

SCIENSH: Sistema de apoio à gestão de pesquisa e inovação baseado em cientometria

SCIENSH: management support system research and innovation based on scientometry

Adrielson Ferreira Justino^{1,2} , Fabiano Paulo Nhoatto¹ , Éfren Lopes de Souza¹ , Fábio Manoel França Lobato^{1,2} 

¹Instituto de Engenharia e Geociências – Universidade Federal do Oeste do Pará (UFOPA)

Santarém, Pará – Brasil

²Centro de Ciências Tecnológicas – Universidade Estadual do Maranhão (UEMA)
São Luis, Maranhão – Brasil

{fabio.lobato, efren.souza}@ufopa.edu.br

Abstract. *Scientometrics is a study that provides indicators and analyses to measure the technical-scientific contribution of the most diverse areas, identifying potentialities and opportunities. This study proved especially relevant for the Universidade Federal Oeste do Pará (UFOPA). It was observed great difficulty consulting and analyzing data on its technical-scientific productions; knowing these data is vital for any institution because several decisions can be made from them. Thus, this work aims to use scientometrics to improve research management at UFOPA through software that consolidates and disseminates the Institution's production data quickly and centrally. This software collects and analyzes data directly from the Lattes Platform and Google Scholar, generating results that help in the management of research and innovation and the evaluation of techno-scientific policies, providing greater transparency, social inclusion, and data-oriented decision-making capacity. This work aims to generate local impacts in the UFOPA environment where the software was developed and implemented; the idea is that it allows managers to evaluate and make decisions and that researchers and society, in general, can interact and obtain the knowledge produced by the Institution.*

Keywords. *Public management; Data-oriented decision making; Transparency; Scientometrics; Information visualization.*

Resumo. *A cientometria é um estudo que fornece indicadores e análises para mensurar a contribuição técnico-científica das mais diversas áreas, identificando potencialidades e oportunidades. Esse tipo de estudo se mostrou especialmente relevante para a Universidade Federal Oeste do Pará (UFOPA), onde se observou grande dificuldade de consultar e analisar dados sobre as suas produções técnico-científicas. Conhecer esses dados é de vital importância para qualquer instituição, pois a partir deles diversas tomadas de decisão podem ser feitas. Dessa forma, o objetivo deste trabalho é usar cientometria para melhorar a gestão da pesquisa na UFOPA por meio de um software que consolida e difunde os dados de produção da Instituição de forma rápida e centralizada. Esse software coleta e analisa dados diretamente da Plataforma Lattes e Google Scholar, gerando resultados que auxiliam na gestão da pesquisa e inovação e na avaliação de políticas tecnocientíficas, dando maior transparência, inserção social e capacidade de tomada de decisão orientada a dados. Este trabalho visa gerar impactos locais, no ambiente da UFOPA onde o software foi desenvolvido e implantado, possibilitando que gestores avaliem e tomem decisões e que pesquisadores e a sociedade de modo geral possam interagir e obter o conhecimento produzido pela Instituição.*

Palavras-Chave. *Gestão pública; Decisão orientada a dados; Transparência; Cientometria; Visualização da Informação.*

1. Introdução

A utilização de ferramentas avançadas de Tecnologia da Informação e Comunicação (TIC) na gestão pública auxiliam e permitem maior agilidade, rapidez, transparência e eficiência aos processos em diversos setores [Dal Molin and da Costa 2018]. Desta forma, promovem aprimoramento da utilização de recursos públicos e facilitam as atividades desenvolvidas por gestores e servidores públicos. Um exemplo de utilização nesse contexto são os sistemas de informação que têm sido requisitados por instituições governamentais para apoiar seu trabalho em diferentes áreas, como gerenciamento e monitoramento de atividades [Fernandes et al. 2016].

No âmbito da gestão de pesquisa e inovação em Instituições de Ensino Superior (IES), torna-se imperativo o uso de métodos para avaliação de sua produção técnico-científica. Visto que, estas têm como base o princípio de indissociabilidade entre ensino, pesquisa e extensão e que os professores devem desenvolver atividades nestes três eixos [Moher et al. 2018, Carpenter et al. 2014]. Neste sentido, a análise e extração de dados de produção científica, identificação de padrões bibliométricos e mensuração da ciência são tópicos relevantes na área de Bibliometria e Cientometria [Mena-Chalco and Júnior 2013, Mingers and Leydesdorff 2015, Nhoatto et al. 2021].

Os dados estatísticos elaborados por meio dos estudos bibliométricos mensuram a contribuição do conhecimento científico derivado das publicações em determinadas áreas, representando tendências de uma determinada pesquisa e identificação de temas para novas investigações [Zeng et al. 2017, Dias and Moita 2018]. Adentrando especificamente no âmbito da pesquisa, a cientometria é o campo de estudo que em-

prega análise quantitativa para se mensurar o impacto e a produtividade de pesquisas científicas mediante parâmetros e indicadores baseados em modelos matemáticos [Parra et al. 2019]. Por meio desses indicadores que se guia o processo de investigação e mensuração da ciência [Parra et al. 2019]. Eles são úteis no planejamento e avaliação de políticas públicas voltadas para pesquisas [Santos and Kobashi 2009]. Com eles é possível ainda identificar relações de colaboração, de forma a vincular dois temas normalmente abordados separadamente: análise de redes sociais e cientometria [Digiampietri et al. 2012, Brito et al. 2016].

Existem inúmeras formas de utilização das técnicas cientométricas, das quais podem permitir a identificação de: (i) autores mais relevantes na área; (ii) autores com mais publicações em determinada área; (iii) maturidade de um certo tema ou área de um autor, grupo ou instituição; (iv) tendências de áreas de pesquisa; (v) produtividade de autores; (vi) conexões de autores e redes de colaboração; (viii) autores importantes para as redes de pesquisa, ou seja, autores que conectam e produzem pesquisas com diversos grupos; (ix) artigos relevantes para iniciar novas pesquisas; (x) árvores genealógicas de autores por orientações ou citações. Esses indicadores podem ajudar a mensurar a contribuição dada por cada país ao acervo mundial de pesquisa, além de localizar geograficamente as pesquisas por autor, assunto e acompanhar o crescimento em cada área da ciência – exemplos desses indicadores incluem: a quantidade de trabalhos, quantidade de coautorias, quantidade de publicações por autor, por instituição ou país, quantidade de patentes, quantidade de artigos publicados e fator de impacto [Zeng et al. 2017].

Sendo assim, a aplicação de indicadores é uma necessidade inequívoca na otimização de recursos, sempre limitados, que se destinam ao fomento da ciência e tecnologia [Silva and Bianchi 2001]. Conforme relata [Lopes 2012], a busca pela excelência em áreas de pesquisa e a competição pelos recursos financeiros de agências de fomento são fatores que impulsionam a realização destes estudos e a adoção de métricas cientométricas. Além disso, editais para financiamento de projetos de pesquisa por agências de amparo à produção de pesquisas no Brasil consideram dados dos currículos cadastrados na plataforma Lattes como uma das formas de avaliação [Dias et al. 2018]. A Plataforma Lattes (PL) possui uma vasta fonte de informação para a criação e análise de redes sociais de pesquisadores, tornando-a uma fonte extremamente rica para análise da produção científica brasileira [Balancieri et al. 2005].

Este tipo de estudo se revela pertinente em particular para Universidade Federal do Oeste do Pará (UFOPA), que apresentou dificuldades na gestão da pesquisa e inovação. Nesta Instituição também foi observado grande dificuldade de consultar e analisar dados sobre as suas produções técnico-científicas devido a multitudes de produções técnico-científicas desenvolvidas e docentes envolvidos. Especificamente a UFOPA conta com 485 docentes ativos, mais de 70 mil produções e média 4.769 produções anuais nos últimos dez anos. Desta forma, todo o estudo é baseado no contexto da UFOPA, sendo essa aplicada como estudo de caso deste trabalho. Sendo assim, este estudo visa otimizar a gestão da pesquisa e inovação da instituição com o desenvolvimento de uma ferramenta computacional que coleta, processa e disponibiliza informações sobre as produções tecnocientíficas dos pesquisadores da instituição e permita a tomada de decisões orientada a dados. E que permita auxiliar na tomada de decisões referente aos editais de pesquisa da

instituição. Entender esses dados é de fundamental importância à qualquer instituição, pois por meio deles diversas tomadas de decisão podem ser realizadas.

A gestão da pesquisa na instituição é gerenciada pela Pró-Reitoria de Pesquisa, Pós-graduação e Inovação Tecnológica (PROPPIT), mais especificamente pela Diretoria de Pesquisa (DPE). Esta unidade gerencia a pesquisa por meio de editais de fomento e políticas públicas. Para isso utiliza programas como Stella Experta para analisar as produções da instituição e melhorar sua estratégia, visando otimizar recursos e melhorar resultados em prol da produção de conhecimento.

Considerando a necessidade de avaliação da produção técnico-científica de pesquisadores com intuito de fomentar a pesquisa na instituição, juntamente com a inviabilidade de análise manual dado o volume de informações e a variedade de atividades desenvolvidas, é patente a necessidade de filtros temporais no cálculo de índices de produtividade dos docentes para distribuição de recursos de editais de fomento, além de outros filtros como área ou seleção de docentes pertencentes a um grupo de pesquisa etc.

Tal funcionalidade é difícil de ser articulada com o uso de planilhas eletrônicas, uma vez que os dados ficam dispersos (*e.g.*: artigos publicados por pesquisadores da instituição em uma planilha e membros de grupo de pesquisa em outra). Com isso, requer-se a utilização de ferramentas computacionais para auxiliar na coleta e análise de dados da Plataforma Lattes e prover indicadores quantitativos e visualização da produtividade e desempenho de instituições e pesquisadores. Destaca-se ainda que a análise manual é uma tarefa demorada e altamente suscetível a erros [Alves et al. 2012, Cesso and Ferraz 2017].

A integração de sistemas de diferentes áreas de atuação do governo pode ser útil para os entes governamentais, visto que o cruzamento de dados pode fornecer informações relevantes que não podem ser obtidas pela análise de sistemas individualmente. Nesse sentido, o uso dessas ferramentas proporciona vários benefícios para os usuários, que vão desde a automatização, praticidade e maior velocidade na execução de tarefas até a redução de custos [Pressman and Maxim 2016, Silveira and Nunes 2013].

Nesse cenário, encontra-se uma gama de *softwares* disponíveis, a exemplo do ScriptLattes¹, Stela Experta² e Intellectus³. O próprio CNPq disponibiliza às Instituições a ferramenta Lattes Extrator⁴ que permite acesso aos dados da plataforma Lattes para que implementem ferramentas de análises de suas produções. No entanto, algumas ferramentas não podem ser acessadas ou distribuídas livremente, outras carecem de alguma funcionalidade, como a geração de redes de colaboração, perfil do pesquisador ou indicadores específicos de uma instituição [Meglioransi 2019, Nhoatto et al. 2021].

À luz dos fatos apresentados, o presente estudo tem como objetivo automatizar o processo de coleta e análise da produção técnico-científica dos docentes por meio de uma ferramenta computacional denominada: Sciensh. Para isso, tal ferramenta combina análise cientométrica e análise de redes sociais, tendo como base os dados obtidos a par-

¹Disponível em: <http://scriptlattes.sourceforge.net>. Acesso em 06 de julho de 2021.

²Disponível em: <http://site.stelaexperta.com.br>. Acesso em 06 de julho de 2021.

³Disponível em: <https://intellectus.ufrn.br>. Acesso em 06 de julho de 2021.

⁴Disponível em: <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. Acesso em 06 de julho de 2021.

tir da Plataforma Lattes. Além disso, o Sciensh funciona em sincronia com os sistemas institucionais – Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) e Intellectus, reaproveitando os investimentos já feitos e melhorando a integração institucional.

O restante deste artigo está organizado como segue. Na Seção 2 são discutidos os materiais e métodos utilizados para o desenvolvimento do estudo. Em sequência, na Seção 3 são apresentados os resultados obtidos, por meio da demonstração do artefato produzido à luz do método *Design Science Research* (DSR). Por fim, na Seção 4 as considerações finais serão apresentadas.

2. Materiais e Métodos

Este trabalho é uma pesquisa aplicada de caráter exploratório e descritivo. Seu objetivo está menos voltado para o desenvolvimento de teorias e mais voltado para a aplicação imediata numa realidade circunstancial [Gil 2008]. Foi utilizado o método DSR, que é uma abordagem bem aceita em projetos de pesquisa de sistemas de informação quando o objetivo final é o desenvolvimento de um produto, estratégias e serviços [Peffer et al. 2007, Hevner et al. 2010].

O DSR consiste na construção de conhecimento científico voltado à inovação por meio de artefatos, com a intenção de resolver problemas reais, com contribuição científica de caráter prescritivo [Dresch et al. 2015]. Na literatura vários estudos demonstram as contribuições teóricas e práticas resultantes da aplicação deste método [Paxiúba and Lima 2019, Caldas et al. 2018]. O DSR é composto de seis etapas postas em sequências de acordo com a definição utilizada por [Peffer et al. 2007], tal como disposto na Figura 1.

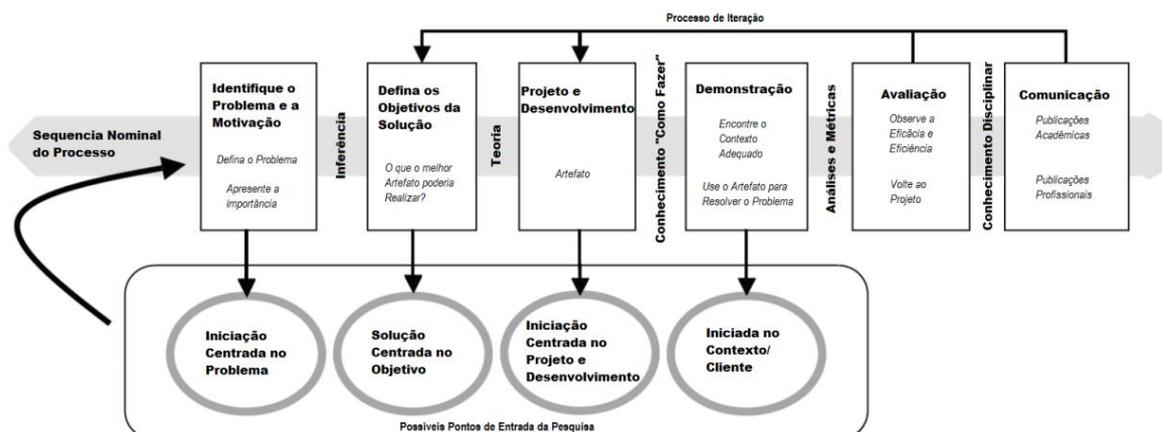


Figura 1. Etapas do DSR.

A Figura 1 apresenta as seis etapas do DSR, a saber: (i) Identificação do problema e motivação de sua solução; (ii) Definição dos objetivos da solução; (iii) Projeto e desenvolvimento do artefato; (iv) Demonstração do artefato quanto a resolução de uma ou mais instâncias dos problemas levantados; (v) Avaliação da eficiência da solução na resolução dos problemas; e, por fim, (vi) Comunicação dos resultados técnico-científicos

alcançados. Importante notar que este é um processo iterativo e que permite início em uma das quatro primeiras fases.

A fase inicial consiste no entendimento de negócio e visa a identificação e a motivação do problema, juntamente com setor responsável pela pesquisa da UFOPA. Considerando que eram vários desafios apresentados, selecionaram-se os mais relevantes, tal como destacado na seção 2.1. A partir do estudo da relevância e viabilidade, dois foram selecionados e passou-se para a definição dos objetivos. Importante notar que a definição dos objetivos foi realizada conjuntamente com uma revisão da literatura e levantamento do estado da prática a fim de investigar técnicas e procedimentos para geração de indicadores de produção, conforme a seção 2.2.

Na fase de projeto e desenvolvimento (seção 3.1), definiu-se que o artefato a ser desenvolvido será a implementação de uma ferramenta computacional para coletar e permitir analisar os dados das produções e geração dos indicadores. Após a implementação do artefato, ocorreu a demonstração do mesmo como solução para as classes de problemas levantadas (seção 3.4). Nesta etapa ocorreu a implantação do sistema para testes e correção de eventuais falhas na implementação. A avaliação da solução foi executada por meio de análises qualitativas e quantitativas e do desenvolvimento e aplicação de questionários de satisfação aos *stakeholders* envolvidos (seção 3.5).

Em conclusão, a fase final consiste na comunicação contendo a apresentação da solução desenvolvida e a análise dos impactos gerados pela ferramenta para identificar os perfis e tendências de pesquisa, além da divulgação nos veículos de comunicação institucionais para a ciência da comunidade (seção 3.5). Esta etapa também foi realizada por meio da publicação do artigo em periódico e no registro de programa de computador junto ao Instituto Nacional de Propriedade Industrial (INPI).

2.1. Identificação do problema e motivação de sua solução

Um estudo recente de prospecção tecnológica realizado por [Nhoatto et al. 2021], trouxe à baila o cenário das principais ferramentas de extração e produção de indicadores bibliométricos, conforme mostra a Tabela 1.

O estudo identificou que nenhuma das ferramentas de extração analisadas atendeu todas as funcionalidades necessárias para avaliação cienciométrica de pesquisadores. Desta forma, por meio do desenvolvimento deste projeto pretendemos sanar a necessidade de uma nova solução tecnológica, que atenda funcionalidades indispensáveis para o auxílio em tomadas de decisão na gestão de pesquisas, principalmente no que tange à visualização da informação de forma interativa [Nhoatto et al. 2021].

A partir do problema que foi apresentado pela DPE, os requisitos do sistema foram passados pela PROPPIT, onde alguns desafios operacionais e de gestão foram elencados em conjunto aos servidores atuantes na DPE. A instituição apresentou algumas especificidades nos processos de pontuação dos docentes inscritos no edital de pesquisa 02 de 2020⁵. Além disso, também observaram-se os seguintes problemas na gestão de suas

⁵Disponível em: <http://www.ufopa.edu.br/media/file/site/proppit/documentos/2021/578384e5368d4d14f2696814c6ef2864.pdf>. Acesso em 06 de julho de 2021.

Tabela 1. Análise comparativa dos sistemas de extração e produção de indicadores bibliométricos

	Funcionalidades	SL	LL	LT	eL	SS	SE
(1)	Produções Bibliográficas por grupo	+	+	+	+	+	+
(2)	Indicadores do membro do grupo	-	-	+	-	+	+
(3)	Propriedade intelectual	+	+	+	-	+	+
(4)	Orientações	+	+	+	+	+	+
(5)	Prêmios científicos/Artísticos	+	+	+	-	-	+
(6)	Gráficos quantitativos de produção	+	+	-	+	+	+
(7)	Grafos de Coautoria	+	+	+	+	-	-
(8)	Gráficos interativos	-	-	-	-	+	+
(9)	Livre Acesso	+	+	+	+	+	-

SL -

ScriptLattes; LL - LucyLattes; LT - Lattices; eL - e-Lattes; SS - Sistema Somos; SE - Stela Experta. **Fonte:** Elaborado por [Nhoatto et al. 2021].

pesquisas: i) dificuldade de coleta e organização dos dados de pesquisa da instituição; ii) falta integração com sistemas institucionais; e iii) pouca transparência das produções técnico-científicas [Justino et al. 2023].

Por meio de entrevistas com os gestores e servidores identificou-se a necessidade de uma nova abordagem que pudesse substituir o sistema utilizado pela instituição, a fim de que a ferramenta proposta pudesse auxiliá-los. Inicialmente, o problema estava na dificuldade em se quantificar um índice de produtividade para distribuição de bolsas de pesquisa por meio de editais de fomento interno. Desse modo, em busca de agilidade e viabilidade da automatização de processos, os gestores do setor apresentaram uma demanda que diz respeito à dificuldade de mapeamento das produções feitas pelos pesquisadores da universidade.

Esse mapeamento visa quantificar e identificar as áreas de pesquisa das produções da universidade, descobrindo tendências, tópicos pesquisados, produções relevantes, pesquisadores relevantes, conexões entre pesquisadores e projetos desenvolvidos, por exemplo. Além de publicizar esses dados para a instituição e para a sociedade. Foi realizada análise de competidores, mapeada às funcionalidades e constatado o que poderia ser melhorado. Assim, a demanda por uma forma de visualização, publicidade e conhecimento destas informações foi tratada por meio de uma ferramenta computacional que unifica e automatiza a atualização dessas informações.

2.2. Definição dos objetivos da solução

Nesta etapa, iniciou-se um estudo sobre como as universidades federais do Brasil divulgam dados sobre suas pesquisas. No estudo realizado por [Nhoatto et al. 2021] foi efetuado um levantamento dos meios de divulgação e os tipos de informações divulgadas por outras universidades. Em sua pesquisa os autores observaram que a maioria das instituições divulgam seus dados de pesquisa por meio de relatórios e portais.

Também foi levantado por [Nhoatto et al. 2021] um estudo sobre os tipos de informações que essas IFES divulgam, para identificar os itens publicizados e que demonstram importância. Em que foi observado que os meios de divulgação empregados dificultam a visualização por serem estáticos. Consequentemente, tal formato não permite filtros de informações e atualização dos dados para maior transparência e interatividade. Algumas poucas universidades utilizam portais interativos que solucionam, parcialmente, estas últimas questões apresentadas.

A partir destas constatações, a proposta consistiu na implementação de um *software* que se integrasse com os editais cadastrados no SIGAA e buscasse os docentes participantes quantificando suas produções conforme um modelo de pontuação previamente cadastrado no próprio *software*. Essa solução visa sanar tais lacunas melhorando o gerenciamento de índices personalizados com base na produção dos pesquisadores e reduzir o tempo para apresentar os resultados dos editais.

Visando a publicidade e visualização dos dados de pesquisa, foi proposto a implementação de buscas e filtros para auxiliar a avaliação da pesquisa institucional, abalizando o processo de tomada de decisão, com acesso simples e flexível à produção técnico-científica institucional. Também, com a integração de dados da plataforma Lattes, *Google Scholar*, plataforma Sucupira e sistemas institucionais e utilizando algoritmos de fusão de dados unificar essas bases e exibir os dados por meio de *dashboards* objetivos e atualizados. Além de permitir compartilhar essa nova base para outras aplicações institucionais.

3. Resultados e discussões

Nesta seção são apresentadas as fases de impletação, demonstração e avaliação da ferramenta. Na qual, a partir dos materiais e métodos apresentados, consoante aos objetivos definidos, se propôs o desenvolvimento de uma ferramenta interativa, denominada de *Sciensh*, capaz de auxiliar na avaliação da pesquisa institucional, abalizando o processo de tomada de decisão, com acesso simples e flexível à produção técnico-científica institucional.

3.1. Projeto e desenvolvimento do artefato

Baseada nos dados da plataforma Lattes, *Google Scholar*, plataforma Sucupira e sistemas institucionais, essa ferramenta utiliza algoritmos de fusão de dados para unificar essas bases e exibir os dados por meio de *dashboards* objetivos e atualizados. Além de permitir compartilhar essa nova base para outras aplicações institucionais. Na Figura 2 é apresentado o fluxo da estrutura do sistema desenvolvido.

No fluxo disposto na Figura 2 podemos observar a coleta de dados obtida de diferentes bases de dados, como Plataforma Lattes, sistemas institucionais, *Google Scholar* e plataforma Sucupira, como demonstrado na seção 3.1.1. O processo é iniciado com a consulta de todos os docentes cadastrados e ativos no sistema interno, utilizando o CPF como critério de seleção. Para isso, o CNPq possibilita às instituições a extração de dados da Plataforma Lattes, incluindo a base de Currículos Lattes e o Diretório dos Grupos de

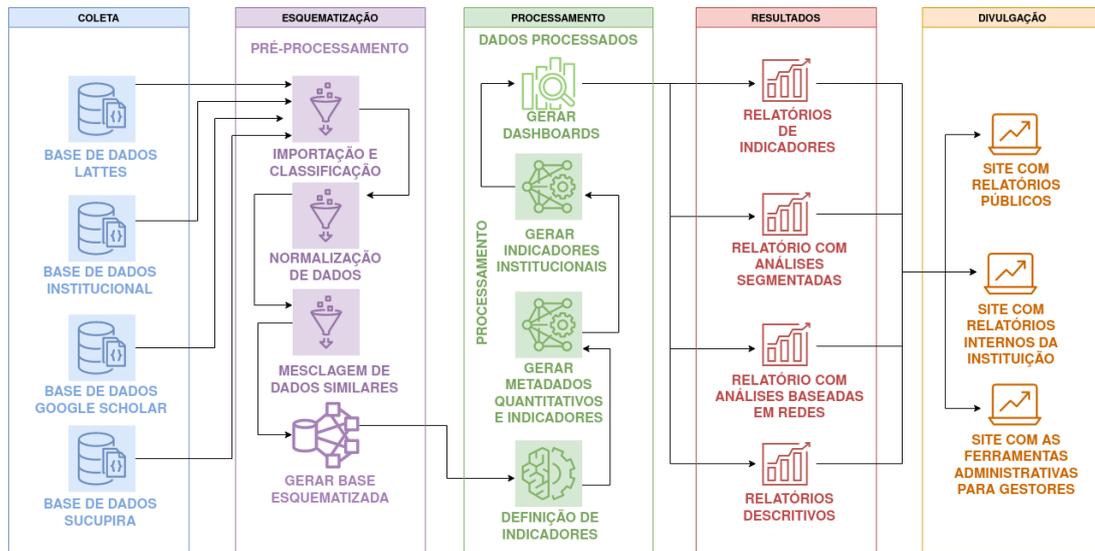


Figura 2. Fluxo do software desenvolvido.

Fonte: Elaborado pelos autores.

Pesquisa por meio da ferramenta Lattes Extrator ⁶.

Por meio do sistema Intellectus os currículos são importados, é realizado um parse e construído o banco de dados. Além disso, a plataforma Sucupira é utilizada para coletar o Qualis mais atual. Com o banco de dados construído, inicia-se o processo de fusão de dados, considerando também as páginas do *Google Scholar* previamente mapeadas com os docentes que confirmaram o email institucional @ufopa.edu.br em seu cadastro, ressalta-se que as tecnologias utilizadas neste projeto estão descritas na seção 3.2.

Após a coleta são efetuados processos de classificação e normalização dos dados e mesclagem de dados similares para obter uma fusão das informações obtém-se uma nova base esquematizada, conforme descrito na seção 3.3. No processamento das bases, são utilizados os identificadores das produções. Sendo assim, são mapeadas todas as produções dos docentes e verificasse se constam na base. A partir disso é possível identificar e quantificar indicadores, dados quantitativos e descritivos sobre pesquisadores e instituições, para então gerar os *dashboards* para expor e gerir esses dados, permitindo gerar relatórios. Na seção 3.5 é apresentada a descrição detalhada da etapa de divulgação com foco na publicidade das informações por meio de páginas na internet e intranet.

3.1.1. Integração da fonte de Dados

Foram utilizadas diferentes fontes de dados para prover uma maior diversidade informacional. Além da Plataforma Lattes, o *Google Scholar* e os Sistemas Institucionais foram integrados, tal como descrito a seguir.

⁶Disponível em: <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. Acesso em 06 de julho de 2021.

3.1.2. Plataforma Lattes

A Plataforma Lattes foi disponibilizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com a finalidade de unificar o cadastro nacional de pesquisadores. Em 2012, a Plataforma Lattes contava com mais de dois milhões de currículos cadastrados [Alves et al. 2012]. É a plataforma com maior e mais relevante repositório curricular de pesquisadores acadêmicos da Ciência Brasileira [Lane 2010].

Plataformas com as mesmas características não são encontradas em outras partes do mundo, o que se tem disponível são bases de dados referenciais, que permitem extração de fontes de informação secundárias. O Brasil é o único país que vincula obrigatoriamente as informações constantes em uma plataforma curricular com as solicitações de fomento aos órgãos competentes. Para tal, além da concessão de bolsas e auxílios à pesquisa, há também a avaliação da qualidade dos cursos de Pós-Graduação *Stricto Sensu*. A coleta de dados em larga escala da plataforma Lattes foi restringida pelo CNPq, permitindo somente esta funcionalidade a instituições e pesquisadores que solicitem autorização especial.

Em função de ser uma plataforma de inserção obrigatória para pesquisadores que necessitem participar de editais de fomento à pesquisa, possui forte potencial como fonte de informações unificadas viabilizando o desenvolvimento deste projeto, permitindo análises de dados e automatização de processos, uma vez que a instituição conquistou permissão do CNPq para coletar dados da Plataforma Lattes por meio de seu *web service* [Leite et al. 2011, Mugnaini et al. 2011]. Ressalta-se que a instituição precisou atender os seguintes requisitos: i) Ser Instituição de Ensino e Pesquisa; ii) Possuir um IP dedicado; e iii) Desenvolver solução tecnológica para integração com o *web service* do Lattes Extrator.

3.1.3. Google Scholar

O *Google Scholar* é uma ferramenta gratuita que possibilita a busca de trabalhos acadêmicos de diversas categorias, como artigos de congressos, teses e dissertações, bem como artigos de periódicos de acesso aberto ou pago, em vários idiomas, esses materiais estão disponíveis em repositórios na *web* ou em sites acadêmicos em que é possível verificar a quantidade de vezes que foram citados em outras publicações acadêmicas. [Caregnato 2011].

Essa plataforma tem possibilidade de ser utilizada como referência principal em índices de citações para estudos bibliométricos [Noruzi 2005], em mensuração do impacto das pesquisas [Vaughan and Shaw 2008] e periódicos de acesso aberto [Mukherjee 2009]. Seu processo de coleta de dados foi realizada utilizando um *software* que rastreia a *web* e é capaz de identificar automaticamente os campos que compõem os documentos científicos e suas respectivas referências. Com isso, surge uma ferramenta sem precedentes, que oferece gratuitamente dados bibliográficos e milhões de artigos na íntegra [Jacsó 2010].

Foi empregada a seção *Profiles*⁷ da plataforma *Google Scholar*, que lista pesquisadores vinculados a uma instituição específica, neste caso, pesquisadores com e-mails @ufopa.edu.br validados. Por meio dessa página, é possível visualizar os perfis dos pesquisadores, incluindo suas publicações, citações, índice h, colaborações e outras informações relevantes à pesquisa. Com isso, optou-se por utilizar dados provenientes do *Google Scholar* para obter índices de impacto e citações das pesquisas dos docentes da instituição. Permitindo assim um gerenciamento e visualização de maneira ágil e unificada das produções e pesquisadores mais relevantes na instituição. Para identificar os pesquisadores tanto no Lattes quanto no *Google Scholar*, utilizam-se os dados armazenados nos sistemas institucionais.

3.1.4. Sistemas Institucionais

Um dos diferenciais deste trabalho é a obtenção de uma solução que consiga integrar dados institucionais, objetivando automatizar a sincronização de dados e fazer jus aos recursos já investidos. Assim, alguns dados que não se pode obter pelas fontes anteriormente citadas devem ser obtidas por meio dos sistemas institucionais. A UFOPA utiliza um sistema acadêmico bem difundido pelas IFES brasileiras, no caso o sistema SIGAA da instituição UFRN. Segundo site da UFRN⁸, 73 instituições utilizam seus sistemas.

Os dados do SIGAA englobam informações sobre os docentes, servidores técnicos e discentes da instituição, suas unidades de lotação, inícios e termos de vínculo, facilitando assim a alimentação de maneira automática da ferramenta proposta. Além disso, o módulo de pesquisa do SIGAA permite o cadastro e gerenciamento de editais de pesquisa. Assim, a coleta de dados buscou os inscritos nestes editais para produzir resultados de pontuações. É por meio do Sistema Intellectus⁹, sistema da UFRN, que se obtém os dados dos currículos Lattes importados dos servidores do CNPq. Este sistema conecta à base de dados do SIGAA e busca os docentes da instituição para poder obter seus currículos.

3.2. Tecnologias

Primeiramente a escolha da linguagem de programação para o desenvolvimento do projeto foi Python¹⁰, baseou-se na escolha de outros projetos disponíveis, como o *ScriptLattes*, visando facilitar a integração entre as ferramentas e aprendizado [Mena-Chalco and Júnior 2013]. Além de ser uma linguagem bastante utilizada no meio científico, com muitas bibliotecas especializadas em análises de dados e aprendizado de máquina, o que expande o horizonte para trabalhos futuros. Uma dessas bibliotecas de análises de dados utilizada neste projeto para o processamento dos dados foi a biblioteca Pandas¹¹.

⁷Disponível em: https://scholar.google.com/citations?view_op=view_org&hl=en&org=2173319500607027007. Acesso em 29 de julho de 2021.

⁸Disponível em: <https://portalcooperacao.info.ufrn.br/pagina.php?a=parceiros>. Acesso em 29 de julho de 2021.

⁹Disponível em: <https://sistemas.ufrn.br/>. Acesso em 29 de julho de 2021.

¹⁰Disponível em: <https://www.python.org/>. Acesso em 29 de julho de 2021.

¹¹Disponível em: <https://pandas.pydata.org/>. Acesso em 29 de julho de 2021.

O artefato faz comunicação com o sistema gerenciador de banco de dados Postgres¹² para sincronização com dados institucionais do Sistema SIGAA. Mas para armazenamento de dados de sua base própria foi escolhido o SGBD MongoDB¹³, pois apresenta vantagens de escalabilidade, manipulação da quantidade massiva dos dados, flexibilidade e facilidade para consultas [Souza and de Oliveira 2019].

Para a entrega dos dados por meio de páginas *web*, o *framework* escolhido foi o FastAPI¹⁴. Essa escolha se definiu devido tecnologias atualizadas deste *framework*, oferecendo ferramentas que agilizam o desenvolvimento e melhoram a manutenção e desempenho, como documentação automática usando o Swagger¹⁵, funções assíncronas e tipagem de dados com retornos amigáveis para validações.

Consoante a tecnologia padrão do FastAPI, MongoDB e os formatos de dados do Intellectus, utilizou-se o formato JSON para armazenar e distribuir a base de dados. A tecnologia *JavaScript Object Notation* (JSON) é uma sintaxe, baseada em Javascript, para serialização de objetos, matrizes, números, *strings*, booleanos, e *NULL*¹⁶. Para facilitar a disponibilização da base desenvolvida pelo artefato, foi implementado utilizando o FastAPI, *web services* com rotas para acesso a projetos, artigos, pesquisadores entre outros, isto feito utilizando a especificação OpenAPI¹⁷. Juntamente com o Swagger permitem fácil visualização e testes para conexão do artefato com outros sistemas.

3.3. Implementação

Após a definição das tecnologias utilizadas, o passo seguinte foi o entendimento do negócio e das dificuldades já existentes. Para começar o desenvolvimento foi necessário fazer um mapeamento dos itens do currículo Lattes. A primeira coisa a ser feita foi a importação dos currículos no formato JSON e armazená-los no banco MongoDB. Observou-se uma grande dificuldade para o mapeamento e manutenção dos caminhos específicos de cada item na estrutura de arquivo. Então foi desenvolvida um algoritmo com buscas recursivas em um JSON, que agilizou e viabilizou o desenvolvimento da obtenção dos itens nos currículos.

Assim foi possível cadastrar vários modelos de pontuação, de forma genérica, permitindo buscas de itens por chave-valor e comparando-os com as condições especificadas no modelo, para então contabilizá-los. Os dados do modelo são cadastrados no banco e uma lógica de negócio implementa a validação do modelo com o currículo.

Nesta etapa uma grande dificuldade foi encontrada para obter uma base confiável de dados foi a duplicidade das produções, pois a base Lattes é uma base de currículos, ou seja, os registros são focados na individualidade de cada autor. Assim, registros de coautorias dentro da instituição são coletados uma vez para cada autor, gerando redundância.

¹²Disponível em: <https://www.postgresql.org/>. Acesso em 29 de julho de 2021.

¹³Disponível em: <https://docs.mongodb.com/manual/>. Acesso em 29 de julho de 2021.

¹⁴Disponível em: <https://fastapi.tiangolo.com/>. Acesso em 29 de julho de 2021.

¹⁵Disponível em: <https://swagger.io/>. Acesso em 29 de julho de 2021.

¹⁶Disponível em: https://developer.mozilla.org/pt-BR/docs/Web/JavaScript/Reference/Global_Objects/JSON. Acesso em 29 de julho de 2021.

¹⁷Disponível em: <https://github.com/OAI/OpenAPI-Specification>. Acesso em 29 de julho de 2021.

Então, desenvolveu-se um algoritmo para o tratamento de redundâncias que permite a detecção de produções acadêmicas iguais ou similares, conforme Algoritmo 1. Por uma questão de performance decidiu-se comparar apenas os primeiros trinta caracteres para depois comparar o texto completo.

Algoritmo 1: Algoritmo utilizado para identificar para identificar produções similares

Entrada: itemBuscado, listaDeProducoes, atributo
Saída: itemsSimilares

```

1 início
2   itemsSimilares = []
3   para cada indice, itemComparado em listaDeProducoes faça
4     se itemComparado['nomeAutorAscii'] == itemBuscado['nomeAutorAscii'] então
5       |   continuar
6     fim
7     se itemComparado não contém atributo então
8       |   continuar
9     fim
10    se ratio(itemComparado[atributo][:30],itemBuscado[atributo][:30]) > 0.8 então
11      |   se ratio(itemComparado[atributo],itemBuscado[atributo]) > 0.90 então
12        |   adiciona itemComparado a itemsSimilares
13      fim
14    fim
15  fim
16  retorna itemsSimilares
17 fim

```

Outro problema encontrado foi em relação a não completude de informações nos currículos, uma vez que nem todos os campos do Lattes são obrigatórios e não se tem um identificador obrigatório para cada item, o que dificulta a fusão dos dados entre currículos. Essas informações registradas não mantêm um padrão e, em muitos casos, estão incompletas ou em campos trocados. Por exemplo registros com nomes de autores abreviados, ausência do identificador, títulos com atributos de formatação e ausência de acentos. Assim, a obtenção de um base unificada representa um problema não-trivial para uma melhor análise das produções.

Nesse contexto, optou-se pela comparação dos nomes das produções, porém devido a inconsistências, como erros de digitação ou falta de padronização na escrita de cada currículo, não é eficaz comparar textos diretamente. Para isso, a solução adotada foi a busca de proximidade de cadeias de caracteres. Para a definição de similaridade entre registros foi escolhido o algoritmo de distância de *Levenshtein*. Esse método compara duas cadeias de caracteres e retorna um valor correspondente à quantidade de alterações para um texto se transformar em outro, como inserções, eliminações ou substituições de caracteres [Mena-Chalco and Júnior 2013].

Assim, dois itens foram considerados iguais se possuísem um coeficiente de similaridade superior a 90%, diferentemente dos 80% utilizados por [Mena-Chalco and Júnior 2013]. Esta decisão foi baseada em uma avaliação qualitativa com servidores da DPE. A Tabela 2 foi construída com base nos currículos coletados, e

possibilita observar a quantidade de registros encontrados por tipo de produção e a quantidade obtida após fusão de dados, juntamente com o percentual de itens similares obtido. Por fim, este procedimento executa comparações dois a dois entre todas as produções de um conjunto de dados segmentado por tipo de produção.

Tabela 2. Quantitativo pós-fusão de dados

Tipo de produção	Registros	Únicos	Similares (%)
Artigos Completos	5.223	4.554	12,80
Trabalhos em eventos	13.225	12.214	7,64
Livros	455	399	12,30
Capítulos de livro	1.464	1.337	8,67
Patentes	29	25	13,79
<i>Softwares</i>	86	84	2,32
Projetos	4.467	3.770	15,60
Projetos de Pesquisa	3.005	2.608	13,21
Projetos de Extensão	1.061	820	22,71
Projetos de Desenvolvimento	93	73	21,50
Projetos de Ensino	62	57	8,06

Fonte: Elaborado pelos autores.

Para melhorar os resultados foram aplicadas técnicas para normalizar textos antes de compará-los. Os procedimentos adotados para o pré-processamento foram: (i) conversão para caixa baixa; (ii) remoção de espaços em branco extras; (iii) remoção de *tags* Linguagem de Marcação de Hipertexto, do inglês, *HyperText Markup Language* (HTML); (iv) remoção de elementos de formatação; e (v) remoção de caracteres não padronizados de acordo com o *American Standard Code for Information Interchange* (ASCII).

Um dos benefícios do uso desta estratégia é a geração de uma base de dados mais completa após fusão de registros de produções similares (vinda de currículos diferentes), pois as informações de registros similares são mescladas, assim informações existentes em um registro complementam as que faltavam em outros. Para casos em que ocorreram divergências de informações foi implementada e aplicada uma função, conforme Algoritmo 2 para identificar o autor principal ou responsável pela produção, e então utiliza-se o Algoritmo 3 para mesclar os dados dos registros dando prioridade para o registro do autor principal, caso seja docente da instituição, caso contrário utiliza-se como referência o primeiro registro de um autor da instituição.

Outro ponto, relevante foi a fusão de dados junto à base *Google Scholar*. Para isso utilizou-se a biblioteca *scholarly*¹⁸, escrita em Python, para a coleta de dados. Para efetuar a fusão dos dados, uma vez que o identificador *Google* do pesquisador não é encontrado no currículo Lattes, utilizou-se o nome completo do pesquisador e nome em citações, juntamente com uma validação do domínio de e-mail igual a “ufopa.edu.br”. É importante ressaltar que a validação foi feita somente com o e-mail de domínio da ufopa,

¹⁸Disponível em: <https://pypi.org/project/scholarly/>. Acesso em 29 de julho de 2021.

Algoritmo 2: ALGORITMO UTILIZADO PARA IDENTIFICAR RESPONSÁVEL PELA PRODUÇÃO.

Entrada: dados
Saída: Autor principal da produção

```

1 início
2   autorPrincipal = vazio
3   para cada dado em dados faça
4     se dado contém 'participantes' então
5       para cada participante em dado faça
6         se participante contém 'ordemdeautoria' então
7           se participante['ordemdeautoria'] == 1 então
8             autorPrincipal = participante
9             break
10          fim
11         fim
12       senão
13         autorPrincipal = participante
14         break
15       fim
16     fim
17   fim
18 fim
19 retorna autorPrincipal
20 fim
  
```

Algoritmo 3: Algoritmo utilizado para fusão de dados de registros similares.

Entrada: producoesSimilares, curriculos
Saída: dados de produção mesclado

```

1 início
2   se existem 'participantes' em producoesSimilares então
3     registroMesclado = vazio responsavel = obterAutorPrincipal(dados)
4     indiceRegistroResponsavel = vazio se responsavel não é vazio então
5       identificar o indice do responsavel na lista producoesSimilares se
6         indiceRegistroResponsavel não é vazio então
7           registroMesclado = producoesSimilares[indiceRegistroResponsavel]
8           remove registro de producoesSimilares
9         fim
10      fim
11    fim
12  para cada dado em dados faça
13    mesclar objetos dado e registroMesclado
14  fim
15 retorna registroMesclado
16 fim
  
```

sendo assim, caso o e-mail principal não fosse @ufopa.edu.br, o pesquisador não seria considerado na fusão dos dados.

Por fim, foi criada uma arquitetura baseada em OpenAPI para que a base compilada pudesse ser acessada como serviço *web*, garantindo transparência e flexibilidade de dados já processados pelo sistema. Além disso, é possível estender novas exibições por meio de *iframe*¹⁹, uma funcionalidade adicionada que permite que *dashboards* e relatórios possam ser gerados externamente, como por exemplo em Microsoft PowerBI, Apache Superset e Metabase, e disponibilizadas por páginas internas no sistema centralizando os dados em um único portal.

A construção dos *dashboards* foi baseada em boas práticas para mostrar conhecimento por meio de dados, fornecendo visualizações eficientes para a tomada de decisões, conforme orienta [Knafllic 2019]. Algumas destas orientações foram aplicadas para entregar conhecimento utilizando de infográficos simples e objetivos, evitando problemas comuns que dificultam a transmissão da mensagem. Por exemplo, evitar gráficos em formato de pizza quando existem muitos dados ou pouca diferença entre os valores, evitar gráficos 3D caso a terceira dimensão não apresente uma nova informação e eliminar elementos apenas visuais que não colaboram com a passagem da mensagem. Foram adicionadas também funções de exportações de dados para cada gráfico, objetivando manter a integridade das informações, assim é possível verificar os dados reais sem distorções visuais que alguns gráficos podem gerar. E para melhorar a visualização dos dados *dashboards* também foram implementados para exibição em dispositivos móveis.

3.4. Demonstração

A proposta de solução foi apresentada para a DPE. Nesta etapa ocorreu a implantação do sistema para testes e correção de eventuais falhas na implementação. Desta forma pôde-se demonstrar a validade do artefato. Uma versão do protótipo está disponibilizada para utilização da PROPPIT. Nesta versão, o processo de coleta de dados está executando de forma automática e disponibilizado por páginas públicas e internas, permitindo a visualização das produções via *dashboards*, conforme Figura 3.

A Figura 3 apresenta a contabilização de produções por ano e por Qualis, identificando os autores mais ativos e as áreas de conhecimento Capes mais pesquisadas. É possível identificar os institutos que mais geram produções, palavras-chave com base nos títulos e descrições das produções. Além de fornecer estatísticas como média, desvio padrão, mediana, total e quantidade no último ano por tipo de produção.

Além de *dashboards* quantitativos, é possível obter o conhecimento detalhando das produções por meio de páginas descritivas, permitindo vários filtros, ordenamentos e funções de exportação destes dados, conforme a Figura 4.

Para facilitar a procura de pesquisadores, foi criado um *dashboard* (Figura 5) com o perfil individual. Nele é possível identificar pesquisadores por área de conhecimento, áreas de interesses, palavras-chave de suas produções, ordenamento por índices de produção, seus coautores, estatísticas de produtividades, entre outros.

¹⁹Um elemento HTML que incorpora outra página HTML na página atual. fonte: <https://developer.mozilla.org/pt-BR/docs/Web/HTML/Element/iframe>

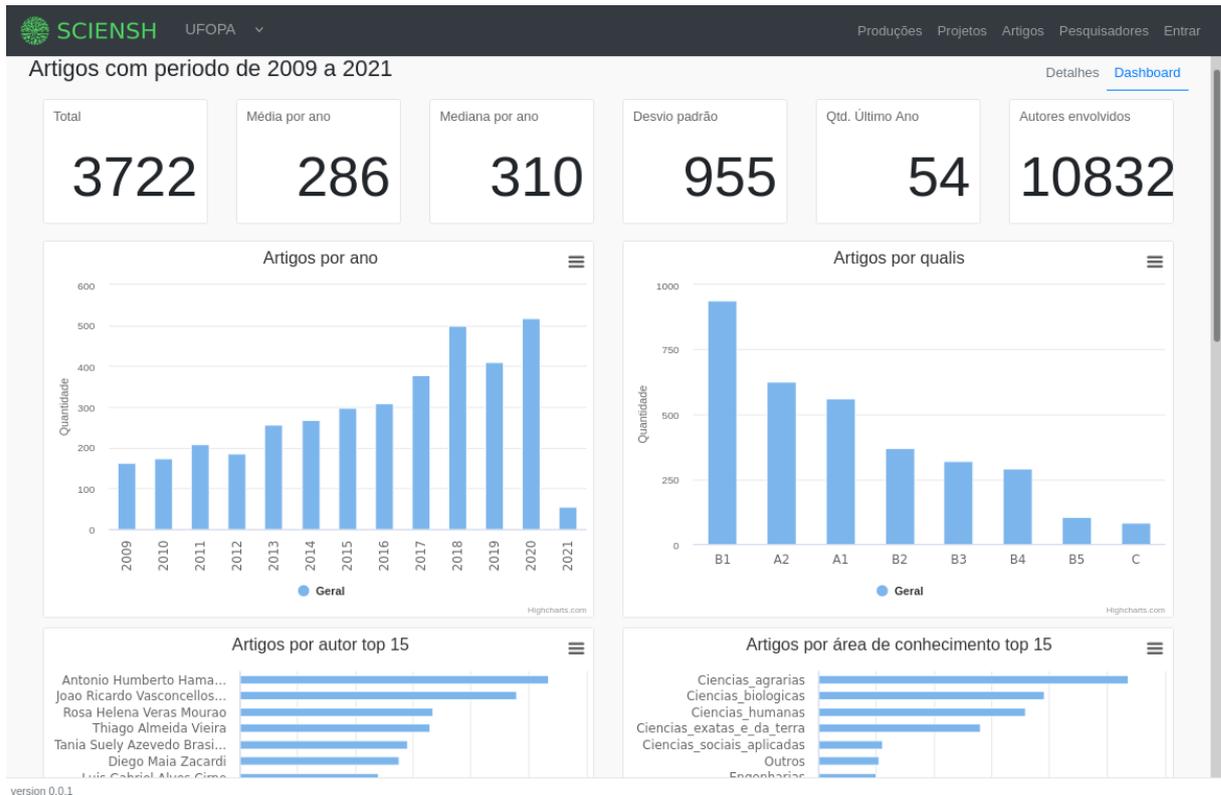


Figura 3. Dashboard com perfil das produções da instituição.

Fonte: Elaborado pelo autor.

Aborda-se também a quantificação e estratificação de vários tipos de produção, artigos, trabalhos em eventos, prêmios, orientações, registros de patentes, registros de *softwares* e projetos. Essas quantificações foram segmentadas por pesquisador gerando índices individuais e facilitando ordenações. Por exemplo, quantitativo de artigos completos, de livros, capítulos, trabalhos em eventos, índice-H e citações, por exemplo.

Necessidades de transparência foram solucionadas lançando mão de páginas públicas, como a página da instituição, mostrada na Figura 6. Este requisito permite ver o perfil da instituição filtrando por palavras-chave, nomes pesquisadores, títulos de projetos, artigos, pesquisadores mais citados, maiores financiadores, entre outros. Visando ampliar alcance e visibilidade das produções da instituição, são mostrados *links* de cada artigo completo para buscas dos arquivos no *Google Scholar* e *links* diretos dos perfis dos pesquisadores na Plataforma Lattes.

Para avaliar a produtividade da instituição foi desenvolvido um gráfico segmentado por ano e tipo de produção, capaz de mostrar o quanto foi produzido de cada tipo de produção ao longo do tempo. Com a utilização da Plataforma Lattes como base de dados para mensurar a produtividade dos pesquisadores da instituição permitiu-se descentralizar o trabalho e garantir a validade das informações, uma vez que os próprios pesquisadores são os responsáveis pelo cadastro. Isso permitiu fornecer anuário e relatórios dinâmicos de forma automatizada.

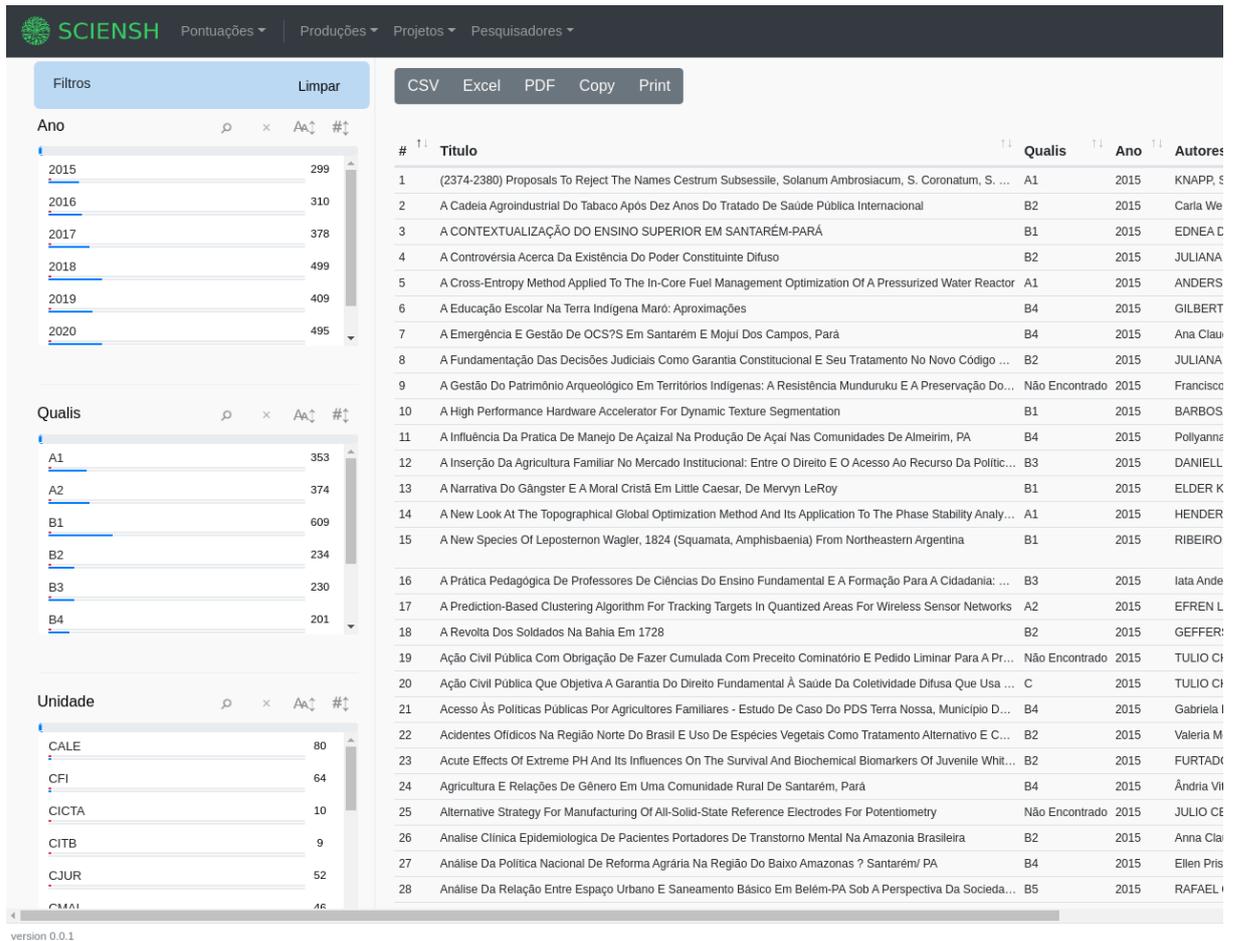


Figura 4. Filtros e exibições descritivas

Fonte: Elaborado pelos autores.

3.5. Avaliação e Comunicação

A etapa de avaliação consistiu na resposta dos *stakeholders* sobre as soluções propostas e demonstradas, com o intuito de determinar se o artefato supera os desafios apresentados. Para tal, foi necessário estabelecer uma forma de avaliação e coleta de informações para, se possível, adaptações da solução oferecida. Assim, a condução desta etapa se executou em uma abordagem quantitativa-qualitativa.

Durante a apresentação do sistema, perante ao desafio que consiste na necessidade de visualizar e publicizar as produções da instituição. Foram coletados os *feedbacks* dos *stakeholders* as soluções do sistema foram aprovadas e alguns novos requisitos foram solicitados para tornar a solução mais completa.

Dentre os requisitos solicitados estão: acesso diferenciado para perfis de pesquisador, gestor, administrador; permissão para simulação da própria pontuação conforme modelos de pontuações utilizados na instituição; *dashboards* de livros; *dashboards* de prêmios; *dashboards* de artigos publicados em conferências; padronização da nomenclatura das áreas de conhecimento; vitrine virtual de produtos contendo os interesses nego-

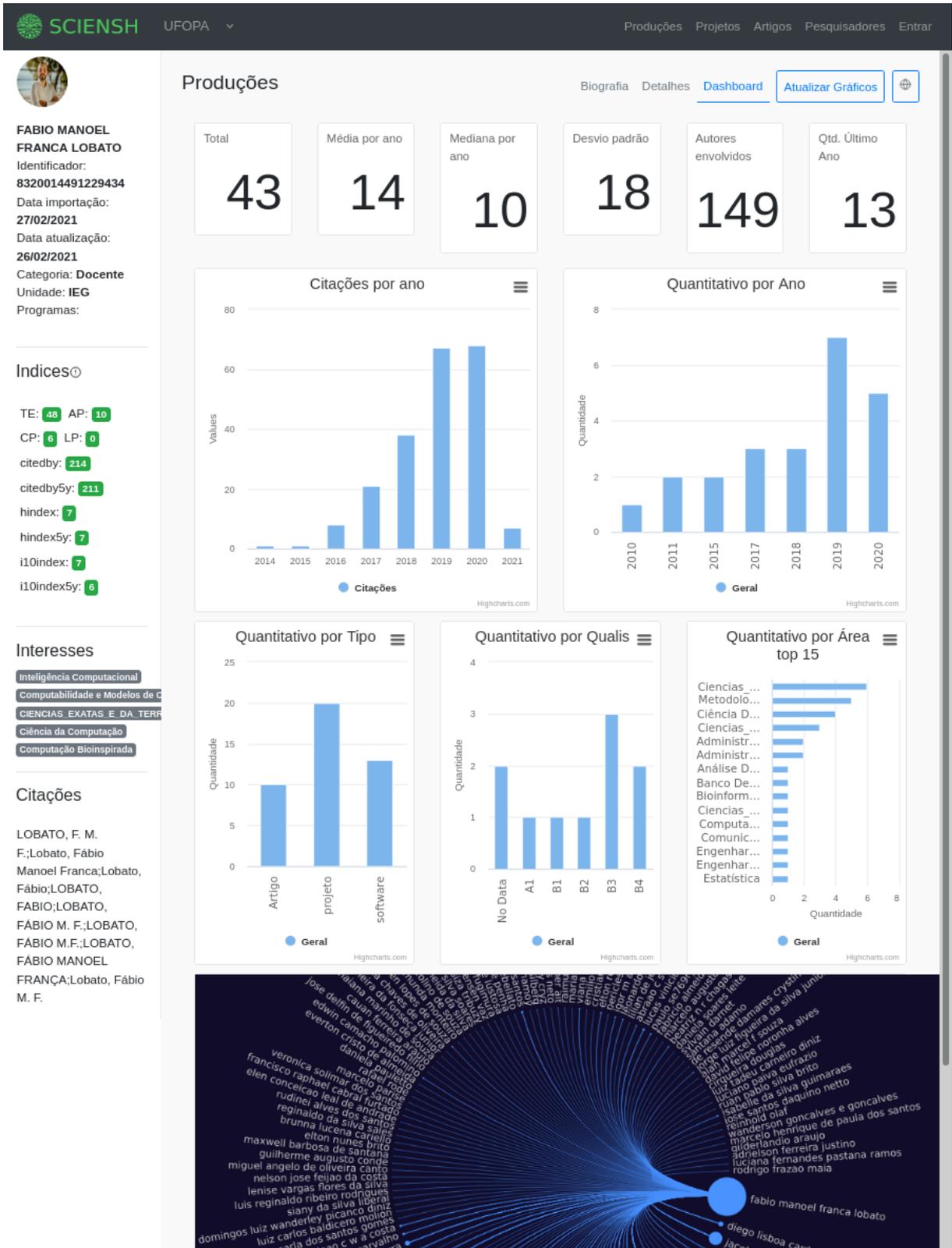


Figura 5. Perfil do pesquisador.
 Fonte: Elaborado pelos autores.




518
Pesquisadores


2769
Projetos


66
Patentes e Registros


3722
Artigos

Venha nos conhecer

Através de dados coletados automaticamente da Plataforma Lattes mapeamos nosso perfil como instituição fomentadora de conhecimento

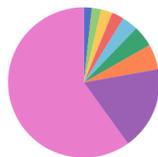
Artigos Pesquisadores Projetos Financiadores

<p>PEDRO HENRIQUE SALOMÃO GANANÇA A2</p> <p>a case report of human intoxication due to a snakebite by the episthegylhous dipsadid thamnodynastes lanei bailey, thomas & silva-jr, 2005</p>	<p>CARLILE LANZIERI JUNIOR B2</p> <p>a idade média imaginada pelos adeptos da retórica do choque de civilizações em diferentes momentos dos séculos xx e xxi</p>
<p>SOARES, LUCAS DE VASCONCELOS None</p> <p>a organização do trabalho pedagógico no oeste do pará: discussões no contexto pandêmico</p>	<p>DIEGO MAIA ZACARDI B3</p> <p>a pesca artesanal na região dos lagos: o caso do município de pracuúba, amapá</p>
<p>JOÃO PAULO SOARES DE CORTES None</p> <p>a statistical-based reach scale classification for the lower tapajós river channel, eastern amazonia</p>	<p>MEDEIROS, HELOISE MICHELLE NUNES A2</p> <p>alternative tourism and environmental impacts: perception of residents of an extractive reserve in the brazilian amazonia</p>
<p>GRACIENE DO SOCORRO TAVEIRA FERNANDES None</p> <p>antibiotic resistance in the superficial mucosa microbiota in an amazonian fish, mapará (hypophthalmus spp.) / resistência a antibióticos na microbiota do muco superficial em peixe amazônico, mapará (hypophthalmus spp.)</p> <p>MICROBIOLOGIA AMBIENTAL MICROBIOLOGIA CIÊNCIAS BIOLÓGICAS MICROBIOLOGIA APLICADA</p>	<p>BARROS, ISABELLA OLIVEIRA B1</p> <p>assessment of donkey (equus asinus africanus) whole blood stored in cpda-1 and cpd/ag-m blood bags</p>

[Ver mais](#)

Acompanhe nossa produção ao longo dos anos

Desde 2009 estamos investindo em nossas pesquisas, pesquisadores e infraestrutura pra garantir maior produtividade e qualidade nos conhecimentos produzidos em nossa região.



Quem financia nossa produção de conhecimento?

O financiamento da pesquisa no Brasil se dá por meio de diferentes sistemas e instituições de fomento, que estão ligadas direta ou indiretamente aos ministérios brasileiros. Há também leis de incentivo fiscal e fomento à inovação, financiamentos empresariais e institucionais.

Quer conhecer a UFOPA?

Por meio de análises da plataforma Lattes produzimos o perfil da nossa instituição

Figura 6. Página pública da instituição.

Fonte: Elaborado pelos autores.

ciais da instituição e os serviços que a instituição oferece, tais como transferência de tecnologia, patentes e softwares; mecanismo de troca de mensagens para captar interessados em financiar, desenvolver ou apoiar projetos com a instituição. A etapa de comunicação consistiu no registro do artefato (*software*) construído junto ao Instituto Nacional de Propriedade Industrial.

Após a etapa de desenvolvimento dos novos requisitos identificados foi aplicado um questionário de satisfação na escala *Likert*, que consiste numa escala para mensuração da concordância com as afirmações propostas [Filardi and Traina 2008].

O formulário aplicado foi desenvolvido utilizando a escala *Likert* de cinco pontos invertida, conforme análises de [Dalmoro and Vieira 2013], pois, em média, apresenta maior precisão, facilidade e agilidade em ser respondida. Este formulário contém 10 afirmações, a saber: 1) Calcular pontuações personalizadas de indicadores cientométricos; 2) Permitir maior agilidade na gestão da pesquisa e inovação na instituição; 3) Identificar os perfis e tendências de pesquisa da instituição; 4) Gerar relatórios dinâmicos das produções dos pesquisadores e suas relações de cooperação; 5) Prover informações e auxiliar no planejamento e a avaliação de políticas científicas; 6) Disponibilizar dados das atividades de pesquisa à comunidade; 7) Permitir a difusão das produções da instituição, principalmente para a sociedade; 8) Permitir maior eficiência no processo de análise da produtividade dos pesquisadores; 9) Identificar a conexão de pesquisadores na instituição e 10) Mensurar itens relevantes para a Gestão e a comunidade.

O resultado de satisfação dos *stakeholders* em relação as premissas constantes no questionário foi positivo, concordando plenamente e parcialmente com os requisitos levantados nos estudos. Ressalta-se que, no enfoque de TIC como ferramenta, há uma clara separação entre o artefato e o contexto de uso, os quais são compreendidos como entidades discretas que podem ser separadas e estudadas de forma independente [Araujo et al. 2017].

3.6. Impactos do estudo

Após a etapa de demonstração e avaliação o sistema passou a ser utilizado por usuários reais a título de gestores temos reitor da instituição, pró-reitores, diretor de pesquisa além da equipe operacional da DPE. Na qual, o sistema desenvolvido serviu como auxílio em reuniões de planejamento com os programas de pós-graduação, utilizada para mapa de pesquisa e identificação de pesquisadores em temas específicos. Ressalta-se também, que a ferramenta teve impacto direto a todos os professores que submetem aos editais de pesquisa da instituição, pois automatiza o processo do cálculo das pontuações que anteriormente eram calculadas de forma manual e individualmente por cada docente.

O desenvolvimento deste tipo de artefato pode alterar processos na gestão institucional. Sumarizamos alguns dos impactos do estudo a seguir:

- Resignificar a área e provocar modificações profundas na forma como são conduzidos processos internos de gestão;
- Possibilidade de escala para outras instituições de ensino, pesquisa e desenvolvimento e inovação - como centros de pesquisa, agências de fomento e fóruns de pró-reitores;

- Possibilita a avaliação e planejamento estratégico de grupos de pesquisa e programas de pós-graduação;
- Por considerar também a produção técnica, é possível construir uma vitrine de produtos e facilitar a inserção de pesquisadores no ambiente produtivo.

4. Conclusões

Visando resolver problemas relacionados à organização dos dados de pesquisa da instituição e transparência das produções técnico-científicas, este estudo apresentou o processo de projeto e construção do *SCIENSH: Sistema de apoio à gestão da pesquisa e inovação baseado em cientometria* foi apresentado. O artefato foi desenvolvido e resultou no Registro de Programa de Computador junto ao INPI, caracterizando-o como produção técnica. E também, a biblioteca para buscas recursivas em arquivos JSON, formato que permite explorar profundamente os currículos cadastrados na plataforma Lattes.

O Sciensh usa diversas fontes de dados, tais como a plataforma Lattes, o *Google Scholar*, o SIGAA e o Sucupira, permitindo assim, a fusão de dados e maior completude de dados cientométricos a serem disponibilizados aos gestores institucionais. As principais funcionalidades do Sciensh incluem a geração de *dashboards* e indicadores institucionais, disponibilização de metadados e também a customização de indicadores a fim de se adequarem à realidade institucional. Além disso, utilizam-se técnicas de análises baseadas em redes e em mineração de texto para prover informações com maior acuidade e de fácil entendimento, possibilitando a tomada de decisões baseadas em dados. Destaca-se ainda que o Sciensh provê três níveis de acesso, um portal público, outro para membros internos da instituição e um específico para os gestores.

Dos impactos esperados do projeto, vale mencionar a melhora na eficiência da gestão por meio do processo de mapeamento das áreas e tendências de pesquisa, integração com dados institucionais, acompanhamento da produtividade e possibilidade de avaliação e direcionamento das pesquisas e políticas públicas da instituição. Com o término desse projeto, será criada uma base de dados única das produções acadêmicas da instituição que visa divulgar o atual estado da produção acadêmica para auxiliar em tomadas de decisões. Vale ressaltar ainda que o enfoque do projeto até o momento consistiu no desenvolvimento do artefato e para as próximas etapas almeja-se a construção de um modelo de negócios que possibilite a transferência de tecnologia para outras instituições e monetização do produto desenvolvido.

Algumas ameaças à validade do estudo também precisam ser mencionadas. Por exemplo, embora o Sciensh automatize e forneça relatórios com fusão de dados das produções, os resultados referem-se aos dados cadastrados na Plataforma Lattes, sendo recomendada uma verificação cuidadosa nos relatórios automaticamente gerados para que possíveis erros de preenchimento ou ausências de dados sejam corrigidos. Ressalta-se, que os dados quantitativos produzidos por ferramentas similares a este protótipo, precisam ser analisado com cuidado, levando em consideração contextos sociais, econômicos e regionais para não haver dissociação com a realidade e por fim contribuir de forma produtiva com a gestão do conhecimento.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)-DT-308334/2020; e pela Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) PRONEM-FAPESPA/CNPq nº 045/2021. Estendemos nossos agradecimentos aos revisores, cujas contribuições foram essenciais para melhorar a qualidade do trabalho.

Referências

- [Alves et al. 2012] Alves, A. D., Yanasse, H. H., and Soma, N. Y. (2012). Lattesminer: uma linguagem de domínio específico para extração automática de informações da plataforma lattes. *XII Workshop de Computação Aplicada - WORCAP*.
- [Araujo et al. 2017] Araujo, R., Fornazin, M., and Pimentel, M. (2017). Uma análise sobre a produção de conhecimento científico nas pesquisas publicadas nos primeiros 10 anos da isys (2008-2017). *iSys-Brazilian Journal of Information Systems*, 10(4):45–65.
- [Balancieri et al. 2005] Balancieri, R., Bovo, A. B., Kern, V. M., Pacheco, R. C. d. S., and Barcia, R. M. (2005). A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na plataforma lattes. *Ciência da informação*, 34(1).
- [Brito et al. 2016] Brito, A. G. C. d., Quoniam, L., and Mena-Chalco, J. P. (2016). Exploração da plataforma lattes por assunto: proposta de metodologia. *Transinformação*, 28:77–86.
- [Caldas et al. 2018] Caldas, L. V., Jacob, A. F., Silva, S. S., Pontes, F. A., and Lobato, F. M. (2018). Development of a social network for research support and individual well-being improvement. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 383–386. IEEE.
- [Caregnato 2011] Caregnato, S. E. (2011). Google acadêmico como ferramenta para os estudos de citações: avaliação da precisão das buscas por autor. *Pontodeacesso*, 5(3):72–86.
- [Carpenter et al. 2014] Carpenter, C. R., Cone, D. C., and Sarli, C. C. (2014). Using publication metrics to highlight academic productivity and research impact. *Academic emergency medicine*, 21(10):1160–1172.
- [Cesso and Ferraz 2017] Cesso, M. V. and Ferraz, R. R. N. (2017). Uso de ferramentas computacionais para gestão estratégica da produtividade em pesquisa de um programa de pós-graduação stricto sensu: Uma revisão sistemática da literatura. *Proceedings of the 14th CONTECSI International Conference on Information Systems and Technology Management*, 14.
- [Dal Molin and da Costa 2018] Dal Molin, T. C. and da Costa, F. A. M. (2018). A tecnologia da informação aplicada à gestão pública. *Revista de contabilidade Dom Alberto*, 7(13).
- [Dalmoro and Vieira 2013] Dalmoro, M. and Vieira, K. M. (2013). Dilemas na construção de escalas tipo likert: o número de itens e a disposição influenciam nos resultados? *Revista gestão organizacional*, 6(3).

- [Dias et al. 2018] Dias, T. M., Moreira, T. H., and Dias, P. M. (2018). Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do cnpq. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. SBC.
- [Dias and Moita 2018] Dias, T. M. R. and Moita, G. F. (2018). Um retrato da produção científica brasileira baseado em dados da plataforma lattes. *Brazilian Journal of Information Science: research trends*, 12(4):62–74.
- [Digiampietri et al. 2012] Digiampietri, L. A., Mena-Chalco, J. P., Pérez-Alcázar, J. J., Tuesta, E. F., Delgado, K. V., Mugnaini, R., and Silva, G. S. (2012). Minerando e caracterizando dados de currículos lattes. In *Anais do I Brazilian Workshop on Social Network Analysis and Mining*. SBC.
- [Dresch et al. 2015] Dresch, A., Lacerda, D. P., and Júnior, J. A. V. A. (2015). *Design science research: método de pesquisa para avanço da ciência e tecnologia*. Bookman Editora.
- [Fernandes et al. 2016] Fernandes, E., Silva, M. A., and Cagnin, M. I. (2016). Sigs-s: a web application and a mobile application for social and health care data management. *iSys-Brazilian Journal of Information Systems*, 9(1):81–100.
- [Filardi and Traina 2008] Filardi, A. L. and Traina, A. J. M. (2008). Montando questionários para medir a satisfação do usuário: avaliação de interface de um sistema que utiliza técnica de recuperação de imagens por conteúdo. In *Embrapa Territorial-Artigo em anais de congresso (ALICE)*. In: Simpósio brasileiro de fatores de humanos em sistemas computacionais, 8.
- [Gil 2008] Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.
- [Hevner et al. 2010] Hevner, A., Chatterjee, S., and Iivari, J. (2010). Twelve theses on design science research in information systems. *Design research in information systems: Theory and practice*, pages 43–62.
- [Jacsó 2010] Jacsó, P. (2010). Metadata mega mess in google scholar. *Online Information Review*.
- [Justino et al. 2023] Justino, A., Nhoatto, F., Souza, E., and Lobato, F. (2023). Análise orientada a dados como auxílio para tomada de decisão em gestão de pesquisa. *Revista da CGU*, 15(27).
- [Knafllic 2019] Knafllic, C. N. (2019). *Storytelling com Dados: Um guia sobre visualização de dados para profissionais de negócios*. Alta Books.
- [Lane 2010] Lane, J. (2010). Let's make science metrics more scientific. *Nature*, 464(7288):488–489.
- [Leite et al. 2011] Leite, P., Mugnaini, R., and Leta, J. (2011). A new indicator for international visibility: exploring brazilian scientific community. *Scientometrics*, 88(1):311–319.

- [Lopes 2012] Lopes, G. R. (2012). *Avaliação e recomendação de colaborações em redes sociais acadêmicas*. PhD thesis, Universidade Federal do Rio Grande do Sul.
- [Meglioransi 2019] Meglioransi, M. D. (2019). Um sistema para coleta e apresentação de indicadores bibliométricos da plataforma lattes. *Universidade Federal do Pampa*.
- [Mena-Chalco and Júnior 2013] Mena-Chalco, J. P. and Júnior, C. (2013). Prospecção de dados acadêmicos de currículos lattes através de scriptlattes. *Bibliometria e Cientometria: reflexões teóricas e interfaces*. São Carlos: Pedro & João, pages 109–128.
- [Mingers and Leydesdorff 2015] Mingers, J. and Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246(1):1–19.
- [Moher et al. 2018] Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., and Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS biology*, 16(3):e2004089.
- [Mugnaini et al. 2011] Mugnaini, R., Leite, P., and Leta, J. (2011). Fontes de informação para análise de internacionalização da produção científica brasileira. *PontodeAcesso*, 5(3):87–102.
- [Mukherjee 2009] Mukherjee, B. (2009). Do open-access journals in library and information science have any scholarly impact? a bibliometric study of selected open-access journals using google scholar. *Journal of the American Society for Information Science and Technology*, 60(3):581–594.
- [Nhoatto et al. 2021] Nhoatto, F., Justino, A., de Souza, E. L., and Lobato, F. M. F. (2021). Prospecção de ferramentas para gestão de pesquisa e desenvolvimento em instituições de ciência e tecnologia. *Revista E-Tech: Tecnologias para Competitividade Industrial-ISSN-1983-1838*, 15(1).
- [Noruzi 2005] Noruzi, A. (2005). Google scholar: The new generation of citation indexes. *Walter de Gruyter GmbH & Co. KG*.
- [Parra et al. 2019] Parra, M. R., Coutinho, R. X., and Pessano, E. F. C. (2019). Um breve olhar sobre a cienciometria: origem, evolução, tendências e sua contribuição para o ensino de ciências. *Revista Contexto & Educação*, 34(107):126–141.
- [Paxiúba and Lima 2019] Paxiúba, C. M. C. and Lima, C. P. (2019). *Um Modelo Conceitual para Trabalhar Emoções e Aprendizagem Utilizando Expressões Faciais*. PhD thesis, Universidade Federal do Oeste do Pará.
- [Peppers et al. 2007] Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.
- [Pressman and Maxim 2016] Pressman, R. and Maxim, B. (2016). *Engenharia de Software-8ª Edição*. McGraw Hill Brasil.
- [Santos and Kobashi 2009] Santos, R. N. M. d. and Kobashi, N. Y. (2009). Bibliometria, cientometria, infometria: conceitos e aplicações. *Tendências da Pesquisa Brasileira em Ciência da Informação*.

- [Silva and Bianchi 2001] Silva, J. A. d. and Bianchi, M. d. L. P. (2001). Cientometria: a métrica da ciência. *Paidéia (Ribeirão Preto)*, 11(21):5–10.
- [Silveira and Nunes 2013] Silveira, D. S. and Nunes, M. A. S. N. (2013). Avaliação de usabilidade de interfaces computacionais: uma prospecção tecnológica. *Revista GEINTEC-Gestão, Inovação e Tecnologias*, 3(3):111–119.
- [Souza and de Oliveira 2019] Souza, E. C. and de Oliveira, M. R. (2019). Comparativo entre os bancos de dados mysql e mongodb: quando o mongodb é indicado para o desenvolvimento de uma aplicação. *Revista Interface Tecnológica*, 16(2):38–48.
- [Vaughan and Shaw 2008] Vaughan, L. and Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74(2):317–330.
- [Zeng et al. 2017] Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., and Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, 714:1–73.