

GT-8 – Informação e Tecnologia

ISSN 2177-3688

PROPOSIÇÃO DE UM FRAMEWORK DE COLETA E TRATAMENTO DE DADOS SOBRE A PRODUÇÃO TÉCNICA BRASILEIRA.

PROPOSITION OF A FRAMEWORK FOR COLLECTING AND PROCESSING DATA ON BRAZILIAN TECHNICAL PRODUCTION.

Raulivan Rodrigo da Silva - Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG),

Thiago Magela Rodrigues Dias - Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG),

Washington Luís Ribeiro de Carvalho Segundo - Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Modalidade: Trabalho Completo

Resumo: Com a Revolução Industrial e o advento do capitalismo, surgiu a sociedade tecnológica, onde o desenvolvimento tecnológico desempenha um papel importante. No século XXI, é possível observar avanços significativos em várias áreas do conhecimento, com novos dispositivos e aplicações sendo lançados constantemente. Diante disso, as organizações precisam acompanhar essas mudanças para manter sua competitividade. Uma forma de acompanhar o desenvolvimento tecnológico é por meio de documentos de patentes, que contêm valiosas informações sobre tendências e novas tecnologias. Além disso, programas de computador registrados oferecem proteção legal aos desenvolvedores, incentivando a inovação e a criação de novos *softwares*. No entanto, acessar e manipular grandes volumes de dados de patentes não é uma tarefa fácil. Existem diversos repositórios que disponibilizam acesso aos dados, mas lidar com um alto volume de informações ainda é um desafio. Isso limita análises mais abrangentes, como a avaliação do cenário tecnológico de um país ou instituição. Para enfrentar esse desafio, é proposto um framework que viabilize a coleta e o tratamento de grandes volumes de dados de patentes e programas de computador depositados no Brasil. No processo metodológico é descrito um conjunto de estratégias para a coleta e tratamento de dados sobre patentes e programas de computador com registro. Resultando na construção de uma rica base de dados sobre a produção técnica brasileira, composta por 903.979 patentes e 25.092 programas de computador com registro. Isso permitirá análises mais amplas e abrangentes, fornecendo informações valiosas para pesquisadores, empresas e instituições interessadas no acompanhamento tecnológico.

Palavras-chave: patente; software; registro; produção técnica.

Abstract: With the Industrial Revolution and the advent of capitalism, the technological society emerged, where technological development plays an important role. In the 21st century, it is possible to observe significant advances in several areas of knowledge, with new devices and applications being constantly launched. Given this, organizations need to keep up with these changes to maintain their competitiveness. One way to keep up with technological development is through patent documents, which contain valuable information on trends and new technologies. In addition, registered computer programs offer developers legal protection, encouraging innovation and the

creation of new software. However, accessing and manipulating large volumes of patent data is not an easy task. There are several repositories that provide access to data, but dealing with a high volume of information is still a challenge. This limits more comprehensive analysis, such as assessing the technological scenario of a country or institution. To face this challenge, a framework is proposed that enables the collection and processing of large volumes of data from patents and computer programs deposited in Brazil. The methodological process describes a set of strategies for collecting and processing data on registered patents and computer programs. Resulting in the construction of a rich database on Brazilian technical production, consisting of 903,979 patents and 25,092 registered computer programs. This will allow broader and more comprehensive analysis, providing valuable information for researchers, companies and institutions interested in technological monitoring.

Keywords: patent; software; record; technical production.

1 INTRODUÇÃO

Com a eclosão da Revolução Industrial e o advento do capitalismo, emergiu a sociedade tecnológica, termo este, que se refere a um contexto social em que o desenvolvimento tecnológico desempenha um papel importante, influenciando significativamente em diversos aspectos da sociedade. Neste contexto, Mazzola e Armellini (2011) complementam que o desenvolvimento econômico de um país é fundamentado na aplicação do conhecimento na geração e distribuição de riquezas.

Nessa perspectiva, o século XXI tem se destacado por grandes avanços tecnológicos nas mais diversas áreas do conhecimento e mais do que nunca, a rapidez na evolução destas tecnologias tem sido visível. Diariamente novos dispositivos, aplicações, meios digitais permeiam o mercado, trazendo versões melhores de recursos e funcionalidades que até então conhecíamos ou apresentando novas soluções. Em consequência disso, as organizações desse mercado estão empenhadas em realizar um monitoramento constante de suas atividades e da viabilidade de seus produtos e serviços oferecidos, sendo necessário para tanto, implementar inovações que fidelizem ou aumentem a base de clientes (AMADEI; TORKOMIAN, 2009).

Certamente existem diversas formas de acompanhar o desenvolvimento tecnológico, como leitura de notícias, publicações especializadas, participação em conferências e eventos tecnológicos, acompanhamento de pesquisas acadêmicas, dentre outros. No entanto, a principal fonte de dados sobre desenvolvimento tecnológico são os documentos de patentes, de acordo com Instituto Nacional da Propriedade Industrial (INPI) cerca de 70% do conhecimento tecnológico é divulgado exclusivamente por esse tipo de documento (BRASIL, 2015). Calzolaio *et al.* (2018) concordam que os dados de patentes contêm informações valiosas para análises técnicas, além de serem consideradas a principal forma de

compartilhar conhecimento tecnológico. Além disso, a compreensão do estado da técnica da arte por meio de documentos de patentes, consequentemente apresenta um cenário mais assertivo a respeito de tendências tecnológicas, setores promissores, bem como, a possibilidade de novas tecnologias (NASCIMENTO; SPEZIALI, 2020).

Do mesmo modo, programas de computador com registro, ou seja, um *software* que recebeu proteção legal por meio de um registro de propriedade intelectual no INPI, por força da Lei de Direito Autoral (Lei nº 9.610/98), da Lei de *Software* (Lei nº 9.609/98) e do Decreto nº 2.556/98 (BRASIL, 1998), pode ser interpretado como um incentivo à inovação e à criação de novos programas de computador, pois a proteção legal oferecida pelo registro encoraja os desenvolvedores a investirem tempo, esforços e recursos na criação de *software*, pois mediante o registro, receberam reconhecimento e consequentemente direitos sobre sua criação devidamente assegurados (BRASIL, 2021). Assim, podem armazenar dados relevantes para o acompanhamento tecnológico.

1.1 Caracterização do problema

De acordo com dados levantados pelo INPI, são depositadas anualmente cerca de 28 mil patentes, constituindo assim, um rico arcabouço de informações tecnológicas em diversas áreas do conhecimento (BRASIL, 2022). No entanto, ter acesso aos dados de documentos de patentes não é considerada uma tarefa trivial. Apesar de existirem diversos repositórios que disponibilizam o acesso pago ou gratuito aos dados de patente, como o INPI e Espacenet, manipular grande volume de dados ainda é um fato limitador. Isso dificulta análises mais amplas, que abrangem um alto volume de dados, como por exemplo, avaliar o cenário de um país ou conjunto de instituições.

Nessa conjuntura, Brandão (2016) destaca sete bases de dados que disponibilizam a consulta de documentos de patentes gratuitamente pela internet, dentre elas: “Pesquisa em Propriedade Industrial” mantido pelo INPI; “Latipat” para buscas em documentos de patentes da América Latina; “Espacenet” mantido pelo Escritório Europeu de Patentes (EPO); e “Patentscope” que é a ferramenta de buscas da Organização Mundial da Propriedade Intelectual (OMPI). Já Pires, Ribeiro e Quintella (2020) em seu trabalho intitulado “Sistemas de Busca de Patentes: análise comparativa entre Espacenet, Patentscope, Google *Patents*, Lens, *Derwent Innovation Index* e *Orbit Intelligence*”, realizaram uma análise comparativa entre seis ferramentas de consulta de documentos de patentes, analisando 13 características

distintas tais como, tipo de acesso, número de autoridades de patentes coberta, tamanho da base e exportação de dados. Logo, concluíram que todas as ferramentas analisadas possuem limitações, porém a ferramenta “*Orbit Intelligence*” se destaca como a melhor ferramenta de consulta de documentos de patentes, no entanto, a ferramenta não é gratuita.

Mediante ao exposto, é evidente a necessidade de alternativas que viabilizem a análise de grandes volumes de dados de patentes, bem como, de outros recursos tecnológicos, como por exemplo programas de computador devidamente registrados.

1.2 Objetivo

Logo, este trabalho tem como principal objetivo, a proposição de um *framework* que viabilize a coleta e tratamento de grandes volumes de dados sobre patentes depositadas no Brasil e programas de computador registrados no INPI.

2 FUNDAMENTAÇÃO TEÓRICA

A área da Propriedade Intelectual (PI) está diretamente relacionada aos elementos abordados neste trabalho, refere-se a direitos legais que são atribuídos a criações da mente humana. Ela abrange diversas formas de expressão criativa e inovação, protegendo os interesses dos criadores e detentores dessas criações. De acordo com Bonetti (2010) a Propriedade Intelectual se divide em três grandes grupos, “direito autoral”, “propriedade industrial” e “proteção sui generis”. Neste mesmo cenário Quoniam, Kniess e Mazieri (2014) destacam que patentes, parte fundamental deste estudo, pertence ao grupo “propriedade industrial”, junto com marcas, desenhos industriais e indicações geográficas de origem.

Considera-se a patente um documento formal, que é expedido por uma repartição pública, pela qual se conferem e se reconhecem direitos de propriedade e uso exclusivo para uma invenção amplamente descrita (BONETTI, 2010). Trata-se de um privilégio concedido aos idealizadores sejam eles pessoas física ou jurídica detentores do direito de invenção seja de produtos, processos de fabricação, ou até mesmo aperfeiçoamento de algum produto ou processo já existente.

Neste contexto, existe uma área dentro da Bibliometria para descrever os estudos métricos de patentes, a “Patentometria”. A Patentometria faz parte do grupo que constitui a Bibliometria, pois para a análise de documentos de patentes, são utilizadas adaptações dos indicadores aplicados à produção de outros tipos de documentos (FAGUNDES *et al.*, 2014).

Assim, a Patentometria objetiva analisar informações contidas em documentos de patente buscando identificar possíveis relacionamentos, padrões e tendências, entre outros.

3 METODOLOGIA

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

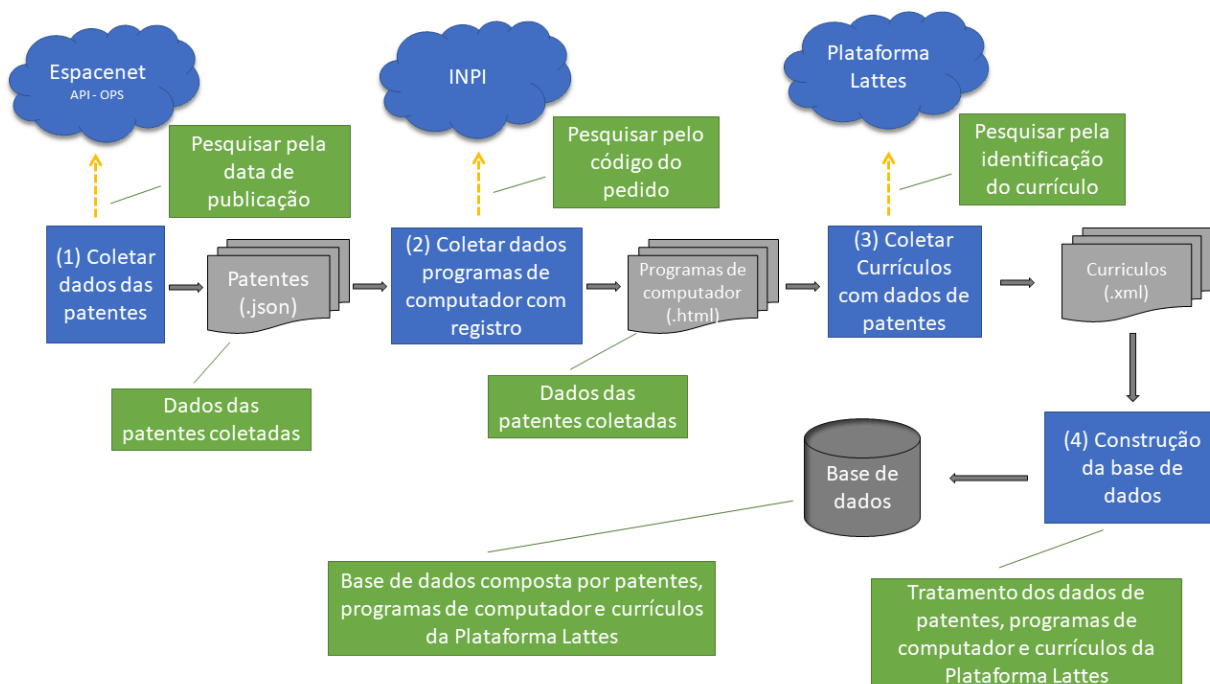
O processo metodológico consiste na descrição de um conjunto ferramental para coleta e tratamento de documentos de patentes e programas de computador depositados/registrados no Brasil, doravante chamados de Produção Técnica (PT).

Dito isto, no intuito de obter informações que normalmente não são registradas nos documentos referentes a PT, foram considerados os dados provenientes de currículos registrados na Plataforma Lattes que possuem informações sobre a PT. Os currículos registrados na Plataforma Lattes são fontes de dados relevantes por possuírem informações complementares sobre os proponentes da PT, uma vez que os pesquisadores registram em seus currículos, seus históricos acadêmicos, publicações científicas, instituições de vínculo, dentre outras informações que podem serem utilizadas para os mais diversos tipos de análise (LOPES, 2022). Atualmente a Plataforma Lattes é composta por mais de 8,1 milhões de currículos desde a sua criação em 1999. Este crescimento expressivo da base curricular da Plataforma Lattes é um dos fatores que destacam sua relevância neste estudo (GÜNTHER *et al.*, 2020).

3.1 Estratégia de coleta

Toda a estratégia proposta foi dividida em quatro etapas distintas, a saber, (1) coleta dos dados de patentes no repositório da Espacenet; (2) coleta dos dados de programa de computador com registro no INPI; (3) a coleta de dados dos currículos da Plataforma Lattes que possuem referências a depósito de patentes e/ou programas de computador com registro; e por fim, e por fim, (4) a construção da base de dados local. A Figura 1 apresenta o esquema geral elaborado para o processo de coleta e construção da base de dados.

Figura 1 – Visão geral da coleta de dados



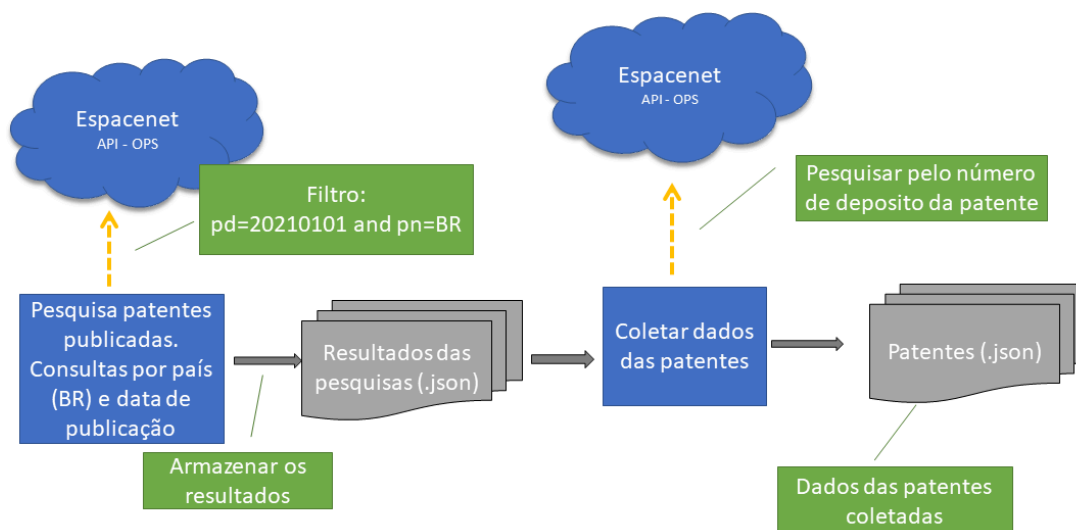
Fonte: Elaboração do autor

Os dados de patentes foram coletados no repositório da Espacenet. A Espacenet disponibiliza um serviço chamado *Open Patent Services* (OPS), um serviço web que fornece acesso aos dados armazenados no banco de dados do *European Patent Office* (EPO) por meio de *Application Programming Interface* (API). Esta possibilidade consequentemente torna viável o desenvolvimento de aplicativos e robôs de extração automática para baixar grandes volumes de dados. Tais fatores motivaram a escolha da Espacenet.

Logo, baseado na documentação da OPS¹, foi implementado um algoritmo, utilizando a linguagem de programação Python, que realiza requisições à API informando como critério de seleção a data de publicação e o país, efetuando uma requisição para cada dia dentro do período de janeiro de 1900 a dezembro de 2022, objetivando obter o maior número de patentes. A coleta foi realizada entre os meses de janeiro a abril do ano de 2023. A Figura 2 apresenta o processo de coleta de patentes na Espacenet.

Figura 2 – Coleta na Espacenet

¹ OPS. Open Patent Services RESTful Web Services. 1.3.18. Disponível em: <https://www.epo.org/searching-for-patents/data/web-services/ops.html>.



Fonte: Elaboração do autor

Para coletar os dados referente a programas de computador com registro, foi definida a ferramenta BuscaWeb disponibilizada pelo INPI, em que é possível realizar a consulta de programas de computador com registro de forma gratuita. Ao acessar a ferramenta pelo método identificação anônima, foi selecionada a opção de “Programa de computador”, em que é apresentada uma página contendo as opções de pesquisa, apresentando como critérios de busca o número do pedido, ou termos que possam conter no título do programa, no nome dos autores etc. Ao informar no campo “Contenha o Número do Pedido” o número “BR512021001270-4” por exemplo, e selecionar a opção “pesquisar”, o sistema retorna uma página com a listagem programas de computadores que correspondem aos critérios de pesquisa, neste exemplo retorna um único programa de nome “TRCAPP (Treinamento Rítmico Corporal Aplicativo)”. Ao selecionar o mesmo, é carregada a página com os detalhes do programa como data de depósito, linguagem de programação, titular, inventores, dentre outras informações, inclusive acesso ao arquivo PDF de registro.

Assim sendo, objetivando construir processos computacionais para a coleta automática, utilizou-se da URL da página de detalhes. No exemplo anterior, a URL que dá acesso a página de detalhes do programa “BR512021001270-4” é “<https://busca.inpi.gov.br/pePI/servlet/ProgramaServletController?Action=detail&CodPedido=32339&SearchParameter=>”, na qual é possível observar a variável “CodgPedido” recebe o valor “32339”, é o número interno de identificação do registro do programa de computador. Logo, a estratégia definida consiste em realizar requisições a URL, alterando o valor da variável “CodgPedido”.

E por fim, foi realizada a coleta de currículos registrados na Plataforma Lattes que possuem informações de patentes ou programas de computador com registro. O processo de coleta e seleção dos dados curriculares da Plataforma Lattes foi realizado por meio do *framework LattesDataXplorer* (DIAS, 2016). A coleta dos currículos foi realizada em janeiro de 2023.

De posse dos dados da PT e currículos da Plataforma Lattes, foi estruturado o repositório local de dados. Para facilitar a organização e recuperação das informações foi adotado o uso de um banco de dados relacional. O banco de dados definido foi o SQLite, por se tratar de um banco de dados que não tem a necessidade de um servidor, além de não demandar nenhum tipo de configuração.

Assim, os dados coletados na Espacenet foram organizados em 7 tabelas, primeira é a tabela “PATENTS” onde são armazenadas as informações básicas de uma patente, como sua identificação, família, título e resumo. A próxima é a tabela “PERSONS”, onde são armazenadas todas as pessoas envolvidas (depositantes e inventores). Em sequência vem a tabela “APPLICANTS” que estabelece o relacionamento do tipo depositante entre as tabelas “PATENTS” e “PERSONS”. Por sua vez a tabela “INVENTORS” representa o relacionamento do tipo inventor entre as tabelas “PATENTS” e “PERSONS”. A tabela “CLASSIFICATIONS” armazena as classificações recebidas pela patente. A tabela “IDENTIFICATIONS” armazena todas as formas de se identificar uma patente, armazenando seus diferentes números de identificação, quando existir. E por fim, a tabela “PERSON_CITATIONS” onde é armazenado as diferentes formas em que uma pessoa é referenciada, ou seja, diferentes grafias de um mesmo nome.

Já os dados coletados no INPI, foram organizados em 2 tabelas, a primeira a tabela “REGISTERED_SOFTWARE” contendo todas as informações coletadas do programa de computador; e a tabela “REGISTERED_SOFTWARE_CREATOR” com os nomes de todos os autores do programa de computador.

E por fim, os dados coletados na Plataforma Lattes foram organizados em 9 tabelas. A primeira “LATTES_IDENTIFICATION” onde são armazenadas as informações de identificação do currículo. A tabela “LATTES_PATENTS” armazena as informações sobre a patente informada na Plataforma Lattes e possui relacionamento com a tabela “LATTES_PATENTS_AUTHORS” onde são armazenados os inventores informados nos currículos. A tabela “LATTES_SOFTWARE” armazena as informações dos programas de

computador com registro e possui relacionamento com a tabela “LATTES_SOFTWARE_AUTHORS” onde é armazenado os autores do programa. A tabela “LATTES_OCCUPATION_AREA” armazena as áreas de atuação informada nos currículos. A tabela “LATTES_FORMATION” armazena a formação do proprietário do currículo. A tabela “LATTES_AREA_OF_EXPERTISE” armazena as áreas de atuação informadas nos currículos. E por fim, a tabela “LATTES_ADDRESS” armazena o endereço profissional informado nos currículos. Todo o processo de tratamento dos dados e construção da base de dados, foi codificada usando a linguagem de programação Python.

Mediante as estratégias de coleta e tratamento de dados apresentadas, foi possível construir uma base de dados em que os dados foram organizados de forma segura, consistente e confiável, a qual se caracteriza como uma rica fonte de dados sobre a PT.

4 RESULTADOS

Atendendo o objetivo principal deste trabalho, obteve-se como resultado um conjunto de scripts escritos na linguagem de programação Python que constituem um framework para coleta e tratamento de dados sobre a PT. Logo, foi compilado um pacote que será disponibilizado no *Python Package Index*², o repositório oficial de pacotes do Python, assim que concluído o processo de registro junto ao INPI.

Em sequência, com o intuito de explorar a base de dados obtida neste trabalho o Quadro 1 caracteriza as entidades constituídas, bem como seus respectivos quantitativos.

Quadro 1 – Caracterização da base de dados

Entidade	Total
Patentes (Espacenet) – Global (PEG)	903.979
Patentes (Espacenet) – Único - PEU	797.747
Programas de computador com registro (INPI) - PCI	25.092
Currículos da Plataforma Lattes analisados	8.054.300
Currículos da Plataforma Lattes (com patentes e softwares registrado) - CPL	35.072
Currículos da Plataforma Lattes (com patentes)	28.581

² <https://pypi.org/>

Currículos da Plataforma Lattes (com softwares registrados)	8.479
Patentes (Plataforma Lattes) – Global - PGPL	76.619
Patentes (Plataforma Lattes) – Único - PUPL	57.281
Programas de computador com registro (Plataforma Lattes) – Global - PCGPL	19.264
Programas de computador com registro (Plataforma Lattes) – Único - PCUPL	14.708
Patentes do Lattes identificadas na Espacenet – Global	44.856
Patentes do Lattes identificadas na Espacenet – Único	18.678

Fonte: Elaboração do autor

Conforme o Quadro 1, mediante as estratégias propostas neste trabalho foi possível recuperar 903.979 registros de patentes oriundos da Espacenet (PEG), considerando apenas a última publicação de cada patente, obtém-se o total de 797.747 patentes únicas (PEU). O termo global refere-se ao conjunto todo, e o termo único, por sua vez refere-se a registros distintos. Prosseguindo, a fins de comparação, em junho de 2023 o INPI possuía 921.055 registros de patentes publicadas disponíveis para consulta em seu portal de pesquisa, o conjunto de PEU representa aproximadamente 87% do conjunto de patentes disponível no INPI. Já em relação aos dados de programas de computador com registro coletados no INPI (PCI), foi possível coletar 25.092 registros.

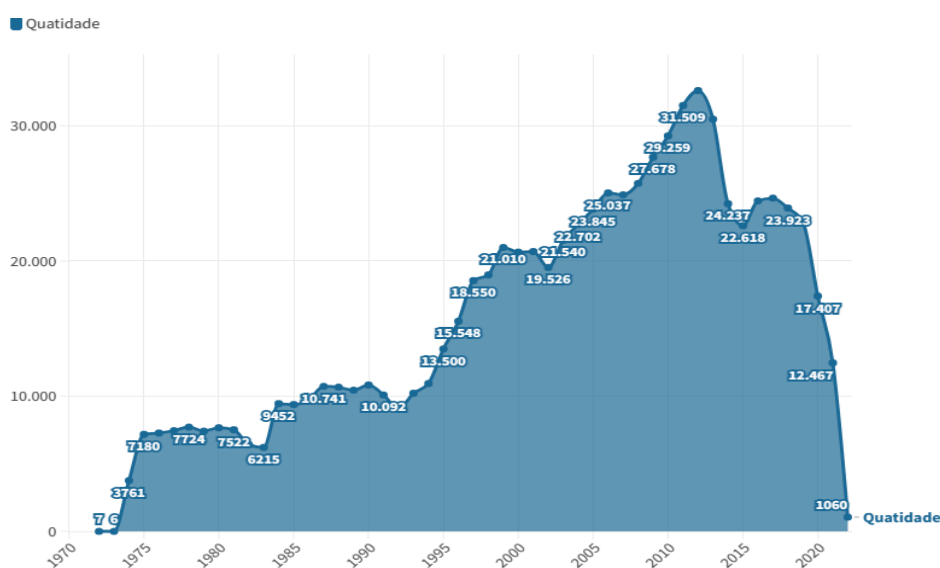
Do mesmo modo, após analisar toda a base curricular da Plataforma Lattes, foram coletados 35.072 currículos da Plataforma Lattes (CPL), desse montante 28.581 currículos possuem registros de patentes, já 8.479 currículos possuem ocorrências de programas de computador com registro. Aprofundando no conjunto de dados coletado na Plataforma Lattes, foi recuperado 76.619 registro de patentes informados nos currículos da Plataforma Lattes (PGPL). Avaliando apenas registros únicos, considerando como critério título e ano da patente, obtém-se o total de 57.281 registros de patentes (PUPL). Em relação aos programas de computador com registro informados nos currículos da Plataforma Lattes (PCGPL), foi coletado 19.264 registros, sendo que desse montante 14.708 são registros únicos, considerando o título e o ano como critério (PCUPL).

Em sequência, a fim de relacionar os dados coletados, foi possível identificar 44.856 registros do conjunto de dados PGPL no conjunto de dados PEG, sendo assim consideradas patentes certificadas, devido a identificação em uma base de dados internacional utilizando

como critério número de depósito e o ano das patentes. Contudo, 18.678 das patentes certificadas são registros únicos.

Como resultado, objetivando explorar os dados coletados, algumas análises foram elaboradas, considerando o período de 1970 a 2022. Inicialmente, para apresentar um panorama da evolução tecnológica nacional, foi realizada uma análise dos depósitos anuais de patentes. O Gráfico 1 apresenta a evolução temporal no total de depósito de patentes, ressaltando que as patentes que não contém a informação de ano de depósito foram desconsideradas.

Gráfico 1 – Evolução temporal do depósito de patentes por ano



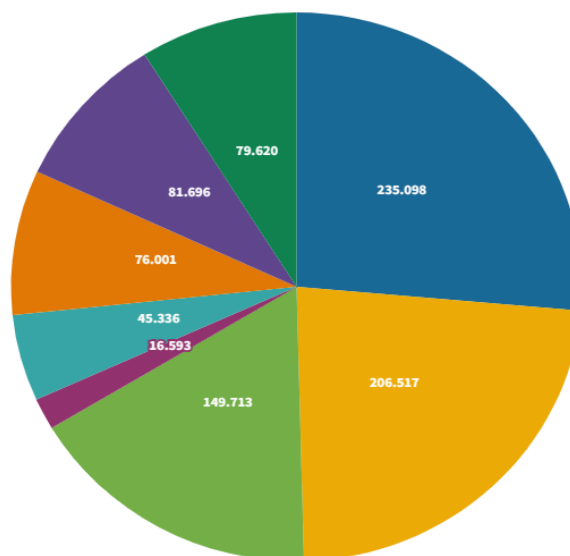
Fonte: Elaboração do autor

Desde 1972 até 2021 no Brasil são realizados em média 15.933,52 depósitos de patentes por ano. O ano com maior número de patentes depositadas é o ano de 2012, contendo um total de 32.608 patentes depositadas. Destaca-se ainda que houve um crescimento contínuo no número de depósitos entre os anos de 1992 e 2012, podendo ser considerado um período de grande desenvolvimento tecnológico.

Cada patente de acordo com sua natureza e finalidade recebe uma classificação de acordo com o sistema internacional de classificação de patentes, com base nestas classificações é possível compreender quais áreas do conhecimento tem gerado o maior número de patentes. Para compreender melhor este cenário foi compilado o Gráfico 2.

Gráfico 2 – Patentes por classificação

■ A-Necessidades humanas ■ B-Operações de processamento (...) ■ C-Química; metalurgia ■ D-Têxteis; papel
■ E-Construções fixas ■ F-Engenharia mecânica (...) ■ G-Física ■ H-Eletricidade

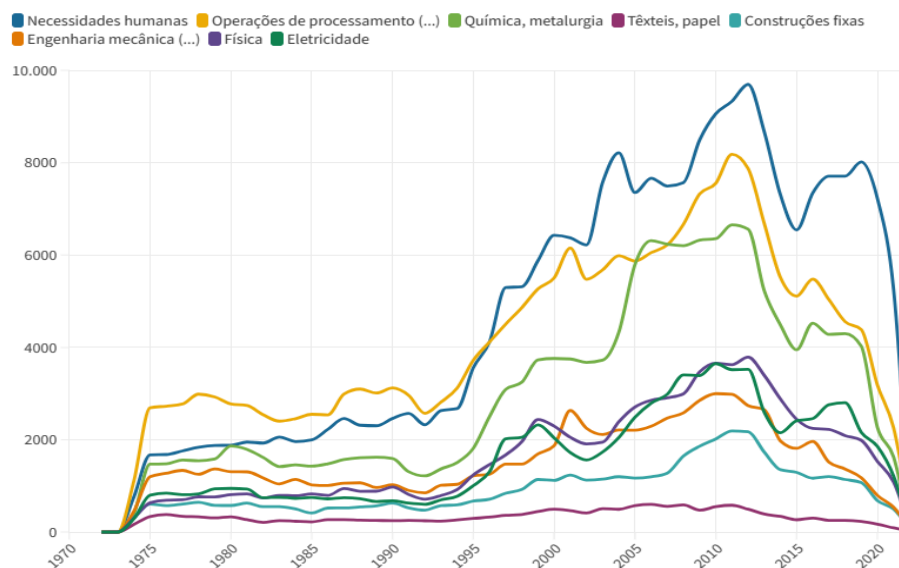


Fonte: Elaboração do autor

Analisando o Gráfico 2, em um contexto geral, as patentes brasileiras em sua maioria estão concentradas em três classificações, a saber: A – Necessidades humanas, B - Operações de processamento e transportes e C – Química, Metalurgia. Essas classificações incluem áreas que impulsionam a economia do Brasil, dentre elas agricultura, indústria, serviços, tecnologia da informação, energia, mineração, entre outras. Esses setores desempenham papéis importantes na geração de empregos, na geração de receita e no desenvolvimento econômico do país.

Já em um contexto temporal, as classificações podem indicar as áreas de tecnologia que estão recebendo maior atenção e investimento, bem como áreas emergentes, observando o total de depósito de patentes por ano. O Gráfico 3 apresenta o total de patentes por ano para cada tipo de classificação de patente.

Gráfico 3 – Total de depósito de patentes por ano e por classificação

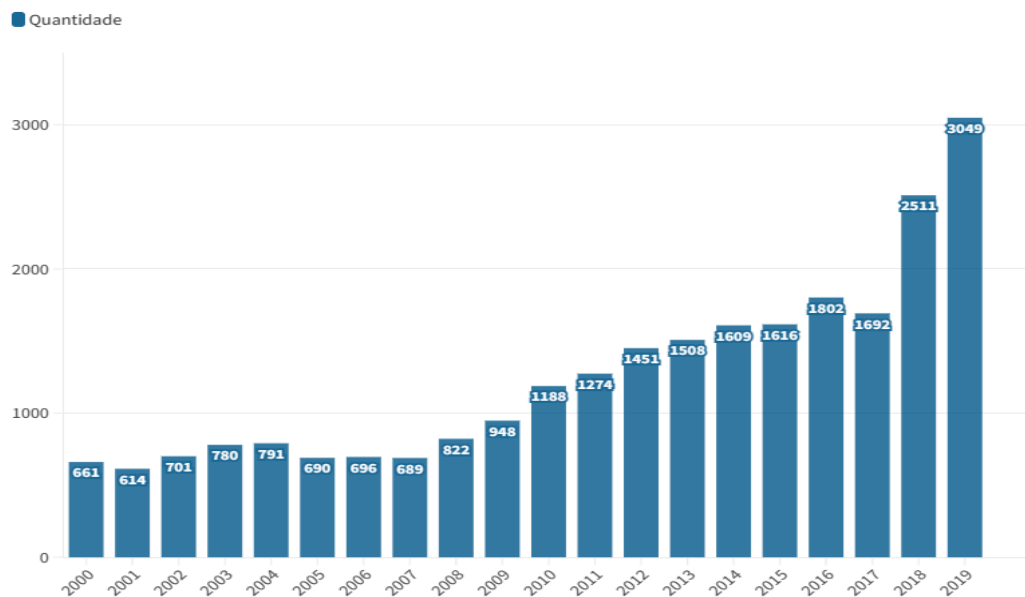


Fonte: Elaboração do autor

Analisando o panorama apresentado pelo Gráfico 3, é possível observar que na primeira metade da década de 70, as áreas relacionadas a operações de processamento e transporte se destacam como áreas emergentes, com grande crescimento em um curto intervalo de tempo, liderando o ranking de depósito de patentes até o ano de 1995. Contudo, após o ano de 1995 as patentes classificadas como “Necessidades Humanas” se destacam.

Uma análise relevante sobre PT, consiste em avaliar o número de registros de programa de computador por ano, conforme é apresentado pelo Gráfico 4, considerando registros realizado entre os anos de 2000 e 2019.

Gráfico 4 – Total de registro de programas de computador por ano.



Fonte: Elaboração do autor

Importante destacar que o crescimento de registro de programas de computador ao longo dos anos, acompanha o mesmo período de crescimento no depósito de patentes o que justifica considerar dados de programas de computador com registro na compreensão do cenário tecnológico.

Ainda explorando a base de dados constituída neste trabalho, de acordo com os dados coletados dos currículos da Plataforma Lattes, a maioria dos proponentes da PT são doutores.

5 CONSIDERAÇÕES FINAIS

Diante do exposto, os documentos de patentes e os programas de computador com registros são caracterizados portadores de informações cruciais sobre desenvolvimento tecnológico. No entanto, acessar e manipular grandes volumes de dados de patentes e programas de computador pode ser um desafio. Apesar da existência de repositórios online que disponibilizam acesso, tanto gratuitos quanto pagos, todos apresentam limitações quanto a manipulação de grandes volumes de dados. Diante disso, é evidente a necessidade de alternativas que facilitem tais análises. Portanto, objetivando contribuir com alternativas a este problema, propõe-se um *framework* que visa viabilizar a coleta e o tratamento de dados sobre a produção técnica brasileira. Portanto, através das estratégias definidas neste trabalho foi possível compilar um *framework*, atingindo o objetivo proposto.

Consequentemente, a construção de uma base de dados local que viabiliza análises de diversas magnitudes.

Os Autores esperam que essa abordagem possa contribuir com o avanço científico e tecnológico, incentivando a adoção dos dados coletados e disponibilizados pelo *framework* para diversas novas pesquisas.

REFERÊNCIAS

AMADEI, José Roberto Plácido; TORKOMIAN, Ana Lúcia Vitale. As patentes nas universidades: análise dos depósitos das universidades públicas paulistas. **Ciência da Informação**, v. 38, n. 2, p. 9-18, 2009.

BONETTI, Esther Aquemi. **Inovação e Propriedade Intelectual: Guia para o Docente**, 2010.

BRANDÃO, F. G. **Democratização da informação a partir do uso de repositórios digitais institucionais: da comunicação científica às informações tecnológicas de patentes**. Dissertação (Mestrado) — Universidade Regional Integrada do Alto Uruguai e das Missões. 2016.

BRASIL. Lei Nº 9.609. 19 fev. 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19609.htm. Acesso em: 20 jun. 2023.

BRASIL. Busca de patentes. 08 abr. 2015. Disponível em: <https://www.gov.br/inpi/pt-br/assuntos/informacao/busca-de-patentes>. Acesso em: 20 jun. 2023.

BRASIL. Depósitos e registros de programa de computador crescem em 2021. 2021. Disponível em: <https://bit.ly/46fsU5A>. Acesso em: 22 jun. 2023.

BRASIL. Estatísticas. 2022. Disponível em: <https://www.gov.br/inpi/pt-br/central-de-conteudo/estatisticas/estatisticas>. Acesso em: 20 jun. 2023.

CALZOLAIO, Aziz Eduardo; MATEI, Ana Paula; POHLMANN, Jaime; MENDINA, Heitor José Cademartori; FORGIARINI, Deivid Illecki; GARCIA, Alexandre de Souza. Mapeamento dos registros de propriedade intelectual (patentes) na universidade federal do rio grande do sul. **Revista Brasileira de Gestão e Inovação**, v. 6, n. 1, p. 44–70, 2018. Disponível em: <http://www.uces.br/etc/revistas/index.php/RBGI/article/view/5860>. Acesso em: 17 jun. de 2023.

DIAS, Thiago Magela Rodrigues. **Um Estudo da Produção Científica Brasileira a partir de Dados da Plataforma Lattes**. 181 f. Tese (Doutorado em Modelagem Matemática e Computacional) — Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.

FAGUNDES, Mariana; GARCIA, Pauli; MOTTA, Gustavo. Perfil tecnológico da CSN: um estudo patentométrico. **RAI : Revista de Administração e Inovação**. 2014.

GÜNTHER, Leonardo Lima; PACHECO, Roberto Carlos dos Santos; VARVAKIS, Gregorio; KERN, Vinícius Medina. Análise do sistema de currículo lattes segundo o modelo CESM: perspectivas para um sistema de informação para a e-science. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 10, n. 1, p. 107-130, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/pgc/article/view/46092>. Acesso em: 19 jun. de 2023.

LOPES, Ricardo Cortez. Currículo Lattes: um estudo sobre algumas representações sociais. **PragMATIZES - Revista Latino-Americana de Estudos em Cultura**, Niterói, RJ, v. 12, n. 22, p. 663-694, mar. 2022.

MAZZOLA, Claudio Renato Ribeiro de Mello; ARMELLINI, Fabiano. Desenvolvimento tecnológico do Brasil no cenário mundial: um estudo exploratório por meio de indicadores de patentes. *In*: XXXI Encontro nacional de engenharia de produção, 31, 2011, Belo Horizonte. **Anais [...]**. Disponível em: https://abepro.org.br/biblioteca/enegep2011_tn_sto_142_896_19395.pdf Acesso em: 16 jun. 2023.

NASCIMENTO, Raphael; SPEZIALI, Marcelo. Patentometria: a utilização de dados contidos em patentes como mecanismo de análise da predominância tecnológica dos nits. *in*: IV Encontro Internacional de Gestão, Desenvolvimento e Inovação, 4, 2020. **Anais [...]**. Disponível em: <https://periodicos.ufms.br/index.php/EIGEDIN/article/view/11705> Acesso em: 17 jun. 2023.

PIRES, Edilson Araújo; RIBEIRO, Nubia Moura; QUINTELLA, Cristina M. Sistemas de Busca de Patentes: análise comparativa entre Espacenet, Patentscope, Google Patents, Lens, Derwent Innovation Index e Orbit Intelligence. **Cadernos de Prospecção**, v. 13, n. 1, p. 13, 2020. DOI: 10.9771/cp.v13i1.35147.

QUONIAM, C. T. L.; KNISS, M. R.; MAZIERI. A patente como objeto de pesquisa em ciências da informação e comunicação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 19, n. 39, p. 243– 268, abr. 2014.

SERRANO, B.; JUNIOR, J. A. G. Redes de inovação: mapeamento de inventores de patentes em uma empresa do setor de cosméticos. **Revista GEPROS. Gestão da Produção, Operações e Sistemas**, v. 09, n. 1, p. 101–113, jan 2014.