

# ESTRATÉGIA PARA CERTIFICAÇÃO DE PATENTES EM REPOSITÓRIOS INTERNACIONAIS

*Strategy for patent certification in international repositories*  
*Estrategia para la certificación de patentes en repositorios internacionales*



Raulivan Rodrigo da Silva  
Mestre em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, Minas Gerais, Brasil.  
Professor do ensino básico, técnico e tecnológico, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Divinópolis, Minas Gerais, Brasil.  
Lattes: <http://lattes.cnpq.br/6465763991846426>  
Orcid: <https://orcid.org/0000-0002-2740-1045>



Thiago Magela Rodrigues Dias  
Doutor em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, Minas Gerais, Brasil.  
Professor do ensino básico, técnico e tecnológico, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Divinópolis, Minas Gerais, Brasil.  
Lattes: <http://lattes.cnpq.br/4687858846001290>  
Orcid: <https://orcid.org/0000-0001-5057-9936>



Washington Luís Ribeiro de Carvalho Segundo  
Doutor em Informática, Universidade de Brasília (UnB), Brasília, Distrito Federal, Brasil.  
Coordenador Técnico da Área de Tratamento, Análise e Disseminação da Informação Científica, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasília, Distrito Federal, Brasil.  
Lattes: <http://lattes.cnpq.br/9453481318889500>  
Orcid: <https://orcid.org/0000-0003-3635-9384>

## Resumo

**Introdução:** A Plataforma Lattes é uma valiosa fonte de informação para os mais diversos tipos de análise, entretanto, devido ao grande volume de dados e o preenchimento das informações ser de responsabilidade do próprio indivíduo, pode eventualmente ocasionar inconsistências nos dados inseridos, o que enfatiza a necessidade de mecanismos de validação das informações. **Metodologia:** O processo metodológico foi dividido em duas partes, em que a primeira é caracterizada pela coleta e construção da base de dados local com dados de patentes depositadas no Brasil e currículos da Plataforma Lattes e a segunda parte consiste na descrição do processo de certificação das patentes informadas nos currículos da Plataforma Lattes. **Resultados:** Foi construída uma base de dados local composta por 903.979 registros de patentes oriundos da Espacenet e 76.619 registros de patentes distribuídos em 28.581 currículos coletados na Plataforma Lattes, todos organizados e estruturados

em um banco de dados relacional. Em relação ao processo de certificação das patentes registradas na Plataforma Lattes, foi possível certificar 44.348 registros de patentes, aproximadamente 58% das patentes foram certificadas.

**Conclusão:** Cerca de 1% de todos os currículos da Plataforma Lattes, possuem informações sobre o depósito de patentes, base composta por mais de 7 milhões de currículos (2022). Deste montante, nem todos puderam ser validados e certificados na Espacenet devido à inconsistência nos dados registrados, evidenciando a necessidade de mecanismos de validação e certificação dos dados patentários.

**Palavras-chave:** patente; Plataforma Lattes; certificação; indicadores.

### Abstract

**Introduction:** The Lattes Platform is a valuable source of information for the most diverse types of analysis, however, due to the large volume of data and filling in the information is the responsibility of the individual, it may eventually cause inconsistencies in the data entered, which emphasizes the need information validation mechanisms. **Methodology:** The methodological process was divided into two parts, in which the first is characterized by the collection and construction of the local database with data from patents filed in Brazil and curricula from the Lattes Platform and the second part consists of the description of the certification process of the informed patents in the curricula of the Lattes Platform. **Results:** A local database was built, consisting of 903,979 patent records from Espacenet and 76,619 patent records distributed in 28,581 CVs collected on the Lattes Platform, all organized and structured in a relational database. Regarding the certification process of patents registered on the Lattes Platform, it was possible to certify 44,348 patent registrations, approximately 58% of the patents were certified. **Conclusion:** About 1% of all CVs on the Lattes Platform contain information on the filing of patents, a base made up of more than 7 million CVs (2022). Of this amount, not all could be validated and certified on Espacenet due to inconsistency in the recorded data, highlighting the need for validation and certification mechanisms for patent data.

**Keywords:** patent; Lattes Platform; certification; indicators.

### Resumen

**Introducción:** La Plataforma Lattes es una valiosa fuente de información para los más diversos tipos de análisis, sin embargo, debido al gran volumen de datos y al hecho de que el llenado de la información es responsabilidad del individuo, eventualmente puede ocasionar inconsistencias en los datos ingresados, lo cual enfatiza la necesidad de mecanismos de validación de la información. **Metodología:** El proceso metodológico se dividió en dos partes, en la que la primera se caracteriza por la recolección y construcción de la base de datos local con datos de patentes registradas en Brasil y currículos de la Plataforma Lattes y la segunda parte consiste en la descripción del proceso de certificación de las patentes informadas en los planes de estudio de la Plataforma Lattes. **Resultados:** Se construyó una base de datos local, compuesta por 903.979 registros de patentes de Espacenet y 76.619 registros de patentes distribuidos en 28.581 CV recopilados en la Plataforma Lattes, todos organizados y estructurados en una base de datos relacional. En cuanto al proceso de certificación de patentes registradas en la Plataforma Lattes, se logró certificar 44.348 registros de patentes, aproximadamente el 58% de las patentes fueron certificadas. **Conclusión:** Alrededor del 1% de todos los CV en la Plataforma Lattes contienen información sobre la presentación de patentes, una base compuesta por más de 7 millones de CV (2022). De esta cantidad, no todo pudo ser validado y certificado en

*Espacenet debido a la inconsistencia en los datos registrados, destacando la necesidad de mecanismos de validación y certificación de datos de patentes.*

**Palabras clave:** *patentar; Plataforma de Lattes; certificación; indicadores.*

# 1. Introdução

---

No ambiente acadêmico é uma prática comum a publicação de artigos científicos, podendo inclusive, ser considerada a principal forma pelo qual é disseminado grande parte do conhecimento científico com a comunidade. A publicação de artigos é tão importante no cenário acadêmico que também é utilizada como métrica para avaliar a produtividade dos pesquisadores e utilizada como critério de avaliação em diversos concursos e processos seletivos (FERREIRA *et al.*, 2023).

Neste contexto, assim como a publicação de artigos, as patentes são a principal forma de compartilhamento do conhecimento tecnológico, além da possibilidade de gerar benefícios financeiros para os inventores, instituições e até mesmo para o país, contribuindo diretamente com o desenvolvimento tecnológico. Do mesmo modo que a publicação de artigos, o depósito de patentes também pertence ao ambiente acadêmico, uma vez que grande parte das invenções patenteadas foram geradas em universidades ou com a participação de universidades sejam elas públicas ou privadas (RIBEIRO; FREY; AZEVEDO, 2022).

Nessa conjuntura, como mecanismo de registro de produção técnica e científica tem-se os currículos da Plataforma Lattes. A qual é mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) tendo como objetivo de integrar em uma única plataforma bases de currículos acadêmicos de instituições públicas e privadas (LOPES, 2022).

## 1.1 Definição do problema

Mediante ao apresentado a Plataforma Lattes pode ser considerada uma valiosa fonte de informação para os mais diversos tipos de análise, entretanto, devido ao grande volume de dados e o preenchimento das informações ser de responsabilidade do próprio indivíduo, pode eventualmente ocasionar inconsistências nos dados inseridos (DIGIAMPIETRI *et al.*, 2012). O que enfatiza a necessidade de mecanismos de validação das informações contidas nos currículos da Plataforma Lattes, a fim de contribuir com a consistência dessa importante base de dados. Assim como o processo de certificação das Teses e Dissertações informadas nos currículos da Plataforma Lattes realizadas por meio do Oasisbr - Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (BRASIL, 2022).

## 1.2 Objetivos

Logo, este trabalho tem como principal objetivo propor uma estratégia de certificação de patentes informadas nos currículos da Plataforma Lattes em um repositório internacional de patentes.

## 2. Procedimentos Metodológicos

---

O processo metodológico foi dividido em duas partes, a primeira consiste na construção de um repositório local de dados, já na segunda, a definição do processo de certificação das patentes.

Para a construção do repositório local de dados, *a priori* foi definida a coleta de dados de patentes depositadas no Brasil disponibilizados na base de dados da Espacenet. A Espacenet é um serviço de pesquisa amplo e consistente de cobertura mundial que oferece acesso gratuito a informações sobre patentes desde os anos de 1782 até os dias atuais, contendo dados de mais de 120 milhões de documentos de patentes de diversos países (ESPACENET, 2023).

A escolha da Espacenet foi feita devido à sua capacidade de disponibilizar um serviço web que oferece acesso ao banco de dados do EPO (*European Patent Office*) por meio de API (*Application Programming Interface*), chamado OPS (*Open Patent Services*). O que torna viável o desenvolvimento de algoritmos computacionais que possibilitem a consulta e a coleta dos dados de forma mais eficiente.

Logo, baseado na documentação da OPS<sup>1</sup>, foram implementadas rotinas computacionais para a coleta dos dados de interesse via API, considerando o período de depósito das patentes de janeiro de 1900 a dezembro de 2022.

Como resultado, a API de busca retorna uma lista contendo o número de depósito das patentes que atenderam aos filtros. De posse desta lista, para cada patente da lista foi realizada uma requisição ao serviço “*publication*”, informando como parâmetro o número da patente, por exemplo “BR 9104913A”, para obter mais informações da patente, tais como título, resumo, classificação, inventores, depositantes, entre outras. A coleta foi realizada entre os meses de janeiro a abril do ano de 2023.

Dando sequência no processo de aquisição dos dados, a próxima etapa foi coletar currículos registrados na Plataforma Lattes que possuem informações de patentes. O processo de coleta e seleção dos dados curriculares da Plataforma Lattes foi realizado por meio do *framework* LattesDataXplorer (DIAS, 2016). A coleta dos currículos foi realizada em janeiro de 2023.

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

---

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

Após a construção do repositório local de dados, foi realizada a certificação das patentes informadas nos currículos da Plataforma Lattes. O processo de certificação consiste em identificar a patente informada pelo pesquisador em seu currículo no conjunto de patentes recuperada na Espacenet, utilizando como critério de identificação o número de depósito e o título da patente. Mediante ao exposto, a seguinte estratégia foi desenvolvida:

- I. Recuperar o número de depósito da patente no currículo.
- II. Exemplo: BR 2020 12345-9B2
- III. Remover caracteres especiais, tais como . - # , / \ " ' ( ) e espaços.
- IV. Exemplo : BR 2020 12345-9 (B2) BR2020123459B2
- V. Remover os seguintes termos: SIGILO, PROTOCOLO, REGISTRO, N°, INPI, EM ANDAMENTO, DEPÓSITO.
- VI. Exemplo: : N° BR2020123459B2 BR2020123459B2
- VII. Verificar se a quantidade de caracteres restantes é maior que 4, caso seja inferior a 4 caracteres, encerra-se o processo, considerando a patente como não certificada.
- VIII. Pesquisar a patente na base de dados, caso localize ir para o passo 10, caso contrário ir para o passo 6.
- IX. Caso o número inicie com BR, pesquise novamente desconsiderando os 3 últimos dígitos, caso localize ir para o passo 10, caso contrário ir para o passo 7.
- X. Exemplo: BR2020123459B2 BR202012345
- XI. Caso o número inicie com os prefixos CI, DI, MU ou PI, pesquisar incluindo o BR no início e desconsiderar os 3 últimos dígitos, caso localize ir para o passo 10, caso contrário ir para o passo 8.
- XII. Exemplo: PI2020123459B2 BRPI202012345
- XIII. Caso o número inicie com os prefixos CI, DI, MU ou PI, pesquisar substituindo o prefixo por BR e desconsiderar os 3 últimos dígitos, caso localize ir para o passo 10, caso contrário ir para o passo 9.
- XIV. Exemplo: PI2020123459B2 BR202012345
- XV. Caso não inicie com BR, pesquisar incluindo o BR e desconsiderar os 3 últimos dígitos. Considerando se uma ou mais palavras do título da patente informado na Plataforma Lattes, removendo *stopwords*, contém no título da patente na Espacenet. Caso localize ir para o passo 10, caso contrário a patente é considerada como não certificada.

**XVI.** É removido as *stopwords* e as acentuações do título da patente informada no currículo, bem como o título de cada patente localizada. Posteriormente o título da patente informada no currículo é comparado com o título de cada patente localizada, calculando a similaridade dos títulos usando o método *Levenshtein*. Considerando como certificada somente as patentes com similaridade de título maior ou igual a 85%.

Logo, após todas as etapas descritas, foi possível certificar as patentes que estão nos currículos cadastrados na Plataforma Lattes, agregando diversas outras informações, bem como, ter acesso à página *web* do registro da referida patente na Espacenet, que inclui outras informações como documentos e imagens da invenção.

### 3. Resultados

Como resultado, mediante as estratégias definidas neste trabalho foi possível coletar 903.979 registros de patentes da base de dados Espacenet. Já com relação aos dados de patentes dos currículos da Plataforma Lattes, foi possível coletar 76.619 registros de patentes distribuídos em 28.581 currículos (Quadro 1).

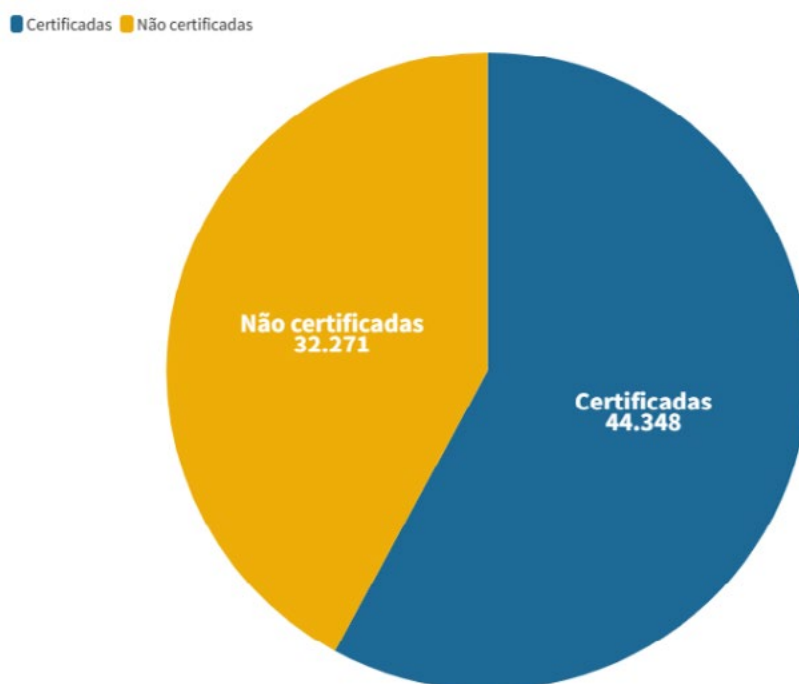
**Quadro 1 – Quantitativos dos dados coletados**

Tipo de registro	Quantidade
Patentes coletadas na Espacenet	903.979
Patentes coletadas na Plataforma Lattes	76.619
Currículos da Plataforma Lattes com registros de patentes	28.581

Fonte: Elaboração do autor (2023)

Por sua vez, contemplando o objetivo principal, foi possível certificar 44.348 do total de 76.619 registros de patentes coletados na Plataforma Lattes, um percentual de aproximadamente 58% das patentes. A Figura 1 apresenta um gráfico comparativo com as patentes certificadas e não certificadas.

**Figura 1 – Patentes certificadas e patentes não certificadas.**



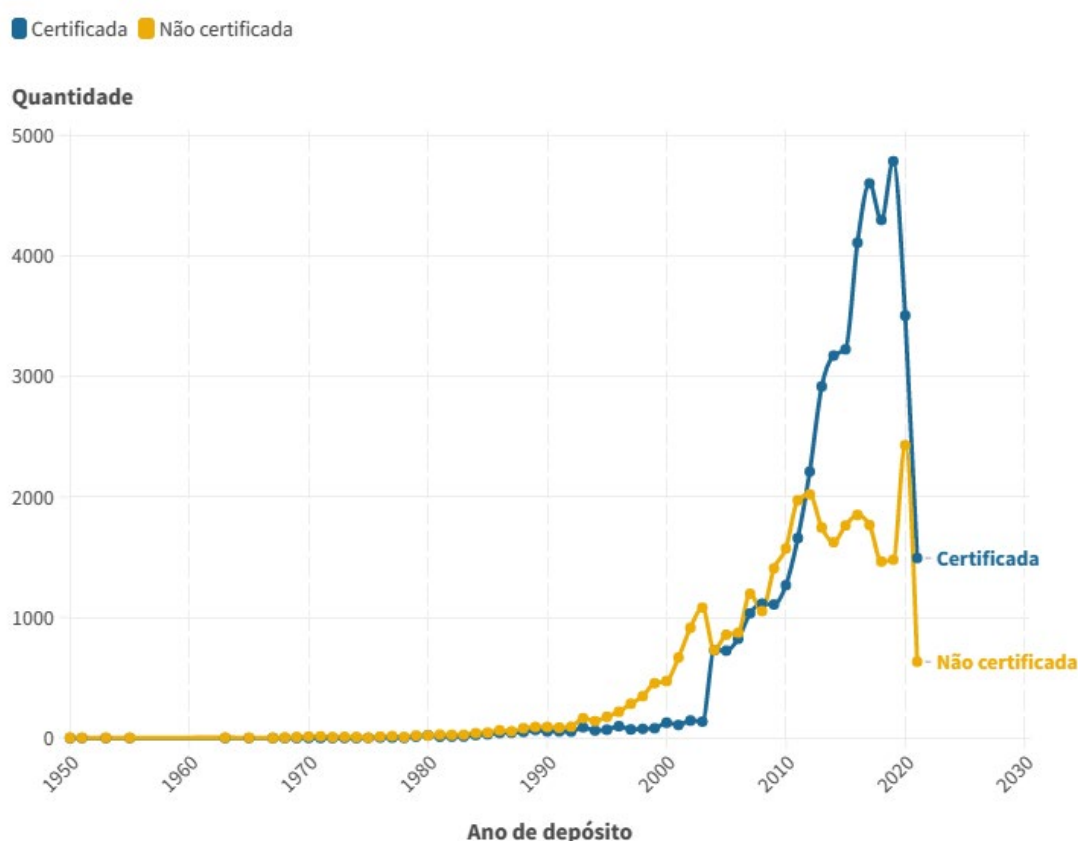
Fonte: Elaboração do autor (2023)



Esse resultado aponta a necessidade de se ter um certificador de dados de patentes informado na Plataforma Lattes, visto o alto número de patentes não certificadas. Com a certificação das patentes é possível aumentar a consistência dos dados que são informados pelo próprio autor.

Com intuito buscar compreender o motivo de quase 42% dos registros de patentes não serem certificadas, a Figura 2 apresenta o total de depósito de patentes por ano, considerando apenas as patentes informadas nos currículos, mostrando uma série para as certificadas e outra série para as não certificadas.

**Figura 2 – Patentes certificadas e patentes não certificadas.**



Fonte: Elaboração do autor (2023)

É possível observar que entre os anos 1950 e 1990, não há uma diferença considerável entre o total de patentes certificadas e não certificadas, porém, entre os anos de 1990 e 2002, o número de patentes não certificadas aumentou consideravelmente. Considerando o fato de o INPI possuir dois formatos distintos para definição do número de depósito das patentes, um primeiro utilizado para patentes mais antigas e um segundo formato para patentes depositadas a partir 02 de janeiro de 2012 visando atender o padrão internacional identificação de patentes, que é inclusive o padrão adotado pela Espacenet. O que implica que todas as patentes depositadas até o ano de 2011 no Brasil, e que estão disponíveis na Espacenet, tivessem sua numeração alterada buscando atender o padrão internacional (SILVA, 2022). Mediante ao apresentado, é possível explicar o motivo de que entre os anos de 1990 e 2002 aumentou o número de patentes não certificadas e posterior as 2011 esse cenário inverteu, o número de patentes certificadas ultrapassou o número de patentes não certificadas.

No período de 1950 a 2011 foram certificadas 10.019 (39,30%) e não certificadas foram 15.476 (60,70%) registros de patentes. Já entre os anos de 2012 e 2021 foram certificadas 34.329 (67,16%) e não certificadas foram 16.785 (32,84%). Vale ressaltar que se somar todos os valores obtém-se um total de 76.609, faltando assim, 10 registros de patentes do quantitativo informado no Quadro 1, esses 10 registros faltantes foram patentes não certificadas que não tiveram o ano informado no currículo da Plataforma Lattes, por isso não entraram na análise temporal.

A estratégia proposta neste trabalho certifica apenas patentes depositadas no Brasil, por tanto, é importante verificar se existem patentes que não foram depositadas no Brasil no conjunto das não certificadas. Logo, foi possível identificar que 7.977 registros de patentes não certificadas não foram depositados no Brasil, justificando a sua não certificação mediante as estratégias definidas.

Em continuidade, após a identificação de patentes não depositadas no Brasil o conjunto de patentes não certificadas passa a ser composto por 24.294 registros, ou seja, cerca de 32% os quais apresentam diversas inconsistências que justificam a sua não identificação, como por exemplo, as patentes de números de depósito "62/1781"; "N.Ref. 05385"; "PI ? 0006594-3"; "n.05059911-9" e "INPI PI0705367" que após a aplicação da estratégia não geraram números de depósitos que pudessem ser certificados e nem pelos seus títulos. Outro exemplo foram as patentes de número de depósito "WO2010054452", "US2009246833-A1" que não atendem ao padrão brasileiro de numeração de patente e foram informadas pelos seus proponentes como patentes brasileiras.

## 4. Considerações Finais

---

Com a certificação das informações inseridas nos currículos é possível validar que os trabalhos científicos, patentes, softwares, dentre outros, foram inseridos de maneira correta, evitando possíveis informações incorretas.

Por essa razão, a certificação das patentes informadas nos currículos da Plataforma Lattes se destaca como um fator que promove maior credibilidade e gera confiança quanto a veracidade da informação. Ademais, conforme apontam os resultados o número de patentes informados nos currículos da Plataforma Lattes vem aumentando ao longo dos anos, acredita-se que esse percentual de currículos com informações de patente venha aumentar.

## Referências

---

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Conselho Nacional de Desenvolvimento Científico e Tecnológico. **Oasisbr passa a ser utilizado para certificação de Teses e Dissertações no Currículo Lattes:** Teses e dissertações do Lattes agora contam com o selo de certificação de autenticidade do Oasisbr. 12 set. 2022. Disponível em: <http://bit.ly/3P9jjHm>. Acesso em: 4 maio 2023.

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

DIGIAMPIETRI, Luciano Antonio; MENA-CHALCO, Jesus Pascual; PÉREZ-ALCAZAR, José de Jesús; TUESTA, Esteban Fernandez; DELGADO, Karina Valdivia; MUGNAINI, Rogério; SILVA, Gabriela S. Minerando e Caracterizando Dados de Currículos Lattes. *In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRAS-NAM)*, 1., 2012, Curitiba. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2012. ISSN 2595-6094.

ESPACENET. **Espacenet patent search.** 2023. Disponível em: <https://worldwide.espacenet.com/patent/>. Acesso em: 27 abr. 2023.

FERREIRA, Chrystiano de Campos; PIRES, Rafael Cardoso; FERREIRA, Herika Rangel; MEDEIROS, Fernanda Fanciani; ZAMBRANA, Diego Eusebio Quispe; LOPES, Wagner Bachelos; KORTE, Rodolfo Luis. Como escrever e publicar um artigo científico: uma revisão da literatura. **Revista JRG de Estudos Acadêmicos**, Brasília, DF, v. 6, n. 12, p. 216-225, jan./jul. 2023. Disponível em: <https://revistajrg.com/index.php/jrg/article/view/496/539>. Acesso em: 27 abr. 2023.

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

RIBEIRO, Mariana Eleutério; FREY, Irineu Afonso; AZEVEDO, Paola. Classificação das Patentes em Universidades Federais na Escala TRL (Technology Readiness Level): estudo de caso a partir da Norma ISO 16290:2013. **Cadernos de Prospecção**, Salvador, BA, v. 15, n. 1, p. 117-130, jan./mar. 2022. DOI: <https://doi.org/10.9771/cp.v15i1.42173>.

SILVA, Raulivan Rodrigo da. **Coleta e tratamento de dados sobre a produção técnica brasileira:** um estudo baseado em patentes. 2022. 77 f. Dissertação (Mestrado em Modelagem Matemática e Computacional) — Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2022.