



UNIVERSIDADE DE BRASÍLIA – UNB
CAMPUS GAMA – FGA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

**WEB SCRAPING NA EXTRAÇÃO E COMBINAÇÃO SISTEMÁTICA DE CONTEÚDOS:
FERRAMENTA AUXILIAR EM PROCESSOS DE
PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I).**

JESSICA SOUSA OLIVEIRA

ORIENTADOR: DR. MÁRIO FABRÍCIO FLEURY ROSA



UNIVERSIDADE DE BRASÍLIA – UNB

FACULDADE UNB GAMA – FGA



**WEB SCRAPING NA EXTRAÇÃO E COMBINAÇÃO SISTEMÁTICA DE CONTEÚDOS:
FERRAMENTA AUXILIAR EM PROCESSOS DE
PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I).**

JESSICA SOUSA OLIVEIRA

ORIENTADOR: DR. MÁRIO FABRÍCIO FLEURY ROSA

DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA BIOMÉDICA

PUBLICAÇÃO: 171A/2023
BRASÍLIA/DF, JUNHO DE 2023

UNIVERSIDADE DE BRASÍLIA - UNB
FACULDADE UNB GAMA - FGA
PROGRAMA DE PÓS-GRADUAÇÃO

**WEB SCRAPING NA EXTRAÇÃO E COMBINAÇÃO SISTEMÁTICA DE CONTEÚDOS:
FERRAMENTA AUXILIAR EM PROCESSOS DE
PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I).**

JESSICA SOUSA OLIVEIRA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA BIOMÉDICA.

APROVADA POR:

DR. MÁRIO FABRÍCIO FLEURY ROSA
(ORIENTADOR)

DR. ADSON FERREIRA DA ROCHA
(EXAMINADOR INTERNO)

DR. PAULO ROBERTO DOS SANTOS
(EXAMINADOR EXTERNO)

DRA. SUÉLIA DE SIQUEIRA RODRIGUES FLEURY ROSA
(EXAMINADOR SUPLENTE)

BRASÍLIA/DF, JUNHO DE 2023

FICHA CATALOGRÁFICA

OLIVEIRA, JESSICA S.

Web Scraping na extração e combinação sistemática de conteúdos: ferramenta auxiliar em processos de Pesquisa, Desenvolvimento e Inovação (PD&I).

[Distrito Federal], 2023.

User Field page_minus_preamble = 70p., 210 x 297 mm (FGA/UnB Gama, Mestrado em Engenharia Biomédica, 2023).

Dissertação de Mestrado em Engenharia Biomédica, Faculdade UnB Gama, Programa de Pós-Graduação em Engenharia Biomédica.

- | | |
|-------------------------|-----------------------------|
| 1. Lei da Inovação | 2. Raspagem da Web |
| 3. Pesquisas Acadêmicas | 4. Translação de Tecnologia |
| I. FGA UnB | II. Título (série) |

REFERÊNCIA

Oliveira, Jessica S. (2023). *Web Scraping* na extração e combinação sistemática de conteúdos: ferramenta auxiliar em processos de Pesquisa, Desenvolvimento e Inovação (PD&I). Dissertação de mestrado em Engenharia Biomédica, Publicação 171A/2023, Programa de Pós-Graduação, Faculdade UnB Gama, Universidade de Brasília, Brasília, DF, 70p.

CESSÃO DE DIREITOS

Autor: Jessica Sousa Oliveira

Título: *Web Scraping* na extração e combinação sistemática de conteúdos: ferramenta auxiliar em processos de Pesquisa, Desenvolvimento e Inovação (PD&I).

Grau: Mestre

Ano: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender essas cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

210007532@aluno.unb.br

Brasília, DF – Brasil

DEDICATÓRIA

Ao meu esposo e melhor amigo, que, além de sempre me apoiar, ilumina o meu caminho com seu amor, me reergue quando tropeço e caminha ao meu lado, sem jamais me abandonar.

AGRADECIMENTOS

Agradeço a Deus por toda proteção, discernimento, evolução e força que Ele me concedeu. Sou grata também por todas as pessoas que Ele colocou em meu caminho, afinal todas possuem algo para nos ensinar. Gratidão por Ele ter me conduzido a todo momento e por ter me auxiliado a enxergar o meu potencial, do qual muitas vezes duvidei.

E para eu parar de duvidar de mim mesma, tenho sempre o meu esposo ao meu lado. Ele é a única pessoa que, a cada instante, me apoia e me ajuda a mostrar ao mundo sempre a minha melhor versão, a pessoa que eu realmente sou, sem medo e sem dúvidas. Sem ele, eu provavelmente nem teria ingressado nesse Mestrado. Hélder, minha vida, obrigada por tudo e por tanto.

Agradeço ao meu orientador, Professor Doutor Mário Fabrício Fleury Rosa, por toda atenção, compreensão, paciência e dedicação incansáveis durante os anos de desenvolvimento do meu trabalho. Sou grata por todo o aprendizado que tive por meio das orientações e dos diálogos engrandecedores.

Gratidão imensa aos membros de minhas bancas examinadoras, Professora Suélia, Professor Paulo Roberto, Professora Glécia e Professora Aldira, por todo o apoio que me deram, pela leitura cuidadosa do documento que tiveram em mãos e por todas as valiosas contribuições que permitiram a melhoria desta dissertação.

Por fim, agradeço ao Programa de Pós-Graduação em Engenharia Biomédica, pela excelente equipe que a compõem, em especial agradeço à Denise e ao Anderson, pela atenção e presteza de sempre.

“Se você quer que coisas ruins parem de acontecer com você, então pare de aceitar qualquer porcaria e exija algo mais.”

Cristina Yang
(Grey’s Anatomy)

RESUMO

Introdução: O Marco Legal da Ciência, Tecnologia e Inovação tem como propósito a contribuição para com o desenvolvimento sustentável do Brasil, coordenando e alocando recursos com vistas a garantir produtos ou estudos científicos, tecnológicos, industriais e comerciais no país. Para que as instituições públicas de ensino superior possam usufruir dos direitos garantidos por este marco, é necessário que evidenciem e tornem públicas suas contribuições para a sociedade incluindo a translação do desenvolvimento científico para usufruto social. Nesse viés, evidencia-se a demanda por uma busca sistemática que enfatize as iniciativas científicas e tecnológicas desenvolvidas por estas instituições que, posteriormente, culminaram em produtos técnicos e tecnológicos para além das publicações, mas principalmente patentes e/ou transferência tecnológicas. **Objetivos:** Surge então, como uma solução tecnológica viável, o desenvolvimento de uma aplicação que automatize esse processo. Este trabalho visa entregar uma ferramenta de mineração de dados, baseada em *Web Scraping*, que realize uma coleta capaz de extrair e combinar conteúdo da web de maneira sistemática e prática, buscando autores de projetos desenvolvidos na Universidade de Brasília que, posteriormente, translacionaram seus resultados para patentes e/ou transferências tecnológicas. **Metodologia:** Para tanto, utilizou-se uma abordagem aplicada, visando solucionar um problema específico que envolve interesses locais. Com base em revisões bibliográficas e em produtos de pesquisas semelhantes, espera-se que o *scraper* facilite a extração de informações acerca de projetos de pesquisa depositados em determinado repositório *online* e que, ao combinar os dados obtidos com os que foram minerados de uma base de patentes e/ou transferências tecnológicas, comprove o retorno de estudos, produções e pesquisas para o desenvolvimento econômico e social do país. **Resultados:** Como resultado da execução da ferramenta no repositório do Instituto Nacional da Propriedade Industrial, obteve-se 783 nomes de inventores de patentes; e a aplicação no repositório da Universidade de Brasília retornou 53.704 autores de projetos desenvolvidos na universidade. Após a comparação das tabelas geradas, considera-se 3.244 registros referentes aos autores que depositaram alguma patente, o que auxilia na análise das pesquisas que se tornaram inovações. **Conclusão:** O desfecho aponta que a ferramenta desenvolvida é funcional ao que se destina, considerando-se também que o cruzamento de informações é totalmente viável, podendo ser incorporado em análises e/ou aperfeiçoamento de processos que maximizem os indicadores de inovação gerados pelas universidades mitigando o Marco Legal da Ciência, Tecnologia e Inovação.

Palavras-chave: lei da inovação, raspagem da *web*, pesquisas acadêmicas, translação de tecnologia.

ABSTRACT

Introduction: The Legal Framework for Science, Technology, and Innovation aims to contribute to the sustainable development of Brazil by coordinating and allocating resources to ensure scientific, technological, industrial, and commercial products or studies in the country. For public institutions of higher education to benefit from the rights guaranteed by this framework, it is necessary for them to demonstrate and make public their contributions to society, including the translation of scientific development for social benefit. In this regard, there is a demand for a systematic search that emphasizes the scientific and technological initiatives developed by these institutions, which subsequently result in technical and technological products beyond publications, but mainly patents and/or technology transfer. **Objectives:** Therefore, the development of an application that automates this process emerges as a viable technological solution. This work aims to deliver a data mining tool based on web scraping that performs systematic and practical data collection by extracting and combining web content, seeking authors of projects developed at the University of Brasília who have subsequently translated their results into patents and/or technology transfers. **Methodology:** To this end, an applied approach was used, aiming to solve a specific problem involving local interests. Based on literature reviews and similar research products, it is expected that the scraper will facilitate the extraction of information about research projects deposited in a specific online repository and, by combining the obtained data with those mined from a database of patents and/or technology transfers, demonstrate the return of studies, productions, and research to the country's economic and social development. **Results:** As a result of running the tool on the repository of the National Institute of Industrial Property, we obtained 783 patent inventors' names. The application on the repository of the University of Brasília returned 53,704 authors of projects developed at the university. After comparing the generated tables, we identified 3,244 records corresponding to authors who filed patents, which assists in the analysis of research that led to innovations. **Conclusion:** The outcome indicates that the developed tool is functional for its intended purpose, considering that the cross-referencing of information is entirely feasible and can be incorporated into analyses and/or process improvements that maximize the innovation indicators generated by universities, mitigating the Legal Framework for Science, Technology, and Innovation.

Keywords: innovation law, data mining, academic research, technology translation.

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 Hipótese	4
1.2 Objetivos.....	5
1.2.1 Objetivo Geral	5
1.2.2 Objetivos Específicos	5
1.3 Justificativa	5
2 FUNDAMENTAÇÃO TEÓRICA.....	9
2.1 Protocolo HTTP: Cliente e servidor	9
2.2 Páginas <i>web</i> e repositórios digitais	11
2.3 Propriedade intelectual	14
2.4 Patentes	16
2.4.1 Natureza das patentes	18
2.4.2 Classificação das patentes	19
2.5 Interface de Programação de Aplicação	21
2.6 <i>Web Scraping</i>	21
2.7 Aspectos legais e éticos do <i>Web Scraping</i>	22
3 MATERIAIS E MÉTODOS	23
3.1 Visão geral das tecnologias utilizadas	23
3.2 Python	25
3.3 Beautiful Soup	26
3.4 Selenium	26
3.5 Pandas e NumPy	27
3.6 Sistema Gerenciador de Banco de Dados MySQL.....	27
3.7 MySQL Connector.....	29
3.8 Máquinas Virtuais	30
4 RESULTADOS E DISCUSSÕES	32
4.1 Aquisição dos dados	32
4.1.1 Coleta de dados no repositório da UnB.....	33
4.1.2 Coleta de dados no INPI.....	34
4.2 Tratamento e organização dos dados	37
4.3 Comparação dos dados de interesse	39
4.4 Discussões.....	43
5 CONSIDERAÇÕES FINAIS.....	48
5.1 Trabalhos futuros	50
5.2 Limitações encontradas	50

REFERÊNCIAS BIBLIOGRÁFICAS	51
APÊNDICE I	62
APÊNDICE II.....	63
APÊNDICE III.....	64
APÊNDICE IV	65
ANEXO I	66
ANEXO II.....	67
ANEXO III	68

LISTA DE FIGURAS

Figura 1.1. Principais diferenças entre a mineração de dados e a busca manual.....	7
Figura 2.1. Transação HTTP.	9
Figura 2.2. Métodos HTTP.....	10
Figura 2.3. Exemplo de código fonte HTML.	11
Figura 2.4. Modalidades e subáreas da Propriedade Intelectual.	16
Figura 2.5. Exemplo de classificação IPC.	20
Figura 3.1. Visão geral das tecnologias utilizadas.	24
Figura 3.2. Principais linguagens de programação do IEEE Spectrum 2022.	25
Figura 3.3. Principais bancos de dados, de acordo com o <i>DB-Engines Ranking</i>	28
Figura 3.4. Concepção de uma estrutura de virtualização.	31
Figura 4.1. Visão geral da coleta de dados.....	33
Figura 4.2. Pseudocódigo da coleta de dados no repositório da UnB.....	34
Figura 4.3. Diagrama de entidade e relacionamento do banco de dados local criado.	38
Figura 4.4. Primeira sintaxe criada para comparar as tabelas da UnB e do INPI.....	40
Figura 4.5. Segunda sintaxe criada para comparar as tabelas da UnB e do INPI.....	41
Figura 4.6. Repetição de registro nas tabelas do banco de dados local.	44
Figura 4.7. Representação dos primeiros registros presentes no RIUnB.....	45

LISTA DE TABELAS E GRÁFICOS

Tabela 4.1 – Teste de coleta manual de dados na base de dados do INPI.	35
Tabela 4.2 – Quantitativo referente aos nomes registrados nas bases garimpadas.	42

LISTA DE NOMENCLATURAS E ABREVIações

API	<i>Application Programming Interface</i> (Interface de Programação de Aplicação).
BCE	Biblioteca Central.
BNDES	Banco Nacional de Desenvolvimento Econômico e Social.
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
CDT	Centro de Apoio ao Desenvolvimento Tecnológico da Universidade de Brasília.
CMS	<i>Content Management System</i> (Sistema de Gestão de Conteúdos).
CNIPA	<i>China National Intellectual Property Administration</i> (Administração Nacional de Propriedade Intelectual da China).
CNPJ	Cadastro Nacional da Pessoa Jurídica.
CPC	<i>Cooperative Patent Classification</i> (Classificação Cooperativa de Patentes).
CSV	<i>Comma-separated values</i> (Valores Separados por Vírgula).
CT&I	Ciência, Tecnologia e Inovação.
EPO	<i>European Patent Office</i> (Escritório Europeu de Patentes).
HTML	<i>HyperText Markup Language</i> (Linguagem de Marcação de Hipertexto).
HTTP	<i>Hypertext Transfer Protocol</i> (Protocolo de Transferência de Hipertexto).
IBM	<i>International Business Machines Corporation</i> (Corporação Internacional de Máquinas de Negócios).
IEEE	<i>Institute of Electrical and Electronics Engineers</i> (Instituto de Engenheiros, Eletricistas e Eletrônicos).
IES	Instituição de Ensino Superior.
INPI	Instituto Nacional da Propriedade Industrial.
IPC	<i>International Patent Classification</i> (Classificação Internacional de Patente).
JPO	<i>Japan Patent Office</i> (Escritório Japonês de Patentes).
LGPD	Lei Geral de Proteção de Dados Pessoais.
MLCTI	Marco Legal da Ciência, Tecnologia e Inovação.
NIT	Núcleo de Inovação Tecnológica.
OMPI	Organização Mundial da Propriedade Intelectual.
OMR	<i>Object Relational Mapping</i> (Mapeamento Objeto Relacional).
ONU	Organização das Nações Unidas
PCT	<i>Patent Cooperation Treaty</i> (Cooperação em Matéria de Patentes).
PD&I	Pesquisa, Desenvolvimento e Inovação.
PSF	<i>Python Software Foundation</i> .
PTS	Pesquisa Translacional em Saúde.

RAM	<i>Random Access Memory</i> (Memória de Acesso Randômico).
RFC	<i>Request for Comments</i> (Pedido para Comentários).
RI	Repositório Institucional.
RIUnB	Repositório Institucional da Universidade de Brasília.
SGBD	Sistema Gerenciador de Banco de Dados.
SO	Sistema Operacional.
SQL	<i>Structured Query Language</i> (Linguagem de Consulta Estruturada).
TT	Transferência Tecnológica.
UnB	Universidade de Brasília.
URL	<i>Uniform Resource Locator</i> (Localizador Uniforme de Recursos).
USPTO	<i>United States Patent and Trademark Office</i> (Instituto de Marcas e Patentes dos Estados Unidos).
VM	<i>Virtual Machine</i> (Máquina Virtual).
WIPO	<i>World Intellectual Property Organization</i> (Organização Mundial da Propriedade Intelectual).
WS	<i>Web Scraping</i> (Raspagem da Web).
WSL	<i>Windows Subsystem for Linux</i> (Subsistema Windows para Linux).

1 INTRODUÇÃO

O Marco Legal da Ciência, Tecnologia e Inovação (MLCTI), legislação federal promulgada para regulamentar as atividades de sua área de foco, tem por finalidade contribuir para o desenvolvimento sustentável do país, por meio de uma melhor coordenação e alocação de recursos financeiros, humanos e materiais, visando garantir produtos ou estudos científicos, tecnológicos, industriais e comerciais no Brasil. Tem como foco geral implantar boas práticas de gestão e políticas que estimulem a pesquisa e o desenvolvimento nas áreas de ciência e tecnologia, inovando e estabelecendo mecanismos dedicados a viabilizar os investimentos nessas áreas (BRASIL, 2016; BRASIL, 2018).

Segundo Rosa (2022), a importância do Marco Legal da Inovação está entre variáveis não excludentes sobre o setor de Pesquisa, Desenvolvimento e Inovação (PD&I) brasileiro:

Quando pensamos sobre o espectro da inovação no Brasil, vemos a chance de explorar algumas dimensões. Primeiro, sob a ótica cultural, podemos inferir aspectos que se debruçam acerca do modelo norte-americano Research Universities, interpelado pela teoria do quadrante de pasteur presente nas pautas mais progressistas do fazer científico nacional; caso queiramos explorar a inovação pela visão do setor de Ciência, Tecnologia e Inovação (CT&I) pautadas pelo Estado Nacional, ações desenhadas, como FINEP, BNDES, Procis, PPSUS, CNPq, Fundação de Apoio à Pesquisa (FAP), entre outras, nos orientamos para o centro do domínio da quadrupla hélice. Ajustando nosso foco, podemos especular inovações na seara jurídica; a institucionalização do Marco Legal Regulatório da Inovação representado pelas Lei de Inovação Tecnológica e a Lei do Bem, que contribuem com a harmonização dos fazeres e práticas que envolvem o processo de PD&I, integrando universidades, indústrias e/ou iniciativas privadas com políticas públicas. (ROSA, 2022)

O MLCTI, em perspectiva, contribui para que os fazeres e práticas de PD&I realizados na universidade operem o processo de Pesquisa Translacional em Saúde (PTS), ou seja, a transformação do desenvolvimento científico e tecnológico realizados pelas universidades em produtos capazes de serem consumidos pela demanda constante da sociedade (ROSA, 2022). Além disso, de acordo com Barreto *et al.* (2019), a translação é considerada bem-sucedida quando a pesquisa de base e/ou aplicada é convertida em um produto com potencial comercial, que atenda às demandas da população nos sistemas de saúde.

Para que as instituições públicas de ensino superior sejam capazes de realizar a translação de suas pesquisas para a sociedade, são necessários os instrumentos dispostos no MLCTI. Entretanto, para que possam usufruir dos direitos garantidos por este marco, é importante que apresentem e operem facilitadores que contribuam para o ecossistema de translação de forma clara, a fim de que se tornem evidentes o retorno de suas produções e pesquisas para o desenvolvimento econômico e social do país, competência essa atribuída ao Núcleo de Inovação Tecnológica (NIT), que, conforme a Lei Nº 10.973/2004, em seu artigo 16, tem por finalidade gerir a política de inovação da instituição de ensino, como segue

I - zelar pela manutenção da política institucional de estímulo à proteção das criações, licenciamento, inovação e outras formas de transferência de tecnologia; II - avaliar e classificar os resultados decorrentes de atividades e projetos de pesquisa para o atendimento das disposições desta Lei; III - avaliar solicitação de inventor independente para adoção de invenção na forma do art. 22; IV - opinar pela conveniência e promover a proteção das criações desenvolvidas na instituição; V - opinar quanto à conveniência de divulgação das criações desenvolvidas na instituição, passíveis de proteção intelectual; VI - acompanhar o processamento dos pedidos e a manutenção dos títulos de propriedade intelectual da instituição; VII - desenvolver estudos de prospecção tecnológica e de inteligência competitiva no campo da propriedade intelectual, de forma a orientar as ações de inovação da ICT; VIII - desenvolver estudos e estratégias para a transferência de inovação gerada pela ICT; IX - promover e acompanhar o relacionamento da ICT com empresas, em especial para as atividades previstas nos arts. 6º a 9º; X - negociar e gerir os acordos de transferência de tecnologia oriunda da ICT. (BRASIL, 2004, art. 16)

Além disso, as instituições de ensino superior (IES) devem seguir as diretrizes do MLCTI implementando políticas, práticas e/ou facilitadores que permitam o acesso à informação e aos conhecimentos científicos, visando o amadurecimento do ecossistema e maximizando incorporação para uso social (BRASIL, 2016). Algumas das práticas alinhadas a essas diretrizes são: criar e implementar uma política de inovação que oriente as atividades de pesquisa, desenvolvimento tecnológico e transferência de tecnologia, e promover a interação com o setor produtivo, visando a geração de resultados práticos para a sociedade; criar e fortalecer os NITs; incentivar seus pesquisadores a protegerem a propriedade intelectual resultante de suas pesquisas; estabelecer parcerias com empresas e outros atores do setor produtivo, buscando a colaboração em projetos de pesquisa e desenvolvimento tecnológico, seja por meio de convênios, contratos de pesquisa, ou outros instrumentos; e, por fim, oferecer programas de formação e capacitação que preparem os pesquisadores para atuarem na área de

ciência, tecnologia e inovação, e isso inclui estimular a participação em projetos de pesquisa, promover a interdisciplinaridade e incentivar a criação de startups e empreendedorismo. (BRASIL, 2004; BRASIL, 2018).

Dentre os meios de divulgação e/ou materialização de conhecimentos científico e tecnológico que as IES normalmente utilizam estão a publicação de artigos científicos e o depósito de patentes. Estes possuem pontos em comum, tais como: necessitam da validação feita por avaliadores; e proporcionam meios de registro de autoria ou propriedade do conhecimento. Entretanto, esses meios possuem duas diferenças importantes, que são o direito à propriedade e o direito de acesso e uso, incluindo exploração comercial. Enquanto os artigos científicos, divulgados por revistas, periódicos e afins, não restringem o uso do conhecimento contido nos mesmos, desde que citados de maneira apropriada, as patentes são obtidas por meio de depósito do documento específico para tal, os direitos sobre seu conteúdo são concedidos a quem deposita, não necessariamente ao autor, e o uso por terceiros envolve pagamentos ao detentor dos direitos sobre a patente (MUELLER; PERUCCHI, 2014).

De sobremaneira, a patente significa, em muitos casos, a perspectiva em transladar o conhecimento para a iniciativa privada, como amplamente divulgado. E essa etapa translacional, ou seja, da bancada para a indústria, significa coerência com a perspectiva do MLCTI, uma vez que a pesquisa pode prosperar para a transferência tecnológica (TT). De acordo com Almeida e Pinheiro (2020), a TT é um processo essencial para conectar a pesquisa científica e os avanços tecnológicos à aplicação prática no mundo real, desempenhando um papel crucial no avanço da sociedade, impulsionando a inovação, promovendo o crescimento econômico e resolvendo desafios sociais e ambientais. Conforme descrito por Brasil (2023), envolve a transferência de conhecimento, tecnologia, métodos e recursos de uma entidade (como uma universidade, instituto de pesquisa ou empresa) para outra, que pode utilizar e desenvolver essas informações, onde o objetivo principal é converter ideias e descobertas em produtos, processos ou serviços tangíveis que beneficiem a sociedade como um todo.

Neste ínterim, em virtude da necessidade de evidenciar como a Universidade de Brasília (UnB) contribui ativamente com PD&I, surge a demanda de efetuar uma busca cruzada em bases de dados, que permita que seja mensurado o número de projetos de PD&I desenvolvidos na UnB, presentes no repositório da instituição, vinculados às patentes depositadas pela mesma.

Ou seja, lançar luz nessas duas categorias, a saber: produto publicado (projetos de pesquisa e artigos) e produto patenteado (depósito de patentes), pode ser uma ação contributiva nas políticas de inovação das universidades afim de aumentar seus indicadores de translação. Entende-se que uma busca humanizada dessas informações requereria demasiado tempo, além do fato de que o cruzamento manual estaria sujeito a falhas, uma vez que a quantidade de trabalhos e patentes a ser verificada é expressiva. Nesse viés, o desenvolvimento de uma ferramenta que automatize este processo surge como uma solução tecnológica viável, para que não apenas a UnB, em suas diversas ações de PD&I, seja favorecida, mas também outras instituições que necessitem de um modo de quantificar suas contribuições ativas para com a sociedade, medidas por essas duas categorias.

A partir da base de dados do Instituto Nacional de Pesquisa e Inovação (INPI), única fonte pública de armazenamento referente às patentes e demais temáticas relacionadas à propriedade industrial no Brasil, e do repositório institucional da UnB, que dispõe de todos os trabalhos desenvolvidos por seus discentes em todas as áreas de estudo e formação, compreendeu-se a necessidade de buscar e combinar conteúdos específicos da *web* de modo a economizar tempo e custos, poupando todo e qualquer trabalho manual excessivo e possibilitando entregar resultados com menor índice de eventuais falhas humanas.

Nesse contexto, a presente pesquisa de mestrado procurou construir uma ferramenta, baseada em *Web Scraping*, capaz de minerar o bloco de produções acadêmicas e científicas realizadas pela UnB, indexadas em seu repositório, e as informações referentes às patentes depositadas no INPI, de forma a evidenciar os autores de pesquisas e inventores de patentes presentes em ambas as bases, a fim de facilitar a identificação de quais estudos transladaram da fase conceitual para a fase de proteção intelectual, visto que a patente é considerada a antessala da transferência tecnológica que, por sua vez, antecede o processo de produção em escala industrial.

1.1 HIPÓTESE

Uma ferramenta de coleta de dados, capaz de extrair informações da base do INPI e do repositório da UnB, permite a comparação entre produtos acadêmicos e registros de patentes,

através dos autores e inventores de estudos, pesquisas e inovações. Com ela, é possível identificar iniciativas de PD&I que resultaram exitosamente em proteções intelectuais e/ou registros de patentes e auxiliar na tomada de decisão de gestores da UnB para a política de PD&I da instituição?

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Desenvolver uma ferramenta de mineração de dados que seja capaz de extrair registros oriundos dos repositórios do INPI e da UnB, permitindo a combinação dessas informações de forma a auxiliar pesquisadores, cientistas e gestores a encontrar e evidenciar as áreas e segmentos que mais converteram a fase de pesquisa de base em patentes.

1.2.2 Objetivos Específicos

- Criar na linguagem de programação Python, o código de extração das informações referentes aos trabalhos desenvolvidos existentes no repositório da UnB e das patentes disponibilizadas na base de dados do INPI;
- Modelar um banco de dados local com tabelas específicas para o armazenamento e consulta dos dados oriundos das aplicações criadas;
- Exportar, para as tabelas criadas, os dados extraídos pelas aplicações desenvolvidas;
- Elaborar um algoritmo interno ao banco de dados que resulte no cruzamento dos registros obtidos para evidenciar, por meio dos autores, os projetos convertidos em depósitos de patentes, possibilitando melhor análise das informações extraídas;
- Análise das informações extraídas por meio de arquivo CSV a ser gerado ao final do cruzamento dos dados.

1.3 JUSTIFICATIVA

Com a ascensão e disseminação da internet e a grande quantidade de dados gerados por diferentes tipos de usuários, houve um aumento exponencial de informações produzidas

diariamente em várias plataformas e em todos os âmbitos da sociedade moderna. Esse aumento significativo tornou imperativo o desenvolvimento de técnicas de processamento e análise de informações, afinal, a popularização da internet em todo o mundo, transformou-a no maior repositório de dados e informações já conhecido (BHARDWAJ et al., 2021).

Uma das técnicas que auxilia na análise de dados é a mineração. Ao tentar conceituar o termo mineração de dados, é possível perceber que ele varia de acordo com o campo de atuação de cada autor, sendo assim, serão apresentados os que são considerados mais expressivos: no campo da Estatística, do Aprendizado de Máquina e de Banco de Dados (CAMILO; SILVA, 2009). À luz da Estatística, mineração de dados é conceituada como a análise de grandes conjuntos de dados, no intuito de encontrar relações inesperadas e de resumi-los de uma maneira que permaneçam úteis e compreensíveis ao proprietário dos mesmos (HAND *et al.*, 2001). Na perspectiva do aprendizado de máquina, Fayyad *et al.* (1996) explica que é um passo no processo da realização da análise de dados e na aplicação de algoritmos que produzem um conjunto de padrões determinados. Por fim, com vistas ao banco de dados, a mineração une diversas técnicas da área da computação e estatística, visando a extração de informações em massa do banco (CABENA *et al.* 1998).

A coleta automatizada de dados da Internet pode ser considerada tão antiga quanto a própria Internet e o termo "raspagem" já existe há muito mais tempo do que a *web*. Antes de "raspagem da *web*" se popularizar como um termo, uma prática conhecida como "raspagem de tela" já era bem estabelecida como uma forma de extração de dados de uma representação visual, onde, nos primórdios da computação, muitas vezes se reduziam a simples terminais baseados em texto. Assim como hoje, as pessoas daquela época também estavam interessadas em "raspar" grandes quantidades de texto de terminais pré-estabelecidos e armazenar estes conteúdos para uso posterior (BROUCKE; BAESSENS, 2018). Atualmente, um dos métodos mais importantes e predominantes de extrair dados relevantes da rede mundial de computadores é o *Web Scraping*. A raspagem da *web* ganhou popularidade significativa devido à facilidade que oferece na extração de informações de páginas da *web* e na apresentação destas em um formato estruturado, sem intervenção manual (BHARDWAJ *et al.*, 2021).

Importante ressaltar que Mayer-Schönberger e Cukier (2013), Mitchell (2018) e Grus (2019) argumentam que a coleta automatizada de dados é essencial para lidar com a grande

quantidade de informações geradas diariamente na Internet, além de permitir descobertas e *insights* que seriam impossíveis com a coleta manual. Munzert *et al.* (2015) corrobora com essa perspectiva e destaca a eficiência e precisão desse tipo de coleta, aspectos que podem ser grandes diferenciais atualmente. É possível compreender as principais diferenças entre a mineração de dados e a busca manual com a Figura 1.1.

	Mineração de dados	Busca manual
Velocidade	Processa grandes quantidades de dados em questão de segundos ou minutos.	Dependendo do caso, pode levar horas ou dias.
Escala	É capaz de lidar com grandes conjuntos de dados.	É limitada pelo tempo e recursos disponíveis.
Precisão	Encontra padrões e tendências escondidos nos dados.	Pode ser subjetiva e suscetível a erros.
Personalização	Permite criação de consultas e análises personalizadas, de acordo com necessidades.	Geralmente limitada ao que o usuário consegue encontrar e compreender.

Figura 1.1. Principais diferenças entre a mineração de dados e a busca manual.

Ao minerar o conteúdo disponível na base de dados do INPI e no repositório da UnB, a ferramenta criada auxiliará pesquisadores e cientistas, principalmente aqueles na área de gestão, a cruzarem os dados obtidos, de modo a proporcionar a análise e o levantamento de informações sobre a aplicação dos estudos e pesquisas já realizados e patenteados, visando auxiliar no desenvolvimento de tecnologias e tratamentos inovadores que, diante do cenário social, poderão ser de mais fácil acesso às pessoas ou de melhor adesão pelos profissionais aos quais as criações serão voltadas.

A autora que desenvolveu a ferramenta de mineração acredita que ela poderá ser aplicada por gestores da UnB para que avaliem dois eixos: quantidade de patentes registradas corretamente pelos setores competentes e quais pesquisas realizadas permitem complementos e/ou atualizações. Mostra-se de grande importância, nos anos de pesquisa da instituição, ter o quantitativo correto e atualizado sobre as pesquisas convertidas em depósitos de patentes, uma vez que essa informação é um fator primordial para o reconhecimento da capacidade de

produção de inovações e para a obtenção de recursos e parcerias, uma vez que a quantidade e a qualidade das patentes depositadas podem demonstrar seu potencial de inovação e de transferência de tecnologia para o setor produtivo, o que pode atrair investimentos e parcerias com empresas e órgãos governamentais (BRASIL, 2016; BRASIL, 2018).

Além disso, obter de forma rápida as informações cruzadas com as pesquisas já criadas, auxilia na busca de complementos e atualizações futuras em trabalhos já produzidos, o que pode gerar economia de tempo e de recursos, promovendo uma qualidade diferenciada em estudos futuros, o que é de extrema relevância no que se refere ao compromisso da instituição para com a sociedade, e possibilita que seja verificado se as pesquisas produzidas estão adequadas ou não com as demandas mais urgentes da sociedade em sua plenitude.

Diante do exposto, além das suas contribuições para a análise de dados referentes à inovação na UnB, a pesquisa em tela se encontra em consonância com a Lei nº 10.973/2004, também conhecida como Lei da Inovação, que estabelece normas de incentivo ao progresso no setor produtivo, bem como estimula a pesquisa, o desenvolvimento tecnológico e a inovação no país, promovendo, dessa forma, a transferência de conhecimento e tecnologia entre as universidades, centros de pesquisa e o setor empresarial (BRASIL, 2004), pois, ao desenvolver uma ferramenta que permite o fácil acesso e a análise de dados sobre estudos, pesquisas e patentes, o trabalho contribui diretamente para a valorização e o aproveitamento do conhecimento científico, tecnológico e contemporâneo gerado na instituição acadêmica em foco.

Por fim, ao simplificar a identificação de tecnologias e pesquisas com potencial de aplicação prática, a ferramenta proposta promove a transferência de tecnologia e a aproximação entre o meio acadêmico e o setor produtivo, facilitando a cooperação entre esses diferentes atores, com o propósito de impulsionar a competitividade, o crescimento econômico e a melhoria da qualidade de vida da população.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo, descreve-se o Protocolo de Transferência de Hipertexto (HTTP, do inglês *Hypertext Transfer Protocol*), que é utilizado para obter e enviar dados na *web*; a definição de páginas *web*, de repositórios digitais, tal como o repositório institucional da UnB, e bases de dados; discorre-se sobre propriedade intelectual e patentes; trata brevemente sobre Interface de Programação de Aplicação (API, do inglês *Application Programming Interface*); apresenta a técnica *Web Scraping* e, por fim, apresenta os aspectos legais e éticos do seu uso.

2.1 PROTOCOLO HTTP: CLIENTE E SERVIDOR

A *web* possui regras que estabelecem a comunicação e a troca de informações entre dispositivos. Uma das formas que define o acesso de dados é o HTTP, que opera na camada de aplicação, definido no *Request for Comments* (RFC, Pedido para Comentários, em português) 1945¹ e no RFC 2616², executado entre o cliente, que faz as requisições, e o servidor, que atende às solicitações (KUROSE; ROSS, 2013).

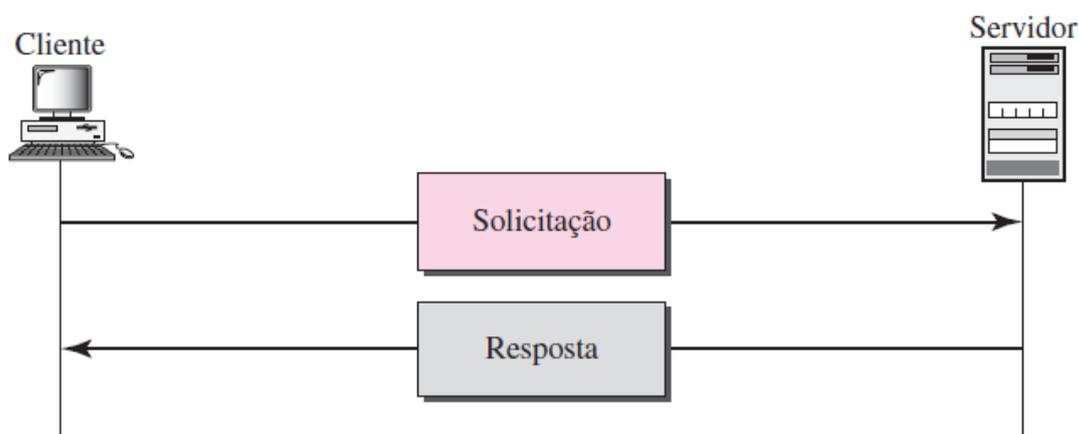


Figura 2.1. Transação HTTP.
Fonte: (FOROUZAN, 2010).

¹ < <https://www.rfc-editor.org/rfc/rfc1945> >

² < <https://www.rfc-editor.org/rfc/rfc2616> >

A Figura 2.1 apresenta uma concepção da interação entre a comunicação cliente-servidor. O cliente consiste em três partes: um controlador, um programa cliente e um interpretador. O controlador recebe as entradas do teclado ou do mouse e usa os programas clientes para acessar os documentos. Após isto, o controlador exibe o documento na tela, utilizando um dos interpretadores disponíveis. Em contrapartida, o servidor está do outro lado do canal de comunicação e, sempre que recebe uma solicitação, o documento correspondente é localizado e enviado ao cliente rapidamente. Em outras palavras, o fluxo da interação se inicia no cliente, através da entrada de dados e comandos. Após o comando inicial, as informações são capturadas e uma solicitação é enviada ao servidor. Por fim, a resposta retornada é apresentada ao cliente por meio do interpretador (FOROUZAN, 2010).

De acordo com Tanenbaum, Feamster e Wetherall (2021), o HTTP é um protocolo simples, que aceita operações denominadas métodos, ou, como também são chamados, verbos, apresentados na Figura 2.2, que são diferentes das solicitações simples de uma página Web.

<i>Método(Comando)</i>	<i>Ação</i>
GET	Solicita um documento ao servidor
HEAD	Solicita informações sobre um documento, mas não o documento em si
POST	Envia informações do cliente para o servidor
PUT	Envia um documento do servidor para o cliente
TRACE	Ecoa uma solicitação que chega
CONNECT	Reservado
OPTION	Solicita detalhamento sobre opções disponíveis

Figura 2.2. Métodos HTTP.
Fonte: (FOROUZAN, 2010).

Um dos métodos mais comuns é o GET, utilizado quando o cliente solicita um objeto e o mesmo é passado no cabeçalho da requisição, podendo ser visualizado no campo do URL (*Uniform Resource Locator* - Localizador Uniforme de Recursos). Outro método muito utilizado é o POST, onde o cliente continua solicitando um objeto ao servidor, mas dependente

do seu pedido, como, por exemplo, quando um usuário fornece palavras-chave a um site buscador (KUROSE; ROSS, 2013).

2.2 PÁGINAS WEB E REPOSITÓRIOS DIGITAIS

Uma página *web*, uma das respostas que o servidor HTTP pode fornecer, é um documento escrito com o uso da Linguagem de Marcação de Hipertexto (HTML, do inglês *Hypertext Markup Language*). Essa linguagem tem por responsabilidade demarcar a estrutura de uma página *web*, ou seja, organizar os elementos desta, de forma que eles se conectem e resultem em algo coerente. Estes elementos HTML, também conhecidos por *tags*, são utilizados para informar ao navegador que tipo de elemento foi construído, podendo ser um título, uma imagem, um link, um botão, entre outros (MOURA; HENRIQUE, 2021). Na Figura 2.5 é possível visualizar um código fonte HTML simples (à esquerda) e sua interpretação por meio de um navegador da *web* (à direita), ficando compreensível a finalidade de algumas *tags* da linguagem HTML.



Figura 2.3. Exemplo de código fonte HTML.
Fonte: autoria própria.

Uma página *web* pode ser desenvolvida para ser estática ou dinâmica, aquela quando o arquivo HTML é hospedado no servidor *web* apenas para leitura, e esta permitindo que a página receba entradas dos usuários, atualize o HTML e mostre conteúdos personalizados para cada

usuário, sem que seja necessário modificações manuais (BENTO, 2014). É o conjunto destas páginas *web* que compõe um site.

Existem dois tipos de sites muito utilizados no meio acadêmico: os repositórios institucionais (RI) e as bases de dados. Os RI possuem papel fundamental no livre acesso à produção científica de um país, sendo uma espécie de biblioteca digital que tem por escopo guardar, preservar e garantir livre acesso, via internet, à produção científica e acadêmica de uma determinada instituição (MARCONDES; SAYÃO, 2009).

O RI em foco para a pesquisa é o da UnB, regulamentado pela Resolução da Reitoria de Nº 0101/2013³, que estabelece as diretrizes para a Política de Informação do Repositório Institucional da Universidade de Brasília (RIUnB) e determina a Biblioteca Central (BCE) como principal responsável pelo desenvolvimento, implementação, manutenção, hospedagem e atualização do RIUnB, com o objetivo de preservar e armazenar toda a produção científica em formato digital. Além disso, a resolução visa preservar a produção científica da instituição, bem como ampliar sua visibilidade; fomentar o intercâmbio com outras instituições de ensino, pesquisa e extensão, visando promover uma maior colaboração e cooperação entre elas; e, além dos outros benefícios citados na resolução, objetiva também acelerar o desenvolvimento das pesquisas produzidas no âmbito da UnB e otimizar a gestão de investimentos nas mesmas.

Já as bases de dados são coleções eletrônicas com alta capacidade de armazenamento e organizadas de tal maneira que a consulta é confiável, fácil e rápida. Estas reúnem revistas científicas, artigos e demais documentos pertinentes à sua finalidade, e podem ser multitemáticas, englobando diversas áreas de conhecimento, ou específicas, focando em determinado assunto (BIBLIOTECA CENTRAL, 2022). Para o trabalho em tela, as bases de dados de maior interesse, no momento, são as que disponibilizam a consulta de documentos de patentes, em especial, a do INPI, que é a única focada em patentes depositadas no Brasil, disponibilizando assim o acesso a pedidos de patentes que não estão em sigilo, patentes já concedidas e patentes arquivadas. Além desta, existem diversas bases mundiais, destacando-se algumas, a saber:

³ < [https://repositorio.unb.br/documentos/Resolucao da Política de Informacao do RIUnB.pdf](https://repositorio.unb.br/documentos/Resolucao_da_Politica_de_Informacao_do_RIUnB.pdf) >

- **Espacenet**⁴, mantido pelo Escritório Europeu de Patentes (EPO, do inglês *European Patent Office*), amplamente utilizado por sua popularidade para busca de documentos de patentes, possui mais de 140 milhões de documentos de patentes de todo o mundo;
- **Escritório Japonês de Patentes**⁵ (JPO, do inglês *Japan Patent Office*), para as buscas das patentes japonesas;
- **Latipat**⁶, voltado para buscas de patentes da América Latina e Espanha;
- **Patentscope**⁷, ferramenta de buscas da Organização Mundial da Propriedade Intelectual (OMPI) que oferece acesso gratuito a documentos de diversos países, tais como Austrália, China, Colômbia, Estados Unidos, Índia, Japão, México e Brasil, e inclui os depósitos realizados via Tratado de Cooperação em matéria de Patentes (PCT, do inglês *Patent Cooperation Treaty*);
- **Administração Nacional de Propriedade Intelectual da China**⁸ (CNIPA, do inglês *China National Intellectual Property Administration*), para buscas de patentes depositadas na República Popular da China;
- **Instituto de Marcas e Patentes dos Estados Unidos**⁹ (USPTO, do inglês *United States Patent and Trademark Office*), para buscas em patentes depositadas nos Estados Unidos da América (EUA, do inglês *United States of America*);
- **Google Patents**¹⁰, mantido pelo Google, inclui mais de 120 milhões de publicações de patentes de mais de 100 escritórios em todo o mundo. Possui, de acordo com a própria plataforma, patentes de todo o mundo, além de artigos e livros técnicos, organizados e rotulados automaticamente.

⁴ <<https://worldwide.espacenet.com/patent/>>

⁵ <<https://www.jpo.go.jp/e/>>

⁶ <<https://lp.espacenet.com/>>

⁷ <<https://patentscope.wipo.int/search/en/search.jsf>>

⁸ <<https://pss-system.cponline.cnipa.gov.cn/conventionalSearch>>

⁹ <<https://ppubs.uspto.gov/pubwebapp/static/pages/landing.html>>

¹⁰ <<https://patents.google.com/>>

2.3 PROPRIEDADE INTELECTUAL

Buscando uma melhor compreensão do assunto em análise, é necessária a conceituação em separado das palavras “propriedade” e “intelectual”. Define-se “propriedade”, no dicionário da língua portuguesa, como “algo que se possui, se tem a posse exclusiva” (PROPRIEDADE, 2023). Conforme a definição de Serrano (2021), em seu dicionário jurídico, é o “direito legítimo ou poder de gozar e dispor de uma coisa com exclusão” e o “domínio ou direito legítimo de dispor ou transferir os bens”. Dessa forma, é possível compreender que o detentor da propriedade possui liberdade, desde que não transgrida à lei, para utilizá-la da maneira que lhe convier e impedir que terceiros a utilizem. Já o adjetivo “intelectual” significa “possuir dotes de espírito, de inteligência” (FERREIRA, 2010). Sendo assim, ao juntar ambas as palavras e analisar o termo “propriedade intelectual”, pode-se concluir que o mesmo consiste em possuir direitos sobre tudo o que é considerado inventividade e/ou conhecimento, ou seja, de acordo com Jungmann e Bonetti (2010a), o termo trata das “criações do espírito humano e dos direitos de proteção dos interesses dos criadores sobre suas criações”.

A Organização Mundial da Propriedade Intelectual (WIPO, do inglês *World Intellectual Property Organization*) foi criada em 1967, pela Organização das Nações Unidas (ONU), e entrou em vigor em 1970. Foi motivada pela necessidade de se estabelecer uma estrutura internacional para a proteção dos direitos de propriedade intelectual, portanto, trabalha para facilitar a cooperação nessa área, estabelecendo padrões globais e oferecendo serviços de registro e gerenciamento, além de fornecer treinamento e assistência técnica aos países membros, definindo como propriedade intelectual:

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

De acordo com Jungmann e Bonetti (2010), a propriedade intelectual pode ser dividida em três grandes grupos: direito autoral, propriedade industrial e proteção *sui generis*, e as mesmas podem ser compreendidas a seguir:

- **Direito autoral:** protege obras intelectuais como livros, músicas, filmes, softwares, entre outros, ou seja, as formas de expressão da obra e não somente as ideias isoladas. Confere ao autor dois tipos de direitos: o moral, que é irrenunciável e se refere ao direito de ter seu nome vinculado à obra; e o patrimonial, que regula a utilização econômica da obra e pode ser negociado. Independe do registro formal e abrange os países signatários da Convenção de Berna, da qual o Brasil faz parte, sendo consolidado na Lei nº. 9.610/98, que interpreta restritivamente os negócios jurídicos sobre direitos autorais (DUARTE; PEREIRA, 2009).
- **Propriedade industrial:** amparada pela Lei 9.279, de 14 de maio de 1996, é um conjunto de direitos que se referem às criações industriais e comerciais, tais como patentes, marcas, desenhos industriais, indicações geográficas e segredos de negócio. Visa proteger a inovação e a criatividade de indivíduos e empresas, incentivando assim o desenvolvimento tecnológico e econômico do país, estabelecendo as regras para a transferência de tecnologia, proteção de programas de computador, concessão e proteção de patentes, bem como para o combate à concorrência desleal (BRASIL, 1996).
- **Proteção *sui generis*:** engloba a topografia de circuito integrado, as variedades de plantas cultivadas, os conhecimentos tradicionais e o acesso ao patrimônio genético. Vale ressaltar que essa proteção e exclusividade de novas variedades vegetais incentivam empresas a investirem em agricultura, horticultura e reflorestamento, gerando retorno financeiro e benefícios para a sociedade (JUNGMANN; BONETTI, 2010b)

Para melhor compreensão do que cada grupo da propriedade intelectual engloba, a Figura 2.4 ilustra suas modalidades e subáreas de maneira clara e resumida.

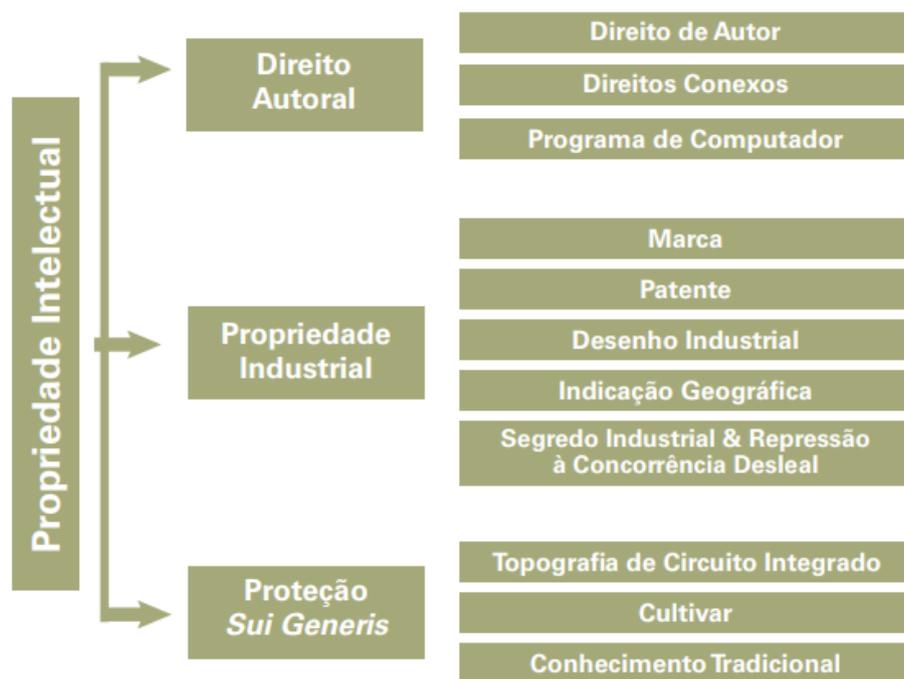


Figura 2.4. Modalidades e subáreas da Propriedade Intelectual.
 Fonte: (JUNGMANN; BONETTI, 2010a).

Nesse contexto, para melhor compreensão da parte fundamental deste estudo, faz-se necessário compreender melhor o que são as patentes, suas principais características, naturezas e classificações.

2.4 PATENTES

Invenção é uma criação intelectual que resulta da solução de um problema através da descoberta de uma nova tecnologia, processo, serviço ou produto, que não foi pensado por ninguém. Geralmente envolve a combinação de conhecimentos já existentes para criar algo e é parte integral da inovação e da criatividade humanas, o que a caracteriza como uma propriedade intangível, volátil e impossível de ser retirada de seu inventor. Ao se materializar, a invenção pode provocar avanços industriais e/ou sociais, sendo assim altamente valorizada, podendo então se tornar um bem econômico. Por esta razão criou-se a patente, forma que a sociedade encontrou para assegurar a posse deste tipo de bem (FRANÇA, 2000).

Gusberti *et al.* (2014) afirmam que a maneira mais utilizada para se comparar o desempenho de países, empresas e instituições de pesquisa no tocante ao desenvolvimento tecnológico é a quantidade de patentes que possuem. Os autores também afirmam que, no que tange ao desenvolvimento tecnológico, a parceria entre universidades e os setores produtivos é de extrema relevância, uma vez que a completividade entre as diversas áreas de conhecimento favorece a criação de tecnologias com maior potencial de aplicação comercial. Além disso, a colaboração entre os setores produtivos e as universidades também possibilita a realização de treinamentos e a atualização de conhecimento dos profissionais destas áreas, promovendo melhores condições de competitividade no mercado.

Conforme é exposto pelo INPI, autarquia federal vinculada ao Ministério da Economia que é responsável pelo aperfeiçoamento, disseminação e gestão do sistema brasileiro de concessão e garantia de direitos de propriedade intelectual para a indústria (BRASIL, 2021), conceder uma patente é um ato administrativo declarativo, que emite um documento oficial conferindo ao seu titular, ou aos seus sucessores, direitos exclusivos para produzir, fabricar, vender ou usar uma descoberta inovadora por um determinado período de tempo, onde terceiros só poderão explorar a patente com permissão do titular, mediante uma licença, porém, após o período de vigência, a patente se torna domínio público, podendo então ser usufruída por qualquer pessoa da sociedade (INPI, 2021). Vale ressaltar a importância da garantia de que os inventores obtenham alguma remuneração lucrativa de seu trabalho. De acordo com o INPI:

A intenção é que durante a vigência da patente, o titular seja recompensado pelos esforços e gastos despendidos na sua criação. Sendo assim, a patente pode ser considerada uma forma de incentivar a contínua renovação tecnológica, estimulando o investimento das empresas para o desenvolvimento de novas tecnologias e a disponibilização de novos produtos para a sociedade (INPI, 2021).

Importante salientar que existem regras específicas para o pedido de depósito de patente, incluindo seu processamento e publicação. De acordo com a Lei 9279/96, publicada por Brasil (1996), são estabelecidos os critérios de patenteabilidade, evidenciando que é patenteável toda invenção que apresenta os requisitos de novidade, atividade inventiva e aplicação industrial. Vale ressaltar que para obter uma patente, o inventor precisa submeter um pedido ao INPI, o qual analisa o mérito da invenção e verifica se atende aos requisitos legais. Em seguida, o

pedido é publicado no Diário Oficial da União, permitindo que terceiros apresentem contestações. Após esse período, se não houver impugnações, a patente é concedida.

2.4.1 Natureza das patentes

É prevista, pela Lei da Propriedade Industrial, em seu artigo 2º, inciso I, a concessão de patentes de invenção e de modelo de utilidade. De acordo com INPI (2015a), é possível descrevê-las como se segue:

- **Patente de invenção:** é a mais comum e protege uma solução técnica para um problema específico. Deve ser nova, envolver atividade inventiva e ter aplicação industrial. Uma invenção pode ser um produto ou processo que é novo e não óbvio, ou seja, que não é uma simples combinação de ideias já existentes, como, por exemplo, um novo medicamento ou um processo de produção mais eficiente para um produto já existente. O prazo para a vigência de uma patente de invenção é de 20 anos, a contar a partir da data de seu depósito.
- **Patente de modelo de utilidade:** protege uma solução técnica para um problema específico, assim como a patente de invenção, onde a diferença é que a patente de modelo de utilidade se aplica a objetos que tenham uma utilidade prática, como ferramentas ou dispositivos mecânicos. Podemos usar como exemplo um novo tipo de oxímetro ou de aferidor de pressão arterial. Importante ressaltar que a patente de modelo de utilidade exige menos atividade inventiva do que a patente de invenção. Geralmente, a patente de modelo de utilidade é válida por 15 anos, a partir da data em que é depositada.

A lei também prevê a concessão de Certificado de Adição de Invenção, documento que permite que um inventor ou detentor de uma patente modifique ou acrescente novos elementos a uma invenção já patenteada, sendo assim, uma forma de proteção de patente secundária, que se destina a aprimorar a patente original sem prejudicar sua validade. O certificado pode ser solicitado a qualquer momento durante a vigência da patente original e deve, obrigatoriamente, estar relacionado à mesma invenção ou modelo de utilidade protegido. Seu objetivo é incentivar

a melhoria contínua da tecnologia, permitindo que os inventores modifiquem ou aprimorem suas invenções patenteadas sem precisar obter uma nova patente.

2.4.2 Classificação das patentes

As patentes concedidas possuem classificação que variam de acordo com sua natureza, seguindo o sistema de Classificação Internacional de Patente (IPC, do inglês *International Patent Classification*). A versão mais recente desse sistema está disponível em inglês no *site* da WIPO¹¹ e em português no site do INPI¹². Essa classificação visa a introdução de uma ferramenta que, de maneira eficaz, busque e recupere documentos de patentes; organize-os, facilitando o acesso à informação; sirva como base para a divulgação seletiva de dados; e venha prover uma estrutura para a elaboração de estatísticas que facilitem a avaliação do desenvolvimento tecnológico nas mais diversas áreas (OMPI, 2020).

A IPC é uma classificação hierárquica que divide o conhecimento tecnológico em oito grandes seções, que são representadas pelas letras de A a H. Além disso, cada seção é dividida em classes, as classes em subclasses e cada subclasse em grupos, que podem, inclusive, conter subgrupos. Sobre as seções, cada uma possui um título e uma letra como código específico, conforme especificado a seguir:

- **A** - Necessidades Humanas,
- **B** - Operações de Processamento; Transporte,
- **C** - Química e Metalurgia,
- **D** - Têxteis e Papel,
- **E** - Construções Fixas,
- **F** - Engenharia Mecânica; Iluminação; Aquecimento; Armas; Explosão,
- **G** – Física, e,
- **H** – Eletricidade.

¹¹<<https://www.wipo.int/publications/en/details.jsp?id=4656&plang=EN>>

¹²<<http://ipc.inpi.gov.br/classificacoes/ipc/ipcpub/?notion=scheme&version=20230101&symbol=none&menulanguage=pt&lang=pt&viewmode=f&fipcpc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>>

A classificação que uma patente recebe é composta por uma combinação de letras e números, como a classificação A61B10/04 (WIPO, 2023). A Figura 2.5 pode auxiliar no entendimento.

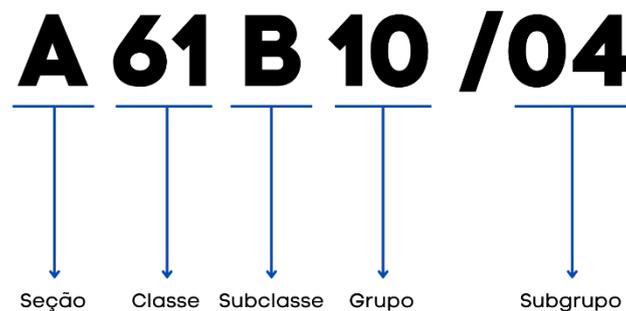


Figura 2.5. Exemplo de classificação IPC.
Fonte: autoria própria.

De forma detalhada, a classificação apresentada na Figura 2.5 trata do que segue:

- **Seção A:** Necessidades humanas.
- **Classe 61:** Ciência médica ou veterinária; higiene.
- **Subclasse B:** Diagnóstico; cirurgia; identificação.
- **Grupo 10:** Outros métodos ou instrumentos para diagnósticos, p. ex. para diagnóstico por vacinação; Determinação do sexo; Determinação do período de ovulação; Instrumentos para raspagem da garganta [2006.01].
- **Subgrupo 04:** Instrumentos para endoscopia [2006.01].

O sistema IPC não é o único de classificação de patentes. De acordo com o INPI (2015b), há a Classificação Cooperativa de Patentes (CPC, do inglês *Cooperative Patent Classification*). É um sistema criado pelo Escritório Europeu de Patentes, em parceria com o Escritório Norte Americano de Patentes, baseado na IPC, porém com mais detalhes acerca da classificação. Enquanto este possui em torno de 70 mil grupos, aquele possui em torno de 200 mil registrados. Vale ressaltar que a CPC é adotada pelo INPI, em conjunto com a IPC, desde 2014.

2.5 INTERFACE DE PROGRAMAÇÃO DE APLICAÇÃO

Rodrigues (2017) explica que a API é uma estrutura de regras e protocolos que viabiliza uma comunicação transparente e eficaz entre determinado conjunto de dados, por dois ou mais sistemas de informação, sem depender da plataforma, do tipo de acesso, de quais padrões utiliza para o intercâmbio dos dados e, além disso, disponibiliza, na sua origem, a documentação necessária para o entendimento das partes acerca do seu modo de operacionalização.

2.6 WEB SCRAPING

De acordo com MITCHELL (2018), coletar dados é uma ação quase tão antiga quanto a própria internet, a qual teve sua origem no início da década 60 e que serve para interconectar inúmeros dispositivos computacionais ao redor do mundo (KUROSE, 2013). Essa interconexão auxilia na tramitação de dados e informações, além de ter papel fundamental nas comunicações. Dados como textos, imagens, sons e vídeos estão trafegando por esta rede mundial e fica cada vez mais complexo, por conta da quantidade quase imensurável de informações, encontrar rapidamente dados específicos para algumas pesquisas.

Web Scraping (WS, traduzido significa raspagem da *web*) é a coleta de dados realizada por qualquer meio que não seja um programa interagindo com uma API ou por intermédio de uma pessoa utilizando um navegador (MITCHELL, 2018). Esta ferramenta pode ser entendida como o processo de extrair e combinar conteúdos de interesse da *web* de maneira sistemática. Neste método, uma espécie de robô imita a interação entre a Internet e o usuário, acessando sites de acordo com a necessidade, e analisando o conteúdo com o intuito de encontrar e extrair dados de interesse para, posteriormente, estruturá-los da maneira desejada pelo usuário (GLEZ-PEÑA et al., 2014). CALÒ (2014) explica que esta técnica é útil quando é necessário monitorar dados, ou até mesmo realizar a coleta massiva destes na *web*, que já se encontram disponíveis e distribuídos apenas em páginas de determinados sites, pois, conforme MITCHELL (2018) explica, os *web scrapers* conseguem acessar locais que as usuais ferramentas de pesquisa da atualidade não conseguem.

Vale salientar que é importante não confundir *Web Scraping* com *Web Crawling*. TAVARES e CUNHA (2021) afirmam que, mesmo que as duas ferramentas automatizem os processos, o *scraping* é bem mais assertivo em suas buscas. O *Web Crawling* é o processo que varre a Internet extraindo os dados para indexação, ou seja, faz o download em massa de páginas *web* e monta um índice que assegura a recuperação de qualquer documento ou informação no exato momento que o usuário busca um assunto na base escolhida, enquanto o *Web Scraping* coleta determinadas informações de um site específico e processa estes dados rapidamente, entregando o resultado solicitado.

Ressalta-se que é bem possível encontrar ferramentas de *Web Scraping* prontas, mas é necessário alertar que, atualmente, com as linguagens de programação e suas bibliotecas disponíveis, é muito mais prático desenvolver um programa para tal fim. Soluções prontas possuem algumas desvantagens, como o fato de que, em geral, permitem poucos ajustes para suprir as necessidades de quem as usa e, além disso, muitas delas não são gratuitas (AGOSTINHO, 2021).

2.7 ASPECTOS LEGAIS E ÉTICOS DO *WEB SCRAPING*

É importante salientar que a prática em si não é ilegal e que não há regulamentação específica para serviços e práticas de *Web Scraping* (MAGRI, 2022). Embora seja uma excelente ferramenta para empresas e pesquisadores, também pode ser usado de maneira inapropriada, tornando-se uma violação de direitos dos titulares de dados pessoais e do princípio da necessidade da Lei Geral de Proteção de Dados Pessoais. De acordo com Mitchell (2018), o *Web Scraping* também pode esbarrar em limitações que envolvem propriedade intelectual, tais como marcas registradas, direitos autorais e patentes. Além disso, esta prática pode caracterizar invasão de bens, fraude e abuso de computadores (MITCHELL, 2018).

A ética envolvida na prática de *Web Scraping* se refere ao esforço para coletar dados de maneiras aceitáveis. Mitchell (2018) explica que não desviar das medidas de segurança de um site para coletar as informações contidas no mesmo e se atentar aos termos de serviço, quando houver, são detalhes que diferenciam uma simples coleta de uma ilegalidade.

3 MATERIAIS E MÉTODOS

O presente trabalho caracteriza-se, quanto a sua abordagem, como uma pesquisa quantitativa, uma vez que seus resultados permitirão que seja mensurado o número de produtos acadêmicos e científicos, presentes no repositório UnB, vinculados às patentes depositadas pela instituição. Fonseca (2002) esclarece que a pesquisa quantitativa é objetiva, considerando tão somente que a realidade só será realmente compreendida com base em dados brutos que foram recolhidos com instrumentos neutros e padronizados. Já quanto a sua natureza, trata-se de uma pesquisa aplicada, já que tem por base a aplicação prática de conhecimentos, visando solucionar um problema específico que envolve interesses locais (GERHARDT; SILVEIRA, 2009). Quanto aos seus objetivos, é uma pesquisa descritiva, pois visa capturar e apresentar resultados de determinado cenário que podem ser expressos em números (CASTRO, 1976, apud OLIVEIRA, 2011). Por fim, tratando-se dos procedimentos, é considerada uma pesquisa bibliográfica, já que se baseia em soluções referenciadas e de funcionalidade atestada por outros trabalhos, além de que, de acordo com Fonseca (2002), trabalhos científicos costumam ter por princípio uma pesquisa bibliográfica para que o pesquisador possa ter noção do que já foi estudado sobre o assunto.

3.1 VISÃO GERAL DAS TECNOLOGIAS UTILIZADAS

O *scraper* foi desenvolvido em uma máquina virtual Linux, com Sistema Operacional (SO) Ubuntu, construída especificamente para o projeto. Nela foram instaladas, via linha de comando, as ferramentas pertinentes a construção do sistema. É possível visualizá-las na Figura 3.1 e todas serão detalhadas nas subseções a seguir. Importante salientar que, visando melhor interação entre as ferramentas, o *MySQL* e o *MySQL Workbench* interagem com a linguagem de programação *Python* por meio do driver *MySQL Connector*.

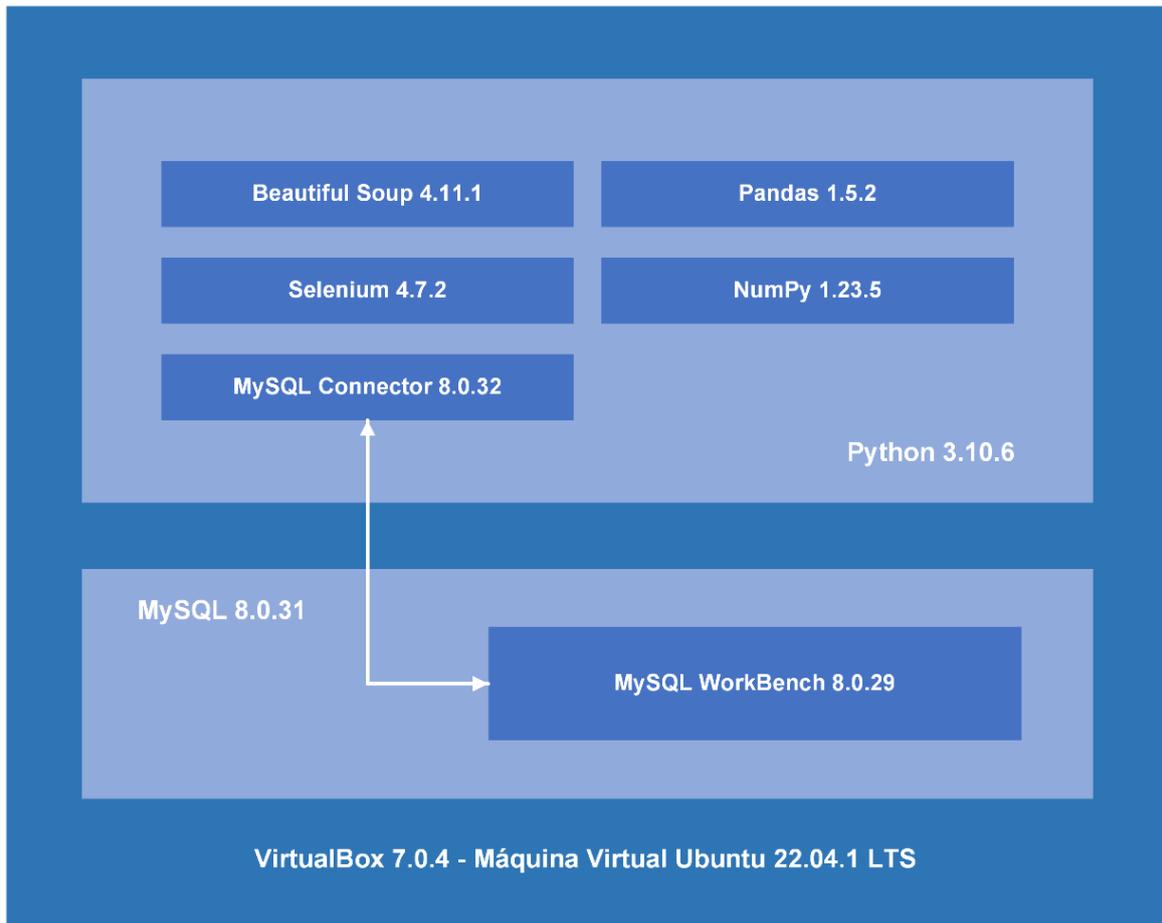


Figura 3.1. Visão geral das tecnologias utilizadas.
Fonte: autoria própria.

De maneira resumida, o *MySQL* disponibiliza o serviço de armazenamento e consulta de dados; o *Python* possui o *Beautiful Soup*, para facilitar a extração de conteúdos específicos nas páginas *web* a serem mineradas; o *Selenium* possibilita criar o *script* de automatização do navegador *web*; o driver *MySQL Connector*, permite a comunicação com o banco de dados *MySQL*; o *Selenium*, que contém todo seu conjunto de instruções de automatização, de forma a receber os comandos necessários para controlar o navegador *web* a fim de realizar as buscas e os acessos às páginas requisitadas; o *Pandas*, para organizar, manipular e analisar dados de forma eficiente; e o *NumPy*, para facilitar a manipulação e o armazenamento dos dados coletados das páginas de patentes.

3.2 PYTHON

Por conta de sua fácil sintaxe e de sua velocidade, pode ser considerada a linguagem de programação mais utilizada para o desenvolvimento de projetos de automação e *scripting* (DIVINO, 2021). Além disso, conforme a Figura 3.2 ilustra, *Python* está no topo do ranking de linguagens de programação mais utilizadas do mundo em 2022.

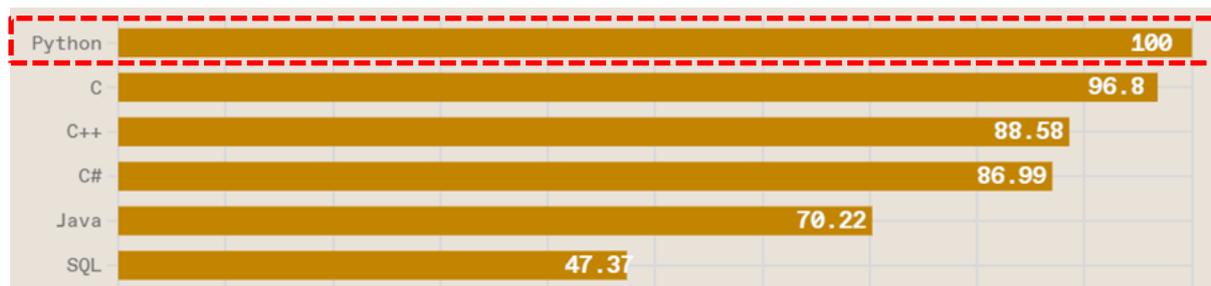


Figura 3.2. Principais linguagens de programação do IEEE Spectrum 2022.
Fonte: (CASS, 2022).

Isso se dá por ser uma linguagem poderosa e flexível (CRUZ, 2015), além de ser clara e objetiva (MENEZES, 2017). É considerada uma linguagem de fácil aprendizado, simples e muito produtiva, o que justifica seu crescimento ao longo dos anos, junto aos mais diversos perfis de profissionais, no meio científico e acadêmico, no uso em várias áreas da computação, como inteligência artificial, bancos de dados, biotecnologia, animações, aplicativos móveis e como plataforma *web*. CRUZ (2015) explica que *Python* é uma linguagem interpretada de alto nível e que suporta múltiplos paradigmas de programação, além disso, possui tipagem dinâmica e forte, escopo léxico e gerenciamento automático de memória. O autor afirma também que ela possui algumas estruturas de dados embutidas em sua sintaxe o que aumenta bastante a expressividade do código, além de possuir *batteries included*, expressão que indica que o *Python* possui uma vasta biblioteca padrão com inúmeros utilitários poderosos. A nível de história, foi criado no início dos anos 1990, por Guido van Rossum, e em 2001 criou-se a *Python Software Foundation*, uma organização sem fins lucrativos criada especificamente para deter a propriedade intelectual relacionada ao *Python* (DOCUMENTAÇÃO, 2022). Importante salientar que todas as versões do *Python* são de código aberto, ou seja, possuem desenvolvimento baseado no compartilhamento e melhoria colaborativa, além de ser gratuito.

3.3 BEAUTIFUL SOUP

De acordo com sua própria documentação, *Beautiful Soup* é uma biblioteca *Python* de extração de dados de arquivos HTML, ou seja, é uma ferramenta que recebe o conteúdo em HTML, interpreta-o e permite que seja possível buscar e extrair os dados contidos no mesmo (DOCUMENTAÇÃO..., 2004-2015). Essa documentação explica também que a ferramenta é capaz de transformar um documento HTML complexo em uma árvore de objetos *Python*, o que significa que ela pode, de certa forma, estruturar o conteúdo de tal forma que facilite o trabalho do programador, tornando a árvore mais legível e de fácil navegabilidade. Em suma, o *Beautiful Soup* tenta dar sentido ao que não faz sentido, ajudando a formatar e a organizar a web confusa, fazendo correções em um código HTML mal formatado e apresentando objetos *Python* que podem ser facilmente percorridos (MITCHELL, 2018). A ferramenta foi escolhida para auxiliar na extração de dados devido aos métodos de busca que proporciona, possibilitando encontrar a informação desejada por meio do nome da *tag* ou dos atributos de um elemento HTML. Isso auxilia no acesso a conteúdos mais específicos de uma página *web*, por meio da classe de um elemento de divisão, por exemplo.

3.4 SELENIUM

É um *framework* que possui várias ferramentas e bibliotecas que possibilitam e suportam a automação de navegadores da web, ou seja, oferece recursos avançados para interação com elementos da página, como localizar elementos por meio de seletores CSS ou XPath, o que viabiliza a criação de sistemas *Web Scraping* (DOCUMENTAÇÃO, 2022). Torna possível, ou pelo menos mais simples, controlar o navegador web por meio de uma sequência de códigos, preencher formulários, simular cliques em links e botões, além de coletar textos, ou seja, praticamente simula o comportamento do usuário, lidando, portanto, de maneira natural com os sites, assim como qualquer navegador (SAHIN, 2022). É útil para o projeto em questão, haja vista a necessidade de interação com as páginas da base de dados do INPI e do repositório da UnB, antes de ser possível acessar as informações de interesse, e o *Selenium* poupa esse trabalho, uma vez que o mesmo automatiza vários processos.

3.5 PANDAS E NUMPY

As bibliotecas *Pandas* e *NumPy* são duas das mais importantes bibliotecas de análise de dados e computação científica em Python. São de código aberto e amplamente utilizadas em áreas como finanças, biologia, física e engenharia, mas possuem funções e recursos ligeiramente diferentes (MENDES, 2023).

NumPy (abreviação de "Numerical Python"), é uma biblioteca para cálculo científico que fornece estruturas de dados para trabalhar com matrizes multidimensionais e funções matemáticas que operam nesses *arrays*. Esta biblioteca é útil para realizar operações numéricas em grandes conjuntos de dados, como cálculos matemáticos e estatísticos, processamento de sinais, álgebra linear, entre outros. Já o *Pandas* é uma biblioteca projetada para trabalhar com dados tabulares, oferecendo ferramentas para manipulação, transformação e análise de dados, bem como funções para lidar com dados faltantes. Tem como vantagem o fato de não exigir muitas linhas de código para suas operações e suporta um extenso conjunto de operações para análise exploratórias de dados, possibilitando, inclusive, cálculos estatísticos. Vale salientar que o *Pandas* consegue lidar com grandes conjuntos de dados de maneira rápida e eficiente e foi feito especificamente para o Python (PANDAS, 2023; NUMPY, 2022; WILLEMS, 2023; CHUGH, 2022).

Para o projeto desenvolvido, a biblioteca *NumPy* foi utilizada para criar um *arrays* uni e multidimensionais, a fim de permitir melhor manipulação e armazenamento dos dados coletados da base de dados do INPI. Já a *Pandas* auxiliou na criação de uma estrutura de dados tabular, permitindo a organização e conversão dos dicionários (estruturas de dados que permitem armazenar e acessar dados de forma flexível e eficiente) para posterior exportação.

3.6 SISTEMA GERENCIADOR DE BANCO DE DADOS MYSQL

Para o processamento de arquivos pré-formatados, como o CSV, extrair informações que dependem de relacionamento e/ou cruzamento de dados, exigiria maior esforço computacional e a construção de algoritmos específicos para a coleta das informações necessárias. Com o intuito de sanar este problema e suscitar novas estratégias de processamento e cruzamento de

informações, será implementado um Sistema Gerenciador de Banco de Dados (SGBD), que nada mais é que um programa específico para criar, armazenar, organizar e acessar dados a partir de um banco de dados, livrando o desenvolvedor, ou o usuário final, da tarefa de entender onde e como os dados estão de fato armazenados, separando assim as visões lógica e física dos mesmos, onde a visão lógica trata de como os dados serão vistos pelo usuário, ao passo que a visão física mostra como eles realmente são (LAUDON; LAUDON, 2014). De acordo com Elmasri e Navalhe (2019), as principais vantagens do uso de um SGBD são: controle de redundância, restrição de acesso não autorizado, técnicas de processamento eficiente de consulta, recuperação de falhas e restrições de integridade.

Para esse projeto, foi escolhido o SGBD *MySQL*, uma vez que, de acordo com a classificação gerada pela *DB-Engines*, apresentada na Figura 3.3, está em segundo lugar no *ranking* de setembro de 2022. A popularidade, por esse *ranking*, tem como parâmetros o número de menções do sistema em sites, o interesse geral pelo seu uso, a frequência de discussões técnicas relativas ao mesmo, a quantidade de ofertas de emprego relacionadas à sua aplicação, o número de perfis em redes profissionais em que é mencionado e a sua relevância nas redes sociais. A posição em destaque do *MySQL* indica também que, por ser muito utilizado, as chances de se encontrar soluções de modo mais rápido para eventuais problemas ou dúvidas que possam surgir durante a utilização do mesmo são maiores.

Rank			DBMS	Database Model	Score		
Apr 2023	Mar 2023	Apr 2022			Apr 2023	Mar 2023	Apr 2022
1.	1.	1.	Oracle +	Relational, Multi-model	1228.28	-33.01	-26.54
2.	2.	2.	MySQL +	Relational, Multi-model	1157.78	-25.00	-46.38
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	918.52	-3.49	-19.94
4.	4.	4.	PostgreSQL +	Relational, Multi-model	608.41	-5.41	-6.05
5.	5.	5.	MongoDB +	Document, Multi-model	441.90	-16.89	-41.48
6.	6.	6.	Redis +	Key-value, Multi-model	173.55	+1.10	-4.05
7.	7.	↑ 8.	IBM Db2	Relational, Multi-model	145.49	+2.57	-14.97
8.	8.	↓ 7.	Elasticsearch	Search engine, Multi-model	141.08	+2.01	-19.76
9.	9.	↑ 10.	SQLite +	Relational	134.54	+0.72	+1.75
10.	10.	↓ 9.	Microsoft Access	Relational	131.37	-0.69	-11.41

Figura 3.3. Principais bancos de dados, de acordo com o *DB-Engines Ranking*.
Fonte: (SOLID IT, 2023).

De acordo com sua documentação, o *MySQL* é muito rápido, de processamento múltiplo, multiusuário e robusto, além de ser duplamente licenciado, ou seja, os usuários podem optar

por utilizá-lo como um produto de código livre ou podem adquirir uma licença comercial padrão da *Oracle* (DOCUMENTAÇÃO, 2022). A Figura 3.3 mostra que o *MySQL* faz parte dos bancos de dados de modelo relacional, sendo este o principal motivo de sua escolha para compor este projeto, além de sua boa integração com o *Python*, uma vez que o modelo relacional possui uma maneira intuitiva e direta de representar dados em tabelas, além de sua regra de integridade que especifica que linhas duplicadas não são permitidas, eliminando, dessa forma, o potencial de informações errôneas inseridas no banco de dados (ORACLE, 2022).

Visando facilitar o manuseio do SGBD, será utilizado o *MySQL Workbench*, ferramenta gráfica criada para trabalhar com servidores e bancos de dados *MySQL*, que facilita a criação e o gerenciamento de conexões, permitindo que sejam criados modelos do esquema do banco de dados graficamente, além de auxiliar na administração das instâncias do servidor *MySQL* e tornar simples o processo de migração de dados e de servidores, proporcionando alta produtividade, uma vez que não é necessário dominar a Linguagem de Consulta Estruturada (SQL, do inglês *Structured Query Language*). O *MySQL Workbench* possui a versão gratuita e a edição comercial, esta última que disponibiliza alguns recursos extras, como ferramentas de auditoria e segurança (ORACLE, 2022).

3.7 MYSQL CONNECTOR

Sabe-se que são inúmeros os sistemas que precisam gravar dados durante seu uso, seja para processamento posterior ou para leitura como, por exemplo, um sistema de prontuário eletrônico que precisa guardar dados dos pacientes, da agenda dos médicos, dentre outros. Nesse viés, é justificável que muitas linguagens de programação tenham sua própria API de banco de dados. Para esse projeto será utilizado o *MySQL Connector*, um *driver* desenvolvido pela *Oracle Corporation*, que permite a comunicação entre a linguagem de programação *Python* e o banco de dados *MySQL*.

O uso do *MySQL Connector* com o *Python* é comum em projetos de grande volume de dados, onde a ferramenta funciona como um adaptador, conectando a aplicação desenvolvida ao banco e possibilitando a manipulação de dados de forma segura e ágil, por meio da execução de comandos SQL e da recuperação dos resultados de consultas. Vale ressaltar que entre as

vantagens que ele possui estão a facilidade de uso, sendo de fácil instalação e configuração; a segurança que o mesmo dispõe para proteger os dados, permitindo, inclusive, a criação de *scripts* seguros para manipulação de dados, evitando assim vulnerabilidades que possam comprometer a segurança do sistema; e a compatibilidade com diversas versões do Python e do *MySQL*, o que torna a sua utilização versátil, permitindo que os desenvolvedores escolham a versão que melhor se adapte às necessidades de seus projetos. Por outro lado, a necessidade de aprendizado da sintaxe SQL para executar consultas pode ser considerada uma desvantagem para desenvolvedores que não possuem experiência em SQL. Além disso, apesar de ser compatível com várias versões do *MySQL*, o *MySQL Connector* não suporta todos os recursos avançados do banco de dados, podendo limitar a funcionalidade de certos projetos que exigem recursos específicos do *MySQL* (BAWEJA, 2023; ORACLE, 2023).

3.8 MÁQUINAS VIRTUAIS

Uma máquina virtual (VM, do inglês *Virtual Machine*) pode ser compreendida como um clone quase perfeito de uma máquina real, haja vista que não possui acesso direto aos recursos de *hardware*, sendo esse intermediado e controlado pela máquina anfitriã, com suas características físicas e lógicas definidas através de parâmetros baseados na infraestrutura da máquina hospedeira. O conceito de virtualização é antigo, iniciou na década de 1960, por conta de inovações da *International Business Machines Corporation* (IBM), mas os avanços dessa tecnologia permitiram que as VMs se popularizassem com os mais diversos objetivos (MAZIERO, 2019). De acordo com Laureano (2004), as vantagens de utilizar VMs são várias, mas destacam-se a facilidade para aperfeiçoar e testar novos softwares e sistemas, o auxílio no ensino prático de sistemas operacionais e programação e a facilidade no gerenciamento, migração e replicação de aplicações e sistemas operacionais.

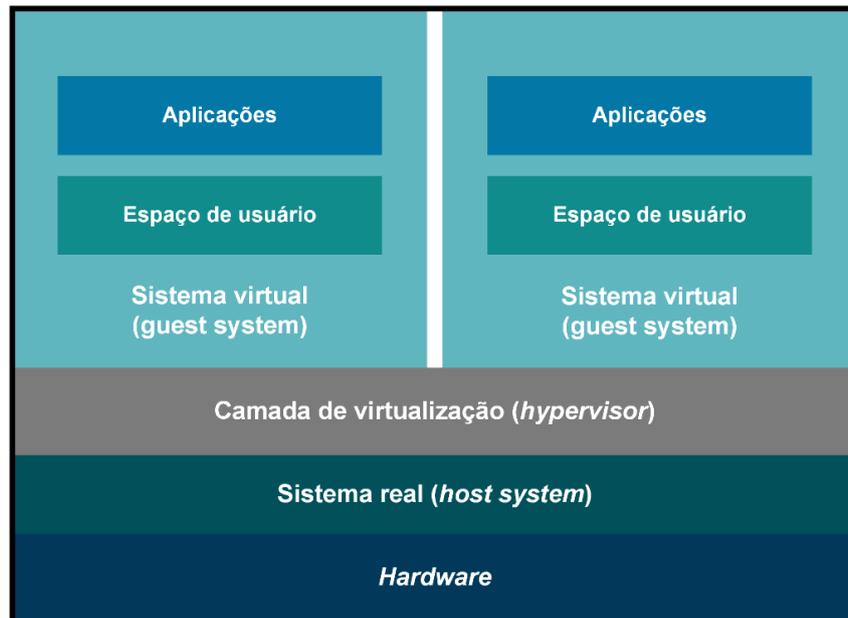


Figura 3.4. Concepção de uma estrutura de virtualização.
 Fonte: (MAZIERO, 2019, adaptado).

Na Figura 3.4 há a representação gráfica da concepção de uma estrutura de virtualização, onde a camada de *hardware* representa a parte física da máquina hospedeira; a camada de sistema real (*host system*) traz o sistema operacional que gerencia os recursos da máquina anfitriã; a camada de virtualização (*hypervisor*) que realiza a mediação e a criação da interface virtual a partir da interface real; e as camadas do sistema virtual (*guest system*), executado sobre a camada de virtualização (MAZIERO, 2019). Para o trabalho em tela, foi utilizado o *software VirtualBox*, programa de virtualização gratuito que possui suporte para os principais sistemas operacionais da atualidade. Esta ferramenta permite a instalação e execução de quantas VMs forem necessárias ao usuário, uma vez que o limite será estabelecido de acordo com o número de processadores, o espaço livre no disco de armazenamento e na memória de acesso randômico (RAM, do inglês *Random Access Memory*) disponíveis na máquina *host* (ORACLE, 2004-2022).

4 RESULTADOS E DISCUSSÕES

Nas próximas seções descreve-se em detalhes os sistemas baseados em *Web Scraping* desenvolvidos para realizar a coleta de trabalhos oriundos do repositório institucional da UnB, e para recuperar os depósitos de patentes realizados por discentes da instituição, disponíveis na base de dados do INPI, com o intuito de cruzar essas informações e evidenciar os trabalhos produzidos que geraram inovações. Em seguida, discorre-se sobre o tratamento e a organização dos dados coletados, a comparação das informações de interesse e, por fim, as discussões levantadas durante a pesquisa e as análises.

4.1 AQUISIÇÃO DOS DADOS

Para a construção do banco de dados local que fundamentou esta pesquisa, foi realizada a coleta de informações referentes a todos os produtos acadêmicos e científicos indexados no repositório da UnB, sem a aplicação de quaisquer filtros ou recortes temporais, visando abranger todas as áreas de estudos e pesquisas ao longo do tempo. De posse dos dados coletados e visando cruzar informações para encontrar as patentes vinculadas aos autores destes trabalhos, realizou-se a mineração de informações referentes aos registros de patentes depositados no INPI, única base em língua portuguesa e focada no Brasil, onde o depositante é, exclusivamente, vinculado ao Cadastro Nacional da Pessoa Jurídica (CNPJ) da UnB, sem aplicação de nenhum outro filtro para a requisição das informações. Na Figura 4.1 é possível ter uma compreensão mais clara do que a coleta de dados pretende entregar.

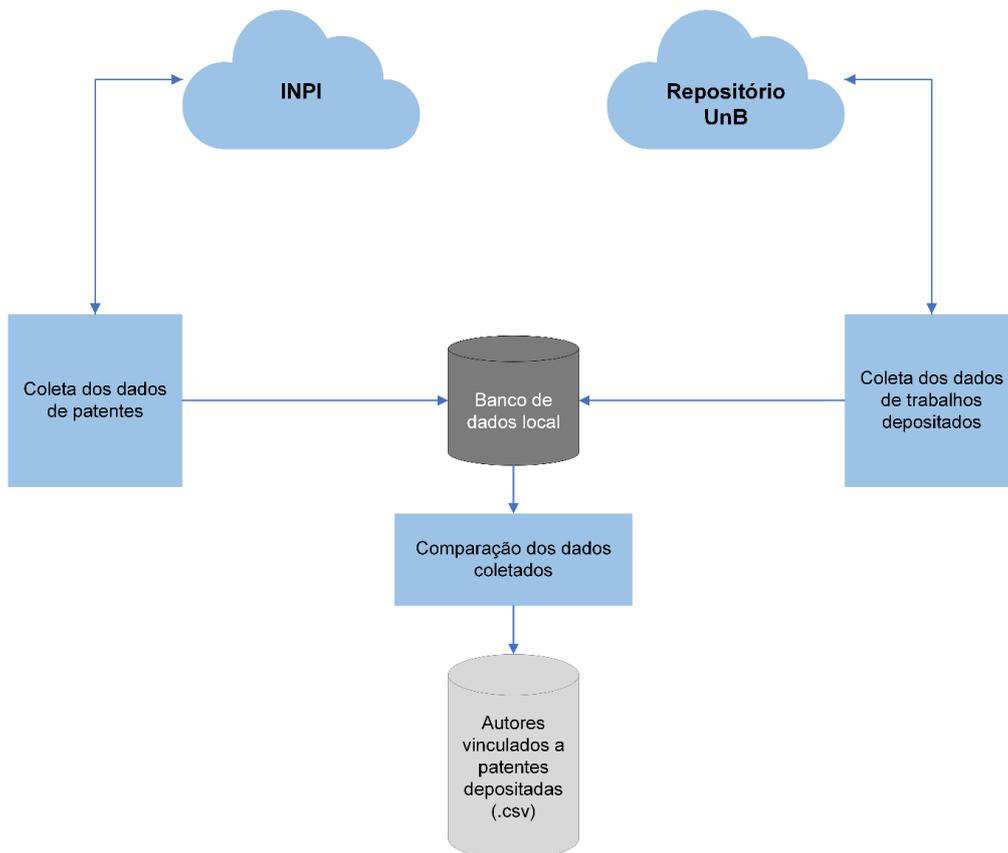


Figura 4.1. Visão geral da coleta de dados.
Fonte: autoria própria.

Conforme Figura 4.1, foi criado um banco de dados local, que recebeu os dados coletados do repositório da UnB e da base de dados do INPI. Após isso, realizou-se os procedimentos necessários para a comparação das informações e, ao final, foi entregue um arquivo apresentando os autores que contribuíram como inventores de patentes depositadas.

4.1.1 Coleta de dados no repositório da UnB

O repositório institucional da UnB¹³ possui uma boa construção de página, onde suas *tags* do código HTML estão bem definidas, o que facilitou o desenvolvimento do *scraper* nesta base. A URL informada ao algoritmo, para iniciar a coleta de dados, abrange todo o conteúdo disponível no repositório, no total de 42.103 registros, distribuídos em 4.211 páginas que exibem 10 linhas por vez.

¹³ <<https://repositorio.unb.br/>>

```

1 algoritmo ScraperUnB
2
3 max_pages = máximo de páginas a serem percorridas pelo scraper
4 current_page = página que o scraper está analisando
5
6 com repositórioUnB.csv aberto, em modo de escrita, como uma variável arquivo
7   cabeçalho = ['Data de Publicação', 'Data de Defesa', 'Título', 'Autor(es)',
8               'Orientador(es)', 'Coorientador(es)']
9   escreva cabeçalho
10
11   enquanto current_page for menor ou igual a max_pages
12     current_url = página de resultados visitada
13
14     para cada página de resultados encontre o atributo 'tr' e faça
15       publicacao = obtenha 'td' Data de publicação
16       defesa = obtenha 'td' Data de defesa
17       titulo = obtenha 'td' Título
18       autor = obtenha 'td' Autor(es)
19       orientador = obtenha 'td' Orientador(es)
20       coorientador = obtenha 'td' Coorientador(es)
21
22       info = [publicacao, defesa, titulo, autor, orientador, coorientador]
23       escreva info na variável arquivo
24
25     aguarde 10 segundos
26     some +10 à current_page
27
28     fim
29   fim
30 fim

```

Figura 4.2. Pseudocódigo da coleta de dados no repositório da UnB.

Fonte: autoria própria.

O pseudocódigo apresentado pela Figura 4.2 esboça o algoritmo desenvolvido e não foi encontrada nenhuma limitação para o desenvolvimento do *scraper* para o repositório.

4.1.2 Coleta de dados no INPI

Para coletar os dados de patentes no INPI, utilizou-se sistema Busca *Web*¹⁴, no qual é possível realizar a consulta gratuita de informações de patentes concedidas informando *login* e senha, ou por acesso anônimo. O que difere as formas de identificação é o acesso aos documentos extras referentes à patente, quando se opta por informar o *login* e senha, porém, para atingir o objetivo deste trabalho, o acesso anônimo é o suficiente. Após acessar a ferramenta pelo método identificação anônima, selecionou-se a opção de “Patentes”, em que é

¹⁴ < <https://busca.inpi.gov.br/pePI/> >

apresentada uma página contendo opções de pesquisa e, em seguida, foi selecionada a opção “pesquisa avançada” para exibir mais critérios de pesquisa.

No campo "CPF/CNPJ do Depositante" foi inserida a chave referente à UnB e selecionada a opção "pesquisar", ao passo que o sistema retorna uma listagem de 260 patentes, distribuídas em 7 páginas, exibindo até 40 registros em cada. No intuito de exemplificar melhor a dimensão do volume dos dados, foram realizadas coletas manuais das informações de interesse registradas na página de detalhes de 04 (quatro) patentes, seguindo a seguinte ordem de passos:

1. Clique no link da patente apresentada na lista de resultados da pesquisa;
2. Coleta individual dos campos de interesse (09, no total);
3. Cópia de cada campo, individualmente, para um editor de planilhas.

Esta tarefa foi cronometrada e o tempo despendido para a coleta não considerou a formatação das informações no editor de planilhas. Importante ressaltar que o tempo de carregamento da página de detalhe da patente é um fator significativo para o resultado da coleta, tanto manual, quanto automatizada. Os resultados dos testes podem ser visualizados na Tabela 3.1, onde, para as coletas 01 e 02, utilizou-se o modo de extensão de vídeo do computador em uso, com monitor extra, e nas coletas 03 e 04 foi simulada a situação da existência de apenas um monitor disponível para a atividade.

Tabela 4.1 – Teste de coleta manual de dados na base de dados do INPI.

NÚMERO DA COLETA	TEMPO GASTO
01	1min e 13s
02	1min e 06s
03	1min e 35s
04	1min e 48s

Fonte: autoria própria.

Com base em um cálculo de média aritmética simples, o processo manual de coleta das informações de uma única página de detalhes leva em torno de 1min e 25s, o que para todas as páginas apresentadas na lista de resultados levaria mais de 6 horas, equivalendo, praticamente, a 1 (um) dia de trabalho comercial, isso sem computar mudanças de páginas e correções de formatação. Para atenuar este tempo e otimizar a coleta dos dados, foi proposto um *scraper*, desenvolvido com as tecnologias apresentadas na seção 3.2, para viabilizar uma mineração de dados, composta por 05 etapas dispostas a seguir:

1. Acessar o sistema Busca *Web* e realizar o *login* anônimo para realizar a pesquisa;
2. Acessar a seção dedicada aos registros de patentes;
3. Selecionar a pesquisa avançada, informar no campo “CPF/CNPJ do Depositante” a chave referente à UnB e disparar o evento de pesquisa;
4. Percorrer a toda a listagem de patentes apresentada na página de resultados:
 - a) Para cada patente, acessar a página de detalhes;
 - b) Analisar o conteúdo das tabelas na página e recuperar as informações: “Número do pedido”, “Data de depósito”, “Data de publicação”, “Classificação ICP”, “Título”, “Resumo”, “Nome do Depositante” e “Nome do Inventor”
 - c) Armazenar as informações recuperadas em um arquivo CSV;
 - d) Voltar à listagem de patentes;
5. Repetir a etapa 4 para todas as páginas de resultados da pesquisa (260 registros).

Durante os testes do sistema desenvolvido, foi possível identificar uma limitação na abordagem estabelecida, pois, por motivos de segurança da plataforma do INPI, as credenciais, mesmo que não seja no modo anônimo, expiram depois de um determinado período ou na troca de seção, sendo necessário fazer a requisição novamente à base de dados, em ambos os casos. Para que as credenciais não expirem durante a coleta é necessário ter o cuidado de, durante a automatização, manter a navegação sempre na primeira seção requisitada, uma vez que os *links* que levam à página de detalhes das patentes são liberados a partir do *login* efetuado no início da rotina.

Outra limitação percebida na base de dados do INPI se deu por conta de sua estrutura diferenciada, que utiliza chamadas pré-formatadas que trazem resultados absolutos da pesquisa realizada, o que sugere o uso de um Sistema de Gestão de Conteúdos (CMS, do inglês *Content Management System*), uma plataforma de desenvolvimento de sites que auxilia na organização e administração de conteúdos, onde o administrador alimenta o banco de dados com o que ele deseja disponibilizar ao usuário e a aplicação entrega o conteúdo em uma página *web* automaticamente estruturada (ARIANE, 2022). Isso dificultou o desenvolvimento do *scraper*, mas não impossibilitou que fosse realizada a extração dos dados e a armazenagem dos mesmos em um arquivo CSV.

4.2 TRATAMENTO E ORGANIZAÇÃO DOS DADOS

No que diz respeito à extração de dados do repositório da UnB, os resultados foram armazenados diretamente no banco de dados local com êxito, como pode ser observado no Apêndice I. Entretanto, foi necessário realizar ajustes para separar os nomes dos autores dos trabalhos, criando colunas na tabela de acordo com a quantidade de nomes identificados e alocando cada um deles em uma coluna específica. Adicionalmente, foi preciso converter o formato do nome dos autores, orientadores e coorientadores de "sobrenome, nome" para "nome sobrenome", a fim de torná-los compatíveis com os nomes de inventores presentes na tabela de patentes do INPI, que são armazenados nesse formato. Essa medida foi necessária para permitir a combinação de dados posterior sem erros.

Com relação aos dados de patentes extraídos da base do INPI, Apêndice II, a mineração obteve sucesso, mas a estrutura da página dificultou o armazenamento direto dos dados no banco de dados local. Dessa forma, os dados extraídos foram gravados em um arquivo CSV. Antes de inserir manualmente os dados no banco de dados local, foi necessário tratar o campo "nomeInventor" que, assim como na tabela do repositório da UnB, continha todos os nomes juntos em um único campo. Utilizando o editor de planilhas disponível na máquina virtual, aplicou-se a função "linha para colunas" para separar adequadamente os nomes. Somente após essa etapa, os dados foram armazenados no banco de dados.

Para ambas as tabelas, já no banco de dados local, os campos que estavam sem nenhum dado, ou seja, vazios, foram definidos como “NULL”. Importante salientar que, de acordo com Hernandez (2021) definir campos vazios como “NULL” em uma tabela de banco de dados ajuda a economizar espaço em disco, garante a consistência dos dados, aumenta a compatibilidade com outros sistemas e facilita a escrita de consultas precisas e eficientes. Além disso, foi criada, em cada tabela, uma coluna para a definição da chave primária das mesmas, visando a integridade dos dados e a pesquisa e recuperação com eficiência (HERNANDEZ, 2021). A Figura 4.3, que representa o diagrama de entidade e relacionamento do banco de dados local criado, nos permite uma visualização geral do tratamento e da organização dos dados, que serão relatados a seguir.

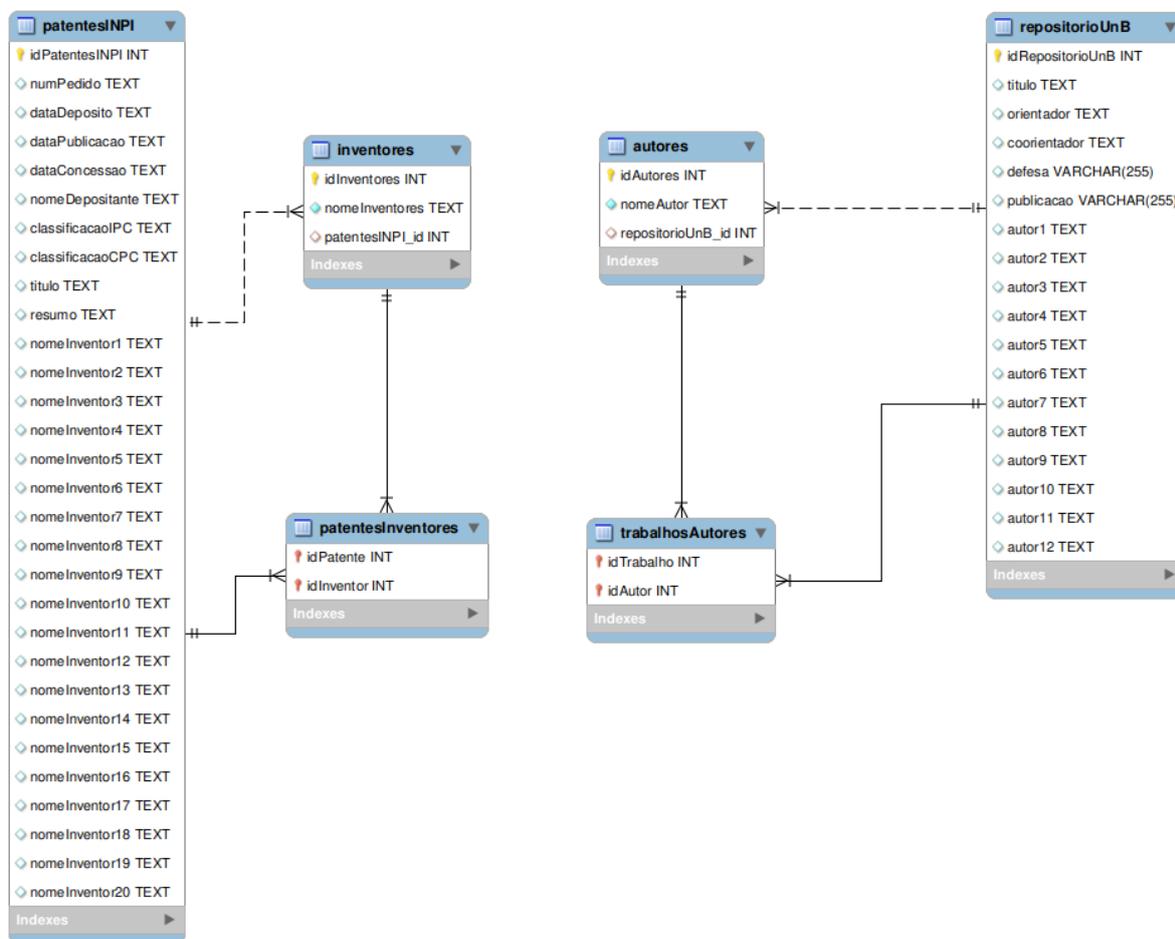


Figura 4.3. Diagrama de entidade e relacionamento do banco de dados local criado.
Fonte: autoria própria.

No cenário da pesquisa em tela, comparar os dados referentes aos nomes dos autores e inventores pode ser trabalhoso, já que existem muitos campos envolvidos. No banco de dados local criado, a tabela “repositorioUnB” possui 12 campos de autores e a tabela “patentesINPI”, tem 20 campos de inventores, o que significa que seria necessário comparar 240 campos (12 x 20) para identificar quais pessoas estão em ambas as tabelas. Visando facilitar essa comparação, duas novas tabelas foram criadas. A primeira, nomeada “autores”, foi preenchida com todos os autores de trabalhos presentes no repositório da UnB, bem como com todos os orientadores e coorientadores. A segunda, de nome “inventores”, armazena todos os inventores presentes nas patentes depositadas pela UnB no INPI. Visando a padronização dos campos a serem comparados, os campos que armazenam nomes, em ambas as tabelas, foram atualizados para que os nomes estejam todos em letra minúscula. Isso se dá pelo fato de que muitos bancos de dados são *case sensitive*, ou seja, diferenciam letras maiúsculas de minúsculas. Isso significa que, se os campos não estiverem padronizados, as consultas que envolvem a combinação de dados de diferentes tabelas podem não retornar resultados precisos ou podem retornar resultados incompletos.

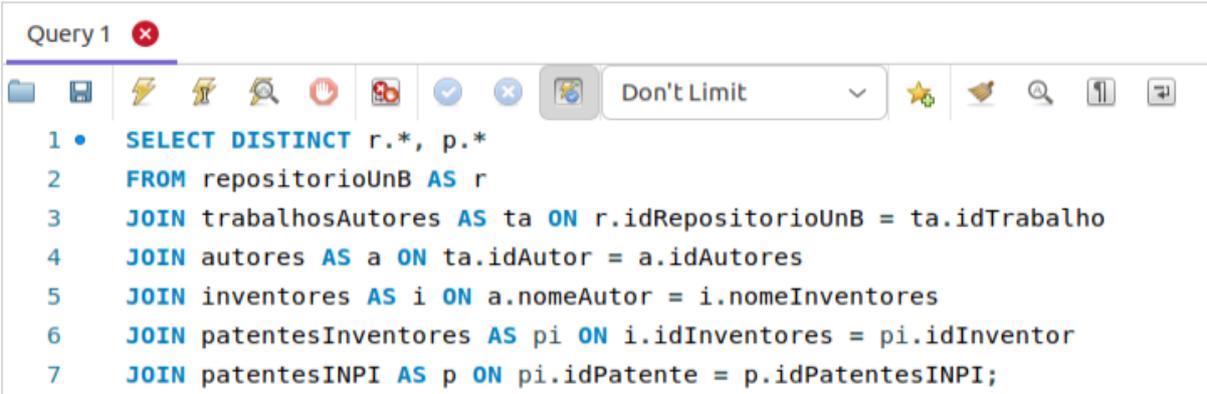
Após essa padronização, foram criadas tabelas adicionais que relacionassem os autores e inventores com os trabalhos e patentes, respectivamente. Essa reorganização visa a possibilidade de execução de consultas mais complexas e específicas com maior eficiência, uma vez que as informações necessárias estarão disponíveis de forma mais estruturada e organizada. Além disso, a criação de tabelas adicionais de relacionamento pode facilitar a manutenção dos dados, uma vez que informações redundantes podem ser evitadas e a integridade dos dados pode ser mantida.

4.3 COMPARAÇÃO DOS DADOS DE INTERESSE

Os dados coletados no INPI, vinculados ao CNPJ da UnB como depositante, totalizaram 260 registros de patentes, contendo 783 nomes de inventores. Esse número relacionado aos nomes pode ser menor, devido à falta de padronização no preenchimento dos pedidos de patentes, acarretando nomes "duplicados" nos registros, por falta ou excesso de caracteres, como, por exemplo, um acento ou até mesmo um dos sobrenomes. Vale lembrar que o

depositante é a pessoa física ou jurídica que tem os direitos de propriedade intelectual sobre a patente, já o inventor é o pesquisador que desenvolveu a patente. No tocante aos registros referentes ao repositório da UnB foram contabilizados 42.103 trabalhos depositados e 53.704 autores registrados, existindo inconsistências no tocante ao preenchimento dos dados, o que pode resultar em um número de autores maior ou menor que o apresentado.

Para comparar as tabelas do repositório institucional da UnB e a base de dados do INPI, foi utilizada a função “SELECT” do MySQL, retornando todos os trabalhos da tabela “repositorioUnB” e todas as patentes da tabela “patentesINPI” que estão relacionadas aos inventores, conforme sintaxe apresentada na Figura 4.4.



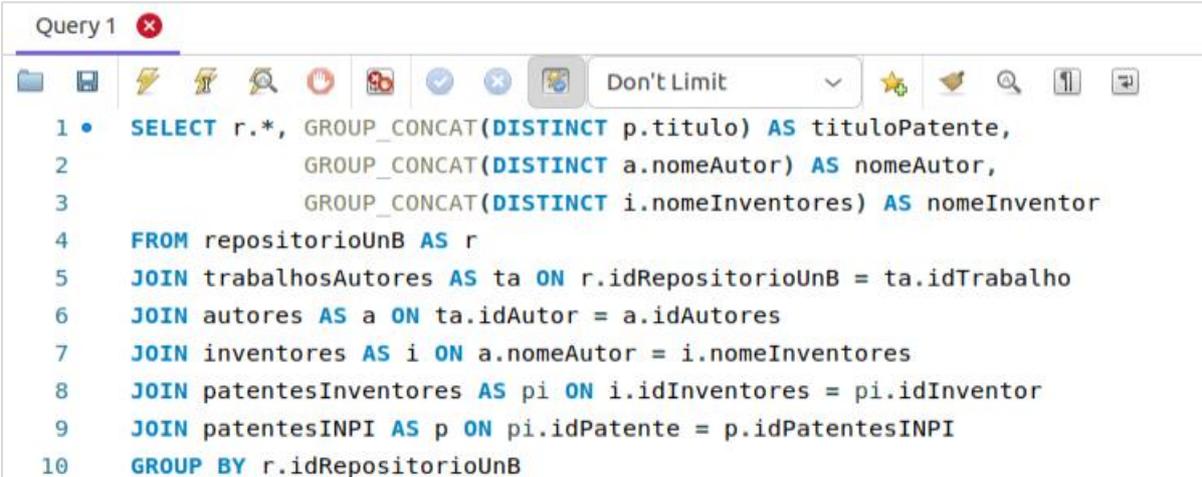
```
Query 1
1 • SELECT DISTINCT r.*, p.*
2 FROM repositorioUnB AS r
3 JOIN trabalhosAutores AS ta ON r.idRepositorioUnB = ta.idTrabalho
4 JOIN autores AS a ON ta.idAutor = a.idAutores
5 JOIN inventores AS i ON a.nomeAutor = i.nomeInventores
6 JOIN patentesInventores AS pi ON i.idInventores = pi.idInventor
7 JOIN patentesINPI AS p ON pi.idPatente = p.idPatentesINPI;
```

Figura 4.4. Primeira sintaxe criada para comparar as tabelas da UnB e do INPI.
Fonte: autoria própria.

A consulta executada utilizou a linguagem SQL para obter informações de um banco de dados, através do uso de diversas cláusulas, que estabeleceram relações entre tabelas do banco, visando a seleção de informações específicas. Primeiramente, a cláusula “SELECT” seleciona as informações que serão exibidas na consulta. Neste caso, a palavra-chave “DISTINCT” garante que não haverá duplicação de informações. A partir daí, a cláusula “FROM” especifica a tabela principal da consulta, "repositorioUnB". Em seguida, diversas cláusulas “JOIN” estabelecem as relações entre as tabelas "trabalhosAutores", "autores", "inventores", "patentesInventores" e "patentesINPI". As relações estabelecidas entre estas tabelas garantem que apenas as informações relacionadas serão exibidas. Por exemplo, a relação estabelecida entre a tabela "inventores" e "autores" assegura que somente autores que possuem informações de inventores correspondentes serão exibidos na consulta. A cláusula “ON” especifica as

colunas nas quais as relações entre as tabelas serão estabelecidas. Além disso, foram utilizados os operadores "=" e "AND" para garantir que as condições estabelecidas sejam atendidas. Ao final, a consulta retorna as informações das tabelas "repositorioUnB" e "patentesINPI", garantindo que apenas os registros relacionados através das demais tabelas sejam exibidos. Essa consulta, que pode ser parcialmente visualizada no Apêndice III, retorna um arquivo CSV com 3.244 linhas, onde os registros se repetem conforme a quantidade de reincidência do nome do autor ou do inventor na tabela. É possível, no próprio arquivo, com o filtro automático das colunas, separar, por exemplo, todos os registros referentes a um autor ou inventor específico na tabela, evidenciando assim tudo o que ele produziu de trabalho na UnB e todas as patentes que levam seu nome no registro. Isso facilita, por exemplo, a identificação não só dos trabalhos que viraram patentes, mas também a análise do quanto uma determinada pessoa pôde contribuir com a inovação enquanto esteve na UnB.

Uma outra sintaxe foi desenvolvida para que a consulta fosse mais específica (Figura 4.5), sem repetições de registros, retornando assim 1.155 linhas. Inicialmente, é selecionada a tabela "repositorioUnB" (r) que contém informações sobre os trabalhos acadêmicos, como o título, autor, data de publicação, entre outras. Em seguida, ocorrem os "JOINS" com as tabelas "trabalhosAutores" (ta), "autores" (a) e "inventores" (i) que possuem informações sobre os autores dos trabalhos e suas respectivas patentes. Nesses "JOINS", são estabelecidas as conexões entre as tabelas a partir das chaves estrangeiras presentes nestas.



```
Query 1
1 • SELECT r.*, GROUP_CONCAT(DISTINCT p.titulo) AS tituloPatente,
2     GROUP_CONCAT(DISTINCT a.nomeAutor) AS nomeAutor,
3     GROUP_CONCAT(DISTINCT i.nomeInventores) AS nomeInventor
4 FROM repositorioUnB AS r
5 JOIN trabalhosAutores AS ta ON r.idRepositorioUnB = ta.idTrabalho
6 JOIN autores AS a ON ta.idAutor = a.idAutores
7 JOIN inventores AS i ON a.nomeAutor = i.nomeInventores
8 JOIN patentesInventores AS pi ON i.idInventores = pi.idInventor
9 JOIN patentesINPI AS p ON pi.idPatente = p.idPatentesINPI
10 GROUP BY r.idRepositorioUnB
```

Figura 4.5. Segunda sintaxe criada para comparar as tabelas da UnB e do INPI.
Fonte: autoria própria.

O “JOIN” com a tabela "patentesInventores" (pi) estabelece a conexão entre os autores e suas respectivas patentes. Por fim, o “JOIN” com a tabela "patentesINPI" (p) busca informações sobre as patentes registradas no INPI, como o título da patente, a data de registro, entre outras. Ao final da consulta, é utilizada a função “GROUP BY” para agrupar os resultados com base no ID do trabalho acadêmico presente na tabela "repositorioUnB". O resultado da consulta está parcialmente apresentado no Apêndice IV. Essa função permite que sejam agrupados em uma única linha os autores e suas respectivas patentes, evitando a repetição dos registros e tornando a consulta mais clara. Dessa forma, a sintaxe apresentada possibilita buscar informações importantes sobre os trabalhos acadêmicos e suas patentes, o que pode ser de grande utilidade para pesquisas científicas, inventores e empresas que busquem informações sobre patentes registradas na Universidade de Brasília.

Visando quantificar os nomes presentes na tabela “autores” que também constam na tabela “inventores”, uma vez que este dado é o único passível de comparação, realizou-se uma nova consulta no banco de dados, utilizando-se da mesma lógica das consultas anteriores, e o quantitativo retornado consta na Tabela 4.2 a seguir.

Tabela 4.2 – Quantitativo referente aos nomes registrados nas bases garimpadas.

Nº DE AUTORES REGISTRADOS NO RIUNB	Nº INVENTORES REGISTRADOS NO INPI	Nº DE PESSOAS REINCIDENTES EM AMBAS AS TABELAS
53.704	783	423

Fonte: autoria própria.

Os dados acima proporcionam a análise da relação entre o número de autores registrados no RIUnB e o de inventores registrados no INPI, que pode fornecer *insights* sobre a interseção entre a produção acadêmica e a produção de patentes na instituição. O número de autores registrados no RIUnB é de 53.704, enquanto o de inventores registrados no INPI é de 783. Ao analisar o quantitativo de reincidências, que totalizam 423 pessoas, percebemos que cerca de 54,03% dos inventores estão engajados em práticas inovadoras, enquanto os demais estão vinculados a outras atividades e estudos na UnB. Calculando a proporção entre esses dois valores, observamos que os inventores registrados no INPI representam, aproximadamente, 1,46% dos autores registrados no RIUnB. Além disso, é possível calcular a diferença absoluta

entre os dois números. Nesse caso, há uma diferença de 52.921 entre a quantidade de autores registrados e de inventores, indicando a magnitude da discrepância entre as duas atividades, sugerindo a existência de uma oportunidade para incentivar mais pesquisadores a se envolverem com atividades de inovação e patentes. Importante ressaltar que essa análise é apenas quantitativa, fornecendo somente uma visão geral e ampla da relação entre a produção acadêmica e o registro de patentes na UnB.

4.4 DISCUSSÕES

Em princípio, foram pesquisadas soluções com o uso da técnica de *web crawler*, contudo a mesma foi descartada, em virtude das limitações referentes à sua aplicação, uma vez que a mesma faz a varredura e a extração apenas para a indexação dos dados. De acordo com (STREHL, 1998), a principal finalidade de um serviço de indexação é assegurar a recuperação de qualquer documento ou informação quando um usuário busca um assunto na base escolhida.

No tocante aos testes com a técnica de *Web Scraping*, houve a tentativa de construir o *scraper* da base do INPI com o *framework Scrapy*, amplamente usado para rastrear sites e extrair dados estruturados de suas páginas, podendo assim ser utilizado para uma ampla variedade de propósitos, inclusive para processamento ou armazenamento de informações, mineração de dados e para monitoramentos e testes automatizados (SCRAPY, 2022). Porém, devido a complexa estruturação da página do INPI, a ferramenta foi descartada.

Já na infraestrutura de desenvolvimento, foi testada a interação com o *Docker*, plataforma *open source* que permite a criação e a administração de ambientes isolados, auxiliando no gerenciamento de toda a infraestrutura de um aplicativo e garantindo que os ambientes de desenvolvimento, homologação e produção contenham os mesmos componentes e versões do aplicativo, minimizando assim os impactos no processo de desenvolvimento e entrega de software (DOCUMENTAÇÃO, 2013). A plataforma trabalha com virtualização a nível de sistema operacional, contudo ainda é capaz de oferecer ambientes leves, pois utiliza recursos do núcleo do SO hospedeiro para executar os *containers*, evitando então a instalação de uma camada de SO completa para execução dos aplicativos, como as VMs tradicionais necessitariam (DELGADO, 2021). Contudo o *Docker* foi descartado em virtude da dificuldade

de interação daquele com os controles lógicos do SO utilizado, já que sua operação é voltada para o Linux e o hospedeiro utiliza Windows, mesmo com a configuração correta do Subsistema Windows para Linux (WSL, do inglês *Windows Subsystem for Linux*), camada de compatibilidade que permite a execução de um ambiente Linux diretamente no Windows, tendo por principal vantagem o consumo mínimo de recursos do computador, quando comparado a uma VM completa (LEARN, 2022).

Durante o tratamento e organização das informações extraídas da base da UnB, foi notado que existiam dados faltantes, como, por exemplo, o nome do orientador e a data de defesa do trabalho presente no repositório. Campos em branco nas tabelas do INPI também foram encontrados, mas a situação justifica-se devido aos prazos de cada processo do pedido de patente. Além disso, os nomes de autores e inventores não seguem um padrão, conforme é apresentado na Figura 4.6, onde é possível notar que a mesma pessoa possui, na situação em tela, dois registros no banco de dados por causa de um hífen a mais em um dos registros (Figura 4.6 A) e de um sobrenome ausente em outro (Figura 4.6 B). A situação se repete em vários outros nomes nas tabelas “autores” e “inventores”, com variações de acentos, sobrenomes e/ou caracteres especiais. Há também os casos em que os nomes estão abreviados, dificultando a identificação do autor do trabalho depositado no repositório durante a consulta no banco de dados e tornando a comparação com a tabela do INPI inviável.

#	nomeAutor	#	nomeInventores
139	adalberto correa lima	702	suzange soares de camargo
140	adalberto corrêa napoleão	703	sonia maria de freitas
141	adalberto corrêa pagani	704	soraya megumi kazuma
142	adalberto corrêa santos	705	spartaco astolfi filho
143	adalberto corrêa souza	706	spartaco astolfi-filho
144	adalberto de (org.) salles-lima	707	stephania graziani barbosa lima
145	adalberto de salles lima	708	suelia de siqueira rodrigues fleury
146	adalberto de salles-lima	709	suelia de siqueira rodrigues fleury rosa
147	adalberto farinasso	710	suzana curi
148	adalberto felinto da cruz júnior	711	suzana moreira avila
149	adalberto josé vilela júnior	712	suzana neiva santos

Figura 4.6. Repetição de registro nas tabelas do banco de dados local.
Fonte: autoria própria.

Outra questão percebida foi em relação a alguns tipos de documentos presentes no repositório que não parecem configurar produtos acadêmicos e/ou científicos, estando também datados em períodos anteriores à fundação da UnB, como é possível visualizar na Figura 4.7. Os nomes de autores presentes nestes registros foram considerados nas análises por terem sido minerados e tratados junto com os demais, trazendo certa inconsistência nos dados comparados, uma vez que não podem ser contabilizados junto com os autores, por exemplo, que não possuem nenhum registro de patente.

Conjunto de itens:					
Data de publicação	Data de defesa	Título	Autor(es)	Orientador(es)	Coorientador(es)
1947	-	O lusitanismo de Eça de Queiroz : ensaio	Nunes, Cassiano	-	-
1961	-	Necrose das folhas da batatinha, causada por toxina de afídios	Costa, A. S.; Carvalho, Ana Maria B.; Costa, Cláudio Lúcio	-	-
1962	-	Contrôle da cigarrinha verde em feijoa	Costa, Cláudio Lúcio; Nagai, Hiroshi; Costa, A. S.	-	-
1964	-	Controle de vira-cabeça do tomateiro pela destruição do vector	Costa, Cláudio Lúcio; Alves, Aldo; Nagai, Hiroshi; Costa, A. S.	-	-
Mar-1964	-	A economia do petróleo e os interesses nacionais	Marinho, Josaphat	-	-
Jun-1964	-	O art. 64 da Constituição e o papel do Senado	Marinho, Josaphat	-	-
Set-1964	-	Reforma da estrutura jurídica	Marinho, Josaphat	-	-
Dez-1964	-	O art. 20 da Constituição e a competência da União e dos Estados	Marinho, Josaphat	-	-
Mar-1965	-	Irretroatividade e retroatividade da lei na Constituição de 1946	Marinho, Josaphat	-	-
Jun-1965	-	Inelegibilidades no Direito brasileiro	Marinho, Josaphat	-	-

Figura 4.7. Representação dos primeiros registros presentes no RIUnB.

Fonte: Repositório Institucional da UnB.

Dessa forma, torna-se necessária uma investigação para compreender como otimizar o preenchimento dos campos do repositório institucional, a fim de prover clareza e veracidade nas informações presentes no mesmo, uma vez que o cruzamento dos dados da UnB com os do INPI pode ter sido comprometido devido à esta eventualidade, já que, por exemplo, os docentes que não foram registrados como orientadores podem ter depositado alguma patente não gerando qualquer evidência, haja vista que o dado não foi passível de comparação. Tal cenário

requisita uma análise minuciosa nos processos de inclusão de documentos no repositório institucional, uma vez que o mesmo é de grande relevância ao exercer papel fundamental no livre acesso à produção científica da UnB, promovendo, ou não, nesse caso, o devido prestígio aos pesquisadores que produziram e promoveram inovações e grandes estudos durante sua permanência na instituição. Vale ressaltar que há a necessidade de padronização também nos pedidos de patentes, para que os nomes dos inventores sejam o mais fidedigno possível.

Outra incongruência notável foi observada em relação ao número de patentes depositadas no INPI em comparação com a quantidade informada pelo Centro de Apoio ao Desenvolvimento Tecnológico da Universidade de Brasília (CDT)¹⁵. É importante ressaltar que esses números apresentam uma leve discrepância e o recorte apresentado pelo documento do CDT¹⁶ possui registros que não constam na base de dados do INPI. De acordo com o INPI, há um total de 260 registros, enquanto o CDT afirma que existem 369 patentes ou pedidos de patentes associados à UnB, sendo 285 delas de nível nacional e 84 de nível internacional. Embora os motivos exatos dessas inconsistências sejam desconhecidos, e não façam parte do foco desta pesquisa, é relevante destacá-los como um aspecto a ser considerado em futuras investigações ou estudos. Essa situação levanta dúvidas sobre a existência de um parâmetro consistente no preenchimento dos formulários de depósito de patente, devido à variação considerável no registro dos nomes dos inventores em diferentes documentos, sugerindo possíveis contradições nos registros. Portanto, é importante analisar se todos os depósitos de patentes da UnB foram registrados com seu CNPJ, a fim de identificar possíveis falhas ou áreas que necessitem de ajustes.

Em síntese, a análise das informações extraídas da base da UnB revelou a presença de dados faltantes e inconsistências nos campos do repositório institucional, assim como nos registros de patentes no INPI. Essas lacunas e incongruências prejudicam a clareza e veracidade das informações, comprometendo a comparação e o cruzamento de dados entre a UnB e o INPI. Além disso, a falta de padronização nos nomes de autores e inventores dificulta a identificação correta e precisa dos responsáveis pelas produções científicas e inovações. Diante desse cenário, é essencial aprimorar a metodologia de preenchimento de dados no repositório da UnB

¹⁵ <<http://pesquisa.unb.br/propriedade-intelectual-e-transferencia-de-tecnologia>>

¹⁶ <http://pesquisa.unb.br/images/NIT/Indicadores_Nupitec_detalhamento_tudo_2_2022.pdf>

e no depósito de patentes, visando fornecer informações confiáveis e úteis à comunidade científica. Essa melhoria permitirá a valorização adequada dos pesquisadores e inovadores, além de gerar dados consistentes que contribuirão para futuras investigações e estudos no campo científico.

5 CONSIDERAÇÕES FINAIS

As ferramentas de mineração de dados têm sido amplamente utilizadas em diversas áreas do conhecimento e, neste estudo, seu uso mostrou-se eficaz na extração de informações dos repositórios do INPI e da UnB, permitindo a análise e comparação entre produtos acadêmicos e registros de patentes. Cabe ressaltar que a relação entre pesquisa acadêmica e inovação é de suma importância para o avanço do conhecimento e para a promoção do desenvolvimento tecnológico.

Nesse contexto, o principal objetivo desta pesquisa foi desenvolver uma ferramenta de mineração de dados utilizando a linguagem de programação Python, com a finalidade de extrair registros do repositório do INPI, focando apenas nos registros de patentes, e da UnB, abrangendo todas as publicações lá presentes. A criação de códigos eficientes possibilitou a obtenção dos dados de interesse de forma precisa e exitosa. Além disso, um banco de dados local foi modelado para permitir o armazenamento e a consulta adequada dos dados extraídos das aplicações desenvolvidas. Com êxito, os dados de interesse foram importados para suas tabelas específicas, garantindo assim a integridade e a facilidade de acesso às informações. No intuito de enriquecer ainda mais a análise dos registros obtidos, foi desenvolvido um algoritmo interno ao banco de dados que possibilitou o cruzamento das informações nele contidas, evidenciando, de forma precisa, os projetos convertidos em depósitos de patentes por meio dos autores envolvidos. Essa abordagem permitiu uma melhor compreensão e análise das informações extraídas, revelando inclusive as áreas e segmentos que mais converteram a fase de pesquisa de base em proteções intelectuais. Por fim, foi realizada uma análise aprofundada das informações mineradas, por meio de um arquivo CSV gerado ao final do processo de cruzamento dos dados. O arquivo forneceu um panorama completo e organizado dos resultados obtidos, permitindo a identificação de tendências e padrões relevantes para pesquisadores, cientistas e gestores.

A ferramenta de coleta de dados desenvolvida demonstrou ser uma abordagem promissora para a comparação entre produtos acadêmicos e registros de patentes, permitindo a identificação de iniciativas de PD&I bem-sucedidas. No entanto, as divergências encontradas

durante o tratamento e organização das informações destacam a importância do aprimoramento do preenchimento dos campos do RIUnB e dos formulários de pedidos de patentes no INPI. A inexistência de alguns dados, as inconsistências nos nomes de autores e inventores, a existência de documentos anteriores à fundação da UnB e a discrepância entre os números de patentes informados pelo CDT e pelo INPI, evidenciam a necessidade de se realizarem as investigações pertinentes e buscas por soluções para essas questões. Ao aprimorar a metodologia de preenchimento de dados e promover a padronização dos registros, será possível fornecer informações mais confiáveis e precisas, garantindo a clareza e veracidade dos dados comparados entre a UnB e o INPI. Essas melhorias são cruciais para adequada valorização dos pesquisadores e inovadores, além de contribuir para a tomada de decisões de gestores da UnB para a política de Pesquisa, Desenvolvimento e Inovação.

Os resultados obtidos confirmam a hipótese estabelecida nesta pesquisa, demonstrando que a ferramenta de mineração de dados desenvolvida é eficaz na extração de informações dos repositórios abordados nesta dissertação. A ferramenta proposta permite a análise das informações extraídas, proporcionando contribuições significativas para a sociedade e para a gestão da propriedade intelectual ao incentivar a colaboração entre a academia e o setor produtivo. Dessa forma, ela pode auxiliar os gestores da UnB em relação à alocação de recursos, identificação de áreas de pesquisa promissoras e estímulo à colaboração entre pesquisadores e setor produtivo. Além disso, a melhoria na gestão da propriedade intelectual facilita a proteção adequada das inovações geradas, promovendo um ambiente propício para o desenvolvimento científico e tecnológico da instituição.

Isto posto, a solução desenvolvida é funcional ao que se destina, bem como o cruzamento de informações é totalmente viável, podendo ambos serem incorporados em análises e/ou aperfeiçoamento de processos que maximizem os indicadores de inovação gerados pelas universidades, atendendo assim ao Marco Legal da Ciência, Tecnologia e Inovação. Desse modo, pode-se considerar que esta pesquisa tem implicações significativas para a área de estudo em questão, destacando a relevância do uso de ferramentas de mineração de dados para subsidiar a tomada de decisão em relação ao direcionamento de recursos para áreas com maior potencial inovador, estabelecimento de parcerias estratégicas e elaboração de estratégias de proteção da propriedade intelectual.

5.1 TRABALHOS FUTUROS

Como proposta para trabalhos futuros, sugere-se viabilizar o armazenamento automático da garimpagem da base de dados do INPI diretamente no banco de dados; criar uma rotina que envie os dados extraídos já tratados, uma vez que a forma de apresentação dos nomes são diferentes em ambas as bases; investigar formas de automatizar o processo de combinação sistemática dos dados, tornando a ferramenta ainda mais eficiente e prática; e, por fim, adicionar um parâmetro que, de forma agendada, inicie a funcionalidade de coleta periodicamente, coletando novas páginas e atualizando as que foram coletadas anteriormente.

5.2 LIMITAÇÕES ENCONTRADAS

É importante reconhecer que esta pesquisa apresentou algumas limitações que podem influenciar a interpretação dos resultados. Durante o tratamento e organização dos dados extraídos do repositório da UnB e da base de dados do INPI, foram observados dados faltantes, presença de documentos no RIUnB que não parecem configurar produtos acadêmicos ou científicos, além de inconsistências nos nomes de autores e inventores. Da mesma forma, foram encontradas discrepâncias entre o número de patentes depositadas no INPI e a quantidade informada pelo CDT. Apesar dessas limitações, os resultados obtidos nesta pesquisa oferecem contribuições valiosas para a compreensão da relação entre pesquisa acadêmica e registros de patentes, sendo importante considerá-las como oportunidades para investigações futuras, a fim de aprimorar as metodologias e superar os desafios encontrados.

REFERÊNCIAS BIBLIOGRÁFICAS

AGOSTINHO, Jackson W. S. **Coleta de dados de egressos via Web Scraping do LinkedIn e do Escavador**. Orientador: Prof.^a Dra. Karin Satie Komati. 2021. 73 p. Monografia (Bacharelado em Sistemas de Informação) - Instituto Federal do Espírito Santo, Espírito Santo, 2021.

ALMEIDA, Diana Beatriz de; PINHEIRO, Helano Diógenes. **A transferência tecnológica: o caminho da inovação para as universidades**. Teresina: EDUFPI, 2020. 111 p. ISBN 978-65-5904-092-6. Disponível em:
https://ufpi.br/arquivos_download/arquivos/BIBLIOTECA/Biblioteca_Comunit%C3%A1ria/EBOOK_Transferencia_Tecnologica_-_Diana_Almeida_e_Helano_Pinheiro.pdf. Acesso em: 19 maio 2023.

ARIANE. **O Que é CMS (Sistema de Gerenciamento de Conteúdo)?**. [S. l.]: Hostinger Tutoriais, 23 dez. 2022. Disponível em: <https://www.hostinger.com.br/tutoriais/o-que-e-cms>. Acesso em: 30 dez. 2022.

BARBOSA, D. B. **Uma introdução à Propriedade Intelectual**. Editora Lumen Juris, 2010. Disponível em:
<<http://sinova.ucdb.br/wpcontent/uploads/biblioteca/Uma%20Introducao%20a%20Propriedade%20Intelectual.pdf>>. Acesso em: 13 março de 2023.

BARRETO, J. O. et al. Pesquisa translacional em saúde coletiva: desafios de um campo em evolução. *Saúde em Debate*, Rio de Janeiro, v. 43, n. especial, p. 4-9, 2019.

BAWEJA, Chaitanya. **Python and MySQL Database: A Practical Introduction**. [S. l.]: Real Python, 2023. Disponível em: <https://realpython.com/python-mysql/#installing-mysql-connectorpython>. Acesso em: 31 mar. 2023.

BIBLIOTECA CENTRAL, Irmão José Otão. **FAQ: O que são bases de dados?**. Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2022. Disponível em: <https://biblioteca.pucrs.br/ufaqs/o-que-sao-bases-de-dados/>. Acesso em: 20 nov. 2022.

BENTO, Evaldo Junior. **Desenvolvimento web com PHP e MySQL**. São Paulo: Casa do Código, 2014. 218 p.

BHARDWAJ, Bhavya et al. **Web Scraping Using Summarization and Named Entity Recognition (NER)**. 7ª Conferência Internacional sobre Sistemas Avançados de Computação e Comunicação (ICACCS), [s. l.], p. 261-265, 19 mar. 2021. DOI 10.1109/ICACCS51430.2021.9441888. Disponível em: <https://ieeexplore-ieee-org.ez54.periodicos.capes.gov.br/document/9441888>. Acesso em: 4 dez. 2022.

BRASIL. **Lei nº 9.279, de 14 de maio de 1996**. Regula direitos e obrigações relativos à propriedade industrial. Brasília, DF, 14 maio 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19279.htm. Acesso em: 23 mar. 2023.

BRASIL. **Lei nº 10.973, de 2 de dezembro de 2004**. Dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências. Brasília, DF, 2004. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.973.htm. Acesso em: 19 maio 2023.

BRASIL. **Lei nº 13.243, de 11 de janeiro de 2016**. Dispõe sobre estímulos ao desenvolvimento científico, à pesquisa, à capacitação científica e tecnológica e à inovação e altera a Lei nº 10.973, de 2 de dezembro de 2004, a Lei nº 6.815, de 19 de agosto de 1980, a Lei nº 8.666, de 21 de junho de 1993, a Lei nº 12.462, de 4 de agosto de 2011, a Lei nº 8.745, de 9 de dezembro de 1993, a Lei nº 8.958, de 20 de dezembro de 1994, a Lei nº 8.010, de 29 de março de 1990, a Lei nº 8.032, de 12 de abril de 1990, e a Lei nº 12.772, de 28 de dezembro de 2012, nos termos da Emenda Constitucional nº 85, de 26 de fevereiro de 2015. [S. l.], 11 jan. 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13243.htm. Acesso em: 29 nov. 2022.

BRASIL. **Decreto nº 9.283, de 7 de fevereiro de 2018.** Regulamenta a Lei nº 10.973, de 2 de dezembro de 2004, a Lei nº 13.243, de 11 de janeiro de 2016, o art. 24, § 3º, e o art. 32, § 7º, da Lei nº 8.666, de 21 de junho de 1993, o art. 1º da Lei nº 8.010, de 29 de março de 1990, e o art. 2º, caput, inciso I, alínea "g", da Lei nº 8.032, de 12 de abril de 1990, e altera o Decreto nº 6.759, de 5 de fevereiro de 2009, para estabelecer medidas de incentivo à inovação e à pesquisa científica e tecnológica no ambiente produtivo, com vistas à capacitação tecnológica, ao alcance da autonomia tecnológica e ao desenvolvimento do sistema produtivo nacional e regional. [S. l.], 7 fev. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/d9283.htm. Acesso em: 29 nov. 2022.

BRASIL. Serviços e Informações do Brasil. [Brasília]: **Instituto Nacional da Propriedade Industrial (INPI)** [2021]. Disponível em: <https://www.gov.br/pt-br/orgaos/instituto-nacional-da-propriedade-industrial#:~:text=Criado%20em%201970%2C%20o%20Instituto,propriedade%20intelectual%20para%20a%20ind%C3%BAstria>. Acesso em: 06 dez. 2022.

BRASIL. Serviços e Informações do Brasil. **Solicitar Transferência de Tecnologia.** [Brasília]: Governo do Distrito Federal, 03 mar. 2023. Disponível em: <https://www.gov.br/pt-br/servicos/transferencia-de-propriedade-intelectual>. Acesso em: 19 mai. 2023.

BROUCKE, Seppe vanden; BAESEENS, Bart. **Practical Web Scraping for Data Science: Best Practices and Examples with Python.** Bélgica: Apress, 2018. 313 p. ISBN 978-1-4842-3582-9.

CABENA, P; HADJINIAN, P; STADLER, R; JAAPVERHEES; ZANASI, A. **Discovering Data Mining: From Concept to Implementation.** Prentice Hall, 1998.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas.** Goiás: Universidade Federal de Goiás, 2009. 28 p. Disponível em: https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF_001-09.pdf. Acesso em: 4 dez. 2022.

CASS, Stephen. **Top Programming Languages 2022**: Python's still No. 1, but employers love to see SQL skills. IEEE Spectrum, 23 ago. 2022. Disponível em: <https://spectrum.ieee.org/top-programming-languages-2022>. Acesso em: 26 set. 2022.

CATIVELLI, Adriana Stefani. **Patentes das Universidades Públicas Brasileiras**: análise das concessões. Orientador: Elaine Rosangela de Oliveira Lucas. 2016. 175 p. Dissertação (Mestrado Profissional em Gestão de Unidades de Informação) - Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Florianópolis, 2016. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/458>. Acesso em: 13 dez. 2022.

CHUGH, Vidhi. **Python pandas tutorial**: The ultimate guide for beginners. [S. l.], 6 nov. 2022. Disponível em: <https://www.datacamp.com/tutorial/pandas>. Acesso em: 14 abr. 2023.

CRUZ, Felipe. **Python**: Escreva seus primeiros programas. São Paulo: Casa do Código, 2015. 287 p. ISBN 9788555190926.

DELGADO, Caio. **Docker**: Do básico à Certificação Docker DCA. [S. l.]: Leanpub, 2021. 286 p. E-book (286 p.).

DEVMEDIA, Artigos. **ORM**: Object Relational Mapper. [S. l.], 2011. Disponível em: <https://www.devmedia.com.br/orm-object-relational-mapper/19056>. Acesso em: 26 set. 2022

DIVINO, Bruno. **Python** - Uma Introdução à Linguagem. [S. l.], 15 out. 2021. Disponível em: <https://www.alura.com.br/artigos/python-uma-introducao-a-linguagem#utilizacao-no-mercado>. Acesso em: 16 ago. 2022.

DOCUMENTAÇÃO **Beautiful Soup**. [S. l.], 2004 - 2015. Disponível em: <https://www.crummy.com/software/BeautifulSoup/bs4/doc.ptbr/index.html>. Acesso em: 12 set. 2022.

DOCUMENTAÇÃO **Docker**. [S. l.], 2013. Disponível em: <https://docs.docker.com/get-started/>. Acesso em: 30 set. 2022.

DOCUMENTAÇÃO **MySQL**. [S. l.], 2022. Disponível em: <https://dev.mysql.com/doc/refman/8.0/en/introduction.html>. Acesso em: 26 set. 2022.

DOCUMENTAÇÃO **Python**: História e Licença. [S. l.]: Python Software Foundation, 2001 - 2022. Disponível em: <https://docs.python.org/pt-br/3/license.html>. Acesso em: 12 set. 2022.

DOCUMENTAÇÃO **Selenium**. [S. l.], 2022. Disponível em: <https://www.selenium.dev/pt-br/documentation/>. Acesso em: 25 set. 2022.

DOWNEY, Allen B. **Pense em Python**: Pense como um cientista da computação. 2ª. ed. São Paulo: Novatec, 2016. 326 p. ISBN 978-85-75227503.

DUARTE, Eliane C. de V. G.; PEREIRA, Edmeire C. **Direito Autoral: Perguntas e Respostas**. Paraná: UFPR, 2009. 164 p. ISBN 978-85-87801-07-4. Disponível em: <http://www.cipead.ufpr.br/wp-content/uploads/2015/03/LivroDireitoAutoral.pdf>. Acesso em: 24 mar. 2023.

ELMASRI, Ramez; NAVALHE, Shamkant B. **Sistemas de banco de dados**. 7ª. ed. [S. l.]: Pearson, 2019. 1156 p. ISBN 978-85-430-2500-1.

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996.

FERREIRA, Aurélio Buarque de H. **Minidicionário Aurélio**: O dicionário da língua portuguesa. 2021. 954 p. ISBN 6557982990.

FRANÇA, R. O. **A patente**. In: CAMPELLO, B. S; CENDÓN, B. V.; KREMER, J. M. (Orgs.). Fontes de informação para pesquisadores e profissionais. Belo Horizonte: Editora UFMG, 2000. cap.12, p.152-182.

FOROUZAN, Behrouz A. **Comunicação de dados e redes de computadores**. 4ª. ed. Porto Alegre: AMGH, 2010. 1145 p. ISBN 978-85-63308-47-4.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6ª. ed. São Paulo: Atlas, 2017. 128 p. ISBN 978-85-97-01292-7.

HAJBA, Gábor László. **Website scraping with Python: Using Beautiful Soup and Scrapy**. Sopron, Hungary: Apress, 2018. 235 p. ISBN 978-1-4842-3924-7.

HAND, D; MANNILA, H; SMYTH, P. **Principles of Data Mining**. MIT Press, 2001.

HERNANDEZ, Michael J. **Database Design for Mere Mortals: 25th Anniversary Edition**. 4ª. ed. USA: Addison-Wesley Professional, 2021. 640 p. ISBN 978-0136788041. E-book (640 p.).

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL - INPI. **Manual de Básico para Proteção por Patentes Invenções, Modelos de Utilidade e Certificados de Adição**. 2021. Disponível em: <<https://www.gov.br/inpi/pt-br/servicos/patentes/guia-basico/ManualdePatentes20210706.pdf>>. Acesso em 7 nov. 2022.

INPI. **Manual para o depositante de patentes**. Brasília, DF. 2015a. Disponível em: https://www.gov.br/inpi/pt-br/assuntos/arquivos-dirpa/ManualparaoDepositantedePatentes23setembro2015_versaoC_set_15.pdf. Acesso em: 21 mar. 2023.

INPI. **Classificação de patentes**. Brasília, DF. 2015b. Disponível em: <https://www.gov.br/inpi/pt-br/assuntos/patentes/classificacao-de-patentes>. Acesso em: 25 mar. 2023.

JUNGMANN, Diana de Mello; BONETTI, Esther Aquemi. **Inovação e propriedade intelectual: guia para o docente**. Brasília, DF: IEL, 2010a. 93 p. ISBN 978-85-7519-389-1. Disponível em: https://www.gov.br/inpi/pt-br/composicao/arquivos/guia_docente_iel-senai-e-inpi.pdf. Acesso em: 21 mar. 2023.

JUNGMANN, Diana de Mello; BONETTI, Esther Aquemi. **Proteção da criatividade e inovação: entendendo a propriedade intelectual: guia para jornalistas**. Brasília, DF: IEL, 2010b. 65 p. ISBN 978-85-87257-48-2. Disponível em: <https://www.gov.br/inpi/pt->

br/servicos/patentes/materiais-de-consulta-e-apoio/guia-para-jornalistas.pdf. Acesso em: 22 mar. 2023.

KOUZIS-LOUKAS, Dimitrios. **Learning Scrapy**: Learn the art of efficient Web Scraping and crawling with Python. Birmingham, Mumbai: Packt Publishing, 2016. 270 p. ISBN 978-1-78439-978-8.

KUROSE, James F.; ROSS, Keith W. **Redes de computadores e a Internet**: uma abordagem top-down. 6ª. ed. São Paulo: Pearson Education do Brasil, 2013. 658 p. ISBN 978-85-430-1443-2.

LAUDON, Kenneth C.; LAUDON, Jane P. **Sistemas de informação gerenciais**. 11ª. ed. [S. l.]: Pearson, 2015. 508 p. ISBN 978-85-430-0585-0.

LAUREANO, Marcos Aurelio Pchek. **Uma Abordagem para a Proteção de Detectores de Intrusão Baseada em Máquinas Virtuais**. 2004. 103 p. Dissertação (Mestre em Informática Aplicada) - Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada., Curitiba, 2004. Disponível em: <http://www.mlaureano.org/projects/vmids/dissert-laureano.pdf>. Acesso em: 9 dez. 2022.

LEARN, Microsoft. **Perguntas frequentes sobre o Subsistema Windows para Linux**. [S. l.]: Microsoft, 2022. Disponível em: <https://learn.microsoft.com/pt-br/windows/wsl/faq>. Acesso em: 4 dez. 2022.

MAGRI, Eduardo. **Web Scraping: legal ou ilegal?**. Estadão, [S. l.], 26 jan. 2022. Política. Disponível em: <https://politica.estadao.com.br/blogs/fausto-macedo/web-scraping-legal-ou-ilegal/#:~:text=Ainda%20n%C3%A3o%20h%C3%A1%20uma%20regulamenta%C3%A7%C3%A3o,nos%20Termos%20de%20Uso%20da>. Acesso em: 20 nov. 2022.

MARCONDES, Carlos H.; SAYÃO, Luis F. **À guisa de introdução**: repositórios institucionais e livre acesso. In: SAYÃO, Luis; TOUTAIN, Lídia B.; ROSA, Flavia G.; MARCONDES, Carlos H. (org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-21. Disponível

em: https://repositorio.ufba.br/bitstream/ufba/473/3/implantacao_repositorio_web.pdf. Acesso em: 30 nov. 2022.

MATTHES, Eric. **Curso intensivo de Python: Uma introdução prática e baseada em projetos à programação**. São Paulo: Novatec, 2016. 592 p. ISBN 978-1- 59327-603-4.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data: A Revolution that Will Transform how We Live, Work, and Think**. New York: Houghton Mifflin Harcourt, 2013. 242 p. ISBN 978-0-544-002-297-7. E-book (242 p.).

MAZIERO, Carlos Alberto. **Sistemas operacionais: conceitos e mecanismos**. Curitiba: DINF - UFPR, 2019. 456 p. ISBN 978-85-7335-340-2. E-book (456 p.).

MENDES, Daniel. **Numpy e Pandas: Bibliotecas básicas para Análise e Ciência de dados**. [S. l.], 18 mar. 2023. Disponível em: <https://www.linkedin.com/pulse/numpy-e-pandas-bibliotecas-b%C3%A1sicas-para-an%C3%A1lise-ci%C3%A4ncia-daniel-mendes/?originalSubdomain=pt>. Acesso em: 29 mar. 2023.

MENEZES, Nilo Ney Coutinho. **Introdução à Programação com Python: Algoritmos e lógica de programação para iniciantes**. 2ª. ed. São Paulo: Novatec, 2017. 325 p. ISBN 978-85-7522-559-2.

MITCHELL, Ryan. **Web Scraping com Python: Coletando mais dados da web moderna**. 2ª. ed. São Paulo: Novatec, 2018. 350 p. ISBN 978-85-7522-734-3.

MOURA, Beatriz; HENRIQUE, Mateus. **O que é o HTML e suas tags? Parte 1: estrutura básica**. [S. l.]: Alura, 14 dez. 2021. Disponível em: <https://www.alura.com.br/artigos/o-que-e-html-suas-tags-parte-1-estrutura-basica>. Acesso em: 18 nov. 2022.

MUELLER, Suzana P. M.; PERUCCHI, Valmira. **Universidades e a produção de patentes: tópicos de interesse para o estudioso da informação tecnológica**. Perspectivas em Ciência da Informação, [online], v. 19, ed. 2, p. 15-36, 26 set. 2014. DOI <https://doi.org/10.1590/1981-5344/1828>. Disponível em: <https://doi.org/10.1590/1981-5344/1828>. Acesso em: 9 dez. 2022.

MUNZERT, Simon et al. **Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining**. 1. ed. Reino Unido: Wiley, 2015. 446 p. ISBN 978-1-118-83481-7. E-book (446 p.).

NUMPY. **NumPy documentation**. [S. l.], 2022. Disponível em: <https://numpy.org/doc/stable/>. Acesso em: 14 abr. 2023.

OMPI, Organização Mundial da Propriedade Intelectual. **Classificação Internacional de Patentes.**, 2020. 51 p. Disponível em: <https://www.ipdec.org/themes/ipdec/material-de-apoio/classificacao-de-patentes.pdf>. Acesso em: 7 dez. 2022.

ORACLE. **O que é um banco de dados relacional (RDBMS)?**. [S. l.], 2022. Disponível em: <https://www.oracle.com/br/database/what-is-a-relational-database/>. Acesso em: 26 set. 2022.

ORACLE, MySQL. **MySQL Workbench Manual: Chapter 1 General Information**. [S. l.]: Oracle, 2022. Disponível em: <https://dev.mysql.com/doc/workbench/en/wb-intro.html>. Acesso em: 26 nov. 2022.

ORACLE, VirtualBox. **User Manual**. [S. l.]: Oracle, 2004-2022. Disponível em: <https://www.virtualbox.org/manual/UserManual.html>. Acesso em: 8 dez. 2022.

ORACLE, MySQL. **MySQL Connector: Python Developer Guide**. [S. l.], 2023. Disponível em: <https://dev.mysql.com/doc/connector-python/en/connector-python-introduction.html>. Acesso em: 30 mar. 2023.

PANDAS. **Pandas documentation**. [S. l.], 3 abr. 2023. Disponível em: <https://pandas.pydata.org/docs/index.html>. Acesso em: 14 abr. 2023.

PROPRIEDADE. In: DICIO, **Dicionário Online de Português**. Porto: 7Graus, 2023. Disponível em: <https://www.dicio.com.br/propriedade/>. Acesso em: 03/03/2023.

RODRIGUES, Fernando de Assis. **Coleta de dados em redes sociais: privacidade de dados pessoais no acesso via Application Programming Interface**. Orientador: Ricardo Cesar Gonçalves Sant 'Ana. 2017. 679 p. Tese (Doutorado em Ciência da Informação) - Faculdade

de Ciência da Informação, Universidade Estadual Paulista, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149768>. Acesso em: 4 dez. 2022.

ROSA, Mário Fabrício Fleury (2022). **Pesquisa, Desenvolvimento & Inovação (PD&I) do Respirador Vesta**: Universidade De Brasília (UnB) como vetor da Pesquisa Translacional em Saúde. Dissertação de mestrado em Saúde Coletiva, Publicação 011A/2021, Programa de Pós-Graduação, Faculdade UnB Ceilândia, Universidade de Brasília, Brasília, DF, 102p.

SAHIN, Kevin. **Web Scraping using Selenium and Python**. [S. l.], 25 ago. 2022. Disponível em: <https://www.scrapingbee.com/blog/selenium-python/>. Acesso em: 26 set. 2022.

SCRAPY 2.6 documentation. [S. l.], 2022. Disponível em: <https://docs.scrapy.org/en/latest/>. Acesso em: 25 set. 2022.

SERRANO, Pablo Jiménez. **Dicionário Jurídico Atualizado**. Volta Redonda/RJ: Editora Jurismestre, 2021. 580 p. ISBN B09HWDWPNL. E-book (580 p.).

SOLID IT. DB-Engines. **DB-Engines Ranking**. [S. l.], abril 2023. Disponível em: <https://db-engines.com/en/ranking>. Acesso em: 17 abr. 2023.

STREHL, L. **Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes**. Ciência da Informação, scielo, v. 27, 09 1998. ISSN 0100-1965.

TANENBAUM, Andrew; FEAMSTER, Nick; WETHERALL, David. **Redes de computadores**. 6ª. ed. Porto Alegre: Pearson, 2021. 1175 p. ISBN 978-85-8260-560-8.

TAVARES, Frederico; CUNHA, Leticia Mello. **Web Scraping, um caso de uso para coletar metadados de artigos científicos publicados na biblioteca digital da IEEE**. 2021. 19 p. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Universidade Federal Fluminense, Rio de Janeiro, 2021. Disponível em: <https://app.uff.br/riuff/handle/1/24056>. Acesso em: 17 set. 2022.

WIECHORK, Karina. **Extração automatizada de dados de documentos em formato PDF: aplicação a grandes conjuntos de exames educacionais**. Orientador: Andrea Schwertner

Charao. 2021. 73 p. Dissertação (Mestrado em Ciência da Computação) - Programa de Pós-Graduação em Ciência da Computação, Santa Maria, RS, 2021. Disponível em: <https://repositorio.ufsm.br/handle/1/23130>. Acesso em: 22 set. 2022.

WILLEMS, Caroline. **Python Numpy Array Tutorial**. [S. l.], 5 fev. 2023. Disponível em: <https://www.datacamp.com/tutorial/python-numpy-tutorial>. Acesso em: 14 abr. 2023.

WIPO. **Sistema de Classificação de patentes**. 2023. Disponível em: <http://ipc.inpi.gov.br/classifications/ipc/ipcpub/?notion=scheme&version=20230101&symbol=none&menulang=pt&lang=pt&viewmode=f&fipopc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=02n&initial=A&cwid=none&tree=no&searchmode=smart>. Acesso em: 22 de mar. de 2023.

YANG, Hongxia. Design and Implementation of Data Acquisition System Based on Scrapy Technology. **2019 2nd International Conference on Safety Produce Informatization (IICSPI)**, Chongqing, China, p. 417-420, 2019. DOI 10.1109/IICSPI48186.2019.9096044. Disponível em: <https://ieeexplore-ieee-org.ez54.periodicos.capes.gov.br/document/9096044>. Acesso em: 25 set. 2022.

ZHANG, Haifei et al. Deep Web Data Integration System Based on Scrapy Framework. **2022 2nd Asia Conference on Information Engineering (ACIE)**, Haikou, China, p. 18-23, 2022. DOI 10.1109/ACIE55485.2022.00012. Disponível em: <https://ieeexplore-ieee-org.ez54.periodicos.capes.gov.br/document/9831725/authors#authors>. Acesso em: 25 set. 2022.

APÊNDICE I

The screenshot displays the MySQL Workbench interface. The main window shows a query result for the 'repositorioUnB' table. The query executed is 'SELECT * FROM db_mestrado.repositorioUnB;'. The result grid shows 20 rows of data. The columns are 'idRepositorioUnB', 'titulo', and 'orientador'. The Action Output pane at the bottom shows the execution of the query, indicating that 45662 rows were returned.

#	idRepositorioUnB	titulo	orientador
1	1		
2	2	A indisciplina que muda a arquitetura : a dinâmica do espaço doméstico no Distrito Federal	Frederico Rosa Borges de Holanda
3	3	Políticas de austeridad y gobernabilidad : el caso de la aplicación del programa de ajuste estructural y de estabilización del FMI en el Brasil	Ramón Cotarelo
4	4	Categorias lexicais e funções na linguagem de especialidade da economia	Enilde Leite de Jesus Faulstich
5	5	A avaliação da performance no canto lírico : uma análise de conteúdo	Cristina de Souza Grossi
6	6	Oportunidades e desafios para o desenvolvimento do turismo em Brasília-DF	Gilson Zehetmeyer Borda
7	7	Da estrutura de expressões nominais quantificadas em posição de tópico	Cilene Aparecida Nunes Rodrigues
8	8	A influência da força do clima no absentismo	Katia Elizabeth Puente-Palacios
9	9	Sobre soluções que mudam de sinal via Teoria de Enlace	Marcelo Fernandes Furtado
10	10	Seleção de plantas antagonistas para manejo de Meloidogyne ethiopica em videira e quivi	Regina Maria Dechechi Gomes Mattos
11	11	Leaf area index and canopy open ness estimation using high spatial resolution image quickbird	
12	12		
13	13	Sob o céu das valquírias : as concepções de heroísmo e honra dos pilotos de caça na Grande Guerra (1914-18)	João Lacerda
14	14	Fraternidade eclética : comendo memórias e construindo identidades	Marcia de Melo Martins Kuyumjian
15	15	Desafios a verticalização agroindustrial familiar de oleaginosas no semiárido baiano: direcionadores de competitividade das usinas de Lapão e Olindina	João Nildo de Souza Vianna
16	16	Hidrodinâmica do escoamento do sangue em microvasos	Francisco Ricardo da Cunha
17	17	Que objetividade para a Ciência da História? : o Índio brasileiro e a revolução francesa à luz da teoria da história, de Rûsen a Hayden White	Estevão Chaves de Rezende Martins
18	18	Mobilidade urbana e cidadania no Distrito Federal : um estudo do programa Brasília integrada	Maria Salete Kern Machado
19	19	Avaliação da memória emocional na doença de Parkinson	Carlos Alberto Bezerra Tomaz
20	20	Clonagem, análise da sentiência do gene p74 e filogenia de um novo vírus isolado da lagarta-do-álamo CondViorrhiza vestinialis	Maria Elita Batista de Castro

Tabela no banco de dados local com os dados garimpados do repositório institucional da UnB.

APÊNDICE II

The screenshot shows the MySQL Workbench interface. The query editor contains the following SQL statement:

```
SELECT * FROM db_mestrado.patentesINPI;
```

The result grid displays the following data:

#	idpatentesINPI	numPedido	dataDeposit	dataPublicacao	dataConcessao	nomeDepositante	classificacaoIPC	classificacaoCPC	titulo
1	1	BR 10 2022 011393 9	09/06/2022			FUNDAÇÃO UNIVERSIDADE DE...			
2	2	BR 10 2022 009997 9	23/05/2022			FUNDAÇÃO UNIVERSIDADE DE...			
3	3	BR 20 2022 006924 2	11/04/2022			FUNDAÇÃO UNIVERSIDADE DE...			
4	4	BR 10 2022 003906 2	03/03/2022			FUNDAÇÃO UNIVERSIDADE DE...			
5	5	BR 10 2022 000149 9	05/01/2022			UNIVERSIDADE ESTADUAL DE ...			
6	6	BR 10 2021 024335 0	01/12/2021			FUNDAÇÃO UNIVERSIDADE DE...			
7	7	BR 10 2021 022720 6	11/11/2021			UNIVERSIDADE FEDERAL DE M...			
8	8	BR 10 2021 021376 0	25/10/2021			UNIÃO BRASILEIRA DE EDUCA...			
9	9	BR 10 2021 021097 4	21/10/2021			FUNDAÇÃO UNIVERSIDADE DE...			
10	10	BR 10 2021 016932 0	26/08/2021			FUNDAÇÃO UNIVERSIDADE DE...			
11	11	BR 10 2021 013032 6 A2	30/06/2021	10/01/2023		UNIVERSIDADE FEDERAL DO ...	A01N 65/28 ; A01P 1/00	A01N 65/28 ; A01P 1/00	COMPOSIÇÃO A BASE DE ROMÃ (PUNICA GRANZ...
12	12	BR 10 2021 012922 0 A2	29/06/2021	03/01/2023		FUNDAÇÃO UNIVERSIDADE DE...	G06F 21/36 ; H04L 9/32	G06F 21/36 ; H04L 9/32	AUTENTICAÇÃO SEMÂNTICA GRÁFICA
13	13	BR 10 2021 011530 0 A2	14/06/2021	27/12/2022		EMBRAPA-EMPRESA BRASILEI...	C12N 15/113 ; A01N 63/60		SILENCIAMENTO DO GENE PARA A PROTEÍNA CA...
14	14	BR 10 2021 011096 1 A2	09/06/2021	20/12/2022		FUNDAÇÃO UNIVERSIDADE DE...	C12N 15/83		VETORES VIRÁIS DERIVADOS DO MATERIAL GEN...
15	15	BR 10 2021 008059 0 A2	27/04/2021	01/11/2022		FUNDAÇÃO UNIVERSIDADE DE...	A61K 36/45 ; A61K 9/107 ; A61P 17/00		COMPOSIÇÃO CONTENDO EXTRATO DO FRUTO...

The interface also shows the table schema for 'patentesINPI' with columns: idpatentesINPI (int AI PK), numPedido (text), dataDeposit (text), dataPublicacao (text), dataConcessao (text), nomeDepositante (text), classificacaoIPC (text), classificacaoCPC (text), titulo (text), resumo (text), nomeInventor1 (text), nomeInventor2 (text), nomeInventor3 (text), and nomeInventor4 (text). The query execution log shows the query completed at 13:55:26, returning 260 rows in 0.0085 seconds.

Tabela no banco de dados local com os dados garimpados da base do INPI.

ANEXO I

Para validar, acesse <http://www.abrascoeventos.org.br/validacao/?cod=6108480>



Certificamos que o trabalho

INTEGRAÇÃO DO MARCO LEGAL DA INOVAÇÃO NO PROCESSO DE PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I) EM SAÚDE.

dos autores: WESLAINE MACEDO GUIMARAES DOS SANTOS; JESSICA SOUSA OLIVEIRA; MÁRIO FABRÍCIO FLEURY ROSA, foi aprovado na modalidade Apresentação Curta Assíncrona, no 13º Congresso Brasileiro de Saúde Coletiva ocorrido de 21 a 24 de novembro de 2022.

24 de novembro de 2022

Rosana Teresa Onocko Campos
Presidente da Abrasco

Isabela Cardoso de Matos Pinto
Presidente do Congresso

Realização:



Apoio:



Certificado de Aprovação no 13º Congresso Brasileiro de Saúde Coletiva.

ANEXO II

Para validar, acesse <http://www.abrascoeventos.org.br/validacao/?cod=6108480>



Certificamos que o trabalho

INTEGRAÇÃO DO MARCO LEGAL DA INOVAÇÃO NO PROCESSO DE PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I) EM SAÚDE.

dos autores: WESLAINE MACEDO GUIMARAES DOS SANTOS; JESSICA SOUSA OLIVEIRA; MÁRIO FABRÍCIO FLEURY ROSA, foi apresentado na modalidade Apresentação Curta Assíncrona, no 13º Congresso Brasileiro de Saúde Coletiva ocorrido de 21 a 24 de novembro de 2022.
24 de novembro de 2022

Rosana Teresa Onocko Campos
Presidente da Abrasco

Isabela Cardoso de Matos Pinto
Presidente do Congresso

Realização:



Apoio:



Certificado de Apresentação no 13º Congresso Brasileiro de Saúde Coletiva.

ANEXO III

☰

Anais 2022
ISSN: 2965-2154

Login (/user/login/ashna)

INTEGRAÇÃO DO MARCO LEGAL DA INOVAÇÃO NO PROCESSO DE PESQUISA, DESENVOLVIMENTO E INOVAÇÃO (PD&I) EM SAÚDE.

SANTOS, WESLAINE MACEDO GUIMARAES DOS (SANTOS, WESLAINE MACEDO GUIMARAES DOS) (/abrascao-2022/autores/weslaine-macedo-guimaraes-dos-santos?lang=pt-br)
OLIVEIRA, JESSICA SOUSA (OLIVEIRA, JESSICA SOUSA) (/abrascao-2022/autores/jessica-sousa-oliveira?lang=pt-br)
ROSA, MÁRIO FABRÍCIO FLEURY (ROSA, MÁRIO FABRÍCIO FLEURY) (/abrascao-2022/autores/mario-fabricio-fleury-rosa?lang=pt-br)
Vol 2, 2022 - 161950
Relato de Pesquisa

☆ (/user/login/ashnazg?destination=/proceedings/100325/_papers/161950/favorite%3Flang%3Dpt-br&lang=pt-br)

COMO CITAR ESSE TRABALHO?

Resumo

A Lei de Inovação é uma pedra fundamental no sentido de propiciar o incentivo à Inovação através dos Núcleos de Inovação Tecnológica (NITs), ligados à Instituição Científica, Tecnológica e de Inovação (ICT), como o Centro de Desenvolvimento Tecnológico (CDT), que tem o objetivo precípuo fomentar o desenvolvimento científico e tecnológico na Universidade de Brasília-UnB.

Objetivos

Analisar quantos projetos de pesquisas se consagraram como patentes, transferência de tecnologia e royalties, sobre produtos em tecnologias para a saúde.

Metodologia

Para essa análise, a coleta de dados será feita por meio de um web crawler focada, que é um programa que analisa e percorre páginas web para criar um índice dos dados de interesse. Com ele o usuário define como entrada um tópico e um conjunto de URLs de partida para orientar a busca através de páginas de interesse, tornando assim a coleta mais incisiva e triando os dados quase que automaticamente.

Resultados

O trabalho em tela, espera alcançar como resultado melhorias para a inovação dos programas de mestrado da Universidade de Brasília, com possibilidades de investimentos, através de parcerias público-privadas, ou ainda por intermédio de editais e programas de fomento, para transacionar suas pesquisas em saúde para a sociedade, com a utilização da Lei de Inovação.

Conclusões/Considerações

Por fim, deve-se averiguar a aplicação da Lei de Inovação, no intuito de incentivar a busca por investimentos, seja de parcerias público-privadas, quer seja por intermédio de Programas e Editais de fomentos, ou ainda através da instituição de novas diretrizes no âmbito UnB, aplicadas aos seus programas de pós-graduação, voltados para a área da saúde.



Compartilhe suas ideias ou dúvidas com os autores!



Sabia que o maior estímulo no desenvolvimento científico e cultural é a curiosidade? Deixe seus questionamentos ou sugestões para o autor!

Faça login para interagir (/user/login/ashnazg?destination=/abrascao-2022/trabalhos/integracao-do-marco-legal-da-inovacao-no-processo-
de-pesquisa-desenvolvimento-e?lang=pt-br)

Eixo Temático

- Eixo 09 - Saúde Coletiva, ciência, tecnologia, inovação e soberania

Publicação no Anais do Congresso Brasileiro de Saúde Coletiva.

Disponível em: <https://proceedings.science/abrascao-2022/trabalhos/integracao-do-marco-legal-da-inovacao-no-processo-de-pesquisa-desenvolvimento-e?lang=pt-br>