

A PLATAFORMA BRCRIS COMO FERRAMENTA PARA CERTIFICAÇÃO DE TESES E DISSERTAÇÕES NO BRASIL

The BrCris Platform as a tool for certification of Theses and Dissertations in Brazil
La Plataforma BrCris como herramienta para la certificación de Tesis y Disertaciones en Brasil



Thiago Magela Rodrigues Dias
Doutor em Modelagem Matemática e Computacional, Centro Federal de Educação
Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil.
Professor, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo
Horizonte, MG, 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, DF, Brasil.
Coordenador Técnico, Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict),
Brasília, DF, Brasil.
Lattes: <http://lattes.cnpq.br/9453481318889500>
ORCID: <https://orcid.org/0000-0003-3635-9384>



Jesús Pascual Mena Chalco
Doutor em Ciências da Computação, Universidade de São Paulo (USP), São Paulo, SP, Brasil.
Professor, Universidade Federal do ABC (UFABC), Santo André, SP, Brasil.
Lattes: <http://lattes.cnpq.br/4727357182510680>
ORCID: <http://orcid.org/0000-0001-7509-5532>



Vivian dos Santos Silva
Doutora em Ciência da Computação, Universität Passau, UNI/Passau, Alemanha.
Pesquisadora, Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, DF,
Brasil.
Lattes: <http://lattes.cnpq.br/6480849016794725>
ORCID: <https://orcid.org/0000-0001-6797-2400>



Tales Henrique José Moreira
Mestre em Modelagem Matemática e Computacional, Centro Federal de Educação
Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil.
Pesquisador, Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, DF,
Brasil.
Lattes: <http://lattes.cnpq.br/6579166915730771>
ORCID: <https://orcid.org/0000-0002-9865-6918>

Resumo

Introdução: Atualmente diversas iniciativas voltadas para a criação de sistemas que gerenciam a produção acadêmica e tecnológica de uma instituição, país ou área do conhecimento têm recebido atenção de diferentes pesquisadores sob diversos enfoques. Tais sistemas são conhecidos pela sigla CRIS (*Current Research Information Systems*) e destinam-se a agregar informações de diversas bases de dados a fim de fornecer relatórios, dados consolidados para que os pesquisadores possam analisar, bem como ser objeto de certificação da informação científica e tecnológica. Portanto, este trabalho apresenta o processo de integração de dados da Plataforma BrCris com o objetivo de fornecer dados confiáveis, para que seja possível realizar o processo de certificação de informações sobre a comunidade científica brasileira. **Metodologia:** Como metodologia para a certificação, um registro coletado na fonte “A” tem um atributo comum com o registro coletado na fonte “B”, podendo ser estabelecida uma ligação entre ambos, com certo grau de confiabilidade. Os outros atributos de registro podem ser mesclados para resultar em um único registro enriquecido, eliminando as réplicas. Todo este processo foi aplicado em um estudo de caso.

Resultados: Como estudo de caso foi certificado um conjunto de aproximadamente 700 mil teses e dissertações de currículos cadastrados na Plataforma Lattes com os dados exportados do BrCris. **Conclusão:** Além dos dados inicialmente certificados sobre o conjunto de teses e dissertações cadastrado nos currículos da Plataforma Lattes, a utilização dos dados agregados da Plataforma BrCris para o processo de certificação proporcionará diversos estudos bibliométricos que a priori seriam extremamente complexos em sua elaboração.

Palavras-chave: integração de dados; certificação; produção científica; BrCris.

Abstract

Introduction: Currently, several initiatives aimed at creating systems that manage the academic and technological production of an institution, country or area of knowledge have received attention from different researchers under different approaches. Such systems are known by the acronym CRIS (*Current Research Information Systems*) and are intended to aggregate information from various databases in order to provide reports, consolidated data for researchers to analyze, as well as to be the object of certification of scientific information and technological. Therefore, this work presents the data integration process of the BrCris Platform with the objective of providing reliable data, so that it is possible to carry out the process of certifying information about the Brazilian scientific community. **Methodology:** As a methodology for certification, a record collected from source “A” has a common attribute with the record collected from source “B”, and a link between both can be established, with a certain degree of reliability. The other record attributes can be merged to result in a single enriched record, eliminating duplicates. This whole process was applied in a case study. **Results:** As a case study, a set of approximately 700,000 theses and dissertations from curricula registered on the Lattes Platform with data exported from BrCris was certified. **Conclusion:** In addition to the initially certified data on the set of theses and dissertations registered in the Lattes Platform curricula, the use of aggregated data from the BrCris Platform for the certification process will provide several bibliometric studies that a priori would be extremely complex in their elaboration.

Keywords: data integration; certification; scientific production; BrCris.

DOI 10.22477/vi.widat.53

Resumen

Introducción: En la actualidad, diversas iniciativas encaminadas a crear sistemas que gestionen la producción académica y tecnológica de una institución, país o área del conocimiento han recibido atención de diferentes investigadores bajo diferentes enfoques. Dichos sistemas se conocen con el acrónimo CRIS (Current Research Information Systems) y tienen por objeto agregar información de diversas bases de datos con el fin de proporcionar informes, datos consolidados para que los investigadores analicen, así como ser objeto de certificación de información científica y tecnológica. . Por lo tanto, este trabajo presenta el proceso de integración de datos de la Plataforma BrCris con el objetivo de proporcionar datos confiables, para que sea posible realizar el proceso de certificación de información sobre la comunidad científica brasileña. **Metodología:** Como metodología para la certificación, un registro recolectado de la fuente "A" tiene un atributo común con el registro recolectado de la fuente "B", y se puede establecer un vínculo entre ambos, con cierto grado de confiabilidad. Los otros atributos de registro se pueden fusionar para dar como resultado un único registro enriquecido, eliminando las réplicas. Todo este proceso se aplicó en un caso de estudio. **Resultados:** Como estudio de caso, se certificó un conjunto de aproximadamente 700 mil tesis y disertaciones de planes de estudios registrados en la Plataforma Lattes con datos exportados de BrCris. **Conclusión:** Además de los datos inicialmente certificados sobre el conjunto de tesis y disertaciones registradas en los planes de estudio de la Plataforma Lattes, el uso de datos agregados de la Plataforma BrCris para el proceso de certificación proporcionará varios estudios bibliométricos que a priori serían extremadamente complejos en su elaboración.

Palabras clave: integración de datos; certificación; producción científica; BrCris .

1. Introdução

A produção do conhecimento científico é um processo que leva tempo e é incremental (HUANG; GLÄNZEL; ZHANG, 2021). Pesquisadores buscam as bases para suas pesquisas no desenvolvimento observado da sua área do conhecimento (estado da arte). A identificação do estado da arte de uma determinada área pode ser feita principalmente pela análise de publicações que concentram as informações a respeito daquele tópico do conhecimento.

O Brasil tem uma parcela relevante na produção científica internacional principalmente em nichos específicos ao atuar em sua economia. O país é líder na produção do conhecimento no contexto de América Latina (COLLAZO-REYES, 2014) e um atrator de talentos no contexto regional (SARAVIA; MIRANDA, 2004). O Brasil tem se destacado pela sua implementação de plataformas digitais de registro nacional do atuar de seus pesquisadores.

O maior destaque internacional foi o da Plataforma Lattes em que se observa o quão necessário, importante e estratégico é ter as informações científicas curriculares disponíveis de forma ampla (LANE, 2010). Similar à Plataforma Lattes existem outras plataformas nacionais que registram parte do atuar acadêmico (Sucupira, Banco de teses e dissertações, Banco de propriedade industrial e Portal Transparência Federal). Embora todas as informações dessas fontes de dados sejam abertas e livres para consulta, a integração e exploração para conhecer de forma ampla a ciência do Brasil ainda é um desafio.

Neste contexto, a informação resultante de pesquisas em Ciência, Tecnologia e Inovação (CT&I) é um dos principais insumos do desenvolvimento econômico e social de um país. É um elemento que orienta a produção sistematizada de conhecimento, não somente o conhecimento teórico, mas também o aplicado.

A informação em CT&I é altamente especializada e difere dos demais tipos de informação. Sua forma de produção é baseada em um método próprio - o método científico - e a disseminação de seus resultados também ocorre por meio de alguns processos específicos, que envolvem avaliação e validação, publicação e recuperação em fontes especializadas (MEADOWS, 1999).

Diversas fontes de informação abertas são desenvolvidas a partir dos resultados de pesquisas em Ciência, Tecnologia e Inovação (CT&I). Documentos técnicos e científicos são normalmente depositados em bases de dados e repositórios, garantindo a preservação, a visibilidade e a recuperação da informação contida.

CRIS (*Current Research Information Systems*) define um sistema de informação sobre todo o ecossistema de produção científica. Todas as informações sobre o ciclo de pesquisa científica estão organizadas em um só lugar, desde a divulgação, passando pelos Projetos, Pesquisadores, Instituições de Pesquisa e Laboratórios, até os *outputs* de uma pesquisa científica, como artigos científicos, teses, dissertações, livros, capítulos de livros, patentes e conjuntos de dados científicos (SIVERTSEN, 2019).

Assim, o BrCris visa estabelecer um modelo único de organização da informação científica de todo o ecossistema de pesquisa brasileira. Entre os agentes desse ecossistema estão pesquisadores, projetos, infraestruturas, laboratórios e instituições de pesquisa, financiadores, além de resultados de pesquisa expressos principalmente por publicações científicas, teses, dissertações, conjuntos de dados científicos, *software* e patentes.

Portanto, com a integração dos dados em um repositório de dados padronizado e devidamente avaliado, pode-se realizar todo um processo de certificação de dados originários de outras fontes ainda não validadas, proporcionando uma visão real da produção científica e tecnológica brasileira.

Os sistemas de certificação de dados caracterizam-se como mecanismos de verificação da origem, veracidade, integridade e confiabilidade dos conjuntos de dados armazenados em diferentes sistemas.

As ações auto declaratórias de um sistema podem ser validadas por um agente denominado “terceiro de confiança” (assinatura auto declaratória versus assinatura certificada), conferindo segurança e veracidade às informações prestadas.

Em fontes de informação de ciência e tecnologia, no entanto, sistemas de certificação de dados ainda são raros no Brasil, tendo em vista os diversos desafios envolvidos no processo de certificação. A ausência de identificadores persistentes que as diversas entidades que compõem todo o ecossistema da investigação científica destacam-se como uma das principais limitações.

Logo, este trabalho tem como objetivo principal apresentar a Plataforma BrCris como mecanismo de certificação de Teses e Dissertações registradas em currículos cadastrados na Plataforma Lattes.

2. Procedimentos Metodológicos

Com a integração dos dados em um repositório de dados padronizado e devidamente avaliado, pode-se realizar todo um processo de certificação de dados originários de outras fontes ainda não validadas, proporcionando uma visão real da produção científica e tecnológica brasileira.

Os sistemas de certificação de dados caracterizam-se como mecanismos de verificação da origem, veracidade, integridade e confiabilidade dos conjuntos de dados armazenados em diferentes sistemas.

Com os dados coletados e já duplicados, classificados e categorizados, eles podem ser posteriormente adaptados e validados, estabelecendo relações com registros de outras fontes. Um registro coletado na fonte “A” tem um atributo comum com o registro coletado na fonte “B”, podendo ser estabelecida uma ligação entre ambos, com certo grau de confiabilidade. Os outros atributos de registro podem ser mesclados para resultar em um único registro enriquecido, eliminando as réplicas. Um esquema de validação pode ser criado para descartar registros malformados, redundantes, inconsistentes ou ambíguos.

No contexto deste trabalho, o primeiro modelo de certificação testado é a integração da Plataforma Lattes com o Oasis.br. Com todo o conjunto de dados do Oasis.br incorporados no BrCris e por meio de diversos desdobramento no âmbito do Projeto BrCris, foi possível criar rotinas computacionais utilizando técnicas de Processamento de Linguagem Natural para a identificação de teses e dissertações declaradas nas sessões de formação acadêmica e orientações concluídas de um determinado currículo cadastrado na Plataforma Lattes, que também constavam no cadastro agregado pelo Oasisbr.

Desta forma, a Oasisbr torna-se o “terceiro de confiança” neste processo, sem a necessidade da pré-existência de um identificador persistente explicitamente atribuído à tese ou dissertação.

3. Resultados

Todo o processo de certificação é baseado em estratégias computacionais testadas e validadas em diversos estudos, por meio da análise de informações autodeclaradas, comparadas com informações inseridas em repositórios, bibliotecas digitais e portais agregados pelo Oasisbr. O selo de certificação Oasisbr é exibido próximo aos títulos das teses ou dissertações no currículo do usuário (FIGURA 1 E FIGURA 2).

Figura 1 – Fragmento de um currículo com uma certificação incluída.

Formação acadêmica/titulação

- 2014 - 2019** Doutorado em Informática (Conceito CAPES 5).
Universidade de Brasília, UnB, Brasil.
com **período sanduíche** em King's College London (Orientador: Maribel Fernández).
Título: Nominal Equational Problems Modulo Associativity, Commutativity and Associativity-Commutativity. Ano de obtenção: 2019.
Orientador: Meuricio Ayala Rincón.
Coorientador: Maribel Fernández.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: Métodos Formais; Reescrita Nominal; Formal Methods; Nominal Rewriting.
Grande área: Ciências Exatas e da Terra / Área: Matemática / Subárea: Álgebra / Especialidade: Lógica Matemática.
- 2008 - 2011** Mestrado em Informática (Conceito CAPES 5).
Universidade de Brasília, UnB, Brasil.
Título: Verificação de Propriedades do Cálculo Lambda_ex em Coq, Ano de Obtenção: 2011.
Orientador: Flávio Leonardo Cavalcanti de Moura.
Palavras-chave: verificação formal; cálculos de substituições explícitas; cálculo lambda ex.
Grande área: Ciências Exatas e da Terra / Área: Ciência da Computação / Subárea: Computação.
Grande Área: Ciências Exatas e da Terra / Área: Matemática / Subárea: Verificação Formal.
- 2007 - 2008** Graduação em Matemática - Licenciatura.
Universidade de Brasília, UnB, Brasil.
- 2002 - 2007** Graduação em Matemática (Bacharelado).
Universidade de Brasília, UnB, Brasil.

Fonte: Os autores (2023).

Figura 2 – Fragmento de um currículo com duas certificações incluídas.

Formação acadêmica/titulação

- 2012 - 2017** Doutorado em Informática (Conceito CAPES 5).
Universidade de Brasília, UnB, Brasil.
Título: Feature-Family-Based Reliability Analysis of Software Product Lines. Ano de obtenção: 2017.
Orientador: Vander Ramos Alves.
Coorientador: Genaina Nunes Rodrigues.
Palavras-chave: Software Product Lines; Reliability Analysis; Model Checking.
Grande área: Ciências Exatas e da Terra
- 2006 - 2009** Mestrado em Engenharia Elétrica (Conceito CAPES 3).
Pontifícia Universidade Católica de Minas Gerais, PUC Minas, Brasil.
Título: Reuso de Processos de Software baseado na Componentização de Processos e Conhecimento. Ano de Obtenção: 2009.
Orientador: Carlos Alberto Marques Pietrobon.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: componentes; processo de software; qualidade de software; Conhecimento; reuso.
Grande área: Ciências Exatas e da Terra
- 2001 - 2006** Graduação em Ciência da Computação.
Pontifícia Universidade Católica de Minas Gerais, PUC Minas, Brasil.
Título: Refactoring como um mecanismo para extração de componentes reutilizáveis.

Fonte: Os autores (2023).

Com o selo é possível obter, de forma rápida e simples, a comprovação documental do título informado e acessar o documento no Oasisbr. O certificado pode ser emitido automaticamente pela Plataforma Lattes ou solicitado manualmente pelo usuário.

Por meio de desenvolvimentos no âmbito do Projeto BrCris, foi possível criar um mecanismo inteligente para identificar teses e dissertações declaradas nas seções de treinamento e orientação concluídas de um determinado currículo Lattes, que também foram incluídas no conjunto de registros agregados pelo Oasisbr. Desta forma, a Oasisbr torna-se o “terceiro de confiança” neste processo, sem a necessidade da pré-existência de um identificador persistente explicitamente atribuído à tese ou dissertação.

Dentre os currículos da Plataforma Lattes, há aproximadamente 1,1 milhão de registros de teses e dissertações, em que 65% (aproximadamente 700 mil) já são passíveis de certificação. Outras 10 mil teses e dissertações defendidas no exterior também foram mapeadas para receber o selo de certificação.

As vantagens do processo de certificação são muitas. Por meio da certificação, é possível verificar que os trabalhos científicos, orientações, participação em bancas, entre outros elementos de fontes autodeclaradas são realmente verdadeiros, evitando informações falsas. Portanto, a certificação acaba sendo um fator que promove maior credibilidade e autoridade ao pesquisador.

4. Considerações Finais

Ressalta-se que os currículos Lattes são de natureza auto declaratória, destacando-se, portanto, a importância dos processos de certificação aplicados a essa base. Uma tese ou dissertação só é considerada documento oficial de titulação, se a versão final, e com as respectivas correções sugeridas pelos avaliadores, for depositada em repositório oficial e de acesso público.

Com base nas informações apresentadas, podemos concluir que a certificação oferecida pelo BrCris, considerando os dados do Oasisbr, é uma forma confiável e eficiente de comprovar a autenticidade de títulos acadêmicos, como teses e dissertações. Todo o processo utiliza estratégias computacionais testadas e validadas para comparar informações autodeclaradas com dados de repositórios, bibliotecas digitais e portais agregados. Com o selo de certificação da Oasisbr, é possível obter comprovação documental rápida e simples dos títulos informados, o que promove maior credibilidade e autoridade ao pesquisador. Além disso, o processo de certificação é capaz de identificar teses e dissertações declaradas nas seções de formação e orientações concluídas de um determinado currículo, sem a necessidade da pré-existência de um identificador persistente explicitamente atribuído à tese ou dissertação.

Referências

COLLAZO-REYES, Francisco. Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. **Scientometrics**, v. 98, p. 197-209, 2014. Disponível em: <https://doi.org/10.1007/s11192-013-1036-2>. Acesso em: 15 maio 2023.

HUANG, Ying; GLÄNZEL, Wolfgang; ZHANG, Lin. Tracing the development of mapping knowledge domains. **Scientometrics**, v. 126, p. 6201-6224, 2021. Disponível em: <https://doi.org/10.1007/s11192-020-03821-x>. Acesso em: 15 maio 2023.

LANE, Julia. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488-489, 2010. Disponível em: <https://doi.org/10.1038/464488a>. Acesso em: 16 maio 2023.

MEADOWS, Arthur Jack. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

SARAVIA, Nancy Gore; MIRANDA, Juan Francisco. Plumbing the brain drain. **Bulletin of the World Health Organization**, Geneva, v. 82, n. 8, p. 608-615, 2004. Disponível em: <https://tinyurl.com/yxrncrjm>. Acesso em: 12 maio 2023.

SIVERTSEN, Gunnar. Developing Current Research Information Systems (CRIS) as data sources for studies of research. In: GLÄNZEL, Wolfgang; MOED, Henk F.; SCHMOCH, Ulrich; THELWALL, Mike (ed.). **Springer handbook of science and technology indicators**. Switzerland: Springer, 2019. p. 667-683. Disponível em: https://doi.org/10.1007/978-3-030-02511-3_25. Acesso em: 18 maio 2023.