira começou no final do século XIX. Segundo a ANFAVEA (2021), o primeiro automóvel desembarcou no Brasil em 1898 (um Peugeot). As primeiras montadoras a se instalarem no país foram as americanas Ford em 1919 e General Motors em 1922. A partir da década de 50 a indústria automobilística brasileira ganhou maior proeminência, com a proibição da importação de autopeças com similar nacional e de veículos completos. Nesta mesma década as empresas alemãs Volkswagen e Mercedes-Benz, e a sueca Scania, começaram suas atividades no Brasil. Nos anos 70 a italiana Fiat (atualmente FCA) se instalou no Brasil e, nas últimas décadas, vieram as japonesas Toyota e Honda, as francesas Renault e PSA, a sul-coreanas Hyundai, e a montadora de veículos premium alemã BMW, entre outras.

De acordo com a (ANFAVEA, 2021) existem no Brasil 26 empresas fabricantes de veículos automotivos e de máquinas agrícolas e rodoviárias. Estas empresas reúnem 61 unidades industriais, divididas em 10 estados e 42 municípios (autoveículos, máquinas

agrícolas e rodoviárias, motores, componentes e outros produtos). A capacidade instalada de produção no Brasil é de 5,05 milhões de veículos automotivos.

O Sindicato Nacional da Indústria de Componentes para Veículos Automotores (SINDIPEÇAS) e a Associação Brasileira da Indústria de Autopeças (ABIPEÇAS) possuem cerca de 484 associados. Estas empresas fornecem componentes para as montadoras brasileiras, para o mercado de reposição no Brasil e exportam para cerca de 180 países (ANFAVEA, 2021).

Henriques e Miguel (2017) afirmam que a modularidade consiste basicamente na divisão de um produto (ou processo) em módulos compostos de várias tarefas, etapas, ou até mesmo atividades de projeto (ou componentes).

Para Daudt e Willcox (2018) a importância da indústria automobilística para a economia pode ser comprovada sob diversos aspectos: entre outras coisas, por ser uma grande geradora de empregos, por apresentar importantes encadeamentos produtivos e por seus investimentos em inovação. Os autores destacam o processo de internacionalização e descentralização da cadeia produtiva da indústria automobilística nos anos 90 e afirmam que o fenômeno da modularidade permitiu a expansão geográfica das atividades e iniciou uma rearticulação produtiva para atender vários mercados. Daudt e Willcox (2018) observam que, como consequência do fenômeno da modularidade, ocorreu uma maior integração da cadeia produtiva, um maior fluxo de investimento entre os países, fortalecimento do mercado regional, e a concentração de grandes empresas fornecedoras de subconjuntos, módulos e sistemas, convencionalmente chamadas de sistemistas (ou camada 1). Os autores citam como exemplo de empresas sistemistas de destaque: Bosch, ZF, Magna, Continental, Denso, Faurecia, entre outras. Para os autores essas empresas, em conjunto com as montadoras, detêm o poder de comando de toda a cadeia produtiva. Tratam-se, em sua maioria, de grandes empresas multinacionais de capital estrangeiro, com grande faturamento mundial, maior capacidade de exportação, práticas modernas de produção e com boa capacidade de engenharia e inovação. Os níveis inferiores do segmento de autopeças (camadas 2 e 3) são compostos de fabricantes de conjuntos, partes e peças. Trata-se de empresas de menor porte, primordialmente de capital nacional, que faturam e empregam menos que os sistemistas. Sendo também mais frágeis tecnicamente, estas empresas não desenvolvem atividades relevantes de engenharia e inovação.

A comissão europeia, por meio do *Economics of Industrial Research and Innovation* (IRI), disponibiliza em seu site dados estatísticos das 2.500 empresas que investem as maiores somas em pesquisa e desenvolvimento (P&D) no mundo. Essas empresas, situadas em 44 países, investiram em 2019 um total de € 823,42 bilhões, o que corresponde a aproximadamente

90% do investimento em P&D financiada por negócios no mundo. Deste total, o setor da indústria automobilística respondeu por € 127,79 bilhões, classificando-se como o terceiro maior setor em P&D, logo atrás das indústrias farmacêuticas e de biotecnologia, e das indústrias de tecnologia em hardware e equipamentos de informática. A Tabela 2 mostra os dez maiores setores industriais em investimento em P&D no mundo em 2019.

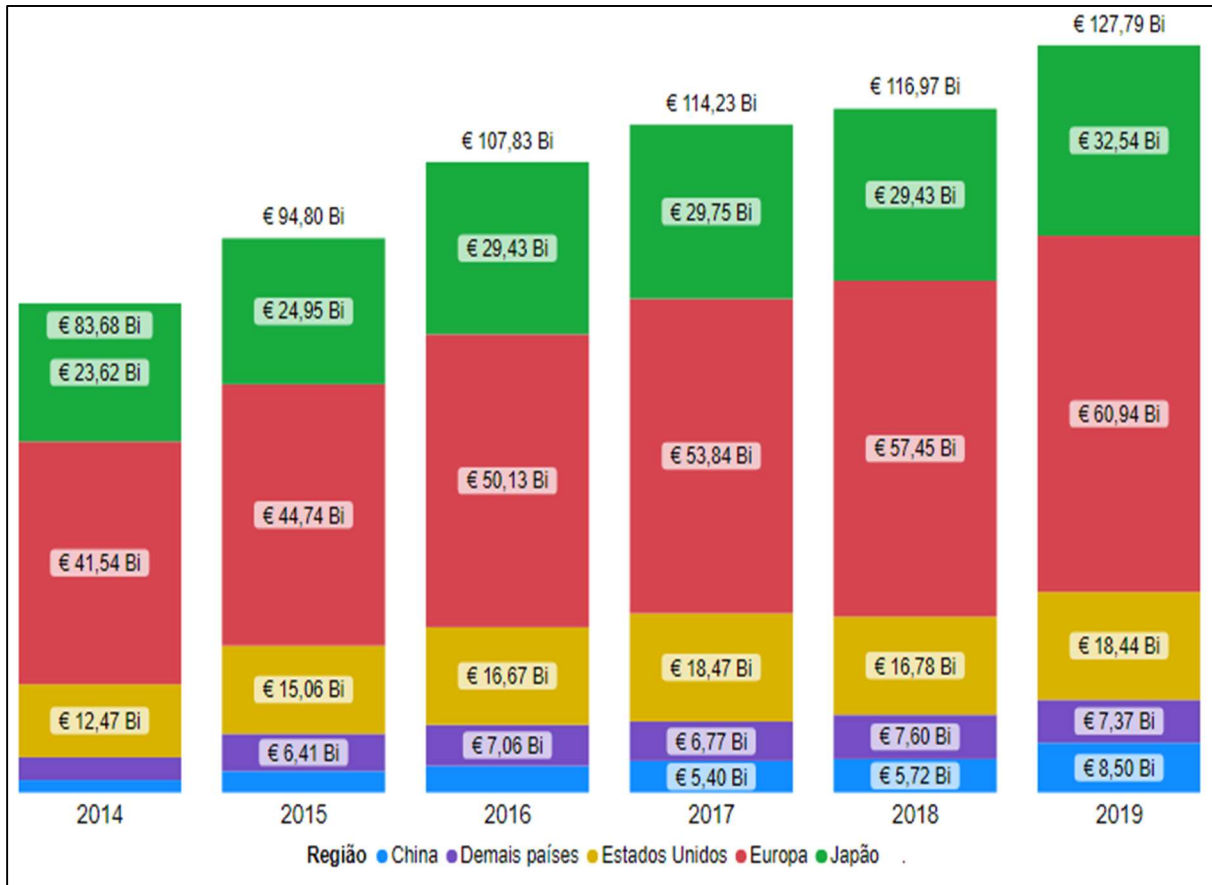
Tabela 2 - Investimentos mundiais em P&D por setor industrial em 2019

Rank	Indústria	Investimentos em P&D 2019 (bilhão)		%
1	Farmacêutica & Biotecnologia	€	153,82	18,68%
2	Tecnologia em Hardware & Equipamentos	€	127,83	15,52%
3	Automóveis & Autopeças	€	127,79	15,52%
4	Software & Serviços de Informática	€	117,71	14,30%
5	Equipamentos Elétricos & Eletrônicos	€	64,16	7,79%
6	Engenharia Industrial	€	29,93	3,63%
7	Químicos	€	22,48	2,73%
8	Indústrias Gerais	€	20,42	2,48%
9	Aeroespacial & Defesa	€	20,20	2,45%
10	Equipamentos & Serviços de Saúde	€	16,56	2,01%
	Outras indústrias	€	122,53	14,88%
Total		€	823,43	100,00%

Fonte: adaptado de IRI (2020)

No período entre 2014 e 2019 os investimentos mundiais em P&D no setor da indústria automobilística cresceram 52,7%. Historicamente as empresas europeias são as que mais investem em P&D. Em 2019 as empresas europeias foram responsáveis por 47,7% dos valores investidos em P&D, seguidas pelas empresas japonesas com 25,5%, as americanas com 14,4% e as chinesas com 6,6%. As empresas localizadas nas demais regiões responderam por 5,8% do total investido em P&D no mundo. Entre as 2.500 empresas listadas pelo IRI como as que mais investiram em 2019 em P&D, 150 pertencem ao setor da indústria automobilística. A Volkswagen, com sede na Alemanha, é a empresa do setor que mais investiu em P&D em 2019, com um total de € 13,64 bilhões, sendo também a quarta maior empresa em investimentos em P&D, independente de setor, no mundo neste ano (IRI, 2020). A Figura 5 mostra os investimentos em P&D no setor da indústria automobilística no período entre 2014 e 2019 por região.

Figura 5 - Investimentos em P&D no setor automotivo por ano e por região



Fonte: adaptado de IRI (2020)

As dez maiores empresas em investimentos em P&D no setor da indústria automobilística responderam por 56,8% do total investido no mundo em 2019. Destas empresas quatro estão sediadas na Alemanha, quatro no Japão e duas nos Estados Unidos, sendo que oito são montadoras de veículos automotores e duas empresas fabricantes de auto peças (IRI, 2020). A Tabela 3 lista as dez maiores empresas em investimentos em P&D no setor da indústria automobilística, o país em que estão sediadas e o total investido em 2019.

Tabela 3 - Dez maiores empresas em investimentos em P&D no setor da indústria automobilística no mundo em 2019

Rank Mundial 2019	Companhia	País	Investimentos em P&D 2019 (bilhão)	%
4	VOLKSWAGEN	Alemanha	€ 13,64	10,7%
10	DAIMLER	Alemanha	€ 9,04	7,1%
13	TOYOTA MOTOR	Japão	€ 8,26	6,5%
15	FORD MOTOR	Estados Unidos	€ 7,16	5,6%
16	BMW	Alemanha	€ 6,89	5,4%

Tabela 3 - Dez maiores empresas em investimentos em P&D no setor da indústria automobilística no mundo em 2019 (cont.)

Rank Mundial 2019	Companhia	País	Investimentos em P&D 2019 (bilhão)		%
18	GENERAL MOTORS	Estados Unidos	€	6,81	5,3%
19	HONDA MOTOR	Japão	€	6,58	5,1%
20	ROBERT BOSCH	Alemanha	€	6,19	4,8%
35	NISSAN MOTOR	Japão	€	4,12	3,2%
38	DENSO	Japão	€	3,92	3,1%
Demais companhias			€	55,17	43,2%
Total			€	127,79	100,0%

Fonte: adaptado de IRI (2020)

Tecnologias automotivas são todas as inovações desenvolvidas pela indústria automobilística e que auxiliam o motorista a conduzir o veículo automotor de maneira segura e eficiente. A seguir descreve-se as tecnologias automotivas utilizadas neste trabalho para basear a coleta dos dados de pesquisa.

1.6.1 Veículos Autônomos

Para Talebpour e Mahmassani (2016) a definição de Veículos Autônomos (AV, do inglês *Autonomous Vehicles*) inclui cinco níveis diferentes de automação: a) o primeiro nível, sem automação, refere-se aos veículos regulares, onde os motoristas são os únicos tomadores de decisão; b) o segundo nível, refere-se a veículos com pelo menos uma função de controle, como por exemplo o controle eletrônico de estabilidade; c) o terceiro nível refere-se a veículos com pelo menos duas funções de controle de veículo, sendo que essas funções de controle trabalham umas com as outras ao mesmo tempo e fornecem mais assistência ao motorista; d) no quarto nível de automação, chamado de automação autônoma limitada, os veículos podem controlar funções essenciais para a segurança do motorista; e) o quinto e último nível de automação, chamado de automação totalmente autônoma, refere-se a veículos que controlam todas as funções de direção em qualquer condição de clima, estrada e tráfego.

1.6.2 Sistemas de Assistência Avançada ao Motorista

Gietelink *et al.* (2006) definem Sistemas de Assistência Avançada ao Motorista (ADAS, do inglês *Advanced Driver Assistance Systems*) como sistemas interligados de controle veicular que usam sensores de ambiente (radares, laser e câmeras) para aumentar o conforto e a segurança ao dirigir um veículo automotor, auxiliando o motorista a reconhecer e reagir a situações potencialmente perigosas. Como um ADAS pode até intervir autonomamente, um veículo equipado com ADAS é popularmente referido como um "veículo inteligente".

Korssen *et al.* (2017) analisam que os ADAS, como por exemplo *adaptive cruiser controllers* (ACC), *lane-keeping assistance systems*, e *collision-avoidance systems*, entre outros, estão cada vez mais presentes nos veículos modernos. Para os autores, no início, os ADAS foram desenvolvidos com o objetivo de aumentar o conforto do motorista, porém atualmente o foco de desenvolvimento dos ADAS é: a) aumentar a segurança dos motoristas no tráfego; b) rendimento do sistema motor/chassi, para reduzir o consumo de combustível e a emissão de gases poluentes; e c) suporte às crescentes tecnologias para desenvolvimento de veículos autônomos. Os diversos sistemas de assistência possibilitados pelos ADAS têm uma forte interação entre si e os demais sistemas veiculares, já não podem mais serem considerados como sistemas independentes. Como consequência os ADAS se tornarão cada vez mais complexos e essenciais para a segurança.

1.6.3 Tecnologias de Segurança Veicular

As Tecnologias de Segurança Veicular (VS, do inglês *Vehicular Safety*) englobam o desenvolvimento de tecnologias, ferramentas e medidas utilizadas na produção e manutenção de veículos, objetivando gerar mais segurança para os motoristas, passageiros e pedestres. A importância da VS fica evidenciada no relatório *Global Status Report on Road Safety 2018*, editado pela Organização Mundial de Saúde (WHO, do inglês *World Health Organization*), onde destaca-se que o número de mortes anuais no trânsito atingiu 1,35 milhão de pessoas no mundo, e que as lesões causadas por acidentes no trânsito são atualmente a principal causa de morte de pessoas na idade entre 5 e 29 anos. As tecnologias de VS podem ser divididas em: a) segurança ativa, cujo propósito é o de evitar acidentes e reduzir as chances de impacto, como por exemplo o freio *Anti-Block System* (ABS) e a distribuição eletrônica de freios; b) segurança passiva, que objetiva a proteção dos ocupantes na ocorrência de acidentes, como por exemplo

os cintos de segurança e os airbags; c) assistência à condução (ADAS), que auxiliam o motorista na condução do veículo com maior segurança, como por exemplo a frenagem automática de emergência; d) assistência à segurança, composta de dispositivos como uma forma de alertar os ocupantes para que utilizem os equipamentos de segurança do automóvel, como o alerta de cinto de segurança; e e) proteção do pedestre, que visa a proteção externa de pedestres em caso de atropelamento, como por exemplo o conjunto para-choque dianteiro, para-lamas e tampa dianteira, projetados especialmente para este fim, e airbag para pedestres.

1.6.4 Informática e Entretenimento Automotivo

Segundo Meixner *et al.* (2017), a história do desenvolvimento da *Human Machine Interface* (HMI) automotiva mostra que o desenvolvimento de novas funcionalidades interativas no carro, como as tecnologias de Informática e Entretenimento Automotivo (AIS, do inglês *Automotive Infotainment Systems*), sempre foram influenciadas pelas novas tecnologias que eram utilizadas pelas pessoas de forma habitual nas suas atividades corriqueiras em suas casas ou em seu trabalho. Como exemplo destas tecnologias citam-se o primeiro rádio automotivo, lançado por volta de 1922, e o primeiro telefone automotivo, de 1952. Para os autores, atualmente a indústria automobilística desenvolve e integra aos veículos funcionalidades inovadoras de AIS, com o objetivo de se manter competitiva no mercado. Como tendências atuais de AIS automotivo Meixner *et al.* (2017) citam aplicativos baseados na internet ou aplicativos de rede social, ao mesmo tempo que citam novos AIS extensíveis, híbridos, adaptáveis ou mesmo personalizados que estão surgindo como tendências futuras.

1.6.5 Cidades Inteligentes (*Smart Cities*)

Bakici; Almirall e Wareham (2013) definem o conceito de *Smart City* como uma cidade que conecta pessoas, informações e objetos por meio de tecnologias avançadas, com o objetivo de criar uma cidade sustentável, verde, com um comércio competitivo e inovador, e com incremento da qualidade de vida da sua população.

Albino; Berardi e Dangelico (2015) afirmam que existem muitas definições do que é uma *Smart City*, sendo que uma gama de variantes conceituais é frequentemente obtida substituindo “*smart*” por adjetivos alternativos, como por exemplo, “*intelligent*” ou “*digital*”.

Para o autor o rótulo “*Smart City*” é um conceito difuso e nem sempre utilizado de forma consistente.

Marsal-Llacuna; Colomer-Llinàs e Meléndez-Frigola (2015) afirmam que as cidades que desejam se tornar *Smart Cities* devem se preparar para disponibilizar muitas informações em tempo real, por meio de dados coletados de sensores, permitindo que os responsáveis pela cidade possam tomar decisões eficientes e centradas nos cidadãos. Para os autores as iniciativas de *Smart Cities* tentam melhorar o desempenho urbano usando dados, informações e tecnologias da informação (TI), para fornecer serviços mais eficientes aos cidadãos, e monitorar e otimizar a infraestrutura existente, com o objetivo de aumentar a colaboração entre os cidadãos e os diferentes atores econômicos, para estimular modelos de negócios inovadores nos setores público e privado.

1.6.6 Mobilidade Inteligente (*Smart Mobility*)

Dentro do conceito de *Smart City*, Albino; Berardi e Dangelico (2015) descrevem o conceito de *Smart Mobility* como sendo o uso de modernas tecnologias de informação e comunicação no transporte público e privado com o objetivo de melhorar o tráfego urbano.

Faria *et al.* (2017) dividem os principais conceitos de *Smart Mobility* em: a) *Driving Safety*, que consiste em fornecer tecnologia para segurança e proteção do motoristas, permitindo que os carros interajam com outros veículos e com a infraestrutura ao seu redor; b) *Smart Lightning Systems*, que usa iluminação urbana com LEDs e sensores, para fornecer melhor iluminação, reduzir o consumo de energia, reduzir congestionamento e melhorar o fluxo do tráfego; c) *Sharing and Urban Mobility*, que são sistemas de transporte que incluem compartilhamento de carros, de bicicletas e o incentivo às caronas; d) *Electric Mobility*, que objetiva a mobilidade sustentável, eficiente, com economia de recursos e favorável ao clima e ao meio ambiente; e) *Green Mobility*, para minimizar o impacto ambiental causado pelo setor de transporte; e f) *Smart Payment Systems*, composto de tecnologias que agilizam as formas de pagamento de parquímetros e estacionamento. A Figura 6 mostra a divisão das áreas de *Smart Mobility* dentro do conceito de *Smart City*.

Figura 6 - Divisão das áreas de *Smart Mobility* dentro do conceito de *Smart City*.



Fonte: adaptado de Faria *et al.* (2017)

2 METODOLOGIA

Esse trabalho de pesquisa, a partir dos objetivos detalhados, pode ser classificado como sendo uma pesquisa empírica, de natureza aplicada com objetivo exploratório.

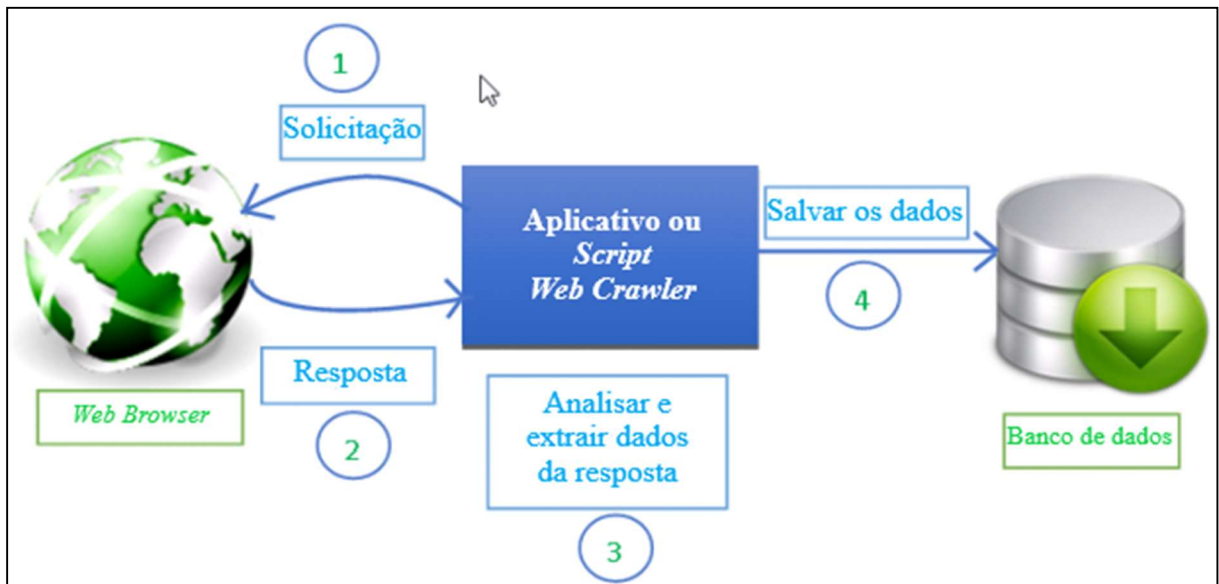
Para atender o objetivo geral da pesquisa, criou-se um programa em *Python* versão 3.8.3 para a coleta de dados, por meio de *web crawler*, de documentos de patentes dos registros do *Derwent Innovations Index* da base de dados *Web of Science* (WoS).

O *Python* é uma linguagem de programação moderna e versátil, tendo o seu foco voltado para a matemática, com ênfase em estatística, o que a torna uma excelente linguagem de programação para os cientistas de dados. Suas enormes bibliotecas são usadas para manipulação de dados e são muito fáceis de aprender, mesmo para um analista de dados iniciante.

Segundo Ren e Ren (2018) *web crawler* é um aplicativo (ou *script*) que extrai automaticamente conteúdo específico em uma página da *Web* de acordo com uma determinada

regra de correspondência. A aplicação mais típica de um aplicativo ou script *web crawler* são os mecanismos de busca, capturando dados da internet e baixando páginas da *Web*. O funcionamento de um *web crawler* pode ser dividido nas seguintes etapas: 1) o aplicativo *web crawler* começa a rastrear os dados a partir de um ou mais conjuntos de localizadores uniformes de recursos da *Internet* (URL, do inglês *Uniform Resource Locator*); 2) os dados solicitados retornam para o aplicativo *web crawler*; 3) o aplicativo *web crawler* analisa e baixa os conteúdos destas páginas URL, extrai as informações solicitadas, e se necessário, rastreia outra URL, colocando-a na fila de solicitação, repetindo o processo até rastrear todas as páginas; 4) os dados com o conteúdo extraído é baixado para um banco de dados. A Figura 7 mostra a estrutura geral de uma *web crawler*.

Figura 7 - Estrutura geral de um *web crawler*



Fonte: adaptado de Jabeen (2019)

Utilizou-se na coleta de dados as ferramentas (bibliotecas) específicas da linguagem de programação *Python* para criação do aplicativo *web crawler*: a) *Scrapy*, versão 2.4.1; e b) *Selenium*, versão 3.17.0.

Scrapy é um *framework open source* de rastreamento e *scraping* da web de alto nível, usada para rastrear sites e extrair dados estruturados de suas páginas, podendo ser usado para uma ampla gama de finalidades, desde mineração de dados até monitoramento e testes automatizados (SCRAPY.ORG, 2020)

Scrapy é um *framework open source* de rastreamento e *scraping* da web de alto nível, usada para rastrear sites e extrair dados estruturados de suas páginas, podendo ser usado para

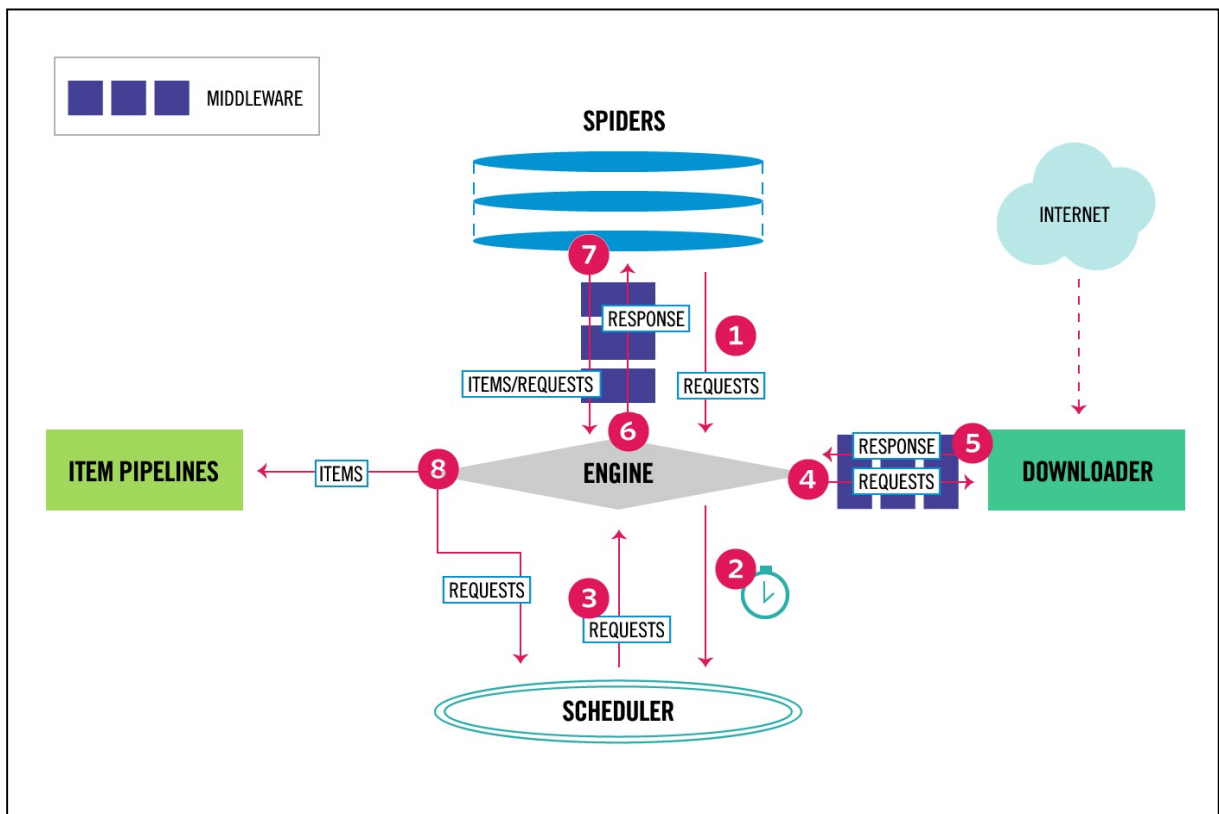
uma ampla gama de finalidades, desde mineração de dados até monitoramento e testes automatizados (SCRAPY.ORG, 2020). O fluxo de dados no sistema é controlado pelo *Scrapy Engine* e pode ser descrito em etapas:

1. O *Scrapy Engine* (ou *bot* de busca) recebe as *Requests* (solicitações) iniciais para rastrear os dados na *web*. O *Scrapy Engine* é responsável por controlar o fluxo de dados entre todos os componentes do sistema e por disparar eventos quando determinadas ações ocorrem.
2. O *Scrapy Engine* agenda as *Requests* no *Scheduler* (agendador) e solicita as próximas *Requests* a serem rastreadas. A função do *Scheduler* é receber as *Requests* do *Scrapy Engine* e enfileirá-las para alimentar posteriormente este mesmo *Scrapy Engine* quando solicitado.
3. O *Scheduler* retorna as próximas *Requests* ao *Scrapy Engine*.
4. O *Scrapy Engine* envia as *Requests* para o *Downloader*, passando pelos *Middlewares*. O *Downloader* é responsável por buscar páginas da *web* e alimentá-las para o *Scrapy Engine* que, por sua vez, as alimenta para os *Spiders*.
5. Assim que a página termina o *download*, o *Downloader* gera uma *Response* (resposta) e a envia para o *Scrapy Engine*, passando pelos *Middlewares*.
6. O *Scrapy Engine* recebe a *Response* do *Downloader* e a envia para o *Spider*, passando pelos *Middlewares*. Os *Middlewares* do *Downloader* são processos específicos que ficam entre o *Scrapy Engine* e o *Downloader*, cuja função é processar as *Requests* quando passam do *Scrapy Engine* para o *Downloader* e as *Responses* que passam do *Downloader* para o *Scrapy Engine*.
7. O *Spider* processa a *Response* e retorna os itens baixados e as novas *Requests* para o *Scrapy Engine*, passando pelo *Middleware*. *Spiders* são classes personalizadas escritas por usuários do *Scrapy* para analisar as respostas e extrair itens delas ou solicitações adicionais a seguir. Os *Middlewares* do *Spider* são processos específicos que ficam entre o *Scrapy Engine* e os *Spiders*, sendo responsáveis por processar as entradas (*Responses*) e as saídas (*Items* e *Requests*) do *Spider*.

8. O *Scrapy Engine* envia os itens processados para as *Pipelines*, envia as *Requests* processadas para o *Scheduler*, e pede as próximas *Requests* para rastreamento. A função das *Pipelines* é processar os itens uma vez que eles tenham sido extraídos pelos *Spiders*.
9. O processo se repete (desde a etapa 1) até que não haja mais *Requests* do *Scheduler*.

A Figura 8 mostra uma visão geral da arquitetura do *Scrapy* com seus componentes e um esboço do fluxo de dados que ocorre dentro do sistema.

Figura 8 - Fluxo de dados e componentes do Scrapy



Fonte: adaptado de scrapy.org (2020)

Selenium é uma ferramenta utilizada para automatização de testes de sistemas criados para a *web*, que permite ao usuário reproduzir rapidamente comandos e ações no ambiente real da aplicação, em função da sua integração direta com o navegador. Gheorghe; Mihai e Dârdală (2018) observam que, no caso de páginas dinâmicas, que incluem questões como redirecionamentos HTTP, cookies ou mudanças dinâmicas no conteúdo (como as páginas *web* contidas na base de dados *Web of Science*), automatizar um navegador é uma abordagem preferencial. Os autores definem o *Selenium* como um software de código aberto, compatível com os sistemas operacionais *Windows*, *Linux* e *MacOS* e que, por meio de seu componente

WebDriver, é capaz de automatizar todos os principais navegadores da web, iniciando uma instância, enviando comandos e recuperando o resultado. O uso do *Selenium*, em conjunto com a linguagem de programação *Python*, se justifica, pois, permite automatizar, de forma simples, o uso do navegador *web* pelo programa em *Python* para informar à página *web* dados como usuário, senha, sintaxe de busca, entre outras ações

2.1 CÓDIGO DE PROGRAMAÇÃO *WEB CRAWLER WOS.PY*

O primeiro passo para o desenvolvimento do *web crawler* para extração dos dados dos documentos de patentes dos registros do *Derwent Innovations Index* da base de dados *WoS* foi a criação da estrutura do *Scrapy*. Para tanto, com o *Python* versão 3.8.3 e o *Scrapy* versão 2.4.1 instalados, digitou-se a seguinte linha de comando no aplicativo *Windows PowerShell*:

```
PS C:\_d> scrapy startproject WOS
```

Utiliza-se o comando *startproject* para gerar um novo projeto *web crawler*. A Figura 9 mostra a estrutura de diretórios e arquivos gerados pelo *Scrapy* após o comando *startproject*.

Figura 9 - Estrutura de diretórios e arquivos gerados pelo *Scrapy* após o comando *startproject*

```

Windows PowerShell
PS C:\_d> scrapy startproject WOS
New Scrapy project 'WOS', using template directory 'c:\programdata\anaconda3\lib\
site-packages\scrapy\templates\project', created in:
  C:\_d\WOS

You can start your first spider with:
  cd WOS
  scrapy genspider example example.com
PS C:\_d> tree /F
Listagem de caminhos de pasta para o volume Windows-SSD
O número de série do volume é 6EE8-E4DC
C:.
├── WOS
│   ├── scrapy.cfg
│   └── WOS
│       ├── items.py
│       ├── middlewares.py
│       ├── pipelines.py
│       ├── settings.py
│       ├── __init__.py
│       └── spiders
│           └── __init__.py

```

Fonte: o autor

Na sequência digitou-se as seguintes linhas de comando no aplicativo *Windows PowerShell*:

```
PS C:\_d> cd WOS
PS C:\_d\WOS> scrapy genspider wos apps.webofknowledge.com
```

Utiliza-se o comando *genspider* para gerar um novo Spider usando *templates* predefinidos no Scrapy. O comando *scrapy genspider wos apps.webofknowledge.com* gerou, no diretório *spiders*, o arquivo *wos.py*, tendo com Request inicial a URL *apps.webofknowledge.com*. A Figura 10 mostra a estrutura de diretórios e arquivos gerados pelo Scrapy após o comando *genspider*.

Figura 10 - Estrutura de diretórios e arquivos gerados pelo Scrapy após o comando *genspider*

```

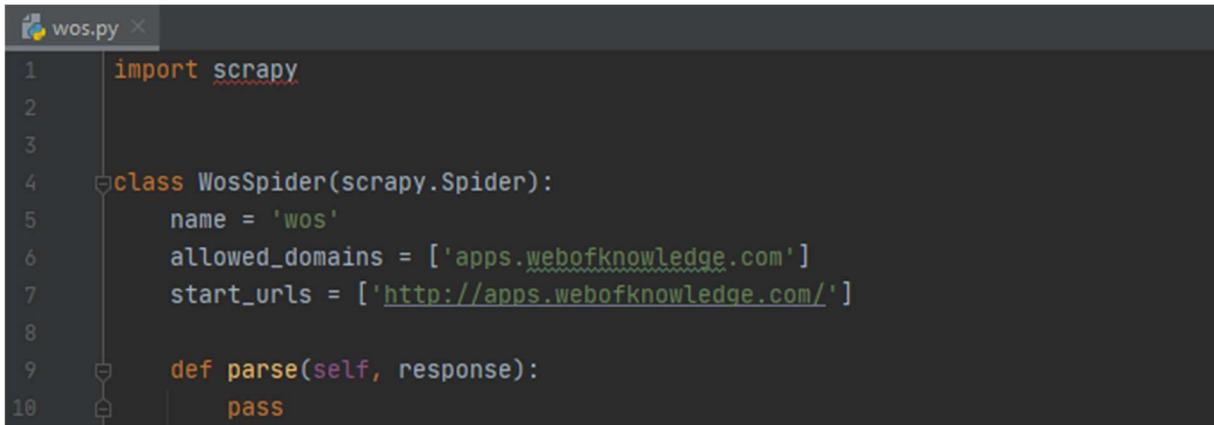
Selecionar Windows PowerShell
PS C:\_d> cd WOS
PS C:\_d\WOS> scrapy genspider wos apps.webofknowledge.com
Created spider 'wos' using template 'basic' in module:
  WOS.spiders.wos
PS C:\_d\WOS> tree /F
Listagem de caminhos de pasta para o volume Windows-SSD
O número de série do volume é 6EE8-E4DC
C:..
| scrapy.cfg
|
|---WOS
|   | items.py
|   | middlewares.py
|   | pipelines.py
|   | settings.py
|   | __init__.py
|   |
|   |---spiders
|   |   | wos.py
|   |   | __init__.py
|   |   |
|   |   |---__pycache__
|   |   |   | __init__.cpython-38.pyc
|   |   |
|   |---__pycache__
|   |   | settings.cpython-38.pyc
|   |   | __init__.cpython-38.pyc

```

Fonte: o autor

A Figura 11 mostra o programa em *Python*, gerado automaticamente pelo Scrapy, após o comando *genspider*.

Figura 11 - Programa *wos.py* gerado automaticamente pelo *Scrapy* após o comando *genspider*



```

1  import scrapy
2
3
4  class WosSpider(scrapy.Spider):
5      name = 'wos'
6      allowed_domains = ['apps.webofknowledge.com']
7      start_urls = ['http://apps.webofknowledge.com/']
8
9      def parse(self, response):
10         pass

```

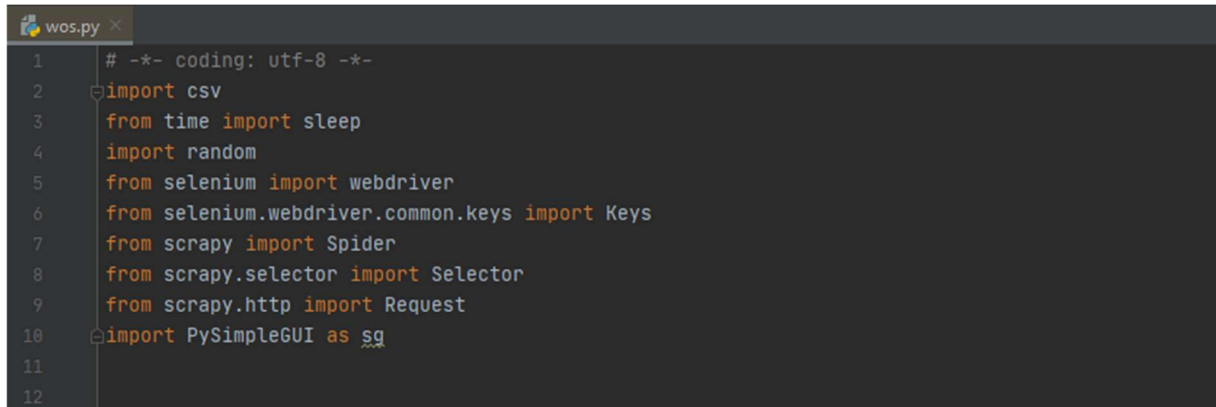
Fonte: o autor

A próxima etapa a ser executada para a criação do sistema *web crawler* é a edição do programa *wos.py*, com a inclusão de códigos para interação e obtenção dos dados das páginas *web* da base de dados *Web of Science*. Na edição do programa *wos.py* utiliza-se o editor de texto *Sublime Text Version 3.2.2, Build 3211*. O programa *web crawler wos.py* final (*Spider*), construído pelo autor, pode ser dividido em: a) abertura das bibliotecas Python; b) classes com a programação das telas para coletar as informações de usuário, senha, sintaxe de busca e tipo de extração de dados; c) tela com a mensagem informativa do sistema d) criação dos arquivos no formato CSV (*Comma Separated Values*); e) interação com o navegador *Google Chrome*, versão 86.0.4240.198; f) processamento da *web crawler* (*Scrape Engine*).

2.1.1 Abertura das bibliotecas Python

As bibliotecas do Python são uma coleção de módulos de *script* acessíveis a um programa Python, cuja função é simplificar o processo de programação e remover a necessidade de reescrever os comandos mais usados. Eles podem ser usados chamando-os no início de um *script*. No início do programa *wos.py* ocorre a abertura das bibliotecas necessárias para a programação da *web crawler*: a) *CSV*, que implementa classes para ler e gravar dados tabulares no formato CSV; b) *Time*, que fornece várias funções relacionadas ao tempo; c) *Selenium*; d) *Scrapy*; e e) *PySimpleGUI*, que é um pacote que permite a criação de interfaces gráficas. A Figura 12 mostra as linhas de código do programa *wos.py* com a importação das bibliotecas utilizadas na criação da *web crawler*.

Figura 12 - Linhas de código do programa *wos.py* com a importação das bibliotecas



```

1  # -*- coding: utf-8 -*-
2  import csv
3  from time import sleep
4  import random
5  from selenium import webdriver
6  from selenium.webdriver.common.keys import Keys
7  from scrapy import Spider
8  from scrapy.selector import Selector
9  from scrapy.http import Request
10 import PySimpleGUI as sg
11
12

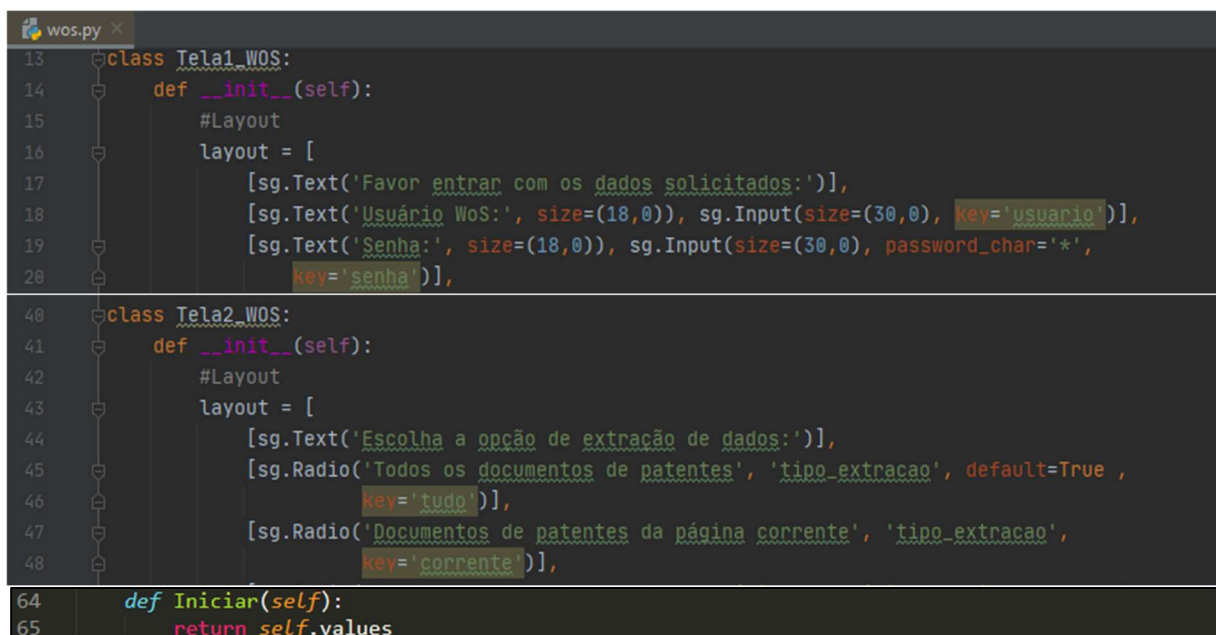
```

Fonte: o autor

2.1.2 Programação das Telas de Interação com o Usuário

Utilizou-se a biblioteca *PySimpleGUI* para criar duas interfaces gráficas com o objetivo de coletar informações necessárias ao funcionamento do sistema, dividida em duas classes: a) *class Tela1_WOS*, cuja função é coletar informações de *Sign In* (usuário e senha da base de dados *WoS*) e sintaxe de busca que será usada na plataforma *Derwent*; e b) *class Tela2_WOS*, cuja função é definir o tipo de extração de dados que será utilizado. A Figura 13 mostra parcialmente as linhas de código do programa *wos.py*, contemplando as classes *class Tela1_WOS* e *class Tela2_WOS*.

Figura 13 - Linhas de código do programa *wos.py* (*class Tela1_WOS* e *class Tela2_WOS*)



```

13 class Tela1_WOS:
14     def __init__(self):
15         #Layout
16         layout = [
17             [sg.Text('Favor entrar com os dados solicitados:')],
18             [sg.Text('Usuário WoS:', size=(18,0)), sg.Input(size=(30,0), key='usuario')],
19             [sg.Text('Senha:', size=(18,0)), sg.Input(size=(30,0), password_char='*',
20             key='senha')],
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40 class Tela2_WOS:
41     def __init__(self):
42         #Layout
43         layout = [
44             [sg.Text('Escolha a opção de extração de dados:')],
45             [sg.Radio('Todos os documentos de patentes', 'tipo-extracao', default=True,
46             key='tudo')],
47             [sg.Radio('Documentos de patentes da página corrente', 'tipo-extracao',
48             key='corrente')],
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64     def Iniciar(self):
65         return self.values

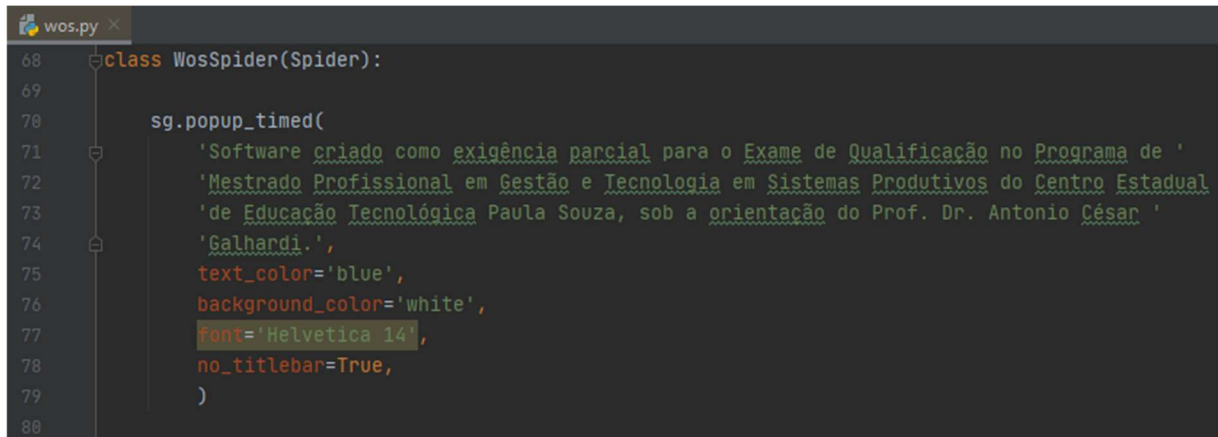
```

Fonte: o autor

2.1.3 Tela de Mensagem Informativa do Programa

No início da classe *class WosSpider*, utilizou-se a biblioteca *PySimpleGUI* para enviar ao usuário uma mensagem informando o objetivo de criação do programa. A Figura 14 mostra as linhas de código do programa *wos.py*, com as etapas de criação da tela informativa.

Figura 14 - Linhas de código do programa *wos.py*, com as etapas de criação da tela informativa do sistema

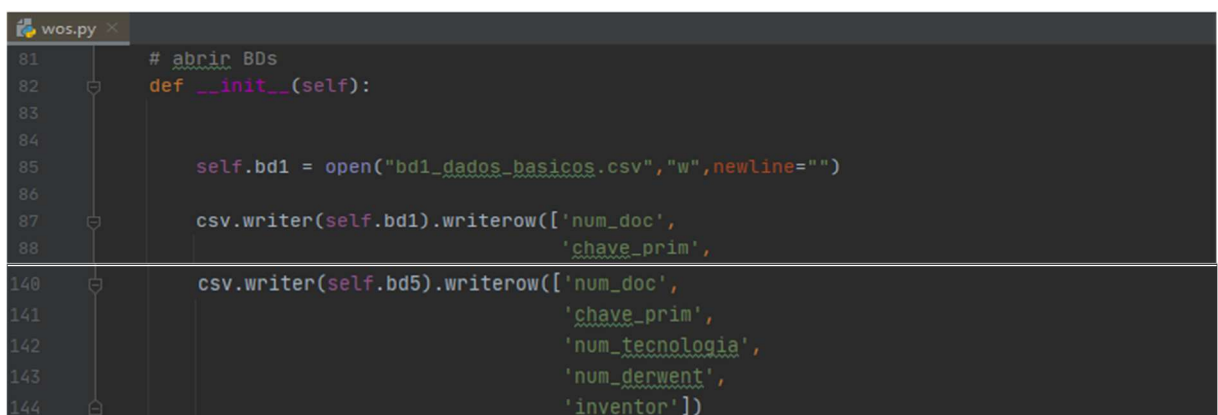


Fonte: o autor

2.1.4 Criação dos Arquivos no Formato CSV

Nesta etapa do programa, com o auxílio da biblioteca *CSV* nativa do Python, foram criados cinco arquivos do tipo *CSV*: a) *bd1_dados_basicos.csv*; b) *bd2_dados_ipc.csv*; c) *bd3_depositantes.csv*; d) *bd4_paises.csv*; e) *bd5_inventores.csv*. A Figura 15 mostra parcialmente as linhas de código do programa *wos.py*, com as etapas de criação dos arquivos tipo *CSV*.

Figura 15 - Linhas de código do programa *wos.py* - etapas de criação dos arquivos tipo *CSV*

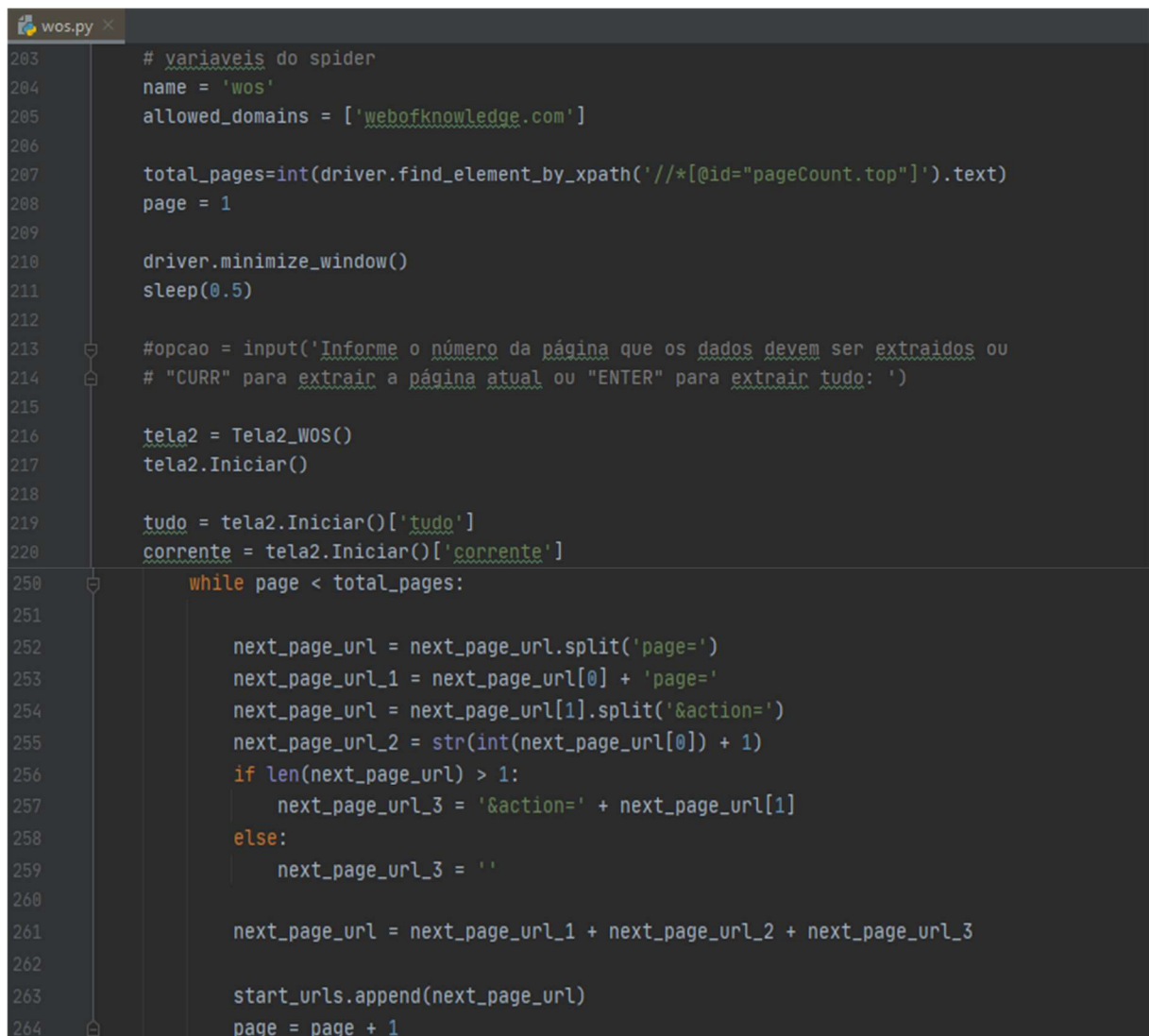


Fonte: o autor

2.1.6 Obtenção das URLs Iniciais (Requests)

O primeiro passo para a obtenção das *URLs* iniciais da *Spider* é coletar junto ao usuário, por meio de uma interface gráfica, o tipo de extração que se deseja executar: a) todos os documentos de patentes encontrados; b) os documentos de patentes de uma determinada página; ou c) os documentos de patentes da página corrente do navegador. Estas opções foram introduzidas no programa visando: a) facilitar a atualização dos dados coletados; ou b) no caso de falha na obtenção dos dados de alguma página *web*, visto que a base de dados *WoS* trabalha com páginas dinâmicas, acarretando eventualmente em erro no carregamento pelo navegador. Na sequência o programa *web crawler* processa a *URLs* que serão acessadas pela *Spider* e as armazena no *Scheduler*. A Figura 17 mostra parcialmente as linhas de código do programa *wos.py*, com as etapas de obtenção das *URLs* iniciais.

Figura 17 - Linhas de código do programa *wos.py* - etapas de obtenção das *URLs* iniciais



```

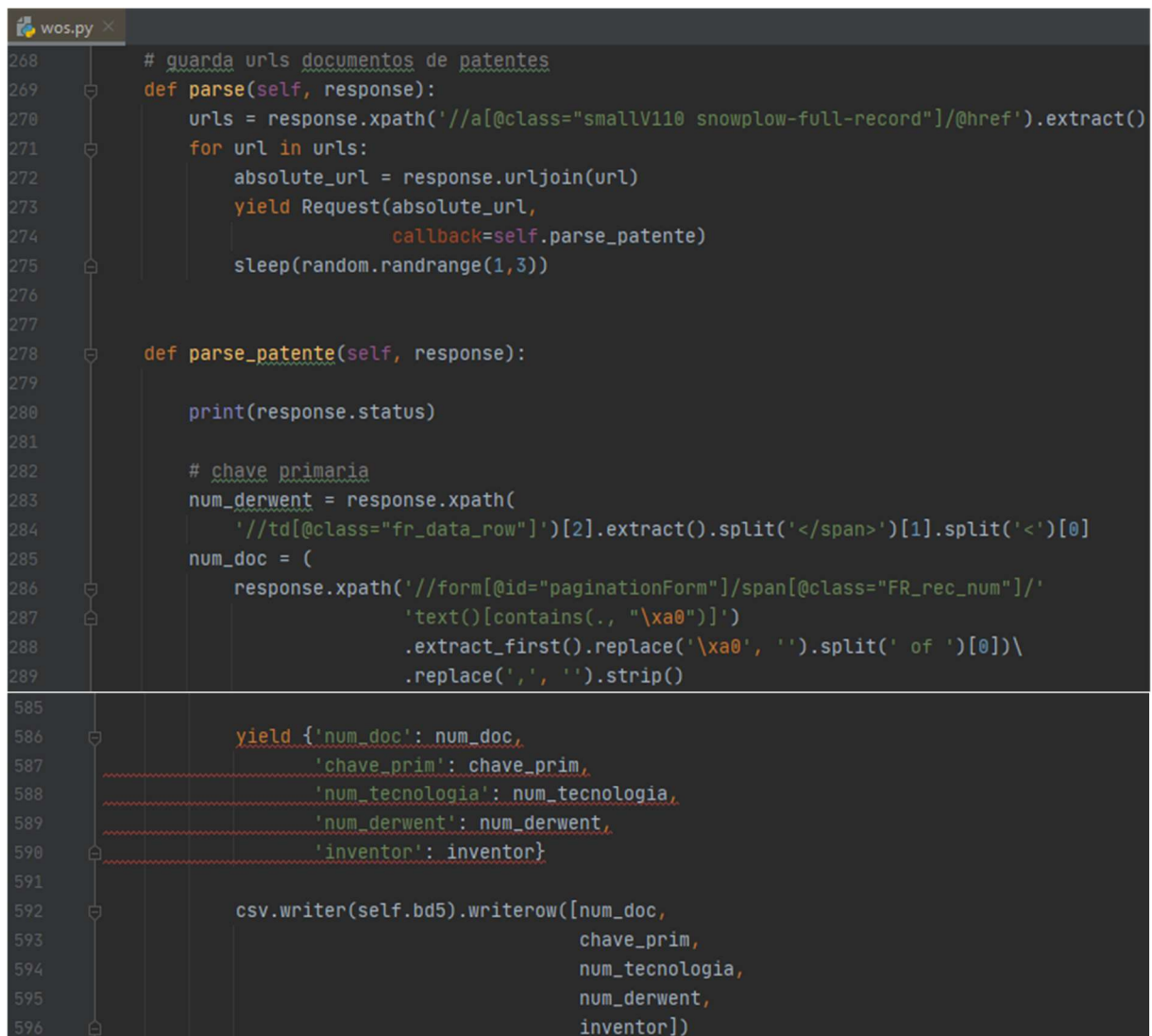
203 # variáveis do spider
204 name = 'wos'
205 allowed_domains = ['webofknowledge.com']
206
207 total_pages=int(driver.find_element_by_xpath('//*[@id="pageCount.top"]').text)
208 page = 1
209
210 driver.minimize_window()
211 sleep(0.5)
212
213 #opcao = input('Informe o número da página que os dados devem ser extraídos ou
214 # "CURR" para extrair a página atual ou "ENTER" para extrair tudo: ')
215
216 tela2 = Tela2_WOS()
217 tela2.Iniciar()
218
219 tudo = tela2.Iniciar()['tudo']
220 corrente = tela2.Iniciar()['corrente']
221
222 while page < total_pages:
223
224     next_page_url = next_page_url.split('page=')
225     next_page_url_1 = next_page_url[0] + 'page='
226     next_page_url = next_page_url[1].split('&action=')
227     next_page_url_2 = str(int(next_page_url[0]) + 1)
228     if len(next_page_url) > 1:
229         next_page_url_3 = '&action=' + next_page_url[1]
230     else:
231         next_page_url_3 = ''
232
233     next_page_url = next_page_url_1 + next_page_url_2 + next_page_url_3
234
235     start_urls.append(next_page_url)
236     page = page + 1
  
```

Fonte: o autor

2.1.7 Processamento do Web Crawler (Scrapy Engine)

Na parte final do programa *web crawler wos.py*, ocorre o funcionamento do *Scrapy Engine*, isto é, o processamento das *Requests* (requisições), das *Responses* (respostas), a obtenção dos dados dos documentos de patente por meio da seleção de elementos do código *HTML*, e a gravação destes dados em arquivos *CSV*. Cada *URL* inicial obtida na etapa anterior, equivalente a uma página resultante da busca na plataforma *Derwent*, possui 50 novas *URLs* (ou *links*) que precisam ser acessados. Cada uma destas *URLs* equivale a um documento de patente. O *Scrapy Engine* abre cada uma destas *URLs*, baixa os dados de documentos de patentes, e grava estes dados em arquivos *CSV*, mantendo este ciclo até acessar a última *URL* de cada *URL* inicial. A Figura 18 mostra parcialmente as linhas de código do programa *wos.py* relativas ao *Scrapy Engine*.

Figura 18 - Linhas de código do programa *wos.py* relativas ao *Scrapy Engine*



```

268 # guarda urls documentos de patentes
269 def parse(self, response):
270     urls = response.xpath('//a[@class="smallV110 snowplow-full-record"]/@href').extract()
271     for url in urls:
272         absolute_url = response.urljoin(url)
273         yield Request(absolute_url,
274                     callback=self.parse_patente)
275         sleep(random.randrange(1,3))
276
277
278 def parse_patente(self, response):
279
280     print(response.status)
281
282     # chave primaria
283     num_derwent = response.xpath(
284         '//td[@class="fr_data_row"]')[2].extract().split('</span>')[1].split('<>')[0]
285     num_doc = (
286         response.xpath('//form[@id="paginationForm"]/span[@class="FR_rec_num"]/'
287             'text()[contains(., "\xa0")]')
288             .extract_first().replace('\xa0', '').split(' of ')[0])\
289             .replace(',', '').strip()
290
291
292     yield {'num_doc': num_doc,
293           'chave_prim': chave_prim,
294           'num_tecnologia': num_tecnologia,
295           'num_derwent': num_derwent,
296           'inventor': inventor}
297
298
299     csv.writer(self.bd5).writerow([num_doc,
300                                   chave_prim,
301                                   num_tecnologia,
302                                   num_derwent,
303                                   inventor])

```

Fonte: o autor

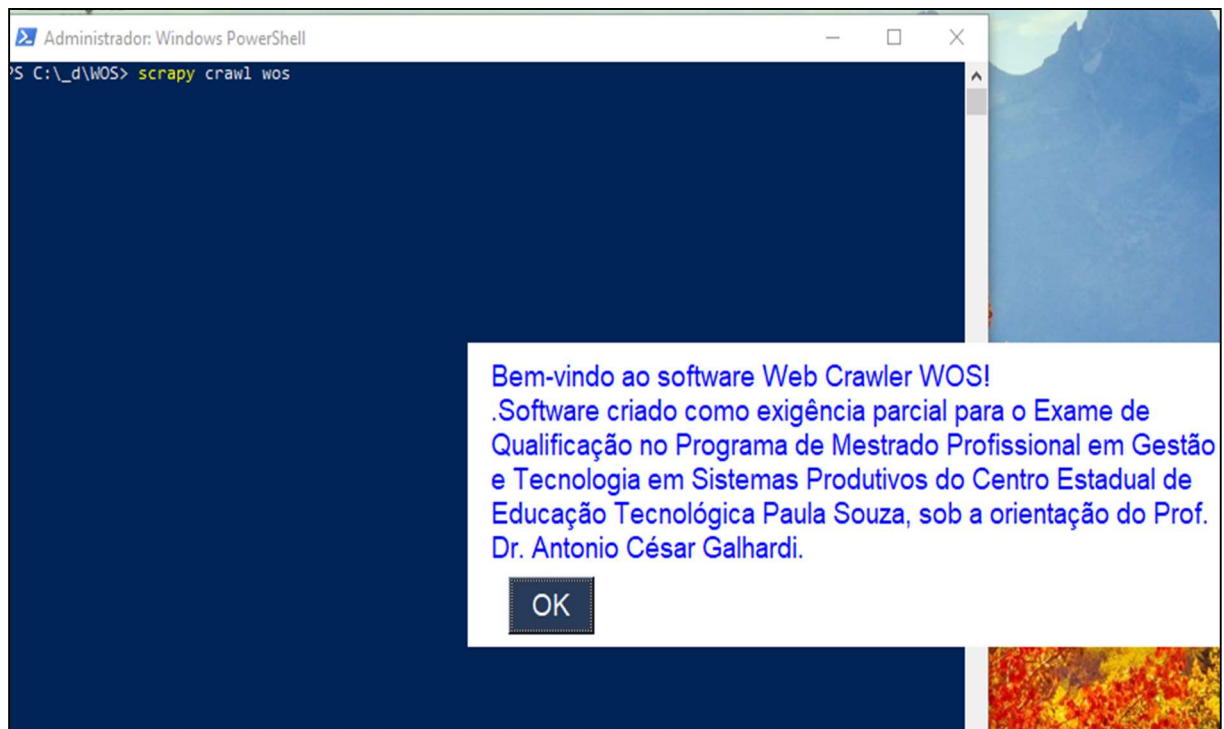
2.2 EXECUÇÃO DO *WEB CRAWLER WOS.PY*

Após a montagem da arquitetura da *web crawler* e a construção do código em *Python* pode-se executar o *Spider*. Para iniciar o sistema *web crawler* digitou-se a seguinte linha de comando no aplicativo *Windows PowerShell*:

```
PS C:\_d\WOS> scrapy crawl wos
```

Ao iniciar o sistema abre a tela de abertura do sistema. A Figura 19 mostra a Tela 1 do *Spider wos*.

Figura 19 - Tela 1 do *Spider wos*



Fonte: o autor

Na sequência o sistema abre a tela de interação com o usuário para coletar os dados de entrada para as páginas web da base de dados WoS. A Figura 28 mostra a Tela 2 do *Spider wos*.

Figura 20 - Tela 2 do *Spider wos*

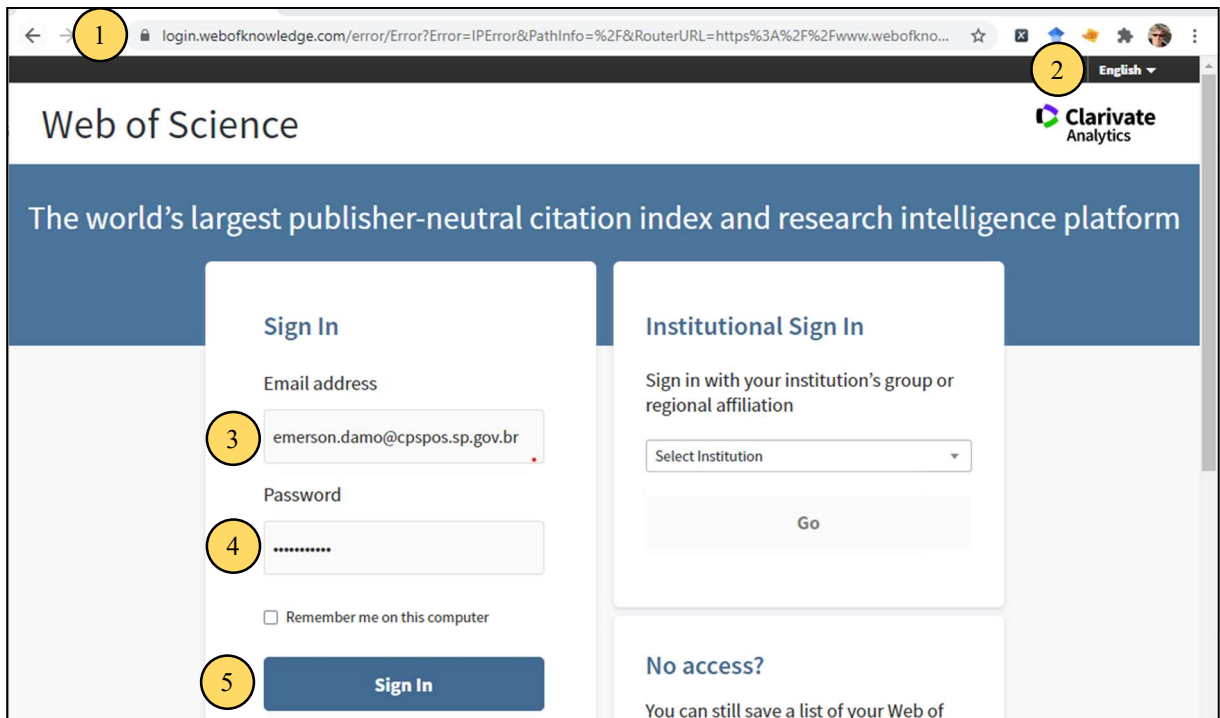
Fonte: o autor

Ao clicar no botão “Enviar dados” o sistema abre o *Google Chrome* e executa autonomamente as seguintes ações:

1. Abre a página web “https://login.webofknowledge.com/”;
2. Altera o idioma da página web de português para inglês;
3. Preenche o campo “*Email address*” com o texto informado no campo “Usuário WoS:” da Tela 2;
4. Preenche o campo “*Password*” com o texto informado no campo “Senha:” da Tela 2;
5. Clica no botão “*Sign In*”, o que acarreta a ação do portal da *WoS* de abrir a página web “apps.webofknowledge.com”;
6. Altera o campo de opção “*Select a database*” para “*Derwent Innovations Index*”;
7. Preenche o campo “*Basic search*” com o texto informado no campo “Sintaxe de Pesquisa:” da Tela 2;
8. Clica no botão “*Search*” o que acarreta a ação do portal da *WoS* de abrir a página web “apps.webofknowledge.com” com o resultado da pesquisa;
9. Altera o campo de opção “*Show*” para “*50 per page*”;
10. Copia a *URL* gerada pela página web (esta *URL* é a *Request* inicial);
11. Minimiza o navegador *Google Chrome*.

A Figura 21 mostra a página web “https://login.webofknowledge.com/” com as ações autônomas de 1 a 5.

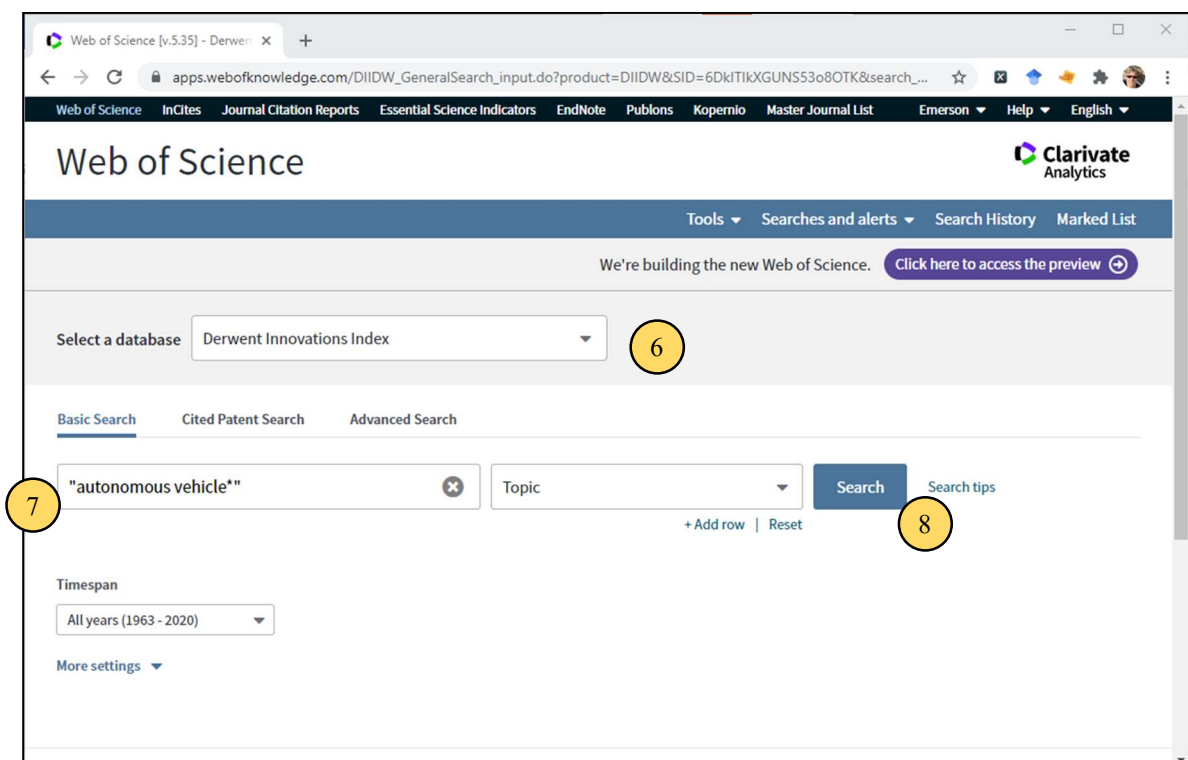
Figura 21 - Página web “login.webofknowledge.com/” com as ações autônomas de 1 a 5



Fonte: o autor

A Figura 22 mostra a página web “apps.webofknowledge.com” com as ações autônomas de 6 a 8.

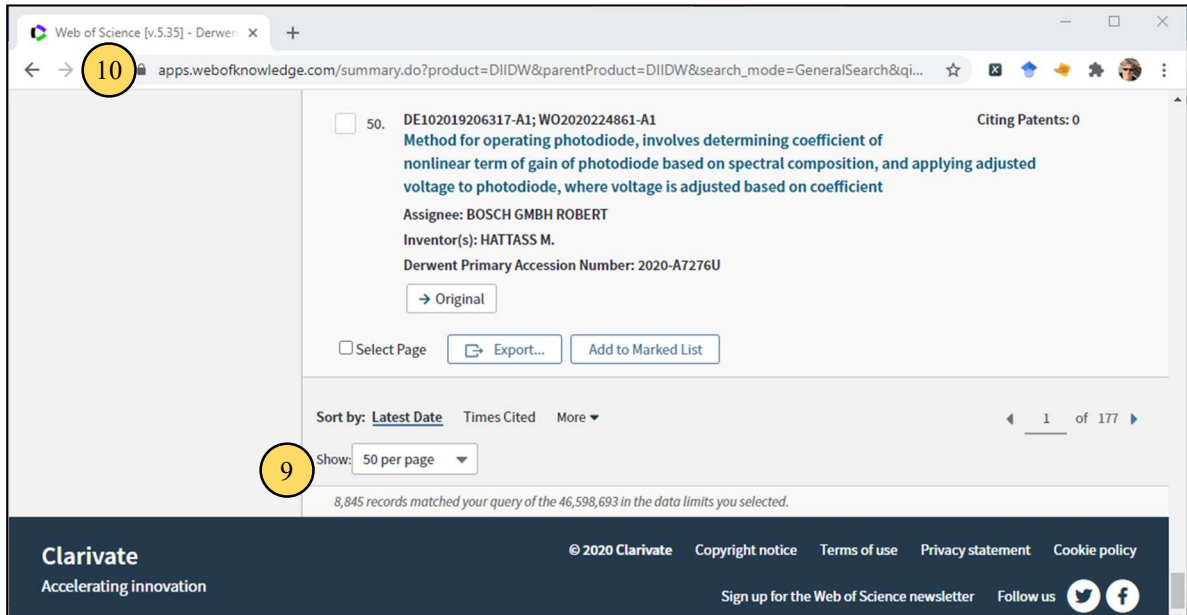
Figura 22 - Página web “login.webofknowledge.com/” com as ações autônomas de 6 a 8



Fonte: o autor

A Figura 23 mostra, na página web “apps.webofknowledge.com” com o resultado da pesquisa, a ação autônoma 9.

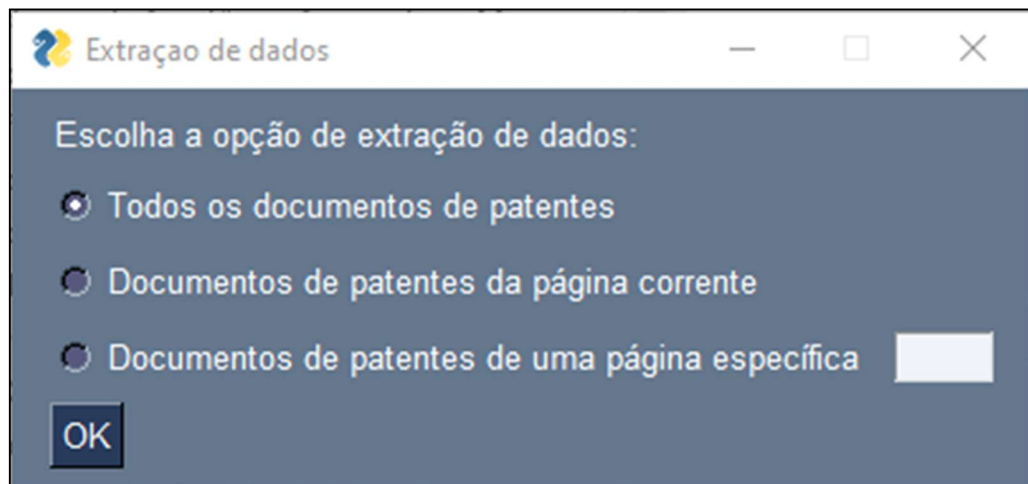
Figura 23 - Página web “login.webofknowledge.com/” com a ação autônoma 9



Fonte: o autor

Logo após o sistema coletar minimizar o navegador, o sistema abre a tela de opção de extração de dados dos documentos de patentes. O usuário pode optar por extrair: a) todos os documentos de patentes encontrados; b) os documentos de patentes de uma determinada página; ou c) os documentos de patentes da página corrente do navegador. A Figura 24 mostra a Tela 3 do *Spider vos*.

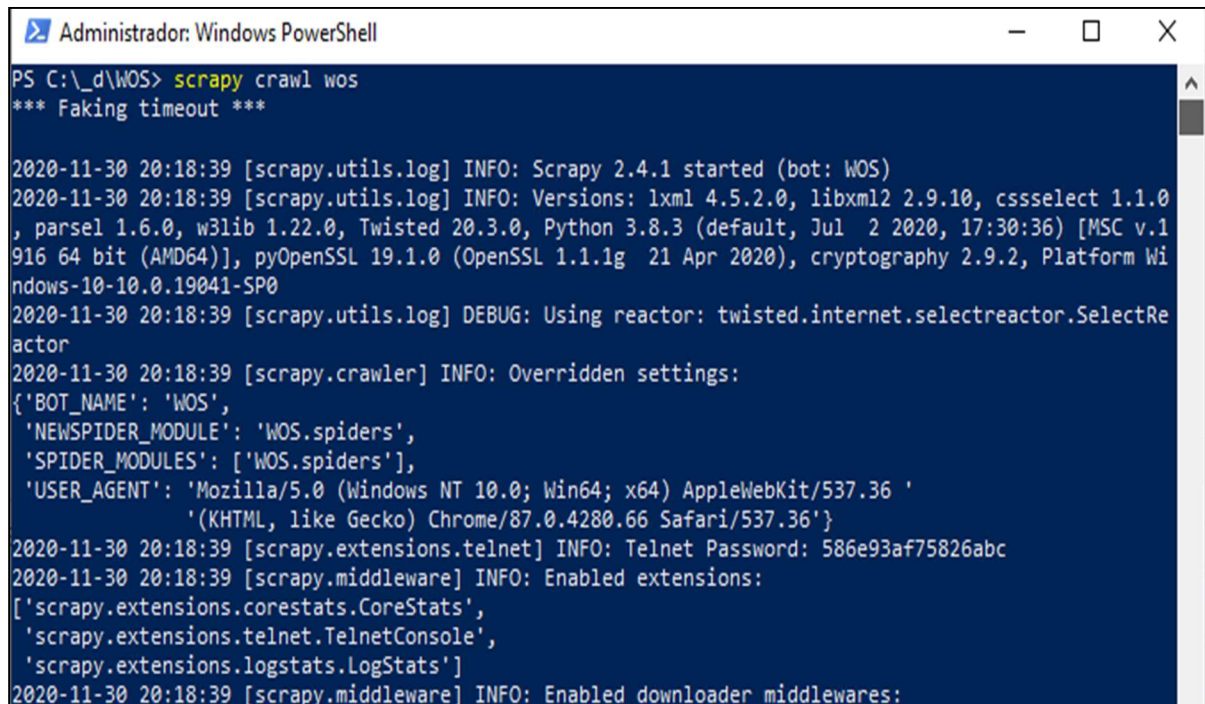
Figura 24 - Tela 3 do *Spider vos*



Fonte: o autor

Ao clicar o botão OK o sistema *web crawler* começa o processamento das *Requests* (requisições), das *Responses* (respostas), a obtenção dos dados dos documentos de patentes, e a gravação destes dados em arquivos *CSV*. Todas as etapas deste processo são visualizadas na tela do aplicativo *Windows PowerShell*. A Figura 25 mostra a tela do *PowerShell* no início do processo *spider engine*.

Figura 25 - Tela do *Windows PowerShell* no início do *spider engine*



```

Administrador: Windows PowerShell
PS C:\_d\WOS> scrapy crawl wos
*** Faking timeout ***

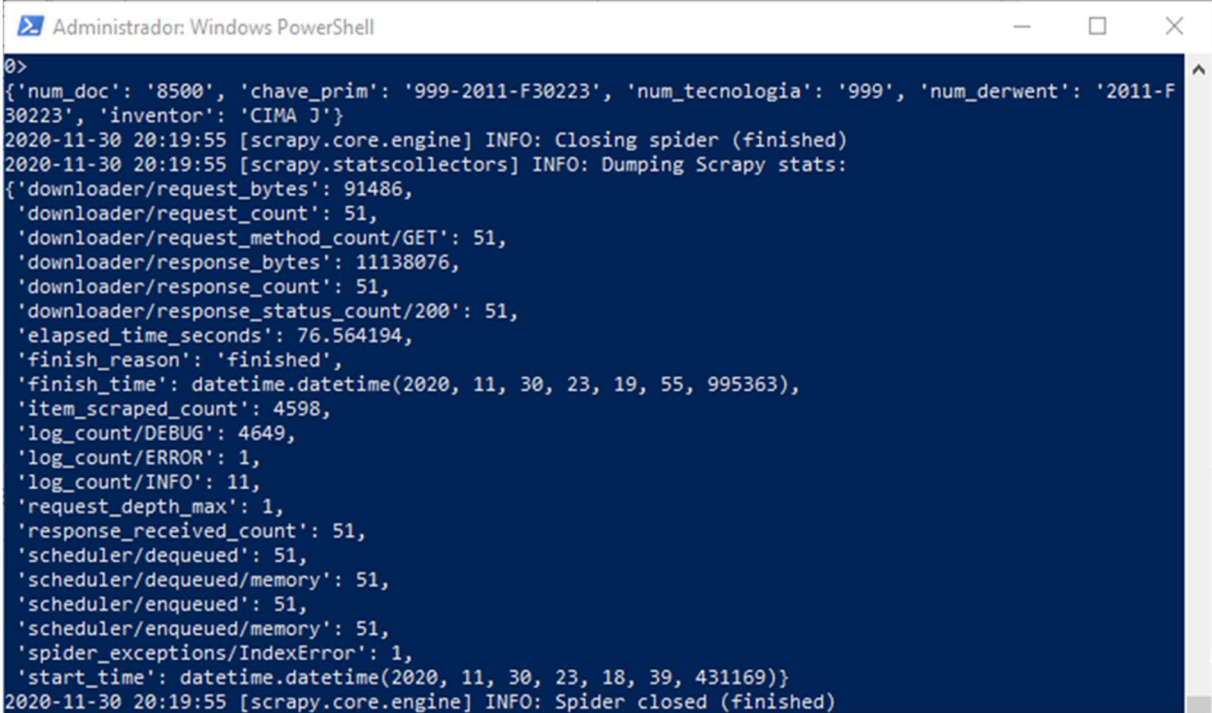
2020-11-30 20:18:39 [scrapy.utils.log] INFO: Scrapy 2.4.1 started (bot: WOS)
2020-11-30 20:18:39 [scrapy.utils.log] INFO: Versions: lxml 4.5.2.0, libxml2 2.9.10, cssselect 1.1.0
, parsel 1.6.0, w3lib 1.22.0, Twisted 20.3.0, Python 3.8.3 (default, Jul  2 2020, 17:30:36) [MSC v.1
916 64 bit (AMD64)], pyOpenSSL 19.1.0 (OpenSSL 1.1.1g  21 Apr 2020), cryptography 2.9.2, Platform Wi
ndows-10-10.0.19041-SP0
2020-11-30 20:18:39 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.selectreactor.SelectRe
actor
2020-11-30 20:18:39 [scrapy.crawler] INFO: Overridden settings:
{'BOT_NAME': 'WOS',
 'NEWSPIDER_MODULE': 'WOS.spiders',
 'SPIDER_MODULES': ['WOS.spiders'],
 'USER_AGENT': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 '
 '(KHTML, like Gecko) Chrome/87.0.4280.66 Safari/537.36'}
2020-11-30 20:18:39 [scrapy.extensions.telnet] INFO: Telnet Password: 586e93af75826abc
2020-11-30 20:18:39 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.corestats.CoreStats',
 'scrapy.extensions.telnet.TelnetConsole',
 'scrapy.extensions.logstats.LogStats']
2020-11-30 20:18:39 [scrapy.middleware] INFO: Enabled downloader middlewares:

```

Fonte: o autor

Ao terminar o processo *spider engine* o sistema *web crawler* mostra um resumo com as informações da quantidade de páginas *web* acessadas (*request count*), quantas páginas retornaram as informações solicitadas (*response count*), o tempo do processo (*elapsed_time_seconds*), e quantos campos foram baixados (*item_scraped_count*). A Figura 26 mostra a tela do *PowerShell* no final do processo *spider engine*. A Figura 27 mostra os arquivos *CSV* gerados após o processo *spider engine*.

Figura 26 - Tela do *Windows PowerShell* no final do *spider engine*

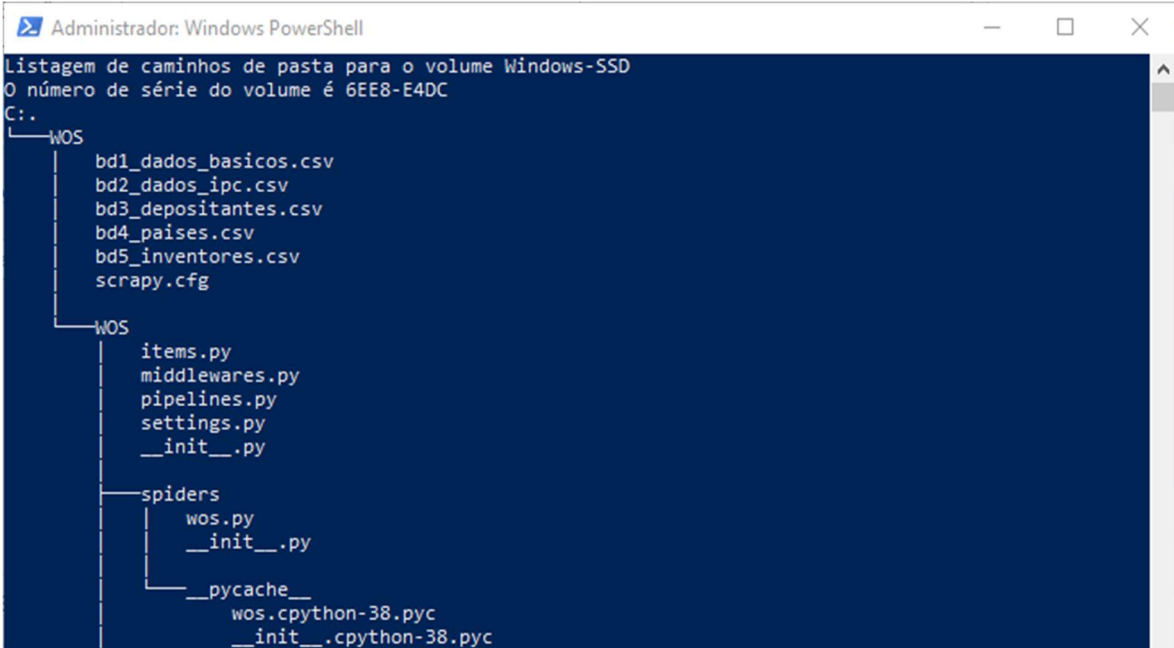


```

0>
{'num_doc': '8500', 'chave_prim': '999-2011-F30223', 'num_tecnologia': '999', 'num_derwent': '2011-F
30223', 'inventor': 'CIMA J'}
2020-11-30 20:19:55 [scrapy.core.engine] INFO: Closing spider (finished)
2020-11-30 20:19:55 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 91486,
 'downloader/request_count': 51,
 'downloader/request_method_count/GET': 51,
 'downloader/response_bytes': 11138076,
 'downloader/response_count': 51,
 'downloader/response_status_count/200': 51,
 'elapsed_time_seconds': 76.564194,
 'finish_reason': 'finished',
 'finish_time': datetime.datetime(2020, 11, 30, 23, 19, 55, 995363),
 'item_scraped_count': 4598,
 'log_count/DEBUG': 4649,
 'log_count/ERROR': 1,
 'log_count/INFO': 11,
 'request_depth_max': 1,
 'response_received_count': 51,
 'scheduler/dequeued': 51,
 'scheduler/dequeued/memory': 51,
 'scheduler/enqueued': 51,
 'scheduler/enqueued/memory': 51,
 'spider_exceptions/IndexError': 1,
 'start_time': datetime.datetime(2020, 11, 30, 23, 18, 39, 431169)}
2020-11-30 20:19:55 [scrapy.core.engine] INFO: Spider closed (finished)
  
```

Fonte: o autor

Figura 27 - Estrutura de diretórios e arquivos gerados após o processo *spider engine*

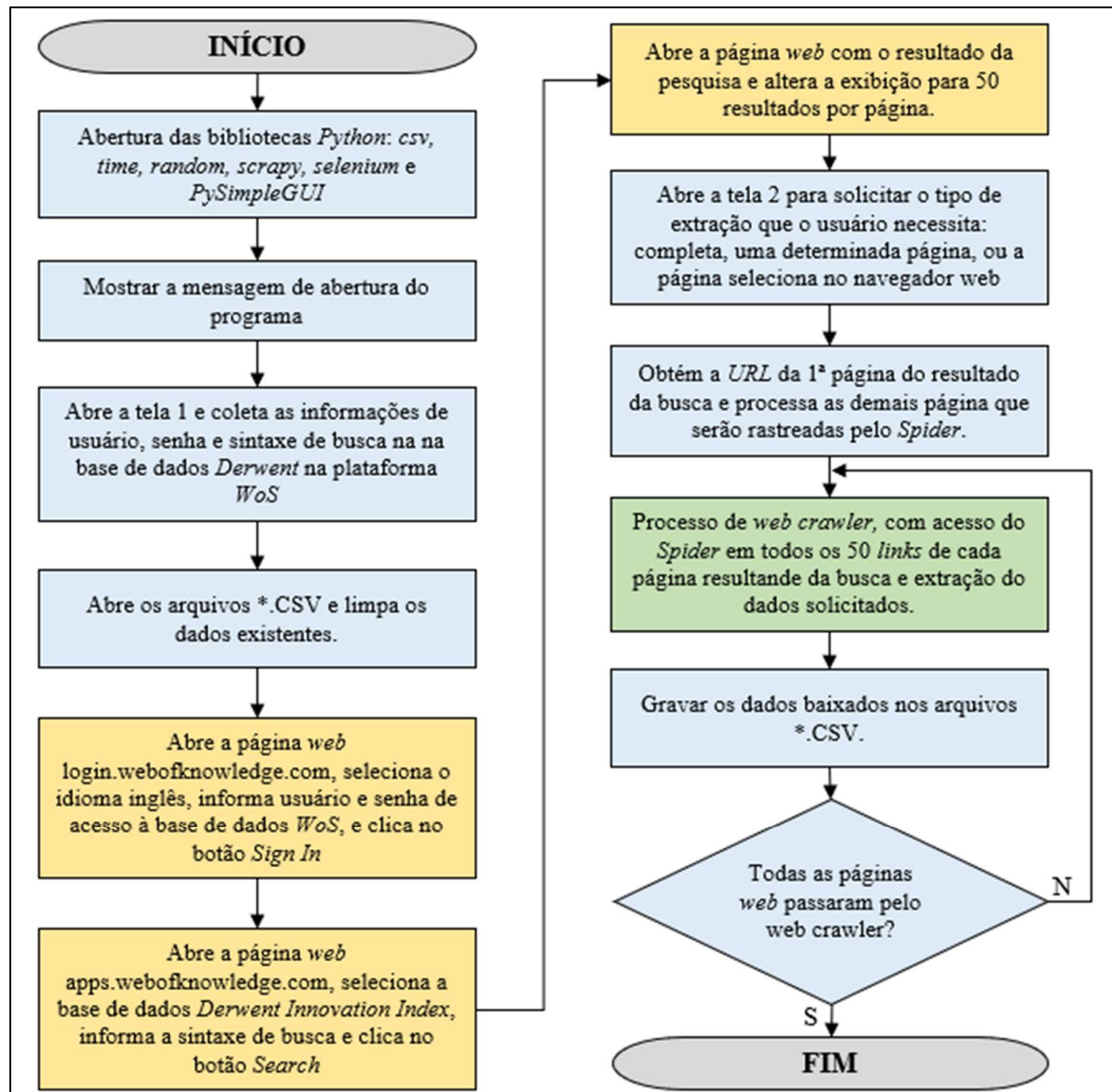


```

Listagem de caminhos de pasta para o volume Windows-SSD
O número de série do volume é 6EE8-E4DC
C:.
├── WOS
│   ├── bd1_dados_basicos.csv
│   ├── bd2_dados_ipc.csv
│   ├── bd3_depositantes.csv
│   ├── bd4_paises.csv
│   ├── bd5_inventores.csv
│   └── scrapy.cfg
├── WOS
│   ├── items.py
│   ├── middlewares.py
│   ├── pipelines.py
│   ├── settings.py
│   └── __init__.py
├── spiders
│   ├── wos.py
│   ├── __init__.py
│   └── __pycache__
│       ├── wos.cpython-38.pyc
│       └── __init__.cpython-38.pyc
  
```

Fonte: o autor

A Figura 28 mostra o fluxograma explicitando as etapas de processamento da *web crawler wos.py*.

Figura 28 - Funcionamento do programa *web crawler wos.py*

Fonte: o autor

2.3 ESTRUTURA DO BANCO DE DADOS

Por meio do programa de *Web Crawler wos.py*, gerou-se cinco arquivos de texto no formato CSV (*Comma Separated Values*), com informações contidas nos documentos de patentes, para as seis tecnologias relevantes à questão de pesquisa: *Vehicular Safety*; *Infotainment*; *Smart Cities*; *Smart Mobility*; *Advanced Driver Assistance Systems*; e

Autonomous Vehicles. A definição das tecnologias relevantes se deu por meio de mapas mentais construídos com informações obtidas pelo autor durante sua atuação na indústria automobilística, como engenheiro do produto, em conversas com especialistas e pesquisas na *internet*. O Quadro 1 apresenta as tecnologias, as *strings* de busca aos registros do *Derwent*, as datas de acesso a plataforma *Derwent* na base de dados *WoS*, e a quantidade de documentos de patentes encontrados em cada busca:

Quadro 1 - Tecnologia, String de busca, data de acesso e quantidade de patentes encontradas na base *WoS*

#	Tecnologia	String de busca no Derwent (WoS)	Data Acesso	Qtde. Documentos
1	Autonomous Vehicles	"autonomous vehicle*" or "autonomous car*"	18/03/2021	10.062
2	ADAS	"advanced driver assistance system*"	15/03/2021	881
3	Smart City	"smart city" or "smart cities"	16/03/2021	1.240
4	Smart Mobility	"smart mobility" or "urban mobility"	16/03/2021	75
5	Vehicular Safety	"vehicular safety"	16/03/2021	194
6	Infotainment	"infotainment system*"	16/03/2021	1.612

Fonte: o autor

A base de dados criada está dividida em: a) *bd1_dados_basicos.csv*; b) *bd2_dados_ipc.csv*; c) *bd3_depositantes.csv*; d) *bd4_paises.csv*; e) *bd5_inventores.csv*. Estes arquivos foram exportados e tratados no software Microsoft Excel for Office 365 MSO, obtendo-se cinco arquivos de base de dados no formato *XLSX*. Os Quadros 2, 3, 4, 5 e 6 mostram os campos criados e as informações contidas em cada campo por banco de dados.

Quadro 2 - Campos criados e descritivo do *bd1_dados_basicos*

bd1_dados_basicos	Descrição
<i>num_doc</i>	Número obtido na busca na plataforma <i>Derwent</i> ; serve de controle para verificar se todos os registros foram importados pelo programa; é removido no tratamento dos dados no Excel.
<i>chave_prim</i>	Criado pela união dos campos <i>num_derwent</i> e <i>num_tecnologia</i> ; usado no relacionamento com os demais bancos de dados.
<i>num_derwent</i>	Número único atribuído pela plataforma <i>Derwent</i> para uma patente ou um conjunto de patentes. Um número <i>Derwent</i> pode conter uma ou mais patentes.
<i>num_tecnologia</i>	Número atribuído pelos autores à tecnologia pesquisada (vide Quadro 2)
<i>número_patente_principal</i>	Número da 1ª patente encontrada no documento
<i>data_publicacao_patente_principal</i>	Data da concessão da patente principal
<i>título_patente</i>	Título das patentes na plataforma <i>Derwent</i>

Quadro 2 - Campos criados e descritivo do bd1_dados_basicos (cont.)

bd1_dados_basicos	Descrição
resumo_patente	Descritivo das patentes na plataforma Derwent
patentes_que_fizeram_citacao	Número de patentes que citaram as patentes deste documento
patentes_citadas_inventor	Número de patentes citadas pelo inventor
patentes_citadas_examinador	Número de patentes citadas pelo examinador
artigos_citados_inventor	Número de artigos citados pelo inventor
artigos_citados_examinador	Número de artigos citados pelo examinador

Fonte: O autor

Quadro 3 - Campos criados e descritivo do bd2_dados_ipc

bd2_dados_ipc	Descrição
num_doc	Número obtido na busca na plataforma Derwent; serve de controle para verificar se todos os registros foram importados pelo programa; é removido no tratamento dos dados no Excel.
chave_prim	Criado pela união dos campos num_derwent e num_tecnologia; usado no relacionamento com os demais bancos de dados.
num_derwent	Número único atribuído pela plataforma Derwent para uma patente ou um conjunto de patentes. Um número Derwent pode conter uma ou mais patentes.
número patente	Número da patente encontrada no documento
data_publicacao	Data da concessão da patente
ipc_principal	Número IPC (International Patent Classification) da patente

Fonte: o autor

Quadro 4 - Campos criados e descritivo do bd3_depositantes

bd3_depositantes	Descrição
num_doc	Número obtido na busca na plataforma Derwent; serve de controle para verificar se todos os registros foram importados pelo programa; é removido no tratamento dos dados no Excel.
chave_prim	Criado pela união dos campos num_derwent e num_tecnologia; usado no relacionamento com os demais bancos de dados.
num_derwent	Número único atribuído pela plataforma Derwent para uma patente ou um conjunto de patentes. Um número Derwent pode conter um ou mais depositantes.
codigo_depositante	Código do depositante da patente
depositante	Nome do depositante da patente

Fonte: o autor

Quadro 5 - Campos criados e descritivo do bd4_países

bd4_países	Descrição
num_doc	Número obtido na busca na plataforma Derwent; serve de controle para verificar se todos os registros foram importados pelo programa; é removido no tratamento dos dados no Excel.
chave_prim	Criado pela união dos campos num_derwent e num_tecnologia; usado no relacionamento com os demais bancos de dados.
num_derwent	Número único atribuído pela plataforma Derwent para uma patente ou um conjunto de patentes. Um número Derwent pode conter uma ou mais patentes atribuídas a um ou mais países.
numero_patente	Número da patente registrada no país
pais	Código do país onde a patente está registrada, conforme a ISSO
tipo	Tipo de registro de patente no país (original ou regional)

Fonte: o autor

Quadro 6 - Campos criados e descritivo do bd5_inventores

bd5_inventores	Descrição
num_doc	Número obtido na busca na plataforma Derwent; serve de controle para verificar se todos os registros foram importados pelo programa; é removido no tratamento dos dados no Excel.
chave_prim	Criado pela união dos campos num_derwent e num_tecnologia; usado no relacionamento com os demais bancos de dados.
num_derwent	Número único atribuído pela plataforma Derwent para uma patente ou um conjunto de patentes. Um número Derwent pode conter um ou mais inventores.
numero_patente	Nome dos inventores descritos no documento

Fonte: o autor

Obteve-se as informações da descrição dos códigos IPC (*International Patents Classification*) no site da WIPO (*World Intellectual Property Organization*), acessado em 17/02/2020. Os dados foram copiados diretamente da WEB e tratados no software Microsoft Excel for Office 365 MSO, obtendo-se um arquivo de base de dados no formato XLSX. Schmoch (2008) estudou a correlação entre os códigos IPC e as áreas tecnológicas da indústria e as sumarizou em uma tabela. Estas correlações foram introduzidas manualmente na base de dados usando o software Microsoft Excel for Office 365 MSO, complementando as informações dos códigos IPC. O Quadro 7 mostra as informações obtidas e o banco de dados gerado.

Quadro 7 - Campos criados e descritivo do bd6_tabela_IPC

bd6_tabela_IPC	Descrição
codigo_ipc	Número IPC (International Patents Classification)
descricao_port	Descrição em português do código IPC
descricao_ingl	Descrição em inglês do código IPC
setor_industrial	Setor industrial do IPC
sub-setor_industrial	Subsetor industrial do IPC

Fonte: o autor

Obteve-se as informações dos códigos de países no site da *International Organization for Standardization* (ISO), acessado em 17/02/2020. Os dados foram copiados diretamente da WEB e tratados no software Microsoft Excel for Office 365 MSO, quando se obteve um arquivo de base de dados no formato XLSX. O Quadro 8 mostra as informações obtidas e o banco de dados gerado.

Quadro 8 - Campos criados e descritivo do bd7_codigo_paises

bd7_codigo_paises	Descrição
Name	Nome do país
Code	Código do país conforme ISO

Fonte: o autor

Por meio do *software Microsoft Excel for Office 365 MSO* criou-se uma base de dados para controle das pesquisas efetuadas à plataforma *Derwent* na base de dados WoS. Os dados relativos a número da tecnologia, *string* de busca utilizada, data de acesso à plataforma, quantidade de documentos encontrados e quantidade de patentes encontradas, foram digitados pelos autores. O Quadro 9 mostra as informações digitadas e o banco de dados gerado.

Quadro 9 - Campos criados e descritivo do bd8_controle_acesso_base_dados

bd8_controle_acesso_base_dados	Descrição
num_tecnologia	Número atribuído pelos autores à tecnologia pesquisada
tecnologia	Nome atribuído pelos autores à tecnologia pesquisada
string_de_busca	Sintaxe de busca utilizada para efetuar a pesquisa à plataforma Derwent na base de dados WoS

Quadro 9 - Campos criados e descritivo do bd8_controle_acesso_base_dados (cont.)

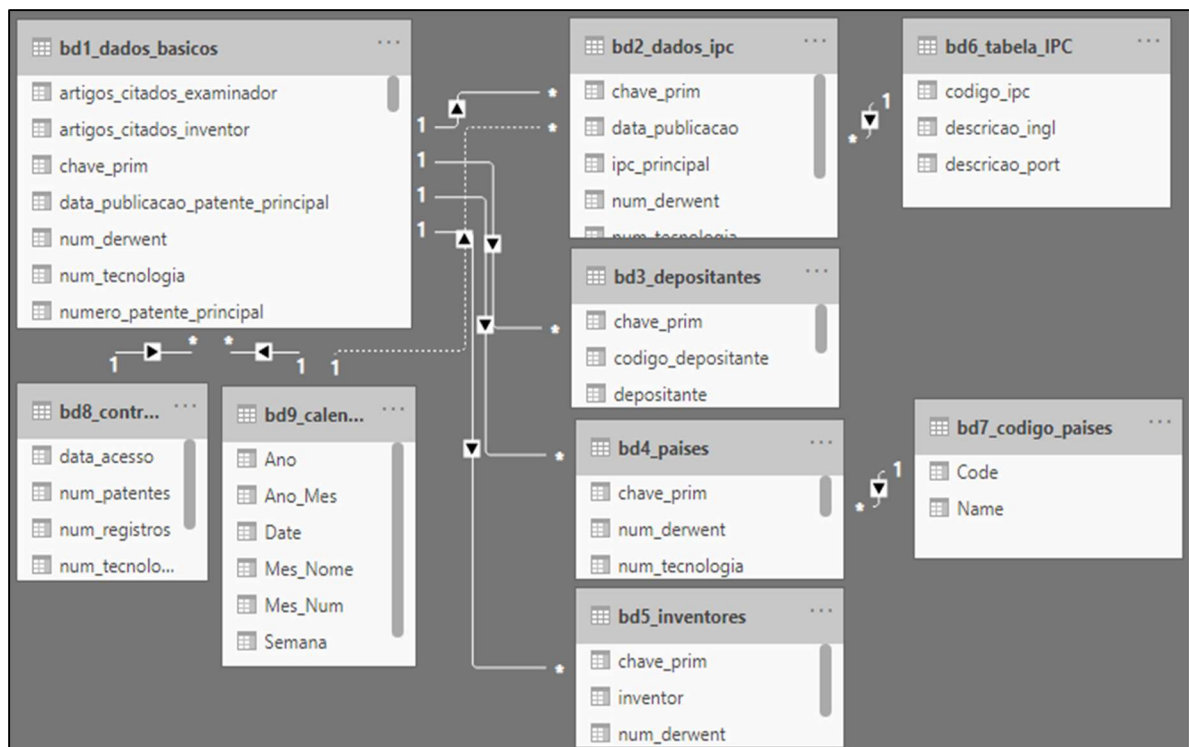
bd8_controle_acesso_base_dados	Descrição
data_acesso	Data de acesso à base de dados
num_registros	Número de documento encontrados
num_patentes	Número de patentes encontradas

Fonte: o autor

Exportou-se os oito bancos de dados para o *software Microsoft Power BI Desktop* versão 2.76.5678.782. O *Power BI* é um conjunto de ferramentas de *business intelligence*, cujo objetivo é analisar dados e compartilhar informações por meio de relatórios e dashboards interativos (KRISHNAN, 2017). O *software* permite a criação de métricas e indicadores personalizados, o que vem de encontro com o objetivo deste estudo.

Utilizando-se a linguagem de programação “M”, nativa do *Power BI*, construiu-se o banco de dados bd9_calendario, considerando a menor e a maior data de publicação do documento de patente no banco de dados “Dados_Complementares”, contendo os campos: [Date]; [Ano]; [Trimestre]; [Semana]; [Mes_Nome]; [Mes_Num]; e [Ano_Mes].

A Figura 29 mostra a estrutura dos bancos de dados no *software Power BI*, incluindo as relações entre campos.

Figura 29 - Estrutura do banco de dados relacional no *software Power BI*

Fonte: o autor.

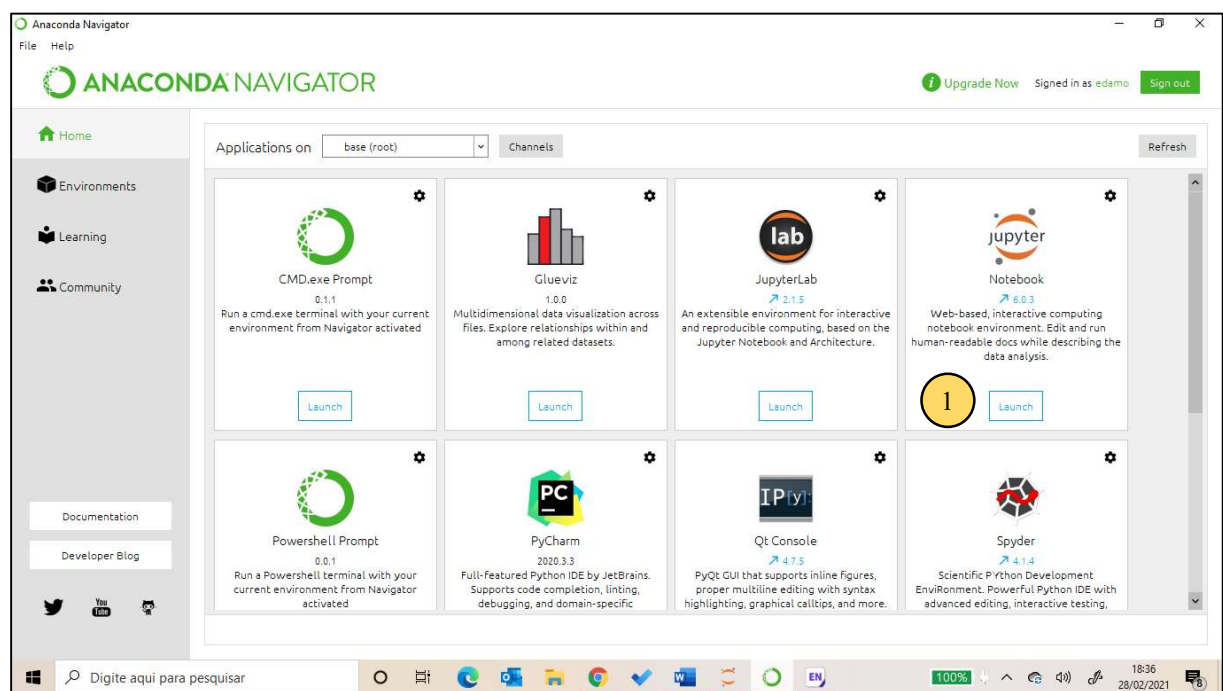
2.4 CRIAÇÃO DO BANCO DE DADOS DO FLUXO DE CONHECIMENTO

Para atender um dos objetivos específicos deste trabalho utilizou-se novamente a técnica de *web crawler*. Por meio do software *Jupyter Notebook*, versão 6.0.3, criou-se um programa em *Python*, versão 3.8.3, para a coleta de dados de documentos de patentes dos registros do *Derwent Innovations Index* da base de dados WoS, mais especificamente os números dos documentos de patentes que fizeram citações aos documentos de patentes coletados na primeira etapa deste trabalho.

De acordo com Perkel (2018) o *Jupyter* é uma ferramenta da *web* interativa gratuita, de código aberto, conhecida como notebook computacional, que os pesquisadores podem usar para combinar código de software, saída computacional, texto explicativo e recursos multimídia em um único documento.

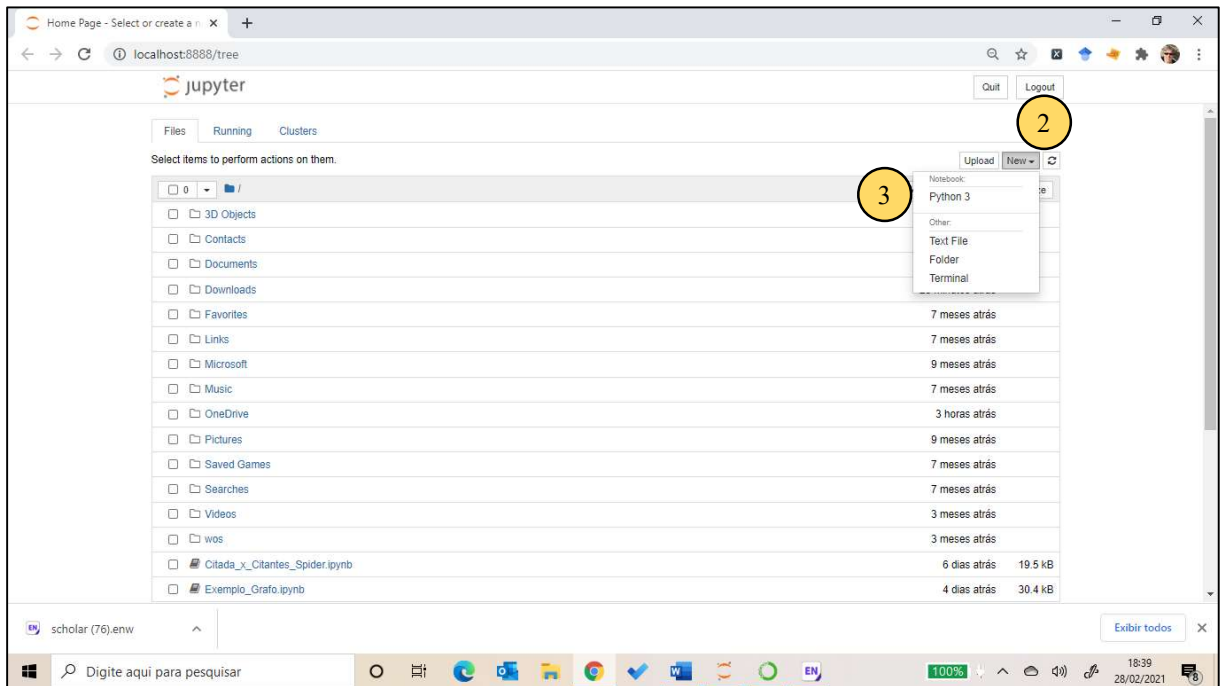
Após a abertura do *software Jupyter*, por meio do *Anaconda Navigator*, cria-se um *notebook* da linguagem *Python 3*. A Figura 30 mostra o *software* Anaconda Navigator. A Figura 31 mostra a tela para a criação de um *notebook* no *software Jupyter*. A Figura 32 mostra o *notebook* criado no *software Jupyter*.

Figura 30 – Tela do Software Anaconda Navigator.



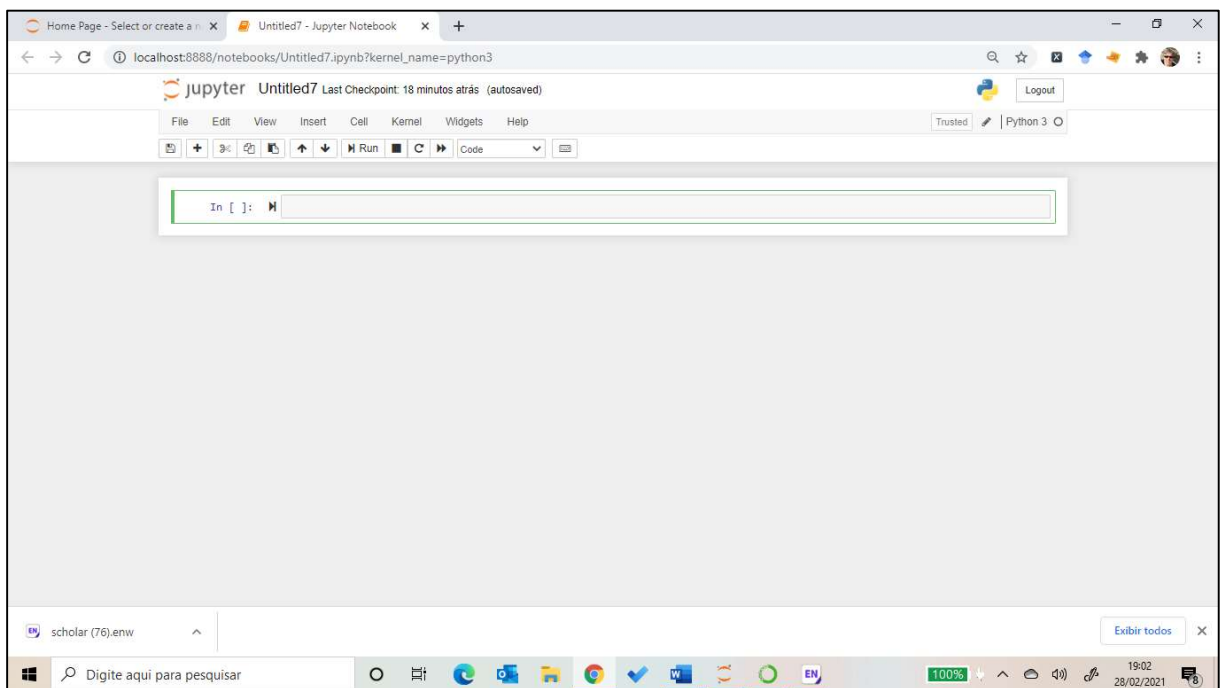
Fonte: o autor

Figura 31 - Tela para a criação de um *notebook* no *software Jupyter*



Fonte: o autor

Figura 32 - Notebook criado no *software Jupyter*



Fonte: o autor

A próxima etapa é a criação do programa de *web crawler*, por meio do notebook criado, com a inclusão de códigos para interação e obtenção dos dados das páginas web da base de

dados *WoS*. O programa *web crawler* fluxo *spider.ipynb* (Spider), construído pelo autor, pode ser dividido em: a) abertura das bibliotecas Python; b) criação dos *dataframes*, por meio da biblioteca Pandas, tendo como base os arquivos *bd1_dados_basicos.CSV* (para os *nodes* nível 1) e *citações_entrada.CSV* (para os *nodes* dos demais níveis); c) interação com o navegador *Google Chrome*, versão 88.0.4324.190; e d) processamento da *web crawler* (*Scrape Engine*) e geração do arquivo *bd9_fluxo_conhecimento.CSV*.

2.4.1 Código de Programação *fluxo_spider.ipynb* – Abertura das bibliotecas Python

A primeira parte do código do programa abre as bibliotecas que serão utilizados pelo programa *web crawler* *fluxo_spider.ipynb* e fornece as configurações ao software *Jupyter* para funcionar com *web crawler*. Utilizou-se as mesmas bibliotecas do programa *wos.py*, *Scrapy*, *Selenium*, *Random* e, adicionalmente, para trabalhar com *dataframes*, a biblioteca *Pandas*.

O *Pandas* é uma biblioteca *Python* que disponibiliza ferramentas para trabalhar com conjuntos de dados estruturados dos campos das estatísticas, finanças, ciências sociais e muitos outros. A biblioteca oferece rotinas intuitivas integradas para executar manipulações de dados e análise de tais conjuntos de dados (MCKINNEY, 2011).

A Figura 33 mostra as linhas de código do programa *fluxo_spider.ipynb* relativas a abertura das bibliotecas *Python* e as configurações ao software *Jupyter* para funcionar como *web crawler*.

Figura 33 - Código do programa *fluxo_spider.ipynb* relativas a abertura das bibliotecas *Python* e as configurações ao software *Jupyter* para funcionar como *web crawler*

```
import random
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import scrapy
from scrapy.http import Request
from scrapy.crawler import CrawlerProcess
import pandas as pd

from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

import platform
platform.python_version()
```

Fonte: o autor

2.4.2 Código de Programação *fluxo_spider.ipynb* – Criação dos *dataframes*

Nesta etapa do código de programação criou-se o *dataframe* de entrada de dados *df1* e o de saída de dados *df2*. Na criação do *dataframe* *df1* usa-se, para a entrada do primeiro nível do fluxo de conhecimento, os arquivos *db1_dados_basico.csv* e *bd2_dados_ipc.csv*, que foram gerados pelo programa *wos.py*. Para processamento da *web crawler* utiliza-se os campos: *num_derwent*, *num_tecnologia*, *patentes_que_fizeram_citacao* e *ipc_principal*. O campo *ipc_principal*, original do arquivo *bd2_dados_ipc.csv*, foi adicionado ao arquivo *db1_dados_basico.csv* por meio da função *PROCV()* do software *Microsoft Excel*. Após o processamento do programa *web crawler fluxo_spider.ipynb* gera-se o *dataframe* *df2* com os campos: *nivel*, *num_tecnologia*, *num_derw_citada*, *IPC_citada*, *num_cit_citada*, *num_derw_citante*, *IPC_citante*, e *num_cit_citante*. A partir do segundo nível do fluxo de conhecimento utiliza-se os dados do *dataframe* *df2* do nível anterior: *num_derw_citante*, *IPC_citante*, e *num_cit_citante*. A Figura 34 mostra as linhas de código do programa *fluxo_spider.ipynb* relativas a criação do *dataframe* de entrada *df1* e do *dataframe* de saída *df2*. Os Quadros 10 e 11 mostram os campos de entrada (*dataframe* *df1*) e de saída (*dataframe* *df2*) conforme nível de fluxo de conhecimento.

Figura 34 - Código do programa *fluxo_spider.ipynb* relativo a criação do *dataframe* de entrada *df1* e do *dataframe* de saída *df2*.

```
global df1
global df2

# Usa para construir o nó 1
#df1 = pd.read_excel("bd1_dados_basicos.csv", sheet_name = "bd1_dados_basicos")
#df1 = df1.query('num_tecnologia == 4').filter(items=['num_derwent_index', 'num_derwent',
#          'num_tecnologia', 'patentes_que_fizeram_citacao', 'IPC'])
#df1.set_index('num_derwent_index', inplace=True)

# Usa para construir os demais nós
df1 = pd.read_csv("citacoes_entrada.csv", sep=';')
df1.set_index('num_derwent_index', inplace=True)

df2 = pd.DataFrame(columns = ['nivel', 'num_tecnologia', 'num_derw_citada', 'IPC_citada', 'num_cit_citada',
                              'num_derw_citante', 'IPC_citante', 'num_cit_citante'])
```

Fonte: O autor

Quadro 10 - Campos de entrada (*dataframe* *df1*) e de saída (*dataframe* *df2*) para o primeiro nível do fluxo de conhecimento.

Nível 1 = primeira execução da <i>web crawler</i> <i>fluxo_spider.ipynb</i>	
<i>Dataframe</i> <i>df1</i>	<i>Dataframe</i> <i>df2</i>
	<i>df2_{n=1}</i> [nivel] → gerado pelo programa

Quadro 10 - Campos de entrada (*dataframe* df1) e de saída (*dataframe* df2) para o primeiro nível do fluxo de conhecimento (cont.)

<i>Dataframe</i> df1	<i>Dataframe</i> df2
bd1_dados_básicos [num_tecnologia]	df2 _{n=1} [num_tecnologia]
bd1_dados_básicos [num_derwent]	df2 _{n=1} [num_derw_citada]
bd1_dados_básicos [IPC_principal]	df2 _{n=1} [IPC_citada]
bd1_dados_básicos [patentes que fizeram citacao]	df2 _{n=1} [num_cit_citada]
	df2 _{n=1} [num_derw_citante] → <i>web crawler</i>
	df2 _{n=1} [IPC_citante] → <i>web crawler</i>
	df2 _{n=1} [num_cit_citante] → <i>web crawler</i>

Fonte: o autor

Quadro 11 - Campos de entrada (*dataframe* df1) e de saída (*dataframe* df2) para o primeiro nível do fluxo de conhecimento.

Nível 2 até n = demais execuções da <i>web crawler</i> <i>fluxo_spider.ipynb</i>	
<i>Dataframe</i> df1; [num_cit_citante] > 0	<i>Dataframe</i> df2
df2 _n [nivel]	df2 _{n+1} [nivel] + 1 → <i>gerado pelo programa</i>
df2 _n [num_tecnologia]	df2 _{n+1} [num_tecnologia]
df2 _n [num_derw_citada]	df2 _{n+1} [num_derw_citada]
df2 _n [IPC_citada]	df2 _{n+1} [IPC_citada]
df2 _n [num_cit_citada]	df2 _{n+1} [num_cit_citada]
df2 _n [num_derw_citante]	df2 _{n+1} [num_derw_citante] → <i>web crawler</i>
df2 _n [IPC_citante]	df2 _{n+1} [IPC_citante] → <i>web crawler</i>
df2 _n [num_cit_citante]	df2 _{n+1} [num_cit_citante] → <i>web crawler</i>

Fonte: o autor

2.4.3 Código de Programação *fluxo_spider.ipynb* – interação com o navegador

O código de programação utilizado no programa *fluxo_spider.ipynb* para interação com o navegador *Google Chrome* é praticamente idêntico ao construído no programa *wos.py*. A única diferença é a inclusão de uma linha para obter os *links* (*start_urls*) das páginas com as informações das patentes que fizeram a citação da patente original. A Figura 35 mostra a linha

Figura 37 - Linhas de 1 a 55 do código do programa `fluxo_spider.ipynb` para a execução do *Scrapy Engine*

Classe object do spider e urls iniciais

```

class FluxoConhecimento(scrapy.Spider):
    name = 'FC'
    allowed_domains = ['webofknowledge.com']
    start_urls = []

    url_corrente = driver.current_url

    for index, row in df1.iterrows():
        #if row['patentes_que_fizeram_citacao'] > 0:
        next_page_url = url_corrente.split('value=')[0] + 'value=' + row['num_derwent'][0:4] \
            + row['num_derwent'][5:11] + '&' + url_corrente.split('&')[-1]

        start_urls.append(next_page_url)

```

Função que executa a 1ª parte do crawler

```

def parse(self, response):

    num_derw_citada = response.request.url.split('value=')[1].split('&parent')[0]
    num_derw_citada = num_derw_citada[0:4] + '-' + num_derw_citada[4:]

    urls = response.xpath('//a[contains(@class,"snowplow-full-record")]/@href').extract()

    for url in urls:
        absolute_url = response.urljoin(url)
        yield Request(absolute_url, callback=self.parse_fluxo, dont_filter=True,
                      headers={'num_derw_citada': num_derw_citada})
        sleep(random.randrange(2,8))

```

Função que executa a 2ª parte do crawler

```

def parse_fluxo(self, response):

    if type(response.request.headers.get('num_derw_citada')) == 'str':
        num_derw_citada = response.request.headers.get('num_derw_citada')
    else:
        num_derw_citada = (response.request.headers.get('num_derw_citada')).decode()

    nivel = 8
    num_tecnologia = df1.loc[num_derw_citada]['num_tecnologia']
    IPC_citada = df1.loc[num_derw_citada]['IPC']
    num_cit_citada = df1.loc[num_derw_citada]['patentes_que_fizeram_citacao']

    IPC_citante = ''

    df2.loc[len(df2)] = [nivel, num_tecnologia, num_derw_citada, IPC_citada,
                        num_cit_citada, num_derw_citante, IPC_citante, num_cit_citante]

```

Processo web crawler e gravação do arquivo `bd9_citacoes.csv`

```

process = CrawlerProcess({'USER_AGENT': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)'})

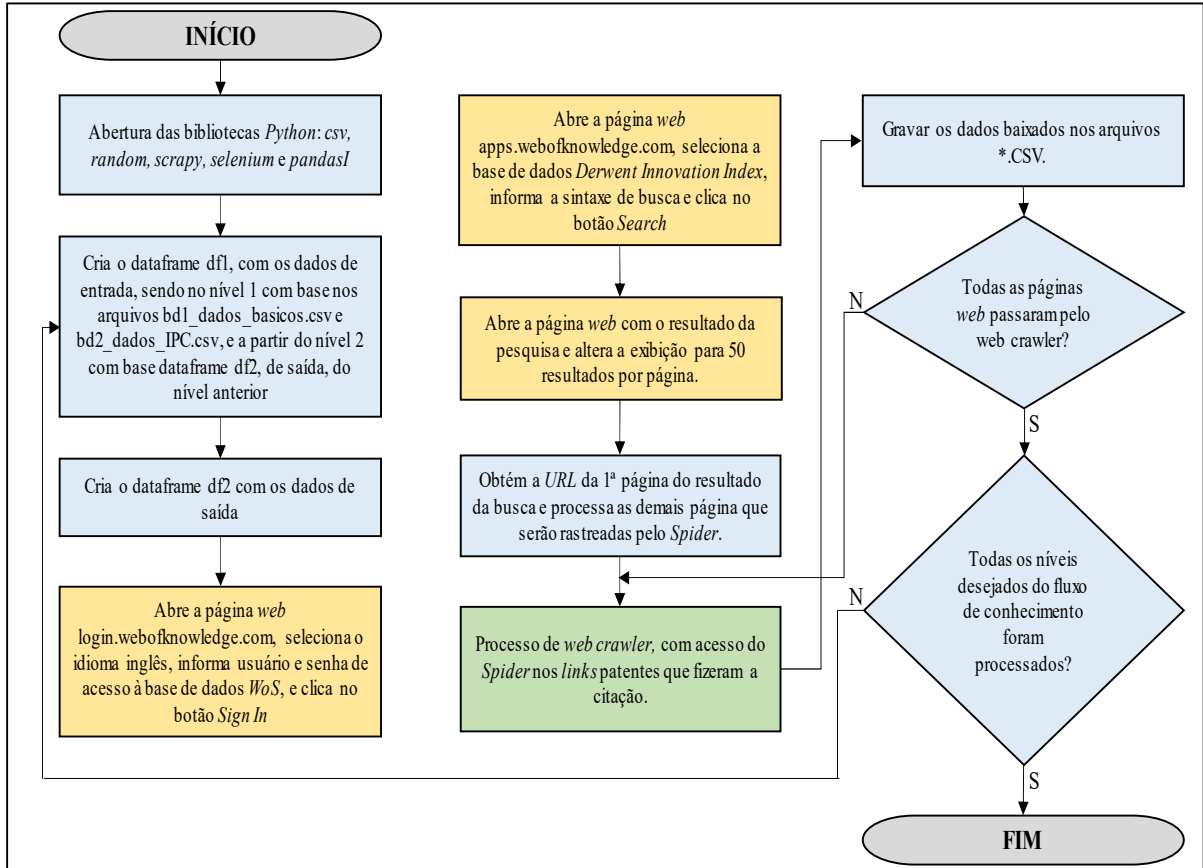
process.crawl(FluxoConhecimento)
process.start()

df2.to_csv('bd9_citacoes.csv', sep=';', encoding='utf-8')

```


A Figura 38 mostra o fluxograma explicitando as etapas de processamento da *web crawler fluxo_spider.ipynb*.

Figura 38 - Funcionamento do programa *web crawler fluxo_spider.ipynb*



Fonte: o autor

2.5 CRIAÇÃO DO MAPA DO FLUXO DO CONHECIMENTO

Para a construção de um mapa do fluxo de conhecimento criou-se um programa em *Python*, versão 3.8.3, por meio do *software Jupyter Notebook*, versão 6.0.3. O programa utiliza como entrada de informações o arquivo *bd9_citacoes.CSV*, criado por meio do programa *fluxo_spider.ipynb*, e tem como objetivo a criação de grafos para criar um mapa do fluxo de conhecimento de uma determinada tecnologia.

Segundo Abreu *et al.* (2007) um grafo é uma estrutura $G = G(V; E)$, constituída por um conjunto finito e não vazio V , cujos elementos são denominados vértices, e um conjunto E de subconjuntos a dois elementos de V , denominados arestas.

A partir da matemática constituída pela teoria de grafos, uma rede pode ser modelada e estruturada por um grafo. A rede representada por um grafo $G = (V, E)$ consiste em um conjunto formado por Vértices (V) e Arestas (E). Em uma análise de rede social, o vértice representa um ator, e uma aresta representa a relação existente entre dois atores. A ligação entre dois vértices pode ser direcionada ou não direcionada (PURIFICAÇÃO, 2015).

O programa *gerar_grafo.ipynb*, criado pelo autor, pode ser dividido em: a) abertura das bibliotecas Python; b) criação dos *Dataframes*, por meio da biblioteca Pandas, tendo como base os arquivos *bd9_fluxo_conhecimento.xlsx*, editado no software Microsoft Excel a partir do arquivo *bd9_citacoes.CSV*; c) construção do grafo por meio da biblioteca *NetworkX*; e d) criação da imagem gráfica do mapa do fluxo de conhecimento por meio da biblioteca *Matplotlib*.

NetworkX é uma biblioteca do Python para exploração e análise de redes e algoritmos de rede. O pacote principal fornece estruturas de dados para representar muitos tipos de redes ou gráficos, incluindo grafos simples, grafos direcionados e grafos com arestas paralelas e loops próprios. Os nós em gráficos *NetworkX* podem ser qualquer objeto *Python* e as arestas podem conter dados arbitrários, sendo que esta flexibilidade torna o *NetworkX* ideal para representar redes encontradas em muitos campos científicos diferentes (HAGBERG; SWART; S CHULT, 2008).

Matplotlib é uma biblioteca do *Python* para plotagem de imagem 2D, voltada principalmente para a visualização de dados científicos, de engenharia e financeiros. *Matplotlib* pode ser usado interativamente a partir do *shell* Python, chamado de *scripts python* ou incorporado em um aplicativo com interface gráfica com o usuário (BARRETT *et al.*, 2005). A Figura 39 mostram as linhas de código do programa *gerar_grafo.ipynb* para a construção do mapa de fluxo de conhecimento.

Figura 39 – Linhas de código do programa *gerar_grafo.ipynb* para a construção do mapa de fluxo de conhecimento

Abertura da bibliotecas e criação do Dataframe

```

import pandas as pd
import networkx as nx
import matplotlib
import matplotlib.pyplot as plt

df = pd.read_excel('bd9_fluxo_conhecimento.xlsx', sheet_name='bd9_network')

```

Criação do grafo (nodes e edges)

```

G = nx.Graph()
for index, row in df.iterrows():
    if row['num_cit_citada'] > 0:
        size = row['num_cit_citada']
    else:
        size = 1

    G.add_node(row['num_derw_citada'], Type=row['tecnologia'], Node_Size=size, Date=row['ano'])
    G.add_node(row['num_derw_citante'], Type=row['tecnologia'], Node_Size=size, Date=row['ano'])
    G.add_edge(row['num_derw_citada'], row['num_derw_citante'])

```

Separação dos nodes por campo tecnológico

```

A_nodes = [n for (n,t) in \
    nx.get_node_attributes(G,'Type').items() if t == 'A']
B_nodes = [n for (n,t) in \
    nx.get_node_attributes(G,'Type').items() if t == 'B']
C_nodes = [n for (n,t) in \
    nx.get_node_attributes(G,'Type').items() if t == 'C']
D_nodes = [n for (n,t) in \
    nx.get_node_attributes(G,'Type').items() if t == 'D']
E_nodes = [n for (n,t) in \
    nx.get_node_attributes(G,'Type').items() if t == 'E']

nx.draw_networkx_nodes(G, pos, nodelist=B_nodes, \
    node_color='c', node_shape='o')
nx.draw_networkx_nodes(G, pos, nodelist=C_nodes, \
    node_color='w', node_shape='s')
nx.draw_networkx_nodes(G, pos, nodelist=D_nodes, \
    node_color='k', node_shape='v')
nx.draw_networkx_nodes(G, pos, nodelist=E_nodes, \
    node_color='g', node_shape='o')
nx.draw_networkx_nodes(G, pos, nodelist=F_nodes, \
    node_color='b', node_shape='s')
nx.draw_networkx_nodes(G, pos, nodelist=G_nodes, \
    node_color='y', node_shape='o')
nx.draw_networkx_nodes(G, pos, nodelist=H_nodes, \
    node_color='r', node_shape='s')

nx.draw_networkx_edges(G, pos, alpha=1, arrows=True, arrowsize=150,
    arrowstyle="Wedge", connectionstyle="arc3")

nx.draw_networkx_labels(G, pos, font_size=8, font_color='k')

plt.title('Fluxo de Conhecimento - Smart Cities', size=20)

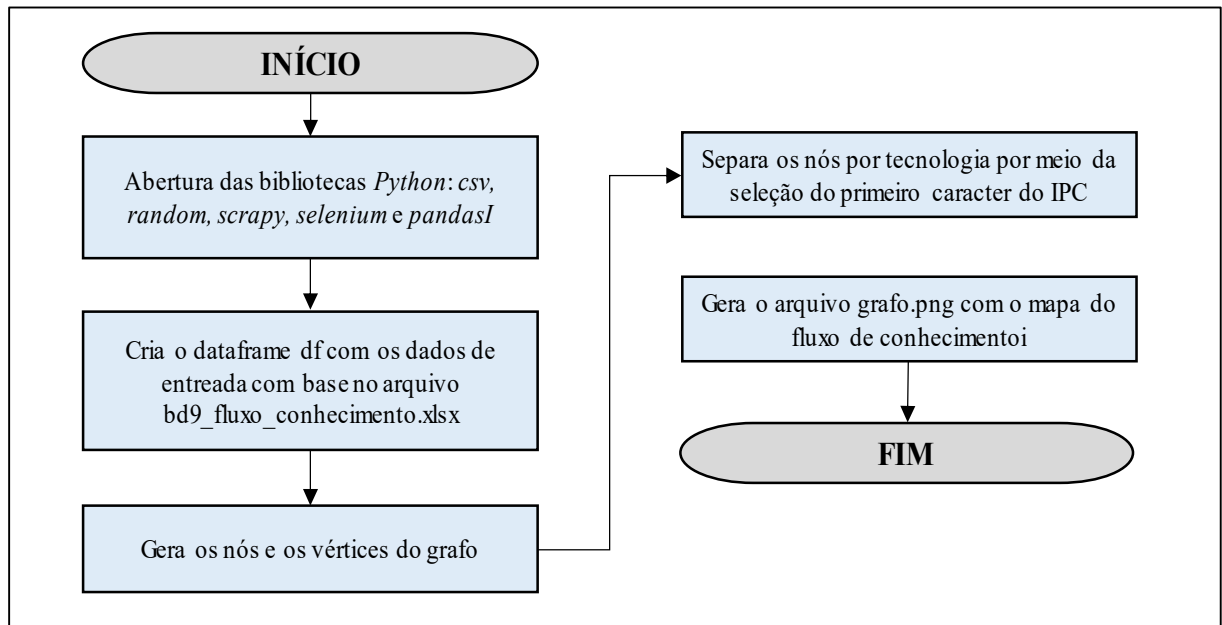
plt.savefig('teste.png', format='png')
plt.show()

```

Fonte: o autor

A Figura 40 mostra o fluxograma explicitando as etapas de processamento do do programa gerar_grafo.ipynb.

Figura 40 - Funcionamento do programa gerar_grafo.ipynb



Fonte: o autor

3 RESULTADOS E DISCUSSÕES

Por meio do software *Power BI* construiu-se *Dashboards* interativos para os bancos de dados gerados pela *web crawler* *wos.py*: a) bd1_dados_basicos; b) bd2_dados_ipc; c) bd3_depositantes; d) bd4_paises; e) bd5_inventores; f) bd6_tabela_IPC; g) bd7_codigo_paises; h) bd8_controle_acesso_base_dados; e i) bd9_calendario. As Figuras 41, 42, 43, 44, 45, 46, 47, 48 e 49 mostram os *Dashboards* criados no software *Power BI* com estes bancos de dados.

Figura 41 - Dashboard Banco de Dados bd1_dados_basicos

bd1_dados_basicos						
Chave Primária	Data Publicação	Nº Derwent	Nº Tecnologia	Patente Principal	Título	Resumo
1-2016-05151Q	21/01/2016	2016-05151Q	1	US2016021178-A1	System of distributed communication of independent autonomous vehicles, has central server to process communication from adjacent ones of set of autonomous vehicles, when error condition is detected in operational mode	Abstract: NOVELTY - The system has an ad-hoc network (112), a wide area network (114) and a centralized communication paths (122) with the set of autonomous vehicles, when an error condition (116) is detected in due to a network failure, a hardware failure, a mechanical failure and an electrical failure. ADVANTAGE - System of distributed communication of independent autonomous vehicles, when error condition is detected, the non-functional vehicle is permitted to enter and safe mode enables other autonomous vehicles and central server to control the vehicle.
1-2019-34548W	18/04/2019	2019-34548W	1	US2019113916-A1	System for driving autonomous vehicle, has processor for accessing safety envelope for vehicle and processing failure in accord with actuator failure in response to determining actuator feedback is divergent from motion control command	Abstract: NOVELTY - The system has a processor for accessing a safety envelope from a sensor (116A). The processor determines the observed path breached from motion control commands in response to determining the safety envelope failure in response to determining the actuator feedback is divergent from the motion control command. ADVANTAGE - The system provides improved diagnostic with external independent information, so that driving speed determines accurately a precise position of the vehicle.
1-2019-37417G	02/05/2019	2019-37417G	1	US2019129431-A1	Autonomous vehicle control system has processor that determines that current location of vehicle is at segment location of one or more segments associated with retrieved feature vector information matched with received feature vector	Abstract: NOVELTY - The autonomous vehicle control system (348) has a processor (348A) that determines that current location of vehicle is at segment location of one or more segments associated with retrieved feature vector information matched with received feature vector. ADVANTAGE - Prevents dangerous mistakes in providing touch input for controlling the vehicle.
1-2018-40363W	24/05/2018	2018-40363W	1	US2018141568-A1	Autonomous vehicle e.g. automobile, has processor generating operational mode for controlling driving function of vehicle and providing operational mode to vehicle control system to	Abstract: NOVELTY - The vehicle (100) has a processor in communication with a driver when the driver is manually driving the vehicle. The processor determines an operational mode for autonomously controlling the vehicle. ADVANTAGE - The vehicle ensures that a driver behavior data structure can be updated.

Fonte: o autor

Figura 42 - Dashboard Banco de Dados bd2_dados_ipc

bd2_dados_ipc					
chave_prim	data_publicacao	num_derwent	num_tecnologia	numero_patente	ipc_principal
5-1977-H6363Y	15/08/1977	1977-H6363Y	5	NL154104-B	A44B
1-1982-J9259E	11/06/1982	1982-J9259E	1	FR2495797-A	G05D
1-1984-252089	07/09/1984	1984-252089	1	FR2541853-A	A01B
1-1985-123364	15/05/1985	1985-123364	1	DE3339894-A	G01C
1-1988-138732	12/01/1988	1988-138732	1	SE8702828-A	G05D
1-1988-352862	07/06/1988	1988-352862	1	SU1400919-A	B60L
1-1988-229675	05/08/1988	1988-229675	1	FR2610427-A	G05D
1-1988-229675	17/08/1988	1988-229675	1	EP278853-A	G05D
1-1988-138732	20/09/1988	1988-138732	1	US4772832-A	G05D
1-1988-229675	01/12/1988	1988-229675	1	ES2003841-A	G05D
1-1989-055751	22/02/1989	1989-055751	1	EP304342-A	B25J
5-1989-145829	05/05/1989	1989-145829	5	WO8903999-A	G01P
5-1989-145829	11/05/1989	1989-145829	5	DE3736294-A	G01P
1-1989-333647	09/11/1989	1989-333647	1	DE3912353-A	G05D
5-1989-349666	29/11/1989	1989-349666	5	EP343579-A	B60R
1-1990-092485	14/03/1990	1990-092485	1	SE8902984-A	G05D
1-1990-092485	22/03/1990	1990-092485	1	DE3930425-A	G05D
5-1990-108705	05/04/1990	1990-108705	5	DE3905677-A	B60R
1-1990-172503	31/05/1990	1990-172503	1	DE3935568-A	B41C
1-1990-172503	20/06/1990	1990-172503	1	GB2225988-A	B41C
5-1989-145829	19/07/1990	1989-145829	5	DE3736294-C	G01P
5-1989-145829	12/09/1990	1989-145829	5	EP386035-A	G01P

Fonte: o autor

Figura 43 - Dashboard Banco de Dados bd3_depositantes

bd3_depositantes				
chave_prim	num_derwent	num_tecnologia	codigo_depositante	depositante
5-1977-H6363Y	1977-H6363Y	5	BULL-Individual	BULLOCK B
1-1982-J9259E	1982-J9259E	1	ONER-C	ONERA OFF NAT ETUD RECH AEROSPATIALE
1-1983-806955	1983-806955	1	AUTO-Soviet Institute	AUTO ELEC EQUIP
1-1983-806955	1983-806955	1	KALU-Soviet Institute	KALUGA AUTO ENG ELE
1-1984-252089	1984-252089	1	DAUD-Individual	DAUDIN J
1-1985-123364	1985-123364	1	VOGE-Individual	VOGEL K H
1-1988-138732	1988-138732	1	TSUC-C	TSUBAKIMOTO CHAIN CO
1-1988-229675	1988-229675	1	PROT-Non-standard	PROTEE GRP INTERET
1-1988-229675	1988-229675	1	PROT-Non-standard	PROTEE GRP INTERET ECONOMIQUE
1-1988-352862	1988-352862	1	BRYA-Soviet Institute	BRYAN MECH ENG WKS
1-1988-352862	1988-352862	1	DIES-Soviet Institute	DIESEL TRAIN RES INST
1-1989-055751	1989-055751	1	WESE-C	WESTINGHOUSE ELECTRIC CORP
5-1989-145829	1989-145829	5	MESR-C	MESSERSCHMITT-BOELKOW-BLOHM GMBH
5-1989-145829	1989-145829	5	MESR-C	MESSERSCHMITT-BOLKOW-BLO
5-1989-145829	1989-145829	5	TELE-C	TEMIC TELEFUNKEN MICROELECTRONIC GMBH
1-1989-333647	1989-333647	1	NSMO-C	NISSAN MOTOR CO LTD
5-1989-349666	1989-349666	5	NPDE-C	NIPPONDENSO CO LTD
5-1989-349666	1989-349666	5	TOYT-C	TOYOTA JIDOSHA KK
1-1990-092485	1990-092485	1	TSUC-C	TSUBAKIMOTO CHAIN CO
5-1990-108705	1990-108705	5	BOSC-C	BOSCH GMBH ROBERT
1-1990-172503	1990-172503	1	DAIT-Non-standard	DAITO DENSHI COLOR
1-1990-306127	1990-306127	1	JUNU-C	JUNGHEINRICH AG
1-1990-306127	1990-306127	1	JUNU-C	JUNGHEINRICH UNTERNEHMENSVERW
1-1990-354706	1990-354706	1	TRAN-Non-standard	TRANSITIONS RES CORP
1-1991-231095	1991-231095	1	DAIS-Non-standard	DAISALUX SA
1-1991-267319	1991-267319	1	SIEI-C	SIEMENS AG
1-1992-017900	1992-017900	1	UYNV-C	STATE UNIV NEW YORK
1-1992-017900	1992-017900	1	UYNV-C	UNIV NEW YORK STATE RES FOUND
1-1992-097003	1992-097003	1	ARCS-Non-standard	ARC SECOND INC
1-1992-097003	1992-097003	1	SPAT-Non-standard	SPATIAL POS SYSTEMS

Fonte: o autor

Figura 44 - Dashboard Banco de Dados bd4_paises

bd4_paises					
chave_prim	num_derwent	num_tecnologia	numero_patente	pais	tipo
1-1982-J9259E	1982-J9259E	1	FR2495797-A	FR	original
1-1983-806955	1983-806955	1	SU987774-A	SU	original
1-1984-252089	1984-252089	1	FR2541853-A	FR	original
1-1985-123364	1985-123364	1	DE3339894-A	DE	original
1-1988-138732	1988-138732	1	SE503584-C2	SE	original
1-1988-138732	1988-138732	1	SE8702828-A	SE	original
1-1988-138732	1988-138732	1	US4772832-A	US	original
1-1988-229675	1988-229675	1	CA1326896-C	CA	original
1-1988-229675	1988-229675	1	DE3889474-G	DE	original
1-1988-229675	1988-229675	1	EP278853-A	AT	regional
1-1988-229675	1988-229675	1	EP278853-A	BE	regional
1-1988-229675	1988-229675	1	EP278853-A	CH	regional
1-1988-229675	1988-229675	1	EP278853-A	DE	regional
1-1988-229675	1988-229675	1	EP278853-A	EP	original
1-1988-229675	1988-229675	1	EP278853-A	ES	regional
1-1988-229675	1988-229675	1	EP278853-A	FR	regional
1-1988-229675	1988-229675	1	EP278853-A	GB	regional
1-1988-229675	1988-229675	1	EP278853-A	GR	regional
1-1988-229675	1988-229675	1	EP278853-A	IT	regional
1-1988-229675	1988-229675	1	EP278853-A	LI	regional
1-1988-229675	1988-229675	1	EP278853-A	LU	regional
1-1988-229675	1988-229675	1	EP278853-A	NL	regional
1-1988-229675	1988-229675	1	EP278853-A	SE	regional
1-1988-229675	1988-229675	1	EP278853-B1	AT	regional
1-1988-229675	1988-229675	1	EP278853-B1	BE	regional
1-1988-229675	1988-229675	1	EP278853-B1	CH	regional
1-1988-229675	1988-229675	1	EP278853-B1	DE	regional
1-1988-229675	1988-229675	1	EP278853-B1	EP	original
1-1988-229675	1988-229675	1	EP278853-B1	ES	regional
1-1988-229675	1988-229675	1	EP278853-B1	FR	regional

Fonte: o autor

Figura 45 - Dashboard Banco de Dados bd5_inventores

bd5_inventores			
chave_prim	num_derwent	num_tecnologia	inventor
1-1982-19259E	1982-19259E	1	FUERTE S V
1-1983-806955	1983-806955	1	AKIMOV V M
1-1983-806955	1983-806955	1	BOT R A
1-1983-806955	1983-806955	1	KONDRATEV V N
1-1984-252089	1984-252089	1	DAUDIN J L
1-1984-252089	1984-252089	1	MONTIGNY G
1-1985-123364	1985-123364	1	VOGEL K H
1-1988-138732	1988-138732	1	OKAZAKI M
1-1988-138732	1988-138732	1	SUDARE M
1-1988-138732	1988-138732	1	TOMIKAWA H
1-1988-229675	1988-229675	1	LEMERCIER P
1-1988-229675	1988-229675	1	TRYSTRAM P
1-1988-352862	1988-352862	1	GRINEVICH V P
1-1988-352862	1988-352862	1	SHARDENKOV E D
1-1989-055751	1989-055751	1	BARRY R F
1-1989-333647	1989-333647	1	HATTORI A
1-1990-092485	1990-092485	1	MINAMI H
1-1990-172503	1990-172503	1	AKOYA K
1-1990-306127	1990-306127	1	CHRISTENSE S
1-1990-306127	1990-306127	1	CHRISTENSEN S
1-1990-306127	1990-306127	1	HAELBICH R
1-1990-306127	1990-306127	1	HAELBICH R P
1-1990-306127	1990-306127	1	HAENEL J
1-1990-306127	1990-306127	1	KRAMER D
1-1990-306127	1990-306127	1	KRIEBISCH T
1-1990-306127	1990-306127	1	KRUEGER B
1-1990-306127	1990-306127	1	PRIDAT V
1-1990-306127	1990-306127	1	SOLSCHEID D
1-1990-306127	1990-306127	1	SOLSCHEID-KOTTMANN D
1-1990-306127	1990-306127	1	THOMANEK G

Fonte: o autor

Figura 46 - Dashboard Banco de Dados bd6_tabela_IPC

bd6_tabela_IPC				
codigo_ipc	descricao_port	descricao_ingl	Setor Industrial	Sub-Sector Industrial
A	NECESSIDADES HUMANAS	HUMAN NECESSITIES		
A01	AGRICULTURA; SILVICULTURA; PECUÁRIA; CAÇA; CAPTURA EM ARMADILHAS; PESCA	AGRICULTURE; FORESTRY; ANIMAL HUSBANDRY; HUNTING; TRAPPING; FISHING		
A01B	TRABALHO DO SOLO EM AGRICULTURA OU SILVICULTURA; PEÇAS, DETALHES OU ACESSÓRIOS DE MÁQUINAS OU IMPLEMENTOS AGRÍCOLAS, EM GERAL (escavação ou cobertura de sulcos ou covas para semeadura, plantio ou adubagem A01C 5/00; máquinas para colheita A01D; aparelhos ceifadeiras conversíveis para trabalhar o solo ou capazes de trabalhar o solo A01D 42/04; ceifadeiras combinados com implementos para tratamento do solo A01D 43/12; trabalho do solo para finalidades de engenharia E01, E02, E21)	SOIL WORKING IN AGRICULTURE OR FORESTRY; PARTS, DETAILS, OR ACCESSORIES OF AGRICULTURAL MACHINES OR IMPLEMENTS, IN GENERAL (making or covering furrows or holes for sowing, planting or manuring A01C 5/00; machines for harvesting root crops A01D; mowers convertible to soil working apparatus or capable of soil working A01D 42/04; mowers combined with soil working implements A01D 43/12; soil working for engineering purposes E01, E02, E21)	Mechanical engineering	Other special machines
A01C	PLANTIO; SEMEADURA; FERTILIZAÇÃO (combinados com o trabalho do solo em geral A01B 49/04; peças, detalhes ou acessórios de máquinas agrícolas ou implementos em geral A01B 51/00-A01B 75/00)	PLANTING; SOWING; FERTILISING (combined with general working of soil A01B 49/04; parts, details or accessories of agricultural machines or implements, in general A01B 51/00-A01B 75/00)	Mechanical engineering	Other special machines
A01D	COLHEITA; CEIFA	HARVESTING; MOWING	Mechanical engineering	Other special machines
A01F	PROCESSAMENTO DO PRODUTO COLHIDO (máquinas combinadas A01D 41/00); ENFARDAMENTO DE PALHA, FENO OU SIMILARES; APARELHO ESTACIONÁRIO OU INSTRUMENTOS MANUAIS PARA FORMAÇÃO OU ENFEIXAMENTO DE PALHA, FENO OU SIMILARES EM MEDAS; CORTE DE PALHA, FENO OU SIMILARES; ARMAZENAMENTO DE PRODUTOS AGRÍCOLAS OU HORTÍCOLAS (disposições para formar ou assentar pilhas em conjunto com colheita A01D 85/00)	THRESHING (combines A01D 41/00); BALING OF STRAW, HAY OR THE LIKE; STATIONARY APPARATUS OR HAND TOOLS FOR FORMING OR BINDING STRAW, HAY OR THE LIKE INTO BUNDLES; CUTTING OF STRAW, HAY OR THE LIKE; STORING AGRICULTURAL OR HORTICULTURAL PRODUCE (arrangements for making or setting stacks in connection with harvesting A01D 85/00)	Mechanical engineering	Other special machines
A01G	HORTICULTURA; CULTIVO DE VEGETAIS, FLORES, ARROZ, FRUTAS, VINHAS, LÚPULOS OU ALGAS; SILVICULTURA; IRRIGAÇÃO (coleta de frutas, vegetais, lúpulos ou similares A01D 46/00; propagação de algas unicelulares C12N 1/12)	HORTICULTURE; CULTIVATION OF VEGETABLES, FLOWERS, RICE, FRUIT, VINES, HOPS OR SEAWEED; FORESTRY; WATERING (picking of fruits, vegetables, hops or the like A01D 46/00; propagating unicellular algae C12N 1/12)	Mechanical engineering	Other special machines
A01H	NOVAS PLANTAS OU PROCESSOS PARA OBTENÇÃO DAS MESMAS; REPRODUÇÃO DE PLANTAS POR MEIO DE TÉCNICAS DE CULTURA DE TECIDOS [5]	NEW PLANTS OR PROCESSES FOR OBTAINING THEM; PLANT REPRODUCTION BY TISSUE CULTURE TECHNIQUES [5]	Chemistry	Food chemistry
A01J	MANUFATURA DE PRODUTOS LÁCTEOS (conservação, pasteurização, esterilização de produtos lácteos A23; para assuntos químicos, ver subclasse A23C)	MANUFACTURE OF DAIRY PRODUCTS (preservation, pasteurisation, sterilisation of milk products A23; for chemical matters, see subclass A23C)	Mechanical engineering	Other special machines

Fonte: o autor

Figura 47 - Dashboard Banco de Dados bd7_codigo_paises

bd7_codigo_paises

Code	Name
AD	Andorra
AE	United Arab Emirates
AF	Afghanistan
AG	Antigua and Barbuda
AI	Anguilla
AL	Albania
AM	Armenia
AO	Angola
AQ	Antarctica
AR	Argentina
AS	American Samoa
AT	Austria
AU	Australia
AW	Aruba
AX	Åland Islands
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BD	Bangladesh
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BH	Bahrain
BI	Burundi
BJ	Benin
BL	Saint Barthélemy
BM	Bermuda
BN	Brunei Darussalam
BO	Bolivia, Plurinational State of
BQ	Bonaire, Sint Eustatius and Saba

Fonte: o autor

Figura 48 - Dashboard Banco de Dados bd8_controle_acesso

bd8_controle_acesso

num_tecnologia	tecnologia	string_de_busca	Dia	Mês	Ano	num_patentes	num_registros
1	Autonomous Vehicles	"autonomous vehicle*" or "autonomous car"	7	junho	2020	16.579	6.545
2	ADAS	"advanced driver assistance system"	25	junho	2020	1.866	701
3	Smart City	"smart city" or "smart cities"	7	junho	2020	2.082	858
4	Smart Mobility	"smart mobility" or "urban mobility"	7	junho	2020	102	54
5	Vehicular Safety	"vehicular safety"	25	junho	2020	487	187
6	Infotainment	"infotainment system"	25	junho	2020	4.122	1.458

Fonte: o autor

Figura 49 - Dashboard Banco de Dados bd9_calendario

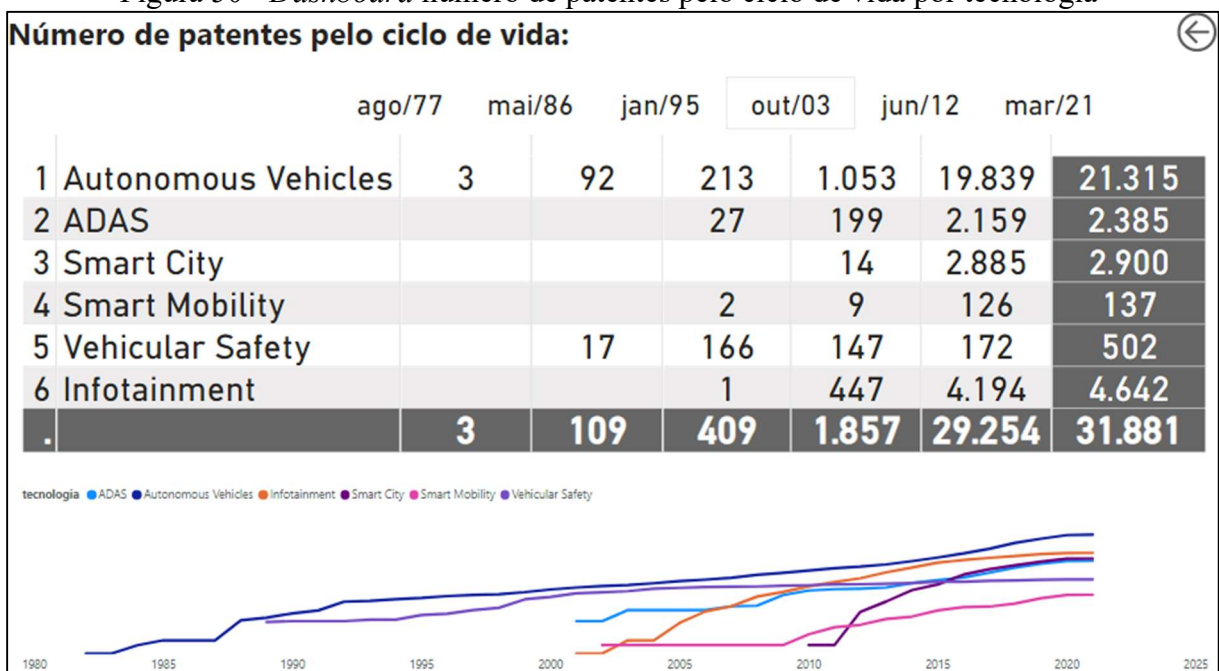
bd9_calendario

Ano	Ano_Mes	Mes_Nome	Mes_Num	Semana	Trimestre
1977	1977-01	jan	1	01	1
1977	1977-01	jan	1	02	1
1977	1977-01	jan	1	03	1
1977	1977-01	jan	1	04	1
1977	1977-01	jan	1	05	1
1977	1977-01	jan	1	06	1
1977	1977-02	fev	2	06	1
1977	1977-02	fev	2	07	1
1977	1977-02	fev	2	08	1
1977	1977-02	fev	2	09	1
1977	1977-02	fev	2	10	1
1977	1977-03	mar	3	10	1
1977	1977-03	mar	3	11	1
1977	1977-03	mar	3	12	1
1977	1977-03	mar	3	13	1
1977	1977-03	mar	3	14	1
1977	1977-04	abr	4	14	2
1977	1977-04	abr	4	15	2
1977	1977-04	abr	4	16	2
1977	1977-04	abr	4	17	2
1977	1977-04	abr	4	18	2
1977	1977-05	mai	5	19	2
1977	1977-05	mai	5	20	2
1977	1977-05	mai	5	21	2
1977	1977-05	mai	5	22	2
1977	1977-05	mai	5	23	2
1977	1977-06	jun	6	23	2
1977	1977-06	jun	6	24	2
1977	1977-06	jun	6	25	2

Fonte: o autor

Por meio do software *Power BI* construiu-se *Dashboards* interativos com mostram as seguintes métricas e indicadores: a) Quantidade de Patentes; b) Ciclo de vida; c) Número de vezes nas quais as patentes foram citadas; d) Número de países nos quais as patentes foram registradas; e) Quantidade de áreas que as patentes afetaram (IPC); e f) Quantidade de proponentes envolvidos com as patentes. As Figuras 50, 51, 52, 53 e 54 mostram os *Dashboards* criados no software *Power BI* com as métricas e indicadores de maturidade.

Figura 50 - Dashboard número de patentes pelo ciclo de vida por tecnologia



Fonte: o autor

Figura 51 - *Dashboard* número de citações das patentes pelo ciclo de vida por tecnologia

Número de citações das patentes pelo ciclo de vida:						
	ago/77	mai/86	jan/95	out/03	jun/12	mar/21
1 Autonomous Vehicles	19	1.457	3.219	3.740	11.668	
2 ADAS			223	666	810	
3 Smart City				21	947	
4 Smart Mobility			5	43	22	
5 Vehicular Safety	0	105	1.425	377	230	
6 Infotainment			0	721	1.992	
.	19	1.562	4.872	5.568	15.669	

Fonte: o autor

Figura 52 - *Dashboard* número de países com patentes registradas por tecnologia

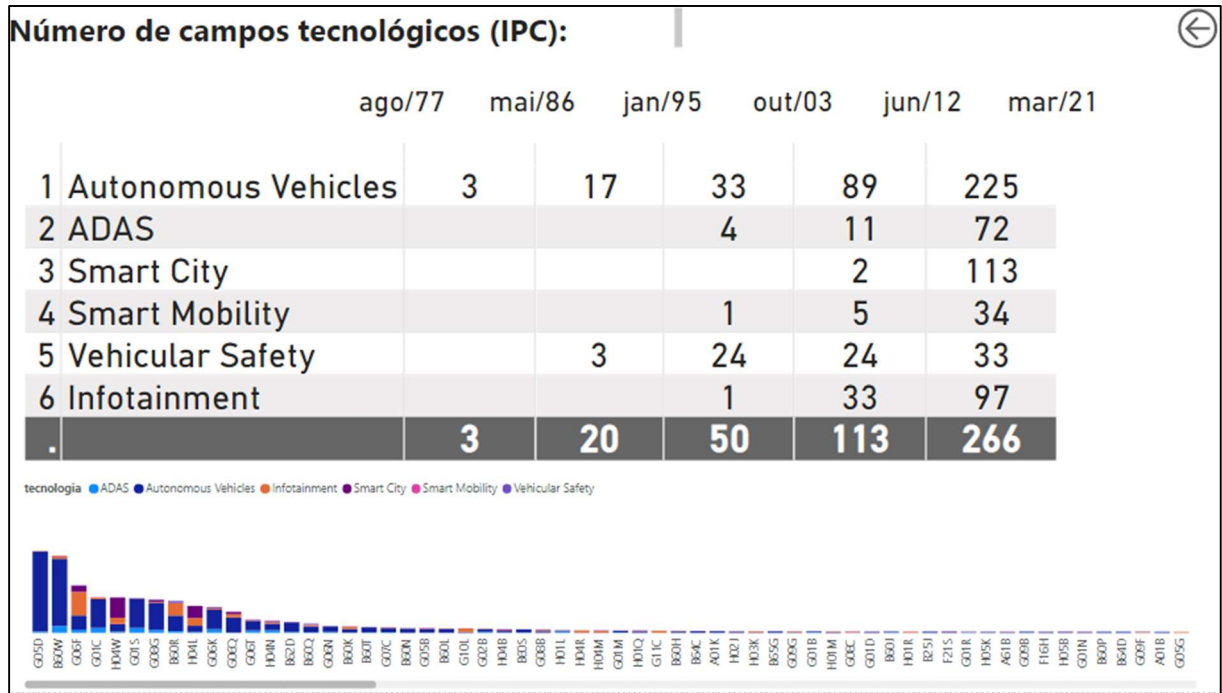
Número de Países com Patentes Registradas:						
	ago/77	mai/86	jan/95	out/03	jun/12	mar/21
1 Autonomous Vehicles	3	15	15	30	38	
2 ADAS			6	15	28	
3 Smart City				4	25	
4 Smart Mobility			1	4	16	
5 Vehicular Safety	1	5	18	14	12	
6 Infotainment			1	20	27	
.	4	15	19	31	41	

tecnologia ● ADAS ● Autono... ● Infotain... ● Smart City ● Smart Mobility ● Vehicular S...

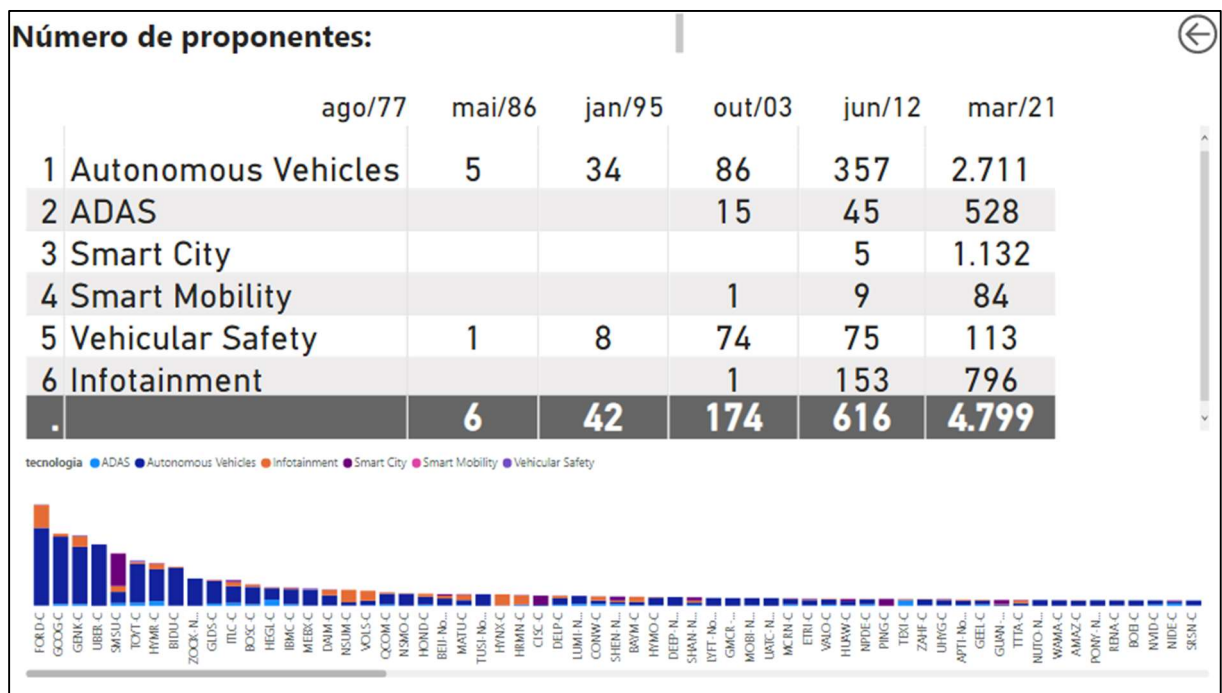
Bing

© 2021 TomTom. © 2021 Microsoft Corporation. Todos os direitos reservados.

Fonte: o autor

Figura 53 - *Dashboard* número de campos tecnológicos (IPC) por tecnologia

Fonte: o autor

Figura 54 - *Dashboard* número de proponentes por tecnologia

Fonte: o autor

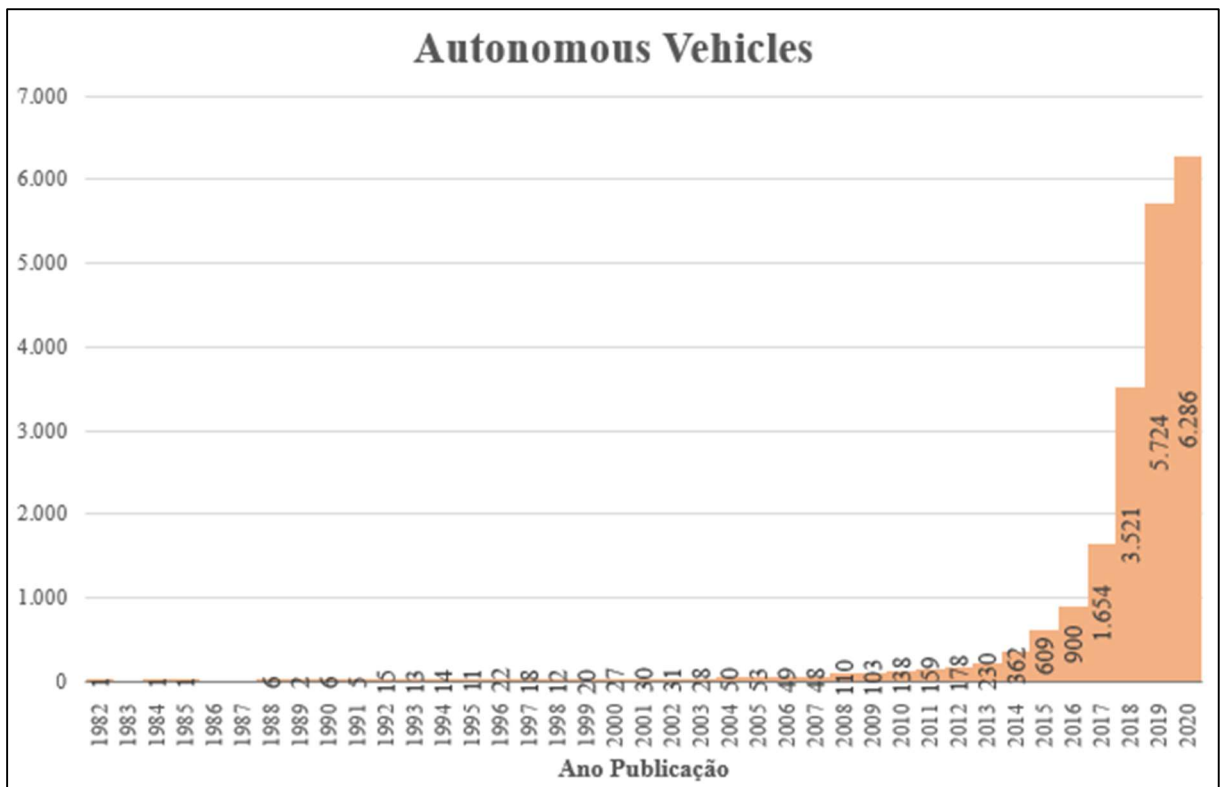
Por meio do *software Microsoft Excel for Office 365 MSO* construiu-se gráficos dinâmicos, para cada tecnologia pesquisada, que mostram as seguintes curvas: a) Quantidade de Patentes por Ano de Publicação; b) Proporção de Patentes por Ano de Publicação por Setor Industrial.

Nas Figuras 55, 57, 59, 61 e 65, referentes à Quantidade de Patentes por Ano de Publicação para as tecnologias *Autonomous Vehicles*, *Advanced Driver Assistance Systems ADAS*, *Smart City*, *Smart Mobility* e *Infotainment System*, observou-se um curva modal semelhante à Curva da Difusão da Inovação proposta por Moore (2014). Na Figura 73, referente à Quantidade de Patentes por Ano de Publicação para a tecnologia *Vehicular Safety*, observou-se uma curva bimodal, podendo caracterizar dois ciclos de inovação, sendo que o primeiro ciclo pode ser associado ao modelo *science-push* e o segundo ciclo ao modelo *market-pull*, conforme proposto por Schmoch (2007).

Nas Figuras 56, 58, 60, 62, 64 e 66, referentes à Proporção de Patentes por Ano de Publicação por Setor Industrial, observou-se, para todas as tecnologias, a presença de pelo menos quatro setores industriais envolvidos, caracterizando assim o transbordamento de conhecimento destas tecnologias pelas áreas tecnológicas da indústria identificadas por Schmoch (2008).

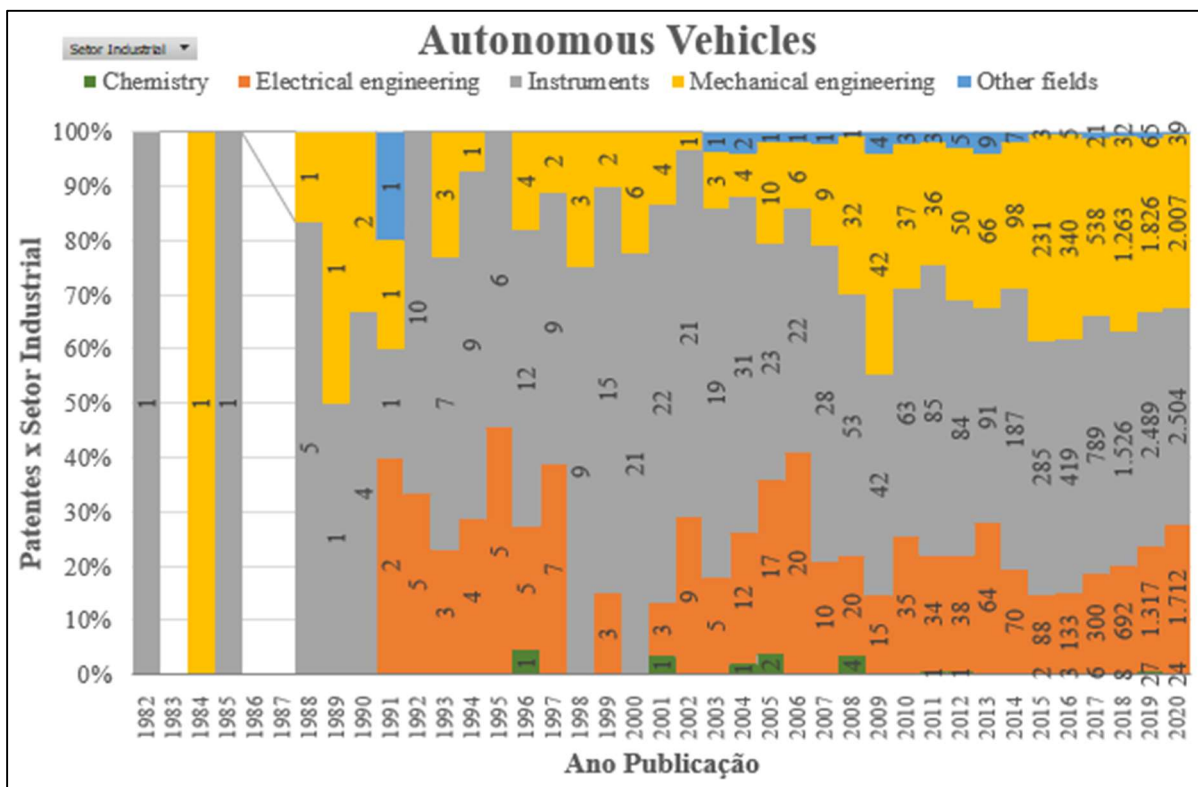
As Figuras 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65 e 66 mostram os gráficos dinâmicos criados no *software Microsoft Excel for Office 365 MSO*.

Figura 55 – Quantidade de Patentes por Ano de Publicação *Autonomous Vehicles*



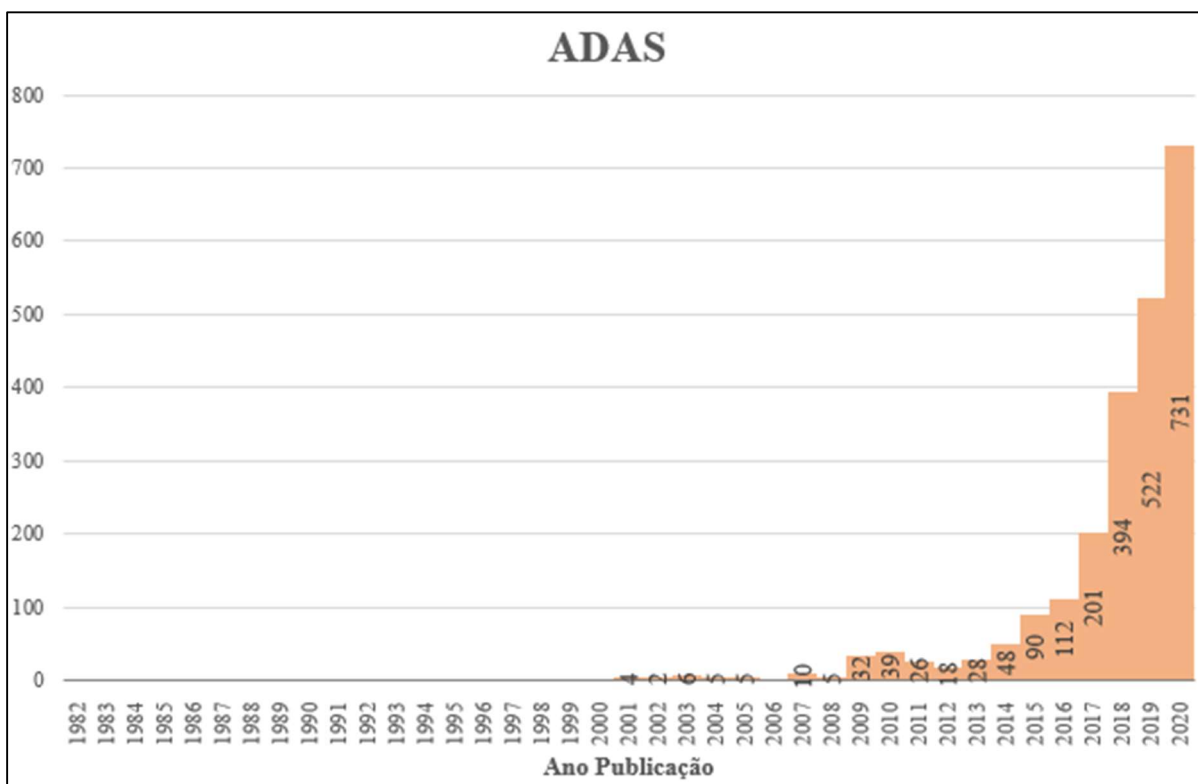
Fonte: o autor

Figura 56 – Proporção de Patentes por Ano de Publicação por Setor Industrial *Autonomous Vehicles*



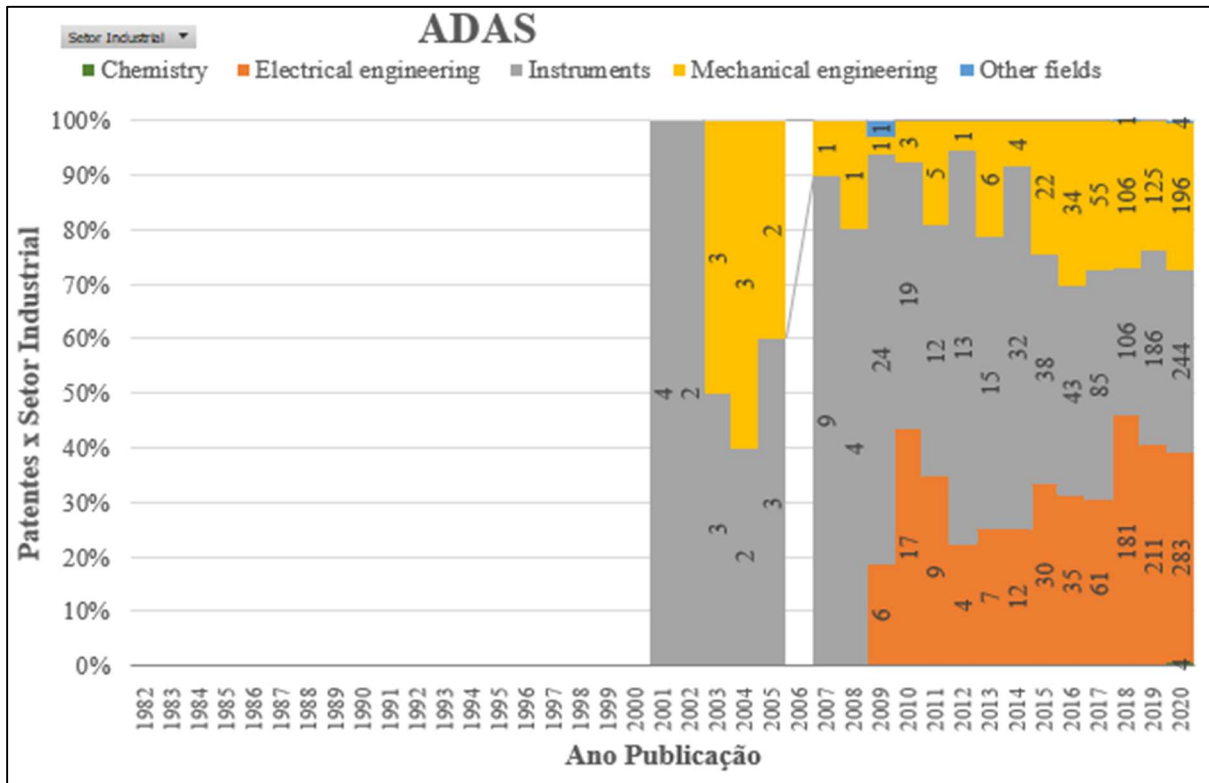
Fonte: o autor

Figura 57 – Quantidade de Patentes por Ano de Publicação *Advanced Driver Assistance Systems (ADAS)*



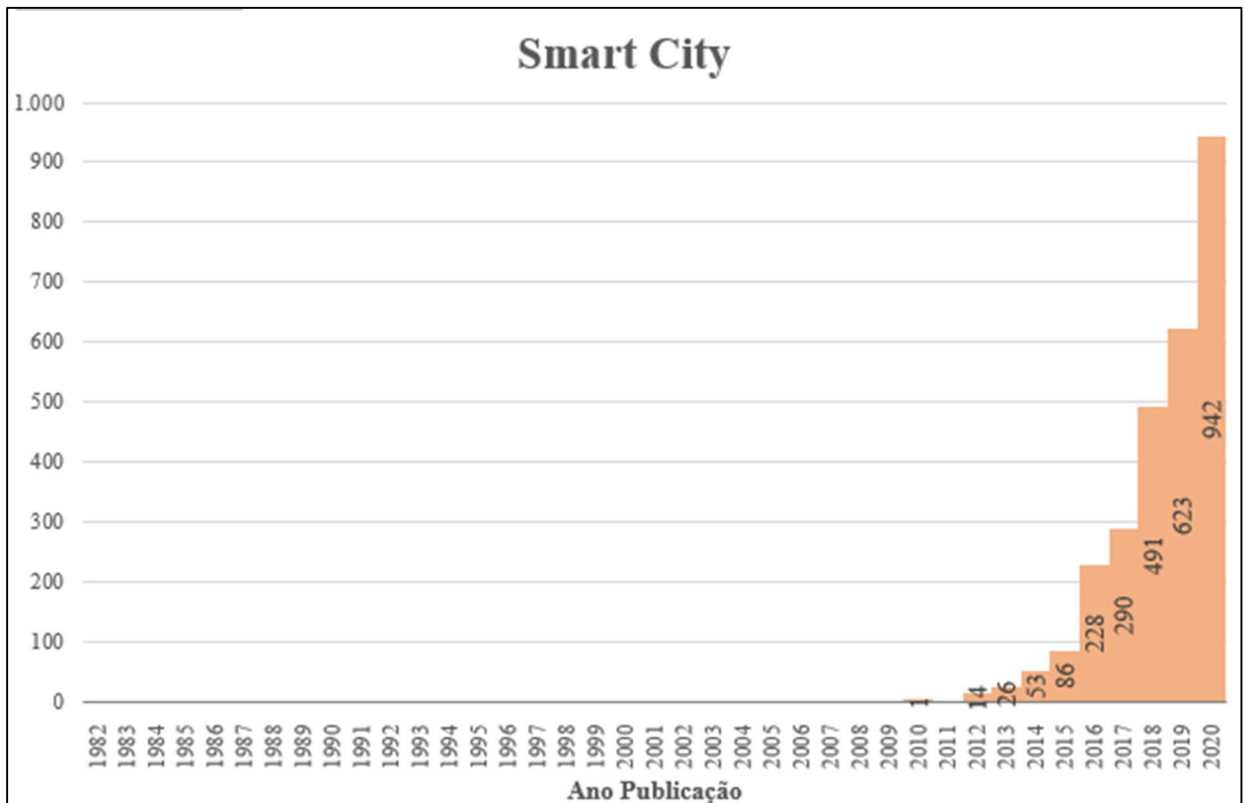
Fonte: o autor

Figura 58 - Proporção de Patentes por Ano de Publicação por Setor Industrial *Advanced Driver Assistance Systems (ADAS)*



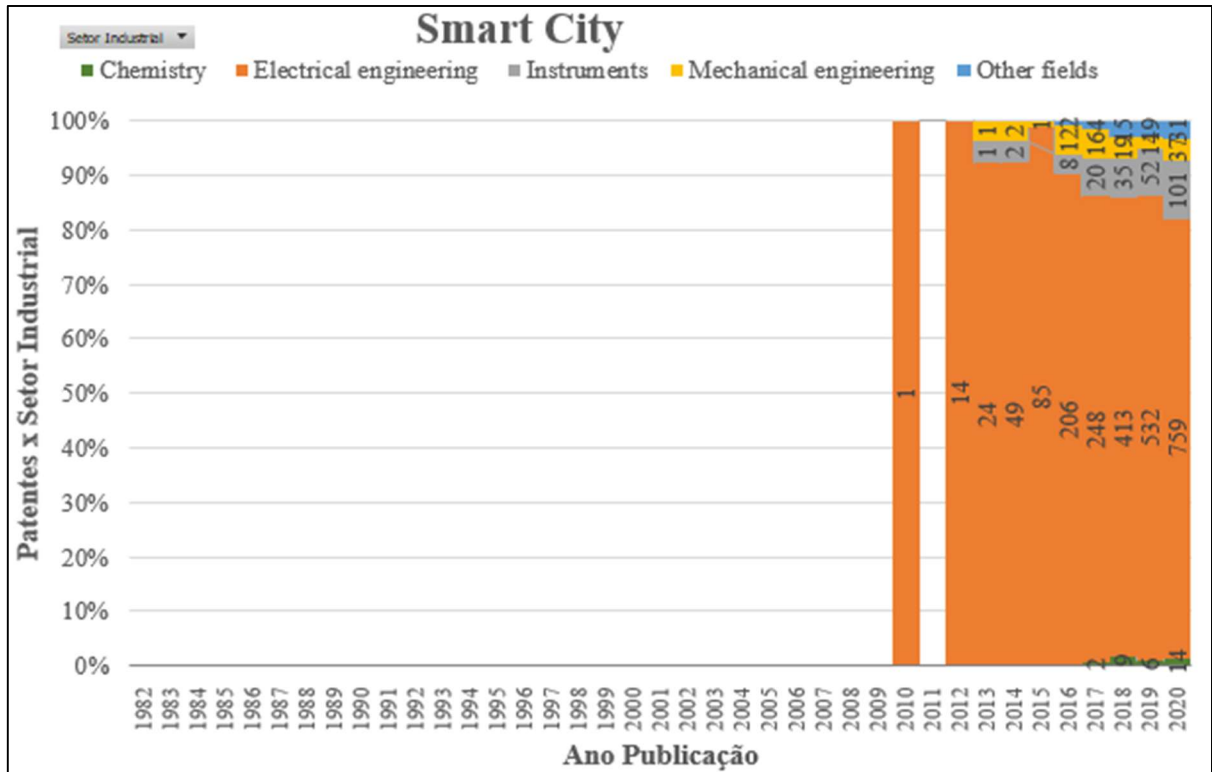
Fonte: o autor

Figura 59 – Quantidade de Patentes por Ano de Publicação *Smart City*



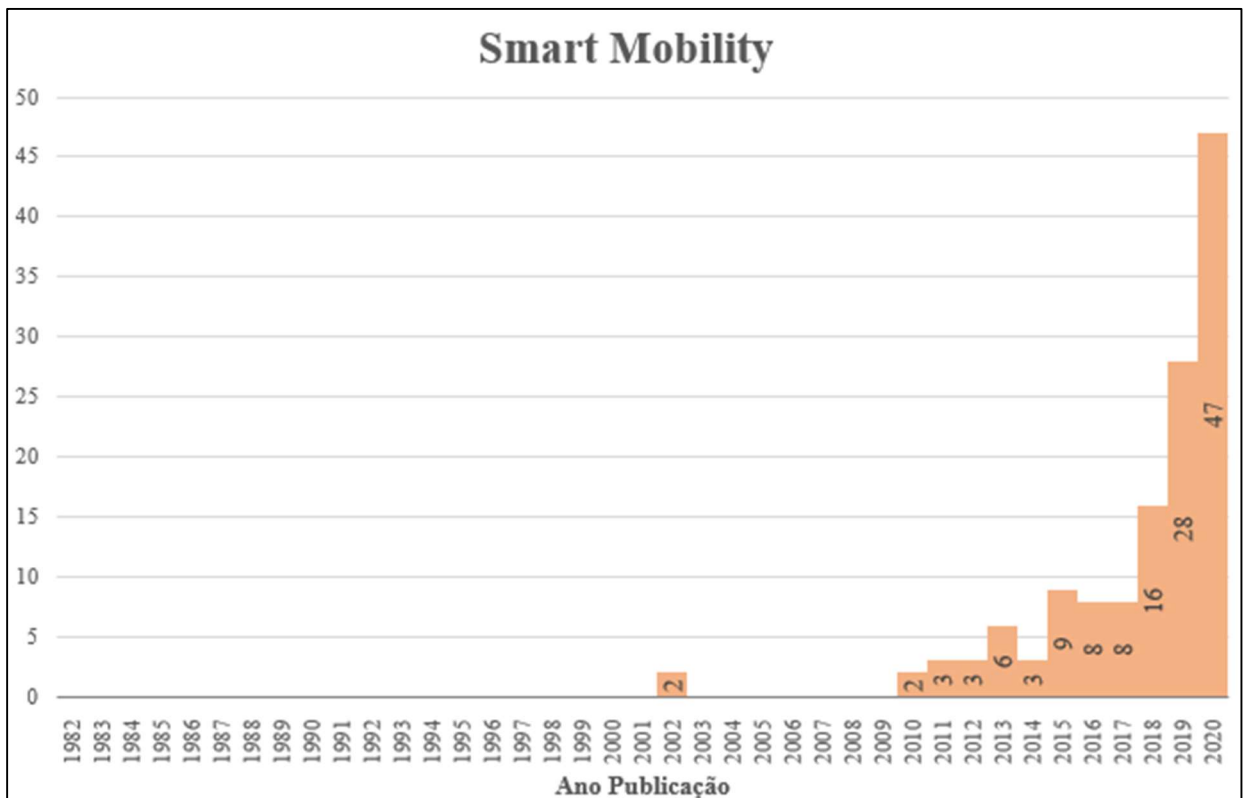
Fonte: o autor

Figura 60 – Proporção de Patentes por Ano de Publicação por Setor Industrial *Smart City*



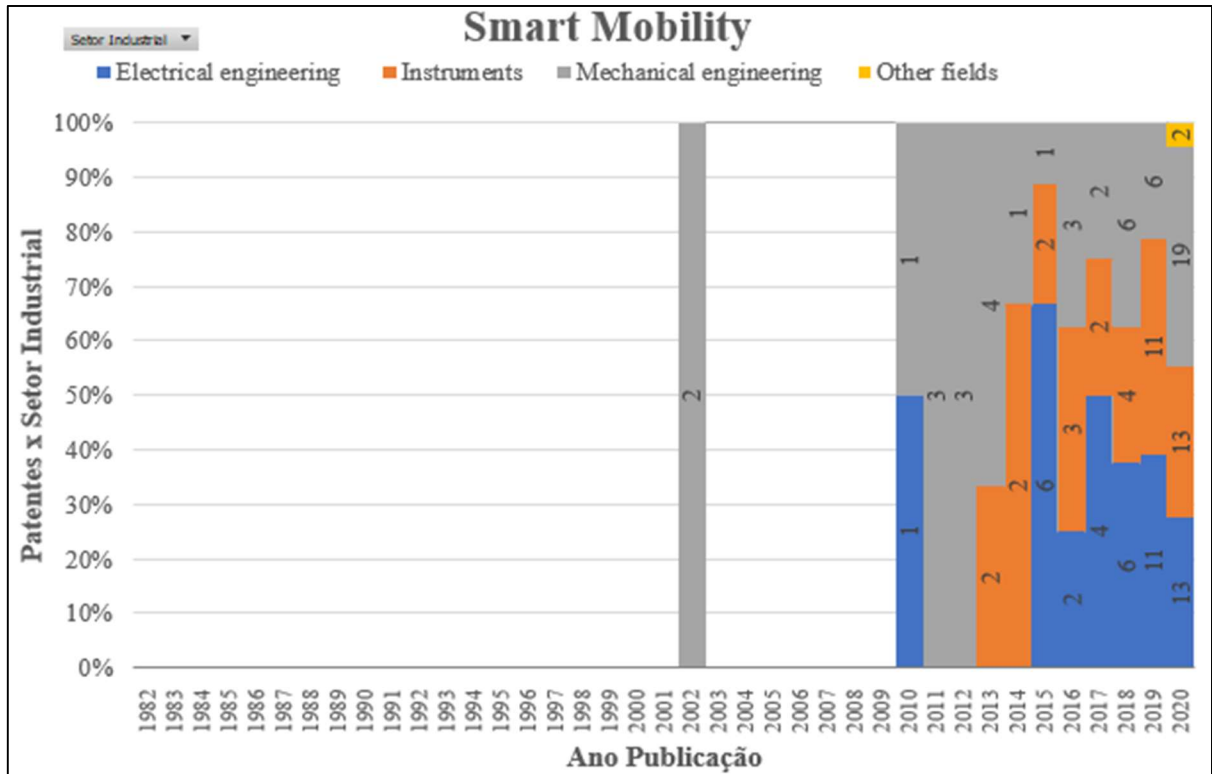
Fonte: o autor

Figura 61 – Quantidade de Patentes por Ano de Publicação *Smart Mobility*



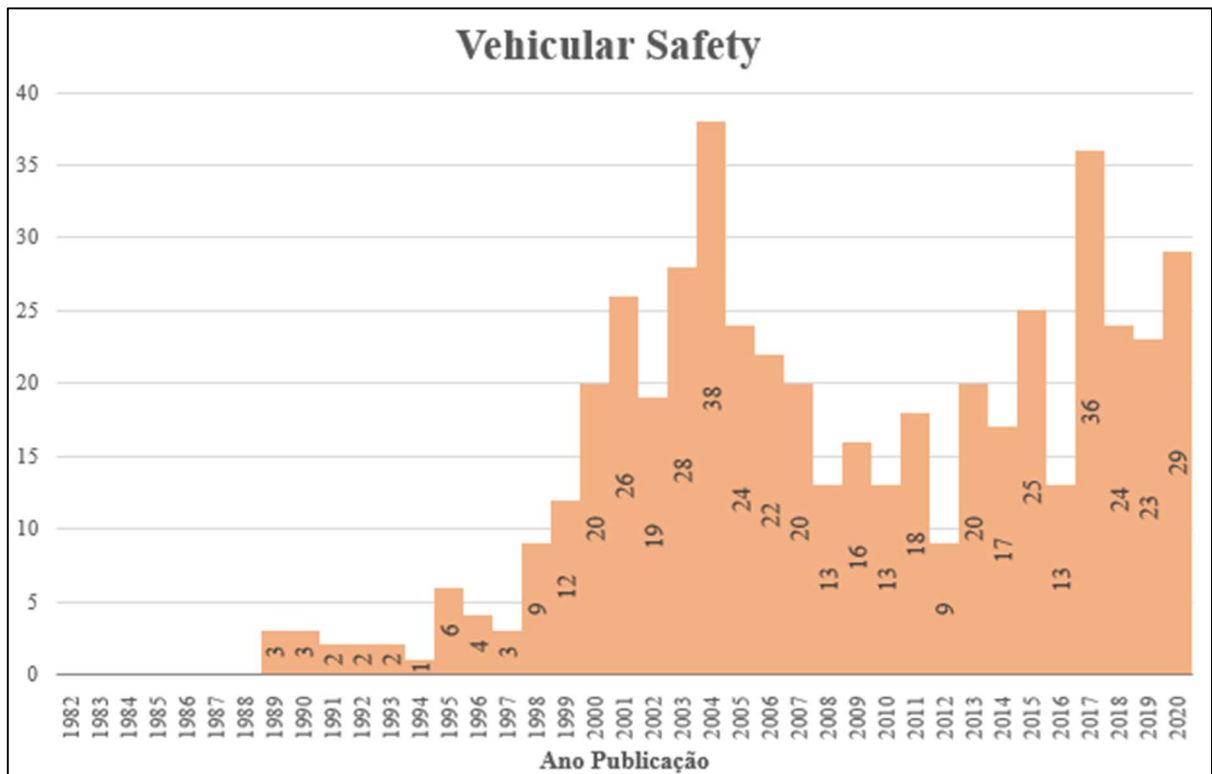
Fonte: o autor

Figura 62 – Proporção de Patentes por Ano de Publicação por Setor Industrial *Smart Mobility*



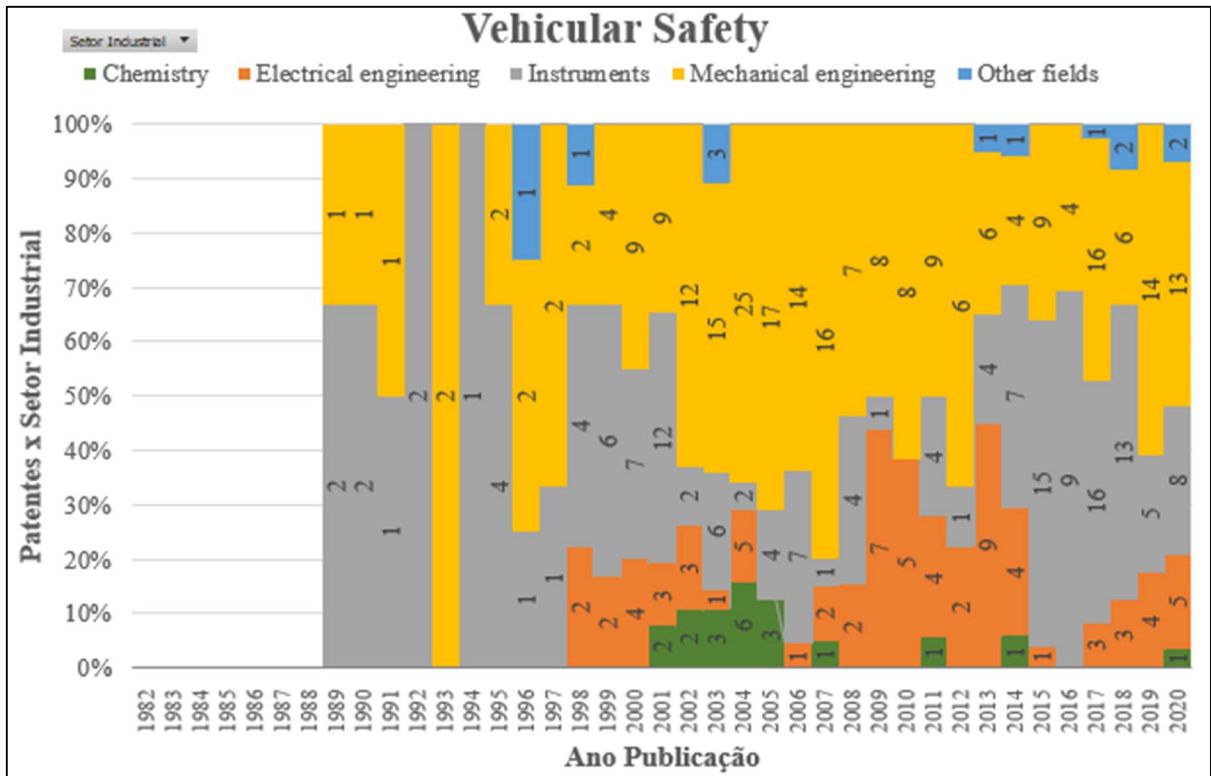
Fonte: o autor

Figura 63 – Quantidade de Patentes por Ano de Publicação *Vehicle Safety*



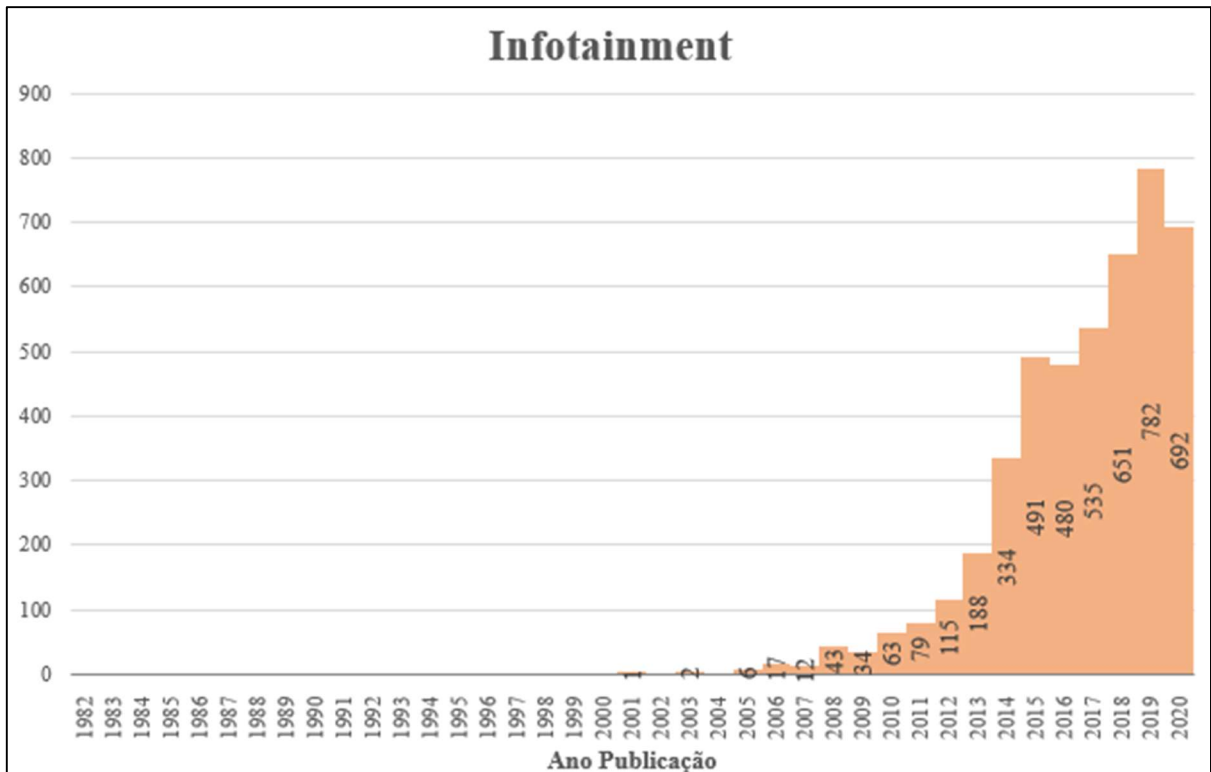
Fonte: o autor

Figura 64 – Proporção de Patentes por Ano de Publicação por Setor Industrial *Vehicular Safety*



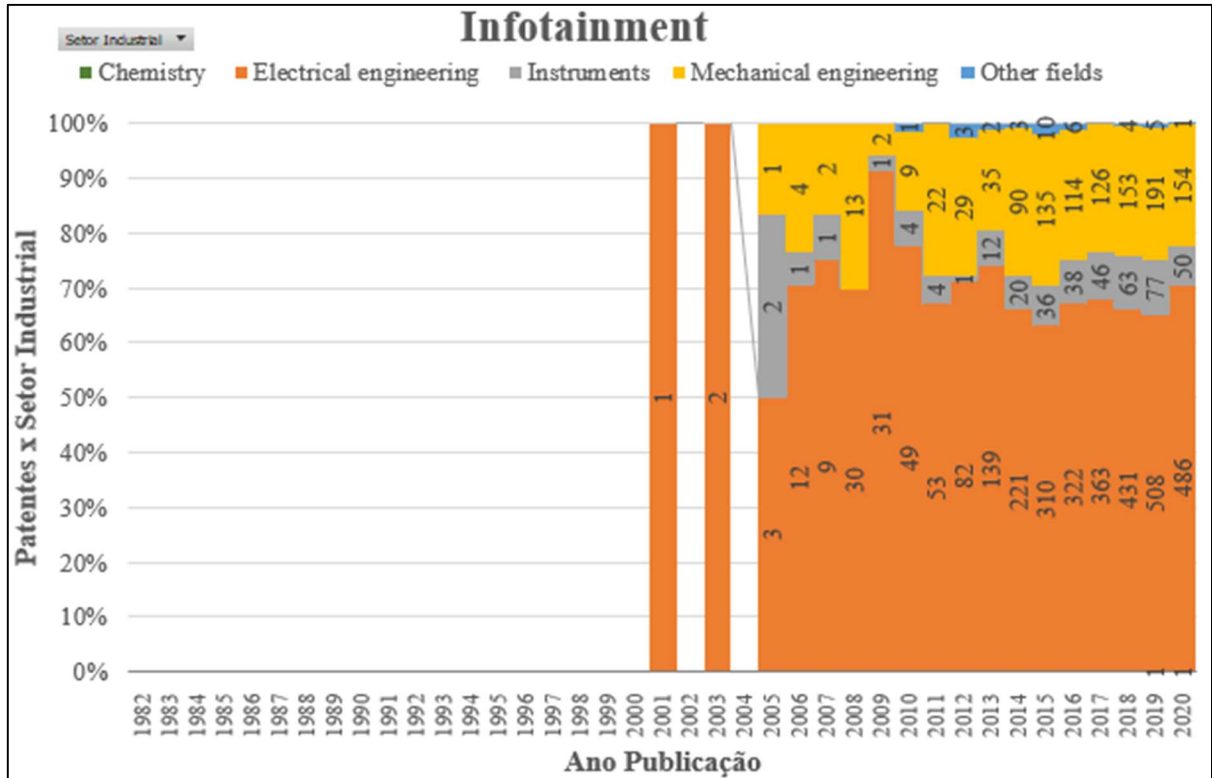
Fonte: o autor

Figura 65 – Quantidade de Patentes por Ano de Publicação *Infotainment System*



Fonte: o autor

Figura 66 – Proporção de Patentes por Ano de Publicação por Setor Industrial *Infotainment System*



Fonte: o autor

A partir dos programas *fluxo_spider.ipynb* e *gerar_grafo.ipynb* gerou-se dois mapas de fluxo de conhecimento considerando a tecnologia *Smart Mobility*: a) Fluxo de conhecimento para as patentes a partir do número *Derwent* 2002-131700, tecnologia *Smart Mobility*, data de pesquisa 26/02/2021, com ciclos de execução até números de patentes citantes serem igual a zero; e b) Fluxo de Conhecimento Total, tecnologia *Smart Mobility*, data de pesquisa 26/02/2021, com ciclos de execução até números de patentes citantes serem igual a zero.

No mapa criado a partir das citações da patente com o número *Derwent* 2002-131700, composto por 61 números *Derwent*, observou-se a existência de três dos oito campos tecnológicos listados pela WIPO: a) B = Operações de Processamento, Transporte (75,4 %); b) H = Eletricidade (19,7 %); e c) G = Física (4,9 %).

No mapa criado a partir das citações de todas as patentes da tecnologia *Smart Mobility*, composto por 260 números *Derwent*, observou-se a existência de seis dos oito campos tecnológicos listados pela WIPO: B = Operações de Processamento, Transporte (53,8 %); b) G = Física (18,8 %); c) H = Eletricidade (8,5 %); d) E = Construções Fixas (7,7 %); e) F =

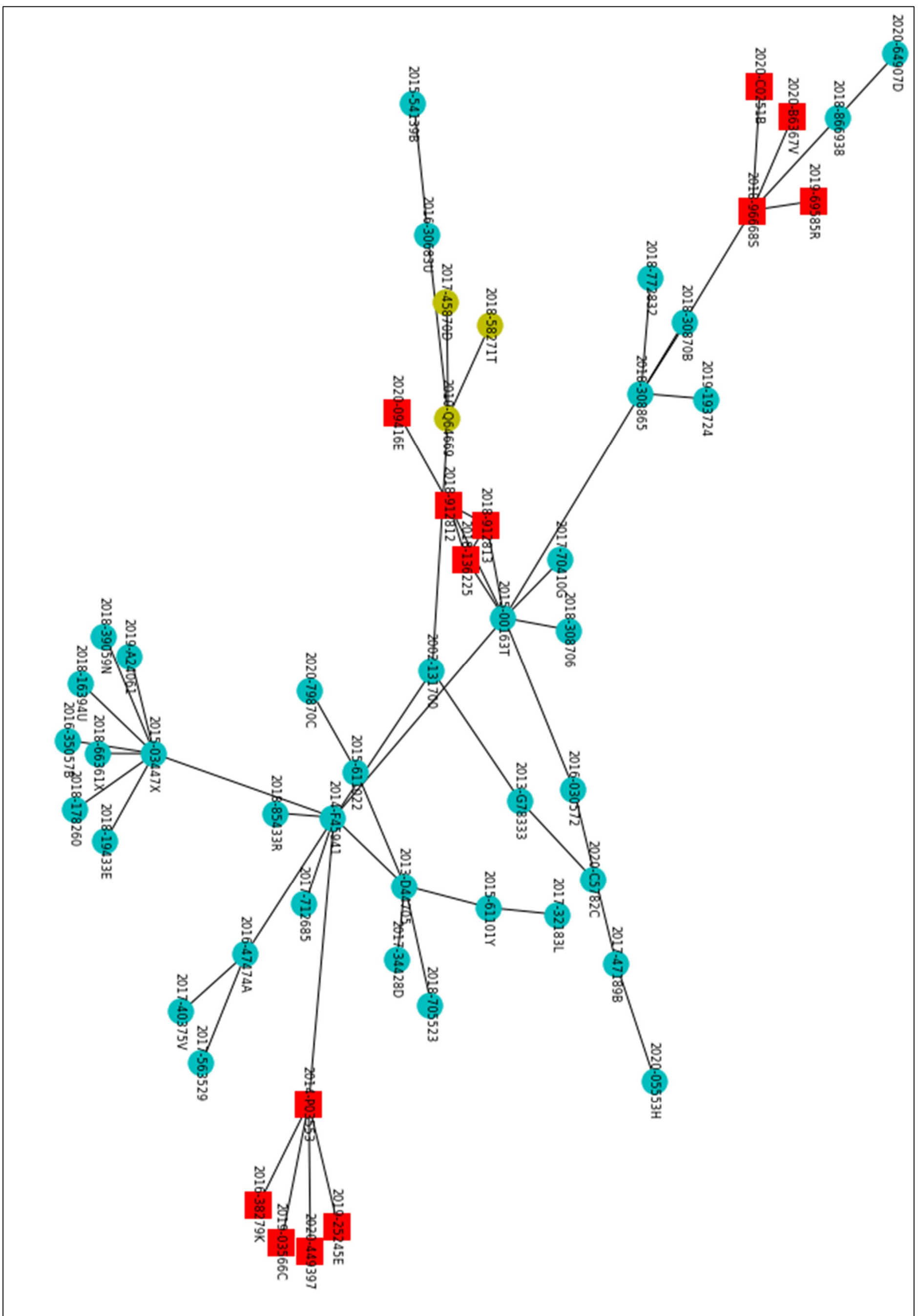
Engenharia Mecânica, Iluminação, Aquecimento, Armas, Explosão (7,3 %); e f) A = Necessidades Humanas (3,8 %).

Neste trabalho limitou-se a construção de mapas de fluxo de conhecimento à tecnologia *Smart Mobility*, em função da quantidade de documentos de patentes com citações presentes nas demais tecnologias objetos deste estudo *versus* a capacidade de processamento do computador usado na extração dos dados.

A Figura 67 mostra o mapa do fluxo de conhecimento das patentes do número *Derwent* 2002-131700, tecnologia *Smart Mobility*.

As Figuras 68 e 69 mostram o mapa do fluxo de conhecimento das patentes da tecnologia *Smart Mobility*.

Figura 67 - Fluxo de conhecimento número Derwent 2002-131700 tecnologia Smart Mobility



Fonte: o autor

Figura 68 - Fluxo de conhecimento das patentes da tecnologia *Smart Mobility* (fase inicial)

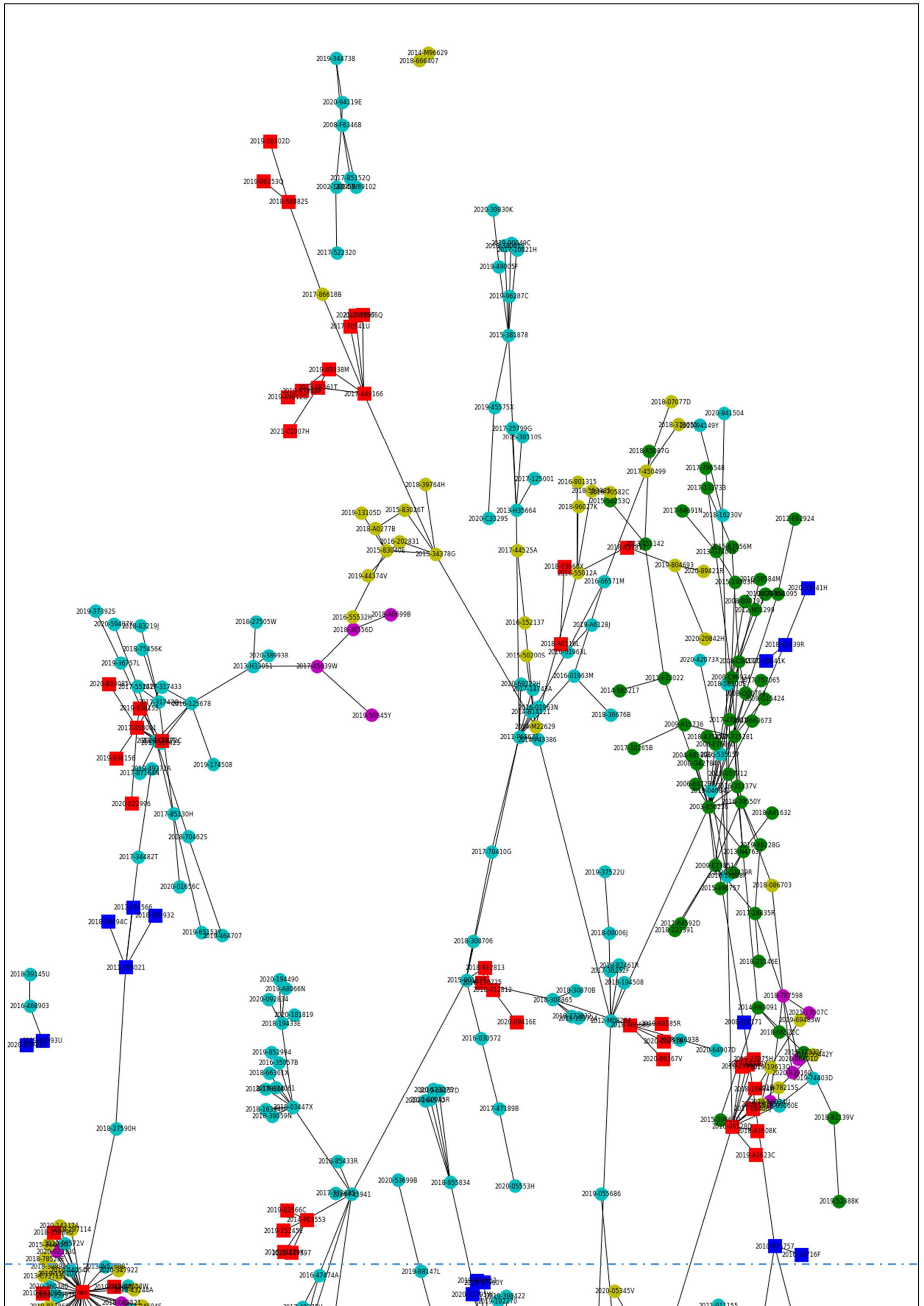
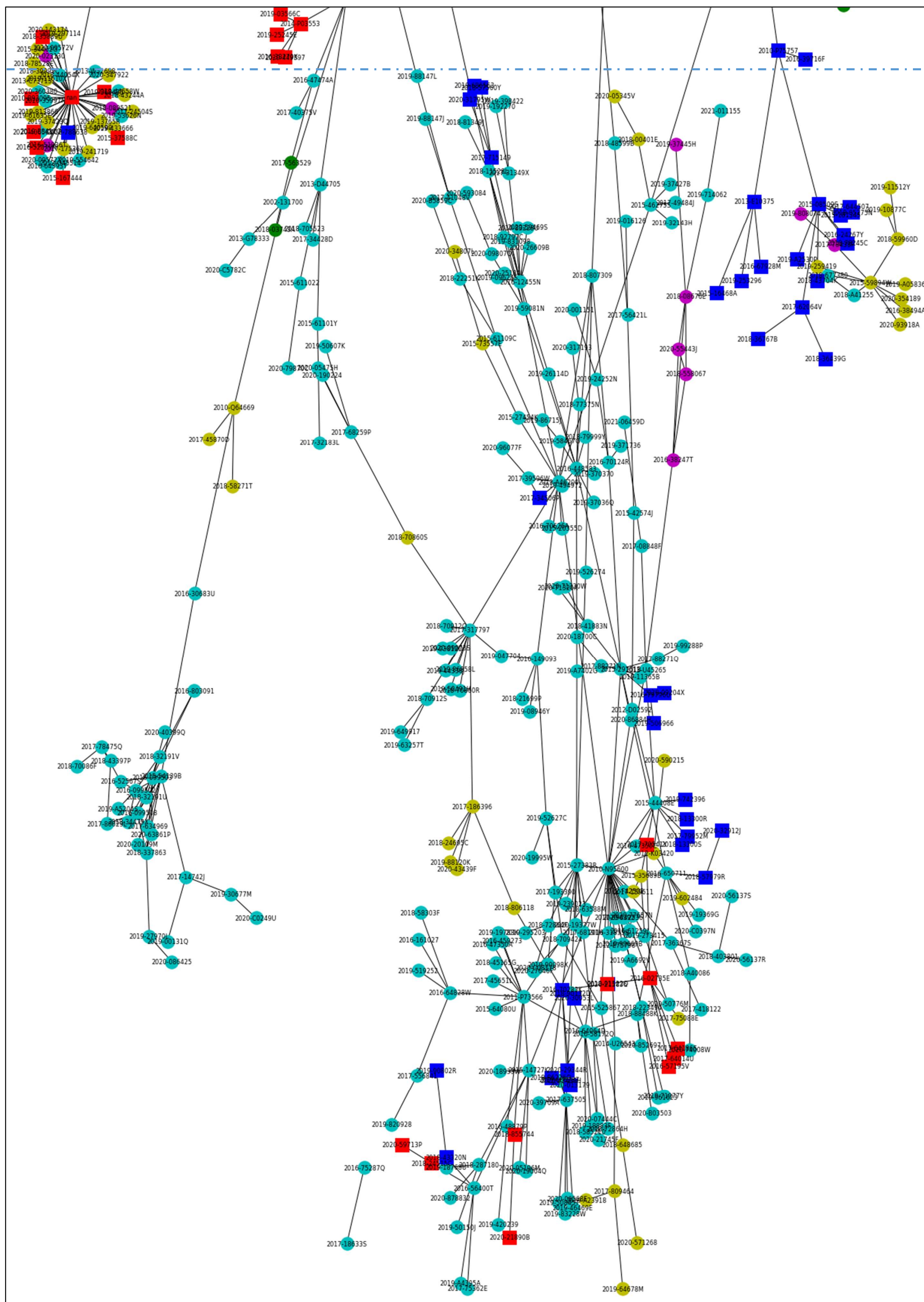
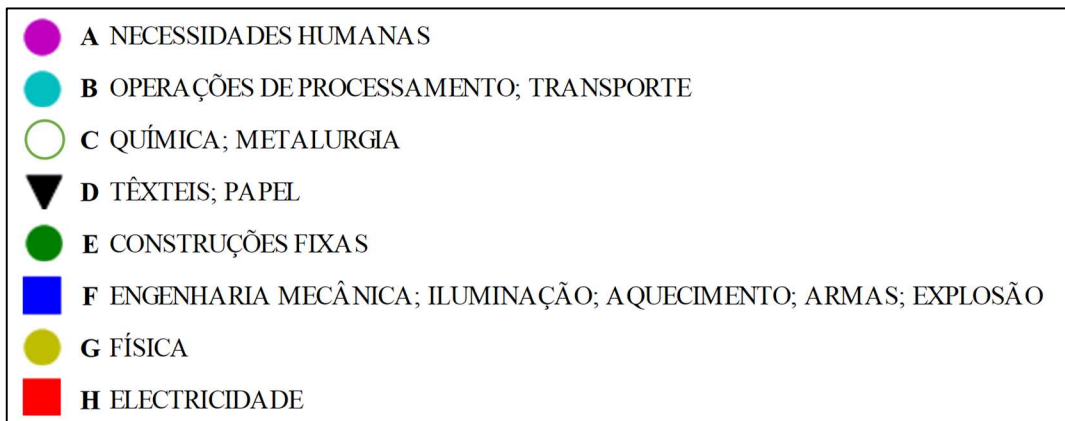


Figura 69 - Fluxo de conhecimento das patentes da tecnologia *Smart Mobility* (fase final)

Fonte: o autor

A Figura 70 mostra a legenda, com as cores e formas dos nós, para os mapas de fluxos de conhecimento mostrados nas Figuras 75, 76A e 76B.

Figura 70 – Legenda Mapa do Fluxo de Conhecimento



Fonte: o autor

4. CONSIDERAÇÕES FINAIS E SUGESTÕES PARA TRABALHOS FUTUROS

Neste trabalho abordou-se o uso de técnicas de *web crawler* no desenvolvimento de um software de prospecção tecnológica, a partir de palavras-chave, de informações dos documentos de patentes contidos no *Derwent Innovatrions Index* na base de dados *Web of Science*. Os dados obtidos foram organizados em bancos de dados relacionais, o que possibilitou a criação de gráficos, ferramentas e de mapas de fluxo de conhecimento, com base nas tecnologias pesquisadas. Como resultado obteve-se, além do software de prospecção tecnológica desenvolvido na linguagem *Python*, gráficos e mapas do fluxo de conhecimento que possibilitaram observar a ocorrência de *spillovers* nas tecnologias apresentadas.

No escopo deste trabalho foram propostos o objetivo geral e quatro específicos, e todos foram integralmente cumpridos, conforme descritos a seguir:

- desenvolver um aplicativo de software para *web crawler*, em linguagem *Python*, que possibilite a prospecção tecnológica, a partir de palavras-chave, de informações nos documentos de patentes da base WoS;
- Organizar em banco de dados relacionais patentes relevantes e tecnologias emergentes na indústria automobilística;
- A partir do banco de dados criar gráficos e ferramentas que possibilite o estudo das tecnologias pesquisadas;
- Desenvolver um *software* para estruturar redes de pesquisa que possibilitem o mapeamento do fluxo de conhecimento das tecnologias pesquisadas;
- Verificar a ocorrência de *spillover* entre as tecnologias pesquisadas.

A criação de um aplicativo *software* para *web crawler*, em linguagem *Python*, objetivo geral deste trabalho, mostrou-se factível. A partir das ferramentas (bibliotecas) específicas da linguagem de programação *Python Scrapy e Selenium*, automatizou-se o processo de busca de informações nos documentos de patentes na plataforma *Derwent Innovations Index* da base de dados *WoS*. No link <https://www.loom.com/share/62f6c74c14c2416dad6f743157fa8ce5> disponibilizou-se um vídeo explicativo que mostra o aplicativo *web cale wos.py*, criado pelo autor, extraíndo dados de documentos de patentes a partir de uma sintaxe de busca.

A maior dificuldade do autor para atender o objetivo geral desta pesquisa foi a complexidade da linguagem de programação utilizada no desenvolvimento das páginas *web* da base de dados *Web os Science*. Por se tratar de páginas *web* dinâmicas, o uso de aplicativos de *web crawl* mais simples, como por exemplo a extensão do navegador *Chrome webcraper.io*, mostrou-se inexequível. Estas dificuldades iniciais só foram superadas após o estudo de conceitos básicos de *Data Science*, o que orientou o autor na aplicação das bibliotecas *Script* e *Selenium* em conjunto com a linguagem de *Python* para conclusão deste trabalho.

A partir da extração dos dados de documentos de patentes das seis tecnologias consideradas relevantes à questão de pesquisa: *Vehicular Safety*; *Infotainment*; *Smart Cities*; *Smart Mobility*; *Advanced Driver Assistance Systems*; e *Autonomous Vehicles*, organizou-se bancos de dados relacionais, com informações características dos documentos de patentes: *num_derwent*, número de patentes, data de publicação, título, resumo, número de patentes que fizeram citação a outras patentes, número de patentes citadas pelo inventor, número de patentes citadas pelo examinador, número de artigos científicos citados pelo inventor, número de artigos científicos citados pelo examinador, informações dos depositantes, inventores e dos países nos quais foram depositados os documentos de patentes.

A organização dos dados em bancos de dados relacionais proporcionou a construção de diversos *dashboards* e gráficos interativos, por meio de softwares com *Power BI* e o *Excel*. Os *dashboards* e gráficos criados facilitaram o estudo das tecnologias pesquisadas neste trabalho e, ao analisar as informações, observou-se que as tecnologias estudadas apresentam patentes relevantes e são emergentes na indústria automobilística. Nos gráficos criados observou-se, nas tecnologias pesquisadas, curvas modais, semelhantes à Curva da Difusão da Inovação, e uma curva bimodal característica de ciclos de inovação associados aos modelos *science-push* e *market-pull*. Evidenciou-se também a presença de diversos setores industriais envolvidos por tecnologia, caracterizando assim transbordamento de conhecimento entre os documentos de patentes. Um maior aprofundamento no estudo das ocorrências observadas nos gráficos não faz parte do escopo deste trabalho e é uma importante sugestão para trabalhos futuros.

A partir dos dados coletados pelo software para *web crawler* criou-se a um banco de dados, utilizando-se o conceito da anterioridade em base de patentes por meio de citações de documentos anteriores, com informações das patentes citadas e citantes para a tecnologia *Smart Mobility*. Os dados disponibilizados possibilitaram a construção de dois mapas de fluxo de conhecimento da tecnologia *Smart Mobility*, sendo um para todas as patentes encontradas na pesquisa desta tecnologia e outro para uma patente específica. Na análise dos dois mapas de

fluxo de conhecimento observou-se a ocorrência de transbordamento de conhecimento por meio da citação de patentes anteriores, classificadas em tecnologias distintas das patentes citantes. Como limitante deste estudo considera-se a impossibilidade do processamento de mapas de fluxo de conhecimento para as demais tecnologias pesquisadas, em função da quantidade de citações de patentes e da capacidade do computador do autor.

Este estudo mostrou-se importante para a compreensão dos temas que envolvem a prospecção tecnológica por meio de documentos de patentes, e de como funciona os mecanismos de propagação da inovação e do fluxo de conhecimento, assuntos estranhos a este autor até o início desta pesquisa. Será gratificante a possibilidade de aprofundar o conhecimento dessas áreas, com o desenvolvimento de trabalhos futuro sobre estes temas.

Como sugestões de trabalhos futuros com base neste estudo pode-se elencar:

- executar uma pesquisa quantitativa que permita comparar os resultados obtidos por esta pesquisa com *softwares* de prospecção tecnológica profissionais disponíveis no mercado, como por exemplo, o software *Orbit Intelligence* da *Questel*;
- pesquisar outras tecnologias por meio da metodologia apresentada neste trabalho;
- aprofundar o estudo da ocorrência de curvas modais e bimodais em gráficos de números de patentes em função tempo, com o objetivo de verificar se existem relações com as teorias da propagação da inovação com os modelos de inovação *science-push* e *market-pull*.
- acrescentar outras variáveis de documentos de patentes na elaboração de mapas de fluxo de conhecimento, como por exemplo informações de inventores, para tornar possível o mapeamento da transferência de conhecimento por relacionamento interpessoal.

REFERÊNCIAS

- ABPI. **O que é Propriedade Intelectual?** Associação Brasileira da Propriedade Intelectual 2020. Disponível em: <https://abpi.org.br/blog/o-que-e-propriedade-intelectual/>. Acesso em: 12/11/2020.
- ABREU, N. M. M. d.; DEL-VECCHIO, R.; VINAGRE, C. *et al.* **Introdução à teoria espectral de grafos com aplicações.** 2007. 25 p., v. Notas em Matemática Aplicada).
- ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. **Journal of urban technology**, 22, n. 1, p. 3-21, 2015.
- ANFAVEA. Anuário Estatístico da Indústria Automobilística Brasileira. Associação Nacional dos Fabricantes de Veículos Automotores. 2021.
- ANTUNES, A.; PARREIRAS, V.; QUINTELLA, C. *et al.* Métodos de Prospecção Tecnológica, Inteligência Competitiva e Foresight: principais conceitos e técnicas. *In.* Salvador - Bahia: Prospecção Tecnológica, 2018. v. 1, (Prospecção Tecnológica).
- BAKICI, T.; ALMIRALL, E.; WAREHAM, J. A smart city initiative: the case of Barcelona. **Journal of the knowledge economy**, 4, n. 2, p. 135-148, 2013.
- BARRETT, P.; HUNTER, J.; MILLER, J. T. *et al.*, 2005, **matplotlib--A Portable Python Plotting Package.** 91.
- BRAGA, E. J.; DE SOUZA, A. R.; DE LIMA SOARES, P. L. *et al.* The role of specification in patent applications: A comparative study on sufficiency of disclosure. **World Patent Information**, 53, p. 58-65, 2018.
- CARVALHO, E. G. d. Inovação tecnológica na indústria automobilística: características e evolução recente. **Economia e Sociedade**, 17, n. 3, p. 429-461, 2008.
- DA SILVA MORAES, E.; RAMOS, J. E. S.; DE SOUZA MELO, A. *et al.* A relação dos dispêndios em P&D e o crescimento econômico do Brasil: uma análise por vetores auto regressivos. **Revista Brasileira de Gestão e Inovação**, 6, n. 1, p. 27-43, 2018.
- DAUDT, G. M.; WILLCOX, L. D. Indústria automotiva = Automotive industry. *In:* PUGA, F. P.; CASTRO, L. B. (Org.). Visão 2035 : Brasil, país desenvolvido : agendas setoriais para alcance da meta. *In.* 1. ed. Rio de Janeiro ed.: Banco Nacional de Desenvolvimento Econômico e Social, 2018. p. 183-208.
- DEARING, J. W.; COX, J. G. Diffusion of innovations theory, principles, and practice. **Health Affairs**, 37, n. 2, p. 183-190, 2018.
- DRUCKER, P. F. The discipline of innovation. **Harvard business review**, 76, n. 6, p. 149-157, 1998.
- FARIA, R.; BRITO, L.; BARAS, K. *et al.*, 2017, **Smart mobility: A survey.** IEEE. 1-8.

FORTEC. **PROSPECÇÃO TECNOLÓGICA: Primeiro passo no caminho da INOVAÇÃO**. 2019. Disponível em: <http://fortec.org.br/2019/08/06/prospeccao-tecnologica-primeiro-passo-no-caminho-da-inovacao/>. Acesso em: 17/02/2021.

GARCEZ JÚNIOR, S. S.; MOREIRA, J. d. J. d. S. O backlog de patentes no Brasil: o direito à razoável duração do procedimento administrativo. **Revista Direito GV**, 13, n. 1, p. 171-203, 2017.

GHEORGHE, M.; MIHAI, F.-C.; DÂRDALĂ, M. J. I. J. o. U.-S. I. Modern techniques of web scraping for data scientists. 11, n. 1, p. 63-75, 2018.

GIETELINK, O.; PLOEG, J.; DE SCHUTTER, B. *et al.* Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations. **Vehicle System Dynamics**, 44, n. 7, p. 569-590, 2006.

HAGBERG, A.; SWART, P.; S CHULT, D. Exploring network structure, dynamics, and function using NetworkX. 2008.

HENRIQUES, F. E.; MIGUEL, P. A. C. Adoção da modularidade em produto e em produção na indústria automotiva: uma análise comparativa em projetos de veículos com participação da engenharia brasileira. **Gestão & Produção**, 24, n. 1, p. 161-177, 2017.

INPI. **Patentes**. Instituto Nacional da Propriedade Industrial 2020. Disponível em: <https://www.gov.br/inpi/pt-br/servicos/perguntas-frequentes/patentes#patente>. Acesso em: 12/11/2020.

IRI. **R&D Monitoring Scoreboard**. 2020. Disponível em: <https://iri.jrc.ec.europa.eu/scoreboard>. Acesso em: 28/10/2020.

JABEEN, H. **Making Web Crawlers Using Scrapy for Python**. 2019. Disponível em: <https://www.datacamp.com/community/tutorials/making-web-crawlers-scrapy-python>. Acesso em: 02/10/2020.

JARDIM, P. E. Spillovers de conhecimento das firmas brasileiras internacionalizadas: um estudo a partir de patentes. *In*: IV ENEI Encontro Nacional de Economia e Inovação, 2019, Campinas, São Paulo.

KLINE, S. Innovation is not a linear process. **Research management**, 28, n. 4, p. 36-45, 1985.

KORSSSEN, T.; DOLK, V.; VAN DE MORTEL-FRONCZAK, J. *et al.* Systematic model-based design and implementation of supervisors for advanced driver assistance systems. **IEEE Transactions on Intelligent Transportation Systems**, 19, n. 2, p. 533-544, 2017.

KRISHNAN, V. Research data analysis with Power BI. **INFLIBNET Centre**, 2017.

MARSAL-LLACUNA, M.-L.; COLOMER-LLINÀS, J.; MELÉNDEZ-FRIGOLA, J. Lessons in urban monitoring taken from sustainable and livable cities to better address the Smart Cities initiative. **Technological Forecasting Social Change**, 90, p. 611-622, 2015.

MAYERHOFF, Z. D. V. L. Uma análise sobre os estudos de prospecção tecnológica. **Cadernos**

de prospecção, 1, n. 1, p. 7-9, 2008.

MCKINNEY, W. pandas: a Foundational Python library for data analysis and statistics. **Python for High Performance Scientific Computing**, 14, n. 9, p. 1-9, 2011.

MEIXNER, G.; HÄCKER, C.; DECKER, B. *et al.* Retrospective and future automotive infotainment systems—100 years of user interface evolution. *In: Automotive user interfaces*: Springer, 2017. p. 3-53.

MOORE, G. A. **Crossing the chasm: Marketing and Selling Disruptive Products to Mainstream**. 3rd. Edition ed. New York (USA): Harper Business Essentials, 2014. 0062292986.

NASCIMENTO, R. d. S. **Patentometria: estudo de múltiplos casos sobre a utilização de dados contidos em patentes como mecanismo de auxílio na gestão da inovação dos NITs**. 2020. - Programa de PósGraduação em Inovação Tecnológica e Propriedade Intelectual, Universidade Federal de Minas Gerais, Minas Gerais.

OECD/EUROSTAT. **Oslo Manual 2018: Guidelines for collecting, reporting and using data on innovation**. 4th Edition ed. Paris: OECD publishing, 2018. (The Measurement of Scientific, Technological and Innovation Activities. 978-92-64-30460-4.

OICA. **International Organization of Motor Vehicle Manufacturers**. 2021. Disponível em: <https://www.oica.net/>. Acesso em: 05/04/2021.

PARANHOS, R. d. C. S.; RIBEIRO, N. M. Importância da prospecção tecnológica em base em patentes e seus objetivos da busca. **Cadernos de Prospecção**, 11, n. 5, p. 1274, 2018.

PERKEL, J. M. Why Jupyter is data scientists' computational notebook of choice. **Nature**, 563, n. 7732, p. 145-147, 2018.

PURIFICAÇÃO, M. A. **UMA ANÁLISE DE REDES DE TRANSBORDAMENTO EM FLUIDOS DE PERFURAÇÃO**. 2015. - Programa de Pós-graduação em Tecnologia, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,, Rio de Janeiro.

QUINTELLA, C. M.; ALMEIDA, B. A.; SANTOS, W. P. C. *et al.* Busca de anterioridade. *In*. Salvador - Bahia: Instituto Federal de Educação, Ciência e Tecnologia da Bahia - IFBA, 2018. v. 1, (Prospecção Tecnológica).

QUINTELLA, C. M.; MEIRA, M.; GUIMARAES, A. K. *et al.* Prospecção Tecnológica como uma Ferramenta Aplicada em Ciência e Tecnologia para se Chegar à Inovação. **Revista virtual de química**, 3, n. 5, p. 406-415, 2011.

RAMOS, A.; ZILBER, S. N. O impacto do investimento na capacidade inovadora da empresa. **RAI Revista de Administração e Inovação**, 12, n. 1, p. 303-325, 2015.

REN, Y.; REN, Y., 2018, **A framework of petroleum information retrieval system based on web scraping with python**. IEEE. 1-6.

ROGERS, E. M. **Diffusion of innovations**. New York (USA): Simon and Schuster, 2010.

1451602472.

SCHMOCH, U. Double-boom cycles and the comeback of science-push and market-pull. **Research policy**, 36, n. 7, p. 1000-1015, 2007.

SCHMOCH, U. Concept of a technology classification for country comparisons. **Final report to the world intellectual property organisation WIPO**, 2008.

SCHUMPETER, J. A. **Teoria do desenvolvimento econômico**. Tradução SCHLÄPFER, L. Rio de Janeiro: Fundo de Cultura, 1961.

SCRAPY.ORG. **Scrapy 2.4 documentation**. 2020. Disponível em: <https://docs.scrapy.org/en/latest/>. Acesso em: 28/11/2020.

TALEBPOUR, A.; MAHMASSANI, H. S. Influence of connected and autonomous vehicles on traffic flow stability and throughput. **Transportation Research Part C: Emerging Technologies**, 71, p. 143-163, 2016.

TIGRE, P. B. Inovação e teorias da firma em três paradigmas. **Revista de economia contemporânea**, 3, p. 67-111, 1998.

TIGRE, P. B. **Gestão da inovação: a economia da tecnologia no Brasil**. . Rio de Janeiro: Campus/Elsevier, 2006.

TRIPPE, A. **Guidelines for preparing patent landscape reports**. Geneva: WIPO, 2015. 2015 p. (Patent landscape reports).

TRIPPE, A. J. Patinformatics: Tasks to tools. **World Patent Information**, 25, n. 3, p. 211-221, 2003.

USPTO. **General information concerning patents**. United States Patent and Trademark Office, 2015. Disponível em: <https://www.uspto.gov/patents/basics>. Acesso em: 12/11/2020.

VILELA JÚNIOR, D. C. Os spillovers e a sua influência na criação de conhecimento em clusters industriais. **Gestao e Desenvolvimento**, 12, n. 1, p. 50-69, 2015.

VON FERSEN, O. **Ein Jahrhundert Automobiltechnik: Personenwagen**. Springer-Verlag, 2013. 3642957722.

WIPO. Convention Establishing the World Intellectual Property Organization. Stockholm: World Intellectual Property Organization 1967.

WIPO. WIPO Statistics Database. : World Intellectual Property Organization 2020a.

WIPO. **World Intellectual Property Indicators 2020**. Geneva: World Intellectual Property Organization (WIPO), 2020b. 237 p. 978-92-805-3201-2.

XIANG, X.-Y.; CAI, H.; LAM, S. *et al.* International knowledge spillover through co-inventors: An empirical study using Chinese assignees' patent data. **Technological Forecasting Social Change**, 80, n. 1, p. 161-174, 2013.