



WTICIFES/2022

Workshop de Tecnologia da Informação e Comunicação das Instituições Federais de Ensino Superior

14ª edição | V. 1

Organizadoras

Ligia Maria Carvalho Sousa (UNILAB)
Eunice Pereira dos Santos Nunes (UFMT)

**Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)**

Workshop de tecnologia da informação e comunicação das instituições federais de ensino superior [livro eletrônico] : v. 1 / organização Lígia Maria Carvalho Sousa, Eunice Pereira dos Santos Nunes. -- 14. ed. -- Brasília, DF : Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior - Andifes, 2022. PDF.

Vários autores.

Bibliografia.

ISBN 978-85-67619-02-6

1. Institutos federais de educação, ciência e tecnologia - Brasil 2. Tecnologia da informação e da comunicação - Uso - Brasil I. Sousa, Lígia Maria Carvalho. II. Nunes, Eunice Pereira dos Santos.

22-116662

CDD-004.6072081

Este periódico foi especialmente editado a partir de conteúdos desenvolvidos para o Workshop de Tecnologia da Informação e Comunicação das Instituições Federais de Ensino Superior (WTICIFES) apresentados durante a 14ª Edição do Workshop de Tecnologia da Informação e Comunicação das Instituições Federais de Ensino Superior do Brasil, realizado em Aracaju, entre 05 e 07 de julho de 2022, promovido pelo Colégio de Gestores de Tecnologia da Informação e Comunicação das IFES (CGTIC) e organizado pela Universidade Federal de Sergipe(UFS).

Aracaju- SE
2022

TODOS OS DIREITOS RESERVADOS

Permitida a reprodução total ou parcial desde que citada a fonte (XIII Workshop de Tecnologia da Informação e Comunicação das Instituições Federais de Ensino Superior, Aracaju-SE, Brasil).

Na ausência da indicação da fonte ou expressa autorização da Instituição, é proibida a sua reprodução total ou parcial, por qualquer meio ou processo, especialmente por sistemas gráficos, microfílmicos, fotográficos, reprográficos, fonográficos ou videográficos. Vedada a memorização e/ou recuperação total ou parcial, bem como a inclusão de quaisquer partes desta obra em qualquer sistema de processamento de dados. Essas proibições aplicam-se também às características da obra e à sua editoração. A violação dos direitos autorais é punível como crime (art. 184 e §§, do Código Penal, cf. Lei no. 6.895, de 17-12-1980) com pena de prisão e multa, conjuntamente com busca e apreensão e indenizações diversas (arts. 122, 123, 124 e 126, da Lei no. 5.988 de 14-12-1973, Lei dos Direitos Autorais).

Software Prod+: Eficiência e Automatização na descoberta da produção intelectual UNIFESP

Alexsandro Cardoso Carvalho¹, Tiago Rodrigo Marçal Murakami¹,
Virgínia M. L. Leite¹, Lidiane Cristina da Silva¹

¹ Superintendência de Tecnologia da Informação

Universidade Federal de São Paulo (UNIFESP) – São Paulo – SP – Brazil

{alexandro.carvalho, tiago.murakami, virginia.lapoian, lidiane.cristina}@unifesp.br

Resumo. *O artigo visa expor a experiência de implementação do Prod+Unifesp. O software processa os dados de produção científica do currículo Lattes e as disponibiliza de forma sistematizada para a universidade. Para isso, a UNIFESP firmou acordo com a CNPq para a disponibilização da plataforma Lattes como uma base de dados acadêmicos. O Prod+Unifesp apresenta, como resultados iniciais, duas dashboards e um catálogo das produções. Dessa maneira, o Prod+Unifesp mostra-se solução acessível e possibilita a gestão do conhecimento focada na produção intelectual da instituição.*

Palavra-Chave: Prod+, produção acadêmica, gestão do conhecimento.

1. Introdução

No Brasil, os programas de Pós-graduação *stricto sensu* (PPGs) são coordenados pela agência de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A CAPES observa e fiscaliza os programas, a formação de novos pesquisadores, a produção científica e a qualificação dessa produção. O cenário motiva os PPGs a buscar compreender a situação e posicionamento da produção realizada pela instituição no cenário global, nacional e local. Essa observação interfere na governança universitária em âmbito institucional [THIENGO et al. 2018].

Para armazenar e centralizar estes dados, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) desenvolveu a plataforma Lattes. A plataforma armazena informações dos pesquisadores brasileiros e cada registro individual é conhecido como currículo Lattes. Estes são um importante histórico das atividades científicas, acadêmicas e profissionais dos pesquisadores cadastrados na plataforma [AMORIN 2003]. O Lattes influencia também a avaliação de mérito e competência dos pleitos de financiamento [SANTOS 2017].

O desafio de manter e gerenciar essas informações demandam técnicas de Gestão do Conhecimento (GC). A GC é definida pela mistura de informações internas e externas, transformando-a em conhecimento acessível. Dessa forma, a GC é um processo gerido de forma centralizada por um órgão responsável pela compilação, síntese e integração de experiências e projeto para o desenvolvimento do trabalho [FUKUNAGA 2017].

A Universidade Federal de São Paulo (UNIFESP), fundada em 1933, conta com sete *campi* em seis cidades no estado de São Paulo, possuindo 57 cursos de pós-graduação *stricto sensu*. De acordo com o *ranking Times Higher Education* a UNIFESP se posiciona entre as melhores universidades federais brasileiras. Na edição de 2019 a UNIFESP ficou atrás apenas de duas universidades estaduais e uma privada. No quesito de citações, a UNIFESP se encontrou em 2019 em primeiro lugar nas universidades brasileiras.

Para manter e gerenciar essas informações, tornou-se necessário encontrar uma ferramenta que extraísse e tabulasse os dados dos currículos. Até 2015, utilizava-se o ScriptLattes. O ScriptLattes é um *software* livre desenvolvido por Jesús Mena-Chalco e Roberto Cesar Junior, projetado para identificar relacionamentos entre pesquisadores e para sistematizar as informações das atividades dos pesquisadores através da criação de relatórios. [MENA-CHALCO 2009]. Contudo, após 2015, o CNPq implantou o *captcha* que evita a extração em lotes com uso indevido dos dados em sites não autorizados [CORREA et al. 2017]¹.

Frente ao desafio de buscar uma ferramenta a um custo acessível, a UNIFESP tomou conhecimento do Coletaprod. Disponível como *software* livre o Coletaprod foi desenvolvido para o Sistema Integrado de Bibliotecas da Universidade de São Paulo, com a finalidade de auxiliar as equipes das Bibliotecas no acompanhamento da produção intelectual de seus docentes [MURAKAMI 2020]. O Coletaprod tem como principal característica processar as seções de dados da produção científica do currículo Lattes e disponibilizá-las ao usuário e alimentar outros repositórios de gestão do conhecimento. O Coletaprod foi desenvolvido na linguagem PHP e utiliza um banco de dados noSQL. Como complementar, está integrado com a ferramenta de *Business Intelligence* Kibana.

A partir do código do ColetaProd a Unifesp realizou modificações no código fonte, implementou novas funcionalidades, aperfeiçoou funções e criou documentação e identidade visual própria e após consulta aos usuários do foi definido para este software o nome de Prod+. O lançamento da versão 1.0, disponível em <https://unifesp.br/prodmais/>, ocorreu em evento público com a presença do sr. Nelson Sass, reitor da Unifesp, e do sr. Milton Shintaku, coordenador de Articulação, Geração e Aplicação de Tecnologia (CoTec) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)².

O objetivo deste estudo de caso é apresentar a experiência de desenvolvimento e implementação do Prod+ na UNIFESP para entender a produção científica institucional.

¹ Outras possíveis ferramentas, como Somos [CIT 2020] e Stela Experta [ESTÁCIO 2020], não se mostraram soluções viáveis em razão dessas por serem *softwares* com licenças proprietárias com alto custo de licenciamento.

²

<https://www.unifesp.br/eventos-antiores/item/5468-analise-da-producao-academica-sem-complicacao-lancamento-do-software-prod>

2. Metodologia

Para obter acesso aos dados da Plataforma Lattes, primeiramente, a UNIFESP firmou um Protocolo de Cooperação Técnica [CNPq(b) 2020] com o CNPq, tornando possível que esses dados fossem consumidos via API *Simple Object Access Protocol* (SOAP). São realizadas requisições, enviando o número do CPF do pesquisador e havendo cadastro na plataforma Lattes é retornado o identificador Lattes. Esse identificador é uma sequência numérica individual e única de 16 dígitos.

Posteriormente o currículo obtido é descompactado e processado pelo Prod+Unifesp. O arquivo descompactado utiliza o formato *xml*, que consiste em dados hierárquicos, agrupados e identificados por rótulos. Ao fim desse processo, os dados processados são encapsulados com a tecnologia *json* e armazenados em um banco de dados noSQL Elasticsearch, que disponibiliza para consultas e cruzamento de informações. Assim, de forma dinâmica, conforme os dados disponíveis e a necessidade, há o cruzamento desses dados com outras bases de dados que também utilizam o mesmo processo SOAP em APIs próprias. Esse cruzamento valida e checka dados obtidos na plataforma Lattes. A disponibilização das informações é realizada por uma interface de consulta e também utiliza técnicas de *Business Intelligence*, gerando *dashboards* dinâmicos para os usuários.

3. Resultados

Com o Prod+Unifesp, foram processados 1.727 currículos Lattes computando 229.381 registros de produção intelectual. Criou-se três módulos (i) catálogo de produções dos currículos, (ii) *dashboard* com dados pessoais dos currículos e (iii) *dashboard* com dados de produções dos currículos.



Uma ferramenta de busca da produção científica de
pesquisadores da UNIFESP.

Figura 1. tela inicial

O catálogo de produções dos currículos tem diferentes opções de filtragem, possibilitando o refinamento dos resultados. Também existem opções para exportar a produção bibliográfica para a plataforma Open Researcher and Contributor ID (ORCID) e está em desenvolvimento uma página do pesquisador que será lançada em breve.

Os *dashboards* com dados pessoais dos currículos e o *dashboard* com dados de produções dos currículos apresentam as mesmas funcionalidades. Ambos os *dashboards* trazem infográficos dinâmicos que também funcionam como filtros e apresentam diversas informações. Os *dashboards* com dados pessoais dos currículos trazem ao usuário a titulação dos autores, o gênero, a nacionalidade entre outras informações que possam ser utilizadas para identificar comportamentos, origem e/ou analisar o histórico dos pesquisadores, traçando o seu perfil profissional. A figura 2 mostra essa interface.

O *dashboard* de dados de produções dos currículos traz os dados das publicações e pode-se observar os infográficos divididos por data de publicação, produção por universidade, produção por departamento, tipo de produção, por idioma, nuvem de palavras-chaves, grande área de conhecimento e por autores. A figura 3 exibe essa interface.

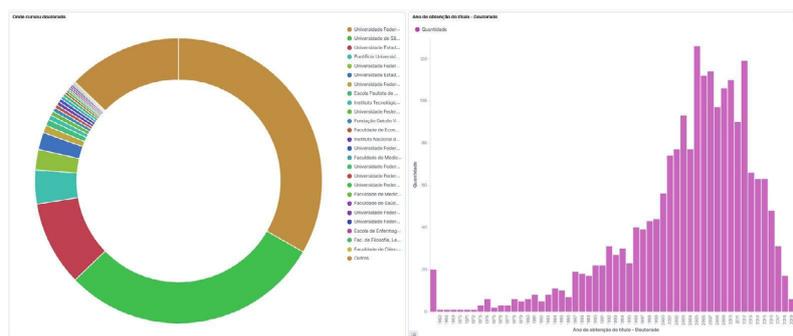


Figura 2. infográfico apresentado nos *dashboards* com dados pessoais.



Figura 3. infográfico apresentado nos *dashboard* com dados da produção intelectual.

O Prod+Unifesp foi licenciado como software livre e está disponível em repositório público³ sendo possível também o estabelecimento de acordo de cooperação sem ônus financeiro para transferência de tecnologia para as instituições de ensino superior que

³ <https://github.com/unifesp/prodmais>

desejarem utilizar o software, além da possibilidade de colaboração e desenvolvimento conjunto de aprimoramentos.

4. Conclusão

O PROD+UNIFESP mostra-se uma solução acessível e possibilita a gestão do conhecimento focada na produção intelectual da instituição. O Prod+UNIFESP criou uma ponte de informações que possibilita a análise da produção científica e o comportamento dos pesquisadores institucionais, permitindo que a UNIFESP compreenda as suas tendências científicas e o seu impacto na esfera nacional e internacional. Tal afirmação pode ser verificada na plataforma Google Analytics, onde os relatórios mostram um quantitativo de aproximadamente três mil usuários no mês.

Ademais, o Prod+Unifesp utilizou a plataforma Lattes como uma das bases de dados. Com o auxílio do Prod+Unifesp, foi possível consolidar os dados do Lattes e cruzá-los com outras fontes, como CrossRef.

Referências

- Amorin, C. V. (2003). “Curriculum vitae organization: The Lattes software platform”. *Pesquisa Odontológica Brasileira*, p. 17(1):18–22.
- CIT (2020). “Somos UNIFESP”, <http://www.somos.unifesp.br/>, Março.
- CNPq(b) (2020). “Acordos institucionais”, <https://tinyurl.com/w8h8ot8>, Março.
- Correa, T. S., Suzuki, M. B., Cintra, R. P., and Costa, L. S. F. (2017). “O fim do ScriptLattes? Uma análise de suas funcionalidades, alternativas para o presente e perspectivas para o futuro”, <https://tinyurl.com/utl2rrm>, Março.
- Estácio (2020). “Stela experta”, <https://tinyurl.com/szkg642>, Março.
- Fukunaga, M. S. (2017). “Gestão do conhecimento: conceitos e definições”. *Sociedade Brasileira de Gestão do Conhecimento*, pages 1–23, <https://tinyurl.com/wcw3537>, Março.
- Mena-Chalco, J.P., Junior, R. M. C. (2009). “scriptLattes: an open-source knowledge extraction system from the Lattes platform”. *Journal of the Brazilian Computer Society*, p. 15(4):31–39.
- Murakami, T. R. M. (2020). “Coletaprod”, <http://doi.org/10.5281/zenodo.3633209>, Março.
- Santos, S. D. (2017). “Método De Agrupamento Hierárquico a partir de currículos acadêmicos”. In: *Universidade Nove de Julho. Dissertação [Mestrado]*.
- THE (2019). “Emerging economies university rankings 2019”, <https://tinyurl.com/qvsvvdu>, Março.
- Thiengo, L. C., Bianchetti, L., and Mari, C. L. (2018). “Rankings acadêmicos e universidades de classe mundial: relações, desdobramentos e tendências”, <https://tinyurl.com/sz5rxb5>, Março.