



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

GUILHERME EIJI ENOMOTO HANTANI

**FASTLATTES: UMA FERRAMENTA PARA EXTRAÇÃO DE DADOS
ACADÊMICOS DE CURRÍCULOS LATTES**

**BELÉM
2022**



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

GUILHERME EIJI ENOMOTO HANTANI

**FASTLATTES: UMA FERRAMENTA PARA EXTRAÇÃO DE DADOS
ACADÊMICOS DE CURRÍCULOS LATTES**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação. Faculdade de Computação. Instituto de Ciências Exatas e Naturais.

Orientador: Prof. Dr. Reginaldo Cordeiro Dos Santos Filho

**BELÉM
2022**

Hantani, Guilherme

FastLattes: Uma ferramenta para extração de dados acadêmicos de currículos Lattes/ GUILHERME EIJI ENOMOTO HANTANI. – BELÉM, 2022.

41 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Reginaldo Cordeiro Dos Santos Filho

Monografia – UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO, 2022.

1. Lattes. 2. Extração de Dados. 3. Python. 3. Automatização. I. Título.

GUILHERME EIJI ENOMOTO HANTANI

**FASTLATTES: UMA FERRAMENTA PARA EXTRAÇÃO DE
DADOS ACADÊMICOS DE CURRÍCULOS LATTES**

Trabalho de Conclusão de Curso apresentado
como requisito parcial para a obtenção do tí-
tulo de Bacharel em Ciência da Computação.
Faculdade de Computação. Instituto de Ciências
Exatas e Naturais.

Data da Defesa: 20 de Dezembro de 2022

Conceito:

Banca Examinadora

Prof. Dr. Reginaldo Cordeiro Dos Santos

Filho

Faculdade de Computação - UFPA
Orientador

Prof. Dra. Marcelle Pereira Mota

Faculdade de Computação - UFPA
Membro da Banca

Prof. Dra. Regiane Silva Kawasaki Francês

Faculdade de Computação - UFPA
Membro da Banca

BELÉM
2022

AGRADECIMENTOS

Primeiramente, gostaria de agradecer especialmente meus amados pais e irmã, que me deram todo o amor e apoio possível durante o período do curso.

Ao meu orientador, prof. Reginaldo, por todos os ensinamentos, orientações e assistência recebidas durante a elaboração deste trabalho e minha caminhada acadêmica.

À toda a minha família, que compreendeu minha ausência durante os períodos difíceis enquanto me dedicava a realização deste trabalho.

Aos meus amigos do conselho, pela amizade incondicional e os vários momentos de lazer que me ajudavam em tempos difíceis.

Aos amigos que fiz dentro da universidade, pela amizade e bons momentos vividos, sempre me apoiando durante essa longa jornada.

Por fim, a todos que não foram mencionados, porém contribuíram de alguma forma para minha jornada acadêmica.

*“A necessidade é a mãe da inovação.”
(Platão)*

RESUMO

Devido a grande quantidade de currículos Lattes existente na plataforma Lattes, combinado com o fato de que a plataforma possui poucos recursos para visualização de dados quantitativos, muitas instituições, como programas de Pós-graduação, precisam fazer a contabilização de trabalhos científicos dos pesquisadores de maneira manual, formando um processo bem cansativo, trabalhoso e propenso a erros. Dessa forma, o objetivo deste trabalho é desenvolver uma sistema web que seja capaz de extrair, de maneira automatizada, dados e informações de um conjunto de currículos Lattes, e apresente informações estatísticas por meio de gráficos visuais gerados pelo mesmo. Para isso, foram definidas duas listas de requisitos, funcionais e não-funcionais, além de um diagrama de casos de uso UML, e também, foram selecionadas 5 bibliotecas externas do *Python* para auxiliar no desenvolvimento da ferramenta. O sistema web também conta com outro recurso visual, nuvens de palavras, a qual conseguem capturar a frequência com que algumas palavras importantes se repetem no currículo do pesquisador, informando ao usuário sobre o perfil de atuação do pesquisador. A aplicação está disponível publicamente e pode ser utilizada tanto para analisar um currículo particular de um pesquisador quanto para realizar análises conjuntas, considerando, por exemplo, um colegiado inteiro de um programa de pós-graduação.

Palavras-chave: lattes, extração de dados, python, automatização.

ABSTRACT

Due to the big quantity of existing Lattes resumes in the Lattes platform, combined with the fact that the platform disposes of few resources to visualization of quantitative data, many institutions, such as postgraduate programs, need to make the accounting of scientific works by hand, leading to an exhausting and laborious process, prone to mistakes. Thus, the objective of this paper is developing an web system capable of extracting, in an automatic way, data and information from an set of Lattes resumes, and presenting statistic info through visual graphs generated by it. For that, two requirements lists were defined, functional and not-functional, in addition to a case-use UML diagram, and also, 5 Python external libraries were selected to assist in the development of the tool. The web system also counts with another visual resource, word clouds, which are capable of capturing the frequency of which some important words repeat themselves on the resume of the researcher, informing the user about the occupational profile of the researcher. The application is also publicly available and can be used to analyze an particular resume of a researcher, or to do joint analyzes, considering, for example, an whole postgraduate program collegiate.

Keywords: lattes, data extraction, python, automation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de Casos de Uso do Sistema.	21
Figura 2 – Funcionamento da estrutura da aplicação.	25
Figura 3 – Tela inicial da aplicação.	27
Figura 4 – Menu de seleção de página.	27
Figura 5 – Página específica do pesquisador.	28
Figura 6 – Gráfico de Trabalhos em Eventos.	30
Figura 7 – Gráfico de Artigos.	30
Figura 8 – Gráfico de Participações em Bancas.	31
Figura 9 – Gráfico de Orientações Completas.	32
Figura 10 – Gráfico de Orientações em Progresso.	32
Figura 11 – Gráfico de Participações em Projetos.	33
Figura 12 – Gráfico de Trabalhos em Eventos de 2010 a 2020.	33
Figura 13 – Gráfico específico de Trabalhos em Eventos.	34
Figura 14 – Gráfico específico de Artigos.	35
Figura 15 – Gráfico específico de Participações em Bancas.	35
Figura 16 – Gráfico específico de Orientações Completas.	36
Figura 17 – Gráfico específico de Orientações em Progresso.	36
Figura 18 – Gráfico específico de Participações em Bancas.	37
Figura 19 – Nuvem de Palavras Geral de todos os pesquisadores do PPGCC..	38
Figura 20 – Nuvem de Palavras Específica do professor Reginaldo.	39

LISTA DE TABELAS

Tabela 1 – Tabela de Requisitos Funcionais.	18
Tabela 2 – Tabela de Requisitos Não-Funcionais.	19

LISTA DE ABREVIATURAS E SIGLAS

CNPq	Conselho Nacional de Desenvolvimento Científico
XML	eXtensible Markup Language
FUNDEP	Fundação de Apoio da UFMG
PPGCC	Programa de Pós-Graduação em Ciência da Computação
HTML	HyperText Markup Language
UFPA	Universidade Federal do Pará
UFOP	Universidade Federal de Ouro Preto

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Motivação	13
1.2	Justificativa	14
1.3	Objetivos	14
1.3.1	Objetivo Geral	14
1.3.2	Objetivos Específicos	14
1.4	Trabalhos Relacionados	15
2	METODOLOGIA	17
2.1	Requisitos	17
2.1.1	Requisitos funcionais	17
2.1.2	Requisitos não-funcionais	19
2.2	Diagrama UML	20
2.3	Tecnologias Utilizadas	21
2.3.1	Bibliotecas do Python	22
3	CODIFICAÇÃO DA APLICAÇÃO <i>FASTLATTES</i>	23
4	RESULTADOS	26
4.1	Apresentação do <i>FastLattes</i>	26
4.2	Estudo de Caso	28
4.2.1	Base de Dados	28
4.2.2	Experimentos	29
4.2.2.1	Dados Gerais	29
4.2.2.2	Dados Específicos	33
4.2.2.3	Nuvens de palavras	37
5	CONCLUSÃO	40
	REFERÊNCIAS	41

1 INTRODUÇÃO

A plataforma Lattes foi desenvolvida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e se mostra uma importante base de dados brasileira de currículos de grupos de pesquisa e instituições compiladas em um único sistema de informações. O currículo Lattes se tornou um padrão nacional para documentação de pesquisas científicas e tecnológicas de diversos estudantes e pesquisadores pelo País. O padrão Currículo Lattes surgiu em agosto de 1999, desde então, o mesmo vem sendo utilizado em diversas universidades, institutos, centros de pesquisa e fundações de amparo à pesquisa de diversos estados como uma ferramenta de avaliação de pesquisadores, professores e alunos (CNPQ, 2022a). Para utilizar a plataforma Lattes, deve-se acessar o site oficial da mesma, selecionar a opção de “Buscar currículo”, inserir o nome do responsável pelo currículo a ser acessado, e visualizar o currículo, assim, é possível visualizar tanto as informações do pesquisador no formato de página *web*, quanto realizar o *download* do arquivo que contém tais informações no formato *eXtensible Markup Language* (XML). Entretanto um dos problemas que surge com esse *download* é a existência de *Recaptchas* (GOOGLE, 2022), que geralmente são utilizadas para confirmar que o indivíduo que está realizando aquela solicitação de *download* é um ser humano mesmo, e não um robô, algo que atrapalha bastante a automatização dessa tarefa.

A plataforma Lattes, no dia 07 de dezembro de 2022, apresenta um número total de 7.093.754 currículos cadastrados (CNPQ, 2022b), possuindo uma base de dados bem extensa. Além disso, os dados armazenados estão na forma textual, ou seja, por exemplo, projetos de pesquisa apresentam um texto contendo o período de duração, título e descrição do projeto, e participantes do mesmo, dessa forma, caso haja o interesse de contabilizar o número de projetos presentes no currículo Lattes, isso deve ser feito de maneira manual, verificando os projetos um por um. O trabalho manual para extrair dados de currículos Lattes, somados a grande quantidade de currículos existentes, devido à padronização do mesmo em território nacional, se mostra uma tarefa bem trabalhosa e cansativa, principalmente quando há a necessidade de analisar a produção técnica e científica de múltiplos pesquisadores. Ademais, a coleta de informações dentro desses Lattes se mostra uma tarefa de suma importância para analisar os dados referentes a pesquisadores de um grupo ou uma instituição.

Algo comum de programas de pós-graduação no Brasil, é que os mesmos passem por um processo de avaliação interno e externo de tempos em tempos. Estas avaliações consistem em contabilizar um conjunto de estatísticas quantitativas que tentarão capturar a qualidade do programa, considerando um período de tempo de normalmente quatro anos. Gestores acadêmicos e pesquisadores precisam manualmente contabilizar as suas produções e publicações científicas, participações em projetos de pesquisa, colaborações nacionais e internacionais, orientações de discentes em todo o espectro *stricto sensu*, entre outras. Devido o processo ser trabalhoso e demandar bastante tempo, uma ferramenta que fosse capaz de extrair boa parte destas estatísticas,

de maneira automatizada, seria de grande valia para acelerar o cumprimento do trabalho e evitar a contabilização de erros manuais. É neste contexto que o sistema web que será descrito neste trabalho de conclusão de curso será abordado.

1.1 Motivação

Como mencionado anteriormente, Muitas instituições, em especial programas de pós-graduação de faculdades públicas, frequentemente tem a necessidade de realizar avaliações da produção científica e técnica de seus pesquisadores, tanto de maneira individual quanto conjunta, de todos os mesmos, ou grupos específicos. Entretanto, atualmente a plataforma Lattes não possui uma interface amigável para a coleta e análise de dados quantitativos do currículos existentes, havendo a necessidade de tal tarefa ser realizada de maneira manual, algo bem trabalhoso, além dos resultados estarem fadados a serem afetados pelo erro humano. Ademais, com a necessidade da avaliação de múltiplos currículos de maneira conjunta, ainda surge a questão de que a compilação dos dados, seja em tabelas ou gráficos, gera uma camada a mais de dificuldade. Dessa forma, existe um número considerável de trabalhos feitos em uma tentativa de automatizar tal tarefa, entretanto, com a grande diversidade de ferramentas desenvolvidas existentes, cada uma possui pontos negativos que afetam a usabilidade para o usuário e a disponibilidade da mesma. Algo que afetou algumas ferramentas mais antigas foi a adoção de *Recaptchas* pela plataforma Lattes, uma vez que ferramentas como o “*ScriptLattes*” (MENA-CHALCO; JUNIOR, 2009) e “*LattesMiner*” (ALVES; YANASSE; SOMA, 2012) eram baseadas na automatização do *download* de currículos Lattes no formato *HyperText Markup Language* (HTML), e a adoção de *Recaptchas* inviabilizou o funcionamento de tais ferramentas do jeito que estavam.

Outro ponto negativo de algumas aplicações dessa área, são a presença de necessidade da intervenção manual do usuário em tarefas que poderiam ser automatizadas. Como as ferramentas “*lucyLattes*” (TIEPPO, 2019) e “*LattesMiner*”, aos quais requerem que o usuário insira manualmente os nomes dos pesquisadores, aos quais terão seus currículos Lattes analisados.

Algo que faz falta em certos trabalhos, como no “*lucyLattes*” e no artigo de Oliveira e Silva (OLIVEIRA; SILVA, 2021), são a utilização de representação visuais, como gráficos, dos dados coletados e compilados, algo que não vai impedir o uso da ferramenta, entretanto, afeta de maneira negativa a experiência de quem utiliza as mesmas.

Ademais, poucos são os trabalhos que disponibilizam, a partir de um grupo de currículos Lattes, dados coletivos e individuais dos mesmos, uma funcionalidade extremamente útil para dar flexibilidade de uso à ferramenta. Alguns trabalhos que apresentam tal função são a “Somos UFMG” (VASCONCELOS et al., 2017) e a “UFOP Ativa” (LANA, 2019), entretanto, a “Somos UFMG” é uma ferramenta paga, enquanto que a “UFOP Ativa” foi feita pensando no uso apenas de um grupo seletivo de pesquisadores.

Dessa forma, existe a necessidade do desenvolvimento de uma ferramenta que englobe a

maior parte dos pontos positivos de outros trabalhos e minimize a existência dos pontos negativos mencionados.

1.2 Justificativa

O *FastLattes* é uma ferramenta que: realiza a coleta e compilação dos dados de um conjunto de currículos Lattes, gerando e apresentando gráficos em um *Dashboard* para facilitar a visualização do usuário; automatiza o máximo de tarefas possíveis para evitar o trabalho manual de quem utilizar a aplicação; apresenta relatórios tanto sobre os dados gerais de todos os currículos presentes no sistema, ou dados específicos de um pesquisador; e pode ser utilizada por qualquer pessoa que tenha interesse, por ter código aberto e ser gratuita. Dessa forma, a ferramenta seria de grande auxílio para a área acadêmica, uma vez que a mesma reduziria esforços relacionados a coleta e organização de dados de currículos Lattes e traria resultados mais precisos.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral deste trabalho de conclusão de curso consiste em apresentar o desenvolvimento de uma aplicação em *Python* que, sendo rodada localmente, tenha capacidade de realizar uma análise automatizada de dados referentes a produções científicas, por parte de múltiplos docentes ou pesquisadores de faculdades, contidos em arquivos do formato XML referentes à currículos Lattes.

1.3.2 Objetivos Específicos

Já os objetivos específicos deste trabalho são:

- Realizar um Levantamento bibliográfico acerca de ferramentas já existentes;
- Desenvolver uma aplicação web capaz de lidar com as bases de dados da plataforma Lattes;
- Selecionar bibliotecas do *Python* para auxiliar o desenvolvimento da aplicação;
- Desenvolver a aplicação de maneira “genérica” para que possa ser utilizada por qualquer indivíduo/instituição que possua interesse;
- Desenvolver a visualização da informação por intermédio de gráficos;
- Disponibilizar publicamente o código-fonte;

- Realizar um estudo de caso com os pesquisadores do programa de pós-graduação em ciência da computação da UFPA.

1.4 Trabalhos Relacionados

Através de uma levantamento bibliográfico, foi possível observar alguns trabalhos existentes na área de extração de dados de currículos Lattes, cada um com suas diferenças. Dessa forma, foram selecionados trabalhos e ferramentas relacionadas que serão brevemente abordadas a seguir.

No trabalho de Oliveira e Silva (OLIVEIRA; SILVA, 2021), foi desenvolvido um protótipo de Extrator de produções acadêmicas, em *Python*, onde o foco gira em torno de calcular a taxa de produção científica do Campus Tianguá no Instituto Federal de Educação, Ciência e Tecnologia do Ceará. Tal taxa de produção nada mais seria do que o número de produções realizadas dividido pelo número total de pesquisadores. Além disso, o protótipo desenvolvido no trabalho possui a funcionalidade de baixar automaticamente os currículos Lattes, entretanto, devido ao fato da necessidade de resolver a *Recaptcha*, há um custo monetário atrelado a esse *download* automático. Além disso, os dados apresentados são bem gerais, apresentando apenas números brutos categorizados, sem informação sobre o ano das produções ou qualquer gráfico para facilitar a visualização dos mesmos. Ademais, vale ressaltar que o mesmo é apresentado como uma ferramenta incompleta, justamente por ser um protótipo inicial.

Outro trabalho encontrado foi o da ferramenta “lucyLattes” em *Python* (TIEPPO, 2019), a qual possui a capacidade de realizar a extração, organização e geração de um relatório com gráficos, referentes aos dados contidos em currículos Lattes. o *download* dos currículos é feito de maneira manual para serem inseridos na ferramenta. Entretanto, existem alguns inconvenientes nesta ferramenta, como a necessidade de adicionar manualmente os IDs e nomes dos currículos em um arquivo de texto. Além disso, o usuário não pode interagir com o resultado final, pois é gerado apenas um relatório estático.

Outra ferramenta desenvolvida, a “Somos UFMG” (VASCONCELOS et al., 2017), se mostra uma das aplicações mais completas para a extração de informação de currículos Lattes, possuindo uma interface interativa e simples, apresentando diversos dados relevantes em relação a produção acadêmica de pesquisadores. Entretanto, a ferramenta é paga, e para ser utilizada, necessita que a instituição interessada entre em contato com a Fundação de Apoio da UFMG (FUNDEP) para obter o direito de uso do sistema, dificultado a acessibilidade da mesma.

Uma das ferramentas mais famosas desenvolvidas, ao qual tem uma função semelhante a desde trabalho, foi a “*ScriptLattes*” (MENA-CHALCO; JUNIOR, 2009). Tal ferramenta foi desenvolvida em 2009, e possui diversas funcionalidades, como o *download* automático de currículos Lattes, apresentando dados quantitativos de trabalhos científicos, gráficos de barra com o número de trabalhos por ano, mostrando a localização geográfica dos pesquisadores

através da integração com o *Google Maps*, e grafos referentes a colaboração entre os pesquisadores presentes. Porém, além da ferramenta não apresentar dados quantitativos específicos de pesquisadores, a mesma foi afetada pela implementação de *Recaptchas* para o *download* de currículos Lattes, parando de funcionar no estado em que estava.

Mais uma aplicação desenvolvida foi o “*LattesMiner*” (ALVES; YANASSE; SOMA, 2012), a qual de maneira similar a outras ferramentas, também extrai dados gerais de um grupo de pesquisadores através dos currículos Lates. De maneira similar ao “*ScriptLattes*”, a ferramenta também realiza o *download*, de maneira automática, dos currículos Lattes de um grupo de pesquisadores, onde os dados extraídos são armazenados em um banco de dados que pode ser acessado pelo usuário. Entretanto, um dos pontos negativos do trabalho, se da pelo fato de que não há nenhuma menção de gráficos, que sejam gerados pela ferramenta, para facilitar a visualização dos dados extraídos. Ademais, este trabalho não faz nenhuma referência a *Recaptchas*, o que leva a entender que o mesmo foi desenvolvido antes da adoção das mesmas na plataforma Lattes, e conseqüentemente, teve o seu funcionamento afetado, assim como o “*ScriptLattes*”.

Outro sistema, o “UFOP Ativa” (LANA, 2019), foi desenvolvido com o objetivo de extrair e organizar dados do currículo Lattes de servidores da Universidade Federal de Ouro Preto (UFOP), onde os dados eram armazenados em um banco de dados, e apresentados através de um *Dashboard* em uma página *Web*, utilizando gráficos e textos referentes tanto ao conjunto de currículos de todos os servidores, quanto a servidores específicos. Ademais, os dados eram coletados e atualizados automaticamente pelo sistema, entretanto, isso só foi possível devido ao autor do trabalho haver solicitado, à presidência do CNPq, o acesso aos dados da plataforma Lattes. É notável que este é um trabalho bem completo, tanto no quesito de automatização, quanto de funcionalidades, entretanto, o ponto negativo do mesmo é que ele está limitado apenas para o uso da UFOP, sem que a ferramenta esteja disponível para outras instituições interessadas.

Mais um trabalho encontrado foi o da ferramenta “*LattesPlan*” (VASCONCELOS, 2019), ao qual de maneira semelhante as outras, também é uma ferramenta para a extração de dados de currículos Lattes, apresentando principalmente dados de classificação da Qualis. Entretanto, a mesma possui a desvantagem de gerar apenas planilhas contendo as informações extraídas, sem utilizar de recursos visuais melhores, como gráficos, tornando mais difícil a visualização por parte do usuário.

Por fim, uma ferramenta que já está bem estabelecida no mercado é a “*Stela Experta*” (STELATEK, 2005), a mesma permite a avaliação da performance de diversos pesquisadores, realizando a comparação dos mesmos com a média de programas de pós-graduação por todo o país, dessa forma, é possível fazer um monitoramento constante da qualidade de um programa de pós-graduação específico. Entretanto, apesar de ser uma ferramenta bem completa e automatizada, o uso da mesma não é de fácil acesso, devido a sua utilização estar atrelada a um custo monetário relativamente alto.

2 METODOLOGIA

Neste capítulo, será descrito com mais detalhes o processo de desenvolvimento do projeto da aplicação, onde foram determinados requisitos funcionais e não-funcionais, assim como um diagrama de casos de uso UML, para direcionar o desenvolvimento da aplicação *FastLattes*. Ademais, também serão apresentadas as tecnologias utilizadas para auxiliar na codificação da ferramenta, que por fim, foi utilizada em testes experimentais com uma base de dados englobando o currículo Lattes de professores que compõem os docentes do Programa de Pós-Graduação em Ciência da Computação (PPGCC).

2.1 Requisitos

Nesta seção do trabalho, serão apresentadas duas listas de Requisitos definidas para o desenvolvimento do sistema de extração de dados de currículos Lattes, uma de requisitos funcionais na Tabela 1, e outra de requisitos não-funcionais na Tabela 2.

2.1.1 Requisitos funcionais

Os requisitos funcionais tem como objetivo definir quais as funcionalidades, que serão utilizadas pelo usuário, um sistema deve ter. Além disso, para cada requisito funcional, também está atrelada uma prioridade, que define o quão importante a funcionalidade é para o sistema, e também um status de produção, que define se a funcionalidade foi aprovada para ser desenvolvida, se está em andamento, ou se foi descartada.

Na Tabela 1, foram definidos um total de 10 requisitos funcionais, aos quais estão numerados de RF 01 até RF 10. Ademais, também foram definidas as suas prioridades para que requisitos mais urgentes tenham maior atenção na hora da implementação, por exemplo, o RF 01 define que “Deve-se permitir a inserção de novos pesquisadores”, ou seja, o usuário que está utilizando o sistema deve conseguir adicionar os currículos Lattes que tenha baixado, para que possa se obter os dados referentes a pesquisadores de interesse, o mesmo está definido com prioridade alta, pois essa funcionalidade é a base do funcionamento do sistema. Outros requisitos funcionais importantes são o RF 03 e RF 04, pois os mesmos determinam que o sistema deve gerar gráficos para a visualização dos dados de interesse, tanto para o grupo de pesquisadores, quanto para um pesquisador específico, sendo mais uma funcionalidade que define o sistema como um todo. Ademais, também existem requisitos menos relevantes, como o RF 07 e RF 08, pois, apesar de definirem funcionalidades que possuem grande relevância para o sistema, que nesse caso é a geração de Nuvens de Palavras, o sistema ainda cumpriria seu objetivo principal caso não fossem implementadas.

Requisitos Funcionais			
Nº	Descrição	Prioridade	Status
RF 01	Deve-se permitir a inserção de novos pesquisadores	Alta	Aprovado
RF 02	Deve-se permitir a deleção de pesquisadores existentes	Média	Aprovado
RF 03	Deve-se gerar gráficos referente a dados gerais quantitativos, por ano, do Lattes de pesquisadores, como: <ul style="list-style-type: none"> • Trabalhos em eventos; • Artigos; • Participações em Bancas (Mestrado, Doutorado, Exame de qualificação e Graduação); • Orientações Concluídas e Em Progresso; • Projetos (Pesquisa, Extensão, Ensino, Desenvolvimento e Outros). 	Alta	Aprovado
RF 04	Deve-se gerar gráficos referente a dados apresentados no RF 03, para cada pesquisador cadastrado no sistema.	Alta	Aprovado
RF 05	Deve-se apresentar dados pessoais, caso estejam presente no currículo Lattes, dos pesquisadores, como: <ul style="list-style-type: none"> • Nome; • Endereço institucional; • ID Orcid. 	Alta	Aprovado
RF 06	Deve-se permitir que o usuário possa alterar o intervalo de anos dos trabalhos apresentados nos gráficos gerados pelo RF 03 e RF 04	Alta	Aprovado
RF 07	Deve-se gerar uma Nuvem de Palavras referente aos títulos de Trabalhos em eventos e Artigos de todos os pesquisadores	Média	Aprovado
RF 08	Deve-se gerar uma Nuvem de Palavras referente aos títulos de Trabalhos em eventos e Artigos de um pesquisador específico	Média	Aprovado
RF 09	Deve-se permitir que o usuário escolha se quer gerar, ou não, a Nuvem de Palavras no <i>Dashboard</i>	Média	Aprovado
RF 10	O usuário deve poder alterar a página dos gráficos sendo gerados (Todos os pesquisadores, ou um pesquisador específico)	Alta	Aprovado

Tabela 1 – Tabela de Requisitos Funcionais.

2.1.2 Requisitos não-funcionais

Requisitos não-funcionais estão relacionados com aspectos da aplicação na questão de usabilidade, desempenho, disponibilidade, entre outros. Diferentemente de requisitos funcionais, aos quais definem funcionalidades de maneira direta, requisitos não-funcionais abordam aspectos de qualidade do sistema para o usuário. Apesar de não estarem diretamente relacionados com funcionalidades, os mesmos são de suma importância para direcionar como a aplicação será desenvolvida.

Foram definidos 7 requisitos não-funcionais numerados de RNF 01 até RNF 07, para a ferramenta *FastLattes*. Tais requisitos, de maneira similar aos funcionais, possuem prioridades atreladas aos mesmos, por exemplo, O RNF 04 define que o sistema “Deve apresentar um *Dashboard* com todos os gráficos gerados, permitindo que haja uma interação do usuário com o mesmo”, ou seja, todos os gráficos gerados pela ferramenta devem estar compilados em um *Dashboard* interativo que facilite a utilização por parte do usuário, algo importante a ser considerado. Outros requisitos não-funcionais são definidos com prioridade “média”, dessa forma, possuem grande importância para o sistema, mais ainda existem requisitos mais relevantes, como é o caso do RNF 03, em que diz que o sistema “Deve seguir as melhores práticas de usabilidade de software”, em outras palavras, a aplicação deve ser desenvolvida da melhor forma possível para facilitar o uso da mesma para o usuário final. Por fim, foi definido um requisito não-funcional com prioridade “baixa”, que foi o RNF 02, onde define que o sistema “Deve ser, em sua maioria, em PT-BR”, algo importante, devido ao grande uso do currículo Lattes no Brasil, entretanto, acaba ficando atrás, em comparação com outros requisitos não-funcionais, no quesito de relevância.

Requisitos Não-Funcionais			
Nº	Descrição	Prioridade	Status
RNF 01	Deve ser otimizado para funcionar em máquinas com menos poder de processamento	Média	Aprovado
RNF 02	Deve ser, em sua maioria, em PT-BR	Baixa	Aprovado
RNF 03	Deve seguir as melhores práticas de usabilidade de software.	Média	Aprovado
RNF 04	Deve apresentar um <i>Dashboard</i> com todos os gráficos gerados, permitindo que haja uma interação do usuário com o mesmo	Alta	Aprovado
RNF 05	Deve funcionar em qualquer sistema operacional	Média	Aprovado
RNF 06	Deve funcionar em qualquer navegador web	Alta	Aprovado
RNF 07	Deve ser desenvolvido na linguagem <i>Python</i>	Alta	Aprovado

Tabela 2 – Tabela de Requisitos Não-Funcionais.

2.2 Diagrama UML

Para representar as funcionalidades disponíveis para o usuário no sistema, foi criado um diagrama de casos de uso, que tem como objetivo facilitar a visualização das maneiras com que o usuário pode interagir com o sistema *FastLattes*. Foi utilizado o programa “Astar UML” para o desenvolvimento do diagrama, que pode ser observado na Figura 1. Foram ilustradas 8 funcionalidades no diagrama, onde algumas possuem relações específicas entre si por exemplo, o caso de uso “Adicionar Pesquisador”, que determina possibilidade de poder adicionar múltiplos novos currículos Lattes dentro do sistema através de um mecanismo de Upload de arquivos zip, possui uma relação “*extend*” com o caso de uso “Alterar Página a ser visualizada”, ao qual pode ser descrita como a funcionalidade de alterar a página atual, podendo selecionar mostrar dados gerais de todos os currículos Lattes, ou de um currículo específico. Ou seja, a relação entre os dois casos de uso determina que, caso seja adicionado um pesquisador no sistema, o usuário poderá alterar a página sendo visualizada.

Outra relação seria a “*include*”, onde está presente entre os casos de uso “Adicionar Pesquisador” e “Visualizar Gráficos”, o último é definido como a possibilidade do usuário poder visualizar gráficos referentes ao número de trabalhos e atividades acadêmicas por ano, de todos os currículos Lattes juntos, ou de um currículo específico. Dessa forma, a relação entre os dois implica que, ao adicionar um pesquisador, obrigatoriamente o usuário irá visualizar gráficos gerais, levando em conta a inclusão do novo pesquisador no sistema.

Mais uma relação seria entre os casos de uso “Ativar/Desativar a geração da Nuvem de Palavras”, que é descrito como a capacidade do usuário poder ativar ou desativar a funcionalidade do sistema de gerar uma nuvem de palavras, e “Visualizar Nuvem de Palavras”, na qual significa que o usuário deve poder visualizar uma nuvem de palavras baseada no título de trabalhos e artigos de todos os currículos Lattes, ou de um currículo específico. Pode-se entender, pela ligação “*extend*” entre os dois, que caso o usuário resolva ativar ou desativar a geração das nuvens de palavras, o usuário poderá visualizar a mesma no *Dashboard* do sistema.

Um caso de uso que não possui nenhuma relação com outro é o “Modificar Intervalo de Anos”, ao qual define a funcionalidade de alterar o intervalo de anos dos trabalhos científicos que são levados em consideração nos gráficos gerados. Essa ausência de relações ocorre pois, não importa a forma como o usuário interaja com a ferramenta, a opção de modificar o intervalo de anos sempre vai estar disponível, entretanto, as vezes pode ser que não cause mudanças perceptíveis pelo usuário nos resultados.

Dessa forma, tal diagrama de casos de uso auxilia no entendimento de como o usuário pode utilizar a ferramenta, além de como funcionam as interações entre as funcionalidades presentes na mesma, sendo uma forma simplificada de representar o sistema para quem tenha interesse.

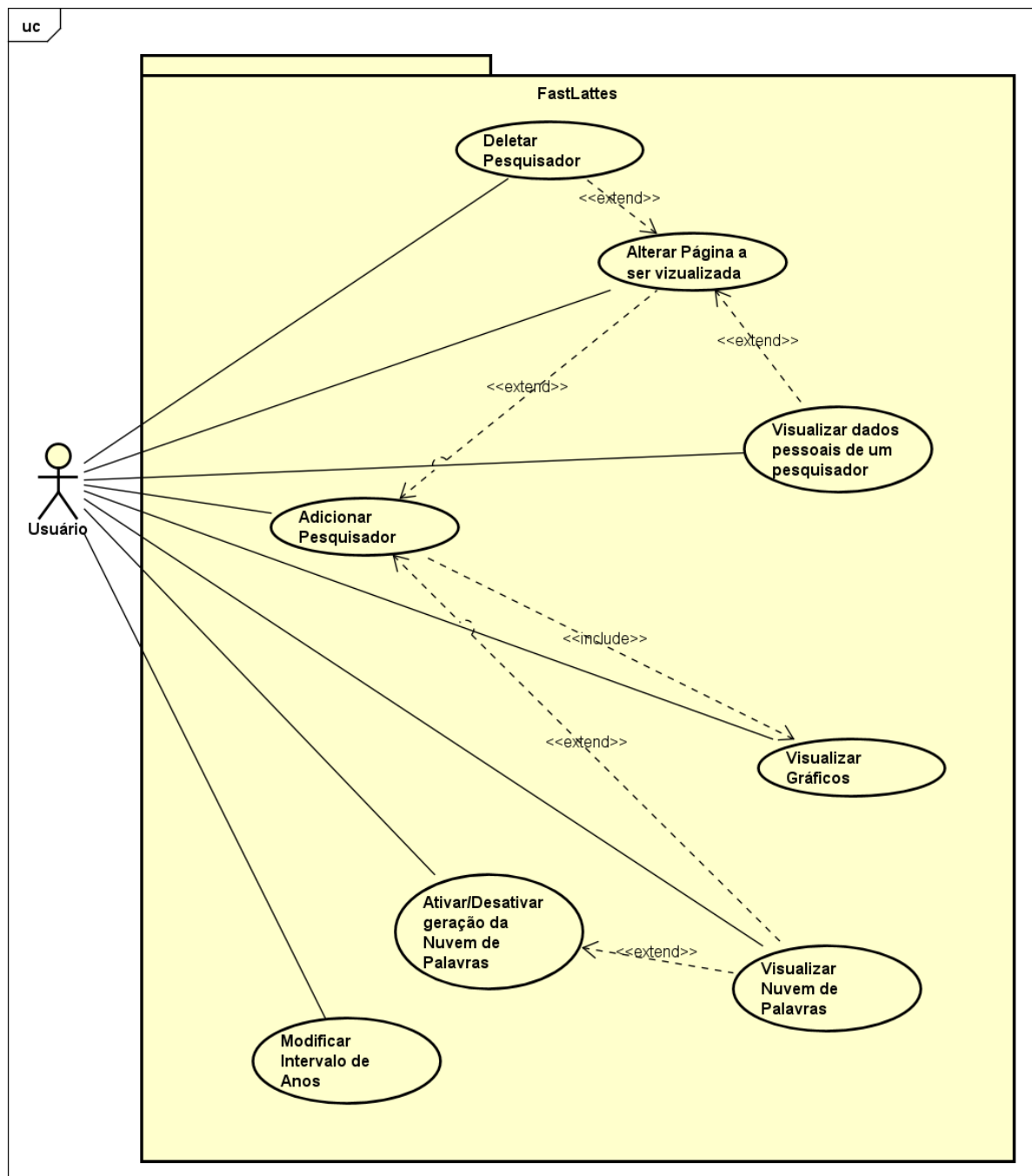


Figura 1 – Diagrama de Casos de Uso do Sistema.

2.3 Tecnologias Utilizadas

Nesta seção, serão descritas as tecnologias utilizadas para a implementação prática da primeira versão do *FastLattes*. Como mencionado anteriormente, o desenvolvimento da ferramenta foi feito 100% na linguagem de programação *Python 3*, ademais, foram utilizadas algumas bibliotecas externas do *Python* que serão explicadas mais para frente, são elas: *Matplotlib*, *Numpy*, *Streamlit*, *Wordcloud* e *Nltk*.

2.3.1 Bibliotecas do Python

Para poder rodar a ferramenta localmente, é de suma importância que as bibliotecas externas estejam disponíveis no projeto, isso por ser feito através do *pip*, que nada mais é do que um sistema de gerenciamento de pacotes no *Python*, dessa forma, as bibliotecas podem ser instaladas utilizando o comando `pip install -r requirements.txt`. O arquivo `requirements.txt` é um arquivo de texto onde encontra-se todos os módulos do *Python* que precisam ser instalados para que a máquina possa rodar a ferramenta.

Primeiramente, a biblioteca *Matplotlib 3.5.3* foi utilizada para a geração de gráficos a serem mostrados no *Dashboard* da ferramenta, pois a mesma tem a capacidade de simplificar todo esse processo de compilação de dados coletados dos currículos Lattes em visualizações gráficas. Outra biblioteca utilizada foi a biblioteca *Numpy 1.23.2*, a qual foi utilizada pois a mesma permite uma flexibilidade maior na manipulação de listas, em comparação às listas nativas do *Python*

A biblioteca *Streamlit 1.12.2* foi o que possibilitou o desenvolvimento do projeto 100% em *Python*, uma vez que a mesma permite criar e rodar um *Webapp* localmente. Assim, a utilização de tal biblioteca foi feita para gerar um *Dashboard* que possa ser aberto em um navegador *Web* qualquer, para apresentar os gráficos gerados pelo *Matplotlib*, além de permitir a implementação de outras funcionalidades mencionadas anteriormente, como o *Upload* de múltiplos currículos Lattes simultaneamente ao sistema.

Por fim, as bibliotecas *Wordcloud 1.8.2.2* e *Nltk 3.7* foram utilizadas de maneira conjunta para gerar as nuvens de palavras dos currículos Lattes no sistema. A *Wordcloud* possui uma função que gera automaticamente uma imagem, a partir de uma lista de palavras contendo os títulos dos trabalhos científicos de cada pesquisador, onde palavras mais frequentes em um texto são representadas por um tamanho maior, e as menos frequentes são mostradas com um tamanho menor, assim, é possível ter uma representação gráfica do perfil dos trabalhos de um pesquisador ou um grupo de pesquisadores. A *Nltk* forneceu a biblioteca *Wordcloud* uma lista de *StopWords*, ou seja, um conjunto de palavras e símbolos que não devem ser levados em consideração na hora da construção da nuvem de palavras, como por exemplo “de”, “para”, “and”, entre outras. Ou seja, a biblioteca *Nltk* auxilia a remoção de tais palavras e símbolos da imagem final gerada.

Ademais, vale ressaltar algumas bibliotecas que, apesar de serem nativas do *Python*, foram de grande relevância para o projeto, como a *xml.etree.ElementTree*, a qual foi utilizada para buscar dados relevantes nos currículos Lattes em formato XML. Outra biblioteca nativa é a *zipfile*, que serve para descompactar arquivos *zip*, o formato inicial padrão ao baixar currículos da plataforma Lattes.

3 CODIFICAÇÃO DA APLICAÇÃO *FASTLATTES*

Durante a implementação da ferramenta *FastLattes*, foram criados 3 arquivos principais para rodar o sistema: `researcher.py`, `functions.py`, e `main.py`. Ademais, também foi criado um diretório de arquivos `lattes` para armazenar arquivos `zip` dos currículos Lattes cadastrados no sistema.

Primeiramente, a classe `Researcher` está codificada no arquivo `researcher.py`, na mesma, são coletados dados relevantes do arquivo XML referente ao currículo Lattes de cada pesquisador, e armazenando os mesmos em variáveis, são eles:

- Nome;
- Nome do arquivo;
- Anos de trabalhos em eventos;
- Anos de artigos;
- Anos de participações em bancas;
- Anos de participações em projetos;
- Anos de orientações completas;
- Anos de orientações em progresso;
- ID Orcid;
- Endereço institucional;
- Instituição;
- Títulos de trabalho.

Outro arquivo presente no repositório é o arquivo `functions.py`, onde estão contidas funções gerais da ferramenta para geração de *plots* de gráficos, nuvens de palavras, inicialização dos pesquisadores, deleção dos pesquisadores, entre outras. Essas funções, na maioria, são responsáveis compilação dos dados brutos em números e representações gráficas mais agradáveis para a visualização por parte do usuário.

O arquivo `main.py` contém a parte do código que realiza toda a estruturação do *Dashboard* através da biblioteca *Streamlit*, com os gráficos, troca de página, intervalo de anos, etc. Além disso, o mesmo chama a função do arquivo `functions.py` para inicializar todos os currículos disponíveis na pasta `lattes`. Dessa forma, para rodar a ferramenta, basta utilizar o comando `streamlit run main.py` no terminal, assim, será aberta uma página no navegador *Web* do usuário, permitindo a interação do mesmo com a aplicação.

Os requisitos funcionais moldaram a codificação de funcionalidades do *Fastlattes*, e todos eles foram implementados no código da ferramenta. No caso de requisitos funcionais, requisitos que envolvem a geração em si de tais gráficos, como o RF 03, RF 04, RF 07 e RF 08, listados na Tabela 1, foram implementados como funções no arquivo `functions.py`, por exemplo, existe a função `generate_works_per_year_graphic`, a qual é uma implementação direta de parte do requisito funcional RF 03, referente a geração de um gráfico de trabalhos em eventos por ano. Outros requisitos que tem relação com a modificação dos currículos armazenados no sistema, como o RF 01 e RF 02, também foram parcialmente implementados no código do `functions.py`,

pois é nele que contém as funções para realizar tais alterações, ademais, todos os requisitos funcionais citados tem relação com a classe `Researcher`, pois é nela que ficam armazenadas as informações dos currículos Lattes de pesquisadores. Entretanto, requisitos funcionais que envolvem uma interação mais direta do usuário com o *dashboard* do sistema, como o RF 05, RF 06, RF 09 e RF 10, ficam a cargo do `main.py`, pois neste arquivo do sistema que o usuário utiliza certas funcionalidades do *dashboard*, como a definição de um intervalo de anos dos trabalhos, a ativação da geração de nuvens de palavras, troca de páginas dos pesquisadores, entre outros.

Sobre os requisitos não-funcionais, listados na Tabela 2, serviram para guiar o desenvolvimento do sistema como um todo, não tendo foco em uma funcionalidade específica. Sobre o RNF 01, a aplicação foi desenvolvida de uma maneira otimizada, de forma que o sistema possa rodar até em *hardwares* mais limitados; no RNF 02, todos os textos que são apresentados no *Dashboard* do sistema foram escritos em português brasileiro; o RNF 03 também foi levado em consideração na hora de codificar o sistema, pois o mesmo deve ser de fácil utilização por parte do usuário, de forma que o *dashboard*, no geral, seja bem simples de visualizar e interagir; o RNF 04, está presente no fato de que toda a parte de geração de gráficos foi codificada para ser apresentada no *dashboard*; o RNF 05 é cumprido juntamente por ser desenvolvido em *python*, uma linguagem que tem suporte para os principais sistemas operacionais, como *Mac*, *Windows* e *Linux*; o RNF 06 é resolvido pelo uso da biblioteca *streamlit*, pois a mesma tem suporte para a maioria dos navegadores existentes; e o RNF 07 se refere a escolha do *python* para o desenvolvimento da ferramenta, pois tal linguagem de programação se apresenta como uma das mais utilizadas na atualidade (TIOBE, 2022).

Por fim, na Figura 2, é possível observar uma representação simplificada do funcionamento da ferramenta. Primeiramente, o usuário irá manipular os dados pelo *dashboard* do sistema *web*, podendo adicionar ou remover pesquisadores da base de currículos, depois, o *dashboard web*, que é executado pelo arquivo `main.py` do projeto, irá solicitar a execução de funções contidas no arquivo `functions.py`, assim, as funções executadas incluem a consulta de currículos Lattes armazenados no diretório `lattes` do sistema e a inicialização dos pesquisadores na classe `Researcher`. ademais, as funções retornam gráficos e números, referentes aos dados coletados dos currículos, para o *dashboard*, e por fim, o *dashboard* apresenta para o usuário os resultados.

O código completo está disponível em um repositório público na plataforma *GitHub*, e pode ser acessado através da seguinte url: <<https://github.com/guihantani/FastLattes>> (HANTANI, 2022)

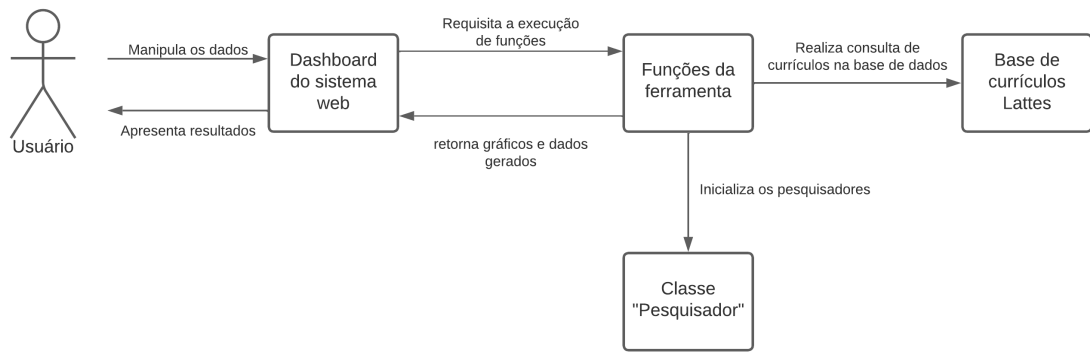


Figura 2 – Funcionamento da estrutura da aplicação.

4 RESULTADOS

4.1 Apresentação do *FastLattes*

Nesta seção, serão apresentados as funcionalidades implementadas no sistema. Primeiramente, na Figura 3, é possível observar a tela inicial da aplicação ao ser rodada pela primeira vez com os currículos Lattes. Na parte a esquerda da imagem, existem um menu com algumas opções de interação aos quais o usuário pode utilizar, como:

- “Selecione a página”: o usuário pode escolher ver os dados gerais referentes a todos os currículos salvos no sistema, ou selecionar uma página específica Figura 4;
- “Ano Mínimo” e “Ano Máximo”: pode ser definido um ano mínimo ou/e um ano máximo para os trabalhos que são levados em consideração nos gráficos e números totais gerados;
- “Gerar Nuvem de Palavras”: esta caixa de seleção define se, no *dashboard* do sistema, será gerada a nuvem de palavras referente a todos os currículos armazenados, ou de um pesquisador específico;
- “Adicione Múltiplos Lattes (Tipo Zip)”: esta opção serve para que o usuário possa adicionar múltiplos arquivos *zip*, referente a currículos Lattes, de uma vez só.

Ademais, ao lado do menu, fica justamente o *dashboard* contendo os dados extraídos e compilados em gráficos gerados pela ferramenta *FastLattes*, aos quais serão demonstrados mais para frente na subseção 4.2.2, onde o usuário pode arrastar a página para baixo e navegar por um total de 7 gráficos gerados, que são:

- “Nuvem de Palavras”: gerada a partir de títulos de periódicos e artigos publicados, caso sua construção esteja ativada;
- “Trabalhos em Eventos”: referente a periódicos publicados, por ano;
- “Artigos”: referente a artigos publicados, por ano;
- “Participações em Bancas”: referente a diferentes tipos de bancas participadas;
- “Orientações Completas”: referente a orientações completas, por ano;
- “Orientações em Progresso”: referente a orientações em progresso, por ano;
- “Participações em Projetos”: referente a diferentes tipos de projetos participados;



Figura 3 – Tela inicial da aplicação.

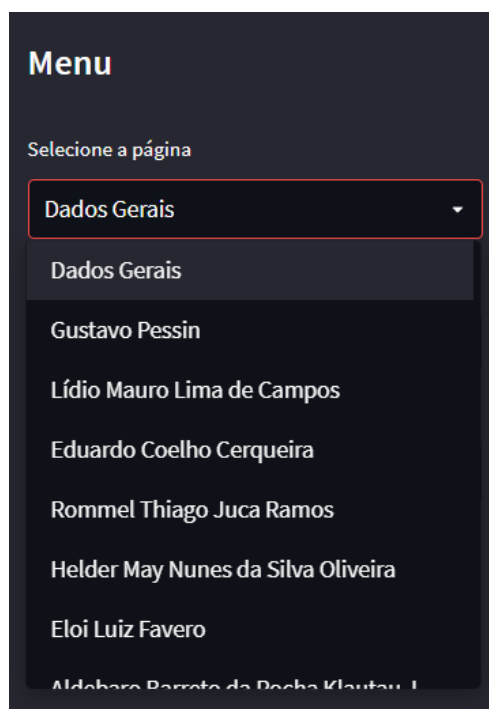


Figura 4 – Menu de seleção de página.

Por fim, é possível acessar a página pesquisador em específico. Primeiramente, ao selecionar a página do pesquisador, é possível observar algumas informações a mais em comparação

com a página de dados gerais, são elas: Nome, ID Orcid, endereço institucional e instituição. Entretanto, a instituição está vazia pois não foi encontrada no arquivo XML, ademais, existe um botão “Deletar Pesquisador”, onde pode ser feito a deleção do mesmo do conjunto de currículos.



Figura 5 – Página específica do pesquisador.

4.2 Estudo de Caso

Nesta seção, serão apresentados alguns estudos de caso do uso da aplicação, através de telas e imagens geradas pelo sistema. Assim como a base de dados utilizada para a realização de experimentos com a ferramenta desenvolvida.

4.2.1 Base de Dados

De acordo com o site (PPGCC, 2022), o Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal do Pará (UFPA), foi fundado em 2005, sendo reconhecido inclusive pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que possui o objetivo de consolidar e expandir a Pós-Graduação no Brasil. Ademais, a CAPES realiza uma avaliação dos programas de pós-graduação por todo o país no quesito de produção científica, dessa forma, neste trabalho foram utilizados o currículo Lattes de 28 professores do PPGCC para o estudo de caso da ferramenta *FastLattes*, aos quais foram baixados

diretamente da plataforma Lattes, no dia 11 de Novembro de 2022. Os professores do PPGCC são os seguintes (PPGCC, 2022):

- Aldebaro Barreto da Rocha Klautau Júnior
- André Figueira Riker
- Antonio Jorge Gomes Abelém
- Bianchi Serique Meiguins
- Carlos Gustavo Resque dos Santos
- Carlos Renato Lisboa Francês
- Claudomiro de Souza de Sales Junior
- Cleidson Ronald Botelho de Souza
- Denis Lima do Rosário
- Eduardo Coelho Cerqueira
- Eloi Luiz Favero
- Fabiola Pantoja Oliveira Araújo
- Filipe de Oliveira Saraiva
- Gustavo Henrique Lima Pinto
- Gustavo Pessin
- Helder May Nunes da Silva Oliveira
- Jefferson Magalhães de Morais
- Joao Crisóstomo Weyl Albuquerque Costa
- Josivaldo de Souza Araújo
- Lídia Mauro Lima de Campos
- Marcelle Pereira Mota
- Nelson Cruz Sampaio Neto
- Reginaldo Cordeiro dos Santos Filho
- Roberto Samarone dos Santos Araújo
- Rommel Thiago Juca Ramos
- Ronnie Cley de Oliveira Alves
- Sandro Ronaldo Bezerra Oliveira
- Vinicius Augusto Carvalho de Abreu

4.2.2 Experimentos

Nesta seção, serão apresentados experimentos realizados utilizando a base de dados mostrada anteriormente, de forma que gráficos e números totais gerados pela ferramenta *FastLattes* serão demonstrados através de *prints* feitos diretamente do *dashboard* da aplicação.

4.2.2.1 Dados Gerais

Primeiramente, foram realizados experimentos referentes aos dados gerais de todos os 28 currículos da base de dados. Sendo o primeiro gráfico gerado, um gráfico de barras sobre “Trabalhos em Eventos” por ano, ou seja, periódicos. Na Figura 6, é possível observar um total de 2480 periódicos nos currículos Lattes do sistema, além disso, o maior número de trabalhos ocorreu em 2019, onde foram encontrados 223 trabalhos em eventos.

Outro gráfico de barras gerado, foi construído a partir de artigos publicados em periódicos,

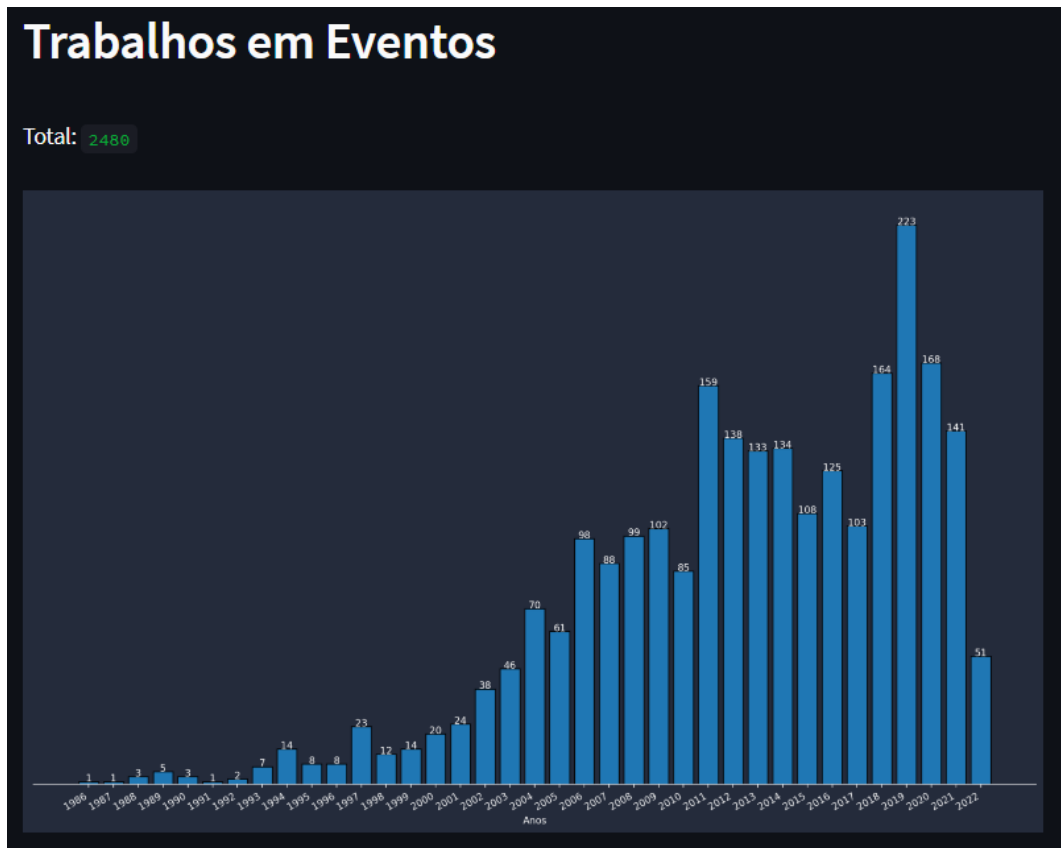


Figura 6 – Gráfico de Trabalhos em Eventos.

assim, na Figura 7, é possível observar um total de 933 artigos encontrados na base de dados, ademais, o ano com maior número de artigos publicados foi 2021, com um total de 103 artigos.

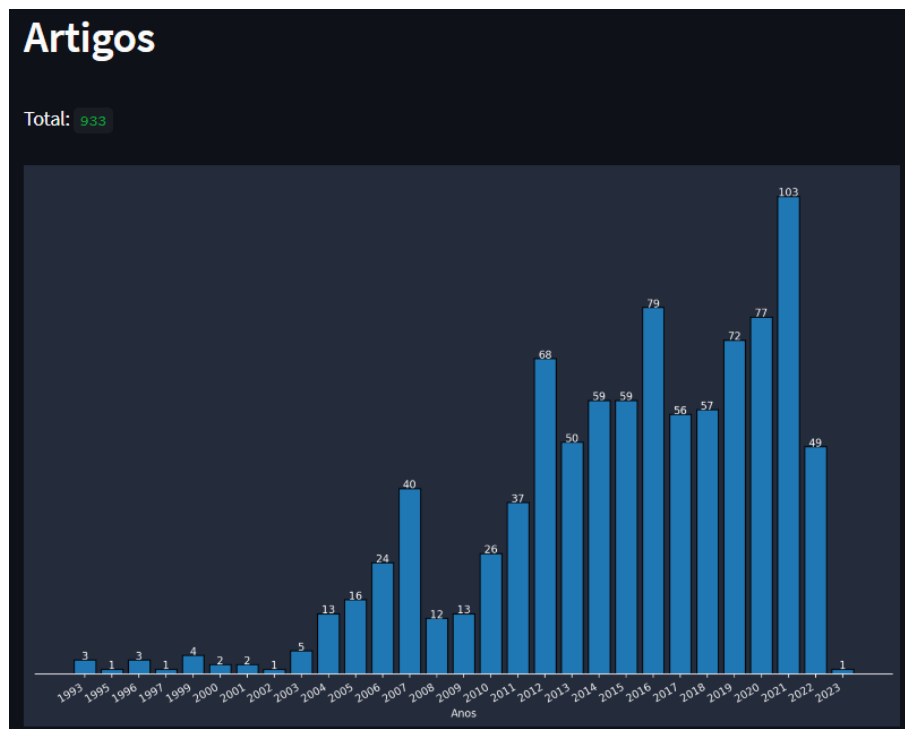


Figura 7 – Gráfico de Artigos.

Mais um gráfico gerado é o gráfico de “Participações em Bancas”, entretanto, diferente dos últimos 2 gráficos apresentados, este gráfico é no formato de pizza, ou seja, é possível observar na Figura 8 o número e porcentagens referentes aos diferentes tipos de Bancas em que os pesquisadores participaram, havendo um total de 1520 bancas participadas, onde o tipo de banca com o maior número de participações por parte dos pesquisadores são as bancas de mestrado, com 518 participações, somando 34,08% de todas as participações.



Figura 8 – Gráfico de Participações em Bancas.

Outro gráfico de barras gerado se refere às “Orientações Completas”, ou seja, orientações realizadas pelos pesquisadores que já foram finalizadas. Na Figura 9, é possível observar um total de 1614 orientações completas, com uma grande concentração no ano de 2019, totalizando 150 orientações completas daquele ano.

Mais um gráfico de barras gerado se refere às “Orientações em Progresso”, que ao contrário das “Orientações Completas”, se refere a orientações que ainda estão ocorrendo até o dia em que o currículo Lattes foi baixado. Na Figura 10, é possível observar que o total de orientações em progresso é de 207, com grande concentração em orientações que iniciaram no ano de 2021.

O último gráfico gerado é semelhante ao de “Participações em Bancas”, entretanto, são apresentadas as “Participações em Projetos”. Na Figura 11, obtém-se um total de 573 participações diferentes em projetos, onde a maior concentração se encontra em “Projetos de Pesquisa”, com um total de 419 participações, referentes a 73,12% das participações totais.

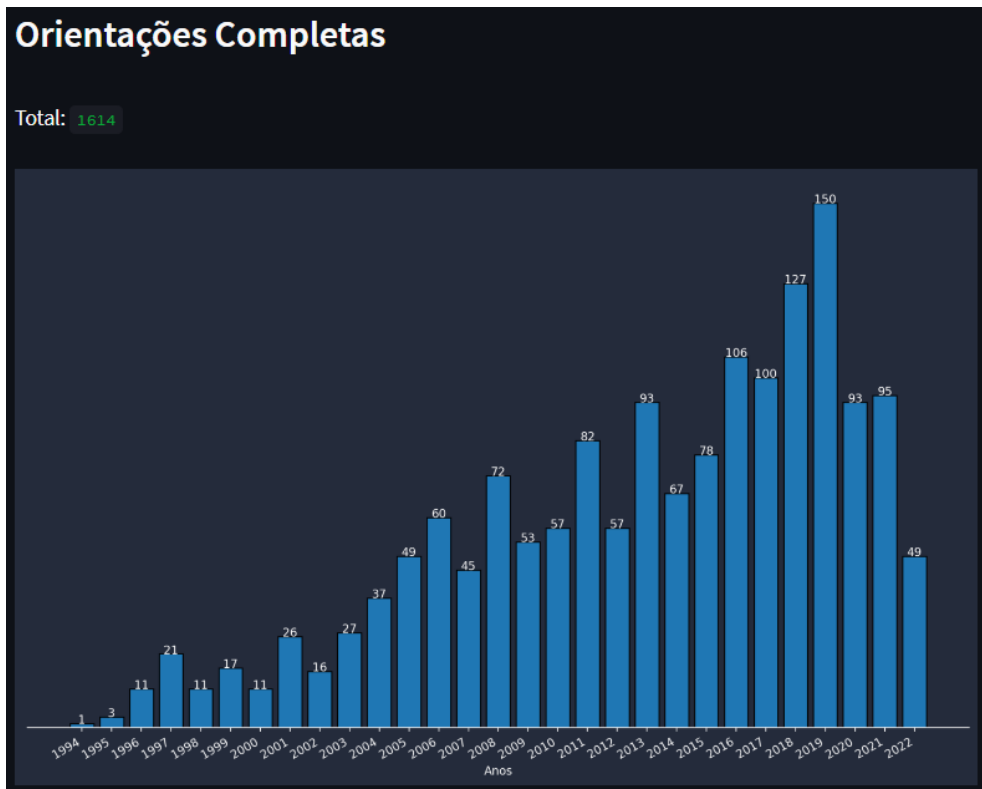


Figura 9 – Gráfico de Orientações Completas.

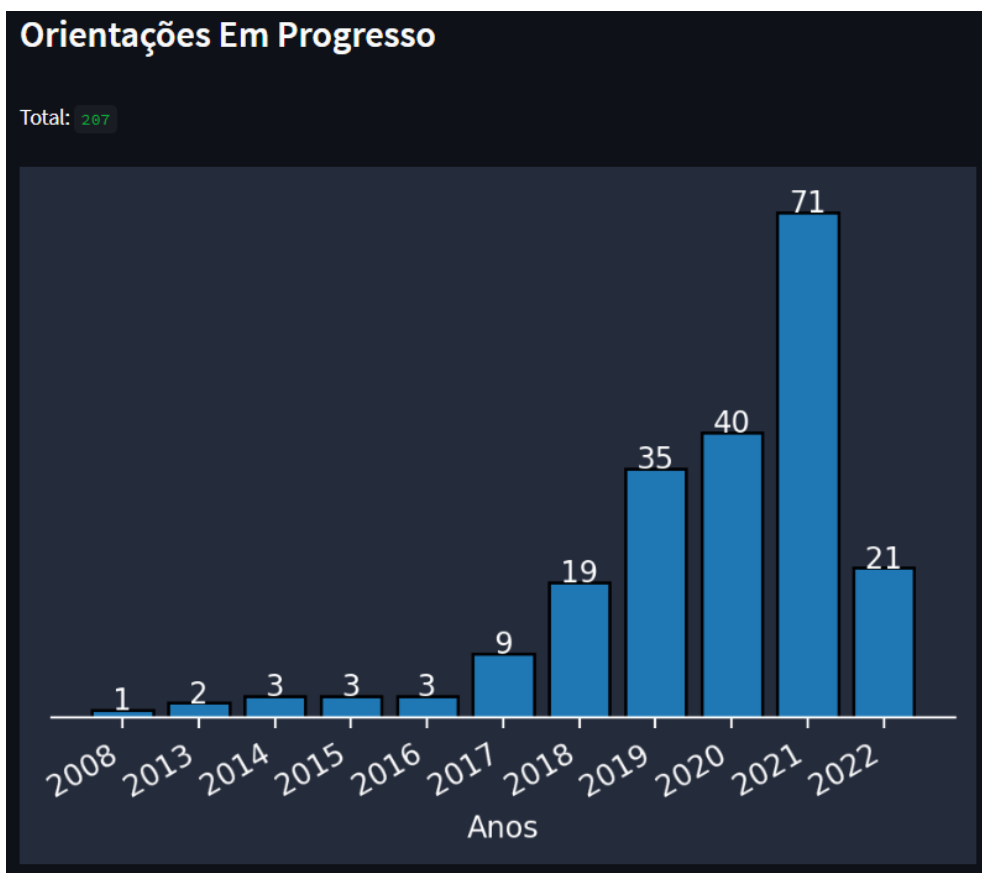


Figura 10 – Gráfico de Orientações em Progresso.

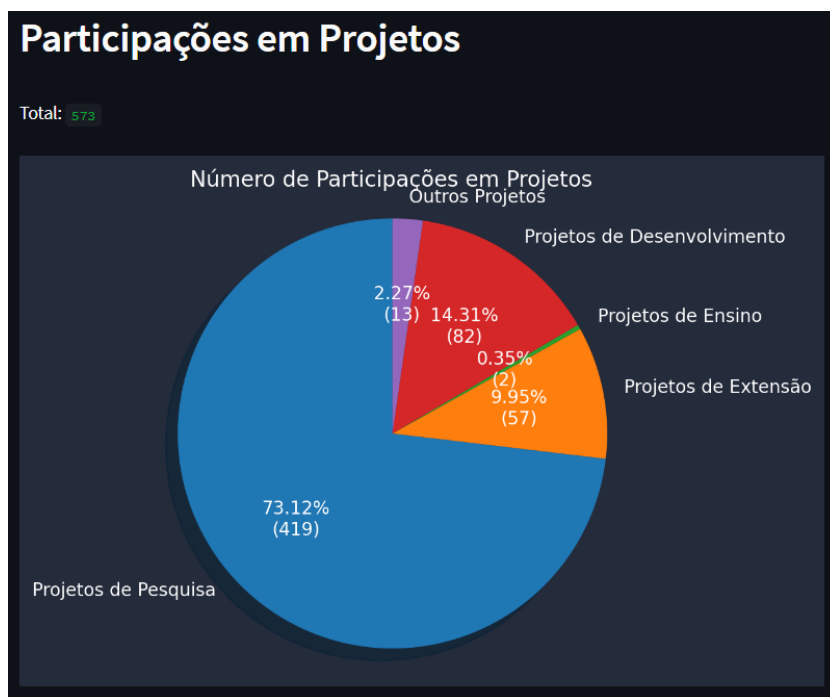


Figura 11 – Gráfico de Participações em Projetos.

Por fim, também foi realizado um experimento para demonstrar os efeitos da funcionalidade de definir intervalos de anos para os gráficos gerados. Na Figura 12, é possível observar que, definindo o ano mínimo para 2010 e o ano máximo para 2020, apenas trabalhos nesse intervalo serão considerados para a geração de gráficos. Em comparação com a Figura 6, onde o número total de trabalhos era 2480, o número total diminuiu para 1540 na Figura 12.

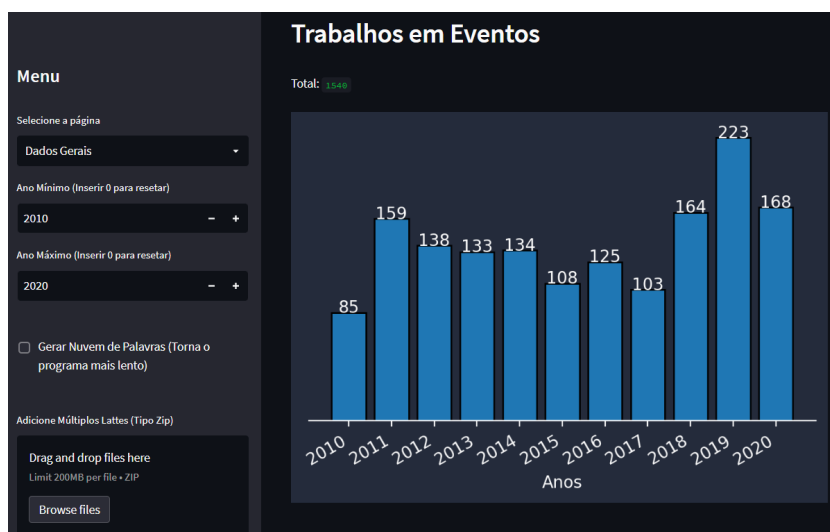


Figura 12 – Gráfico de Trabalhos em Eventos de 2010 a 2020.

4.2.2.2 Dados Específicos

Para realizar os experimentos de um pesquisador em específico, foi selecionado a página do orientador deste trabalho, Reginaldo Cordeiro Dos Santos Filho. Primeiramente, é possível

observar o gráfico de “Trabalhos em Eventos”, onde são totalizados 16 trabalhos, com o maior número de trabalhos nos anos de 2018 e 2021, tendo 4 trabalhos cada.



Figura 13 – Gráfico específico de Trabalhos em Eventos.

Outros dados que podem ser observados são no gráfico de artigos publicados em periódicos, onde na Figura 14, foram encontrados 16 artigos no currículo Lattes específico, com uma maior concentração no ano de 2016, tendo 4 artigos.

Mais um gráfico específico gerado se refere a “Participações em Bancas”. Na Figura 15, podem ser observadas um total de 37 participações em bancas, havendo um maior número de participações em bancas de graduação, totalizando 18 participações, que condizem com 48,65% do total de participações.

Outro gráfico específico gerado apresenta dados de orientações completas por ano. Na Figura 16, é possível observar um total de 18 orientações, com um foco maior no ano de 2022, tendo 7 orientações.

Mais um gráfico específico apresentado pela ferramenta é o de “Orientações em Progresso”. Na Figura 17, é possível notar o total 4 orientações em progresso, além de haver um número maior de orientações no ano de 2022, tendo 3 orientações.

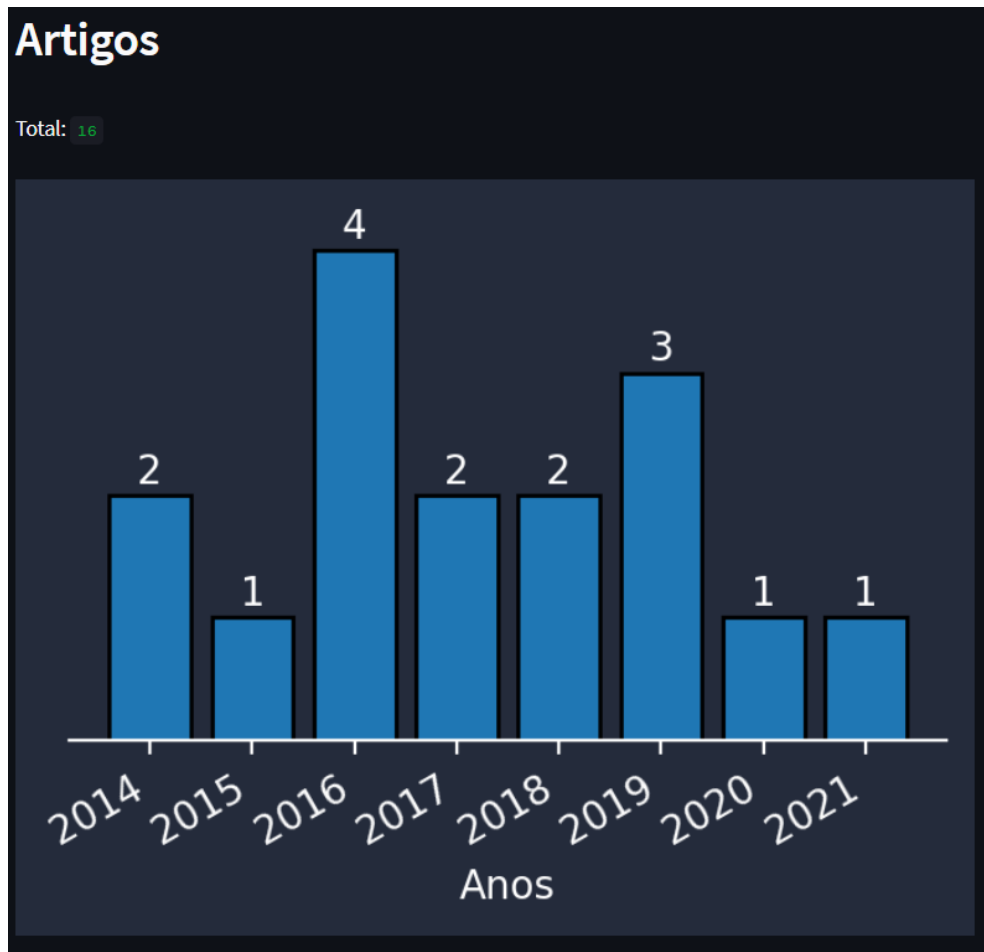


Figura 14 – Gráfico específico de Artigos.

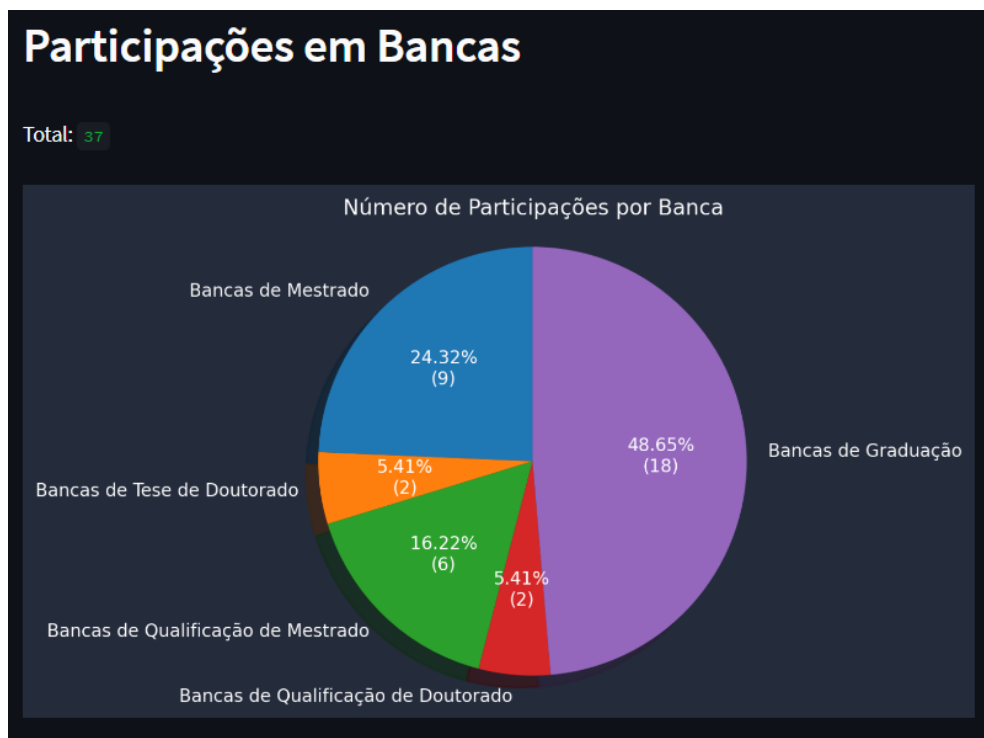


Figura 15 – Gráfico específico de Participações em Bancas.

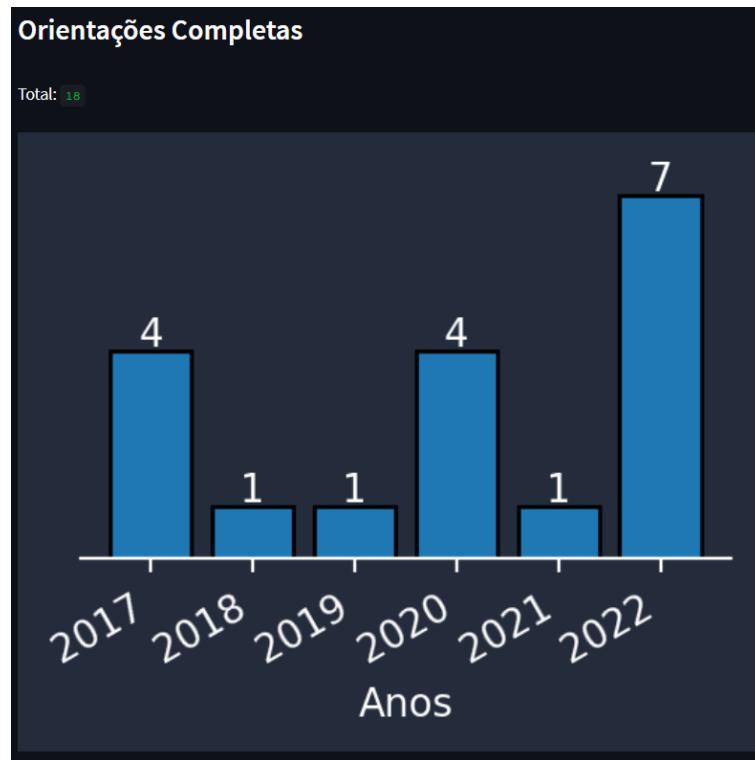


Figura 16 – Gráfico específico de Orientações Completas.



Figura 17 – Gráfico específico de Orientações em Progresso.

Por fim, o gráfico específico da Figura 18 apresenta um total de 13 participações em projetos, com mais concentração em “Projetos de Pesquisa”, tendo um total de 12 participações em projetos de pesquisa.

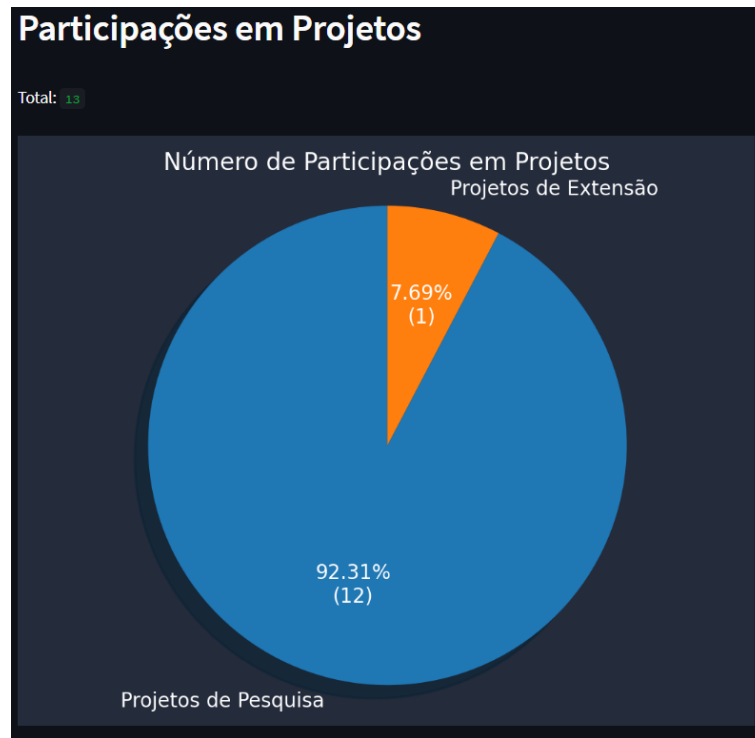


Figura 18 – Gráfico específico de Participações em Bancas.

4.2.2.3 Nuvens de palavras

Nesta seção, será abordada a funcionalidade de gerar Nuvens de Palavras. Como foi mencionado anteriormente no seção 2.3, a nuvem de palavras se caracteriza como uma imagem onde palavras mais frequentes em um texto são representadas por um tamanho maior, e as menos frequentes são mostradas com um tamanho menor. A ferramenta *FastLattes* possui a funcionalidade de gerar tanto uma nuvem de palavras, referentes aos títulos de trabalhos em eventos e artigos, de todos os currículos Lattes presentes no sistema, ou até mesmo de um currículo específico.

Na Figura 19, é possível observar a nuvem de palavras gerada pela aplicação, de acordo com todos os 28 currículos Lattes de pesquisadores do PPGCC no sistema. Pelo tamanho de algumas palavras, é possível notar que algumas aparecem com maior frequência nos títulos de trabalhos, em comparação à outras, como:

- *Network*: pode se assumir a presença uma grande quantidade de trabalhos que abordam a área de redes;
- *Software*: bem provável que haja um grande número de trabalhos sobre *software*, seja desenvolvimento ou análise dos mesmos;
- *System*: possivelmente uma grande presença de trabalhos que abordam o desenvolvimento e análise de sistemas diversos;

5 CONCLUSÃO

Neste trabalho, foi apresentada um sistema Web capaz de extrair informações úteis do currículo Lattes de pesquisadores e apresentá-los em um *dashboard*. A automatização desta tarefa é de muita valia para os gestores acadêmicos que, em sua maioria das vezes, precisam obter estas estatísticas e acabam por fazê-las manualmente.

A versão inicial da ferramenta *FastLattes* desenvolvida possivelmente mostra um potencial de contribuição para análises acadêmicas. Devido a automatização de um processo que atualmente é manual em grande parte, o tempo gasto com tal tarefa acaba sendo reduzido drasticamente. Os dados apresentados na subseção 4.2.2, foram coletados em questão de segundos, uma vez que todos os currículos Lattes já haviam sido baixados. Além disso, as funcionalidades apresentadas permitem que o usuário possa controlar as informações que estão sendo apresentadas para o mesmo, como através da definição de um intervalo de anos para os gráficos.

Outro ponto interessante que foi apresentado no texto foi essa capacidade de gerar nuvens de palavras. Como mencionado anteriormente, as nuvens de palavras facilitam bastante a visualização do perfil acadêmico dos pesquisadores, uma vez que possuindo apenas o currículo Lattes, o sistema já apresenta uma representação gráfica da principais áreas de atuação de um pesquisador em particular, algo muito útil, por exemplo, para discentes de graduação que estejam escolhendo um futuro professor pesquisador como seus respectivos orientadores.

No geral, mesmo ainda havendo espaço para melhorias da ferramenta, a mesma se mostra um facilitador para a área acadêmica, ou para qualquer indivíduo que tenha necessidade de utilizá-la, sendo bem fácil e simples de se utilizar, apenas sendo necessário um nível de conhecimento básico de *Python*.

Para trabalhos futuros, algo relevante seriam melhorias no *Design* do sistema para que fique visualmente mais agradável para o usuário, como a organização do nome dos pesquisadores na seleção de página por ordem alfabética, ou melhoria da paleta de cores utilizada no *dashboard*. Outro ponto interessante seria realizar o *Deploy* do sistema na Internet, uma vez que a versão atual do sistema só funciona executando localmente na máquina. Poderiam também serem exploradas mais áreas para a extração de dados relevantes dentro de currículos Lattes, e posteriormente sendo mostradas visualmente da melhor maneira possível para o usuário. Também poderiam ser implementadas novas funcionalidades com o objetivo de atender mais necessidades de gestores acadêmicos, como: a implementação de um filtro para selecionar um certo grupo de pesquisadores para serem analisados; apresentação de planilhas com informações dos trabalhos utilizados para gerar os gráficos; implementação de um filtro de conferências, para analisar apenas trabalhos de conferências de interesse; e a geração de um gráfico mostrando a interação entre pesquisadores por trabalhos realizados de maneira conjunta. Ademais, seria interessante o desenvolvimento do *dashboard* utilizando linguagens voltadas especificamente para aplicações *web*, fazendo uso da geração de gráficos do *FastLattes* como um *back-end*.

REFERÊNCIAS

- ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. Lattesminer: uma linguagem de domínio específico para extração automática de informações da plataforma lattes. **XII Workshop de Computação Aplicada - WORCAP 2012**, 2012.
- CNPQ. **Lattes - Histórico**. 2022. Disponível em: <<https://memoria.cnpq.br/web/portal-lattes/historico>>. Acesso em: 07 dez 2022.
- CNPQ. **Painel Lattes**. 2022. Disponível em: <<https://painel-lattes.cnpq.br/>>. Acesso em: 07 dez 2022.
- GOOGLE. **Google reCAPTCHA**. 2022. Disponível em: <<https://www.google.com/recaptcha/about/>>. Acesso em: 10 dez 2022.
- HANTANI, G. E. E. **FastLattes**. [S.l.]: GitHub, 2022. <<https://github.com/guihantani/FastLattes>>.
- LANA, M. V. P. Ufop ativa: uma ferramenta para extração e análise de dados acadêmicos de servidores da ufop baseada no currículo lattes. In: UFOP. **Monografia do curso de Ciência da Computação da Universidade Federal de Ouro Preto**. [S.l.], 2019.
- MENA-CHALCO, J. P.; JUNIOR, R. M. C. Scriptlattes: an open-source knowledge extraction system from the lattes platform. **Journal of the Brazilian Computer Society**, Springer, v. 15, n. 4, p. 31–39, 2009.
- OLIVEIRA, A.; SILVA, F. Extrator de produções acadêmicas de pesquisadores do ifce campus tianguá pela plataforma lattes. In: **Anais do XIV Encontro Unificado de Computação do Piauí e XI Simpósio de Sistemas de Informação**. Porto Alegre, RS, Brasil: SBC, 2021. p. 81–88. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/enucompi/article/view/17757>>.
- PPGCC. **PPGCC - Programa de Pós-Graduação em Ciência da Computação**. 2022. Disponível em: <<https://ppgcc.propesp.ufpa.br/index.php/br/>>. Acesso em: 04 dez 2022.
- STELATEK. **Stela Experta**. 2005. Disponível em: <<https://site.stelaexperta.com.br>>. Acesso em: 26 dez 2022.
- TIEPPO, R. **lucyLattes um script para manipular dados da plataforma Lattes**. 2019. Disponível em: <https://rafatieppo.github.io/post/2019_03_13_lucylattes/>. Acesso em: 04 dez 2022.
- TIOBE. **TIOBE Index for December 2022: December Headline: Python, C and C++ competing for language of the year award**. 2022. Disponível em: <<https://www.tiobe.com/tiobe-index/>>. Acesso em: 11 dez 2022.
- VASCONCELOS, A. J. d. et al. **Somos UFMG**. 2017. Disponível em: <<http://somos.ufmg.br>>. Acesso em: 04 dez 2022.
- VASCONCELOS, F. A. M. C. **LattesPlan: Uma Ferramenta para Extração e Organização de Informações de Currículos Lattes**. 58 p. Monografia (Trabalho de Conclusão de Curso) — Faculdade de Computação, Universidade Federal do Pará, Belém, 2019.