

AnyLattes: An Application for Continuous Assessment of Lattes Curriculum Information

Alex Carlos Saraiva Cirilo
Universidade Federal do Pará
Belém, Pará, Brasil
c.alex.cirilo@gmail.com

Isadora Mendes dos Santos
Universidade Federal do Pará
Belém, Pará, Brasil
isadoramsantos@gmail.com

Marcelle Pereira Mota
Universidade Federal do Pará
Belém, Pará, Brasil
mpmota@ufpa.br

Abstract

Context: With the advancement of technology, process automation becomes increasingly necessary, especially for data analysis and organization of information for decision making. **Problem:** The Process of obtaining scores from conferences and journals within publications in Lattes Curriculum demands a lot of time and manual work, in addition to the need for analysis to be done jointly at universities due to the way in which capes evaluates postgraduate programs. **Solution:** AnyLattes – An application for automation, data extraction and organization of information from the Lattes Curriculum to assist in the evaluation of postgraduate programs, in addition to serving as a basis for filling out the Sucupira platform.

SI Theory: This work is influenced by Organizational Information Processing Theory, where organizations need quality information to deal with environmental uncertainty and improve their decision making. **Method:** The research has a prescriptive nature along with the evaluation carried out through proof of concept, in addition to the qualitative approach for analyzing the results. **Summary of results:** In the experiments carried out, AnyLattes showed optimization in the data extraction routine and organization and grouping of information for analysis of postgraduate programs in comparison to commonly used means. **Contributions and Impact in the IS area:** The contribution is to enable an application focused on optimizing, organizing and grouping information extracted from Lattes Curriculum for postgraduate programs, have a view of the entire program and proceed with consistent data in their decision making.

CCS Concepts

- Information systems → Information retrieval; Information systems applications.

Keywords

Lattes curriculums, automation software, extraction, analysis

1 Introdução

A CAPES possui o papel de expandir e consolidar a pós-graduação *stricto sensu* em todos os estados. Dentre as suas competências estão: propagar o acesso e divulgação da produção científica e avaliar o desempenho dos cursos de mestrado e doutorado nacionais a fim de buscar um padrão de excelência acadêmica [8]. Para a avaliação, utiliza-se a Plataforma Sucupira, um sistema de coleta de informações dos programas de pós-graduação em tempo real e estabelecimento de procedimentos de avaliação com transparência para toda comunidade acadêmica [9]. A Plataforma possui um módulo que deve ser preenchido pelos coordenadores de Pós-Graduação, e servindo para o planejamento dos programas.

Diversos programas de Pós-Graduação (PPG) realizam a avaliação de desempenho dos docentes de forma anual ou quadrienal, utilizando a Plataforma Lattes para consultar os currículos dos professores, verificar as publicações válidas e registrar tais dados em planilhas, a fim de calcular as notas e efetuar os cálculos. No entanto, para que a CAPES possa realizar a avaliação, é necessário que essas informações sejam consolidadas de forma integrada, o que implica em um processo demorado e trabalhoso, envolvendo grande esforço manual.

Tal problema se torna importante pela quantidade de dados a serem analisados e calculados. Sem uma forma de automatizar esse processo, a coordenação não consegue analisar todos os dados necessários para garantir o melhor desempenho do programa. A falta de uma visão integral do programa ocorre devido à ausência de agrupamento dessas informações, o que impede a tomada das melhores decisões e o correto lançamento dos dados na Plataforma Sucupira. Isso, por sua vez, dificulta a melhoria do conceito do programa no cenário nacional, conforme a avaliação da CAPES.

Dante do exposto, o objetivo deste trabalho é automatizar a extração de dados, fazer a classificação dos artigos em periódicos e conferências, e realizar o cálculo de notas, possibilitando uma avaliação de forma contínua, através de uma aplicação web denominada AnyLattes. A ferramenta AnyLattes foi criada aplicando o método pesquisa-ação, onde o envolvimento do corpo docente está incluso durante o processo de desenvolvimento e uso da aplicação, mediante testes e implementação de funcionalidades a serem adaptadas conforme a necessidade. A mesma, através do upload do arquivo xml do Currículo Lattes, realiza a leitura dos dados e gera a classificação e suas devidas pontuações conforme o Qualis CAPES. Além disso, todas as informações necessárias podem ser acessadas diretamente pelo navegador, juntamente com todos os cálculos e gráficos de comparações de desempenho entre os docentes. O foco principal é obter uma melhor visualização e o agrupamento das informações processadas, interpretadas e analisadas.

Os resultados da aplicação trazem melhorias para a análise do programa, atendendo aos requisitos necessários com eficiência para aplicar a análise. A rotina, que antes era bastante complexa para a coleta de informações, tornou-se mais ágil, tanto no processamento quanto no agrupamento e visualização dos dados. Com pouca intervenção, os resultados são satisfatórios, pois é possível reunir todas essas informações em tabelas e gráficos dinâmicos em questão de instantes, além de ser de fácil acesso e utilização.

Além desta seção introdutória, na seção 2 são apresentados alguns trabalhos relacionados e suas características. A seção 3 traz a metodologia utilizada para o desenvolvimento do trabalho. A seção 4 descreve a aplicação, com a arquitetura e suas funcionalidades, enquanto a seção 5 traz a demonstração da aplicação. A seção 6

aborda os resultados obtidos após a avaliação da aplicação feita pelos docentes do programa, bem como as configurações que podem ser ajustadas de acordo com a classificação e as pontuações a serem usadas. O artigo termina com a seção 7 apresentando a discussão, e a seção 8 exibe as considerações finais e os trabalhos futuros.

2 Trabalhos Relacionados

A ferramenta intitulada LattesPlain, desenvolvido por Vasconcelos [14] na linguagem Python, extrai os dados do Currículo Lattes a partir da leitura dos arquivos XML, e a leitura da planilha com os eventos e estratos, calcula e armazena-os em uma planilha de Excel para análise e comparação entre os docentes. Apesar desta aplicação exercer tais funções, e apresentar os resultados esperados, algumas funções são manuais, como a busca por títulos publicados por mais de um docente e a função de recalcular as notas dos mesmos. Ademais, a classificação não traz, em alguns casos, o estrato correto, sendo necessário inserir manualmente o valor e a nota, além de ser acessada através do prompt de comando (ou terminal) e exibir os dados em planilhas, tendo que ser acessada apenas no próprio computador onde foi executado.

Hantani [4] traz o FastLattes, ferramenta de extração de dados com finalidade de analisar as orientações e projetos de pesquisa de docentes dentro de um PPG, exibindo-as em um dashboard com os comparativos e dados referentes a periódicos e orientações, além das participações em projetos e bancas de graduação, mestrado e doutorado conforme os períodos informados pelo usuário onde é possível executar no próprio computador.

O trabalho de Mendonça [7], desenvolveu uma extensão para os navegadores que automaticamente classifica os artigos em periódicos identificados nas páginas dos Currículos da Plataforma Lattes do CNPq, de acordo com o Qualis da CAPES. Ele consegue, através da busca pelo nome do pesquisador, extrair as informações dos periódicos e calcular suas médias, apresentando gráficos comparativos e tabelas de pontuações de acordo com o ano desejado. Mesmo atendendo de forma satisfatória a classificação e permitindo a análise, para o problema em questão, esta solução além de não classificar as conferências, apenas permite analisar um currículo por vez, selecionando-os via caixa de seleção. Ou seja, não permite realizar uma análise completa e comparativa entre os docentes desejados, impossibilitando a avaliação contínua de um grupo de docentes, gerando assim um trabalho manual para analisar de forma individual estas informações.

O software ScriptLattes, um script GPL (General Public License), como uma ferramenta projetada para extração e compilação automática de produções técnicas, artísticas e bibliográficas, entre outros tipos de produções, de um grupo de pesquisadores cadastrados na plataforma Lattes. Segundo Mena-Chalco and Junior [6], o ScriptLattes é o pioneiro na prospecção de extensos conjuntos de dados acadêmicos provenientes de Currículos Lattes em formato HTML (*Hypertext Markup Language*). A aplicação baixa automaticamente os currículos de um grupo de pessoas, no formato HTML, compilando as listas de produções, tratando as produções duplicadas e gerando relatórios também em HTML, com listas separadas por títulos e em ordem cronológica. Todavia, mesmo com a extensa quantidade de informações que o ScriptLattes traz, o mesmo

não oferece uma visualização geral das publicações com seus estratos e notas, além de que, os autores relatam que a aplicação é multiplataforma, porém, seu site afirma que a sua execução é principalmente sob o Sistema Operacional Linux, e que quando necessário a instalação em outro ambiente, a página não demonstra ou explica a possibilidade de configuração ou demanda muito esforço técnico.

Simão [11], professor da Universidade de São Paulo (USP), apresenta o APOEMA-PG, sistema utilizado para apoio na análise da Pós-Graduação. Disponibilizado ao público na web, a aplicação utiliza dados públicos da plataforma Sucupira, referente a trabalhos concluídos e publicações, além dos dados referentes aos docentes, os quais unem e permitem a seleção de diversas áreas de conhecimento e dos seus PPG (Programas de Pós-Graduação), efetuando comparativos entre dois ou mais programas, ou mesmo uma visão geral de um único programa. Por utilizar dados públicos, a aplicação apresenta uma visão macro dos programas a serem analisados conforme os anos, além do quantitativo de periódicos e conferências, teses e dissertações e trazem dados de mestrado e doutorado e suas colaborações das opções dos PPG selecionados pelo usuário. Contudo, apesar da visão geral que a aplicação apresenta, a visualização dos gráficos são extremamente grandes e alguns não apresentam legendas ou dados que facilitem a leitura dos mesmos, necessitando de um entendimento maior para detectar o que cada gráfico representa. Além do mais, fica complexo analisar pontos importantes dentro do programa, dificultando as tomadas de decisão, principalmente sobre docentes que precisam publicar em qualis melhores, ou seja, devido a falta de visão micro do PPG, se torna difícil medir quantidade x qualidade, o que afeta diretamente o programa, mesmo com os percentuais que são apresentados. Além de não permitir quaisquer alterações dos dados publicados, sendo necessário lançar as informações na plataforma sucupira para que a aplicação possa analisá-las.

3 Procedimento da Pesquisa

Para se alcançar o objetivo proposto, o método escolhido foi a pesquisa-ação, definida por Thiolent [13], como “um tipo de investigação social com base empírica que é concebida e realizada em estreita associação com uma ação ou com a resolução de um problema coletivo e no qual os pesquisadores e os participantes representativos da situação estejam envolvidos de modo cooperativo ou participativo.”

Inicialmente, após a definição do domínio do problema foram definidos os requisitos para o desenvolvimento, dentre os quais, os principais estão:

- (1) Verificar as publicações, analisar e classificá-las como conferências e periódicos;
- (2) Totalizar as notas dos docentes de acordo com seus estratos, permitindo a divisão das pontuações caso haja vários docentes envolvidos na publicação;
- (3) Salvar as informações em uma base de dados e posteriormente utilizá-las para análise contínua através de gráficos comparativos e tabelas com informações específicas de acordo com os anos e estratos;
- (4) Opção de editar publicação, para corrigir possíveis erros de digitação ou correção de estratos e reclassificá-lo;

- (5) Opção para deletar docente, caso o mesmo seja descredenciado do programa.

Após a análise de requisitos, definiu-se a linguagem de programação Python, junto ao framework Flask e suas bibliotecas, para efetuar a leitura, tratamento dos dados e suas classificações, e posteriormente armazenar estas informações na base de dados criada utilizando o SQLite.

A fase de classificação divide as publicações em duas categorias - Periódicos e Conferências - obtidas através do relatório do Qualis Periódicos e Eventos científicos desenvolvido pela CAPES para estratificar a produção intelectual. São 9 estratos divididos em percentis, sendo A1 o mais elevado, passando pelo A2; A3; A4; B1; B2; B3; B4 e C, que possui o peso zero [1].

A partir da classificação, é gerado o parâmetro para se calcular as notas e salvar estes dados extraídos dos arquivos XML na base de dados, processá-las e disponibilizá-las através de gráficos e tabelas. Ao final do desenvolvimento da aplicação, a mesma passou por testes com um grupo de docentes, para avaliar se as funcionalidades atendiam os objetivos previamente estabelecidos e se poderiam haver problemas que requisitem melhorias a serem implementadas em uma futura versão da aplicação.

4 A Aplicação: AnyLattes

A aplicação apresenta uma interface utilizada pelo ator principal que possibilita a inserção dos arquivos referentes aos currículos, importação dos arquivos Qualis e a visualização dos resultados da extração dos dados obtidos da avaliação de forma contínua, onde o usuário pode visualizar, corrigir dados e o sistema recalcula as médias de acordo com as informações inseridas a cada docente conforme o diagrama de atividades de comportamento do sistema presente na Figura 1.

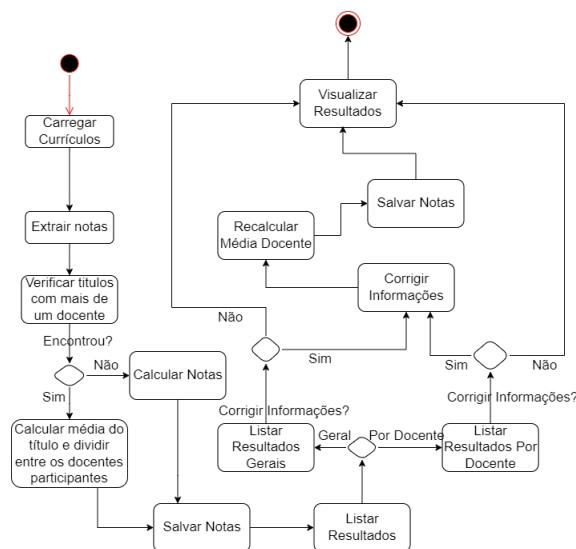


Figure 1: Diagrama de atividade geral do protótipo

4.1 Arquitetura

A arquitetura da aplicação possui um ambiente classificado como servidor, que é composto por classes e componentes que ficam armazenados consumindo recursos e disponibilizando telas de acesso para o usuário conforme requisição. O Ambiente é formado pela base de dados da aplicação, construída em SQLite, e pelos seguintes componentes:

- Classes e bibliotecas em Python;
- Flask;
- SQLite;
- HTML, CSS, Bootstrap e JavaScript;
- Plotly.

A escolha das ferramentas em questão tem o propósito de trazer simplicidade e rapidez à aplicação, além de possuírem fácil manutenção, interpretação e uma documentação bem concisa. A linguagem Python, por ser uma linguagem flexível e de código aberto, e a variedade de bibliotecas tornam-a acessível e de fácil manipulação para a análise de dados complexos, além de ser altamente escalável e rápida para resolução de problemas que não podem ser resolvidos usando outras linguagens de programação [3].

Para trazer uma estrutura leve e com ferramentas e recursos úteis para criação de aplicativos web, o Flask, um microframework feito em Python, que segue o padrão MVC (Model-View-Controller), oferecendo flexibilidade e uma estrutura bem acessível. Mesmo com sua arquitetura simples, permite a criação de aplicações robustas devido ser totalmente personalizável e sem a necessidade de realizar configurações que muitas vezes não serão utilizadas [2]. Possui também um suporte embutido para renderizar templates HTML que permite a criação de páginas web dinâmicas.

Também para trazer leveza, simplicidade e flexibilidade no armazenamento dos dados, utilizou-se o SQLite, um mecanismo de banco de dados SQL integrado transacional independente, ou seja, sem a necessidade de um servidor sem qualquer configuração, uma ferramenta open source e para qualquer finalidade [12].

O Bootstrap, um framework front-end de código aberto para a construção de sites mais rápidos sem a preocupação com comandos básicos e funções adicionais, além de possuir tipos de templates prontos baseados em HTML e CSS, e a utilização de Javascript para diversas funções e componentes, afim de poupar tempo no gerenciamento e trazer responsividade e otimização à interface [5]. Tal ferramenta foi utilizada para compor a interface da aplicação, que permite a inserção de arquivos e exibição de dados, e pode ser acessado via navegador (Chrome, Firefox, Edge, etc.) através do endereço do servidor ou via URL de hospedagem.

O Plotly, biblioteca Python para visualização de gráficos dinâmicos, que podem ser facilmente compartilhados e explorados. Construída sobre a biblioteca Javascript(plotly.js), permite que os usuários python criem visualizações interativas baseadas na web. São salvos em arquivos HTML independentes ou servidas como parte de aplicativos da web usando dashboards [10]. Dentro da aplicação, é utilizada para plotar gráficos no front-end a partir de dados exportados para um arquivo JSON.

4.2 Funcionamento

O funcionamento da aplicação se divide em quatro momentos, que dependem de decisões realizadas do início à conclusão de todo processo. São eles:

- Carregar currículos e informar o período inicial e final;
- Extraír, classificar e processar as informações retiradas do XML;
- Atribuir notas a partir da classificação em periódicos e conferências;
- Exibir as informações classificadas, processadas e organizadas a partir do período informado.

As decisões são iniciadas desde a classificação em periódicos e conferências dos títulos extraídos dos currículos. As informações passam por uma comparação com os arquivos que são disponibilizados pela CAPES e importadas à aplicação, trazendo os nomes dos eventos e revistas científicas, junto dos estratos. Após essa classificação, são atribuídas as notas equivalentes a cada estrato, e são verificados os títulos que possuem publicação por mais de um docente para que sejam divididas as notas proporcionalmente a cada um. Por fim, as informações são salvas na base de dados, sem qualquer intervenção externa durante todo o fluxo. Na Figura 2 são ilustradas as etapas e decisões supracitadas.

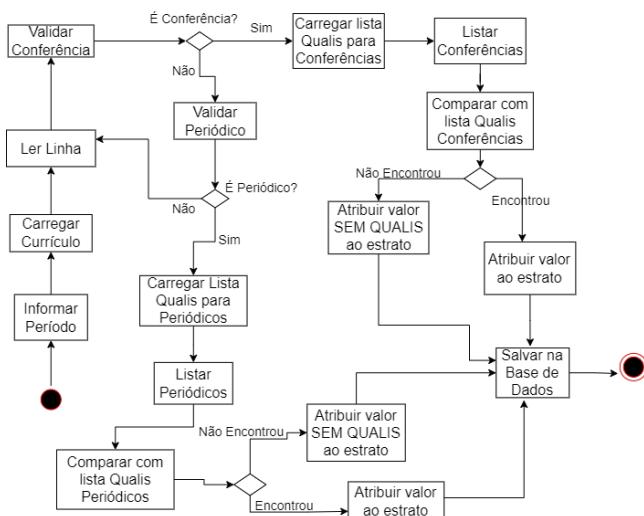


Figure 2: Rotina para extração de notas

Após o processamento das informações, pode-se visualizar as telas em HTML tanto no modo geral (todas as informações dos docentes) quanto no modo específico (selecionando um docente e apresentando suas médias e quantidade de publicações durante os anos), através de tabelas e gráficos dinâmicos. As informações também podem ser ajustadas, caso haja necessidade, com opções de edição, para fins de correção de títulos, estratos e outros dados, e a mesma passará por um reprocessamento caso o estrato mude e sua nota altere, afetando ou não a média do docente.

A extração das notas é iniciada a partir da importação dos currículos e pelo período informado. Primeiro, a aplicação lista os nomes dos docentes dos currículos. O título passa pela comparação com

as listas de classificação do Qualis - arquivos em extensão XLSX para Conferências, utilizando a biblioteca **OpenPyxl** e extensão PDF para Periódicos, utilizando a biblioteca **PyPDF2** - e através de laço de repetição e estrutura de comparação, separa as publicações em uma lista com:

- Tipo de publicação (Periódico ou Conferência);
- Ano do evento;
- Título;
- DOI;
- Sigla;
- Nome do Evento;
- Autores.

Utilizando o nome do evento ou a sigla para identificar artigos de Conferências e o nome ou o ISSN para identificar artigos de Periódicos, pode-se identificar os estratos, que passam por uma comparação em uma tabela salva na base de dados com as notas determinadas para cada estrato e depois salva as alterações. A aplicação permite a divisão proporcional das notas entre os docentes que publicaram o mesmo trabalho, escolha esta que pode ser feita nas configurações.

5 Avaliação

Após toda a codificação, foram realizados testes com um grupo de docentes. Durante os testes foi possível avaliar e apresentar melhorias para este trabalho. Os problemas apontados foram aceitos, corrigidos e serão tratados na demonstração e nos resultados.

5.1 Demonstração do AnyLattes

Durante a demonstração da aplicação, foi utilizado o Sistema Operacional Windows, o navegador Google Chrome, e alguns XML, para testar as funcionalidades e verificar como ela agrega agilidade e facilidade na obtenção das notas, sem tanto trabalho manual. Neste ponto, a aplicação foi avaliada positivamente, pois a única intervenção necessária ocorre no início do processo. A Figura 3 contém a tela de importação que é utilizada caso haja novas tabelas com os eventos e seus respectivos estratos, disponibilizadas pela CAPES para comparação e extração de notas.

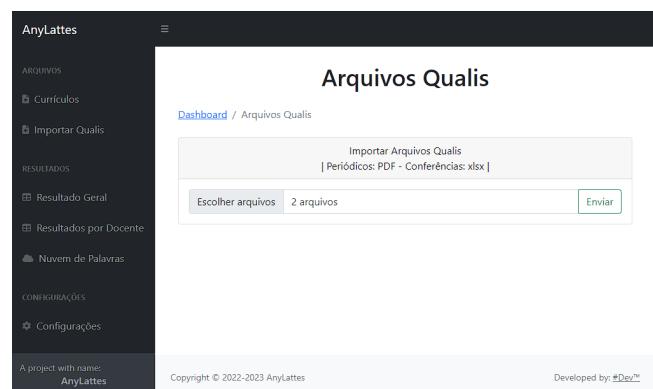


Figure 3: Tela de importação dos arquivos Qualis

A tela principal, visualizada na Figura 4, contém o campo para inserir os arquivos XML dos currículos (que também podem ser compactados em um arquivo ZIP) e o campo para informar o período inicial e final que serão analisados, podendo importar um ou mais currículos simultaneamente. Para o teste foram utilizados 29 currículos no período de 2021 a 2023. Ao término, foram identificados pontos a serem ajustados e acrescentados na aplicação.

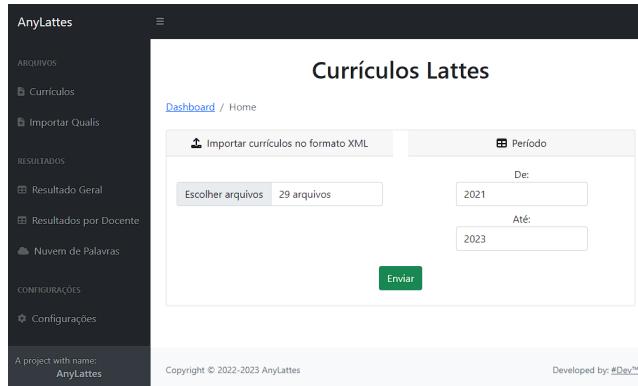


Figure 4: Tela para importação dos currículos

6 Resultados e Versão Atual

Os testes efetuados foram de grande importância, pois através deles, o AnyLattes chega a uma versão que atende com êxito os requisitos anteriormente citados, principalmente por ser uma aplicação de fácil manuseio, e que torna prática uma atividade que demandava extenso trabalho manual.

O AnyLattes demonstrou que é capaz de extrair, organizar os dados, efetuar os cálculos necessários e exibir as informações em uma tela HTML bem estruturada, apresentando as informações necessárias para a avaliação de um PPG. Algumas considerações foram colocadas pelos docentes durante os testes e foram bem aceitas. São elas:

- (1) Salvar a lista do Qualis na base de dados;
- (2) Apresentar um gráfico entre periódico ou conferência;
- (3) Apresentar grafo ou tabela de colaboração;
- (4) Gerar uma Nuvem de palavras mais utilizadas nas publicações de cada docente;
- (5) Apresentar gráficos dos totais de conferências e periódicos de cada docente;
- (6) Apresentar gráficos dos totais de estratos nos anos informados;
- (7) Automatizar a busca dos arquivos XML na plataforma Lattes.
- (8) Possibilidade de reimpor currículos sem perda dos dados.

É importante citar algumas dificuldades encontradas durante o processo de codificação até chegar na versão que o AnyLattes se encontra. Nas publicações em conferências, a falta de um único identificador tornou a comparação complexa, dependendo principalmente do nome do evento ou da sigla para análise. Em alguns casos, os nomes são registrados de forma incorreta, com palavras faltando ou escritos em inglês, mesmo quando o arquivo Qualis apresenta o nome do evento em português, podendo ocorrer também

o inverso. No entanto, isso não ocorre com os periódicos, pois eles possuem o ISSN como identificador, o qual é utilizado pelo arquivo Qualis.

Contudo, a aplicação permite a edição das publicações para efetuar tais correções quando necessárias. Mesmo com as problemáticas apresentadas, a aplicação apresenta os resultados esperados e já aplica algumas das considerações citadas anteriormente. Após todo o processamento feito a partir das importações dos currículos e arquivos Qualis, o AnyLattes divide os resultados em três funcionalidades, as quais serão apresentadas a seguir.

6.1 Resultado Geral

O Resultado geral apresenta uma visão total dos resultados obtidos a partir das importações, demonstrados através de tabelas e gráficos comparativos e dinâmicos, que possibilitam avaliar o desempenho das publicações dos docentes que compõem o programa e extrair os pontos necessários para as tomadas de decisão.

A Figura 5 contém um gráfico com as médias e mediana de todos os docentes do PPG, calculadas ao processar as notas, seja na inserção, edição ou remoção de docentes. O cálculo para obtenção das médias é feito a partir da extração dos dados dos currículos, mais especificamente utilizando os nomes dos eventos, sigla ou o DOI no caso dos Periódicos, em comparação à tabela Qualis, as quais possuem os nomes dos eventos e os estratos equivalentes.

A partir disso, é associado àquele título um estrato, e cada estrato possui uma determinada nota, notas estas que são informadas pela CAPES no documento da área de Ciência da Computação e, dentro da aplicação, são salvas na base de dados. Após estas comparações, são somadas e salvas, para posteriormente efetuar o cálculo para retirada das médias de cada docente, além da mediana que é o valor central das médias já calculadas e postas em ordem crescente. Observa-se que é possível fazer a comparação entre as publicações dos docentes, pois já estão consideradas as produções que possuem coautores de um mesmo título com a nota dividida.

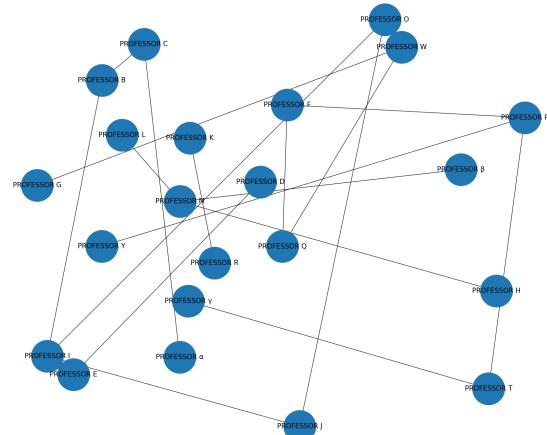


Figure 6: Grafo de colaboração entre docentes

Nas publicações, sejam Periódicos ou Conferências, um docente pode colaborar em um ou mais títulos, agindo como um coautor.

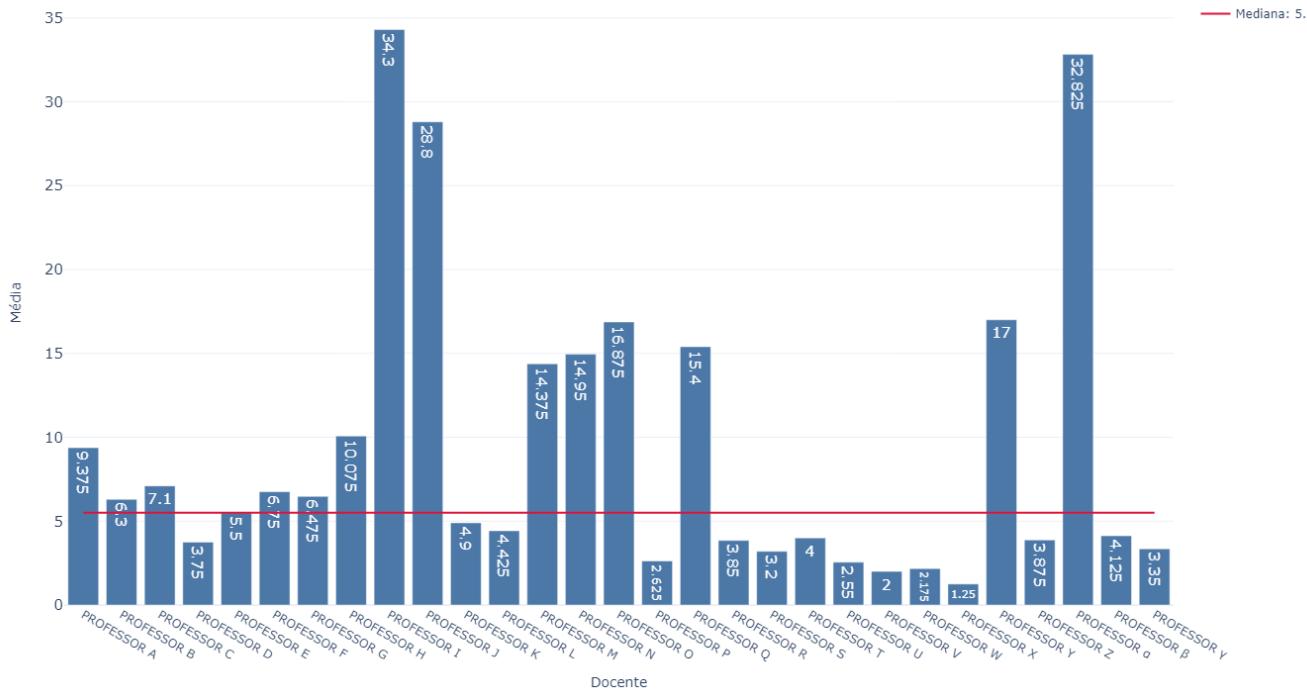


Figure 5: Gráfico das médias e medianas das publicações dos docentes

Ao processar os dados, as notas das publicações que possuem colaboração são divididas igualmente entre os docentes envolvidos e inclusos dentro daquele PPG. Com isso, a aplicação identifica essa coautoria e representa através de um grafo de colaboração, conforme apresentado na Figura 6. Os vértices representam os docentes e estão interligados uns aos outros por meio de arestas que representam colaboração entre os mesmos.

Para ter a visão geral do Programa, é necessário ter a noção da quantidade de publicações feitas durante os anos e uma representação do total de Conferências e Periódicos, com a identificação dos estratos. A aplicação informa quantos e quais estratos foram publicados através do gráfico geral, um gráfico dinâmico que torna possível visualizar o total de cada estrato por ano, representado pelo gráfico de barras, assim como a totalização de Periódicos e Conferências que o programa possui, visualizado em um gráfico de pizza, contendo o percentual entre Conferências e Periódicos naquele mesmo período informado, conforme pode-se visualizar na Figura 7.

Os gráficos contém mais informações como as quantidades e qual barra representa o estrato, ou dentro do gráfico de pizza que identifica em qual parte é representada a conferência ou periódico. Estas informações são acessadas por meio de um rótulo apenas passando com o mouse por cima do gráfico. O gráfico permite a análise do crescimento ou decréscimo do total de produções ao longo dos anos.

Em um Currículo Lattes, é possível extrair quaisquer informações referentes a uma publicação. Cada docente preenche as informações

de acordo com os dados daquela publicação, e para o PPG, a visualização individual, e o agrupamento destes dados é um fator determinante, até mesmo para identificar possíveis melhorias dentro do programa. A tabela com as produções intelectuais mostra as informações referentes a cada título de cada docente e suas devidas notas e estratos, além de direcionar ao DOI da publicação e um Modal que ao clicar sobre o título, traz outras informações importantes daquela publicação, como quais autores estão envolvidos na publicação, o título, o nome do evento e a sigla (Figura 8).

Relação das publicações								
Show:	10	entries	Search:					
Docente	Tipo	Ano	Titúlo	Estrato	Notas	Editor	DOI	DOI
PROFESSOR A	Conferência	2021	autonomous UAV Simulator for Research and Development Applied to 5G Networks	-	0.0			
PROFESSOR A	Conferência	2021	An LSTM-based Approach for Holdover Clock Disciplining in IEEE 1588 PTP Applications	A1	1.0			
PROFESSOR A	Conferência	2021	Analysis of Kalman Filtering for Clock Synchronization in PTP-Unaware Networks	B1	0.5			
PROFESSOR A	Conferência	2021	Reinforcement Learning for Scheduling and Mimo Beam Election Using Caviar Simulation	-	0.0			-
PROFESSOR A	Periódico	2021	Artificial Intelligence for Enhanced Mobility and 5G Connectivity in UAV-Based Critical Missions	A3	0.75			
PROFESSOR A	Periódico	2021	Simulation of machine learning-based 6G systems in virtual worlds	-	0.0			-
PROFESSOR A	Periódico	2021	Clock Synchronization Algorithms Over PTP-Unaware Networks: Reproducible Comparison Using an FPGA Testbed	A3	0.75			
PROFESSOR A	Periódico	2021	Functional Split and Frequency-Domain Processing for Fronthaul Traffic Reduction	A1	1.0			
PROFESSOR A	Periódico	2021	Deep Transfer Learning for Site-Specific Channel Estimation in Low-Resolution mmWave MIMO	A1	1.0			
PROFESSOR A	Conferência	2022	A Coverage-Aware VNF Placement and Resource Allocation Approach for Disaggregated vRANs	A1	1.0			

Figure 8: Tabela Geral de Produção Intelectual

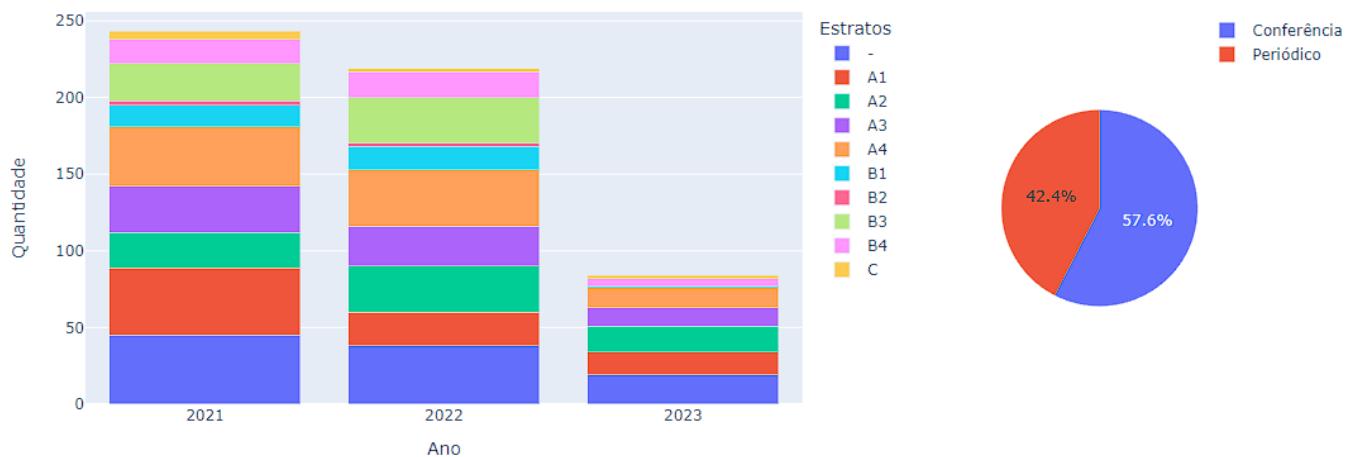


Figure 7: Gráfico total de estratos publicados por ano e Periódico x Conferência

A falta de um identificador para Conferências torna difícil a extração exata e a classificação das publicações, e devido a possíveis erros até mesmo de digitação ou caracteres que podem inviabilizar a análise e por muitas vezes fragmentar o texto dentro do XML, o AnyLattes permite a edição das publicações, possibilitando corrigir informações referentes ao nome do evento e estrato e reprocessá-las. Isso irá influenciar também o gráfico de média e mediana caso haja alteração no Estrato e consequentemente a nota da publicação conforme o exemplo visto na Figura 9.

Editar Publicações

Nome:
PROFESSOR A

Título:
An LSTM-based Approach for Holdover Clock Disciplining in IEEE 1588 PTP Applications

Evento:
GLOBECOM 2021 2021 IEEE Global Communications Conference

DOI : 10.1109/globecom46510.. Sigla: GLOBECOM Estratos: A1

Figure 9: Edição de uma publicação

Para auxiliar o rastreio de publicações que possuem coautores, e manter a organização dos dados e analisá-los nos mínimos detalhes, além do Grafo de colaboração, a aplicação identifica quais docentes publicaram o mesmo título através da tabela de colaboração (Figura 10). a qual apresenta os títulos em ordem crescente, mas sendo possível alterar a ordenação e filtrar qual título for necessário.

Existe um modal que exibe os dados de quais docentes do PPG são colaboradores, além da nota da publicação conforme o estrato e um link para o DOI da publicação (quando houver).

Tabela de Colaboração	
Show 10 entries	Search:
Titúlo:	Colaboradores
A Map Building and Sharing Framework for Multiple UAV Systems	+
A Monitoring System for Electric Vehicle Charging Stations: A Prototype in the Amazon	+
A Survey of Music Visualization Techniques	+
A Survey on Long-Range Wide-Area Network Technology Optimizations	+
A Techno-Economic Framework for Installing Broadband Networks in Rural and Remote Areas	+
A brief review of dashboard visualizations employed to support management or business decisions	+
A low-cost multi-user augmented reality application for data visualization	+
A strategy to the reduction of communication overhead and overfitting in Federated Learning	+
Adaptive Parameters for LoRa-Based Networks Physical-Layer	+
Adaptive priority-aware LoRaWAN resource allocation for Internet of Things applications	+

Figure 10: Títulos que possuem colaboração entre os docentes

6.2 Resultados por Docente

Apesar da visão ampla do programa, ainda é necessário analisar cada membro que compõe o PPG em prol de melhorias, não apenas na quantidade de publicações, mas também no estrato que pode aumentar a média daquele docente. A aplicação permite uma outra visualização dos resultados, a análise feita por cada docente, onde as quatro visualizações disponíveis (por nota, por produção intelectual, gráfico de publicações por ano e nuvem de palavras por docente), são necessárias para a avaliação individual. A aplicação possui uma função de remoção de docente, para caso o mesmo não faça mais parte do programa. Tal função remove o docente selecionado da base de dados, afetando diretamente o gráfico de média e mediana, pois se o mesmo for coautor de outra publicação, isso irá afetar as médias dos remanescentes e a mediana, além da visualização do grafo e da tabela de colaboração.

Na Figura 11 é possível ver um exemplo da análise individual, a visualização por nota, em que se apresentam os estratos, as quantidades que foram publicados e o total das notas de cada estrato naquele ano. Esta visualização pode ser ordenada por estrato, quantidade ou total de notas, além de filtrar quaisquer informações registradas nesta tabela. No final da tabela, apresenta-se a média geral deste docente, média essa que foi calculada a partir da soma de todas as notas dos estratos durante os anos e dividido pela quantidade de publicações feitas. Tal informação é exibida também no gráfico de média e mediana, nos resultados gerais.

The screenshot shows a table titled "Docente" with the following data:

Ano	Estrato	Quantidade	Total Notas
2021	B3	1	0.1
2022	-	4	0.0
2022	A2	1	0.875
2022	A4	2	1.25
2022	B3	3	0.3
2022	B4	2	0.1
2023	-	3	0.0
2023	A2	1	0.875
2023	B1	1	0.5

Below the table, it says "Showing 1 to 9 of 9 entries". At the bottom, there is a summary bar with "Média: 4.0".

Figure 11: Tabela quantitativa de estratos e suas notas

Como foi mostrado no resultado geral, a visão por docente também possui gráficos dinâmicos que expressam a quantidade de estratos registrados durante os anos e quais estratos são, podendo ser filtradas as visualizações destes clicando no estrato necessário, ou mesmo desmarcá-los para exibir todos. A Figura 12 contém o mesmo quantitativo expresso em gráfico de barras durante os anos, mas, é referente ao docente selecionado, com o total de cada estrato obtido naquele ano, e o gráfico de pizza onde é possível visualizar se o docente fez mais publicações em Conferências ou em Periódicos, informações expressas em percentuais equivalentes ao total de publicações.

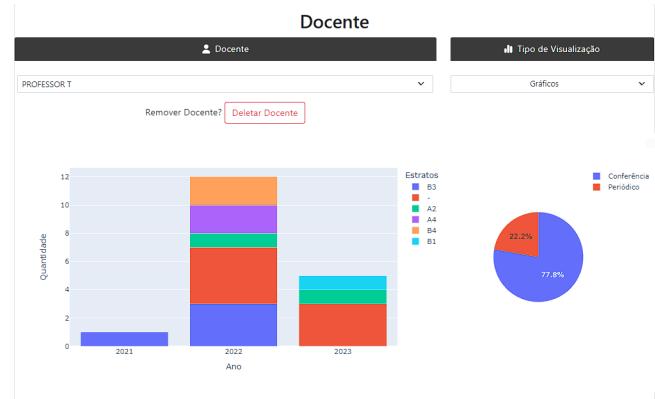


Figure 12: Gráfico quantitativo de publicações por ano e Conferência x Periódico

Como uma forma de organizar os dados dos docentes, a aplicação fornece a tabela das produções intelectuais (Figura 13), onde é possível visualizar cada publicação que foi registrada naquele período informado, com informações de qual tipo de evento se classifica a publicação, o ano, o título, qual o estrato referente a ela, e o DOI (se houver).

É possível editar um título, da mesma forma que pode ser feito no resultado geral, onde possibilita a edição do evento e do estrato referente a este título e caso haja alterações, o sistema reprocessa e reclassifica aquele título e caso haja alterações na nota, o mesmo recalcular a média e a mediana.

É apresentada uma nuvem de palavras referente a palavras mais utilizadas nos títulos por aquele docente, contudo, será explicado melhor a seguir.

The screenshot shows a table titled "Relação das publicações" with the following data:

Tipo	Ano	Título	Estrato	Notas	Ação	DOI
Conferência	2021	A STUDY ON DISCOVERING ACCESSIBILITY ISSUES IN THE SOFTWARE DEVELOPMENT PROCESS	B3	0.1	<input checked="" type="checkbox"/>	DOI
Conferência	2022	Proposta e Avaliação de uma Ferramenta Web para Vocalização de Gráficos de Barras	-	0.0	<input checked="" type="checkbox"/>	DOI
Conferência	2022	Explanation-by-Example Based on Item Response Theory	A4	0.625	<input checked="" type="checkbox"/>	DOI
Conferência	2022	A systematic mapping study on digital game adaptation dimensions	-	0.0	<input checked="" type="checkbox"/>	DOI
Conferência	2022	Accessibility assessment of violations on the slack overflow platform	-	0.0	<input checked="" type="checkbox"/>	DOI
Conferência	2022	TikTok: Qual o Impacto do Crescimento da Plataforma?	B4	0.05	<input checked="" type="checkbox"/>	DOI
Conferência	2022	Uma Análise Comparativa sobre o que Dizem a Grande Midia e os Usuários Comuns no Twitter sobre os Presidentáveis Brasileiros em 2022	B4	0.05	<input checked="" type="checkbox"/>	DOI
Conferência	2022	ADAPTACAO E AUTOMACAO DE TESTE DE CANCELAMENTO PARA AVALIAÇÃO DE COMPORTAMENTO DE EXPLORAÇÃO VISUAL	B3	0.1	<input checked="" type="checkbox"/>	DOI
Conferência	2022	DESIGN CONSIDERATIONS FOR GENOMICS TOOLS BASED ON CIRCULAR VISUALIZATION OF BIOLOGICAL DATA	B3	0.1	<input checked="" type="checkbox"/>	DOI
Conferência	2022	A STUDY ON ACCESSIBILITY ASPECTS OF SOFTWARE DEVELOPMENT IN BRAZIL	B3	0.1	<input checked="" type="checkbox"/>	DOI

Below the table, it says "Showing 1 to 10 of 18 entries".

Figure 13: Produção Intelectual do docente

6.3 Nuvem de Palavras

O AnyLattes também gera uma nuvem de palavras com os termos mais utilizados dentro dos títulos das publicações, em que se pode observar e identificar as áreas de atuação de cada docente. Ela é caracterizada por sua forma de percepção visual para facilitar a

compreensão dos termos mais relevantes dentro de um texto. Na Figura 14, é possível observar a nuvem gerada pela aplicação, a partir das publicações dos 29 XML importados.

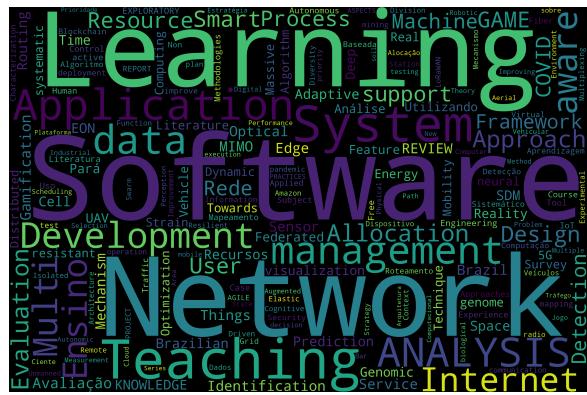


Figure 14: Nuvem geral de Palavras

6.4 Configurações

A Capes utiliza critérios de classificação para cada estrato, através de pontuações. Conforme as mudanças destas pontuações, é possível alterá-las dentro da aplicação e a mesma irá recalcular as notas. A Figura 15, apresenta a tabela de pontuações atualizada e uma opção de correção de notas, a qual proporcionaliza as notas dos docentes que colaboraram em uma mesma publicação. A função é aplicada sobre os títulos que são identificados na tabela de colaboração. Ela primeiramente iguala as pontuações daquele título, e depois divide a nota pela quantidade de docentes participantes.

Pontuações Qualis		Correção de Notas
Estratos	Notas	
A1	1.0	
A2	0.875	
A3	0.75	
A4	0.625	
B1	0.5	
B2	0.2	
B3	0.1	
B4	0.05	
C	0.0	
SEM QUALIS	0.0	

Figure 15: Tabela de Pontuações Qualis e opção para divisão de notas de trabalhos colaborados

7 Discussão

Mediante avaliação implementada pela CAPES, as informações de desempenho dos PPG precisam ser organizadas e claras, pois, é possível verificar a qualidade do programa através de suas publicações. Para tal coleta destas informações, é necessário processar

os dados em um menor tempo e precisão nos resultados obtidos. O AnyLattes se caracteriza pela consistência na extração, na classificação e organização de informações extraídas de Currículos Lattes. O mesmo não se limita a uma área de atuação específica, e pode ser expandido a outros PPG ou a qualquer conjunto de dados de docentes no Brasil devido aos dados possuírem a mesma estrutura (XML) e serem interpretadas facilmente dentro da aplicação para obtenção dos resultados. A partir disso, as informações são exibidas de forma clara, e explicativa devido a precisão dos resultados. Ele facilita a visão geral e específica do programa a ser analisado, além de ser uma aplicação web, pode ser acessada a partir de dispositivos móveis e computadores.

A principal dificuldade foi a classificação das Conferências pela falta de um identificador, todavia, o problema foi contornado com a implementação da edição das publicações. Outro ponto seria o download dos XML dos docentes, já que a plataforma Lattes limita o acesso a estes dados implementando o sistema de Captcha, o qual dificulta o download de forma automatizada. Salvar a tabela Qualis na base de dados para atualizações e agilizar a comparação dos eventos e relação dos estratos também foi considerada uma dificuldade. Contudo, tais alterações serão implementadas em versões futuras.

Caso não houvesse tais dificuldades, o AnyLattes iria dispor do sistema de importação de dados automatizado dos currículos, onde seria possível apenas informar a relação de ID Lattes dos docentes ou o nome dos docentes, e o sistema trataria de consultar na base de dados de forma automática os dados respectivos e importaria, onde hoje, o sistema de Captcha da plataforma Lattes, torna o processo manual. Atualmente, os dados das tabelas Qualis são disponibilizados em arquivos Excel, mas se fossem disponibilizados através de uma API, iria tornar o processo de comparação dos eventos e relacionar aos estratos devidos mais fácil, pois não teria a necessidade de atualizar tais arquivos e processar os dados, simplesmente seria consultada e comparada de maneira mais rápida e concisa.

8 Considerações Finais e Trabalhos Futuros

Este artigo apresentou a aplicação web denominada AnyLattes, desenvolvida para auxiliar coordenações de Programas de Pós-Graduação a efetuarem de forma automatizada a avaliação contínua e análise de desempenho das publicações e participações em eventos ao longo dos anos, classificando Conferências e Periódicos, organizando, calculando e agrupando as informações necessárias para serem utilizadas como base para tomadas de decisão dentro do programa e para o preenchimento anual da Plataforma Sucupira.

Com os resultados dos testes, pôde-se perceber avanços nos objetivos apresentados, a metodologia aplicada foi de grande importância para aquisição de conhecimentos nas áreas de gestão de dados, informação e conhecimento. As correções feitas conforme as solicitações dos docentes, foram essenciais para que a aplicação se diferenciasse, especialmente na resolução de problemas relatados, pela falta de identificadores para relacionar as publicações, na questão organizacional e agrupamento de informações, além da simplicidade e questão visual que viabiliza o uso.

Entre os trabalhos futuros, podemos considerar o armazenamento das tabelas Qualis na base de dados para agilizar a comparação dos eventos e relação dos estratos, onde é necessária toda

uma dependência de biblioteca para percorrer os dados e comparar com o nome do evento referente a publicação, onde pode ser substituído por uma consulta na base de dados, tornando o processo mais rápido e preciso. A automatização do download de currículos diretamente da plataforma Lattes, onde hoje é preciso efetuar o download diretamente da plataforma e fazer o upload dos arquivos para a aplicação, poderá ser feita a busca utilizando o ID Lattes ou mesmo o nome completo do docente, e a partir disso, trazer os dados e importá-los para a aplicação. Por fim, a inclusão de um módulo para avaliar não somente as publicações em eventos, mas também as orientações de discentes seja no nível de graduação, mestrado e doutorado, onde estes dados possuem sua relevância não somente dentro de um PPG, também para a avaliação da CAPES.

Por fim, o código fonte da aplicação e o README com explicação para execução em ambiente Windows, Linux e em container docker, está disponibilizado na plataforma online GitHub (<https://github.com/alexcirilo/AnyLattes>), para que possa servir à sua finalidade, além de seguir a proposta de Software Livre para ser aprimorado por outros desenvolvedores interessados.

References

- [1] CAPES. 2019. Relatório do Qualis Periódicos e Eventos Científicos. Disponível em: https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/qualis_periodico_eventos_cientifico_Ciencia_Computacao.pdf. Acesso em: 29 de abril de 2023.
- [2] Ana Paula de Andrade. 2019. O que é Flask? Disponível em: <https://www.treinaweb.com.br/blog/o-que-e-flask>. acesso em: 16 de outubro de 2023.
- [3] Insight Lab Science Lab Laboratório de Pesquisa em Ciência de Dados na Universidade Federal do Ceará. 2019. Por que o Python é a Linguagem mais adotada na área de Data Science ? Disponível em: <https://www.insightlab.ufc.br/por-que-o-python-e-a-linguagem-mais-adotada-na-area-de-data-science/>. acesso em: 16 de outubro de 2023.
- [4] Guilherme Eiji Enomoto Hantani. 2022. FastLattes: Uma ferramenta para extração de dados acadêmicos de currículos Lattes. Disponível em: <https://github.com/guihantani/FastLattes>. Acesso em : 06 de maio de 2023.
- [5] Andrei L. 2023. Desvendando o Bootstrap: O Que É e Como Usar? Disponível em: <https://www.hostinger.com.br/tutoriais/o-que-e-bootstrap>. acesso em: 16 de outubro de 2023.
- [6] J. P. Mena-Chalco and R. M. Cesar Junior. 2013. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. In *Bibliometria e Cientometria: reflexões teóricas e interfaces (pp.109-128)*, Maria Cristina Piumbato Innocentini Hayashi and Jacqueline Letta (Eds.). São Carlos: Pedro & João.
- [7] Nabor Mendonça. 2023. QLattes, Extensão do Chrome que automaticamente classifica os artigos em periódicos identificados nas páginas dos CVs da Plataforma Lattes do CNPq, de acordo com o Qualis da CAPES. <https://chrome.google.com/webstore/detail/qlattes/cobekobjpobenpjggbpkkllkcfoinen?hl=pt-BR>, acesso em: 13 de maio de 2023.
- [8] Ministério da Educação. 2013. História e Missão CAPES. Disponível em: <https://www.gov.br/capes/pt-br/acesso-a-informacao/institucional/historia-e-missao>. Acesso em: 15 de outubro de 2023.
- [9] Pedro Arcanjo. 2014. Capes Lança Plataforma Sucupira para gestão da pós graduação. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/capes-lanca-plataforma-sucupira-para-gestao-da-pos-graduacao>. Acesso em: 15 de outubro de 2023.
- [10] Plotly. 2023. Plotly Python Overview. Disponível em: <https://plotly.com/python/getting-started/>. acesso em: 16 de outubro de 2023.
- [11] Adenilso Simão. 2021. APOEMA-PG: Sistema de apoio estratégico à medição e análise da Pós Graduação. Disponível em: <http://thanos.icmc.usp.br:40123/apoema/analysis>, acesso em 17 de agosto de 2023.
- [12] SQLite. 2023. About SQLite. Disponível em: <https://www.treinaweb.com.br/blog/o-que-e-flask>. acesso em: 16 de outubro de 2023.
- [13] M Thiolent. 1988. *Metodologia da pesquisa-ação*. São Paulo: Cortez.
- [14] Fábio Alex Mendes Campos Vasconcelos. 2019. LattesPlan: Uma Ferramenta para Extração e Organização de Informações de Currículos Lattes. Disponível em: <https://github.com/fabioalexmcv/LattesPlan>, acesso em: 15 de abril de 2022.